The Robustness of Volcanic Hazard Forecasts under Uncertainty

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Earth Science

at Massey University (Manawatū campus), Aotearoa New Zealand

Emmy Elizabeth Scott

2025

MASSEY UNIVERSITY

TE KUNENGA KI PŪREHUROA

UNIVERSITY OF NEW ZEALAND

Abstract

Forecasting volcanic hazards is crucial for mitigating the impacts of volcanic activity; however, substantial uncertainty exists in model outputs due to both inherent variability in eruptive processes and limited observational data. Computational volcanic hazard models translate physical processes into simulations, but differences in model structure, input requirements, and assumptions can strongly influence forecast outcomes. Furthermore, the effect of prior knowledge about a volcano's eruptive history on forecast accuracy remains poorly understood.

This thesis investigates how model choice, input parameter uncertainty, and volcano-specific knowledge affect the accuracy and usefulness of volcanic hazard forecasts. A structured review and classification of existing models establishes a framework for understanding how differences in model design and computational complexity influence forecasting under uncertainty. Global sensitivity analyses using two volcanic ash transport and dispersion models, Tephra2 and Fall3D, identify which input parameters most strongly influence output variance. Case-study simulations of the 17 June 1996 eruption of Mount Ruapehu, Aotearoa New Zealand, compare forecasts generated with informed versus uninformed input parameter distributions, evaluating both mass-based and impact-based accuracy.

Results show that input parameters such as median grain size, diffusion, and plume shape consistently have the largest influence on forecast variability, while others, such as plume height, demonstrate limited individual influence but become influential through higher-order interactions with other input parameters. Case-study forecasts indicate that reliance on prior eruption data does not systematically improve forecast accuracy and can introduce biases.

These findings provide guidance for prioritising input parameters, selecting suitable models, and balancing complexity with uncertainty to improve forecast accuracy and efficiency. By linking model structure, input sensitivity, and data bias, this work contributes to more accurate, actionable forecasts, supporting decision-making for volcanic hazard mitigation and emergency response.

Acknowledgements

This thesis would not have been possible without the support and advice from countless people, to whom I express my sincere thanks.

Dr Melody Whitehead - Thank you for putting your dislike of ash modelling aside and offering continued support, encouragement, and enthusiasm throughout this project. Your prompt feedback was invaluable, and I'm proud to have finally deciphered (most of) your handwriting (only took four years). I'll now carry a healthy scepticism about everything in science - especially when researchers fail to account for uncertainty.

Thank you also to my co-supervisors:

- Dr Stuart Mead - Thank you for your continued enthusiasm for my thesis. Your help setting up Tephra2 and Fall3D simulations, as well as navigating NeSI, saved me months of extra hardship. Your IT wizardry was a lifeline.

- Prof. Mark Bebbington - Thank you for always finding time to review my work. Ever since I became your student during my Master's, your input and expertise have been invaluable, particularly when I'm stumbling around trying to understand statistics. If only you could have convinced everyone to use the word "robustifying" in my thesis.

- Prof. Jonathan Procter - Thank you for your support in this project, particularly for your assistance in securing funding. Thank you for all the opportunities you have provided for me to present to and connect with researchers and students.

I would also like to thank everyone at Volcanic Risk Solutions for your support, feedback and advice. I would also like to thank my office pals - Maia Kidd, Dan Sturgess, Rae Sanchez, and Joseph Fleming - for your wisdom, lunchtime breaks, sweet treat runs, and unwavering support.

I am deeply grateful for the funding that made this project possible. Thank you to Resilience to Nature's Challenges (RNC) Volcano Program, the Dr Eileen Fair Doctoral Scholarship in Earth Science, and the Massey University Doctoral Conference Grant for their financial support to undertake research, live, and participate in international conferences.

Thank you to my family and friends for always supporting me throughout this thesis and wider academic journey. Your enthusiasm to one day call me "Dr Scott" and "rock doctor" has sustained me.

I owe my greatest appreciation to Jono. Thank you for endless encouragement, advice, patience, and support, especially during nonsensical meltdowns.

A special thanks to Squiggles, who has travelled with me through three cities (I'll try not to take you on another aeroplane) and for being a constant source of companionship and laughs.

Contents

List of Tables

List of Figures

List of Acronyms

1D	One-dimension(al)
2D	Two-dimension(al)
3D	Three-dimension(al)
DEM	Digital elevation model
eFAST	extended Fourier Amplitude Sensitivity Test
ESP	Eruption source parameter
FTT	Fall time threshold
GSA	Global sensitivity analysis
GVP	Global Volcanism Program
IVESPA	The Independent Volcanic Eruption Source Parameter Archive
LSA	Local sensitivity analysis
MER	Mass eruption rate
PDC	Pyroclastic density current
SA	Sensitivity analysis
TEM	Total erupted mass
TGSD	Total grain size distribution
VAAC	Volcanic Ash Advisory Centre
VATDM	Volcanic ash transport and dispersion model
VEI	Volcanic explosivity index
VBP	Volcanic ballistic projectile

Chapter 1

Introduction

Volcanic eruptions pose a constant risk to proximal populations. Eruptions vary from small and effusive to large and explosive, and their hazards (e.g., lava, tephra) may cause catastrophic impacts, as well as economic, societal, and critical lifeline disruptions and loss (Wilson et al. 2014; Loughlin et al. 2015). However, if information on an eruption and its associated hazards can be obtained beforehand, then informed decisions can be made to reduce impacts (Wilson, Dantas, and Cole 2009; Mei et al. 2013). This information is currently provided through long-term hazard assessments and short-term or near-real-time hazard forecasts.

Volcanic hazard forecasts are progressively reliant on computational models that simulate specific hazard dynamics to produce estimates of hazard intensity, duration, and spatial footprints (e.g., tephra, pyroclastic density currents (PDC), and volcanic ballistic projectiles (VBP), Sparks and Aspinall 2004; Poland and Anderson 2020; Papale 2021). Volcanic hazard models are impeded by multiple sources of uncertainty, limitations in knowledge of volcanic behaviour, incomplete and biased data, and the simplifying assumptions required by model resolution and computational constraints. This thesis addresses the core challenge of improving short-term and near-real-time volcanic hazard forecasts by evaluating how model structure, input uncertainty, and prior knowledge shape forecast accuracy under operational constraints. While this thesis provides a general discussion of volcanic hazard models, the detailed analysis and case studies focus specifically on volcanic ash.

This chapter presents the thesis aims and objectives, followed by background information about volcanic eruption phenomena and their hazardous nature, followed by the exploration of

computational volcanic hazard models for both hindcasting and forecasting. It concludes with a summary of the structure of the subsequent chapters.

1.1 Problem Statement and Objectives

Volcanic hazard models are essential tools for forecasting future volcanic hazards and their impacts. However, their accuracy is limited by the difficulty of extrapolating model performance from past-event validation to future forecasting. As Oreskes, Shrader-Frechette, and Belitz (1994) noted "*if a model result is consistent with present and past observational data, there is no guarantee that the model will perform at an equal level when used to predict the future*". This uncertainty remains a central challenge in volcanic hazard forecasting, particularly for short-term, operational applications where data quality, computational resources, and time are limited.

While the primary driving forces behind volcanic hazards, such as gravity, wind advection, and particle settling, are generally well understood (Sparks et al. 1997; Branney and Kokelaar 2002; Sulpizio and Dellino 2008; Bonadonna et al. 2015), substantial uncertainty emerges when these processes are translated into computational models. Both aleatoric uncertainty (inherent variability in eruptive conditions) and epistemic uncertainty (arising from limited knowledge or sparse observations) become significant when eruption source parameters (ESPs) are poorly constrained. Current forecasting-modelling approaches adopt a baseline assumption that if a model can accurately simulate *past* events, then the model can simulate a *future* event (Poland and Anderson 2020), yet this assumption is rarely tested across the range of plausible eruptive scenarios and may not hold under conditions of high uncertainty or limited observational data.

These uncertainties underscore the importance of selecting appropriate models for specific hazard types and forecasting contexts. Volcanic hazard models vary widely in structure, complexity, and outputs, from simple empirical relationships (e.g., energy cone, Malin and Sheridan 1982) to complex numerical equations (e.g., Titan2D, Patra et al. 2005). Even within a single model, choices of ESPs (e.g., plume height or grain size) can substantially alter forecast outcomes. Despite the widespread use of these models, concepts such as model complexity, structure, testing, and the assumptions underlying each model are often ill-defined or inconsistently described across the literature (e.g., Guthke 2017; Höge, Wöhling, and Nowak 2018;

Baartman et al. 2020). No comprehensive analysis has systematically compared these characteristics for all volcanic hazards in a model framework, limiting guidance on model selection and the interpretation of forecast uncertainty. Understanding these factors is therefore key to improving short-term and real-time volcanic hazard forecasts.

The primary aim of the thesis is to understand the factors that influence the accuracy of short-term volcanic hazard forecasts. The above considerations lead to the following three main avenues of research in volcanic hazard forecasting.

1. **Does model choice matter for volcanic hazard forecasting under uncertainty?** This objective examines whether model selection meaningfully influences volcanic hazard forecasts, given the high uncertainty in input parameters and eruptive behaviour. It includes a structured review and classification of volcanic hazard models, identifying how differences in assumptions, input requirements, complexity, computational demands, and representations of uncertainty align with forecasting needs.

2. **Which input parameters most influence volcanic hazard model outputs?** While reliable forecasting requires accounting for input uncertainty, not all input parameters contribute equally to model outputs. During heightened volcanic unrest, there is insufficient time to explore all possible combinations of input parameters in the lead-up to an eruption. Using volcanic ash as an exemplar, this objective investigates the relative importance of input parameters in shaping volcanic ash deposition forecasts and employs global sensitivity analysis to identify which input parameters most strongly influence simulated mass deposition.

3. **To what extent does volcano-specific knowledge improve or limit forecast accuracy?** Over-constraining input parameter estimates may bias forecasts toward previously observed eruptive scenarios, increasing the likelihood of excluding the true future event within its uncertainty range. Conversely, under-constraining inputs reduces bias but yields broader uncertainty bounds, which can make the forecast less actionable. This objective explores the bias-variance trade-off by comparing uninformed (no volcano-specific knowledge) and informed (volcano-specific data such as magma type and eruptive history) input parameter sets in forecasting the 17 June 1996 eruption of Mount Ruapehu, Aotearoa New Zealand.

1.2 Background Information

1.2.1 Volcanic Hazards

A volcanic hazard is defined here as a volcanic process that may result in a negative impact on society or the environment (UNDRR 2017) (Figures 1.1, 1.2). This definition includes hazards that occur during an eruption, such as ashfall, PDC, lava flows, and VBP, but also hazards that may occur both during and between eruptions, including lahars, block and ash flows, volcanic earthquakes, uplift, subsidence, and volcanic gases (Gottsmann, Komorowski, and Barclay 2019).

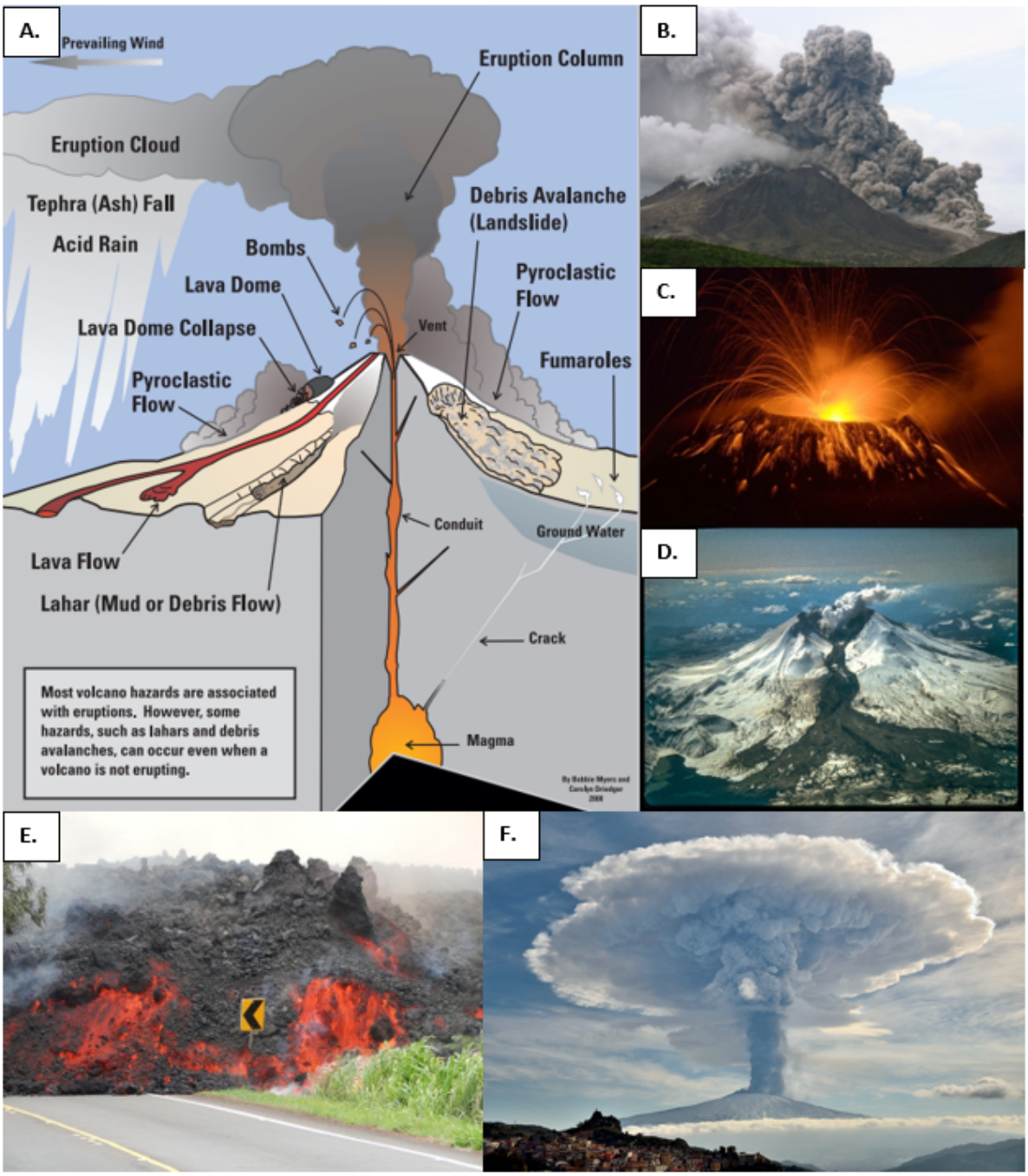


Figure 1.1: (A.) Schematic diagram of volcanic hazards (Myers and Driedger 2008); (B.) Pyroclastic density current, 2010 Soufriere Hills, Montserrat (Stone 2010); (C.) Volcanic ballistic projectiles, 2010 Tungurahua, Ecuador (Bernard 2018); (D.) Lahar, 1982 Mount St. Helens, USA (Casadevall 1982); (E.) Lava flow, 2018 Kīlauea, Hawaii (Hawaiian Volcano Observatory 2019); (F.) Volcanic ash plume, 2015 Mount Etna, Italy (de'Michieli Vitturi and Pardini 2021).

Figure 1.2: (A.) Roof of a mountain hut hit by volcanic ballistic projectiles, 2014 Mount Ontake, Japan (Williams et al. 2017); (B.) Pyroclastic density current impacts, 1997 Soufrière Hills volcano, Montserrat (the black arrow indicates the direction of the flow) (Baxter et al. 2005); (C.) lava inundation between roof structures, 2021 Tajogaite eruption, La Palma, Spain (Biass et al. 2024); (D.) Lahar inundating a village, 2011 Mount Merapi, Indonesia (Jenkins et al. 2015a); (E.) Aeroplane covered with volcanic ash in Bariloche airport, 2011 Puyehue-Cordon Caulle, Chile (Elissondo et al. 2016).

Volcanic hazards can impact areas ranging from those immediately surrounding the vent to regions thousands of kilometres away. VBP, for example, typically land within 5 km of the vent, but can reach up to ∼10 km (Blong 1984). Other hazards, such as lahar and debris flows, may travel up to 100 km, typically confined to river valleys and flood plains (Thouret et al.

2020). The most widespread volcanic hazard is tephra, which can travel hundreds to thousands of kilometres within the atmosphere (e.g., Rose and Durant 2009). The distance a hazard may travel is an important concept in volcanic hazard modelling, as it dictates (in part) how they are modelled and what data are required. For example, suppose a lahar may travel over 100 km from the vent; any lahar model will need to incorporate a digital elevation model (DEM) covering at least 100 km around the volcano to capture the spatial footprint. However, distance alone does not fully characterise hazard impact. Hazard intensity is also important, although its relevance varies among volcanic processes. For example, lava flows are uniformly destructive, with damage determined primarily by whether the flow interacts with something or not (Harris 2015; Jenkins et al. 2017). In contrast, tephra deposition requires quantification of deposit thickness or mass loading, as varying intensities produce distinct impacts on infrastructure, health, and aviation (Wilson et al. 2012b; Jenkins et al. 2015b; Wilson, Jenkins, and Stewart 2015).

Despite volcanic hazards being related to the same fundamental process (volcanic eruptions), hazard behaviours are highly varied. Explosive eruptions produce hazards such as ashfall, PDCs, and VBP, whereas effusive eruptions more commonly produce lava flows (Tilling 1989). The dynamics of each hazard also vary: volcanic ash transport is governed initially by the force of the eruption pushing the material up and out of the volcano (Figure 1.1F), then atmospheric conditions take over, such as wind speed and wind direction, eventually leading to settling and deposition due to gravity (Bonadonna et al. 2015) (Figure 1.2E). On the other hand, lahars are gravity-driven processes that are primarily constrained by topography, particularly low-lying areas and valley floors (Thouret et al. 2020).

1.2.2 Volcanic Hazard Models

Volcanic hazard models simulate specific volcanic hazards by processing a set of input parameters that describe the eruption source, the surrounding landscape, and/or atmospheric conditions. These inputs are run through computational code, typically structured around physical equations or empirical relationships, to generate outputs that estimate the spatial extent, intensity, and/or duration of the hazard (Papale 2017).

The type of data required to run volcanic hazard models depends on the physical processes

being simulated. For tephra and gas dispersion, advection and diffusion are driven by meteorological conditions, making atmospheric data essential (Bonadonna and Costa 2013). Flow models rely on DEMs, which capture topographic constraints on flow paths (Branney and Kokelaar 2002; Sulpizio and Dellino 2008). In practice, both input types require frequent updating. Operational ash deposition forecasts in Aotearoa New Zealand, for example, are typically run over 24-hour periods and run every six hours to incorporate the latest meteorological data (Trancoso et al. 2022). This short forecast window is important because weather forecast accuracy (i.e., the closeness of agreement between a test result and the accepted reference value, ISO 5725-1 1994) declines significantly beyond seven days (Haiden et al. 2021). Similarly, ongoing eruptive activity can substantially alter topography through successive emplacement of deposits, reducing digital elevation model (DEM) accuracy and degrading flow forecast accuracy unless DEMs are regularly updated (De Beni, Cantarero, and Messina 2019).

Beyond the type and quality of data, volcanic hazard models also differ in how they represent uncertainty. Deterministic models produce an individual point measure for an event (i.e., one possible outcome for one set of input parameters) (Manga et al. 2017). In contrast, probabilistic models produce a range of outcomes, often expressed as a distribution, designed to capture uncertainty in both input parameters and resulting outputs (Manga et al. 2017). In practice, probabilistic outputs are frequently generated by executing deterministic models multiple times across a range of plausible input values (e.g., Procter et al. 2010a; Wild et al. 2019; Verolino et al. 2022; Osman et al. 2024). Ensemble forecasting extends this concept further by combining multiple, different (deterministic) models and/or multiple simulations from one model to capture a broader range of potential outcomes, thereby better representing epistemic (knowledge-based) uncertainty (Toth and Kalnay 1993).

1.2.3 Volcanic Hazard Model Use

Most volcanic hazard forecasts rely on models that simulate the transportation of volcanic products (Manga et al. 2017). Forecasts are commonly based on past eruptions and their modelled impacts (Papale 2021), following the premise that past data can be used to make predictions about the future (Pyle 2018). While eruptions occurring today are often well-observed through a combination of direct observation, eyewitness reports, and remote sensing,

the completeness of historical and prehistoric eruption records is much lower. Historic eruptions may have gone unobserved or misclassified (e.g., Wilson et al. 1995), older eruptions may be missing in the geologic record due to deposition and weathering processes (e.g., Kueppers et al. 2019), and small-sized eruptions may be missed in the geologic record (Mead and Magill 2014; Damaschke et al. 2017). For example, minor eruptions of Mount Ruapehu in 2007 would be too small to appear in the geologic record (Kilgour et al. 2010) and therefore would be missed if not directly observed. Even if a complete eruption history existed for a volcano, we do not know the validity of the assumption that past behaviour informs the future (Marzocchi and Woo 2007; Whitehead and Bebbington 2021).

These challenges influence the approach taken in forecasting, which can broadly be divided into deterministic and probabilistic approaches. Deterministic forecasts are based on a single scenario and are essentially a "best guess" of what might happen, while probabilistic forecasts account for uncertainty across a large continuum of scenarios. Deterministic forecasts for volcanic hazards are not particularly achievable, as they would require knowing every aspect of a volcano's internal system (plumbing, magma properties, temperature, etc.) before the event (Marzocchi and Bebbington 2012). Despite this limitation, deterministic forecasts are still used in operational contexts (e.g., in Aotearoa New Zealand, Ecuador, and Mexico; Hurst and Davis 2017; Bernard et al. 2022; Garcia et al. 2023), primarily as they require fewer resources and can provide immediate guidance (even if not fully representative of likely outcomes) for technical monitoring personnel (Folch 2012; Garcia et al. 2023). Probabilistic forecasts, in contrast, represent a range of plausible outcomes by producing probability distributions for future events. They allow decision makers to assess relative likelihoods and explicitly account for "known unknowns" in volcanic systems (Gneiting 2008). They are primarily useful for assessing risk and guiding decisions under uncertainty (e.g., Jenkins et al. 2012; Pardini et al. 2022; Sandri et al. 2024; Williams et al. 2025).

1.2.4 Long-term Forecasts

Long-term forecasts assess eruption potential and hazards over periods of years to decades. Long-term forecasts are derived from geological and historical records, and are often used for land-use planning and risk mitigation methods (Marzocchi and Bebbington 2012). Volcanic

hazard models are used in conjunction with geologic data to develop hazard scenarios, showing relative hazards in different areas around a volcano (e.g., Neri et al. 2015b).

The development of hazard scenarios begins with the identification of plausible eruption types, magnitudes, and styles based on detailed geologic studies at a specific volcano. In most cases, eruption scenarios are developed based on a few studies of eruptions that represent common eruptive styles or relate underlying volcanic processes with analogue eruptive events (Sheldrake 2014). Each scenario generally includes an explicit timeline of eruptive events and is inherently deterministic, as the user defines specific input values needed for simulation (e.g., Scenario A has a plume height of 10 km, Scenario B has a plume height of 20 km; Macedonio, Pareschi, and Santacroce 1990; Barberi et al. 1992). Eruption scenarios become probabilistic when input parameters have a range of variability and rely on a large number of simulations (e.g., Connor et al. 2012; Sandri et al. 2016; Mastin, Van Eaton, and Schwaiger 2020). However, the user still defines lower and higher boundaries as to where the input parameters fall, often from past eruptive events that may not represent any future events.

1.2.5 Short-term Forecasts

Short-term forecasts are at the scale of hours to months (Whitehead and Bebbington 2021), and are primarily used in the context of volcanic unrest and crisis management (Marzocchi and Bebbington 2012). Compared to long-term forecasting, there is less literature on volcanic hazard models and short-term forecasting, likely because short-term hazard forecasts must contend with high levels of uncertainty with limited time to collect or assimilate relevant data.

Short-term hazard forecasting is closely intertwined with eruption forecasting. Both hazard and eruption forecasts rely on changes in monitoring data around a volcano (e.g., seismicity, deformation, gas fluxes) as the primary indicators of volcanic activity (Manga et al. 2017; Poland and Anderson 2020). Most short-term hazard forecasts are embedded within broader eruption forecasting tools, such as event trees (e.g., Marzocchi, Sandri, and Selva 2008; Lindsay et al. 2010; Selva et al. 2014; Scott et al. 2022).

While researching the differences between long- and short-term hazard forecasting, it was observed that, in many respects, there was little difference between the two. Both forecast types use scenarios based on the geologic history of the volcano (e.g., low, medium, and high eruption

sizes; Martinez Montesinos et al. 2022). For flow and projectile models, input parameters are the same for long- and short-term forecasting, as DEM and other input parameters, such as mass and flow rate, do not generally change, and cannot be measured before an event occurs. Atmospheric dispersion and deposition models, on the other hand, rely on different meteorological information for short- and long-term forecasts, which change the output of a volcanic hazard model (Folch, Costa, and Basart 2012).

1.2.6 Near-real-time Forecasts

Near-real-time forecasts (also known as syn-eruptive forecasts) are produced either immediately before or during a volcanic eruption, typically within a timeframe of minutes to hours (Manga et al. 2017; Merz et al. 2020). Unlike long- or short-term forecasts, near-real-time forecasting benefits from the fact that the eruption has already begun; therefore, ESPs (e.g., plume height, mass eruption rate, etc.) can be directly incorporated into models without the need to rely solely on pre-eruption scenarios (Engwell et al. 2024). However, these observations remain subject to uncertainty and interpretation. During the 2010 Eyjafjallajökull eruption in Iceland, remote sensing played a key role in observing ESPs. For example, LiDAR (Light Detection and Ranging) measurements were used to estimate tephra mass concentrations in the atmosphere (Gasteiger et al. 2011), while satellite imagery was used to estimate tephra volume, and weather radar was used to track plume height (Gudmundsson et al. 2012). As ESPs evolve, both forecasts and hazard maps are updated during ongoing activity (Manga et al. 2017).

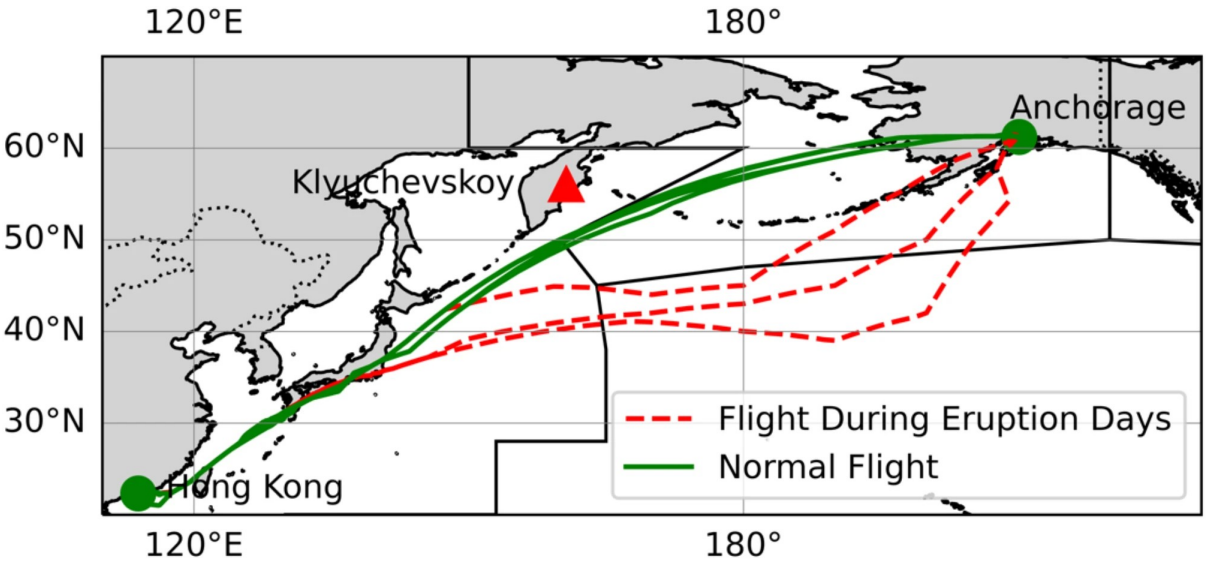


Figure 1.3: Flight data during the Klyuchevskoy eruptive activity (dashed red lines) and during times of volcanic quiescence (green lines). The black lines outline the different Volcanic Ash Advisory Centre (VAAC) areas of responsibility. Figure from Delbrel et al. (2025).

Applications in Volcanic Hazard Management

Currently, near-real-time forecasting is most commonly applied to tephra dispersal and deposition. Volcanic Ash Advisories provide flight path information in real time (Webley, Stunder, and Dean 2009; Engwell et al. 2021), resulting in diversions off normal flight routes (Figure 1.3). Near-real-time lava flow modelling is emerging, as demonstrated during the 2014/15 Fogo (Cappello et al. 2016a), 2017 Mount Etna (Cappello et al. 2019), 2021 La Palma (Marquez, Paredes, and Llorente 2022), and 2022 Mauna Loa eruptions (Hyman et al. 2024).

In the early stages of an eruption, authorities often adopt conservative strategies due to high uncertainty and low risk tolerance. Pre-defined evacuation zones are a common approach: for example, during the January 2020 Taal eruption in the Philippines, a 14 km evacuation zone was implemented (Martinez-Villegas et al. 2022); during the 2010 eruption of Mount Merapi, a 20 km zone was applied (Mei et al. 2013); and in Aotearoa New Zealand, Auckland and Taranaki Mounga have pre-defined evacuation zones of 5 km and 15 km, respectively (Tomsen et al. 2014; Horsfall 2019; Wild et al. 2021). In these scenarios, even highly accurate hazard forecasts may not meaningfully alter decision-making because authorities prioritise minimising worst-case outcomes (Marzocchi and Woo 2007; Woo 2008). However, this conservative stance may not solely reflect forecast unreliability, but rather a structural preference for precaution in the face

of epistemic uncertainty and high consequence. In some contexts, near-real-time forecasts have demonstrated operational value. For instance, lava flow forecasts were generated in near-real time during the 2021 La Palma eruption and provided information to local civil protection to contribute to the emergency management to reduce the risk of disaster (Martín-Raya et al. 2023).

Over prolonged eruptive episodes, however, societal pressure to resume normal activities or return to affected areas can increase reliance on updated hazard information (e.g., Bird, Gísladóttir, and Dominey-Howes 2011; Hicks and Few 2015). In such cases, short-term or near-real-time forecasts become vital for decision-making. For example, during the long-lived eruption of Soufrière Hills (1995 - 2013), access to Plymouth varied over time, with restrictions applied and lifted according to hazard evolution (Francis et al. 2000; Wilkinson 2015).

Challenges in Forecasting and Model Validation

Near-real-time volcanic hazard forecasting is fundamentally constrained by limited opportunities for model validation. Unlike weather forecasting, which benefits from continuous data streams and frequent events, volcanic eruptions are rare, short-lived, and often irregular. This limits the availability of observational data needed to calibrate and test models (Papale 2017) (Section 1.4). Eruptions are inherently variable, as eruptive style and intensity can shift abruptly, eruptions can pause and resume without warning, and even small meteorological changes (e.g., wind direction) and topography can drastically alter outcomes (Merz et al. 2020). While weather forecasts face similar uncertainties in initial conditions and model approximations (Ehrendorfer 1997; Bauer, Thorpe, and Brunet 2015), they benefit from near-continuous observational data that captures both system state and model outputs. Volcanic hazards lack this data continuity. For example, Mount Ruapehu has had ash-generating eruptions for only 0.03% of the time in the last 50 years (six eruptive days) (Houghton, Latter, and Hackett 1987; Cronin et al. 1997; Kilgour et al. 2010), while even the comparatively frequently active Mount Etna has erupted only ∼1% over the past five decades (around 183 days) (Global Volcanism Program 2025b).

One proposed way to address this challenge is data assimilation, i.e., the process of incorporating real-time observational data (e.g., satellite imagery) into volcanic hazard models. While widely used in numerical weather prediction (e.g., Rabier 2005; Cheng et al. 2017; Geer et al.

2018), data assimilation applications in volcanic hazard models remain limited and are primarily constrained to deterministic models (Poland and Anderson 2020), particularly volcanic ash transport and dispersion models (VATDMs). Recent studies have begun to demonstrate its potential: ensemble-based data assimilation methods have been used to assimilate satellite-derived ash concentration measurements, improving near-real-time ash dispersion forecasts at Eyjafjallajökull (Fu et al. 2017; Weng et al. 2025), Mount Etna (Pardini et al. 2020), and Kasatochi (Chai et al. 2017). Despite this progress, continuous data assimilation in volcanic hazard forecasting is still under development, owing to the complexity of eruption processes and the need for rapid assimilation cycles. This limitation highlights the value of complementary approaches, such as identifying influential input parameters (Section 1.2.7) and exploring whether constraining them with prior knowledge can enhance model performance and forecast accuracy when data assimilation is not yet operationally feasible.

1.2.7 Sensitivity Analysis

Sensitivity analysis (SA), defined as "the study of how the uncertainty of a model can be apportioned to different sources of uncertainty in the model input" (Saltelli 2002), is used to identify which input parameters most strongly influence model outputs. Two main types exist: local sensitivity analysis (LSA), which varies one input at a time while keeping others fixed, and global sensitivity analysis (GSA), which varies all inputs simultaneously across their ranges to capture nonlinear and interactive effects (Saltelli et al. 2008). While GSA is widely used in hydrology, climate, and ecological modelling (e.g., Lidén and Harlin 2000; Gupta, Liu, and Wagener 2008; Ruano et al. 2012), GSA remains underutilised in volcanology.

In volcanic hazard modelling, LSAs are relatively common (Table 1.1), likely because they are computationally inexpensive and straightforward to implement (Saltelli et al. 2004). However, only three GSAs have been conducted to date: two for VATDMs (TEPHRA and HYSPLIT) (Scollo et al. 2008; Pardini et al. 2022), and one for a lava flow model (MAGFLOW) (Bilotta et al. 2012). All three studies relied on input parameter ranges derived from specific past eruptions (Mount Ruapehu and Mount Etna), limiting their ability to evaluate model behaviour across the full range of plausible future scenarios. This makes it difficult to assess which inputs are most influential in a real forecasting context, where future conditions are unknown.

Table 1.1: Non-exhaustive list of local sensitivity analysis studies for different volcanic hazard models

	Literature	Model	Volcano Studied
Volcanic ash transport and dispersion models	Komorowski et al. (2008)	HAZMAP	1530 AD Soufrière (Guadeloupe)
	Peterson and Dean (2008)	PUFF	2000 Hekla, 2001 Cleveland, future Mount St Helens
	Scollo, Folch, and Costa (2008)	Fall3D, HAZMAP, TEPHRA	1990 and 2002/03 Mount Etna
	Webley, Stunder, and Dean (2009)	HYSPLIT and PUFF	1992 Mount Spurr
	Devenish et al. (2012), Beckett et al. (2015), Harvey et al. (2018), Osman et al. (2020), and Beckett et al. (2022)	NAME	2010 Eyjafjallajökull
	Folch, Costa, and Basart (2012)	Fall3D	2010 Eyjafjallajökull
	Kylling et al. (2015)	FLEXPART	2010 Eyjafjallajökull and 2011 Grímsvötn
	Liu et al. (2015)	FLEXPART	1996 Mount Ruapehu
	Buckland et al. (2022)	Ash3D	7.7 ka Mount Mazama
	James, Dacre, and Harvey (2024)	NAME	2019 Raikoke
Lava flow models	Rongo et al. (2008)	SCIARA	2001 Mount Etna
	Tarquini and Favalli (2013)	DOWNFLOW	Future Mount Etna
	Bilotta et al. (2019)	MAGFLOW	N/A
	Zeinalova et al. (2025)	GPUFLOW (MAGFLOW)	2015 Mount Etna
Mass flow models	Stevens, Manville, and Heron (2003)	LAHARZ	Mount Ruapehu and Taranaki Mounga (Mount Taranaki)
	Stinton et al. (2004) and Sheridan et al. (2005)	Titan2D	1963 Mount Rainier
	Rupp et al. (2006), Dalbey et al. (2008), Capra et al. (2011), and Neglia et al. (2021)	Titan2D	1991 Colima
	Charbonnier and Gertisser (2012)	Titan2D and VolcFlow	2006 Merapi
	Oramas-Dorta et al. (2012)	Titan2D	Arenal
	Stefanescu, Bursik, and Patra (2012)	Titan2D	Mammoth Mountain
	Ogburn and Calder (2017)	Titan2D, VolcFlow, LAHARZ, ΔH/L	1996–2010 Soufrière Hills
Volcanic ballistic projectile models	Tsunematsu et al. (2014)	Ballista	N/A
	Biass et al. (2016a)	Great Balls of Fire	1888–1890 La Fossa
	Bertin (2017)	3-D	1984–1993 Lascar

This gap is problematic. Without GSA applied in a forecasting context, the dominant controls on model output remain poorly quantified, and the influence of input uncertainty on forecast realism is unclear. Over a decade ago, Bonadonna et al. (2012) recommended that all volcanic ash models undergo SA to support ensemble modelling and prioritise input data collection. Yet, despite this recommendation, only one GSA study has emerged since (Pardini et al. 2022).

These gaps in understanding how input parameters influence model outputs directly motivate the second research objective of this thesis (conducting a global sensitivity analysis to identify the most influential inputs for volcanic hazard forecasting under uncertainty). Additionally, understanding input parameter sensitivities informs decisions around model choice (Objective One) and the extent to which volcano-specific knowledge should be used to constrain forecasts (Objective Three).

1.2.8 The Bias-Variance Trade-Off

A central concept for this thesis is the bias-variance trade-off, a fundamental challenge in all modelling efforts involving uncertainty (Geman, Bienenstock, and Doursat 1992). In volcanology, this trade-off is especially pronounced due to the complex, variable nature of volcanic systems (Sparks and Cashman 2017). Forecasting volcanic hazards compounds this difficulty, as models must balance the inclusion of detailed physical processes against the need to remain robust under uncertain and often incomplete input data (Papale 2021). This balance becomes even more critical for short-term and near-real-time forecasting (Section 1.2.5), where timely hazard estimates are required despite high uncertainty and limited observational data (Papale 2017).

The bias-variance trade-off underpins the accuracy of volcanic hazard forecasts. Bias refers to the difference between the average prediction of a model and the correct value being predicted, while variance describes the variability of model predictions at a given data point or value (Friedman 1997). If a model is too simple, it will underfit the data and exhibit high bias (Figure 1.4); if too complex, it will overfit and display high variance (Figure 1.4). The ideal model has low bias and low variance, but in practice, reducing one often increases the other (e.g., Dunant 2021).

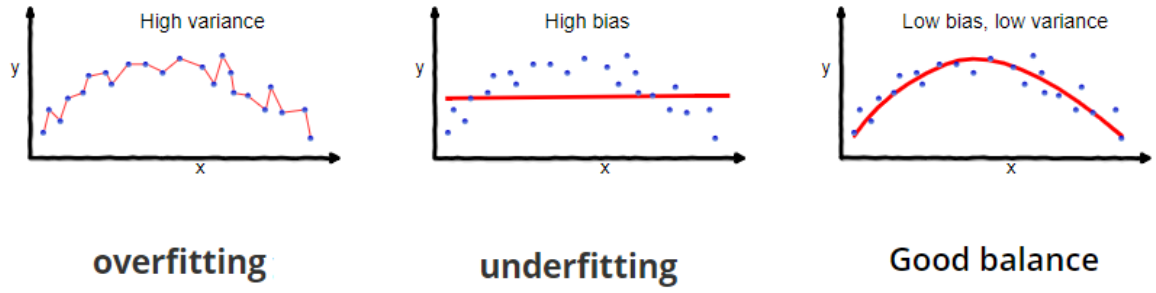


Figure 1.4: Three graphs representing an overfitted model, which has a high variance, and an underfitted model, which has a high bias; and a model that has a good bias-variance balance, which is a low bias and low variance (Singh 2018).

For volcanic hazard forecasting, this trade-off is particularly difficult to manage. Evaluating under- and over-fitting typically requires test and training data, yet for future events (which we are interested in modelling), such data do not exist. Opportunities for observational validation are far more limited than in fields such as weather forecasting (Section 1.2.6).

Managing this trade-off requires complementary strategies. GSA (Section 1.2.7) helps identify influential input parameters and parameter interactions, addressing epistemic uncertainty and guiding model simplification or refinement. While hindcasting (retrospective forecasts) provides a mechanism to evaluate model bias by comparing model outputs against historical eruptive events, it allows calibration while highlighting potential overfitting or underfitting.

1.2.9 Volcanic Ash in the Aotearoa New Zealand Context

Volcanic ash is one of the most disruptive volcanic hazards and can be transported and deposited hundreds, or even thousands of kilometres from the volcano, inhibiting air travel (Casadevall 1994), disrupting and damaging infrastructure (Wilson et al. 2012b), and can cause injury and death (Spence et al. 2005).

Volcanic ash, like many volcanic hazards, can be mitigated with the help of hazard modelling. VATDMs are used to estimate the spatial footprint of atmospheric ash concentrations, ashfall deposit mass, and other deposit characteristics. The first VATDM was proposed by Suzuki (1983), who conducted a numerical analysis on tephra diffusion and dispersion to create a theoretical model for tephra dispersion. Since then, a variety of methods have been developed to simulate and study ash transportation and deposition (Searcy, Dean, and Stringer 1998;

Draxler and Hess 1998; Bonadonna et al. 2005; Connor and Connor 2006; Costa, Macedonio, and Folch 2006). This thesis focuses on VATDMs as an exemplar of volcanic hazard models, specifically in Chapters Three, Four and Five.

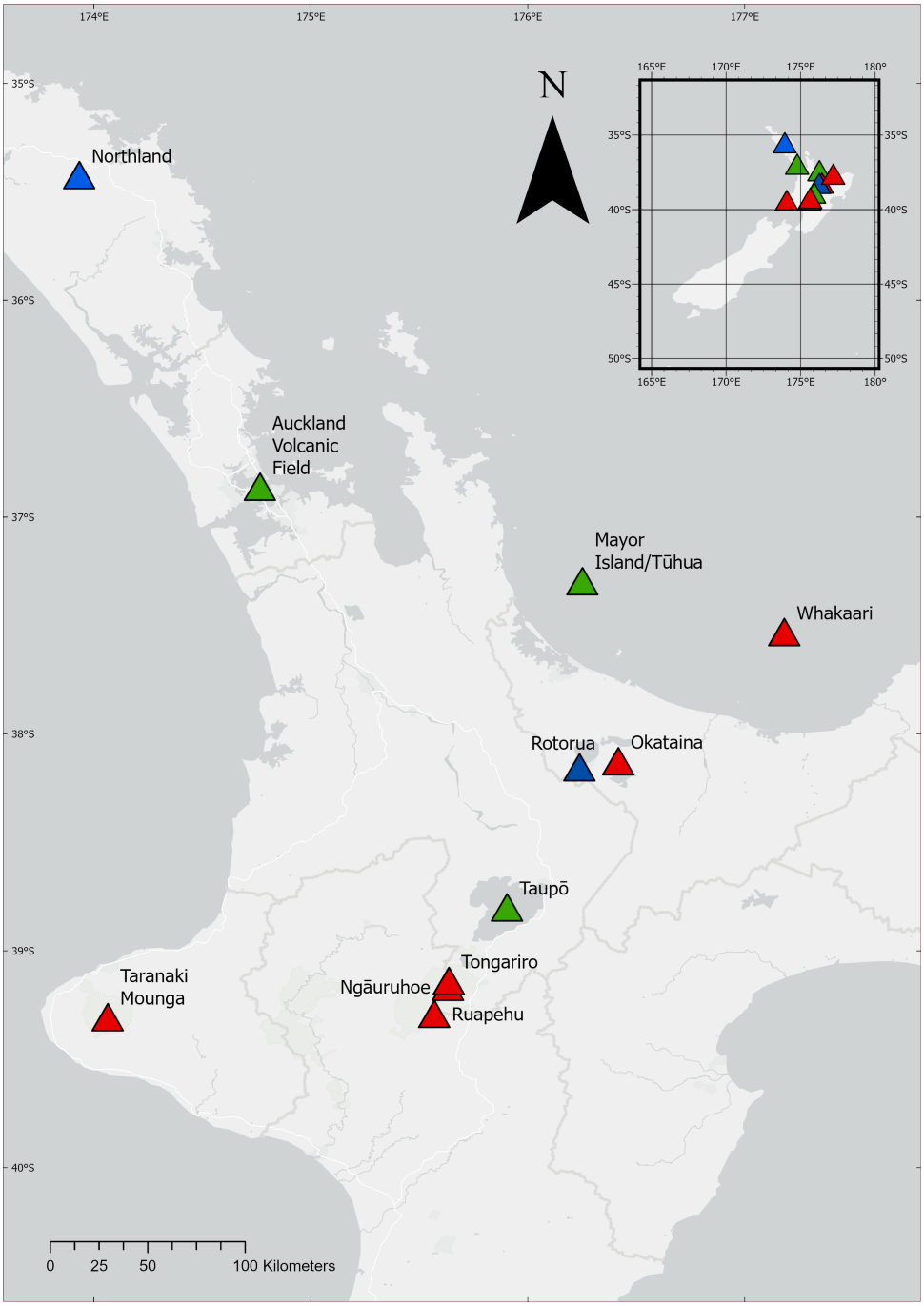


Figure 1.5: Volcanoes monitored in Aotearoa New Zealand (11 shown, the 12th is the Kermadec Islands). Those in red have erupted within the last 200 years, in green within the last 4000 years, and in blue within the last 100,000 years.

In Aotearoa New Zealand, understanding and modelling ashfall is particularly important

given the country's active volcanic landscape. Here, there are six different volcanic regions, five of which are located in the North Island (Heming 1980; Wright 1994; Wilson et al. 1995; Hopkins et al. 2020; Cronin et al. 2021). Within these volcanic regions, 12 volcanoes are currently monitored by Aotearoa New Zealand's volcano monitoring service GeoNet (within the wider Earth Sciences New Zealand) (Figure 1.5). All of these volcanoes are capable of producing widespread ashfall that can cause significant impacts to infrastructure, agriculture, and the wider economy (e.g., Thompson et al. 2017; Barker et al. 2019; McDonald et al. 2024; Weir et al. 2024; Porter et al. 2025).

Over the past 50 years, there have been seven periods of large ash-producing eruptions at three of Aotearoa New Zealand's volcanoes: Whakaari (1975 - 2000, 2012, 2016, 2019, 2022 - current) (Kilgour et al. 2021), Mount Ngāuruhoe (1973 - 1975) (Nairn and Self 1978), and Mount Ruapehu (1995/96) (Cronin et al. 1998). In the next 50 years, there is estimated to be a 30 - 50% probability of an explosive eruption at Taranaki Mounga (Turner et al. 2008; Damaschke, Cronin, and Bebbington 2018), a 10% probability of an eruption in the Auckland Volcanic Field (National Emergency Management Agency 2025), and a high likelihood of further eruptive activity at both Whakaari and Mount Ruapehu (Willis 2014; National Emergency Management Agency 2025). This thesis focuses on Taranaki Mounga and Mount Ruapehu as case study examples.

1.3 Thesis Outline

This thesis is composed of seven chapters and follows a thesis-by-publication format. Chapters Two, Three, and Four are manuscripts either published or under review. Each chapter directly addresses one or more of the three research objectives introduced in Section 1.1.

- *Chapter One* outlines the motivation for this research, situates it within the broader field of volcanic hazard forecasting, and presents the three guiding objectives.
- *Chapter Two* addresses the first objective by providing the first synthesis of computational volcanic hazard models. It provides a volcanic hazard model ontology based on model inputs, complexity, structure, testing, and treatment of assumptions. This chapter establishes the conceptual foundation necessary for understanding how different components of volcanic hazard models influence forecast accuracy (Scott, Whitehead, and Procter 2026).
- *Chapter Three* addresses the second objective by quantifying the influence of input parameters on model outputs using a GSA of two VATDMs, Tephra2 and Fall3D. The chapter identifies which input parameters exert the greatest control over simulated ash deposition, thereby clarifying which input parameters must be prioritised during heightened unrest when time for exploration of input uncertainty is limited. This analysis provides the quantitative basis for improving input parameter selection and reducing uncertainty in forecasting (Scott et al. 2025).
- *Chapter Four* builds on these results by examining the third objective. Using the 17 June 1996 eruption of Mount Ruapehu as a test case, the chapter compares VATDM forecasts produced using uninformed and informed input parameter distributions based on past data. This chapter provides a clear assessment of the bias-variance trade-off inherent in short-term and near-real-time forecasting and challenges the assumptions about the benefits of increased knowledge (Scott et al., under review).
- *Chapter Five* extends the analysis of Chapter Four by evaluating the accuracy of the modelled total grain size distribution. This chapter tests whether VATDMs can reproduce observed fine ash patterns. By showing that total grain size distribution forecasts

remain systematically limited, this chapter further addresses the third objective and reveals fundamental constraints in current forecasting and modelling practice.

- *Chapter Six* synthesises the findings of Chapters Two through Five, demonstrating how each chapter contributes to the research objectives and how, collectively, they provide a coherent answer to the overarching question of how to improve short-term volcanic hazard forecast accuracy under uncertainty. This chapter additionally outlines research limitations and recommends areas of future research that can build upon the work presented in this thesis.

- *Chapter Seven* provides the overall conclusion and key contributions of the thesis.

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the student and the student's main supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the student's contribution as indicated below in the Statement of Originality.

Student name:	Emmy Elizabeth Scott
Name and title of main supervisor:	Dr Melody Whitehead, Senior Research Officer and Senior Lecturer
In which chapter is the manuscript/published work?	Chapter Two

Describe the contribution that the student and members of the supervisory team have made to the manuscript/published work:[1]

Emmy Scott: Conceptualisation, Investigating, Writing - original draft
Melody Whitehead: Conceptualisation, Writing - review and editing
Jonathan Procter: Conceptualisation, Writing - review and editing

Please select one of the following three options:

(●) **The manuscript/published work is published or in press**

Please provide the full reference of the research output:

Scott, E., Whitehead, M., & Procter, J. (2026). Exploring the role of model classification, complexity, and selection in volcanic hazard forecasting. Computers & Geosciences, 207. https://doi.org/https://doi.org/10.1016/j.cageo.2025.106070

( ) **The manuscript is currently under review for publication**

Please provide the name of the journal:

( ) **It is intended that the manuscript will be published, but it has not yet been submitted to a journal**

Student's signature:	Emmy Scott Digitally signed by Emmy Scott Date: 2025.11.12 11:03:24 +13'00'	Main supervisor's signature:	Melody Whitehead Digitally signed by Melody Whitehead Date: 2025.11.12 12:12:25 +13'00'

This form should be placed at the beginning of each relevant thesis chapter.

[1] Refer to the Massey University Publishing and Authorship guidelines (OneMassey for staff, Stream for students) and/ or Contributor Roles Taxonomy (CRediT) guidelines for guidance.

Chapter 2

Exploring the role of model classification, complexity, and selection in volcanic hazard forecasting

Chapter One introduced the motivation and key research questions, outlining the challenges of improving volcanic hazard forecasts under uncertainty. This chapter provides a structured literature review of current volcanic hazard models, examining their structure, input requirements, complexity, and suitability for short-term forecasting, contributing to Objective One of this thesis: *Does model choice matter for volcanic hazard forecasting under uncertainty?*.

Chapter Two has been published as Scott, E., Whitehead, M., & Procter, J. (2026). Exploring the role of model classification, complexity, and selection in volcanic hazard forecasting. *Computers & Geosciences*, *207*.
https://doi.org/10.1016/j.cageo.2025.106070. (Accepted: 24 October 2025; Published: 30 October 2025).

Chapter Two reproduces the published version with only formatting modifications.

Abstract

This review examines the current landscape of computational volcanic hazard models, focusing on their creation and application, for a diverse set of end-users' short-term and long-term forecasting requirements. We provide a comprehensive classification of volcanic hazard models, categorising them according to their theoretical foundations. This is central to understanding the diversity of hazard characterisation and simulation approaches, from empirical models to computationally demanding physics-based numerical models. The classification framework helps contextualise the strengths and limitations of different models and their suitability for specific forecasting demands. We discuss the fundamental principles behind model construction, considering factors such as input parameters, conceptual frameworks, and the incorporation of uncertainties. We also synthesise existing literature on model testing, covering aspects such as model verification, validation, calibration, and benchmarking, and provide a systematic and transparent framework for model selection, considering data availability, computational constraints, and specific forecasting needs. We explore the balance between model complexity, computational efficiency, and accuracy, addressing the uncertainties inherent in both input parameters and model processes. A key focus is the role of input parameters in forecasting and the need to select models that are detailed enough to capture essential hazard dynamics, yet simple enough to minimise error and computational costs.

2.1 Introduction

Volcanoes are complex and unique systems governed by underlying non-linear dynamics (Cashman and Biggs 2014), with potentially catastrophic eruptive hazards (Brown et al. 2017). Forecasting the spatial extent and intensity of these hazards, whether in the long-term, short-term, or near-real-time, currently relies on the use of computational models (Bayarri et al. 2009). Volcanic hazard models have to deal with both the aleatoric uncertainty (natural inherent variability), as well as the epistemic uncertainties that arise from data and knowledge limitations specific to a given volcano (Bayarri et al. 2009; Rougier and Beven 2013). Together, these factors create substantial and often poorly constrained uncertainties in hazard and eruption forecasting (Sparks 2003; Poland and Anderson 2020).

Computational volcanic hazard models simulate the transport and deposition of potentially hazardous eruption products such as tephra (volcanic ash), debris flow/lahars, pyroclastic density currents (PDCs), lava flows, and volcanic ballistic projectiles (VBPs) (e.g., Folch 2012; Kavanagh, Engwell, and Martin 2018; Pardini et al. 2024). These models use empirical relationships and/or numerical physics-based equations to simplify inherently complex natural processes (Renschler 2005). Models vary in complexity, balancing trade-offs between (e.g.) computational cost, data requirements, the level of physical detail incorporated, the knowledge available for a specific volcano, and time constraints. Additionally, the degree of uncertainty that can be tolerated by the end-user, shaped by the type of decision being made, such as short-term evacuation versus long-term land-use planning, as well as the specific purpose and context of use, can influence the choice of model (Loucks and van Beek 2017; Crawford et al. 2018). While complex models are often assumed to provide more accurate forecasts than simpler models, there is little empirical evidence to support this assumption (Petropoulos et al. 2022). Green and Armstrong (2015) found that for economic models, model complexity may even harm forecast accuracy, particularly when data are limited, as complexity increases uncertainty.

Despite the increasing number of volcanic hazard models available, model selection is still highly subjective and often based purely on model familiarity or institutional preference (e.g., Hurst and Davis 2017) rather than situational suitability (Jackson, Trebitz, and Cottingham 2000; Jakeman, Letcher, and Norton 2006). This lack of clear guidance can result in the selection of models that may not be the most accurate for a given application. We evaluate how model characteristics influence forecasting performance across different timescales, from near-real-time operational contexts to long-term planning scenarios.

This review is intentionally focused on volcanic hazard simulation models - that is, models that are used once a hazard is assumed to occur. As such, the review inherently addresses conditional hazard modelling. A detailed discussion of temporal occurrence rates and hazard likelihood estimation methods is beyond the scope of the review, but the reader is directed to (e.g.,) Bebbington (2009), Marzocchi, Sandri, and Selva (2010), and Bebbington and Jenkins (2019).

This work delineates the volcanic hazard model space for end-users over five key themes. In Section 2.2, we provide an overview of model classifications and their theoretical/conceptual

foundations. Section 2.3 examines how factors such as conceptual models, model structure, input parameters, and uncertainty influence model design. Model complexity is examined in Section 2.4 in relation to its impacts on output accuracy and forecasting capability within volcanic hazard models. As the reliability and accuracy of volcanic hazard models depend on rigorous testing, Section 2.5 examines model verification, validation, calibration, and benchmarking methodologies. In Section 2.6, we discuss model selection, considering practical criteria such as data availability, computational constraints, and end-user needs, intertwined with model complexity.

2.1.1 Definitions

Terminology varies across disciplines, including within volcanology. Thus, we define here key terms for clarity (Table 2.1).

The focus here is specifically on computational models. Any reference to the word 'model' is specifically referring to computational models unless otherwise specified.

2.2 Model Classification

Models for simulating volcanic eruption phenomena originated in the 1950s, with theoretical works by Prandtl (1954) and Morton, Taylor, and Turner (1956) looking at turbulent gravitational flows. These papers became the foundations of many flow models that were developed in the 1980s (e.g., Malin and Sheridan 1982; Sheridan and Malin 1983; Armienti and Pareschi 1987; Valentine and Wohletz 1989), which were primarily used to understand and reconstruct the spatial extents and dynamics of past eruptions. Around the same time, models for ashfall and lava flow also began to emerge, e.g., ash - numerical analysis by Suzuki (1983) and Sparks (1986); and lava flows - analytical and empirical analysis by Dragoni, Bonafede, and Boschi (1986) and Young and Wadge (1990). In the 1990s and 2000s, advances in computational power and improved understanding of volcanic processes enabled the development of more complex and computationally intensive numerical models (Pitman et al. 2003; Costa et al. 2016; Papale 2021).

All volcanic hazard models are ultimately grounded in the fundamental laws of physics, including the conservation of mass, momentum, and energy (Kavanagh, Engwell, and Martin

Table 2.1: Key terms and their definitions

Term	Definition	Reference
Accuracy	A measure of closeness between a model's output and the observed or actual value	This publication
Aleatoric uncertainty	The uncertainty in eruption phenomena, observations, and data that comes from the inherent variability of a natural system	Marzocchi et al. (2004)
Benchmarking	An exercise to compare the performance of many models when simulating a specific case with a known solution	Dietterich et al. (2017)
Calibration	The manipulation of model parameters to align with observations from a specific event or dataset	Oreskes, Shrader-Frechette, and Belitz (1994)
Computational model	The use of computers to simulate volcanic systems	This publication
Conceptual framework	The definition of scope, key concepts, and theoretical foundations of a system, theory, or phenomenon	This publication
Conceptual model	A schematic or qualitative description of a system, theory, or phenomenon and its components or relationships	Thalheim (2011)
Deterministic model	A model that produces an individual point measure for an event	Manga et al. (2017)
Empirical model	A model developed from observational data (these may include statistical models)	This publication
Epistemic uncertainty	The uncertainty in eruption phenomena, observations, and data that comes from incomplete knowledge or conceptual limitations	Marzocchi et al. (2004)
Input parameter	Data provided to the model, used to drive model behaviour	This publication
Long-term forecast	An assessment of the eruption likelihood and related phenomena over the scale of years to decades	Marzocchi and Bebbington (2012)
Model	A computational representation of volcanic eruption phenomena	This publication
Near-real time forecast	An assessment of the eruption likelihood and related phenomena for the immediate future (also known as syn-eruptive forecasts)	This publication
Numerical model	A model that solves governing equations using numerical methods (these may include physical and analytical models)	Larson (2005)
Ontology	A formal, unambiguous description of domain-specific knowledge, providing standardised definitions for terminology within a given field	Hofmann, Palii, and Mihelcic (2011) and Masseroli (2019)
Output	Refers to the results produced by the model	This publication
Probabilistic model	A model that produces a range of possible outcomes, often expressed as distributions	Manga et al. (2017)
Short-term forecast	An assessment of the eruption likelihood and related phenomena over the scale of hours to months	Marzocchi and Bebbington (2012)
Simulation	The execution of a model within a computational environment	This publication
Validation	The demonstration that a system of equations and their numerical approximation reasonably represents the physical conceptual model for the real-world process it is intended to simulate	Esposti Ongaro et al. (2020)
Verification	The demonstration that a model implementation accurately represents the conceptual description and solution of the model	Oberkampf and Trucano (2002)

2018). These laws underpin the governing equations that define volcanic flow dynamics, atmospheric dispersion, particle settling, and other key processes, even when simplifications are applied due to model structure, computational demand, or data constraints.

Volcanic hazard models (Table 2.2) can broadly be defined into three categories based on the governing force for the phenomena being modelled: (1) Atmospheric/weather driven models for volcanic ash and gases, (2) Gravity-driven models for lava flows, pyroclastic density currents (including surges and block-and-ash flows), lahars (including debris flows), and debris avalanches, and (3) Momentum-driven models for projectile motions such as volcanic ballistic projectiles (VBPs).

2.2.1 Atmospheric Dispersion and Deposition Models

Atmospheric dispersion and deposition models simulate how volcanic ash and gases are transported through and settle from the atmosphere. Their primary outputs are atmospheric particle concentrations and/or mass loading on the ground. Both tephra and volcanic gas dispersal are governed by the advection-diffusion-sedimentation (ADS) equation, which governs the movement of particles in the atmosphere (Tsunematsu et al. 2011). Dimensionality plays a key role in input parameter requirements, with higher-dimensional models typically requiring more detailed input parameter data (Cao et al. 2021).

Two-dimensional (2D) Eulerian models, such as Tephra2 and HAZMAP, apply a simplified Gaussian solution to the ADS equation. These models assume a constant diffusion rate and wind fields that may vary with height but are spatially uniform across the computational domain, which limits their accuracy for weak plumes, long-range dispersal, or changing meteorology (Pfeiffer, Costa, and Macedonio 2005; Folch 2012; Bonadonna et al. 2015).

Three-dimensional (3D) models offer enhanced spatial resolution and the ability to simulate more complex atmospheric interactions compared to 2D models. Lagrangian models (e.g., PUFF, NAME) track particles individually via random-walk diffusion (Schwaiger, Denlinger, and Mastin 2012), while Eulerian models (e.g., Fall3D, Ash3D) solve the full ADS equation on a fixed grid. These models can resolve detailed wind and turbulence interactions but are computationally intensive (Folch 2012).

Table 2.2: Examples list of different volcanic hazard models and their characteristics. Dimension: x = horizontal coordinate along east-west axis, y = horizontal coordinate in north-south axis, z = vertical coordinate (elevation or height in 3D space), t = time. Model type: E = Empirical, N = Numerical. Frame of reference: E = Eulerian, L = Lagrangian, H = Hybrid. Model Approach: D = Deterministic, P = Probabilistic.

Volcanic Hazard Model	Dimension	Model Type	Frame of Reference	Model Approach	Volcanic Phenomena
Atmospheric Dispersion and Deposition Models					
HYSPLIT (Draxler and Hess 1997)	3D (x,y,z,t)	N	H	D	Tephra, Gas
PUFF (Searcy, Dean, and Stringer 1998)	3D (x,y,z,t)	N	L	D	Tephra
HAZMAP (Macedonio, Costa, and Longo 2005)	2D (x,y,z)	N	E	D	Tephra
NAME III (Jones et al. 2007)	3D (x,y,z,t)	N	L	D	Tephra, Gas
Tephra2 (Connor, Connor, and Bonadonna 2008)	2D (x,y,z)	N	E	D	Tephra
Fall3D (Folch, Costa, and Macedonio 2009)	3D (x,y,z,t)	N	E	D	Tephra, Gas
Ash3D (Schwaiger, Denlinger, and Mastin 2012)	3D (x,y,z,t)	N	E	D	Tephra
Volcanic Flow Models					
Energy Cone (ΔH/L) (Malin and Sheridan 1982)	1D (x,y)	E	E	D	Mass Flows
SCIARA (Crisci et al. 1986)	2D (x,y,z)	N	E	D	Lava
FLOW-3D (Kover 1995)	3D (x,y,z,t)	N	E	D	Lava, Mass Flows
LaharZ (Schilling 1998)	1D (x,y)	E	E	D	Mass Flows
FLOWGO (Harris and Rowland 2001)	1D (x,y,z,t)	N	L	D	Lava
PDAC (Neri et al. 2003; Esposti Ongaro et al. 2007)	2D/3D (x,y,z,t)	N	E	D	Mass Flows
DOWNFLOW (Favalli et al. 2005)	1D (x,y)	N	E	P	Lava
Titan2D (Patra et al. 2005)	2D (x,y,z,t)	N	E	D	Mass Flows
MAGFLOW (Vicari et al. 2007)	2D (x,y,z)	N	E	D	Lava
VolcFlow (Kelfoun and Vargas 2016)	2D (x,y,z,t)	N	E	D	Lava, Mass Flows
MOLASSES (Richardson 2016)	2D (x,y,z)	N	E	D	Lava
MrLavaLoba (de'Michieli Vitturi and Tarquini 2018)	2D (x,y,z)	E	L	P	Lava
IMEX_SfloW2D (de'Michieli Vitturi et al. 2019)	2D (x,y,z,t)	N	E	D	Mass Flows
EC_MapProb & **Box_MapProb** (Aravena et al. 2022)	1D (x,y)	N	E	P	Mass Flows
Volcanic Ballistic Projectile (VBP) Models					
Eject! (Mastin 2001)	2D (x,y,)	N	L	D	VBP
LPAC (de'Michieli Vitturi et al. 2010)	2D (x,y,t)	N	L	D	VBP
Alatorre-Ibargüengoitia, Delgado-Granados, and Dingwell (2012)	2D (x,y)	N	L	D	VBP
Ballista (Tsunematsu et al. 2014)	3D (x,y,z)	N	L	D	VBP
Great Balls of Fire (Biass et al. 2016a)	2D (x,y,t)	N	L	D	VBP

Output accuracy across all models depends heavily on the quality of eruption source parameters (ESPs) and meteorological input data (primarily wind direction and speed) (Bursik 2001; Bonadonna and Costa 2013). 2D models typically use static wind profiles, while 3D models use full meteorological datasets (e.g., Numerical Weather Prediction (Haupt et al. 2017) and observational reanalysis (Hersbach et al. 2020)). Atmospheric dispersion models always require ESPs such as plume height, total erupted mass, and grain size distribution, often estimated from deposit mapping (e.g., Longchamp et al. 2011; Bonadonna and Costa 2012), remote sensing (both after the fact and in real-time) (e.g., Prata and Grant 2001), or inversion techniques (e.g., Volentik et al. 2010; Mannen 2014).

2.2.2 Flow Models

Volcanic flows are gravity-driven and strongly influenced by topography (e.g., obstacles, breaks, surface roughness) (Branney and Kokelaar 2002; Sulpizio and Dellino 2008; Dietterich et al. 2015). 1D flow models simulate channelised flows constrained to a single downslope direction, typically to estimate runout distance. Simple 1D models use digital elevation models (DEMs), and sometimes volume, as input parameters (Schilling 1998). These models range in complexity: some, such as DOWNFLOW, incorporate probabilistic input distributions to account for uncertainty, while others, such as FLOWGO, include detailed rheological properties such as cooling, viscosity changes, and crystallisation (Harris and Rowland 2001).

2D models, particularly cellular automata models (e.g., MAGFLOW, SCIARA, MOLASSES), are widely used for probabilistic hazard assessments (Herault et al. 2009; Gallant 2016). Cellular automata models use a simple set of rules (often not derived directly from physical laws) to distribute fluids from a central cell (i.e. a grid point) to a neighbouring one, depending on the cell's relative elevation and sometimes flow viscosity and temperature, until all cells meet the criteria defined by the model (e.g., when the total volume defined by the user has been added to the flow) (Kavanagh, Engwell, and Martin 2018; Hyman, Dietterich, and Patrick 2022). They perform well on gentle terrain or in depressions, but lack vertical detail and mainly estimate deposit extent (Cordonnier, Lev, and Garel 2015).

Depth-averaged models approximate shallow water equations by averaging flow properties (energy, mass, momentum) over depth (e.g., Titan2D, VolcFlow) (Cordonnier, Lev, and Garel

2015; Hyman, Dietterich, and Patrick 2022). These models are suited for simulating flow dynamics such as arrival times, especially for concentrated PDCs where flow properties remain relatively uniform, but are slower to run as they involve more detailed computations (Dufek 2016; Hyman, Dietterich, and Patrick 2022).

3D models are thermodynamically coupled models that use computational fluid dynamics (CFD) principles to simulate heat transfer and account for various viscous flow behaviours, incorporating the complex interactions between temperature, velocity, and viscosity (Neri et al. 2015a). These models have higher computational requirements but are particularly suitable for long-lived, cooling-limited lava flows (when the rate of cooling and solidification is the primary control of the flow's length) (Dietterich et al. 2017).

All volcanic flow models use DEMs, but for hindcasting, modern DEMs must be adjusted to represent pre-eruption terrain (Procter et al. 2010b). This can be worked around, although impractically, by subtracting thicknesses of past flow deposits based on detailed stratigraphy of the deposit (e.g., Daag 2003; Dietterich et al. 2021). For forecasting, using a DEM of the current landscape is generally more applicable. However, this approach becomes challenging during multi-flow events, where DEMs must be dynamically updated to improve simulation accuracy (e.g., De Beni, Cantarero, and Messina 2019). DEM resolution (both grid size and vertical accuracy) also significantly affects model simulations, especially for small-scale features that influence flow paths (Huggel et al. 2008; Joyce et al. 2009). For example, in simulations of mass flows at Nevado del Ruiz, Colombia, a 10 m DEM successfully constrained flows within steep valley walls (∼70 m relief), whereas a coarser 30 m DEM resulted in unrealistic lateral spread due to its inability to resolve sharp topographic boundaries (Deng et al. 2019).

Typical model inputs include volume, effusion rate, viscosity, and density. Lava flow viscosity is a key factor in determining the shape and extent of the flow. Lava viscosity can be measured through laboratory measurements (e.g., Sehlke et al. 2014; Kolzenburg et al. 2016), estimated in the field (e.g., Chevrel et al. 2018; Chevrel, Pinkerton, and Harris 2019), and/or inferred from empirical relationships based on temperature (e.g., Pinkerton and Norton 1995; Ishihara, Iguchi, and Kamo 1990; Giordano and Dingwell 2003). PDCs and lahars have more complex, variable rheologies, which are harder to model. PDCs range from dense pyroclastic flows (which are topographically controlled) to dilute surges (which are capable of overriding obstacles) (Sparks

1976; Burgisser and Bergantz 2002). Their behaviour is poorly understood due to complexities in gas-particle interactions, flow mechanics, and generation processes (Lube et al. 2020; Jones et al. 2024). Lahars are water-saturated flows categorised by high densities and velocities. Their properties, such as particle concentration, temperature, and bulk rheology, vary significantly in space and time due to processes such as flow bulking (erosion and incorporation of debris) and debulking (Thouret et al. 2020). The rheological properties for both PDCs and lahar are derived from laboratory experiments (e.g., Roche et al. 2004; Lube et al. 2015; Jones et al. 2024), field-based depositional analyses (e.g., Sparks 1976; Branney and Kokelaar 2002; Dumaisnil et al. 2010; Bernard et al. 2014), and remote sensing techniques (e.g., Kumagai et al. 2009; Bosa et al. 2021; Macorps 2021; Bosa et al. 2024), while direct studies on the rheology of propagating PDCs are sparse (Sulpizio et al. 2014; Delannay et al. 2017).

2.2.3 Volcanic Ballistic Projectile Models

Volcanic ballistic projectiles (VBPs) range from a few centimetres to tens of metres in diameter and are large enough to follow ballistic trajectories through the atmosphere. Most VBP models are 2D and use simplified gravity and drag equations (e.g., Eject!, Great Balls of Fire, Alatorre-Ibargüengoitia, Delgado-Granados, and Dingwell 2012) (Table 2.2). The LPAC VBP model differs by applying the Basset-Boussenesq-Ossen equation to calculate Lagrangian particle acceleration under set assumptions (e.g., constant drag coefficient, no lift forces) (Maxey and Riley 1983; Crowe et al. 1998).

The Ballista model (Tsunematsu et al. 2016) is one of the few 3D VBP models, considering multiple particles and their collisions in 3D space (Tsunematsu et al. 2014). Ejected particles follow a parabolic trajectory until they reach the ground or collide with another airborne particle (Tsunematsu et al. 2014).

All VBP models require inputs for block properties (e.g., density, size) and ejection parameters (e.g., initial velocity, angle), which are either measured (e.g., Alatorre-Ibargüengoitia, Delgado-Granados, and Dingwell 2012; Fitzgerald et al. 2014; Tsunematsu et al. 2016) or assumed (e.g., Chouet, Hamisevicz, and McGetchin 1974; Patrick et al. 2007; de'Michieli Vitturi et al. 2010; Kilgour et al. 2010). Because the travel distance of projectiles is primarily influenced by ejection parameters, these are often estimated using short-term exposure time (e.g.,

Chouet, Hamisevicz, and McGetchin 1974; Blackburn, Wilson, and Sparks 1976; Ripepe, Rossi, and Saccorotti 1993; Edwards et al. 2017) or thermal videos (e.g., Patrick et al. 2007; Capponi et al. 2016).

2.3 Model Construction

Model construction has been well defined across various fields, including ecology (e.g., Jackson, Trebitz, and Cottingham 2000; Grimm et al. 2014) and hydrology (e.g., Gupta, Liu, and Wagener 2008; Clark and Kavetski 2010; Gupta et al. 2012; Anderson, Woessner, and Hunt 2015). In the context of volcanic hazards, Renschler (2005) highlights the importance of structuring model development to balance environmental processes, data limitations, and uncertainty. Here we apply the following five defining principles to volcanic hazard models: (1) identifying model purpose and scope, (2) building a conceptual model, (3) formulating model structure, (4) defining model input parameters, and (5) model documentation and evaluation, as outlined in previous studies.

2.3.1 Model Purpose and Scope

Volcanic phenomena models generally fall into two overlapping categories: (1) process-based models, that aim to improve understanding of the physical processes governing volcanic phenomena (e.g., plume dynamics, flow rheology) (e.g., Mastin 2007; Piombo and Dragoni 2009; Hoffman et al. 2023), and (2) hazard models, that aim to simulate the spatial extent and intensity of volcanic phenomena (e.g., ashfall distribution, lahar inundation, lava flow paths) (Table 2.2). While hazard models often incorporate simplified process-based components, their primary goal is to support forecasting and decision-making. For example, users can integrate the 1D model FPLUME (Folch, Costa, and Macedonio 2016), an eruption column model based on buoyant plume theory (BPT), to calculate the vertical distribution of mass within an eruption column, which can then be used to estimate the atmospheric or deposited mass of ash.

When defining a model's scope, developers must consider both the intended outputs and the target end-users (Renschler 2005; Jakeman, Letcher, and Norton 2006; Topcu and Mesmer 2018). Process-based models typically produce outputs that contribute to advancing scientific knowledge (e.g., Mastin 2007; Cerminara, Esposti Ongaro, and Berselli 2016; Folch, Costa,

and Macedonio 2016; Dioguardi and Mele 2018), whereas hazard models generate outputs that inform decision-making, such as hazard assessments, mitigation strategies, and evacuations (Costa and Macedonio 2005; Folch 2012). In particular, developers must carefully consider the model's temporal and spatial scope, scale, and resolution, particularly for models designed to support decision-making (Jakeman, Letcher, and Norton 2006).

Hazard model outputs can be communicated to decision-makers in ways that require minimal understanding of volcanic processes or the model itself, which can introduce significant risks (Barclay et al. 2008; Procter et al. 2021). Engaging end-users in the development process allows for valuable feedback on model assumptions and limitations, ultimately improving model reliability and usability (Jakeman, Letcher, and Norton 2006; Robinson and Brooks 2024).

2.3.2 Conceptual Model

Once the scope and purpose of the model is defined, the system being modelled must be conceptualised (Banks 1999; Seppelt 2003; Thalheim 2019). This includes the system's physical boundaries, such as computational domain limits (e.g., simulations will not run over areas larger than 100 km^2) and atmospheric boundaries (e.g., processes above 70 km may not be simulated), as well as physical and behavioural laws that must be obeyed (e.g., momentum, conservation of mass) (Seppelt 2003; Gupta, Liu, and Wagener 2008). The conceptual model allows developers to determine which variables should be included, the level of detail required, and the role of prior knowledge and assumptions about different processes (Section 2.3.3) (Butts et al. 2004; Jakeman, Letcher, and Norton 2006; Robinson and Brooks 2024). This process initially begins with qualitative questions such as: *what is known about the system? Are there available data, instrumentation, or monitoring that can be used to quantify aspects of the system?*

In volcanic hazard modelling, many computational models are built upon conceptual models developed in the late 20th century, such as those for eruption plume dynamics (Suzuki 1983; Woods 1988; Holasek, Woods, and Self 1996), pyroclastic flow dynamics (Fisher 1966; Sparks 1976; Branney and Kokelaar 1992), and lava flow dynamics (Walker 1973; Dragoni, Bonafede, and Boschi 1986). Conceptual models are continually updated as new data and understanding become available (Butts et al. 2004; Anderson, Woessner, and Hunt 2015), including during model calibration (Section 2.5.3) and sensitivity analysis.

2.3.3 Model Structure

Developers must translate the features identified in the conceptual model into specific mathematical representations to build the model structure. This begins with selecting governing equations that represent the relevant process dynamics and their interactions (Gupta et al. 2012) - such as the ADS equation used in atmospheric dispersion and deposition models (Folch 2012; Bonadonna et al. 2015), or the Navier-Stokes equations for flow models (Cordonnier, Lev, and Garel 2015). While these governing equations are typically dictated by the physical processes involved, model developers must make critical decisions about how to simplify, structure, or adapt these equations that balance accuracy, computational efficiency, and available data.

These decisions give rise to different model structures (e.g., empirical or numerical), each offering different levels of complexity (Section 2.4.2). Even within the same model family, structures may vary in detail, with more complex models incorporating additional mechanisms (e.g., particle aggregation in ash models, Folch et al. 2010; Hoffman et al. 2023).

In practice, model structure is often influenced by factors beyond scientific rationale, including institutional preference, model familiarity, and established acceptance within the technical community (Jakeman, Letcher, and Norton 2006; Jackson, Trebitz, and Cottingham 2000). Advances in high-performance computing and model optimisation are helping to reduce some of these limitations, enabling more complex models to be used in ensemble or probabilistic frameworks (e.g., Folch, Mingari, and Prata 2022; Martinez Montesinos et al. 2022; Massaro et al. 2023; Hyman et al. 2024). However, careful structural choices remain essential, especially when short-term and near-real-time forecasts are generated quickly and with limited input parameter data.

2.3.4 Model Inputs

Determining how model input parameters will be estimated occurs alongside the development of model structure (Jakeman, Letcher, and Norton 2006). As the model structure is developed, it must be translated into a discretised form suitable for computer implementation. Discretisation not only affects model fidelity and computational demand, but also shapes which input parameters are required, how they are estimated, and what can be held constant or inferred from others (Seppelt 2003; Jakeman, Letcher, and Norton 2006).

In atmospheric dispersion and deposition models, for instance, ESPs such as plume height, duration, and particle characteristics (i.e., size, shape) define the initial conditions for simulations (Engwell et al. 2024). Other input parameters, such as the vertical mass distribution in the plume or diffusion coefficients, arise from how the governing equations (e.g., the ADS equation) are simplified and parameterised (Bonadonna et al. 2005; Macedonio, Costa, and Longo 2005; Pfeiffer, Costa, and Macedonio 2005; Schwaiger, Denlinger, and Mastin 2012). Similarly, in flow models, discretised forms of the momentum equations require specification of parameters such as basal friction angle, which controls when and how a flow will begin to move downslope (Procter et al. 2010b).

The process of estimating these parameters differs depending on the forecasting stage. After an event, observed deposits can be used to retrospectively calibrate or validate input parameters (Section 2.5.5). Before an event, the model relies on past eruptive data at a volcano (e.g., Barsotti et al. 2018; Spiller et al. 2020; Titos et al. 2022), analogous event/volcano data (e.g., Sheldrake 2014; Clarke et al. 2020; Tennant et al. 2021) and/or expert elicitation/opinion (e.g., Tadini et al. 2021; Bernard et al. 2024). During an event, real-time data can constrain key input parameters (e.g., Vicari et al. 2011; Scollo et al. 2019; Hyman, Dietterich, and Patrick 2022; Pardini et al. 2022).

2.3.5 Model Uncertainty

Uncertainties in model development occur at every stage of model construction and should be explicitly acknowledged. A commonly adopted framework in volcanic hazard modelling is to classify these uncertainties into two broad types: epistemic and aleatoric. Epistemic uncertainty is often seen as reducible through better data or understanding, while aleatoric uncertainty is considered irreducible and should be managed in a fully structured, probabilistic manner (Marzocchi et al. 2004; Beven et al. 2018; Marzocchi, Selva, and Jordan 2021). Some recent probabilistic hazard modelling approaches explicitly address both types of uncertainty using doubly stochastic frameworks (e.g., Neri et al. 2015b; Bevilacqua 2016), where epistemic and aleatoric uncertainties are separately characterised and propagated.

However, this distinction has been criticised for oversimplifying the nature of uncertainty in complex natural systems. In practice, what is initially characterised as aleatoric may later

be better understood and reclassified as epistemic, highlighting the limitations of a rigid dichotomy. As discussed by Marzocchi and Jordan (2014) and Marzocchi, Selva, and Jordan (2021), such classifications are model-dependent, and the reducibility of uncertainty often reflects the assumptions and scope of the model itself rather than any intrinsic property of the volcanic system. They further propose that a more robust framework includes a third category, ontological error, that accounts for the unknown unknowns: behaviours of the system that fall entirely outside of the model's representational assumptions.

A critical component of model development and evaluation, particularly in forecasting, is the quantification of uncertainty in both model structure and input parameters. This is especially important when forecasts are used to inform emergency management personnel responsible for interpreting hazard information (Renschler 2005; Bayarri et al. 2009; Bonadonna et al. 2012; Folch 2012). Model users, whether they are scientists, policymakers, or emergency managers, must be informed about the limitations and assumptions in both the model and its input parameters (Sparks et al. 2013). Failing to communicate these uncertainties can lead to inappropriate or ineffective decision-making (Renschler 2005). For operational users in particular, understanding the order of magnitude of uncertainty is often more useful than knowing the precise value of a model output (Gupta, Liu, and Wagener 2008; Crawford et al. 2018).

These uncertainties relate to both the model's accuracy in representing physical processes and its ability to forecast. For example, technical uncertainty (a subset of epistemic uncertainty) includes not only potential coding errors (Walker et al. 2003), but also unavoidable approximations and assumed relationships that may not always hold (e.g., grain size - magma viscosity relationships, Costa, Pioli, and Bonadonna 2016). These uncertainties can be further unpacked by identifying where they arise within the modelling process. In volcanic hazard modelling, key sources of uncertainty include:

- **Governing equations**: physical processes are often represented using idealised or simplified equations (Section 2.2), but the extent to which these fully capture the real behaviour of volcanic phenomena, especially under varying conditions, is uncertain.
- **Model resolution (spatial and temporal)**: models must be discretised in time and space for computational implementation, yet coarse resolutions may fail to capture key dynamics, while fine resolutions are computationally expensive and may still not guarantee

accuracy (Section 2.2).

- **Input data**: key input parameters may be poorly constrained, especially for long- and short-term forecasting (Section 2.3.4). Epistemic uncertainty dominates when such input parameters are based on limited or past observations. Past data is often incomplete, as historic eruptions may be unobserved or misclassified (e.g., Wilson et al. 1995), pre-historic eruptions may not exist in the geologic record due to deposition and weathering processes (e.g., Kueppers et al. 2019), and small sized eruptions may be missed in the geologic record (Mead and Magill 2014). The certainty of direct observations can also be affected by many factors, such as bad weather conditions, low frequency of satellite passages, poor vertical resolution (radar measurements), or hazardous conditions (e.g., Prejean and Brodsky 2011; Oddsson et al. 2012; Dürig et al. 2018; Lube et al. 2020).

- **Input parameter relationships**: there are some key input parameter relationships, such as the linear relationship between plume height and total erupted mass/mass eruption rate (Mastin et al. 2009; Aubry et al. 2022), that are often incorporated into volcanic hazard modelling. These types of relationships are complicated by a variety of real-world factors, including but not limited to total grain size distribution, wind velocity, the plume water content, and aggregation (e.g., Degruyter and Bonadonna 2012; Girault et al. 2014).

- **Volcano and eruption heterogeneity**: volcanic systems are heterogeneous, and model assumptions calibrated on one volcano or eruption may not apply to another volcano or eruption (Cashman and Biggs 2014). Site-specific behaviours (e.g., unique conduit geometries, wind profiles, terrain) introduce both epistemic and aleatoric uncertainty, challenging generalisation, and transferability.

- **Plausible input parameter values**: while model developers often assume that users will input physically realistic values, some lack internal checks to prevent implausible values. For example, atmospheric dispersion and deposition model Tephra2 allows users to apply negative plume heights (intentionally or otherwise) without producing an explicit error. This introduces technical uncertainty as the model may proceed with invalid assumptions and generate misleading outputs.

2.3.6 Model Documentation

A critical aspect of model construction is documentation, which should be maintained throughout every stage of development. Comprehensive documentation provides a clear rationale for model selection, including the chosen model family (e.g., empirical vs. numerical), structural features (e.g., governing equations), and input parameter data sources. This ensures that model assumptions and design choices are transparent to both current users and future developers.

Volcanic hazard models are typically well-documented, detailing their utility, assumptions, limitations, and areas for potential improvement (e.g., de'Michieli Vitturi et al. 2010; Mastin et al. 2013; Kelfoun and Vargas 2016; Folch et al. 2020). In operational or decision-making contexts, it helps that users understand model constraints and do not misinterpret outputs.

The documentation process should continue during model testing (Section 2.5). This includes during recording test cases, software version upgrades, and any model refinements based on performance or expert review. Thorough documentation also supports the identification of the provenance of unexpected behaviours or errors, particularly when models are developed in open-source platforms such as GitHub (`https://github.com/`).

2.4 Model Complexity

Research into model complexity is vast, yet no strict or unified definitions exist within or across different disciplines (Brooks and Tobias 1996; Guthke 2017; Höge, Wöhling, and Nowak 2018; Baartman et al. 2020). Various recurring themes emerge in definitions of complexity, such as an increased number of mechanisms or processes (e.g., Bal and Rein 2013; Larsen et al. 2016; Baatz et al. 2018; Getz et al. 2018), an increased number of input parameters (e.g., Rickles, Hawe, and Shiell 2007; Bal and Rein 2013), and greater computational or data requirements (e.g., Zeigler, Muzy, and Kofman 2019).

In the context of forecasting, especially for volcanic hazards, understanding model complexity is crucial. While volcanic systems are inherently complex, this does not mean that models used to forecast hazardous behaviour must be equally complex. While it is often assumed that more complex models produce more accurate forecasts, evidence from environmental modelling suggests that increased complexity can reduce accuracy due to higher error and uncertainty (Oreskes, Shrader-Frechette, and Belitz 1994; Snowling and Kramer 2001; Srikrishnan and

Keller 2021). This can be partly explained by the bias-variance trade-off, a core issue in model development. Overly simplistic models are prone to high bias, defined as the difference between a model's average prediction and the true value, while overly complex models are prone to high variance, which captures the spread of predictions around the true value (Friedman 1997).

A recurring challenge in forecasting is balancing model complexity with the inherent uncertainties in both input parameters and model processes. Ultimately, the primary outputs required for volcanic hazard forecasting are spatial extent and hazard intensity (e.g., ash density at a location, lahar inundation thickness). If both simple and complex models yield comparable outputs, and unless input parameters are well constrained (which is next to impossible), increasing complexity by adding more parameters does not provide a meaningful advantage for forecasting (Scollo, Folch, and Costa 2008). However, complex models remain essential in research and hindcasting (Sparks and Aspinall 2004). Achieving the right balance between simplicity and complexity is essential in volcanic hazard forecasting; however, this challenge is not unique to volcanology, as similar issues arise in weather and climate forecasting (Scher and Messori 2019), hydrology (Doherty and Christensen 2011), and ecology (Evans et al. 2013; Ward et al. 2014). Ultimately, selecting a model that is both reliable and appropriate for its intended use requires careful evaluation of complexity in terms of structure, data requirements, and usability (Höge, Wöhling, and Nowak 2018).

The following sections explore how model complexity influences volcanic hazard modelling, both conceptually and practically, and highlight the trade-offs between realism, usability, and computational capacity.

2.4.1 Volcanic Hazard Model Complexity

A recent study by Malmborg et al. (2024) presents a comprehensive framework for defining model complexity in ecological models, which serves as the basis for defining complexity in volcanic hazard models. Similarly to volcanic systems, ecological systems exhibit inherent complexity (Levin 1998; May 2019; Riva et al. 2023), yet the term "complex" remains inconsistently used across studies (Malmborg et al. 2024).

Following Malmborg et al. (2024), model complexity can be defined into four categories:

- **Model Class Complexity:** The mathematical framework that defines a model's structure and governs the interaction between its input parameters and outputs.
- **Input Parameter Complexity**: Compromises of several components of complexity, including the number of input parameters in a model and input parameter-parameter relationships.
- **Input Data Complexity**: Refers to the resolution, variability, and uncertainty of the data used to populate model input parameters. This includes how difficult the data are to obtain, process, or interpret.
- **Computational Complexity**: Includes concepts such as the computing resources required to run a model and the time required to complete a model simulation.

These categories together form a conceptual framework to facilitate a systematic assessment of a given model's outputs and complexities is applied here to volcanic hazard models.

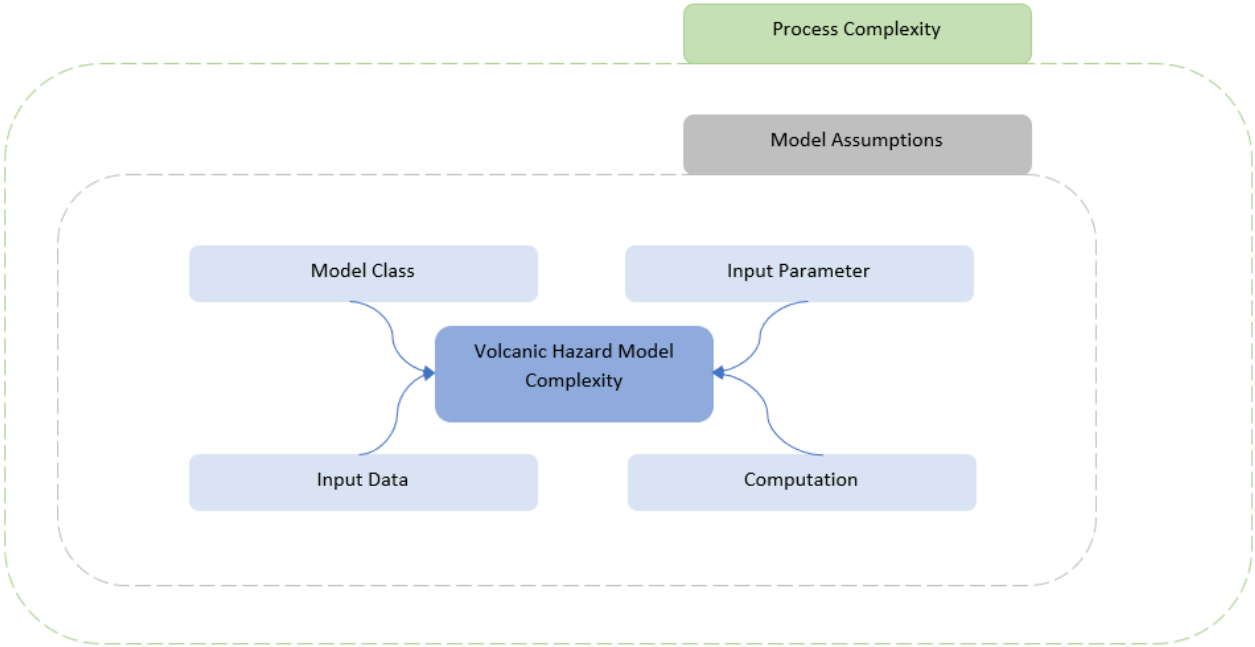


Figure 2.1: Volcanic hazard model complexity can be decomposed into four primary categories: model class complexity, input parameter complexity, input data complexity, and computation complexity. Model assumptions influence the four facets of model complexity. Process complexity refers to the complexity of the volcanic hazard being modelled, which inherently defines and shapes the facets of model complexity.

Table 2.3: Different indicators of model complexity in volcanic hazard model literature.

What makes models complex?	Directly Referenced	Indirectly Referenced
Number of parameters	Volcanic ballistic projectile model[1]	Hazard mapping[2], General hazard modelling[3,4]
Variability each input parameter introduces	Volcanic ballistic projectile model[1]	
Dimensionality		Tephra dispersion and deposition models[5,6,7], Lava flow model[8], Mass flow model[9]
Run time		Lava flow model[8], Mass flow model[9], Hazard assessment[10], General hazard modelling[4]
Accounting for multiple physical process		Mass flow models[9,11,12,13]
Computational Requirements		Hazard mapping[2]

[1] Núñez-Corrales and Brenes-André (2023), [2] Felpeto, Martí, and Ortiz (2007), [3] Sparks and Aspinall (2004), [4] Douglas (2007), [5] Scollo, Folch, and Costa (2008), [6] Poulidis and Iguchi (2021), [7] Pardini et al. (2024), [8] Harris (2013), [9] Esposti Ongaro et al. (2020), [10] Marzocchi, Sandri, and Selva (2010), [11] Neri and Macedonio (1996), [12] Hooper and Mattioli (2001), [13] Ogburn and Calder (2017).

For comparison, various descriptions of complex models found in volcanic hazard modelling studies are shown in Table 2.3.

2.4.2 Model Class Complexity

Different model classes involve inherent trade-offs between realism, generality, and precision (Jehn et al. 2019). In volcanic hazard modelling, primary model classes are defined as empirical and numerical (Section 2.1.1, Figure 2.2).

Empirical models provide a comparatively simpler alternative to heavily parameterised numerical models. However, empirical models are not inherently less complex, as they encompass a wide range of phenomenological modelling techniques and inputs (e.g., volcanic flow models MrLavaLoba, LAHARZ). One of the most well-known empirical models in volcanic hazard modelling, the Energy Cone, simulates a flow using only the release height and run-out distance (the Heim coefficient) (Malin and Sheridan 1982; Sheridan and Macías 1995). In contrast, LAHARZ utilises an empirical relationship between volume and planimetric area of inundation derived from a scaling analysis of 27 documented lahar paths from nine volcanoes (Schilling 1998).

Additionally, some modelling approaches blend in elements of both empirical and numerical models, forming 'hybrid' approaches. In volcanic hazard modelling, numerical models can be updated or calibrated using new information derived from empirical models. An example of

this occurs in atmospheric dispersion and deposition model Fall3D, where the input parameter mass eruption rate (MER) can be determined through embedded empirical models (e.g., Mastin et al. 2009; Degruyter and Bonadonna 2012; Woodhouse et al. 2013).

This demonstrates that complexity is not strictly tied to model class, as both empirical and numerical models can range from simple to complex, depending on their formulation, assumptions, and data requirements.

2.4.3 Input Parameter Complexity

Model complexity increases as additional parameters are introduced, whether by incorporating more covariates or accounting for interactions between them (Malmborg et al. 2024). While adding parameters can enhance model flexibility, it also introduces new challenges. A higher number of input parameters increases the dimensionality of the input space, which can make calibration more difficult and lead to greater uncertainty if those parameters are poorly constrained (Parry 1996; Sparks and Aspinall 2004; Felpeto, Martí, and Ortiz 2007). Additionally, complex models are generally more prone to overfitting, especially when data are limited.

Beyond the total number of input parameters, the nature and distribution of these input parameters also influence model complexity. For example, linear relationships are often easier to generalise and interpret, while non-linear forms (e.g., exponential or quadratic) increase a model's sensitivity to input variation, potentially amplifying forecast error. Similarly, the choice of which input parameters to actively estimate versus those held constant can alter the effective complexity. In the atmospheric dispersion and deposition model Fall3D, for instance, users may choose to enable particle aggregation. Doing so introduces up to seven additional parameters (Folch, Mingari, and Prata 2022), expanding the model structure and increasing its sensitivity to those specific input parameter values.

2.4.4 Input Data Complexity

While input parameter complexity concerns the number and interactions of input parameters within a model, input data complexity relates to the characteristics of the data used to quantify those parameters - its resolution, availability, quality, and source. Due to the hazardous nature of volcanic eruptions and their products, collecting certain data can be challenging due to

physical limitations or the inherent complexity of the natural process being measured.

This issue is particularly emphasised in literature on PDCs (Lube et al. 2020), as PDCs are extremely hazardous, accounting for one-third of volcanic fatalities and injuries worldwide (Auker et al. 2013). The limited optical depths of PDCs prevent visual observations from capturing their internal dynamics (Dufek 2016). Moreover, the intense forces within these flows can destroy structures and instruments deployed to study them (Valentine 1998).

Given these challenges, estimating input parameter data may necessitate the use of proxies rather than direct measurement. Such proxies can include post-depositional field-based observation and sedimentation techniques (Bonadonna et al. 2015; Brown and Andrews 2015) and analogue experiments (Holasek, Woods, and Self 1996; Rowley et al. 2014; Poppe et al. 2022), or through theoretical modelling (Alatorre-Ibargüengoitia et al. 2010; Doronzo et al. 2011; Folch, Costa, and Macedonio 2016). However, due to the infrequent, complex, and hazardous nature of volcanic eruptions, it is challenging to gather high-resolution or repeatable observations across a wide range of eruption conditions. As a result, the statistical uncertainty associated with many key input parameters can be substantial (Parry 1996).

The spatial and temporal resolution of input parameters also adds to model complexity by introducing greater heterogeneity in space and time. Higher resolution data requires more memory, computational power, and/or storage (see Section 2.4.5). As a result, processing and analysing larger volumes of data can take significantly more time. An example of this is the resolution of meteorological data used in atmospheric dispersion and deposition modelling. For modelling particle dispersion and deposition, a larger modelling area and higher resolution of both the meteorological data and the simulation grid will increase simulation and processing times (Parra 2019; Poulidis, Takemi, and Iguchi 2019; Poulidis and Iguchi 2021). The same is true for digital elevation models and their resolution, used in volcanic flow models (Capra et al. 2011; Marquez, Paredes, and Llorente 2022).

Using these models for forecasting further increases the complexity of input data. Whether short- or long-term forecasting, in the absence of direct observations, the data needed to run these models is usually based on prior data, including past eruptions (e.g., Mastin et al. 2009; Aubry et al. 2021; Deligne 2021) and expert judgement (Folch 2012; Engwell et al. 2024). However, historic data is often incomplete, as historic eruptions may be unobserved or misclassified

(e.g., Wilson et al. 1995), pre-historic eruptions may not exist in the geologic record due to deposition or weathering processes (e.g., Kueppers et al. 2019), and small sized eruptions may be missed in the geologic record (e.g., Mead and Magill 2014; Damaschke et al. 2017). When models require many input parameters, uncertainty inevitably increases due to the number of assumptions that must be made about both eruptive behaviour and data quality. To account for this uncertainty, input parameters can be represented as probability distributions rather than single values (probabilistic models, Section 2.2), and (deterministic) models are run multiple times using a range of sampled values. This strategy helps capture the plausible range of outcomes, but also increases the computational demand (Section 2.4.5).

2.4.5 Computational Complexity

Computational cost (e.g., run time, memory usage, energy consumption) is a critical factor when evaluating the complexities of different models. Numerical models, particularly high-dimensional ones, tend to have higher computational cost (Folch 2012). Probabilistic hazard assessments, which rely on volcanic hazard model outputs, typically require hundreds to several thousands of simulations to capture the full range of natural variability in input parameters (e.g., n = 1000; Jenkins et al. (2012) and Volentik and Houghton (2015), n = 10 000; Clarke et al. (2020), n = 19 200; Jenkins et al. (2015b)). While restricting the input parameter space to representative scenarios can reduce this burden, such shortcuts risk introducing bias or underestimating uncertainty (Sandri et al. 2016; Selva et al. 2018). Ensemble modelling approaches help manage this challenge by combining outputs from multiple models or multiple runs of the same model under different conditions to better capture uncertainty (e.g., Holland et al. 2020; Plu et al. 2021; Folch, Mingari, and Prata 2022; Hyman, Dietterich, and Patrick 2022).

Model parallelisation is commonly employed to improve computational efficiency (e.g., Esposti Ongaro et al. 2007; D'Ambrosio et al. 2013; de la Cruz et al. 2016; Martínez-Sepúlveda, Martí, and López-Saavedra 2024), but this approach often requires the use of additional libraries (e.g., MPI, OpenMP), thereby increasing the overall complexity of the model. Importantly, computational efficiency also informs model selection, particularly in forecasting contexts. Simpler

models, which require fewer input parameters and lower computational cost, can generate forecasts more efficiently, making them especially useful for time-sensitive operational forecasting where decision-makers must rapidly assess a wide range of hazardous outcomes (Bayarri et al. 2009; Folch 2012; Dietterich et al. 2017; Ogburn and Calder 2017).

Another important strategy for reducing computational cost is the use of statistical emulators, also known as surrogates. These are statistical models trained to approximate the outputs of more computationally expensive models (Bayarri et al. 2009; Bayarri et al. 2015). Once trained, emulators can generate rapid outputs, dramatically reducing the number of full model runs needed for probabilistic forecasting (e.g., Monte Carlo simulations, Rutarindwa et al. 2019; Yang et al. 2020; Mead, Procter, and Bebbington 2023).

Ultimately, appropriate model complexity should be determined by the specific forecasting or decision-making context rather than the level of detail alone. Overly complex models can become computationally impractical, and the tendency to equate detail with realism can undermine a model's usefulness (Salt 2008). Effective models strike a balance between bias and variance, realism and interpretability, and theoretical accuracy and operational practicality. Model complexity is also inherently shaped by the quality and availability of input data, limiting the feasibility of high-detail approaches in many real-world scenarios. Crucially, probabilistic forecasting provides a framework to incorporate uncertainty regardless of model complexity.

2.5 Model Testing

Model testing assesses the viability of a model for the purpose for which it was developed. Model testing is composed of model verification, validation, calibration, and benchmarking. Model testing is an integral part of making sure a model is scientifically robust - i.e., reproducible, replicable, and generalised with clear definitions (Larson 2005; Bollen et al. 2015). However, the concept of model testing is not straightforward.

It has been argued that models cannot truly be verified or validated in an absolute sense. As Oreskes, Shrader-Frechette, and Belitz (1994) states, "models can only be evaluated in relative terms, and their predictive value is always open to question". Even when model outputs match observations, this does not necessarily mean the model structure is correct; it could simply reflect compensating errors or coincidental agreement. Therefore, the primary value of models

is often considered to be heuristic: they help improve understanding and support decision making, rather than offering definitive predictions.

These issues are particularly relevant when evaluating forecasts produced by probabilistic models or deterministic models applied probabilistically. In theory, probabilistic forecasts could be tested by comparing predicted and observed hazard phenomena over many repeated events. But in volcanology, where eruptions are infrequent and each event is unique, this is rarely possible. A Bayesian perspective offers an alternative, framing probabilities as expressions of belief that can be updated with new data or knowledge. Under this approach, probabilistic forecasts are not validated in the classical sense. Instead, they are assessed based on how well they incorporate existing knowledge, characterise uncertainty, and are compared to alternative models or expert judgement (Rougier and Beven 2013).

Given these challenges, testing volcanic hazard models should be viewed as a multi-step, iterative process. Rather than seeking to "prove" a model is right or wrong, the goal should be to evaluate whether it is fit for purpose. This also requires transparency about assumptions, simplifications, and known limitations, and careful attention to how uncertainty is handled (Section 2.3.5).

The following subsections provide a practical overview of how model testing has been applied in volcanic hazard modelling.

2.5.1 Verification

Model verification is not often explicitly mentioned in volcanic hazard model literature, indicating that there is an assumption that if a model is validated, then it is also verified (Esposti Ongaro et al. 2020). A thorough literature search revealed only five examples of volcanic hazard models being explicitly verified: atmospheric dispersion and deposition models Tephra2 (Connor, Connor, and Saballos 2011), Ash3D (Schwaiger, Denlinger, and Mastin 2012), Fall3D (Folch et al. 2020; Prata et al. 2021), VBP model Eject! (Mastin 2001), and mass flow model IMEX_SfloW2D (de'Michieli Vitturi et al. 2019). This pattern is not unique to volcanology; in many scientific modelling domains, verification is often conflated with validation (Augusiak, Van den Brink, and Grimm 2014). While model verification in volcanology rarely involves direct comparison to analytical solutions or theoretical benchmarks, because such solutions often

do not exist for complex volcanic processes, verification is still possible by testing whether the model implementation behaves as intended, given the conceptual model (e.g., conserving mass, producing physically plausible transport behaviour).

2.5.2 Validation

Validation exercises for volcanic hazard models are primarily conducted using past eruptions (Table 2.4) (hindcasting, Section 2.5.5). Model validation is a continuous process - as the greater the number and diversity of validated eruptions, the more probable that the model's equations and conceptualisation are not flawed (Oreskes, Shrader-Frechette, and Belitz 1994; Trucano et al. 2006; Esposti Ongaro et al. 2020).

However, it is technically infeasible to validate a model against future, unknown hazardous events. This is particularly challenging in long-term forecasting, where the rarity and unpredictability of eruptions limit opportunities for model validation (Mason, Pyle, and Oppenheimer 2004; Sheldrake et al. 2017). Similar issues have been acknowledged in other hazard domains, such as earthquake forecasting, where efforts such as the Regional Earthquake Likelihood Models (RELM) project - that aims to produce and evaluate models of earthquake potential for California, USA - have highlighted the difficulty of evaluating model performance on long timescales due to the scarcity of observed events (Schorlemmer et al. 2007; Schorlemmer et al. 2010).

Furthermore, the various components of volcanic hazard models (Section 2.7.1), along with model complexity (Section 2.4), can be explored to assess their influence on a model's ability to produce accurate forecasts.

2.5.3 Calibration

Table 2.5 provides examples of calibration efforts across various volcanic hazard models. In many volcanic hazard studies, the distinction between calibration and validation is blurred, and both are often conducted simultaneously (e.g., Bonadonna et al. 2005; Andronico et al. 2008; Wantim et al. 2013; Biass et al. 2016a). For instance, input parameters may be adjusted iteratively until model outputs align with observations - a process that may serve both to calibrate uncertain model parameters and validate the model. This process is typically not about making extreme changes, but rather refining uncertain parameters based on observational constraints or expert

Table 2.4: Examples of validation cases for different volcanic hazard models.

Volcanic Hazard Model	Validated Eruption(s)	Reference
Atmospheric Dispersion and Deposition Models		
HYSPLIT	2008 Kasatochi	Crawford et al. (2016)
PUFF	1992 Mount Spurr/Crater Peak, 1994 Klyuchevosky volcano, & 1994 Rabaul caldera	Searcy, Dean, and Stringer (1998)
HAZMAP	79AD, 1631 & 1994 Vesuvius, 2002-03 Mount Etna	Andronico et al. (2008) and Macedonio, Costa, and Folch (2008)
NAME III	2010 Eyjafjallajökull	Grant et al. (2012)
Tephra2	1992 Cerro Negro	Connor, Connor, and Saballos (2011)
Fall3D	2008 Chaitén, 2010 Eyjafjallajökull, 2011 Puyehue-Cordon Caulle, & 2019 Raikoke (SO_2 cloud),	Folch, Costa, and Basart (2012), Osores et al. (2013), and Prata et al. (2021)
Ash3D	1992 Mount Spurr/Crater Peak	Schwaiger, Denlinger, and Mastin (2012)
Volcanic Flow Models		
Energy Cone (ΔH/L)	1980 Mount St. Helens	Malin and Sheridan (1982)
SCIARA	1991 & 2001 Mount Etna	Crisci et al. (2004)
FLOW-3D	1982 El Chichon	Macías et al. (2008)
LaharZ	1998 Sarno, 1997 & 2001 Popocatepetl	Dorta et al. (2007) and Muñoz-Salinas et al. (2009)
FLOWGO	1984 Mauna Loa, 1997 Pu'u 'Ō'ō & 1998 Mount Etna	Harris and Rowland (2001)
PDAC	1997 Soufrière Hills	Clarke et al. (2002)
DOWNFLOW	1991-93 & 2001 Mount Etna	Favalli et al. (2005)
Titan2D	Soufrière Hills	Widiwijayanti et al. (2004)
MAGFLOW	2001 Mount Etna	Vicari et al. (2007)
VolcFlow	2006, 2010 Tungurahua	Kelfoun et al. (2009) and Kelfoun and Vargas (2016)
MOLASSES	2012-13 Kamchatka	Richardson (2016) and Kubanek et al. (2015)
MrLavaLoba	2021 Fagradalsfjall	Pedersen et al. (2022)
IMEX_SfloW2D	Synthetic eruption Vesuvius	de'Michieli Vitturi et al. (2019)
EC_MapProb & Box_MapProb	2008-09 Chaitén	Aravena et al. (2020)
Volcanic Ballistic Projectile Models		
Eject!	2016-2017 Bogoslof & 2022 Mount Etna	Waythomas and Mastin (2020) and Costa et al. (2023)
LPAC	1997 Soufriere Hills	de'Michieli Vitturi et al. (2010)
Alatorre-Ibargüengoitia, Delgado-Granados, and Dingwell (2012)	1998 & 2003 Popocatepetl	Alatorre-Ibargüengoitia, Delgado-Granados, and Dingwell (2012)
Ballista	2014 Mount Ontake	Tsunematsu et al. (2016)
Great Balls of Fire	1888 - 90 La Fossa (Vulcano Island)	Biass et al. (2016a)

Table 2.5: Examples of calibration cases for different volcanic hazard models.

Volcanic Hazard Model	Eruption	Reference
Atmospheric Dispersion and Deposition Models		
HYSPLIT	N/A	N/A
PUFF	2018 & 2019 Sakurajima	Tanaka and Iguchi (2019) and Tanaka, Nakamichi, and Iguchi (2020)
HAZMAP	2002-03 Mount Etna	Andronico et al. (2008)
NAME III	Post 2010 Eyjafjallajökull	Beckett et al. (2017)
Tephra2	1315 Kaharoa (Mount Tarawera) & June 1996 Mount Ruapehu	Bonadonna et al. (2005)
Fall3D	2013 Mount Etna, 1980 Mount St. Helens, & 1992 Crater Peak	Folch et al. (2010) and Prata et al. (2021)
Ash3D	N/A	N/A
Volcanic Flow Models		
Energy Cone (ΔH/L)	Tungurahua, Soufriere Hills & Arenal Volcano	Toyos et al. (2007) and Aravena et al. (2024)
SCIARA	2001 & 2006 Mount Etna	D'Ambrosio et al. (2005) and Spataro et al. (2015)
FLOW-3D	N/A	N/A
LaharZ	Original Calibration of Nine Volcanoes	Iverson, Schilling, and Vallance (1998)
FLOWGO	Mount Cameroon	Wantim et al. (2013)
PDAC	N/A	N/A
DOWNFLOW	1977 & 2002 Nyiragonga, Mount Etna, 2014-15 Fogo	Favalli et al. (2009), Tarquini and Favalli (2011), and Richter et al. (2016)
Titan2D	2006 Merapi	Charbonnier and Gertisser (2012)
MAGFLOW	N/A	N/A
VolcFlow	2006 Merapi	Charbonnier and Gertisser (2012)
MOLASSES	N/A	N/A
MrLavaLoba	N/A	N/A
IMEX_SfloW2D	2022 Mount Etna	Zuccarello et al. (2025)
EC_MapProb & Box_MapProb	2070 cal yr BP El Misti	Aravena et al. (2022)
Volcanic Ballistic Projectile Models		
Eject!	N/A	N/A
LPAC	N/A	N/A
Alatorre-Ibargüengoitia and Delgado-Granados (2006)	1994 Popocatepetl	Alatorre-Ibargüengoitia and Delgado-Granados (2006)
Ballista	2012 Te Maari	Fitzgerald et al. (2014)
Great Balls of Fire (GBF)	La Fossa	Biass et al. (2016a)

judgement.

In some cases, such as in studies using the atmospheric transport and deposition model Ash3D, the term "calibration" is rarely used. Instead, the literature refers to "input parameter adjustment" or similar terminology (Mastin, Van Eaton, and Durant 2016). For other models, such as the atmospheric transport and dispersion model HYSPLIT and VBP model Eject!, published studies often focus on calibrating observational tools (e.g., radar, thermal imagery, or camera measurements) while using the model outputs as a reference or validation target (e.g., Corradini et al. 2016; Pardini et al. 2020; Costa et al. 2023).

2.5.4 Benchmarking

In tephra modelling, a benchmarking exercise was carried out during the 2010 International Association of Volcanology and Chemistry of the Earth's Interior (IAVCEI) workshop on Ash Dispersal Forecast and Civil Aviation (Bonadonna et al. 2011a; Bonadonna et al. 2012). 12 atmospheric dispersion and deposition models were benchmarked against the Hekla 2000 eruption in Iceland (Hoskuldsson et al. 2007). This exercise highlighted substantial variability in model outputs, primarily due to differences in model structure and model input parameters and assumptions, but did not identify a single best-performing model. These findings underscore that model performance is potentially eruption- or volcano-specific, and no single model can be universally applied across all scenarios.

For lava flows, Cordonnier, Lev, and Garel (2015) conducted five benchmarking exercises with seven flow models simulating (1) a dam-break flow, (2) inclined viscous isothermal spreading, (3) axisymmetric cooling and spreading, (4) a split flow experiment, and (5) a natural case. Dietterich et al. (2017) expanded this by incorporating additional experimental data and assessing both accuracy and CPU efficiency. They found that while most models performed similarly for short-duration, isoviscous flows, thermal processes became increasingly important for longer-lived, cooling-limited lava flows. This suggests that models incorporating thermal dynamics may offer improved performance for simulating longer-duration eruptions and cooling-limited lava flows.

For PDC, Esposti Ongaro et al. (2020) proposed - though did not implement - a set of experimental configurations to test PDC models, including: (1) large-scale, dilute, turbulent,

polydisperse gravity current over an incline, (2) large-scale axisymmetric polydisperse gravity current from jet collapse, (3) concentrated, fluidised/non-fluidised granular current over an incline, (4) turbulent gas-particles flows with buoyancy reversal, and (5) interaction of stratified gas-particle gravity currents with obstacles. In contrast, Gueugneau et al. (2021) conducted a benchmarking exercise focused on four concentrated PDC models, with five test scenarios that included flat slopes, bends, slope breaks, obstacles, and constriction, which highlighted the complexity of simulating PDC dynamics across varied terrain.

One study used laboratory experiments to evaluate the performance of VBP models, comparing analytical and numerical implementations under simplified conditions (e.g., no topography, consistent initial conditions across simulations) (Bertin 2017). Although the authors referred to this as benchmarking, the exercise more closely resembled a sensitivity analysis, focused on internal comparisons of model behaviour rather than performance against other VBP models or real-world data.

Benchmarking is the least frequently conducted type of model testing for volcanic hazard models.

2.5.5 Past Event Testing (Hindcasting)

Hindcasts, or retrospective forecasts, are constructed after an event has occurred. The advantage of running a model retrospectively is that input parameters are better constrained (to the best of the monitoring data and sampling techniques). Hindcasting is an integral part of model testing, as it can be used to see whether a model can correctly simulate a historical event. At all four stages of model testing - verification, validation, calibration, and benchmarking - hindcasts can, and are, used to test model accuracy (e.g., Widiwijayanti et al. 2004; Scollo, Del Carlo, and Coltelli 2007; Procter, Cronin, and Sheridan 2012; Dietterich et al. 2017). If a model can accurately simulate a past event, then the model is considered to be an accurate and credible representation of the phenomenon modelled. However, this is only the case if the observations used to evaluate the hindcast were not also used in calibrating the model, as this can lead to overfitting (e.g., Schorlemmer et al. 2018). Hindcasting can also be applied to simulate past behaviours at any volcano for applications such as reconstructing hazard records (e.g., Johnston et al. 2012; Jenkins et al. 2020; Tennant et al. 2021) and hazard assessment

(e.g., hazard maps) (e.g., Bonasia et al. 2011; Michaud-Dubuy et al. 2019).

However, even when using optimal or carefully calibrated input parameters, different models can produce different results when simulating the same eruption. Scollo, Folch, and Costa (2008) ran a comparative parametric study of different atmospheric dispersion and deposition models (Fall3D, HAZMAP, and TEPHRA) for two plausible eruption scenarios from Mt. Etna, similar to the 2002-03 and 1990 eruptions, and found that there were differences in model outputs between all three models. The study found that each model responded differently to changes in key input parameters, such as erupted mass, column height, column model, bulk granulometry, particle shape, and settling velocity. These differences highlight how structural and technical variations between models can influence outcomes, even when input parameters are held constant or tuned to observed events.

Hindcasting also provides insights into a model's uncertainty - if a model output were accurate, there would exist a set of input parameters that could perfectly reproduce past observations. However, if a model exhibits large errors in hindcasting, its ability to accurately simulate future events becomes questionable.

2.5.6 Near-Real-Time Forecast Evaluation

Forecast testing evaluates a model's performance against future events and is an important component of model evaluation in domains with continuous or frequent data streams, such as weather forecasts (e.g., Thornes and Stephenson 2001; Gilleland and Roux 2015), economic market forecasting (e.g., Pyo et al. 2017; Iregui, Núñez, and Otero 2021), and electricity price modelling (e.g., Yamin, Shahidehpour, and Li 2004; Weron 2014). In these fields, routine forecasts are continuously updated, tested, and refined through near-immediate feedback from observations.

In contrast, forecast testing for volcanic hazard models remains extremely limited due to the infrequent and highly variable nature of eruptions. Moreover, not all eruptions produce the same hazard types, as some generate ash clouds but not lava flows, so evaluating a specific hazard model requires an eruption of the relevant style and magnitude. This lack of continuous observational feedback impedes systematic calibration and iterative correction of volcanic hazard models.

Despite these challenges, emerging efforts are attempting to evaluate model accuracy for near-real-time forecasts. For example, Folch, Mingari, and Prata (2022) assessed the accuracy of ensemble-based deterministic and probabilistic simulations in the Fall3D model using post-event observations of SO_2 clouds (from the July 2018 Ambae eruption) and ash clouds and ashfall (from the April 2015 Calbuco eruption). While conducted retrospectively, this study highlights a promising direction towards more accurate near-real-time forecasts.

Forecast evaluation is also supported by a broad body of statistical literature that offers tools for assessing both deterministic and probabilistic forecasts. This includes the use of proper scoring rules, such as the continuous ranked probability score (CRPS), which evaluate the calibration and precision of probabilistic forecasts (e.g., Gneiting and Raftery 2007; Gneiting 2011; Gneiting and Katzfuss 2014). These methods are widely used in fields such as weather and earthquake forecasting (e.g., Brehmer et al. 2024), and provide a theoretical framework for comparing volcanic hazard models against observed outcomes. While their direct application to volcanic hazard forecasting is still limited, these approaches offer useful guidance for future efforts to evaluate model performance, particularly as more observational data become available.

2.6 Model Selection

In many scientific disciplines, including weather prediction and earthquake forecasting, the use of multi-model ensembles is becoming standard practice (e.g., Krishnamurti et al. 2000; Rhoades and Gerstenberger 2009; Krishnamurti et al. 2016; Marzocchi, Zechar, and Jordan 2012; Llenos and Michael 2019). These approaches combine outputs from multiple models to provide a more robust representation of uncertainty, often outperforming individual models (Herrmann and Marzocchi 2023). Ensemble approaches offer a clear advantage in volcanic hazard forecasting as well, particularly in capturing uncertainty from the models, input parameters, and data. However, the strength of an ensemble depends on the quality and suitability of the individual models it includes. Poorly chosen or incompatible models can introduce biases and reduce forecast interpretability. Thus, whether adopting a single model or ensemble-based approach, selecting appropriate models remains a foundational step.

Selecting the most appropriate model or set of models for any given purpose is critical, even though, in theory, running all available models might yield the most accurate forecast. In

practice, however, this is rarely feasible due to constraints on time, computational resources, and data availability. Traditional model selection techniques, such as F-tests, Bayesian model averaging, or Akaike's Information Criterion, are valuable tools for evaluating goodness-of-fit in contexts where observed outcomes are known (Ludden, Beal, and Sheiner 1994; Glatting et al. 2007; Steel 2011). However, in forecasting applications, these approaches are often impractical because the true outcome is not yet available. Instead, model selection must also consider factors such as operational constraints and the feasibility of implementation for real-time or decision-making purposes. While formal validation techniques, such as comparing model outputs to observed data (Section 2.5.2), are essential for assessing model accuracy, they are use- and user-dependent and may not always be feasible, particularly for forecasting. The next steps in model selection should involve iterative testing to ensure the chosen model aligns with the intended application. Ultimately, model selection is shaped by its broader context - the who, why, when, and where - each of which influences what model is most suitable:

- **Who is using the model?** Scientists and modellers may prioritise detailed representations of volcanic phenomena; emergency managers may require simpler yet accurate outputs for risk communication and decision making; and policy makers may need technically accurate results that are usable and clear for long-term planning (Figure 2.2) (Doyle et al. 2014; Thompson, Lindsay, and Gaillard 2015; Das et al. 2025). Additionally, some models require specialised datasets that may only be accessible through government agencies or research institutions, which are unlikely to be used in practical decision making (Renschler 2005). These user needs influence which models are considered suitable candidates - whether for individual use or in a broader ensemble forecast.

- **Why is the model being used?** Long-term hazard assessments or hindcasting allow time for detailed modelling and data collection, whereas short-term and syn-eruptive forecasting necessitates models that can run quickly with limited data.

- **When will the model be used?** Similarly to above, short-term and syn-eruptive forecasting impose constraints on data availability, as well as computational efficiency. Conversely, long-term forecasting and hindcasting allow for more time for data collection (e.g., historical data on past eruptions) and model testing (hindcasting past events to refine model accuracy).

- **Where is the model being applied?** The model should be appropriate for both the volcano and the spatial scale of interest. Model users must define the target volcano, which dictates the level of knowledge available about its past activity and hazards. The spatial extent over which a user wants to simulate will also influence model selection. Some atmospheric dispersion and deposition models are designed for local-scale simulations (e.g., Tephra2), while others can simulate hazards at regional or global scales (e.g., Fall3D, Ash3D, HYSPLIT).

These considerations can be synthesised into a broader framework that links model selection to volcano type and modelling purpose. For example, Mount Etna (Italy) is a frequently active basaltic system where key volcanic phenomena include ashfall, lava flows, and VBP (Branca and Carlo 2005; Andronico et al. 2021). Mount Etna benefits from a wealth of observational data for these phenomena (e.g., Barberi et al. 1993; Behncke, Neri, and Nagay 2005; Andronico et al. 2009; Costa et al. 2023), enabling the use of more complex models with numerous input parameters. However, because these frequently active systems often require rapid, short-term forecasts, lower-CPU models that can deliver more timely outputs are often more practical in operational contexts.

In contrast, Taupō volcano (Aotearoa New Zealand) is a large rhyolitic caldera system with a history of widespread ashfall and PDCs (Wilson 1985; Wilson et al. 2006). Taupō volcano erupts infrequently but with high impact, and has no observational eruption data and high epistemic uncertainty (Barker et al. 2020). In such cases, models with fewer input parameters may be preferred to minimise user assumptions, while higher-CPU models that simulate over larger spatial domains or include more detailed processes are feasible due to the longer timescales between unrest and eruption (Phillipson, Sobradelo, and Gottsmann 2013; Sandri, Acocella, and Newhall 2017).

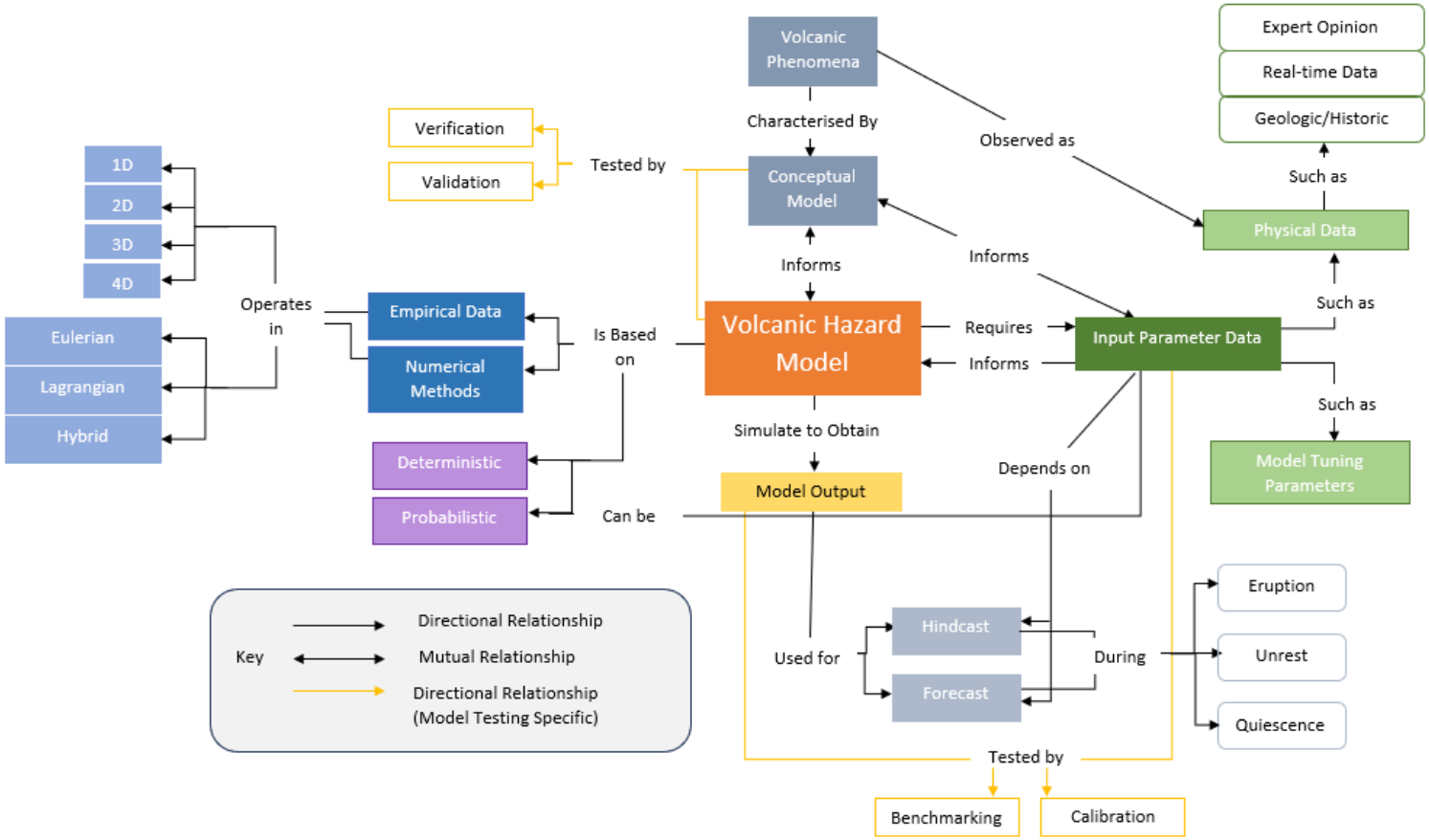
2.7 Discussion

This review of computational volcanic hazard models situates volcanic hazard modelling within the broader context of environmental modelling, where challenges such as model complexity, input parameter uncertainty, and model selection are well recognised - particularly in fields such as hydrology (e.g., Butts et al. 2004; Li, Xu, and Beldring 2015; Guthke 2017; Paul et al.

2021) and ecology (e.g., Tredennick et al. 2021; Riva et al. 2023; Malmborg et al. 2024). Like these disciplines, volcanic hazard modelling often requires balancing physical realism with computational resources and limited data. To help address these challenges, this review introduces a volcanic hazard model ontology that captures key components of model structure, function, and application. This ontology is described in detail below (Section 2.7.1). This review also supports the application and future development of volcanic hazard models for a diverse set of end-users, across forecasting timescales - long-term, short-term and near-real-time.

2.7.1 Volcanic Hazard Model Ontology

In the context of volcanic hazard modelling, ontologies enable consistency in model descriptors, which is essential for transparent comparison and evaluation of different models, as well as improving interpretability, comprehension, and ultimately, trust in modelling outcomes (Broniatowski 2021). In other environmental disciplines, this is accomplished through model ontologies. These often consist of terminologies (nodes) connected through a semantic network (via relationship links). Top-level nodes represent fundamental processes and assumptions in the domain, while sub-level nodes capture more specialised entities and terms (Myer 2018). Example model ontologies exist for systems biology (Lambrix 2004), medicine (Haendel, Chute, and Robinson 2018; Ong et al. 2020), tsunami research (Ramar and Mirnalinee 2012; Babič et al. 2022), and landslide susceptibility (Jackson, Smyth, and Poole 2008; Jung and Chung 2015).



Knowledge Needs per User					
Model Users	Conceptual Model	Model Type	Model Approach	Input Parameter Data	Model Output
Geologist	Relevant	Relevant	Somewhat Relevant	Relevant	Somewhat Relevant
Forecasters/ Modellers	Somewhat Relevant	Relevant	Relevant	Relevant	Relevant
Emergency Managers	Less Relevant	Less Relevant	Relevant	Less Relevant	Relevant

Figure 2.2: Volcanic hazard model ontology. The table at the bottom of the figure shows the relative relevance of different aspects of volcanic hazard models to three user types: geologists, forecasters, and emergency managers. Relevant = Directly impacts the user's ability to use, trust, or interpret the model. Somewhat Relevant = Useful context but not essential for decision-making. Less Relevant = Not typically a focus for this user type.

In volcanology, ontologies have been used to describe geologic features and processes (e.g., Mcguinness et al. 2006; Pulido et al. 2009; Fauziati and Watannabe 2010; Myer 2018), but existing frameworks do not yet incorporate components related to volcanic hazard models.

We present here a volcanic hazard model ontology (Figure 2.2) that connects volcanic processes with modelling components, to facilitate effective model comparison and meaningful knowledge sharing.

This ontology was produced by systematically identifying and documenting key features across multiple volcanic hazard models and incorporates the key features from Sections 2.2, 2.3, 2.5. An example of how different model users may engage with this ontology is illustrated in the table within Figure 2.2, which outlines the relative relevance of key model components, such as conceptual models, model type (i.e., empirical or numerical), model approach (i.e., deterministic or probabilistic), input parameter data, and model outputs, for geologists, forecasters/modellers, and emergency managers.

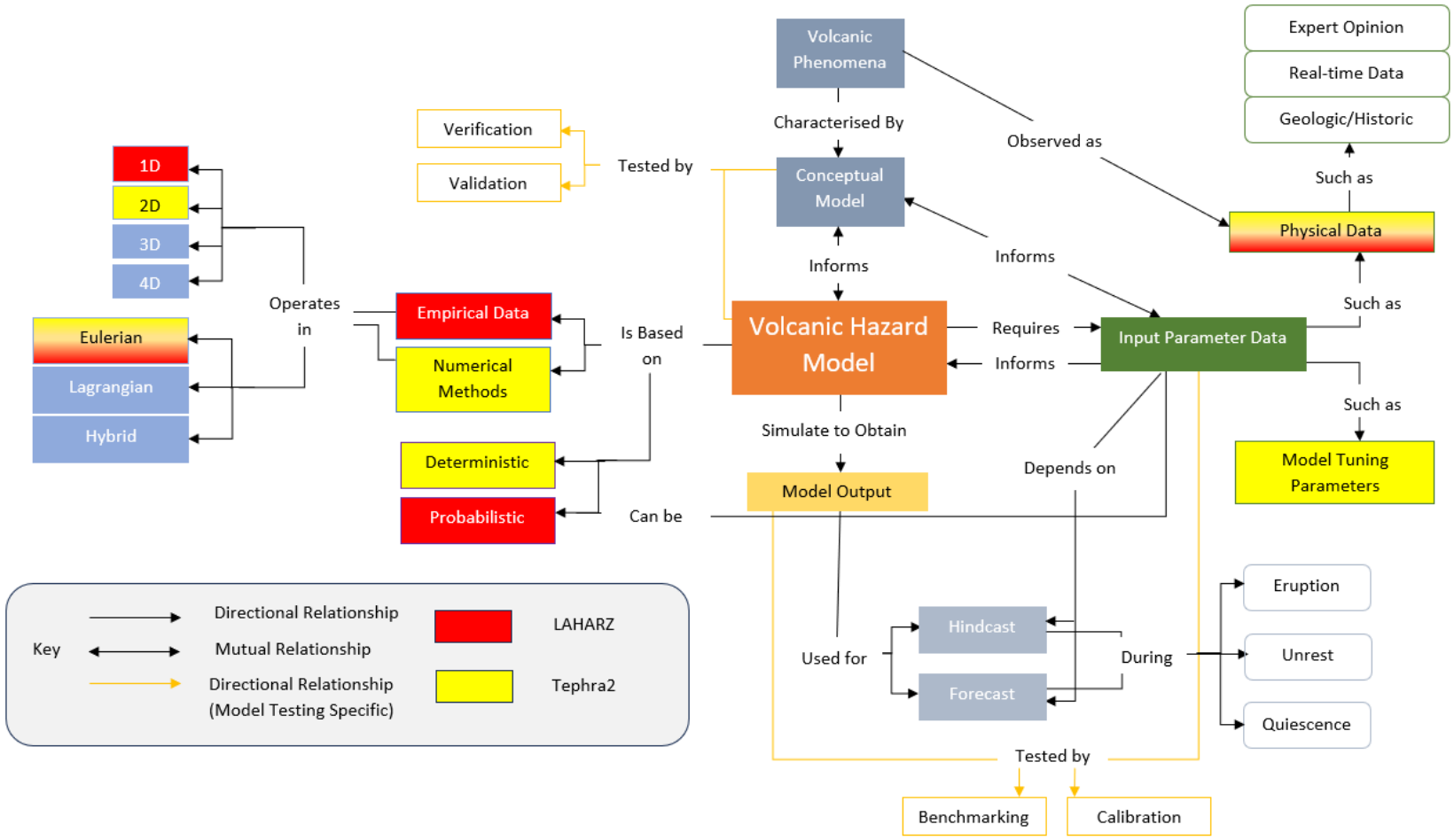


Figure 2.3: Volcanic hazard model ontology with two volcanic hazard model examples - Tephra2 (highlighted yellow) and LAHARZ (red).

An example of model representation with this ontology is an atmospheric dispersal and deposition model, Tephra2, and a flow model, LAHARZ (Figure 2.3). Tephra2 is a deterministic, Eulerian, two-dimensional model that employs numerical methods. The model requires tuning parameters, such as the number of column integration steps (i.e., the number of levels in the eruption plume) (Connor, Connor, and Bonadonna 2008) - and physical data, such as ESPs (e.g., plume height, total grain size distribution, and particle density). These input parameters can inform the conceptual model, while the conceptual model, in turn, guides parameter selection. As discussed in Section 2.5, the Tephra2 conceptual and computational models have undergone verification and validation, while their input parameters and outputs have been calibrated (Table 2.4, 1992 Cerro Negro, Nicaragua; Connor, Connor, and Saballos 2011) and benchmarked (2000 Hekla, Iceland; Hoskuldsson et al. 2007).

In comparison, LAHARZ is a probabilistic, Eulerian, one-dimensional model that employs an empirical relationship. The model primarily requires physical data, including a digital elevation model, flow volume, and channel geometry (i.e., height and length). The LAHARZ conceptual and computational models have also undergone verification and validation, and their input parameters have been calibrated (Iverson, Schilling, and Vallance 1998; Oramas-Dorta et al. 2012). However, unlike Tephra2, LAHARZ has not been subject to benchmarking exercises.

2.7.2 Limitations

One key gap is the lack of benchmarking studies for volcanic hazard models. Among the different types of model evaluation identified in this review, benchmarking was the least frequently documented. Despite an exhaustive literature review, few studies directly compare model outputs under standardised conditions. Notably, there have been no benchmarking exercises for volcanic ballistic projectile models. Another limitation found in the literature is that some models have not been explicitly calibrated against observational data, making their reliability in forecasting contexts uncertain. For forecasting applications, particularly those involving atmospheric dispersion and tephra deposition, more benchmarking efforts are essential. This is because both eruptive and atmospheric conditions exhibit far greater variability than the relatively static terrain parameters involved in volcanic flow modelling.

In addressing model uncertainty, a proposed approach in forecasting is to constrain input

parameter values using databases of ESPs from past eruptions (e.g., Mastin et al. 2009; Aubry et al. 2021; Deligne 2021; Ogburn 2025; Paine and Wadsworth 2025). This method has been primarily applied in tephra dispersal and deposition modelling (e.g., Bonadonna et al. 2011a; Bonadonna et al. 2013). However, there are currently no open-source databases specifically designed for VBP, although initiatives such as the LaMEVE database have aimed to provide global eruption records that could support their future development (e.g., Crosweller et al. 2012). In theory, such databases could enhance forecast accuracy by offering historically informed input parameter estimates. However, it remains unclear whether this strategy meaningfully improves predictive capability, as volcanic eruptions are highly variable and may deviate significantly from a volcano's past behaviour (Whitehead and Bebbington 2021). Testing is needed to evaluate whether database-informed input parameters actually improve forecasting accuracy.

Given these limitations and uncertainties, sensitivity analysis plays a crucial role in model selection and refinement. Sensitivity studies help identify which input parameters most influence model output variance, enabling users to focus efforts on characterising the most impactful sources of uncertainty (e.g., Stevens, Manville, and Heron 2003; Scollo et al. 2008; Bilotta et al. 2012; Devenish et al. 2012; Osman et al. 2020; Scott et al. 2025). This information is particularly valuable in operational settings, where reducing the dimensionality of input parameter space without sacrificing accuracy can make probabilistic forecasting more computationally feasible and actionable.

2.8 Conclusion

By evaluating model construction principles and synthesising insights from existing model approaches, this research contributes to a more systematic and transparent approach to volcanic hazard modelling. This research explores the classification, development, testing, and complexity of volcanic hazard models, with a particular focus on their application in forecasting. It highlights the need to balance model complexity, computational efficiency, and accuracy while acknowledging the uncertainties inherent in both input parameters and model processes. Model construction and selection must consider not only data availability but also the intended application, whether for syn-eruptive forecasting or long-term hazard assessments.

This manuscript adopts a qualitative lens to discuss model complexity, reflecting the diverse

ways complexity manifests in volcanic hazard models. While many quantitative metrics exist, such as spatio-temporal resolution, model dimensionality, and the number of input parameters, none fully capture the multifaceted nature of complexity. Future research should explore integrating these quantitative measures with qualitative insights to develop more comprehensive complexity assessments.

A key finding is the importance of models that are detailed enough to provide meaningful simulations yet simple enough to minimise error. The structured framework for model selection proposed in this research enables an objective selection method, based on the suitability of volcanic hazard models for different objectives.

This review is limited to forward modelling. Other approaches, including inverse modelling and model falsification, fall outside the scope of this work. However, they offer valuable methods for improving model development and testing, and should be explored in future studies (e.g., Bevilacqua et al. 2019).

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the student and the student's main supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the student's contribution as indicated below in the Statement of Originality.

Student name:	Emmy Elizabeth Scott
Name and title of main supervisor:	Dr Melody Whitehead, Senior Research Officer and Senior Lecturer
In which chapter is the manuscript/published work?	Chapter Three

Describe the contribution that the student and members of the supervisory team have made to the manuscript/published work:[1]

Emmy Scott: Writing: original draft, Methodology, Investigation, Data curation, Conceptualisation
Melody Whitehead: Writing - review and editing, Methodology, Conceptualisation
Stuart Mead: Writing - review and editing, Methodology, Conceptualisation
Mark Bebbington: Writing: review and editing, Methodology
Jonathan Procter: Writing: review and editing

Please select one of the following three options:

(●) **The manuscript/published work is published or in press**

Please provide the full reference of the research output:

Scott, E., Whitehead, M., Mead, S., Bebbington, M., & Procter, J. (2025). Global sensitivity analysis of models for volcanic ash forecasting. Journal of Volcanology and Geothermal Research, 466. https://doi.org/10.1016/j.jvolgeores.2025.108393

( ) **The manuscript is currently under review for publication**

Please provide the name of the journal:

( ) **It is intended that the manuscript will be published, but it has not yet been submitted to a journal**

Student's signature:	Emmy Scott Digitally signed by Emmy Scott Date: 2025.11.12 11:04:00 +13'00'	Main supervisor's signature:	Melody Whitehead Digitally signed by Melody Whitehead Date: 2025.11.12 12:13:03 +13'00'

This form should be placed at the beginning of each relevant thesis chapter.

[1] Refer to the Massey University Publishing and Authorship guidelines (OneMassey for staff, Stream for students) and/ or Contributor Roles Taxonomy (CRediT) guidelines for guidance.

Chapter 3

Global sensitivity analysis of models for volcanic ash forecasting

In Chapter Two, I reviewed the current landscape of volcanic hazard models, highlighting their differences, input requirements, and the role of uncertainties in shaping forecast outputs. This review demonstrates that while models can be broadly classified and their structures understood, the relative influence of individual input parameters on forecast accuracy remains unclear. Chapter Three builds on this by quantifying how model inputs influence model outputs, using a global sensitivity analysis of two volcanic ash transport and dispersion models. This approach directly addresses Objective Two: *Which input parameters most influence volcanic hazard model outputs?*

Chapter Three has been published as Scott, E., Whitehead, M., Mead, S., Bebbington, M., & Procter, J. (2025). Global sensitivity analysis of models for volcanic ash forecasting. *Journal of Volcanology and Geothermal Research*, *466*. https://doi.org/10.1016/j.jvolgeores.2025.108393 (Accepted: 6 June 2025; Published: 3 July 2025).

Chapter Three reproduces the published version with only formatting modifications.

Abstract

Volcanic ash is a widespread and destructive volcanic hazard. Timely forecasts for ash deposition help mitigate the risks of volcanic hazards to society. Producing these forecasts requires numerous simulations with varying input parameters to encapsulate uncertainty and accurately capture the actual event to deliver a reliable forecast. However, this process is computationally intensive and there is often insufficient time to run hundreds of thousands of simulations across all feasible combinations of input parameters, especially considering that some input parameters are continuous. Since input parameters used for volcanic hazard models are fundamentally uncertain before (and often also after) an eruption, it is critical to understand which input parameters should be varied to assess worst case scenarios or to produce a realistic range of possible scenarios before an eruption. This research explores the input space of two volcanic ash transport and dispersion models, Tephra2 and Fall3D, in the context of forecasting an unknown future eruption. We use the exemplar of Taranaki Mounga (Mount Taranaki), Aotearoa New Zealand, which has an estimated 30% to 50% chance of an explosive eruption in the next 50 years. We statistically determine how much each input parameter contributes to model output variance through a global sensitivity analysis via Sobol' indices and the extended Fourier Amplitude Sensitivity Test (eFAST). Our findings show that for both Tephra2 and Fall3D, grain size distribution, diffusion, plume shape, and plume duration (Fall3D only) have a substantial first-order impact on model output variance. In contrast, mass, particle density, and plume height have minimal impact in the first-order but become influential when considering parameter-parameter inter-relationships (total order). The results not only enhance our understanding of model sensitivities but also point to improved efficiency in forecasting efforts.

3.1 Introduction

Volcanic eruptions are powerful natural events that can endanger populations. Mitigating the societal impacts of eruption-related hazards can be accomplished through actions and plans based on hazard assessments and forecasts, which are reliant on the output of computational models (e.g., Folch 2012; Selva et al. 2014; Sandri et al. 2016; Pardini et al. 2024). Advancements in computing power have enabled the development of increasingly complex volcanic ash

transport and dispersion models (VATDMs) (Folch 2012). These models take input parameters - such as plume height, mass eruption rate (MER), and wind conditions - and simulate ash dispersal and deposition to produce outputs such as ashfall extent and thickness. More complex models can be developed by increasing the number of tuneable input parameters to capture the physical processes of ash transport and deposition (Höge, Wöhling, and Nowak 2018; Malmborg et al. 2024). However, there is little evidence that more complex models produce more accurate forecasts of future events (Petropoulos et al. 2022). Additionally, more unconstrained input parameters may contribute to increases in uncertainty, and consequently, poorly constrained forecasts (Green and Armstrong 2015; Saltelli et al. 2020).

Volcanic hazard model outputs rely on the accuracy of the input values (Selva et al. 2014; Engwell et al. 2024; Pardini et al. 2024), and are validated through simulating past eruptions (hindcasting) (Bonasia et al. 2012; Michaud-Dubuy, Carazzo, and Kaminski 2021). Successfully replicating a past event is a key step in model validation (Widiwijayanti et al. 2004; Connor and Connor 2006; Scollo, Del Carlo, and Coltelli 2007; Dietterich et al. 2017), based on the assumption that a model which accurately simulates one eruption provides a credible basis for forecasting future activity (Pyle 2018). However, hindcasting often involves extensive tuning of input parameters, and performance on one eruption does not guarantee similar accuracy for others (Scollo, Folch, and Costa 2008; Dietterich et al. 2017; Gueugneau et al. 2021).

During signs of heightened volcanic unrest, there is insufficient time to run hundreds of thousands of simulations of all possible input parameter combinations. Therefore, we want to know before an eruption, how many of the input parameter combinations produce statistically distinguishable results. In other words, we aim to identify which input parameters should be varied to explore worst-case scenarios and to capture the full range of physically plausible outcomes, including extreme but realistic combinations of conditions (i.e., end-member events) before an imminent eruption. Because ash transport and deposition is strongly influenced by wind (Bursik 2001; Bonadonna and Costa 2013; Costa, Pioli, and Bonadonna 2016; Mulena et al. 2016), which varies in real-time, our focus remains on identifying key input parameters that contribute to output variability, rather than atmospheric conditions. Building on previous studies (Scollo et al. 2008; Pardini et al. 2022), we examine input parameter ranges under complete uncertainty acknowledging that the future is fundamentally unknown.

This paper explores the input parameter space of two commonly used VATDMs Tephra2 (Connor, Connor, and Bonadonna 2008) and Fall3D (Folch, Costa, and Macedonio 2009; Folch et al. 2020) for tephra deposition forecasting. We selected these two VATDMs because they differ in the number and style of input parameters, are widely used, and are open-source. Tephra2 is a semi-analytical model with fewer input parameters, making it computationally efficient. In contrast, Fall3D is a more complex Eulerian model that incorporates finer-scale physics and a greater number of input parameters, enabling higher-resolution simulation.

3.2 Volcanic Ash Transport and Dispersion Models

Various computational models have been developed to simulate the transport and deposition of volcanic ash. The following sections provide an overview of the two VATDM used in our study, Tephra2 and Fall3D.

3.2.1 Tephra2

Tephra2 is a numerical, two-dimensional VATDM that simulates deposited tephra mass. It is designed for fast computation, as the reference grid can be straightforward (e.g., one single coordinate, where one mass value is produced), and the wind is homogeneous over the whole reference grid.

Table 3.1: Sampled input parameter ranges for Tephra2 and Fall3D. All input parameters are sampled with a uniform distribution excluding spatial resolution, meteorological data, mass, and fall time threshold. The methods used to determine the number of simulations are described in Section 3.3.3.

Parameter Inputs	**Tephra2**	**Fall3D**	**Reference**
Spatial resolution	Specific points at 16 cardinal directions at 5 km, 10 km, and 30 km from the vent	0.125° × 0.125°	
Vertical resolution	37 Pressure levels, 0 to 32 km		Hersbach et al. (2020)
Meteorological data	ERA5 reanalysis		Hersbach et al. (2020)
Median grain size (phi)	–4 to 6		Pistolesi et al. (2015) and Pioli, Bonadonna, and Pistolesi (2019)
Standard deviation grain size (phi)	0.3 to 4.5		Pioli, Bonadonna, and Pistolesi (2019)
Plume height (m, top of plume)	1000 to 40000*		IVESPA database Aubry et al. (2021)
Plume shape (vertical distribution of ash)	Beta distribution $Alpha^{(T2)}$: 0.1 to 10 Beta: 0.1 to 10	Suzuki distribution $Alpha^{(F3D)}$: 3 to 5 Lambda: 1 to 10	Beta – University of South Florida (n.d.) and Wiejaczka and Giachetti (2024) Suzuki – Sulpizio et al. (2012) and Parra et al. (2016)
Density (kg/m^3)	Pumice: 450 to 2000 Lithic: 1000 to 4000	Minimum: 450 to 1000 Maximum: 2000 to 4000	IVESPA database Aubry et al. (2021), University of South Florida (n.d.)
Mass (kg) / MER (kg/s)	2.5×10^9 to 1×10^{13}	Kept constant using Mastin et al. (2009) model	Min/max from Mastin et al. (2009) database
Diffusion	Diffusion coefficient: 100 to 100,000	Horizontal diffusion (m/s^2): 300 to 10,000	Diffusion coefficient: max in Volentik et al. (2010) Horizontal diffusion: range in Byun and Schere (2006)
Fall time threshold	100 to 150,000	N/A	
Plume duration (hours)	N/A	0.1 to 12	Most eruptions <10 hrs in Aubry et al. (2021)
Number of simulations	eFAST – 5000 × 12 winds = 60,000 Sobol' – 265,500 × 12 winds = 3.18 Million	eFAST – 1800 × 12 winds = 21,600 Sobol' – N/A	

Of the fourteen input parameters in Tephra2 (Connor and Hagdorn 2018), four are kept constant throughout the simulations: the grain size range was kept at -7 to 7 phi (block/bomb to extremely fine ash) with 15 particle steps, resulting in one phi per bin; column steps (number of levels in the eruption plume) were kept constant at 100, as recommended by the Tephra2 developers (Connor, Connor, and Saballos 2011); and the Eddy constant (diffusion of fine particles) is always fixed at 0.04 (Suzuki 1983). The remaining ten input parameters are sampled from informed distributions (Table 3.1). We sample mass using log-uniform sampling and use the Mastin et al. (2009) equation (Equation 3.1) to ensure a realistic relationship between plume height (i.e., the top of the plume) and total erupted mass. As mass is needed in Tephra2, we converted volume to mass assuming a density of 2500 kg/m^3, which is a typical density for lithics (Carey and Sparks 1986; Bonadonna, Ernst, and Sparks 1998). We also sample values around the regression line, where H is in km, and $Z \sim N(0, 2)$, to obtain a range of plume heights that account for natural variability.

$$H = 25.9 + 6.64\text{x}log_{10}(\text{Volume}) + Z \tag{3.1}$$

In Tephra2, the vertical distribution of ash within the eruption column is controlled by the plume height (i.e., the top of the plume) and the plume shape input parameters of the beta distribution (alpha$^{(T2)}$ and beta), which determine how the erupted mass is distributed vertically within the column. Different combinations of these parameters control where the majority of ash is released within the plume (e.g., near the vent, mid-column, or near the plume top). This parameterisation is important because particles released at different heights experience different wind speeds and directions, which can influence the resulting ash dispersal and deposition patterns.

We additionally sampled the fall time threshold input parameter dependent on the diffusion coefficient. During preliminary investigations, we found that certain combinations of fall time threshold (FTT) and diffusion coefficient produced unrealistic diffusion patterns (i.e., larger particles diffuse more than smaller particles) (Supplementary Material, Figure B.3). To address this model-inherent issue, we used an equation derived from best-fit combinations of these input

parameters, as identified in the literature (Equation 3.2):

$$FTT = 0.638\text{xDiffusion}^{(1.0792)} \tag{3.2}$$

3.2.2 Fall3D

Fall3D is a numerical, three-dimensional VATDM, that simulates tephra mass both in the atmosphere and on the ground. Compared to Tephra2, Fall3D is more computationally intensive, as it calculates particle transport and deposition across a three-dimensional domain.

Fall3D contains more than 20 input parameters, along with numerous sub-parameters (see Folch et al. (2023) for a full list of input parameters). For this study, we sample nine key input parameters (Table 3.1). All the Fall3D input parameters were sampled uniformly. We excluded certain input parameters from our sampling, as some input parameters depended on multiple sub-parameter choices. For instance, one of the plume shape (i.e., the vertical distribution of ash in the eruption column) distribution options (named "PLUME"), which uses the one-dimensional buoyant plume theory model FPLUME-1.0 (Folch, Costa, and Macedonio 2016), requires an additional seventeen input parameter values or yes/no decisions. To facilitate insights into behaviour between the two VATDM, input parameter distributions were aligned. For example, we kept the total grain size distribution (TGSD) as a Gaussian distribution, which is the only option available in Tephra2, even though Fall3D provides five different input options (Gaussian, BiGaussian, Weibull, BiWeibull, Custom, and Estimate from column height and magma viscosity) (Folch et al. 2023).

Fall3D contains categorical variables that cannot be incorporated into global sensitivity analysis (GSA) like eFAST and Sobol' indices (Saltelli et al. 2008) (Section 3.3.1), as they are either non-numeric or do not have continuous ranges. Thus, a separate strategy was required to explore the influence of these categorical variables. Specifically, we examined distribution (categorised as a categorical variable because multiple distribution types are available), terminal velocity model, aggregation, flux limiter, and horizontal and vertical turbulence models. The definitions of each categorical variable can be found in Folch, Costa, and Macedonio (2009) and Folch et al. (2020). To be able to compare plume shape distributions, specifically Suzuki and Top-Hat, we have normalised the two distributions i.e., the distributions have the same mean and standard deviation. Plume shapes are sampled at 5, 10, and 20 km plume heights. To avoid

unnecessary redundancy, we reduced the sampling of terminal velocity models from seven to three distinct models. This decision was based on preliminary testing, where we observed that seven models could be grouped into three categories based on their similar outputs, allowing us to focus on the most representative models (supplementary material, Figure B.4).

All the other input parameters, including simulation and plume duration, are kept constant. To avoid ignoring these potentially highly influential input parameters, we applied a local sensitivity analysis (LSA) (Figure 3.1) to identify any individual or combinations of categorical variables that appeared to have a large effect on the output variance.

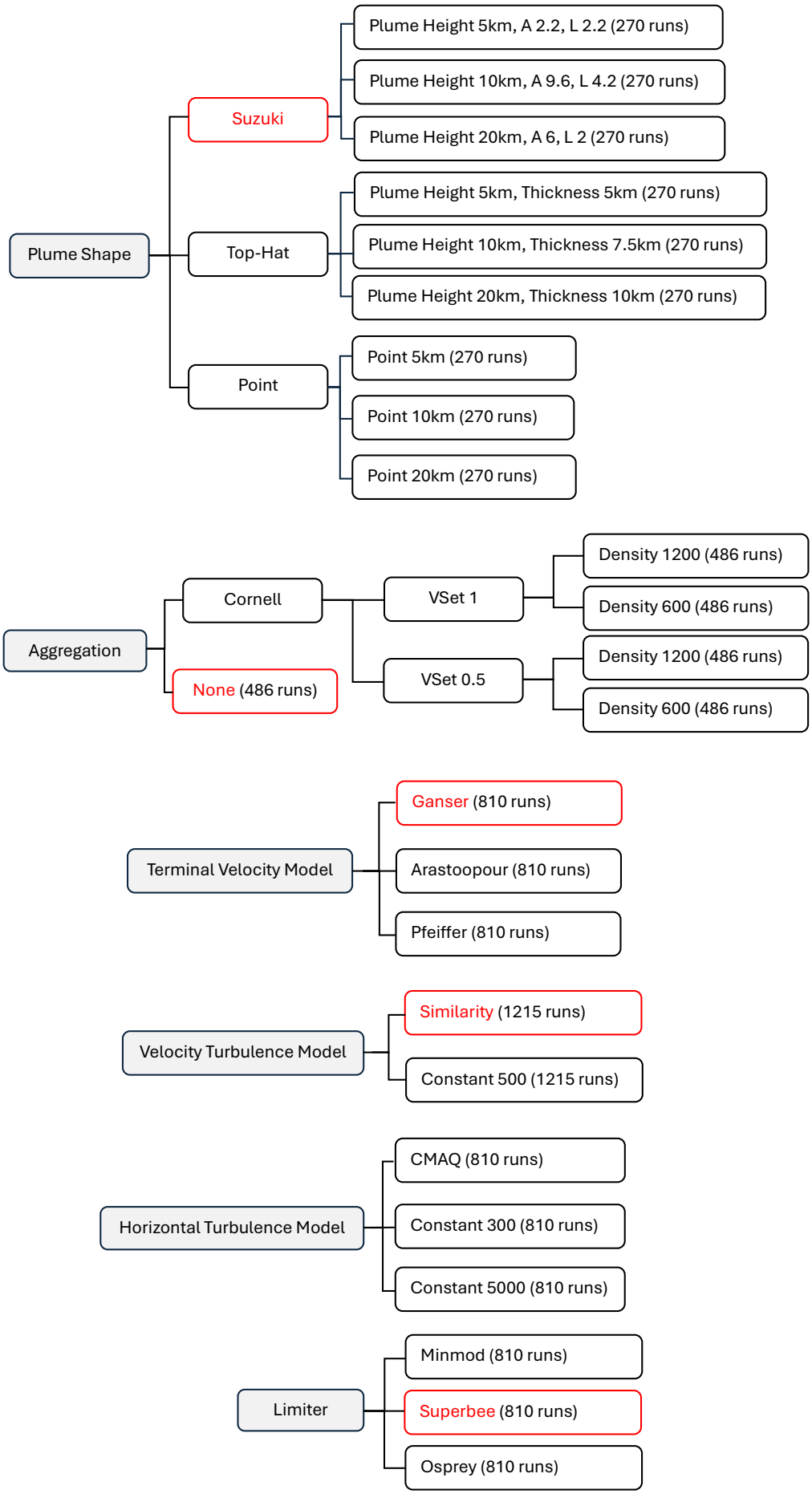


Figure 3.1: Schematic of Fall3D categorical variables, totalling 2430 unique combinations. The red text indicates input variables that were kept constant in the GSA analyses as a result of the preliminary work.

We excluded the MER models from the categorical sampling because, although MER is a key driver of the total amount of deposited mass, total erupted mass is a function of both MER and eruption duration. The correlation between MER and deposited mass is therefore only approximate unless duration is held constant. To reduce the influence of this correlation, we kept the MER input parameter consistent with the Mastin model in our GSA analysis based on findings from Dürig, Schmidt, and Dioguardi (2023).

3.3 Methods

This section outlines our methodological approach, beginning with an introduction to global sensitivity analysis. We then provide an overview of the meteorological conditions examined, followed by the practical implementation of the methodology.

3.3.1 Global Sensitivity Analysis

Sensitivity analysis is a necessary step in understanding the relationship between input parameters and model outputs for complex computational models (Saltelli 2002). Global sensitivity analysis (GSA) examines how much of the variation in model outputs can be attributed to different input parameters across an input parameter space range (Saltelli et al. 2004). GSA stems from the analysis of variance (ANOVA), which is used to determine if an input parameter includes values that provide statistically different outputs (Girden 1992). While ANOVA statistically identifies significant relationships between individual input parameters and model outputs (Montgomery 2017), GSA extends this concept by quantifying the contributions of each input variable - both individually and in combination - to the overall variability of the model output (Mara 2009).

Several studies have conducted sensitivity analyses on VATDMs, focusing specifically on tephra deposition (Table 3.2). From these four studies, total erupted mass (TEM)/mass eruption rate (MER) and plume height emerge as influential input parameters, while plume shape and the TGSD are influential in local sensitivity analysis (LSA), whereby one input parameter is changed at a time, but not as influential in GSA. These previous studies focused on limited subsets of input parameters. In contrast, this study applies GSA to the full input parameter space in a forecasting context (i.e., under no prior assumptions), capturing the effects of complex

Table 3.2: Influential and non-influential input parameters identified in sensitivity analyses of volcanic ash transport and dispersion models for tephra deposition.

Input Parameter	**Scollo, Folch, and Costa (2008)**	**Komorowski et al. (2008)**	**Scollo et al. (2008)**	**Pardini et al. (2022)**
Sensitivity Analysis Type	Local Sensitivity Analysis	Local Sensitivity Analysis	Global Sensitivity Analysis	Global Sensitivity Analysis
Number of Simulations	96 over three VATDM	93	7168	N/A
VATDM Explored	Fall3D, HAZMAP, TEPHRA	HAZMAP	TEPHRA	HYSPLIT
Total Erupted Mass/Mass Eruption Rate	Influential	Influential	Influential	N/A
Plume Height	Influential	Influential	Influential	Most Influential
Plume Shape	Influential	N/A	Non-influential	N/A
Total Grain Size Distribution	Influential	Influential	Influential	Least Influential
Particle Density	N/A	N/A	Non-influential	N/A
Wind Speed and Direction	N/A	Influential	N/A	Influential

interactions that may significantly influence model output.

We apply the two main GSA methods: Sobol' indices (Sobol' 1990) and the extended Fourier Amplitude Sensitivity Test (eFAST) (Saltelli, Tarantola, and Chan 1999). Both Sobol' indices and eFAST provide first-order (S_i^{S} for Sobol', S_i^{F} for eFAST) and total-order (T_i^{S} for Sobol', T_i^{F} for eFAST) indices for each input parameter. Consider a model for a scalar $y = f(x)$, where x is a vector of k probabilistic input parameters. Sobol' indices are defined as:

$$S_i^{\mathrm{S}} = V[E(y|x_i)]/V(y)$$

which represents the proportion of the total variation attributable to variations in x_i. The total-order index is given by:

$$T_i^{\mathrm{S}} = 1 - V[E(y|x_{\sim i})]/V(y)$$

The eFAST method is analogous to Sobol' indices but variance is estimated as Z_i^2 via Fourier coefficients as described in the Appendix (Appendix B). The first-order index is expressed as:

$$S_i^{\mathrm{F}} = Z_i^2/Z_{total}^2$$

while the total-order index is given by:

$$T_i^{\mathrm{F}} = 1 - (Z_{\sim i}^2 / Z_{total}^2)$$

where $Z_{\sim i}$ represents the summed complementary set of parameters (i.e., all parameters except x_i). See Appendix B for a more in-depth explanation.

For both Sobol' and eFAST, larger S_i and T_i values indicate a larger influence on model output variance. Given the high dimensionality and number of input parameters in the models, we use a 5% significance threshold for S_i rather than considering the relative number of input parameters per model (Puy et al. 2023). Sobol' indices provide a more comprehensive analysis by also providing second- and higher-order indices (Saltelli et al. 2008) (Appendix B). eFAST requires substantially fewer simulation runs to explore the entire input space compared to Sobol' indices, while still providing valuable sensitivity information (Saltelli et al. 2008). This meant that eFAST could be used for both our GSA of Tephra2 and Fall3D, while Sobol' indices is only applied to Tephra2.

3.3.2 Meteorological Conditions

It is known that meteorological conditions play a major role in the deposition of ash (Bursik 2001; Bonadonna and Costa 2013; Mulena et al. 2016). No equivalent method exists for the stochastic sampling of wind, thus, we follow the advice of Phillips et al. (2023) and consider 12 meteorological conditions, corresponding to the 12 synoptic patterns in Aotearoa New Zealand (Kidson 2000). These weather types are nationwide patterns that fall into three regimes: troughs (low-pressure systems), the trough group, high-pressure systems to the north and west-east flow in the south, the zonal group, and blocking patterns with high-pressure systems to the south, the blocking group.

Using stratified sampling (Cochran 1977), we treated the 12 Kidson types as strata and selected one 24-hour period per Kidson type from February 2016 (Figure 3.2). Each selected period was chosen to ensure that the Kidson type remained consistent throughout. We aggregate the GSA results from the 12 Kidson types to produce one overall result, as to provide a general overview of input parameter influence across different meteorological conditions. As different meteorological conditions may influence our GSA analysis, we also explore GSA outputs in both

weak and strong wind fields (Figure 3.2). Since the maximum eruption duration in our study is 12 hours or less, we allowed an additional 12 hours for finer grained particles to deposit. For Tephra2, the first hour in each Kidson type was used as the model does not consider temporal variability in weather.

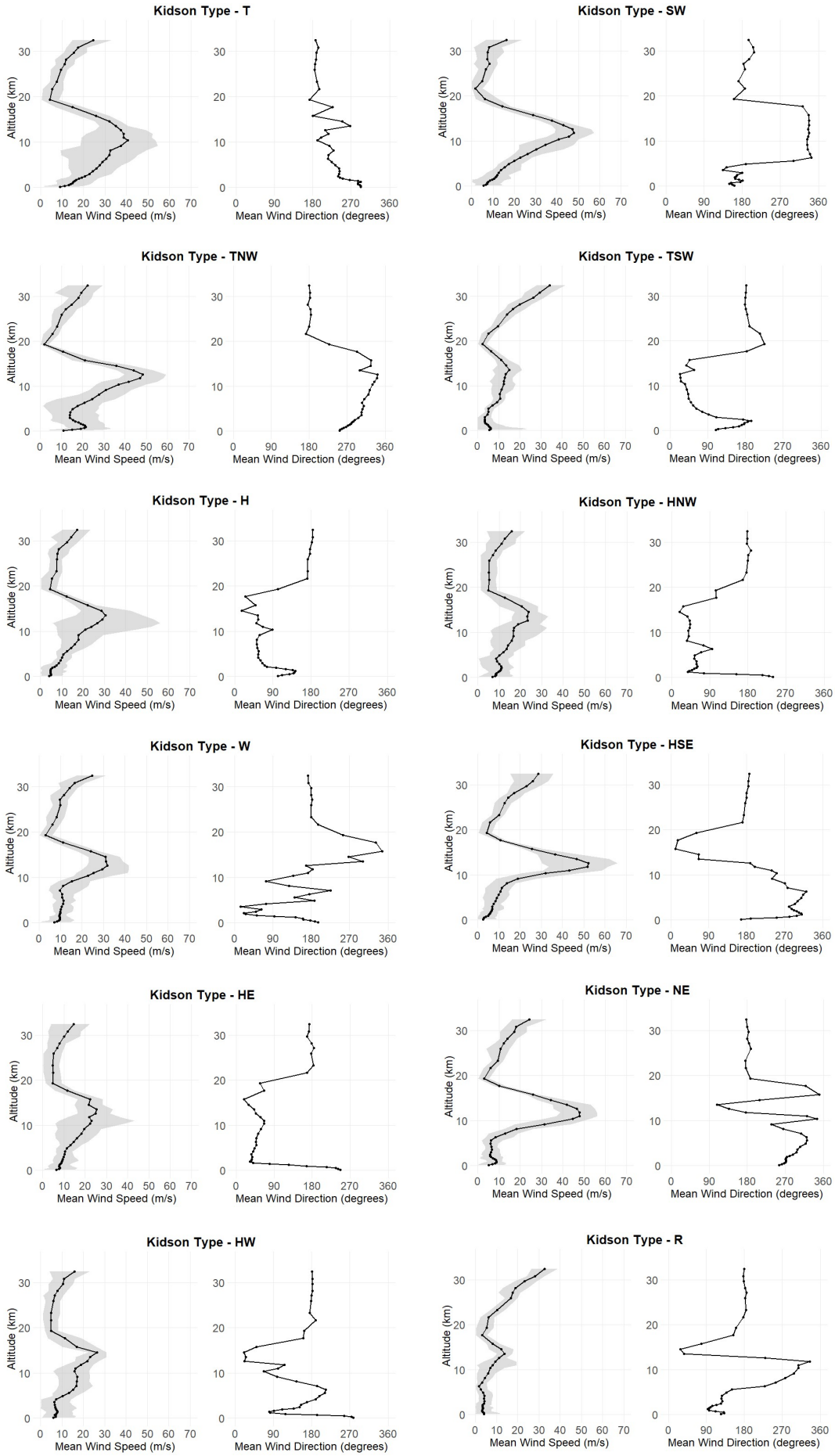
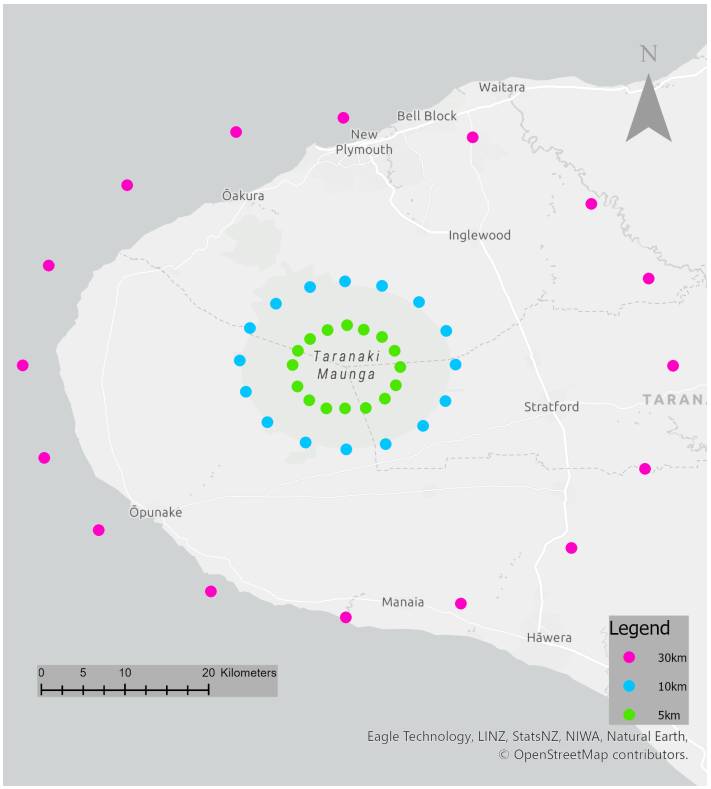


Figure 3.2: Mean wind speeds (over all areas and time) and mean wind directions (from North) at different altitudes for the 12 24-hour long Kidson Types at Taranaki Mounga (39.2968° S, 174.0634° E), calculated over the selected 24-hour periods. Grey areas highlight the minimum and maximum wind speeds at different altitudes calculated over the selected 24-hour periods. Trough group: T, SW, TNW, TSW. Zonal group: H, HNW, W. Blocking group: HSE, HE, NE, HW, R.

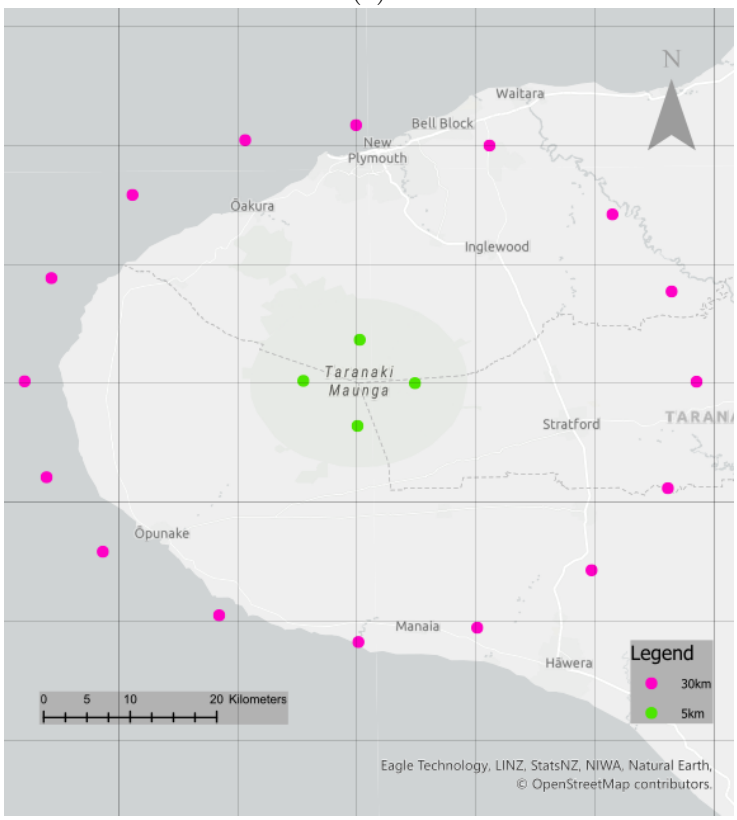
Tephra2 uses minimal meteorological data, requiring only a text file of atmospheric height, wind speed, and direction that is assumed constant across the computational domain (Connor, Connor, and Saballos 2011). Fall3D uses more complex meteorological data such as numerical weather prediction and reanalysis datasets, interpolating the required meteorological data (wind field, air temperature, friction velocity, atmospheric boundary layer height, and Monin-Obukhov length (turbulence height)) from the original grid driving the meteorological models to the computational domain (Folch et al. 2020). For this study, we use the European Centre for Medium-Range Weather Forecasts (ECMWF) hourly ERA5 reanalysis dataset, which spans from 1979 onwards with a vertical resolution of 37 pressure levels (Hersbach et al. 2020). For Fall3D, we use the ERA5 reanalysis data in NetCDF format with a spatial grid at half the resolution (0.125° x 0.125°) (Folch, Costa, and Basart 2012) (Figure 3.3a). For Tephra2, we extract the necessary data from the ERA5 reanalysis data and run simulations on a spatial "grid" of the sampled locations and eruptive vent (Figure 3.3b).

3.3.3 Practical Implementation

We constrain our approach to an imminent eruption (i.e., meteorological conditions are known and assumed to be as accurate as possible) at Taranaki Mounga (Mount Taranaki), Aotearoa New Zealand where there is an estimated 30-50% chance of an explosive eruption in the next 50 years (Damaschke, Cronin, and Bebbington 2018). We sample deposit densities at locations across the region within 30 km of the vent, at 5 km, 10 km, and 30 km distances in 16 cardinal directions, with Fall3D sampling restricted to four cardinal directions at 5 km and no sampling at 10 km due to the grid size (Figure 3.3). A similar approach of using 16 cardinal directions has been employed in previous studies (Macedonio, Costa, and Folch 2008). While Taranaki Mounga is used here as a case study, our study can be applied to any volcano worldwide, provided that meteorological data are available.



(a)



(b)

Figure 3.3: Sampling locations at Taranaki Mounga (Mount Taranaki), Aotearoa New Zealand. (a) Tephra2 sampling locations at 5 km (green), 10 km (blue), and 30 km (purple) from the vent in 16 cardinal directions. (b) Fall3D sampling locations at 5 km in 4 cardinal directions (green) and 30 km in 16 cardinal directions (purple). The grey grid represents the ERA5 spatial resolution at 0.125° x 0.125°. Both models do not differentiate between tephra deposition on land or water.

For our GSA, we use the `sensobol` package in R for the Sobol' indices (Puy et al. 2021), and investigate up to third-order interactions. The the number of simulation runs generated for third-order Sobol' indices follows: $(N(k+2))+(k!/(2!(k-2)!)\text{ x }N)+(k!/(3!(k-3)!)\text{ x }N)$, where N is the initial sample size of the base sample matrix (typically varying from a few hundred to a few thousand Saltelli et al. 2008), and k is the number of input parameters. For Tephra2, we use $N = 1500$, resulting in 265,500 runs (per wind). We do not conduct Sobol' indices GSA on Fall3D due to computational limitations. Running the model a sufficient number of times to obtain a valid GSA result would have required several months worth of simulations.

We use the `SALib` package in Python to conduct our eFAST GSA (Herman and Usher 2017). For eFAST, the number of simulations run is N x k, where N is the number of samples to generate and k is the number of input parameters. For Tephra2 we use $N = 500$ ($k = 10$, resulting in 5000 runs per wind) and for Fall3D $N = 200$ ($k = 9$, resulting in 1800 runs per wind). The different values of N between eFAST and Sobol' for Tephra2 is due to Sobol' indices needing more repetitions to obtain second- and third-order indices. The different number of simulation runs between Tephra2 and Fall3D is based on the computational time it takes to run Fall3D. To conduct GSA on Fall3D, we used the Aotearoa New Zealand eScience Infrastructure (NeSI) High-Performance Computing (HPC) system.

3.4 Results

Here we show the results of the GSA analysis sampling tephra deposition at three distances and in 16 cardinal directions around Taranaki Mounga for Tephra2 (Figure 3.3a), and two distances in four and 16 cardinal directions for Fall3D (Figure 3.3b) based on 12 different wind scenarios (Figure 3.2).

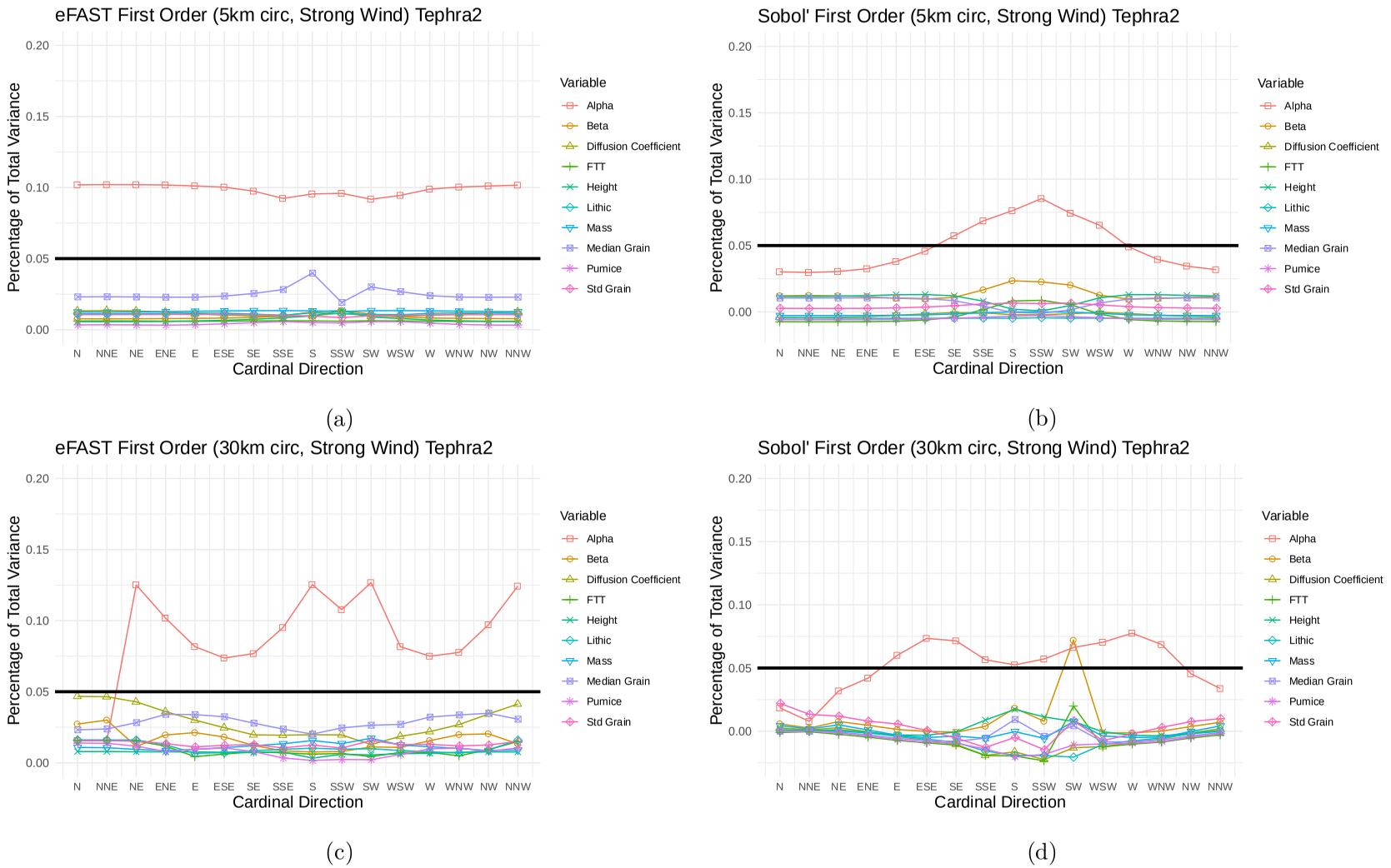


Figure 3.4: S_i values from a strong wind scenario for Tephra2. Each coloured line and point shape represents an input parameter examined in the GSA analysis. Points indicate S_i values at each cardinal direction. Points above the solid black line represent significant S_i values (more than 5%). (a) 5 km Sampling distance via eFAST; (b) 5 km Sampling distance via Sobol' indices; (c) 10 km Sampling distance via eFAST; (d) 10 km Sampling distance via Sobol' indices; (e) 30 km Sampling distance via eFAST; (f) 30 km Sampling distance via Sobol' indices.

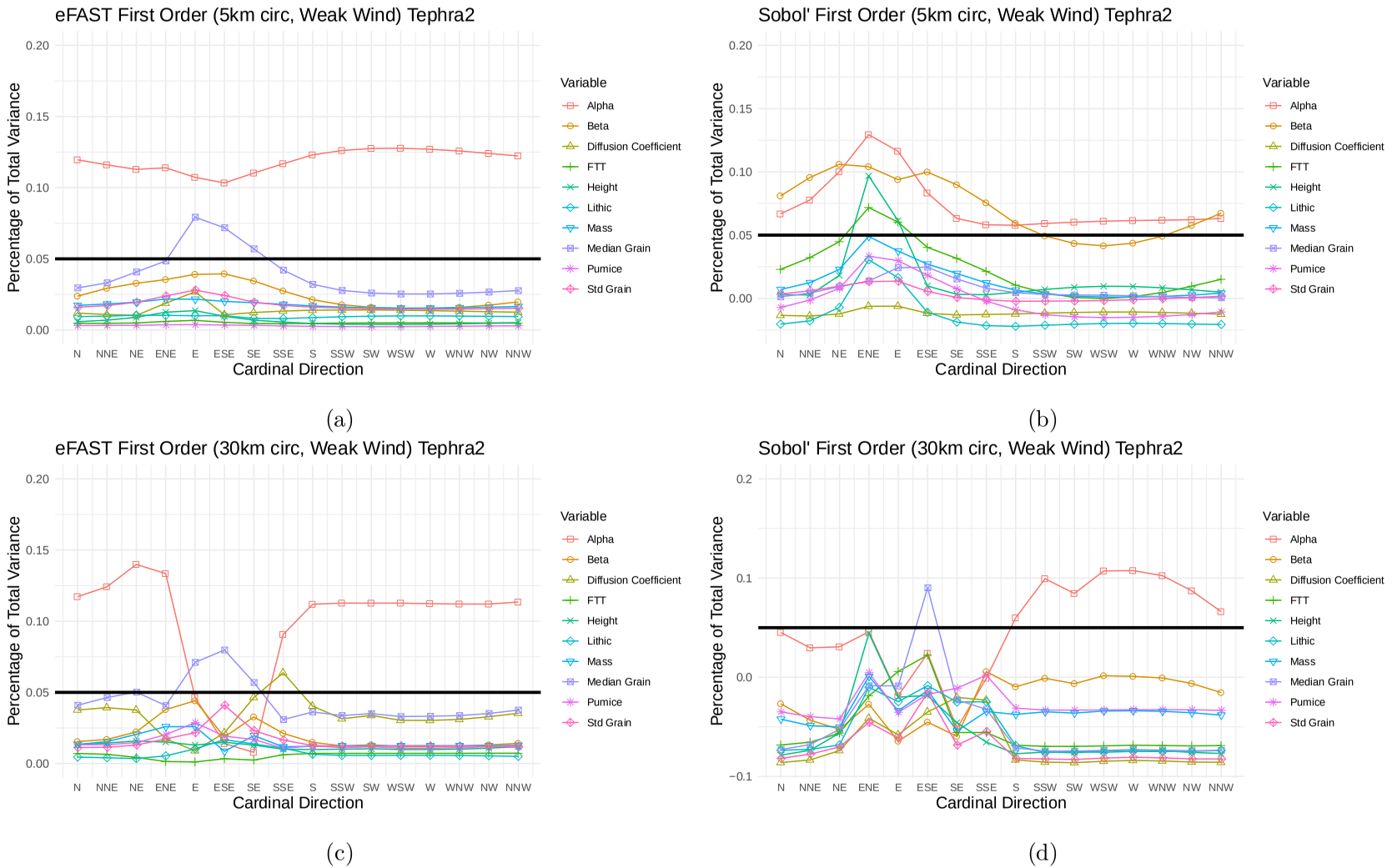


Figure 3.5: S_i values from a weak wind scenario for Tephra2. Each coloured line and point shape represents an input parameter examined in the GSA analysis. Points indicate S_i values at each cardinal direction. Points above the solid black line represent significant S_i values (more than 5%). (a) 5 km Sampling distance via eFAST; (b) 5 km Sampling distance via Sobol' indices; (c) 30 km Sampling distance via eFAST; (d) 30 km Sampling distance via Sobol' indices.

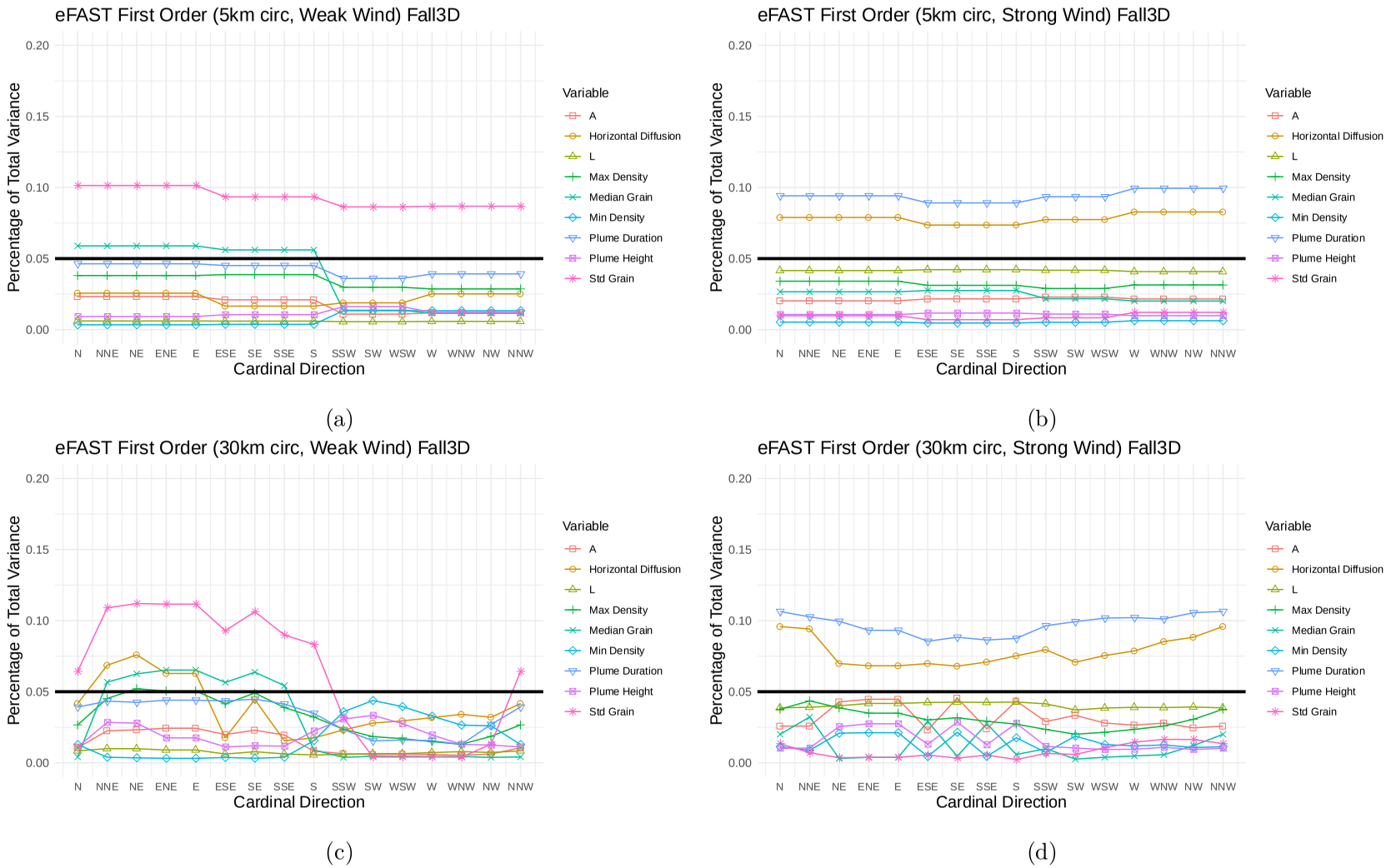


Figure 3.6: S_i values via eFAST from a weak and strong wind scenario for Fall3D. Each coloured line and point shape represents an input parameter examined in the GSA analysis. Points indicate S_i values at each cardinal direction. Points above the solid black line represent significant S_i values (more than 5%). (a) Weak wind at 5 km sampling distance; (b) Strong wind at 5 km sampling distance; (c) Weak wind at 30 km sampling distance; (d) Strong wind at 30 km sampling distance.

3.4.1 Weak vs Strong Wind

We first look at the Sobol' indices and eFAST outputs for two wind fields that convey a weak and strong wind (referred to here as a weak and strong wind scenario; Kidson type HE and T respectively (Figure 3.2)).

For Tephra2 and Fall3D, we found that the output variance is concentrated over fewer input parameters in strong wind fields, while more input parameters become significant when the wind is weaker (Figure 3.4,3.5,3.6). For Tephra2, only plume shape (alpha$^{(T2)}$ and beta) remains significant across both Sobol' and eFAST indices in the strong wind scenario (Figure 3.4). For Fall3D, plume duration and horizontal diffusion are the only parameters deemed significant under the strong wind scenario (Figure 3.6).

In the weak wind scenario for Tephra2, plume shape (alpha$^{(T2)}$) is consistently significant, especially off axis of the major direction of tephra deposition (ENE to ESE), while its influence decreases with distance (Figure 3.4). The diffusion coefficient remains significant across all distances in eFAST but only in distal areas (30 km) for Sobol' indices. Plume height, fall time threshold (FTT), plume height, and plume shape (beta) are significant in proximal areas (5 km) in Sobol' indices, but lose significance distally. In eFAST the diffusion coefficient remains significant in distal areas.

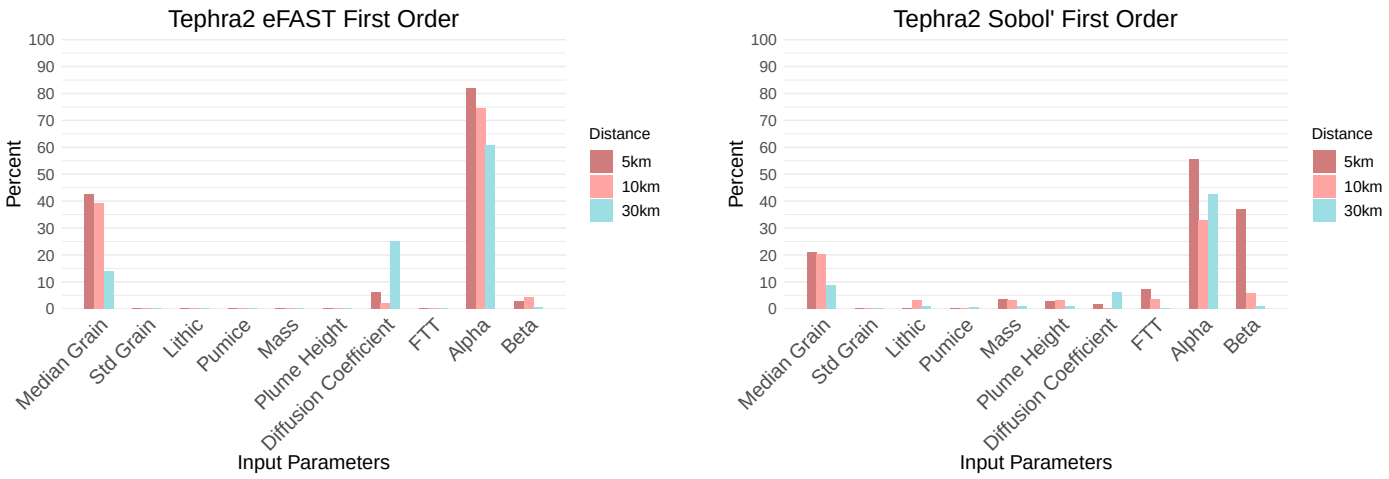
For Fall3D, in weak wind conditions, median and standard deviation grain size are the only significant parameters in proximal areas, while the diffusion coefficient gains significance in distal regions (Figure 3.6).

3.4.2 Tephra2

In the following sections, we look at the first-order sensitivity indices: S_i^{S} (Sobol') and S_i^{F} (eFAST), as well as the total-order indices: T_i^{S} (Sobol') and T_i^{F} (eFAST), over the whole GSA space, providing insights into the average trends over all meteorological conditions. Additionally, we also examine the S_i^{S}, S_i^{F}, T_i^{S} and T_i^{F} values in the major direction of tephra deposition, selecting values for the direction closest (out of 16 cardinal directions) to the predominant wind direction. It is important to investigate whether the results of our GSA vary in these regions of heightened tephra deposition.

First-Order Indices

The results of the GSA looking at the first-order indices (S_i) for Tephra2 for all sampling locations are shown in Figure 3.7.



(a) (b)

Figure 3.7: Proportion of first-order S_i outputs for Tephra2 of input parameters that reach a 5% significance threshold. All winds are included. (a) S_i via eFAST incorporating 16 cardinal directions and sampling points. (b) S_i via Sobol' indices incorporating 16 cardinal directions and sampling points.

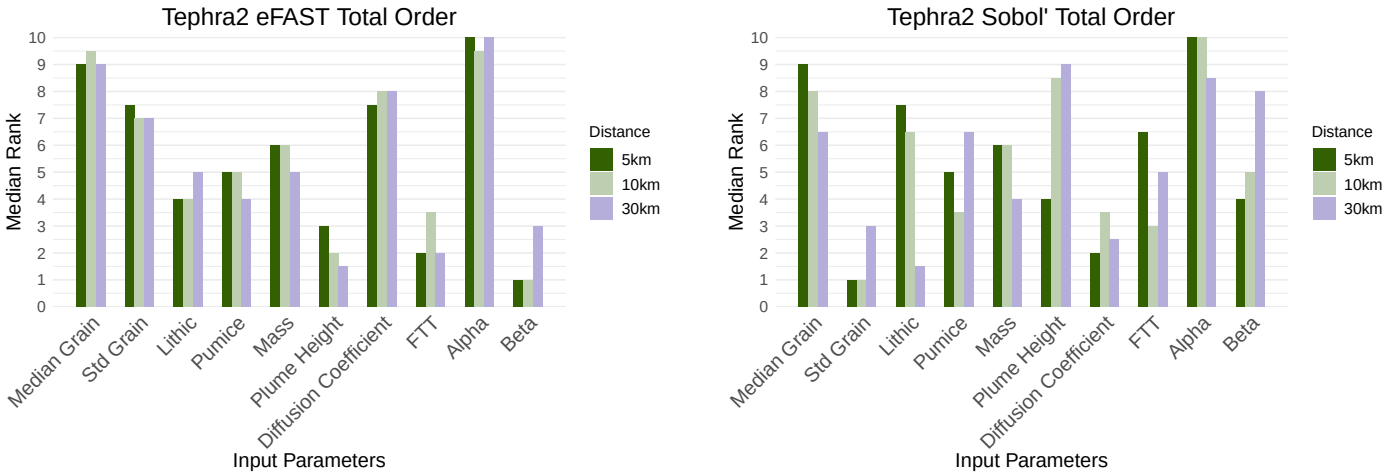
The most notable difference is in plume shape alpha$^{(T2)}$ (which controls how much mass is contributed to the upper part of the eruption column), which significantly influences output variance in 61-82% of simulations in eFAST but only 33-55% in Sobol' indices. eFAST shows a decline in the significance of median grain size and plume shape alpha$^{(T2)}$ with distance, whereas Sobol' indices reveal no consistent trend for plume shape alpha$^{(T2)}$.

Grain size standard deviation, lithic and pumice density, mass, plume height, and FTT are not significant first-order input parameters in either GSA method, suggesting their influence on tephra deposition is indirect or dependent on interactions with more dominant parameters like plume shape.

Examining the significant input parameters along the main tephra deposition axis, median grain size and plume shape (alpha$^{(T2)}$ and beta) remain the primary drivers of output variance.

Total-Order Indices

For the purpose of our analysis, the T_i are discussed in terms of rank, ranging from one (representing the lowest T_i value) to ten (representing the highest T_i value), as we investigate ten input parameters in Tephra2. Higher T_i indicates a greater influence on model output variance when considering interactions with all other input parameters.



(a) (b)

Figure 3.8: Total-order T_i outputs for Tephra2. Fractional values result from averaging over multiple winds. All winds are included. (a) The median rank of T_i^{F} via eFAST incorporating 16 cardinal directions. (b) The median rank of T_i^{S} via Sobol' indices incorporating 16 cardinal directions.

Analysis of total-order indices (T_i) in Tephra2 revels consistent trends (Figure 3.8). Median grain size and plume shape (alpha$^{(\mathrm{T2})}$) consistently rank highest.

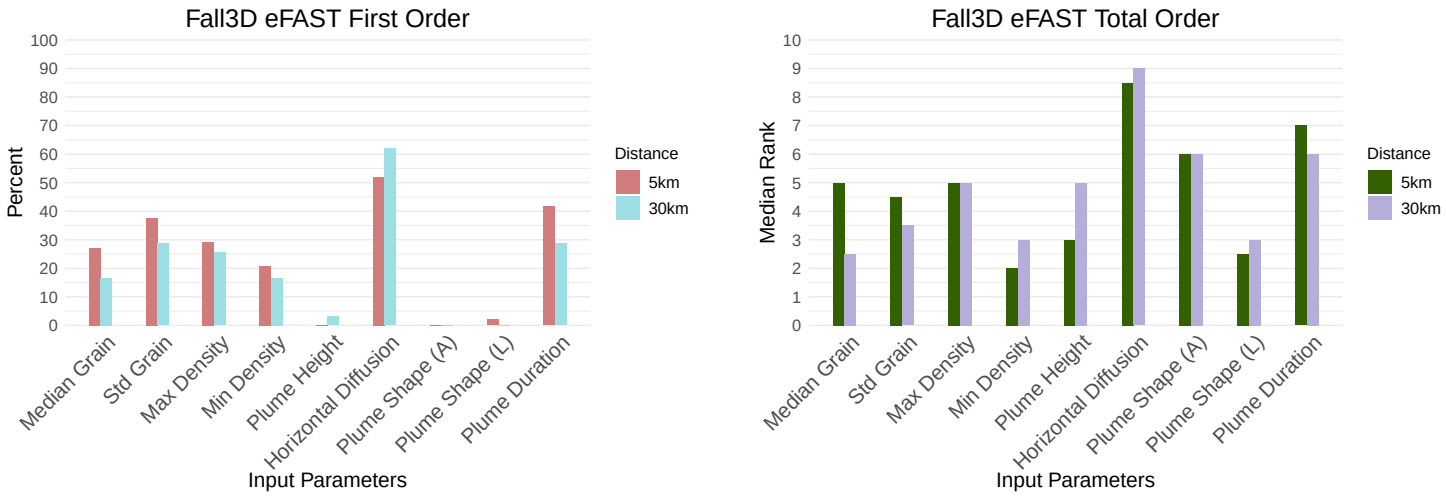
Discrepancies between eFAST and Sobol' indices appear for grain size standard deviation, plume height, diffusion coefficient, FTT, and plume shape (beta). In eFAST, grain size standard deviation and diffusion coefficient rank relatively high (7-7.5 and 7.5-8, respectively) but drop in Sobol' indices (1-3 and 2-3.5). Conversely, FTT and plume shape (beta) increase in rank in Sobol' indices, shifting from 2-3.5 to 3-6.5 and from 1-3 to 4-8, respectively. These trends persist when considering T_i values specifically along the main axis of tephra deposition.

3.4.3 Fall3D

For Fall3D, we look at the first-order sensitivity indices S_i^{F} (eFAST) and total-order indices T_i^{F}, over the whole GSA space, providing insights into the average trends over all meteorological

conditions. Additionally, we also examine the S_i^{F} and T_i^{F} values in the major direction of tephra deposition. This selection is based on four cardinal directions at a 5 km sampling distance or 16 cardinal directions at a 30 km sampling distance.

First and Total-Order Indices



(a) (b)

Figure 3.9: Proportion of first-order S_i and total-order T_i outputs for Fall3D. Fractional T_i values result from averaging over multiple winds. All winds are included. (a) S_i^{F} via eFAST incorporating all cardinal directions and sampling points. (b) The median rank of T_i^{F} via eFAST summed in all cardinal directions.

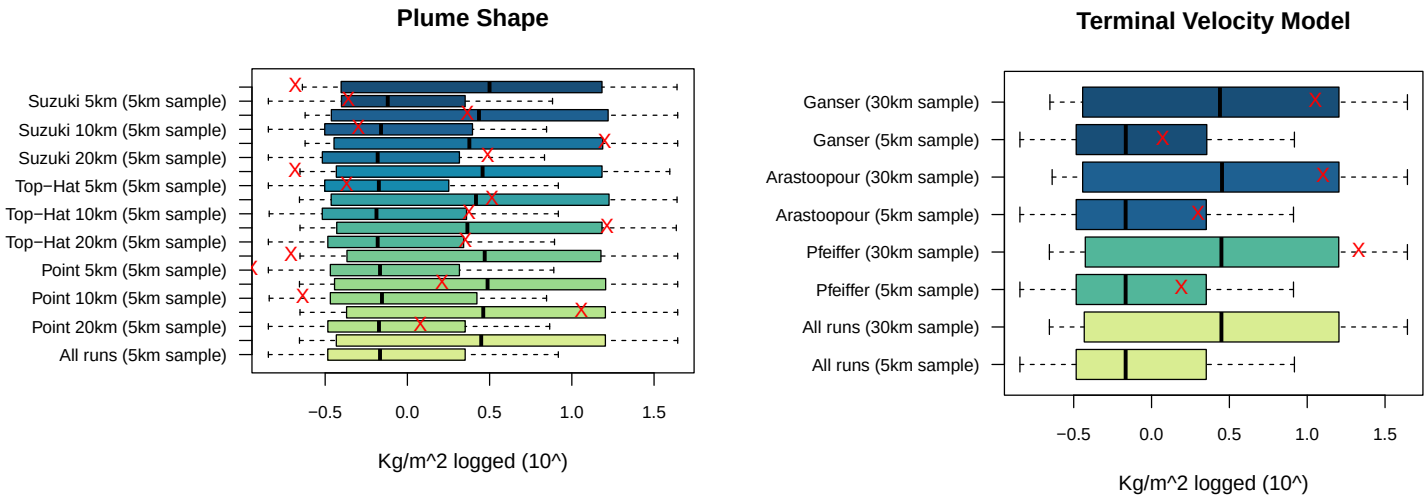
For Fall3D, first-order (S_i) outputs indicate that horizontal diffusion is the most significant input parameter, with its influence increasing with distance from the vent (Figure 3.9). Median and standard deviation grain size, minimum and maximum density, and plume duration also contribute notably to output variance. In contrast, plume height and plume shape have minimal influence. These trends remain consistent when considering only sampling locations along the main axis of tephra deposition.

Total-order indices (T_i) for Fall3D also show that horizontal diffusion is the most significant input parameter (Figure 3.9). Plume shape parameters alpha$^{(\mathrm{F3D})}$ and lambda gain more influence compared to their S_i values, with alpha$^{(\mathrm{F3D})}$ becoming a top-ranked input parameter, though lambda remains among the least influential, similar to minimum density.

When considering only the major axis of tephra deposition, ranking trends remain similar. However, plume height decreases in rank, alongside the low ranking of maximum density and

plume shape lambda.

There is no overall variation in deposit density distribution when considering categorical variables in Fall3D when input parameters are perturbed along the major tephra deposition axis. However, plume shape distributions introduce slight variations in median deposit densities (Figure 3.10).



(a) (b)

Figure 3.10: Categorical simulation outputs with logged deposit density in the major direction of tephra deposition. Red crosses are individual runs on a base run where only the categorical input is changed.

Individual runs based on a baseline scenario, where only categorical inputs were altered, reveal that the largest variations in deposit density (red crosses) occur within the plume shape distribution category, particularly for plume heights exceeding 10 km. Minor variations also appear in individual runs for terminal velocity models (Figure 3.10), while aggregation, vertical and horizontal turbulence models, and limiter categories show no significant density variation within their respective categories.

3.5 Discussion

This section discusses the key findings in our GSA analysis, including an exploration of higher-order Sobol' indices. We also examine the role of categorical variables, address the study's limitations, and consider the use of our VATDMs Tephra2 and Fall3D.

3.5.1 GSA

The objective of this study is to understand which input parameters contribute the most and least amount of variance in model outputs, which in turn can correctly represent uncertainty in forecasts, as well as reduce the number of simulations needed to produce probabilistic forecasts or ensembles. Our GSA analysis has shown that input parameters total grain size distribution, plume shape and diffusion (diffusion coefficient and horizontal diffusion) have significant first- and total-order influences on the output variance in both Tephra2 and Fall3D. In addition, plume duration (only an option in Fall3D) was also a significant input parameter. The other input parameters, mass (Tephra2 only), particle density, plume height, and FTT (Tephra2 only), were found to have only total-order influences on output variance for Tephra2 and Fall3D. These findings highlight that when designing forecasts, particular attention should be paid to significant input parameters as they contribute most to the variance in model outputs. These results differ from previous research, as Scollo et al. (2008) (Table 3.2) found that in VATDM TEPHRA, using Sobol' indices, lithic and pumice densities and plume shape were not significant input parameters.

We conducted additional testing in our GSA by examining different sensitivity thresholds for S_i^{S} and S_i^{F} — 3% and 10% — compared to the 5% threshold originally used, to assess whether this would affect our conclusions. Results showed that even though the frequency of significant input parameters increased at 3% significance and decreased at 10%, the overall pattern was the same.

Our GSA results could inform the design of an optimal sampling or simulation strategy when the number of simulation runs is constrained. Both first- and total-order influences of input parameters can be utilised to identify which input parameters may be excluded from sampling designs. First-order indices, which correspond to the direct influence of individual input parameters, are particularly advantageous because their uncertainty is easier to quantify compared to the more complex and less well-understood interactions between input parameters.

3.5.2 Sobol' Second and Third-order Indices

If the first-order interactions at a sampling point sum to one, it indicates no interaction between input parameters (Sobol' 1990). If they do not sum to one, second and third-order interactions

account for the remaining proportion of the total-output variance. We found that the sum of first-order indices is lower (around 0.1) in stronger wind fields and along the major deposition axis. This indicates that interactions between input parameters play a larger role in controlling ash deposition.

Conversely, in weaker wind fields and opposite the major direction of tephra deposition, the sum of the first-order interaction terms is larger (around 0.5). This suggests that model behaviour in weaker winds is more predictable and may require fewer simulations to characterise.

For both strong and weak wind fields, the sum of second- and third-order interactions decreases with distance from the vent. This trend likely reflects the influence of input parameters, such as diffusion, which become increasingly more significant at further distances from the vent. At greater distances, the mass of ash and the prevalence of finer-grained particles introduce more variability, potentially leading to complex interactions beyond the third order.

We also examined which input parameters and their interactions had the largest index values. By selecting the top five index values for each Kidson Type (12), by each sampling distance (3), and for each order (3), we identified a total of 540 high-value indices (5 x 12 x 3 x 3 =540). The first-order input parameter, plume shape alpha$^{(T2)}$, had the highest count among these, accounting for 14.6% (79/540). In comparison, 32% of the largest indices were related to median grain size, including 9.4% (51/540) as first-order contributions and the remainder from second- and third-order interactions with particle density, plume shape (alpha$^{(T2)}$ and beta), diffusion coefficient, plume height, FTT, standard deviation of grain size, and mass, ranging from 9% to 0.19%.

3.5.3 Categorical Variables

We found that particle aggregation did not significantly affect the outputs of Fall3D. This differs from expectations, as most fine volcanic ash tends to settle out as particle aggregates (Carey and Sigurdsson 1982), often leading to secondary thickening of deposits (e.g., Tsuji, Nishizaka, and Ohnishi 2020). Aggregation in Fall3D has been validated with the 17 September 1992 Mount Spurr and the 18 May 1980 Mount St. Helens eruptions, which showed strong agreement with observed deposit values (Folch et al. 2010). So for hindcasting, the use of aggregation models may improve model output accuracy. However, for a future unknown eruption, the extent of

particle aggregation is uncertain, making it unclear whether particle aggregation model input parameter options will improve model output accuracy.

One possible explanation for the discrepancy between our results and previous studies is the consistency of the TGSD and particle density values used in our simulations. Research has shown that aggregation is sensitive to factors such as particle density, the TGSD of non-aggregated particles, and the specific aggregation model applied (Poret et al. 2017; Beckett et al. 2022).

Plume shape significantly influenced the outputs of Fall3D, with variations within individual plume shape distributions creating differences in deposit density. This finding aligns with the results of our GSA, which identified plume shape as a significant input parameter impacting output variance in both Tephra2 and Fall3D.

The other categorical input parameters: limiter, horizontal and vertical turbulence models, and terminal velocity models do not significantly affect output deposit variations. This suggests that their selection does not strongly influence the overall model output. Having multiple options within each categorical variable may still provide flexibility to achieve a closer fit during hindcasting by allowing adjustments to better match observed data for a specific eruption. However, while greater model expertise can help inform input parameter use, multiple options within categorical variables still introduce uncertainty for forecasting, as these choices may not accurately represent future conditions.

3.5.4 Limitations

We represented the variation in wind in the Taranaki region of Aotearoa New Zealand via Kidson types. However, the use of only 12 unique winds is unlikely to have captured the full range of meteorological conditions - which is impossible to do so, as wind is continuous - meaning that our GSA may not have fully captured input parameter interactions. The Kidson types are also based on country-wide patterns, and may not reflect all meteorological patterns seen at regional to local scales. A logical extension to this work would therefore be to expand the wind input variability, either through multiple days under each Kidson type or by stratifying wind sampling with location-specific patterns.

A limitation of this study is that the GSA focused only on tephra deposition. As VATDMs

such as Fall3D are also used to forecast atmospheric ash concentrations (e.g., for aviation safety, Garcia et al. 2023; Folch, Mingari, and Prata 2022; Beckett et al. 2024), the relative influence of input parameters may differ for atmospheric ash loading.

Due to time and computational constraints, the GSA for Fall3D used a fixed MER model (i.e., Mastin et al. 2009). Varying the MER model in future GSA may significantly influence the results and warrants further investigation.

Both eFAST and Sobol' indices analyses identified the same key input parameters that significantly influence output variance in both Tephra2 and Fall3D. eFAST was used for both Tephra2 and Fall3D due to its computational efficiency. While Sobol' indices can provide more detailed insights into higher-order interactions, the occurrence of negative Sobol' indices suggests that eFAST is a more robust approach, particularly when analysing models with a high-dimensional input parameter space.

Sobol' indices produced negative S_i^{S} values (55% of the time). Negative S_i^{S} values can occur due to sampling variability due to an insufficient sample size (Saltelli et al. 2010). Even with over 260,000 simulations per wind field in Tephra2, negative S_i^{S} indicate that the number of samples to generate should be increased as S_i^{S} by definition should not be negative. However, increasing the number of simulations further may not get rid of negative values entirely, due to the high-dimensional input parameter space. Scollo et al. (2008) also encountered this limitation, reporting negative Sobol' indices with only 4,608 simulation runs.

3.5.5 Considerations of the Models used (Tephra2 and Fall3D)

The type of GSA analysis presented in this paper can be applied to any volcanic hazard model. We decided to test Tephra2 and Fall3D as they have different benefits and limitations, especially in forecasting. Tephra2 is computationally efficient, with a single simulation typically taking less than a second, allowing thousands of simulations to be run in minutes on a standard computer. In contrast, Fall3D is more computationally intensive; a single simulation took between 10 and 18 minutes for short-duration eruptions (less than 12 hours) with coarse grid sizes, even when using batch processing on a supercomputer system (NeSI). These differences in simulation time are largely due to the broader application of Fall3D, which simulates both ash deposition and atmospheric concentrations and incorporates large meteorological datasets.

Tephra2 simplifies the user experience by directly embedding different physical processes, such as turbulence and diffusion calculations within the code. For example, atmospheric diffusion is assumed to be Gaussian over the whole computational domain (Connor, Connor, and Bonadonna 2008). Other notable assumptions of the model include a Gaussian particle size distribution, particles of every size are well-mixed throughout the entire eruption plume, and a layered atmosphere where windspeed and direction are constant within horizontal layers but allowed to vary between them (Connor, Connor, and Bonadonna 2008). In addition, Tephra2 considers a consistent wind profile, which means that over large domains the winds cannot be accurately represented by a single profile. These simplifications enable the model to achieve short computing times, allowing thousands of simulations to be run in just a few minutes on standard computers.

Fall3D, by contrast, provides users the flexibility to define and manipulate more aspects of the model compared to Tephra2, such as particle shape and limiters. Additionally, Fall3D is usually used for simulating atmospheric concentrations of volcanic ash - a major concern for Volcanic Ash Advisory Centers (VAACs) (Beckett et al. 2024) - in addition to deposition.

3.6 Conclusions

While VATDM models such as Tephra2 and Fall3D can retrospectively simulate past eruption behaviours, identifying the key input parameters that significantly influence output variance in the context of forecasting is crucial for improving hazard forecast reliability. During or immediately before an eruption, VATDM users are constrained in the number of simulations that can be run in a fast-evolving situation. Our GSA revealed that grain size distribution, diffusion, and plume duration input parameters have the most significant impact on model output variance. Contrary to previous studies, our analysis demonstrated that the plume shape input parameters also significantly affect model output variance.

Therefore, when time is limited, VATDM users should focus on running simulations that prioritise varying these influential input parameters. Identifying these key input parameters allows users to direct efforts towards quantifying their variability, including understanding the correlations between them, which will improve the overall accuracy of the simulations. By

more accurately quantifying distributions of significant input parameters, and leveraging limited simulation resources to best exploit these, more reliable hazard forecasts and better preparedness strategies can be created during volcanic eruptions. Additionally, this sensitivity information can also underpin the development of long-term tephra hazard assessments, where pre-established input parameter ranges are established in advance. As thousands of simulations are required to conduct a GSA, alternative strategies for assessing the importance of VATDM input parameters may include the use of emulators, which can be particularly useful for models with large computational demands (e.g., Harvey et al. 2018).

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the student and the student's main supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the student's contribution as indicated below in the Statement of Originality.

Student name:	Emmy Elizabeth Scott
Name and title of main supervisor:	Dr Melody Whitehead, Senior Research Officer and Senior Lecturer
In which chapter is the manuscript/published work?	Chapter Four

Describe the contribution that the student and members of the supervisory team have made to the manuscript/published work:[1]

Emmy Scott: Conceptualisation, Methodology, Data curation, Writing - original draft
Melody Whitehead: Conceptualisation, Methodology, Writing - review and editing
Mark Bebbington: Conceptualisation, Methodology, Writing - review and editing
Stuart Mead: Conceptualisation, Methodology

Please select one of the following three options:

○ **The manuscript/published work is published or in press**

Please provide the full reference of the research output:

◉ **The manuscript is currently under review for publication**

Please provide the name of the journal:
Bulletin of Volcanology

○ **It is intended that the manuscript will be published, but it has not yet been submitted to a journal**

Student's signature:	Emmy Scott Digitally signed by Emmy Scott Date: 2025.11.12 11:04:32 +13'00'	Main supervisor's signature:	Melody Whitehead Digitally signed by Melody Whitehead Date: 2025.11.12 12:13:40 +13'00'

This form should be placed at the beginning of each relevant thesis chapter.

[1] Refer to the Massey University Publishing and Authorship guidelines (OneMassey for staff, Stream for students) and/ or Contributor Roles Taxonomy (CRediT) guidelines for guidance.

Chapter 4

Impact of input parameter uncertainty on tephra deposition forecasts: insights from modelling of the 17 June 1996 Mount Ruapehu eruption

Building on the sensitivity analysis in Chapter Three, this chapter evaluates how volcano-specific knowledge of input parameters affects forecast accuracy, using the 17 June 1996 eruption of Mount Ruapehu as a case study. This directly addresses Objective Three (*To what extent does volcano-specific knowledge improve or limit forecast accuracy?*) by testing whether informed input parameter distributions improve short-term volcanic ash forecasts.

Chapter Four has been submitted and is currently under peer review at the Bulletin of Volcanology, as Scott, E., Whitehead, M., Bebbington, M., Mead, S. (2025). Impact of input parameter uncertainty on tephra deposition forecasts: insights from modelling of the 17 June 1996 Mount Ruapehu eruption.

Abstract

Volcanic ash is a destructive and widespread hazard associated with volcanic activity. To support hazard mitigation, forecasts of ash deposition, from long-term scenario planning to near-real-time response, are typically generated by using volcanic ash deposition models such as Tephra2. These models require eruption source parameters such as plume height, total erupted mass, and the total grain size distribution. A common way to inform these input parameters is to use prior knowledge of a volcano's magma composition(s) or past eruptions. This practice assumes that informed input parameter distributions improve forecast accuracy, but this assumption has not been systematically tested. In this study, we evaluate the role of informed versus uninformed input parameters in forecast accuracy using the 17 June 1996 eruption of Mount Ruapehu, Aotearoa New Zealand, as a case study. Forecasts were evaluated in two ways: as mass forecasts, comparing simulated to observed tephra mass, and impact-based forecasts, assessing whether forecasted ashfall exceeded defined impact thresholds. We show that forecasting accuracy was poor overall, with only 2.8% of the simulations achieving $\geq$50% impact-based accuracy. Moreover, forecast performance did not differ systematically between informed and uninformed input parameter distributions, highlighting that reliance on past eruption data can introduce a bias that reduces, rather than improves, forecast performance.

4.1 Introduction

Forecasting the spatial extent and intensity of volcanic hazards is a major goal in volcanology (Sparks 2003; Poland and Anderson 2020). Forecasts are hindered by aleatoric (inherent variability) and epistemic (knowledge-related) uncertainties, which arise from limited data, incomplete understanding of hazard (and eruption) dynamics, and volcano-specific variability (Cashman and Biggs 2014).

Volcanic hazard forecasts should, by necessity, be probabilistic, reflecting the intrinsic uncertainty and natural variability of eruptive processes (Marzocchi and Bebbington 2012; Poland and Anderson 2020). To align within a probabilistic framework, forecast accuracy refers to the degree to which observed data are consistent with the forecast distribution. This definition applies differently depending on whether the forecast is continuous (e.g., mass) or categorical

(e.g., impact thresholds) (Section 4.3).

Hazard forecasts rely on empirical and/or numerical models that simulate the transportation and deposition of volcanic products such as tephra (Manga et al. 2017; Pardini et al. 2022). While post-event modelling, or hindcasting, allows model input parameters to be reasonably well-constrained through direct observations (e.g., plume height, Pailot-Bonnétat et al. 2020; Barnie et al. 2023), or post-eruption analyses (e.g., total grain size distribution, Corradini et al. 2016; Freret-Lorgeril et al. 2021), estimating input parameter ranges before an eruption remains problematic but unavoidable.

Volcanic processes are inherently variable and uncertain, so perfect forecasts are neither possible nor the goal. Instead, probabilistic forecasting aims to capture the range of plausible outcomes, such that the eventual reality falls within the forecast distribution, ideally in the more probable portions. The challenge then is to refine input parameter ranges without introducing systematic bias, thereby enhancing the usefulness of model outputs for decision-making.

Probabilistic approaches represent input parameters as probability density functions (PDFs) (e.g., Sandri et al. 2016; Folch, Mingari, and Prata 2022; Trancoso et al. 2022; Jenkins et al. 2024), or through model ensembles (e.g., Dare, Smith, and Naughton 2016; Plu et al. 2021; Pardini et al. 2022), accounting for both epistemic and aleatory uncertainties (Rougier and Beven 2013). Model ensembles may improve forecasting (Bonadonna et al. 2013), by running multiple simulations with varying input parameters, helping to quantify uncertainty and identify the most probable outcomes (e.g., Harvey et al. 2020; Capponi et al. 2022; Folch, Mingari, and Prata 2022).

A key assumption in probabilistic forecasting is that using PDFs for input parameters improves forecast accuracy and credibility (Pyle 2018). In practice, however, the construction of these PDFs is itself uncertain, as they may take many forms (e.g., uniform, Gaussian, or empirically derived distributions), and the choice of form is often based on limited data on past eruptions (e.g., IVESPA database, Aubry et al. 2021) and expert judgement (e.g., Neri et al. 2008; Jenkins et al. 2024).

Here, we focus specifically on tephra deposition. Forecast accuracy in tephra deposition modelling has significant downstream implications, influencing critical decisions such as evacuations, road closures, and airspace restrictions (e.g., Wilson et al. 2012a; Blake 2016; Guffanti

and Tupper 2015). The consequences of forecast errors vary: a forecast of 3 mm when 2 mm occurs is unlikely to have serious impacts, whereas a forecast of 3 mm when 2000 mm falls could lead to severe disruptions and increased risk to life (e.g., Spence et al. 2005; Wilson, Jenkins, and Stewart 2015). It is not unusual for tephra deposition forecasts to deviate by up to a factor of 5 from observed ground deposits, even during hindcasts, where input parameters are constrained through observations and post-eruptive studies, e.g., HAZMAP (Campi Flegrei: Costa et al. 2008) and Fall3D (Vesuvius, 2011 Cordon Caulle, 2018 Kusatsu-Shirane: Folch et al. 2010; Collini et al. 2013; Tomii et al. 2020). Assessing the potential variability in forecast outcomes is therefore essential to ensure that hazard forecasts remain reliable and accurate for effective hazard mitigation.

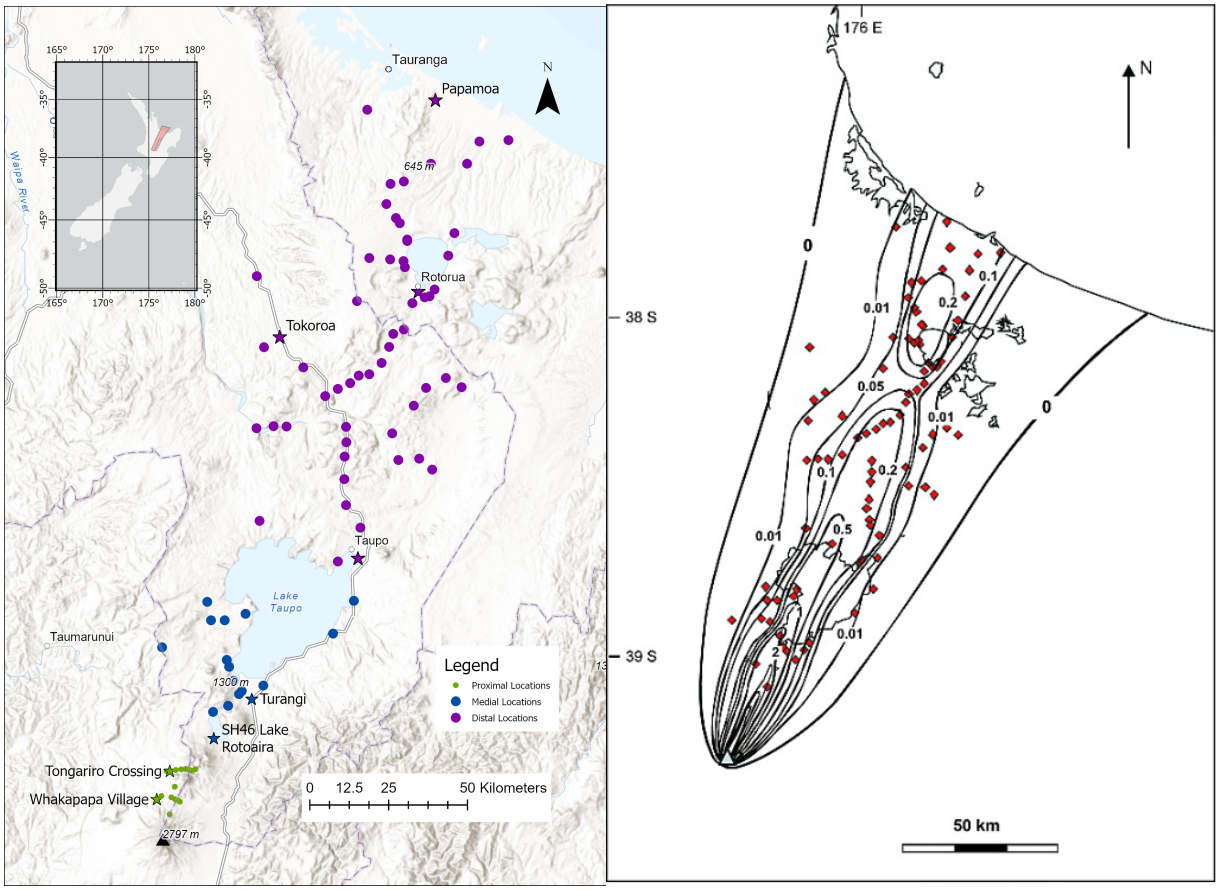
We investigate how input parameter value ranges in the volcanic ash transport and dispersal model (VATDM) Tephra2 (Connor, Connor, and Bonadonna 2008) affect tephra deposition forecasts for an imminent eruption. We compare the performance of uninformed input parameter distributions (assuming no prior knowledge) and informed input parameter distributions (based on factors such as magma type and previous eruption characteristics) against actual deposit data of the 17 June 1996 eruption at Mount Ruapehu, Aotearoa New Zealand (Bonadonna, Phillips, and Houghton 2005). This approach is possible because we have both meteorological and deposit data from the eruption, with the latter serving to assess our forecasts.

4.2 17 June 1996 Eruption of Mount Ruapehu

Mount Ruapehu is a frequently active stratovolcano located in the southern Taupō Volcanic Zone in the North Island of Aotearoa New Zealand, with 40 eruptions in the last 80 years (Kilgour et al. 2013) (Figure 4.1). Since forming in 200 ka, Mount Ruapehu has erupted with a wide range of styles and sizes (Houghton, Latter, and Hackett 1987; Leonard et al. 2021). Volcanic activity in recent millennia has come from a single vent, Te Wai ā-Moe/Crater Lake (Hackett and Houghton 1989; Hales 2000), primarily erupting small magma volumes ($\leq$VEI 3) (Kilgour et al. 2010).

Several notable ash-producing eruptions occurred around 1995/96, including events on 11 and 14 October 1995 and 17 June 1996 (Leonard et al. 2021). The 11 October eruption produced an estimated tephra volume of 0.025 km^3, while the 14 October eruption produced 0.005 km^3

(Cronin et al. 1997). The 17 June 1996 event, with an estimated tephra volume of 0.005 km^3, was the last in this sequence to produce widespread ash deposition. A later eruption in 2007 produced ashfall over a limited area of 2.5 km^2, confined to proximal regions and not dispersed downwind (Kilgour et al. 2010).



(a) (b)

Figure 4.1: (a) Sampling locations from the 17 June 1996 Mount Ruapehu eruption (Bonadonna, Phillips, and Houghton 2005). Star symbology represents key sampling locations (i.e., population centres and important road networks) used in assessing output mass accuracy. (b) Isomass contours (kg/m^2) of the 17 June 1996 Mount Ruapehu eruption from Bonadonna, Phillips, and Houghton (2005).

On the morning of June 17 1996, Mount Ruapehu erupted, producing a lahar down the Whangaehu Valley and causing significant ashfall hundreds of kilometres downwind of the volcano (Scott and Neall 2016) (Figure 4.1). This eruption produced a well-known "bent-over" plume (Bonadonna, Phillips, and Houghton 2005), which is an eruption plume that is wind-dominated rather than buoyancy-dominated (Carey and Bursik 2000; Dürig et al. 2023). The

eruption led to the closure of several airports, multiple injuries (primarily from falls while clearing ash from roofs), and livestock deaths due to fluorine-contaminated pasture (Johnston et al. 2000; Scott and Neall 2016). The economic cost of the eruption was over $NZ 130 million (in 1995/1996 value, $US 104 million) (Wilson et al. 2012b).

4.2.1 Previous Modelling with Volcanic Ash Transport and Dispersion Models

Several studies have focused on the dynamics of the June 1996 eruption at Mount Ruapehu, covering impacts (Cronin et al. 1998; Johnston et al. 2000; Cronin et al. 2003), plume properties (Prata and Grant 2001), volume and total grain size distributions (Bonadonna and Houghton 2005), and sedimentation dynamics (Bonadonna, Phillips, and Houghton 2005; Klawonn et al. 2012; Klawonn et al. 2014).

During the eruption, the VATDM ASHFALL (Hurst and Hurst 1994) was used to generate ashfall forecasts (referred to as predictions in their paper) to support civil defence decision-making. Hurst and Turner (1999) compared these predictions to observed ashfall distributions and concluded that the model provided an "*adequate indication of the ashfall distribution for emergency response purposes*" (Hurst and Turner 1999, p. 620), with forecasted ash thickness in the most affected areas falling within a factor of two (Hurst and Scott 1998). ASHFALL used standard eruption source parameters (ESPs) such as plume height, total erupted mass (TEM), and grain size distribution. However, the estimation of these input parameters, both in 1996 and today, relied on analogue eruptions or observations made after the eruption was underway. Although ASHFALL produced forecasts considered adequate at the time, it is now defunct and has been replaced by the VATDM HYSPLIT for operational forecasting in Aotearoa New Zealand (Hurst and Davis 2017; Trancoso et al. 2022).

Turner and Hurst (2001) used the June 1996 eruption to calibrate and test the HYPACT (Walko, Tremback, and Bell 1995) and ASHFALL VATDMs, finding that the HYPACT and ASHFALL models produce "reasonable", but not exact, agreement with the observed ash deposition. They also found, through local sensitivity testing, that TEM was the primary control on ash deposition for this eruption. Scollo et al. (2008) conducted a global sensitivity analysis using Sobol' indices (Sobol' 1990) on the June 1996 eruption to identify the most influential

input parameters in the VATDM TEPHRA (Bonadonna et al. 2005). They found that TEM, plume height, fall time threshold (FTT, when the diffusion model changes from linear to power law), and diffusion coefficient were key parameters, while particle density and plume shape (i.e., the vertical distribution of ash within a plume) had little to no effect on model outputs.

Table 4.1: Values of ESPs found within the literature from the 17 June 1996 eruption at Mount Ruapehu, Aotearoa New Zealand. ESP Values in bold with an asterisk indicate our assumed most accurate value.

Eruption Source Parameter	Value	How it was measured/calculated	Reference
Total Erupted Mass	4.16×10^9 kg (a)	(a) Best fit based on comparison between field data and model outputs from the TEPHRA model (b) presumed value by Bonadonna and Houghton (2005) multiplied by assumed density of 1150	(a) Scollo et al. (2008) (b) Bonadonna and Houghton (2005), cited by Scollo et al. (2008)
	4.6×10^9 kg (b)		
	5×10^9 kg	Value used in plume modelling	Bonadonna, Phillips, and Houghton (2005)
	4.51×10^9* kg	**Average volume estimates multiplied by deposit density (1100, Cronin et al. 1998)**	**Aubry et al. (2021)***
Particle Density	**Uncompacted bulk density 1100 kg/m^3***	**Field measurements**	**Cronin et al. (1998)***
	Pumice Density: 1100 kg/m^3, Lithic density: 2650 kg/m^3	Value used in plume modelling	Bonadonna, Phillips, and Houghton (2005)
Total Grain Size Distribution	Uniform distribution: 10 - 150μm	Assumed for modelling	Turner and Hurst (2001)
	Mean (overall): 0.7 - 3.70 phi, Standard deviation (overall): 0.59 - 2.47 phi	Assumed Gaussian distribution on field measurements	Cronin et al. (2003)
	Median: 0.8 phi, Standard deviation: 2.4 phi	Voronoi Tessellation	Bonadonna and Houghton (2005)

(Continued)

Eruption Source Parameter	Value	How it was measured/calculated	Reference
	Coarse grain mean: -3.02 phi, Fine grain mean: 0.04 phi, Coarse grain standard deviation: 1.06 phi, Fine grain standard deviation: 1.94 phi	Bi-Gaussian distribution fitting from field data	Costa, Pioli, and Bonadonna (2016)
	Median grain size: -0.6495 phi*	**Unknown**	**Aubry et al. (2021)***
	Standard Deviation: 2.46 phi*	**Unknown**	**Aubry et al. (2021)***
Plume Height	"several kilometres"	Visual observations	GeoNet (1996a) and GeoNet (1996b)
	6.7 - 7.4 km, up to 9 km	Infra-red satellite images	Hurst and Turner (1999)
	7 km	Best-fit match against ashfall reports and satellite images, from modelling	Turner and Hurst (2001)
	7.5 - 8.5 km a.s.l	Satellite images	Prata and Grant (2001)
	7 - 10 km	Unknown	Cronin et al. (2003)
	7.5 km a.s.l*	**Unknown**	**Aubry et al. (2021)***
	7 km	Visual observation	Deligne (2021)
Diffusion Coefficient	1 $m^2 s^{-1}$	Sensitivity test TEPHRA	Bonadonna et al. (2005)
	18158 m^2 s^{-1}*	**Average value found in literature**	
Fall time Threshold	180 seconds	Sensitivity test TEPHRA	Bonadonna et al. (2005)
	7840.03* seconds	**Dependent on diffusion coefficient**	**Scott et al. (2025)**

(Continued)

Eruption Source Parameter	Value	How it was measured/calculated	Reference
Plume Shape	**Beta Distribution** **Alpha: 4*** **Beta: 1***	**Unknown**	**Trancoso et al. (2022)***

A wide variety of ESPs have been suggested for the 17 June 1996 eruption (Table 4.1). Although this eruption was directly observed and occurred less than 30 years ago, a considerable range remains in the reported ESP values.

It is clear that multiple combinations of ESPs can be used to best fit a simulation to a tephra deposit after an eruption (hindcasting). This discrepancy in ESPs has been documented in past studies; for instance, Pfeiffer, Costa, and Macedonio (2005) found that best-fit ESPs varied depending on whether proximal or distal ashfall deposits were considered. Such discrepancies suggest that even well-constrained hindcasts may not uniquely identify the "true" ESPs of an eruption. One possible explanation is that the ESPs do not fully describe the complexity of a given eruption, or that some input parameters may not remain constant through an event, e.g., eruptions are unlikely to have constant eruption rates (Sparks 1986). Consequently, before an eruption, when ESPs are highly uncertain, it remains unclear how different input parameter ranges influence model outputs.

To guide our analysis, we selected a single representative value for each key ESP (marked bold with an asterisk in Table 4.1). These values were chosen where possible from direct measurements or well-documented estimates, rather than from plume modelling or best-fit hindcasts. They provide a consistent reference point for comparison across our results and discussion.

4.3 Methodology

For this study, we use the VATDM Tephra2, specifically focusing on tephra deposition (i.e., ashfall) in the context of an imminent eruption, assuming meteorological conditions are known, essentially envisioning ourselves in early June 1996 and forecasting ashfall for 17 June 1996.

We conduct batch simulations for Tephra2 to investigate the influence of ESP ranges on producing accurate forecasts. Given that it is infeasible to simulate every possible combination of input parameters (because the input parameters are continuous) (Oberkampf, Trucano, and Hirsch 2004), we use Latin Hypercube Sampling (Stein 1987) to generate 1000 simulations per input parameter distribution (total 7000).

We use the European Centre for Medium-Range Weather Forecasts (ECMWF) hourly ERA5 reanalysis dataset, which spans from 1979 onwards with a vertical resolution of 37 pressure levels (Hersbach et al. 2020). We extracted the ERA5 reanalysis data (wind speed, wind direction,

vertical height) for the morning of 17 June 1996. Simulations were run on a spatial grid of 1 km x 1 km, as increasing spatial resolution does not significantly impact simulation duration. A limitation of using ERA5 data is that it is derived from a global model at relatively coarse spatial resolution and may not capture local-scale wind variability within the computational domain.

4.3.1 Uninformed Sampling Distribution

Before an eruption, ESPs are fundamentally unknown (Engwell et al. 2024). This uncertainty can be framed between two end-member cases: (1) complete knowledge of ESPs, and (2) complete ignorance, where all values across an extremely broad range are considered equally likely (i.e., uninformed sampling distributions extending from zero to infinity). In practice, neither of these extremes is realistic. Complete knowledge is infeasible due to the inherent aleatoric variability of volcanic processes, while total ignorance is implausible because we do have broad constraints on plausible ESP ranges from past eruptions (e.g., plume height typically maxes out at around 50 km, Wilson et al. 1978; Aubry et al. 2021). Forecasting approaches must therefore balance between these extremes by incorporating available knowledge in a way that reduces uncertainty without introducing unjustified bias. Without additional constraints, a wide range of ESPs must be explored - not simply to cover the sample space, but to capture the resulting variability in forecast outputs. Here, we construct an uninformed sampling distribution based on broad theoretical minimum and maximum ranges. We adopt input parameter bounds from Scott et al. (2025), including the dependent relationships between Tephra2 input parameters such as plume height (i.e., top of the plume) and TEM, and between diffusion coefficient and FTT. These ranges are applied consistently, except where meteorological conditions and grid resolution differ. For grain size, we use a minimum and maximum value of 9 and -5 phi.

4.3.2 Informed Sampling Distribution

For most volcanoes, some knowledge and/or data are available on past eruptions and magma composition that may help constrain future eruption characteristics. This information can be used to build informed input parameter distributions by incorporating factors such as volcano style, eruptive history, and expert judgement.

Here, we inform our sampling distributions using past data from two ESP databases: the IVESPA and the Deligne (2021) databases. The IVESPA database by Aubry et al. (2022) contains ESP data from 134 eruptive events from 1902 to 2016, associated with 45 different volcanoes. The database by Deligne (2021) contains ESP data from 213 eruptive events from prehistoric time to 2019, associated with 37 different volcanoes, with a focus on Aotearoa New Zealand volcanoes.

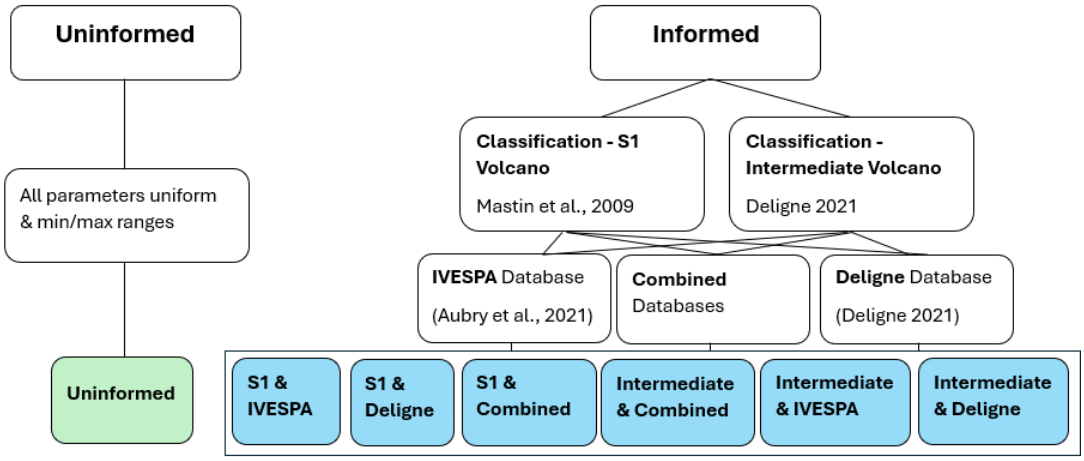


Figure 4.2: Schematic of the uninformed and informed input parameter distribution sampling.

We filter these two datasets in two ways (Figure 4.2). The first is per Mastin et al. (2009), who categorised all Holocene volcanoes from the Smithsonian database according to the dominant magma type and historical eruptive episodes. Mount Ruapehu falls under the S1 classification, which corresponds to a volcano with silicic magma, including andesitic volcanoes, and has produced historic plume heights of less than 6 km or VEI 2. The second filter is per Deligne (2021), who categorised volcanoes by their major magma composition. Mount Ruapehu is classed as an intermediate composition, and as such, we filter out volcanoes with mafic and silicic compositions. We use the subset S1 and intermediate composition data, further removing eruptions from Mount Ruapehu post-1996 since we aim to treat the 17 June 1996 event as a true forecast (i.e., forecasting before an eruption). However, we could not exclude all eruptions from S1 and intermediate magma volcanoes after 1996 entirely due to the limited data available (especially for the Deligne (2021) database, where only 0.5 to 0.8% of categorised eruptions were post-1996) (Table 4.2).

Table 4.2: Number of unique volcanoes and eruptions used in Tephra2 input parameter sampling, grouped by database and magma classification.

Database	**S1 Classification**		**Intermediate Classification**	
	Volcanoes	**Eruptions**	**Volcanoes**	**Eruptions**
IVESPA	8	15	31	87
Deligne (2021)	6	25	24	86
Combined	11	37	45	153

Additionally, we wanted to investigate S1 and intermediate magma compositions with a database that combines the IVESPA and Deligne (2021) databases. If both the IVESPA and Deligne (2021) databases contained the same eruption, we kept the data from the Deligne (2021) database.

We specifically inform Tephra2 inputs TEM and plume height (i.e., top of the plume) (Figure 4.3). The Deligne (2021) database contains the mass eruption rate (MER) and eruption duration. To obtain TEM, assuming a constant MER, we multiplied MER by the eruption duration. To characterise the distributions of plume height and TEM, we fit four distributions - Weibull, Gamma, Log-normal, and Gaussian - to each ESP dataset and selected the best-fit distribution based on log-likelihood values, which reflect goodness-of-fit (Table 4.3; TEM was log-transformed prior to fitting to allow Gaussian sampling). All other Tephra2 input parameters, including diffusion and total grain size distribution (TGSD), were retained as uninformed, uniform distributions because they are not represented in the IVESPA or Deligne (2021) databases.

Table 4.3: Informed input parameter sampling for Tephra2 using the IVESPA (Aubry et al. 2021), Deligne (2021) (Deligne 2021), and combined databases. Input parameters include plume height and total erupted mass (TEM), grouped by classification: S1 and Intermediate.

Input Parameter	**Database**	**Distribution Type**	**S1 Classification**	**Intermediate Classification**
Plume Height (km)	IVESPA	Log-normal	MeanLog: 2.37, SDLog: 0.243	MeanLog: 2.47, SDLog: 0.433
	Deligne (2021)	Weibull	Shape: 3.314, Scale: 21.479	Shape: 1.88, Scale: 18.125
	Combined	Weibull	Shape: 2.388, Scale: 18.397	Shape: 1.957, Scale: 16.379
Total Erupted Mass (TEM, log-transformed kg)	IVESPA	Gaussian (S1), Gamma (Intermediate)	Mean: 9.628, SD: 0.966	Shape: 72.534, Rate: 7.198
	Deligne (2021)	Weibull	Shape: 13.743, Scale: 12.33	Shape: 11.613, Scale: 12.112
	Combined	Weibull (S1), Gamma (Intermediate)	Shape: 8.78, Scale: 11.798	Shape: 52.449, Rate: 4.909

4.3.3 Mass Accuracy

To evaluate how closely our simulations match the observed deposit, model output values (kg/m^2) are compared to real deposit data from the 17 June 1996 eruption, with 91 field measurements (Bonadonna, Phillips, and Houghton 2005) (Figure 4.1). For each field measurement, the percentile of the observed deposit within the simulated distribution was calculated. Percentiles were grouped into the following bins: <2.5%, 2.5 - 4.9%, 5 - 39.9%, 40 - 59.9%, 60 - 94.9%, 95 - 97.4%, and >97.5%. For example, if the simulated deposit masses were normally distributed between 2 and 10 kg/m^2, and the observed value at a given location was 9.9 kg/m^2, this measurement would fall near the 95 - 97.4% bin.

This approach allows systematic quantification of forecast bias and accuracy. Observations near the median percentile (40 - 59.9%) indicate a good agreement with simulations, while values

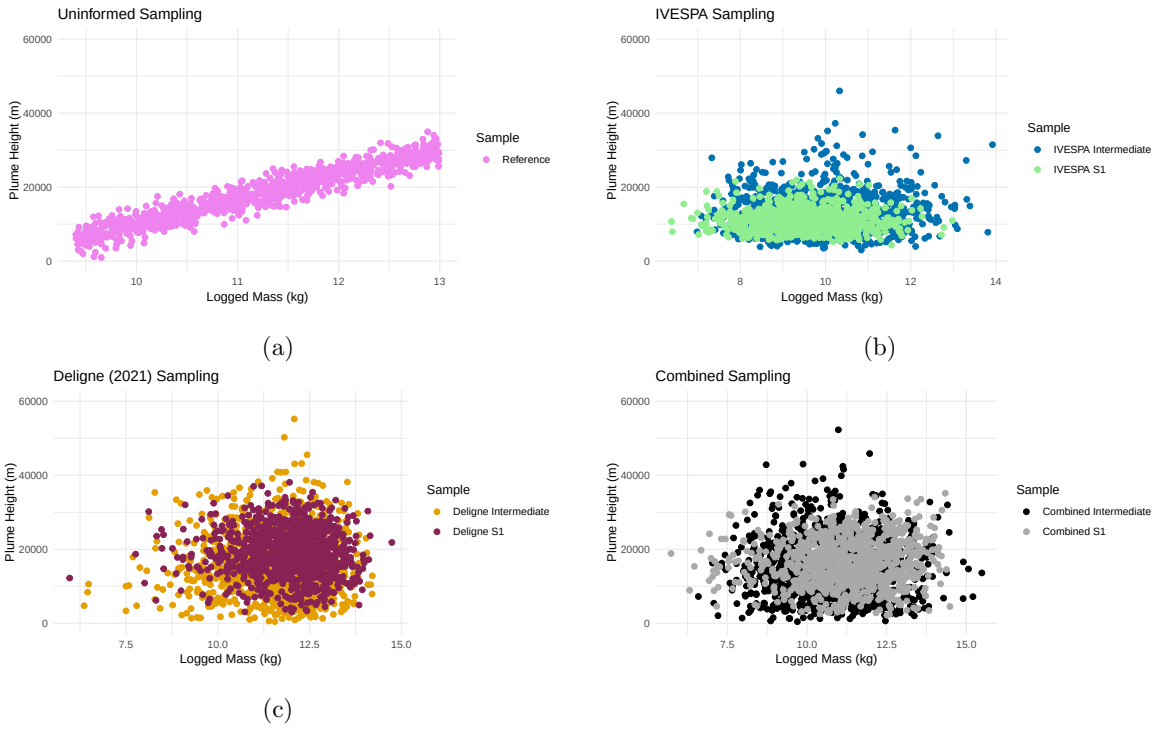


Figure 4.3: Input parameter sampling of total erupted mass (logged) and plume height derived from the reference sampling distribution and from the IVESPA, Deligne (2021), and combined databases.

in the lower and upper percentile bins reveal persistent under- or over-prediction, respectively. Extreme percentiles (<2.5%, >97.5%) highlight simulation sets that fail to capture the observed deposit. By categorising observations in this way, we can assess the spatial patterns in model performance and identify systematic biases across different sampling distributions.

To evaluate overall calibration, we first calculate, for each observation, its percentile within the corresponding simulation distribution (i.e., the proportion of simulations that forecast a value less than or equal to the observed value). The collection of these percentiles across all observations forms the distribution of simulated percentiles. An ideal, unbiased forecast would produce percentiles that are uniformly distributed between 0 and 100%. Departures from uniformity indicate systematic bias. We formally assess this using the Kolmogorov-Smirnov (K-S) test: a statistically significant result (low p-value) suggests that the forecast percentiles deviating from uniformity are therefore not well-calibrated.

4.3.4 Impact-Based Thresholds

Rather than focusing solely on the accuracy of mass outputs, we also focus on the utility of forecasts for practical, real-time emergency management decisions that require impact-based metrics. We do this via key tephra deposition impacts on the built and natural environment (Table 4.4).

Table 4.4: Key tephra mass threshold levels and their impacts on the built and natural environment. Values from Houghton et al. (2006), Jenkins et al. (2014), Wilson et al. (2014), Jenkins et al. (2015b), and Blake et al. (2017).

Threshold Level	**Impact Threshold (kg/m^2)**	**Impact**
0	<0.0001	No impact.
1	0.1	Closure of airports.
2	1	Markings covered at airports and on roads. Increased skid risk.
3	10	Damage to power, water, and communication networks. Severe impact on roads and airport closures. Impact on crops, depending on growth stage.
4	100	Collapse of weakest roofs. Structural damage to power lines.
5	150	Collapse of long-span roofs.
6	300	Collapse of high-quality metal roofs.
7	400	Collapse of high-quality tile roofs.
8	700	Collapse of high-quality concrete roofs.
9	1800	Near-total vegetation kill.
10	>2500	Total vegetation kill.

For every simulation and every sampling location (Figure 4.1), observed and simulated tephra mass are assigned to threshold categories (0 to 10, Table 4.4), with a 10% buffer applied so that mass values within 10% are assigned to the closest category. For example, a mass of 149 kg/m^2 would normally fall into threshold level 4, but is assigned threshold level 5 because it falls within 10% of that threshold's value.

Forecast accuracy is then evaluated using a "Hit or Miss" system: a hit occurs when the simulated and observed values fall into the same threshold category; otherwise, it is classified as a miss. Unlike mass forecasts, which support percentile-based diagnostics, impact-based

forecasts are assessed through classification agreement and spatial patterns of hit rates across the sampling domain. While hit rates are higher for lower impact thresholds, which occur more frequently, we use a strict categorical match system to avoid artificially inflating accuracy through over-prediction. A cumulative approach, where any exceedance counts as a hit, would obscure accuracy, especially given the large impact mass differences between higher thresholds, where exceeding a value by even a small margin would imply a vastly different impact. Although over-prediction may be preferable to under-prediction, our method prioritises precision in impact classification to preserve the integrity of forecast evaluation.

4.4 Results and Discussion

We first look at how accurate our simulated mass is compared to the observed mass of the 17 June 1996 Mount Ruapehu deposit (Figure 4.1) and then examine how using different sampling distributions - IVESPA S1, IVESPA Intermediate, Deligne (2021) S1, Deligne (2021) Intermediate, Combined S1, and Combined Intermediate - influence the resulting forecasts (Figure 4.2).

4.4.1 Mass

Overall, forecast performance was poor across all sampling distributions. Simulated deposit masses were often biased high relative to observations with probability distributions extending several orders of magnitude above the true values. In the most extreme case, an observed mass of 0.5 kg/m^2 was simulated as more than 275,000 kg/m^2.

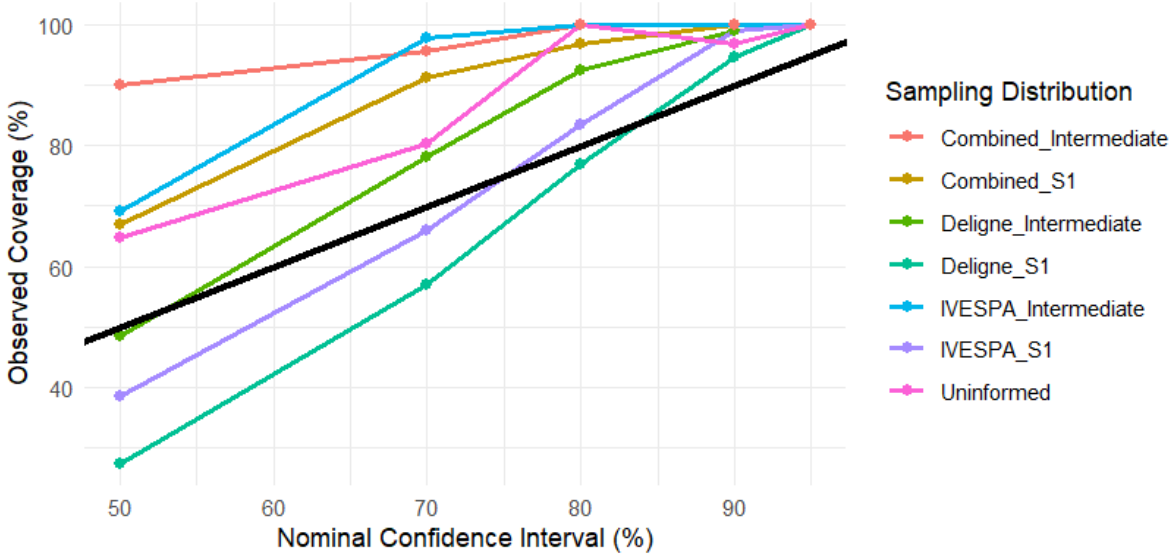


Figure 4.4: Forecast calibration across different input parameter sampling distributions. The black line indicates the 1:1 line.

Observed coverage across sampling distributions generally exceeded nominal confidence intervals. This is shown by the data points being above the 1:1 reference line of perfect calibration, which indicates that most forecasts were conservative (Figure 4.4). For instance, both the Combined Intermediate and IVESPA Intermediate sampling distributions captured the majority of observed deposit mass across sampling locations, even at the 50% confidence interval. These forecasts, therefore, aligned with observations at that level, though they tended to overestimate the range of plausible outcomes. In contrast, the Deligne (2021) S1 and IVESPA S1 sampling distributions captured fewer observed deposit masses at the 50% and 70% intervals, making their forecasts more often inconsistent with observations and indicative of underestimated uncertainty. At the broadest (95%) confidence interval, however, all sampling distributions achieved the full range (0-100%).

Percentiles of observed tephra mass relative to simulated distributions reveal systematic biases across the sampling distributions (Figure 4.5). Under a well-calibrated probabilistic forecast, observed percentiles should be uniformly distributed across the 0 - 100% range, reflecting appropriate uncertainty and no systematic bias. However, the distributions analysed here deviate significantly from uniformity, as confirmed by Kolmogorov-Smirnov (K-S) tests ($p < 0.001$ for all sampling distributions). The uninformed sampling distribution showed a strong tendency toward under-prediction, with percentiles concentrated between 5 - 59% and a median of 31%, especially at proximal and medial sites (18% and 17%, respectively). The IVESPA S1 sampling distribution produced the highest percentiles overall (median 79%), but still failed the uniformity test ($p = 4.96\text{x}10^{-12}$), indicating over-prediction and reduced spread. The Deligne (2021) S1 sampling distribution performed worst, with percentiles tightly clustered below 40% and several sites falling below the 5th percentile - far from the expected uniform spread. Even when all simulations were combined, the distribution remains skewed, with most percentiles between 5 and 59% and a median of 42%. These results demonstrate that while input parameter choices shift the direction of bias, none of the sampling strategies produced a robust forecast, i.e., the spread of percentiles consistently failed to span the full 0 - 100% range.

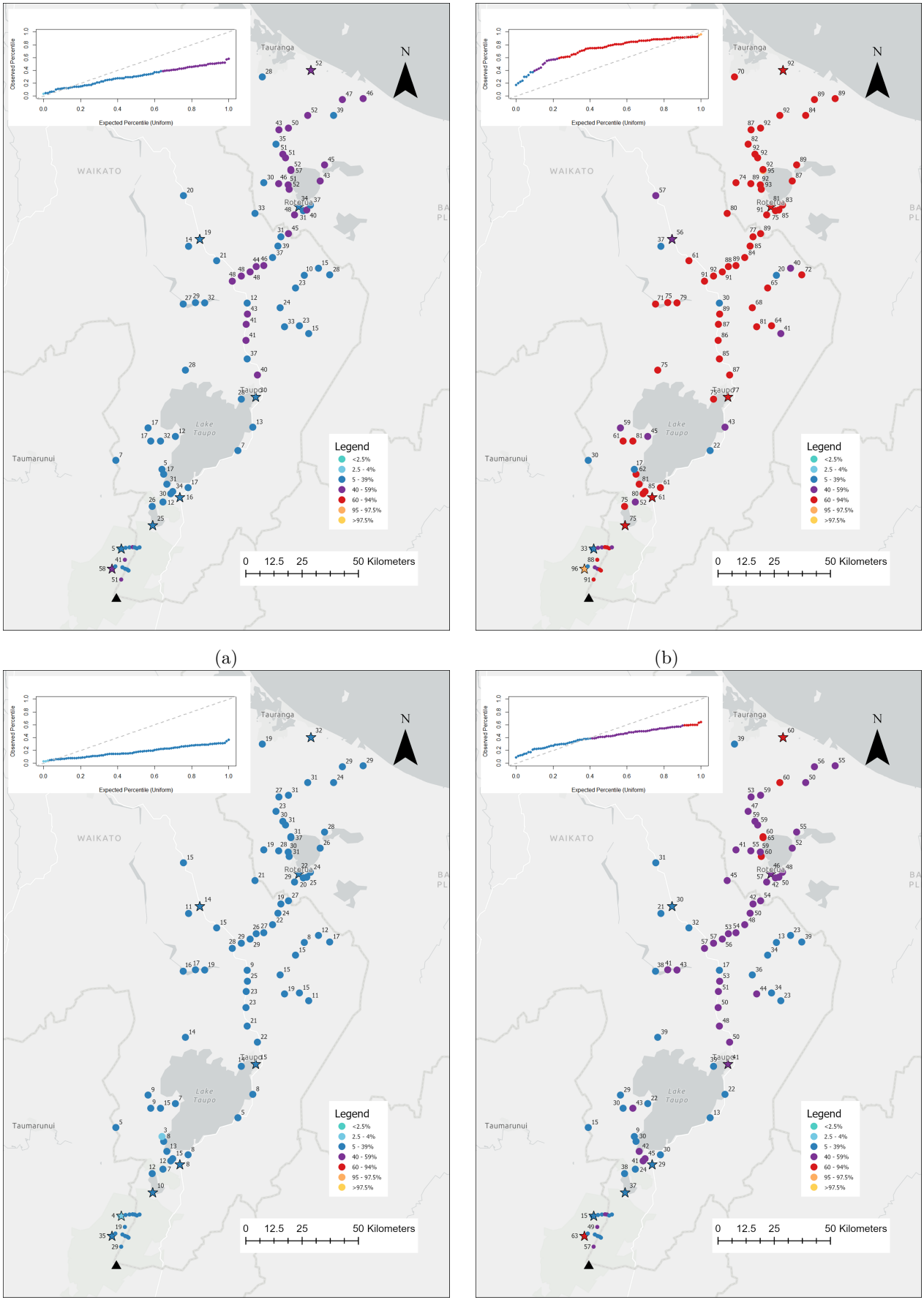


Figure 4.5: Percentiles of observed tephra mass relative to simulated distributions: (a) uninformed, (b) IVESPA S1, (c) Deligne (2021) S1, and (d) all simulations from all distributions.

4.4.2 Impact Thresholds

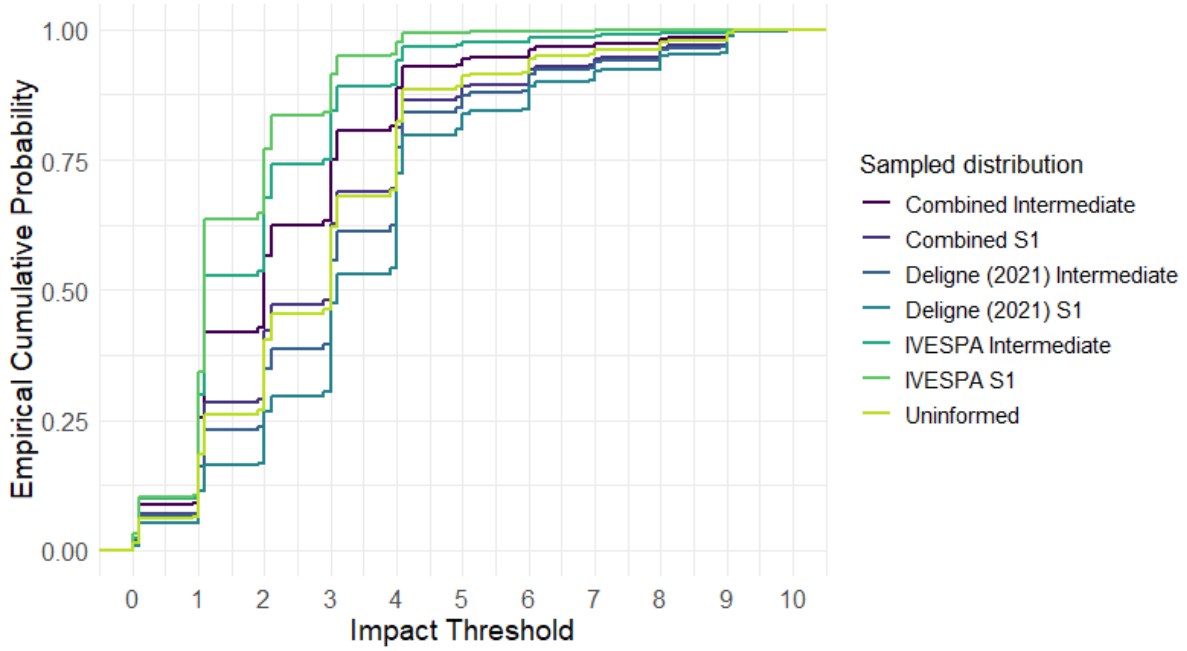


Figure 4.6: Empirical cumulative probability function plots of simulated impact thresholds across all sample locations.

Compared to mass forecasts, the range of impact thresholds across all sampling distributions spanned the full 0 to 10 range. This indicates a more robust representation of uncertainty in impact-based forecasts, though this outcome partly reflects the way the thresholds themselves were defined. Empirical cumulative distribution functions (eCDF) reveal distinct patterns across sampling distributions (Figure 4.6). All sampled distributions show a steep rise between thresholds 2 and 5, indicating that the majority of threshold values cluster within this band. For all sampling distributions, the curve is steeper than the uniform reference line, reflecting a bias towards lower thresholds and a lack of high-threshold values. These patterns highlight how exceedance probabilities can be strongly influenced by the choice of input parameter sampling distribution, which has implications for defining and applying damage state thresholds (e.g., Wilson et al. 2014; Craig et al. 2016).

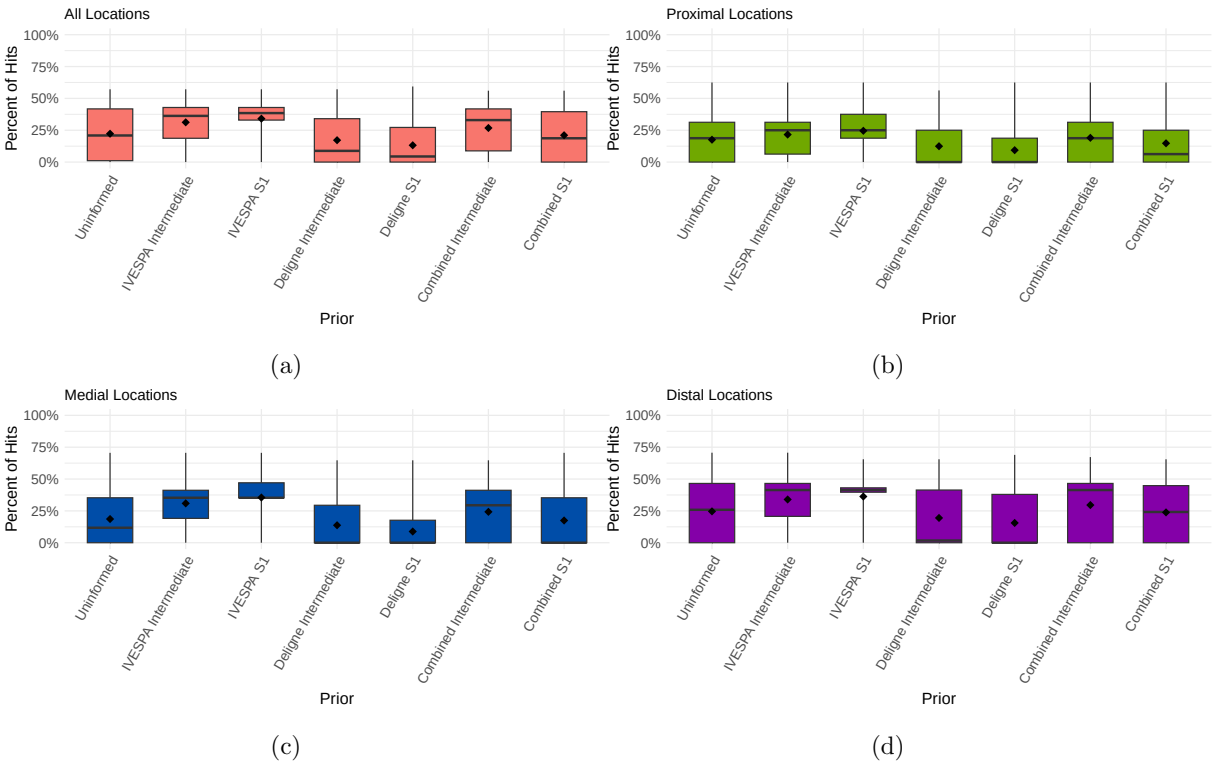
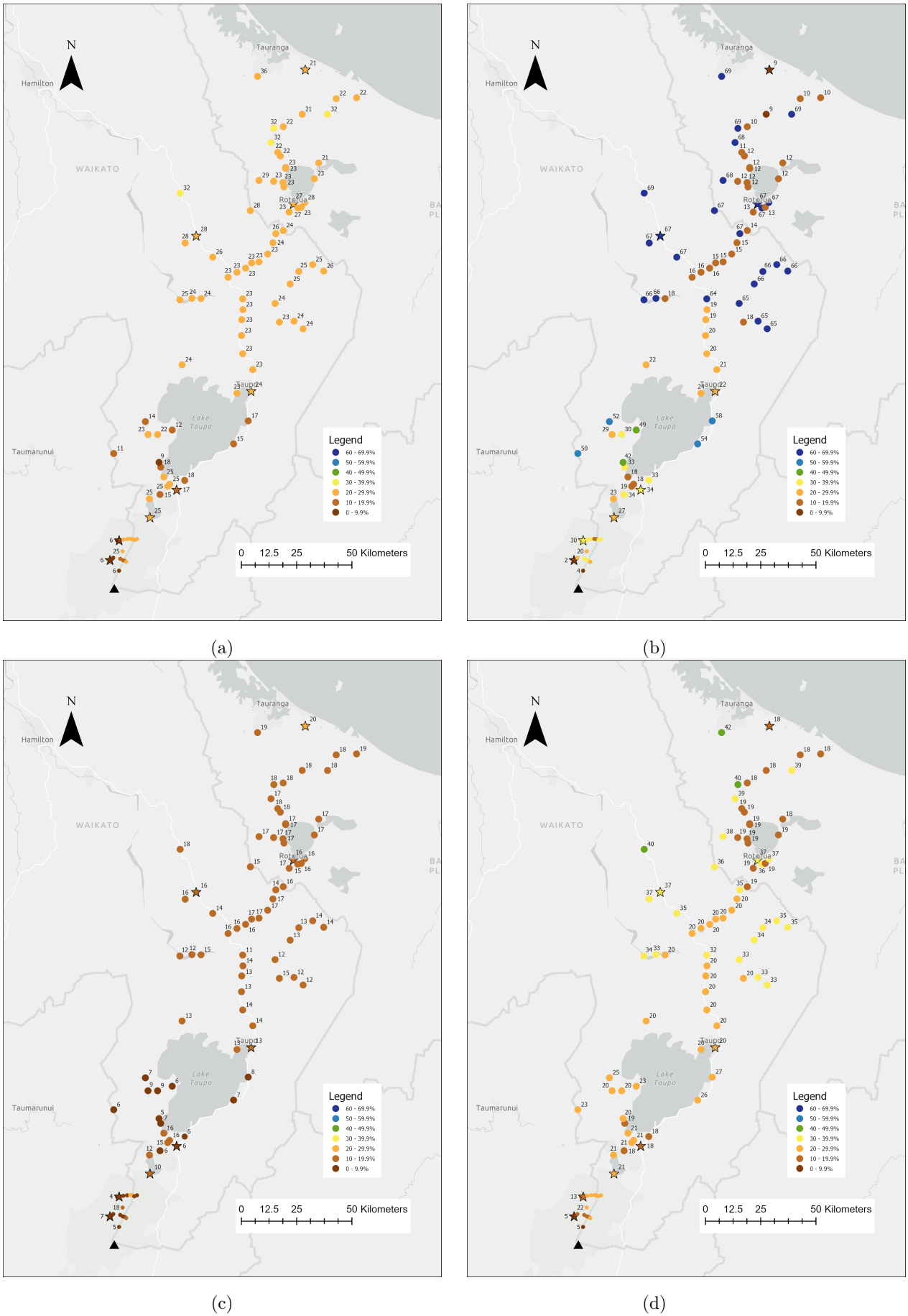


Figure 4.7: Boxplots showing the percentage of hit locations for each distribution, over (a) the whole sampling space (out of 91); (b) proximal locations (out of 16); (c) medial locations (out of 17); and (d) distal locations (out of 58). The horizontal line indicates the median, and the dot represents the mean.

Figure 4.7 shows that, across individual boxplots for each sampling distribution, the IVESPA Intermediate and IVESPA S1 sampling distributions consistently produce more hit locations across all spatial sampling areas than any other distribution. In contrast, the Deligne (2021) Intermediate and Deligne (2021) S1 sampling distributions produce the fewest hits across all spatial sampling areas. Notably, the inter-quartile range for the IVESPA S1 sampling distribution narrows with increasing distance from the vent, resulting in greater consistency in hit performance at distal locations. However, no single simulation achieves a 100% hit rate in any location, reinforcing that reliable forecasts require ensembles of many runs rather than reliance on any single run.

When examining the best-performing locations (Figure 4.8), both the combined simulation results - aggregating outputs from all simulations from all sampling distributions - and the results from the Deligne (2021) S1 sampling distribution show the least successful forecast locations are situated south of Lake Taupō, within our proximal and medial zones (i.e., within 80 km of the vent). For the IVESPA S1 sampling distribution, a few locations east and west



(a) (b)

(c) (d)

Figure 4.8: Median hit rate for all sampling locations, looking at (a) uninformed, (b) IVESPA S1, (c) Deligne S1, and (d) all simulations.

of Lake Taupō show higher hit rates. Notably, the most hit locations tend to be on the fringes of the sampling domain (near the edges of the tephra deposit). For the IVESPA S1 sampling distribution, distal locations exhibit the highest hit rates overall, although there is considerable spatial variability, with areas of high and low hit rates interspersed. This observed pattern in hit rates is partly explained by how forecast error interacts with the structure of the impact thresholds. Although our impact thresholds are categorical, the underlying tephra mass errors scale with the size of the deposit, so a small percentage error in proximal zones can translate into large differences in mass. Importantly, higher threshold categories span wider mass ranges, making them less sensitive to error. In contrast, lower thresholds have narrower ranges and are therefore more likely to be misclassified by even small deviations.

Table 4.5: Percentage of simulations that met or exceeded a 50% hit rate. The column "All simulations" includes the full set of 7000 simulations, while each sampling distribution column is based on 1000 simulations.

Location	All Simulations	Uninformed	IVESPA Intermediate	IVESPA S1	Deligne (2021) Intermediate	Deligne (2021) S1	Combined Intermediate	Combined S1
All	2.8	2.9	3.4	3.8	1.4	1.4	3.3	3.3
Proximal	0.6	0.4	0.6	1.1	0.4	0.4	0.9	0.4
Medial	8.3	6.5	12.3	15.6	4.9	2.8	9.3	6.5
Distal	17.6	20.2	19.8	17.3	14.8	11.2	20.3	19.5

We further examine our impact-based forecasts by looking at the simulations that most clearly matched the observed deposit, i.e., simulations that met or exceeded a 50% hit rate, so were more accurate than not across all sampling locations. While a 50% or more threshold is not intended as a definitive measure of forecast accuracy, it provides a pragmatic filter for identifying input parameter combinations that yield reasonable spatial agreement with observed impacts. Looking at our most accurate simulations, forecasts performed poorly when evaluated against impact thresholds, particularly in proximal areas (Table 4.5). These findings align with previous studies using Tephra2 that experience reduced accuracy in model outputs in proximal areas to the vent (Connor, Connor, and Saballos 2011; Magill et al. 2015), especially for a bent-over plume (Bonadonna et al. 2005).

When comparing sample distribution types, clear differences in forecast performance emerged. Simulations using the uninformed sampling distribution consistently overpredicted mass, which is unsurprising given that the uninformed range spans large eruption sizes (9.87×10^{12} kg),

whereas the 1996 Mount Ruapehu event was a relatively small eruption ($4.51 \text{x} 10^9$ kg). The IVESPA sampling distributions, by contrast, tended to underpredict mass but yielded the most accurate impact-based forecasts, particularly across medial and distal locations. The Deligne (2021) sampling distributions also led to overpredictions of mass but produced the least accurate impact-based forecasts, indicating that simply overestimating tephra mass does not result in improved impact-based forecasts. The Combined sampling distributions similarly overpredicted mass, but the impact-based forecast performance was comparable to the uninformed distribution.

The effect of including prior information on forecast accuracy was mixed. Overall, there was no consistent improvement from informed sampling distributions over the uninformed baseline. This reflects the bias-variance trade-off: introducing prior information narrows the input parameter space and reduces variance, but also introduces bias. Our results suggest that prior information has the potential to enhance forecasts when it is appropriately constrained (shown by the results of the IVESPA sampled distributions for Mount Ruapehu), but also highlight the challenge that even well-characterised past data may not fully capture the conditions of a specific eruption.

4.4.3 Useful Probability Space

Something to consider for forecasting is the useful probability space of an input parameter, i.e., the range of values of an input parameter that produce accurate outputs. We explore the input parameter ranges for simulations where there was a hit rate of 50% or more (i.e., forecasts that are more accurate than not).

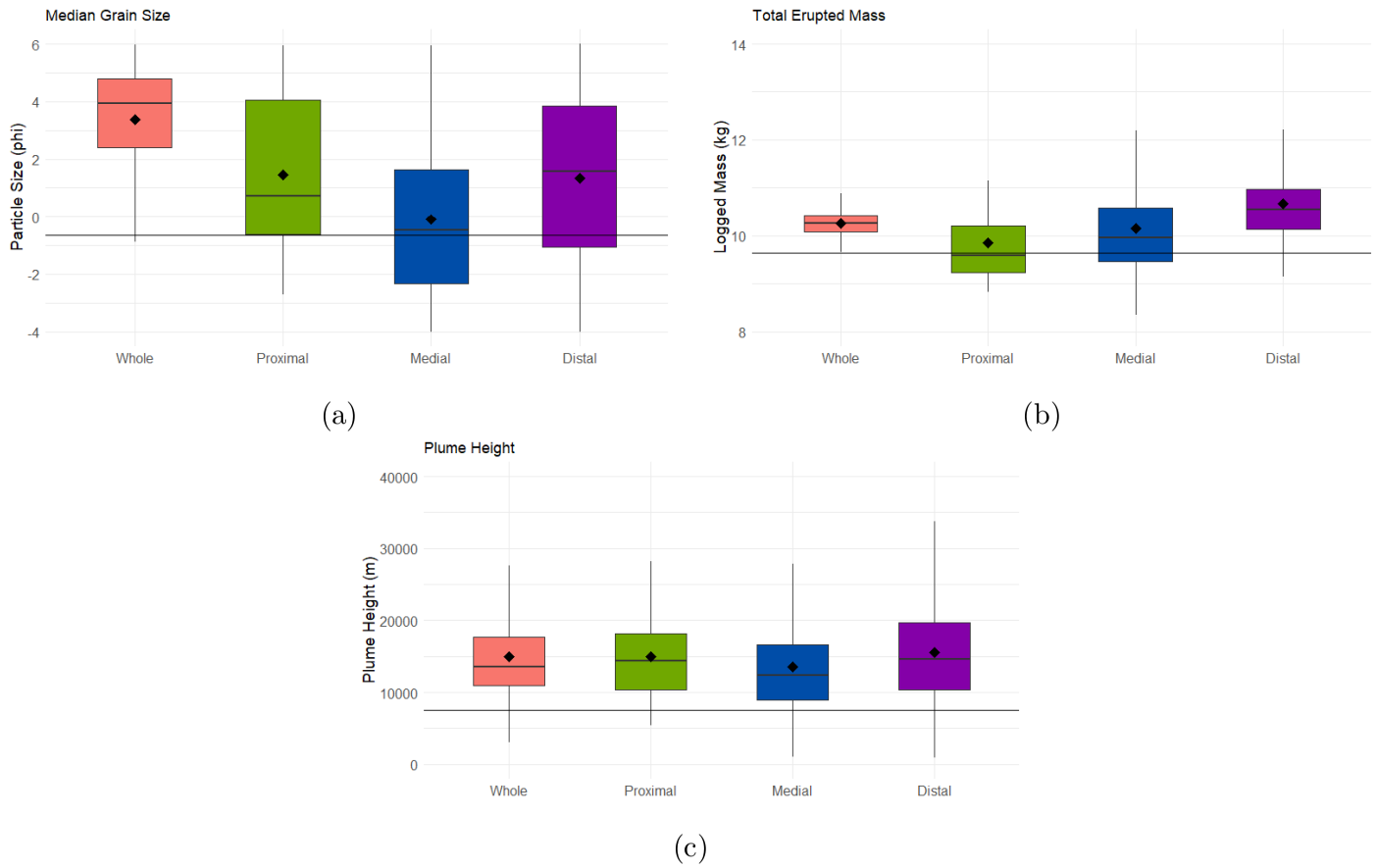


Figure 4.9: Boxplots of (a) median grain size (phi), (b) total erupted mass (kg), and (c) plume height (m) input parameter ranges of simulations that have a hit rate of 50% or more. The dot within the boxplots represents the mean. The solid black line represents the observed value.

From Scott et al. (2025), we know that median grain size is one of the most influential input parameters in output variance. Given its importance, we examined how this input parameter behaved in our forecasts (Figure 4.9a). Across the whole sampling domain, the best value (which we take as the median value) is 4 phi (very fine particles). However, in both proximal and distal locations, the best value is found to be 1 phi, and in medial zones, the best value is -1 phi (coarser particles). When comparing the best-fit values from the forecasts to the observed deposit value, the medial zone represents the observed value well; however, there is a large difference between the full domain best-fit value and the observed deposit value of 4.6 phi (best fit: 4, observed: -0.65)) (i.e., the best-fit value is much finer than the observed deposit value).

This discrepancy likely reflects Tephra2's assumption of a vertically rising plume (Mannen et al. 2020). In reality, bent-over plumes strongly influence grain size patterns. For example, for this eruption at Mount Ruapehu, observations show a rapid decrease in lapilli (>-1 phi) within 10 km of the vent, dominance of lapilli and coarse ash (-1 to 2 phi) between 10 and 30 km, and progressively finer ash (> 2 to 3 phi) deposition beyond 30 km (Bonadonna, Phillips, and Houghton 2005). Such behaviour contrasts with strong, vertically rising plumes, where coarse particles are not dominant throughout the deposit (e.g., Bursik et al. 1992). Because Tephra2 does not simulate bent-over plume dynamics, the spatial variability we observe in our best median grain size values is likely an artefact of this simplification rather than a reflection of true eruptive processes.

Looking at the input parameters that had informed sampling distributions (TEM and plume height), TEM showed notable variations across sampling zones (Figure 4.9b). The best-fit TEM across the full domain (10.26; $1.8\text{x}10^{10}$ kg/m^2) exceeded the observed deposit (9.65; $4.51\text{x}10^{9}$ kg/m^2), indicating that Tephra2 required larger mass inputs to reproduce the deposit. Proximal and median zones were closer to observations (9.59 and 9.97, respectively), while distal sites required substantially higher values (10.54).

Plume height was more consistent, with best-fit values ranging from 12 to 15 km across sampling zones (Figure 4.9c). The observed plume height (7.5 km) lay near the lower quartile of all sampling locations, showing again that the model requires larger plume heights to reproduce the deposit, especially at distal sites where uncertainty was the largest.

These patterns align with previous Tephra2 inversion studies, which similarly identified

plume height as poorly constrained and TEM as both easier to constrain and more influential on output variance (Scollo et al. 2008; Volentik et al. 2010; Magill et al. 2015; Jenkins et al. 2024; Scott et al. 2025). Importantly, although plume height remains uncertain, its relatively limited influence on output variance means that this uncertainty may not strongly impact forecast accuracy (Scott et al. 2025).

This mismatch in plume height, TEM, and median grain size complicates short-term forecasting, as the measured values are not always the values with which the model performs best. Since the true value is not, in this case, the one that gives the most accurate forecast, it creates uncertainty when making real-time decisions.

For the remaining input parameters - diffusion coefficient, plume shape (alpha and beta), standard deviation of grain size, and pumice and lithic density - there was no clear pattern in the values associated with more accurate forecasts. Accurate forecasts were generated across a broad range of values, regardless of sampling distribution or spatial location. This suggests that these input parameters did not strongly constrain model performance, and much of the sampled input parameter space contributed to more accurate forecasts.

4.4.4 Etna and Villarrica

To extend the analysis beyond a single case study, we apply the same methodology to two additional eruptions: the 28 February 2021 eruption of Mount Etna (data from Mereu et al. (2025)) and the 3 March 2015 eruption at Villarrica, Chile (data from Romero et al. (2018)). Mount Etna is classified as an M1 volcano by Mastin et al. (2009), which is a basaltic volcano with historic plume heights of less than 5 km and/or less than VEI 2. Mount Etna is classified by Deligne (2021) as mafic in composition. Villarrica is classified as an S1 volcano (Mastin et al. 2009), like Mount Ruapehu (silicic volcano with plume heights less than 5 km and/or less than VEI 2); through Deligne (2021), it is classified as a Mafic composition volcano. Further information about input parameter sampling and distributions can be found in Tables 4.6,4.7.

Table 4.6: Informed input parameter sampling for Mount Etna produced from the IVESPA (Aubry et al. 2021) and Deligne (2021) (Deligne 2021) databases.

Database	Input Parameter	Mastin et al. (2009) Classification - M1	Deligne (2021) Classification - Mafic Magma	Distribution	Unique Volcanoes	Unique Eruptions
IVESPA	Plume Height (km)	MeanLog: 1.826 SDLog: 0.326	MeanLog: 1.98 SDLog: 0.496	Log-normal (M1, Mafic)	Mafic - 12 M1 - 3	Mafic - 38 M1 - 16
	Total Erupted Mass (kg) Logged	Shape: 12.87 Scale: 8.76	Shape: 45.38 Scale: 4.589	Weibull (M1, Mafic)	Mafic - 12 M1 - 3	Mafic - 39 M1 - 16
Deligne (2021)	Plume Height (km)	Shape: 2.43 Scale: 6.41	Shape: 1.42 Scale: 10.24	Weibull (M1, Mafic)	Mafic - 12 M1 - 1 (Mount Etna)	Mafic - 54 M1 - 17
	Total Erupted Mass (kg) Logged	MeanLog: 2.176 SDLog: 0.052	MeanLog: 2.26 SDLog: 0.109	Log-normal (M1, Mafic)	Mafic - 11 M1 - 1 (Mount Etna)	Mafic - 71 M1 - 36

Table 4.7: Informed input parameter sampling for Villarrica from the IVESPA (Aubry et al. 2021) and Deligne (2021) (Deligne 2021) databases.

Database	Input Parameter	Mastin et al. (2009) Classification - S1	Deligne (2021) Classification - Mafic Magma	Distribution	Unique Volcanoes	Unique Eruptions
IVESPA	Plume Height (km)	MeanLog: 2.36 SDLog: 0.251	MeanLog: 1.98 SDLog: 0.496	Log-normal (S1, Mafic)	Mafic - 12 S1 - 7	Mafic - 38 S1 - 15
	Total Erupted Mass (kg) Logged	Mean: 9.66 SD: 0.957	Shape: 45.38 Scale: 4.589	Gaussian (S1), Weibull (Mafic)	Mafic - 12 S1 - 7	Mafic - 39 S1 - 15
Deligne (2021)	Plume Height (km)	Shape: 1.66 Scale: 17.056	Shape: 1.42 Scale: 10.24	Weibull (S1, Mafic)	Mafic - 12 S1 - 7	Mafic - 54 S1 - 33
	Total Erupted Mass (kg) Logged	Shape: 12.35 Scale: 12.24	MeanLog: 2.26 SDLog: 0.109	Weibull (S1), Log-normal (Mafic)	Mafic - 11 S1 - 5	Mafic - 71 S1 - 17

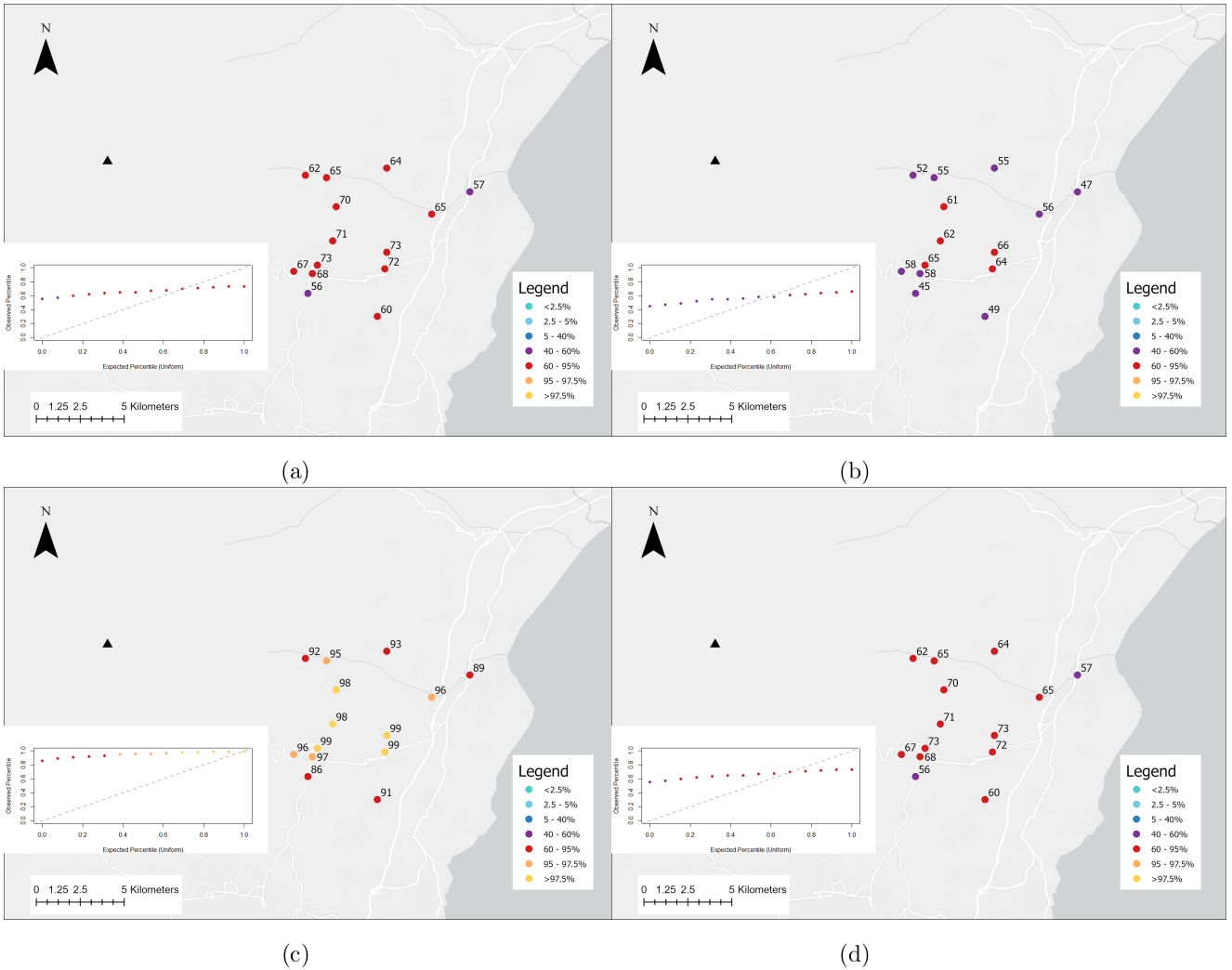


Figure 4.10: Percentiles of observed tephra mass for the 28 February 2021 Mount Etna eruption relative to simulated distributions: (a) uninformed, (b) IVESPA Mafic, (c) Deligne (2021) M1, and (d) all simulations from all sampling distributions.

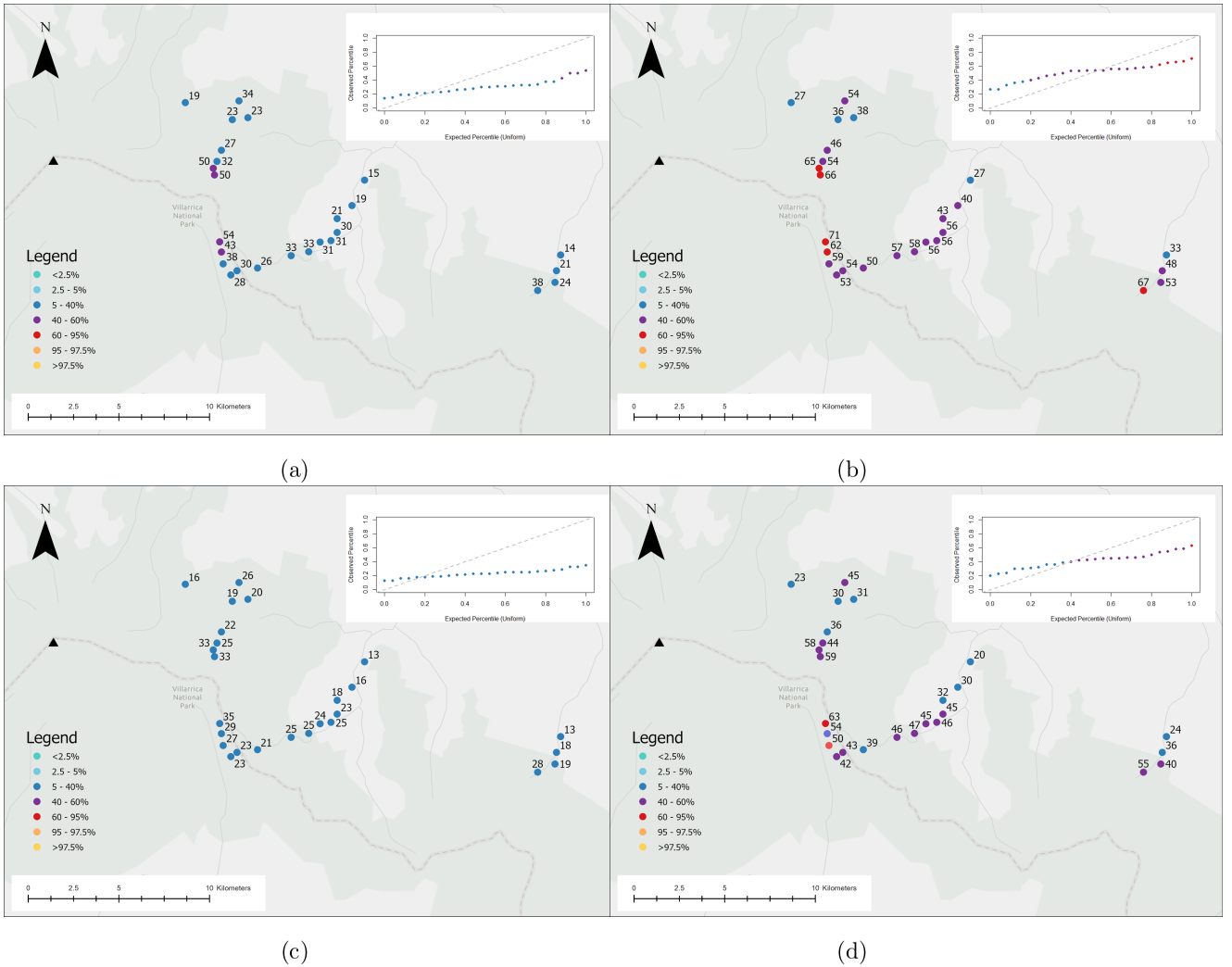


Figure 4.11: Percentiles of observed tephra mass for the 3 March 2015 Villarrica eruption relative to simulated distributions: (a) uninformed, (b) IVESPA Mafic, (c) Deligne (2021) S1, and (d) all simulations from all sapling distributions.

For Mount Etna, all sampling distributions show significant departures from uniformity (p-values ranging from 10^{-10} to 0.0047), indicating that the simulated percentiles are systematically biased rather than evenly distributed across the percentile space (Figure 4.10). In contrast, results for Villarrica are more mixed. While several sampling distributions reject uniformity, the IVESPA S1 distribution ($p = 0.052$) and the Deligne (2021) Mafic distribution ($p = 0.1168$) show no strong evidence against uniformity, suggesting that simulations for these cases are more evenly distributed relative to the observed mass (Figure 4.11).

The uninformed sampling distribution consistently overestimates deposit mass, with average percentiles of approximately 22% for Mount Etna and 24% for Villarrica. In comparison, the informed distributions generally shift closer to the observed values for Mount Etna, with average percentiles ranging between 57 - 95% (Figure 4.10. Improvements are less consistent at Villarrica, where informed distributions produce average percentiles between 13 - 33% (Figure 4.11). These results highlight that while informed priors can improve forecast performance in some cases, they do not consistently outperform the reference sampling baseline.

Another interesting observation is that, due to the limited size of the ESP database, the Deligne (2021) M1 sampling distribution was comprised of data only from Mount Etna. Percentiles of observed tephra mass were on the high side compared to other informed and uninformed distributions, ranging between 86 and 99%. This suggests that even for a well-studied volcano like Mount Etna, relying on past eruption data may not reliably capture the distribution of outcomes for a specific event.

Figure 4.12: Median hit rate for all sampling locations for the 28 February 2021 Mount Etna eruption, looking at (a) uninformed, (b) IVESPA M1, (c) Deligne (2021) Mafic, and (d) all simulations.

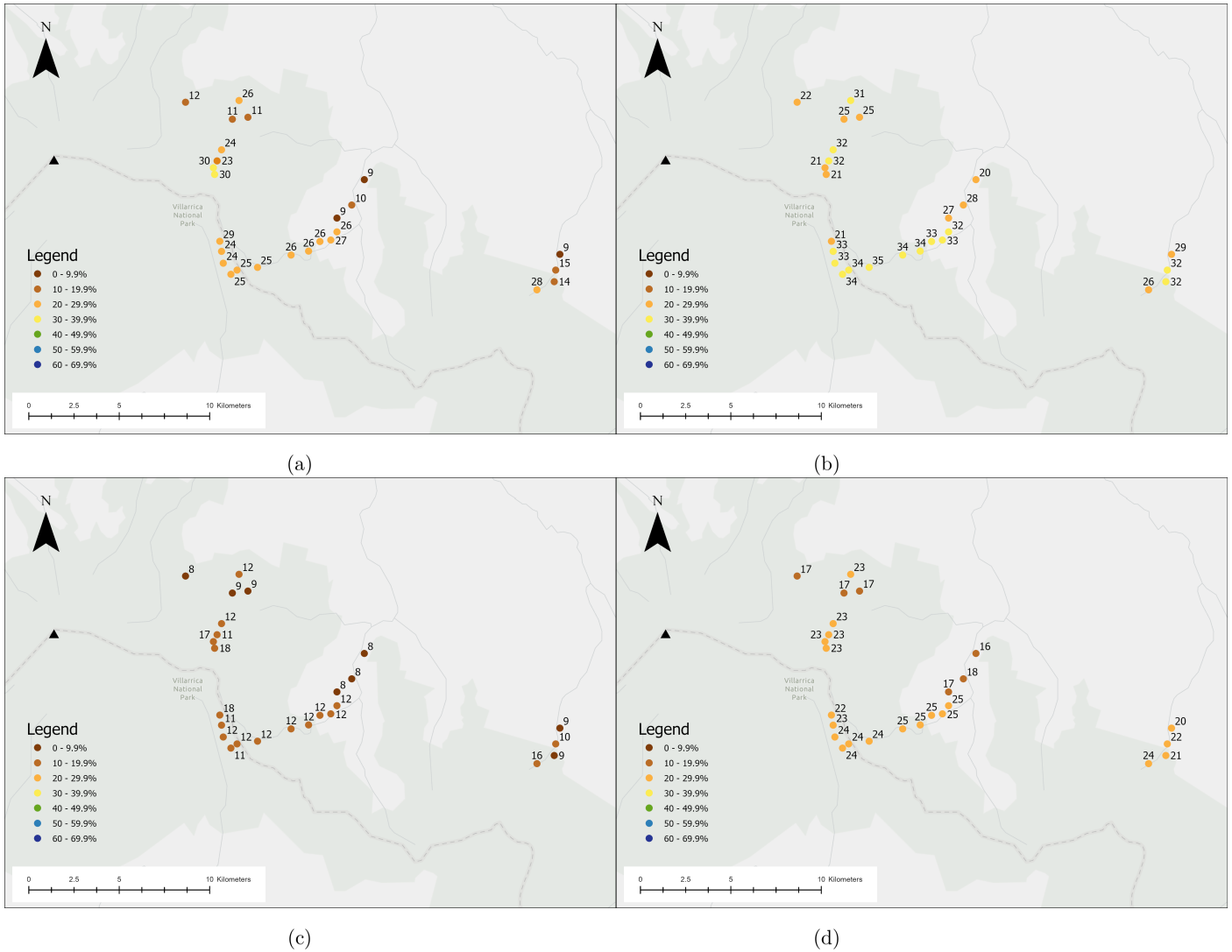


Figure 4.13: Median hit rate for all sampling locations for the 3 March 2015 Villarrica eruption, looking at (a) uninformed, (b) Deligne (2021) Mafic, (c) Deligne (2021) S1, and (d) all simulations.

For Mount Etna, the reference sampling distribution produces broadly consistent hit rates across most locations, typically around 20%, although one location shows a lower value near 10% (Figure 4.12). The IVESPA M1 sampling distribution performs worst overall, with most locations exhibiting relatively low hit rates (6 to 17%). This outcome is notable given that approximately 88% of the data used to construct the IVESPA M1 distribution is derived from past Mount Etna eruptions. In contrast, the Deligne (2021) Mafic distribution consistently produces the highest hit rates across the sampling domain.

At Villarrica, the spatial pattern of hit rates differs somewhat from both Mount Ruapehu and Mount Etna. For the reference distribution, the most successful forecasts tend to occur within the central portion of the deposit closer to the vent, while lower hit rates are observed further downwind along the dispersal axis at further distances (15 km) from the vent (Figure 4.13). As with Mount Etna, the Deligne (2021) Mafic distribution produces the highest hit rates overall, whereas the Deligne (2021) S1 distribution performs worst across most sampling locations.

4.5 Limitations

4.5.1 Eruption Source Databases

The difference in sampling distribution performance likely stems from the construction of the two ESP databases. The IVESPA database focuses exclusively on historic eruptions, with ESPs derived from independently constrained observations, such as using separate methods for estimating plume height and MER. In contrast, the Deligne (2021) database includes both historic and prehistoric eruptions, and some ESPs are derived from model outputs or geological reconstructions that lack independent validation. Additionally, 36% of the Deligne (2021) database consists of eruptions from Aotearoa New Zealand, which may introduce a regional bias (although this did not positively influence the results in this case). Like most eruption databases, both IVESPA and Deligne (2021) are skewed toward larger eruptions, those that are either well-preserved geologically or well-documented.

4.5.2 Classification Systems

In this study, we varied only TEM and plume height because the IVESPA and Deligne (2021) databases lack comprehensive data for other key Tephra2 input parameters. While these databases are widely used to support volcanic hazard assessments and forecasts (e.g., Michaud-Dubuy, Carazzo, and Kaminski 2021; Pardini et al. 2022; Trancoso et al. 2022; Michaud-Dubuy et al. 2024), they often omit input parameters that significantly influence model outcomes (Scott et al. 2025). For instance, input parameters such as median grain size, plume shape (described by the alpha and beta input parameters, which control the vertical ash distribution in the plume), and the diffusion coefficient are either sparsely represented or completely absent. For plume shape and the diffusion coefficient specifically, these input parameters are commonly inferred through inversion modelling, rather than measured directly (e.g., Bonadonna, Macedonio, and Sparks 2002; Bonasia et al. 2010; Mannen 2014), limiting their presence in ESP databases.

Even when data are available, their applicability can be limited. The IVESPA database provides TGSDs for 29 eruptions, where only one was available for a volcano in our S1 category (Tungurahua, median grain size 2.4 phi). Notably, this value differs from the operational value used in Aotearoa New Zealand (0.82 phi), as well as from the best recorded and simulated values from the 1996 Ruapehu eruption (-0.6 and 4 phi, respectively). However, given that only a single TGSD was available for the S1 category, informed sampling was not feasible. Consequently, we treated TGSD input parameters (median grain size and standard deviation of grain size) as uninformed distributions.

Likewise, although the IVESPA database includes bulk density values, we excluded them from our simulations for two reasons: (a) Tephra2 requires separate estimates for pumice and lithic densities; and (b) the Deligne (2021) database lacks any density data, making cross-database comparisons difficult.

To address these gaps, we compiled values for otherwise undocumented input parameters from published Tephra2 modelling studies. These literature-based values were then compared to: (a) best-measured values from the 17 June 1996 eruption (Table 4.1); (b) the median value for each input parameter across all simulations that achieved a 50% or more accuracy in impact-based forecasts across the full spatial domain - hereafter referred to as "best-fit" values; and (c) operational values currently used in Aotearoa New Zealand (Hurst and Davis 2017; Trancoso

et al. 2022) (Figure 4.14). This comparison allows us to evaluate whether the values commonly applied in the literature, operationally, and our own forecasts are consistent, or whether key mismatches exist that could explain differences in forecast performance.

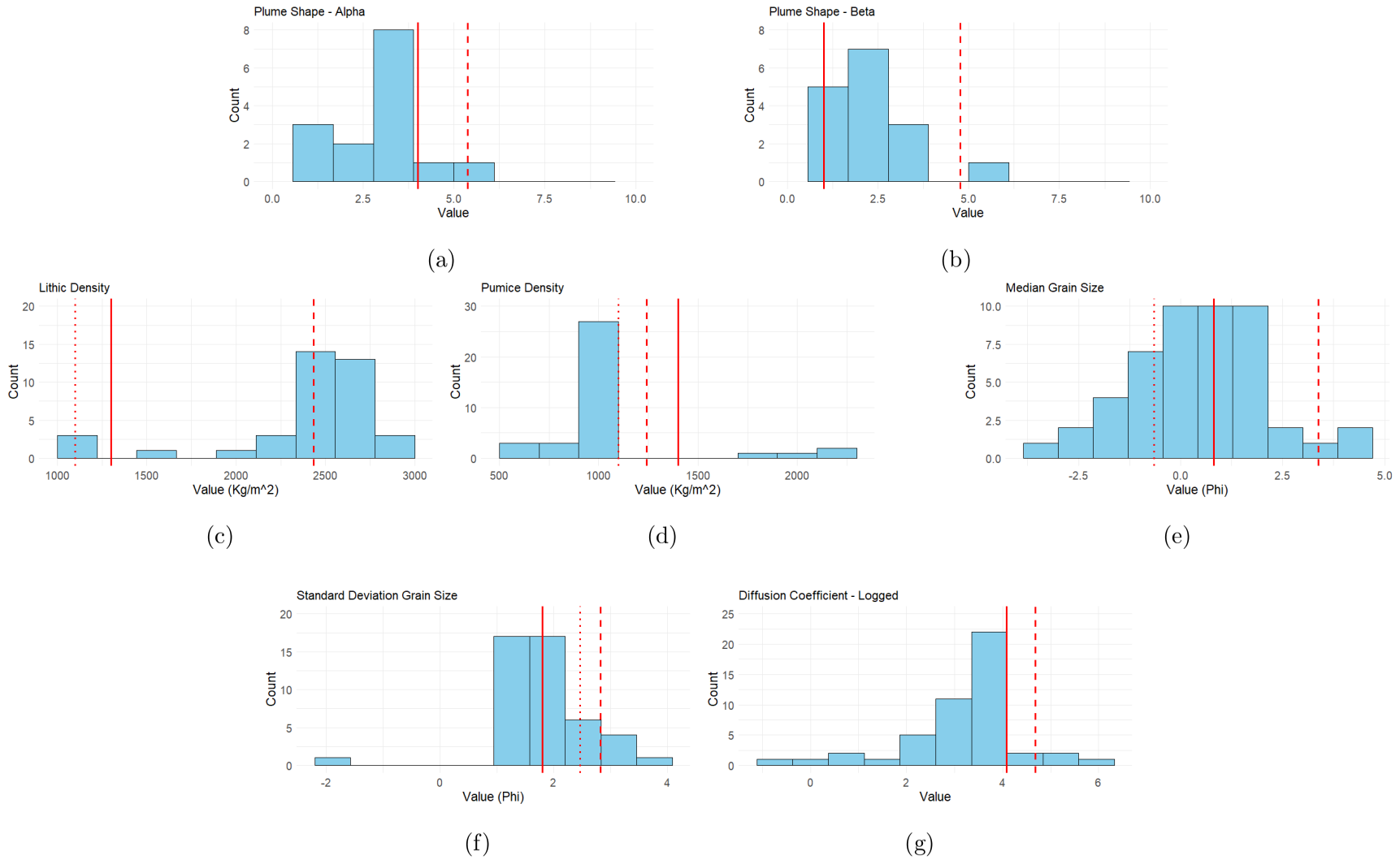


Figure 4.14: Distributions of Tephra2 input parameter values compiled from the literature (a) Plume shape alpha, (b) Plume shape beta, (c) Lithic density (kg/m^3), (d) Pumice density (kg/m^3), (e) Median grain size (phi), (f) Standard deviation of grain size (phi), (g) Diffusion coefficient. Red vertical lines key reference values: dotted lines represent the best-recorded values from the 17 June 1996 eruption (Table 4.1; only available for panels c,d,e,f); solid lines show values used in operational forecasting in Aotearoa New Zealand (Hurst and Davis 2017; Trancoso et al. 2022); and dashed lines indicate the best-fit values from our simulations that achieved $\geq$ 50% accuracy in impact-based forecasts across the full spatial domain.

Figure 4.14 compares literature-derived input parameter values against operational and empirical benchmarks. Most input parameters fall within broad ranges reported in the literature, though plume shape beta shows a clear bimodal distribution. Our best-fit value for plume shape beta lies further along the distribution tail, indicating a more bottom-heavy plume with relatively more mass concentrated at lower elevations.

The TGSD distributions show broader variation (Figure 4.14). Median grain size follows a near-normal distribution in the literature, with the operational value (0.82 phi) close to the centre. Our best-fit value (4 phi), however, lies at the upper end of the distribution. These differences in TGSD values likely reflect the difficulty in accurately reconstructing TGSDs from deposit data (Bonadonna and Houghton 2005; Bonadonna, Biass, and Costa 2015). TGSDs are highly sensitive to the spatial pattern of sampling locations, a well-documented source of uncertainty (e.g., Bonadonna and Houghton 2005; Volentik et al. 2010), and this was evident in our own findings (Figure 4.9). Jenkins et al. (2024) and Scott et al. (2025) found that the median grain size is one of the most influential input parameters, which underscores the need for greater care in the TGSD values selected for modelling.

In Aotearoa New Zealand, TGSD values used operationally (median grain size = 0.82 phi, standard deviation grain size = 1.8) for forecasting ash deposition from andesitic volcanoes, such as Mount Ruapehu, are derived from the Vesuvius 79 AD eruption. This is problematic for a couple of reasons: (1) Vesuvius has a history of much larger eruptions (hence it is classified as an S2 volcano, Mastin et al. 2009), and (2) the 79AD eruption is much larger than any previous eruption from Mount Ruapehu (including the other volcanoes in Aotearoa New Zealand classified as intermediate volcanoes, Hurst and Davis 2017). The use of such a dissimilar TGSD introduces potential bias. Improving TGSD sampling and field reconstruction, and using analogues from eruptions that are similar in magnitude, magma composition, and eruptive behaviour to the volcano being forecast, would likely reduce bias and improve the accuracy of ashfall forecasts.

4.6 Conclusions

Accurate ashfall forecasting is essential for effective hazard mitigation. Traditionally, such forecasts are informed by prior knowledge of a volcano, particularly its eruptive history and

magma composition. However, this assumption that prior information leads to more accurate forecasts has, until now, remained untested. In this study, we evaluated the accuracy of both mass-based and impact-based forecasts using an uninformed sampling distribution and informed sampling distributions by volcano classifications at Mount Ruapehu. Our findings reveal that biasing model inputs with prior information does not consistently improve forecast accuracy. This challenges prevailing assumptions in volcanic hazard modelling and highlights the need for more dynamic, context-specific approaches to forecast development, such as weak and strong plume morphologies.

A key limitation of this study is that it is based on a single VATDM. While our methodology and findings provide useful insights into the forecasting capabilities of Tephra2, extending this research to other VATDMs would also allow users to evaluate the consistency and comparative performance of different models in forecasting contexts.

Future work should particularly examine plume morphology. The IVESPA database includes classifications on plume morphology (i.e., weak or strong plumes). Given that plume morphology directly influences ash dispersal patterns (Koyaguchi and Ohno 2001; Bonadonna and Phillips 2003), incorporating this information into input parameter selection could improve forecast accuracy. Specifically, exploring whether ESPs associated with historic weak or strong plumes lead to different forecast outcomes may reveal whether plume morphology is a more appropriate basis for scenario development than broader volcano classifications based on eruptive history or magma composition. Because plume morphology can often be anticipated in near-real time, using forecasted wind profiles and regional probabilities of plume type occurrence (Ishii and Iriyama 2024), future simulations that account for likely plume morphology on the day of eruption may enable more accurate forecasts. Complementing such an approach with expert elicitation may further strengthen forecasts under uncertainty (e.g., Bernard et al. 2024; Engwell et al. 2024).

Chapter 5

The Importance and Influence of the Total Grain Size Distribution

Building on Chapter Four, which examined how prior knowledge influences mass forecast accuracy, this chapter uses those simulations to assess the accuracy of grain size forecasts and further investigates how input parameters affect the spatial and granulometric distribution of tephra forecasts.

5.1 Introduction

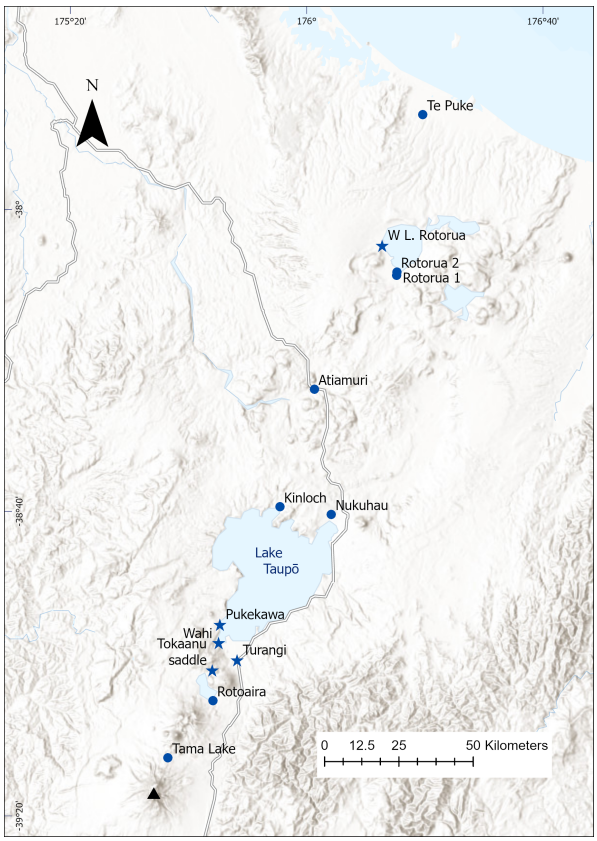
The sensitivity of Tephra2 outputs to median grain size inputs using the 17 June 1996 eruption of Mount Ruapehu was explored in Chapter Four. It showed that the most accurate total grain size distribution (TGSD) inputs for producing the outputs closest to the sampled value varied depending on sampling distance from the vent. However, it did not assess how well the model simulates the full particle size distribution of the tephra deposits.

Pushing the investigation further, this chapter focuses on the grain size outputs generated by Tephra2 and how they compare to observed deposit data. The 17 June 1996 Mount Ruapehu eruption is used as a case study, which represents one of the largest and most spatially extensive sample datasets ever collected from a single tephra deposit (Bonadonna and Houghton 2005; Pioli, Bonadonna, and Pistolesi 2019), comparing modelled TGSDs with those derived from field measurements provided by Shane Cronin (personal communications) (Figures 5.1a, 5.2). Attention is placed on the fine ash fraction as an illustrative example, since fine ash (<63 μm,

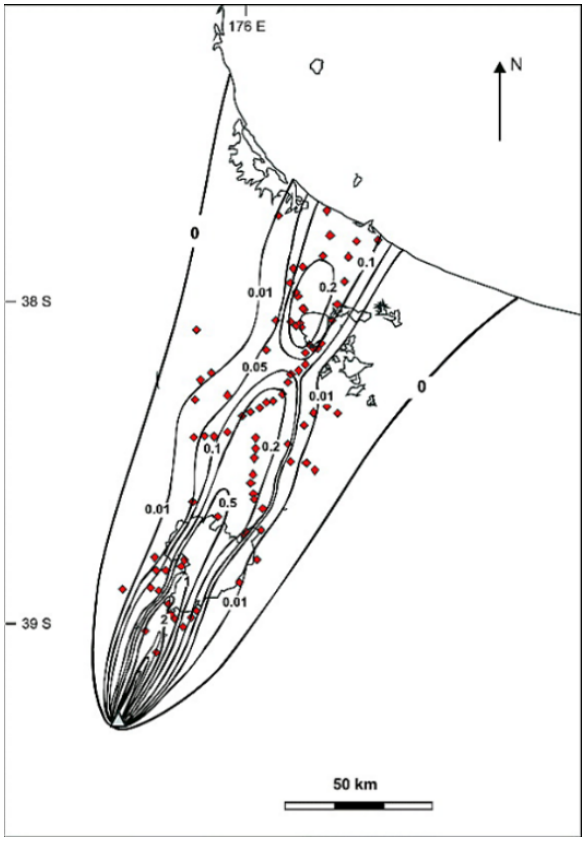
>∼4 phi) is more damaging (compared to coarse ash), causing electrical flashovers (Wardman et al. 2012; Wilson et al. 2014) and respiratory diseases (Forbes et al. 2003; Horwell and Baxter 2006; Horwell 2007). While only looking at the fine ash fraction simplifies the continuous TGSD into a categorical fraction, it provides a practical way to visualise model-data differences. This comparison allows for a detailed assessment of model performance and provides insight into whether model outputs accurately reflect the natural variability of deposits.

5.2 Observed versus Simulated Fine Ash Fraction

Tephra2 is one of only three volcanic ash transport and dispersion models (VATDMs) (with Ash3D and Fall3D) that output TGSD for each simulated location (Bonadonna 2010), providing a direct comparison between modelled and observed values.



(a)



(b)

(c)

Figure 5.1: (a) Total grain size sampling locations from the 17 June 1996 Mount Ruapehu eruption (data from Shane Cronin). Stars represent locations that have both grain size data and mass data from Chapter Four; (b) Isomass contours of the deposit from Bonadonna, Phillips, and Houghton (2005); (c) Photo of the 17 June 1996 eruption, showing a very prominent bent-over plume (photo credit: Lloyd Homer (photographer) and Earth Sciences New Zealand, formerly GNS Science, VML ID: 3547).

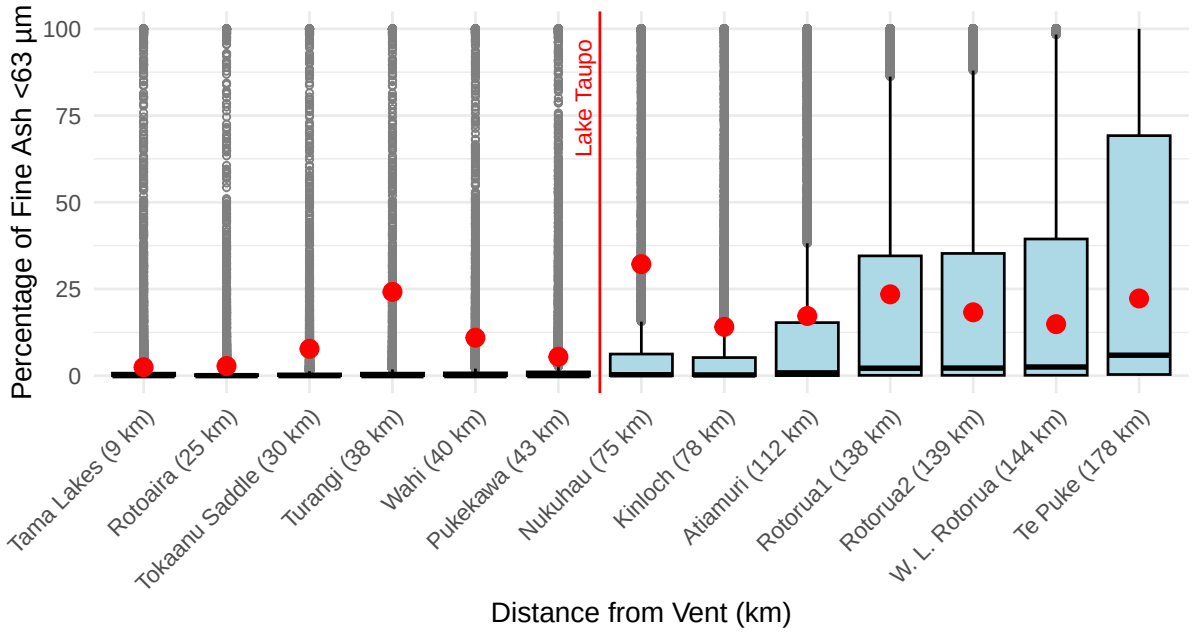


Figure 5.2: Percentage of fine ash (<63 μm, >$\sim$4 phi) at each sampling location of the observed data (red dots) and the range of simulated data from Tephra2 (boxplots, n = 7000). Simulated data were derived from the 17 June 1996 Mount Ruapehu eruption case study (see Chapter Four). Sampling sites are ordered from most proximal to most distal relative to Mount Ruapehu. The red vertical line indicates the location of Lake Taupō. Grey dots are considered outliers.

In a typical eruption, we expect the proportion of fine ash to increase with distance from the vent, since coarser particles settle out sooner due to their greater mass (Bonadonna and Houghton 2005; Rose and Durant 2009). The observed data broadly reflect this pattern (Figure 5.2). Proximal and medial sites have lower fine ash percentages (around 2.4 to 10.9%), except Turangi, which is higher at 24%. Distal sites beyond Lake Taupō show higher fractions of fine ash, mostly >15%, though Nukuhau stands out at 32%.

The simulated data also capture this general trend. However, the observed fine ash percentages are larger than the majority of simulations at proximal locations, such as Turangi (eCDF = 0.96) and Tokaanu Saddle (eCDF = 0.92) (Table 5.1), indicating that simulations tend to underestimate fine ash deposition near the vent. At distal sites past Lake Taupō, observed values are closer to the median of the simulated distributions, suggesting better agreement between simulations and observations at greater distances.

Of the three locations with both deposit mass (Chapter Four, Bonadonna, Phillips, and Houghton 2005) and fine ash data, Turangi stands out with a relatively high fine ash percentage

Table 5.1: Empirical cumulative distribution function (eCDF) values representing the proportion of simulated fine ash percentages (<63μm, >$\sim$4 phi) that are less than or equal to the observed values. Values in brackets are distances from the vent.

	Observed Percentage	CDF Value
Tama Lakes (9 km)	2.44	0.83
Rotoaira (25 km)	2.78	0.89
Tokaanu Saddle (30 km)	7.76	0.92
Turangi (38 km)	24.16	0.96
Wahi (40 km)	10.95	0.92
Pukekawa (43 km)	5.46	0.86
Kinloch (75 km)	14.08	0.83
Nukuhau (78 km)	32.15	0.89
Atiamuri (112 km)	17.21	0.76
Rotorua1 (138 km)	23.45	0.71
Rotorua2 (139 km)	18.27	0.69
W. L. Rotorua (144 km)	14.86	0.66
Te Puke (178 km)	22.26	0.62

(24.2%) despite a moderate mass (0.318 kg/m^2), indicating that finer particles can dominate even when mass deposition is low. Conversely, Wahi has the highest deposition mass (1.36 kg/m^2) but the lowest fine ash content (10.9%), indicating coarser ash deposition. W. L. Rotorua shows both low mass (0.266 kg/m^2) and a moderate fine ash content (14.9%), consistent with distal sites where finer ash fractions dominate despite lower total deposit mass. These examples underscore that mass alone is not a reliable predictor of particle size distribution.

For this example eruption (Table 5.1), the mismatch between observed and simulated fine ash fractions, particularly in proximal locations, reflects a broader issue in TGSD reconstruction, consistent with prior findings that TGSD often deviates from field observation (e.g., Pioli, Bonadonna, and Pistolesi 2019; Wiejaczka 2023). Although TGSD is only one metric, Chapter Four also showed that reproducing/forecasting output mass is similarly problematic.

5.3 Discussion

Examining potential causes for the disparities between observed and simulated data, particularly at Turangi, suggests that local wind patterns influenced by Lake Taupō are likely contributing factors (Poulidis et al. 2017; Poulidis et al. 2018). These influences are not well captured by

the Tephra2 simulations, which tend to underestimate fine ash near the vent (Figure 5.2, Table 5.1).

This mismatch highlights a key limitation in how fine ash transport is parameterised in VATDM. Across all simulations, spatial trends in fine ash fractions were inconsistently reproduced: at the most distal site (Te Puke), 70% of the simulations yielded a higher proportion of fine ash than coarse ash, compared to 97.6% at the most proximal site (Tama Lakes). This means that Tephra2 almost always predicts fine ash dominance near the vent but becomes substantially more variable at greater distances. Importantly, this pattern persists even when all input parameters were sampled uniformly (uniform sampling distribution; Chapter Four), indicating that the discrepancy reflects a systematic limitation in the model's treatment of fine ash transportation and deposition, rather than random variability or errors in the observed data.

This behaviour reflects a limitation in Tephra2's treatment of particle settling, where sedimentation is treated as a continuous process governed by terminal settling velocity, without accounting for the prolonged suspension of particles due to wind variability and turbulence.

These patterns corroborate with other findings that Tephra2 inadequately represents proximal deposition processes (Connor, Connor, and Bonadonna 2008), which in turn affects TGSD (and mass) realism. Comparable issues have been identified in other VATDMs, such as Ash3D, which tends to over-represent coarse ash in distal locations, potentially due to assumptions in drag coefficients and terminal velocity characterisations (Buckland et al. 2022). These limitations underscore the need to refine how TGSD is both parameterised and interpreted within VATDM. While TGSD outputs are not currently used in operational forecasting, improving their accuracy would enhance hazard assessments by enabling better impact characterisations (e.g., particle size-specific effects on aviation, infrastructure, and health) and providing an additional metric for model verification and validation (Chapter Two).

5.4 Influence of Plume Shape

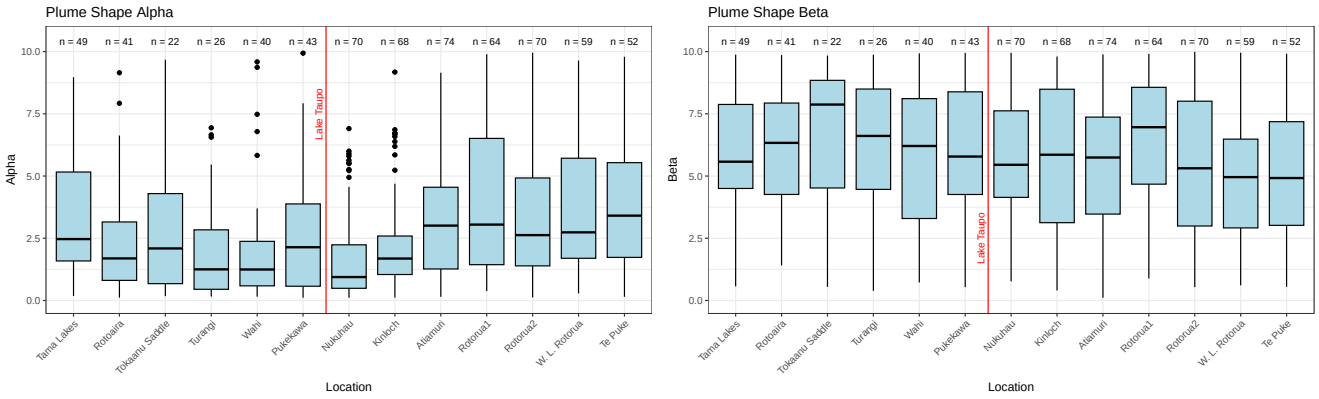


Figure 5.3: Boxplots of plume shape alpha and beta input parameter ranges of simulations (7000) that were within 5% of the observed percentage of fine ash (<63μm). Values above each boxplot indicate the number of simulations contributing to each location ($n = x$).

When examining the distributions of input parameters associated with simulations that produced TGSD outputs closest to observations, no clear spatial trends were observed in most input parameters. However, plume shape parameters alpha and beta show distinct patterns that diverge from those seen in Chapter Four (Figure 5.3). For simulations with accurate mass outputs (Chapter Four), the input values of alpha and beta were approximately normally distributed across the full sampling range (0.1 to 10), suggesting no strong sensitivity to specific values. In contrast, simulations with more accurate TGSD outputs tend to cluster around the lower alpha values (<5) across all sampling locations, while beta values sit at 5 or above across all sampling locations (Figure 5.3).

In physical terms, alpha controls the vertical spread of ash in the plume, while beta governs the horizontal spread. When alpha < beta, the plume expands more horizontally than vertically, resulting in a bottom-heavy particle concentration (Figure 5.4). This configuration is often associated with lower plume height, strong wind shear, or reduced buoyancy, all of which limit vertical lifting and promote lateral spreading (Carey and Sparks 1986; Bursik 2001). The 17 June 1996 event featured a well-documented bent-over plume, producing a relatively narrow deposit (Figure 5.1). In the simulations, the plume shape input parameters exert a strong influence on the vertical distribution of particles, leaving a clear signature on the output TGSD. While these effects are less distinct in the output mass, plume shape input parameters remain

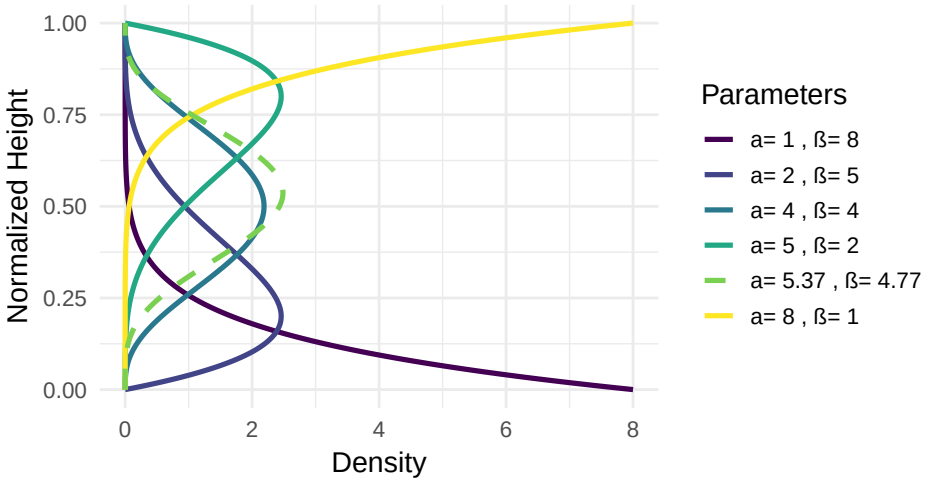


Figure 5.4: Beta distributions for vertical ash column profiles. The green dashed line represents the best-fit value of alpha and beta from Chapter Four.

very influential, reinforcing the findings from Chapter Three.

5.5 Implications

These results raise important questions on the most useful aspects of TGSD for ash modelling and field studies. While TGSD is often described in terms of its distribution (Pioli, Bonadonna, and Pistolesi 2019), including median and standard deviation, the median grain size emerges as the key input parameter influencing model outputs (Chapter Three). Given that deposit datasets are often spatially limited and may underrepresent fine fractions, it may be more practical to focus on ensuring that TGSD measurements are sufficient to accurately constrain the median grain size, even if the full distribution cannot be completely characterised, particularly since individual outcrops rarely reflect the complete TGSD (Beckett et al. 2015; Rossi, Bonadonna, and Degruyter 2019; Pioli, Bonadonna, and Pistolesi 2019).

Unfortunately, TGSD is one of the most difficult to constrain of all ESPs (Bonadonna, Biass, and Costa 2015; Pioli, Bonadonna, and Pistolesi 2019). Due to technical limitations, no instruments, such as radars and satellites, can currently provide comprehensive information on the TGSD during an eruption (Bonadonna et al. 2011b). However, there are some promising advances through in-situ and airborne sampling techniques (e.g., Elissondo et al. 2016; Thivet et al. 2025). Reconstruction of TGSD is challenging even after an eruption and often inconsistent

due to variable sampling density, integration techniques, and the differential exposure and preservation of deposits (Bonadonna and Houghton 2005; Bonadonna, Biass, and Costa 2015). Most grain size studies underrepresent the fine ash fraction (e.g., Rose and Durant 2011; Durant 2015), and proximal samples have a large effect on the resulting TGSD, as they reflect the most mass (Bonadonna and Houghton 2005; Andronico et al. 2014).

Because of all these challenges, TGSD used in modelling is typically either assumed (Mastin et al. 2009; Folch 2012) or derived from statistical models such as those proposed by Costa, Pioli, and Bonadonna (2016). These statistical approaches infer a TGSD by fitting parametric models (e.g., Weibull or log-normal) to deposit data across multiple sampling sites, producing a single, eruption-wide distribution. In the case of the 17 June 1996 eruption, Cronin et al. (2003) applied Gaussian fits to grain-size frequency data at individual sampling sites, obtaining local modal grain sizes between 0.7 and 3.7 phi. In contrast, Bonadonna and Houghton (2005) used Voronoi tessellation to integrate the data spatially, estimating an eruption-wide median grain size of 2.7 phi. Together, these examples illustrate that different methods capture different scales of variability and that a single TGSD may not always be the most informative metric within ash deposition modelling.

Recognising the limitations of assuming a single, eruption-wide TGSD, other studies have sought to derive empirical relationships between eruptive conditions and expected grain-size distributions. One example is the approach by Costa, Pioli, and Bonadonna (2016), who fitted a bi-Gaussian distribution with coarse and fine modes at -3.02 and 0.04 phi, respectively. Their empirical TGSD model estimates grain-size distributions based on plume height and magma viscosity, using bi-Gaussian fits to deposit data. This method provides rapid TGSD approximations for short-term and near-real-time forecasting when direct deposit data are unavailable, but it is only valid within the tested range (2 to 30 km plume heights, 10^2 to 10^8 Pa s viscosities). Its application is limited by large uncertainties (>50% for median grain size and >40% for distribution across bins).

Moreover, Chapter Four demonstrated spatial variability in the input median grain size across proximal and distal locations, as well as discrepancies in the observed deposit median grain size and the modelled best fit. These findings suggest that applying a single median grain size value uniformly across the model domain may oversimplify the deposit structure and

obscure key dispersal dynamics, e.g., Figure 5.1a. It should be noted that Tephra2 does not simulate key processes such as particle aggregation or bent-over plumes (Klawonn et al. 2014; Mannen et al. 2020), which were observed during the 17 June 1996 eruption, further limiting the realism of TGSD inputs. Such variability likely reflects eruptive processes that evolve over time and space, including particle aggregation, plume bending, and wind shear.

These constraints highlight the need for VATDMs to evolve toward more dynamic frameworks that accommodate heterogeneous particle release and changing eruptive conditions. Incorporating temporally and spatially variable TGSD inputs may improve forecast accuracy and better reflect the complexity of real-world deposits.

These findings challenge the traditional use of a single, eruption-wide TGSD as a representative input. The strong spatial variability observed within the 17 June 1996 Mount Ruapehu deposit suggests that TGSD is not a fixed property of an eruption but the result of multiple interacting processes that vary through space and time. Consequently, the concept of a single TGSD may be inappropriate for forecasting purposes. Future work should therefore focus on developing parameterisations that capture the variability and evolution of grain-size distributions throughout an eruption, rather than reducing them to a static value.

Chapter 6

Discussion

This chapter presents a summary of the body of work this PhD represents. It is drawn from the research, discussions, and conclusions presented in Chapters One through Five. This chapter is structured by first presenting a thesis synopsis, followed by a research synthesis, and then by a discussion of research limitations and recommendations for future research.

This thesis highlights the considerable challenges of short-term forecasting and the progress still needed to produce reliable, accurate volcanic hazard forecasts. Using volcanic ash as an exemplar, this thesis demonstrates that improving forecasts is not simply a matter of refining model structure or collecting more data. Instead, it requires a deeper understanding of how operational constraints shape the balance between model complexity and accuracy (Chapter Two), how input parameters influence model outputs (Chapter Three), and how prior information can bias forecasts (Chapter Four).

6.1 Thesis Synopsis

This section provides an overview of each thesis research chapter and its contributions to short-term volcanic hazard forecasting.

This thesis aims to understand the factors that influence the accuracy of short-term volcanic hazard forecasts by examining how model choice, input parameters, and prior knowledge shape forecast accuracy. This was achieved by:

1. Collating and elucidating information about the structure, testing, ontology, and complexity of volcanic hazard models, bringing clarity to a previously unsynthesised and

conceptually fragmented field (Chapter Two);

2. Directly improving ash deposition forecasting through identifying the key input parameters that influence model output variance (Chapter Three);

3. Challenging the assumption that increased knowledge about a volcano's past behaviour improves volcanic hazard forecast accuracy, and demonstrating how such assumptions may bias future forecasts (Chapter Four); and

4. Assessing the accuracy and limitations of modelled grain-size forecasts, demonstrating how TGSD parameterisation and plume shape influence spatial patterns in simulated fine ash deposition (Chapter Five).

The literature review in Chapter One identified several research gaps, including: the lack of conceptual clarity around complexity, structure, testing, and guidance for selecting appropriate volcanic hazard models; the scarcity of GSA applied to volcanic hazard models and therefore limited understanding of which input parameters most strongly influence ash deposition forecasting; and the assumption that volcano-specific knowledge reliably improves forecast accuracy.

This thesis addressed these gaps through three research objectives:

6.1.1 Objective One: Does Model Choice Matter for Volcanic Hazard Forecasting Under Uncertainty?

This objective was achieved by undertaking a literature review of the current state of volcanic hazard modelling, addressing a knowledge gap by providing clear definitions for model complexity, model structure, model testing, and model selection. Before this work, no framework existed to make sense of how volcanic hazard models differ, how they are constructed, how uncertainty propagates within them, or how complexity should be interpreted. Because of this, I developed an ontology of volcanic hazard modelling, which defined model complexity, model structure, model testing, and key gaps (e.g., the lack of GSA and inconsistent testing practices across different volcanic hazard models), and used this synthesis to provide practical guidance for selecting appropriate model types under different levels of uncertainty. Through this work, I found that:

Model choice directly affects forecast accuracy under high uncertainty.

Different volcanic hazard models represent the same underlying physical relationships in distinct ways. They may be deterministic or probabilistic, empirical or numerical, and one- or multi-dimensional, all of which influence how uncertainty is propagated (Chapter Two). When input parameter uncertainty is high, as it often is before an eruption, the balance between physical realism, computational cost, and the way uncertainty is handled becomes more important (Chapter Two). In some cases, high uncertainty can outweigh the benefits of increased complexity because complex models amplify uncertainty further due to poorly constrained input parameters. As a result, the most suitable model is not necessarily the most complex, but the one that is transparent about its assumptions and best matched to available data and the decision context (e.g., Akaike's Information Criterion, Chapter Two). This is important for operational forecasting because it means that investing resources in better data (e.g., monitoring) will be more valuable than investing in more complex code.

For the first time, my thesis establishes an ontology showing that short-term forecasting requires explicit consideration of who will use the model, why it is being used, when and where it will be applied, and whether the necessary data exist to support it (Chapter Two). This means moving away from selecting models simply because they are readily available, and instead evaluating how model structure, complexity, and data availability affect forecast accuracy. It also requires recognising that adding more input parameters does not inherently improve forecast accuracy: model structure and the availability of operationally measurable inputs all shape the quality of the final forecast. As shown in Chapter Three, the GSA identified that only a subset of input parameters meaningfully control forecast variance. This finding underscores that additional model complexity provides little benefit for accuracy, because the low-influence input parameters it adds contribute minimal information to the forecast while potentially amplifying epistemic uncertainty. Additionally, whenever the model structure is modified or new inputs are introduced, a new GSA is required to determine how these changes influence forecast accuracy.

Model Class and Computational Complexity

Model complexity influences how volcanic processes are represented and how uncertainty propagates (Chapter Two). More complex models can represent volcanic processes in greater detail by incorporating interactions such as 2D versus 3D dynamics or more realistic flow rheologies (Chapter Two) (e.g., Sparks and Aspinall 2004), which may either improve or hinder forecast accuracy depending on data quality and uncertainty (Chapter Four). Increased computational demand and higher data requirements, as well as the inclusion of input parameters that cannot be feasibly estimated, can limit the usefulness for short-term or near-real-time forecasting. In situations where both aleatoric and epistemic uncertainty are high, simpler models may provide more reliable decision support because they minimise the potential for overfitting highly uncertain data. Although advances in high-performance computing and the development of tools such as emulators (e.g., Bayarri et al. 2015) now make it feasible to run more complex models in near-real-time contexts (e.g., Fall3D for operational forecasting, Folch, Mingari, and Prata 2022), greater model complexity does not necessarily translate into more accurate forecasts. This underscores that model suitability, rather than complexity, governs forecast performance. For example, in Chapters Four and Five, Tephra2 (a simplified, 2D model) was unable to accurately forecast a bent-over plume under strong wind conditions, as the model does not account for bent-over plume geometries and aggregation, illustrating how model class can constrain forecast accuracy.

Input Parameter and Input Data Complexity

Crucially, I found that the most influential input parameters for ash deposition forecasts are the ones for which reliable data are hardest to obtain or are non-measurable (Chapter Four). For example, if a model is highly sensitive to TGSD inputs but reliable TGSD data are unavailable, then small errors or assumptions in those inputs propagate directly into the model results, producing large variability in outputs. Therefore, models that contain these types of input parameters may not be suitable for operational use or may need to be run across such a broad range that their forecasts become too uncertain to be actionable. This is not necessarily a flaw in the model itself, but a reflection on the limitations in available data and knowledge. The goal of forecasting is not to eliminate uncertainty (an outcome that is neither realistic nor achievable),

but to characterise it accurately (acknowledging what is not known) and then express it in terms of exceedance probabilities and confidence levels that are meaningful to decision-makers and communities (e.g., volcanic ash advisories for aviation, ICAO 2023).

Even sampling input parameters across wide ranges may not produce realistic forecast bounds if the underlying estimates are poorly constrained (Chapters Four and Five). Increased model complexity cannot compensate for uncertain or unmeasurable input parameters; instead, it amplifies epistemic uncertainty, reducing the practical utility of forecasts. Input parameters that represent real, observable phenomena (e.g., plume height, MER) support more transparent model interpretation and better uncertainty quantification, since they can be directly measured in practice. However, even these seemingly straightforward input parameters, such as plume height, can vary widely depending on the measurement method, timing, and atmospheric conditions. As shown in Table 4.1, plume height estimates for the 17 June 1996 Mount Ruapehu eruption span several kilometres (Chapter Four). This illustrates how past data shape both model structure and input data complexity, yet remain unreliable (Chapter Four). Models selected based on their performance in past eruptions may therefore underperform in future events, as poorly constrained input parameters amplify uncertainty rather than reduce it.

In contrast, tuning input parameters that do not clearly correspond to physical processes (e.g., the FTT in Tephra2) can yield better hindcast fits, but can also obscure model behaviour (Chapter Three). For example, reported values for FTT in the literature range widely, from 50 to 200,000 (Johnston et al. 2012; Fernandez-Turiel et al. 2019), with 5000 being the most common (e.g., Johnston et al. 2012; Biass et al. 2016b; Tennant et al. 2021; Osman et al. 2024). This combination of a large, unconstricted range and its low sensitivity (Chapter Three) suggests that such a parameter may be unnecessary in VATDM. More broadly, this thesis highlights the concerns of over-reliance on input parameters of limited sensitivity, particularly when they are used to artificially "tune" hindcasts to fit observed deposits. Such input parameters should therefore be used sparingly and clearly documented, especially in operational forecasting.

6.1.2 Objective Two: Which Input Parameters Most Influence Volcanic Hazard Model Outputs?

This objective addressed the knowledge gap regarding which input parameters most strongly influence volcanic ash deposition forecasts under uncertain conditions. This objective was achieved by conducting a GSA (Sobol' Indices and eFAST) for two VATDM Tephra2 and Fall3D. This work found that:

Only a subset of input parameters - plume shape, median grain size, plume duration, diffusion, and erupted mass - meaningfully contribute to output variance.

Chapter Three demonstrated that although VATDMs include many input parameters (e.g., Appendix A), only a subset meaningfully contribute to output variance. In Tephra2 and Fall3D, input parameters such as plume shape, median grain size, plume duration, diffusion, and erupted mass contributed the most to output variance, while input parameters such as particle density and plume height were consistently less influential. This partially aligns with, and significantly extends, previous GSA work on VATDM, which had only explored limited eruptive conditions (1996 Ruapehu, 1997 and 2021 Mount Etna) (Scollo et al. 2008; Pardini et al. 2022). By exploring the full uncertainty space rather than eruption-specific scenarios, Chapter Three provides a more complete and generalised assessment of input parameter influence across all plausible conditions.

These findings have important implications for forecasting during volcanic crises. When time and computational capacity are limited, varying only the most influential input parameters allows for more efficient probabilistic forecasts. For instance, if an eruption is expected to occur within the next 24 hours at a particular volcano, and only 1000 simulations can be run, a forecaster could prioritise varying plume shape, median grain size, diffusion, and erupted mass, rather than varying all input parameters. Under strong wind conditions, deposition becomes primarily controlled by plume shape and diffusion (Chapter Three), which further reduces the number of influential input parameters that need to be varied. Focusing on the most relevant input parameters improves transparency, interpretability, and practical usability in time-sensitive situations.

Despite the robustness of the main findings, practical considerations limited the full characterisation of all interactions. My GSA incorporated two dependent relationships - one between plume height and erupted mass, and another between FTT and diffusion. The latter dependence was necessary due to a structural (model class) error in Tephra2, whereby certain combinations of FTT and the diffusion coefficient would produce unrealistic diffusion patterns (i.e., large particles diffuse more than smaller particles). However, because the sampled values for both pairs of input parameters spanned wide ranges, these dependencies are unlikely to have materially influenced the results; testing them independently produced the same conclusions.

Additionally, only first- and total-order indices were explored in Chapter Three (Sobol' indices), as numerous negative second- and third-order indices indicated insufficient simulations, constrained by time and computational resources. Within these limitations, total-order indices provided the most defensible and informative basis for assessing overall input parameter influence. These findings also highlight that GSA is best conducted outside of crisis time, as even with substantial computational resources (NESI, Aotearoa New Zealand's supercomputer network), a full understanding of all input parameter interactions remains challenging and is time- and resource-heavy. Future work should therefore target these higher-order interactions through more computationally intensive designs, such as increasing the number of simulations or adopting emulators (e.g., Harvey et al. 2018; Salter, Webster, and Saint 2024), to better resolve subtle but potentially important dependencies among input parameters.

The wider implications of these results extend beyond Tephra2 and Fall3D. Similar patterns are likely to be observed in other VATDMs, particularly in Eulerian frameworks. Lagrangian and Hybrid models differ in how they simulate particle trajectories, but because they incorporate many of the same ESPs (Chapter Two), the key input parameters identified here are also expected to strongly influence deposition forecasts in those models.

Implications for Other Volcanic Hazards

Across all volcanic hazards, the dominant control of hazard footprint is typically the transport mechanism (wind, gravity, or momentum), with ESPs playing a secondary role (Chapter Two). In VATDMs, wind speed and direction are the primary controls on how far and in what direction ash is dispersed (Bursik 2001; Bonadonna and Costa 2013; Mulena et al. 2016). My GSA results

in Chapter Three demonstrated that different input parameters become influential under strong versus weak wind regimes (Section 6.1.2), highlighting that the relative importance of ESPs shifts in response to atmospheric conditions.

For other volcanic hazards, particularly flow models, topography (and therefore gravity) is the main driver of hazard propagation (Branney and Kokelaar 2002; Sulpizio and Dellino 2008; Dietterich et al. 2015) (Chapter Two, Appendix A). In these cases, ESPs such as erupted mass or viscosity are secondary to topography, because slope and terrain exert strong control on travel distance (Chapter Two). The one GSA applied to a (lava) flow model, MAGFLOW, for the 2006 Mount Etna eruption, identified water content and solidus temperature (i.e., viscosity) as the most influential input parameters (Bilotta et al. 2012). However, this study did not explore variability across DEMs of multiple volcanoes, leaving open questions about how topography interacts with ESPs to shape forecast outcomes - such as the effects of slope angle, surface roughness, valley confinement, and resolution of digital terrain/surface models - affect forecast accuracy and modulate the influence of ESPs. These patterns suggest that the influence of input parameters on volcanic hazard forecasts is context-dependent, emphasising the importance of GSA for each hazard type.

6.1.3 Objective Three: To What Extent Does Volcano-Specific Knowledge Improve or Limit Forecast Accuracy?

This objective addresses a commonly held, but largely untested, assumption that increased knowledge about past eruptions improves future forecasts (Poland and Anderson 2020; Tierz 2020; Martí 2024). This established assumption rests on the premise that eruptive histories are sufficiently representative of future behaviour, thereby biasing forecasts in a direction assumed to be informative. This presumes that volcanic systems exhibit some degree of stationary or cyclic recurrence, yet in reality, volcanoes evolve through time, meaning that even a near-complete eruptive record may not fully capture future variability (Mead and Magill 2014). This is despite the fact that eruptive processes can change over time, with eruption style, intensity, and frequency subject to variability (Sparks and Aspinall 2004; Bebbington and Jenkins 2019). Forecasts based on historical data may reinforce patterns that no longer reflect current system dynamics, thereby undermining forecast accuracy (e.g., see Vesuvius for an example of cyclic

behaviour, Scandone, Cashman, and Malone 2007). Through retrospective forecast analysis of the 17 June 1996, 3 March 2015, and 28 February 2021 eruptions of Mount Ruapehu, Villarrica (Chile), and Mount Etna, I found that:

Historical knowledge does not reliably improve forecast accuracy and may bias outputs.

This research demonstrated that increased knowledge about a volcano's eruptive history or composition does not always lead to more accurate tephra deposition forecasts. In some cases, it may even bias outputs away from observed outcomes (Chapters Four and Five). While some forecasts performed better than the uninformed baseline (Ruapehu - IVESPA sampled distributions), others performed significantly worse (Ruapehu - Deligne (2021) sampled distributions) (Chapter Four). Interestingly, the Mount Etna example shows that the IVESPA Mafic sampled distribution, which included up to 83% of Mount Etna eruptions within the sampling space, produced more successful forecasts than the uninformed baseline. In contrast, forecasts based only on previous Mount Etna eruptions (Deligne (2021) M1 sampled distribution) still performed better than the uninformed baseline, but were less accurate than the IVESPA Mafic sampling distribution. This demonstrates that more knowledge does not automatically translate to greater accuracy. Instead, the relevance, representativeness, and quality of the data are key. This provides substantial evidence that lower uncertainty is not inherently desirable and may reflect misplaced confidence in input parameters that are not fully representative of future behaviour. Conversely, higher uncertainty may provide a more realistic depiction of the limits of available data and knowledge, even though it may seem less precise, producing forecasts that are more transparent and accurate under genuinely uncertain conditions.

These findings also highlight the importance of evaluating model performance with metrics that reflect real decision needs, not just deposit thickness (Chapter Two). Mass-based forecast accuracy was often low (Chapter Four), and millimetre-level thresholds are highly sensitive to small differences in key input parameters (Chapter Three), meaning that thickness-based accuracy alone can give a misleading impression of model performance. By incorporating impact-based metrics, this research provided a more realistic and operationally relevant picture of forecast accuracy. Impact-focused assessments better capture the consequences of over- or

under-prediction and align more closely with how forecasts are used by emergency managers and infrastructure planners in Aotearoa New Zealand (Doyle et al. 2014; Das et al. 2025).

This thesis highlights a critical and continuing challenge: although past eruptions provide valuable information, the inherent variability and evolving behaviour of volcanic systems limit the extent to which historical information can reliably inform forecasts (Bebbington and Jenkins 2019). In one example, forecasts based solely on previous eruptions at the same volcano (e.g., Mount Etna) outperformed the uninformed baseline. However, incorporating data from analogue volcanoes also produced strong forecast performance, in some cases exceeding that of a single-volcano approach. This suggests that expanding datasets can improve forecast accuracy, provided the additional data are sufficiently representative of the target system. Because a single example from Mount Etna is insufficient to determine a general rule, it remains uncertain which approach consistently yields better forecast accuracy. However, identifying what constitutes 'representative' data remains challenging. Similar patterns have been observed for eruption forecasting more broadly, where limiting analogue datasets based on volcano morphology or composition are not significantly more informative than using the full dataset (Bebbington and Jenkins 2022). These observations underscore the need for further testing and research, particularly as additional eruptions are observed and higher-quality, volcano-specific data become available.

Beyond the quantity and quality of past data, my findings also identify a mismatch between input parameters that are most often constrained by past data and those that exert the greatest influence on model outputs (Chapter Three). Input parameters such as plume height are frequently documented (e.g., Table 4.1, Chapter Four), but have proved to be relatively unimportant for output variance (Chapter Three), while input parameters such as plume shape and diffusion, which strongly influence deposition forecasts, remain poorly constrained historically (Chapters Four and Five). This imbalance shows how reliance on past records can shape input parameter choices in ways that do not necessarily improve forecast accuracy (Chapter Two).

This is further complicated by the use of empirical relationships, such as those used in Chapter Three, to estimate plume height and erupted mass (Mastin et al. 2009), which assume that past eruptive behaviour from other volcanoes is transferable to future scenarios at your volcano of interest. For the GSA and the uninformed sampling distribution in Chapters Four

and Five, I deliberately adopted the empirical relationship from Mastin et al. (2009). However, an important finding was that in the most accurate forecasts in Chapter Four (from Mount Ruapehu, i.e., hits of 50% or more) generated using the informed sampling distributions (where no explicit relationship between height and mass was imposed), plume height and erupted mass were effectively uncorrelated (r = -0.06, R^2 = 0). This stands in sharp contrast to both the Mastin et al. (2009) relationship (r = 0.77) and the IVESPA relationship (r = 0.54). These results highlight that relying on average relationships from past eruptions may misrepresent future behaviour, because assumed relationships can fail to capture the actual variability of key input parameters at a given volcano, potentially leading to biased or less accurate forecasts. This reinforces the need for forecasting approaches that prioritise constraining the plausible range of input parameters.

6.2 Thesis Synthesis

Despite major methodological advances, the driver of forecast accuracy in volcanic hazard models has remained unclear until now. We care about forecast accuracy because incorrect volcanic hazard forecasts can lead to inappropriate advice, unnecessary disruption, or damaging underestimation of risk. Volcanic hazard forecasting has made major advances in quantifying and communicating uncertainty through probabilistic forecasting frameworks (e.g., Osman et al. 2024; Sandri, Tierz, and Loughlin 2025), improved understanding of ESP processes (e.g., Bonadonna, Biass, and Costa 2015; Engwell et al. 2024; Pardini et al. 2024), and the increasing adoption of ensemble-based approaches (e.g., Capponi et al. 2022; Folch, Mingari, and Prata 2022; Hyman et al. 2024). These developments have improved how forecasts express uncertainty. Still, before this thesis, there had remained a lack of understanding of what governs forecast accuracy in volcanic hazard models.

Compared with other environmental modelling areas, such as ecology and hydrology, which have well-established literature around GSA (e.g., Pianosi et al. 2016; Pianosi and Wagener 2016; Malchow and Hartig 2024) and model evaluation (e.g., Butts et al. 2004; Larsen et al. 2016; Getz et al. 2018), volcanic hazard models had previously lacked a coherent ontology of model structure and complexity, had undergone minimal GSA, and had never been systematically tested to determine whether past eruptive behaviour reliably constrains short-term forecasting.

This absence of foundational understanding meant that it was difficult to assess whether volcanic hazard models were sufficiently accurate, appropriately complex, or even justified for operational use. As a result, advances in model complexity often outpaced the ability to determine whether adding more processes, input parameters, or computational detail genuinely improved forecast accuracy.

Although it is often assumed that more complex models produce more accurate forecasts (Robinson 2023), evidence from hydrology modelling shows that added complexity can worsen accuracy by overfitting noisy or poorly constrained inputs, whereas models grounded in reliable data perform better (e.g., Schoups, van de Giesen, and Savenije 2008; Finger et al. 2015; Orth et al. 2015). This thesis demonstrates that the same principle holds for volcanic hazard models: forecast accuracy is limited not by physics or computational complexity, but by the quality, reliability, and measurability of the data used to drive key input parameters.

Using volcanic ash deposition as an exemplar, this thesis identified three factors that control the accuracy of short-term volcanic ash forecasts: (1) the structure and complexity of the volcanic hazard model, (2) the sensitivity and influence of key input parameters, and (3) the extent and reliability of the knowledge and data used to constrain those input parameters.

My findings show that increasing model complexity, whether by adding more input parameters or adopting more computationally demanding formulations, has limited benefit for improving short-term forecast accuracy. This is because variance is controlled primarily by a small set of influential input parameters (Chapter Three), and when the plausible range of these input parameters is wide or weakly informed by past data, they dominate forecast accuracy and limit the benefit of any new gains from added physics or detail to the model structure (Chapter Two). Together, these results demonstrate a clear bias-variance trade-off: added complexity can reduce structural bias by correcting errors introduced by overly simple model assumptions (e.g., Tephra2 in Chapters Four and Five), but only when the influential input parameters are measurable and well constrained. Attempts to narrow input parameter ranges using past eruptive behaviour reduced variance but often increased bias by forcing forecasts toward patterns found in historical datasets that did not fully represent the range of plausible future behaviour (Chapters Four and Five).

Historical eruptive behaviour helps shape volcanic hazard model development, but it does

not reliably constrain the key input parameters that control forecast variance. Consequently, relying on past behaviour can mislead forecasts and degrade accuracy. In practice, only a few ESPs - plume height, MER, and erupted mass - are commonly estimated in real time, and even these measurements are unavailable for many volcanoes and eruptions (Engwell et al. 2024). Plume height, in particular, is widely assumed to be one of the most important ESPs simply because it is required to initialise every VATDM and can often be measured, which creates a false confidence in model outputs (Engwell et al. 2024). However, the results of this thesis show that, although plume height is operationally essential (e.g., for flight path divergence), it is among the least influential input parameters for ash deposition forecasts. When influential input parameters cannot be constrained, increasing model complexity simply expands the volume of unknowns and amplifies epistemic uncertainty.

By identifying the most influential input parameters and assessing the reliability of past knowledge, this thesis demonstrates how decisions around model choice (which itself is influenced by complexity), the sensitivity of key inputs, and the constraints imposed by volcano-specific data knowledge together determine forecast accuracy. These findings demonstrate that prioritising the measurement and monitoring of volcanic hazards to accurately estimate key input parameters improves forecast accuracy far more than building increasingly complex models.

6.2.1 Implications for Ensemble Forecasting

This thesis focused on understanding factors that control forecast accuracy in individual volcanic hazard models. Ensemble modelling can complement this approach by providing a framework to capture epidemic uncertainty; however, its effectiveness depends on how well the ensemble represents the plausible range of eruptive conditions. Ensemble modelling is often presented as a solution for capturing epistemic uncertainty (Folch 2012; Poland and Anderson 2020; Folch, Mingari, and Prata 2022). However, ensembles alone do not address the full range of variability in volcanic hazards, because they can only sample from the input parameters provided. When key ESPs are poorly constrained, unknown, or misrepresented, even a large ensemble may fail to capture some plausible eruptive conditions. Therefore, the utility and accuracy of an ensemble depend not only on how many simulations are run, but on whether the ensemble members meaningfully capture the plausible range of eruptive conditions and account for the dominant

sources of forecast variance (Chapters Three and Four).

In addition, ensemble accuracy depends on selecting models that are structurally appropriate for the forecast problem. Ensemble forecasts cannot reach their full potential accuracy if the underlying models or the data used to drive them are poorly suited to the eruption context (Chapter Two). In volcanology, no studies have examined how model structure or complexity influences ensemble accuracy, even though we already know that different volcanic hazard models - each with distinct model structures - can produce different levels of output accuracy when given identical inputs (Scollo, Folch, and Costa 2008; Bonadonna et al. 2011a; Bonadonna et al. 2012; Cordonnier, Lev, and Garel 2015; Dietterich et al. 2017; Gueugneau et al. 2021). What remains unknown is how these structural differences propagate through ensemble forecasts and affect ensemble accuracy. My research begins to address this gap, but further work is needed to understand how different model structures and their key input parameters influence ensemble forecasts. In hydrology, this issue has been examined (Butts et al. 2004; Krueger et al. 2010; Spieler and Schütze 2024), demonstrating that model structure is a critical determinant of ensemble accuracy, and the same principles likely apply to volcanic hazard models.

6.3 Limitations

6.3.1 Data Constraints

A significant limitation in volcanic hazard modelling arises from the availability, quality, and usefulness of ESP data, as volcanoes do not erupt frequently, and their eruptive histories extend far beyond the timescales of instrumental measurement or human observation (e.g., the world's first volcano observatory was constructed in 1841 on the flank of Vesuvius, Tilling 2022). To address this, Mastin et al. (2009) developed a global compilation of estimated input parameters, including plume height, eruption duration, MER, erupted volume, and TGSD for 1535 volcanoes worldwide. These values are not derived from specific eruptions but are generalised estimates based on volcano type, magma composition, and eruptive history, meaning the database is most appropriate for simulating hypothetical scenarios or volcanoes lacking observational data.

The limitations of generalised estimates are evident at Mount Ruapehu. For example, Mastin et al. (2009) classify Mount Ruapehu as S1, indicating historic eruptions less than VEI 2. However, the estimated eruptive volume is 0.003 km^3 (around VEI 1), whereas historical eruptions have ranged up to VEI 5 (between 0.3 - 0.6 km^3) (Donoghue et al. 1995; Pardo et al. 2012). This discrepancy highlights that the database reflects average behaviour rather than the full distribution of eruptive variability, and can therefore underestimate the hazard potential of individual volcanoes.

The Global Volcanism Program (GVP) Holocene database provides the most comprehensive global record of eruption histories, including eruption duration and VEI for thousands of eruptions. However, while useful for general assessments of eruption frequency and magnitude, it lacks the full suite of ESPs required for quantitative hazard modelling (e.g., Table 4.1). Furthermore, the GVP's definition of an eruption - where events separated by three months or less are classified as a single eruption - may not realistically represent the episodic nature of many volcanic systems. This is problematic for hazard modelling because episodic activity can produce multiple discrete ash-producing events within a single eruptive episode, meaning that treating them as one eruption obscures the temporal variability that drives hazard exposure.

In contrast, more recent databases such as LaMEVE (Large Magnitude Explosive Volcanic Eruptions) (Crosweller et al. 2012) and IVESPA (Aubry et al. 2021) provide eruption-specific

ESPs. The LaMEVE database contains ESPs, primarily deposit volumes, for over 1800 Quaternary large-magnitude eruptions (VEI 4 or greater) across 480 volcanoes, while IVESPA includes detailed records for 134 eruptions across 45 volcanoes. IVESPA, in particular, offers a wider breadth of ESPs, including plume height, mass, MER, TGSD, and atmospheric conditions, with associated data uncertainties explicitly reported. These uncertainties reflect both the variability in observational data and the methodological constraints inherent in specific parameter estimation. For example, MER cannot be directly observed and is unlikely to remain constant throughout an eruption (Sparks 1986; Degruyter and Bonadonna 2012). Its estimation typically relies on one-dimensional plume modelling (e.g., Mastin 2014), empirical relationships between MER and plume height (e.g., Sparks 1986; Mastin et al. 2009; Aubry et al. 2022), or analytical expressions that incorporate eruption dynamics and atmospheric conditions (e.g., Degruyter and Bonadonna 2012). These approaches involve assumptions about steady-state flow, entrainment rates, and plume buoyancy, which all introduce aleatoric uncertainty into the derived MER values. Although satellite, infrasound, and radar sensors have been used to estimate MER in near-real time (e.g., Freret-Lorgeril et al. 2021; Mereu et al. 2023; Michaud-Dubuy and Gouhier 2025), such methods remain limited to a small number of well-monitored volcanoes (e.g., Mount Etna) and are not yet operationally widespread.

In Chapter Four, we found that the choice of database directly influenced forecast accuracy. Input parameter ranges informed by the IVESPA database produced more accurate forecasts than those based on uninformed ranges, whereas the Deligne (2021) database reduced forecast accuracy. This contrast likely reflects differences in the underlying data: IVESPA is based solely on observed eruption records, whereas Deligne (2021) includes older geologic data. The results raise a broader question explored throughout the thesis about the reliability of past eruptive data. The stronger performance of IVESPA suggests that recent, well-documented eruptions can improve forecast accuracy - not because recent activity is known to be more representative (indeed, this remains uncertain (Marzocchi and Bebbington 2012)) - but because modern measurements provide more accurate and complete estimates of key input parameters (Mead and Magill 2014). Thus, the apparent benefit stems from data quality rather than any inherent predictiveness of recent eruptive behaviour.

Uncertainty in ESP Parameterisation

Beyond data limitations, epistemic uncertainty also arises from how models define and use ESPs. While each model specifies a core set of required input parameters, choices made by the user, such as which parameters to vary, simplify, or hold constant, can substantially affect model behaviour and forecast accuracy (e.g., the difference in best-fit values of median grain size depending on sampling location, Chapter Four). While IVESPA mitigates some of these issues through peer-reviewed sourcing and dual volcanologist quality control, the inherent limitations in both data and parameterisation must still be critically evaluated when applying these datasets to hazard modelling, as they describe how volcanoes have behaved in the past rather than how they will behave in the future.

Biases in ESP Database Coverage

Despite the improvements offered by databases such as IVESPA, the number of eruptions they cover remains small relative to the global eruption record. As of August 2025, the GVP reports 9902 documented Holocene eruptions, 3883 of which have occurred since 1900 (Global Volcanism Program 2025a). Using the IVESPA database as an example (as it is the database with the largest range of ESPs), only 70 unique eruptions between 1902 and 2016 are included, compared to 3,619 eruptions listed in GVP over the same period. This represents just 1.9% of recorded eruptions. These 70 eruptions are also skewed towards more recent events (average year 1995, median 2000). While the volcanoes are distributed relatively evenly across regions of volcanic activity (Figure 6.1), South and Central America account for the highest number of individual volcanoes (with 11 and 7, respectively). The volcanoes with the greatest number of eruptions in the IVESPA database are Mount Etna (Europe, 6 eruptions), Bezymianny (Russia, 5), Cerro Negro (Central America, 5), and Hekla (Europe, 5) (Figure 6.1). The proportion of eruptions with high-quality, eruption-specific ESPs is therefore extremely small and spatio-temporally biased. For example, volcanoes located in remote or logistically challenging regions, such as parts of Russia, Antarctica and Africa, remain poorly represented due to limited accessibility, monitoring infrastructure, and sampling opportunities.

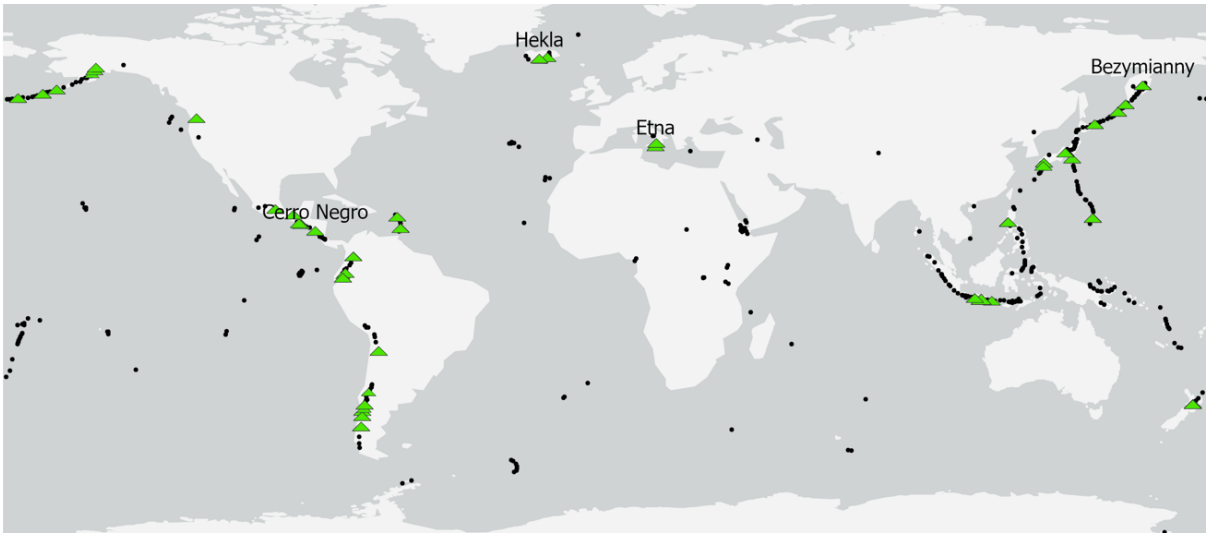


Figure 6.1: Global distribution of volcanoes from the IVESPA database (green triangles) (Aubry et al. 2021). Black dots show volcanoes that have erupted within the same temporal window as the IVESPA dataset, according to the Global Volcanism Program.

However, a comprehensive, standardised ESP database, particularly one that spans multiple volcanic hazards and supports operational modelling, remains unavailable. Thus, initiatives such as the GVP are expanding their coverage of eruption data; these resources are not yet integrated into a unified, centralised framework suitable for hazard modelling. ESP information often remains fragmented across literature, observatory reports, and database entries, requiring manual synthesis and interpretation (Aubry et al. 2021). Platforms such as GVP are valuable for tracking eruption chronology, but less effective for ESPs, highlighting a persistent gap in accessible centralised information. High-quality ESP estimations also require consistent and continuous monitoring during eruptions, detailed post-eruption fieldwork, and synthesis of diverse datasets, which can be incomplete, unpublished, or inaccessible (Aubry et al. 2021). Furthermore, standardising input parameter definitions (i.e., whether plume height is above vent or sea level) and quantifying uncertainties remain substantial barriers.

This thesis showed that if ESPs such as TGSD could be obtained during an eruption, they would enable substantial improvements in hazard forecasting (Chapters Three, Four and Five). If a complete database of ESPs were available, defined here as one containing temporally detailed records of ESPs and associated atmospheric and ground conditions across multiple eruptions and volcanoes, it would substantially enhance model testing and calibration (Chapter Two). Such a database would enable more realistic input distributions, clearer characterisation

of input parameter variability, and more confident identification of dominant controls under different eruption scenarios (e.g., the influence of plume shape under strong wind conditions, as discussed in Chapters Four and Five). It would also allow for more thorough testing of informed versus uninformed input parameter distributions (Chapter Four), potentially improving forecast accuracy and reducing reliance on assumed distributions. Beyond calibration, such a resource could also inform the development of new modelling approaches. For example, temporally and spatially resolved TGSDs could allow models to incorporate dynamic grain-size distributions that vary over the course of an eruption, instead of assuming a fixed value (Chapter Four).

While a complete database would not eliminate uncertainty, it would allow it to be quantified and communicated more transparently. This, in turn, would support more robust decision-making under operational contexts, particularly where rapid forecasting is required, but input data are limited or evolving. However, developing such a database is currently infeasible. Many ESPs (such as TGSD) cannot be directly measured during eruptions and are inferred from deposits, which often only represent a portion of an event. Capturing the temporal evolution of ESPs such as plume height, MER, or TGSD across multiple eruptions and volcanoes would require continuous, high-resolution monitoring sustained over many decades, which is an unrealistic prospect given that most volcanoes erupt infrequently and observation capabilities remain limited, especially in under-monitored regions such as Africa and Indonesia (Aspinall et al. 2011). In fact, up to 45% of historically active volcanoes are currently unmonitored (Brown et al. 2015). Although advances in satellite imagery are beginning to ease these constraints by providing global, near-real-time coverage of eruptive activity (Poland et al. 2020).

6.3.2 Empirical Relationships

Uncertainty in volcanic hazard modelling arises not only from limited data but also from the interdependence of input parameters. As discussed in Chapters Three and Four, sampling techniques needed to be constrained as some input parameters are dependent on each other. Random sampling without constraints can generate unrealistic combinations (for example, a very large plume height with a negligible erupted mass). To reduce such artefacts, empirical relationships were used to constrain input parameter values to ranges consistent with real-world eruption behaviour. Several empirical relationships link erupted mass (or MER) to plume height

in explosive eruptions (e.g., Carey and Sigurdsson 1989; Wilson et al. 1978; Mastin et al. 2009; Aubry et al. 2022), generally showing that higher plumes are associated with larger erupted masses, reflecting the intuitive link between eruption intensity and plume height. For the GSA in Chapter Three and uninformed sampling from Chapters Four and Five, the plume height-mass relationship used was derived from Mastin et al. (2009). This linear model was fitted to data from 29 volcanic eruptions (representing only 0.97% of eruptions compared to GVP eruptions in the same period) and, despite the relatively small dataset, remains the last published direct relationship between plume height and erupted volume (and therefore mass). To account for variability in eruption conditions and the limitations of the original dataset, an additive term $Z \sim N(0, 2)$ was included in the plume height-mass relationship for the GSA analysis. This adjustment allowed the GSA to explore a more realistic range of eruption conditions, while preserving the underlying positive relationship between plume height and erupted mass.

Although an alternative fit was explored using IVESPA data, it was not used in my analyses within Chapters Three, Four, and Five. While the IVESPA database contains more eruption records overall, the subset with both well-documented plume heights and associated mass values is still relatively small (132 eruptions). This limited overlap means that an IVESPA-derived relationship would have been based on a similarly narrow dataset. In my analyses, using the Mastin et al. (2009) relationship restricted the mass input parameter space to the minimum and maximum values present in the dataset - a constraint that would not have been significantly alleviated by using the IVESPA database, and does not affect the results within Chapters Three, Four, or Five.

This limitation reflects the broader philosophical challenge in Chapter Two: the difference between numerical and empirical modelling approaches, particularly in terms of model class, input parameter, and input data complexity. If reliable data exist only for a limited set of input parameters (e.g., plume height), then building numerical models that heavily rely on additional uncertain inputs or unmeasurable inputs (e.g., diffusion) risks introducing more noise than signal. If reliable data are abundant, then empirical models constrained by well-supported input parameters may offer more robust and interpretable outputs than complex models populated with poorly constrained variables. This trade-off is especially relevant when designing ensemble simulations, where the inclusion of uncertain input parameters, alongside more certain ones,

can obscure dominant controls and dilute forecast clarity by reducing the practical utility of the model for decision-making or hazard communication.

While this work has advanced understanding of how input parameter sensitivity, volcano-specific knowledge, and model choice interact to shape forecast accuracy, fundamental limitations remain. Data constraints, incomplete eruption records, and the inherent variability of volcanic systems mean that forecasts will always carry epistemic uncertainty. Consequently, even well-parameterised models cannot guarantee predictive accuracy, but they can provide probabilistic insights that are meaningful for decision-makers. Importantly, prioritising influential, measurable input parameters is applicable beyond volcanic ashfall, informing the development of robust forecasting approaches for other volcanic and natural hazards where rapid, evidence-based decisions are required.

6.4 Future Directions

While this thesis has gone some way toward addressing the challenges of accurate short-term volcanic hazard forecasting, it has also revealed several new avenues of research, particularly relating to TGSD, ESP data, and the application of similar approaches to other hazards.

This thesis identified TGSD, specifically median grain size, as a particularly influential VATDM input parameter and an area requiring more targeted research (Chapters Three, Four, and Five). Improvements are needed in TGSD sampling and data availability, especially during eruptions, where accurate characterisation is difficult. Fine particles remain in the atmosphere and deposit at greater distances over longer timescales. In contrast, coarse particles are often inaccessible within exclusion zones and can only be sampled after the eruption. As a result, forecasts frequently rely on poorly constrained or assumed TGSD inputs, introducing uncertainty in both mass distribution (Chapter Four) and the proportion of potentially hazardous fine ash (Chapter Five). Rather than proposing new sampling strategies, this thesis highlights the need for model frameworks that can accommodate uncertainty in input parameters such as TGSD. For example, varying input TGSD spatially may offer a more realistic and accurate mass and TGSD forecasts (Chapters Four and Five). Future work should therefore focus on refining model structures and input strategies to better reflect TGSD uncertainty, especially in operational contexts where sampling data are limited or unavailable (e.g., Costa, Pioli, and

Bonadonna 2016).

TGSD highlights how uncertainty in a single input can strongly influence forecast outcomes, but similar challenges arise across other volcanic hazard models. There is a clear need to extend GSA methodologies to other volcanic hazard models, particularly flow and VBP models, where application has so far been limited. To date, only one GSA has been conducted for a lava flow model, MAGFLOW (Bilotta et al. 2012), and no GSAs have been conducted for volcanic mass flows or VBP models. Expanding GSA to other volcanic hazard models represents a critical opportunity to improve the accuracy of volcanic hazard forecasting and model transparency. By demonstrating how GSA can be applied effectively to VATDMs, this thesis provides a practical framework that can be adapted to other model types. Given the multitude of distinct volcanic hazard models (e.g., Chapter Two, Table 2.1), scaling GSA methodologies across the entire range of volcanic hazard models will entail significant computational time and effort.

Beyond technical future work, the findings of my thesis point toward a broader conceptual challenge for future research. My finding that increased knowledge of a volcano's eruptive history or composition does not always improve forecast accuracy challenges the assumption that more information leads to better forecasts (Chapter Four). Similarly, the observation that broader, more uncertain input ranges sometimes yield more accurate forecasts than narrowly defined ones questions the view that reducing uncertainty should be the primary goal of modelling/forecasting. Instead, my findings suggest that embracing a realistic level of uncertainty, rather than overconfidence in limited data, may lead to forecasts that are more transparent and ultimately more useful for decision-making. Together, these results invite deeper reflection on what constitutes "useful" knowledge in volcanic hazard forecasting and how uncertainty can be understood, represented, and communicated more effectively.

The approaches developed and tested in this thesis offer a blueprint for systematically addressing the challenges of accurate short-term volcanic hazard forecasting. They provide a pathway toward accurate, uncertainty-informed forecasts across all volcanic hazards. These findings not only advance VATDM-based ashfall forecasting but also establish principles that can be applied to other volcanic hazard models, supporting more transparent and accurate hazard assessments and forecasts in the future.

Chapter 7

Conclusion

Given that volcanic hazards are inherently uncertain and currently unknowable before an eruption, accurate short-term forecasting requires approaches that explicitly acknowledge these limitations and optimise reliability. This thesis has examined these challenges in short-term volcanic hazard forecasting from multiple perspectives: the conceptual (Chapter Two), the computational (Chapter Three), and the practical (Chapters Four and Five).

Chapter Two introduces a generalised ontology for volcanic hazard models, helping users better understand their position within the broader modelling landscape and identify key areas for improvement in model construction, testing, and usability. This framework clarifies how models differ in theoretical foundations, computational demands, and suitability for specific forecasting tests. It also underscores the importance of selecting a model that balances complexity with usability - those detailed enough to capture essential hazard dynamics, yet simple enough to remain interpretable and robust under data constraints. Models with fewer input parameters are not only more efficient but also (generally) more transparent, especially when data availability is limited. As model complexity increases, sensitivity testing becomes more difficult, heightening the risk of compounding uncertainty through untested input interactions. The overarching goal is therefore to identify models that achieve the greatest explanatory power with the fewest and best-constrained inputs (consistent with Occam's razor).

For tephra deposition forecasting specifically, findings from Chapter Three move us closer to reliable forecasts by identifying which input parameters most strongly influence model outputs. Using GSA (Sobol' indices and eFAST), I show that for VATDMs Tephra2 and Fall3D, plume

shape, diffusion, median grain size, and plume duration (Fall3D only) have substantial first-order effects on model variance. In contrast, input parameters such as particle density and plume height exert little direct influence but become more important through higher-order interactions with other input parameters. These results show that VATDM outputs are driven by a few dominant input parameters, supporting a more targeted approach to model calibration and operational forecasting.

Chapter Four further demonstrates that current forecasting methods are likely biased due to insufficient data, even for well-studied volcanoes such as Mount Etna. Using the 17 June 1996 eruption of Mount Ruapehu as a case study, I show that forecast accuracy was poor overall, and that simulations informed by past eruption data did not systematically outperform those based on uninformed, uniform sampling of input parameters. In fact, some informed input parameter ranges performed worse, suggesting that relying solely on historical data may introduce bias without improving predictive skill. This is particularly problematic when forecasting unknown future eruptions, where the temptation to bias input parameters based on past events can lead to misleading results. The second example of Mount Etna illustrates this risk clearly: forecasts based on previous Mount Etna eruptions produced highly inaccurate outcomes, despite the volcano's extensive observational record (almost 3,500 years, Mulargia, Tinti, and Boschi 1985). These findings reinforce the need for flexible, eruption-specific approaches that account for natural variability and uncertainty, rather than overfitting to historical precedent.

Taken together, the three main studies presented in this thesis support the following major conclusions:

1. **Model complexity is a trade-off**. The ontology framework shows that, in short- and near-real-time forecasting contexts with limited data, simpler models with fewer inputs and coarser spatial or temporal resolution often provide the most reliable and actionable results. Selecting a more complex model could be considered when sufficient observational data exist to constrain its input parameters. In such cases, complexity is not inherently problematic - provided the model's structure is well understood. GSA plays a critical role in both model selection and evaluation, as it helps identify which input parameters drive output variance and clarifies whether a model's complexity is warranted for short- and near-real-time forecasting. Importantly, this analysis can be conducted during quiescence,

allowing forecasters to assess model suitability before an eruption occurs. When the structure, assumptions, and sensitivities of a complex model are known and supported by data, its use can be justified. Sensitivity testing should therefore be a routine part of model evaluation, not only to guide calibration and uncertainty quantification but also to inform model selection itself.

2. **Not all input parameters are equally important**. Sensitivity analysis reveals that only a subset of input parameters, such as plume shape, diffusion, duration, and median grain size, substantially influence model outputs. Prioritising these inputs by varying and constraining the most influential inputs, while sampling the less influential input parameters more sparsely, will allow for decreased computational demand and calibration efforts. This targeted approach improves forecast efficiency by reducing the dimensionality of the input parameter space, as fewer input parameters require intensive sampling for probabilistic forecasts or ensembles.

3. **Avoid biasing forecasts when uncertain**. Forecasts based on past eruption data do not consistently outperform those based on uninformed distributions. In some cases, they perform worse. When the future is unknown, introducing bias without justification can degrade forecast accuracy.

This thesis shows that understanding model complexity, prioritising influential input parameters, and carefully considering the limits of past data together enable more informed, reliable, and accurate short-term volcanic hazard forecasts.

Appendix A

Volcanic Hazard Models

A.1 Atmospheric Dispersion and Deposition Models

A.1.1 Model Dimensionality

The dimensionality of a model governs the number and complexity of input parameters (Cao et al. 2021). Before an eruption, the eruption source parameters (ESPs) are unknown. ESPs such as plume height, grain size, and mass eruption rate (MER) are commonly estimated from empirical expressions relating plume height to MER (i.e., a larger MER generally results in a higher plume) (e.g., Carey and Sigurdsson 1982; Mastin et al. 2009; Aubry et al. 2022), or through parameter calibration or inversion (e.g., Fee et al. 2017; Poret et al. 2017; Perttu et al. 2020). These relationships are expressed as one-dimensional (1D) models that can be used within or in conjunction with higher-dimensional volcanic ash transport and dispersion models (VATDM) (e.g., within Fall3D, Folch et al. 2020). 1D models show how plume dynamics change in response to variations in specific input parameters, rather than their ability to simulate specific eruptions.

Two-dimensional (2D) models are based on a simplified analytical solution, in the form of a Gaussian function (Macedonio, Costa, and Longo 2005), of three-dimensional (3D) advection-diffusion-sedimentation equations (ADS; the governing equation that simulates particle movement in the atmosphere, Tsunematsu et al. 2011) (Folch 2012; Bonadonna et al. 2015). The ADS equation is simplified by introducing some approximations, such as: (1) constant horizontal diffusion; (2) negligible vertical diffusion and vertical wind velocity; (3) horizontal components

of the wind field are homogeneous (i.e., steady in both XYZ); and (4) negligible particle-particle interaction (e.g., aggregation) (Pfeiffer, Costa, and Macedonio 2005; Bonadonna et al. 2015). The advantages of 2D models include their simplicity of physical parameterisation and, therefore, the fast computational speeds (Folch 2012; Bonadonna et al. 2015). However, 2D models need to be used with caution under certain circumstances, such as (1) weak plumes, as dispersion from low plumes primarily occurs in the lower troposphere, where winds are variable and vertical components may not be negligible; (2) near-source regions, as topography and plume dynamics are neglected; (3) long-range particle dispersal, as wind fields over large areas are usually not homogenous and are likely to change with time; and (4) dispersal from long-lasting eruptions, as wind conditions vary with time (Folch 2012; Bonadonna et al. 2015). Examples of 2D ADDM include HAZMAP (Macedonio, Costa, and Longo 2005; Pfeiffer, Costa, and Macedonio 2005) and Tephra2 (Bonadonna et al. 2005; Connor, Connor, and Bonadonna 2008).

Table A.1: Common volcanic atmospheric dispersion and deposition models.

Model	**Dimensionality**	**Model Type**	**Reference System**	**Hazard**
Tephra2[1]	2D	Numeric	Eulerian	Tephra
HAZMAP[2]	2D	Numeric	Eulerian	Tephra
Fall3D[3]	3D	Numeric	Eulerian	Tephra, Gas
PUFF[4]	3D	Numeric	Lagrangian	Tephra
Ash3D[5]	3D	Numeric	Eulerian	Tephra
HYSPLIT[6]	3D	Numeric	Hybrid	Tephra, Gas
NAME[7]	3D	Numeric	Lagrangian	Tephra, Gas

[1] Connor, Connor, and Bonadonna (2008), [2] Macedonio, Costa, and Longo (2005), [3] Folch, Costa, and Macedonio (2009), [4] Searcy, Dean, and Stringer (1998), [5] Schwaiger, Denlinger, and Mastin (2012), [6] Draxler and Hess (1997), [7] Jones et al. (2007)

Some of the limitations of 2D VATDM can be overcome by solving the ADS equation numerically, i.e., in 3D, to obtain time-dependent airborne particle concentrations and ground deposit thickness/load (Folch 2012). These 3D models are most suited for short-term and near-real-time forecasting (for which the meteorological data is available) and the study of past events (where ESPs were observed, or at least can be relatively well-defined from known deposits and observations) (Bonadonna et al. 2015). Commonly, 3D models are used by Volcanic Ash Advisory Centres (VAACs) to forecast the dispersal and airborne concentrations of volcanic ash clouds in quasi-real-time, mitigating aviation risk (Bonadonna et al. 2015). Due to the nature of 3D models (i.e., they require numerical simulation in three spatial dimensions), they require

considerable computing time, and may limit the user to a small number of possible simulations (Pfeiffer, Costa, and Macedonio 2005; Folch 2012; Bonadonna et al. 2015). Popular 3D models include Fall3D (Folch, Costa, and Macedonio 2009; Folch, Mingari, and Prata 2022), VOLCALPUFF (Barsotti, Neri, and Scire 2008), and Ash3D (Schwaiger, Denlinger, and Mastin 2012; Mastin et al. 2013).

A.1.2 Frame of Reference

For VATDMs, the frame of reference determines how physical processes are represented in space and time. These frames of reference can be Lagrangian, Eulerian, or Hybrid.

Lagrangian models, or Lagrangian particle tracking, are random walk models that describe the dispersion of a particle in turbulent flow and calculate the trajectories of each individual particle (Hurley 1994). These Lagrangian models follow particles around and all equations within the model then relate to each specific particle, e.g., PUFF (Tanaka 1994; Searcy, Dean, and Stringer 1998) and NAME (Jones et al. 2007; Müller et al. 2013; Turner et al. 2014). The observer, in this case, moves along with the parcel of air and describes the characteristics of its motion and location. Compared to Eulerian models, Lagrangian models are less computationally intensive and exhibit much less numerical diffusion (Schwaiger, Denlinger, and Mastin 2012). However, Lagrangian models require a large number of particles (hundreds of thousands to millions) to be released to calculate dispersal effectively (Boybeyi and Raman 1995). The more particles, the longer it takes to run a model. Operational simulations, such as those conducted by VAAC, typically employ Lagrangian-based models to provide forecasts to the aviation community of ash cloud extent and movement, normally focusing on the long-range dispersal of ash (Folch 2012).

Eulerian models are primarily advection-diffusion type models whereby the trajectory of tephra is solved through the observer's reference system, which is located on a fixed point on the reference grid, e.g., Ash3D (Schwaiger, Denlinger, and Mastin 2012; Mastin et al. 2013), Tephra2 (Bonadonna et al. 2005; Connor, Connor, and Bonadonna 2008), and Fall3D (Costa, Macedonio, and Folch 2006; Folch et al. 2020). The evolution of atmospheric particle concentrations can be calculated at every point in the reference grid. Eulerian models are better suited to particle

dispersion at larger distances from the source, as they cannot properly parameterise sub-grid-scale processes, leading to artificial diffusion (especially near the source) (Boybeyi and Raman 1995).

Hybrid VATDMs utilise both Eulerian and Lagrangian methods in model calculations (Draxler and Hess 1997; Draxler and Hess 1998; Draxler 1999; Hurst and Davis 2017). In Hybrid models, advection-diffusion calculations are made in a Lagrangian framework, while particle concentrations are calculated on a fixed grid (Eulerian) (Draxler and Hess 1997). This approach allows the model to properly capture both the horizontal dispersion and the change in wind properties with height (Tadini et al. 2020). The primary Hybrid VATDM used is HYSPLIT (Draxler and Hess 1998; Stein et al. 2015).

A.1.3 Model input Parameters

In VATDMs, there are three types of input parameter data:

1. **Meteorological Data**: provides wind velocity, wind direction, atmospheric pressure, air temperature, humidity, and atmospheric heights (levels) to models.

2. **Eruption Source Parameters (ESPs)**: including (but not limited to) eruption duration, particle density, total erupted mass (TEM), mass eruption rate (MER), plume height, and the total grain size distribution (TGSD).

3. **Simulation Conditions**: unmeasurable quantities that models use to run simulations, such as the diffusion coefficient.

Table A.2: Common input parameters in atmospheric dispersion and deposition models. 'x' indicates that the model uses the input parameter; blank cells indicate that the model does not use the input parameter.

Model Input Parameters	**Tephra2**	**HAZMAP**	**Fall3D**	**PUFF**	**Ash3D**	**HYSPLIT**	**NAME**
Horizontal Wind Advection	x	x	x	x	x	x	x
Vertical Wind Advection			x	x	x	x	x
Eruption Duration (hours)			x	x	x		
Total Erupted Mass/Volume (kg, m^3)	x	x			x		
Mass Eruption/Flow Rate (kg/s^{-1})			x				
Plume Height (km, m)	x	x	x	x	x		
Plume Shape	x	x	x	x	x		
Total Grain Size Distribution (phi)	x	x	x		x		
Wet Deposition			x		x	x	x
Aggregation			x				
Particle Density (kg/m^3)	x	x	x		x		
Particle Shape	x	x	x	x	x	x	
Diffusion Coefficient	x	x			x		
Gas Species		x				x	x
Number of Ash Particles				x			

Meteorological Data:

One primary input that dictates ash and gas movement is meteorological data. The meteorological conditions strongly influence ash and gas transport and deposition during an eruption (Bonadonna and Costa 2013), especially wind direction and speed (Bursik 2001; Mulena et al. 2016). On small scales, wind shapes local eddy structures, while at larger scales, wind influences the entire plume trajectory (Bursik 2001). Eruptions occurring during near-zero wind conditions would result in a perfect circle-shaped deposition around the plume as all the ash and gas would be diffused rather than advected (Suzuki 1983). There are four types of meteorological datasets:

- **Numerical Weather Prediction (NWP)**: NWP computer models process current weather observations to forecast future weather (e.g., Haupt et al. 2017).
- **Observational Reanalysis**: meteorological data that consists of physical parameters that are measured directly through instrumentation and then synthesised to create a new dataset (e.g., Hersbach et al. 2020), most commonly used in the simulation of prehistoric events (Buckland et al. 2022).
- **Real-time**: meteorological datasets that record current meteorological and atmospheric data.
- **Historical**: climate data - not commonly used in VATDMs, as for some models, the data does not consist of enough elements needed to model ash dispersal.

Errors in meteorological data can propagate into forecasts created by VATDMs (Kristiansen et al. 2012; Bonadonna and Costa 2013; Poulidis et al. 2018). This is especially true for NWP data, where error growth inevitably increases with time into the future (Dacre et al. 2016). The resolution of meteorological data needed within a model depends mainly on the specific problem being considered (Bonadonna and Costa 2013). To keep computational costs low, lower resolution meteorological datasets are better suited to modelling particle dispersion with global coverage. Higher resolution meteorological datasets are better suited to modelling particle dispersion at local and regional levels, as they can capture the complexity of the local meteorological conditions (Bonadonna and Costa 2013).

Eruptions Source Parameters (ESPs):

The second type of input parameters for VATDMs are ESPs; ESPs are quantified primarily from field data (e.g., Longchamp et al. 2011; Bonadonna and Costa 2012), direct observation (e.g., Ripepe et al. 2013), remote sensing (e.g., Prata and Grant 2001), and/or inversion techniques (e.g., Volentik et al. 2010; Mannen 2014). ESP include:

- **Total Erupted Mass/Volume**

Erupted mass defines how much tephra or gas is injected from a volcano, and thus how it will be dispersed and eventually deposited (Turner and Hurst 2001; Bonadonna and Costa 2013). However, it is also one of the most difficult input parameters to derive accurately from field data (sampling of past deposits), even for eruptions that are observed in real-time (Bonadonna and Costa 2013; Aubry et al. 2021). The best estimates of volume (primarily tephra) are obtained by rapid, extensive mapping within a few days of an eruption (Mastin et al. 2009). However, this can be impractical due to the location of the volcano, continuous eruptions, and other volcanic hazards. The calculation of mass and volume is not straightforward. Issues include: (1) non-linearity of the functions linking area and thickness, and the integration of functions is subject to large errors due to limitations on the amount of field data available, especially in distal areas where ash is thin and easily re-distributed; (2) A general lack of data especially for prehistorical eruptions; (3) The reliability and accuracy of data in the immediate vicinity and distal areas of the eruptive vent due to burial, pyroclastic flow contributions, vent collapse, erosion, and where tephra may fall into the sea (Fierstein and Nathenson 1992; Bonadonna and Houghton 2005; Engwell, Sparks, and Aspinall 2013; Aubry et al. 2021). As volume and mass calculations are not straightforward, estimates derived from field data should be treated as minimum values (Bonadonna and Costa 2013).

- **Mass Eruption/Flow Rate**

MER (average eruption rate in kg/s^{-1}) is another key ESP. Typically, MER is obtained from plume height through field-based and empirical relationships (e.g., Carey and Sigurdsson 1982; Sparks 1986; Wilson and Walker 1987). However, these methodologies to obtain MER are not perfect (e.g., Mastin et al. 2009; Degruyter and Bonadonna 2012). For example, the semi-empirical formula by Wilson and Walker (1987) only holds for circular-vent plumes less than

35 km height, with plume temperatures of about 800°C (appropriate for andesitic magmas). In contrast, the model of Sparks (1986) is applied to plumes with temperatures between 400°C and 1000°C and for both temperature and tropical atmospheres. On the other end, Mastin et al. (2009) considered twenty-eight eruptions to obtain an empirical trend for MER and plume height, but disregarded eruption dynamics, magma temperature and composition. The relationship between MER and plume height is complicated by a variety of real-world factors, including but not limited to TGSD, wind velocity, the plume water content, and aggregation (e.g., Degruyter and Bonadonna 2012; Girault et al. 2014).

- **Eruption Duration**

Eruption duration is an important ESP as it is intrinsically linked to MER (e.g., Wilson and Walker 1987). There are a variety of methods to determine eruption duration, including: TEM/volume divided by the total MER (e.g., Romero et al. 2016), image analysis from thermal and visual cameras (Calvari et al. 2018), ground-based radar (e.g., Freret-Lorgeril et al. 2018), satellite observations (e.g., Pistolesi et al. 2015), analysis of infrasound signals and volcanic tremor (e.g., Sato, Fukui, and Shimbori 2018), and visual reports from local witnesses (e.g., Van Eaton et al. 2016). There are several challenges associated with defining eruption duration, particularly in the literature. Eruption duration may be defined as the period of time when a volcano is erupting fresh volcanic material (e.g., Gunn et al. 2013). However, eruptions can show distinct phases and styles, some of which contain ash, and some of which do not, and are often interspersed by periods of quiescence (Bebbington and Jenkins 2019).

- **Plume Height**

An accurate description of plume height is crucial as it is needed to calculate MER and understand the overall plume dynamics (Scollo et al. 2008). Plume height can be obtained in near real time through direct observation (e.g., Folch et al. 2008; Carn et al. 2009; Major and Lara 2013), ground-based instruments (e.g., LiDAR (Light Detection and Ranging), radar, and thermal cameras, Prata and Grant 2001; Pailot-Bonnétat et al. 2020). Plume height can also be obtained retrospectively, through field studies of eruption deposits, i.e., using isopleth maps derived from tephra stratigraphy to calculate plume height (e.g., Carey et al. 1990; Alfano et al. 2011; Rossi, Bonadonna, and Degruyter 2019). Like many ESPs, the certainty of plume height

estimates can be affected by many factors, such as bad weather conditions, low frequency of satellite passages, poor vertical resolution (radar measurements) (e.g., Prejean and Brodsky 2011; Oddsson et al. 2012; Dürig et al. 2018), resulting in plume height differing up to several kilometres from each other (Ripepe et al. 2013). It has been noted by Deligne (2021) and Aubry et al. (2021), who compiled extensive ESP databases, that literature surrounding plume heights often fails to specify whether the plume height is classed as height above vent or height above sea level. It may also be difficult to provide a single plume height for an eruption, as plume heights can fluctuate over the course of an eruption (Deligne 2021).

- **Plume Shape**

The vertical distribution of ash in an eruptive column (plume shape - can also be referred to as source term) can play a significant role in the deposition of ash, as it dictates the mass per unit time release of ash at different heights. How the ash is distributed in an ash cloud can be considered as a point (e.g., Hurst and Davis 2017), linearly (e.g., Yang et al. 2020), or through other empirical distributions of mass along the column, such as the Suzuki distribution (Suzuki 1983), uniform distribution (Bonadonna, Macedonio, and Sparks 2002), Beta distribution (Folch 2012), Buoyant Plume Theory (BPT) (Sparks et al. 1997; Bursik 2001), and other distributions such as Poisson and exponential (Draxler and Hess 1997; Searcy, Dean, and Stringer 1998). Identifying the best source term to use within a model can be difficult, especially when considering strong and weak plumes (Sparks et al. 1997). Strong plumes are associated with wide but elongate deposits, lasting a few seconds to hours and can reach up to 15 km above the eruptive vent. In contrast, weak plumes generate narrower deposits, last from days to months, and reach a few kilometres (Sparks et al. 1997; Bonadonna et al. 2015).

- **Total Grain Size Distribution**

Atmospheric dispersal and deposition models can model a wide range of particle sizes, from very fine ash (<63 μm, >4 phi) to ballistics (>64 mm, <-6 phi) (Osman et al. 2020). For tephra modelling, models are strongly dependent on the TGSD of tephra, defined as the grain-size distribution of the whole eruptive mixture expelled during an explosive eruption (Pioli, Bonadonna, and Pistolesi 2019). For example, if the eruption is dominated by smaller particles, then the plume will travel further. If it is dominated by larger particles, then the particles

will fall out sooner, and the plume will not travel as far (Turner and Hurst 2001). The TGSD within a plume is related to the sorting processes acting on the particles, mostly resulting from magma fragmentation (Bonadonna et al. 2015). Generally, the grain size distribution of particles within the eruptive plume decreases with height, i.e., large blocks and bombs (>64 mm, <-6 phi) are deposited near the vent, while the plume and umbrella clouds are dominated by lapilli (2 to 64 mm, -1 to -6 phi) and ash (<2 mm, >-1 phi) respectively (Bonadonna et al. 2015). Determining the TGSD from an eruption commonly relies on detailed characteristics of tephra deposits, which are often incomplete and lack data, especially below 63μm (>4 phi) (Bonadonna and Houghton 2005; Girault et al. 2014; Aubry et al. 2021).

A.1.4 Model Output

Table A.3: Common output parameters in atmospheric dispersion and deposition models. 'x' indicates that the model produces an output; blank cells indicate that the model does not produce an output.

Model Outputs	**Tephra2**	**HAZMAP**	**Fall3D**	**PUFF**	**Ash3D**	**HYSPLIT**	**NAME**
Tephra Load/Accumulation (kg/m^2)	x	x	x	x	x	x	x
Track Points			x	x		x	x
Trajectory of Specific Tracers				x		x	x
Deposit Thickness (mm)					x		
Ash/Gas Cloud Concentrations (mg/m^-3)			x	x	x	x	x
Ash/Gas Cloud Area/Volume (km^2)				x	x	x	
Age of Particle				x		x	
Hazard Probability		x					x

Common outputs from atmospheric dispersion and deposition models include tephra load/accumulation (kg/m^2), deposit thickness (mm), ash/gas cloud concentrations, and ash/gas cloud area/volume (km^3) (Table A.3). However, output data are not consistent for all atmospheric dispersion and deposition models (i.e., Fall3D will output atmospheric particle concentrations, while Tephra2 will not).

Table A.4: Assumptions in atmospheric dispersion and deposition models.

Model Assumptions	Tephra2	HAZMAP	Fall3D	PUFF	Ash3D	HYSPLIT	NAME
Particles Settle at Their Terminal Velocity		x	x	x	x		x
Isotropic Horizontal Diffusion	x	x					
Aggregates are Round					x		
Turbulent Velocity Constant with Height						x	
Vertical Mixing Diffusivity = Heat Coefficient						x	
Ash Disperses Passively						x	
Vertical Column Extends Above Vent	x						
Gaussian Grain Size Distribution	x						
Spherical Particles	x			x			x
Vertical Wind Velocity Negligible	x						
Wind Speed/Direction Constant in Domain	x						
Wind Field Constant During Eruption	x						
Eruption Conditions Constant per Phase			x				
Linear Interpolation (Density & Shape)			x				
Wet Deposition Below PBL			x				

A.1.5 Model Assumptions

Table A.4 summarises the assumptions made by different atmospheric dispersion and deposition models. While some models explicitly define key physical processes, others rely on implicit assumptions. For example, HAZMAP, Fall3D, PUFF, and Ash3D explicitly assume that particles settle at their terminal velocity, whereas other models, such as Tephra2 and HYSPLIT, do not explicitly state that assumption. Some assumptions, such as spherical particles (Tephra2, PUFF, NAME) and Gaussian grain size distributions (Tephra2), are simplifications used to improve computational efficiency, but may not always reflect real-world conditions.

A.2 Flow Models

A.2.1 Model Dimensionality

1D models are primarily channelised flows, where the fluid advances downslope in a single direction, as opposed to free flows that can spread across a slope. The simplest 1D models use digital elevation models (DEM) and sometimes volume as input parameters, as knowledge of the thermal and rheological properties of flows and effusion rate is not required (e.g., LAHARZ, Schilling 1998). There is a subset of 1D models referred to as stochastic models, which describe

Table A.5: Common volcanic flow models.

Model	Dimensionality	Model Type	Model Approach	Volcanic Flow Hazard
DOWNFLOW[1]	1D	Numeric	Stochastic 1D	Lava
MrLavaLoba[2]	1D	Empiric	Stochastic 1D	Lava
LaharZ[3]	1D	Empiric	1D	PDC, Lahar
Energy Cone (ΔH/L)[4]	1D	Empiric	1D	PDC
FLOWGO[5]	1D	Numeric	1D	Lava
MAGFLOW[6]	2D	Numeric	CA	Lava
SCIARA[7]	2D	Numeric	CA	Lava
MOLASSES[8]	2D	Numeric	CA	Lava
Titan2D[9]	2D	Numeric	DAM	PDC, Lahar
VolcFlow[10]	2D	Numeric	DAM	Lava, PDC
GPUSH[11]	3D	Numeric	CDF	Lava, PDC
FLOW-3D	3D	Numeric	CDF	Lava, PDC, Lahar

[1] Favalli et al. (2005), [2] de'Michieli Vitturi and Tarquini (2018), [3] Schilling (1998), [4] Malin and Sheridan (1982), [5] Harris and Rowland (2001), [6] Cappello et al. (2016b), [7] Crisci et al. (1986), [8] Richardson (2016), [9] Patra et al. (2005), [10] Kelfoun and Vargas (2016), [11] Zago et al. (2019)

the system's state not by unique values but rather probability distributions. In principle, most numerical models could be used as stochastic models, such as DOWNFLOW (Favalli et al. 2005) and MrLavaLoba (de'Michieli Vitturi and Tarquini 2018). More complex 1D models exist, primarily FLOWGO, a lava flow model that tracks lava flow advancement based on its rheological properties (Harris and Rowland 2001).

One of the most popular 2D model approaches is cellular automata (CA) modelling (Cordonnier, Lev, and Garel 2015), commonly used in probabilistic hazard assessments (e.g., Herault et al. 2009; Gallant 2016). CA models use a simple set of rules (often not derived directly from physical laws) to distribute fluids from a central cell (i.e. a grid point) to a neighbouring one, depending on the cell's relative elevation - until all cells meet the criteria defined by the model. CA models account for mass conservation, non-Newtonian flow behaviour, viscosity variation with temperature, and radiative heat exchange with the atmosphere (Kavanagh, Engwell, and Martin 2018). Complex behaviours such as Bingham flow stopping mechanics can be incorporated into CA models, allowing for more realistic lava flow deposits (Hyman, Dietterich, and Patrick 2022). CA models are particularly good at simulating lava flows over very gentle terrain or for modelling how flows fill depressions (Hyman, Dietterich, and Patrick 2022). However, 2D models lack the detailed vertical structure of lava flows, and generally only attempt to simulate the final deposit thickness and extent since they do not contain fluid kinematics (Cordonnier, Lev, and Garel 2015; Hyman, Dietterich, and Patrick 2022).

Another approach is depth-averaged models (DAM), which applies the Navier-Stokes equations derived from the principles of computational fluid dynamics (CFD) (Cordonnier, Lev, and Garel 2015). DAM models exploit the large separation between vertical and horizontal gradients within a flow, averaging energy, mass, and momentum over the flow's depth (Cordonnier, Lev, and Garel 2015; Hyman, Dietterich, and Patrick 2022). DAM models capture flow kinematics (e.g., arrival times), which are crucial for hazard forecasting. However, they are significantly slower to run compared to empirical and CA models (Hyman, Dietterich, and Patrick 2022). DAM models are steady-state, meaning they do not account for changes in flow behaviour over time (Cordonnier, Lev, and Garel 2015). They can also only represent the cooling effects in relation to a depth-averaged temperature, which does not adequately capture the thermal stratification typically observed in lava flows. For pyroclastic density current (PDC) hazard, DAM models are best suited for concentrated block-and-ash flows and collapsing lava domes rather than more dilute flows (Dufek 2016).

3D models are thermodynamically coupled rheological models, using full CDF principles, that simulate heat transfer and account for a range of viscous flow behaviours (Neri et al. 2015a). These models have higher computational requirements but are particularly suitable for long-lived, cooling-limited flows (Dietterich et al. 2017).

A.2.2 Model Input Parameters

For volcanic flow models, the main input parameters can be classified into three categories:

1. **Topography**
2. **Physical properties of the flow**
3. **Simulation Conditions**

Table A.6: Common input parameters in volcanic flow models. 'x' indicates that the model uses the input parameter; blank cells indicate that the model does not use the input parameter.

Model Input Parameter	**MrLavaLoba**	**DOWNFLOW**	**FLOWGO**	**MAGFLOW**	**SCIARA**	**MOLASSES**	**FLOW-3D**	**TITAN2D**	**VolcFlow**	**LAHARZ**	**Energy Cone**
DEM	x	x	x	x	x	x	x	x	x	x	x (Height and Slope)
Starting Coordinates	x	x	x	x	x	x	x	x	x	x	
Volume (m^3)	x				x	x	x	x		x	
Effusion Rate/Velocity (m/s)			x	x		x	x	x	x		
Flow Thickness (m)					x	x	x	x			
Flow Viscosity			x	x			x		x		
Flow Density (kg/m^3)			x		x				x		
Basal Friction Angle							x	x	x		
Internal Friction Angle								x	x		
Eruption Temperature			x	x	x						
Solidification Temperature			x	x	x						
Number of Flows	x										
Time Steps / Max Run Time (seconds)								x	x		

Topography

One of the primary input parameter categories of any volcanic flow model is topography. Topography includes input parameters such as digital elevation models (DEMs) and basal friction angle of flows.

- **DEM**

Flows are highly sensitive to topographic changes, such as obstacles or breaks in slopes (Branney and Kokelaar 2002; Sulpizio and Dellino 2008; Dietterich et al. 2015). Most volcanic flow models require a DEM of the volcano and surrounding topography to simulate flow emplacement (Capra et al. 2011; Cordonnier, Lev, and Garel 2015). A DEM is a digital model of topography composed of X, Y coordinate pairs with assigned elevation values (Z) (Florinsky 2012). DEMs can be generated using various methods, including satellite imagery (e.g., ASTER - Advanced Spaceborne Thermal Emission and Reflection Radiometer) (e.g., Fujisada, Urai, and Iwasaki 2011), LiDAR (e.g., Raber et al. 2007), and vector-based data such as contour lines from topographic maps (e.g., Szypuła 2019).

For hindcasting, a major challenge in using DEMs is that they typically represent present-day topography rather than pre-event terrain conditions (Procter et al. 2010b). This can be worked around, although impractically, by subtracting thicknesses of past flow deposits based on detailed stratigraphy of the deposit (e.g., Daag 2003; Dietterich et al. 2021). For forecasting, using a DEM of the current landscape is generally more applicable. However, this approach becomes challenging during multi-flow events, where DEMs must be dynamically updated to improve simulation accuracy (e.g., De Beni, Cantarero, and Messina 2019).

- **Basal Friction Angle**

Another key input parameter common in volcanic flow models is the basal friction angle. This input parameter represents the minimum slope angle at which a flow begins to move from a static position and is one of the primary factors influencing flow runout (Sulpizio et al. 2010; Capra et al. 2011; Breard et al. 2020). However, the basal friction angle has not been directly related to physical parameters that can be measured in the field or laboratory (Procter et al. 2010b). Instead, it is typically calibrated to known values by matching simulated inundation areas with observed travel times (e.g., Neglia et al. 2021).

Physical Flow Properties

Another main type of input parameter category is physical flow properties. Eruption conditions may include volume, effusion rate, temperature, viscosity, and density of a flow. Different volcanic flow models contain different input parameters of physical flow properties depending on the type of volcanic flow they simulate.

- **Lava Rheology**

Lava flows are highly dense, with viscosity being a key factor in determining their shape and extent. The viscosity of lava can be measured through laboratory measurements (e.g., Sehlke et al. 2014; Kolzenburg et al. 2016), estimated in the field (e.g., Chevrel et al. 2018; Chevrel, Pinkerton, and Harris 2019), and/or inferred from empirical relationships based on temperature (e.g., Pinkerton and Norton 1995; Ishihara, Iguchi, and Kamo 1990; Giordano and Dingwell 2003).

- **Lahar and PDC Rheology**

PDC and lahar exhibit complex and poorly constrained flow rheologies. PDCs are typically classified into two end-member types: pyroclastic flows, which are highly concentrated and topographically controlled, and pyroclastic surges, which are dilute, turbulent currents capable of overriding obstacles (Sparks 1976). Modelling of PDCs is challenging due to contrasting behaviours of these components - while dense pyroclastic flows follow topographic lows, dilute surges are less constrained by terrain (Sparks 1976; Burgisser and Bergantz 2002). Additionally, PDC dynamics remain one of the least understood volcanic hazards due to complexities in gas-particle interactions, flow mechanics, and generation processes (Lube et al. 2020; Jones et al. 2024).

Lahars, in contrast, are water-saturated flows characterised by high density and velocity. Their properties, such as particle concentration, temperature, and bulk rheology, vary significantly in space and time due to processes such as flow bulking (erosion and incorporation of debris) and debulking (Thouret et al. 2020).

The rheological properties for both PDCs and lahar are derived from laboratory experiments (e.g., Roche et al. 2004; Lube et al. 2015; Jones et al. 2024), field-based depositional analyses

(e.g., Sparks 1976; Branney and Kokelaar 2002; Dumaisnil et al. 2010; Bernard et al. 2014), and remote sensing techniques (e.g., Kumagai et al. 2009; Bosa et al. 2021; Macorps 2021; Bosa et al. 2024), while studies on the rheology of propagating PDCs are sparse (Sulpizio et al. 2014; Delannay et al. 2017).

- **Volume**

Flow volume is another critical input parameter in volcanic flow models, with a direct relationship to the size of the hazard footprint. Volume estimates are typically derived by comparing pre- and post-event topography through mapping techniques (e.g., Stevens, Wadge, and Murray 1999; Muñoz-Salinas et al. 2009; Kubanek et al. 2015; Flynn, Williams, and Ramsey 2024).

- **Effusion Rate**

The effusion rate, similarly to volume, plays a crucial role in determining a flow's final extent and velocity. Higher effusion rates typically result in flows that travel further and faster (Walker 1973; Harris, Dehn, and Calvari 2007). Effusion rates for lava flows can be measured directly through thermal imaging (using cameras or satellites) (e.g., Oppenheimer et al. 1993; Harris et al. 1997; Harris et al. 2005; Calvari et al. 2010), changes in DEM over the course of flow emplacement (e.g., Poland 2014), or inferred through volume-based measurements (e.g., Zebker et al. 1996; Behncke and Neri 2003; Andronico et al. 2005).

A.2.3 Model Output

Table A.7: Common output parameters in volcanic flow models. 'x' indicates that the model produces an output; blank cells indicate that the model does not produce an output; '?' indicates that the model produces an output, but it was not specified in the literature.

Model Output Parameter	**MrLavaLoba**	**DOWNFLOW**	**FLOWGO**	**MAGFLOW**	**SCIARA**	**MOLASSES**	**FLOW-3D**	**TITAN2D**	**VolcFlow**	**LAHARZ**	**Energy Cone**
Runout (m)	x		x	x	?	?	x	x	x	x	x
Inundated area (km^2, m^2)	x		x	x	x	x	x	x	x	x	
Thickness (cm, m)	x		x	?	?	x	x	x	x		
Maximum velocity (m/s) vs time (seconds)							x	x	x		
Average velocity (m/s) vs time (seconds)			x				x	x			
Velocity (m/s) at maximum thickness (m) vs time (seconds)							?	x			
Flow passage count per point		x									
Effusion rate			x								
Viscosity			x								

There are a variety of outputs produced by volcanic flow models (Table A.7). All volcanic flow models, except for DOWNFLOW, generate or are assumed to generate runout distances for the simulated flows. In addition, all models, except for the Energy Cone model, produce an output of inundated area. Eight of the eleven models also calculate flow thickness. DOWNFLOW stands out as an outlier in this context, as its sole output is to record each time a flow (lava) passes over specific points.

A.2.4 Model Assumptions

Table A.8: Assumptions in volcanic flow models.

Model Assumption	**MrLavaLoba**	**DOWNFLOW**	**FLOWGO**	**MAGFLOW**	**SCIARA**	**MOLASSES**	**FLOW-3D**	**Titan2D**	**VolcFlow**	**LAHARZ**	**Energy Cone**
Conservation of Mass/Volume	x		x			x	x	x	x	x	x
Flows with Several Pulses Treated as Single Event		x						x		x	
Constant Rheology in Vertical Profile			x					x	x		
Constant Effusion Rate			x								
Flow Stops When Cooled Enough to Impede Motion			x	x							
Flow Starts as Ellipsoidal Pile	x							x			
Channelised Flow			x								
Flow Density Insensitive to Temperature			x								
Each Cell Has Constant Dimensions					x					x	
Shallow, Incompressible Flow Behaviour								x			
Ignores Topography											x

For volcanic flow models, model assumptions are primarily based on the governing equations. For example, depth-averaged models such as Titan2D and VolcFlow assume that the vertical properties of a flow, such as density and velocity, are reasonably represented by their depth-averaged values at each location (Table A.8). While this simplifies the modelling of the flow, it may overlook fine-scale stratification within the flow.

A.3 Volcanic Ballistic Projectile Models

A.3.1 Model Simulation

Table A.9: Common volcanic ballistic projectile models.

Model	Dimensionality	Model Type
Eject![1]	2D	Numeric
Alatorre-Ibargüengoitia and Delgado-Granados (2006)	2D	Numeric
LPAC[2]	2D	Numeric
Ballista[3]	3D	Numeric
Great Balls of Fire (GBF)[4]	2D	Numeric

[1] Mastin (2001), [2] de'Michieli Vitturi et al. (2010), [3] Tsunematsu et al. (2014) and Tsunematsu et al. (2016), [4] Biass et al. (2016a)

Most commonly, VBP models are two-dimensional (Table A.9). Models such as Eject!, Great Balls of Fire, and Alatorre-Ibargüengoitia, Delgado-Granados, and Dingwell (2012) use simplified drag and gravity equations to calculate the movement and deposition of VBPs. The LPAC VBP model, unlike others, does not consider the initial ejection velocity or ejection angle. Instead, it uses a simplified BBO (Basset-Boussenesq-Ossen) equation (Maxey and Riley 1983; Crowe et al. 1998), which expresses the Lagrangian acceleration of a particle as the sum of the forces exerted on it along its trajectory. All equations in these models are solved numerically using fourth-order Runge Kutta integration (Wilson 1972), where calculations continue until the vertical position of the VBP reaches a specified landing elevation.

The Ballista Model by Tsunematsu et al. (2016) is one of the few 3D VBP models, considering multiple particles and their collisions in 3D space (Tsunematsu et al. 2014). Particles are ejected following a parabolic trajectory until they reach the ground or collide with another airborne particle (Tsunematsu et al. 2014).

Table A.10: Common input parameters in volcanic ballistic projection models. 'x' indicates that the model uses the input parameter; blank cells indicate that the model does not use the input parameter.

Model Input Parameter	**Eject!**	**Alatorre-Ibargüengoitia, Delgado-Granados, and Dingwell (2012)**	**LPAC**	**Ballista**	**Great Balls of Fire**
Grain Size (phi)	x	x	x	x	x
Grain Density (kg/m^3)	x	x	x	x	x
Ejection Velocity (m/s)	x	x	x	x	x
Ejection Angle	x	x	x	x	x
Ejection Azimuth					x
Number of Particles				x	x
Takeoff Elevation (m)	x		x	x	
Wind Speed	x	x			x
Wind Direction					x
Time Step (seconds)			x		x
Atmospheric Pressure					x
Temperature	x				x
Drag Coefficient		x	x		
Reduced Drag Radius	x				x

A.3.2 Model Input Parameters

There are two main input parameter categories found in all volcanic ballistic projectile (VBP) models:

1. **Block Properties**: including the density and size of the ejected volcanic ballistic projectile.

2. **Ejection Properties**: including ejection angle and initial velocity.

Block Properties

One of the main input parameters for VBP models is the ballistic projectile itself. All VBP models take into account the density and size of the ballistic projectile. The size of these projectiles ranges from a few centimetres to tens of metres in diameter (Nairn and Self 1978; Andronico, Scollo, and Cristaldi 2015; Tsunematsu et al. 2016), while the density of VBP can range from around 600 to 2700 kg m^{-3} (Taddeucci et al. 2017). Most commonly, VBP properties such as density and size used in hazard models can be either measured (e.g., Alatorre-Ibargüengoitia, Delgado-Granados, and Dingwell 2012; Fitzgerald et al. 2014; Tsunematsu et al. 2016) or assumed (e.g., Chouet, Hamisevicz, and McGetchin 1974; Patrick et al. 2007; de'Michieli Vitturi et al. 2010; Kilgour et al. 2010).

Ejection Properties

The ejection properties of the VBP directly influence how far the projectile will travel (Bertin 2017). The main ejection properties considered in VBP models are the initial velocities and ejection angles. Ejection angles and velocities are often extracted from photos with short-term exposure time (e.g., Chouet, Hamisevicz, and McGetchin 1974; Blackburn, Wilson, and Sparks 1976; Ripepe, Rossi, and Saccorotti 1993; Edwards et al. 2017) and thermal videos (e.g., Patrick et al. 2007; Capponi et al. 2016).

Other Input Parameters

In addition to the primary VBP properties, other input parameters also play a role in the modelling process, although not all models require them. For example, the atmosphere through which the VBP travels impacts its transportation (Mastin 2001). A key parameter in this regard is the drag coefficient, which quantifies the resistance of a projectile as it moves through the air. The drag coefficient can be measured experimentally (Bagheri and Bonadonna 2016), or estimated from field-based observations (Taddeucci et al. 2017). Some models, such as Eject! and Great Balls of Fire, introduce an arbitrary distance (ranging from tens to hundreds of metres, Mastin 2001) from the vent over which there is a reduced drag coefficient. For simplicity, models such as LPAC and Ballista assume a constant drag coefficient throughout the VBPs' trajectory (Taddeucci et al. 2017).

A.3.3 Model Output

The output of VBP models is largely similar across all the VBP models discussed here (Table A.11). The primary differences are that Ballista and Great Balls of Fire models can simulate multiple particles, and that the LPAC model can also output the flight times for each simulated particle.

Table A.11: Common output parameters in volcanic ballistic projection models. 'x' indicates that the model produces an output; blank cells indicate that the model does not produce an output.

Model Output	**Eject!**	**Alatorre-Ibargüengoitia, Delgado-Granados, and Dingwell (2012)**	**LPAC**	**Ballista**	**Great Balls of Fire**
Maximum Height (km, m)	x	x	x	x	x
Distance (m)	x	x	x	x	x
Particle Trajectory	x	x	x	x	x
Maximum Velocity (m/s)	x	x			x
Exit Velocity (m/s)			x		
Flight Time (seconds)			x		
Single Particle	x	x	x		
Multi-Particle				x	x

A.3.4 Model Assumptions

Table A.12: Assumptions in volcanic ballistic projection models.

Model Assumption	**Eject!**	**Alatorre-Ibargüengoitia, Delgado-Granados, and Dingwell (2012)**	**LPAC**	**Ballista**	**Great Balls of Fire**
Constant Drag Coefficient	x		x	x	
Spherical Particles			x	x	x
Constant Air Density	x				
Flow Fields Affect Particle Trajectory			x		
Parabolic Trajectories				x	
Equal Velocity Between the Fluid and Particle Phases at the Conduit Exit	x				
Positive Values of Thermal Lapse Rate Imply Decreasing Temperature with Elevation	x				
Ejection of VBP Occurs when the Acceleration of Caprock is 8% of Initial Acceleration		x			

Model assumptions in VBP models are primarily based on the particle shape (spherical) and the drag coefficient (air resistance forces on the VBP) (Table A.12). Eject!, LPAC, and Ballista

assume a constant drag coefficient, whereas Great Balls of Fire and Ballista explicitly assume spherical particles, which simplifies aerodynamic calculations but may not fully capture the irregular shapes of real projectiles (Sork et al. 2025).

Appendix B

Global Sensitivity Analysis

B.1 Appendix B

B.1.1 Sobol' Indices

Sobol' indices (Sobol' 1990) is a Monte Carlo-based calculation that quantifies the contribution of each input variable to the output variance of a model. Given a model in the form:

$$y = f(x), x = (x_1, x_2, \ldots, x_k) \in R^k \tag{B.1}$$

where y is a scalar output and $x_1, \ldots, x_k$ are k independent input parameters described by probability distributions, we can calculate the proportion of total variance conveyed to y by each input parameter x_i, known as the first-order effect (S_i^{S}).

$$S_i^{\mathrm{S}} = V[E(y|x_i)]/V(y) \tag{B.2}$$

where

$$V[y] = \sum_{i=1}^{N} (y_i - \bar{y})^2/(N-1) \tag{B.3}$$

where N is the sample size, y_i is the output from the ith combination of inputs, and $\bar{y}$ is the mean output across all combinations of input parameters, and $E(y|x_i)$ is the expected value of y given a fixed value of x_i

We can also extract the proportion of variance contributed by the interaction between pairs of inputs (second-order effect, S_{ij}), triplets of input parameters (third-order effect, S_{ijl}), etc.,

up to the kth-order interaction, e.g.,:

$$S_{ij} = V[E(y|x_i, x_j)] - V[E(y|x_i)] - V[E(y|x_j)]/V(y) \tag{B.4}$$

For a model with three input parameters, the total variance can be expressed as, for example: $S_1 + S_2 + S_3 + S_{1,2} + S_{1,3} + S_{2,3} + S_{1,2,3} = 1$.

S_i^{S} can then be used to rank input parameters by their relative first-order contribution to model output variation. When $\sum_{i=1}^{k} S_i^{\mathrm{S}} = 1$, the model is additive, i.e., the total variance (y) can be fully decomposed as the sum of first-order effects, indicating that there is no interaction between input parameters. However, this is rarely the case in complex models (Puy et al. 2023), and the sum of the first-order indices usually falls well short of total model output variance.

Total-order indices T_i^{S}, which measure the first-order effects of x_i together with its interactions with all the other input parameters, provide information on the non-additive features of a model (Saltelli et al. 2010).

$$T_i^{\mathrm{S}} = 1 - V[E(y|x_{\sim i})]/V(y) \tag{B.5}$$

where $V[E(y|x_{\sim i})]$ is the portion of variance that is explained given all input parameters except x_i. For a three-parameter model, $T_1 = S_1 + S_{1,2} + S_{1,3} + S_{1,2,3}$.

B.1.2 eFAST - extended Fourier Amplitude Sensitivity Test

eFAST is a variance decomposition method, analogous to ANOVA (Cukier, Levine, and Shuler 1978). A transformation (sinusoidal) function, $x = f(j), j = 1, 2, \ldots, M$, is used to convert input parameter values to values along a search curve, based on the sample number M (number of repetitions per input combination) (Figure B.1) - this essentially turns a probability density function into a signal that can then be traced back to the output variance. Using Fourier coefficients, eFAST decomposes the variance of the model output, determining what fraction of the variance can be explained by the variation in each input parameter.

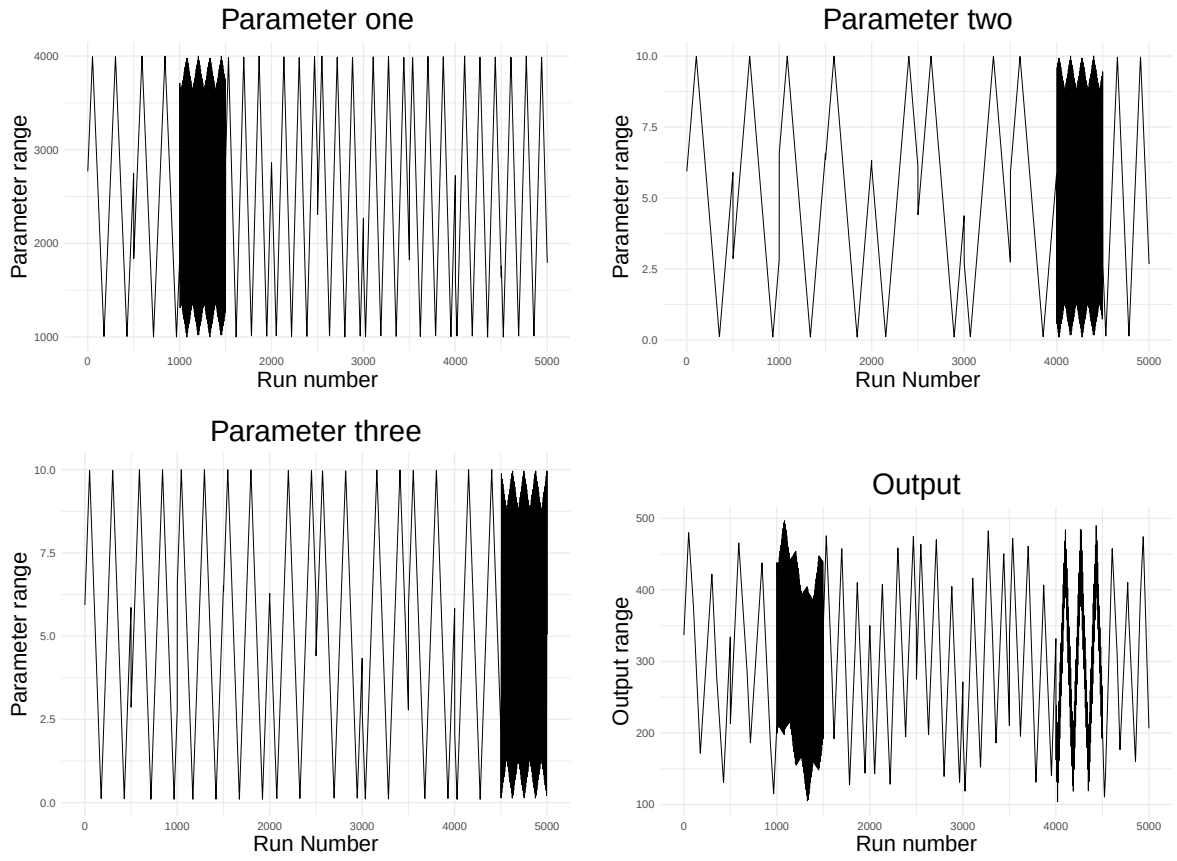


Figure B.1: Visualisation of the eFAST search curve, which shows the decomposition of output variance based on the spectral analysis of input parameters. Results show that input parameter one strongly contributes to the output variance through high spectral amplitude, input parameter two contributes moderately with a smaller amplitude, and input parameter three shows no contribution, as its spectral amplitude remains negligible.

The variance contribution of each input parameter (Equation B.3) (i) is estimated from the Fourier coefficients and its unique frequency (and harmonics of that frequency):

$$Z_i^2 = 2(A_j^2 + B_j^2) \tag{B.6}$$

where A_j and B_j are the Fourier coefficients corresponding to the input parameter's frequency.

The total variance of the model output is then computed as the sum of these variance estimates across all input parameters:

$$Z_{total}^2 = \sum_i Z_i^2 \tag{B.7}$$

Finally, the first-order sensitivity index (S_i^{F}) for a given input parameter is calculated as the

variance attributed to that input parameter divided by the total variance of the model output:

$$S_i^{\mathrm{F}} = Z_i^2 / Z_{total}^2 \tag{B.8}$$

To calculate the total-order sensitivity index (T_i^{F}) of a given input parameter, eFAST calculates the summed sensitivity index contributions of the entire complementary set of input parameters (i.e., all input parameters except x_i). T_i^{F} is then derived as the proportion of variance remaining after accounting for the contribution of this complementary set:

$$T_i^{\mathrm{F}} = 1 - Z_{\sim i} \tag{B.9}$$

where $Z_{\sim i}$ represents the contribution of the complementary set to the total variance:

$$Z_{\sim i} = Z_{total} - Z_i^2 \tag{B.10}$$

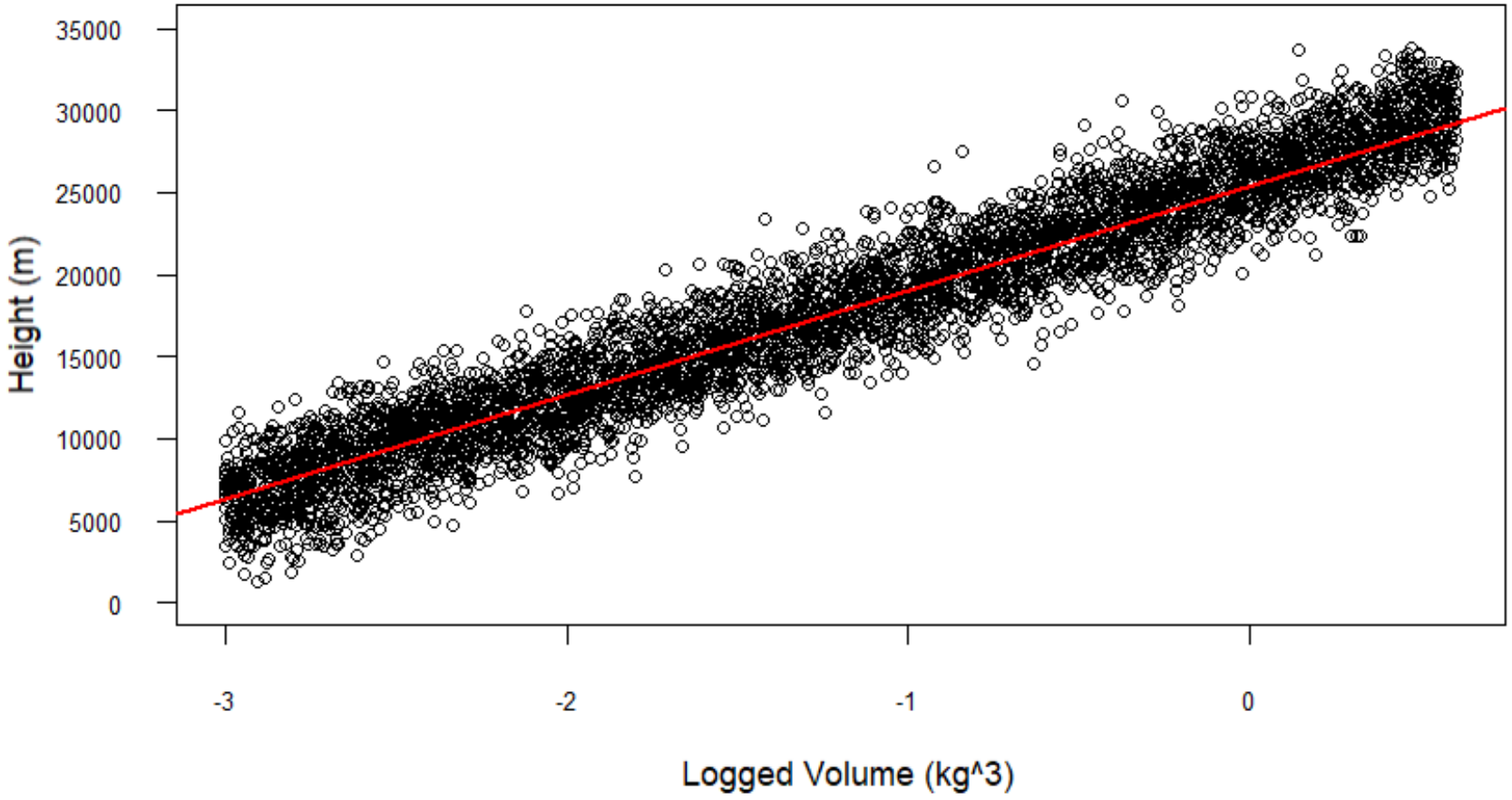


Figure B.2: Sampled points of logged volume (km^3) (via eFAST) against plume height (m). The red line is the equation from Mastin et al. (2009) (Equation 3.1), while each black dot correspond to a sample.

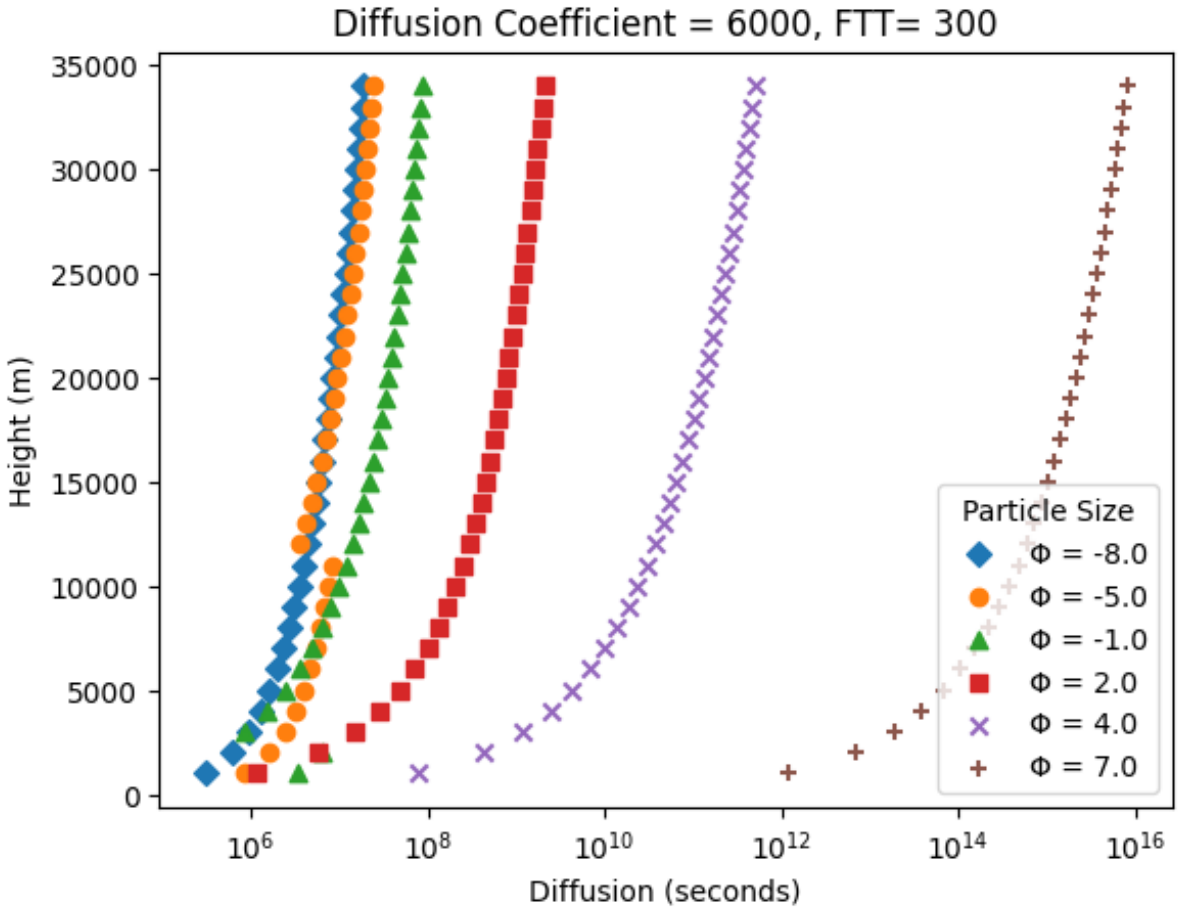


Figure B.3: Plot showing the variation of different particle sizes (phi) for a diffusion coefficient of 6000 and a fall time threshold (FTT) of 300. Note how the diffusion of the orange particle (8 phi) is smaller at 12,000m height compared to 11,000m. Based off Bonadonna et al. (2005).

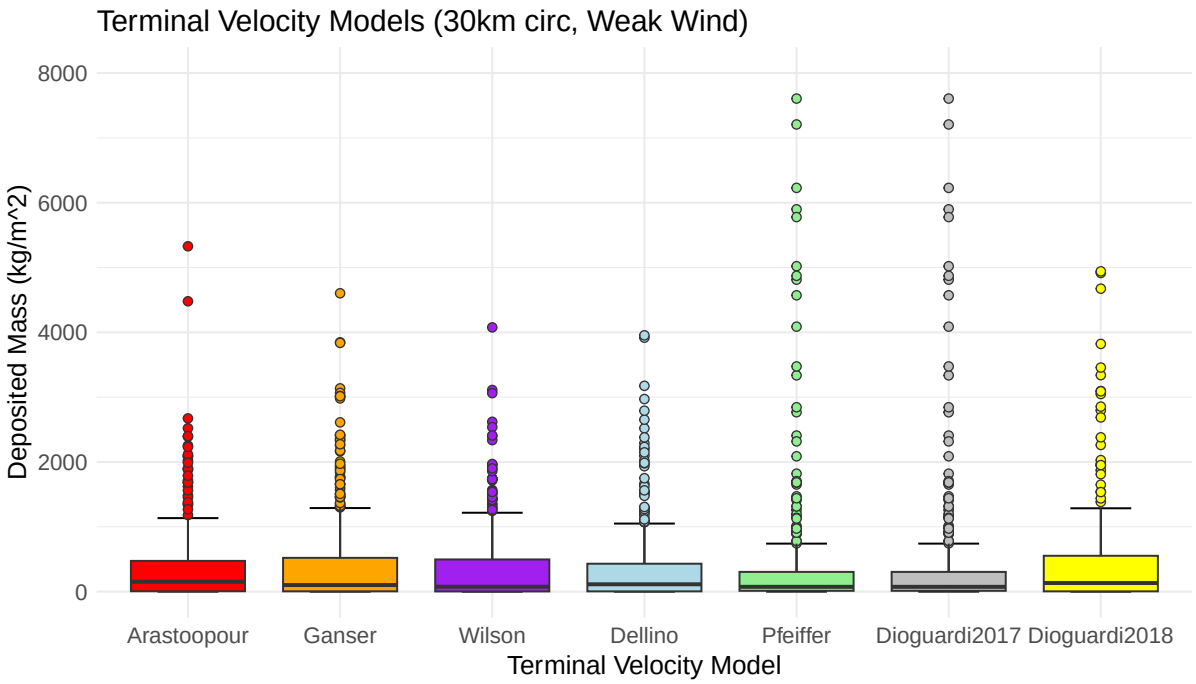


Figure B.4: Boxplots of deposited tephra mass (30 km sampling distance, weak wind) for each of the seven terminal velocity models in Fall3D. Data are from an additional set of simulations (2400), separate from the global sensitivity analysis.

Bibliography

Alatorre-Ibargüengoitia, M A and H Delgado-Granados (2006). "Experimental determination of drag coefficient for volcanic materials: Calibration and application of a model to Popocatépetl volcano (Mexico) ballistic projectiles". In: *Geophysical Research Letters* 33.11. ISSN: 0094-8276. DOI: https://doi.org/10.1029/2006GL026195.

Alatorre-Ibargüengoitia, M A, H Delgado-Granados, and D B Dingwell (2012). "Hazard map for volcanic ballistic impacts at Popocatépetl volcano (Mexico)". English. In: *Bulletin of Volcanology* 74.9, pp. 2155–2169. ISSN: 0258-8900. DOI: 10.1007/s00445-012-0657-2.

Alatorre-Ibargüengoitia, M A et al. (2010). "Energy consumption by magmatic fragmentation and pyroclast ejection during Vulcanian eruptions". In: *Earth and Planetary Science Letters* 291.1-4, pp. 60–69. ISSN: 0012-821X. DOI: https://doi.org/10.1016/j.epsl.2009.12.051.

Alfano, F et al. (2011). "Tephra stratigraphy and eruptive volume of the May, 2008, Chaitén eruption, Chile". In: *Bulletin of Volcanology* 73.5, pp. 613–630. ISSN: 1432-0819. DOI: 10.1007/s00445-010-0428-x.

Anderson, M P, W W Woessner, and R J Hunt (2015). *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*. 2nd ed. Academic Press. ISBN: 0080916384. DOI: https://doi.org/10.1016/C2009-0-21563-7.

Andronico, D, S Scollo, and A Cristaldi (2015). "Unexpected hazards from tephra fallouts at Mt Etna: The 23 November 2013 lava fountain". In: *Journal of Volcanology and Geothermal Research* 304, pp. 118–125. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2015.08.007.

Andronico, D et al. (2005). "A multi-disciplinary study of the 2002–03 Etna eruption: insights into a complex plumbing system". In: *Bulletin of Volcanology* 67, pp. 314–330. ISSN: 0258-8900. DOI: https://doi.org/10.1007/s00445-004-0372-8.

Andronico, D et al. (2008). "The 2002–03 Etna explosive activity: Tephra dispersal and features of the deposits". In: *Journal of Geophysical Research: Solid Earth* 113.B4. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2007JB005126.

Andronico, D et al. (2009). "Monitoring ash emission episodes at Mt. Etna: The 16 November 2006 case study". In: *Journal of Volcanology and Geothermal Research* 180.2, pp. 123–134. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2008.10.019.

Andronico, D et al. (2014). "Representivity of incompletely sampled fall deposits in estimating eruption source parameters: a test using the 12–13 January 2011 lava fountain deposit from Mt. Etna volcano, Italy". In: *Bulletin of Volcanology* 76.10. DOI: 10.1007/s00445-014-0861-3.

Andronico, D et al. (2021). "The 1986–2021 paroxysmal episodes at the summit craters of Mt. Etna: Insights into volcano dynamics and hazard". In: *Earth-Science Reviews* 220. ISSN: 0012-8252. DOI: https://doi.org/10.1016/j.earscirev.2021.103686.

Aravena, A et al. (2020). "Tree-Branching-Based Enhancement of Kinetic Energy Models for Reproducing Channelization Processes of Pyroclastic Density Currents". In: *Journal of Geophysical Research: Solid Earth* 125.7. ISSN: 2169-9313. DOI: https://doi.org/10.1029/2019JB019271.

Aravena, A et al. (2022). "Calibration strategies of PDC kinetic energy models and their application to the construction of hazard maps". In: *Bulletin of Volcanology* 84.3, p. 29. ISSN: 1432-0819. DOI: 10.1007/s00445-022-01538-8.

Aravena, A et al. (2024). "Probabilistic, scenario-based hazard assessment for pyroclastic density currents at Tungurahua volcano, Ecuador". In: *Bulletin of Volcanology* 86.10, p. 84. ISSN: 1432-0819. DOI: 10.1007/s00445-024-01768-y.

Armienti, P and M T Pareschi (1987). "Automatic reconstruction of surge deposit thicknesses. Applications to some Italian volcanoes". In: *Journal of Volcanology and Geothermal Research* 31.3, pp. 313–320. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(87)90074-6.

Aspinall, W P et al. (2011). *Volcano Hazard and Exposure in Track II Countries and Risk Mitigation Measures - GFDRR Volcano Risk Study*. Tech. rep. Bristol University Cabot Institute and NGI Norway: NGI Report 20100806, p. 309.

Aubry, T et al. (2021). "The Independent Volcanic Eruption Source Parameter Archive (IVESPA, version 1.0): A new observational database to support explosive eruptive column model validation and development". In: *Journal of Volcanology and Geothermal Research* 417. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2021.107295.

Aubry, T et al. (2022). "New insights into the relationship between mass eruption rate and volcanic column height based on the IVESPA dataset". In: *Geophysical Research Letters*. DOI: https://doi.org/10.1029/2022GL102633.

Augusiak, J, P J Van den Brink, and V Grimm (2014). "Merging validation and evaluation of ecological models to 'evaludation': A review of terminology and a practical approach". In: *Ecological Modelling* 280, pp. 117–128. ISSN: 0304-3800. DOI: https://doi.org/10.1016/j.ecolmodel.2013.11.009.

Auker, M R et al. (2013). "A statistical analysis of the global historical volcanic fatalities record". In: *Journal of Applied Volcanology* 2.1, p. 2. ISSN: 2191-5040. DOI: 10.1186/2191-5040-2-2.

Baartman, J E M et al. (2020). "On the complexity of model complexity: Viewpoints across the geosciences". In: *CATENA* 186. ISSN: 0341-8162. DOI: https://doi.org/10.1016/j.catena.2019.104261.

Baatz, R et al. (2018). "Steering operational synergies in terrestrial observation networks: Opportunity for advancing Earth system dynamics modelling". In: *Earth System Dynamics* 9.2, pp. 593–609. ISSN: 2190-4979. DOI: https://doi.org/10.5194/esd-9-593-2018.

Babič, F et al. (2022). "Review of Tools for Semantics Extraction: Application in Tsunami Research Domain". In: *Information* 13.1. ISSN: 2078-2489. DOI: 10.3390/info13010004.

Bagheri, G and C Bonadonna (2016). "On the drag of freely falling non-spherical particles". In: *Powder Technology* 301, pp. 526–544. ISSN: 0032-5910. DOI: https://doi.org/10.1016/j.powtec.2016.06.015.

Bal, N and G Rein (2013). "Relevant model complexity for non-charring polymer pyrolysis". In: *Fire Safety Journal* 61, pp. 36–44. ISSN: 0379-7112. DOI: https://doi.org/10.1016/j.firesaf.2013.08.015.

Banks, J (1999). "Introduction to simulation". In: *Proceedings of the 31st conference on Winter simulation: Simulation - a bridge to the future-Volume 1*, pp. 7–13. DOI: https://doi.org/10.1145/324138.324142.

Barberi, F et al. (1992). "Volcanic hazard assessment of Guagua Pichincha (Ecuador) based on past behaviour and numerical models". In: *Journal of Volcanology and Geothermal Research* 49.1, pp. 53–68. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(92)90004-W.

Barberi, F et al. (1993). "The control of lava flow during the 1991–1992 eruption of Mt. Etna". In: *Journal of Volcanology and Geothermal Research* 56.1, pp. 1–34. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(93)90048-V.

Barclay, J et al. (2008). "Framing volcanic risk communication within disaster risk reduction: Finding ways for the social and physical sciences to work together". In: *Geological Society, London, Special Publications*. Vol. 305, pp. 165–177. DOI: 10.1144/SP305.14.

Barker, S J et al. (2019). "Modeling Ash Dispersal From Future Eruptions of Taupo Supervolcano". In: *Geochemistry, Geophysics, Geosystems* 20.7, pp. 3375–3401. ISSN: 1525-2027. DOI: https://doi.org/10.1029/2018GC008152.

Barker, S J et al. (2020). "Taupō: an overview of New Zealand's youngest supervolcano". In: *New Zealand Journal of Geology and Geophysics*, pp. 1–27. ISSN: 11758791. DOI: 10.1080/00288306.2020.1792515.

Barnie, T et al. (2023). "Volcanic plume height monitoring using calibrated web cameras at the Icelandic Meteorological Office: system overview and first application during the 2021 Fagradalsfjall eruption". In: *Journal of Applied Volcanology* 12.1, p. 4. ISSN: 2191-5040. DOI: 10.1186/s13617-023-00130-9.

Barsotti, S, A Neri, and J S Scire (2008). "The VOL-CALPUFF model for atmospheric ash dispersal: 1. Approach and physical formulation". In: *Journal of Geophysical Research: Solid Earth* 113.B3. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2006JB004623.

Barsotti, S et al. (2018). "Assessing Impact to Infrastructures Due to Tephra Fallout From Öræfajökull Volcano (Iceland) by Using a Scenario-Based Approach and a Numerical Model". In: *Frontiers in Earth Science* 6. ISSN: 2296-6463. DOI: 10.3389/feart.2018.00196.

Bauer, P, A Thorpe, and G Brunet (2015). "The quiet revolution of numerical weather prediction". In: *Nature* 525.7567, pp. 47–55. ISSN: 1476-4687. DOI: 10.1038/nature14956.

Baxter, P J et al. (2005). "The impacts of pyroclastic surges on buildings at the eruption of the Soufrière Hills volcano, Montserrat". In: *Bulletin of Volcanology* 67.4, pp. 292–313. ISSN: 1432-0819. DOI: 10.1007/s00445-004-0365-7.

Bayarri, M J et al. (2009). "Using statistical and computer models to quantify volcanic hazards". In: *Technometrics* 51.4, pp. 402–413. ISSN: 0040-1706. DOI: https://doi.org/10.1198/TECH.2009.08018.

Bayarri, M J et al. (2015). "Probabilistic quantification of hazards: a methodology using small ensembles of physics-based simulations and statistical surrogates". In: *International Journal for Uncertainty Quantification* 5.4, pp. 297–325. ISSN: 2152-5099. DOI: 10.1615/Int.J.UncertaintyQuantification.2015011451.

Bebbington, M S (2009). "Volcanic Eruptions: Stochastic Models of Occurrence Patterns". In: *Encyclopedia of Complexity and Systems Science*. Ed. by Robert A Meyers. New York, NY: Springer New York, pp. 1–58. ISBN: 978-3-642-27737-5. DOI: 10.1007/978-3-642-27737-5{_}580-2.

Bebbington, M S and S F Jenkins (2019). "Intra-eruption forecasting". In: *Bulletin of Volcanology* 81.6, p. 34. ISSN: 1432-0819. DOI: 10.1007/s00445-019-1294-9.

— (2022). "Intra-Eruption Forecasting Using Analogue Volcano and Eruption Sets". In: *Journal of Geophysical Research: Solid Earth* 127.6. ISSN: 2169-9313. DOI: https://doi.org/10.1029/2022JB024343.

Beckett, F et al. (2017). "Quantifying the mass loading of particles in an ash cloud remobilized from tephra deposits on Iceland". In: *Atmospheric Chemistry and Physics* 17.7, pp. 4401–4418. ISSN: 1680-7316. DOI: https://doi.org/10.5194/acp-17-4401-2017.

Beckett, F M et al. (2015). "Sensitivity of dispersion model forecasts of volcanic ash clouds to the physical characteristics of the particles". In: *Journal of Geophysical Research: Atmospheres* 120.22, pp. 611–636. ISSN: 2169-897X. DOI: https://doi.org/10.1002/2015JD023609.

Beckett, F M et al. (2022). "Modelling the size distribution of aggregated volcanic ash and implications for operational atmospheric dispersion modelling". In: *Atmospheric Chemistry and Physics* 22.5, pp. 3409–3431. DOI: 10.5194/acp-22-3409-2022.

Beckett, F M et al. (2024). *VAAC Model Setup Tables 2023*. Tech. rep. DOI: 10.5281/zenodo.10671098.

Behncke, B and M Neri (2003). "The July–August 2001 eruption of Mt. Etna (Sicily)". In: *Bulletin of Volcanology* 65, pp. 461–476. ISSN: 0258-8900. DOI: https://doi.org/10.1007/s00445-003-0274-1.

Behncke, B, M Neri, and A Nagay (2005). "Lava flow hazard at Mount Etna (Italy): New data from a GIS-based study". In: *Kinematics and Dynamics of Lava Flows*. Ed. by Michael Manga and Guido Ventura. Geological Society of America. ISBN: 9780813723969. DOI: 10.1130/0-8137-2396-5.189.

Bernard, B (2018). "Rapid hazard assessment of volcanic ballistic projectiles using long-exposure photographs: insights from the 2010 eruptions at Tungurahua volcano, Ecuador". In: *Volcanica* 1, pp. 49–61. DOI: 10.30909/vol.01.01.4961.

Bernard, B et al. (2022). "Forecasting and communicating the dispersion and fallout of ash during volcanic eruptions: lessons from the September 20, 2020 eruptive pulse at Sangay volcano, Ecuador". In: *Frontiers in Earth Science* Volume 10. ISSN: 2296-6463. DOI: https://doi.org/10.3389/feart.2022.912835.

Bernard, B et al. (2024). "Developing hazard scenarios from monitoring data, historical chronicles, and expert elicitation: a case study of Sangay volcano, Ecuador". In: *Bulletin of Volcanology* 86.8, p. 68. ISSN: 1432-0819. DOI: 10.1007/s00445-024-01754-4.

Bernard, J et al. (2014). "Pyroclastic flow erosion and bulking processes: comparing field-based vs. modeling results at Tungurahua volcano, Ecuador". In: *Bulletin of Volcanology* 76.9, p. 858. ISSN: 1432-0819. DOI: 10.1007/s00445-014-0858-y.

Bertin, D (2017). "3-D ballistic transport of ellipsoidal volcanic projectiles considering horizontal wind field and variable shape-dependent drag coefficients". In: *Journal of Geophysical Research: Solid Earth* 122.2, pp. 1126–1151. ISSN: 2169-9313. DOI: https://doi.org/10.1002/2016JB013320.

Beven, K J et al. (2018). "Epistemic uncertainties and natural hazard risk assessment–Part 1: A review of different natural hazard areas". In: *Natural Hazards and Earth System Sciences* 18.10, pp. 2741–2768. ISSN: 1561-8633. DOI: https://doi.org/10.5194/nhess-18-2741-2018.

Bevilacqua, A (2016). "Pyroclastic density current invasion maps". In: *Doubly Stochastic Models for Volcanic Hazard Assessment at Campi Flegrei Caldera*. Ed. by Andrea Bevilacqua. Pisa:

Scuola Normale Superiore, pp. 61–94. ISBN: 978-88-7642-577-6. DOI: 10.1007/978-88-7642-577-6{_}3.

Bevilacqua, A et al. (2019). "Probabilistic forecasting of plausible debris flows from Nevado de Colima (Mexico) using data from the Atenquique debris flow, 1955". In: *Natural Hazards Earth System Science* 19.4, pp. 791–820. ISSN: 1684-9981. DOI: 10.5194/nhess-19-791-2019.

Biass, S et al. (2016a). "Great Balls of Fire: A probabilistic approach to quantify the hazard related to ballistics — A case study at La Fossa volcano, Vulcano Island, Italy". In: *Journal of Volcanology and Geothermal Research* 325, pp. 1–14. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2016.06.006.

Biass, S et al. (2016b). "TephraProb: a Matlab package for probabilistic hazard assessments of tephra fallout". In: *Journal of Applied Volcanology* 5.1, p. 10. ISSN: 2191-5040. DOI: 10.1186/s13617-016-0050-5.

Biass, S et al. (2024). "The spatiotemporal evolution of compound impacts from lava flow and tephra fallout on buildings: lessons from the 2021 Tajogaite eruption (La Palma, Spain)". In: *Bulletin of Volcanology* 86.2, p. 10. ISSN: 1432-0819. DOI: 10.1007/s00445-023-01700-w.

Bilotta, G et al. (2012). "Sensitivity analysis of the MAGFLOW Cellular Automaton model for lava flow simulation". In: *Environmental Modelling & Software* 35, pp. 122–131. ISSN: 1364-8152. DOI: https://doi.org/10.1016/j.envsoft.2012.02.015.

Bilotta, G et al. (2019). "Influence of topographic data uncertainties and model resolution on the numerical simulation of lava flows". In: *Environmental Modelling & Software* 112, pp. 1–15. ISSN: 1364-8152. DOI: https://doi.org/10.1016/j.envsoft.2018.11.001.

Bird, D K, G Gísladóttir, and D Dominey-Howes (2011). "Different communities, different perspectives: issues affecting residents' response to a volcanic eruption in southern Iceland". In: *Bulletin of Volcanology* 73.9, pp. 1209–1227. ISSN: 1432-0819. DOI: 10.1007/s00445-011-0464-1.

Blackburn, E A, L Wilson, and R S J Sparks (1976). "Mechanisms and dynamics of strombolian activity". In: *Journal of the Geological Society* 132.4, pp. 429–440. ISSN: 0016-7649. DOI: https://doi.org/10.1144/gsjgs.132.4.0429.

Blake, D M (2016). "Impacts of volcanic ash on surface transportation networks: considerations for Auckland City, New Zealand." PhD thesis. Christchurch: University of Canterbury.

Blake, D M et al. (2017). "Investigating the consequences of urban volcanism using a scenario approach II: insights into transportation network damage and functionality". In: *Journal of Volcanology and Geothermal Research* 340, pp. 92–116. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2017.04.010.

Blong, R J (1984). *Volcanic hazards: a sourcebook on the effects of eruptions*. English. Sydney: Academic Press. ISBN: 9780121071806;0121071804;

Bollen, K et al. (2015). *Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science*. Tech. rep. URL: https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf.

Bonadonna, C (2010). *Ash Dispersal Forecast and Civil Aviation Workshop - Model Definition Document*. Tech. rep. URL: https://www.unige.ch/sciences/terre/CERG-C/application/files/5414/7732/2216/Model-Document-Geneva10.pdf.

Bonadonna, C, S Biass, and A Costa (2015). "Physical characterization of explosive volcanic eruptions based on tephra deposits: Propagation of uncertainties and sensitivity analysis". In: *Journal of Volcanology and Geothermal Research* 296, pp. 80–100. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2015.03.009.

Bonadonna, C and A Costa (2012). "Estimating the volume of tephra deposits: a new simple strategy". In: *Geology* 40.5, pp. 415–418. ISSN: 1943-2682. DOI: 10.1130/G32769.1.

— (2013). "Modeling tephra sedimentation from volcanic plumes". In: *Modeling Volcanic Processes*, pp. 173–202. DOI: 10.1017/CBO9781139021562.009.

Bonadonna, C, G G J Ernst, and R S J Sparks (1998). "Thickness variations and volume estimates of tephra fall deposits: the importance of particle Reynolds number". In: *Journal of Volcanology and Geothermal Research* 81.3, pp. 173–187. ISSN: 0377-0273. DOI: https://doi.org/10.1016/S0377-0273(98)00007-9.

Bonadonna, C and B F Houghton (2005). "Total grain-size distribution and volume of tephra-fall deposits". In: *Bulletin of Volcanology* 67.5, pp. 441–456. ISSN: 1432-0819. DOI: 10.1007/s00445-004-0386-2.

Bonadonna, C, G Macedonio, and R S J Sparks (2002). "Numerical modelling of tephra fallout associated with dome collapses and Vulcanian explosions: application to hazard assessment on Montserrat". In: *Geological Society, London, Memoirs* 21.1, pp. 517–537. DOI: 10.1144/GSL.MEM.2002.021.01.23.

Bonadonna, C and J C Phillips (2003). "Sedimentation from strong volcanic plumes". In: *Journal of Geophysical Research: Solid Earth* 108.B7. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2002JB002034.

Bonadonna, C, J C Phillips, and B F Houghton (2005). "Modeling tephra sedimentation from a Ruapehu weak plume eruption". In: *Journal of Geophysical Research: Solid Earth* 110.B8. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2004JB003515.

Bonadonna, C et al. (2005). "Probabilistic modeling of tephra dispersal: Hazard assessment of a multiphase rhyolitic eruption at Tarawera, New Zealand". English. In: *Journal of Geophysical Research* 110.B3. DOI: 10.1029/2003JB002896.

Bonadonna, C et al. (2011a). "Ash Dispersal Forecast and Civil Aviation Workshop - Model Benchmark Document". In: URL: https://vhub.org/resources/505.

Bonadonna, C et al. (2011b). "Tephra sedimentation during the 2010 Eyjafjallajökull eruption (Iceland) from deposit, radar, and satellite observations". In: *Journal of Geophysical Research: Solid Earth* 116.B12. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2011JB008462.

Bonadonna, C et al. (2012). "Future developments in modelling and monitoring of volcanic ash clouds: outcomes from the first IAVCEI-WMO workshop on Ash Dispersal Forecast and Civil Aviation ". English. In: *Bulletin of volcanology* 74.1, pp. 1–10. DOI: 10.1007/s00445-011-0508-6.

Bonadonna, C et al. (2013). *2nd IUGG-WMO workshop on Ash Dispersal Forecast and Civil Aviation - Consensual Document*. Tech. rep. URL: https://theghub.org/resources/3343.

Bonadonna, C et al. (2015). "Tephra Dispersal and Sedimentation". English. In: *The Encyclopedia of Volcanoes*. Second. Elsevier Inc, pp. 587–597. ISBN: ; DOI: 10.1016/B978-0-12-385938-9.00033-X.

Bonasia, R et al. (2010). "Numerical inversion and analysis of tephra fallout deposits from the 472 AD sub-Plinian eruption at Vesuvius (Italy) through a new best-fit procedure". In:

Journal of Volcanology and Geothermal Research 189.3-4, pp. 238–246. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2009.11.009.

Bonasia, R et al. (2011). "Tephra fallout hazard assessment for a Plinian eruption scenario at Volcán de Colima (Mexico)". In: *Journal of Volcanology and Geothermal Research* 203.1, pp. 12–22. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2011.03.006.

Bonasia, R et al. (2012). "Numerical simulation of tephra transport and deposition of the 1982 El Chichón eruption and implications for hazard assessment". In: *Journal of Volcanology and Geothermal Research* 231-232, pp. 39–49. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2012.04.006.

Bosa, A et al. (2021). "Tracking secondary lahar flow paths and characterizing pulses and surges using infrasound array networks at Volcán de Fuego, Guatemala". In: *Volcanica* 4.2, pp. 239–256. ISSN: 2610-3540. DOI: https://doi.org/10.30909/vol.04.02.239256.

Bosa, A R et al. (2024). "Dynamics of rain-triggered lahars and destructive power inferred from seismo-acoustic arrays and time-lapse camera correlation at Volcán de Fuego, Guatemala". In: *Natural Hazards*. ISSN: 1573-0840. DOI: 10.1007/s11069-024-06926-1.

Boybeyi, Z and S Raman (1995). "Simulation of elevated long-range plume transport using a mesoscale meteorological model". In: *Atmospheric Environment* 29.16, pp. 2099–2111. ISSN: 1352-2310. DOI: https://doi.org/10.1016/1352-2310(94)00288-V.

Branca, S and P D Carlo (2005). "Types of eruptions of Etna volcano AD 1670–2003: implications for short-term eruptive behaviour". In: *Bulletin of Volcanology* 67.8, pp. 732–742. ISSN: 1432-0819. DOI: 10.1007/s00445-005-0412-z.

Branney, M J and B P Kokelaar (1992). "A reappraisal of ignimbrite emplacement: progressive aggradation and changes from particulate to non-particulate flow during emplacement of high-grade ignimbrite". In: *Bulletin of Volcanology* 54, pp. 504–520. ISSN: 0258-8900. DOI: https://doi.org/10.1007/BF00301396.

— (2002). "Pyroclastic density currents and the sedimentation of ignimbrites". In: Geological Society of London. ISBN: 1862391246. DOI: https://doi.org/10.1144/GSL.MEM.2003.027.

Breard, E C P et al. (2020). "The basal friction coefficient of granular flows with and without excess pore pressure: implications for pyroclastic density currents, water-rich debris flows, and

rock and submarine avalanches". In: *Journal of Geophysical Research: Solid Earth* 125.12. ISSN: 2169-9313. DOI: https://doi.org/10.1029/2020JB020203.

Brehmer, J R et al. (2024). "Enhancing the Statistical Evaluation of Earthquake Forecasts—An Application to Italy". In: *Seismological Research Letters* 96.3, pp. 1966–1988. ISSN: 0895-0695. DOI: 10.1785/0220240209.

Broniatowski, D A. (2021). *Psychological Foundations of Explainability and Interpretability in Artificial Intelligence*. Tech. rep. Gaithersburg, MD: National Institute of Standards and Technology. DOI: 10.6028/NIST.IR.8367.

Brooks, R J and A M Tobias (1996). "Choosing the best model: Level of detail, complexity, and model performance". In: *Mathematical and Computer Modelling* 24.4, pp. 1–14. ISSN: 0895-7177. DOI: https://doi.org/10.1016/0895-7177(96)00103-3.

Brown, R J and G D M Andrews (2015). "Deposits of pyroclastic density currents". In: *The Encyclopedia of Volcanoes*. Ed. by H Sigurdsson. 2nd edn. Elsevier, pp. 631–648. DOI: https://doi.org/10.1016/B978-0-12-385938-9.00036-5.

Brown, S K et al. (2015). "Global volcanic hazard and risk". In: *Global Volcanic Hazards and Risk*. Ed. by Susan C Loughlin et al. Cambridge: Cambridge University Press, pp. 81–172. ISBN: 9781107111752. DOI: DOI:10.1017/CBO9781316276273.004.

Brown, S K et al. (2017). "Volcanic fatalities database: analysis of volcanic threat with distance and victim classification". In: *Journal of Applied Volcanology* 6.1, p. 15. ISSN: 2191-5040. DOI: 10.1186/s13617-017-0067-4.

Buckland, H M et al. (2022). "Modelling the transport and deposition of ash following a magnitude 7 eruption: the distal Mazama tephra". In: *Bulletin of Volcanology* 84.9, p. 87. ISSN: 1432-0819. DOI: 10.1007/s00445-022-01593-1.

Burgisser, A and G W Bergantz (2002). "Reconciling pyroclastic flow and surge: the multiphase physics of pyroclastic density currents". In: *Earth and Planetary Science Letters* 202.2, pp. 405–418. ISSN: 0012-821X. DOI: https://doi.org/10.1016/S0012-821X(02)00789-6.

Bursik, M (2001). "Effect of wind on the rise height of volcanic plumes". In: *Geophysical Research Letters* 28.18, pp. 3621–3624. ISSN: 0094-8276. DOI: https://doi.org/10.1029/2001GL013393.

Bursik, M I et al. (1992). "Sedimentation of tephra by volcanic plumes: I. Theory and its comparison with a study of the Fogo A plinian deposit, Sao Miguel (Azores)". In: *Bulletin of Volcanology* 54.4, pp. 329–344. ISSN: 1432-0819. DOI: 10.1007/BF00301486.

Butts, M B et al. (2004). "An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation". In: *Journal of Hydrology* 298.1, pp. 242–266. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2004.03.042.

Byun, D and K L Schere (2006). "Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System". In: *Applied Mechanics Reviews* 59.2, pp. 51–77. ISSN: 0003-6900. DOI: 10.1115/1.2128636.

Calvari, S et al. (2010). "The 2007 Stromboli eruption: Event chronology and effusion rates using thermal infrared data". In: *Journal of Geophysical Research: Solid Earth* 115.B4. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2009JB006478.

Calvari, S et al. (2018). "Paroxysmal Explosions, Lava Fountains and Ash Plumes at Etna Volcano: Eruptive Processes and Hazard Implications". In: *Frontiers in Earth Science* 6. ISSN: 2296-6463. DOI: 10.3389/feart.2018.00107.

Cao, Z et al. (2021). "Simulating the Transport and Dispersal of Volcanic Ash Clouds With Initial Conditions Created by a 3D Plume Model ". In: *Frontiers in Earth Science* 9. DOI: https://doi.org/10.3389/feart.2021.704797.

Cappello, A et al. (2016a). "Lava flow hazard modeling during the 2014–2015 Fogo eruption, Cape Verde". In: *Journal of Geophysical Research: Solid Earth* 121.4, pp. 2290–2303. ISSN: 2169-9313. DOI: https://doi.org/10.1002/2015JB012666.

Cappello, A et al. (2016b). "MAGFLOW: a physics-based model for the dynamics of lava-flow emplacement". In: *Geological Society, London, Special Publications* 426.1, pp. 357–373. ISSN: 1786204363. DOI: https://doi.org/10.1144/sp426.16.

Cappello, A et al. (2019). "Satellite-driven modeling approach for monitoring lava flow hazards during the 2017 Etna eruption". In: *Annals of Geophysics* 62.2. ISSN: 2037-416X. DOI: https://doi.org/10.4401/ag-7792.

Capponi, A et al. (2016). "Recycled ejecta modulating Strombolian explosions". In: *Bulletin of Volcanology* 78, pp. 1–13. ISSN: 0258-8900. DOI: https://doi.org/10.1007/s00445-016-1001-z.

Capponi, A et al. (2022). "Refining an ensemble of volcanic ash forecasts using satellite retrievals: Raikoke 2019". In: *Atmospheric Chemistry and Physics* 22.9, pp. 6115–6134. DOI: 10.5194/acp-22-6115-2022.

Capra, L et al. (2011). "The importance of digital elevation model resolution on granular flow simulations: a test case for Colima volcano using TITAN2D computational routine". In: *Natural Hazards* 59.2, pp. 665–680. ISSN: 1573-0840. DOI: 10.1007/s11069-011-9788-6.

Carey, S and M Bursik (2000). "Volcanic Plumes". In: *Encyclopedia of Volcanoes*. Ed. by H Sigurdson et al. San Diego, California: Academic Press. Chap. 32, pp. 527–554. DOI: https://doi.org/10.1016/B978-0-12-385938-9.00032-8.

Carey, S and H Sigurdsson (1989). "The intensity of plinian eruptions". In: *Bulletin of Volcanology* 51, pp. 28–40. ISSN: 0258-8900. DOI: https://doi.org/10.1007/BF01086759.

Carey, S and R S J Sparks (1986). "Quantitative models of the fallout and dispersal of tephra from volcanic eruption columns". In: *Bulletin of Volcanology* 48.2, pp. 109–125. ISSN: 1432-0819. DOI: 10.1007/BF01046546.

Carey, S et al. (1990). "Variations in column height and magma discharge during the May 18, 1980 eruption of Mount St. Helens". In: *Journal of Volcanology and Geothermal Research* 43.1, pp. 99–112. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(90)90047-J.

Carey, S N and H Sigurdsson (1982). "Influence of particle aggregation on deposition of distal tephra from the May 18, 1980, eruption of Mount St. Helens volcano." In: *Journal of Geophysical Research* 87.B8, pp. 7061–7072. ISSN: 01480227. DOI: 10.1029/JB087iB08p07061.

Carn, S A et al. (2009). "The unexpected awakening of Chaitén Volcano, Chile". In: *Eos, Transactions, American Geophysical Union* 90.24, pp. 205–206. DOI: 10.1029/2009EO240001.

Casadevall, T (1982). *lahar originating in the Mount St. Helen*. URL: https://www.usgs.gov/media/images/lahar-dark-deposit-snow-originating-mount-st-helen.

Casadevall, T J (1994). *Volcanic ash and aviation safety; proceedings of the First international symposium on Volcanic ash and aviation safety*. ENGLISH. Tech. rep. DOI: 10.3133/b2047. URL: http://pubs.er.usgs.gov/publication/b2047.

Cashman, K and J Biggs (2014). "Common processes at unique volcanoes—A volcanological conundrum". In: *Frontiers in Earth Science* 2. DOI: 10.3389/feart.2014.00028.

Cerminara, M, T Esposti Ongaro, and L C Berselli (2016). "ASHEE-1.0 a compressible, equilibrium - Eulerian model for volcanic ash plumes". English. In: *Geoscientific Model Development* 9.2, pp. 697–730. DOI: 10.5194/gmd-9-697-2016.

Chai, T et al. (2017). "Improving volcanic ash predictions with the HYSPLIT dispersion model by assimilating MODIS satellite retrievals". In: *Atmospheric Chemistry and Physics* 17.4, pp. 2865–2879. ISSN: 1680-7316. DOI: https://doi.org/10.5194/acp-17-2865-2017.

Charbonnier, S J and R Gertisser (2012). "Evaluation of geophysical mass flow models using the 2006 block-and-ash flows of Merapi Volcano, Java, Indonesia: Towards a short-term hazard assessment tool". In: *Journal of Volcanology and Geothermal Research* 231-232, pp. 87–108. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2012.02.015.

Cheng, W Y Y et al. (2017). "Short-term wind forecast of a data assimilation/weather forecasting system with wind turbine anemometer measurement assimilation". In: *Renewable Energy* 107, pp. 340–351. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2017.02.014.

Chevrel, M O, H Pinkerton, and A J L Harris (2019). "Measuring the viscosity of lava in the field: A review". In: *Earth-Science Reviews* 196. ISSN: 0012-8252. DOI: https://doi.org/10.1016/j.earscirev.2019.04.024.

Chevrel, M O et al. (2018). "The viscosity of pāhoehoe lava: In situ syn-eruptive measurements from Kilauea, Hawaii". In: *Earth and Planetary Science Letters* 493, pp. 161–171. ISSN: 0012-821X. DOI: https://doi.org/10.1016/j.epsl.2018.04.028.

Chouet, B, N Hamisevicz, and T R McGetchin (1974). "Photoballistics of volcanic jet activity at Stromboli, Italy". In: *Journal of Geophysical Research* 79.32, pp. 4961–4976. ISSN: 0148-0227. DOI: https://doi.org/10.1029/JB079i032p04961.

Clark, M P and D Kavetski (2010). "Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes". In: *Water Resources Research* 46.10. ISSN: 0043-1397. DOI: https://doi.org/10.1029/2009WR008894.

Clarke, A B et al. (2002). "Transient dynamics of vulcanian explosions and column collapse". In: *Nature* 415.6874, pp. 897–901. ISSN: 1476-4687. DOI: 10.1038/415897a.

Clarke, B A et al. (2020). "Probabilistic Volcanic Hazard Assessment for Pyroclastic Density Currents From Pumice Cone Eruptions at Aluto Volcano, Ethiopia". In: *Frontiers in Earth Science* 8. ISSN: 2296-6463. DOI: https://doi.org/10.3389/feart.2020.00348.

Cochran, W G (1977). *Sampling techniques*. 3rd ed. New York: John Wiley & Sons. ISBN: 8126515244.

Collini, E et al. (2013). "Volcanic ash forecast during the June 2011 Cordón Caulle eruption". In: *Natural Hazards* 66.2, pp. 389–412. ISSN: 1573-0840. DOI: 10.1007/s11069-012-0492-y.

Connor, L J and C B Connor (2006). "Inversion is the key to dispersion: understanding eruption dynamics by inverting tephra fallout". In: *Statistics in Volcanology* 1. Ed. by H M Mader et al., p. 0. DOI: 10.1144/IAVCEI001.18.

Connor, L J, C B Connor, and C Bonadonna (2008). *Forecasting tephra dispersion using TEPHRA2*. URL: https://vhub.org/resources/574/download/Tephra2_Manual.pdf.

Connor, L J, C B Connor, and A Saballos (2011). "Tephra2 Users Manual". In: *University of South Florida, Tampa, FL, accessed Sept* 24, p. 2017. URL: https://theghub.org/resources/756/download/Tephra2_Users_Manual.pdf.

Connor, L J and M Hagdorn (2018). *Tephra2*. URL: https://github.com/geoscience-community-codes/tephra2.git.

Connor, L J et al. (2012). "Probabilistic approach to modeling lava flow inundation: a lava flow hazard assessment for a nuclear facility in Armenia". In: *Journal of Applied Volcanology* 1.1, p. 3. ISSN: 2191-5040. DOI: 10.1186/2191-5040-1-3.

Cordonnier, B, E Lev, and F Garel (2015). "Bechmarking lava-flow models". In: *Geological Society, London, Special Publications* 426. DOI: 10.1144/SP426.7.

Corradini, S et al. (2016). "A multi-sensor approach for volcanic ash cloud retrieval and eruption characterization: The 23 November 2013 Etna lava fountain". In: *Remote Sensing* 8.1, p. 58. ISSN: 2072-4292. DOI: https://doi.org/10.3390/rs8010058.

Costa, A and G Macedonio (2005). "Computational modeling of lava flows: A review". In: *Special papers-Geological Society of America* 396, p. 209. ISSN: 0072-1077. DOI: https://doi.org/10.1130/0-8137-2396-5.209.

Costa, A, G Macedonio, and A Folch (2006). "A three-dimensional Eulerian model for transport and deposition of volcanic ashes". In: *Earth and Planetary Science Letters* 241.3, pp. 634–647. ISSN: 0012-821X. DOI: https://doi.org/10.1016/j.epsl.2005.11.019.

Costa, A, L Pioli, and C Bonadonna (2016). "Assessing tephra total grain-size distribution: Insights from field data analysis". In: *Earth and Planetary Science Letters* 443, pp. 90–107. ISSN: 0012-821X. DOI: https://doi.org/10.1016/j.epsl.2016.02.040.

Costa, A et al. (2008). "Tephra fallout hazard assessment at the Campi Flegrei caldera (Italy)". In: *Bulletin of Volcanology* 71.3, p. 259. ISSN: 1432-0819. DOI: 10.1007/s00445-008-0220-3.

Costa, A et al. (2016). "Results of the eruptive column model inter-comparison study". In: *Journal of Volcanology and Geothermal Research* 326, pp. 2–25. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2016.01.017.

Costa, G et al. (2023). "Modeling the trajectories of ballistics in the summit area of Mt. Etna (Italy) during the 2020–2022 sequence of lava fountains". In: *Geosciences* 13.5, p. 145. ISSN: 2076-3263. DOI: https://doi.org/10.3390/geosciences13050145.

Craig, H et al. (2016). "Impacts to agriculture and critical infrastructure in Argentina after ashfall from the 2011 eruption of the Cordón Caulle volcanic complex: an assessment of published damage and function thresholds". In: *Journal of Applied Volcanology* 5.1, p. 7. ISSN: 2191-5040. DOI: 10.1186/s13617-016-0046-1.

Crawford, A M et al. (2016). "Initializing HYSPLIT with satellite observations of volcanic ash: A case study of the 2008 Kasatochi eruption". In: *Journal of Geophysical Research: Atmospheres* 121.18. ISSN: 2169-897X. DOI: https://doi.org/10.1002/2016JD024779.

Crawford, M H et al. (2018). "Risk modelling as a tool to support natural hazard risk management in New Zealand local government". English. In: *International Journal of Disaster Risk Reduction* 28, pp. 610–619. ISSN: 2212-4209. DOI: 10.1016/j.ijdrr.2018.01.011.

Crisci, G M et al. (1986). "Lava Flow Simulation Bv A Discrete Cellular Model: First Implementation". In: *International Journal of Modelling and Simulation* 6.4, pp. 137–140. ISSN: 0228-6203. DOI: 10.1080/02286203.1986.11759975.

Crisci, G M et al. (2004). "The simulation model SCIARA: the 1991 and 2001 lava flows at Mount Etna". In: *Journal of Volcanology and Geothermal Research* 132.2, pp. 253–267. ISSN: 0377-0273. DOI: https://doi.org/10.1016/S0377-0273(03)00349-4.

Cronin, S J et al. (1997). "Impact of Ruapehu ash fall on soil and pasture nutrient status 1. October 1995 eruptions". In: *New Zealand Journal of Agricultural Research* 40.3, pp. 383–395. ISSN: 0028-8233. DOI: https://doi.org/10.1080/00288233.1997.9513260.

Cronin, S J et al. (1998). "Agronomic impact of tephra fallout from the 1995 and 1996 Ruapehu Volcano eruptions, New Zealand". In: *Environmental Geology* 34.1, pp. 21–30. ISSN: 1432-0495. DOI: 10.1007/s002540050253.

Cronin, S J et al. (2003). "Environmental hazards of fluoride in volcanic ash: a case study from Ruapehu volcano, New Zealand". In: *Journal of Volcanology and Geothermal Research* 121.3, pp. 271–291. ISSN: 0377-0273. DOI: https://doi.org/10.1016/S0377-0273(02)00465-1.

Cronin, S J et al. (2021). "The geological history and hazards of a long-lived stratovolcano, Mt. Taranaki, New Zealand". In: *New Zealand Journal of Geology and Geophysics*, pp. 1–23. ISSN: 0028-8306. DOI: 10.1080/00288306.2021.1895231.

Crosweller, H S et al. (2012). "Global database on large magnitude explosive volcanic eruptions (LaMEVE)". In: *Journal of Applied Volcanology* 1.1, p. 4. ISSN: 2191-5040. DOI: 10.1186/2191-5040-1-4.

Crowe, C et al. (1998). *Multiphase flows with Droplets and Particles*. 2nd ed. Vol. 906. CRC Press. DOI: https://doi.org/10.1201/b11103.

Cukier, R I, H B Levine, and K E Shuler (1978). "Nonlinear sensitivity analysis of multiparameter model systems". In: *Journal of Computational Physics* 26.1, pp. 1–42. ISSN: 0021-9991.

D'Ambrosio, D et al. (2005). "Parallel genetic algorithms for calibrating cellular automata models: Application to lava flows". In: *NUOVO CIMENTO-SOCIETA ITALIANA DI FISICA SEZIONE C* 28.2, p. 115. ISSN: 0390-5551. DOI: 10.1393/ncc/i2005-10178-1.

D'Ambrosio, D et al. (2013). "Efficient application of GPGPU for lava flow hazard mapping". In: *The Journal of Supercomputing* 65.2, pp. 630–644. ISSN: 1573-0484. DOI: 10.1007/s11227-013-0949-0.

Daag, A S (2003). "Modelling the erosion of the pyroclastic flow deposits and the occurrences of Lahars at Mt. Pinatubo, Philippines". PhD thesis. Enschede: International Institute for Geo-Information Science and Earth Observation.

Dacre, H F et al. (2016). "How accurate are volcanic ash simulations of the 2010 Eyjafjallajökull eruption?" In: *Journal of Geophysical Research: Atmospheres* 121.7, pp. 3534–3547. ISSN: 2169-897X. DOI: https://doi.org/10.1002/2015JD024265.

Dalbey, K et al. (2008). "Input uncertainty propagation methods and hazard mapping of geophysical mass flows". In: *Journal of Geophysical Research: Solid Earth* 113.5. DOI: 10.1029/2006JB004471.

Damaschke, M, S J Cronin, and M S Bebbington (2018). "A volcanic event forecasting model for multiple tephra records, demonstrated on Mt. Taranaki, New Zealand". In: *Bulletin of Volcanology* 80, pp. 1–14. ISSN: 0258-8900. DOI: https://doi.org/10.1007/s00445-017-1184-y.

Damaschke, M et al. (2017). "A 30,000 yr high-precision eruption history for the andesitic Mt. Taranaki, North Island, New Zealand". In: *Quaternary Research* 87.1, pp. 1–23. ISSN: 0033-5894. DOI: 10.1017/qua.2016.11.

Dare, R A, D H Smith, and M J Naughton (2016). "Ensemble Prediction of the Dispersion of Volcanic Ash from the 13 February 2014 Eruption of Kelut, Indonesia". English. In: *Journal of Applied Meteorology and Climatology* 55.1, pp. 61–78. DOI: https://doi.org/10.1175/JAMC-D-15-0079.1.

Das, M et al. (2025). "The communication of volcano information in New Zealand – a narrative review". In: *New Zealand Journal of Geology and Geophysics*, pp. 1–18. DOI: 10.1080/00288306.2025.2454558.

De Beni, E, M Cantarero, and A Messina (2019). "UAVs for volcano monitoring: A new approach applied on an active lava flow on Mt. Etna (Italy), during the 27 February–02 March 2017 eruption". In: *Journal of Volcanology and Geothermal Research* 369, pp. 250–262. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2018.12.001.

De la Cruz, R et al. (2016). "Optimization of atmospheric transport models on HPC platforms". In: *Computers & Geosciences* 97, pp. 30–39. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2016.08.019.

De'Michieli Vitturi, M and F Pardini (2021). "PLUME-MoM-TSM 1.0.0: a volcanic column and umbrella cloud spreading model". In: *Geoscientific Model Development* 14, pp. 1345–1377. DOI: 10.5194/gmd-14-1345-2021.

De'Michieli Vitturi, M and S Tarquini (2018). "MrLavaLoba: A new probabilistic model for the simulation of lava flows as a settling process". In: *Journal of Volcanology and Geothermal Research* 349, pp. 323–334. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2017.11.016.

De'Michieli Vitturi, M et al. (2010). "Lagrangian modeling of large volcanic particles: Application to Vulcanian explosions". In: *Journal of Geophysical Research: Solid Earth* 115.B8. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2009JB007111.

De'Michieli Vitturi, M et al. (2019). "IMEX_SfloW2D 1.0: a depth-averaged numerical flow model for pyroclastic avalanches". In: *Geoscientific Model Development* 12.1, pp. 581–595. DOI: 10.5194/gmd-12-581-2019.

Degruyter, W and C Bonadonna (2012). "Improving on mass flow rate estimates of volcanic eruptions". In: *Geophysical Research Letters* 39.16. DOI: 10.1029/2012GL052566.

Delannay, R et al. (2017). "Granular and particle-laden flows: from laboratory experiments to field observations". In: *Journal of Physics D: Applied Physics* 50.5. ISSN: 0022-3727. DOI: 10.1088/1361-6463/50/5/053001.

Delbrel, J et al. (2025). "An investigation of changes to commercial aircraft flight paths during volcanic eruptions". In: *Journal of Applied Volcanology* 14.1, p. 2. ISSN: 2191-5040. DOI: 10.1186/s13617-025-00150-7.

Deligne, N I (2021). *Mass eruption rate, column height, and duration dataset for volcanic eruptions.* Tech. rep. GNS Science Report. DOI: 10.21420/P18W-7674.

Deng, F et al. (2019). "High-resolution DEM generation from spaceborne and terrestrial remote sensing data for improved volcano hazard assessment — A case study at Nevado del Ruiz, Colombia". In: *Remote Sensing of Environment* 233. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2019.111348.

Devenish, B J et al. (2012). "Sensitivity analysis of dispersion modeling of volcanic ash from Eyjafjallajökull in May 2010". In: *Journal of Geophysical Research: Atmospheres* 117.D20. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2011JD016782.

Dietterich, H R et al. (2015). "Diverting lava flows in the lab". In: *Nature Geoscience* 8.7, pp. 494–496. ISSN: 1752-0908. DOI: 10.1038/ngeo2470.

Dietterich, H R et al. (2017). "Benchmarking computational fluid dynamics models of lava flow simulation for hazard assessment, forecasting, and risk management". In: *Journal of Applied Volcanology* 6.1, p. 9. ISSN: 2191-5040. DOI: 10.1186/s13617-017-0061-x.

Dietterich, H R et al. (2021). "Lava effusion rate evolution and erupted volume during the 2018 Kīlauea lower East Rift Zone eruption". In: *Bulletin of Volcanology* 83.4, p. 25. ISSN: 1432-0819. DOI: 10.1007/s00445-021-01443-6.

Dioguardi, F and D Mele (2018). "PYFLOW_2.0: a computer program for calculating flow properties and impact parameters of past dilute pyroclastic density currents based on field data". In: *Bulletin of Volcanology* 80.3, p. 28. ISSN: 1432-0819. DOI: 10.1007/s00445-017-1191-z.

Doherty, J and S Christensen (2011). "Use of paired simple and complex models to reduce predictive bias and quantify uncertainty". In: *Water Resources Research* 47.12. ISSN: 0043-1397. DOI: https://doi.org/10.1029/2011WR010763.

Donoghue, S L et al. (1995). "Magma mingling in an andesite pyroclastic flow of the Pourahu Member, Ruapehu volcano, New Zealand". In: *Journal of Volcanology and Geothermal Research* 68.1, pp. 177–191. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(95)00012-J.

Doronzo, D M et al. (2011). "Numerical simulation of pyroclastic density currents using locally refined Cartesian grids". In: *Computers & Fluids* 44.1, pp. 56–67. ISSN: 0045-7930. DOI: https://doi.org/10.1016/j.compfluid.2010.12.006.

Dorta, D O et al. (2007). "Empirical modelling of the May 1998 small debris flows in Sarno (Italy) using LAHARZ". English. In: *Natural Hazards* 40.2, pp. 381–396. ISSN: 0921030X. DOI: http://dx.doi.org/10.1007/s11069-006-0035-5.

Douglas, J (2007). "Physical vulnerability modelling in natural hazard risk assessment". In: *Natural Hazards and Earth System Sciences* 7.2, pp. 283–288. ISSN: 1561-8633. DOI: DOI:10.5194/nhess-7-283-2007.

Doyle, E E H et al. (2014). "Uncertainty and decision making: Volcanic crisis scenarios". In: *International Journal of Disaster Risk Reduction* 10, pp. 75–101. ISSN: 2212-4209. DOI: https://doi.org/10.1016/j.ijdrr.2014.07.006.

Dragoni, M, M Bonafede, and E Boschi (1986). "Downslope flow models of a Bingham liquid: Implications for lava flows". In: *Journal of Volcanology and Geothermal Research* 30.3, pp. 305–325. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(86)90059-4.

Draxler, R and G Hess (1997). "Description of the HYSPLIT_4 modelling system". In: *NOAA Tech. Mem. ERL ARL-224*.

— (1998). "An overview of the HYSPLIT_4 modeling system for trajectories, dispersion, and deposition". In: *Australian Meteorological Magazine* 47.4, pp. 295–308. ISSN: 00049743.

Draxler, R R (1999). *HYSPLIT-4 user's guide*. English. United States. URL: https://www.osti.gov/biblio/20005929%20pbd:%20Jun%201999.

Dufek, J (2016). "The fluid mechanics of pyroclastic density currents". In: *Annual Review of Fluid Mechanics* 48.1, pp. 459–485. ISSN: 0066-4189. DOI: 10.1146/annurev-fluid-122414-034252.

Dumaisnil, C et al. (2010). "Hydraulic, physical and rheological characteristics of rain-triggered lahars at Semeru volcano, Indonesia". In: *Earth Surface Processes and Landforms* 35.13, pp. 1573–1590. ISSN: 0197-9337. DOI: https://doi.org/10.1002/esp.2003.

Dunant, A (2021). "Are We Missing the Target? A Bias-Variance Perspective on Multi-Hazard Risk Assessment". In: *Frontiers in Earth Science*. DOI: https://doi.org/10.3389/feart.2021.685301.

Durant, A J (2015). "RESEARCH FOCUS: Toward a realistic formulation of fine-ash lifetime in volcanic clouds". In: *Geology* 43.3, pp. 271–272. ISSN: 0091-7613. DOI: 10.1130/focus032015.1.

Dürig, T, L S Schmidt, and F Dioguardi (2023). "Optimizing mass eruption rate estimates by combining simple plume models". In: *Frontiers in Earth Science* 11. ISSN: 2296-6463. DOI: https://doi.org/10.3389/feart.2023.1250686.

Dürig, T et al. (2018). "REFIR- A multi-parameter system for near real-time estimates of plume-height and mass eruption rate during explosive eruptions". In: *Journal of Volcanology and*

Geothermal Research 360, pp. 61–83. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2018.07.003.

Dürig, T et al. (2023). "Quantifying the Effect of Wind on Volcanic Plumes: Implications for Plume Modeling". In: *Journal of Geophysical Research: Atmospheres* 128.2. ISSN: 2169-897X. DOI: https://doi.org/10.1029/2022JD037781.

Edwards, M J et al. (2017). "Evolution of a small hydrothermal eruption episode through a mud pool of varying depth and rheology, White Island, NZ". In: *Bulletin of Volcanology* 79, pp. 1–16. ISSN: 0258-8900. DOI: https://doi.org/10.1007/s00445-017-1100-5.

Ehrendorfer, M (1997). "Predicting the uncertainty of numerical weather forecasts: A review". In: *Meterorologische Zeitschrift* 6, pp. 147–183. ISSN: 0941-2948. DOI: 10.1127/metz/6/1997/147.

Elissondo, M et al. (2016). "Chronology and impact of the 2011 Cordón Caulle eruption, Chile". In: *Natural Hazards and Earth System Sciences* 16.3, pp. 675–704. ISSN: 1561-8633. DOI: doi:10.5194/nhess-16-675-2016.

Engwell, S et al. (2021). "Near-real-time volcanic cloud monitoring: insights into global explosive volcanic eruptive activity through analysis of Volcanic Ash Advisories". In: *Bulletin of Volcanology* 83.2, p. 9. ISSN: 1432-0819. DOI: 10.1007/s00445-020-01419-y.

Engwell, S et al. (2024). "Characterising, quantifying, and accessing eruption source parameters of explosive volcanic eruptions for operational simulation of tephra dispersion: a current view and future perspectives". In: *Bulletin of Volcanology* 86.7, p. 67. ISSN: 1432-0819. DOI: 10.1007/s00445-024-01706-y.

Engwell, S L, R S J Sparks, and W P Aspinall (2013). "Quantifying uncertainties in the measurement of tephra fall thickness". In: *Journal of Applied Volcanology* 2.1. DOI: 10.1186/2191-5040-2-5.

Esposti Ongaro, T et al. (2007). "A parallel multiphase flow code for the 3D simulation of explosive volcanic eruptions". In: *Parallel Computing* 33.7, pp. 541–560. ISSN: 0167-8191. DOI: https://doi.org/10.1016/j.parco.2007.04.003.

Esposti Ongaro, T et al. (2020). "A framework for validation and benchmarking of pyroclastic current models". In: *Bulletin of Volcanology* 82.6, p. 51. ISSN: 1432-0819. DOI: 10.1007/s00445-020-01388-2.

Evans, M R et al. (2013). "Do simple models lead to generality in ecology?" In: *Trends in Ecology & Evolution* 28.10, pp. 578–583. ISSN: 0169-5347. DOI: https://doi.org/10.1016/j.tree.2013.05.022.

Fauziati, S and K Watannabe (2010). "Ontology of Volcano System and Volcanic Hazards Assessment". In: *International Journal of Geoinformatics* 6.4, pp. 49–62. ISSN: 1686-6576.

Favalli, M et al. (2005). "Forecasting lava flow paths by a stochastic approach". In: *Geophysical Research Letters* 32.3. ISSN: 0094-8276. DOI: https://doi.org/10.1029/2004GL021718.

Favalli, M et al. (2009). "Lava flow hazard at Nyiragongo volcano, D.R.C." In: *Bulletin of Volcanology* 71.4, pp. 363–374. ISSN: 1432-0819. DOI: 10.1007/s00445-008-0233-y.

Fee, D et al. (2017). "Eruption mass estimation using infrasound waveform inversion and ash and gas measurements: Evaluation at Sakurajima Volcano, Japan". In: *Earth and Planetary Science Letters* 480, pp. 42–52. ISSN: 0012-821X. DOI: https://doi.org/10.1016/j.epsl.2017.09.043.

Felpeto, A, J Martí, and R Ortiz (2007). "Automatic GIS-based system for volcanic hazard assessment". In: *Journal of Volcanology and Geothermal Research* 166.2, pp. 106–116. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2007.07.008.

Fernandez-Turiel, J L et al. (2019). *Dataset of SEM images, modelled isopach map and topographic profiles, radiocarbon ages and data of parameters of Tephra2 and AshCalc codes of Holocene volcanic ashes of NW Argentina*. DOI: 10.20350/digitalCSIC/8629. URL: https://core.ac.uk/outputs/195572515/?source=2.

Fierstein, J and M Nathenson (1992). "Another look at the calculation of fallout tephra volumes". In: *Bulletin of Volcanology* 54.2, pp. 156–167. ISSN: 1432-0819. DOI: 10.1007/BF00278005.

Finger, D et al. (2015). "The value of multiple data set calibration versus model complexity for improving the performance of hydrological models in mountain catchments". In: *Water Resources Research* 51.4, pp. 1939–1958. ISSN: 0043-1397. DOI: https://doi.org/10.1002/2014WR015712.

Fisher, R V (1966). "Mechanism of deposition from pyroclastic flows". In: *American Journal of Science* 264.5, pp. 350–363. DOI: 10.2475/ajs.264.5.350.

Fitzgerald, R H et al. (2014). "The application of a calibrated 3D ballistic trajectory model to ballistic hazard assessments at Upper Te Maari, Tongariro". In: *Journal of Volcanology and Geothermal Research* 286, pp. 248–262. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2014.04.006.

Florinsky, I V (2012). "Chapter 3 - Digital Elevation Models". In: *Digital Terrain Analysis in Soil Science and Geology*. Ed. by Igor V Florinsky. Boston: Academic Press, pp. 31–41. ISBN: 978-0-12-385036-2. DOI: https://doi.org/10.1016/B978-0-12-385036-2.00003-1.

Flynn, I T W, D B Williams, and M S Ramsey (2024). "Quantifying volumes of volcanic deposits using time-averaged ASTER digital elevation models". In: *Science of Remote Sensing* 10. ISSN: 2666-0172. DOI: https://doi.org/10.1016/j.srs.2024.100179.

Folch, A (2012). "A review of tephra transport and dispersal models: Evolution, current status, and future perspectives". In: *Journal of Volcanology and Geothermal Research* 235-236, pp. 96–115. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2012.05.020.

Folch, A, A Costa, and S Basart (2012). "Validation of the FALL3D ash dispersion model using observations of the 2010 Eyjafjallajökull volcanic ash clouds". In: *Atmospheric Environment* 48, pp. 165–183. ISSN: 1352-2310. DOI: https://doi.org/10.1016/j.atmosenv.2011.06.072.

Folch, A, A Costa, and G Macedonio (2009). "FALL3D: A computational model for transport and deposition of volcanic ash". In: *Computers & Geosciences* 35.6, pp. 1334–1342. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2008.08.008.

— (2016). "FPLUME-1.0: An integral volcanic plume model accounting for ash aggregation". In: *Geoscientific Model Development* 9.1, pp. 431–450. ISSN: 1991-9603. DOI: 10.5194/gmd-9-431-2016.

Folch, A, L Mingari, and A T Prata (2022). "Ensemble-Based Forecast of Volcanic Clouds Using FALL3D-8.1". In: *Frontiers in Earth Science* 9. ISSN: 2296-6463. DOI: https://doi.org/10.3389/feart.2021.741841.

Folch, A et al. (2008). "An automatic procedure to forecast tephra fallout". In: *Journal of Volcanology and Geothermal Research* 177.4, pp. 767–777. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2008.01.046.

Folch, A et al. (2010). "A model for wet aggregation of ash particles in volcanic plumes and clouds: 2. Model application". In: *Journal of Geophysical Research: Solid Earth* 115.B9. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2009JB007176.

Folch, A et al. (2020). "FALL3D-8.0: a computational model for atmospheric transport and deposition of particles, aerosols and radionuclides – Part 1: Model physics and numerics". English. In: *Geoscientific Model Development* 13.3, pp. 1431–1458. ISSN: 1991962X. DOI: http://dx.doi.org/10.5194/gmd-13-1431-2020.

Folch, A et al. (2023). *FALL3D*. DOI: 10.5281/zenodo.7802290.

Forbes, L et al. (2003). "Volcanic ash and respiratory symptoms in children on the island of Montserrat, British West Indies". In: *Occupational and Environmental Medicine* 60.3, p. 207. DOI: 10.1136/oem.60.3.207.

Francis, P et al. (2000). "The 1995–1998 eruption of the Soufriére Hills volcano, Montserrat, WI". In: *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 358.1770, pp. 1619–1637. DOI: 10.1098/rsta.2000.0607.

Freret-Lorgeril, V et al. (2018). "Mass Eruption Rates of Tephra Plumes During the 2011–2015 Lava Fountain Paroxysms at Mt. Etna From Doppler Radar Retrievals". In: *Frontiers in Earth Science* 6. ISSN: 2296-6463. DOI: 10.3389/feart.2018.00073.

Freret-Lorgeril, V et al. (2021). "Examples of multi-sensor determination of eruptive source parameters of explosive events at Mount Etna". In: *Remote Sensing* 13.11, p. 2097. ISSN: 2072-4292. DOI: https://doi.org/10.3390/rs13112097.

Friedman, J H (1997). "On Bias, Variance, 0/1—Loss, and the Curse-of-Dimensionality". In: *Data Mining and Knowledge Discovery* 1.1, pp. 55–77. ISSN: 1573-756X. DOI: 10.1023/A:1009778005914.

Fu, G et al. (2017). "Data assimilation for volcanic ash plumes using a satellite observational operator: a case study on the 2010 Eyjafjallajökull volcanic eruption". In: *Atmospheric Chemistry and Physics* 17.2, pp. 1187–1205. DOI: 10.5194/acp-17-1187-2017.

Fujisada, H, M Urai, and A Iwasaki (2011). "Advanced methodology for ASTER DEM generation". In: *IEEE Transactions on Geoscience and Remote Sensing* 49.12, pp. 5080–5091. ISSN: 0196-2892. DOI: DOI:10.1109/TGRS.2011.2158223.

Gallant, E (2016). "Lava Flow Hazard Assessment for the Idaho National Laboratory, Idaho Falls, and Pocatello, Idaho, U.S.A." PhD thesis. ISBN: 978-1-369-43947-2.

Garcia, A R et al. (2023). "An automated ash dispersion forecast system: case study Popocatépetl volcano, Mexico". In: *Journal of Applied Volcanology* 12.1, p. 9. ISSN: 2191-5040. DOI: 10.1186/s13617-023-00135-4.

Gasteiger, J et al. (2011). "Volcanic ash from Iceland over Munich: mass concentration retrieved from ground-based remote sensing measurements". In: *Atmospheric Chemistry and Physics* 11.5, pp. 2209–2223. DOI: 10.5194/acp-11-2209-2011.

Geer, A J et al. (2018). "All-sky satellite data assimilation at operational weather forecasting centres". In: *Quarterly Journal of the Royal Meteorological Society* 144.713, pp. 1191–1217. ISSN: 0035-9009. DOI: https://doi.org/10.1002/qj.3202.

Geman, S, E Bienenstock, and R Doursat (1992). "Neural Networks and the Bias/Variance Dilemma". In: *Neural Computation* 4.1, pp. 1–58. DOI: 10.1162/neco.1992.4.1.1.

GeoNet (1996a). *Science Alert Bulletin RUA-1996/09 - Ruapehu Volcano*. URL: https://www.geonet.org.nz/vabs/6Q0ToV8fgAkIASoMwwQCSa.

— (1996b). *Science Alert Bulletin RUA-1996/10 - Ruapehu Volcano*. URL: https://www.geonet.org.nz/vabs/6SUp6E10HemwuK0M6QOMs0.

Getz, W M et al. (2018). "Making ecological models adequate". In: *Ecology Letters* 21.2, pp. 153–166. ISSN: 1461-023X. DOI: https://doi.org/10.1111/ele.12893.

Gilleland, E and G Roux (2015). "A new approach to testing forecast predictive accuracy". In: *Meteorological Applications* 22.3, pp. 534–543. ISSN: 1350-4827. DOI: https://doi.org/10.1002/met.1485.

Giordano, D and D Dingwell (2003). "Viscosity of hydrous Etna basalt: implications for Plinian-style basaltic eruptions". In: *Bulletin of Volcanology* 65, pp. 8–14. ISSN: 0258-8900. DOI: https://doi.org/10.1007/s00445-002-0233-2.

Girault, F et al. (2014). "The effect of total grain-size distribution on the dynamics of turbulent volcanic plumes". In: *Earth and Planetary Science Letters* 394, pp. 124–134. ISSN: 0012-821X. DOI: https://doi.org/10.1016/j.epsl.2014.03.021.

Girden, E R (1992). *ANOVA: Repeated measures*. 84. Newbury Park: Sage. ISBN: 0803942575. DOI: https://doi.org/10.4135/9781412983419.

Glatting, G et al. (2007). "Choosing the optimal fit function: comparison of the Akaike information criterion and the F-test". In: *Medical physics* 34.11, pp. 4285–4292. ISSN: 0094-2405. DOI: 10.1118/1.2794176.

Global Volcanism Program (2025a). *[Database] Volcanoes of the World (v. 5.3.0; 17 Jul 2025). Distributed by Smithsonian Institution, compiled by Venzke, E.* DOI: https://doi.org/10.5479/si.GVP.VOTW5-2025.5.3.

— (2025b). *Etna (211060)*. URL: https://volcano.si.edu/volcano.cfm?vn=211060.

Gneiting, T (2008). "Probabilistic forecasting". In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pp. 319–321. ISSN: 0964-1998.

— (2011). "Making and Evaluating Point Forecasts". In: *Journal of the American Statistical Association* 106.494, pp. 746–762. ISSN: 0162-1459. DOI: 10.1198/jasa.2011.r10138.

Gneiting, T and M Katzfuss (2014). "Probabilistic forecasting". In: *Annual Review of Statistics and Its Application* 1.1, pp. 125–151. ISSN: 2326-8298. DOI: https://doi.org/10.1146/annurev-statistics-062713-085831.

Gneiting, T and A E Raftery (2007). "Strictly proper scoring rules, prediction, and estimation". In: *Journal of the American statistical Association* 102.477, pp. 359–378. ISSN: 0162-1459. DOI: https://doi.org/10.1198/016214506000001437.

Gottsmann, J, J.-C. Komorowski, and J Barclay (2019). "Volcanic Unrest and Pre-eruptive Processes: A Hazard and Risk Perspective". In: *Volcanic Unrest : From Science to Society*. Ed. by Joachim Gottsmann, Jürgen Neuberg, and Bettina Scheu. Cham: Springer International Publishing, pp. 1–21. ISBN: 978-3-319-58412-6. DOI: 10.1007/11157{_}2017{_}19.

Grant, A L M et al. (2012). "Horizontal and vertical structure of the Eyjafjallajökull ash cloud over the UK: a comparison of airborne lidar observations and simulations". In: *Atmospheric Chemistry and Physics* 12.21, pp. 10145–10159. ISSN: 1680-7324. DOI: https://doi.org/10.5194/acp-12-10145-2012.

Green, K C and J S Armstrong (2015). "Simple versus complex forecasting: The evidence". In: *Journal of Business Research* 68.8, pp. 1678–1685. ISSN: 0148-2963. DOI: https://doi.org/10.1016/j.jbusres.2015.03.026.

Grimm, V et al. (2014). "Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE". In: *Ecological Modelling* 280, pp. 129–139. ISSN: 0304-3800. DOI: https://doi.org/10.1016/j.ecolmodel.2014.01.018.

Gudmundsson, M T et al. (2012). "Ash generation and distribution from the April-May 2010 eruption of Eyjafjallajökull, Iceland". In: *Scientific Reports* 2.1, p. 572. ISSN: 2045-2322. DOI: 10.1038/srep00572.

Gueugneau, V et al. (2021). "Synthetic benchmarking of concentrated pyroclastic current models". In: *Bulletin of Volcanology* 83.11, p. 75. ISSN: 1432-0819. DOI: 10.1007/s00445-021-01491-y.

Guffanti, M and A Tupper (2015). "Volcanic Ash Hazards and Aviation Risk". In: *Volcanic Hazards, Risks and Disasters*. Ed. by John F Shroder and Paolo Papale. Hazards and Disasters Series. Boston: Elsevier, pp. 87–108. ISBN: 978-0-12-396453-3. DOI: https://doi.org/10.1016/B978-0-12-396453-3.00004-6.

Gunn, L S et al. (2013). "Forecasting the duration of volcanic eruptions: an empirical probabilistic model". In: *Bulletin of Volcanology* 76.1, p. 780. ISSN: 1432-0819. DOI: 10.1007/s00445-013-0780-8.

Gupta, H V, Y Liu, and T Wagener (2008). "Reconciling theory with observations: Elements of a diagnostic approach to model evaluation". English. In: *Hydrological Processes* 22.18, pp. 3802–3813. ISSN: 08856087. DOI: 10.1002/hyp.6989.

Gupta, H V et al. (2012). "Towards a comprehensive assessment of model structural adequacy". In: *Water Resources Research* 48.8. ISSN: 0043-1397. DOI: https://doi.org/10.1029/2011WR011044.

Guthke, A (2017). "Defensible Model Complexity: A Call for Data-Based and Goal-Oriented Model Choice". In: *Groundwater* 55.5, pp. 646–650. ISSN: 0017-467X. DOI: https://doi.org/10.1111/gwat.12554.

Hackett, W R and B F Houghton (1989). "A facies model for a quaternary andesitic composite volcano: Ruapehu, New Zealand". In: *Bulletin of Volcanology* 51.1, pp. 51–68. ISSN: 1432-0819. DOI: 10.1007/BF01086761.

Haendel, M A, C G Chute, and P N Robinson (2018). "Classification, Ontology, and Precision Medicine." In: *The New England Journal of Medicine* 379.15, pp. 1452–1462. ISSN: 1533-4406. DOI: 10.1056/NEJMra1615014.

Haiden, T et al. (2021). *Evaluation of ECMWF forecasts, including the 2021 upgrade*. Tech. rep. 884. ECMWF. DOI: 10.21957/90pgicjk4.

Hales, T C (2000). "The geology of the Summit Area, Mt Ruapehu." PhD thesis. University of Canterbury.

Harris, A et al. (2005). "Lava effusion rates from hand-held thermal infrared imagery: an example from the June 2003 effusive activity at Stromboli". In: *Bulletin of Volcanology* 68, pp. 107–117. ISSN: 0258-8900. DOI: https://doi.org/10.1007/s00445-005-0425-7.

Harris, A J and S Rowland (2001). "FLOWGO: a kinematic thermo-rheological model for lava flowing in a channel". In: *Bulletin of Volcanology* 63.1, pp. 20–44. ISSN: 1432-0819. DOI: 10.1007/s004450000120.

Harris, A J L (2013). "Lava flows". In: *Modeling Volcanic Processes: The Physics and Mathematics of Volcanism*. Ed. by Sarah A Fagents, Tracy K P Gregg, and Rosaly M C Lopes. Cambridge: Cambridge University Press, pp. 85–106. ISBN: 9781139021562. DOI: DOI:10.1017/CBO9781139021562.005.

— (2015). "Chapter 2 - Basaltic Lava Flow Hazard". In: *Volcanic Hazards, Risks and Disasters*. Ed. by John F Shroder and Paolo Papale. Hazards and Disasters Series. Boston: Elsevier, pp. 17–46. ISBN: 978-0-12-396453-3. DOI: https://doi.org/10.1016/B978-0-12-396453-3.00002-2.

Harris, A J L, J Dehn, and S Calvari (2007). "Lava effusion rate definition and measurement: a review". In: *Bulletin of Volcanology* 70.1, pp. 1–22. ISSN: 1432-0819. DOI: 10.1007/s00445-007-0120-y.

Harris, A J L et al. (1997). "Low-cost volcano surveillance from space: case studies from Etna, Krafla, Cerro Negro, Fogo, Lascar and Erebus". In: *Bulletin of Volcanology* 59, pp. 49–64. ISSN: 0258-8900. DOI: https://doi.org/10.1007/s004450050174.

Harvey, N J et al. (2018). "Multi-level emulation of a volcanic ash transport and dispersion model to quantify sensitivity to uncertain parameters". In: *Natural Hazards and Earth System Sciences* 18.1, pp. 41–63. DOI: 10.5194/nhess-18-41-2018.

Harvey, N J et al. (2020). "The Impact of Ensemble Meteorology on Inverse Modeling Estimates of Volcano Emissions and Ash Dispersion Forecasts: Grímsvötn 2011". In: *Atmosphere* 11.10. ISSN: 2073-4433. DOI: 10.3390/atmos11101022.

Haupt, S E et al. (2017). "1 - Principles of meteorology and numerical weather prediction". In: *Renewable Energy Forecasting*. Ed. by George Kariniotakis. Woodhead Publishing, pp. 3–28. ISBN: 978-0-08-100504-0. DOI: https://doi.org/10.1016/B978-0-08-100504-0.00001-9.

Hawaiian Volcano Observatory (2019). *Volcano Watch - How do lava flows cool and how long does it take?* URL: https://www.usgs.gov/observatories/hvo/news/volcano-watch-how-do-lava-flows-cool-and-how-long-does-it-take.

Heming, R F (1980). "Patterns of Quaternary basaltic volcanism in the northern North Island, New Zealand". In: *New Zealand Journal of Geology and Geophysics* 23.3, pp. 335–344. ISSN: 0028-8306. DOI: 10.1080/00288306.1980.10424143.

Herault, A et al. (2009). "Forecasting lava flow hazards during the 2006 Etna eruption: using the MAGFLOW cellular automata model". In: *Computers & Geosciences* 35.5, pp. 1050–1060. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2007.10.008.

Herman, J and W Usher (2017). "SALib: An open-source Python library for sensitivity analysis". In: *Journal of Open Source Software* 2.9, p. 97. ISSN: 2475-9066. DOI: doi:10.21105/joss.00097.

Herrmann, M and W Marzocchi (2023). "Maximizing the forecasting skill of an ensemble model". In: *Geophysical Journal International*. ISSN: 0956-540X. DOI: 10.1093/gji/ggad020.

Hersbach, H et al. (2020). "The ERA5 global reanalysis". In: *Quarterly Journal of the Royal Meteorological Society* 146.730, pp. 1999–2049. ISSN: 0035-9009. DOI: https://doi.org/10.1002/qj.3803.

Hicks, A and R Few (2015). "Trajectories of social vulnerability during the Soufrière Hills volcanic crisis". In: *Journal of Applied Volcanology* 4.1, p. 10. ISSN: 2191-5040. DOI: 10.1186/s13617-015-0029-7.

Hoffman, D W et al. (2023). "A one-dimensional volcanic plume model for predicting ash aggregation". In: *Journal of Geophysical Research: Solid Earth* n/a. ISSN: 2169-9313. DOI: https://doi.org/10.1029/2023JB027002.

Hofmann, M, J Palii, and G Mihelcic (2011). "Epistemic and normative aspects of ontologies in modelling and simulation". In: *Journal of Simulation* 5.3, pp. 135–146. ISSN: 1747-7778. DOI: https://doi.org/10.1057/jos.2011.13.

Höge, M, T Wöhling, and W Nowak (2018). "A Primer for Model Selection: The Decisive Role of Model Complexity". In: *Water Resources Research* 54.3, pp. 1688–1715. ISSN: 0043-1397. DOI: https://doi.org/10.1002/2017WR021902.

Holasek, R E, A W Woods, and S Self (1996). "Experiments on gas-ash separation processes in volcanic umbrella plumes". In: *Journal of Volcanology and Geothermal Research* 70.3, pp. 169–181. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(95)00054-2.

Holland, L et al. (2020). "Two ensemble approaches for forecasting sulfur dioxide concentrations from Kīlauea Volcano". In: *Weather and Forecasting* 35.5, pp. 1923–1937. ISSN: 0882-8156. DOI: https://doi.org/10.1175/WAF-D-19-0189.1.

Hooper, D M and G S Mattioli (2001). "Kinematic Modeling of Pyroclastic Flows Produced by Gravitational Dome Collapse at Soufriere Hills Volcano, Montserrat". In: *Natural Hazards* 23.1, pp. 65–86. ISSN: 1573-0840. DOI: 10.1023/A:1008130605558.

Hopkins, J L et al. (2020). "Auckland Volcanic Field magmatism, volcanism, and hazard: a review". In: *New Zealand Journal of Geology and Geophysics*, pp. 1–22. ISSN: 0028-8306. DOI: 10.1080/00288306.2020.1736102.

Horsfall, S (2019). "Improving Evacuation Decision-Making in a Taranaki Volcanic Crisis through Analogue Case Studies". PhD thesis. University of Canterbury.

Horwell, C J (2007). "Grain-size analysis of volcanic ash for the rapid assessment of respiratory health hazard". In: *Journal of Environmental Monitoring* 9.10, p. 1107. ISSN: 1464-0325. DOI: 10.1039/b710583p.

Horwell, C J and P J Baxter (2006). "The respiratory health hazards of volcanic ash: a review for volcanic risk mitigation". In: *Bulletin of Volcanology* 69.1, pp. 1–24. ISSN: 1432-0819. DOI: 10.1007/s00445-006-0052-y.

Hoskuldsson, A et al. (2007). "The millennium eruption of Hekla in February 2000". In: *Bulletin of Volcanology* 70, pp. 169–182. ISSN: 0258-8900. DOI: https://doi.org/10.1007/s00445-007-0128-3.

Houghton, B F, J H Latter, and W R Hackett (1987). "Volcanic hazard assessment for Ruapehu composite volcano, Taupo volcanic zone, New Zealand". In: *Bulletin of Volcanology* 49, pp. 737–751. ISSN: 0258-8900. DOI: https://doi.org/10.1007/BF01079825.

Houghton, B F et al. (2006). "Proximal tephra hazards: Recent eruption studies applied to volcanic risk in the Auckland volcanic field, New Zealand". In: *Journal of Volcanology and Geothermal Research* 155.1-2, pp. 138–149. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2006.02.006.

Huggel, C et al. (2008). "Evaluation of ASTER and SRTM DEM data for lahar modeling: A case study on lahars from Popocatépetl Volcano, Mexico". In: *Journal of Volcanology and Geothermal Research* 170.1, pp. 99–110. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2007.09.005.

Hurley, P (1994). "PARTPUFF—A Lagrangian Particle–Puff Approach for Plume Dispersion Modeling Applications". In: *Journal of Applied Meteorology (1988-2005)* 33.2, pp. 285–294. ISSN: 08948763, 15200450. DOI: https://doi.org/10.1175/1520-0450(1994)033{\%}3C0285:PLPPAF{\%}3E2.0.CO;2.

Hurst, A W and T Hurst (1994). *ASHFALL, a computer program for estimating volcanic ash fallout: report and users guide*. Institute of Geological & Nuclear Sciences. ISBN: 0478088140.

Hurst, A W and R Turner (1999). "Performance of the program ASHFALL for forecasting ashfall during the 1995 and 1996 eruptions of Ruapehu volcano". In: *New Zealand Journal of Geology and Geophysics* 42.4, pp. 615–622. ISSN: 0028-8306. DOI: 10.1080/00288306.1999.9514865.

Hurst, T and C Davis (2017). "Forecasting volcanic ash deposition using HYSPLIT". In: *Journal of Applied Volcanology* 6.1, p. 5. ISSN: 2191-5040. DOI: 10.1186/s13617-017-0056-7.

Hurst, T and B J Scott (1998). *Ash predictions, how successful were they?* Tech. rep. Institute of Geological & Nuclear Sciences Limited. URL: https://www.naturalhazards.govt.nz/assets/Publications-Resources/4575-Ash-predictions-how-successful-were-they-compressed.pdf.

Hyman, D M R, H R Dietterich, and M R Patrick (2022). "Toward Next-Generation Lava Flow Forecasting: Development of a Fast, Physics-Based Lava Propagation Model". In: *Journal*

of Geophysical Research: Solid Earth 127.10. ISSN: 2169-9313. DOI: https://doi.org/10.1029/2022JB024998.

Hyman, D M R et al. (2024). "Real-time lava flow forecasting during the 2022 Mauna Loa eruption response". In: *Bulletin of Volcanology* 87.1, p. 3. ISSN: 1432-0819. DOI: 10.1007/s00445-024-01783-z.

ICAO (2023). *International Civil Aviation Organization Meteorology Panel Roadmap for International Airways Volcano Watch (IAVW) in support of international air navigation*. URL: https://www.icao.int/sites/default/files/METP/Documents/IAVW-Roadmap-vs-5.1-07-Feb-2023.pdf.

Iregui, A M, H M Núñez, and J Otero (2021). "Testing the efficiency of inflation and exchange rate forecast revisions in a changing economic environment". In: *Journal of Economic Behavior & Organization* 187, pp. 290–314. ISSN: 0167-2681. DOI: https://doi.org/10.1016/j.jebo.2021.04.037.

Ishihara, K, M Iguchi, and K Kamo (1990). "Numerical Simulation of Lava Flows on Some Volcanoes in Japan". In: *Lava Flows and Domes: Emplacement Mechanisms and Hazard Implications*. Ed. by Jonathan H Fink. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 174–207. ISBN: 978-3-642-74379-5. DOI: 10.1007/978-3-642-74379-5{_}8.

Ishii, K and Y Iriyama (2024). "Do Seasonality and Latitude Dictate the Formation of Strong or Weak Volcanic Eruption Plumes?" In: *Geophysical Research Letters* 51.20. ISSN: 0094-8276. DOI: https://doi.org/10.1029/2023GL106674.

ISO 5725-1 (1994). *Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions*. URL: https://www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-1:v1:en.

Iverson, R M, S P Schilling, and J W Vallance (1998). "Objective delineation of lahar-inundation hazard zones". In: *GSA Bulletin* 110.8, pp. 972–984. ISSN: 0016-7606. DOI: 10.1130/0016-7606(1998)110<0972:ODOLIH>2.3.CO;2.

Jackson, L, C Smyth, and D Poole (2008). "Hazardmatch: an application of artificial intelligence to landslide susceptibility mapping, Howe Sound area, British Columbia". In: Quebec City: 4th Canadian Conference on Geohazards.

Jackson, L J, A S Trebitz, and K L Cottingham (2000). "An introduction to the practice of ecological modeling". In: *BioScience* 50.8, pp. 694–706. ISSN: 1525-3244. DOI: https://doi.org/10.1641/0006-3568(2000)050[0694:AITTPO]2.0.CO;2.

Jakeman, A J, R A Letcher, and J P Norton (2006). "Ten iterative steps in development and evaluation of environmental models". In: *Environmental Modelling & Software* 21.5, pp. 602–614. ISSN: 1364-8152. DOI: https://doi.org/10.1016/j.envsoft.2006.01.004.

James, L, H Dacre, and N Harvey (2024). "How dependent are quantitative volcanic ash concentration and along-flight dosage forecasts to model structural choices?" In: *Meteorological Applications*. ISSN: 1469-8080. DOI: https://doi.org/10.1002/met.70003.

Jehn, F U et al. (2019). "Trade-offs between parameter constraints and model realism: a case study". In: *Scientific Reports* 9.1. ISSN: 2045-2322. DOI: 10.1038/s41598-019-46963-6.

Jenkins, S F et al. (2012). "Regional ash fall hazard I: a probabilistic assessment methodology". In: *Bulletin of Volcanology* 74.7, pp. 1699–1712. ISSN: 1432-0819. DOI: 10.1007/s00445-012-0627-8.

Jenkins, S F et al. (2014). "Volcanic risk assessment: Quantifying physical vulnerability in the built environment". In: *Journal of Volcanology and Geothermal Research* 276, pp. 105–120. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2014.03.002.

Jenkins, S F et al. (2015a). "Developing building-damage scales for lahars: application to Merapi volcano, Indonesia". In: *Bulletin of Volcanology* 77.9, p. 75. ISSN: 1432-0819. DOI: 10.1007/s00445-015-0961-8.

Jenkins, S F et al. (2015b). "Volcanic ash fall hazard and risk". In: *Global Volcanic Hazards and Risk*. Cambridge University Press Cambridge. Chap. 3, pp. 173–222. DOI: https://doi.org/10.1017/CBO9781316276273.005.

Jenkins, S F et al. (2017). "Damage from lava flows: insights from the 2014–2015 eruption of Fogo, Cape Verde". In: *Journal of Applied Volcanology* 6.1, p. 6. ISSN: 2191-5040. DOI: 10.1186/s13617-017-0057-6.

Jenkins, S F et al. (2020). "Reconstructing eruptions from historical accounts: Makaturing c. 1765, Philippines". In: *Journal of Volcanology and Geothermal Research* 404. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2020.107022.

Jenkins, S F et al. (2024). "Assessing volcanic hazard and exposure in a data poor context: Case study for Ethiopia, Kenya, and Cabo Verde". In: *Progress in Disaster Science* 23. ISSN: 2590-0617. DOI: https://doi.org/10.1016/j.pdisas.2024.100350.

Johnston, D M et al. (2000). "Impacts of the 1945 and 1995–1996 Ruapehu eruptions, New Zealand: an example of increasing societal vulnerability". In: *Geological Society of America Bulletin* 112.5, pp. 720–726. ISSN: 1943-2674. DOI: https://doi.org/10.1130/0016-7606(2000)112<720:IOTARE>2.0.CO;2.

Johnston, E N et al. (2012). "Reconstructing the tephra dispersal pattern from the Bronze Age eruption of Santorini using an advection–diffusion model". In: *Bulletin of Volcanology* 74.6, pp. 1485–1507. ISSN: 1432-0819. DOI: 10.1007/s00445-012-0609-x.

Jones, A et al. (2007). "The U.K. Met Office's Next-Generation Atmospheric Dispersion Model, NAME III BT - Air Pollution Modeling and Its Application XVII". In: ed. by Carlos Borrego and Ann-Lise Norman. Boston, MA: Springer US, pp. 580–589. ISBN: 978-0-387-68854-1. DOI: https://doi.org/10.1007/978-0-387-68854-1{_}62.

Jones, T J et al. (2024). "Identifying rheological regimes within pyroclastic density currents". In: *Nature Communications* 15.1, p. 4401. ISSN: 2041-1723. DOI: 10.1038/s41467-024-48612-7.

Joyce, K E et al. (2009). "Remote sensing data types and techniques for lahar path detection: A case study at Mt Ruapehu, New Zealand". In: *Remote Sensing of Environment* 113.8, pp. 1778–1786. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2009.04.001.

Jung, H and K Chung (2015). "Ontology-driven slope modeling for disaster management service". In: *Cluster Computing* 18.2, pp. 677–692. ISSN: 1573-7543. DOI: 10.1007/s10586-015-0424-1.

Kavanagh, J L, S L Engwell, and S A Martin (2018). "A review of laboratory and numerical modelling in volcanology". In: *Solid Earth* 9.2, pp. 531–571. DOI: 10.5194/se-9-531-2018.

Kelfoun, K and S V Vargas (2016). "VolcFlow capabilities and potential development for the simulation of lava flows". In: *Detecting, Modelling and Responding to Effusive Eruptions*. Ed. by A J L Harris et al. Vol. 426. Geological Society of London, pp. 337–343. ISBN: 9781862397361. DOI: 10.1144/SP426.8.

Kelfoun, K et al. (2009). "Testing the suitability of frictional behaviour for pyroclastic flow simulation by comparison with a well-constrained eruption at Tungurahua volcano (Ecuador)". In: *Bulletin of Volcanology* 71.9, pp. 1057–1075. ISSN: 1432-0819. DOI: 10.1007/s00445-009-0286-6.

Kidson, J W (2000). "An analysis of New Zealand synoptic types and their use in defining weather regimes". In: *International Journal of Climatology* 20.3, pp. 299–316. ISSN: 0899-8418. DOI: https://doi.org/10.1002/(SICI)1097-0088(20000315)20:3<299::AID-JOC474>3.0.CO;2-B.

Kilgour, G et al. (2010). "The 25 September 2007 eruption of Mount Ruapehu, New Zealand: Directed ballistics, surtseyan jets, and ice-slurry lahars". In: *Journal of Volcanology and Geothermal Research* 191.1, pp. 1–14. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2009.10.015.

Kilgour, G et al. (2013). "Small volume andesite magmas and melt–mush interactions at Ruapehu, New Zealand: evidence from melt inclusions". In: *Contributions to Mineralogy and Petrology* 166.2, pp. 371–392. ISSN: 1432-0967. DOI: 10.1007/s00410-013-0880-7.

Kilgour, G et al. (2021). "Whakaari/White Island: a review of New Zealand's most active volcano". In: *New Zealand Journal of Geology and Geophysics* 64.2-3, pp. 273–295. ISSN: 0028-8306. DOI: https://doi.org/10.1080/00288306.2021.1918186.

Klawonn, M et al. (2012). "Novel inversion approach to constrain plume sedimentation from tephra deposit data: Application to the 17 June 1996 eruption of Ruapehu volcano, New Zealand". In: *Journal of Geophysical Research: Solid Earth* 117.B5. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2011JB008767.

Klawonn, M et al. (2014). "Constraining particle size-dependent plume sedimentation from the 17 June 1996 eruption of Ruapehu Volcano, New Zealand, using geophysical inversions". In: *Journal of Geophysical Research: Solid Earth* 119.3, pp. 1749–1763. ISSN: 2169-9313. DOI: https://doi.org/10.1002/2013JB010387.

Kolzenburg, S et al. (2016). "In situ thermal characterization of cooling/crystallizing lavas during rheology measurements and implications for lava flow emplacement". In: *Geochimica et Cosmochimica Acta* 195, pp. 244–258. ISSN: 0016-7037. DOI: https://doi.org/10.1016/j.gca.2016.09.022.

Komorowski, J-C et al. (2008). "Reconstruction and analysis of sub-plinian tephra dispersal during the 1530 A.D. Soufrière (Guadeloupe) eruption: Implications for scenario definition and hazards assessment". In: *Journal of Volcanology and Geothermal Research* 178.3, pp. 491–515. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2007.11.022.

Kover, T P (1995). "Application of a digital terrain model for the modeling of volcanic flows: a tool for volcanic hazard determination". PhD thesis. SUNY at Buffalo, p. 62.

Koyaguchi, T and M Ohno (2001). "Reconstruction of eruption column dynamics on the basis of grain size of tephra fall deposits: 2. Application to the Pinatubo 1991 eruption". In: *Journal of Geophysical Research: Solid Earth* 106.B4, pp. 6513–6533. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2000JB900427.

Krishnamurti, T N et al. (2000). "Multimodel Ensemble Forecasts for Weather and Seasonal Climate". English. In: *Journal of Climate* 13.23, pp. 4196–4216. DOI: https://doi.org/10.1175/1520-0442(2000)013<4196:MEFFWA>2.0.CO;2.

Krishnamurti, T N et al. (2016). "A review of multimodel superensemble forecasting for weather, seasonal climate, and hurricanes". In: *Reviews of Geophysics* 54.2, pp. 336–377. ISSN: 8755-1209. DOI: https://doi.org/10.1002/2015RG000513.

Kristiansen, N I et al. (2012). "Performance assessment of a volcanic ash transport model mini-ensemble used for inverse modeling of the 2010 Eyjafjallajökull eruption". In: *Journal of Geophysical Research: Atmospheres* 117.D20. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2011JD016844.

Krueger, T et al. (2010). "Ensemble evaluation of hydrological model hypotheses". In: *Water Resources Research* 46.7. ISSN: 0043-1397. DOI: https://doi.org/10.1029/2009WR007845.

Kubanek, J et al. (2015). "Lava flow mapping and volume calculations for the 2012–2013 Tolbachik, Kamchatka, fissure eruption using bistatic TanDEM-X InSAR". In: *Bulletin of Volcanology* 77.12, p. 106. ISSN: 1432-0819. DOI: 10.1007/s00445-015-0989-9.

Kueppers, U et al. (2019). "Biased volcanic hazard assessment due to incomplete eruption records on ocean islands: an example of Sete Cidades Volcano, Azores". In: *Frontiers in Earth Science* 7, p. 122. ISSN: 2296-6463. DOI: https://doi.org/10.3389/feart.2019.00122.

Kumagai, H et al. (2009). "Seismic tracking of lahars using tremor signals". In: *Journal of Volcanology and Geothermal Research* 183.1-2, pp. 112–121. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2009.03.010.

Kylling, A et al. (2015). "A model sensitivity study of the impact of clouds on satellite detection and retrieval of volcanic ash". In: *Atmospheric Measurement Techniques* 8.5, pp. 1935–1949. DOI: 10.5194/amt-8-1935-2015.

Lambrix, P (2004). "Ontologies in Bioinformatics and Systems Biology". In: *Artificial Intelligence Methods And Tools For Systems Biology*. Ed. by Werner Dubitzky and Francisco Azuaje. Dordrecht: Springer Netherlands, pp. 129–145. ISBN: 978-1-4020-2865-6. DOI: 10.1007/978-1-4020-5811-0{_}8.

Larsen, L G et al. (2016). "Appropriate complexity landscape modeling". In: *Earth-Science Reviews* 160, pp. 111–130. ISSN: 0012-8252. DOI: https://doi.org/10.1016/j.earscirev.2016.06.016.

Larson, M (2005). "Numerical Modeling". In: *Encyclopedia of Coastal Science*. Ed. by Maurice L Schwartz. Dordrecht: Springer Netherlands, pp. 730–733. ISBN: 978-1-4020-3880-8. DOI: 10.1007/1-4020-3880-1{_}232.

Leonard, G S et al. (2021). "Ruapehu and Tongariro stratovolcanoes: a review of current understanding". In: *New Zealand Journal of Geology and Geophysics* 64.2-3, pp. 389–420. ISSN: 0028-8306. DOI: 10.1080/00288306.2021.1909080.

Levin, S A (1998). "Ecosystems and the biosphere as complex adaptive systems". In: *Ecosystems* 1, pp. 431–436. ISSN: 1432-9840. DOI: https://doi.org/10.1007/s100219900037.

Li, H, C Y Xu, and S Beldring (2015). "How much can we gain with increasing model complexity with the same model concepts?" In: *Journal of Hydrology* 527, pp. 858–871. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2015.05.044.

Lidén, R and J Harlin (2000). "Analysis of conceptual rainfall–runoff modelling performance in different climates". In: *Journal of Hydrology* 238.3, pp. 231–247. ISSN: 0022-1694. DOI: https://doi.org/10.1016/S0022-1694(00)00330-9.

Lindsay, J et al. (2010). "Towards real-time eruption forecasting in the Auckland Volcanic Field: application of BET_EF during the New Zealand National Disaster Exercise 'Ruaumoko'".

In: *Bulletin of Volcanology* 72.2, pp. 185–204. ISSN: 1432-0819. DOI: 10.1007/s00445-009-0311-9.

Liu, J et al. (2015). "Validation of ash cloud modelling with satellite retrievals: a case study of the 16–17 June 1996 Mount Ruapehu eruption". In: *Natural Hazards* 78.2, pp. 973–993. ISSN: 1573-0840. DOI: 10.1007/s11069-015-1753-3.

Llenos, A L and A J Michael (2019). "Ensembles of ETAS models provide optimal operational earthquake forecasting during swarms: Insights from the 2015 San Ramon, California swarm". In: *Bulletin of the Seismological Society of America* 109.6, pp. 2145–2158. DOI: 10.1785/0120190020.

Longchamp, C et al. (2011). "Characterization of tephra deposits with limited exposure: the example of the two largest explosive eruptions at Nisyros volcano (Greece)". In: *Bulletin of Volcanology* 73.9, pp. 1337–1352. ISSN: 1432-0819. DOI: 10.1007/s00445-011-0469-9.

Loucks, D P and E van Beek (2017). "System Sensitivity and Uncertainty Analysis". In: *Water Resource Systems Planning and Management: An Introduction to Methods, Models, and Applications*. Ed. by Daniel P Loucks and Eelco van Beek. Cham: Springer International Publishing, pp. 331–374. ISBN: 978-3-319-44234-1. DOI: 10.1007/978-3-319-44234-1{_}8.

Loughlin, S C et al. (2015). "An introduction to global volcanic hazard and risk ". In: *Global Volcanic Hazards and Risk*, pp. 1–80. DOI: 10.1017/CBO9781316276273.003.

Lube, G et al. (2015). "Synthesizing large-scale pyroclastic flows: Experimental design, scaling, and first results from PELE". In: *Journal of Geophysical Research: Solid Earth* 120.3, pp. 1487–1502. ISSN: 2169-9313. DOI: https://doi.org/10.1002/2014JB011666.

Lube, G et al. (2020). "Multiphase flow behaviour and hazard prediction of pyroclastic density currents". In: *Nature Reviews Earth & Environment* 1.7, pp. 348–365. ISSN: 2662-138X. DOI: 10.1038/s43017-020-0064-8.

Ludden, T M, S L Beal, and L B Sheiner (1994). "Comparison of the Akaike Information Criterion, the Schwarz criterion and the F test as guides to model selection". In: *Journal of Pharmacokinetics and Biopharmaceutics* 22.5, pp. 431–445. ISSN: 0090-466X. DOI: 10.1007/BF02353864.

Macedonio, G, A Costa, and A Folch (2008). "Ash fallout scenarios at Vesuvius: Numerical simulations and implications for hazard assessment". In: *Journal of Volcanology and Geothermal Research* 178.3, pp. 366–377. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2008.08.014.

Macedonio, G, A Costa, and A Longo (2005). "A computer model for volcanic ash fallout and assessment of subsequent hazard". In: *Computers & Geosciences* 31.7, pp. 837–845. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2005.01.013.

Macedonio, G, M T Pareschi, and R Santacroce (1990). "Renewal of explosive activity at Vesuvius: models for the expected tephra fallout". In: *Journal of Volcanology and Geothermal Research* 40.4, pp. 327–342. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(90)90112-S.

Macías, J L et al. (2008). "Hazard map of El Chichón volcano, Chiapas, México: Constraints posed by eruptive history and computer simulations". In: *Journal of Volcanology and Geothermal Research* 175.4, pp. 444–458. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2008.02.023.

Macorps, E (2021). "Field and Remote Sensing Analysis of the 2015 Pyroclastic Density Currents at Colima (Mexico) and Calbuco (Chile) Volcanoes: Implications for Hazard Assessment and Crisis Management". English. PhD thesis. University of South Florida, p. 340. ISBN: 9798535509617.

Magill, C et al. (2015). "Simulating a multi-phase tephra fall event: inversion modelling for the 1707 Hoei eruption of Mount Fuji, Japan". In: *Bulletin of Volcanology* 77.9, p. 81. ISSN: 1432-0819. DOI: 10.1007/s00445-015-0967-2.

Major, J J and L E Lara (2013). "Overview of Chaitén Volcano, Chile, and its 2008-2009 eruption". In: *Andean Geology* 40.2, pp. 196–215. DOI: 10.5027/andgeoV40n2-a01.

Malchow, A-K and F Hartig (2024). "Calibration, sensitivity and uncertainty analysis of ecological models–a review". In: *Ecology Letters*. DOI: DOI:10.22541/au.173090741.12160653/v1.

Malin, M C and M F Sheridan (1982). "Computer-Assisted Mapping of Pyroclastic Surges". In: *Science* 217.4560, pp. 637–640. ISSN: 00368075, 10959203. DOI: DOI:10.1126/science.217.4560.637.

Malmborg, C A et al. (2024). "Defining model complexity: An ecological perspective". In: *Meteorological Applications* 31.3. ISSN: 1350-4827. DOI: https://doi.org/10.1002/met.2202.

Manga, M et al. (2017). *Volcanic eruptions and their repose, unrest, precursors, and timing.* National Academies Press. DOI: https://doi.org/10.17226/24650.

Mannen, K (2014). "Particle segregation of an eruption plume as revealed by a comprehensive analysis of tephra dispersal: Theory and application". In: *Journal of Volcanology and Geothermal Research* 284, pp. 61–78. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2014.07.009.

Mannen, K et al. (2020). "Simulations of Tephra Fall Deposits From a Bending Eruption Plume and the Optimum Model for Particle Release". In: *Journal of Geophysical Research: Solid Earth* 125.6. ISSN: 2169-9313. DOI: https://doi.org/10.1029/2019JB018902.

Mara, T (2009). "Extension of the RBD-FAST method to the computation of global sensitivity indices". In: *Reliability Engineering & System Safety* 94, pp. 1274–1281. DOI: 10.1016/j.ress.2009.01.012.

Marquez, M, C Paredes, and M Llorente (2022). "Attempt to Model Lava Flow Faster Than Real Time: An Example of La Palma Using VolcFlow". In: *GeoHazards* 3.4, pp. 529–562. ISSN: 2624-795X. DOI: 10.3390/geohazards3040027.

Martí, J (2024). "From rest to eruption: How we should anticipate volcanic eruptions". In: *Natural Hazards* 1.1, p. 31. ISSN: 2948-2100. DOI: 10.1038/s44304-024-00033-8.

Martín-Raya, N et al. (2023). "A lava flow simulation experience oriented to disaster risk reduction, early warning systems and response during the 2021 volcanic eruption in Cumbre Vieja, La Palma". In: *Natural Hazards*. ISSN: 1573-0840. DOI: 10.1007/s11069-023-05989-w.

Martinez Montesinos, B et al. (2022). "On the feasibility and usefulness of high-performance computing in probabilistic volcanic hazard assessment: An application to tephra hazard from Campi Flegrei". In: *Frontiers in Earth Science*. DOI: 10.3389/feart.2022.941789.

Martínez-Sepúlveda, M, J Martí, and M López-Saavedra (2024). "VOLCANBOX: A software platform for volcanic hazard assessment". In: *International Journal of Disaster Risk Reduction* 113. ISSN: 2212-4209. DOI: https://doi.org/10.1016/j.ijdrr.2024.104817.

Martinez-Villegas, M M et al. (2022). "Perspectives on the 12 January 2020 Taal Volcano eruption: An analysis of residents' narrative accounts". In: *Frontiers in Earth Science* Volume 10 - 2022. ISSN: 2296-6463. DOI: 10.3389/feart.2022.923224.

Marzocchi, W and M S Bebbington (2012). "Probabilistic eruption forecasting at short and long time scales". English. In: *Bulletin of Volcanology* 74.8, pp. 1777–1805. ISSN: 0258-8900. DOI: http://dx.doi.org/10.1007/s00445-012-0633-x.

Marzocchi, W and T H Jordan (2014). "Testing for ontological errors in probabilistic forecasting models of natural systems". In: *Proceedings of the National Academy of Sciences* 111.33, pp. 11973–11978. DOI: 10.1073/pnas.1410183111.

Marzocchi, W, L Sandri, and J Selva (2008). "BET_EF: a probabilistic tool for long- and short-term eruption forecasting". In: *Bulletin of Volcanology* 70.5, pp. 623–632. ISSN: 1432-0819. DOI: 10.1007/s00445-007-0157-y.

— (2010). "BET_VH: a probabilistic tool for long-term volcanic hazard assessment". English. In: *Bulletin of Volcanology* 72.6, pp. 705–716. ISSN: 0258-8900. DOI: http://dx.doi.org/10.1007/s00445-010-0357-8.

Marzocchi, W, J Selva, and T H Jordan (2021). "A unified probabilistic framework for volcanic hazard and eruption forecasting ". English. In: *Natural Hazards and Earth System Sciences* 21.11, pp. 3509–3517. DOI: 10.5194/nhess-21-3509-2021.

Marzocchi, W and G Woo (2007). "Probabilistic eruption forecasting and the call for an evacuation". In: *Geophysical Research Letters* 34.22. ISSN: 0094-8276. DOI: 10.1029/2007GL031922.

Marzocchi, W, J Zechar, and T Jordan (2012). "Bayesian Forecast Evaluation and Ensemble Earthquake Forecasting". In: *The Bulletin of the Seismological Society of America* 102, pp. 2574–2584. DOI: 10.1785/0120110327.

Marzocchi, W et al. (2004). "Quantifying probabilities of volcanic events: The example of volcanic hazard at Mount Vesuvius". In: *Journal of Geophysical Research: Solid Earth* 109.B11. ISSN: 0148-0227. DOI: 10.1029/2004JB003155.

Mason, B G, D M Pyle, and C Oppenheimer (2004). "The size and frequency of the largest explosive eruptions on Earth". In: *Bulletin of Volcanology* 66.8, pp. 735–748. ISSN: 0258-8900. DOI: https://doi.org/10.1007/s00445-004-0355-9.

Massaro, S et al. (2023). "Assessing long-term tephra fallout hazard in Southern Italy from Neapolitan volcanoes". In: *Natural Hazards and Earth System Sciences Discussions* 2023, pp. 1–35. DOI: 10.5194/nhess-2023-3.

Masseroli, M (2019). "Biological and Medical Ontologies: Introduction". In: *Encyclopedia of Bioinformatics and Computational Biology*. Ed. by Shoba Ranganathan et al. Oxford: Academic Press, pp. 813–822. ISBN: 978-0-12-811432-2. DOI: https://doi.org/10.1016/B978-0-12-809633-8.20395-6.

Mastin, L G (2001). *A simple calculator of ballistic trajectories for blocks ejected during volcanic eruptions*. Tech. rep. Reston, VA: U.S. Geological Survey, p. 13. DOI: 10.3133/ofr0145.

— (2007). "A user-friendly one-dimensional model for wet volcanic plumes". In: *Geochemistry, Geophysics, Geosystems* 8.3. ISSN: 1525-2027. DOI: https://doi.org/10.1029/2006GC001455.

— (2014). "Testing the accuracy of a 1-D volcanic plume model in estimating mass eruption rate". In: *Journal of Geophysical Research: Atmospheres* 119.5, pp. 2474–2495. ISSN: 2169-897X. DOI: https://doi.org/10.1002/2013JD020604.

Mastin, L G, A R Van Eaton, and A J Durant (2016). "Adjusting particle-size distributions to account for aggregation in tephra-deposit model forecasts ". English. In: *Atmospheric Chemistry and Physics* 16.14, pp. 9399–9420. DOI: 10.5194/acp-16-9399-2016.

Mastin, L G, A R Van Eaton, and H F Schwaiger (2020). *A probabilistic assessment of tephra-fall hazards at Hanford, Washington, from a future eruption of Mount St. Helens*. English. Tech. rep. Reston, VA, p. 54. DOI: 10.3133/ofr20201133.

Mastin, L G et al. (2009). "A multidisciplinary effort to assign realistic source parameters to models of volcanic ash-cloud transport and dispersion during eruptions". In: *Journal of Volcanology and Geothermal Research* 186.1, pp. 10–21. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2009.01.008.

Mastin, L G et al. (2013). *User's guide and reference to Ash3d—A three-dimensional model for Eulerian atmospheric tephra transport and deposition*. English. Tech. rep. Reston, VA, p. 25. DOI: 10.3133/ofr20131122. URL: http://pubs.er.usgs.gov/publication/ofr20131122.

Maxey, M R and J J Riley (1983). "Equation of motion for a small rigid sphere in a nonuniform flow". In: *The Physics of Fluids* 26.4, pp. 883–889. ISSN: 0031-9171. DOI: https://doi.org/10.1063/1.864230.

May, R M (2019). *Stability and complexity in model ecosystems*. Princeton University Press. ISBN: 0691206910. DOI: https://doi.org/10.2307/j.ctvs32rq4.

McDonald, N J et al. (2024). "Quantifying economic risks to dairy farms from volcanic hazards in Taranaki, New Zealand". In: *EGUsphere* 2024, pp. 1–42. DOI: https://doi.org/10.5194/egusphere-2024-3619.

Mcguinness, D et al. (2006). "Towards a Reference Volcano Ontology for Semantic Scientific Data Integration". In: *AGU Spring Meeting Abstracts*.

Mead, S and C Magill (2014). "Determining change points in data completeness for the Holocene eruption record". In: *Bulletin of Volcanology* 76.11, p. 874. ISSN: 1432-0819. DOI: 10.1007/s00445-014-0874-y.

Mead, S R, J Procter, and M Bebbington (2023). "Probabilistic volcanic mass flow hazard assessment using statistical surrogates of deterministic simulations". In: *Computers & Geosciences* 178. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2023.105417.

Mei, E T W et al. (2013). "Lessons learned from the 2010 evacuations at Merapi volcano". In: *Journal of Volcanology and Geothermal Research* 261, pp. 348–365. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2013.03.010.

Mereu, L et al. (2023). "A New Radar-Based Statistical Model to Quantify Mass Eruption Rate of Volcanic Plumes". In: *Geophysical Research Letters* 50.7. ISSN: 0094-8276. DOI: https://doi.org/10.1029/2022GL100596.

Mereu, L et al. (2025). "Estimating the mass of tephra accumulated on roads to best manage the impact of volcanic eruptions: the example of Mt Etna, Italy". In: *Natural Hazards and Earth System Sciences* 25.6, pp. 1943–1962. DOI: 10.5194/nhess-25-1943-2025.

Merz, B et al. (2020). "Impact Forecasting to Support Emergency Management of Natural Hazards". In: *Reviews of Geophysics* 58.4. ISSN: 8755-1209. DOI: https://doi.org/10.1029/2020RG000704.

Michaud-Dubuy, A, G Carazzo, and E Kaminski (2021). "Volcanic hazard assessment for tephra fallout in Martinique". In: *Journal of Applied Volcanology* 10.1, p. 8. ISSN: 2191-5040. DOI: 10.1186/s13617-021-00106-7.

Michaud-Dubuy, A and M Gouhier (2025). "Validation of Near Real-Time Retrieval of Plume Mass Eruption Rates: The Case of the 2021 Eruption of La Soufrière, St Vincent". In: *Geochemistry, Geophysics, Geosystems* 26.10. DOI: https://doi.org/10.1029/2025GC012498.

Michaud-Dubuy, A et al. (2019). "Impact of wind direction variability on hazard assessment in Martinique (Lesser Antilles): The example of the 13.5kacalBP Bellefontaine Plinian eruption of Mount Pelée volcano". In: *Journal of Volcanology and Geothermal Research* 381, pp. 193–208. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2019.06.004.

Michaud-Dubuy, A et al. (2024). "Tephra fallout hazard assessment for a hydrovolcanic eruptive scenario in Mayotte". In: *Scientific Reports* 14.1. ISSN: 2045-2322. DOI: 10.1038/s41598-024-83266-x.

Montgomery, D C (2017). *Design and analysis of experiments*. Hoboken, NJ: John Wiley & Sons. ISBN: 1119113474.

Morton, B R, G Taylor, and J S Turner (1956). "Turbulent Gravitational Convection from Maintained and Instantaneous Sources". In: *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 234.1196, pp. 1–23. ISSN: 00804630. DOI: https://doi.org/10.1098/rspa.1956.0011.

Mulargia, F, S Tinti, and E Boschi (1985). "A statistical analysis of flank eruptions on Etna volcano". In: *Journal of Volcanology and Geothermal Research* 23.3, pp. 263–272. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(85)90037-X.

Mulena, G C et al. (2016). "Examining the influence of meteorological simulations forced by different initial and boundary conditions in volcanic ash dispersion modelling". In: *Atmospheric Research* 176-177, pp. 29–42. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2016.02.009.

Müller, E H et al. (2013). "Parallelisation of the Lagrangian atmospheric dispersion model NAME". In: *Computer Physics Communications* 184.12, pp. 2734–2745. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2013.06.022.

Muñoz-Salinas, E et al. (2009). "Lahar flow simulations using LAHARZ program: Application for the Popocatépetl volcano, Mexico". In: *Journal of Volcanology and Geothermal Research* 182.1, pp. 13–22. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2009.01.030.

Myer, J (2018). "The volcanism ontology (VO): Semantic modeling of volcanic eruptions and volcanoes". PhD thesis. Georgia State University. DOI: https://doi.org/10.57709/12492932.

Myers, B and C Driedger (2008). *Geologic Hazards at Volcanoes*. URL: https://pubs.usgs.gov/gip/64/.

Nairn, I A and S Self (1978). "Explosive eruptions and pyroclastic avalanches from Ngauruhoe in February 1975". In: *Journal of Volcanology and Geothermal Research* 3.1-2, pp. 39–60. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(78)90003-3.

National Emergency Management Agency (2025). *New Zealand's riskscape*. URL: https://www.civildefence.govt.nz/assets/Uploads/documents/publications/Busy-Few-Years_Hazardscape-A3s-17-April-2025.pdf.

Neglia, F et al. (2021). "Shallow-water models for volcanic granular flows: A review of strengths and weaknesses of TITAN2D and FLO2D numerical codes". In: *Journal of Volcanology and Geothermal Research* 410. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2020.107146.

Neri, A and G Macedonio (1996). "Physical Modeling of Collapsing Volcanic Columns and Pyroclastic Flows". In: *Monitoring and Mitigation of Volcano Hazards*. Ed. by Roberto Scarpa and Robert I Tilling. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 389–427. ISBN: 978-3-642-80087-0. DOI: 10.1007/978-3-642-80087-0{_}12.

Neri, A et al. (2003). "Multiparticle simulation of collapsing volcanic columns and pyroclastic flow ". English. In: *Journal of Geophysical Research - Solid Earth* 108.B4. DOI: 10.1029/2001JB000508.

Neri, A et al. (2008). "Developing an Event Tree for probabilistic hazard and risk assessment at Vesuvius". In: *Journal of Volcanology and Geothermal Research* 178.3, pp. 397–415. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2008.05.014.

Neri, A et al. (2015a). "Pyroclastic Density Current Hazards and Risk". In: *Volcanic Hazards, Risks and Disasters*, pp. 109–140. ISBN: 9780123964533. DOI: 10.1016/B978-0-12-396453-3.00005-8.

Neri, A et al. (2015b). "Quantifying volcanic hazard at Campi Flegrei caldera (Italy) with uncertainty assessment: 2. Pyroclastic density current invasion maps". In: *Journal of Geophysical Research: Solid Earth* 120.4, pp. 2330–2349. ISSN: 2169-9313. DOI: https://doi.org/10.1002/2014JB011776.

Núñez-Corrales, S and J Brenes-André (2023). "BALISTICA: A software suite for ballistic motion with applications to geophysics research and education". In: *Software Impacts* 16. ISSN: 2665-9638. DOI: https://doi.org/10.1016/j.simpa.2023.100488.

Oberkampf, W L and T G Trucano (2002). "Verification and validation in computational fluid dynamics". In: *Progress in Aerospace Sciences* 38.3, pp. 209–272. ISSN: 0376-0421. DOI: https://doi.org/10.1016/S0376-0421(02)00005-2.

Oberkampf, W L, T G Trucano, and C Hirsch (2004). "Verification, validation, and predictive capability in computational engineering and physics". In: *Applied Mechanics Reviews* 57.5, pp. 345–384. ISSN: 0003-6900. DOI: https://doi.org/10.1115/1.1767847.

Oddsson, B et al. (2012). "Monitoring of the plume from the basaltic phreatomagmatic 2004 Grímsvötn eruption—application of weather radar and comparison with plume models". In: *Bulletin of Volcanology* 74.6, pp. 1395–1407. ISSN: 1432-0819. DOI: 10.1007/s00445-012-0598-9.

Ogburn, S E (2025). *FlowDat: Volcanic Mass Flow Database*. DOI: https://doi.org/10.5066/P13CMDC2.

Ogburn, S E and E S Calder (2017). "The Relative Effectiveness of Empirical and Physical Models for Simulating the Dense Undercurrent of Pyroclastic Flows under Different Emplacement Conditions". In: *Frontiers in Earth Science* 5. ISSN: 2296-6463. DOI: 10.3389/feart.2017.00083.

Ong, E et al. (2020). "Modelling kidney disease using ontology: insights from the Kidney Precision Medicine Project." In: *Nature reviews. Nephrology* 16.11, pp. 686–696. ISSN: 1759-507X. DOI: 10.1038/s41581-020-00335-w.

Oppenheimer, C et al. (1993). "Infrared image analysis of volcanic thermal features: Lascar Volcano, Chile, 1984–1992". In: *Journal of Geophysical Research: Solid Earth* 98.B3, pp. 4269–4286. ISSN: 0148-0227. DOI: https://doi.org/10.1029/92JB02134.

Oramas-Dorta, D et al. (2012). "Pyroclastic flow hazard at Arenal volcano, Costa Rica: scenarios and assessment". In: *Journal of Volcanology and Geothermal Research* 247-248, pp. 74–92. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2012.07.015.

Oreskes, N, K Shrader-Frechette, and K Belitz (1994). "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences". In: *Science* 263.5147, pp. 641–646. ISSN: 00368075, 10959203. DOI: 10.1126/science.263.5147.641.

Orth, R et al. (2015). "Does model performance improve with complexity? A case study with three hydrological models". In: *Journal of Hydrology* 523, pp. 147–159. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2015.01.044.

Osman, S et al. (2020). "Sensitivity of Volcanic Ash Dispersion Modelling to Input Grain Size Distribution Based on Hydromagmatic and Magmatic Deposits". In: *Atmosphere* 11.6. ISSN: 2073-4433. DOI: 10.3390/atmos11060567.

Osman, S et al. (2024). "Probabilistic hazard analyses for a small island: methods for quantifying tephra fall hazard and appraising possible impacts on Ascension Island". In: *Bulletin of Volcanology* 86.10, p. 82. ISSN: 1432-0819. DOI: 10.1007/s00445-024-01771-3.

Osores, M S et al. (2013). "Validation of the FALL3D model for the 2008 Chaitén eruption using field and satellite data". In: *Andean Geology* 40.2, p. 262. DOI: http://doi.org/10.5027/andgeoV40n2-a05.

Pailot-Bonnétat, S et al. (2020). "Plume Height Time-Series Retrieval Using Shadow in Single Spatial Resolution Satellite Images". In: *Remote Sensing* 12.3951, p. 3951. ISSN: 12233951. DOI: 10.3390/rs12233951.

Paine, A R and F B Wadsworth (2025). "Large explosive eruptions may be dominated by pyroclastic flows instead of buoyant plumes: insights from a global data compilation". In: *Journal of Applied Volcanology* 14.1, p. 3. ISSN: 2191-5040. DOI: 10.1186/s13617-025-00151-6.

Papale, P (2017). "Rational volcanic hazard forecasts and the use of volcanic alert levels". In: *Journal of Applied Volcanology* 6.1, p. 13. ISSN: 2191-5040. DOI: 10.1186/s13617-017-0064-7.

— (2021). "Some relevant issues in volcanic hazard forecasts and management of volcanic crisis". In: *Forecasting and planning for volcanic hazards, risks, and disasters*. Vol. 2. Elsevier, pp. 1–24. ISBN: 978-0-12-818082-2. DOI: https://doi.org/10.1016/B978-0-12-818082-2.00001-9.

Pardini, F et al. (2020). "Ensemble-Based Data Assimilation of Volcanic Ash Clouds from Satellite Observations: Application to the 24 December 2018 Mt. Etna Explosive Eruption". In: *Atmosphere* 11.4. ISSN: 2073-4433. DOI: 10.3390/atmos11040359.

Pardini, F et al. (2022). "Real-time probabilistic assessment of volcanic hazard for tephra dispersal and fallout at Mt. Etna: the 2021 lava fountain episodes". In: *Bulletin of Volcanology* 85.1, p. 6. ISSN: 1432-0819. DOI: 10.1007/s00445-022-01614-z.

Pardini, F et al. (2024). "Dynamics, Monitoring, and Forecasting of Tephra in the Atmosphere". In: *Reviews of Geophysics* 62.4. ISSN: 8755-1209. DOI: https://doi.org/10.1029/2023RG000808.

Pardo, N et al. (2012). "Andesitic Plinian eruptions at Mt. Ruapehu: quantifying the uppermost limits of eruptive parameters". In: *Bulletin of Volcanology* 74.5, pp. 1161–1185. ISSN: 1432-0819. DOI: 10.1007/s00445-012-0588-y.

Parra, R (2019). "Influence of spatial resolution in modeling the dispersion of volcanic ash in Ecuador". In: *WIT Transactions on Ecology and the Environment*. Ed. by G Passerini et al. Vol. 236. Southhampton: WIT Press, pp. 67–78. DOI: 10.2495/AIR190071.

Parra, R et al. (2016). "Eruption Source Parameters for forecasting ash dispersion and deposition from vulcanian eruptions at Tungurahua volcano: Insights from field data from the July 2013 eruption". In: *Journal of Volcanology and Geothermal Research* 309, pp. 1–13. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2015.11.001.

Parry, G W (1996). "The characterization of uncertainty in Probabilistic Risk Assessments of complex systems". In: *Reliability Engineering & System Safety* 54.2, pp. 119–126. ISSN: 0951-8320. DOI: https://doi.org/10.1016/S0951-8320(96)00069-5.

Patra, A K et al. (2005). "Parallel adaptive numerical simulation of dry avalanches over natural terrain". In: *Journal of Volcanology and Geothermal Research* 139.1, pp. 1–21. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2004.06.014.

Patrick, M R et al. (2007). "Strombolian explosive styles and source conditions: insights from thermal (FLIR) video". In: *Bulletin of Volcanology* 69, pp. 769–784. ISSN: 0258-8900. DOI: https://doi.org/10.1007/s00445-006-0107-0.

Paul, P K et al. (2021). "Selecting hydrological models for developing countries: Perspective of global, continental, and country scale models over catchment scale models". In: *Journal of Hydrology* 600. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2021.126561.

Pedersen, G B M et al. (2022). "Lava flow hazard modelling during the 2021 Fagradalsfjall eruption, Iceland: Applications of MrLavaLoba". In: *Natural Hazards and Earth System Sciences Discussions* 2022, pp. 1–38. ISSN: 2195-9269. DOI: https://doi.org/10.5194/nhess-23-3147-2023.

Perttu, A et al. (2020). "Estimates of plume height from infrasound for regional volcano monitoring". In: *Journal of Volcanology and Geothermal Research* 402. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2020.106997.

Peterson, R A and K G Dean (2008). "Forecasting exposure to volcanic ash based on ash dispersion modeling". In: *Journal of Volcanology and Geothermal Research* 170.3, pp. 230–246. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2007.10.003.

Petropoulos, F et al. (2022). "Forecasting: theory and practice". In: *International Journal of Forecasting* 38.3, pp. 705–871. ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2021.11.001.

Pfeiffer, T, A Costa, and G Macedonio (2005). "A model for the numerical simulation of tephra fall deposits". In: *Journal of Volcanology and Geothermal Research* 140.4, pp. 273–294. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2004.09.001.

Phillips, J et al. (2023). "Quantifying uncertainty in probabilistic volcanic ash hazard forecasts, with an application to weather pattern based wind field sampling". In: *Bulletin of Volcanology* 85.11, p. 68. ISSN: 1432-0819. DOI: 10.1007/s00445-023-01664-x.

Phillipson, G, R Sobradelo, and J Gottsmann (2013). "Global volcanic unrest in the 21st century: An analysis of the first decade". In: *Journal of Volcanology and Geothermal Research*

264, pp. 183–196. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2013.08.004.

Pianosi, F and T Wagener (2016). "Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis". In: *Hydrological Processes* 30.22, pp. 3991–4003. ISSN: 0885-6087. DOI: https://doi.org/10.1002/hyp.10968. URL: https://doi.org/10.1002/hyp.10968.

Pianosi, F et al. (2016). "Sensitivity analysis of environmental models: A systematic review with practical workflow". In: *Environmental Modelling & Software* 79, pp. 214–232. ISSN: 1364-8152. DOI: https://doi.org/10.1016/j.envsoft.2016.02.008.

Pinkerton, H and G Norton (1995). "Rheological properties of basaltic lavas at sub-liquidus temperatures: laboratory and field measurements on lavas from Mount Etna". In: *Journal of Volcanology and Geothermal Research* 68.4, pp. 307–323. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(95)00018-7.

Pioli, L, C Bonadonna, and M Pistolesi (2019). "Reliability of Total Grain-Size Distribution of Tephra Deposits". In: *Scientific Reports* 9.1. ISSN: 2045-2322. DOI: 10.1038/s41598-019-46125-8.

Piombo, A and M Dragoni (2009). "Evaluation of flow rate for a one-dimensional lava flow with power-law rheology". In: *Geophysical Research Letters* 36.22. ISSN: 0094-8276. DOI: https://doi.org/10.1029/2009GL041024.

Pistolesi, M et al. (2015). "Complex dynamics of small-moderate volcanic events: the example of the 2011 rhyolitic Cordón Caulle eruption, Chile". In: *Bulletin of Volcanology* 77.1, p. 3. ISSN: 1432-0819. DOI: 10.1007/s00445-014-0898-3.

Pitman, E B et al. (2003). "Computing granular avalanches and landslides". In: *Physics of Fluids* 15.12, pp. 3638–3646. DOI: 10.1063/1.1614253.

Plu, M et al. (2021). "An ensemble of state-of-the-art ash dispersion models: towards probabilistic forecasts to increase the resilience of air traffic against volcanic eruptions". English. In: *Natural Hazards and Earth System Sciences* 21.10, pp. 2973–2992. DOI: 10.5194/nhess-21-2973-2021.

Poland, M P (2014). "Time-averaged discharge rate of subaerial lava at Kīlauea Volcano, Hawai'i, measured from TanDEM-X interferometry: Implications for magma supply and

storage during 2011–2013". In: *Journal of Geophysical Research: Solid Earth* 119.7, pp. 5464–5481. ISSN: 2169-9313. DOI: https://doi.org/10.1002/2014JB011132.

Poland, M P and K R Anderson (2020). "Partly Cloudy With a Chance of Lava Flows: Forecasting Volcanic Eruptions in the Twenty-First Century". In: *Journal of Geophysical Research: Solid Earth* 125.1. ISSN: 2169-9313. DOI: https://doi.org/10.1029/2018JB016974.

Poland, M P et al. (2020). "Forecasting, Detecting, and Tracking Volcanic Eruptions from Space". In: *Remote Sensing in Earth Systems Sciences* 3.1, pp. 55–94. ISSN: 2520-8209. DOI: 10.1007/s41976-020-00034-x.

Poppe, S et al. (2022). "Analog experiments in volcanology: towards multimethod, upscaled, and integrated models". In: *Bulletin of Volcanology* 84.5, p. 52. ISSN: 1432-0819. DOI: 10.1007/s00445-022-01543-x.

Poret, M et al. (2017). "Modelling tephra dispersal and ash aggregation: The 26th April 1979 eruption, La Soufrière St. Vincent". In: *Journal of Volcanology and Geothermal Research* 347, pp. 207–220. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2017.09.012.

Porter, H et al. (2025). "A new volcanic multi-hazard impact model for water supply systems: Application at Taranaki Mounga, Aotearoa New Zealand". In: *International Journal of Disaster Risk Reduction* 116. ISSN: 2212-4209. DOI: https://doi.org/10.1016/j.ijdrr.2024.105113.

Poulidis, A P and M Iguchi (2021). "Model sensitivities in the case of high-resolution Eulerian simulations of local tephra transport and deposition". In: *Atmospheric Research* 247. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2020.105136.

Poulidis, A P, T Takemi, and M Iguchi (2019). "Experimental high-resolution forecasting of volcanic ash hazard at Sakurajima, Japan". In: *Journal of Disaster Research* 14.5, pp. 786–797. ISSN: 1881-2473. DOI: 10.20965/jdr.2019.p0786.

Poulidis, A P et al. (2017). "Orographic effects on the transport and deposition of volcanic ash: A case study of Mount Sakurajima, Japan". In: *Journal of Geophysical Research: Atmospheres* 122.17, pp. 9332–9350. ISSN: 2169-897X. DOI: https://doi.org/10.1002/2017JD026595.

Poulidis, A P et al. (2018). "Meteorological Controls on Local and Regional Volcanic Ash Dispersal". In: *Scientific Reports* 8.1, p. 6873. ISSN: 2045-2322. DOI: 10.1038/s41598-018-24651-1.

Prandtl, L (1954). *The Essentials of Fluid Mechanics*. London: Blackie & Son Limited.

Prata, A J and I F Grant (2001). "Retrieval of microphysical and morphological properties of volcanic ash plumes from satellite data: Application to Mt Ruapehu, New Zealand". In: *Quarterly Journal of the Royal Meteorological Society* 127.576, pp. 2153–2179. ISSN: 0035-9009. DOI: https://doi.org/10.1002/qj.49712757615.

Prata, A T et al. (2021). "FALL3D-8.0: a computational model for atmospheric transport and deposition of particles, aerosols and radionuclides - Part 2: Model validation". English. In: *Geoscientific Model Development* 14.1, pp. 409–436. DOI: 10.5194/gmd-14-409-2021.

Prejean, S G and E E Brodsky (2011). "Volcanic plume height measured by seismic waves based on a mechanical model". In: *Journal of Geophysical Research: Solid Earth* 116.B1. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2010JB007620.

Procter, J et al. (2021). "A review of lahars; past deposits, historic events and present-day simulations from Mt. Ruapehu and Mt. Taranaki, New Zealand". In: *New Zealand Journal of Geology and Geophysics* 64.2-3, pp. 479–503. ISSN: 0028-8306. DOI: 10.1080/00288306.2020.1824999.

Procter, J N, S J Cronin, and M F Sheridan (2012). "Evaluation of Titan2D modelling forecasts for the 2007 Crater Lake break-out lahar, Mt. Ruapehu, New Zealand". In: *Geomorphology* 136.1, pp. 95–105. ISSN: 0169-555X. DOI: https://doi.org/10.1016/j.geomorph.2011.05.001.

Procter, J N et al. (2010a). "Lahar hazard assessment using Titan2D for an alluvial fan with rapidly changing geomorphology: Whangaehu River, Mt. Ruapehu". In: *Geomorphology* 116.1, pp. 162–174. ISSN: 0169-555X. DOI: https://doi.org/10.1016/j.geomorph.2009.10.016.

Procter, J N et al. (2010b). "Mapping block-and-ash flow hazards based on Titan 2D simulations: a case study from Mt. Taranaki, NZ". In: *Natural Hazards* 53.3, pp. 483–501. ISSN: 1573-0840. DOI: 10.1007/s11069-009-9440-x.

Pulido, J R G et al. (2009). "On the Finding Process of Volcano-Domain Ontology Components Using Self-Organizing Maps". In: *Advances in Self-Organizing Maps*. Ed. by José C Príncipe and Risto Miikkulainen. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 255–263. ISBN: 978-3-642-02397-2. DOI: https://doi.org/10.1007/978-3-642-02397-2{_}29.

Puy, A et al. (2021). *sensobol: an R package to compute variance-based sensitivity indices*. URL: https://cran.r-project.org/package=sensobol.

Puy, A et al. (2023). "Models with higher effective dimensions tend to produce more uncertain estimates". In: *Science Advances* 8.42. DOI: 10.1126/sciadv.abn9450.

Pyle, D M (2018). "What Can We Learn from Records of Past Eruptions to Better Prepare for the Future?" In: *Observing the Volcano World: Volcano Crisis Communication*. Ed. by Carina J Fearnley et al. Cham: Springer International Publishing, pp. 445–462. ISBN: 978-3-319-44097-2. DOI: 10.1007/11157{_}2017{_}5.

Pyo, S et al. (2017). "Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the Korean stock markets". In: *PLOS ONE* 12.11. DOI: https://doi.org/10.1371/journal.pone.0188107.

Raber, G T et al. (2007). "Impact of LiDAR nominal post-spacing on DEM accuracy and flood zone delineation". In: *Photogrammetric Engineering & Remote Sensing* 73.7, pp. 793–804. ISSN: 0099-1112. DOI: https://doi.org/10.14358/PERS.73.7.793.

Rabier, F (2005). "Overview of global data assimilation developments in numerical weather-prediction centres". In: *Quarterly Journal of the Royal Meteorological Society* 131.613, pp. 3215–3233. ISSN: 0035-9009. DOI: https://doi.org/10.1256/qj.05.129.

Ramar, K and T T Mirnalinee (2012). "An ontological representation for Tsunami early warning system". In: *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012)*, pp. 93–98.

Renschler, C S (2005). "Scales and uncertainties in using models and GIS for volcano hazard prediction". In: *Journal of Volcanology and Geothermal Research* 139.1, pp. 73–87. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2004.06.016.

Rhoades, D A and M C Gerstenberger (2009). "Mixture Models for Improved Short-Term Earthquake Forecasting". In: *Bulletin of the Seismological Society of America* 99.2A, pp. 636–646. ISSN: 0037-1106. DOI: 10.1785/0120080063.

Richardson, J A (2016). "Modeling the Construction and Evolution of Distributed Volcanic Fields on Earth and Mars". PhD thesis. University of South Florida.

Richter, N et al. (2016). "Lava flow hazard at Fogo Volcano, Cabo Verde, before and after the 2014–2015 eruption". In: *Natural Hazards and Earth System Sciences* 16.8, pp. 1925–1951. ISSN: 1561-8633. DOI: https://doi.org/10.5194/nhess-16-1925-2016.

Rickles, D, P Hawe, and A Shiell (2007). "A simple guide to chaos and complexity". In: *Journal of Epidemiology & Community Health* 61.11, pp. 933–937. ISSN: 0143-005X. DOI: 10.1136/jech.2006.054254.

Ripepe, M, M Rossi, and G Saccorotti (1993). "Image processing of explosive activity at Stromboli". In: *Journal of Volcanology and Geothermal Research* 54.3-4, pp. 335–351. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(93)90071-X.

Ripepe, M et al. (2013). "Ash-plume dynamics and eruption source parameters by infrasound and thermal imagery: The 2010 Eyjafjallajökull eruption". In: *Earth and Planetary Science Letters* 366, pp. 112–121. ISSN: 0012-821X. DOI: https://doi.org/10.1016/j.epsl.2013.02.005.

Riva, F et al. (2023). "Toward a cohesive understanding of ecological complexity". In: *Science Advances* 9.25. DOI: 10.1126/sciadv.abq4207.

Robinson, S (2023). "Exploring the relationship between simulation model accuracy and complexity". In: *Journal of the Operational Research Society* 74.9, pp. 1992–2011. ISSN: 0160-5682. DOI: 10.1080/01605682.2022.2122740.

Robinson, S and R Brooks (2024). "Assumptions and simplifications in discrete-event simulation modelling". In: *Journal of Simulation*, pp. 1–18. ISSN: 1747-7778. DOI: 10.1080/17477778.2024.2407369.

Roche, O et al. (2004). "Experimental study of gas-fluidized granular flows with implications for pyroclastic flow emplacement". In: *Journal of Geophysical Research: Solid Earth* 109.B10. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2003JB002916.

Romero, J E et al. (2016). "Eruption dynamics of the 22–23 April 2015 Calbuco Volcano (Southern Chile): Analyses of tephra fall deposits". In: *Journal of Volcanology and Geothermal Research* 317, pp. 15–29. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2016.02.027.

Romero, J E et al. (2018). "Tephra From the 3 March 2015 Sustained Column Related to Explosive Lava Fountain Activity at Volcán Villarrica (Chile)". In: *Frontiers in Earth Science* 6. ISSN: 2296-6463. DOI: https://doi.org/10.3389/feart.2018.00098.

Rongo, R et al. (2008). "Lava Flow Hazard Evaluation Through Cellular Automata and Genetic Algorithms: an Application to Mt Etna Volcano". In: *Fundamenta Informaticae* 87.2, pp. 247–267. DOI: 10.3233/FUN-2008-87208.

Rose, W I and A J Durant (2009). "El Chichón volcano, April 4, 1982: volcanic cloud history and fine ash fallout". In: *Natural Hazards* 51.2, pp. 363–374. ISSN: 1573-0840. DOI: 10.1007/s11069-008-9283-x.

— (2011). "Fate of volcanic ash: Aggregation and fallout". In: *Geology* 39.9, pp. 895–896. ISSN: 0091-7613. DOI: 10.1130/focus092011.1.

Rossi, E, C Bonadonna, and W Degruyter (2019). "A new strategy for the estimation of plume height from clast dispersal in various atmospheric and eruptive conditions". In: *Earth and Planetary Science Letters* 505, pp. 1–12. ISSN: 0012-821X. DOI: https://doi.org/10.1016/j.epsl.2018.10.007.

Rougier, J and K J Beven (2013). "Model and data limitations: the sources and implications of epistemic uncertainty". In: *Risk and Uncertainty Assessment for Natural Hazards*. Vol. 40. Cambridge University Press, New York, pp. 40–63. DOI: https://doi.org/10.1017/CBO9781139047562.004.

Rowley, P J et al. (2014). "Experimental study of dense pyroclastic density currents using sustained, gas-fluidized granular flows". In: *Bulletin of Volcanology* 76, pp. 1–13. ISSN: 0258-8900. DOI: https://doi.org/10.1007/s00445-014-0855-1.

Ruano, M V et al. (2012). "An improved sampling strategy based on trajectory design for application of the Morris method to systems with many input factors". In: *Environmental Modelling & Software* 37, pp. 103–109. ISSN: 1364-8152. DOI: https://doi.org/10.1016/j.envsoft.2012.03.008.

Rupp, B et al. (2006). "Computational modeling of the 1991 block and ash flows at Colima Volcano, Mexico". In: *Neogene-Quaternary Continental Margin Volcanism: A perspective from México*. Ed. by C Siebe, J L Macias Gerardo, and J Aguirre-Diaz. The Geological Society of America. ISBN: 9780813724027. DOI: https://doi.org/10.1130/2006.2402(11).

Rutarindwa, R et al. (2019). "Dynamic Probabilistic Hazard Mapping in the Long Valley Volcanic Region CA: Integrating Vent Opening Maps and Statistical Surrogates of Physical Models of Pyroclastic Density Currents". In: *Journal of Geophysical Research: Solid Earth* 124.9, pp. 9600–9621. ISSN: 2169-9313. DOI: https://doi.org/10.1029/2019JB017352.

Salt, J D (2008). "The seven habits of highly defective simulation projects". In: *Journal of Simulation* 2.3, pp. 155–161. ISSN: 1747-7778. DOI: doi:10.1057/jos.2008.7.

Saltelli, A (2002). "Making best use of model evaluations to compute sensitivity indices". In: *Computer Physics Communications* 145.2, pp. 280–297. ISSN: 0010-4655. DOI: https://doi.org/10.1016/S0010-4655(02)00280-1.

Saltelli, A, S Tarantola, and K P.-S. Chan (1999). "A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output". In: *Technometrics* 41.1, pp. 39–56. ISSN: 0040-1706. DOI: 10.1080/00401706.1999.10485594.

Saltelli, A et al. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models*. Vol. 1. Wiley Online Library. ISBN: 0-470-87093-1. DOI: 10.1002/0470870958.

Saltelli, A et al. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons. ISBN: 0470725176. DOI: 10.1002/978047072518.

Saltelli, A et al. (2010). "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index". In: *Computer physics communications* 181.2, pp. 259–270. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2009.09.018.

Saltelli, A et al. (2020). "Five ways to ensure that models serve society: a manifesto". In: *Nature* 582, pp. 482–484. DOI: doi.org/10.1038/d41586-020-01812-9.

Salter, J M, H N Webster, and C Saint (2024). "Technical note: Exploring parameter and meteorological uncertainty via emulation in volcanic ash atmospheric dispersion modelling". In: *Atmospheric Chemistry and Physics* 24.10, pp. 6251–6274. DOI: 10.5194/acp-24-6251-2024.

Sandri, L, V Acocella, and C Newhall (2017). "Searching for patterns in caldera unrest". In: *Geochemistry, Geophysics, Geosystems* 18.7, pp. 2748–2768. ISSN: 1525-2027. DOI: https://doi.org/10.1002/2017GC006870.

Sandri, L, P Tierz, and S C Loughlin (2025). "From Eruptive Histories to Volcano Monitoring: Probabilistic Eruption Forecasting and Volcanic Hazard Assessment at Varying Temporal and Spatial Scales". In: *Modern Volcano Monitoring*. Ed. by Zack Spica and Corentin Caudron. Cham: Springer Nature Switzerland, pp. 365–397. ISBN: 978-3-031-86841-2. DOI: 10.1007/978-3-031-86841-2{_}12.

Sandri, L et al. (2016). "Beyond eruptive scenarios: Assessing tephra fallout hazard from Neapolitan volcanoes". English. In: *Scientific Reports* 6. ISSN: 20452322. DOI: 10.1038/srep24271.

Sandri, L et al. (2024). "Lahar events in the last 2000 years from Vesuvius eruptions – Part3: Hazard assessment over the Campanian Plain". In: *Solid Earth* 15.4, pp. 459–476. DOI: 10.5194/se-15-459-2024.

Sato, E, K Fukui, and T Shimbori (2018). "Aso volcano eruption on October 8, 2016, observed by weather radars". In: *Earth, Planets and Space* 70.1, p. 105. ISSN: 1880-5981. DOI: 10.1186/s40623-018-0879-4.

Scandone, R, K V Cashman, and S D Malone (2007). "Magma supply, magma ascent and the style of volcanic eruptions". In: *Earth and Planetary Science Letters* 253.3-4, pp. 513–529. ISSN: 0012821X. DOI: 10.1016/j.epsl.2006.11.016.

Scher, S and G Messori (2019). "Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground". In: *Geoscientific Model Development* 12.7, pp. 2797–2809. DOI: 10.5194/gmd-12-2797-2019.

Schilling, S P (1998). "LAHARZ; GIS programs for automated mapping of lahar-inundation hazard zones". ENGLISH. In: *Open-File Report*. DOI: 10.3133/ofr98638. URL: http://pubs.er.usgs.gov/publication/ofr98638.

Schorlemmer, D et al. (2007). "Earthquake likelihood model testing". In: *Seismological Research Letters* 78.1, pp. 17–29. DOI: 10.1785/gssrl.78.1.17.

Schorlemmer, D et al. (2010). "First results of the regional earthquake likelihood models experiment". In: *Pure and Applied Geophysics*, pp. 5–22. DOI: https://doi.org/10.1007/s00024-010-0081-5.

Schorlemmer, D et al. (2018). "The Collaboratory for the Study of Earthquake Predictability: Achievements and Priorities". In: *Seismological Research Letters* 89.4, pp. 1305–1313. ISSN: 0895-0695. DOI: 10.1785/0220180053.

Schoups, G, N C van de Giesen, and H H G Savenije (2008). "Model complexity control for hydrologic prediction". In: *Water Resources Research* 44.12. ISSN: 0043-1397. DOI: https://doi.org/10.1029/2008WR006836.

Schwaiger, H F, R P Denlinger, and L G Mastin (2012). *Ash3d; a finite-volume, conservative numerical model for ash transport and tephra deposition*. English. Tech. rep. B4. Washington, DC. DOI: 10.1029/2011JB008968.

Scollo, S, P Del Carlo, and M Coltelli (2007). "Tephra fallout of 2001 Etna flank eruption: Analysis of the deposit and plume dispersion". In: *Journal of Volcanology and Geothermal Research* 160.1, pp. 147–164. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2006.09.007.

Scollo, S, A Folch, and A Costa (2008). "A parametric and comparative study of different tephra fallout models". English. In: *Journal of Volcanology and Geothermal Research* 176.2, pp. 199–211. DOI: 10.1016/j.jvolgeores.2008.04.002.

Scollo, S et al. (2008). "Sensitivity analysis and uncertainty estimation for tephra dispersal models". In: *Journal of Geophysical Research: Solid Earth* 113.B6. ISSN: 0148-0227. DOI: https://doi.org/10.1029/2006JB004864.

Scollo, S et al. (2019). "Near-Real-Time Tephra Fallout Assessment at Mt. Etna, Italy". In: *Remote Sensing* 11.24. ISSN: 2072-4292. DOI: 10.3390/rs11242987.

Scott, B J and V E Neall (2016). "The Volcano Crises at Ruapehu - 1995 and 1996". In: *Volcanic Hazards and Emergency Management in the Southwest Pacific*. Ed. by P W Taylor. Suva, pp. 119–125.

Scott, E, M Whitehead, and J Procter (2026). "Exploring the role of model classification, complexity, and selection in volcanic hazard forecasting". In: *Computers & Geosciences* 207. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2025.106070.

Scott, E et al. (2025). "Global sensitivity analysis of models for volcanic ash forecasting". In: *Journal of Volcanology and Geothermal Research* 466. ISSN: 0377-0273. DOI: 10.1016/j.jvolgeores.2025.108393.

Scott, E. et al. (2022). "Development of a Bayesian event tree for short-term eruption onset forecasting at Taupō volcano". In: *Journal of Volcanology and Geothermal Research* 432. ISSN: 03770273. DOI: 10.1016/j.jvolgeores.2022.107687.

Searcy, C, K Dean, and W Stringer (1998). "PUFF: A high-resolution volcanic ash tracking model". In: *Journal of Volcanology and Geothermal Research* 80.1, pp. 1–16. ISSN: 0377-0273. DOI: https://doi.org/10.1016/S0377-0273(97)00037-1.

Sehlke, A et al. (2014). "Pahoehoe to 'a'a transition of Hawaiian lavas: an experimental study". In: *Bulletin of Volcanology* 76, pp. 1–20. ISSN: 0258-8900. DOI: https://doi.org/10.1007/s00445-014-0876-9.

Selva, J et al. (2014). "Probabilistic short-term volcanic hazard in phases of unrest: A case study for tephra fallout". In: *Journal of Geophysical Research: Solid Earth* 119.12, pp. 8805–8826. ISSN: 2169-9313. DOI: https://doi.org/10.1002/2014JB011252.

Selva, J et al. (2018). "Sensitivity test and ensemble hazard assessment for tephra fallout at Campi Flegrei, Italy". In: *Journal of Volcanology and Geothermal Research* 351, pp. 1–28. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2017.11.024.

Seppelt, R (2003). *Computer-based environmental management*. John Wiley & Sons. ISBN: 352730732X. DOI: DOI:10.1002/9783527611515.

Sheldrake, T (2014). "Long-term forecasting of eruption hazards: A hierarchical approach to merge analogous eruptive histories". In: *Journal of Volcanology and Geothermal Research* 286, pp. 15–23. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2014.08.021.

Sheldrake, T E et al. (2017). "Understanding causality and uncertainty in volcanic observations: An example of forecasting eruptive activity on Soufrière Hills Volcano, Montserrat". In: *Journal of Volcanology and Geothermal Research* 341, pp. 287–300. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2017.06.007.

Sheridan, M F and J L Macías (1995). "Estimation of risk probability for gravity-driven pyroclastic flows at Volcan Colima, Mexico". In: *Journal of Volcanology and Geothermal Research* 66.1, pp. 251–256. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(94)00058-0.

Sheridan, M F and M C Malin (1983). "Application of computer-assisted mapping to volcanic hazard evaluation of surge eruptions: Vulcano, lipari, and vesuvius". In: *Journal of Volcanology and Geothermal Research* 17.1, pp. 187–202. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(83)90067-7.

Sheridan, M F et al. (2005). "Evaluating Titan2D mass-flow model using the 1963 Little Tahoma Peak avalanches, Mount Rainier, Washington". In: *Journal of Volcanology and Geothermal Research* 139.1, pp. 89–102. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2004.06.011.

Singh, S (2018). *Understanding the Bias-Variance Tradeoff*. URL: https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229#:~:text=What%20is%20bias%3F,on%20training%20and%20test%20data..

Snowling, S D and J R Kramer (2001). "Evaluating modelling uncertainty for model selection". In: *Ecological Modelling* 138.1, pp. 17–30. ISSN: 0304-3800. DOI: https://doi.org/10.1016/S0304-3800(00)00390-2.

Sobol', I M (1990). "On sensitivity estimation for nonlinear mathematical models". In: *Matematicheskoe modelirovanie* 2.1, pp. 112–118. ISSN: 0234-0879.

Sork, A et al. (2025). "Size matters: a new view of the relationship between shape and size for molten volcanic ballistics". In: *Bulletin of Volcanology* 87.8, p. 61. ISSN: 1432-0819. DOI: 10.1007/s00445-025-01845-w.

Sparks, R S J (1976). "Grain size variations in ignimbrites and implications for the transport of pyroclastic flows". In: *Sedimentology* 23.2, pp. 147–188. ISSN: 0037-0746. DOI: https://doi.org/10.1111/j.1365-3091.1976.tb00045.x.

— (1986). "The dimensions and dynamics of volcanic eruption columns". In: *Bulletin of Volcanology* 48.1, pp. 3–15. ISSN: 1432-0819. DOI: 10.1007/BF01073509.

— (2003). "Forecasting volcanic eruptions". In: *Earth and Planetary Science Letters* 210.1-2, pp. 1–15. DOI: 10.1016/S0012-821X(03)00124-9.

Sparks, R S J and W P Aspinall (2004). "Volcanic activity: frontiers and challenges in forecasting, prediction and risk assessment". In: *The State of the Planet: Frontiers and Challenges in Geophysics*. Vol. 150. American Geophysical Union Washington, DC, pp. 359–371. DOI: https://doi.org/10.1029/150GM28.

Sparks, R S J and K V Cashman (2017). "Dynamic Magma Systems: Implications for Forecasting Volcanic Activity". In: *Elements* 13.1, pp. 35–40. ISSN: 1811-5209. DOI: 10.2113/gselements.13.1.35.

Sparks, R S J et al. (1997). *Volcanic plumes*. Wiley. ISBN: 0471939013.

Sparks, R S J et al. (2013). "Risk and uncertainty assessment of volcanic hazards". In: *Risk and Uncertainty Assessment for Natural Hazards*. Ed. by Jonathan Rougier, Steve Sparks, and Lisa J Hill. Cambridge: Cambridge University Press, pp. 364–397. ISBN: 9781139047562. DOI: DOI:10.1017/CBO9781139047562.012.

Spataro, D et al. (2015). "The new SCIARA-fv3 numerical model and acceleration by GPGPU strategies". In: *The International Journal of High Performance Computing Applications* 31.2, pp. 163–176. ISSN: 1094-3420. DOI: 10.1177/1094342015584520.

Spence, R J S et al. (2005). "Residential building and occupant vulnerability to tephra fall". In: *Natural Hazards Earth System Sciences* 5.4, pp. 477–494. ISSN: 1684-9981. DOI: 10.5194/nhess-5-477-2005.

Spieler, D and N Schütze (2024). "Investigating the Model Hypothesis Space: Benchmarking Automatic Model Structure Identification With a Large Model Ensemble". In: *Water Resources Research* 60.7. ISSN: 0043-1397. DOI: https://doi.org/10.1029/2023WR036199.

Spiller, E T et al. (2020). "Volcanic Hazard Assessment for an Eruption Hiatus, or Post-eruption Unrest Context: Modeling Continued Dome Collapse Hazards for Soufrière Hills Volcano". In: *Frontiers in Earth Science* Volume 8 - 2020. ISSN: 2296-6463. DOI: https://doi.org/10.3389/feart.2020.535567.

Srikrishnan, V and K Keller (2021). "Small increases in agent-based model complexity can result in large increases in required calibration data". In: *Environmental Modelling & Software* 138. ISSN: 1364-8152. DOI: https://doi.org/10.1016/j.envsoft.2021.104978.

Steel, M F (2011). "Bayesian model averaging and forecasting". In: *Bulletin of EU and US Inflation and Macroeconomic Analysis* 200, pp. 30–41.

Stefanescu, E R, M Bursik, and A K Patra (2012). "Effect of digital elevation model on Mohr-Coulomb geophysical flow model output". In: *Natural Hazards* 62.2, pp. 635–656. ISSN: 1573-0840. DOI: 10.1007/s11069-012-0103-y.

Stein, A F et al. (2015). "NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System". English. In: *Bulletin of the American Meteorological Society* 96.12, pp. 2059–2077. DOI: 10.1175/BAMS-D-14-00110.1.

Stein, M (1987). "Large Sample Properties of Simulations Using Latin Hypercube Sampling". In: *Technometrics* 29.2, pp. 143–151. ISSN: 00401706. DOI: 10.2307/1269769.

Stevens, N F, V Manville, and D W Heron (2003). "The sensitivity of a volcanic flow model to digital elevation model accuracy: experiments with digitised map contours and interferometric SAR at Ruapehu and Taranaki volcanoes, New Zealand ". English. In: *Journal of Volcanology and Geothermal Research* 119.1, pp. 89–105. DOI: 10.1016/S0377-0273(02)00307-4.

Stevens, N F, G Wadge, and J B Murray (1999). "Lava flow volume and morphology from digitised contour maps: a case study at Mount Etna, Sicily". In: *Geomorphology* 28.3, pp. 251–261. ISSN: 0169-555X. DOI: https://doi.org/10.1016/S0169-555X(98)00115-9.

Stinton, A J et al. (2004). "Incorporation of variable bed friction into TITAN2D mass-flow model: Application to Little Tahoma peak avalanches (Washington)". In: *Acta vulcanologica: Journal of the National Volcanic Group of Italy: 16, 1/2, 2004*, pp. 1000–1011. ISSN: 1724-0425.

Stone, J (2010). *Pyroclastic flows move fast and destroy everything in their path*. URL: https://www.usgs.gov/programs/VHP/pyroclastic-flows-move-fast-and-destroy-everything-their-path.

Sulpizio, R and P Dellino (2008). "Sedimentology, depositional mechanisms and pulsating behaviour of pyroclastic density currents". In: *Developments in Volcanology* 10, pp. 57–96. ISSN: 1871-644X. DOI: https://doi.org/10.1016/S1871-644X(07)00002-2.

Sulpizio, R et al. (2010). "Predicting the block-and-ash flow inundation areas at Volcán de Colima (Colima, Mexico) based on the present day (February 2010) status". In: *Journal of Volcanology and Geothermal Research* 193.1, pp. 49–66. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2010.03.007.

Sulpizio, R et al. (2012). "Hazard assessment of far-range volcanic ash dispersal from a violent Strombolian eruption at Somma-Vesuvius volcano, Naples, Italy: implications on civil

aviation". In: *Bulletin of Volcanology* 74.9, pp. 2205–2218. ISSN: 1432-0819. DOI: 10.1007/s00445-012-0656-3.

Sulpizio, R et al. (2014). "Pyroclastic density currents: state of the art and perspectives". In: *Journal of Volcanology and Geothermal Research* 283, pp. 36–65. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2014.06.014.

Suzuki, T (1983). "A Theoretical Model for Dispersion of Tephra". In: *Arc Volcanism: Physics and Tectonics*, pp. 95–113.

Szypuła, B (2019). "Quality assessment of DEM derived from topographic maps for geomorphometric purposes". In: *Open Geosciences* 11.1, pp. 843–865. ISSN: 2391-5447. DOI: DOI:10.1515/geo-2019-0066.

Taddeucci, J et al. (2017). "In-flight dynamics of volcanic ballistic projectiles". In: *Reviews of Geophysics* 55.3, pp. 675–718. ISSN: 8755-1209. DOI: https://doi.org/10.1002/2017RG000564.

Tadini, A et al. (2020). "Quantifying the Uncertainty of a Coupled Plume and Tephra Dispersal Model: PLUME-MOM/HYSPLIT Simulations Applied to Andean Volcanoes". English. In: *Journal of Geophysical Research. Solid Earth* 125.2, n/a. DOI: 10.1029/2019JB018390.

Tadini, A et al. (2021). "Eruption type probability and eruption source parameters at Cotopaxi and Guagua Pichincha volcanoes (Ecuador) with uncertainty quantification". In: *Bulletin of Volcanology* 83.5, p. 35. ISSN: 1432-0819. DOI: 10.1007/s00445-021-01458-z.

Tanaka, H L (1994). "Development of a prediction scheme for volcanic ash fall from Redoubt Volcano, Alaska, in Casadevall, T.J., Volcanic ash and aviation safety; proceedings of the first international symposium on volcanic ash and aviation safety". In: *U.S. Geological Survey Bulletin* 2047, pp. 283–291.

Tanaka, H L and M Iguchi (2019). "Numerical Simulations of Volcanic Ash Plume Dispersal for Sakura-Jima Using Real-Time Emission Rate Estimation ". English. In: *Journal of Disaster Research* 14.1, pp. 160–172. DOI: 10.20965/jdr.2019.p0160.

Tanaka, H L, H Nakamichi, and M Iguchi (2020). "PUFF Model Prediction of Volcanic Ash Plume Dispersal for Sakurajima Using MP Radar Observation". In: *Atmosphere* 11.11. DOI: 10.3390/atmos11111240.

Tarquini, S and M Favalli (2011). "Mapping and DOWNFLOW simulation of recent lava flow fields at Mount Etna". In: *Journal of Volcanology and Geothermal Research* 204.1-4, pp. 27–39. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2011.05.001.

— (2013). "Uncertainties in lava flow hazard maps derived from numerical simulations: The case study of Mount Etna". In: *Journal of Volcanology and Geothermal Research* 260, pp. 90–102. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2013.04.017.

Tennant, E et al. (2021). "Reconstructing eruptions at a data limited volcano: A case study at Gede (West Java)". In: *Journal of Volcanology and Geothermal Research* 418. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2021.107325.

Thalheim, B (2011). "The Science of Conceptual Modelling". In: *Database and Expert Systems Applications*. Ed. by Abdelkader Hameurlain et al. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 12–26. ISBN: 978-3-642-23088-2. DOI: https://doi.org/10.1007/978-3-642-23088-2{_}2.

— (2019). "Conceptual Models and Their Foundations". In: *Model and Data Engineering*. Ed. by Klaus-Dieter Schewe and Neeraj Kumar Singh. Springer International Publishing, pp. 123–139. ISBN: 978-3-030-32065-2. DOI: https://doi.org/10.1007/978-3-030-32065-2{_}9.

Thivet, S et al. (2025). "In situ volcanic ash sampling and aerosol–gas analysis based on UAS technologies (AeroVolc)". In: *Atmospheric Measurement Techniques* 18.12, pp. 2803–2824. DOI: 10.5194/amt-18-2803-2025.

Thompson, M A, J M Lindsay, and J C Gaillard (2015). "The influence of probabilistic volcanic hazard map properties on hazard communication". In: *Journal of Applied Volcanology* 4.1, p. 6. ISSN: 2191-5040. DOI: 10.1186/s13617-015-0023-0.

Thompson, M A et al. (2017). "Quantifying risk to agriculture from volcanic ashfall: a case study from the Bay of Plenty, New Zealand". In: *Natural Hazards* 86.1, pp. 31–56. ISSN: 1573-0840. DOI: 10.1007/s11069-016-2672-7.

Thornes, J E and D B Stephenson (2001). "How to judge the quality and value of weather forecast products". In: *Meteorological Applications* 8.3, pp. 307–314. DOI: 10.1017/S1350482701003061.

Thouret, J.-C. et al. (2020). "Lahars and debris flows: Characteristics and impacts". In: *Earth-Science Reviews* 201. ISSN: 0012-8252. DOI: https://doi.org/10.1016/j.earscirev.2019.103003.

Tierz, P (2020). "Long-Term Probabilistic Volcanic Hazard Assessment Using Open and Non-open Data: Observations and Current Issues". In: *Frontiers in Earth Science* 8. ISSN: 2296-6463. DOI: https://doi.org/10.3389/feart.2020.00257.

Tilling, R I (1989). "Volcanic hazards and their mitigation: Progress and problems". In: *Reviews of Geophysics (1985)* 27.2, pp. 237–269. DOI: 10.1029/RG027i002p00237.

— (2022). "14 The role of monitoring in forecasting volcanic events". In: *Monitoring Active Volcanoes: Strategies, Procedures and Techniques*. Ed. by B McGuire, C R J Kilburn, and J Murray. London: Taylor & Francis, p. 369. DOI: https://doi.org/10.4324/9781003327080.

Titos, M et al. (2022). "Long-term hazard assessment of explosive eruptions at Jan Mayen (Norway) and implications for air traffic in the North Atlantic". In: *Natural Hazards and Earth System Sciences* 22.1, pp. 139–163. ISSN: 1561-8633. DOI: https://doi.org/10.5194/nhess-22-139-2022.

Tomii, Y et al. (2020). "Estimation of volcanic ashfall deposit and removal works based on ash dispersion simulations". In: *Natural Hazards* 103.3, pp. 3377–3399. ISSN: 1573-0840. DOI: 10.1007/s11069-020-04134-1.

Tomsen, E et al. (2014). "Evacuation planning in the Auckland Volcanic Field, New Zealand: a spatio-temporal approach for emergency management and transportation network decisions". In: *Journal of Applied Volcanology* 3.1, p. 6. ISSN: 2191-5040. DOI: 10.1186/2191-5040-3-6.

Topcu, T G and B L Mesmer (2018). "Incorporating end-user models and associated uncertainties to investigate multiple stakeholder preferences in system design". In: *Research in Engineering Design* 29.3, pp. 411–431. ISSN: 1435-6066. DOI: 10.1007/s00163-017-0276-1.

Toth, Z and E Kalnay (1993). "Ensemble Forecasting at NMC: The Generation of Perturbations". English. In: *Bulletin of the American Meteorological Society* 74.12, pp. 2317–2330. DOI: 10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.

Toyos, G P et al. (2007). "A GIS-based methodology for hazard mapping of small volume pyroclastic density currents". In: *Natural Hazards* 41, pp. 99–112. ISSN: 0921-030X. DOI: https://doi.org/10.1007/s11069-006-9026-9.

Trancoso, R et al. (2022). "Towards real-time probabilistic ash deposition forecasting for Aotearoa New Zealand". In: *Journal of Applied Volcanology* 11. DOI: https://doi.org/10.1186/s13617-022-00123-0.

Tredennick, A T et al. (2021). "A practical guide to selecting models for exploration, inference, and prediction in ecology". In: *Ecology* 102.6. ISSN: 0012-9658. DOI: https://doi.org/10.1002/ecy.3336.

Trucano, T G et al. (2006). "Calibration, validation, and sensitivity analysis: What's what". In: *Reliability Engineering & System Safety* 91.10, pp. 1331–1357. ISSN: 0951-8320. DOI: https://doi.org/10.1016/j.ress.2005.11.031.

Tsuji, T, N Nishizaka, and K Ohnishi (2020). "Influence of particle aggregation on the tephra dispersal and sedimentation from the October 8, 2016, eruption of Aso volcano". In: *Earth, Planets and Space* 72.1, p. 104. ISSN: 1880-5981. DOI: 10.1186/s40623-020-01233-y.

Tsunematsu, K et al. (2011). "Comparison of Two Advection-Diffusion Methods for Tephra Transport in Volcanic Eruptions". In: *Communications in Computational Physics* 9. DOI: 10.4208/cicp.311009.191110s.

— (2014). "A numerical model of ballistic transport with collisions in a volcanic setting". In: *Computers & Geosciences* 63, pp. 62–69. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2013.10.016.

Tsunematsu, K et al. (2016). "Estimation of ballistic block landing energy during 2014 Mount Ontake eruption". In: *Earth, Planets and Space* 68, pp. 1–11. DOI: https://doi.org/10.1186/s40623-016-0463-8.

Turner, M B et al. (2008). "Developing probabilistic eruption forecasts for dormant volcanoes: a case study from Mt Taranaki, New Zealand". In: *Bulletin of Volcanology* 70.4, pp. 507–515. ISSN: 1432-0819. DOI: 10.1007/s00445-007-0151-4.

Turner, R and T Hurst (2001). "Factors Influencing Volcanic Ash Dispersal from the 1995 and 1996 Eruptions of Mount Ruapehu, New Zealand". In: *Journal of Applied Meteorology and*

Climatology 40.1, pp. 56–69. ISSN: 08948763, 15200450. DOI: https://doi.org/10.1175/1520-0450(2001)040<0056:FIVADF>2.0.CO;2.

Turner, R et al. (2014). "The use of Numerical Weather Prediction and a Lagrangian transport (NAME-III) and dispersion (ASHFALL) models to explain patterns of observed ash deposition and dispersion following the August 2012 Te Maari, New Zealand eruption". In: *Journal of Volcanology and Geothermal Research* 286, pp. 437–451. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2014.05.017.

UNDRR (2017). *Terminology*. URL: https://www.undrr.org/drr-glossary/terminology.

University of South Florida (n.d.). "Tephra2 Forward Modeling Tool". In: *Geoscience Community Codes* (). URL: https://gscommunitycodes.usf.edu/geoscicommunitycodes/public/tephra2/forward-model.php.

Valentine, G A (1998). "Damage to structures by pyroclastic flows and surges, inferred from nuclear weapons effects". In: *Journal of Volcanology and Geothermal Research* 87.1-4, pp. 117–140. ISSN: 0377-0273. DOI: https://doi.org/10.1016/S0377-0273(98)00094-8.

Valentine, G A and K H Wohletz (1989). "Numerical models of Plinian eruption columns and pyroclastic flows". In: *Journal of Geophysical Research: Solid Earth* 94.B2, pp. 1867–1887. ISSN: 0148-0227. DOI: https://doi.org/10.1029/JB094iB02p01867.

Van Eaton, A R et al. (2016). "Volcanic lightning and plume behavior reveal evolving hazards during the April 2015 eruption of Calbuco volcano, Chile". In: *Geophysical Research Letters* 43.7, pp. 3563–3571. DOI: 10.1002/2016GL068076.

Verolino, A et al. (2022). "Assessing volcanic hazard and exposure to lava flows at remote volcanic fields: a case study from the Bolaven Volcanic Field, Laos". In: *Journal of Applied Volcanology* 11.1, p. 6. ISSN: 2191-5040. DOI: 10.1186/s13617-022-00116-z.

Vicari, A et al. (2007). "Modeling of the 2001 lava flow at Etna volcano by a cellular automata approach". In: *Environmental Modelling & Software* 22.10, pp. 1465–1471. ISSN: 1364-8152. DOI: https://doi.org/10.1016/j.envsoft.2006.10.005.

Vicari, A et al. (2011). "Near-real-time forecasting of lava flow hazards during the 12–13 January 2011 Etna eruption". In: *Geophysical Research Letters* 38.13. ISSN: 0094-8276. DOI: https://doi.org/10.1029/2011GL047545.

Volentik, A C M and B F Houghton (2015). "Tephra fallout hazards at Quito International Airport (Ecuador)". In: *Bulletin of Volcanology* 77.6, p. 50. ISSN: 1432-0819. DOI: 10.1007/s00445-015-0923-1.

Volentik, A C M et al. (2010). "Modeling tephra dispersal in absence of wind: Insights from the climactic phase of the 2450BP Plinian eruption of Pululagua volcano (Ecuador)". In: *Journal of Volcanology and Geothermal Research* 193.1, pp. 117–136. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2010.03.011.

Walker, G P L (1973). "Mount Etna and the 1971 eruption-lengths of lava flows". In: *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 274.1238, pp. 107–118. ISSN: 0080-4614. DOI: https://doi.org/10.1098/rsta.1973.0030.

Walker, W E et al. (2003). "Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support". In: *Integrated Assessment* 4.1, pp. 5–17. ISSN: 1389-5176. DOI: 10.1076/iaij.4.1.5.16466.

Walko, R L, C J Tremback, and M J Bell (1995). "HYPACT: The Hybrid Particle and Concentration Transport Model". In: *User's guide* 13. URL: http://www.atmet.com/html/docs/hypact/hyp_ug.pdf.

Wantim, M N et al. (2013). "Numerical experiments on the dynamics of channelised lava flows at Mount Cameroon volcano with the FLOWGO thermo-rheological model". In: *Journal of Volcanology and Geothermal Research* 253, pp. 35–53. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2012.12.003.

Ward, E J et al. (2014). "Complexity is costly: a meta-analysis of parametric and non-parametric methods for short-term population forecasting". In: *Oikos* 123.6, pp. 652–661. ISSN: 0030-1299. DOI: https://doi.org/10.1111/j.1600-0706.2014.00916.x.

Wardman, J B et al. (2012). "Potential impacts from tephra fall to electric power systems: a review and mitigation strategies". In: *Bulletin of Volcanology* 74.10, pp. 2221–2241. ISSN: 02588900. DOI: 10.1007/s00445-012-0664-3.

Waythomas, C F and L G Mastin (2020). "Mechanisms for ballistic block ejection during the 2016–2017 shallow submarine eruption of Bogoslof volcano, Alaska". In: *Bulletin of Volcanology* 82.2, p. 13. ISSN: 1432-0819. DOI: 10.1007/s00445-019-1351-4.

Webley, P W, B J B Stunder, and K G Dean (2009). "Preliminary sensitivity study of eruption source parameters for operational volcanic ash cloud transport and dispersion models — A case study of the August 1992 eruption of the Crater Peak vent, Mount Spurr, Alaska". In: *Journal of Volcanology and Geothermal Research* 186.1, pp. 108–119. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2009.02.012.

Weir, A M et al. (2024). "Approaching the challenge of multi-phase, multi-hazard volcanic impact assessment through the lens of systemic risk: application to Taranaki Mounga". In: *Natural Hazards*. ISSN: 1573-0840. DOI: 10.1007/s11069-023-06386-z.

Weng, Z et al. (2025). "Using EnKF Data Assimilation to Improve Predictions of Volcanic Ash Dispersion". In: *Journal of Geophysical Research: Atmospheres* 130.8. ISSN: 2169-897X. DOI: https://doi.org/10.1029/2024JD042215.

Weron, R (2014). "Electricity price forecasting: A review of the state-of-the-art with a look into the future". In: *International Journal of Forecasting* 30.4, pp. 1030–1081. ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2014.08.008.

Whitehead, M G and M S Bebbington (2021). "Method selection in short-term eruption forecasting". In: *Journal of Volcanology and Geothermal Research* 419. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2021.107386.

Widiwijayanti, C et al. (2004). "Validation of TITAN2D flow model code for pyroclastic flows and debris avalanches at Soufrière Hills Volcano, Montserrat, BWI". In: *AGU Fall Meeting Abstracts*. Vol. -1, p. 1428.

Wiejaczka, J (2023). "High-Resolution Grain-Size Distributions: Insight Into Tephra Dispersal and Sedimentation During Plinian Eruptions". PhD thesis. University of Oregon.

Wiejaczka, J and T Giachetti (2024). "Determining the umbrella cloud geometry of unwitnessed silicic explosive eruptions: A case study from Mount Mazama (Oregon, United States)". In: *Journal of Volcanology and Geothermal Research*. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2024.108015.

Wild, A J et al. (2019). "Probabilistic volcanic impact assessment and cost-benefit analysis on network infrastructure for secondary evacuation of farm livestock: A case study from the dairy industry, Taranaki, New Zealand". In: *Journal of Volcanology and Geothermal*

Research 387. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2019.106670.

Wild, A J et al. (2021). "Modelling spatial population exposure and evacuation clearance time for the Auckland Volcanic Field, New Zealand". In: *Journal of Volcanology and Geothermal Research* 416. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2021.107282.

Wilkinson, E (2015). "Beyond the volcanic crisis: co-governance of risk in Montserrat". In: *Journal of Applied Volcanology* 4.1, p. 3. ISSN: 2191-5040. DOI: 10.1186/s13617-014-0021-7.

Williams, G T et al. (2017). "Buildings vs. ballistics: Quantifying the vulnerability of buildings to volcanic ballistic impacts using field studies and pneumatic cannon experiments". In: *Journal of Volcanology and Geothermal Research* 343, pp. 171–180. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2017.06.026.

Williams, S et al. (2025). "Incorporating Eruption Source Parameter and Meteorological Variability in the Generation of Probabilistic Volcanic Ash Hazard Forecasts". In: *Journal of Geophysical Research: Atmospheres* 130. ISSN: 2169-897X. DOI: https://doi.org/10.1029/2024JD042280.

Willis, G (2014). *Managing Natural Hazard Risk in New Zealand - towards More Resilient Communities*. Tech. rep. Local Government New Zealand and Regional Councils. URL: https://www.wcrc.govt.nz/repository/libraries/id:2459ikxj617q9ser65rr/hierarchy/Documents/Publications/Regional%20Plans/Regional%20Policy%20Statement/Review%20Resources/Managing%20natural%20hazard%20risk%20in%20New%20Zealand%2C%20LGNZ%20%282014%29.pdf.

Wilson, C J N (1985). *The Taupo Eruption, New Zealand II. The Taupo Ignimbrite*. London. DOI: 10.1098/rsta.1985.0020.

Wilson, C J N et al. (1995). "Volcanic and structural evolution of Taupo Volcanic Zone, New Zealand: a review". In: *Journal of Volcanology and Geothermal Research* 68.1, pp. 1–28. ISSN: 0377-0273. DOI: https://doi.org/10.1016/0377-0273(95)00006-G.

Wilson, C. J.N. et al. (2006). "The 26.5 ka Oruanui eruption, Taupo Volcano, New Zealand: Development, characteristics and evacuation of a large rhyolitic magma body". In: *Journal of Petrology* 47.1, pp. 35–69. ISSN: 00223530. DOI: 10.1093/petrology/egi066.

Wilson, L (1972). "Explosive volcanic eruptions-II the atmospheric trajectories of pyroclasts". In: *Geophysical Journal International* 30.4, pp. 381–392. ISSN: 1365-246X. DOI: https://doi.org/10.1111/j.1365-246X.1972.tb05822.x.

Wilson, L and G P L Walker (1987). "Explosive volcanic eruptions - VI. Ejecta dispersal in plinian eruptions: the control of eruption conditions and atmospheric properties". In: *Geophysical Journal of the Royal Astronomical Society* 89.2, pp. 657–679. DOI: 10.1111/j.1365-246X.1987.tb05186.x.

Wilson, L et al. (1978). "The control of volcanic column heights by eruption energetics and dynamics". In: *Journal of Geophysical Research: Solid Earth* 83.B4, pp. 1829–1836. ISSN: 0148-0227. DOI: https://doi.org/10.1029/JB083iB04p01829.

Wilson, T, A Dantas, and J Cole (2009). "Modelling livestock evacuation following a volcanic eruption: An example from Taranaki volcano, New Zealand". English. In: *New Zealand Journal of Agricultural Research* 52.1, pp. 99–110. ISSN: 11758775. DOI: 10.1080/00288230909510493.

Wilson, T M, S Jenkins, and C Stewart (2015). "Impacts from volcanic ash fall". In: *Volcanic Hazards, Risks and Disasters*. Ed. by P Papale. Elsevier, pp. 47–86. DOI: https://doi.org/10.1016/B978-0-12-396453-3.00003-4.

Wilson, T M et al. (2012a). "Short- and long-term evacuation of people and livestock during a volcanic crisis: lessons from the 1991 eruption of Volcán Hudson, Chile". In: *Journal of Applied Volcanology* 1.1, p. 2. ISSN: 2191-5040. DOI: 10.1186/2191-5040-1-2.

Wilson, T M et al. (2012b). "Volcanic ash impacts on critical infrastructure". In: *Physics and Chemistry of the Earth, Parts A/B/C* 45-46, pp. 5–23. ISSN: 1474-7065. DOI: https://doi.org/10.1016/j.pce.2011.06.006.

Wilson, T M et al. (2014). "Volcanic hazard impacts to critical infrastructure: A review". In: *Journal of Volcanology and Geothermal Research* 286, pp. 148–182. DOI: 10.1016/j.jvolgeores.2014.08.030.

Woo, G (2008). "Probabilistic criteria for volcano evacuation decisions". In: *Natural Hazards* 45.1, pp. 87–97. ISSN: 1573-0840. DOI: 10.1007/s11069-007-9171-9.

Woodhouse, M J et al. (2013). "Interaction between volcanic plumes and wind during the 2010 Eyjafjallajökull eruption, Iceland". In: *Journal of Geophysical Research: Solid Earth* 118.1, pp. 92–109. ISSN: 2169-9313. DOI: https://doi.org/10.1029/2012JB009592.

Woods, A W (1988). "The fluid dynamics and thermodynamics of eruption columns". In: *Bulletin of Volcanology* 50.3, pp. 169–193. ISSN: 1432-0819. DOI: 10.1007/BF01079681.

Wright, I C (1994). "Nature and tectonic setting of the southern Kermadec submarine arc volcanoes: An overview". In: *Marine Geology* 118.3, pp. 217–236. ISSN: 0025-3227. DOI: https://doi.org/10.1016/0025-3227(94)90085-X.

Yamin, H Y, S M Shahidehpour, and Z Li (2004). "Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets". In: *International Journal of Electrical Power & Energy Systems* 26.8, pp. 571–581. ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2004.04.005.

Yang, Q et al. (2020). "Novel statistical emulator construction for volcanic ash transport model Ash3d with physically motivated measures". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 476.2242. DOI: 10.1098/rspa.2020.0161.

Young, P and G Wadge (1990). "FLOWFRONT: Simulation of a lava flow". In: *Computers & Geosciences* 16.8, pp. 1171–1191. ISSN: 0098-3004. DOI: https://doi.org/10.1016/0098-3004(90)90055-X.

Zago, V et al. (2019). "Preliminary validation of lava benchmark tests on the GPUSPH particle engine". In: *Annals of Geophysics* 62.2. ISSN: 2037-416X. DOI: https://doi.org/10.4401/ag-7870.

Zebker, H A et al. (1996). "Analysis of active lava flows on Kilauea volcano, Hawaii, using SIR-C radar correlation measurements". In: *Geology* 24.6, pp. 495–498. ISSN: 1943-2682. DOI: https://doi.org/10.1130/0091-7613(1996)024<0495:AOALFO>2.3.CO;2.

Zeigler, B.P, A Muzy, and E Kofman (2019). *Theory of Modeling and Simulation*. Third. Elsevier. ISBN: 9780128133705. DOI: 10.1016/C2016-0-03987-6.

Zeinalova, N et al. (2025). "Influence of rheological parameters on lava flow morphology inferred from numerical modelling". In: *International Journal of Earth Sciences*. ISSN: 1437-3262. DOI: 10.1007/s00531-025-02513-2.

Zuccarello, F et al. (2025). "Dynamics and hazards of pyroclastic avalanches at Etna volcano (Italy)". In: *Annals of Geophysics* 68.1. ISSN: 2037-416X. DOI: doi:10.4401/ag-9158.