JGR Solid Earth

RESEARCH ARTICLE
10.1029/2022JB024343

Intra-Eruption Forecasting Using Analogue Volcano and Eruption Sets

Mark S. Bebbington[1] and **Susanna F. Jenkins**[2]

[1]Volcanic Risk Solutions, Massey University, Palmerston North, New Zealand, [2]Earth Observatory of Singapore, Asian School of the Environment, Nanyang Technological University, Singapore, Singapore

Key Points:

- Intra-eruption forecasting models developed for informative analogue eruption subsets on the basis of volcano type and composition
- Mix of informative and null (global frequency) analogs, accounting for transitions not in the data set, proposed for eruption forecasting
- Neither morphology nor composition based analogs, or dynamically adapting analogs, outperform using all of the phase data as an analogue

Correspondence to:
M. S. Bebbington,
m.bebbington@massey.ac.nz

Citation:
Bebbington, M. S., & Jenkins, S. F. (2022). Intra-eruption forecasting using analogue volcano and eruption sets. *Journal of Geophysical Research: Solid Earth*, *127*, e2022JB024343. https://doi.org/10.1029/2022JB024343

Received 7 MAR 2022
Accepted 10 JUN 2022

Abstract Forecasting the likely style and chronology of activity within an eruption is a complex issue that has received far less attention than forecasting the onset and/or the magnitude. By developing a global data set of coded phases (discrete styles of activity within previous eruptions), we can model the resulting data using a semi-Markov chain. Given enough data, we were able to examine the question of whether analogue-based strategies for subsetting the data can improve forecasting performance of phase chronology and style within ongoing eruptions. This work required inclusion of a "null analogue" element to ensure no surprises, that is, phase transitions or durations that were not in the data set and hence cannot be predicted. We have significantly expanded, and made available, our curated data set on eruption phases, which now contains 2670 eruptions (6871 phases), of which 56% are multi-phase. This increases the data set by 283% and includes 95% of Holocene eruptions with text descriptions. We find that, with the notable exception of shields, limiting the analogue set on the basis of volcano morphology and/or composition is not significantly more informative than using the entire data set. Dynamically adjusting the data limits by eliminating eruptions without the observed phase as the eruption progresses provided little benefit, although subsetting on the basis of VEI may have some utility. At the individual volcano level, non-analogue models can outperform the entire data set, if the target volcano has relatively unique behavior and/or a large enough record of phased eruptions.

Plain Language Summary Volcanic eruptions can consist of multiple phases, separated either by quiescence, or by a switch to a different style of activity, e.g., from explosive to effusive. We present a data set of 6871 phases from 2670 historical eruptions, extending a previously curated set of 698 eruptions. We update a probabilistic model for forecasting eruption progression through multiple states. The new data and model are then used to investigate if ongoing eruptions can be better forecast by using only a subset of the data. In particular we consider using (a) only data from volcanoes with similar composition (e.g., basalts), (b) only data from volcanoes of the same type (e.g., stratovolcanoes), (c) only data from volcanoes with the same composition and type, (d) only data from the volcano concerned. Except for shield volcanoes, none of these limitations outperform the entire data set, although using data from only the target volcano may be effective if there is enough of it, or the behavior of the target volcano is relatively unique (e.g., Stromboli).

1. Introduction

Understanding the temporal and behavioral evolution of an eruption, from onset to conclusion, has received relatively little attention to date. The focus instead has been on characterizing and forecasting system reactivation after a long period of quiescence, that is, the onset of a new eruption (Marzocchi & Bebbington, 2012; Newhall, 2000; Poland & Anderson, 2020; Sparks, 2003; Whitehead & Bebbington, 2021). However, nearly two-thirds of volcanic fatalities are recorded more than 1 month after eruption onset (Simkin et al., 2001), highlighting the importance of forecasting eruptive activity beyond the initial eruption onset. Volcanic eruptions can continue for years or even decades (e.g., Kokelaar (2002); Thordarson and Larsen (2007)), encompassing bouts of activity that range from major explosive eruptions through effusive activity to phases of quiescence. Forecasting when each style of activity will begin and end, the associated hazards (e.g., (Bonadonna et al., 2005; Jenkins et al., 2008; Magill et al., 2015)) and, critically, when the eruption will end, are key goals of emergency management (Jolly et al., 2014).

A key limitation to success in forecasting of any type is the availability of data on which to base a forecast (e.g., Bebbington (2013); Thompson et al. (2015); Wright et al. (2019); Tierz (2020)). Most volcanoes have a limited record, which may be strongly time-varying and, moreover, the coverage may be uneven in time. A common

approach for navigating such issues in practice is the reliance on analogue volcanoes (Marzocchi et al., 2004; Neri et al., 2008; Newhall & Pallister, 2015), whose properties are considered similar enough that their forecasts can be overlaid on the target volcano.

A decided advantage to intra-eruption forecasting (Bebbington & Jenkins, 2019) is that only the sequence of phases is needed, exploiting the idea of common processes at different volcanoes (Cashman & Biggs, 2014). The time between eruptions, which may include missing eruptions (Wang & Bebbington, 2012), is irrelevant. Hence, provided the phases are preserved in the geological record (Voloschina et al., 2021), even geological data of individual eruptions may be able to be used, with perhaps some modification or grouping of the phase types.

Previously, using data provided by the Smithsonian Institution's Global Volcanism Program (GVP, 2013) (GVP) and Sarah Ogburn (USGS), we developed a data set of 698 multi-phase eruptions dated between 79CE and 2015 from 187 volcanoes around the world. Each discrete style of activity was coded into one of nine "states" based upon eruption descriptions (Bebbington & Jenkins, 2019). This work built upon an earlier empirical study of intra-eruptive activity by Jenkins et al. (2007). The expanded and improved Bebbington and Jenkins (2019) data set of eruptions and their phases was used to construct a semi-Markov chain model that forecast the probabilities for future eruptive states, based on the duration and style of preceding phases. Here we update our data set by adding another 1972 eruptions, including 1180 single-phase eruptions, and define an additional state ("minor explosion"). Using this expanded data set of 2670 multi- and single-phase eruptions, we re-calculate our probabilities of intra-eruption phase transition and further investigate how, or if, forecasts can be improved using subsets of the data that represent different volcano or eruption "analogues".

In what follows, we first briefly summarise the intra-eruption phases model of Bebbington and Jenkins (2019) before detailing the updates to our data set and resulting forecast model. We then develop a suite of forecast models that draw only from subsets of the data set, where the data subset is considered a set of analogues based on characteristics of the volcano (e.g., volcano type, major rock composition) or eruption (e.g., observed phase types). We retrospectively evaluate the effectiveness of the analogue subsets in forecasting activity during individual eruptions by considering the change in log-likelihood from a reference forecast which uses all available data. We aim to identify what analogue, or combination of analogues, best forecasts eruption chronologies and durations to provide an indication of what analogues would perform best in a true forecasting environment. Such information can be used to provide evidence-based forecasts of intra-eruption activity and durations that can support emergency management prior to and during a future volcanic crisis.

1.1. Recap of the Intra-Eruption Model

The intra-eruption model uses a semi-Markov chain to forecast the probability of transitioning to a certain phase of activity, given knowledge of the current state and its duration, as well as the duration of any preceding or subsequent quiescence. By using a Markov chain, the volcano can make repeated visits to the same state, avoiding the high data demands of an indeterminate branching structure, for example, as in an event tree. This allows for far more complicated behavior to be modeled, but at the cost of accepting the memoryless Markov property. This means that the probability of a future state depends only upon the current state (augmented by its duration and that of preceeding or following quiescences) and not upon how long the eruption has been in progress, the number of phases or (apart from the current one) their type.

Mathematically, the model considers time in days. At any day, t, the volcano is in one of a number of discrete activity states. For the purposes of this paper these are State 1 = Effusive Activity (Eff), 2 = Effusive and Explosive Activity (Eff + Exp), 3 = Continuous Explosions (Cts Exp), such as Strombolian Activity, 4 = Intermittent Explosions (Int Exp), 5 = Minor Eruption (Minor E), 6 = Major Eruption (Major E), 7 = Plinian Eruption (Plinian E), 8 = Minor Explosion (Min Exp). See Table 1 for details. The volcano may also be in (post-eruptive-phase) quiescence, which is treated by including the quiescence length as an additional dimension to the duration of the preceding eruptive state.

The data provide $\Pr(D_i = d | i \to j)$, where D_i is the duration in state i, and $i \to j$ indicates that the subsequent transition is from state i to state j. We can augment this by including the quiescence after state i, obtaining (in obvious notation) $\Pr(D_i = d, Q = q | i \to j)$, where Q is the quiescence between states i and j. This gives the probability that eruptive activity will transition from state i to j as a function of duration in state i and, if applicable, the duration of interphase (post-state i phase) quiescence q.

Table 1
State Definitions, Modified From Bebbington and Jenkins (2019)

State	Description	Notes
1	Effusive	Solely effusive activity, including lava extrusions (domes), lava effusions (flow), fountains and spatter cones. Satellite thermal anomalies were assumed to be effusive, unless stated otherwise.
2	Effusive and explosive	Activity that specifically describes an effusive and explosive component, but that cannot be further divided into discrete phases, either because activity was contemporaneous, for example, lava dome growth and phreatic explosion, or because descriptions weren't detailed enough. Effusive activity includes all of those in State 1; explosive activity includes phreatic through magmatic
3	Continuously explosive	Explosive activity described as continuous or Strombolian.
4	Intermittently explosive	Explosive activity described as intermittent or Vulcanian, or with a date range exceeding 2 days without mention of a major explosion. Intervals between explosions range from minutes to days. See Bebbington and Jenkins (2019), Figure 9.
5	Minor explosive eruption	c. < 10 km column height. Where the dates of individual explosive eruptions were given, but the size was unclear, minor eruptions were assumed.
6	Major explosive eruption	c. 10–20 km column height. Many VEI 3 and most VEI 4 eruptions will have at least one.
7	Plinian explosive eruption	c. > 20 km column height. Most VEI 5 and all VEI 6+ eruptions will have at least one.
8	Minor explosion	A single explosion of the type described in State 4
-	Deformation	No eruptive activity, but explicit mention of deformation.
(9)	Quiescence	More than 1 day between states 1–8

The phase duration and quiescence times are modeled as log-normal random variables (Bebbington & Lai, 1996), treating the logarithm of the times as normally distributed (Bebbington & Jenkins, 2019). If we let $f_{ij}(d)$ be the probability density for $D_i = d|i \to j$, then

$$f_{ij}(d) = \frac{1}{d\sigma_{ij}\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_{ij}^2}(\ln d - \mu_{ij})^2\right) \quad (1)$$

where μ_{ij} and σ_{ij} are estimated as the mean and standard deviation of the relevant observed (and log-transformed) phase durations.

The probability densities $h_{ij}(q)$ for the event '$Q = q|i \to j$' are handled similarly to f, with the addition of an atom at 0 representing the possibility of no intervening quiescence, that is,

$$h_{ij}(q) = q_{ij}^{(0)}\delta_{\{0\}}(q) + \frac{1 - q_{ij}^{(0)}}{q\sigma_{ij}\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_{ij}^2}\exp(\ln q - \mu_{ij})^2\right) \quad (2)$$

where $q_{ij}^{(0)}$ is the observed probability of there being no quiescence between the respective eruptive phases, and $\delta_A(q)$ is the mathematical delta function with the properties that $\delta_A(q) = 0$ if $q \notin A$, and $\int\delta_A(q)\mathrm{d}q = 1$. The parameters μ_{ij} and σ_{ij} are estimated as the mean and standard deviation of the appropriate log-transformed quiescences.

A problem arises when, conditional on the state transition, some duration or quiescence distributions are exemplified by very few (often one) such transitions, meaning that we can not simultaneously estimate both μ and σ. In the case of very few observations, they may be too clustered for numerical stability, with the resulting distribution approaching a point mass, which is very unrealistic. Bebbington and Jenkins (2019) dealt with this by setting the minimum coefficient of variation as $1/n$, where n is the number of observed transitions, meaning that $\sigma = \sqrt{\log(1 + 1/n^2)}$ as a minimum. The subsetting required below to evaluate different analogue strategies meant that the results of this substitution caused undue sensitivity to quiescence lengths of single eruptions. Hence we have adopted here a different fix, adding nominal quiescence lengths of one and 90 days to each set of observed quiescences before calculating the mean and variance. This both controls the variance, and reflects the

GVP definition that an eruption can include periods of "up to 90 days surface quiet". The latter is particularly useful in modeling the question of whether an eruption is actually over.

Bebbington and Jenkins (2019) demonstrated that phase durations and quiescence lengths can be assumed to be independent, in which case we have

$$\Pr(i \to j | D = d, Q = q) = \frac{p_{ij} f_{ij}(d) h_{ij}(q)}{\sum_k p_{ik} f_{ik}(d) h_{ik}(q)}. \quad (3)$$

where p_{ij} is the probability that the current eruptive state i is succeeded by the eruptive state j (there may be an intervening quiescence), and similarly for Q_1, with h replacing g. These are the joint probabilities, that is, assuming that we know d and q. In reality, we will know one of these, and know that the other will be at least as long as that observed to the present. It is then easy to calculate the conditional probabilities

$$\Pr(i \to j | D_i > d, Q_0 = q) = \frac{p_{ij} g_{ij}(q) \int_d^\infty f_{ij}(u)\mathrm{d}u}{\sum_k p_{ik} g_{ik}(q) \int_d^\infty f_{ik}(u)\mathrm{d}u} \quad (4)$$

or

$$\Pr(i \to j | D = d, Q_1 > q) = \frac{p_{ij} f_{ij}(d) \int_q^\infty h_{ij}(u)\mathrm{d}u}{\sum_k p_{ik} f_{ik}(d) \int_q^\infty h_{ik}(u)\mathrm{d}u}, \quad (5)$$

which give the probability that the next eruptive state will be state j, given that the previous eruptive state was i, and the observed previous and current durations (one eruptive phase, the other quiescence). Equation 4 can be used for day-to-day forecasting during an eruptive phase, similarly Equation 5 can be applied during a quiescence.

2. Data and Model Updates

The work presented here builds upon the work of Bebbington and Jenkins (2019) by expanding the data set from 698 to 2670 eruptions. In the process, many small (mainly VEI 1) eruptions were found to consist of one or more discrete explosions, which did not amount to the equivalent of a "minor eruption". Hence a new state was retroactively defined for these "minor explosions".

2.1. New Data

The original data, kindly provided to us by the Smithsonian Institution's Global Volcanism Program (GVP, 2013) and Sarah Ogburn (USGS), included 698 eruptions, with information on the eruption VEI, start and end date (where available) and, critically, the eruption description. While these eruptions had already been decomposed into phases, these did not always reflect the eruption description, and so we recoded the entire data set into discrete styles of activity in a consistent manner using the eruption descriptions provided by the GVP record. The states of activity are described in Table 1, and an example of our coding can be found in Bebbington and Jenkins (2019). Where we could identify a more precise record from published sources, including the GVP bulletins, we used these sources to supplement the GVP eruption description in more accurately assigning the phase states and durations. The coding produced 2775 eruptive phases and 1795 quiescences (Bebbington & Jenkins, 2019).

We have since added another 1972 eruptions from the GVP Holocene eruption record that contained an eruption description comprehensive enough to identify the chronology of eruptive activity (phases). These 1972 eruptions were defined as "single-phase" in the original data set, and thus were not included in the analysis of Bebbington and Jenkins (2019). On closer examination, 796 proved to have multiple phases, leading to a total of 1494 multi-phase eruptions in our new data set, leaving 1176 single-phase eruptions. This represents a near four-fold increase in the number of eruptions in our data set, from 698 to 2670. This represents c. 30% of the eruptions recorded in the GVP, but nearly 95% of those with text descriptions of the eruption. A total of 56% of the eruptions in our data set have identifiable multi-phase signatures. The final data set has 6871 eruptive phases (increased from 2785) and c. 3848 quiescences (increased from c.1520) from 2670 eruptions at 353 volcanoes between 79CE and 2015.

Table 2
Metadata for Expanded Phases Data set

State	Original data (Bebbington & Jenkins, 2019)		Expanded data (this paper)	
	Number	Proportion	Number	Proportion
Effusive	476	0.104	1020	0.095
Effusive and Explosive	407	0.089	845	0.079
Continuous Explosions	136	0.030	319	0.030
Intermittent Explosions	812	0.178	2008	0.187
Minor Explosion	314	0.069	990	0.092
Minor Eruption	460	0.101	1381	0.129
Major Eruption	143	0.031	274	0.026
Plinian Eruption	27	0.006	34	0.003
Quiescence	1795	0.393	3848	0.359
TOTAL	4570		10,719	

2.2. States of Activity

The original data were coded into nine discrete "states" that described the activity during that phase: (1) Effusive, (2) Effusive and explosive, (3) Continuously explosive, (4) Intermittently explosive, (5) Minor explosive eruption, (6) Major explosive eruption, (7) Plinian explosive eruption, (8) Deformation, and (9) Quiescence (see Table 1 for state descriptions). A further two nominal states of (0) Eruption start, and (10) Eruption end were included to support our modeling. The original data contained only eight instances of "deformation" states; we decided to not include this state going forward as it is strongly influenced by the monitoring network in place and does not greatly affect the hazard posed by an eruption. This resulted in four multi-phase eruptions being recoded as single-phase in this analysis.

Upon analysis of the new expanded data set, we identified many VEI 1 eruptions that have multiple "Minor explosive eruption" states that appear to have been small, short duration, single explosions. For this study, we have therefore introduced a new derived state: "Minor explosion". States previously classified as "Minor explosive eruption" have been reclassified as "Minor explosion" in the following circumstances:

1. If the eruption VEI is less than 2.
2. If the eruption VEI is 2 *and* there is no mark for ash production or pyroclastic flow on the phase.
   (a) *Except* where it is the only phase in the eruption.

This applies to 990 phases across 582 eruptions from 174 volcanoes. The new state of "Minor explosion" has durations measured in minutes or less, but is treated as nominally having a constant 0.5 days duration in order to avoid a divide by zero error in the calculations. Thus the new data have been coded into nine discrete states, as before, but with the removal of "Deformation" and the addition of "Minor explosion".

Table 3
Eruptive Phases Per Eruption (Updated), and as a Proportion of All Phases

Phases per eruption	Original data (Bebbington & Jenkins, 2019)		Expanded data (this paper)	
	Number	Proportion	Number	Proportion
1	4	0.001	1180	0.172
2	252	0.091	615	0.090
3	180	0.065	352	0.051
4	101	0.036	200	0.029
5	62	0.022	115	0.017
6	29	0.010	65	0.009
7	21	0.008	42	0.006
8	4	0.001	16	0.002
9	5	0.002	14	0.002
10	5	0.002	17	0.002
11	9	0.003	11	0.002
12	6	0.002	9	0.001
13	3	0.001	5	0.001
14	3	0.001	6	0.001
15	3	0.001	6	0.001
>15	11	0.004	17	0.002

2.3. Summary of Data Updates

The number and proportion of each state in our original (Bebbington & Jenkins, 2019) and updated (this study) data set is given in Table 2. The change in the proportion of states is relatively minor, with the largest changes a 3% decrease in the proportion of quiescent phases, a 2% increase in the proportion of minor explosions and a 1% decrease in the proportion of effusive and explosive phases. With the addition of only seven Plinian phases to our updated data set, the proportion of Plinian phases has halved to 0.3% and this state remains the least common ($n = 34$), by nearly one order of magnitude. The majority of multi-phase eruptions (78%) contain just two or three eruptive phases (up from 62% in the original data) although a small number of eruptions still contain more than 15 eruptive phases (i.e., not including quiescent phases) (Table 3).

Figure 1 shows how the cumulative number of eruptions in the data set with a given number of phases varies over time. The sharp change in slope of the "all eruptions" curve reflects the limitation in the data set of single phase eruptions to post-1900. Apart from this, the curves parallel each other, indicating that the data set is time-homogeneous, conditional on the inclusion of an eruption. Post-1900, the average number of phases per eruption is consistent, as indicated by the alignment of the total eruptions and total phases curves. The necessity of accepting that the Plinian eruptions are more prevalent among the earlier data was discussed by Bebbington and Jenkins (2019).

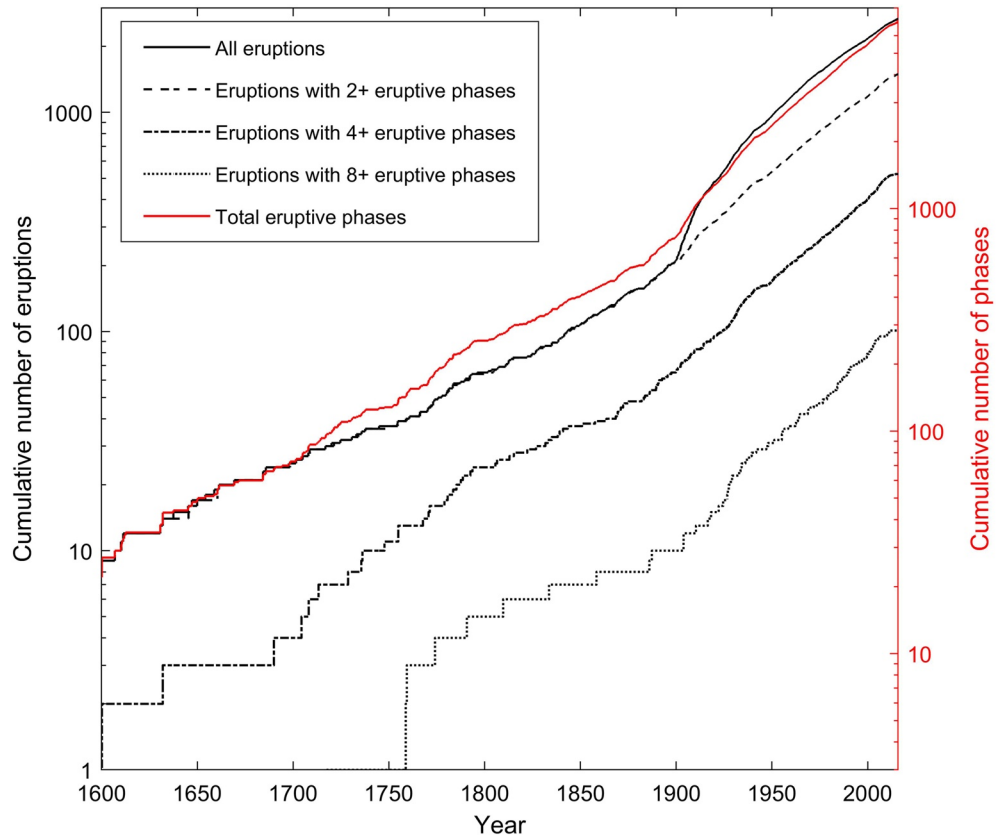

Figure 1. Cumulative numbers of eruptions and phases over time. The nine eruptions between 79 and 1600 CE are not shown.

The distributions of phase durations are shown in Figure 2. We see the same distributional shapes as in Bebbington and Jenkins (2019), with a decrease in the mean durations of effusive and explosive, continuously explosive and quiescence phases.

2.4. Updated Forecast Model

The new transition matrix, which dictates the probability of transitioning from one state of activity to another, based upon our updated phase data is as below, where the current state is the row and the next state is the column. We can see that the proportional decrease in major and Plinian phases (Table 3), relative to the original data, means that in our expanded data set phases are less likely to transition to such behavior and instead are more likely to transition to eruption end (Table 4). With the exception of minor explosions and minor eruptions, activity is now less likely to continue with more of the same behavior, after an intervening quiescent period, although for some states (effusive, intermittently explosive, minor explosions) if the eruption does not end, a continuation with the same behavior is still the most likely outcome. As before, major and Plinian eruptions are unlikely to herald the end of an eruption but rather the start of further large explosive eruptions. Intermittent and minor explosions are the most likely states to transition to eruption end as well as being amongst the most common behavior at the start of an eruption.

The most significant difference from Bebbington and Jenkins (2019) is that these probabilities are not conditional on the eruption being multiphase. Hence, we are now in a position to forecast entire eruption sequences, conditional only on there being an eruption, rather than the eruption being multi-phase.

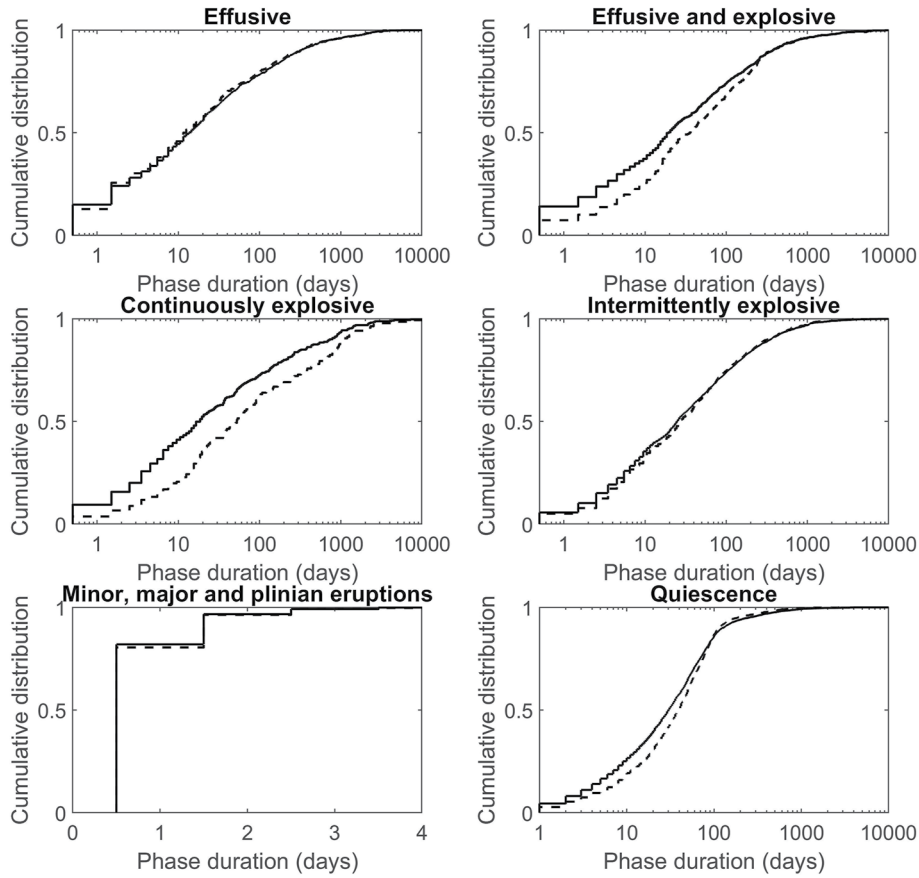

Figure 2. Phase durations for each of the states. The updated data (this paper) are shown by the solid line, the original data (Bebbington & Jenkins, 2019) by the dashed line. The new state "Minor Explosion" is assigned a nominal duration of 0.5 days.

3. Analogue Forecasting

With understanding of volcanoes and consequent forecasting bedeviled by the general lack of data, particularly at potentially awakening volcanoes (Loughlin et al., 2015), that is, those without recent (or historical) eruptions, the only generally applicable (Whitehead & Bebbington, 2021) forecasting technique has been the application of

Table 4
Phase Transition Matrix (Updated)

	Effusive	Eff + Exp	Cts Exp	Int Exp	*Min Exp*	Minor E	Major E	Plinian	End
Start	0.104	0.103	0.042	0.324	*0.171*	0.216	0.037	0.004	0
Effusive	0.275	0.109	0.039	0.105	*0.039*	0.067	0.018	0.001	0.347
Eff + Exp	0.131	0.149	0.056	0.189	*0.015*	0.047	0.036	0.002	0.374
Cts Exp	0.154	0.157	0.078	0.207	*0.025*	0.088	0.041	0.003	0.248
Int Exp	0.068	0.091	0.025	0.182	*0.042*	0.083	0.029	0.003	0.476
Min Exp	0.039	0.014	0.013	0.119	*0.335*	0.049	*0*	*0*	0.429
Minor E	0.053	0.036	0.016	0.157	*0.041*	0.307	0.030	0.002	0.358
Major E	0.179	0.131	0.026	0.339	0	0.099	0.033	0.026	0.168
Plinian	0.147	0.059	0.059	0.441	0	0.059	0.176	0.059	0

Note. Red/blue text indicates probabilities that have increased/decreased more than 20% (compared to original data with minor explosion introduced). Green are possibilities not represented in the original data.

expert judgment (Aspinall et al., 2003; Bebbington et al., 2018; Hincks et al., 2014; Newhall et al., 1996). Expert judgment has since been partially exemplified in the use of analogue volcanoes (Marzocchi et al., 2004; Newhall & Pallister, 2015; Tierz et al., 2019) or eruptions. Analogues are important concepts in volcanology because forecasting at most volcanoes suffers from a lack of data; eruption records are often scant, which makes the evidence base from which we can forecast likely future behavior too limited to be useful. In these cases data, or conceptual models encapsulating these data, from analogue volcanoes are typically used to supplement the available data set on the assumption that they will behave in a similar way (Bebbington, 2014; Newhall et al., 2017; Rodado et al., 2011; Sheldrake, 2014; Whelley et al., 2015). The first use of "analogue volcano" in the scientific literature seems to have been by Miller et al. (1999), where Mt Dutton was identified as an analog of Unzen, based on physical and petrological similarities to the 1990-95 Unzen eruption, although a variety of analogies and similarities were used during the 1991 unrest of Pinatubo (Punongbayan et al., 1996).

A useful analogue volcano typically needs to possess a long and comprehensive eruption record, and there are thus only small numbers of truly useful analogs. As we are looking at intra- and not inter-eruption forecasting, we are only interested in individual eruptions, and hence we require analogs that have well detailed, rather than necessarily long, eruption records.

By sourcing data from multiple analogue volcanoes, the data set available for forecasting can be enlarged to possibly improve forecasts for those volcanoes with an eruption record, and to enable forecasts for volcanoes where little or nothing is known about about their past behavior, particularly the variability within past eruptions. This is an approach similar to the pooling of data for the understanding of VEI (Bebbington, 2014; Sheldrake, 2014), dome building (Sheldrake et al., 2016) and PDCs (Ogburn et al., 2016). The obvious question in applying this approach to our study is whether subdividing the data provides a better forecast, and if it does, what is the best way to subdivide the data, that is, what makes the best analogue? For example, does a model sourcing phase data only from andesitic calderas produce a more informative forecast for andesitic calderas than one sourced from all data, from all calderas, or from all andesitic volcanoes? A unique possibility provided by our phase model is to subdivide the data according to eruption as well as volcano analogs, in that the analogue set can be progressively updated as the eruption progresses.

In what follows, we compare forecast models developed using already defined analogue subsets for Aluto (Tierz et al., 2020) and Pinatubo (Newhall et al., 2017; Whelley et al., 2015), before exploring how we might define analogue subsets for the wider data set and, critically, how we can evaluate the performance of each forecast model.

3.1. Examples: Aluto and Pinatubo

To illustrate the effects of analogue selection, we can simulate eruptions from models produced by an analogue subset, and from the full data set, as was the practice suggested by Bebbington and Jenkins (2019). Repeating this 10,000 times for each model allows us to generate comparable summary statistics for each model.

As part of their analysis of the hazard from Aluto (Central Ethiopia), Tierz et al. (2020) provided a list of the top 100 analogs as selected from the GVP using the VOLCANS procedure of Tierz et al. (2019). As many of these volcanoes have scant eruption records themselves, only 19 of these are represented in our data set, with a total of 94 eruptions. Restricted to these eruptions, our transition matrix is as shown in Table 5, which should be compared with that from the full data set in Table 4. We can see that an eruption forecast using these analogs is much more likely to start with effusive activity or combined effusive and explosive activity, rather than purely explosive activity. No Plinian activity has been recorded for the analogue eruptions and so there is a zero likelihood of such activity when we consider the analogue model alone. A reduced number of phases amongst the analogue set means that the probability of transitioning to eruption end is greater for the majority of states, excepting major eruptions and intermittent explosions, which are both more likely to transition to intermittent explosions.

From 10,000 simulations of the model developed using these analogs, we get the summary statistics shown in Figure 3 (left: Tierz et al. (2020), and center: All). We see that while the eruption duration distributions are remarkably similar, using all the data increases the average number of phases, and the number of eruptive phases. And, of course Table 5, using only the Tierz et al. (2020) analogs does not produce Plinian eruptions, and

Table 5
Phase Transition Matrix (Using Only Aluto Analogs: n = 94 Eruptions)

	Effusive	Eff + Exp	Cts Exp	Int Exp	Min Exp	Minor E	Major E	Plinian	End
Start	0.245	0.170	0.032	0.170	0.149	0.181	0.053	0	0
Effusive	0.053	0.132	0.053	0	0.053	0.026	0	0	0.684
Eff + Exp	0.267	0.033	0.067	0.200	0	0	0.033	0	0.400
Cts Exp	0.222	0	0	0.111	0	0	0.111	0	0.556
Int Exp	0.024	0.119	0.024	0.214	0.024	0.024	0.119	0	0.452
Min Exp	0.045	0	0	0.045	0.227	0	0	0	0.682
Minor E	0.028	0.056	0	0.083	0	0.333	0.056	0	0.444
Major E	0	0.071	0.071	0.429	0	0.357	0	0	0.071
Plinian	0	0	0	0	0	0	0	0	0

downweights the likelihood of minor explosions and pure effusive behavior. Conversely, major explosions and combined effusive and explosive behavior are more prevalent when the Tierz et al. (2020) analogs are considered.

Previous studies have suggested a number of analogs for Pinatubo (Philippines): Colo (one eruption), Inielika ($n = 2$ eruptions), Kelimutu ($n = 1$) and Semeru ($n = 22$) (Whelley et al., 2015) and Redoubt ($n = 6$), Augustine ($n = 6$), Mt St Helens ($n = 4$) (Newhall et al., 2017). Considering only eruptions in our data set from these volcanoes, the transition matrix is shown in Table 6 (again, compare with Table 4). The outstanding feature of Table 6 is its sparsity, with 45% of the transitions being impossible. We will return to this point later. Also, continuous explosions and Plinian eruptions are always followed by Minor eruption and Major eruption respectively.

Simulating 10,000 sets of eruptions gives the summary statistics in Figure 3 (right). Unlike the previous example with Aluto, the eruption duration is much longer for Pinatubo when using only the analogue eruptions. Moreover,

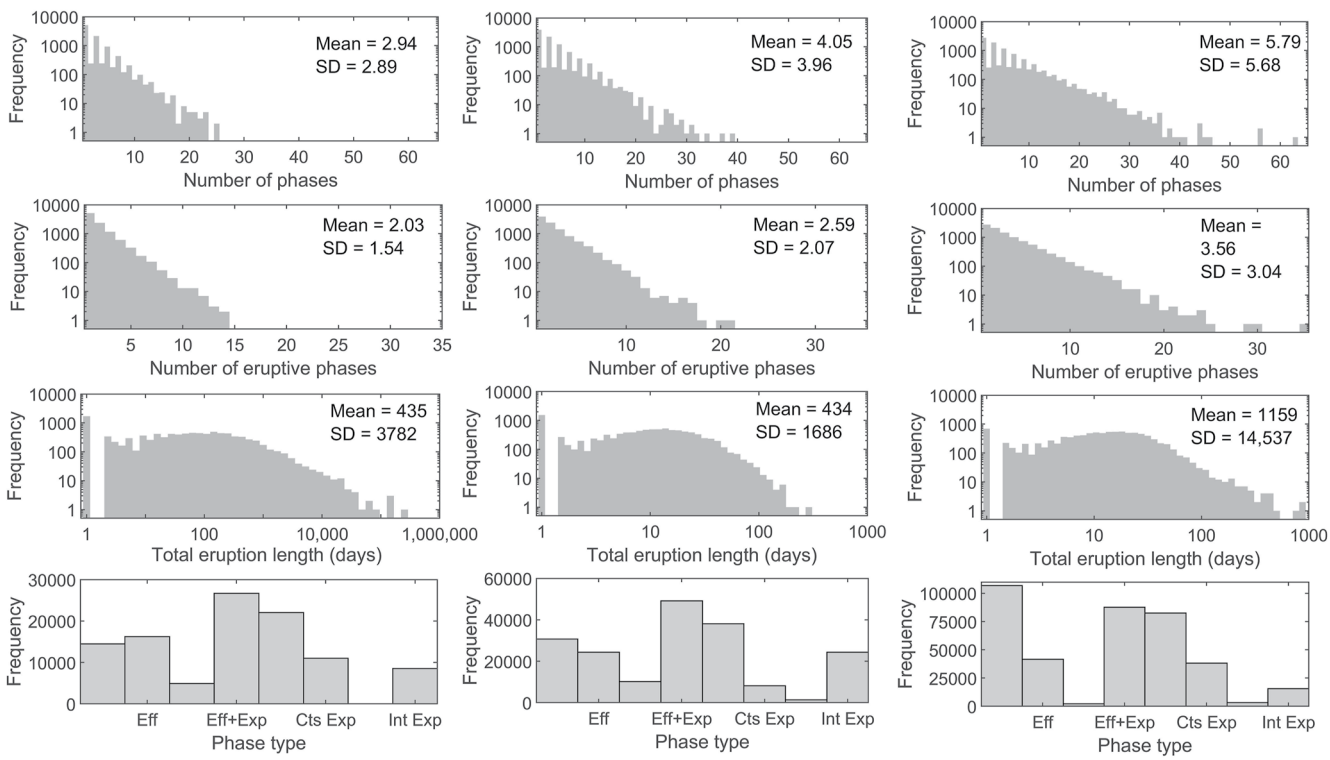

Figure 3. Statistics associated with simulated eruptions considering the model based on left: Tierz et al. analogs for Aluto, center: no analogue subsection, that is, the entire data set, right: Pinatubo analogs. Note that the horizontal scale is the same in all three panels.

Table 6
Phase Transition Matrix (Using Only Pinatubo Analogs: n = 42)

	Effusive	Eff + Exp	Cts Exp	Int Exp	Min Exp	Minor E	Major E	Plinian	End
Start	0.048	0.071	0.024	0.310	0.048	0.381	0.119	0	0
Effusive	0.471	0.088	0	0.088	0	0.088	0.029	0	0.235
Eff + Exp	0.188	0.063	0	0.188	0.125	0.125	0	0	0.313
Cts Exp	0	0	0	0	0	1	0	0	0
Int Exp	0.146	0.073	0	0.220	0	0.024	0.073	0.024	0.439
Min Exp	0.000	0.167	0	0.333	0.333	0	0	0	0.167
Minor E	0.083	0.139	0	0.167	0	0.306	0.083	0	0.222
Major E	0.267	0	0	0.333	0	0.133	0.133	0	0.133
Plinian	0	0	0	0	0	0	1	0	0

in contrast to Aluto, the number of simulated eruptive phases decreases when all data are used, that is, the analogue model suggests that more eruptive phases are likely for Pinatubo than using all of the eruptive data would suggest. Using the analogue set, effusive behavior replaces intermittent explosions as the most likely state and minor explosions become less prevalent. There is also an increase in the likelihood of a Plinian eruption or major eruption.

As there are clearly differences in forecasts when analogue subsets are used, the question arises: which model provides the better analogue when forecasting activity for a single eruption?

3.2. Defining Possible Analogue Sets

The significant increase in the number of phases, eruptions and volcanoes in our data set offers the opportunity to subset the data by analogue volcanoes (or more precisely, their eruptions), recalculate transition matrices and generate targeted (or *analogue*) forecast models. This prompts the question of how to subset the data, which is a question of what makes the activity of one volcano similar to another, that is, what makes an analogue?

Typically, analogs have been defined (Miller et al., 1999; Pantazidis et al., 2019; Peltier et al., 2015; Sheldrake, 2014) around physical attributes (Hone et al., 2007) and petrological makeup (Bas et al., 1986). Bayesian approaches (Bebbington, 2014; Sheldrake, 2014, 2016) have selected data that generally respects this approach. In tailoring this to the GVP definitions, we will hence adopt a definition of an analogue set which has five classes:

1. Eruptions from all volcanoes
2. Eruptions from only volcanoes with the same primary rock composition
3. Eruptions from only volcanoes with the same type (morphology) class (stratovolcano, etc.)
4. Eruptions from only volcanoes with the same composition and type
5. Eruptions from only the target volcano

The idea is not that the volcano composition or type is necessarily particularly relevant to a particular eruption, but that these are ways of characterizing the range of behaviors that might be expected. Note - we use only the first main composition listed. The composition must be that of the volcano rather than the eruption to avoid prevision. We do not consider the tectonic setting, to avoid an undesirably large number of permutations, and because we expect this to be correlated with composition and type. In some cases (e.g., Aluto and Pinatubo) we will also consider exogenously defined analogue sets for the purposes of comparison.

4. "Best" Analogs

As our aim is to determine a general strategy for the selection of analogue volcanoes in intra-eruption forecasting, and we are fortunate enough to have a relatively large amount of data, we will evaluate analogs via a "leave-one-out" approach. That is, we will delete the eruption to be forecast from our data, and then subset the data according to our strategy, and see how well the resulting model forecasts the omitted eruption. If we repeat this for all

Table 7
Best Analogs for Eruptions at Two Basaltic Shields (Piton de La Fournaise (PdF) and Ambrym) and Two Andesitic Calderas (Asosan and Krakatau)

Analogue definition	All eruptions	Only same composition	Only same type	Only same type and composition	Only same volcano
No. analogue euptions	2669	817	280	207	77 (PdF)
	2669	817	280	207	35 (Ambryn)
	2669	1502	270	189	69 (Aso)
	2669	1502	270	189	36 (Krakatau)
Best analogue					
(PdF)	3	6	3 [2]	15 [3]	51 [9]
(Ambryn)	6	7	3 [1]	10 [2]	10 [7]
(Asosan)	4	9	12 [1]	22 [4]	23 [6]
(Krakatau)	10	4 [1]	6 [4]	5 [6]	12 [9]

Note. The first four rows show number of global eruptions in each category, the latter four rows the number of eruptions from that volcano that are best forecast using that analogue set. Numbers in square brackets refer to the number of eruptions in that analogue subset for which there is no otherwise observed transition or quiescence in the omitted eruption being forecast.

2670 eruptions, we can then consider which strategies are significantly better in a statistical sense (Marzocchi & Jordan, 2014).

4.1. Evaluating Performance

Our definition of the best analogue is the one whose use best forecasts the quantity of interest at the target volcano. Here that quantity is the phase sequence and timing of an eruption. If, for example, we were trying to forecast the onset of eruption, a different set of analogs might well be chosen.

Evaluating the performance of the analogue subset in forecasting our intra-eruption progression can easily be done by means of the log-likelihood; the difference in log-likelihoods between two different models is equivalent to the information gain (Bebbington, 2005). We will represent an eruptive sequence as

$$\text{start}, S_1, D_1, Q_{1,2}, S_2, D_2, Q_{2,3}, \ldots, Q_{n-1,n}, S_n, D_n, \text{end}$$

where S_1, …, S_n are the eruptive states, D_1, …, D_n are their durations, and the Q's are the quiescences between eruptive phases, which may be zero. Then the log-likelihood is:

$$\log L = p_{\text{start},S_1} p_{S_1,S_2} f_{S_1,S_2}(D_1)\, h_{S_1,S_2}(Q_{1,2}) \cdots p_{S_n,\text{end}} f_{S_n,\text{end}}(D_n). \quad (6)$$

As the log likelihood is a summation of $O(n_{ph})$ terms, where n_{ph} is the number of phases in the eruption, we will use log L/n_{ph} as our measure of fitness, as this weights all eruptions equally, which is appropriate as the number of phases is not known a priori.

However, when developing an analogue forecast model the data may be limited to (e.g.,) only eruptions from volcanoes with the same rock composition, or to the target volcano itself. As the target eruption is removed, this may result in there being no eruptions in the analogue set in some cases. It may also cause there to be no analogue eruption with a required transition and/or quiescence pattern. Some analogue sets are more vulnerable to not having a likelihood than others. For example, if we use a type-based analogue, 4% of the eruptions contain at least one transition or quiescence that is not in the analogue model, and hence cannot be forecast. More generally, any set of analogue volcanoes needs to incorporate sufficient variability (Rodado et al., 2011). As an example of this problem in even the best cases, we will consider two basaltic shields, Piton de la Fournaise and Ambrym, and two andesitic calderas, Asosan and Krakatau (Table 7).

Table 7 shows, first, the number of analogue eruptions using each possible analogue class as specified above. The subsequent lines show how many eruptions are best forecast by each analogue set. The numbers in square brackets are the number of eruptions for which that analogue set does not have either a required transition or quiescence,

Table 8
Null Transition Matrix

	Effusive	Eff + Exp	Cts Exp	Int Exp	Min Exp	Minor E	Major E	Plinian	End
ALL	0.107	0.089	0.033	0.210	0.104	0.145	0.029	0.004	0.280

and hence we cannot calculate the log-likelihood in (6). We see that Piton de la Fournaise is a good analogue for itself, with 66% of its eruptions best forecast using this set, whereas Ambrym seems equally well (or poorly) modeled by any class of analogue, with less than 30% of eruptions best forecast by itself. However, we see that even data from Piton de la Fournaise, with 78 eruptions, still cannot forecast nine of its own eruptions. Asosan and Krakatau also present contrasting behaviors, with the former better forecast by itself (33% of eruptions) or other andesitic calderas (31% of eruptions), whereas the latter is best forecast by itself (32% of eruptions) and by the entire set of eruptions (27% of eruptions), regardless of volcano. Note that the incidence of unobserved types of transitions is greater for this pair of andesitic calderas than for the previous pair of basaltic shields.

4.2. Null Analogs

The large number of eruptions that are not forecastable, even for the well-recorded examples in the previous section, indicates the necessity of devising a method for constructing robust analogue models. To do this we need to augment the transition matrix and duration distributions obtained from a set of analogue eruptions in such a way that there are no zero transition probabilities, and that all possible durations and quiescences have non-zero support. We will do this by first constructing a null analogue, that is, one without any "tuning" to individual volcanoes. This will serve two purposes: As a benchmark to assess if the analogue sets provide any information (forecasting skill), and second as the required augmentation to the model derived from an analogue set.

Our null analogue will not consider any information in the phase eruption data beyond the overall distributions. That is, it will assume that the next phase is independent of the current and past information, and is assigned randomly according to the overall prevalence of that state in the complete data set. Let π be the vector containing the proportion of each eruptive phase in the entire set of eruption records. Then the transition matrix of the null analogue is

$$T^{(0)} = 1\pi'$$

where 1 is a vector of ones. This corresponds to independent (of previous state) transitions. The duration in eruptive State i is then modeled as a lognormal random variable with parameters $\mu_i^{(0)} = \text{mean}\,(\log D_i)$ and $\sigma_i^{(0)} = \text{SD}\,(\log D_i)$, where D_i are the observed durations (from the entire data set) in State i. Again, this corresponds to independent durations. Post-quiescent durations are handled similarly, including the probability of no quiescence. Again, for consistency, we add nominal quiescence lengths of one and 90 days to the observed data before calculating the mean and variance. The null transition matrix is given in Table 8 (all rows identical).

4.3. Combining Null and Non-Null Analogue Models

Investigation disclosed that in some cases, the null analogue can perform better than the analogue subsets. Hence the most effective forecast model may use a combination of the two (a null and non-null analogue model), which will also cover cases in which transitions are missing.

Various weightings were considered for quantifying the influence of the null analogue model in our combined model forecast, with the weight attached to the null analogue, $w^{(0)}$, being both constant and dependent on the number of eruptions, n_a, in the analogue set. The weight given to the non-null analogue is $1 - w^{(0)}$, and the alternatives considered are shown in Figure 4. Note that $w^{(0)} = 1$ when $n_a = 0$ in all schemes. Besides possessing the essential monotonicity with number of eruptions, the candidates in Figure 4 exhibit a wide range of decay behaviour with the size of the analogue set.

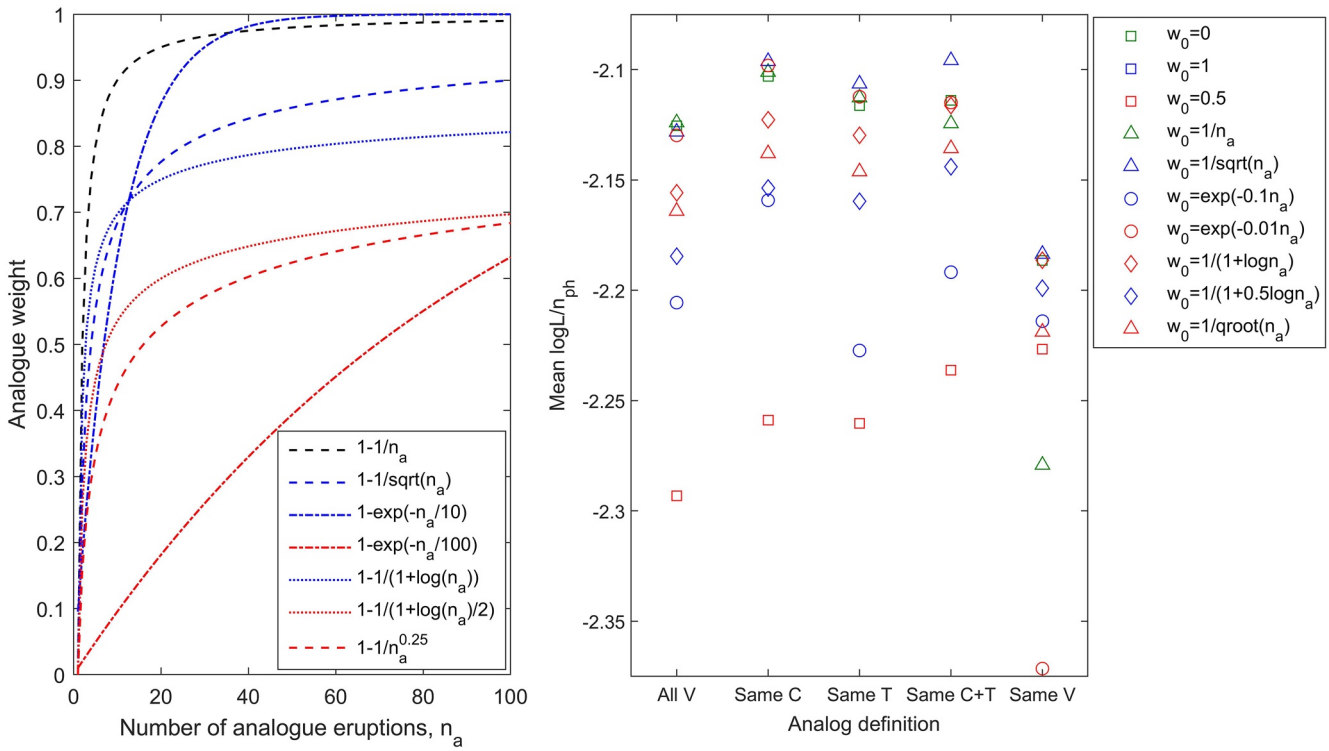

Figure 4. Weighting schemes examined for null analogue as part of combined analogue. Shown (left) is the weight assigned to the actual data depending on the number of analogue eruptions in that data set and, (right) the average log L normalized by number of phases for each analogue set definition; all values for $w_0 = 1$ (null model) are −2.53.

The *combined analogue model* is then assembled as a linear combination of the null analogue and the non-null analogue data model. The transition matrix is thus $T = w^{(0)}T^{(0)} + (1 - w^{(0)})T^{(A)}$ where $T^{(A)}$ is the transition matrix assembled from the analogue eruptions. Similarly

$$\mu = w^{(0)}\mu^{(0)} + \left(1 - w^{(0)}\right)\mu^{(A)}, \sigma = \sqrt{w^{(0)}(\sigma^{(0)})^2 + (1 - w^{(0)})\,(\sigma^{(A)})^2}$$

(weighted mixture of both mean and variance) for both eruptive state durations and quiescence duration, while

$$p_0 = w^{(0)}p_0^{(0)} + \left(1 - w^{(0)}\right)p_0^{(A)}$$

where p_0 is the probability of having no post-eruptive state quiescence. Note that in all cases the analogue terms (those with superscript (A)) are dependent on the previous eruptive state as well as the current one. Overall, the weighting scheme $w_0 = 1/\sqrt{n}$ appeared to be the most robust, having the best performance across all the analogue set definitions, except that of "all volcanoes", where it was narrowly outperformed by $w_0 = 1/n$. It is worth noting that it also outperforms the non-null analogue when using any restriction of the analogue set, even when the unforecasted eruptions are removed before averaging. This underscores the need for any forecasting model to contain a margin for "surprises". Below we will use $w_0 = 1/\sqrt{n}$.

As an example, for the Aluto analogue above, we have 94 eruptions, so $w^{(0)} = 1/\sqrt{94} = 0.103$ and the resulting combined analogue transition matrix is given in Table 9, which can be compared with that in Table 5. We see that the zeros have been replaced by small, non-zero, probabilities, but that the remainder of the transition probabilities are little affected.

Table 9
Combined Transition Matrix (Aluto Example)

	Effusive	Eff + Exp	Cts Exp	Int Exp	Min Exp	Minor E	Major E	Plinian	End
Start	0.231	0.162	0.032	0.174	0.144	0.177	0.050	0.0004	0.029
Effusive	0.059	0.128	0.051	0.022	0.058	0.038	0.003	0.0004	0.642
Eff + Exp	0.251	0.039	0.064	0.201	0.011	0.015	0.033	0.0004	0.388
Cts Exp	0.210	0.009	0.003	0.121	0.011	0.015	0.103	0.0004	0.528
Int Exp	0.033	0.116	0.025	0.214	0.032	0.036	0.110	0.0004	0.434
Min Exp	0.051	0.009	0.003	0.062	0.214	0.015	0.003	0.0004	0.641
Minor E	0.036	0.059	0.003	0.096	0.011	0.314	0.053	0.0004	0.427
Major E	0.011	0.073	0.067	0.406	0.011	0.335	0.003	0.0004	0.093
Plinian	0.011	0.009	0.003	0.022	0.011	0.015	0.003	0.0004	0.029

5. Results

We used as our measure the improvement in log-likelihood (normalized by the number of phases), $\Delta \log L/n_{ph} = (\log L - \log L^{(0)})/n_{ph}$, where $\log L$ is the model log-likelihood for a given eruption, and log; $L^{(0)}$ the corresponding log-likelihood using the null analogue. Both the non-null analogue data only model and the combined model outperform the null (independent phases) analogue. There is almost no qualitative difference between their performance except when only the target volcano is used in the data, where the combined model is less effective. However, it does produce an answer for all 2670 eruptions, whereas almost 41% of the eruptions are not forecastable using only the non-null data in this case.

5.1. Combined Versus Non-Null Analogs

Figure 5 shows the performances of each analogue class across the data set (leave one out validation) when using the combined null and non-null analogue model (left). The analogue class of "all eruptions" is used as the baseline to which the others are compared. Thus negative values indicate that the use of the entire data set is better than the restriction. For comparison we also show the corresponding figures for "pure", that is, without a null component, analogs (right). Remember that the cases where the forecast scored zero probability are not shown. We observe that, using only the non-null analogs rather than the combined model, all four analogue definitions do better than the reference (all data) analogue on average, this being particularly pronounced when using only other eruptions from the target volcano. However, of the 2670 eruptions, 979 did not achieve a score using the target volcano analogue due to transitions or quiescences not in the analogue set, while a further 109 occurred at volcanoes with only one observed eruption. Even when using the entire data set as the analogue model, 12 of the eruptions were unforecasted. For the combined model, we see similar spread as in the non-null model, but much less evidence that any restriction to the analogue set provides greater forecasting power than using all the data.

Figure 6 shows the 30 volcanoes with the greatest number of eruptions in the data set, and what number are unforecasted using each analogue definition. We see that none of these volcanoes (and in fact, none of the 343 in the data set) have enough eruptions to completely forecast their own activity. The 98 eruptions (most in the data set) for Etna, for example, include 12 which have at least one feature not present in the other 97. Hence we confirm that some form of augmented analogue, such as the combined analogue proposed here, is necessary for robust forecasting.

5.2. Restrictions to Type and/or Composition

Accepting that a combined analogue must be used, we can move on to asking on which analogue volcanoes the non-null portion of the analogue model should be based. Figures 7 and 8 break the likelihood improvement down by the type and composition, respectively, of the target volcano. Note that (small, $n \leq 6$) classes, such as crater rows, volcanic fields, maars and trachyte, where the eruptions are the only one at a given volcano, are omitted.We see from Figure 7 that shields as a class are well forecast by limiting the analogue set by composition, type, both, or to themselves, in line with the results in Table 7. Domes, and subglacial/submarine volcanoes, on the other

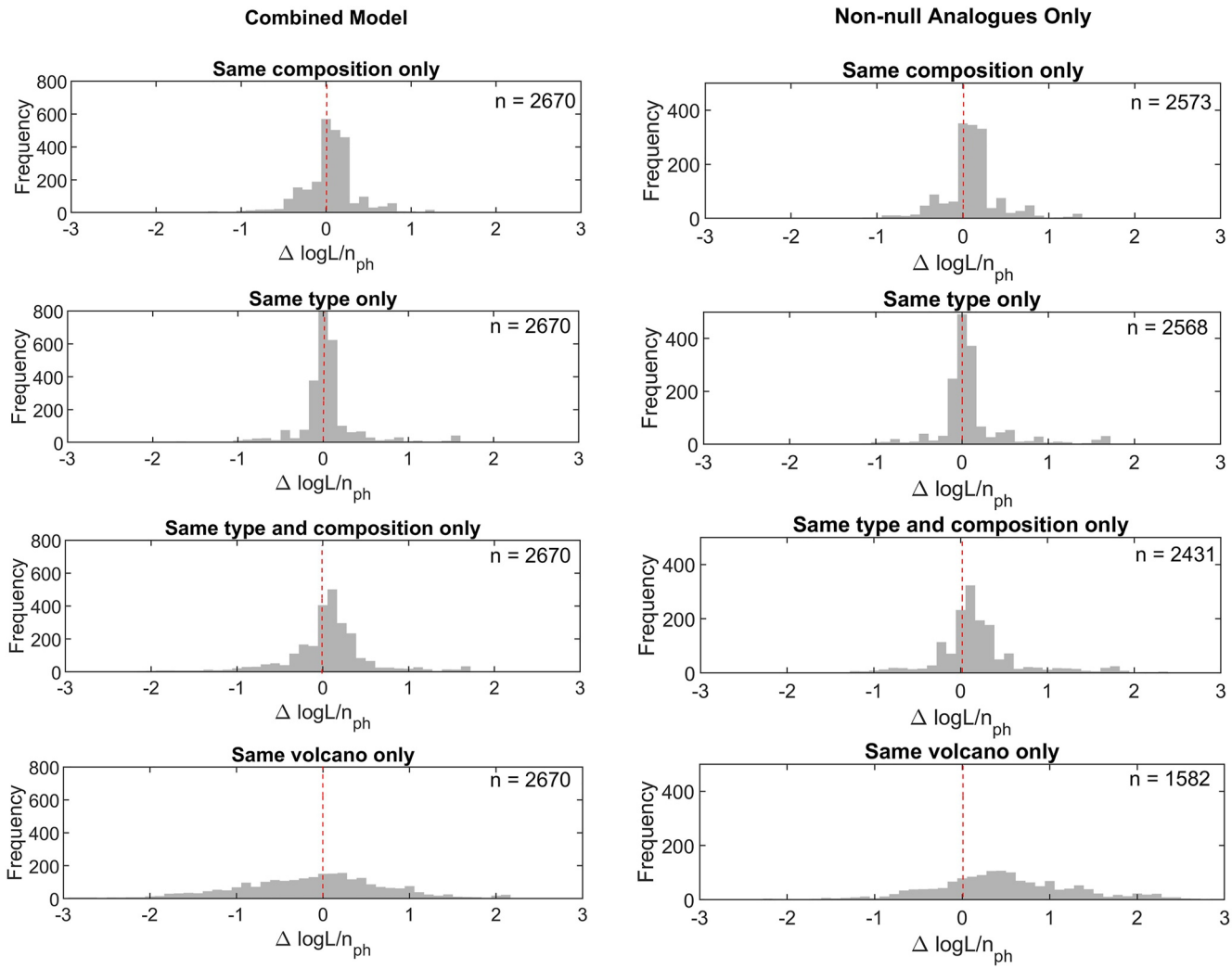

Figure 5. The difference in log-likelihood (normalized by number of phases) in reproducing target eruptions between a reference forecast using all data and when the forecast is based on eruptions from upper) volcanoes of the same volcano composition, upper middle) volcanoes of the same type, lower middle) volcanoes of the same type and composition, and lower) the same volcano. A Δ log L/n_{ph} greater than 0 suggests that the model using eruptions from the restricted analogs is performing better than that using all data (the reference analogue). The value of n is the number of eruptions (out of 2670) represented. Those not shown have a score of −inf. The left panels show the mixed analogue including a null component, while the right panels are without the null component.

hand, are best forecast by using all the data, possibly because there are relatively few eruptions of this type in the data set. The most surprising result, however, is that individual stratovolcanoes (and possibly calderas, complex volcanoes and cones) are relatively poor forecasters of their own behavior, or to put it another way, are potentially fruitful prospects for using analogue data.

From Figure 8 we see that Vesuvius (the only phono-tephrite volcano in our data) is a good predictor of its own behavior, due to its regular behavior (Carta et al., 1981; Santacroce, 1983). Otherwise, unlike the type in Figure 7, we see that no composition is significantly better forecast by restricting the data to that composition. In particular, dacites and rhyolites are poorly forecast by restricting the analogue set, either by composition or to the same volcano. This may reflect a greater variability in eruptions of silicic volcanoes. Oddly, rhyolites and trachyandesites appear to be better forecast by restricting the analogue by volcano type, which we assume to be a small sample quirk.

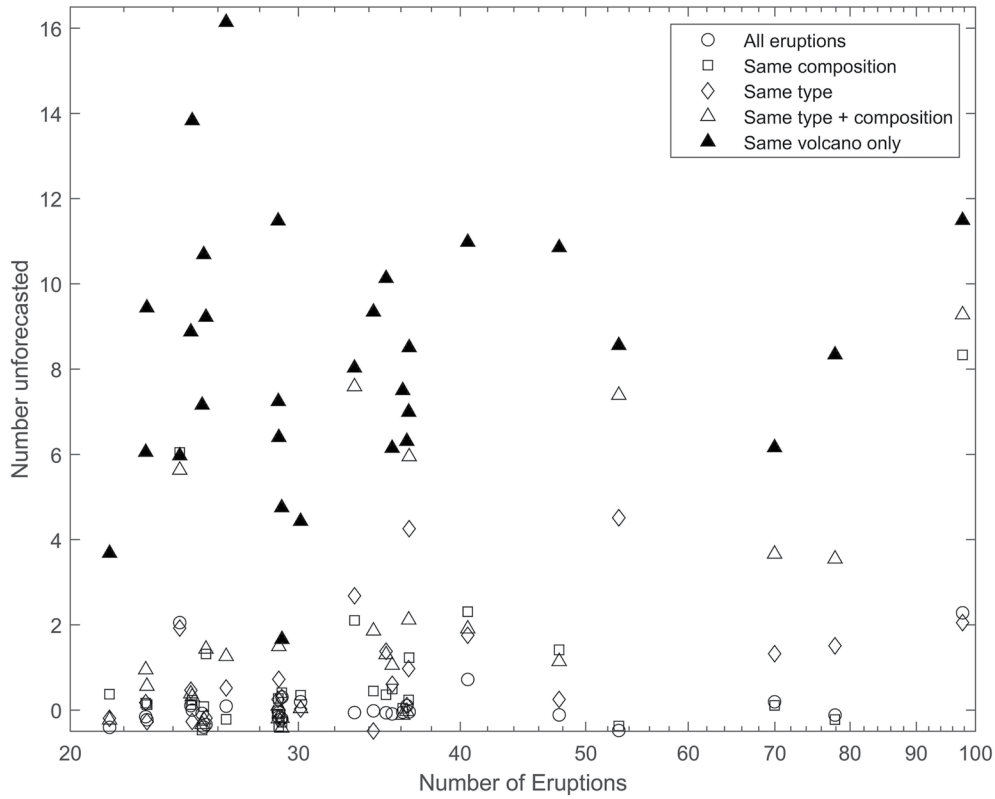

Figure 6. Number of eruptions and how many are unforecasted for the 30 volcanoes with the greatest number of eruptions in our data set. Data have been vertically jittered by up to ±0.25 to improve visibility.

5.3. Dynamic Analogue Sets

One further possibility is to dynamically adjust the analogue set as the eruption progresses. There are probably a number of ways that this can be done, but the one we elected to examine was to remove from the analogue set after each phase, and before forecasting the next phase, any eruption that did not include a phase of that class. Note that this can only be done in the combined analogue paradigm, to ensure that the rapid shrinkage in the analogue data set doesn't make almost every eruption unforecastable. Thus the weighting of the null part of the analogue increases each time the analogue set is pruned. Depending on the analogue definition, an average of 41%–53% of phases in the data set remain after the first observed phase, and after five phases this average is 13%–26%. One interesting observation is that the most data are retained when using only the same volcano, and the most removed when using all the data. Figure 9 shows the difference in mean log L/n_{ph} between the dynamic and static analogs, for each analogue definition, for all 245 volcanoes with more than one eruption in the data set. In general, we see that the performance is worse than the non-dynamic alternative, with the exception of certain volcanoes. Stromboli, in particular, benefits from a dynamic analogue in some cases, largely because it is renowned for Strombolian (continuously explosive) behavior, and the analogue set rapidly contracts only to eruptions featuring this state. Its trachyandesite composition is shared by nine other volcanoes in the data set, among which Tengger, Ioto and Bogoslof have more eruptions. Stromboli appears to differ considerably from these, which means that they also benefit from the dynamic variant by the removal of Stromboli eruptions (which provide a great number of phases). Dynamic forecasts of eruptions at Stromboli and Mount St. Helens perform poorly when other stratovolcanoes are considered as analogs because of the unique nature of Stromboli, and the fact that the phase data for St. Helens is dominated by the dome building behavior of the 1980–1986 eruption. The 24 eruptions in the data set for Vesuvius include the 79AD event and every eruption since CE 1631, but only five of these contain minor eruption phases. This is a very low prevalence relative to the data as a whole (Table 2). Moreover, many of the eruptions in the data set consist of only minor eruption states. Hence, for most of the Vesuvius eruptions, that part of the data set is immediately eliminated in the dynamic analogue, improving the likelihood of 19 out of the 24 eruptions, which appears to be the source of the gain in forecasting power for Vesuvius.

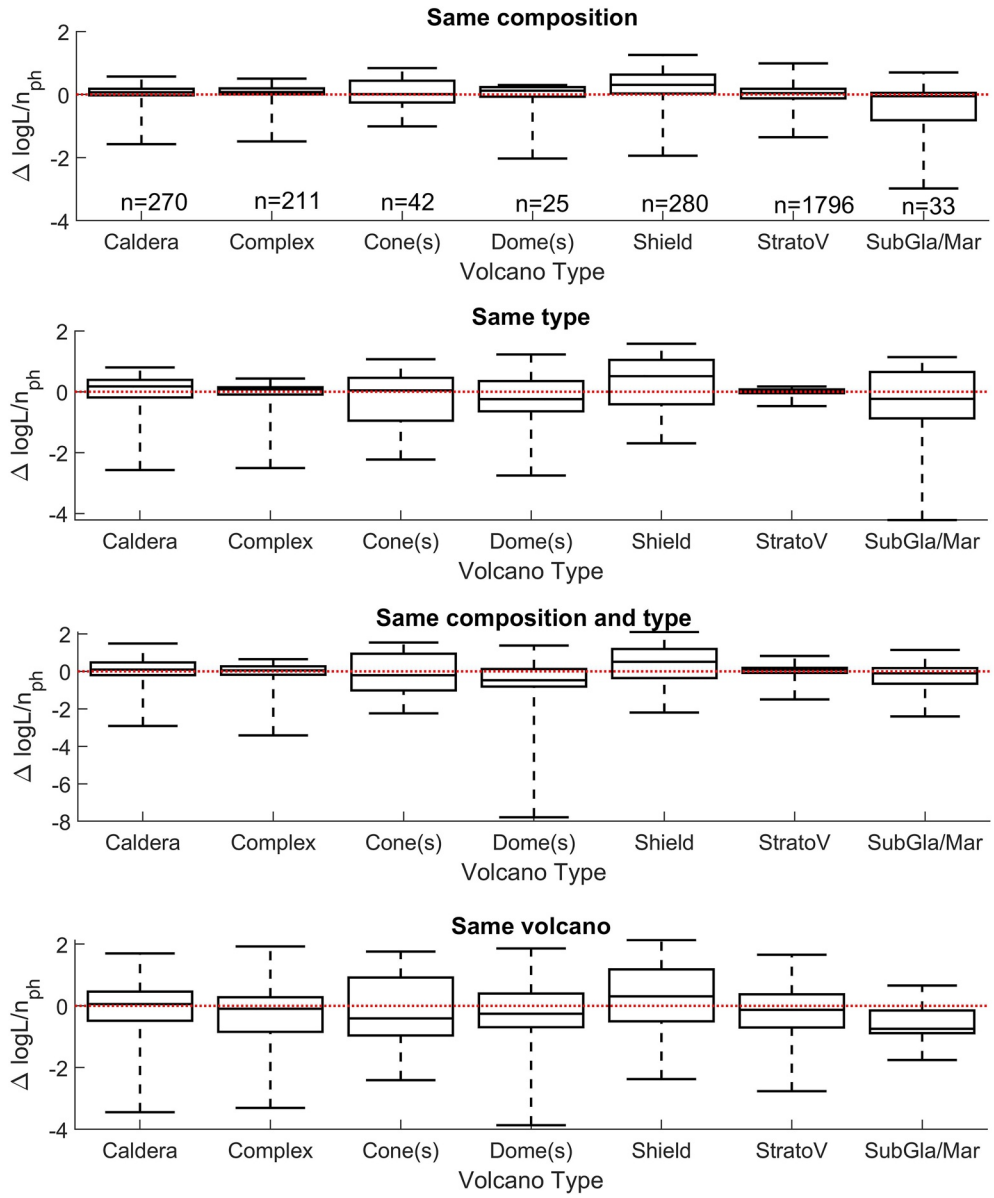

Figure 7. The difference in log-likelihood (normalized by number of phases and broken down by target volcano morphology) in reproducing target eruptions between a reference forecast using all data and when the forecast is based on eruptions from upper) volcanoes of the same volcano composition, upper middle) volcanoes of the same type, lower middle) volcanoes of the same type and composition, and lower) the same volcano. A Δ log L/n_{ph} greater than 0 suggests that the model using eruptions from the restricted analogs is performing better than that using all data (the reference analogue).

5.4. Use of Analogs in Forecasting

Finally, let's look at an example of using analogs to forecast the 1991 eruption of Mt Pinatubo (Newhall et al., 1996). Pinatubo is a dacitic stratovolcano, and hence its analogue sets are all eruptions (n = 2669), eruptions from dacitic volcanoes (54), eruptions from stratovoclanoes (1797), eruptions from dacitic stratovolcanoes (21) and other eruptions from Pinatubo (2). Note that we ignore the fact that many of the eruptions in the data set occurred after 1991. In particular, two post-1991 (1992 and 1993) eruptions from Pinatubo are potentially included in the analogue sets. The 1991 Pinatubo eruption is not used. We also consider an analogue set of the best 25 (to get sufficient eruptions) analogue volcanoes identified using the default weighting from VOLCANS (Tierz et al., 2019), which includes eruptions from Toya (n = 6), Kujusan (1), Puyehue-Cordon Caulle (5), Huaynaputina (1), Kelut (11), Galunggung (3), Krishimayama (15), Cerro Hudson (3), Spurr (1), Avachinsky

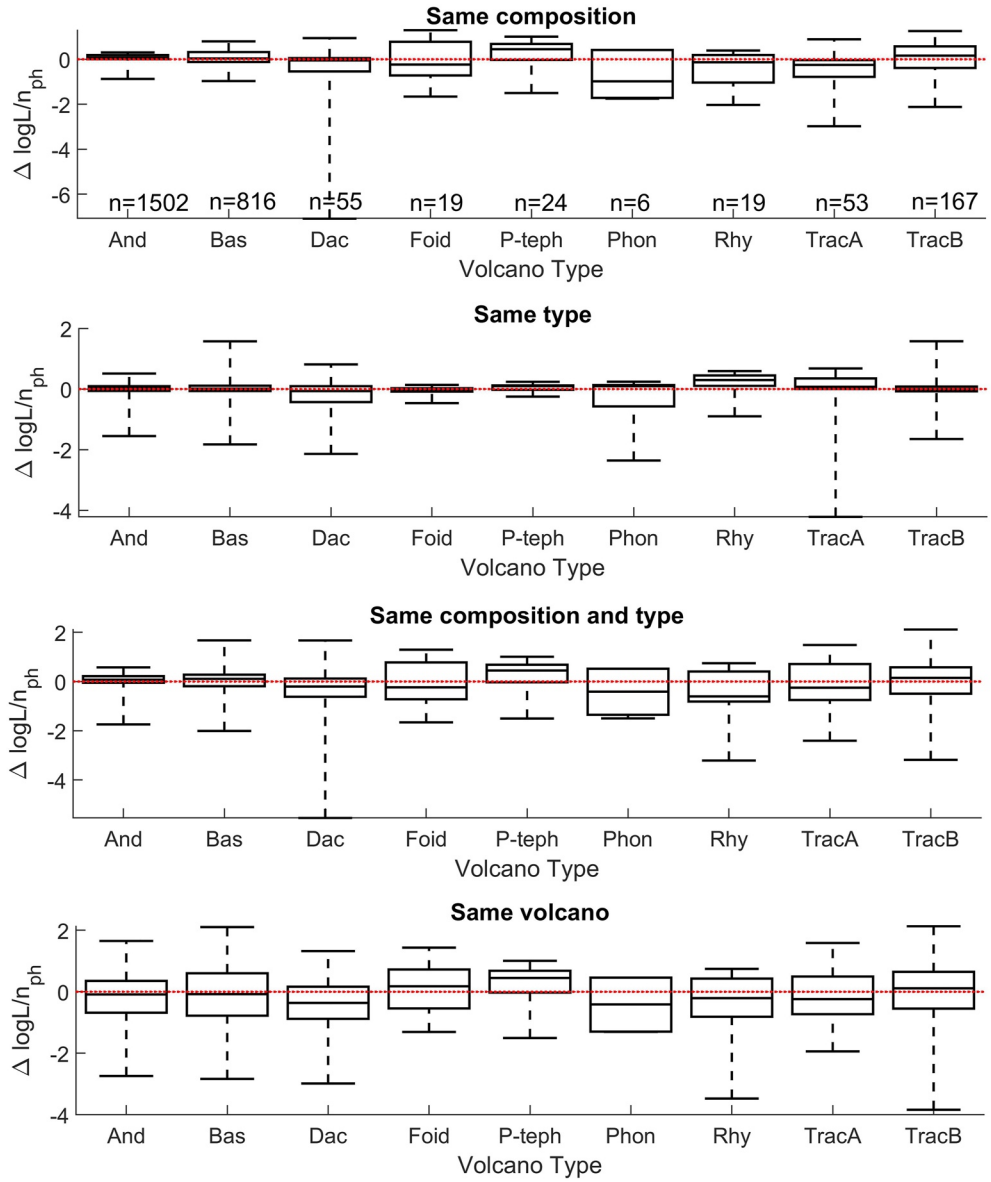

Figure 8. The difference in log-likelihood (normalized by number of phases and broken down by target volcano composition) in reproducing target eruptions between a reference forecast using all data and when the forecast is based on eruptions from upper) volcanoes of the same volcano composition, upper middle) volcanoes of the same type, lower middle) volcanoes of the same type and composition, and lower) the same volcano. A Δ log L/n_{ph} greater than 0 suggests that the model using eruptions from the restricted analogs is performing better than that using all data (the reference analogue).

(8), Gorely (4), and White Island (18), and the set of analogs collectively suggested by Whelley et al. (2015) and Newhall et al. (2017) listed in Section 3.1. The analogs were used to forecast probabilities of Plinian and Major eruptions, and eruption end, on a daily basis (Figure 10). We see that the forecasts without any analogue input are pretty much those of the null model, and very poor. As stratovolcanoes constitute approximately two thirds of the data set, it is not surprising that their forecast is very similar to that using all the data. These are the only two of the examined analogs to estimate a significant probability of the Plinian eruption preceding it and have, along with the Whelley/Newhall analogue, the lowest probability of a Plinian eruption during the long quiesence. The dacitic stratovolcanoes analogue behaves, undesirably, in the opposite fashion. The Whelley/Newhall analogue has the highest probability of a Major eruption during the long quiescence, while VOLCANS has the lowest.

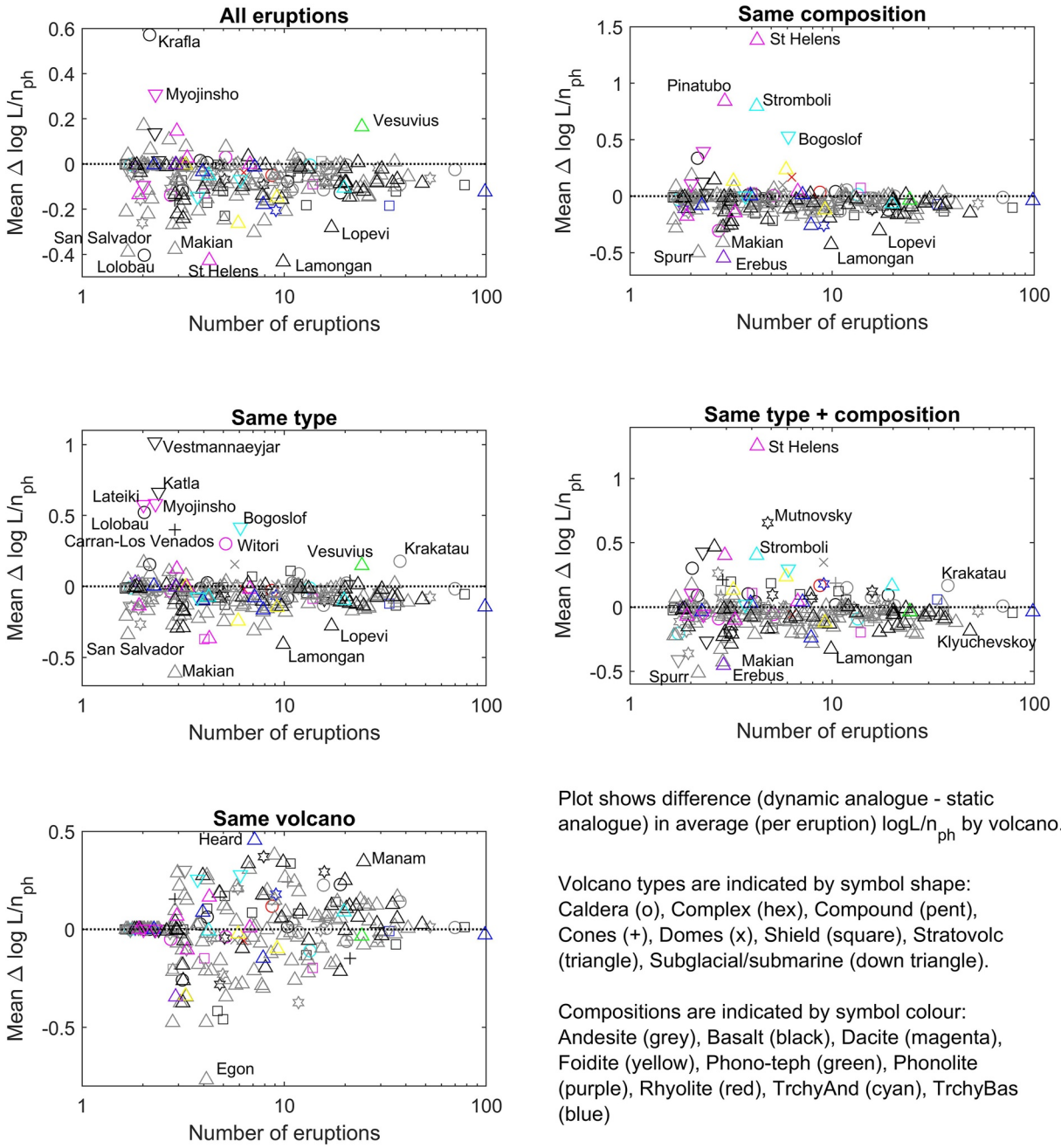

Figure 9. The improvement in mean log *L* by moving to a dynamic analogue. Volcano types are indicated by symbol shape: Caldera (o), Complex (hex), Compound (pent), Cones (+), Domes (x), Shield (square), Stratovolc (triangle), Subglacial (down triangle). Compositions are indicated by symbol color: Andesite (gray), Basalt (black), Dacite (magenta), Foidite (yellow), Phono-teph (green), Phonolite (purple), Rhyolite (red), TrchyAnd (cyan), TrchyBas (blue). A symbol above the dotted line indicates that the dynamic analogue outperforms the static analogue.

6. Discussion

Our discussion will be focused around how individual volcanoes perform as targets, extracting general lessons from the results, how the coverage of the phases data set aligns with the volcanoes it might need to be applied to, and what further investigations might be useful.

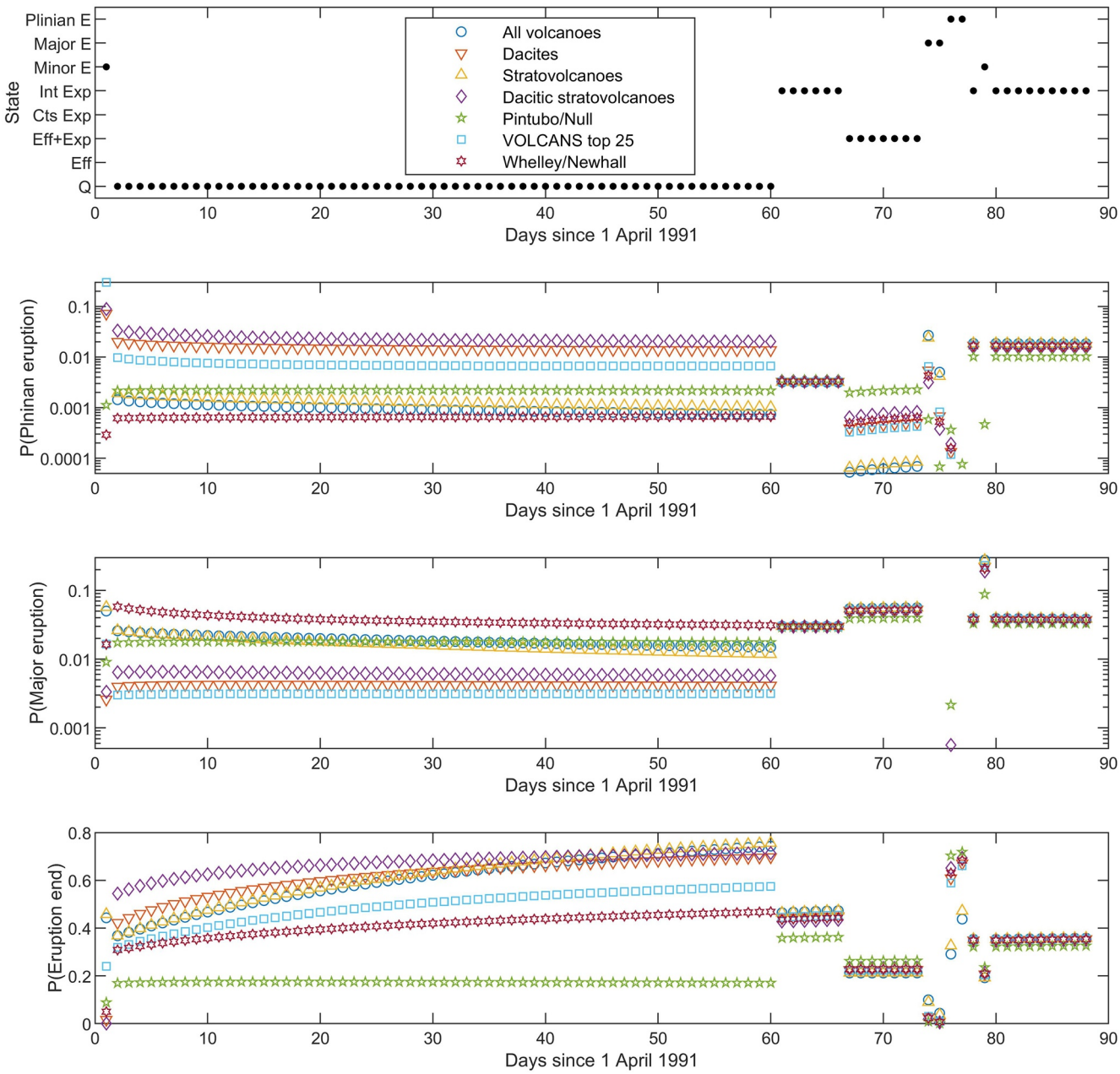

Figure 10. The 1991 eruption from Pinatubo, state (top), estimated probabilities of the next stage being Plinian Eruption (upper middle), Major Eruption (lower middle) and eruption end (bottom).

6.1. Looking at Individual Volcanoes

In an attempt to dig further into which volcanoes might be influencing the analogue performance, we have plotted the log-likelihood gains (over using all of the data) in Figure 11 for 140 of the volcanoes (all those with at least five eruptions in the data set). The outlying volcanoes are annotated. We see a number of shields with positive loglikelihood gain, and also that complexes seem to be poorly forecast by other complexes. Using the same composition results in relatively small gains compared to larger losses, although we see that volcanoes with greater numbers of eruptions buck this trend. Restricting to type shows a much narrower bulk of the distribution, with clearer distinction between good, bad and indifferent. Here, and in the restriction to type and composition, shields exhibit the bulk of the improvement. Restriction to the target volcano suggests that number of data may

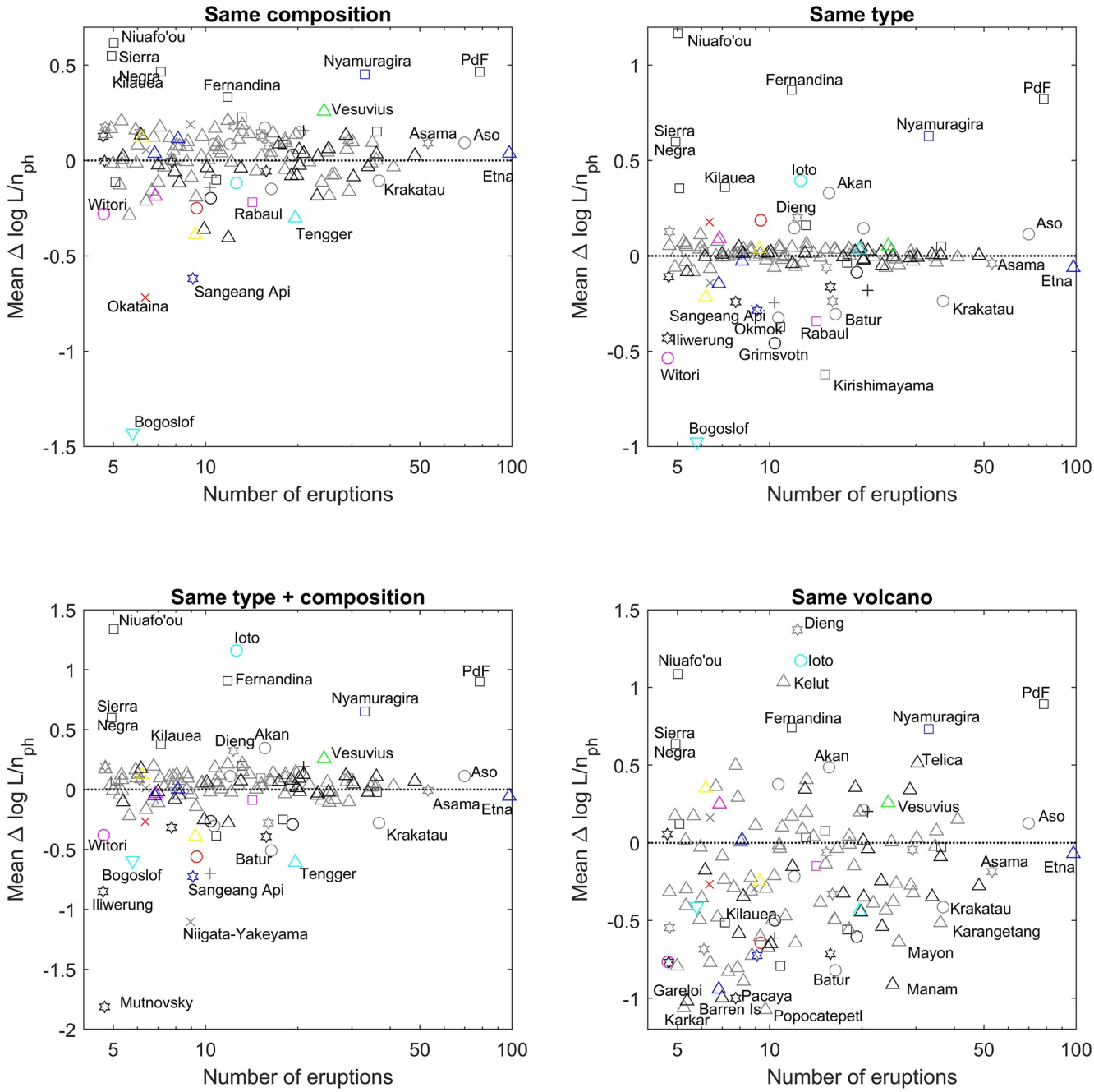

Figure 11. The difference in log-likelihood (normalized by number of phases and broken down by target volcano type and composition) in reproducing target eruptions between a reference forecast using all data and when the forecast is based on eruptions from top left) volcanoes of the same volcano composition, top right) volcanoes of the same type, bottom left) volcanoes of the same type and composition, and bottom right) the same volcano. A Δ log L/n_{ph} greater than 0 suggests that the model using eruptions from the restricted analogs is performing better than that using all data (the reference analogue). Volcano types are indicated by symbol shape: Caldera (o), Complex (hex), Cones (+), Domes (x), Shield (square), Stratovolc (triangle), Subglacial (down triangle). Compositions are indicated by symbol color: Andesite (gray), Basalt (black), Dacite (magenta), Foidite (yellow), Phono-teph (green), Rhyolite (red), TrchyAnd (cyan), TrchyBas (blue).

be a consideration - as the record for a volcano grows, it may improve as a forecaster of its own behavior. This is desirable behavior (Decker, 1986; Pallister et al., 2019), and provides some validation of our methods.

To clarify the forecasting picture, we have examined which analogue definition scores highest in predicting each eruption. Figure 12 shows, for 140 of the volcanoes (all those with at least five eruptions in the data set), the proportion of their eruptions that are best forecast by each analogue definition. Note that there can be ties, not

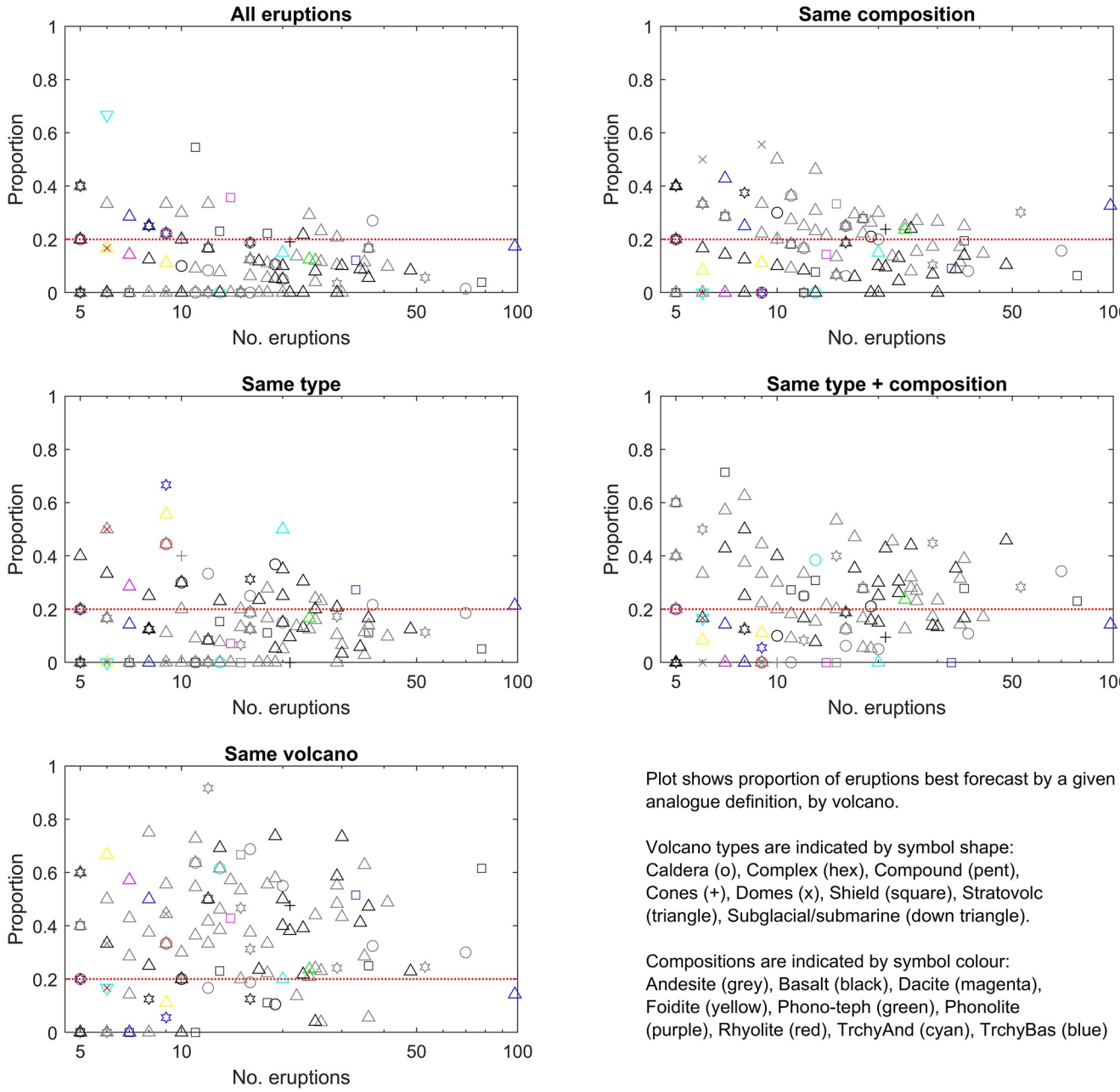

Figure 12. The proportion of eruptions from each volcano that are best forecast by a given analogue definition. Volcano types are indicated by symbol shape: Caldera (o), Complex (hex), Cones (+), Domes (x), Shield (square), Stratovolc (triangle), Subglacial (down triangle). Compositions are indicated by symbol color: Andesite (gray), Basalt (black), Dacite (magenta), Foidite (yellow), Phono-teph (green), Rhyolite (red), TrchyAnd (cyan), TrchyBas (blue).

least because the analogs can be the same in certain cases of unique types or compositions. A reference line is shown at 0.2, the proportion expected if all analogue definitions are equally informative. This shows that volcanoes with greater numbers of eruptions are more prone to be best forecast by themselves, and that the utility of all of the other analogue definitions decrease with number of eruptions. Hence limiting the analogue set to the same volcano is a viable strategy provided the record has c. 20 or more eruptions, and that using the whole data set may not be optimal if the record has more than c. 30 eruptions.

6.2. So, How Well Does it All Work?

We have used the log-likelihood as a measure of the forecast accuracy. This is a consistent measure, as shown by the increase in utility of restricting to the target volcano as the record gets larger. However, we need to emphasize that many eruption sequences evolve in atypical (for the volcano or "class" of volcanoes) fashion. The all volcanoes analogue is most robust to this. In particular, while the log-likelihood is a ratio of probabilities measure, the risk is a product of probability and other factors, and so does not necessarily align with the log-likelihood. It is generally far more dangerous to miss the unlikely than it is to misestimate the probability of the likely.

With that caveat, we first observe in general terms that domes, and those volcanoes with fewer data, in particular complexes and a large number of basaltic and andesitic stratovolcanoes, are best forecast using the same type and composition. However, Figure 11 throws up a number of potentially interesting observations. A number of shields are quite well forecast by restricting the data to only shields, which makes sense as eruptions from shields are, by definition, mainly effusive, and hence largely limited to a restricted set of phase types. However, within that class, we note that Niuafo'ou is always well forecast by *any analogue restriction* despite having only 5 eruptions. In contrast to the shields, some volcanoes such as Krakatau and Etna, with large numbers of eruptions, are poorly forecast by any analogue restriction, including their own record. This perhaps signals that, apart from shields, any volcano with a large number of eruptions exhibits a wide range of behaviors, and may tend toward some sort of "average volcano". Kelut, with fewer eruptions in the data set, of significantly unusual character, illustrates this point, as it is undeniably best forecast by itself. Kelut is also somewhat unusual in that its temporal record is Poisson-like (Cronin & Bebbington, 2008), and its VEI distribution has a significant excess of larger events (Bebbington, 2014).

Figure 11 is also interesting in that while there are a number of volcanoes indicated that perform well independently of the analogue definition, the ones that perform significantly poorly differ from analogue definition to analogue definition. For example, Popocatepetl is a poor forecaster of itself, but any other analogue set performs reasonably. We note also that being a shield is not necessarily a guarantee of good analogue performance, as Rabaul and Okmok (for example) are not well forecast by other shields, perhaps due to their large calderas.

The oddities (such as Vesuvius) that exhibit a predominate and unusual eruptive style are well forecast by themselves, but in general, composition subsetting is not helpful, as the composition that matters is that of the eruption, not the volcano (Cassidy et al., 2018; Sides et al., 2014), and volcanoes produce multiple compositions and there is no robust means for forecasting the next composition.

Somewhat surprisingly, the experiment with dynamic analogs shows that the observed activity during an ongoing eruption seems to offer little information beyond that contained in the knowledge of the current eruptive state and its elapsed duration. This indicates that the Markovian assumption underlying our model is robust.

6.3. Possible Sensitivity of the Results to Coverage of Composition and Type

It is worthwhile comparing the coverage of the phases data set against the GVP catalog. Table 10 shows the number of eruptions in the phases data set, broken down by the GVP categories for type and composition of the volcano concerned, against the number of volcanoes in that category within the GVP catalog. This shows the amount of data available in our data set (number of eruptions) to forecast eruptions at certain volcanoes, relative to the proportion of volcanoes that might need forecasts. A low proportion of eruptions and high proportion of volcanoes in the same category suggests that the phases data set would have relatively few data upon which to base a forecast, a high proportion of eruptions and low proportion of volcanoes, the opposite. Andesite stratovolcanoes are by far the best represented class in our phases data set (and in the GVP), comprising 41% of the total eruptions in the phases data set, and and 33% of the volcanoes in the GVP. At the other end of the scale, the GVP includes 16 rhyolite stratovolcanoes: 10 in Ethiopia, 1 in Turkey, 2 in Alaska, 1 in Iceland, 1 in Greece, 1 in Italy, with most recent eruptions in 1820, 1650 and 1477. There are also 4 rhyolite shields, most recent eruption 1891 (Pantelleria). We have zero examples of these potentially catastrophic eruptions in our phases data set. The robustness of using all of the phases data set to forecast eruptions is particularly useful when considering underrepresented classes of volcano.

Given the coverage mismatch between volcanoes in the GVP and the eruptions in the phases data set, we did not investigate in the analogue context restricting the data set to only eruptions after a certain date, as was considered

Table 10
Number and Percentage of Eruptions (E) Within the Phases Data Set, Following the Global Volcanism Program (GVP) Categories for Type and Composition of the Volcano Producing Them, Relative to the Number and Percentage of Volcanoes (V) of the Same Type and Composition Within the GVP Catalog

Type		Andesite	Basalt	Dacite	Foidite	Phono-tephrite	Phonolite	Rhyolite	Trachy-andesite	Trachy-basalt	Trachyte	Percentage
Caldera	E	189	45	8				11	12		1	9.98
	V	33	11	15	1			18	2		3	6.80
Complex	E	164	35				2	1		9		7.92
	V	32	9	2	1		1	3		1	1	4.10
Compound	E	6										0.23
	V	8	2	1						1		0.98
Cone	E	12	26					1	1		2	1.58
	V	26	53	2	6	1		5	6	20	3	10.00
Fissures	E		4									0.15
	V	0	21						1	1		1.89
Domes	E	18						6	1			0.94
	V	14	3	9				9	1		1	3.03
Shield	E	20	207	18						34	1	10.51
	V	16	117	5			2	5	1	10	12	13.77
Strato	E	1086	491	22	19	24	4		26	123	2	67.43
	V	403	27	42	4	2	3	16	12	17	15	44.34
Submarine	E	7	8	7					10			1.20
	V	20	48	12			2	3	3	3	1	7.54
Fields	E	1										0.04
	V	11	46	2				2	3	16		6.56
Maars	E		1									0.04
	V	2	8		2							0.98
Percentage	E	56.40	30.66	2.06	0.71	0.90	0.23	0.71	1.88	6.23	0.23	
	V	46.31	28.28	7.38	1.15	0.25	0.66	5.00	2.38	5.66	2.95	

by Bebbington and Jenkins (2019). There have been several interventions in monitoring and recording during that time, including the increasing use of satellites and, indeed, the inception of the Bulletin of the Global Volcanism Network (https://volcano.si.edu/reports_bgvn.cfm).

6.4. Future Ideas

Several avenues of further investigation are possible. One of the most promising is to somehow augment the model to account for the elapsed time since the last eruption (Bebbington, 2014), on the assumption that closed-vent versus open-vent volcanoes (Whelley et al., 2015) will exhibit different eruptive styles when they reawaken. A possible development would include a partially observable Markov state (Aspinall et al., 2006; Bebbington, 2007) to account for conduit conditions, but a careful treatment of potentially missing eruptions will be needed.

For the purpose of eruption scenario creation for civil defense (Weir et al., 2022) or economic risk analysis (McDonald et al., 2017), a subsetting to the desired eruption VEI could be useful. In the case of forecasting an ongoing eruption that has already reached a certain VEI, limiting the analogue set to eruptions of at least that VEI would be indicated. A particularly valuable augmentation would be to incorporate monitoring data (Jenkins et al., 2019; McCausland et al., 2019) as a modulating factor in the transition matrix. This would include unrest (during quiescence) within the model, sharpening the forecasting aspect. This could then be extended to further expand the scenario design framework.

More pedestrian options are to look at further combining analogs, for example, considering weighted analogs from all volcanoes and the same volcano.

7. Conclusions

This study builds upon that of Bebbington and Jenkins (2019), which compiled a data set of 698 multi-phase eruptions from 187 volcanoes around the world. We expand the data set and make it freely available, with the updated data set consisting of 2670 single- and multi-phase eruptions from 353 volcanoes, which accounts for nearly 95% of eruptions with text descriptions recorded in the GVP. We introduce a new state "minor explosion" in coding our phases, to cover small, short duration, single explosions (typically produced during VEI 1 eruptions). The (semi-) Markov chain forecast model previously developed is re-applied to the expanded data set and to subsets of the data set that might represent volcano or eruption analogs, for example, of the same composition and volcano type, to establish in what situations analogs can provide more informative forecasts than using the entire data set.

We found that the possibility of rare transitions and absent intervening quiescences requires the introduction of a "null analogue" (simply representing the distribution of phases across the nine different states) element to avoid "surprises" in these respects. With rare exceptions (mainly shields), subsetting the data set by morphology and/or dominant composition does not provide an analogue which improves on using all of the data. This suggests that the phases decomposition captures almost all of the eruption information in a volcano-agnostic form, missing only the information unique to the target volcano. When an analogue is required, we generally have no idea of the latter, but the best way to account for it is to use the entire data set. In the case of volcanoes with sufficient data, or consistently unique behavior, analogs do not significantly improve on using the record of the target volcano alone. But in these cases, a recourse to analogs is unlikely.

There is value in further investigating how intra-eruption forecasts can be used prior to and during an eruption to support crisis management and preparedness.

Data Availability Statement

The curated phases data set is available from Bebbington and Jenkins (2022), and accessible at the following URL: https://doi.org/10.21979/N9/X1LHOT

Acknowledgments
We thank Ang Pei-Shan and Chai Min Wei for distilling some of the data. Careful reading by Tom Sheldrake and an anonymous reviewer improved the organization of the paper. This work is supported by New Zealand Resilience to Nature's Challenges Volcano Programme, Grant GNS-RNC047 (MB), and by the Earth Observatory of Singapore (SJ) via its funding from the National Research Foundation Singapore and the Singapore Ministry of Education under the Research Centres of Excellence initiative. This work comprises EOS contribution number 419. Open access publishing facilitated by Massey University, as part of the Wiley - Massey University agreement via the Council of Australian University Librarians.

References

Aspinall, W. P., Carniel, R., Jaquet, O., Woo, G., & Hincks, T. (2006). Using hidden multi-state Markov models with multi-parameter volcanic data to provide empirical evidence for alert level decision-support. *Journal of Volcanology and Geothermal Research*, *153*(1–2), 112–124. https://doi.org/10.1016/j.jvolgeores.2005.08.010

Aspinall, W. P., Woo, G., Voight, B., & Baxter, P. J. (2003). Evidence-based volcanology: Application to eruption crises. *Journal of Volcanology and Geothermal Research*, *128*(1–3), 273–285. https://doi.org/10.1016/s0377-0273(03)00260-9

Bas, L. J. M., Maitre, L. W. R., Streckeisen, A., & Zanettin, B. (1986). A chemical classification of volcanic rocks based on the total alkali-silica diagram. *Journal of Petrology*, *27*(3), 745–750. https://doi.org/10.1093/petrology/27.3.745

Bebbington, M. S. (2005). Information gains for stress release models. *Pure and Applied Geophysics*, *162*(12), 2299–2319. https://doi.org/10.1007/s00024-005-2777-5

Bebbington, M. S. (2007). Identifying volcanic regimes using hidden Markov models. *Geophysical Journal International*, *171*(2), 921–942. https://doi.org/10.1111/j.1365-246x.2007.03559.x

Bebbington, M. S. (2013). Assessing probabilistic forecasts of volcanic eruption onsets. *Bulletin of Volcanology*, *75*(12), 783. https://doi.org/10.1007/s00445-013-0783-5

Bebbington, M. S. (2014). Long-term forecasting of volcanic explosivity. *Geophysical Journal International*, *197*(3), 1500–1515. https://doi.org/10.1093/gji/ggu078

Bebbington, M. S., & Jenkins, S. F. (2019). Intra-eruption forecasting. *Bulletin of Volcanology*, *81*(6), 34. https://doi.org/10.1007/s00445-019-1294-9

Bebbington, M. S., & Jenkins, S. F. (2022). Eruption phases [Dataset]. 202203-07),DRNTU. https://doi.org/10.21979/N9/X1LHOT

Bebbington, M. S., & Lai, C. (1996). On nonhomogeneous models for volcanic eruptions. *Mathematical Geology*, *28*(5), 585–600. https://doi.org/10.1007/bf02066102

Bebbington, M. S., Stirling, M., Cronin, S., Wang, T., & Jolly, G. (2018). National-level long-term eruption forecasts by expert elicitation. *Bulletin of Volcanology*, *80*(6), 56. https://doi.org/10.1007/s00445-018-1230-4

Bonadonna, C., Connor, C. B., Houghton, B. F., Connor, L., Byrne, M., Laing, A., & Hincks, T. K. (2005). Probabilistic modeling of tephra dispersal: Hazard assessment of a multiphase rhyolitic eruption at Tarawera, New Zealand. *Journal of Geophysical Research*, *110*(B3), B03203. https://doi.org/10.1029/2003jb002896

Carta, S., Figari, R., Sartoris, G., Sassi, R., & Scandone, R. (1981). A statistical model for Vesuvius and its volcanological implications. *Bulletin of Volcanology*, *44*(2), 129–151. https://doi.org/10.1007/bf02597700

Cashman, K., & Biggs, J. (2014). Common processes at unique volcanoes—A volcanological conundrum. *Frontiers in Earth Science*, *2*, 28. https://doi.org/10.3389/feart.2014.00028

Cassidy, M., Manga, M., Cashman, K., & Bachmann, O. (2018). Controls on explosive-effusive volcanic eruption styles. *Nature Communications*, *9*, 1–16. https://doi.org/10.1038/s41467-018-05293-3

Cronin, S. J., & Bebbington, M. S. (2008). Does regional tectonism control the rate of volcanism along convergent margins? In *IAVCEI 2008 general assembly*. Reykjavik.

Decker, R. W. (1986). Forecasting volcanic eruptions. *Annual Review of Earth and Planetary Sciences*, *14*(1), 267–291. https://doi.org/10.1146/annurev.ea.14.050186.001411

GVP (2013). In E. Venzke (Ed.), *Volcanoes of the world, v. 4.0*. Smithsonian Institution. Provided to us June 27 2016. https://doi.org/10.5479/si.GVP.VOTW4-2013

Hincks, T. K., Komorowski, J.-C., Sparks, R. S. J., & Aspinall, W. P. (2014). Retrospective analysis of uncertain eruption precursors at La Soufriere volcano, Guadeloupe, 1975-77: Volcanic hazard assessment using a Bayesian belief network approach. *Journal of Applied Volcanology*, *3*, 1–26. https://doi.org/10.1186/2191-5040-3-3

Hone, D. W. E., Mahony, S. H., Sparks, R. S., & Martin, K. T. (2007). Cladistic analysis applied to the classification of volcanoes. *Bulletin of Volcanology*, *70*(2), 203–220. https://doi.org/10.1007/s00445-007-0132-7

Jenkins, S. F., Goldstein, H., Bebbington, M. S., Sparks, R. S. J., & Koyaguchi, T. (2019). Forecasting explosion respose intervals with a non-parametric Bayesian survival model: Application to Sakura-jima volcano, Japan. *Journal of Volcanology and Geothermal Research*, *381*, 44–56. https://doi.org/10.1016/j.jvolgeores.2019.04.008

Jenkins, S. F., Magill, C. R., & McAneney, J. (2007). Multi-stage volcanic events: A statistical investigation. *Journal of Volcanology and Geothermal Research*, *161*(4), 275–288. https://doi.org/10.1016/j.jvolgeores.2006.12.005

Jenkins, S. F., Magill, C. R., McAneney, J., & Hurst, T. (2008). Multistage volcanic events: Tephra hazard simulations for the Okataina volcanic center, New Zealand. *Journal of Geophysical Research*, *113*(F4), F04012. https://doi.org/10.1029/2007jf000787

Jolly, G. E., Keys, H. J. R., Procter, J. N., & Deligne, N. I. (2014). Overview of the co-ordinated risk-based approach to science and management response and recovery for the 2012 eruptions of Tongariro volcano, New Zealand. *Journal of Volcanology and Geothermal Research*, *286*, 184–207. https://doi.org/10.1016/j.jvolgeores.2014.08.028

Kokelaar, B. P. (2002). Setting, chronology and consequences of the eruption of Soufriere Hills volcano, Montserrat (1995-1999). *Geological Society, London, Memoirs*, *21*(1), 1–43. https://doi.org/10.1144/gsl.mem.2002.021.01.02

Magill, C. R., Mannen, K., Connor, L., Bonadonna, C., & Connor, C. (2015). Simulating a multi-phase tephra fall event: Inversion modelling for the 1707 Hoei eruption of Mount Fuji, Japan. *Bulletin of Volcanology*, *77*(9), 81. https://doi.org/10.1007/s00445-015-0967-2

Marzocchi, W., & Bebbington, M. S. (2012). Probabilistic eruption forecasting at short and long time scales. *Bulletin of Volcanology*, *74*(8), 1777–1805. https://doi.org/10.1007/s00445-012-0633-x

Marzocchi, W., & Jordan, T. H. (2014). Testing for ontological errors in probabilistic forecasting models of natural systems. *Proceedings of the National Academy of Sciences*, *111*(33), 11973–11978. https://doi.org/10.1073/pnas.1410183111

Marzocchi, W., Sandri, L., Gasparini, P., Newhall, C. G., & Boschi, E. (2004). Quantifying probabilities of volcanic events: The example of volcanic hazard at Mount Vesuvius. *Journal of Geophysical Research*, *109*(B11), B11201. https://doi.org/10.1029/2004jb003155

McCausland, W. A., Gunawan, H., White, R. A., Indrastuti, N., Patria, C., Suparman, Y., et al. (2019). Using a process-based model of pre-eruptive patterns to forecast evolving eruptive styles at Sinabung Volcano, Indonesia. *Journal of Volcanology and Geothermal Research*, *38*, 253–266. https://doi.org/10.1016/j.jvolgeores.2017.04.004

McDonald, G. W., Cronin, S. J., Kim, J.-H., Smith, N. J., Murray, C. A., & Procter, J. N. (2017). Computable general equilibrium modelling of economic impacts from volcanic event scenarios at regional and national scale, Mt Taranaki, New Zealand. *Bulletin of Volcanology*, *79*(12), 87. https://doi.org/10.1007/s00445-017-1171-3

Miller, T. P., Chertkoff, D. G., Eichelberger, J. C., & Coombs, M. L. (1999). Mount Dutton volcano, Alaska: Aleutian arc analog to Unzen volcano, Japan. *Journal of Volcanology and Geothermal Research*, *89*(1–4), 275–301. https://doi.org/10.1016/s0377-0273(99)00004-9

Neri, A., Aspinall, W. P., Cioni, R., Bertagnini, A., Baxter, P. J., Zuccaro, G., et al. (2008). Developing an event tree for probabilistic hazard and risk assessment at Vesuvius. *Journal of Volcanology and Geothermal Research*, *178*(3), 397–415. https://doi.org/10.1016/j.jvolgeores.2008.05.014

Newhall, C. G. (2000). Volcano warnings. In H. Sigurdsson, B. Houghton, S. McNutt, H. Rymer, & J. Stix (Eds.), *Encyclopedia of Volcanoes* (pp. 1185–1197). Academic Press.

Newhall, C. G., Costa, F., Ratdomopurbo, A., Venezky, D. Y., Widiwijayanti, C., Nang, T. Z. W., et al. (2017). WOVOdat—An online growing library of worldwide volcanic unrest. *Journal of Volcanology and Geothermal Research*, *345*, 184–199. https://doi.org/10.1016/j.jvolgeores.2017.08.003

Newhall, C. G., Daag, A. S., Jr, D. G. F., Hoblitt, R. P., McGeehin, J., Umbal, J. V., et al. (1996). Eruptive history of Mount Pinatubo. In C. G. Newhall, & R. S. Punongbayan (Eds.), *Fire and Mud* (pp. 165–195). PHIVOLCS and University of Washington Press.

Newhall, C. G., & Pallister, J. S. (2015). Using multiple data sets to populate probabilistic volcanic event trees. In P. Papale (Ed.), *Volcanic Hazards, Risks, and Disasters* (pp. 203–232). Elsevier.

Ogburn, S. E., Berger, J., Calder, E. S., Lopes, D., Patra, A., Pitman, E. B., et al. (2016). Pooling strength amongst limited datasets using hierarchical Bayesian analysis, with application to pyroclastic density current mobility metrics. *Statistics in Volcanology*, *2*, 1–26. https://doi.org/10.5038/2163-338x.2.1

Pallister, J. S., Papale, P., Eichelberger, J., Newhall, C. G., Mandeville, C., Nakada, S., et al. (2019). Volcano observatory best practices (VOBP) workshops—A summary of findings and best-practice recommendations. *Journal of Applied Volcanology*, *8*(1), 2. https://doi.org/10.1186/s13617-019-0082-8

Pantazidis, A., Baziotis, I., Solomonidou, A., Manoutsoglou, E., Palles, D., Kamitsos, E., et al. (2019). Santorini volcano as a potential Martian analogue: The Balos Cove Basalts. *Icarus*, *325*, 128–140. https://doi.org/10.1016/j.icarus.2019.02.026

Peltier, A., Poland, M. P., & Staudacher, T. (2015). Are Piton de la Fournaise (La Reunion) and Kilauea (Hawaii) really "analog volcanoes". *Geophysical Monograph Series*, *208*, 507–531.

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*, *125*(1), e2018JB016974. https://doi.org/10.1029/2018jb016974

Punongbayan, R. S., Newhall, C. G., Bautista, M. L. P., Garcia, D., Harlow, D. H., Hoblitt, R. P., et al. (1996). Eruption hazard assessments and warnings. In C. G. Newhall & R. S. Punongbayan (Eds.), *Fire and Mud* (p. 67–85). PHIVOLCS and University of Washington Press.

Rodado, A., Bebbington, M. S., Noble, A., Cronin, S., & Jolly, G. (2011). On selection of analog volcanoes. *Mathematical Geosciences*, *43*(5), 505–519. https://doi.org/10.1007/s11004-011-9345-6

Santacroce, R. (1983). A general model for the behavior of the Somma-Vesuvius volcanic complex. *Journal of Volcanology and Geothermal Research*, *17*(1–4), 237–248. https://doi.org/10.1016/0377-0273(83)90070-7

Sheldrake, T. E. (2014). Long-term forecasting of eruption hazards: A hierarchical approach to merge analogous eruptive histories. *Journal of Volcanology and Geothermal Research*, *286*, 15–23. https://doi.org/10.1016/j.jvolgeores.2014.08.021

Sheldrake, T. E., Sparks, R. S. J., Cashman, K. V., Wadge, G., & Aspinall, W. P. (2016). Similarities and differences in the historical records of lava dome-building volcanoes: Implications for understanding magmatic processes and eruption forecasting. *Earth-Science Reviews*, *160*, 240–263. https://doi.org/10.1016/j.earscirev.2016.07.013

Sides, I. R., Edmonds, M., Maclennan, J., Swanson, D. A., & Houghton, B. F. (2014). Eruption style at Kilauea Volcano in Hawai'i linked to primary melt composition. *Nature Geoscience*, *7*(6), 464–469. https://doi.org/10.1038/ngeo2140

Simkin, T., Siebert, L., & Blong, R. J. (2001). Disasters. volcano fatalities - Lessons from the historical record. *Science*, *291*(5502), 255. https://doi.org/10.1126/science.291.5502.255

Sparks, R. S. J. (2003). Forecasting volcanic eruptions. *Earth and Planetary Science Letters*, *210*(1–2), 1–15. https://doi.org/10.1016/s0012-821x(03)00124-9

Thompson, M. A., Lindsay, J. M., Sandri, L., Biass, S., Bonadonna, C., Jolly, G., & Marzocchi, W. (2015). Exploring the influence of vent location and eruption style on tephra fall hazard from the Okataina Volcanic Centre, New Zealand. *Bulletin of Volcanology*, *77*(5), 38. https://doi.org/10.1007/s00445-015-0926-y

Thordarson, T., & Larsen, G. (2007). Volcanism in Iceland in historical time: Volcano types, eruption styles and eruptive history. *Journal of Geodynamics*, *43*(1), 118–152. https://doi.org/10.1016/j.jog.2006.09.005

Tierz, P. (2020). Long-term probabilistic volcanic hazard assessment using open and non-open data: Observations and current issues. *Frontiers in Earth Science*, *8*, 257. https://doi.org/10.3389/feart.2020.00257

Tierz, P., Clarke, B., Calder, E. S., Dessalegn, F., Lewi, E., Yirgu, G., et al. (2020). Event trees and epistemic uncertainty in long-term hazard assessment of rift volcanoes: The example of Aluto (Central Ethiopia). *Geochemistry, Geophysics, Geosystems*, *21*(10), e2020GC009219. https://doi.org/10.1029/2020gc009219

Tierz, P., Loughlin, S. C., & Calder, E. S. (2019). VOLCANS: An objective, structured, reproducible method for identifying sets of analogue volcanoes. *Bulletin of Volcanology*, *81*, 12. https://doi.org/10.1007/s00445-019-1336-3

Voloschina, M., Bebbington, M. S., Lube, G., & Procter, J. (2021). Probabilistic modelling of multi-phase eruptions found in geological records: An example from Mt Ruapehu, New Zealand. *Journal of Volcanology and Geothermal Research*, *416*, 107273. https://doi.org/10.1016/j.jvolgeores.2021.107273

Wang, T., & Bebbington, M. S. (2012). Estimating the likelihood of an eruption from a volcano with missing onsets in its record. *Journal of Volcanology and Geothermal Research*, *243–244*, 14–23. https://doi.org/10.1016/j.jvolgeores.2012.06.032

Weir, A. M., Mead, S., Bebbington, M. S., Wilson, T. M., Beaven, S., Gordon, T., & Campbell-Smart, C. (2022). A modular framework for the development of multi-hazard, multi-phase volcanic eruption scenarios. *Journal of Volcanology and Geothermal Research*, *427*, 107557. https://doi.org/10.1016/j.jvolgeores.2022.107557

Whelley, P., Newhall, C. G., & Bradley, K. (2015). The frequency of explosive volcanic eruptions in Southeast Asia. *Bulletin of Volcanology*, *77*, 1–11. https://doi.org/10.1007/s00445-014-0893-8

Whitehead, M., & Bebbington, M. S. (2021). A method selection tool for short-term volcano forecasting. *Journal of Volcanology and Geothermal Research*, *419*, 107386. https://doi.org/10.1016/j.jvolgeores.2021.107386

Wright, H. M. N., Pallister, J. S., McCausland, W. A., Griswold, J. P., Andreastuti, S., Budianto, A., et al. (2019). Construction of probabilistic event trees for eruption forecasting at Sinabung volcano, Indonesia 2013-14. *Journal of Volcanology and Geothermal Research*, *382*, 233–252. https://doi.org/10.1016/j.jvolgeores.2018.02.003