Understanding causality and uncertainty in volcanic observations: An example of forecasting eruptive activity on Soufrière Hills Volcano, Montserrat

T.E. Sheldrake [a,b,*], W.P. Aspinall [a], H.M. Odbert [a], G. Wadge [c], R.S.J. Sparks [a]

[a] School of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK
[b] Department of Earth Sciences, University of Geneva, rue des Maraîchers 13, Geneva CH-1205, Switzerland
[c] Department of Meteorology, University of Reading, Reading RG6 6AL, UK

ARTICLE INFO

ABSTRACT

Following a cessation in eruptive activity it is important to understand how a volcano will behave in the future and when it may next erupt. Such an assessment can be based on the volcano's long-term pattern of behaviour and insights into its current state via monitoring observations. We present a Bayesian network that integrates these two strands of evidence to forecast future eruptive scenarios using expert elicitation. The Bayesian approach provides a framework to quantify the magmatic causes in terms of volcanic effects (i.e., eruption and unrest). In October 2013, an expert elicitation was performed to populate a Bayesian network designed to help forecast future eruptive (in-)activity at Soufrière Hills Volcano. The Bayesian network was devised to assess the state of the shallow magmatic system, as a means to forecast the future eruptive activity in the context of the long-term behaviour at similar dome-building volcanoes. The findings highlight coherence amongst experts when interpreting the current behaviour of the volcano, but reveal considerable ambiguity when relating this to longer patterns of volcanism at dome-building volcanoes, as a class. By asking questions in terms of magmatic causes, the Bayesian approach highlights the importance of using short-term unrest indicators from monitoring data as evidence in long-term forecasts at volcanoes. Furthermore, it highlights potential biases in the judgements of volcanologists and identifies sources of uncertainty in terms of magmatic causes rather than scenario-based outcomes.

1. Introduction

Important decisions can rest on forecasting when a volcano will next erupt: in the short-term, whether to evacuate people at risk (e.g., Surono et al., 2012) and in the long-term, for land-use planning and to improve resilience to eruptive activity (e.g., Marzocchi et al., 2004). Two different approaches can be taken: methods for short-term forecasting usually focus on interpreting patterns of geophysical monitoring observations (e.g., Sparks, 2003; Sobradelo and Martí, 2015); and, for long-term analysis, inferences may be based on a statistical characterisation of the eruptive history of a volcano (e.g., Bebbington, 2008, 2013) or the absence of short-term unrest indicators (e.g., Wadge and Aspinall, 2014). Importantly, short-term forecasts are usually only utilised once volcanic unrest is observed, and where monitoring data

are assumed to be anomalous and indicative of precursory eruptive behaviour (Marzocchi and Bebbington, 2012).

Generally, volcanic forecasting is traditionally based on statistical inference, that is, analysing past observations to calculate the probability of future events, using either or both empirical data and expert opinion. Statistical inference can be particularly tricky to implement, however, when data are sparse or the forecasting timescales are long. Nevertheless, there are instances when understanding volcanic unrest is important for making long-term forecasts: such as during long periods of dormancy when occasional episodes of unrest are observed that are not contemporaneous with eruptive activity (e.g., Santorini 2011–12; Stiros et al., 2010; Aspinall and Woo, 2014); or when there are very low or no obvious signs of unrest but considerable uncertainty as to whether an eruption is finished, and it is desirable to assess the likelihood that the volcano has entered a state of repose that is characteristic of a long-duration cessation in eruptive activity (e.g., Soufrière Hills Volcano; Wadge and Aspinall, 2014).

As the understanding of volcanic behaviour has evolved over recent decades, it is questions involving reverse causality that have motivated scientific research and provided the main basis for development of

* Corresponding author at: Department of Earth Sciences, University of Geneva, rue des Maraîchers 13, Geneva CH-1205, Switzerland.
E-mail address: thomas.sheldrake@unige.ch (T.E. Sheldrake).

conceptual models for magmatic systems (Cashman and Sparks, 2013; Sparks and Ashman, 2016). For example, why does a volcano deform during phases of lava extrusion (Odbert et al., 2015) or why do mafic enclaves occur in andesitic lavas (Barclay et al., 2010)? Thus it makes sense to design forecasting models using a similar rationale. Recently, there has been growing interest in the application of Bayesian Networks (shortened to Bayes Net, or also known as Belief Nets or Bayesian Belief Networks) to incorporate such causal inference into forecasting models (Aspinall et al., 2003; Aspinall and Woo, 2014; Hincks et al., 2014), as an alternative to more traditional approaches based on statistical inference.

To investigate the applicability of a causal inference approach, we combine empirical evidence about long-term patterns of activity with short-term geophysical and geochemical observations at Soufrière Hills Volcano (SHV). SHV has been in a continuous state of unrest since 1992, characterised by intermittent phases of eruptive activity (Sparks and Young, 2002; Wadge et al., 2014), although since March 2010 there has been a pause in eruptive activity at the surface and an increasing interest in projecting when it might next erupt.[1] SHV provides an excellent example of where statistical inference is limited when forecasting the long-term behaviour. Causal inference provides an opportunity to combine monitoring and phenomenological observations with magmatic and volcanic theory to aid long-term forecasts. However, the application of causal inference can be more widespread than simply this one application or just long-term forecasting, and so first we provide a summary of the main differences between statistical and causal inference within volcanology.

2. Causal-statistical dichotomy in volcanology

Forecasts of eruptive activity are commonly based on the estimation of relationships between observations and events using statistical inference. For example, this could be the correlation between observations of unrest such as deformation, and volcanic eruptions (Phillipson et al., 2013). This is because it is often easier to study the 'effects' of magmatic processes, such as unrest and eruptive activity, and quantify the relationship between these effects to estimate the likelihood of future events, rather than identify the magmatic 'causes' of particular scenarios. Within volcanology, this framework of statistical inference is commonly implemented in eruption forecasting using event trees (e.g., BET, Marzocchi et al., 2008; HASSET, Sobradelo et al., 2013), where the logic is entirely forward as the tree represents the progression through a sequence of conditional relationships (Fig. 1a–b).

Causal inference is based on an analysis of potential future events under changing conditions, which requires a level of judgement (Pearl, 2009). Causal inference can be graphically displayed using a Bayesian network, which consists of arcs (arrows) identifying the direction of causality between nodes, which represent 'effects', 'causes' and 'outcomes' (Fig. 1c–d). Nodes represent a state space (e.g., types of eruptive activity) with, in principle, no limit on the possible number of states a node can take. The states of a node can be modelled using continuous probability distributions between zero and one, and thus represent discrete random variables. Such a network design enables magmatic 'causes' to be quantified explicitly (Fig. 1c) or implicitly (Fig. 1d), based entirely on the observation of 'effects'. The application of Bayes Rule to such relationships is the mathematical concept that allows inferences about 'causes' to be made rationally on the basis of evidence from 'effects', where C is the 'cause' and E is the 'effect':

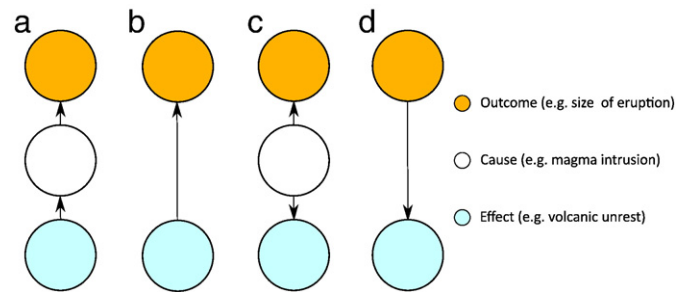$$\Pr(C|E) = \Pr(C)\frac{\Pr(E|C)}{\Pr(E)} \tag{1}$$

Fig. 1. Examples of the direction of reasoning in terms of 'causes', 'effects' and 'outcomes' in: (a + b) event trees where 'causes' are implicit in the 'effects' and so reasoning is from 'effects' to 'outcomes', either (a) indirectly or (b) directly; (c + d) a Bayesian network with reasoning from 'causes' to 'effects', where (c) magmatic processes are explicitly identified, or (d) where the effects implicitly represent magmatic processes. Arrows represent the flow of information and nodes represent 'observable', 'latent' and 'outcome' states.

Whilst Bayes theorem is used in event trees to quantify 'effects' using a joint distribution of observed variables (i.e. at a specific node), the structure of an event tree itself is not Bayesian. Consequently, an event tree is not a Bayesian network, although it can be mapped into a Bayesian network (e.g., Aspinall et al., 2003). In both event trees and Bayesian networks, the state-set of a discrete node must be both mutually exclusive and exhaustive (e.g., erupt vs not erupt) and arcs cannot contain complete circular pathways. However, in a Bayesian network, monitoring data are modelled as 'effects' (Fig. 1c–d), whereas in event trees they indirectly represent a 'cause' (e.g., inferred magmatic unrest) (Fig. 1b) or used to quantify the probability of a casual process, based on a conditional relationship (Fig. 1a).

In a volcanological context, Bayesian networks use physical, chemical and other observational evidence to understand the dynamics and relationships between different magmatic and volcanic processes with the aim of ultimately forecasting eruptive behaviour probabilistically (Aspinall and Woo, 2014; Hincks et al., 2014). This is achieved by designing a model to capture, as comprehensively as feasible, how magmatic processes modulate observational data (e.g., monitoring data, volcanic eruptions). *Observable* and *outcome* nodes can be linked via *latent* nodes, which represent unobservable or inaccessible magmatic processes (Fig. 1c), therefore allowing Bayesian networks to be designed based on a physical model for the magmatic system (Hincks et al., 2014). This structuring represents reverse-causal reasoning, as information acquired about 'effects' (e.g., eruptive activity or volcanic unrest) is used to constrain the likely 'cause' (i.e., the driving magmatic process). Crucially, however, the conditional relationships between the nodes can be based on forward causality (Gelman and Imbens, 2013), that is, the probability of observing 'effects' (e.g., seismicity) given specific 'causes' (e.g., magma intrusion or no magma intrusion).

Traditionally, the utilisation of Bayesian networks in disciplines such as the medical sciences is performed to quantify conditional relationships when plenty of data are available, and where 'causes' can be empirically linked to 'effects' (e.g., via post-mortems). However, in complex natural systems such as volcanoes this is not the case, as 'causes' (i.e., magmatic processes) for past events cannot be unequivocally diagnosed and are only inferred. This has implications for how a Bayesian network should be designed and how conditional probabilities are quantified: one approach is to populate a network using uncertainty judgements elicited from the understanding of experts using formalised techniques (Aspinall and Woo, 2014; Hincks et al., 2014).

When a Bayesian network does not implicate a physical model it does not require *latent* nodes, and so the physical causes become implicit in the *outcomes*. Thus, arcs point directly from *outcome* nodes (e.g., eruption scenario) to *observation* nodes (e.g., seismicity) in a logic tree structure (Fig. 1d) (e.g., Aspinall and Woo, 2014). This means that conditional probabilities between nodes can be quantified directly by a combination of observational data (Aspinall et al., 2003) and expert judgement (Aspinall and Woo, 2014).

3. Application: Soufrière Hills Volcano

3.1. Geological background

Soufrière Hills Volcano (SHV) is an andesitic dome-building volcano on the island of Montserrat in the Caribbean (Fig. 2). In 1995 it erupted following three years of seismic unrest and several centuries of dormancy, which included ~30-year cycles of elevated seismicity (Young et al., 1998; Sparks and Young, 2002; Odbert et al., 2014). The eruption of SHV has involved five distinct phases of lava extrusion over the subsequent fifteen years, monitored by the Montserrat Volcano Observatory (MVO) (Wadge et al., 2010, 2014). The phases of activity were characterised by the repeated formation of Peléean andesitic lava domes that grew and collapsed and, intermittently, vulcanian explosions (Sparks and Young, 2002; Wadge et al., 2014) until February 2010, when, after an explosion and flank collapse, the latest large dome stopped growing and thereafter only experienced minor mass-wasting (Wadge et al., 2014). Between these phases of activity there were pauses when no significant lava extrusion occurred, and eruptive activity was limited to isolated ash venting or small explosive events (Norton et al., 2002; Cole et al., 2014).

Each time eruptive activity at SHV paused, and the volcano entered a state of repose, there was a focus on forecasting if, when, and how eruptive activity would restart. As a corollary, there was also an increasing impetus to develop criteria to identify when the eruption had ceased, which became synonymous with identifying a long-term pause in eruptive activity (Wadge and Aspinall, 2014). Since 1997, a Scientific Advisory Committee (SAC) has met in conjunction with MVO, annually to biannually (including during pauses of eruptive activity), to discuss the on-going state of the volcano and to forecast how eruptive activity may evolve in the future. Within the role of the SAC, identifying when the eruption had ceased was based on meeting simultaneously a set of on-going criteria defined in terms of monitoring metrics relating to three key observables (Fig. 3), all three to be sustained jointly over a year, or longer (Wadge and Aspinall, 2014): (1) absence of low frequency seismicity associated with active magmatic processes; (2) average
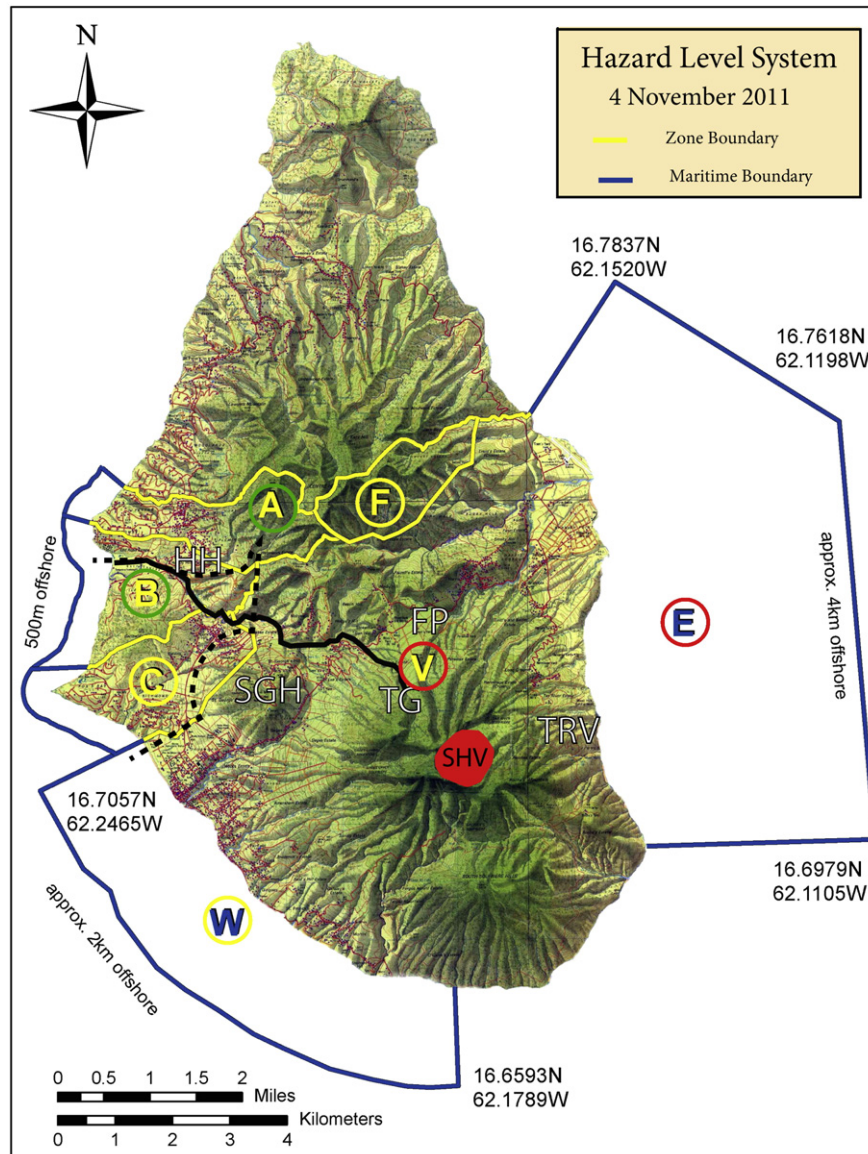

Fig. 2. Modified map of Hazard Level System on Montserrat from Cole et al. (2011) used at the time of this study (i.e. 2013). Soufrière Hills Volcano (SHV) is located in the south of the island, with hazard zones (indicated by yellow and blue outlined areas) encompassing over half the island area. At the time of this study the Hazard Alert System was at Level Two, which meant that areas with letters in red circles were open only to 'essential workers', yellow circles allowed 'daytime access', and green circles were 'unrestricted access'. See MVO website for further details and more up-to-data hazard map (http://www.mvo.ms).
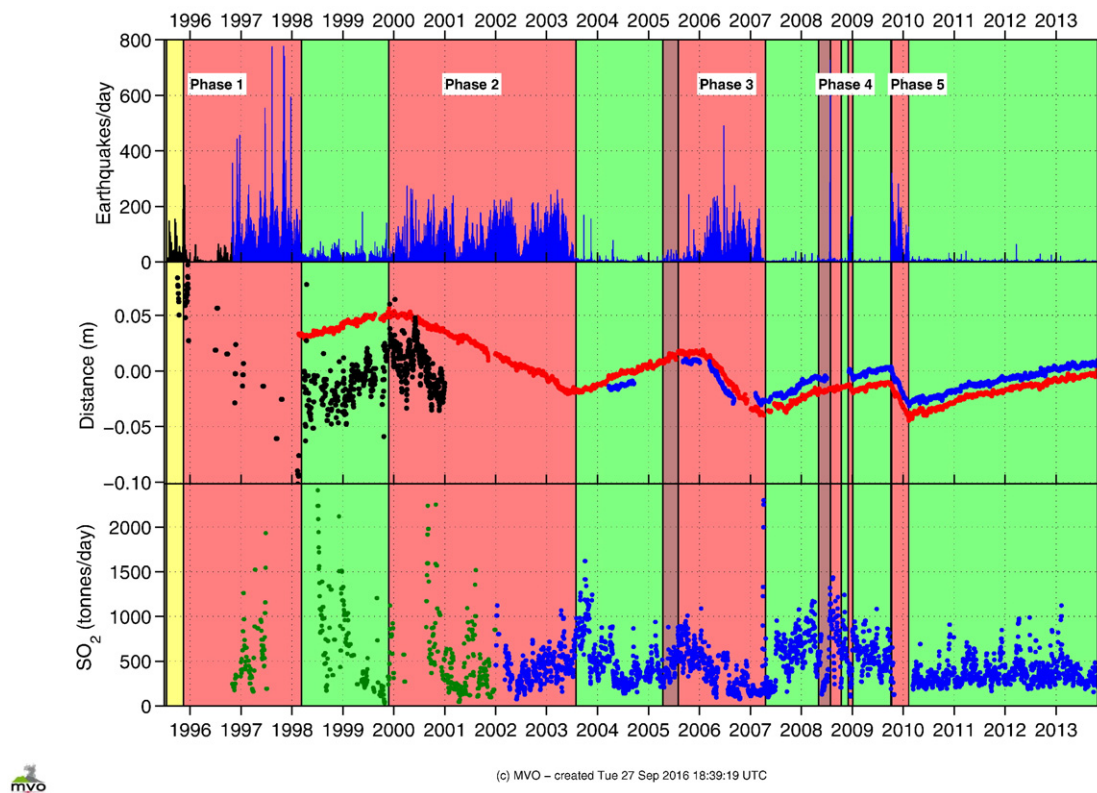
Fig. 3. Summary of the three main monitoring observables as used by Montserrat Volcano Observatory, and provided to the Scientific Advisory Committee as a Report for the SAC October 2013 meeting (Stewart et al., 2013). Pink bars are phases of dome-extrusion and green bars are pauses in eruptive activity. The top plot panel shows the pattern of total seismicity, with most seismicity occurring during periods of dome-growth. The middle panel represents the radial deformation of the cGPS station MVO1, 7.6 km from SHV. The bottom panel is a time series of SO_2 fluxes, with COSPEC measurements in green and DOAS measurements in blue.

SO_2 fluxes below the measurement detection level (50 tonnes per day; Edmonds et al., 2003) and (3) absence of discernible surface deformation of the volcanic edifice from continuous GPS data, that could be interpreted as reflecting deep magmatic pressurisation.

As time has passed since the latest active phase ended (February 2010), the question of when and whether eruptive activity will resume became more pressing, with public perception of volcanic hazard diminishing and demands to develop economic enterprises increasing. The current period of quiescence at SHV at the time of this study (October 2013) was two times longer than any of the previous pauses since 1995, and as it continues (now almost four times longer; September 2016) the pressure to identify evidence on which to base a forecast of future eruptive activity increases.

3.2. Formulating a model

We utilise aspects of Bayesian Networks and Bayes theorem to forecast long-term eruptive activity at SHV, with the aim of quantifying causality between monitoring observations, magmatic processes and eruptive activity. Two different models were investigated in separate workshops: one including students and researchers at the University of Bristol (February 2013) and one on Montserrat that included observatory staff and senior academic researchers (October 2013). This paper reports the findings of the second workshop.

We forecast eruptive activity being renewed in four different scenarios: (1) less than one year (reflecting the traditional forecasting timeframe of the SAC); (2) one to five years (timescales that are similar to previous SHV eruptive cycles); (3) five to thirty years (a timescale relevant to civic planning); (4) and greater than thirty years. Each one of these intervals is mutually exclusive and their respective probabilities add up to unity as it is assumed that SHV will erupt at some point in the future.

Forecasting the renewal of eruptive activity in specific time intervals allows us to utilise the odds formulation of Bayes theorem (Eq. 2). The left-hand side of this equation gives the posterior odds for eruption in a given time interval (i.e., the eruption forecast); the first term on the right-hand side represents the prior probability of eruptive activity being renewed, before monitoring data are incorporated as evidence; and the second term is the likelihood ratio, which quantifies the evidence for the volcano erupting versus not erupting (¬signifies negative evidence), given the monitoring data that are available.

$$odds(erupt|obs) \sim odds(erupt)\frac{Pr(obs|erupt)}{Pr(obs|\neg erupt)} \qquad (2)$$

The application of Bayesian networks to forecast over the timescales considered in this study (i.e., years to decades) presents a series of challenges. Firstly, the inclusion of an unknown state or unobservable process via a *latent* node (e.g., Hincks et al., 2014) is based on conceptual models and equivalent magmatic processes that are generally poorly constrained over timescales of years to decades. Furthermore, empirically linking *observation* nodes directly to *outcome* nodes (e.g., Aspinall et al., 2003) over these longer timescales is problematic due to the relatively short duration that monitoring data covers and, therefore, absence of a statistically significant empirical evidence base.

The Bayesian network is formulated so that estimates of the prior probability for eruptive activity incorporate various strands of empirical evidence in a systematic way (Fig. 4a; Section 3.3). Specifically, the Bayesian network is designed to account for the dynamic nature of the volcano: how it has behaved in the past and the implications of this for how it might behave in the future. These observations are updated, using a likelihood function, based on a synthesis of current monitoring observations (Fig. 4b; Section 3.4), to calculate the posterior probability of eruptive activity. This conceptual approach provides a structured
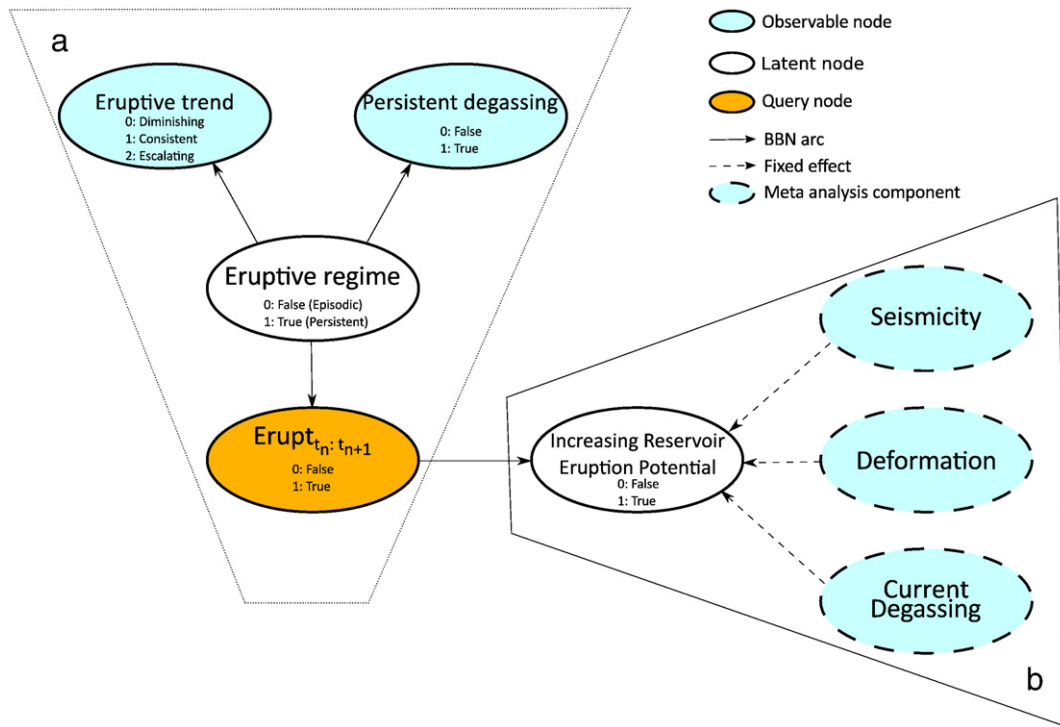
Fig. 4. A Bayesian network for the synthesis of multiple strands of evidence (blue ellipses) at SHV: (a) represents the calculation of the prior probability for whether SHV will erupt based on empirical evidence for long-term magmatic processes; and (b) represents the likelihood model incorporating a fixed-effects meta-analysis, which is used to update the prior probabilities calculated in (a), based on current levels of volcanic unrest. Each node represents a state and the arrows represent reasoning in the network.

framework for interpreting observations of volcanic unrest, reflecting evolving conditions in the upper crustal environment, and relating them to the longer pattern of eruptive activity.

Data is incorporated into the model using expert judgements via an elicitation procedure (Section 3.5). Once actual evidence is entered into the Bayesian network via *observable* nodes (a process known as instantiation), the prior probabilities for the *query* node are converted to their posterior odds form (Eq. 3).

$$odds(erupt) \sim \frac{Pr(erupt)}{Pr(1-erupt)} \tag{3}$$

3.3. Prior probability for eruptive activity

There is increasing evidence that over years to decades the broad spectrum of volcanic activity that is observed at dome-building volcanoes cannot be fully described by a conceptual model focussed only on the transfer of magma to upper crustal regions, and instead must also account for the dynamic destabilisation of multiple regions of magma and magmatic fluids throughout the crust (Christopher et al., 2015; Sparks and Ashman, 2016). It is non-trivial, however, to constrain the extent to which any single magmatic process contributes to monitoring observations made at the surface. Consequently, we characterise the long-term behaviour of SHV as either *persistent* or *episodic* (Sheldrake et al., 2016).

Persistent and episodic regimes are characterised by different patterns of eruptive activity (Sheldrake et al., 2016). In a persistent regime, the duration a volcano spends in a state of eruption is broadly similar to the time it spends in a state of repose. By contrast, in an episodic regime, the duration a volcano spends in a state of eruption is orders of magnitude shorter than the time it spends in a state of repose. Importantly, absolute timescales for persistent and episodic volcanism vary between different volcanoes as a result of variations in the physical, thermal and chemical structure of magmatic systems. At SHV, a persistent

regime has existed since the onset of the eruption in 1995, although over longer century timescales eruptive activity has been episodic. Consequently, any prior probability of when SHV may erupt in the future can be based on whether SHV is judged to have remained in a persistent regime, or has entered a state of repose more characteristic of an episodic regime, as was the case in the centuries prior to 1995.

The judgement of whether SHV remains in a persistent regime, or has entered a state of repose more characteristic of an episodic regime is guided by phenomenological (long-term measures of eruption rate and magmatic degassing) observations from both SHV and other dome-building volcanoes (Sheldrake et al., 2016). Long-term measures of eruption rate (i.e. lava extrusion rate, tephra emission rate) document changes in the explosivity, duration, frequency and/or style of eruptive activity. The eruption rate may diminish, remain consistent, or even escalate, depending upon whether SHV remains in a persistent or episodic regime. Generally, however, over timescales for episodic volcanism, which vary between volcanoes, the eruption rate diminishes during the course of an eruption. The pattern of magmatic degassing varies markedly between a persistent and episodic regime. At volcanoes that remain in a persistent regime degassing is sustained between phases of eruptive activity, and can be decoupled fully from the extrusion of magma. In contrast, over timescales for episodic volcanism, a state of repose is characterised by negligible degassing, albeit fluxes can remain high for several years following an eruption.

Whether SHV remains in a persistent regime is indicative (although not explicitly) of the underlying long-term magmatic processes, and so acts like a *latent* node in the Bayesian network. In this sense, we can interpret *Eruptive regime* as a causal state node. This network formulation allows information about 'causes' (i.e., the eruptive regime) to be quantified using observations of current 'effects' (the pattern of degassing and eruption rate) to quantify future 'effects' (i.e., *outcomes*). This section of the network has a diverging structure, with arcs from the *latent* node to the *observable* nodes and the *query* node (Fig. 4a) that are quantified using a conditional probability table (CPT).

3.3.1. Latent nodes

3.3.1.1. Eruptive regime. (ER) is a Boolean node with two states: (i) *true,* if SHV remains in a persistent regime; or (ii) *false,* if it has entered a state of repose characteristic of an episodic regime.

3.3.2. Observable nodes

3.3.2.1. Persistent degassing. (PD) is a Boolean node with two states: (i) *true*, if the degassing behaviour of SHV since it last erupted in February 2010 is characteristic of a persistent regime; or (ii) false, if the degassing behaviour of SHV since it last erupted in February 2010 is characteristic of a volcano entering a state of repose characteristic of an episodic regime. This node requires probabilities that the current degassing behaviour would be observed, given that SHV either remains in a persistent regime or enters a state of repose characteristic of an episodic regime. Importantly, these two questions are not mutually exclusive when enumerating their probabilities.

3.3.2.2. Eruptive trend. (ET) has three states: (i) *escalating*; (ii) *consistent*; or (iii) *diminishing*, which characterise the current long-term eruption rate at SHV. The node requires the probabilities (which may be different) for SHV remaining in a persistent regime, given each of these three trends occurring. Furthermore, the Bayesian network requires the probability that the current behaviour of SHV is characterised by each of the three trends. Relevant evidence is not explicitly identified but includes direct observations of previous eruptive activity, such as the pattern in the frequency or intensity of explosive activity, or changes in long-term measures of eruption rate, such as lava extrusion.

3.3.3. Query node

3.3.3.1. Erupt. (Er) is a Boolean node with two outcomes: (i) *yes* or (ii) *no*, depending upon whether, or not, SHV is forecast to erupt in the given time interval of concern.

3.4. Likelihood of eruptive activity

The eruption at SHV has generated a rich, multi-parameter dataset, recorded and updated regularly at the MVO, which has enabled detailed investigation of many volcanic processes. Here, we consider how eruptive processes are manifest in three core monitoring datasets: seismicity; degassing; and deformation (Fig. 3; Supplementary material). Whilst other physical parameters are routinely recorded at MVO (e.g., lava effusion rate, surface rockfall activity, thermal imagery, ground strain, etc), these three core datasets have the longest, most consistent record (Wadge et al., 2014).

The monitoring observations are used to characterise the current state of the upper crustal magmatic system, and ultimately calculate the likelihood of each of the eruptive scenarios. Given that the duration over which current monitoring data exist at SHV (~20 years) is similar to the timescales of the processes we are seeking to characterise, direct reasoning from current observable 'effects' of magmatic processes (e.g., monitoring data) to future 'effects' (i.e., eruptive outcomes) cannot be performed empirically, as there is little evidence to determine a background level of unrest.

There is considerable evidence at many dome-building volcanoes that magma storage zones are found in upper crustal regions (Sheldrake et al., 2016, and references therein). Indeed, at SHV there is evidence for multiple zones (Elsworth et al., 2008; Mattioli et al., 2010; Hautmann et al., 2014) with pressure sources extending from the shallow crust (~5 km depth) to the middle crust (10 to 15 km depth), and at least four geochemically distinctive magma compositions (Christopher et al., 2014). Numerical modelling of strain and deformation signals associated with eruptive activity indicates that magmatic processes over an extended depth range are influential in controlling

eruptive activity at SHV (Elsworth et al., 2008; Gottsmann and Odbert, 2014; Hautmann et al., 2014). This said, it is difficult to characterise these different magmatic sources and related processes in any detail, let alone distinguish between their individual contributions to eruptive activity over the timescales we are interested in. Ideas on how magmatic systems work and how they control volcanism are changing (Sparks and Ashman, 2016), so developing a Bayesian Network model which is strongly linked to particular conceptual models is problematic. Consequently, we refrain from explicitly incorporating these processes into our model. Instead, we characterise the quasi-static state of the shallow volcanic system using a *latent* node, called *Increasing Reservoir Eruption Potential* (*I-REP*; Fig. 4b). This term represents all processes in the magmatic system that would increase the likelihood for SHV to erupt (e.g., second boiling, magma injection). This allows experts to interpret evidence collectively, from both SHV and other dome-building volcanoes, without being constrained to identify individual magmatic processes that are specific to particular conceptual or numerical models.

3.4.1. Reservoir eruption potential

The rationale for defining the *latent* node *I-REP* is that it provides a basis for combining multiple monitoring observables into a single state descriptor for the shallow magmatic system. This means each monitoring observable can be interpreted without having to define a priori discrete states (e.g., a background level threshold) or relevant timescales.

When the state of *I-REP* is *true* the likelihood ratio in Eq. 2 has two components:

$$\Pr\left(I\text{--}REP_{TRUE}|erupt_{t_n,t_{n+1}}\right) \tag{4}$$

which represents the probability that *I-REP* is *true*, **given** that SHV will **erupt** in the time interval $t_n;t_{n+1}$, and;

$$\Pr\left(I\text{--}REP_{TRUE}|\neg erupt_{t_n,t_{n+1}}\right) \tag{5}$$

which represents the probability that *I-REP* is *true*, **given** that SHV will **not erupt** in the time interval $t_n;t_{n+1}$.

Alternatively, when the state of *I-REP* is *false*, the likelihood ratio in Eq. 2 has the two following components;

$$\Pr\left(I\text{--}REP_{FALSE}|erupt_{t_n,t_{n+1}}\right) \tag{6}$$

which represents the probability that *I-REP* is *false*, **given** that SHV **will erupt** in the time interval $t_n;t_{n+1}$, and;

$$\Pr\left(I\text{--}REP_{FALSE}|\neg erupt_{t_n,t_{n+1}}\right) \tag{7}$$

which represents the probability that *I-REP* is *false*, **given** that SHV **will not erupt** in the time interval $t_n;t_{n+1}$. In other words, we quantify the probability that an eruption would and would not be associated with upper crustal processes, which in turn constrains our diagnostic model.

We assume that there is only one correct, but unknown, value for the probability of *I-REP* being *true*, which is estimated using fixed-effects meta-analysis. This approach involves calculating the weighted mean for the variable *I-REP*, based on the strength of the evidence (i.e., uncertainty) when interpreting each monitoring observable. As each piece of evidence is given as a probability between 0 and 1, it can be represented using a Beta distribution,

$$\theta_i \sim Beta(\alpha_i, \beta_i) \tag{8}$$

where *i* represents each monitoring variable. Each piece of evidence is weighted according to the ratio of the sum of the parameters of its

Beta distribution and the sum of the parameters of all n Beta distributions,

$$\pi_i = \frac{a_i + \beta_i}{\sum_{i=1}^{n}(a_i + \beta_i)} \tag{9}$$

in which the sum of the parameters for a Beta distribution represents the variance or uncertainty in that observation. As the sum of the parameters increases the variance decreases and hence the uncertainty in interpretation of the monitoring observable decreases. A natural distribution for the weighting parameter π is a Bernoulli distribution,

$$\pi_i \sim Bern(p_i) \tag{10}$$

where p is the resulting weighted probability that $\underline{I\text{-}REP}$ is *true*, given as:

$$\theta_{I-REP_{TRUE}} = (\pi_{GAS} \cdot \theta_{GAS}) + (\pi_{SEIS} \cdot \theta_{SEIS}) + (\pi_{DEF} \cdot \theta_{DEF}) \tag{11}$$

with the converse probability equivalent to the weighted probability for $\underline{I\text{-}REP}$ being false,

$$\theta_{I-REP_{FALSE}} = 1 - \theta_{I-REP_{TRUE}} \tag{12}$$

3.5. Expert elicitation

Expert elicitation is a technique where the judgements of scientists or individuals who have specialist knowledge in a particular subject can be pooled and expressed quantitatively. The approach can be useful in situations where limited quantitative data exist and causal parameters cannot be readily obtained empirically or from the results of numerical models. Consequently, expert elicitation has been previously applied to quantify the judgements of scientists when forecasting events in natural systems (Aspinall, 2010; Aspinall and Cooke, 2013).

Eliciting the judgement of experts allows the collective uncertainty in the group to be quantified using probability distributions for variables of interest. This approach is particularly suited to forecasting applications in volcanology where observational data may be sparse (e.g., few or no historical observations; Martí et al., 2008) or where forward-causal inference is made based upon unobservable (e.g., crustal) processes (Hincks et al., 2014).

We utilised an expert elicitation approach to quantify the judgements of ten experts, which included experienced volcanologists who were full-time staff at the observatory or members of the Scientific Advisory Committee. The weighted views of all experts for a number of target questions (Supplementary material) are combined to calculate the optimised Decision-Maker (DM) distribution, which represents the group consensus (Supplementary material). Given that expert elicitation has been used routinely on Montserrat since 1997 (Wadge and Aspinall, 2014) the method was already familiar to the experts who engaged in this study.

3.6. Model fitting

All probability distributions for the target questions are defined on the interval 0 to 1 to account for uncertainty in expert judgement. Hence, we parameterise them as Beta distributions (Eq. 8) based upon the reported median and 90% credible interval. This is performed using the `beta.parms.from.quantiles' function (Bélisle, 2012) in R statistical language (R Development Core Team, 2008). Each component probability is calculated independently and then combined using Bayes rule via a series of deterministic relationships (Fig. 4). The result is a posterior distribution for the probability of SHV erupting in the four time intervals.

The posterior distributions of the *outcome* node and *latent* nodes are estimated according to the probability model specified in Sections 3.2–3.3 using Markov Chain Monte Carlo (MCMC) methods implemented with the WinBUGS software (Lunn et al., 2000), using R statistical language and the R2WinBUGS package (Sturtz et al., 2005).

4. Results

Prior and posterior probabilities for whether SHV will next erupt in each of the four time intervals (*Erupt* node), are presented in Table 1. Prior distributions are conditioned upon state of the *Eruptive regime* and the long-term behaviour SHV, and updated to calculate the posterior distribution, based on evidence for shallow magmatic processes and the current state of the volcano (synthesised using the *latent* variable *I-REP*). Large uncertainties are observed in the forecasts for future eruptive activity, which are a consequence of extensive uncertainties observed in the elicitation results.

Instances of substantial uncertainty are not uncommon in elicitations regarding natural systems, where data and observations can be sparse, processes are stochastic and understanding or predictability is limited (Aspinall and Cooke, 2013; Hincks et al., 2014). Generally, the uncertainty bounds are smaller for performance-weighted solutions (i.e., using calibration weights) than for equal weights combinations (i.e., each expert has the same weight). Therefore, the levels of uncertainty indicated by the performance-weighted solutions are not reflective of either the applicability of our method or reliability of the results. Rather, we consider it reflects the true scientific uncertainty of the issues being explored.

4.1. Long-term behaviour of SHV

The Bayesian network utilises two strands of evidence (*Eruptive trend* and *Persistent degassing* nodes) to infer the long-term eruptive regime of SHV. Given this evidence, the DM considers it more than twice as likely that in October 2013 SHV remained in a persistent regime, rather than having entered a state of repose that is more characteristic of an episodic behaviour pattern (Fig. 5). This is informed mostly by the degassing behaviour of SHV during the period of quiescence since February 2010 (Fig. 5a), rather than the trend in eruptive activity (Fig. 5b), which provides nugatory evidence for inferring either a persistent or episodic regime.

Integral to the performance of the Bayesian network and calculation of a prior distribution is an understanding of the states that comprise the *latent* node *Eruption regime*. To do this we analyse the probability distributions for eruptive activity being renewed in each time interval ($t < 1$; $1 \leq t < 5$; $5 \leq t < 30$; $t \geq 30$ years), conditional on remaining in a persistent regime or entering a state of repose characteristic of an episodic regime (Fig. 6).

Conditional on SHV remaining in a persistent regime, across the group's judgements a similar pattern in probabilities for each state of the *query* node *Erupt* is observed (Fig. 6a), suggesting experts have a clear understanding of the term *persistent* and its implications for future eruptive activity. Within the group, median probabilities decrease as the duration of the forecast increases. A few experts within the group provided higher median probabilities for the second time interval ($1 < t \leq 5$ years), but this is reasonable given that the duration of this interval is four times longer than the first interval ($t < 1$ year). For the third and fourth time intervals, median probabilities across the group of experts are clustered.

Table 1
Prior and posterior mean probabilities for SHV next erupting in each of the four time intervals ($t < 1$; $1 \leq t < 5$; $5 \leq t < 30$; $t \geq 30$), where t is in years, from the Bayesian network. Values in brackets are standard deviations.

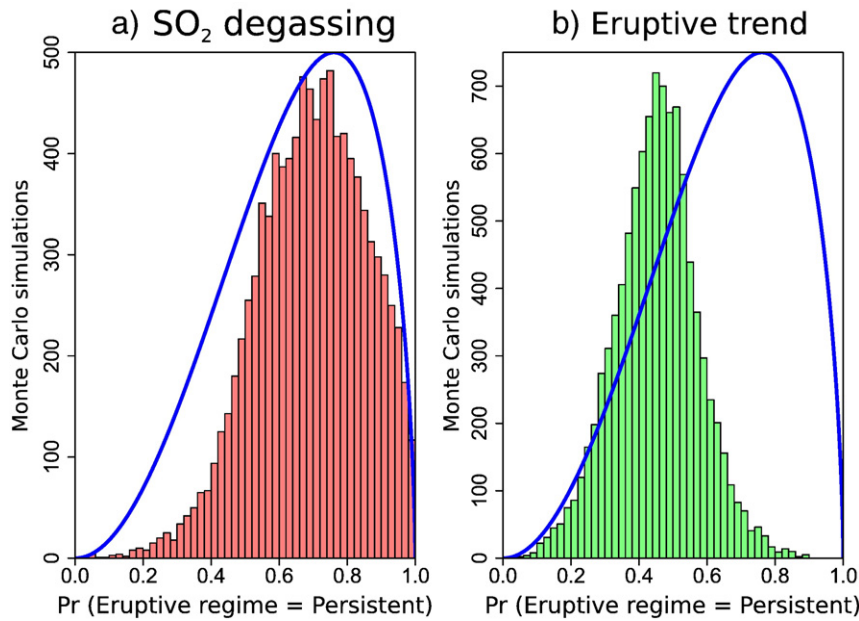|	Erupt($t < 1$)	Erupt($1 \leq t < 5$)	Erupt($5 \leq t < 30$)	Erupt($t \geq 30$)
Prior	0.38 (0.14)	0.38 (0.17)	0.15 (0.09)	0.09 (0.09)
Posterior	0.32 (0.13)	0.35 (0.17)	0.20 (0.12)	0.13 (0.11)

Fig. 5. Marginal posterior distributions for the probability that SHV remains in a persistent regime, conditional on: (a) interpretation of the *Persistent degassing* node (red bars), and (b) the interpretation of the *Eruptive trend* node (green bars). Blue lines represent the joint probability of SHV remaining in a persistent regime given evidence for both the *Eruptive trend* and *Persistent degassing* nodes.

In contrast, there is little consistency in the median probabilities for the *query* node *Erupt*, conditional on being in an *episodic* regime (Fig. 6b), indicating a lack of commonality between experts on what this term represents and entails in the context of event forecasting. This disparity in forecasts for the two regimes reflects the general interpretation of volcanoes as chaotic systems, characterised by intermittency (Odbert et al., 2014) where, in a persistent regime, eruptive activity can be periodic, and in an episodic regime eruptive activity is envisaged as being more random.

To further understand the disparity between responses for a persistent and for an episodic regime, we analyse how the experts interpret the pattern of degassing (*Persistent degassing* node) during the current pause in eruptive activity since February 2010 (Fig. 7). First, we consider question 34, asked as part of the elicitation (Supplementary material):
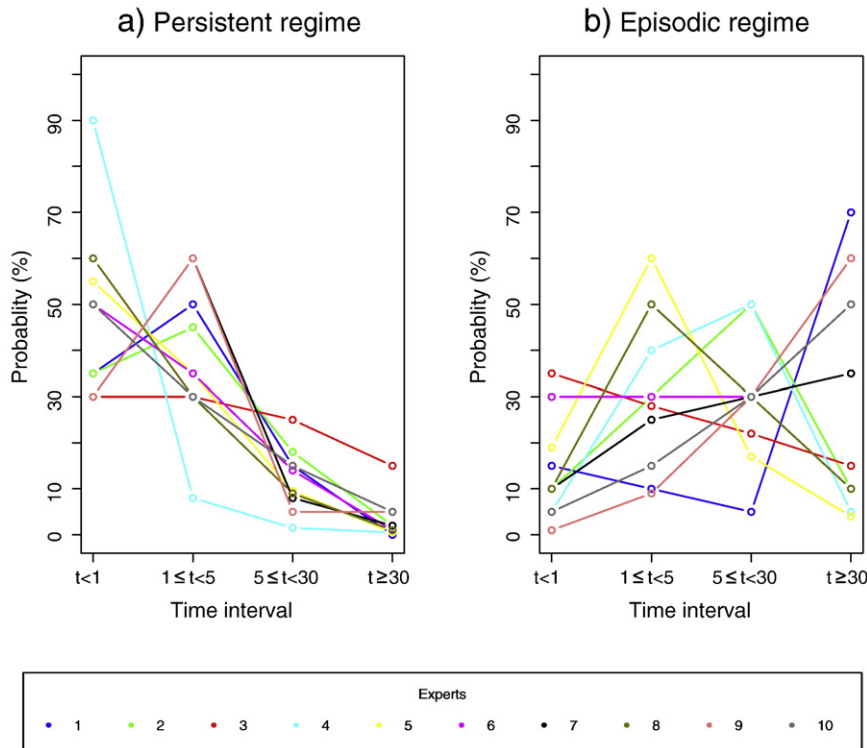

Fig. 6. Median probabilities for each of the experts in the elicitation for the probability of eruptive activity being renewed in the four time intervals ($t < 1$; $1 \leq t < 5$; $5 \leq t < 30$; $t \geq 30$), conditional on SHV: (a) currently remaining in a persistent regime, or (b) entering a state of repose characteristic of an episodic regime, in October 2013. (Questions 17–24: Supplementary material).
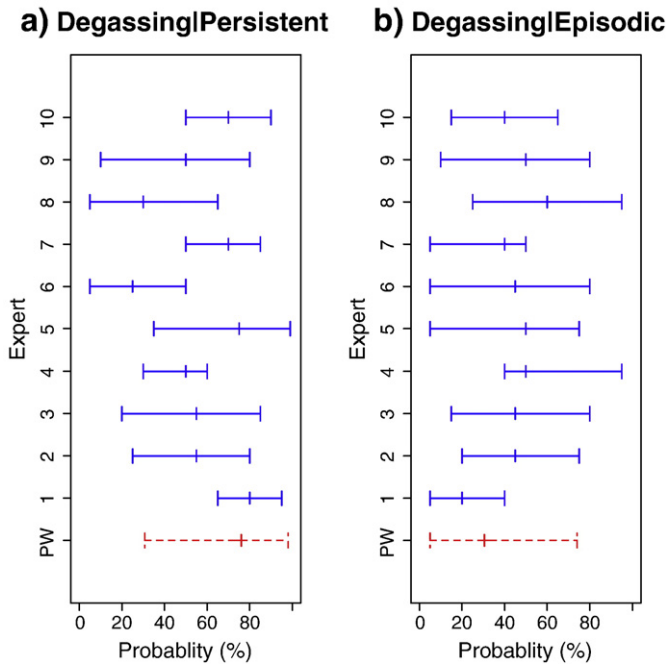
a) Degassing|Persistent

b) Degassing|Episodic

Fig. 7. Probability distribution quantiles (5th, 50th, 95th percentiles) for the 10 experts and the performance-weighted (PW) result for the conditional probabilities for observing the current pause in eruptive activity and accompanying degassing behaviour, given SHV: (a) remains in a persistent regime, or (b) has entered a state of repose characteristic of an episodic regime. (Questions 33–34: Supplementary material).

'Assuming SHV is in an episodic regime, what is the probability of observing the current period of quiescence (over 3 years at the time of the elicitation) accompanied by the observed degassing?'. All participants, except one expert, provide wide uncertainty distributions centred on 50% (Fig. 7b). Either the experts consider the pattern of degassing in an episodic regime as intrinsically ambiguous, or their understanding of an episodic regime is inchoate, and thus attempting to characterise degassing behaviour of a volcano in an episodic regime by elicitation is inherently disordered.

The opposite scenario is presented by question 33: 'Assuming SHV is in a persistent regime, what is the probability of observing the current period of quiescence (over 3 years) accompanied by the observed degassing?'. In comparison to question 34, experts' probability distributions here have smaller credible intervals and are not uniformly centred on 50% (Fig. 7a). Median values range from 25% to 80%, and thus are interpreted as reflecting distinct individual beliefs of the experts rather than a general lack of understanding in the term persistent. This is supported further when comparing the likelihood of being in a persistent regime (Eq. 13), with the equivalent forward causal statement in question 31: 'What is the probability that SHV is in a persistent regime given current levels of degassing at SHV?'. Both the likelihood ratio (LR) and forward causal statement provide similar values for being in a persistent regime, given current levels of degassing (Table 3).

$$LR = \frac{\Pr(Degassing|Persistent)}{\Pr(Degassing|Episodic)} \qquad (13)$$

4.1.1. Interpreting long-term trends

A number of questions regarding the observable nodes of the Bayesian network, including the _eruptive trend_ node, were re-elicited after the SAC meeting, following discussion of the initial results with the experts. This was a consequence of evident incoherence in the answers between experts that did not appear to be due solely to uncertainty. In an attempt to reconcile or remove discrepancies, and to understand the thinking of the experts, the re-elicitation included mutually exclusive questions

regarding both persistent and episodic regimes (e.g., questions 25 to 30; Supplementary material), rather than just considering questions regarding a persistent regime (e.g., questions 4 to 6; Supplementary material), as was performed originally during the SAC meeting.

For an _escalating_ or _consistent_ trend there is still considerable ambiguity in the results between experts, with re-elicited distributions moving in both positive and negative directions, and little coherence across the group (Fig. 8a–b). The results for a _diminishing_ trend convey more conviction, with coherence in the results as credible intervals reduce in size and median values either remaining similar or decreasing following re-elicitation (Fig. 8c).

Whilst ambiguity still remains for an _escalating_ and _consistent_ trend, coherence in the results for a _diminishing_ trend demonstrates the advantages of asking all relevant mutually exclusive questions in an elicitation, rather than presuming the complement of the result from one question can be reliably used to enumerate the probability value of its alternate.

4.2. Current state of the volcano

Coherence in the outcomes of questions 10 to 13 (Supplementary material), concerning future eruptive scenarios and the current state of _I-REP_ (Fig. 9) suggests there is: (1) a common understanding within the group of what the _latent_ parameter _I-REP_ represents; and (2) a common basis for how the monitoring data are interpreted. Expert's conditional probabilities for _I-REP_, given each of the four time intervals, show a similar pattern with the highest median probabilities associated with the shortest durations ($t < 1$ year) and the lowest median probabilities associated with the longest durations ($t \geq 30$ years). Consequently, even a slight deviation from 50% for the value of _I-REP_ influences the posterior timeframe-dependent distributions for SHV erupting. Specifically, the observation of (or lack of) volcanic unrest decreases the probability of SHV erupting in the two shorter time intervals, with probabilities increasing for eruptive activity in the two longer time intervals (Table 1).

The experts provided large uncertainties on their interpretation of the monitoring observables (Fig. 10), and consequently the posterior distribution for _I-REP_ does not deviate significantly from 50% (Fig. 11). The values of the fixed-effects weighting parameters in the meta-analysis were $\pi_{DEF} = 0.38$, $\pi_{SEIS} = 0.21$ and $\pi_{GAS} = 0.41$. To investigate the uncertainty in more detail we perform a paired comparison with probabilistic inversion (Supplementary Material) to quantify how the group ranks (from highest to lowest) the three monitoring observables, in terms of the probability that _I-REP_ is _true_. Given current observations, the deformation is considered to be more diagnostically important for a positive value of _I-REP_ than degassing or seismicity (Fig. 12). However, levels of agreement in the ranking of the observables amongst the group are modest, with a coefficient of concordance of 0.38 and coefficient of agreement of 0.23. Importantly, the null hypothesis – that the group as a whole provided random, unstructured responses on the ranking of observables – can be rejected with a p-value of 0.02.

Median probabilities for _I-REP_, given current observations of deformation, appear to cluster into two groups (Fig. 10): one group judged the state of _I-REP_ to be true (Experts 1, 6 and 9) whereas the other group were unsure (Experts 2,33,4,5,7 and 8). In the elicitation literature this is referred to as a 'two schools of thought' ambiguity (Hincks et al., 2014). In the discussion at the SAC meeting there was a general consensus that, based on the previous observations of deformation at SHV, inflation equated to the _true_ state of _I-REP_. However, some participants suggested the trend in far-field deformation could be 'flattening out' whereas others interpreted it to be still increasing. As a consequence, we suggest the trend in deformation was interpreted differently within the group of experts, which resulted in the apparent dichotomous clustering in median probabilities observed in Fig. 10.
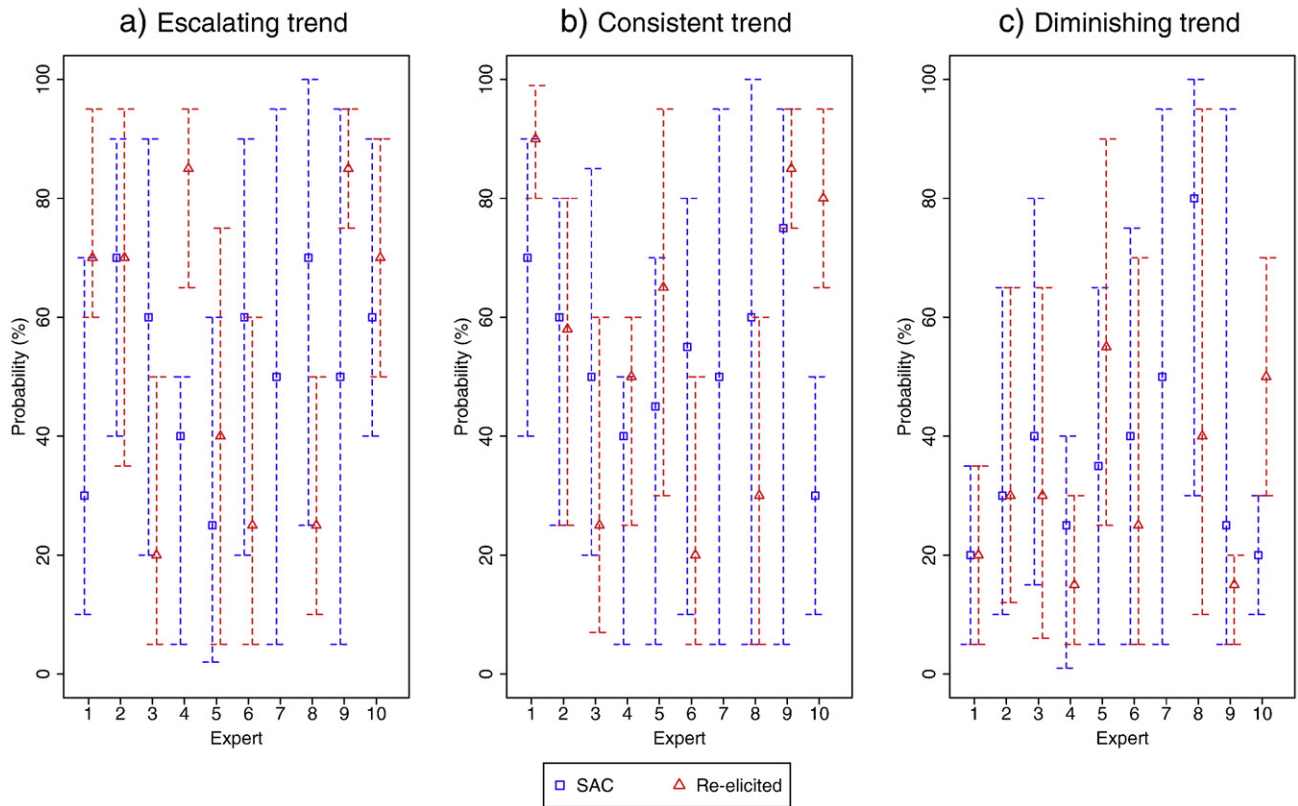
Fig. 8. Probability distribution quantiles (5th, 50th, 95th percentiles) for the 10 experts for observing: (a) an *escalating* trend; (b) a *consistent* trend, and (c) a *diminishing* trend, conditional on SHV remaining in a persistent regime. Blue squares represent answers from the original SAC meeting in October 2013 (Questions 4–6: Supplementary material) and red triangles the re-elicited versions (Questions 25–30: Supplementary material). All experts except one participated in the re-elicitation.

5. Discussion

5.1. Forecasting long-term eruptive activity at SHV

Large uncertainties in eruptive forecasts (Table 1) stem mainly from hesitancy in drawing firm inferences from the long-term behaviour of
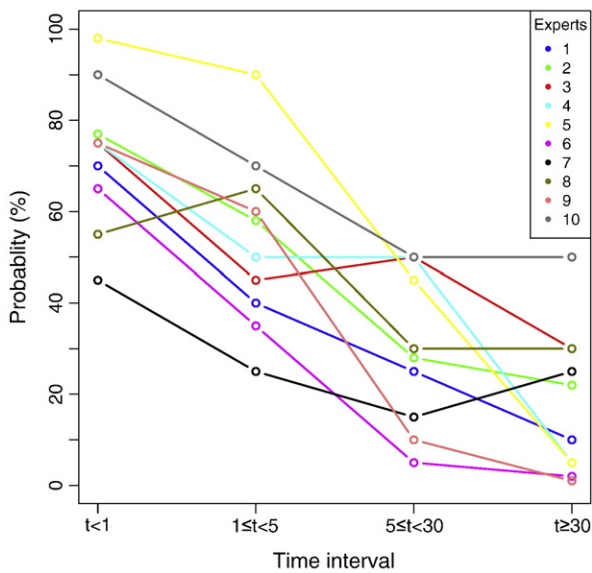

Fig. 9. Elicited median probabilities for the 10 experts for the conditional probability that the state of *I-REP* is *true* conditional on SHV next erupting in each one of the four time intervals $t < 1$; $1 \leq t < 5$; $5 \leq t < 30$; $t \geq 30$). (Questions 10–13: Supplementary material).

SHV and implications of the sustained degassing since February 2010. Thus, it is likely that this uncertainty will remain until there is a marked change in SO_2 flux or there is a significant improvement in understanding of the nature of magmatic degassing, and its long-term implications. Furthermore, the considerable expert uncertainty surrounding the direct informativeness of existing strands of monitoring data suggests that getting close to a probability high enough to satisfy the proposition that the eruption has stopped (however defined) with significant confidence from an evidence-based DM synthesis of judgements and existing data is improbable within several years at least. The propagation of uncertainty through Bayesian networks highlights many human beings' inherent tendency to underestimate uncertainty in complex systems (Best et al., 2013), suggesting that our model, despite the data limitations, provides a rational approach to constraining such uncertainty for long-range volcanic forecasting at SHV, or other volcanoes.

5.2. Interpretation of monitoring data

Generally, analyses of monitoring data and other observations of unrest are undertaken to seek and discern signals that may be short-term precursors to eruptive activity (e.g., Sparks, 2003; Marzocchi and Bebbington, 2012). Over longer timescales monitoring observables can be used as negative evidence, where the absence of signals or changes in data mean increased likelihood of no further eruptive activity. In both cases, however, the relationships between specific magmatic processes and monitoring observations are far from perfectly understood. We have endeavoured to overcome this problem in the Bayesian network design by developing the concept of "eruption potential", which allows experts to judge the competition between slow processes (e.g., heat loss) and fast processes (e.g., destabilisation and magma transport) in their interpretation of volcanic unrest.
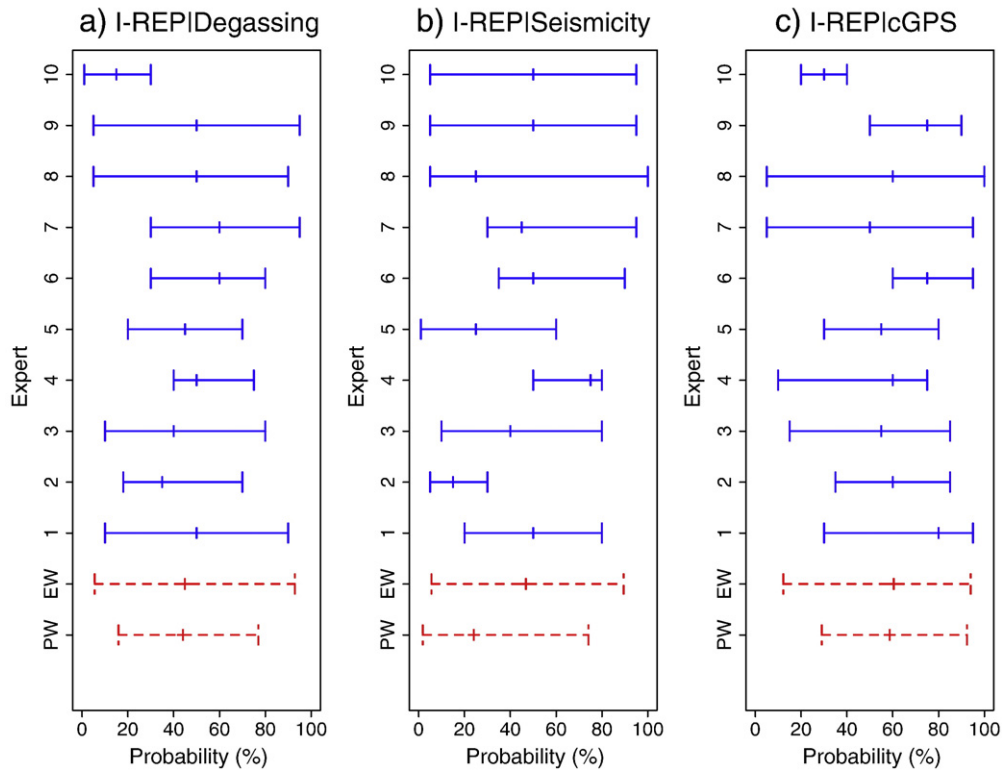
Fig. 10. Probability distribution quantiles (5th, 50th, 95th percentiles) for the 10 experts and the equal weighted (EW) and performance-weighted (PW) group result for the conditional probabilities that the state of *I-REP* is *true*, given each of the three observables: (a) current degassing, (b) seismicity and (c) cGPS. (Questions 14–16: Supplementary material).

It appears, in this case, that monitoring data cannot be used to decide, independently, and deterministically, when a volcano has entered a state of repose characteristic of a long-term pause in eruptive activity, or when an eruption - in this case at SHV - has 'ended'. However, monitoring data still has a significant role in long-term forecasting at volcanoes. Indeed, over half the expert group suggested that if SHV does not erupt for at least the next 30 years there is still a 20% (median probability) chance or greater that the current state of *I-REP* is true (i.e., a state in which shallow crustal processes are increasing the likelihood of SHV erupting): one expert even suggests that this probability is as high as 50% (Fig. 9).

The inclusion of a *latent* node that does not represent a specific magmatic process has likely increased the extent of uncertainty in the model, due to differences in what magmatic processes experts judge as significant for their interpretation of the data. Nevertheless, the concept of "eruption potential" has enabled experts to quantify the state of the magmatic system, without being constrained to any specific conceptual model, and has permitted monitoring data to be used as evidence to update probabilistic forecasts over a range of timescales using Bayes
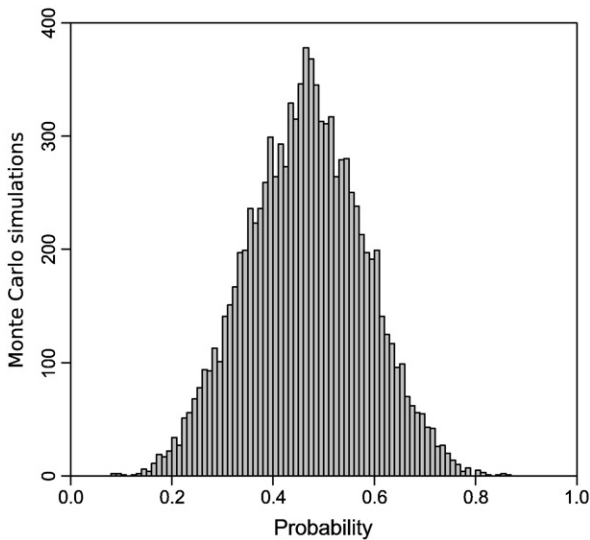

Fig. 11. Weighted distribution for the probability that the state of *I-REP* is *true* given a fixed-effects meta-analysis of the degassing, seismicity and deformation at SHV during the current pause in eruptive activity.
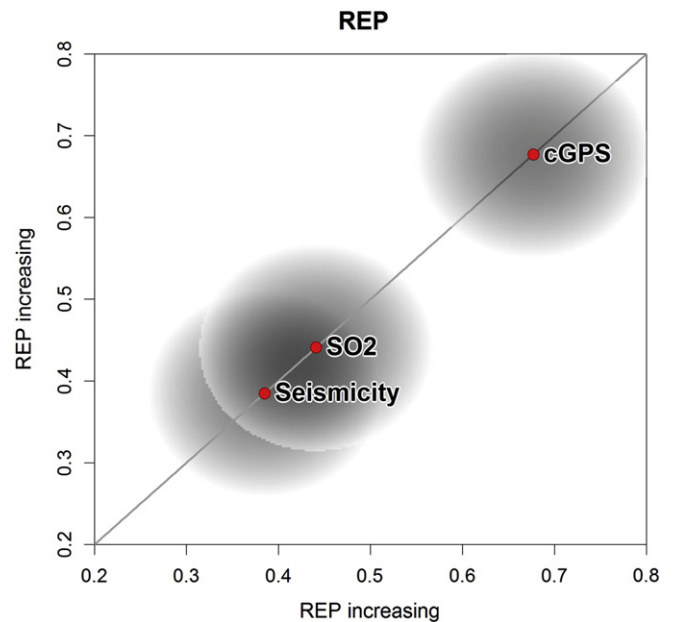

Fig. 12. Paired comparison ranking scores using the median probabilities in Fig. 10. The circles around the ranking scores represent the 90% credible interval for each ranking score.

theorem. Specifically, it appears that interpretation of the shallow magmatic system at SHV depends greatly on the interpretation of the geodetic observations.

5.3. Unique and common volcanic behaviour

Given the discrepancy in the experts ability to interpret either an episodic or persistent regime, it appears that in general experts found it difficult to interpret the current observations at SHV in the wider, generic context of eruptive behaviour at dome-building volcanoes. A Likert-style questionnaire (Likert, 1932) was designed to investigate this statement, with the goal of understanding the rationale behind the scientific understanding that was drawn on, and which observations informed the judgements of the experts. There were five questions to which experts were asked to respond with a number between 1 and 7, with 1 being unimportant and 7 being very important (Table 3). Two extra observers, who were not part of the original elicited group, also participated in this questionnaire.

The results of the questionnaire indicate that there are two schools of thought regarding how useful experts found the two major sources of information. One group (Table 3: B, C, E, I, J) considers the observations that are particular to SHV (Table 3: Q2–4) as being significantly more important than either empirical observations from other volcanoes (Table 3: Q1) or generic models for magmatic processes (Table 3: Q5). In contrast, a second group (Table 3: A, D, F, G, H) consider both sources of evidence as equally useful, albeit with a slightly heavier weighting towards data and scientific understanding that is specific to SHV. Whilst both these groups identify that SHV has unique characteristics, the latter group clearly believes that, over the timescales we are considering, the process of forecasting eruptive activity is based on identifying magmatic processes common to many similar volcanoes. In this regard, we noted that whilst observatory staff tended to fall more in the first school of thought, although not exclusively, academics were more likely to favour the alternative view.

Two experts (K & L) consider empirical evidence from other volcanoes not to be important (Table 3: Q1), yet they posit as exchangeable (i.e. similar) the processes and conceptual models that are so informed by analogues (Table 3: Q5). This represents the conundrum of universal processes versus unique volcanoes (Cashman and Biggs, 2014) and that, whilst empirical data may be used to deduce common magmatic processes using similar conceptual models, scientists still may not consider this information exchangeable when forecasting eruptive activity.

5.4. Biases in expert judgement

In constructing the problem framework, based on the application of Bayes theorem (Eq. 2), each piece of evidence is considered to be conditionally independent. When Bayesian networks are populated by observational data, conditional independence can be justified based solely on the assumptions of those designing the network (Sobradelo and Martí, 2015). When incorporating expert judgement into a Bayesian network, however, the assumption of conditional independence is harder to justify, as experts may not interpret the data as independent, especially when concerning *latent* nodes.

An example of where conditional independence may be invalidated is in the interpretation of the *Eruptive trend* for two experts (numbers 6 & 8), which appear to be influenced by their interpretation of the *Persistent degassing* node Based on the degassing, both experts judge that there is a low probability that SHV remains in a persistent regime (Table 2: column 5). However, this judgement also manifests in the median probabilities for SHV remaining in a persistent regime, given each state (*escalating, consistent and diminishing*) of the node *Eruptive trend*, which fall below 50% in the re-elicited questions (Fig. 8). In comparison, other experts within the group provide median probabilities that are more highly variable across these three states with similar bounds on the uncertainty.

Similar issues of conditional independence are observed when experts interpret the monitoring data. Firstly, there is systematic variation in the probabilities for the node *I-REP*, given each monitoring observation (Fig. 13). By this we mean that experts who provide high median probabilities for SO_2, also provide higher probabilities for seismicity and cGPS; and, vice-versa, where experts who provide low median probabilities for degassing, they also provide lower probabilities for seismicity and deformation. Secondly, one expert (expert 4) interprets the value of *I-REP* given seismicity to be much higher in comparison to the rest of the group (Fig. 13). However, the significance this expert places on seismicity also appears to manifest at other nodes: including in their high belief that SHV remains in a persistent regime (Table 2: column 5) and high belief that SHV will erupt within a year in a persistent regime (Fig. 6a).

5.5. Causality and uncertainty in eruptive forecasts

At the SAC meeting experts were additionally asked: 'Given what has happened up to the present and given current conditions, what is the probability that nothing significant will happen (i.e., no collapse, no restart of dome growth, no magmatic explosion) in the next 12 months?' The performance weighted median estimate for this question was 67%, which is effectively identical to the complement of the posterior value for Pr(erupt <1), which is 0.68 (Table 1). This result is promising as it supports the results of the Bayesian network: however, it might raise the question of whether the added complexity is justified. Nevertheless, adopting a Bayesian approach to the problem provides an opportunity to update beliefs in a logical and rational way, as well as disaggregating the contributions of different factors to understand where the largest uncertainties arise. This provides the potential to direct future research in understanding specific 'causes' and thus reduce uncertainties in future forecasts. Likewise, added sophistication to parts of the model can be used to explore the importance of specific evidence or processes.

The inclusion of unobservable parameters provides much greater flexibility in the model than conventional or deterministic approaches, enabling various different strands of evidence to be incorporated.

Table 2
Comparison of median probabilities for the probability of an *Eruptive regime* based on interpretation of the degassing behaviour of SHV in the current period of repose, calculated using (a) inverse conditional probabilities (columns 2 to 5); and (b) forward-conditional probabilities (column 6). Elicited probabilities (columns 2, 3 and 6) are on a scale of 0 to 100, whereas the likelihood (column 5) is on a scale of 0 to 1. The probabilities originate from the re-elicitation and so no values exist for expert 7.

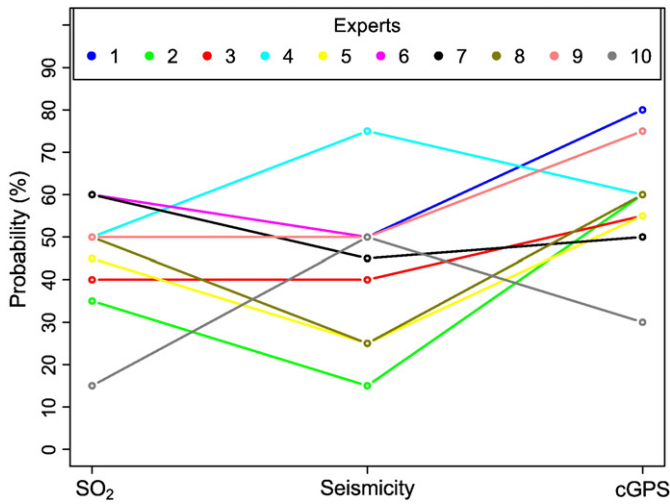| Expert	Pr(degassing|persistent)	Pr(degassing|episodic)	Likelihood ratio	Likelihood(persistent)	Pr(persistent|degassing)
1	80	20	80/20 = 4.00	4.00/(1 + 4.00) = 0.80	80
2	55	45	1.22	0.55	60
3	55	45	1.22	0.55	50
4	50	50	1	0.5	85
5	75	50	1.5	0.6	65
6	25	45	0.56	0.36	30
7	–	–	–	–	–
8	30	60	0.5	0.33	40
9	50	50	1	0.5	50
10	70	40	1.75	0.64	75

Fig. 13. Elicited median values for the probability that the state of *I-REP* is *true*, given each of the three monitoring observables, degassing, seismicity and deformation. (Questions 14–16: Supplementary material).

However, using expert judgement to populate the Bayesian model requires a common understanding of what the *latent* nodes represent. We identify coherence amongst experts when discussion relates to characterising the current state of the volcano (e.g., *I-REP* node), but discussion regarding longer-term magmatic processes provides more ambiguous and less coherent results, especially concerning the *Eruptive trend* node, interpreting what an episodic regime represents (Fig. 6–8), and relating the behaviour of SHV to other dome-building volcanoes (Table 3).

When using expert judgement to populate a Bayesian network it is non-trivial to understand and account for all dependencies in the variables of interest. Furthermore, it is unrealistic to expect experts to always interpret the evidence independently. Nevertheless, it will become important to understand these dependencies as expert elicitation and Bayesian methods become more commonly employed for forecasting eruptive activity. Hence, it will be valuable to explore the role of eliciting correlations between elicited variables (Garthwaite et al., 2005) or perhaps the development of more complex network structures such as vines, which allow dependences between random variables to be accounted for (Bedford and Cooke, 2002).

6. Conclusions

We describe the application of a Bayesian Network for providing scientific support to decision making in relation to a volcano - a complex natural system with sparse data - presenting a case example of how causal inference can be used to design probabilistic forecasting models. To ensure that judgements of experts concerning forecasts of future eruptive behaviour can be elicited tractably, it has been advantageous to keep the Bayesian network model simple, using an empirical approach to analyse evidence for unobservable magmatic processes. This Bayesian approach for combining empirical evidence means criteria (or metrics) for the significance of observables are not required a priori. With the inclusion of *latent* nodes in the Bayesian Network, evidence has been synthesised

in a logical way, based around discussion of magmatic processes; but, importantly, by using the eruption potential concept, inference is not restricted to a single conceptual model for all magmatic systems.

Forecasting volcanic eruptions over long-timescales requires some conceptualisation or accounting for magmatic processes occurring in the vertical direction throughout the crust. However, volcanologists' perspectives when interpreting observations of volcanic unrest are typically dominated by short-term, shallow crustal processes. Here, an alternative methodology is presented with which monitoring data can be used to update beliefs of when a volcano is next going to erupt, without the need to interpret volcanic unrest as either explicitly positive evidence (i.e., precursory) or negative evidence (i.e., baseline).

In volcanology, there is often a lack of adequate available data upon which statistical inference can be based. Causal inference provides a balance between statistically robust methods and a level of pragmatism. Utilising Bayes Rule in the graphical framework of a Bayesian network allows evidence to be interpreted to estimate the probability of 'causes of effects'. It allows various strands of conditionally dependent evidence to be interpreted and synthesised to estimate the probability of 'causes' (i.e., magmatic processes) and thus forecast eruptive activity. Significantly, this allows the sources of uncertainty in probabilistic forecasts to be quantified in terms of the 'causes' rather than 'effects'. Such an approach will also provide the potential to include alternative sources of evidence such as petrological data, where it is easier to interpret data with respect to 'causes' (i.e., long-term magmatic processes) rather than 'outcomes' (i.e., eruptive (in-)activity).

Expert judgement provides a tractable solution to issues of sparse data at a volcano and has the potential to be a relevant approach to quantify 'causes of effects' when forecasting eruptive (in-)activity at volcanoes in general. The findings from our analysis indicate that experts may find it hard to interpret different data strands independently and so Bayesian networks and forecasting models must be carefully designed to account for this.

In this novel approach, magmatic 'causes' have not been explicitly identified in the probabilistic model. Instead evidence is interpreted with regard to long-term patterns of behaviour, informed by empirical data from other volcanoes. The results of our expert elicitation suggest experts found it easier to interpret the evidence for the *status quo* at SHV (i.e., in a persistent regime) and were less confident in relating other patterns of dome-building volcanism to SHV. In a comparable approach, monitoring observables are used to interpret the state of the magma reservoir system rather than identify specific magmatic processes. There is coherence amongst the experts in their interpretation of what this 'latent node' represents, suggesting a similar approach could be employed when analysing observations of unrest to forecast eruptive activity over much shorter timescales than we consider here.

Table 3
Results from a questionnaire on the relevance of the different sources of evidence (i.e. relevant to SHV or dome-building volcanoes in general). Experts are labeled using letters (A–L) as their order is not equivalent to the order in the elicitation results (1 − 10), as this questionnaire was asked anonymously. It also includes the opinions of two extra observers who are not in the elicited group but took part in discussions related to the elicitation.

	A	B	C	D	E	F	G	H	I	J	K	L
1. How important was the information from the analog volcano report and workbook?	5	3	3	5	3	4	5	5	2	1	4	2
2. How important was the monitoring data since 1995?	6	6	6	7	5	5	4	6	7	7	7	7
3. How important was the monitoring data since the current period of quiescence (since Feb 2010)?	6	6	5	7	4	5	2	5	7	7	7	6
4. How important were the discussions on the possible volcanic and geophysical processes at SHV?	6	5	6	6	6	4	5	5	3	6	7	7
5. How important were the science discussions that referred to data and ideas developed at other volcanoes?	5	3	3	5	4	4	5	6	3	2	7	7

Appendix A. Supplementary material

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jvolgeores.2017.06.007.

References

Aspinall, W., 2010. A route to more tractable expert advice. Nature 463:294–295. http://dx.doi.org/10.1038/463294a.

Aspinall, W.P., Cooke, R.M., 2013. Expert elicitation and judgement. Risk and Uncertainty Assessment for Natural Hazards. Cambridge University Press, pp. 64–99 Ch. 4.

Aspinall, W.P., Woo, G., 2014. Santorini unrest 2011–2012: an immediate Bayesian belief network analysis of eruption scenario probabilities for urgent decision support under uncertainty. J. Appl. Volcanol. 3:12. http://dx.doi.org/10.1186/s13617-014-0012-8.

Aspinall, W.P., Woo, G., Voight, B., Baxter, P.J., 2003. Evidence-based volcanology; application to eruption crises. J. Volcanol. Geotherm. Res. 128:273–285. http://dx.doi.org/10.1016/S0377-0273(03)00260-9.

Barclay, J., Herd, R.A., Edwards, B.R., Christopher, T., Kiddle, E.J., Plail, M., Donovan, A., 2010. Caught in the act: implications for the increasing abundance of mafic enclaves during the recent eruptive episodes of the Soufrière Hills Volcano, Montserrat. Geophys, Res. Lett. 37 (L00E09). http://dx.doi.org/10.1029/2010GL042509.

Bebbington, M., 2008. Incorporating the eruptive history in a stochastic model for volcanic eruptions. J. Volcanol. Geotherm. Res. 175:325–333. http://dx.doi.org/10.1016/j.jvolgeores.2008.03.013.

Bebbington, M.S., 2013. Models for temporal volcanic hazard. Stat. Volcanol. 1:1–24. http://dx.doi.org/10.5038/2163-338X.1.1.

Bedford, T., Cooke, R.M., 2002. Vines-a new graphical model for dependent random variables. Ann. Stat. 30:1031–1068. http://dx.doi.org/10.1214/aos/1031689016.

Bélisle, P., 2012. beta.parms.from.quantiles software (version 1.2.2). See. http://www.medicine.mcgill.ca/epidemiology/joseph/pbelisle/BetaParmsFrom Quantiles.html.

Best, N., Ashby, D., Dunstan, F., Foreman, D., McIntosh, N., 2013. A Bayesian approach to complex clinical diagnoses: a case-study in child abuse. J. R. Stat. Soc. 176:53–96. http://dx.doi.org/10.1111/j.1467-985X.2012.01050.x.

Cashman, K., Biggs, J., 2014. Common processes at unique volcanoes-a volcanological conundrum. Front. Earth Sci. 2. http://dx.doi.org/10.3389/feart.2014.00028.

Cashman, K.V., Sparks, R.S.J., 2013. How volcanoes work: a 25-year perspective. Geol. Soc. Am. Bull. http://dx.doi.org/10.1130/B30720.1.

Christopher, T.E., Humphreys, M.C.S., Barclay, J., Genareau, K., de Angelis, S.M., Plail, M., Donovan, A., 2014. Petrological and geochemical variation during the Soufrière Hills eruption, 1995 to 2010. Geol. Soc. Lond. Mem. 39, 317–342.

Christopher, T.E., Blundy, J., Cashman, K., Cole, P., Edmonds, M., Smith, P.J., Sparks, R.S.J., Stinton, A., 2015. Crustal-scale degassing due to magma system destabilization and magma-gas decoupling at Soufrière Hills Volcano Montserrat. Geochem. Geophys. Geosyst. 16:2797–2811. http://dx.doi.org/10.1002/2015GC005791.

Cole, P., Bass, V., Christopher, T., Murrell, C., Odbert, H., Smith, P., Stewart, R., Stinton, A., Syers, R., Williams, P., 2011. MVO scientific report for volcanic activity between 1 May 2011 and 31 October 2011 prepared for the sixteenth meeting of the Scientific Advisory Committee 14–16 November 2011. Open File Report OFR 11-02 available online at. http://www.mvo.ms/pub/.

Cole, P.D., Smith, P., Komorowski, J.-C., Alfano, F., Bonadonna, C., Stinton, A., Christopher, T., Odbert, H.M., Loughlin, S., 2014. Ash venting occurring both prior to and during lava extrusion at Soufrière Hills Volcano, Montserrat, from 2005 to 2010. Geol. Soc. Lond. Mem. 39:71–92. http://dx.doi.org/10.1144/M39.4.

Development Core Team, R., 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria 3-900051-07-0 http://www.R-project.org.

Edmonds, M., Herd, R., Galle, B., Oppenheimer, C., 2003. Automated, high time-resolution measurements of SO_2 flux at Soufrière Hills Volcano Montserrat. Bull. Volcanol. 65: 578–586. http://dx.doi.org/10.1007/s00445-003-0286-x.

Elsworth, D., Mattioli, G., Taron, J., Voight, B., Herd, R., 2008. Implications of magma transfer between multiple reservoirs on eruption cycling. Science 322:246–248. http://dx.doi.org/10.1126/science.1161297.

Garthwaite, P.H., Kadane, J.B., O'Hagan, A., 2005. Statistical methods for eliciting probability distributions. J. Am. Stat. Assoc. 100:680–701. http://dx.doi.org/10.1198/016214505000000105.

Gelman, A., Imbens, G., 2013. Why ask Why? Forward Causal Inference and Reverse Causal Questions. Tech. Rep. 19614. National Bureau of Economic Research.

Gottsmann, J., Odbert, H., 2014. The effects of thermomechanical heterogeneities in island arc crust on time-dependent preeruptive stresses and the failure of an andesitic reservoir. J. Geophys. Res. Solid Earth 119:4626–4639. http://dx.doi.org/10.1002/2014JB011079.

Hautmann, S., Witham, F., Christopher, T., Cole, P., Linde, A.T., Sacks, I.S., Sparks, R.S.J., 2014. Strain field analysis on Montserrat (W.I.) as tool for assessing permeable flow paths in the magmatic system of Soufrière Hills Volcano. Geochem. Geophys. Geosyst. 15:676–690. http://dx.doi.org/10.1002/2013GC005087.

Hincks, T., Komorowski, J.-C., Sparks, S., Aspinall, W., 2014. Retrospective analysis of uncertain eruption precursors at la Soufrière volcano, Guadeloupe, 1975–77: volcanic hazard assessment using a Bayesian Belief Network approach. J. Appl. Volcanol. 3:3. http://dx.doi.org/10.1186/2191-5040-3-3.

Likert, R., 1932. A technique for the measurement of attitudes. Arch. Psychol. 140, 1–55.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. Stats. Comput. 10:325–337. http://dx.doi.org/10.1023/A:1008929526011.

Martí, J., Aspinall, W.P., Sobradelo, R., Felpeto, A., Geyer, A., Ortiz, R., Baxter, P., Cole, P., Pacheco, J., Blanco, M.J., Lopez, C., 2008. A long-term volcanic hazard event tree for Teide- Pico Viejo stratovolcanoes (Tenerife, Canary Islands). J. Volcanol. Geotherm. Res. 178:543–552. http://dx.doi.org/10.1016/j.jvolgeores.2008.09.023.

Marzocchi, W., Bebbington, M.S., 2012. Probabilistic eruption forecasting at short and long time scales. Bull. Volcanol. 74:1777–1805. http://dx.doi.org/10.1007/s00445-012-0633-x.

Marzocchi, W., Sandri, L., Gasparini, P., Newhall, C., Boschi, E., 2004. Quantifying probabilities of volcanic events: the example of volcanic hazard at Vesuvius. J. Geophys. Res. 109:3–20. http://dx.doi.org/10.1029/2004JB003155.

Marzocchi, W., Sandri, L., Selva, J., 2008. BET_EF: a probabilistic tool for long- and short-term eruption forecasting. Bull. Volcanol. 70:623–632. http://dx.doi.org/10.1007/s00445-007-0157-y.

Mattioli, G.S., Herd, R.A., Strutt, M.H., Ryan, G., Widiwijayanti, C., Voight, B., 2010. Long term surface deformation of Soufrière Hills Volcano, Montserrat from GPS geodesy: Inferences from simple elastic inverse models. Geophys. Res. Lett. 37 (L00E13). http://dx.doi.org/10.1029/2009GL042268.

Norton, G.E., Watts, R.B., et al., 2002. Pyroclastic flow and explosive activity at Soufrière Hills Volcano, Montserrat, during a period of virtually no magma extrusion. Geol. Soc. Lond. Mem. 21:467–482. http://dx.doi.org/10.1144/GSL.MEM.2002.021.01.21.

Odbert, H.M., Stewart, R.C., Wadge, G., 2014. Cyclic phenomena at the Soufrière Hills Volcano, Montserrat. Geol. Soc. Lond. Mem. 39:41–60. http://dx.doi.org/10.1144/M39.2.

Odbert, H., Taisne, B., Gottsmann, J., 2015. Deposit loading and its effect on co-eruptive volcano deformation. Earth Planet. Sci. Lett. 413:186–196. http://dx.doi.org/10.1016/j.epsl.2015.01.005.

Pearl, J., 2009. Causality: Models, Reasoning and Inference. 2nd ed. Cambridge University Press, Cambridge, UK.

Phillipson, G., Sobradelo, R., Gottsmann, J., 2013. Global volcanic unrest in the 21st century: an analysis of the first decade. J. Volcanol. Geotherm. Res. 264:183–196. http://dx.doi.org/10.1016/j.jvolgeores.2013.08.004.

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 Sci. Rev. 160:240–263. http://dx.doi.org/10.1016/j.earscirev.2016.07.013.

Sobradelo, R., Martí, J., 2015. Short-term volcanic hazard assessment through Bayesian inference: retrospective application to the Pinatubo 1991 volcanic crisis. J. Volcanol. Geotherm. Res. 290:1–11. http://dx.doi.org/10.1016/j.jvolgeores.2014.11.011.

Sobradelo, R., Bartolini, S., Martí, J., 2013. HASSET: a probability event tree tool to evaluate future volcanic scenarios using Bayesian inference. Bull. Volcanol. 76:770. http://dx.doi.org/10.1007/s00445-013-0770-x.

Sparks, R.S.J., 2003. Forecasting volcanic eruptions. Earth Planet. Sci. Lett. 210:1–15. http://dx.doi.org/10.1016/S0012-821X(03)00124-9.

Sparks, R.S.J., Ashman, K.V., 2016. The implications of magma system models for forecasting volcanic activity. Elements 13, 35–40.

Sparks, R.S.J., Young, S.R., 2002. The eruption of Soufrière Hills Volcano, Montserrat (1995–1999): overview of scientific results. Geol. Soc. Lond. Mem. 21:45–69. http://dx.doi.org/10.1144/GSL.MEM.2002.021.01.03.

Stewart, R., Bass, V., Christopher, T., Cole, P., Dondin, F., Higgins, M., Jospeh, E., Smith, P., Stinton, A., Syers, R., Williams, P., 2013. MVO scientific report for volcanic activity between 13 October 2012 and 30 April 2013. Open File Report OFR 13-06 available online at. http://www.mvo.ms/pub/.

Stiros, S.C., Psimoulis, P., Vougioukalakis, G., Fyticas, M., 2010. Geodetic evidence and modeling of a slow, small-scale inflation episode in the Thera (Santorini) volcano caldera, Aegean Sea. Tectonophysics 494:180–190. http://dx.doi.org/10.1016/j.tecto.2010.09.015.

Sturtz, S., Ligges, U., Gelman, A., 2005. R2WinBUGS: a package for running WinBUGS from R. J. Stat. Softw. 12, 1–16.

Surono, Jousset P., Pallister, J., Boichu, M., Buongiorno, M.F., Budisantoso, A., Costa, F., Andreastuti, S., Prata, F., Schneider, D., Clarisse, L., Humaida, H., SuMartí, S., Bignami, C., Griswold, J., Carn, S., Oppenheimer, C., Lavigne, F., 2012. The 2010 explosive eruption of Java's Merapi volcano - a '100-year' event. J. Volcanol. Geotherm. Res.:121–135 241,242. http://dx.doi.org/10.1016/j.jvolgeores.2012.06.018.

Wadge, G., Aspinall, W.P., 2014. A review of volcanic hazard and risk-assessment praxis at the Soufrière Hills Volcano, Montserrat from 1997 to 2011. Geol. Soc. Lond. Mem. 39: 439–456. http://dx.doi.org/10.1144/M39.24.

Wadge, G., Herd, R., Ryan, G., Calder, E.S., Komorowski, J.C., 2010. Lava production at Soufrière Hills Volcano, Montserrat: 1995–2009. Geophys. Res. Lett. 37 (L00E03). http://dx.doi.org/10.1029/2009GL041466.

Wadge, G., Voight, B., Sparks, R.S.J., Cole, P.D., Loughlin, S.C., Robertson, R.E.A., 2014. An overview of the eruption of Soufrière Hills Volcano, Montserrat from 2000 to 2010. Geol. Soc. Lond. Mem. 39:1–40. http://dx.doi.org/10.1144/M39.1.

Young, S.R., Sparks, R.S.J., Aspinall, W.P., Lynch, L.L., Miller, A.D., Robertson, R.E.A., Shepherd, J.B., 1998. Overview of the eruption of Soufrière Hills Volcano, Montserrat, 18 July 1995 to December 1997. Geophys. Res. Lett. 25:3389–3392. http://dx.doi.org/10.1029/98GL01405.