

# Tephra deposit inversion by coupling Tephra2 with the Metropolis-Hastings algorithm: algorithm introduction and demonstration with synthetic datasets

Qingyuan Yang (✉ [qingyuan.yang@ntu.edu.sg](mailto:qingyuan.yang@ntu.edu.sg))

Nanyang Technological University <https://orcid.org/0000-0002-5631-889X>

**E Bruce Pitman**

University at Buffalo - The State University of New York

**Marcus Bursik**

University at Buffalo - The State University of New York

**Susanna F Jenkins**

Nanyang Technological University

---

## Methodology

**Keywords:** TEPHRA2, Metropolis-Hastings algorithm, Markov Chain Monte Carlo methods, Tephra inversion, Bayesian inversion

**Posted Date:** December 18th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-37455/v2>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on January 26th, 2021. See the published version at <https://doi.org/10.1186/s13617-020-00101-4>.

1           Tephra deposit inversion by coupling *Tephra2* with the  
2           Metropolis-Hastings algorithm: algorithm introduction and  
3           demonstration with synthetic datasets

4           Qingyuan Yang<sup>1,2,\*</sup>, E Bruce Pitman<sup>3</sup>, Marcus Bursik<sup>4</sup> and Susanna F Jenkins<sup>1,2</sup>

5                           <sup>1</sup>Earth Observatory of Singapore, Singapore

6                           <sup>2</sup>Asian School of the Environment, Nanyang Technological University, Singapore

7                           <sup>3</sup>Department of Materials Design and Innovation, University at Buffalo, Buffalo, NY,  
8                           USA

9                           <sup>4</sup>Department of Geology, University at Buffalo, Buffalo, NY, USA

10                           \*qingyuan.yang@ntu.edu.sg

11                           November 11, 2020

12                           **Abstract**

13                           In this work we couple the Metropolis-Hastings algorithm with the volcanic ash transport  
14                           model *Tephra2*, and present the coupled algorithm as a new method to estimate the Eruption  
15                           Source Parameters of volcanic eruptions based on mass per unit area or thickness measurements  
16                           of tephra fall deposits. Outputs of the algorithm are presented as sample posterior distributions  
17                           for variables of interest. Basic elements in the algorithm and how to implement it are introduced.  
18                           Experiments are done with synthetic datasets. These experiments are designed to demonstrate  
19                           that the algorithm works from different perspectives, and to show how inputs affect its perfor-  
20                           mance. Advantages of the algorithm are that it has the ability to i) incorporate prior knowledge;  
21                           ii) quantify the uncertainty; iii) capture correlations between variables of interest in the estimated  
22                           Eruption Source Parameters; and iv) no simplification is assumed in sampling from the posterior  
23                           probability distribution. A limitation is that some of the inputs need to be specified subjectively,

24 which is designed intentionally such that the full capacity of the Bayes' rule can be explored by  
25 users. How and why inputs of the algorithm affect its performance and how to specify them  
26 properly are explained and listed. Correlation between variables of interest in the posterior distri-  
27 butions exists in many of our experiments. They can be well-explained by the physics of tephra  
28 transport. We point out that in tephra deposit inversion, caution is needed in attempting to esti-  
29 mate Eruption Source Parameters and wind direction and speed at each elevation level, because  
30 this could be unnecessary or would increase the number of variables to be estimated, and these  
31 variables could be highly correlated. The algorithm is applied to a mass per unit area dataset  
32 of the tephra deposit from the 2011 Kirishima-Shinmoedake eruption. Simulation results from  
33 *Tephra2* using posterior means from the algorithm are consistent with field observations, sug-  
34 gesting that this approach reliably reconstructs Eruption Source Parameters and wind conditions  
35 from deposits.

## 36 ***Keywords***

37 *Tephra2*, Metropolis-Hastings algorithm, Markov Chain Monte Carlo methods, Tephra inversion, Bayesian  
38 inversion

## 39 **1 Introduction**

40 Quantifying Eruption Source Parameters (ESPs), such as eruption plume height, eruption duration  
41 and variability, and mass eruption rate or total eruption mass, is critical to studies of volcanic eruptions  
42 and their products (Newhall and Self, 1982; Carey and Sparks, 1986; Pieri and Baloga, 1986; Armienti  
43 et al., 1988; Scarpati et al., 1993; Mastin et al., 2009; Stohl et al., 2011; Pouget et al., 2013; Madankan  
44 et al., 2014; Bear-Crozier et al., 2020). Knowing the values of ESPs helps reconstruct pre-historic and  
45 unobserved eruptions, and provides information for the characterization of potential future hazards  
46 (e.g., Suzuki et al., 1983; Carey and Sparks, 1986; Bursik et al., 1992b,a; Bonadonna et al., 1998; Sparks  
47 and Young, 2002; Hildreth, 2004; Bonadonna and Houghton, 2005; Mannen, 2006; Neri et al., 2008;  
48 Jenkins et al., 2008; Bonasia et al., 2010; Jenkins et al., 2012; Bonadonna and Costa, 2012; Bonadonna  
49 et al., 2015; Bevilacqua et al., 2015; Engwell et al., 2015; Yang and Bursik, 2016; Bevilacqua et al., 2018;  
50 Yang et al., 2019; Biass et al., 2019). ESPs are commonly estimated by coupling field observations  
51 with expertise on the process being analyzed. Such expertise could be in the form of a quantitative or  
52 descriptive physical model, an empirical or semi-empirical relationship, or their combination. Suzuki

53 et al. (1983) proposed the first tephra transport and deposition model. Different methods (which  
54 include physical and semi-empirical models) have been proposed to simulate and study the deposition  
55 of volcanic ash (Carey and Sparks, 1986; Bursik et al., 1992b,a; Koyaguchi and Ohno, 2001; Bursik,  
56 2001; Costa et al., 2006; Jones et al., 2007; Folch et al., 2009; Bonadonna et al., 2010; González-Mellado  
57 and De la Cruz-Reyna, 2010; Schwaiger et al., 2012; Suzuki and Koyaguchi, 2013).

58 Estimating ESPs can be treated as an inverse problem (e.g., Tarantola, 2005; Kaipio and Somersalo,  
59 2006), and requires the use of different statistical and engineering techniques (Suzuki et al., 1983;  
60 Carey and Sparks, 1986; Bursik et al., 1992b; Sparks et al., 1992; Mannen, 2006; Klawonn et al.,  
61 2012, 2014; Maeno et al., 2014; Biass et al., 2016; Poret et al., 2017; Koyaguchi et al., 2017; White  
62 et al., 2017; Yang et al., 2019; Mannen et al., 2020). Previous workers have presented different  
63 methods to implement inversion to obtain ESPs from the characteristics of tephra deposits, such as  
64 deposit thickness and grain size. The simplex search algorithm, grid-search method, matrix inversion  
65 with Tikhonov regularization, and a regularized form of the Levenburg-Marquardt algorithm have  
66 been proposed (Connor and Connor, 2006; Klawonn et al., 2012; Johnston et al., 2012; White et al.,  
67 2017; Moiseenko and Malik, 2019; Mannen et al., 2020). The efficiency and ability to characterize  
68 uncertainty with various simplifications (such as those used to avoid solving ill-posed problems) are  
69 the main concerns in proposing these algorithms as alternatives to classical inversion.

70 The challenges in estimating ESPs derive from their (1) high-dimensionality (i.e., too many vari-  
71 ables to be estimated) and (2) limited field observations (e.g., Green et al., 2016). Because of the  
72 ill-posedness of this inversion, it is important to quantify the uncertainty in the process of estimating  
73 ESPs such that we know how certain or uncertain we are about our estimate. In addition, it has been  
74 shown that ESPs influence model prediction through interaction with other ESPs (Scollo et al., 2008).  
75 In tephra inversion, such interactions or coupling could potentially lead to correlated results, i.e.,  
76 estimated ESPs are correlated with one another, which is not always taken into account, or studied  
77 in a systematic and statistically formal manner. Markov Chain Monte Carlo (MCMC) methods have  
78 the ability to quantify inherent uncertainty and address the presence of correlation between variables  
79 of interest in the estimate.

80 In this work, we present and introduce an algorithm that couples the Metropolis-Hastings (M-H)  
81 algorithm (Hastings, 1970), one of the most widely-used MCMC methods, with volcanic ash transport  
82 and deposition model *Tephra2* (Bonadonna et al., 2010; Connor et al., 2011) for the estimation of  
83 ESPs of explosive volcanic eruptions. Advantages in using MCMC methods to estimate ESPs under a

84 Bayesian framework include (1) the estimation can be denoted as a posterior probability distribution,  
85 which enables uncertainty quantification; (2) prior knowledge on ESPs and field observations can be  
86 combined in a statistically formal way to jointly determine the result; (3) correlations among ESPs and  
87 between ESPs and wind conditions can be captured by the algorithm in the estimated result; and (4)  
88 no simplification is assumed in sampling from the posterior probability distribution, which guarantees  
89 that the results are fully Bayesian. Note that these are not unique to the presented algorithm (e.g.,  
90 White et al., 2017). See following sections for comparison between different tephra inversion methods.

91 Because of these advantages, the algorithm presented herein has the potential to better differentiate  
92 and characterize sources of uncertainty, and detect insensitive variables of interest in tephra inversion.  
93 Uncertainty always exists in tephra inversion regardless of the method being used. Given its constant  
94 presence, it is important and better for us to be able to capture and quantify the uncertainty in a  
95 statistically formal manner.

96 We introduce and demonstrate the algorithm in the following way. We introduce the physical  
97 model *Tephra2* (Bonadonna et al., 2010; Connor et al., 2011) first. We then briefly explain Bayes’  
98 rule, which is what the M-H algorithm (Hastings, 1970) solves numerically. An intuitive interpretation  
99 of the M-H algorithm is given. Then we describe in detail the construction and implementation of the  
100 M-H algorithm and specific setups of the presented algorithm. We apply the algorithm to simulated  
101 (synthetic) datasets, i.e., datasets generated from *Tephra2* with known ESPs and wind conditions, to  
102 validate the algorithm. Three sets of experiments are done with different purposes. Note that the  
103 experiments are done to demonstrate that the algorithm is constructed properly for *Tephra2*. There  
104 is no need to demonstrate the validity of Bayes’ rule or the M-H algorithm as they are well-studied  
105 and known to work well in inversion problems (e.g., Hastings, 1970; Berger, 2013)

106 In the discussion section, main advantages and limitations of the algorithm are pointed out. Corre-  
107 lation in the posterior distribution between variables of interest in our experiments is explained by the  
108 physics of tephra transport. Whether a simplified wind profile should be adopted in tephra inversion  
109 is discussed. We apply the algorithm to a dataset consisting of observed mass per unit area data of  
110 the tephra deposit from the well-studied 2011 Kirishima-Shinmoedake eruption to estimate its ESPs.  
111 The results are in general consistent with observations and estimates from previous studies.

112 The algorithm is coded in python scripts, and is published on *vhub* (<https://vhub.org/resources/4614>).  
113 To make the work accessible to a broad audience, we minimize the use of mathematical and statistical  
114 terms in introducing Bayes’ rule and the M-H algorithm. We hope that the algorithm can benefit

115 researchers with interest in estimating ESPs of volcanic eruptions regardless of their backgrounds, and  
116 the text can serve as a tutorial to potential users.

## 117 **2 Volcanic ash transport model *Tephra2***

118 *Tephra2* is a widely-used volcanic ash transport and deposition model (Bonadonna et al., 2010; Connor  
119 et al., 2011). It has been coupled with different statistical and engineering techniques for forward and  
120 inverse modeling of tephra fall deposits and volcanic hazard analysis (Connor and Connor, 2006;  
121 Mannen, 2006; Volentik et al., 2010; Fontijn et al., 2011; Biass et al., 2012; Mannen, 2014; Magill  
122 et al., 2015; Biass et al., 2016, 2017; Takarada, 2017; Wild et al., 2019; Connor et al., 2019; Mannen  
123 et al., 2020; Williams et al., 2020). *Tephra2* assumes that tephra particles with different grain sizes  
124 are released from a vertical column with column radius increasing with height (accounted for by  
125 an additional diffusion term; Suzuki et al., 1983), and their transport is subject to wind advection,  
126 horizontal turbulent diffusion, and falling at terminal velocities. Inputs of *Tephra2* include total  
127 eruption mass and total grain size distribution, and other parameters to characterize the eruptive  
128 column and wind conditions.

129 *Tephra2* gives semi-analytical solution to the advection-diffusion equation, and its output is the  
130 tephra mass per unit area deposited and grain size distribution at user-specified locations. *Tephra2*  
131 assigns the total erupted mass  $M^0$  to grain size bins (in  $\phi$  unit) based on the specified grain size  
132 distribution. The total mass for each grain size is distributed along the eruptive column (discretized  
133 to points). The mass distribution is described by a beta probability density function characterized by  
134  $\alpha$  and  $\beta$ .

135 When tephra particles with grain size  $\phi_j$  released at the height of  $H_i$  (their total mass:  $M_{i,j}$ ) settle  
136 and deposit on the ground, the corresponding spatial distribution of mass per unit area ( $m_{i,j}(x, y)$ )  
137 of the deposit is proportional to a 2D Gaussian function, and can be written as:

$$m_{i,j}(x, y) = M_{i,j} f_{i,j}(x, y), \quad (1)$$

138 where  $(x, y)$  is the spatial coordinates, and  $f_{i,j}(x, y)$  is the 2D Gaussian function with its mean and  
139 variance depending on the grain size  $j$ , released elevation  $H_i$ , wind speed and direction, and parameters  
140 that characterize turbulent diffusion (e.g., turbulent diffusion coefficient). The total mass per unit  
141 area at  $(x, y)$  is the sum of Eq. 1 for all particle sizes released from the eruptive column. As the

142 eruptive column, a line source, is discretized into many point sources in *Tephra2*, the total mass per  
143 unit area can be written as:

$$m(x, y) = \sum_{i=0}^{H_{max}} \sum_{j=\phi_{min}}^{\phi_{max}} M_{i,j} f_{i,j}(x, y). \quad (2)$$

144 To run *Tephra2*, ESPs, wind conditions, and locations of interest need to be specified. *Tephra2*  
145 discretizes the atmosphere into multiple horizontal layers. The number of horizontal layers and their  
146 elevations as well as the corresponding wind speeds and directions need to be specified as wind  
147 conditions by users to run *Tephra2*. See more information on the use and implementation of *Tephra2*  
148 in Connor et al. (2011); Mannen (2014); Connor et al. (2019).

### 149 **3 Methods**

150 Not applicable

## 151 **4 Inversion technique**

### 152 **4.1 Bayes' rule**

153 Simply put, Bayes' rule states that our prior knowledge about certain quantities of interest can be  
154 updated based on new observations. Assuming that the quantities of interest (i.e., ESPs in this study)  
155 is a vector  $\mathbf{x}$ , the prior knowledge on its value can be denoted as a prior probability distribution  $P(\mathbf{x})$ .  
156 Here the prior distributions should be specified in a way such that they truthfully reflect how certain,  
157 or equivalently, how uncertain we are about the values of variables to be estimated. Discussions on  
158 the Bayes' rule and how to properly specify the priors could become either too abstract or technical,  
159 and are too broad to be within the scope of this work. Philosophical, comprehensive, and detailed  
160 discussions on Bayesian statistics can be found in Berger (2013).

161 With a series of observations  $\boldsymbol{\theta}$  (i.e., mass per unit area of tephra deposit on the ground in this  
162 study), our understanding on  $\mathbf{x}$  could be updated, which is denoted by the posterior probability  
163 distribution ( $P(\mathbf{x}|\boldsymbol{\theta})$ ). Bayes' rule is written as:

$$P(\mathbf{x}|\boldsymbol{\theta}) = \frac{P(\boldsymbol{\theta}|\mathbf{x}) \cdot P(\mathbf{x})}{P(\boldsymbol{\theta})}, \quad (3)$$

164 where  $P(\boldsymbol{\theta}|\mathbf{x})$  is the likelihood function. It denotes the probability of observing  $\boldsymbol{\theta}$  given  $\mathbf{x}$ .  $P(\boldsymbol{\theta})$  is

165 the evidence  $P(\boldsymbol{\theta}) = \int P(\boldsymbol{\theta}|\mathbf{x})P(\mathbf{x})d\mathbf{x}$ , and is the total probability of the observations. Here we refer  
166 to probability density as probability for convenience.

167 We can assimilate new observations (e.g. tephra observations) to obtain the posterior distribution  
168 with the help of the likelihood function. In our case, this cannot be done without running the model  
169 *Tephra2*. We use the simplest case with only one variable of interest unknown, say column height, and  
170 one mass per unit area observation ( $\theta_*$ ) to explain the likelihood function. We could apply one value  
171 of column height ( $h_*$ ) to *Tephra2*, and collect the corresponding output  $d(x = h_*)$  with  $d(\cdot)$  denoting  
172 the *Tephra2* output (assume one location of interest in this example). Neglecting model uncertainty,  
173 knowing  $x = h_*$  is equivalent of knowing  $d(x = h_*)$ . Further assuming that the likelihood function  
174 follows a Gaussian distribution, its mean value could be  $d(x = h_*)$ , and its variance needs to be  
175 determined based on our understanding of the data (e.g., the variance should scale with measurement  
176 uncertainty; see Kawabata et al., 2013; Green et al., 2016 for more information on selecting the  
177 likelihood function).

178 If the true column height that generates  $\theta_*$  is 10 km, we expect to see the likelihood function  
179 having a greater value if  $h_*$  is closer to 10 km—the probability of observing  $\theta_*$  is greater when  $h_*$   
180 is closer to 10 km. In general, the likelihood function should have a greater value, if the model  
181 output is similar to the observation. This is also why a Gaussian distribution centered at the model  
182 output can be used as one form of the likelihood function. The scale of the likelihood function, which  
183 is standard deviation in this example, reflects the scale of measurement uncertainty in the present  
184 context. If multiple observations are made, by assuming that each observation is made independently,  
185 the likelihood function could be the product of likelihood function for each observation (as adopted  
186 in this work). Here it should be noted that with multiple observations, each observation could have  
187 different (assumed) measurement uncertainty (e.g., measured thicknesses of tephra deposits from  
188 sediment cores and on land could have different measurement uncertainty; when implementing the  
189 inversion based on the amount of each grain size, measured amount of each grain size could have  
190 various level of measurement uncertainty). This would change the shape of the likelihood function,  
191 and additional attention is needed when constructing the likelihood function in such cases. This is  
192 not considered in the present work, but could be adjusted in the algorithm based on specific needs of  
193 users.

194 Constructing the likelihood function properly requires our knowledge on the observation dataset,  
195 and the form of the likelihood function could vary case by case. See Kawabata et al. (2013); Green

196 et al. (2016); Engwell et al. (2013) for more detailed discussion on how to select the form of the  
197 likelihood function properly and how to quantify measurement uncertainty of tephra fall deposits.

198 To obtain the posterior probability distribution, we need  $P(\mathbf{x})$ ,  $P(\boldsymbol{\theta})$ , and the likelihood function  
199  $P(\boldsymbol{\theta}|\mathbf{x})$ . The prior distribution  $P(\mathbf{x})$ , the likelihood function  $P(\boldsymbol{\theta}|\mathbf{x})$ , and  $P(\boldsymbol{\theta})$  (by definition) need  
200 to be defined beforehand based on prior knowledge about the ESPs and measurement uncertainty.

201 The major difficulty in analytically deriving the posterior distribution comes from the fact that  
202 the likelihood function would become almost certainly non-parametric in practice. This would make  
203 the value of  $P(\boldsymbol{\theta})$  hard to calculate (although it is a constant), and is related to practical issues such  
204 as the high dimensionality of the parameter space. Therefore, the posterior distribution is frequently  
205 obtained through numerical sampling methods.

## 206 4.2 The M-H algorithm

207 MCMC methods, a class of methods that draw samples from a target distribution, can be used to  
208 sample from the posterior distribution based on  $P(\boldsymbol{\theta}|\mathbf{x})P(\mathbf{x})$ , the numerator of the right-hand side  
209 of Eq. 3. In this way, the difficulty in calculating  $P(\boldsymbol{\theta})$  can be avoided. In volcanology, MCMC  
210 methods have been widely adopted for various purposes, such as estimating parameters and initial  
211 conditions of a physical model (which is similar to the goal of the present work), determining ages of  
212 volcanic events, and hazard forecasting (e.g., Green et al., 2016; Anderson et al., 2019; Covey et al.,  
213 2019; Lev et al., 2019; Jenkins et al., 2019; Wang et al., 2020; Liang and Dunham, 2020). Green  
214 et al. (2016) used one MCMC method to estimate volumes of tephra fall deposits based on sparse and  
215 incomplete observations, and their work used a semi-empirical model to characterize tephra thickness  
216 distribution.

217 The procedure adopted in this work, the M-H algorithm, is one of the most widely used MCMC  
218 methods (Hastings, 1970). A brief introduction to the algorithm is given below. More information  
219 about the algorithm can be found in textbooks and published articles (e.g., Chib and Greenberg, 1995;  
220 Andrieu et al., 2003; Kaipio and Somersalo, 2006).

221 The M-H algorithm draws a series of sample points following certain rules. Each sample point  
222 corresponds to one set of ESPs and wind conditions (one value for each variable) that are to be  
223 estimated. These rules are determined by the prior distribution and the likelihood function. With  
224 sufficient draws, it is guaranteed that the distribution of the drawn samples converges or approximates  
225 the target probability distribution, namely the posterior distribution in our case, regardless of the

226 starting sample point. How the algorithm works can be generally described as follow.

227 With a random starting point  $\mathbf{x}_0$  (i.e., the first sample) and corresponding observations  $\boldsymbol{\theta}$ , the  
228 algorithm proposes a new point  $\mathbf{x}_1^*$  using a proposal function that is known and easy to sample. In  
229 this work, we use one of its most common forms, a Gaussian probability density function centered  
230 at the previous point (i.e.,  $\mathbf{x}_0$  for the first draw). The variance of the proposal function needs to be  
231 defined subjectively which will affect the efficiency of the M-H algorithm, and will be illustrated in  
232 later experiments. Proposing a new point  $\mathbf{x}_1^*$  is thus equivalent of drawing one sample from a Gaussian  
233 probability distribution.

234 By calculating and comparing  $P(\boldsymbol{\theta}|\mathbf{x}_0)P(\mathbf{x}_0)$  with  $P(\boldsymbol{\theta}|\mathbf{x}_1^*)P(\mathbf{x}_1^*)$ , the algorithm decides whether  
235 to accept or reject  $\mathbf{x}_1^*$ . Note that we need to know values of the prior ( $P(\mathbf{x}_0)$  and  $P(\mathbf{x}_1^*)$ ) and likelihood  
236 function ( $P(\boldsymbol{\theta}|\mathbf{x}_0)$  and  $P(\boldsymbol{\theta}|\mathbf{x}_1^*)$ ) to do the comparison. The latter requires us to implement *Tephra2*  
237 to obtain values of the likelihood function.

238 If  $\mathbf{x}_1^*$  is rejected (the rejection rule is introduced in the next paragraph),  $\mathbf{x}_1 = \mathbf{x}_0$ . Otherwise,  
239  $\mathbf{x}_1 = \mathbf{x}_1^*$ . The two procedures, namely drawing a new sample point and rejecting or accepting it,  
240 iterate, and after sufficient iterations, a chain of vectors  $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n$  is obtained. By (1) excluding  
241 points with relatively small index (e.g.,  $\mathbf{x}_0, \dots, \mathbf{x}_{499}$ ), and (2) taking points with a fixed interval along  
242 the chain (e.g., only taking  $\mathbf{x}_{500}, \mathbf{x}_{600}, \dots, \mathbf{x}_{9900}, \mathbf{x}_{10000}$ ), the target posterior distribution is obtained  
243 through sampling. The first measure is to make sure that the results are not affected by the value  
244 of the starting point ( $\mathbf{x}_0$ ), and the second is to avoid auto-correlation in the chain (see Chib and  
245 Greenberg, 1995; Andrieu et al., 2003; Kaipio and Somersalo, 2006 for more details).

246 Whether to accept or reject a proposed point follows the rules below:

- 247 • If  $P(\boldsymbol{\theta}|\mathbf{x}_1^*)P(\mathbf{x}_1^*) > P(\boldsymbol{\theta}|\mathbf{x}_0)P(\mathbf{x}_0)$ , then the posterior probability is greater at  $\mathbf{x}_1^*$ , and the  
248 proposed point will be accepted. That is, if the proposal has a higher posterior probability it is  
249 automatically accepted. Following this rule ensures that there will be more samples with greater  
250 posterior probability in the chain.
- 251 • If  $P(\boldsymbol{\theta}|\mathbf{x}_1^*)P(\mathbf{x}_1^*) < P(\boldsymbol{\theta}|\mathbf{x}_0)P(\mathbf{x}_0)$ , then the algorithm accepts  $\mathbf{x}_1^*$  with probability  $P(\boldsymbol{\theta}|\mathbf{x}_1^*)P(\mathbf{x}_1^*)/P(\boldsymbol{\theta}|\mathbf{x}_0)P(\mathbf{x}_0)$ .  
252 This allows the algorithm to occasionally sample points with low or relatively low posterior prob-  
253 ability, in order to explore the entire possible domain of  $\mathbf{x}$ . Following this rule implies that:
  - 254 – If  $P(\boldsymbol{\theta}|\mathbf{x}_1^*)P(\mathbf{x}_1^*)$  is a lot smaller than  $P(\boldsymbol{\theta}|\mathbf{x}_0)P(\mathbf{x}_0)$ , the posterior probability at  $\mathbf{x}_1^*$  is  
255 small, and should be accepted with low probability. This ensures that there will be fewer

256 points with low posterior probability in the chain.

257 – If  $P(\boldsymbol{\theta}|\mathbf{x}_1^*)P(\mathbf{x}_1^*)$  is only slightly smaller than  $P(\boldsymbol{\theta}|\mathbf{x}_0)P(\mathbf{x}_0)$ , the probability of accepting  
258  $\mathbf{x}_1^*$  is relatively greater. In such a case, the algorithm encourages (with high probability of  
259 acceptance) keeping  $\mathbf{x}_1^*$  and further exploring points around (sharing similar values with)  
260  $\mathbf{x}_1^*$ .

261 The second rule is critical to the M-H algorithm. Instead of searching for the values that  
262 maximize the posterior probability (maximum a posteriori estimation), the algorithm functions  
263 through sampling from the target distribution.

## 264 5 Specific setup and running the algorithm

### 265 5.1 Form of the likelihood function

266 In the present version of the algorithm, it is assumed that the likelihood function for each observation  
267 follows a Gaussian distribution with variable  $\log_{10}(\frac{\text{observation}}{\text{model output}})$ . The mean of the distribution  
268 centers at 0 such that the likelihood function peaks when the observation is identical to the model  
269 output. This setup is consistent with previous works. It states that measurement uncertainty scales  
270 with magnitude of the observation (e.g., Connor and Connor, 2006; Kawabata et al., 2013; White  
271 et al., 2017). The standard deviation or scale of the likelihood function, which reflects measurement  
272 uncertainty, needs to be specified by users. Its effect on the results of the algorithm will be examined  
273 in the following experiments. It should be noted that the form of the likelihood function could be  
274 changed to different forms in the algorithm.

### 275 5.2 Two ways to specify the wind profile

276 The algorithm allows for two ways to specify and estimate the wind profile, and users could decide  
277 which way to use, and how to construct the parameterization (i.e., determine what variables are  
278 known, and what are to be estimated for the wind profile). In cases with limited observations, it  
279 is challenging to estimate the wind speed and direction at each elevation (see White et al., 2017  
280 for successful examples). In such cases, a simplified form of the wind profile can be adopted in the  
281 algorithm. It assumes that (1) the wind direction is constant, and does not change with elevation,  
282 and (2) wind speed increases from zero to a certain (a maximum wind speed) value with elevation,

283 and then decreases to zero with elevation linearly. Four variables define such a simplified wind profile.  
284 They are the wind direction, maximum wind speed, elevation that corresponds to the maximum wind  
285 speed, and elevation that the wind speed reaches zero. Users could specify which of the four variables  
286 are to be estimated. Prior distributions of these variables also need to be specified if they are to be  
287 estimated. This wind profile setup is similar to the wind speed profile adopted by Carey and Sparks  
288 (1986).

289 In the second way to specify the wind profile, users could determine whether wind speed and  
290 direction at each elevation level are known (i.e., do not need to be estimated and kept fixed in  
291 implementing the algorithm) or not (i.e., need to be estimated). If the wind speed and direction  
292 at certain or all elevation levels need to be estimated, users need to specify the prior for each of  
293 these variables. For recent eruptions, users could adopt best-fit historically observed wind profiles,  
294 or use them to construct the priors for wind conditions with the non-simplified wind profile. For  
295 example, NOAA NCEP/NCAR REANALYSIS data (Kalnay et al., 1996) provide a large number of  
296 wind profiles globally, based on best-fit wind profiles to observations via weather models (e.g., Connor  
297 et al., 2019; Mannen et al., 2020).

### 298 **5.3 Running the algorithm**

299 To run the algorithm, users first need to prepare input files. Examples of input files can be found in  
300 Table 1 . They have five columns including the “variable name” column. The “initial value” column  
301 specifies the starting value of each variable. If the value of one certain variable is known (not to be  
302 estimated), its initial value will be fixed during the implementation of the algorithm, and the “prior”  
303 column needs to be marked as “Fixed”. For such variables, users could leave the last three columns  
304 blank. To estimate a non-simplified wind profile, users must specify the wind direction and speed at  
305 the specified elevation levels, depending on the form of the assumed profile.

306 For variables to be estimated, their initial values do not affect the results as long as the specified  
307 number of draws is sufficiently large. The only requirements for the initial values are that they need  
308 to be physically possible, and their values would not lead to an error when running *Tephra2*. In  
309 practice, having the initial values located within a meaningful region (for example, means of the prior  
310 distributions) could effectively avoid exploring (sampling) values that are less likely to be accepted.  
311 Forms of prior distributions for these variables need to be defined in the “prior” column. The current  
312 version of the algorithm supports “Gaussian” and “Uniform” distributions. Including more forms

313 (e.g., log-normal distribution and bounded Gaussian distribution) of the prior distribution in the  
314 algorithm will be one goal of our future work for its improvement. If the former is specified, columns  
315 “parameter\_a” and “parameter\_b” should be filled with the mean and standard deviation of the prior  
316 distribution, respectively. Otherwise, the two columns correspond to the minimum and maximum of  
317 the uniform prior distribution. The last column “draw\_scale” specifies the standard deviation of the  
318 proposal function for each variable to be estimated.

319 After the input files are prepared, users need to run all python scripts in order to execute the  
320 algorithm. Other than setting up proper file paths to read input files and observed data, and store  
321 the results, users only need to specify two values, namely the scale of the likelihood function and the  
322 number of draws (length of the chain), to run the algorithm. How to determine the number of draws  
323 is universally challenging when working with MCMC methods. It is preferred to have a large number  
324 of draws so that measures can be done to reduce autocorrelation in the sampled chain (e.g., leave out  
325 samples at the beginning of the chain and taking samples along the chain with a fixed interval as the  
326 final sample distribution; see Chib and Greenberg, 1995; Andrieu et al., 2003; Kaipio and Somersalo,  
327 2006 for more details), and the results are less likely to be divergent. Here we recommend the number  
328 of draws to be ranging from five thousand to one million. This is based on our experiments with the  
329 algorithm. The basic principle is that more number of draws are necessary if more variables are to be  
330 estimated. This is because with more variables to be estimated, the dimensionality of the input space  
331 increases, and more draws are needed to explore the input space (i.e., a lot more possible combinations  
332 of input variable values need to be explored). One million draws are indeed not trivial, but given the  
333 low computational cost, having one million runs with *Tephra2* is not impractical.

334 Users could check for convergence by running two or more separate runs with identical inputs  
335 except for the starting values. If sample distributions from the runs are similar to each other, results  
336 from them converge. Otherwise, users need to increase the number of draws, and run the algorithm  
337 and check again following the same procedure. See Andrieu et al. (2003) and references within for  
338 more information on how to implement the M-H algorithm properly.

339 The primary output from the algorithm is the sampled chain. If 10000 draws are specified, the  
340 sampled chain will be a 10000-by-23 (19 variables for the ESPs plus 4 variables for the wind profile)  
341 matrix if a simplified wind profile is adopted. Otherwise, with the non-simplified wind profile scenario,  
342 the result will be presented as three separate matrices for the eruptive column (10000-by-19 matrix)  
343 and wind speed and direction at each elevation (two 10000-by-40 matrices if the specified elevations

344 are 1000, 2000, ..., 40000 m), respectively. Variables with known and fixed values will remain constant  
345 in the corresponding columns.

346 The algorithm also produces the log-transformed values of the prior probability and likelihood  
347 function of the proposed samples at each draw, log-transformed posterior probability for each accepted  
348 sample on the chain, and the number of acceptances from the run as outputs. They could potentially  
349 help users to debug and adjust parameters to run the algorithm. The ESPs and wind conditions with  
350 the highest posterior probability are mostly likely to reproduce observations. Their values can be  
351 found by examining histograms of resultant samples or finding their medians.

352 The running time of the algorithm depends (roughly linearly based on our experiments) on the  
353 specified number of draws, the running time of *Tephra2*, and the number of observations. On an iMac  
354 with a 3 GHz Intel Core i5 processor, implementing the algorithm with ten thousand draws and 30  
355 observations takes 394 seconds ( $\sim 6.5$  minutes). More details on how to implement the algorithm can  
356 be found in the documentation file of the algorithm on vhub (<https://vhub.org/resources/4614>; Yang  
357 et al., 2020).

## 358 **6 Results**

359 In this section, we generate simulation data using *Tephra2*. The simulated data are treated as field  
360 measurements for testing and validation of the algorithm. It should be noted that the goal of this work  
361 is to present and validate this algorithm and introduce and explain how it works. There is no need  
362 to question Bayes' rule and the adopted M-H algorithm. We avoid using real thickness or mass per  
363 unit area datasets of tephra fall deposits in testing the algorithm. This would avoid additional sources  
364 of uncertainty from affecting the results (e.g., model uncertainty and variable level of post-eruption  
365 erosion and compaction of tephra deposits).

366 It is difficult to validate the algorithm even with synthetic datasets, because (1) outputs of the  
367 algorithm are posterior probability distributions, not individual optimum values. Because of this, it is  
368 difficult to tell that the method works in a single experiment: it is perhaps easier to tell whether the  
369 posterior mean is estimated correctly or not, but how can we check for uncertainty? It is difficult to  
370 calculate the posterior standard deviation analytically; (2) in complex scenarios (i.e., a lot of variables  
371 are to be estimated), it is possible that posterior distributions cannot be updated from the priors.  
372 For example, when the variable to be estimated is not sensitive to model output. In such cases, the

373 not-updated posterior distribution is the right answer, because our prior knowledge on this variable  
374 cannot be updated based on new observations; (3) Whether the prior and posterior distributions are  
375 poorly- or well-constrained is always relative, which depends on numbers of observations and variables  
376 to be estimated and measurement uncertainty.

377 Therefore, we think that comparison is key to demonstrating the validity of the algorithm. By  
378 comparing whether the posterior uncertainty changes accordingly with the inputs and with expect-  
379 tations from Bayes' rule, we examine whether the posterior uncertainty is estimated correctly. For  
380 example, if we increase the assumed measurement error in an experiment, we would expect to have  
381 a greater posterior uncertainty for the ESPs to be estimated in the result. However, because of the  
382 second difficulty in validating the algorithm listed above, the comparison would only be significant  
383 and discernible in simple scenarios (i.e., fewer variables to be estimated).

384 With these concerns, three sets of experiments are done with different purposes. In the first two  
385 sets of experiments, the algorithm with the simplified wind profile is adopted. The experiment in Set  
386 3 works with the non-simplified wind profile.

387 The first set of experiments is designed to test when given different inputs, whether the resultant  
388 posterior distributions from the algorithm would behave consistently with expectations from Bayes'  
389 rule, and to illustrate how specifications of inputs to the algorithm affect the results (sampled posterior  
390 distributions). We thus intentionally keep the scenarios in Set 1 experiments simple, i.e., estimating  
391 column height and total eruption mass with ten or thirty observations. In this way, we know how the  
392 posterior distributions behave given different inputs (e.g., different data quality and priors) based on  
393 Bayes' rule. In these simplified scenarios, we exclude potential impacts from the model (e.g., whether  
394 certain variable is sensitive to the output or not) on the estimated posterior distributions. Therefore,  
395 albeit simplified, Set 1 experiments are considered necessary and strict measures to demonstrate that  
396 the algorithm is constructed properly. In these simplified experiments, we also adjust the number of  
397 input observations (dataset size) and sample site locations to showcase that the algorithm behaves  
398 consistently with the Bayes' rule in different scenarios.

399 In Set 2 experiments, we estimate posterior distributions of eight variables with poorly-constrained  
400 priors (except for column height as it is investigated in Set 1 experiments). These scenarios are more  
401 comparable to a real-world problem of tephra inversion.

402 The experiment in Set 3 is done to show that (a) the algorithm is able to estimate wind speeds and  
403 directions at different elevations when the problem is known to be solvable (with the non-simplified

404 wind profile); and (b) the estimation is not easy due to the number of variables to be estimated (i.e.,  
405 high dimensionality of the input space) for the algorithm. Details and results of the three sets of  
406 experiments are given below.

## 407 **6.1 Set One experiments**

### 408 **6.1.1 Experiment setup**

409 For Set 1, twelve experiments are done. We set column height and total eruption mass as our variables  
410 of interest, and their true values are 15000 m and  $1.88 \times 10^{11}$  kg ( $25.96 = \log(1.88 \times 10^{11})$  log-  
411 transformed kg), respectively. Here we use the natural logarithm, but it can be easily adjusted in the  
412 present version of the algorithm based on the need of its users. The choices would not affect any results  
413 or conclusions in the following experiments. Values of other ESPs and wind conditions used to run  
414 *Tephra2* to generate the “observations” are assumed to be known (i.e., fixed values in the experiments)  
415 and listed in Table 2, and are fixed in all experiments in Set 1. The simplified wind profile is adopted  
416 throughout the 12 experiments. Given purposes of Set 1 experiments, focusing on just two variables  
417 makes it easier for examining and interpreting the results. The first experiment, Experiment # 0, is  
418 used as reference for comparison with results from the rest. For the rest experiments, we just modify  
419 one input of the algorithm or the observation dataset, and keep the others the same as they are in  
420 Experiment # 0. In this way, the impact of each factor on the performance of the algorithm can  
421 be isolated and highlighted. Specified inputs of the 12 experiments as well as changes made in each  
422 experiment are highlighted in Table 3.

423 For the 12 experiments, the M-H algorithm is set to draw 10000 points (10001 points in the chain  
424 including the starting point). After each experiment is finished, we post-process the results by taking  
425 the first 1000 points out of the chain, and collecting samples from the rest of the chain by a 15-points  
426 interval (only taking 1015th, 1030th, ..., 10000th points in the chain) and discarding all other points.  
427 These measures are adopted to prevent the results from being affected by the initial starting point  
428 and autocorrelation as mentioned earlier. For two variables of interest, a chain with 10000 draws is  
429 sufficient enough for the samples to converge to the posterior distribution. Summary of the sampled  
430 posterior distributions is given in Table 4.

### 431 **6.1.2 Reference experiment**

432 In Experiment # 0, ten observations are sampled at localities far from the source vent downwind  
433 ( $\sim 14 - 30$  km from the vent ; sample sites shown in Fig. 1a). Their values range from 50-383  
434 kg/m<sup>2</sup>. Gaussian distributions are assumed for the column height and log-transformed total eruption  
435 mass. Means and standard deviations are 16000 m and 2000 m for column height, and 25.96 and 2  
436 log-transformed kg ( $1.88 \times 10^{11}$  kg) for eruption mass. It is noted that 16000 m is slightly greater  
437 than the specified column height used to generate the dataset, and the prior mean for the log-scaled  
438 eruption mass is identical to the true value (Table 3). The scale of the likelihood function is set to be  
439 0.05, which corresponds to  $\sim 11.6\%$  of relative measurement error (this can be calculated based on  
440 how the likelihood function of the algorithm is defined; see text above).

441 The results are shown and summarized in Table 4 and Fig. 2. The posterior means for the column  
442 height and log-scaled eruption mass are 15338 m and 25.928 log-transformed kg, respectively, and the  
443 corresponding posterior standard deviations (of the samples) are 1067 m and 0.066 log-transformed  
444 kg. Both posterior means are consistent with the true values, and the posterior standard deviations  
445 are smaller than those of the priors (2000 m and 2 log-transformed kg, respectively). The results  
446 suggest that the algorithm works in this simplified case. With ten observations and the presented  
447 algorithm, the consistency between the posterior means and the true values of column height and  
448 total eruption mass and the reduced posterior standard deviations suggest that our knowledge about  
449 the column height and total eruption mass are correctly improved.

### 450 **6.1.3 Effects from specifications required by the M-H algorithm**

451 **Scale of the likelihood function** In Experiment # 1, we increase the scale of the likelihood  
452 function from 0.05 to 0.2 (Table 3), which corresponds to 53.4% of relative measurement error. A  
453 greater scale of the likelihood function implies that a greater measurement uncertainty is assumed  
454 (neglecting model uncertainty). From Bayes' rule, we know that compared with results from the  
455 reference experiment, posterior distributions of Experiment # 1 would not be greatly updated.

456 As expected, the resulting sampled posterior distribution of column height from Experiment #  
457 1 is not significantly different from its prior. Its posterior mean and standard deviation are 15990  
458 m and 1775 m, respectively (prior mean and standard deviation: 16000 m and 2000 m). In this  
459 experiment, observations are made, but because a larger measurement uncertainty is assumed for  
460 them, the algorithm does not trust these measurements as it does in the reference experiment. This is

461 also manifested in the acceptance rate of Experiment # 1, which is 84.0%. What the M-H algorithm  
462 does here is basically accepting or rejecting points based on the prior distribution, and the likelihood  
463 function cannot help determine whether to accept or reject the proposed samples due to the greater  
464 scale specified (i.e., greater measurement uncertainty).

465 The posterior distribution of the log-transformed total eruption mass still centers close to the true  
466 value with lowered standard deviation compared to its prior, suggesting that estimating total eruption  
467 mass is less sensitive to measurement uncertainty. This is consistent with the argument from Scollo  
468 et al. (2008) which states that total eruption mass is a crucial eruption parameter that would greatly  
469 affect *Tephra2* outputs by itself.

470 **Scale of the proposal function** In Experiment # 2, scales of the proposal function are increased  
471 from 500 m and 0.05 log-transformed kg to 2000 m and 0.2 log-transformed kg for column height  
472 and log-transformed total eruption mass, respectively (Table 4). With greater scales of the proposal  
473 functions, the algorithm is more likely to propose a new point that is far from the current point. We  
474 know from the theory of the M-H algorithm that this would not affect the resultant sampled posterior  
475 distributions. This is confirmed through comparing results in Experiments # 0 and 2 which are similar  
476 to each other (Table 4).

477 Nonetheless, scales of the proposal functions would affect the efficiency of the algorithm and  
478 sometimes its performance when the number of draws is not large enough. The algorithm is being too  
479 “adventurous” with proposal functions characterized by greater scales: they tend to explore (propose)  
480 values that are greatly different from the current point. Such values are less likely to be accepted  
481 especially when the current values are characterized by high posterior probability (acceptance rate:  
482 13.6% for Experiment # 2). The greater probability of rejection reduces the efficiency of the algorithm.  
483 Users could adopt the suggested measures (introduced in previous section) to check for convergence  
484 first. If the results converge, there is no need to worry about the greater probability of rejection.  
485 Otherwise, users could lower scales of the proposal functions for the variables of interest to increase  
486 the probability of acceptance, run the algorithm, and check for convergence.

487 **Prior distributions** Experiments # 3 and 4 are designed to test how prior distributions affect the  
488 posterior distributions. The prior means of the column height are set to be 14000 m and 12000 m  
489 with standard deviation of 500 m for both experiments respectively. The prior means for the log-scale  
490 eruption mass are 23.66 log-transformed kg, and the standard deviations are 0.5 log-transformed kg in

491 the two experiments. The prior means are different from the true values of column height (15000 m)  
492 and total eruption mass (25.96 log-transformed kg). In the two experiments, we are incorrect (incorrect  
493 prior means) yet confident (small standard deviations for the priors) in our prior knowledge.

494 Based on Bayes' rule, we know that with the overconfidence in specifying the priors in Experiments  
495 # 3 and 4, the ten observations are probably not sufficient enough to "drag" the posterior distributions  
496 to be centered at the true values. We expect to see the posterior distributions of one or both of the  
497 variables to be centered in between the true values and means of the specified priors.

498 The results are consistent with this expectation, and show that the ten observations are not power-  
499 ful enough to correct both priors in the two experiments. Posterior means of the two experiments are  
500 14283 and 12666 m for the column height (std: 442 and 459 m), and 25.997 and 26.138 log-transformed  
501 kg for the log-scaled total eruption mass (std: 0.048 and 0.059 log-transformed kg), respectively. Ex-  
502 periment # 3 with the specified prior closer to the true value has its posterior distribution of column  
503 height closer to the true value compared to Experiment # 4. The "incorrect" results in the two  
504 experiments represent what we expect to see based on Bayes' rule, and support the validity of the  
505 algorithm.

506 Posterior distributions of total eruption mass in both experiments are centered near the true value  
507 of log-transformed total eruption mass. This again is consistent with the interpretation about the  
508 total eruption mass from Scollo et al. (2008).

509 **Number of input observations** Experiments # 5- 9 share the same specifications with Experi-  
510 ments # 0-4, respectively, except that 20 more samples are included in the input. The 30 measurements  
511 (see Fig. 1 for sample localities) range from 32-383 kg/m<sup>2</sup>. The results are compared pairwise, and  
512 summarized in this section. With more observations, we expect to see the posterior distributions being  
513 improved (either with reduced uncertainty or the posterior means closer to the true values) compared  
514 to Experiments # 0-4 with fewer observations.

515 Comparison between Experiments # 0 and 5 shows that they have similar posterior means that  
516 are consistent with true values of column height and log-transformed total eruption mass. The cor-  
517 responding standard deviations are smaller in Experiment # 5 (685 m and 0.039 log-transformed kg  
518 for column height and log-scaled eruption mass). This confirms that more observations reduce the  
519 uncertainty in the posterior distribution. The same argument can be made for Experiment # 7 as  
520 it has the same specifications as Experiment # 5 except for greater scales of the proposal functions

521 (which would not affect the posterior distribution, and is discussed in Experiment #2).

522 Even with more observations, the posterior distribution (mean: 15786 m; std: 1633 log-transformed  
523 kg) of column height in Experiment # 6 (which assumes a greater measurement error) is not greatly  
524 updated from the prior. This is again due to the greater likelihood scale specified in Experiment #  
525 6. The assumed uncertainty in the measurement is too large that 30 observations are not sufficient  
526 enough to greatly update the prior.

527 Experiments # 8 and 9 with incorrect and confident priors have their column height posterior  
528 means (14483 and 13340 m) slightly closer to the true value compared to results from Experiments #  
529 3 and 4 (14283 and 12666 m). Posterior means of log-transformed eruption mass in the two experiments  
530 are close to the true values (25.986 and 26.070). It can be seen that more observations “drag” the  
531 posterior distributions of column height towards the true value in the two experiments. Results from  
532 experiments # 3, 4, 8, and 9 reflect the “wrestling” between incorrect prior knowledge and informative  
533 observations. The posterior distributions in Experiments # 8 and 9 are also characterized by lower  
534 standard deviations (Table 4). These all conform with our expectations based on Bayes’ rule and  
535 previous study (Scollo et al., 2008) on the sensitivity of total eruption mass in *Tephra2*.

536 **Sample site locations** In Experiments # 10 and 11, ten measurements at proximal (and closely-  
537 spaced) and medial localities are used as input observations, respectively (see sites in Fig. 1a).  
538 Here the proximity is defined relatively, but this would not affect any results or conclusions in these  
539 experiments, as our argument that the algorithm is constructed correctly is demonstrated through  
540 comparison. We only consider cases with sample sites located in the downwind area with respect to  
541 the source vent. This is because the main goal of this experiment is to present and introduce the  
542 algorithm, not to explore under what conditions it would be hard for the algorithm to effectively  
543 update the prior.

544 In Experiment # 10, the sample sites (yellow points in Fig. 1a) are all clustered and located close  
545 to the vent. In theory, they cannot provide too much useful information for the Bayes’ rule to greatly  
546 update the posterior distributions, because the sample sites are too close to each other, and hence  
547 the observations are similar to each other. This is consistent with the results: the sampled posterior  
548 distributions from Experiment # 10 are not greatly updated from the priors. The posterior mean  
549 and standard deviation of the column height (16085 m and 1863 log-transformed kg, respectively)  
550 are similar to those of the prior (Table 4). The posterior mean of the log-transformed eruption mass

551 is 26.106 log-transformed kg, and is characterized by a relatively greater standard deviation (0.280  
552 log-transformed kg) compared to results from other experiments.

553 In Experiment # 11, the sample sites are spatially medial to the source vent, i.e., neither especially  
554 close nor far relative to the entire span of distances (pink points Fig. 1a). The sites are distributed in  
555 a more scattered pattern than are those in Experiment # 10. These observations provide more useful  
556 information than do the proximal observations (Experiment # 10), which can be used to update the  
557 priors. This is consistent with results of Experiment # 11. As shown in Table 4, the posterior means of  
558 the column height and log-scaled eruption mass are consistent with the true values, and the posterior  
559 distributions are also characterized by smaller standard deviations compared to those in Experiment  
560 # 10 and the specified priors.

561 Resultant posterior distributions from Set 1 experiments are consistent with our expectations based  
562 on Bayes' rule, suggesting that the presented algorithm is constructed properly for *Tephra2* in these  
563 simplified scenarios. We hope that these experiments could also help potential users understand how  
564 and why inputs of the algorithm affect its performance.

## 565 **6.2 Set Two experiments**

566 In Set Two, two main experiments plus two supplementary experiments are implemented. The two  
567 main experiments are done to show that the algorithm is able to estimate a set of ESPs and wind-  
568 related variables. In the two main experiments, we find that posterior distributions of diffusion  
569 coefficient and column height are similar to their priors. Two supplementary experiments, which  
570 will be introduced later, are thus proposed to show that the variable diffusion coefficient can be  
571 well characterized in simpler scenarios. For column height, the posterior distribution is not updated  
572 because a relatively strong prior is specified.

573 In the two main experiments, we estimate eight variables of interest with the simplified wind  
574 profile: column height, total eruption mass,  $\alpha$  (which characterizes tephra mass distribution along  
575 the column;  $\beta$  is fixed to be 2 in all Set 2 experiments; see Table 2), median and standard deviation  
576 of total grain size distribution, diffusion coefficient, wind direction, and maximum wind speed. We  
577 do not attempt to estimate more ESPs because other ESPs are in many cases well-constrained (e.g.,  
578 vent coordinates and elevation, maximum and minimum grain size considered). The only difference  
579 between the two main experiments is that different levels of relative measurement error are assumed,  
580 which are 28.9% (Experiment # 1; likelihood scale: 0.12) and 14.0% (Experiment # 2; likelihood

581 scale: 0.06), respectively. These values are set arbitrarily, but since the goal to have two experiments  
582 (two measurement errors) is to see whether they would affect the resultant posterior distributions,  
583 these arbitrary decisions would not affect the motif herein.

584 For all four experiments (two main experiments plus the two supplementary ones) in Set Two, the  
585 ESPs and wind conditions used to generate the “observations” are the same as in Set One experiments,  
586 and are listed in Table 2. All sample sites shown in Fig. 1a are the assumed sample sites, and *Tephra2*  
587 outputs at these sites are assumed to be the “observations”. We assume that we have little knowledge  
588 on the ESPs and wind conditions that are to be estimated in the two main experiments except for  
589 column height, which has a prior Gaussian distribution with mean and standard deviation being 15000  
590 and 2000 m, respectively. This prior is relatively well-constrained in Set 2 main experiments given the  
591 number of variables to be estimated and the amount of observations. All priors are listed in Table 5.

592 In each experiment, five hundred thousand samples are drawn. The first 10000 samples are aban-  
593 doned, and the sample interval is set to be 50 to avoid auto-correlation (for each experiment, three  
594 runs are done to check for convergence). The same implementation is done for the supplementary  
595 experiments.

596 Means and standard deviations of resultant posterior distributions in the two main experiments  
597 of Set Two are listed in Table 5. Most posterior means are at or close to their true values with  
598 greatly reduced uncertainty compared to the corresponding priors. The exceptions are the diffusion  
599 coefficient and column height. For the former, the corresponding posterior means are 6647 and 6248  
600  $\text{m}^2/\text{s}$  (true value:  $5000 \text{ m}^2/\text{s}$ ) with standard deviations of 1877 and  $1524 \text{ m}^2/\text{s}$ , respectively, in the  
601 two experiments. Its sampled posterior distributions also resemble a uniform distribution, which is  
602 how the prior is specified.

603 The posterior distribution of column height is not greatly updated from the prior in the first main  
604 experiment with a greater assumed measurement error (posterior mean and std: 14828 and 2025  
605 m). For the second experiment with a smaller measurement uncertainty, the corresponding posterior  
606 standard deviation becomes slightly smaller (posterior mean and std: 14550 and 1690 m).

607 As a smaller measurement error is assumed in the second main experiment of Set 2, based on Bayes’  
608 rule, we expect to see better-constrained posterior distributions (i.e., with lower posterior standard  
609 deviations) in that experiment. This is confirmed in our results. As shown in Table 5, the posterior  
610 standard deviations in the second experiment of Set 2 are all smaller compared to those from the first  
611 experiment. This is consistent with Bayes’ rule.

612 We also list posterior correlations between variable pairs from the two main experiments in Table 6.  
613 It can be seen that several variable pairs are characterized by high-magnitude correlation, suggesting  
614 that the interaction between ESPs and wind conditions is critical, and should not be ignored in tephra  
615 inversion.

### 616 **6.2.1 Supplementary experiments**

617 From the two main experiments, we find that variables diffusion coefficient and column height are  
618 relatively hard to constrain. Two additional, supplementary experiments are thus done to prove that  
619 the variable diffusion coefficient can be well-estimated in simpler scenarios. We do not experiment  
620 with column height here as experiments in Set 1 have proved that it can be well-estimated in simpler  
621 scenarios.

622 We decide to estimate  $\alpha$  (which characterizes tephra mass distribution along the column), diffusion  
623 coefficient, wind direction, and maximum wind speed of the simplified wind profile in the supplemen-  
624 tary experiments. We assume that all other ESPs and wind conditions are known. The priors of these  
625 variables are the same as in the two main experiments, and are shown in Table 5. The only difference  
626 between the two is the assumed relative measurement errors, which are 13.5% and 6.7%, respectively.  
627 Again, it is through comparison that we validate the algorithm. Exact values of the two would not  
628 affect the motif herein. Similarly, the other three variables are chosen arbitrarily, but would not affect  
629 conclusions or results in these experiments given the goal of the supplementary experiments.

630 Posterior means and standard deviations of the variables to be estimated are shown in Table 5. It  
631 can be seen that variables other than the diffusion coefficient are well-constrained in both experiments.  
632 In terms of diffusion coefficient, the supplementary experiment with a greater measurement uncertainty  
633 has its posterior distribution centered at 5403 with a standard deviation of 1142. In the experiment  
634 with a smaller assumed measurement uncertainty, the resultant posterior distribution finally becomes  
635 close to the true value (posterior mean: 5078), and the posterior standard deviation is also smaller  
636 (543) compared to the other supplementary experiment.

637 Results in the two supplementary experiments suggest that the variable diffusion coefficient can be  
638 well-estimated in simpler scenarios. Comparing posterior distributions from the two suggests that its  
639 posterior distribution is hard to constrain given limited observations compared with the other ESPs.

640 Together with the fact that the M-H algorithm is a generalized probabilistic inversion algorithm,  
641 results from Sets 1 and 2 experiments suggest the validity of the presented algorithm.

### 6.3 Set Three experiment with non-simplified wind profile

Results from one experiment with non-simplified wind profile are presented in this section. The ESPs and wind conditions to generate the synthetic data are shown in Table 2. Change in wind speed and direction with elevation is taken into account in this experiment. The wind speed is specified to increase from 0 on the ground to a maximum of 70 m/s at 16 km, and then decrease to 10 m/s at 24 km. The wind direction is to the north on the ground, and gradually changes to eastwards with elevation.

We set our ESPs of interest to be column height and log-transformed total eruption mass. For the wind profile, we choose to estimate wind directions and speeds at elevation levels 5, 14, and 18 km a.s.l. Wind speeds and directions at other elevations are assumed to be known. This amounts to eight variables: the two ESPs and wind directions and speeds at the three elevation levels.

We do not choose to estimate wind direction and speed at all elevation levels (i.e., estimate the complete wind profile) because that would significantly increase the number of variables to be estimated. In such circumstances, the problem could become hard to solve as the increased dimensionality of the input space makes it hard for the algorithm to draw samples efficiently. At the same time, we know that wind speed and direction only affect wind advection in *Tephra2*. This means that different combinations of wind speeds and directions at various elevations could lead to similar simulated results in *Tephra2*, and that they could remain relatively independent from other ESPs in many cases. This is comparable to vector decomposition: one vector (the total distance each tephra particle travels due to advection in *Tephra2*) can be decomposed to the sum of infinite combinations of two or more vectors (distances each tephra particle travels within two or more elevation layers).

Therefore, an experiment that estimates the wind direction and speed at each elevation cannot be used to justify the method works even if sufficient draws are made. With poorly-constrained priors, we know that the posterior distributions cannot be greatly updated (because we know that a lot of local minima exist, and we are given limited observations given the number of variables to be estimated). On the other hand, we cannot tell whether the priors are specified properly: given the high dimensionality, it would be hard for us to know whether the priors are specified too close to the real wind profile or not. Here, the current experiment is designed to show that the method is able to estimate wind speeds and directions at several elevations when the problem is known to be solvable (i.e., relatively fewer variables are to be estimated). More discussion on the use of simplified and non-simplified wind profiles in tephra inversion is given in the following section.

673 Elevation levels 5, 14, and 18 km are chosen because wind directions at these elevations are 25°,  
674 65° and 85°, respectively. We “collect” tephra mass per unit area at 495 randomly selected sample  
675 sites (Fig. 1b). The dataset size is greater than commonly-seen thickness or mass per unit area  
676 datasets of tephra fall deposits. Estimating wind speed and direction at a few elevations and having  
677 495 observations are not realistic in studies on tephra fall deposits. However, these are adopted such  
678 that we know that the problem is solvable, and can be used to validate the algorithm. The number  
679 of “observations” is chosen arbitrarily, but the value of this number (whether it is 500 or not) would  
680 not affect any results or conclusions from this experiment, because what we need is sufficient amount  
681 of observations. Any number that is above  $\sim 350$  would work.

682 Other ESPs, and wind directions and speeds at other elevations, are kept fixed throughout the  
683 implementation of the algorithm. Fifty thousand runs are done for three times with different starting  
684 points. In each run, the first 2000 runs are discarded to avoid auto-correlation, and further subsetting  
685 with an interval of 50 points along the sample chain are subsetted as the final results. The resultant  
686 sample posterior distributions from the three runs is similar, suggesting that the results converge.  
687 The posterior mean and standard deviation for each variable of interest are listed in Table 7. They  
688 are highly consistent with specified values used to generate the dataset, and the posterior standard  
689 deviations are smaller than those of the priors. The results suggest that the algorithm functions  
690 when a non-simplified wind profile is adopted and the number of variables of interest limited. Greater  
691 uncertainty is obtained for wind direction and speed at higher elevations (i.e., 14 and 18 km). This  
692 is due to the fact that the wind speed at higher elevations is generally greater, and the estimated  
693 uncertainty scales with it.

## 694 7 Discussion

695 In this work, we introduce an algorithm coupling the ash dispersal model *Tephra2* with the Metropolis-  
696 Hastings implementation of MCMC. We validated it with synthetic data generated by *Tephra2*. By  
697 varying the inputs to the algorithm and observation datasets one at a time, we examine and explain  
698 how they affect the performance and efficiency of the algorithm. Three sets of experiments are done.  
699 The first set focuses on simple scenarios (i.e., with simplified wind profile) with two variables of interest  
700 (i.e., total eruption mass and column height) given ten or thirty observations. In these experiments,  
701 the algorithm is shown to work well, and has the ability to quantify the uncertainty in the estimate.

702 Resultant posterior distributions from these experiments are consistent with expectations from Bayes’  
703 rule.

704 In the second set of experiments, we focus on estimating posterior distributions of six ESPs and  
705 two wind-related variables. The results suggest that posterior distributions of most of the variables of  
706 interest are greatly updated from their corresponding priors. While posterior distributions of column  
707 height and diffusion coefficient are similar to their priors, respectively. This is because a relatively  
708 strong prior is specified for column height, and it is harder to constrain the posterior distribution of  
709 diffusion coefficient. The supplementary experiments in Set 2 suggest that diffusion coefficient can be  
710 well estimated in simpler scenarios.

711 In Set 3, we set out to estimate two ESPs, and wind directions and speeds at three elevation  
712 levels given sufficient observations. In this experiment, wind direction and speed are set to vary with  
713 elevation. The results suggest that the algorithm could work well with a non-simplified wind profile.

714 Our discussion here focuses on advantages and limitations of the algorithm, interpreting the pos-  
715 terior correlation between column height and total eruption mass in our experiments, and whether  
716 we should attempt to estimate wind direction and speed at each elevation in tephra inversion when  
717 working with *Tephra2*. The algorithm is then applied to the mass per unit area dataset of the 2011  
718 Kirishima-Shinmoedake tephra deposit to infer the corresponding ESPs.

## 719 **7.1 Advantages and limitations**

720 The main advantages of the algorithm are that it makes use of prior knowledge on a deposit and  
721 eruption, and quantifies the uncertainty in the estimate of ESPs in a statistically formal manner.

722 In studies on tephra fall deposits, previous knowledge plays a critical role in determining the ESPs  
723 and reconstruction of volcanic eruptions (Sparks et al., 1997; Mastin et al., 2009). Such knowledge  
724 has uncertainty within it. How to properly incorporate such uncertainty in the estimated results is  
725 challenging without a probabilistic Bayesian framework. With the algorithm, prior knowledge about  
726 the studied deposit and eruption, and their associated uncertainties, is denoted as the prior probability  
727 distribution, and incorporated in the estimate. Practically speaking, prior knowledge is being used  
728 consistently throughout the implementation of the algorithm. That is, the prior probability helps  
729 determine whether to accept or reject a proposed point in each draw of the algorithm. However, for  
730 a non-probabilistic inversion method, such as gradient methods, prior knowledge might be used only  
731 once in the inversion process—it helps determine the starting point.

732 In tephra deposit inversion, uncertainty in the estimated ESPs comes from the interplay of multiple  
733 sources, which include the uncertainty in the prior, measurement uncertainty, and potential model  
734 uncertainty. Non-probabilistic inversion method helps us find the optimum ESPs that fit well to field  
735 observations, but the uncertainty cannot be quantified.

736 The algorithm samples from the posterior distribution without any presumptions. This means that  
737 (1) its results are fully Bayesian; and (2) more flexibility is given to the users, as they could change  
738 forms of priors and likelihood functions based on their own needs (either by choosing the available  
739 options in the present version of the code or by modifying the code). Therefore, the algorithm can  
740 be used to explore how different factors and components in tephra inversion, such as sample site  
741 distribution and form of the likelihood function, affect the results. White et al. (2017) proposed an  
742 efficient inversion method which quantifies the posterior uncertainty with linear analysis. The linearity  
743 assumption allows the method to operate with efficiency, but the prior and likelihood function in their  
744 method have to be Gaussian such that the method functions properly, and so are the resultant posterior  
745 distributions. This might be less convenient when certain variables are known to have bounds (e.g.,  
746 standard deviation of grain size distribution being greater than zero).

747 Efficiency and the ability to quantify the uncertainty are main motivations driving the development  
748 of different tephra inversion techniques (Connor and Connor, 2006; Klawonn et al., 2012; White et al.,  
749 2017; Mannen et al., 2020), but there is always a tradeoff between these motivations. Therefore, we  
750 think that different tephra inversion methods are equally important, and users should decide which  
751 method to use based on their specific needs. The presented algorithm represents one end of the  
752 spectrum: it only focuses on sampling from the posterior distribution in a statistically formal way  
753 without considering efficiency (i.e., no simplifications are made in the algorithm). This is not possible  
754 without the low computational cost of *Tephra2*.

755 Whether the priors, number of draws, and standard deviations of the proposal functions are  
756 specified properly or not affects the performance of the algorithm. This problem arises as long as  
757 the M-H algorithm is adopted, and there is no definitively correct way to determine some inputs to  
758 the algorithm. Measures to check whether it is used properly are given. The introduction on the  
759 algorithm and experiments in Set 1 are presented such that how each element in the algorithm affects  
760 its performance is given with demonstration.

## 7.2 Correlation in the posterior distribution

Correlations are detected in the posterior distributions in our synthetic experiments (Tables 4 and 6). Here we focus on posterior correlations in Set 1 experiments, as their setups are simpler. Even in such extremely simplified scenarios, as shown below, properly interpreting the correlation is not straightforward. Correlations in these experiments are shown in Table 4, and the bivariate posterior distributions from selected experiments are shown in Fig. 3.

We find that both negative and positive correlations exist between column height and total eruption mass in the sampled posterior distributions, and whether the correlation is positive or not depends on sample site localities. In Experiments # 10 and 11, the correlation is positive (0.986 and 0.831), and for the rest, the correlation (-0.474- -0.856) is negative.

This can be explained by the physics of tephra transport. If observations are made at distal localities (i.e., Experiments # 0-9), the combination of (a) a greater column height, which allows tephra to be dispersed farther downwind and leads to more tephra deposition at distal sites, and (b) a smaller total eruption mass (eruption mass is proportional to tephra thickness/mass per unit area everywhere) leads to results similar to the combination with a lower column height and a greater total eruption mass (within the area where footprints of tephra deposits overlap). Therefore, the correlation is negative in Experiments # 0-9.

A lower column height leads to a thicker tephra deposit at proximal sites because of less interaction with wind, and less time for turbulent diffusion to disperse tephra to distal localities. Total eruption mass is always proportional to tephra thickness and mass per unit area. The above arguments suggest that scenarios with (1) greater column height and greater total eruption mass and (2) lower column height and lower total eruption mass lead to similar tephra thickness or mass per unit area if the observations are all made at proximal sites.

These relationships are consistent with previous studies (e.g., Suzuki et al., 1983; Bonadonna et al., 2005), but how their interaction with sample site locations would affect the correlation between the variables of interest in tephra inversion has not been reported or noted. The correlation in Experiment # 10 is surprisingly high (0.986), which is due to the facts that the sample sites are too close to each other, and the “observed” tephra mass per unit area values at these sites have similar values. To the algorithm, the ten observations in Experiment # 10 are almost the same (both their locations and their observed values). The high correlation reveals that non-unique solutions exist in Experiment # 10.

792 The presence of correlation in the posterior distribution suggests that the interaction of variables  
793 plays a role in tephra inversion. This is consistent with results from sensitivity analysis on *Tephra2* in  
794 the work of Scollo et al. (2008). This finding suggests that the algorithm has the potential to be used  
795 to discover intrinsic relationships (interactions) between variables of interest and wind conditions in  
796 *Tephra2* and other dispersion models, and could thus improve our understanding on different sources  
797 of uncertainty in tephra inversion.

### 798 **7.3 Whether to estimate wind direction and speed at each elevation**

799 The algorithm can be used with either simplified or non-simplified wind profiles. It is always preferred  
800 to have a more exact and detailed understanding on wind conditions in tephra inversion. Estimating a  
801 lot of variables at the same time could be challenging for the algorithm as the number of draws has to  
802 be finite. It is common to estimate six or more ESPs (e.g., column height, total eruption mass, column  
803 mass distribution, mean and standard deviation of grain size distribution, and diffusion coefficient) in  
804 tephra inversion. If wind direction and speed are to be estimated at ten elevations, this adds up to at  
805 least  $26 = 6 + 10 \times 2$  variables of interest. Considering just two values for each variable, this means  
806  $2^{26}$  (which is greater than 60 million) possible combinations for the 26 variables. The number of field  
807 measurements for tephra fall deposits rarely exceeds 300.

808 This problem could be resolved by the method proposed by White et al. (2017), which is able  
809 to estimate a full complement of uncertain ESPs and wind conditions, and provide estimates of  
810 posterior variances. This is owing to the use of regularization techniques in their method that can  
811 accommodate much higher dimensionality. A solution can be found, but how the solution connects  
812 to the wind profile in reality (how well the solution reproduces reality) is another interesting and  
813 different topic. The wind profile only affects advection of tephra dispersal in *Tephra2*. Similar to  
814 vector decomposition, the total effect of advection for tephra particles with a certain grain size could  
815 always be decomposed into different combinations of advection effects at each elevation level. Multiple  
816 or a series of optimum solutions of wind profile are likely to exist. Therefore, we think that it is not  
817 always necessary to estimate the wind speed and direction at each elevation level especially given  
818 extremely sparse observations. As no presumptions are adopted in the present algorithm, allowing  
819 two ways to specify the wind profile in the algorithm could help address questions associated with  
820 wind conditions in tephra inversion, such as under what conditions and how details of the wind profile  
821 affect the inversion results.

## 7.4 Application to the Kirishima-Shinmoedake dataset

In this section, we apply the M-H algorithm to a dataset containing mass of tephra per unit area for the 2011 Kirishima eruption. The goals are to show that *Tephra2* outputs from using posterior means from the algorithm as ESPs and wind conditions would resemble field observations, and that the posterior distributions can be characterized by lower uncertainty compared to their corresponding priors. We refrain from going into details in the physics of tephra transport and how the interaction between variables would affect the estimates. This is because tephra dispersal processes that are not taken into account by *Tephra2* did affect tephra dispersal during the eruption (Mannen et al., 2020). For the same reason, comparing observed and predicted ESPs is not done in the present work (because we know that even if we apply the observed ESPs and wind conditions to *Tephra2*, the prediction would still be different from observations given the presence of model uncertainty). Discussions on model uncertainty are outside the scope of this work. See Mannen et al., 2020 for a detailed, careful, and strict treatment of tephra inversion for this deposit.

The Kirishima-Shinmoedake event took place from 26 to 29 January, 2011, with an eruption of the Shinmoedake volcano. The column height ranged from 6.2-8.6 km above the crater, based on different models and Doppler radar measurements (Shimbori and Fukui, 2012; Maeno et al., 2014). The total eruption mass was estimated to be  $1.8 - 3.1 \times 10^{10}$  kg by Nakada et al. (2013). The mass of tephra erupted from the afternoon of 26 January to the early morning of 27 January, which corresponds to the current dataset, is about  $1.4 - 2.5 \times 10^{10}$  kg (Maeno et al., 2013). The wind was blowing to the southeast, and the wind profile is reported in Hashimoto et al. (2012).

The tephra deposit data we are using are reported in White et al. (2017). Detailed description of the dataset can be found in White et al. (2017); Mannen et al. (2020). Tephra thickness and grain size distributions were measured at 55 locations downwind from the vent. In addition, tephra thickness was measured at another 63 locations. The thickness measurements were converted to mass per unit area (Fig. 4a).

We set the ESPs to be estimated as column height, eruption mass,  $\alpha$  (in this experiment,  $\beta$  is fixed to be one), median and standard deviation of grain size distribution, diffusion coefficient, fall time threshold, and densities of lithic and pumice fragments. The eddy constant is fixed as 0.04. For the wind condition, a simplified wind profile is adopted, to avoid overcomplication of the problem, as discussed before. We assume that the wind speed increases linearly from 0 to 11 km a.s.l., and then decreases with elevation to 24 km a.s.l. This setup is based on the wind speed profile reported

853 in (Hashimoto et al., 2012). Wind direction and maximum wind speed are the two variables to be  
854 estimated for the wind profile. This amounts to 11 variables of interest and 118 mass per unit area  
855 observations for the problem. The priors of these variables are inferred based on Shimbori and Fukui  
856 (2012); Nakada et al. (2013); Miyabuchi et al. (2013); Maeno et al. (2014); White et al. (2017), and  
857 are shown in Table 8. These priors are generally consistent with the priors defined in White et al.  
858 (2017) except that eddy constant is set to be a fixed value in our experiment.

859 We draw fifty thousand samples; this process is repeated three times. Sample distributions from  
860 the three runs are almost identical to each other, suggesting that the results converge. The first 5000  
861 samples are discarded to exclude the impact from the initial starting point. For the rest of the chain,  
862 we collect samples based on a 15-point interval to avoid autocorrelation. Large relative measurement  
863 uncertainty (i.e., 45%) is adopted here. This value is larger than the one (30%) adopted in White  
864 et al. (2017); Engwell et al. (2013). This is because from our experiments, we find that for some  
865 observations with very low magnitude, with 30% of measurement error, the likelihood always goes to  
866 zero. This means that the relative measurement error of these observations has to be greater than 30%  
867 (and assuming the absence of model uncertainty). We could either assign a greater measurement error  
868 to these observations, while keeping the others having 30% of relative measurement error, or assign a  
869 greater measurement error for all observations. Both can be done by the algorithm (but the former  
870 requires slight modification of the current code), but here we prefer the latter. This is consistent with  
871 the main goal of this experiment: to show that the algorithm is able to reproduce observations of a  
872 real tephra fall deposit, and this could avoid justifying how and why we adjust the measurement error  
873 for certain observations.

874 The results are summarized in Table 8. The resultant isopach maps, differences between observa-  
875 tions and simulations, and relative errors are presented in Fig. 4. Posterior means of column height  
876 and total eruption mass are in general consistent with previous studies, and are updated from the  
877 priors with posterior standard deviations being much smaller. The simulated mass per unit area data  
878 from *Tephra2* with posterior means as ESPs and wind conditions are plotted against field observa-  
879 tions in Fig. 5, which suggests that *Tephra2* could generally reproduce field observations based on  
880 estimated results from the algorithm. Posterior means of column height and total eruption mass are  
881 7.3 km and  $9.14 \times 10^9$  kg, respectively. The former is within the range of the observed column height,  
882 and the latter is smaller than estimates from previous work.

883 Posterior distributions of the other ESPs generally lie within commonly-seen ranges (Table 8),

884 and are also altered from the corresponding priors. We note that the posterior mean of the median  
885 of grain size distribution is finer than data reported by Miyabuchi et al. (2013). We think that this  
886 could be explained by the fact that data reported by Miyabuchi et al. (2013) represent the grain size  
887 distribution at certain sample sites, whereas our estimate focuses on the total grain size distribution.

## 888 8 Conclusions

889 In this work, we couple the well-known M-H algorithm with the tephra transport and deposition  
890 model *Tephra2*. The coupled algorithm can be used to infer ESPs of explosive volcanic eruptions  
891 and ambient wind conditions based on thickness or mass per unit area measurements of tephra fall  
892 deposits under a Bayesian framework. It allows users to include their prior knowledge on the eruption  
893 or deposit with field observations in a statistically formal way. The result of the algorithm is presented  
894 as sample posterior distributions for the variables of interest.

895 We introduce the model *Tephra2* and basic elements of Bayes' rule, and present intuitive inter-  
896 pretations on the M-H algorithm. How to implement the coupled algorithm and formats of the input  
897 files are also introduced.

898 The coupled algorithm is validated with three sets of experiments. For the first set, we focus on  
899 two variables of interest. In these experiments, we vary values of the input and size of the synthetic  
900 observation dataset one at a time to show that the algorithm functions consistently with expectations  
901 based on the Bayes' rule, and also to show how inputs affect the performance of the algorithm. In  
902 the second set of experiments, we estimate eight variables of interest with poorly-constrained priors  
903 (except for column height). The results show that the algorithm is able to effectively estimate the  
904 posterior distributions of most of the variables of interest, but the posterior distributions of column  
905 height and diffusion coefficient are similar to their priors. For the former, that is because its prior  
906 is specified to be well-constrained. Supplementary experiments are done to show that the variable  
907 diffusion coefficient can be well-estimated in simpler scenarios. The combination of these experiments  
908 suggests the validity of the algorithm, and indicates that posterior distributions of some ESPs are  
909 harder to constrain compared to others.

910 In the experiment in Set 3, we set eight variables of interest to be estimated, including not only  
911 two ESPs, but also wind directions and speeds at three elevation levels. This experiment is set up to  
912 show that the algorithm works with a complex wind profile.

913 Advantages of the algorithm are that it has the ability to incorporate prior knowledge into the  
914 estimate in a statistically formal way, and to quantify the uncertainty in the estimate, and it captures  
915 correlation between variables of interest in the estimate. Because of these advantages, we think that  
916 the algorithm has the potential to improve our understanding on how different sources of uncertainty  
917 interact and affect the results in tephra inversion. The main limitation of the algorithm is that there  
918 are subjective choices in implementation, which affect its performance. How and why they affect  
919 the algorithm are introduced, and commonly adopted measures and references (Chib and Greenberg,  
920 1995; Andrieu et al., 2003; Kaipio and Somersalo, 2006) on how to specify the inputs properly are  
921 given.

922 Correlations between variables of interest exist in our experiments. Interpretation based on the  
923 physics of tephra transport is given for the correlations in Set 1 experiments (which are extremely  
924 simplified): whether the correlation between column height and total eruption mass is positive or not  
925 depends on sample site locations. A greater column height has a positive and negative relationship  
926 with tephra mass per unit area at distal and proximal sample sites, respectively, and total eruption  
927 mass is always proportional to tephra mass per unit area regardless of sample site location.

928 The algorithm supports specifying and estimating the wind profile in two ways. The first one  
929 takes advantage of a simplified wind profile based on four variables of interest, and assumes that  
930 the wind direction does not change with elevation. The second one allows users to estimate wind  
931 speed and direction at certain or all specified elevation . We argue that users need to be cautious  
932 in choosing how to specify and estimate the wind profile, because the second way could introduce  
933 a lot more variables to be estimated, and might be unnecessary. How to choose the appropriate  
934 way to specify and estimate the wind profile relies on factors such as prior knowledge of weather  
935 conditions and sample site distributions. We think that by experimenting on appropriate synthetic  
936 data, this question can be addressed. We apply the algorithm to the 2011 Kirishima-Shinmoedake  
937 tephra dataset, and the results are in general consistent with observations from previous work. We  
938 hope that the present work benefits future studies that attempt to implement tephra inversion and  
939 quantify the associated uncertainty.

## 940 **List of abbreviations**

941 ESPs: Eruption Source Parameters

942 MCMC methods: Markov Chain Monte Carlo methods

943 M-H algorithm: Metropolis-Hastings algorithm

## 944 **Declarations**

### 945 **Availability of data and materials**

946 The data used in this work are mostly generated from synthesized experiments. How to generate these  
947 data is specified in the text. The dataset of the tephra deposit from the 2011 Kirishima-Shinmoedake  
948 eruption used in this work is from previous publications. Where to find the dataset and corresponding  
949 references are given in the text.

### 950 **Competing interests**

951 There is no competing interest involved in this work.

### 952 **Funding**

953 This work was partially supported by National Science Foundation Hazard SEES grant number  
954 1521855 to G. Valentine, M. Bursik, E.B. Pitman, and A.K. Patra. This work comprises Earth  
955 Observatory of Singapore contribution no. 306. This research is partly supported by the National  
956 Research Foundation Singapore and the Singapore Ministry of Education under the Research Centres  
957 of Excellence initiative (Project Name: Evaluating Unrest and Potential Hazards at Changbaishan  
958 Volcano, China; Project Number: NRF2018NRF-NSFC003ES-010) to S.F. Jenkins and Q. Yang.

### 959 **Authors' contributions**

960 The idea of this manuscript, i.e., coupling the Metropolis-Hastings algorithm with *Tephra2* was pro-  
961 posed by E.B. Pitman. Q. Yang coded the algorithm in python scripts. Q. Yang wrote the manuscript  
962 with inputs from E.B. Pitman, M. Bursik, and S.F. Jenkins.

### 963 **Acknowledgements**

964 We thank Dr. Charles Connor, Dr. Jeremy White, and Dr. Kazutaka Mannen for their construc-  
965 tive and insightful comments on this manuscript. We thank Dr. Charles Connor for handling our

967 **Tables**

Table 1: Examples of input files of the algorithm. Tables on the left and right correspond to input files that specify the ESPs and simplified wind profile, respectively.

Inputs of the algorithm for specifying the ESPs						Inputs of the algorithm for specifying the simplified wind profile					
variable_name	initial_val	prior_type	prior_para_a	prior_para_b	draw_scale	variable_name	initial_val	prior_type	prior_para_a	prior_para_b	draw_scale
column_height	15000	Gaussian	15000	2000	500	wind_direction	45	Gaussian	50	10	0.5
log_m	25.96	Uniform	24.46	27.46	0.15	max_wind_speed	10	Gaussian	12	5	1
alpha	5	Uniform	1	7		elevation_max_speed	10000	Fixed			
beta	2	Fixed				zero_elevation	20000	Fixed			
gs_min	-6	Fixed									
gs_max	6	Fixed									
gs_med	3	Uniform	-1.5	4.5	0.15						
gs_sd	1	Uniform	0.5	3.5	0.15						
vent_x	0	Fixed									
vent_y	0	Fixed									
vent_z	1000	Fixed									
edy_const	0.04	Fixed									
diffu_coef	5000	Uniform	1000	10000	400						
fall_time_thre	1419	Fixed									
lithic_rou	2500	Fixed									
pumice_rou	1100	Fixed									
column_steps	100	Fixed									
particle_bins	12	Fixed									
plume_model	2	Fixed									

Table 2: ESPs and wind conditions used to generate “field observations” for validation of the algorithm. ESPs for Sets 1 and 2 experiments with the simplified wind profile (orange-striped cells) and the Set 3 experiment with the non-simplified wind profile (blue-striped cells), respectively.

Variable name	ESPs		Wind profile specification (Set 3 experiment)					
	Specified value (Sets 1 and 2 experiments)	Specific value (Set 3 experiment)	Elevation (m)	Wind speed (m/s)	Wind direction (°)	Elevation (m)	Wind speed (m/s)	Wind direction(°)
Column Height (m)	15000	20000	1000	0	5	19000	30	90
Total eruption mass (kg)	$1.88 \cdot 10^{11}$	$5 \cdot 10^{11}$	2000	0	5	20000	30	90
$\alpha$	5	1.5	3000	10	15	21000	10	90
$\beta$	2	1	4000	10	15	22000	10	90
Max grain size ( $\phi$ )	-6	-6	5000	20	25	23000	10	90
Min grain size ( $\phi$ )	6	6	6000	20	25	24000	10	90
Median grain size ( $\phi$ )	1	0	7000	30	35	25000	20	90
Standard deviation of	1.5	2	8000	30	35	26000	30	90
Vent coordinates (m)	(0, 0)	(0, 0)	9000	40	45	27000	40	90
Vent elevation a.s.l (m)	1000	1000	10000	40	45			
Eddy constant	0.04	0.04	11000	50	55			
Diffusion coefficient (m <sup>2</sup> /s)	5000	3000	12000	50	55	Wind-related variables (simplified wind profile; Sets 1 and 2 experiments)		
Fall time threshold	1419	1419	13000	60	65			
Lithic density (kg/m <sup>3</sup> )	2500	2500	14000	60	65	Downwind direction (°)		45
Pumice density (kg/m <sup>3</sup> )	1100	1000	15000	70	75	Maximum wind speed (m/s)		10
Column steps	100	100	16000	70	75	Wind speed at 0 km a.s.l. (m/s)		0
Particle stpes	12	12	17000	50	85	Wind speed at 20 a.s.l. (m/s)		0
Plume model	2	2	18000	50	85	Elevation with maxi wind speed (m)		10000

Table 3: Specifications and input observation data used to run Experiments # 0-11 of Set 1. Differences in the specification or input observation data in each experiment compared to Experiment # 0 (as reference experiment; marked as green cells) are highlighted in yellow.

Run #	# of input observations	Column height prior (type, mean, std)	Eruption mass prior (type, mean, std; log-scale)	Scale of proposal function for column height	Scale of proposal function for log(eruption mass)	Likelihood scale	Note
0	10	Gaussian(16000, 2000)	Gaussian( $\log(1.88 \cdot 10^{11})$ , 2)	500	0.05	0.05	Reference run (used for comparison with other simulations).
1	10	Gaussian(16000, 2000)	Gaussian( $\log(1.88 \cdot 10^{11})$ , 2)	500	0.05	0.2	
2	10	Gaussian(16000, 2000)	Gaussian( $\log(1.88 \cdot 10^{11})$ , 2)	2000	0.2	0.05	
3	10	Gaussian(14000, 500)	Gaussian( $\log(1.88 \cdot 10^{10})$ , 0.5)	500	0.05	0.05	
4	10	Gaussian(12000, 500)	Gaussian( $\log(1.88 \cdot 10^{10})$ , 0.5)	500	0.05	0.05	
5	30	Gaussian(16000, 2000)	Gaussian( $\log(1.88 \cdot 10^{11})$ , 2)	500	0.05	0.05	
6	30	Gaussian(16000, 2000)	Gaussian( $\log(1.88 \cdot 10^{11})$ , 2)	500	0.05	0.2	
7	30	Gaussian(16000, 2000)	Gaussian( $\log(1.88 \cdot 10^{11})$ , 2)	2000	0.2	0.05	
8	30	Gaussian(14000, 500)	Gaussian( $\log(1.88 \cdot 10^{10})$ , 0.5)	500	0.05	0.05	
9	30	Gaussian(12000, 500)	Gaussian( $\log(1.88 \cdot 10^{10})$ , 0.5)	500	0.05	0.05	
10	10	Gaussian(16000, 2000)	Gaussian( $\log(1.88 \cdot 10^{11})$ , 2)	500	0.05	0.05	Input observations are at ten sites close to the source vent.
11	10	Gaussian(16000, 2000)	Gaussian( $\log(1.88 \cdot 10^{11})$ , 2)	500	0.05	0.05	Input observations are at ten sites medial to the source vent.

Table 4: Summary of results from Set 1 Experiments # 0- 11.

Run #	# of input observations	Posterior mean (column height; m)	Posterior std (column height)	Posterior mean and std (eruption mass; log-scale kg)	Posterior std (eruption mass; log-scale)	Correlation	Acceptance rate (%)
True value		15000	-	$\log(1.88 \cdot 10^{11})=25.960$	-	-	-
0	10	15338	1067	25.952	0.066	-0.808	53.4
1	10	15990	1775	25.928	0.174	-0.474	84.0
2	10	15497	1106	25.945	0.066	-0.810	13.6
3	10	14283	442	25.997	0.048	-0.650	44.1
4	10	12666	459	26.138	0.059	-0.785	42.0
5	30	15228	685	25.951	0.039	-0.841	36.1
6	30	15786	1633	25.948	0.103	-0.528	77.0
7	30	15285	707	25.950	0.041	-0.832	58.9
8	30	14483	359	25.986	0.032	-0.731	31.1
9	30	13340	388	26.070	0.039	-0.856	28.8
10	10	16085	1863	26.106	0.280	0.986	42.9
11	10	15284	843	25.982	0.068	0.831	51.5

Table 5: True values of the ESPs and wind conditions to be estimated, their assumed priors, and means and standard deviations of the resultant posterior distributions in Set 2 experiments.

Experiment #	Variables of interest	Prior type	True value	Prior mean/ minimum	Prior std/ minimum	Experiment 1		Experiment 2	
						Posterior mean	Posterior std	Posterior mean	Posterior std
Experiments 1 and 2	Column height (m)	Gaussian	15000	15000	2000	14828	2025	14550	1690
	Log(Total eruption mass in kg)	Uniform	25.96	24.46	27.46	26.04	0.13	25.99	0.07
	Alpha	Uniform	5	1	7	4.95	1.23	4.79	1.09
	Median of total grain size distribution ( $\Phi$ )	Uniform	1	-1.5	4.5	1.25	0.47	1.15	0.27
	Std of total grain size distribution ( $\Phi$ )	Uniform	1.5	0.5	3.5	1.58	0.33	1.49	0.21
	Diffusion coefficient ( $m^2/s$ )	Uniform	5000	1000	10000	6647	1877	6248	1523
	Wind direction ( $^\circ$ )	Gaussian	45	50	10	44.76	1.85	44.83	0.89
	Wind speed at 10 km (m/s)	Gaussian	10	12	5	10.18	1.23	10.34	0.76
Supplementary experiment 1	Alpha	Uniform	5	1	7	4.92	0.66		
	Diffusion coefficient ( $m^2/s$ )	Uniform	5000	1000	10000	5403	1142		
	Wind direction ( $^\circ$ )	Gaussian	45	50	10	45.00	0.83		
Supplementary experiment 2	Wind speed at 10 km (m/s)	Gaussian	10	12	5	10	0.42		
	Alpha	Uniform	5	1	7	5.01	0.34		
	Diffusion coefficient ( $m^2/s$ )	Uniform	5000	1000	10000	5078	543		
	Wind direction ( $^\circ$ )	Gaussian	45	50	10	45.00	0.41		
	Wind speed at 10 km (m/s)	Gaussian	10	12	5	10	0.21		

Table 6: Posterior correlation table for results from the two main experiments in Set 2. Blue and pink cells correspond to results from the first (with greater assumed measurement error) and second (with lower assumed measurement error) experiments, respectively. Correlations with magnitude above 0.5 are marked.

	Column height (m)	Log(Total eruption mass in kg)	Alpha	Median of TSGD( $\Phi$ )	Std of TSGD( $\Phi$ )	Diffusion coefficient (kg/m <sup>2</sup> )	Wind direction (°)	Wind speed at 10 km (m/s)
Column height (m)	1	-0.28	-0.03	-0.78	0.01	-0.30	0.06	-0.61
Log(Total eruption mass in kg)	-0.13	1	0.38	0.51	0.69	0.41	-0.15	-0.16
Alpha	-0.04	0.34	1	-0.23	0.69	-0.32	0.05	-0.33
Median of total grain size distribution ( $\Phi$ )	-0.57	0.72	-0.04	1	-0.04	0.40	-0.19	0.38
Std of total grain size distribution ( $\Phi$ )	0.11	0.66	0.55	0.25	1	0.17	0.04	-0.11
Diffusion coefficient (kg/m <sup>2</sup> )	-0.03	0.35	-0.10	0.14	0.37	1	-0.08	0.38
Wind direction (°)	0.05	-0.18	-0.02	-0.17	-0.05	-0.04	1	0.08
Wind speed at 10 km (m/s)	-0.37	-0.37	-0.21	-0.11	0.04	0.14	0.19	1

Table 7: True values, specified prior types and parameters, and posterior means and standard deviations for the Set 3 experiment with the non-simplified wind profile.

Variable name	True value	Prior type	Mean or minimum of the prior	Std or maximum of the prior	Posterior mean	Posterior std
Column height (m)	20000	Gaussian	19000	3000	20149	319
Log(mass in kg)	26.94	Gaussian	26.71	0.5	26.94	0.03
Wind direction at 5 km (°)	25	Uniform	0	50	24.82	1.85
Wind direction at 14 km (°)	65	Uniform	30	80	66.00	3.87
Wind direction at 18 km (°)	85	Uniform	60	130	79.26	4.42
Wind speed at 5 km (m/s)	20	Uniform	0	50	20.28	1.82
Wind speed at 14 km(m/s)	60	Gaussian	55	10	56.97	6.11
Wind speed at 18 km(m/s)	50	Uniform	0	100	48.3	8.34

Table 8: Priors and posterior means and standard deviations from applying the algorithm to the 2011 Kirishima-Shinmoedake eruption tephra mass per unit area dataset. Priors of column height, total eruption mass, and median and standard deviation of grain size distribution are referenced and inferred from (Shimbori and Fukui, 2012; Nakada et al., 2013; Miyabuchi et al., 2013; Maeno et al., 2014; White et al., 2017). Priors of  $\alpha/\beta$  ratio, diffusion coefficient, fall time threshold, and pumice and lithic densities are specified as commonly adopted ranges or maximum ranges possible.

Variable name	Prior type	Prior mean/ prior min	Prior std/ prior max	Posterior mean	Posterior std
Column height (above vent; m)	Gaussian	8000.00	1500.00	7372.00	538.00
Log(mass in kg)	Gaussian	24.73 (54.96*10 <sup>9</sup> kg)	2.30	22.93 (9.14*10 <sup>9</sup> kg)	0.05
Alpha/Beta	Uniform	0.01	4.00	0.401	0.076
Median grain size ( $\phi$ )	Gaussian	-0.25	1.00	2.820	0.210
Std of grain size ( $\phi$ )	Uniform	1.00	4.00	2.040	0.240
Diffusion coefficient (m <sup>2</sup> /s)	Uniform	1000.00	5000.00	2256.00	650.00
Fall time threshold (s)	Uniform	0.00	6000.00	411.00	71.06
Lithic density (kg/m <sup>3</sup> )	Gaussian	2500.00	300.00	2485.89	225.34
Pumice density (kg/m <sup>3</sup> )	Gaussian	1000.00	250.00	985.47	279.78
Maximum wind speed (m/s)	Gaussian	80.00	15.00	65.62	5.76
Wind direction (°)	Gaussian	135.00	20.00	124.60	0.33

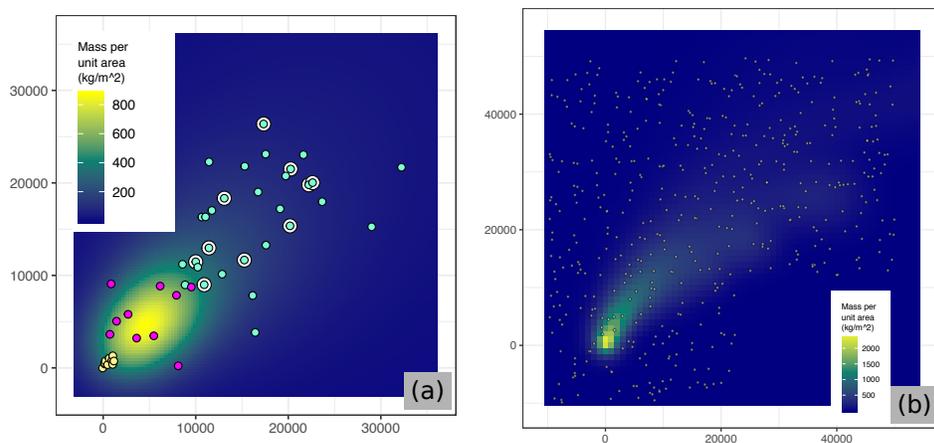


Figure 1: a: mass per unit area distribution used for the validation of Set 1 Experiments # 0-11 and sample site locations. White (larger), turquoise, yellow, and pink dots are the sample site locations used for Experiments # 0-4, 5-9, 10, and 11, respectively. All sites shown in a are used in Set 2 experiments as sample site locations; b: mass per unit area distribution used for the experiment with the non-simplified wind profile. Small white points correspond to sample site locations. Mass per unit area distributions in a and b are in different resolutions. This difference is only for easier visualization (reducing the number of grid points to be plotted in b), and would not affect any arguments or conclusions from this work.

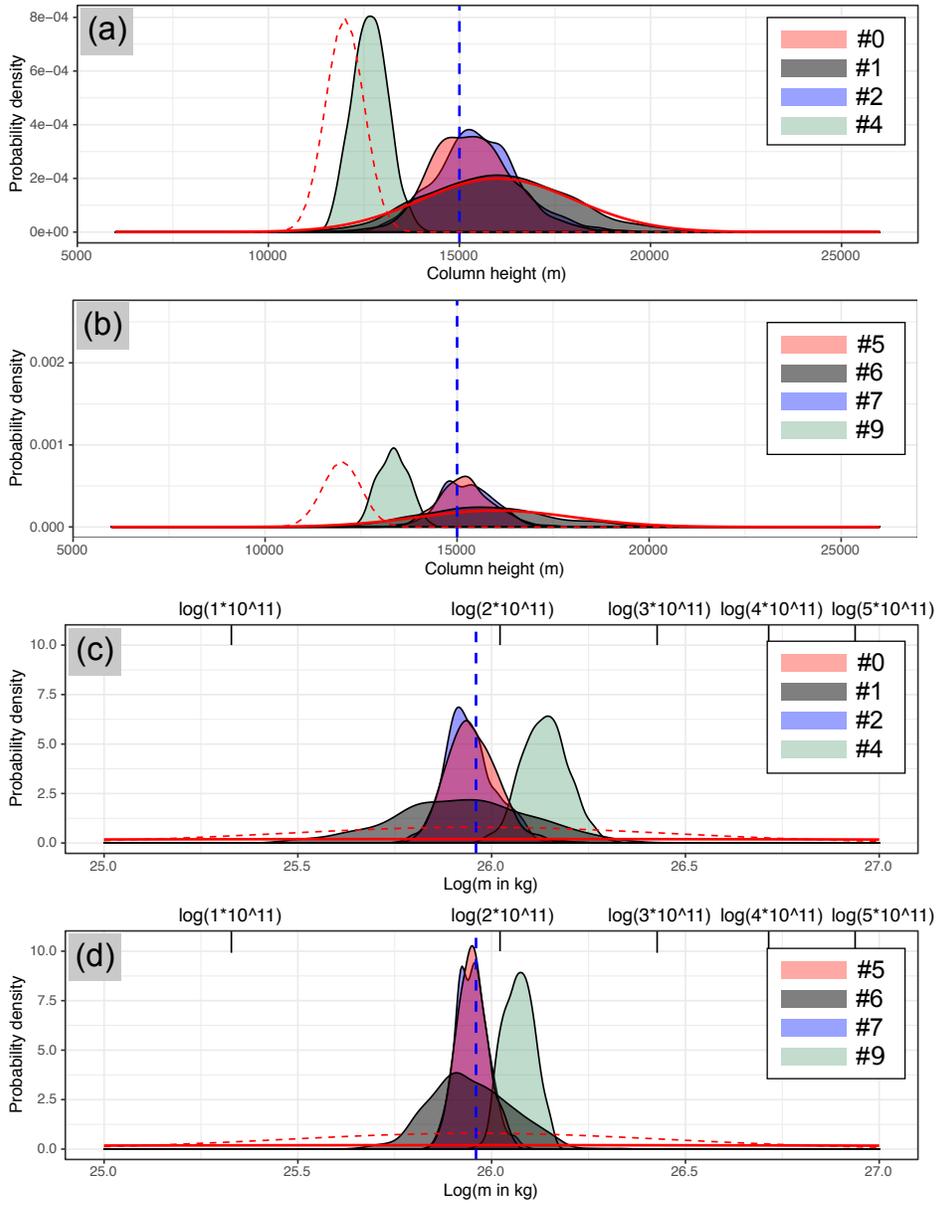


Figure 2: Selected posterior distributions of column height and log-scaled eruption mass. Results from Set 1 Experiments # 0, 1, 2, and 4 are shown in a (column height) and c (log-scaled eruption mass), and results from Set 1 Experiments # 5, 6, 7, and 9 are shown in b (column height) and d (log-scaled eruption mass). The blue dashed lines mark the true values of column height and log-transformed eruption mass used to generate the observation data. The red solid lines correspond to prior distributions assumed for all experiments except for Set 1 Experiments # 4 and 9, and their priors are denoted as red dashed lines.

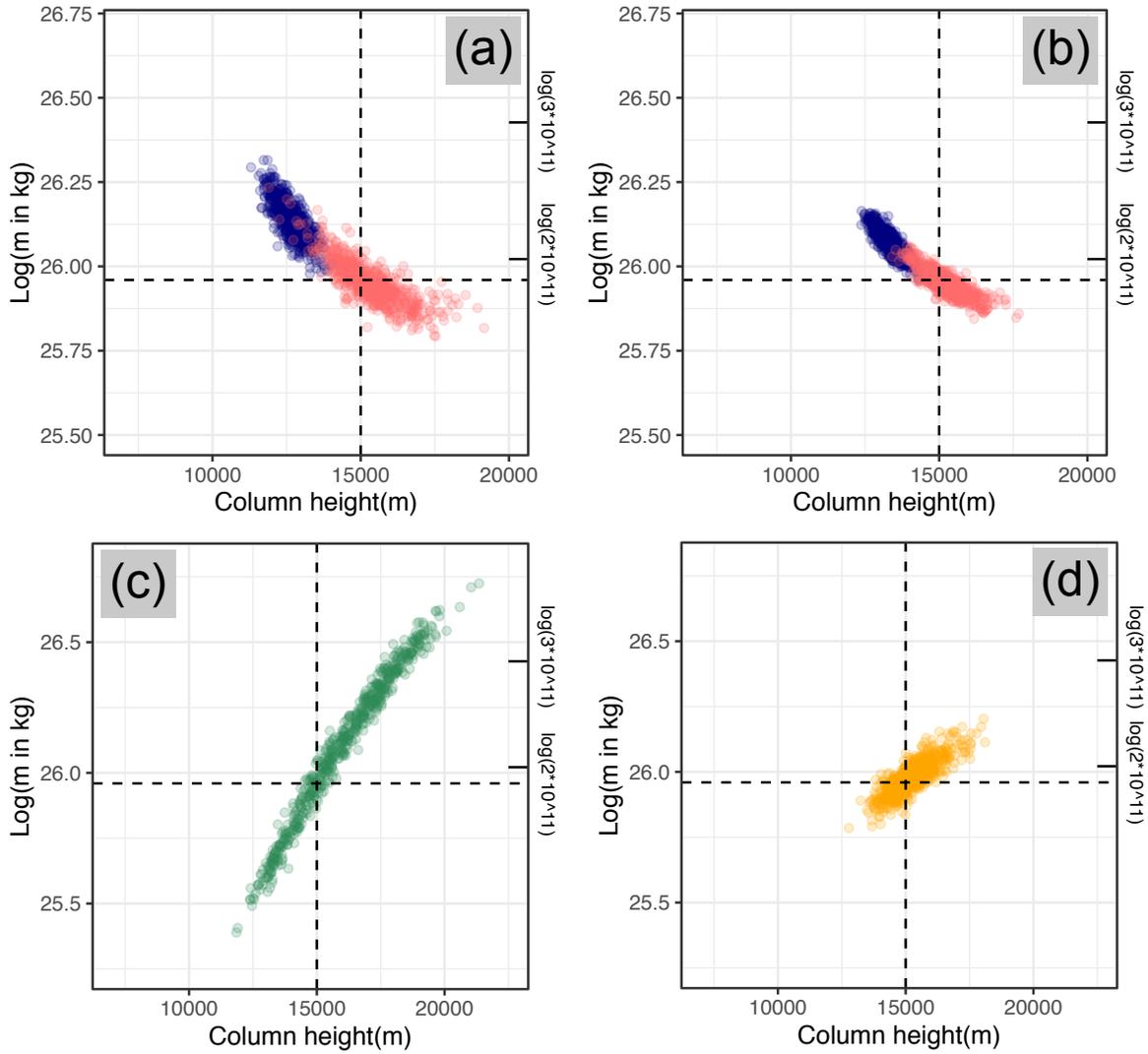


Figure 3: Selected sampled posterior distributions of column height and log-transformed total eruption mass in 2D. Dashed lines mark true values of column height and log-transformed total eruption mass. a: posterior distributions from Set 1 Experiments # 0 (red; reference experiment) and 4 (blue; experiment with incorrect priors); b: posterior distributions from Set 1 Experiments # 5 (red; experiment with 30 observations) and 9 (blue; experiment with 30 observations and incorrect priors). c and d display posterior distributions from Set 1 Experiments # 10 and 11 (experiments with different sample site locations), respectively.

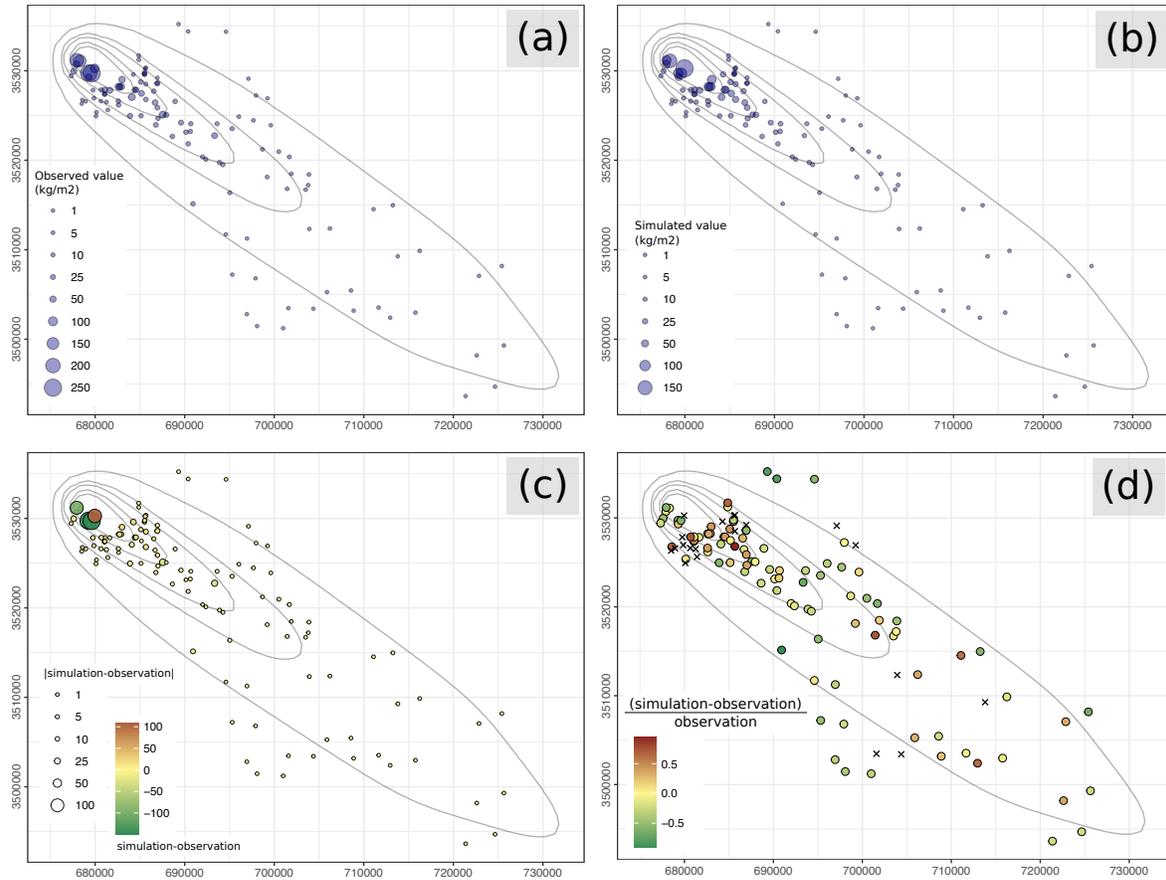


Figure 4: a and b: observed and simulated tephra masses per unit area, respectively, of the 2011 Kirishima-Shinmoedake eruption dataset. In b, the ESPs and wind conditions are determined based on the posterior means. See Table 8 for their values. The corresponding isopachs are also shown in the four subfigures. Levels of these isopachs are 50, 25, 10, 5, 1  $\text{kg}/\text{m}^2$ ; c: absolute difference (point size) and difference (point color) between simulated and observed masses per unit area. d: relative error, i.e.,  $(\text{simulation}-\text{observation})/\text{observation}$  for each sample. Points that have relative error with magnitude greater than 1 are marked as crosses.

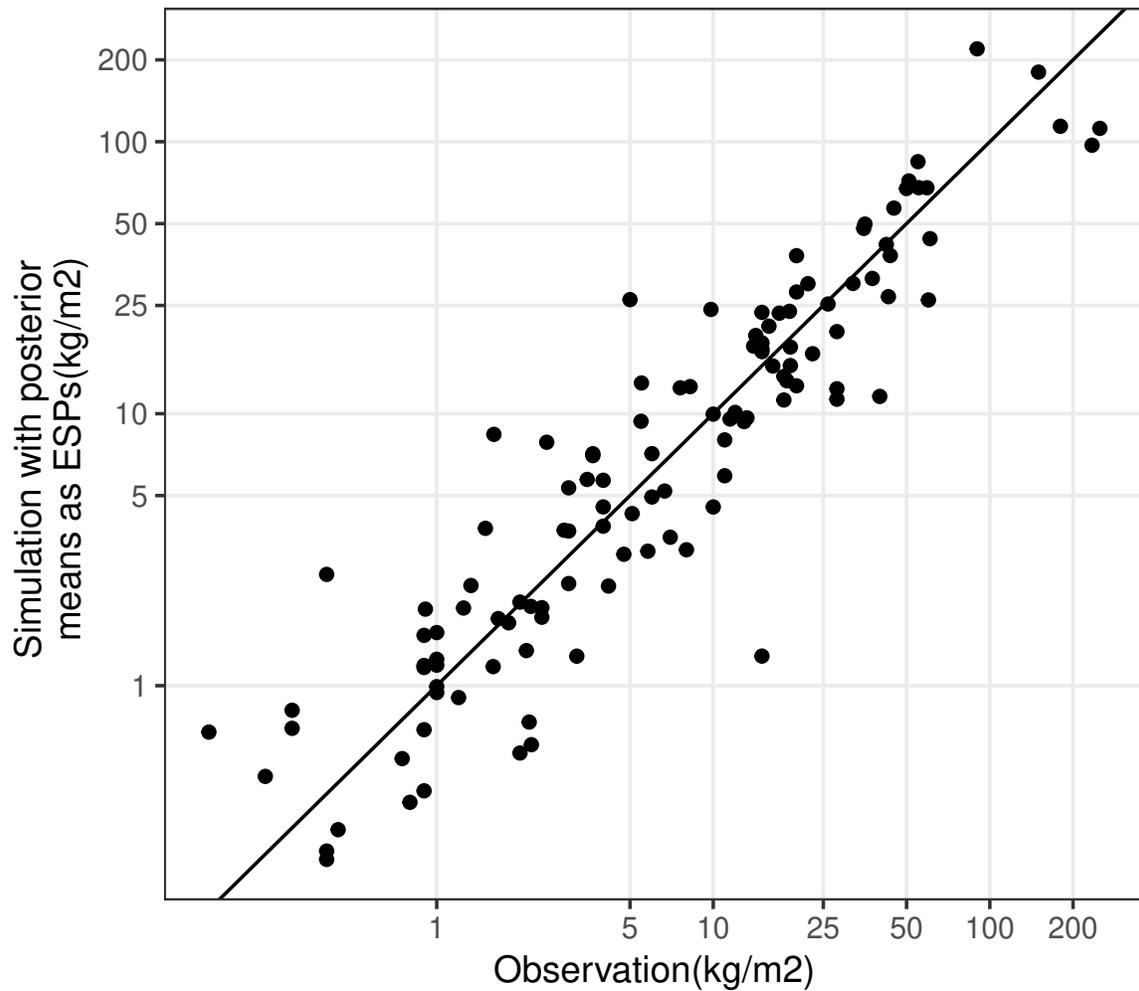


Figure 5: Simulated data from *Tephra2* using posterior means as ESPs and wind conditions plotted against observation data of the tephra deposit from the 2011 Kirishima-Shinmoedake eruption under log-scale.

## 969 References

- 970 Anderson, K. R., Johanson, I. A., Patrick, M. R., Gu, M., Segall, P., Poland, M. P., Montgomery-  
 971 Brown, E. K., and Miklius, A. (2019). Magma reservoir failure and the onset of caldera collapse at  
 972 Kīlauea Volcano in 2018. *Science*, 366(6470).
- 973 Andrieu, C., De Freitas, N., Doucet, A., and Jordan, M. I. (2003). An introduction to mcmc for

974 machine learning. *Machine learning*, 50(1-2):5–43.

975 Armienti, P., Macedonio, G., and Pareschi, M. (1988). A numerical model for simulation of tephra  
976 transport and deposition: Applications to May 18, 1980, Mount St. Helens eruption. *Journal of*  
977 *Geophysical Research: Solid Earth*, 93(B6):6463–6476.

978 Bear-Crozier, A., Pouget, S., Bursik, M., Jansons, E., Denman, J., Tupper, A., and Rustowicz, R.  
979 (2020). Automated detection and measurement of volcanic cloud growth: towards a robust estimate  
980 of mass flux, mass loading and eruption duration. *Natural Hazards*, 101(1):1–38.

981 Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business  
982 Media.

983 Bevilacqua, A., Bursik, M., Patra, A., Bruce Pitman, E., Yang, Q., Sangani, R., and Kobs-Nawotniak,  
984 S. (2018). Late Quaternary eruption record and probability of future volcanic eruptions in the Long  
985 Valley volcanic region (CA, USA). *Journal of Geophysical Research: Solid Earth*.

986 Bevilacqua, A., Isaia, R., Neri, A., Vitale, S., Aspinall, W. P., Bisson, M., Flandoli, F., Baxter, P. J.,  
987 Bertagnini, A., Ongaro, T. E., et al. (2015). Quantifying volcanic hazard at Campi Flegrei caldera  
988 (Italy) with uncertainty assessment: 1. Vent opening maps. *Journal of Geophysical Research: Solid*  
989 *Earth*, 120(4):2309–2329.

990 Biass, S., Bonadonna, C., Connor, L., and Connor, C. (2016). TephraProb: a Matlab package for  
991 probabilistic hazard assessments of tephra fallout. *Journal of Applied Volcanology*, 5(1):10.

992 Biass, S., Bonadonna, C., and Houghton, B. F. (2019). A step-by-step evaluation of empirical methods  
993 to quantify eruption source parameters from tephra-fall deposits. *Journal of Applied Volcanology*,  
994 8(1):1.

995 Biass, S., Frischknecht, C., and Bonadonna, C. (2012). A fast GIS-based risk assessment for tephra  
996 fallout: the example of Cotopaxi volcano, Ecuador-Part II: vulnerability and risk assessment. *Nat-*  
997 *ural hazards*, 64(1):615–639.

998 Biass, S., Todde, A., Cioni, R., Pistolesi, M., Geshi, N., and Bonadonna, C. (2017). Potential impacts  
999 of tephra fallout from a large-scale explosive eruption at Sakurajima volcano, Japan. *Bulletin of*  
1000 *Volcanology*, 79(10):73.

1001 Bonadonna, C., Biass, S., and Costa, A. (2015). Physical characterization of explosive volcanic  
1002 eruptions based on tephra deposits: propagation of uncertainties and sensitivity analysis. *Journal*  
1003 *of Volcanology and Geothermal Research*, 296:80–100.

1004 Bonadonna, C., Connor, C. B., Houghton, B., Connor, L., Byrne, M., Laing, A., and Hincks, T.  
1005 (2005). Probabilistic modeling of tephra dispersal: Hazard assessment of a multiphase rhyolitic  
1006 eruption at Tarawera, New Zealand. *Journal of Geophysical Research: Solid Earth*, 110(B3).

1007 Bonadonna, C., Connor, L., Connor, C. B., and Courtland, L. M. (2010). Tephra2.

1008 Bonadonna, C. and Costa, A. (2012). Estimating the volume of tephra deposits: a new simple strategy.  
1009 *Geology*, 40(5):415–418.

1010 Bonadonna, C., Ernst, G., and Sparks, R. (1998). Thickness variations and volume estimates of tephra  
1011 fall deposits: the importance of particle Reynolds number. *Journal of Volcanology and Geothermal*  
1012 *Research*, 81(3-4):173–187.

1013 Bonadonna, C. and Houghton, B. (2005). Total grain-size distribution and volume of tephra-fall  
1014 deposits. *Bulletin of Volcanology*, 67(5):441–456.

1015 Bonasia, R., Macedonio, G., Costa, A., Mele, D., and Sulpizio, R. (2010). Numerical inversion and  
1016 analysis of tephra fallout deposits from the 472 AD sub-Plinian eruption at Vesuvius (Italy) through  
1017 a new best-fit procedure. *Journal of Volcanology and Geothermal Research*, 189(3-4):238–246.

1018 Bursik, M. (2001). Effect of wind on the rise height of volcanic plumes. *Geophysical Research Letters*,  
1019 28(18):3621–3624.

1020 Bursik, M., Carey, S., and Sparks, R. (1992a). A gravity current model for the May 18, 1980 Mount  
1021 St. Helens plume. *Geophysical research letters*, 19(16):1663–1666.

1022 Bursik, M., Sparks, R., Gilbert, J., and Carey, S. (1992b). Sedimentation of tephra by volcanic  
1023 plumes: I. Theory and its comparison with a study of the Fogo A plinian deposit, Sao Miguel  
1024 (Azores). *Bulletin of Volcanology*, 54(4):329–344.

1025 Carey, S. and Sparks, R. (1986). Quantitative models of the fallout and dispersal of tephra from  
1026 volcanic eruption columns. *Bulletin of volcanology*, 48(2-3):109–125.

1027 Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The american*  
1028 *statistician*, 49(4):327–335.

- 1029 Connor, C. B., Connor, L. J., Bonadonna, C., Luhr, J., Savov, I., and Navarro-Ochoa, C. (2019).  
1030 Modelling tephra thickness and particle size distribution of the 1913 eruption of Volcán de Colima,  
1031 Mexico. In *Volcán de Colima*, pages 81–110. Springer.
- 1032 Connor, L., Connor, C., and Saballos, A. (2011). Tephra2 Users Manual. *University of South Florida*,  
1033 *Tampa, FL, accessed Sept, 24:2017*.
- 1034 Connor, L. J. and Connor, C. B. (2006). Inversion is the key to dispersion: understanding eruption  
1035 dynamics by inverting tephra fallout.
- 1036 Costa, A., Macedonio, G., and Folch, A. (2006). A three-dimensional Eulerian model for transport  
1037 and deposition of volcanic ashes. *Earth and Planetary Science Letters*, 241(3-4):634–647.
- 1038 Covey, J., Horwell, C. J., Rachmawati, L., Ogawa, R., Martin-del Pozzo, A. L., Armienta, M. A.,  
1039 Nugroho, F., and Dominelli, L. (2019). Factors motivating the use of respiratory protection against  
1040 volcanic ashfall: A comparative analysis of communities in Japan, Indonesia and Mexico. *International  
1041 Journal of Disaster Risk Reduction*, 35:101066.
- 1042 Engwell, S., Aspinall, W., and Sparks, R. (2015). An objective method for the production of isopach  
1043 maps and implications for the estimation of tephra deposit volumes and their uncertainties. *Bulletin  
1044 of volcanology*, 77(7):61.
- 1045 Engwell, S., Sparks, R., and Aspinall, W. (2013). Quantifying uncertainties in the measurement of  
1046 tephra fall thickness. *Journal of Applied Volcanology*, 2(1):5.
- 1047 Folch, A., Costa, A., and Macedonio, G. (2009). FALL3D: A computational model for transport and  
1048 deposition of volcanic ash. *Computers & Geosciences*, 35(6):1334–1342.
- 1049 Fontijn, K., Ernst, G. G., Bonadonna, C., Elburg, M. A., Mbede, E., and Jacobs, P. (2011). The ~ 4-ka  
1050 Rungwe Pumice (South-Western Tanzania): a wind-still Plinian eruption. *Bulletin of volcanology*,  
1051 73(9):1353–1368.
- 1052 González-Mellado, A. and De la Cruz-Reyna, S. (2010). A simple semi-empirical approach to model  
1053 thickness of ash-deposits for different eruption scenarios. *Natural Hazards and Earth System Sci-  
1054 ences*, 10(11):2241.

- 1055 Green, R. M., Bebbington, M. S., Jones, G., Cronin, S. J., and Turner, M. B. (2016). Estimation of  
1056 tephra volumes from sparse and incompletely observed deposit thicknesses. *Bulletin of Volcanology*,  
1057 78(4):25.
- 1058 Hashimoto, A., Shimbori, T., and Fukui, K. (2012). Tephra fall simulation for the eruptions at Mt.  
1059 Shinmoe-dake during 26-27 January 2011 with JMANHM. *Sola*, 8:37–40.
- 1060 Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications.
- 1061 Hildreth, W. (2004). Volcanological perspectives on Long Valley, Mammoth Mountain, and Mono  
1062 Craters: several contiguous but discrete systems. *Journal of Volcanology and Geothermal Research*,  
1063 136(3-4):169–198.
- 1064 Jenkins, S., Magill, C., McAneney, J., and Blong, R. (2012). Regional ash fall hazard I: a probabilistic  
1065 assessment methodology. *Bulletin of volcanology*, 74(7):1699–1712.
- 1066 Jenkins, S., Magill, C., McAneney, J., and Hurst, T. (2008). Multistage volcanic events: tephra hazard  
1067 simulations for the Okataina Volcanic Center, New Zealand. *Journal of Geophysical Research: Earth*  
1068 *Surface*, 113(F4).
- 1069 Jenkins, S. F., Goldstein, H., Bebbington, M., Sparks, R., and Koyaguchi, T. (2019). Forecasting  
1070 explosion repose intervals with a non-parametric Bayesian survival model: Application to Sakura-  
1071 jima volcano, Japan. *Journal of Volcanology and Geothermal Research*, 381:44–56.
- 1072 Johnston, E., Phillips, J., Bonadonna, C., and Watson, I. (2012). Reconstructing the tephra dispersal  
1073 pattern from the Bronze Age eruption of Santorini using an advection–diffusion model. *Bulletin of*  
1074 *Volcanology*, 74(6):1485–1507.
- 1075 Jones, A., Thomson, D., Hort, M., and Devenish, B. (2007). The UK Met Office’s next-generation  
1076 atmospheric dispersion model, NAME III. In *Air pollution modeling and its application XVII*, pages  
1077 580–589. Springer.
- 1078 Kaipio, J. and Somersalo, E. (2006). *Statistical and computational inverse problems*, volume 160.  
1079 Springer Science & Business Media.
- 1080 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S.,  
1081 White, G., Woollen, J., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the*  
1082 *American meteorological Society*, 77(3):437–472.

- 1083 Kawabata, E., Bebbington, M. S., Cronin, S. J., and Wang, T. (2013). Modeling thickness variability  
1084 in tephra deposition. *Bulletin of volcanology*, 75(8):738.
- 1085 Klawonn, M., Frazer, L., Wolfe, C., Houghton, B., and Rosenberg, M. (2014). Constraining particle  
1086 size-dependent plume sedimentation from the 17 June 1996 eruption of Ruapehu Volcano, New  
1087 Zealand, using geophysical inversions. *Journal of Geophysical Research: Solid Earth*, 119(3):1749–  
1088 1763.
- 1089 Klawonn, M., Wolfe, C. J., Frazer, L. N., and Houghton, B. F. (2012). Novel inversion approach to  
1090 constrain plume sedimentation from tephra deposit data: Application to the 17 June 1996 eruption  
1091 of Ruapehu volcano, New Zealand. *Journal of Geophysical Research: Solid Earth*, 117(B5).
- 1092 Koyaguchi, T., Anderson, K., and Kozono, T. (2017). Bayesian estimation of analytical conduit-flow  
1093 model parameters from magma discharge rate observed during explosive eruptions. In *AGU Fall*  
1094 *Meeting Abstracts*.
- 1095 Koyaguchi, T. and Ohno, M. (2001). Reconstruction of eruption column dynamics on the basis  
1096 of grain size of tephra fall deposits: 1. Methods. *Journal of Geophysical Research: Solid Earth*,  
1097 106(B4):6499–6512.
- 1098 Lev, E., Conroy, C. J., Birnbaum, J., and Zhan, Y. (2019). Combining probabilistic methods and  
1099 deterministic models to estimate uncertainties in lava flow parameters inferred from observations.  
1100 *AGUFM*, 2019:V34C–02.
- 1101 Liang, C. and Dunham, E. M. (2020). Lava lake sloshing modes during the 2018 Kīlauea Volcano  
1102 eruption probe magma reservoir storativity. *Earth and Planetary Science Letters*, 535:116110.
- 1103 Madankan, R., Pouget, S., Singla, P., Bursik, M., Dehn, J., Jones, M., Patra, A., Pavolonis, M.,  
1104 Pitman, E. B., Singh, T., et al. (2014). Computation of probabilistic hazard maps and source  
1105 parameter estimation for volcanic ash transport and dispersion. *Journal of Computational Physics*,  
1106 271:39–59.
- 1107 Maeno, F., Nagai, M., Nakada, S., Burden, R. E., Engwell, S., Suzuki, Y., and Kaneko, T. (2014).  
1108 Constraining tephra dispersion and deposition from three subplinian explosions in 2011 at Shin-  
1109 moedake volcano, Kyushu, Japan. *Bulletin of Volcanology*, 76(6):823.

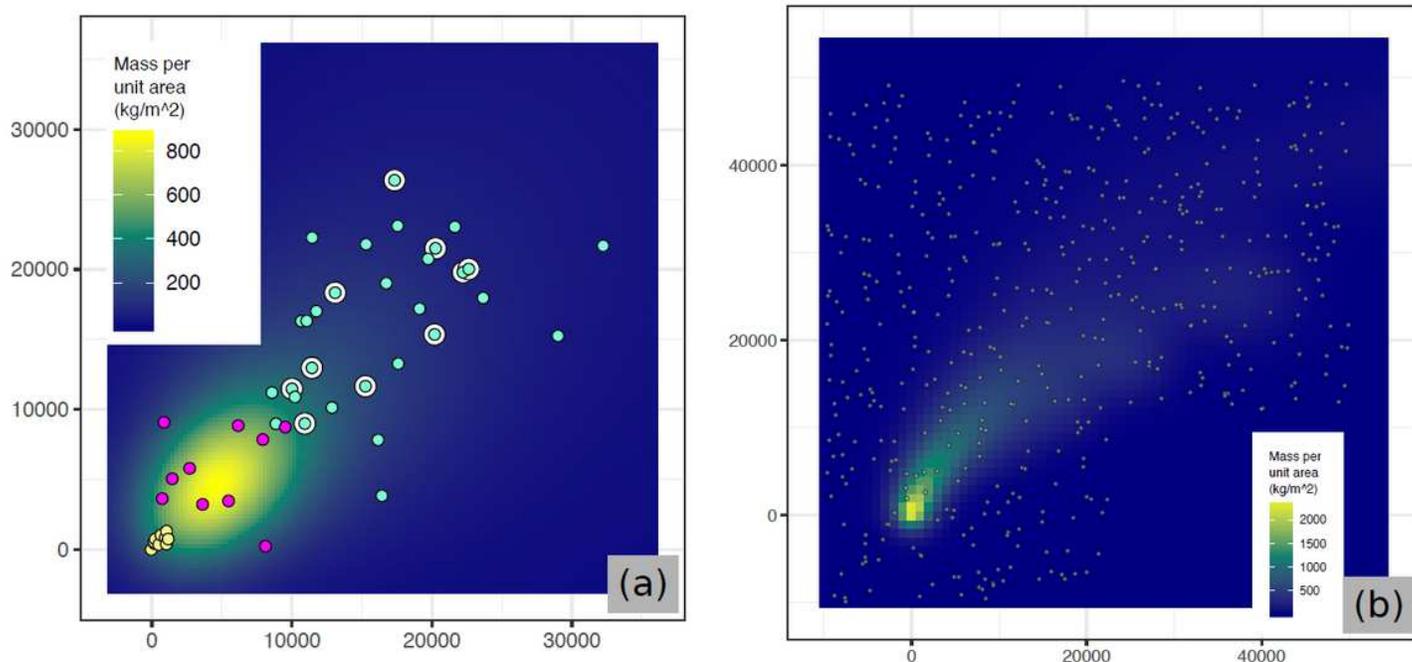
- 1110 Maeno, F., Nakada, S., Nagai, M., and Kozono, T. (2013). Ballistic ejecta and eruption condition  
1111 of the vulcanian explosion of Shinmoedake volcano, Kyushu, Japan on 1 February, 2011. *Earth,*  
1112 *Planets and Space*, 65(6):12.
- 1113 Magill, C., Mannen, K., Connor, L., Bonadonna, C., and Connor, C. (2015). Simulating a multi-phase  
1114 tephra fall event: inversion modelling for the 1707 Hoei eruption of Mount Fuji, Japan. *Bulletin of*  
1115 *Volcanology*, 77(9):81.
- 1116 Mannen, K. (2006). Total grain size distribution of a mafic subplinian tephra, TB-2, from the 1986 Izu-  
1117 Oshima eruption, Japan: An estimation based on a theoretical model of tephra dispersal. *Journal*  
1118 *of Volcanology and Geothermal Research*, 155(1-2):1–17.
- 1119 Mannen, K. (2014). Particle segregation of an eruption plume as revealed by a comprehensive analysis  
1120 of tephra dispersal: theory and application. *Journal of volcanology and geothermal research*, 284:61–  
1121 78.
- 1122 Mannen, K., Hasenaka, T., Higuchi, A., Kiyosugi, K., and Miyabuchi, Y. (2020). Simulations of  
1123 tephra fall deposits from a bending eruption plume and the optimum model for particle release.  
1124 *Journal of Geophysical Research: Solid Earth*, page e2019JB018902.
- 1125 Mastin, L. G., Guffanti, M., Servranckx, R., Webley, P., Barsotti, S., Dean, K., Durant, A., Ewert,  
1126 J. W., Neri, A., Rose, W. I., et al. (2009). A multidisciplinary effort to assign realistic source  
1127 parameters to models of volcanic ash-cloud transport and dispersion during eruptions. *Journal of*  
1128 *Volcanology and Geothermal Research*, 186(1-2):10–21.
- 1129 Miyabuchi, Y., Hanada, D., Niimi, H., and Kobayashi, T. (2013). Stratigraphy, grain-size and compo-  
1130 nent characteristics of the 2011 Shinmoedake eruption deposits, Kirishima Volcano, Japan. *Journal*  
1131 *of volcanology and geothermal research*, 258:31–46.
- 1132 Moiseenko, K. B. and Malik, N. A. (2019). Linear inverse problem for inferring eruption source param-  
1133 eters from sparse ash deposit data as viewed from an atmospheric dispersion modeling perspective.  
1134 *Bulletin of Volcanology*, 81(3):19.
- 1135 Nakada, S., Nagai, M., Kaneko, T., Suzuki, Y., and Maeno, F. (2013). The outline of the 2011 eruption  
1136 at Shinmoe-dake (Kirishima), Japan. *Earth, planets and space*, 65(6):1.

- 1137 Neri, A., Aspinall, W. P., Cioni, R., Bertagnini, A., Baxter, P. J., Zuccaro, G., Andronico, D., Barsotti,  
1138 S., Cole, P. D., Ongaro, T. E., et al. (2008). Developing an event tree for probabilistic hazard and  
1139 risk assessment at Vesuvius. *Journal of volcanology and geothermal research*, 178(3):397–415.
- 1140 Newhall, C. G. and Self, S. (1982). The volcanic explosivity index (VEI) an estimate of explosive  
1141 magnitude for historical volcanism. *Journal of Geophysical Research: Oceans*, 87(C2):1231–1238.
- 1142 Pieri, D. C. and Baloga, S. M. (1986). Eruption rate, area, and length relationships for some Hawaiian  
1143 lava flows. *Journal of Volcanology and Geothermal Research*, 30(1-2):29–45.
- 1144 Poret, M., Corradini, S., Merucci, L., Costa, A., Andronico, D., Montopoli, M., Vulpiani, G., Scollo,  
1145 S., and Freret-Lorgeril, V. (2017). Tephra dispersal and fallout reconstructed integrating field,  
1146 ground-based and satellite-based data: Application to the 23rd November 2013 Etna paroxysm. In  
1147 *AGU Fall Meeting Abstracts*.
- 1148 Pouget, S., Bursik, M., Webley, P., Dehn, J., and Pavolonis, M. (2013). Estimation of eruption  
1149 source parameters from umbrella cloud or downwind plume growth rate. *Journal of volcanology  
1150 and geothermal research*, 258:100–112.
- 1151 Scarpati, C., Cole, P., and Perrotta, A. (1993). The Neapolitan Yellow Tuff—a large volume multiphase  
1152 eruption from Campi Flegrei, southern Italy. *Bulletin of Volcanology*, 55(5):343–356.
- 1153 Schwaiger, H. F., Denlinger, R. P., and Mastin, L. G. (2012). Ash3d: A finite-volume, conservative  
1154 numerical model for ash transport and tephra deposition. *Journal of Geophysical Research: Solid  
1155 Earth*, 117(B4).
- 1156 Scollo, S., Tarantola, S., Bonadonna, C., Coltelli, M., and Saltelli, A. (2008). Sensitivity analysis and  
1157 uncertainty estimation for tephra dispersal models. *Journal of Geophysical Research: Solid Earth*,  
1158 113(B6).
- 1159 Shimbori, T. and Fukui, K. (2012). Time variation of the eruption cloud echo height from Shinmoe-  
1160 dake volcano in 2011 observed by Tanegashima and Fukuoka weather radars. Part II. *Rep. Coord.  
1161 Comm. Predict. Volcan. Erup*, 109:173–178.
- 1162 Sparks, R., Bursik, M., Ablay, G., Thomas, R., and Carey, S. (1992). Sedimentation of tephra by  
1163 volcanic plumes. part 2: controls on thickness and grain-size variations of tephra fall deposits.  
1164 *Bulletin of Volcanology*, 54(8):685–695.

- 1165 Sparks, R. and Young, S. (2002). The eruption of Soufriere Hills Volcano, Montserrat (1995-1999):  
1166 overview of scientific results. *Geological Society, London, Memoirs*, 21(1):45–69.
- 1167 Sparks, R. S. J., Bursik, M., Carey, S., Gilbert, J., Glaze, L., Sigurdsson, H., and Woods, A. (1997).  
1168 *Volcanic plumes*. Wiley.
- 1169 Stohl, A., Prata, A., Eckhardt, S., Clarisse, L., Durant, A., Henne, S., Kristiansen, N. I., Minikin, A.,  
1170 Schumann, U., Seibert, P., et al. (2011). Determination of time-and height-resolved volcanic ash  
1171 emissions and their use for quantitative ash dispersion modeling: the 2010 Eyjafjalla-jökull eruption.  
1172 *Atmospheric Chemistry and Physics*, 11(9):4333–4351.
- 1173 Suzuki, T. et al. (1983). A theoretical model for dispersion of tephra. *Arc volcanism: physics and*  
1174 *tectonics*, 95:113.
- 1175 Suzuki, Y. J. and Koyaguchi, T. (2013). 3D numerical simulation of volcanic eruption clouds during  
1176 the 2011 Shinmoe-dake eruptions. *Earth, Planets and Space*, 65(6):10.
- 1177 Takarada, S. (2017). The Volcanic Hazards Assessment Support System for the Online Hazard As-  
1178 sessment and Risk Mitigation of Quaternary Volcanoes in the World. *Frontiers in Earth Science*,  
1179 5:102.
- 1180 Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*, volume 89.  
1181 *siam*.
- 1182 Volentik, A. C., Bonadonna, C., Connor, C. B., Connor, L. J., and Rosi, M. (2010). Modeling tephra  
1183 dispersal in absence of wind: Insights from the climactic phase of the 2450 BP Plinian eruption of  
1184 Pululagua volcano (Ecuador). *Journal of Volcanology and Geothermal Research*, 193(1-2):117–136.
- 1185 Wang, T., Schofield, M., Bebbington, M., and Kiyosugi, K. (2020). Bayesian modelling of marked  
1186 point processes with incomplete records: volcanic eruptions. *Journal of the Royal Statistical Society:*  
1187 *Series C (Applied Statistics)*, 69(1):109–130.
- 1188 White, J., Connor, C., Connor, L., and Hasenaka, T. (2017). Efficient inversion and uncertainty  
1189 quantification of a tephra fallout model. *Journal of Geophysical Research: Solid Earth*, 122(1):281–  
1190 294.
- 1191 Wild, A., Wilson, T., Bebbington, M., Cole, J., and Craig, H. (2019). Probabilistic volcanic impact  
1192 assessment and cost-benefit analysis on network infrastructure for secondary evacuation of farm

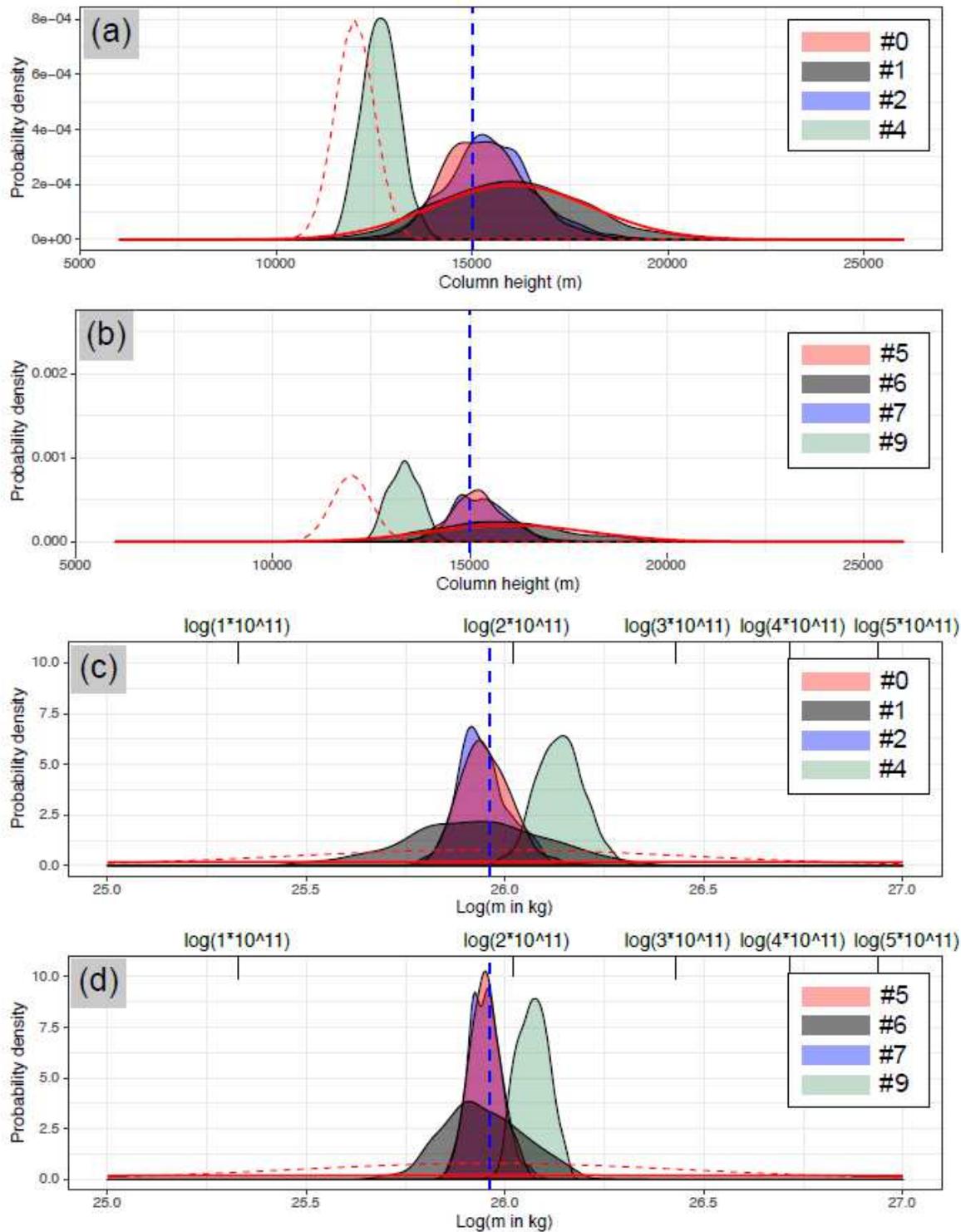
- 1193 livestock: A case study from the dairy industry, Taranaki, New Zealand. *Journal of Volcanology*  
1194 *and Geothermal Research*, 387:106670.
- 1195 Williams, G. T., Jenkins, S. F., Biass, S., Wibowo, H. E., and Harijoko, A. (2020). Remotely assess-  
1196 ing tephra fall building damage and vulnerability: Kelud Volcano, Indonesia. *Journal of Applied*  
1197 *Volcanology*.
- 1198 Yang, Q. and Bursik, M. (2016). A new interpolation method to model thickness, isopachs, extent,  
1199 and volume of tephra fall deposits. *Bulletin of Volcanology*, 78(10):68.
- 1200 Yang, Q., Bursik, M., and Pitman, E. B. (2019). A new method to identify the source vent location of  
1201 tephra fall deposits: development, testing, and application to key Quaternary eruptions of Western  
1202 North America. *Bulletin of Volcanology*, 81(9):51.
- 1203 Yang, Q., Pitman, E. B., Bursik, M. I., and Jenkins, S. (2020). Tephra inversion with Tephra2 and  
1204 the Metropolis-Hastings algorithm.

# Figures



**Figure 1**

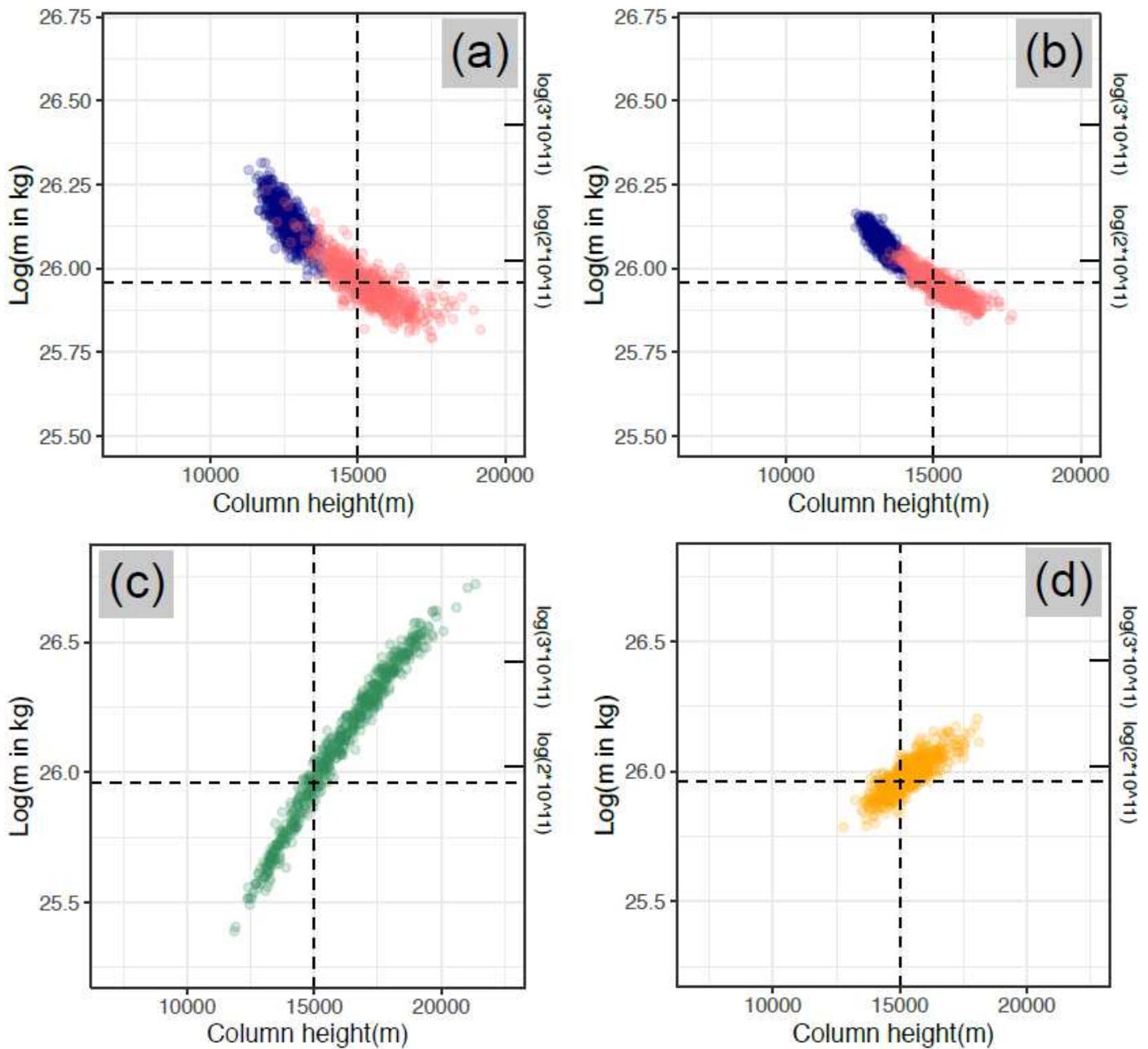
a: mass per unit area distribution used for the validation of Set 1 Experiments # 0-11 and sample site locations. White (larger), turquoise, yellow, and pink dots are the sample site locations used for Experiments # 0-4, 5-9, 10, and 11, respectively. All sites shown in a are used in Set 2 experiments as sample site locations; b: mass per unit area distribution used for the experiment with the non-simplified wind profile. Small white points correspond to sample site locations. Mass per unit area distributions in a and b are in different resolutions. This difference is only for easier visualization (reducing the number of grid points to be plotted in b), and would not affect any arguments or conclusions from this work.



**Figure 2**

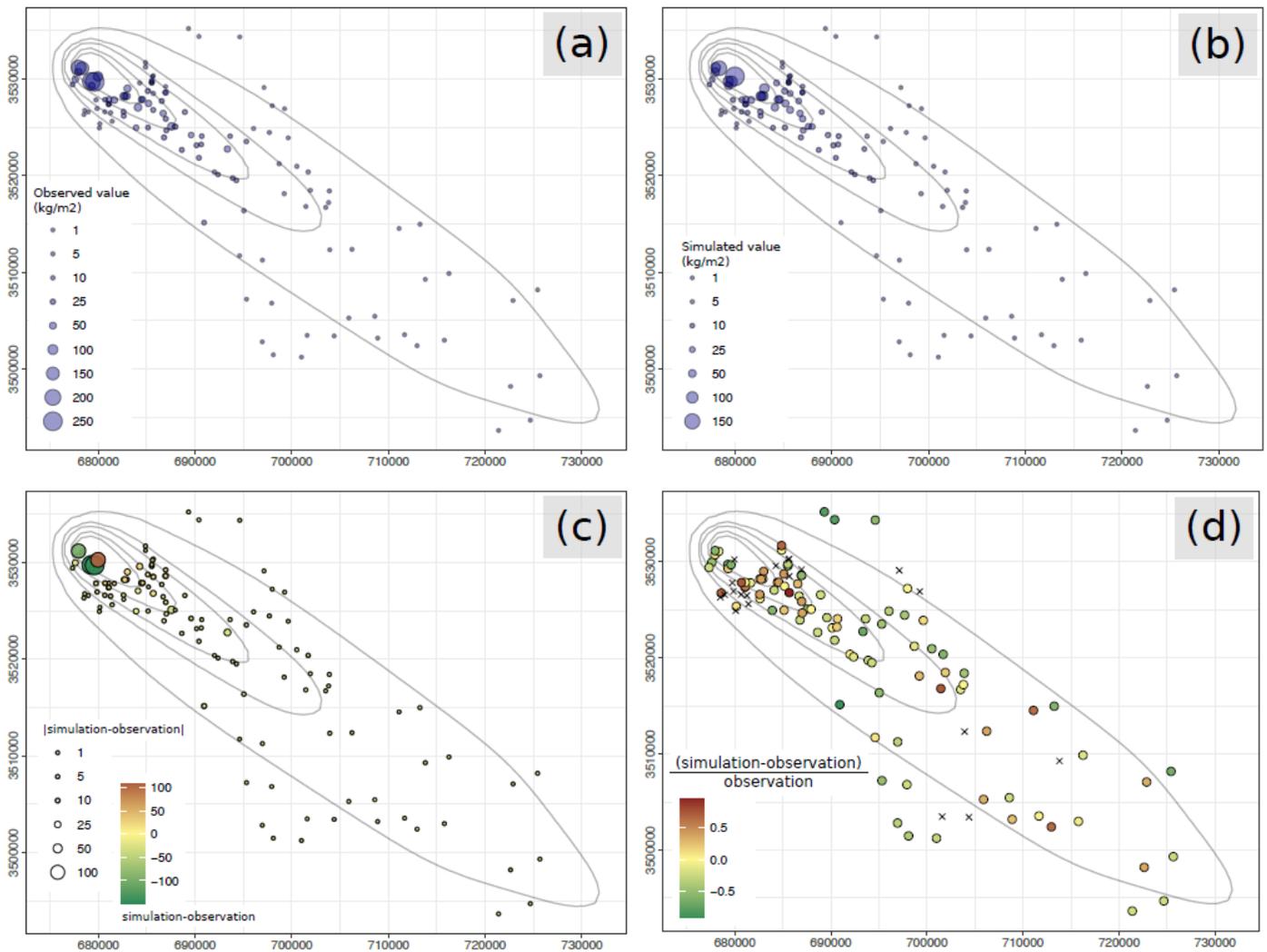
Selected posterior distributions of column height and log-scaled eruption mass. Results from Set 1 Experiments # 0, 1, 2, and 4 are shown in a (column height) and c (log-scaled eruption mass), and results from Set 1 Experiments # 5, 6, 7, and 9 are shown in b (column height) and d (log-scaled eruption mass). The blue dashed lines mark the true values of column height and log-transformed eruption mass used to

generate the observation data. The red solid lines correspond to prior distributions assumed for all experiments except for Set 1 Experiments # 4 and 9, and their priors are denoted as red dashed lines.



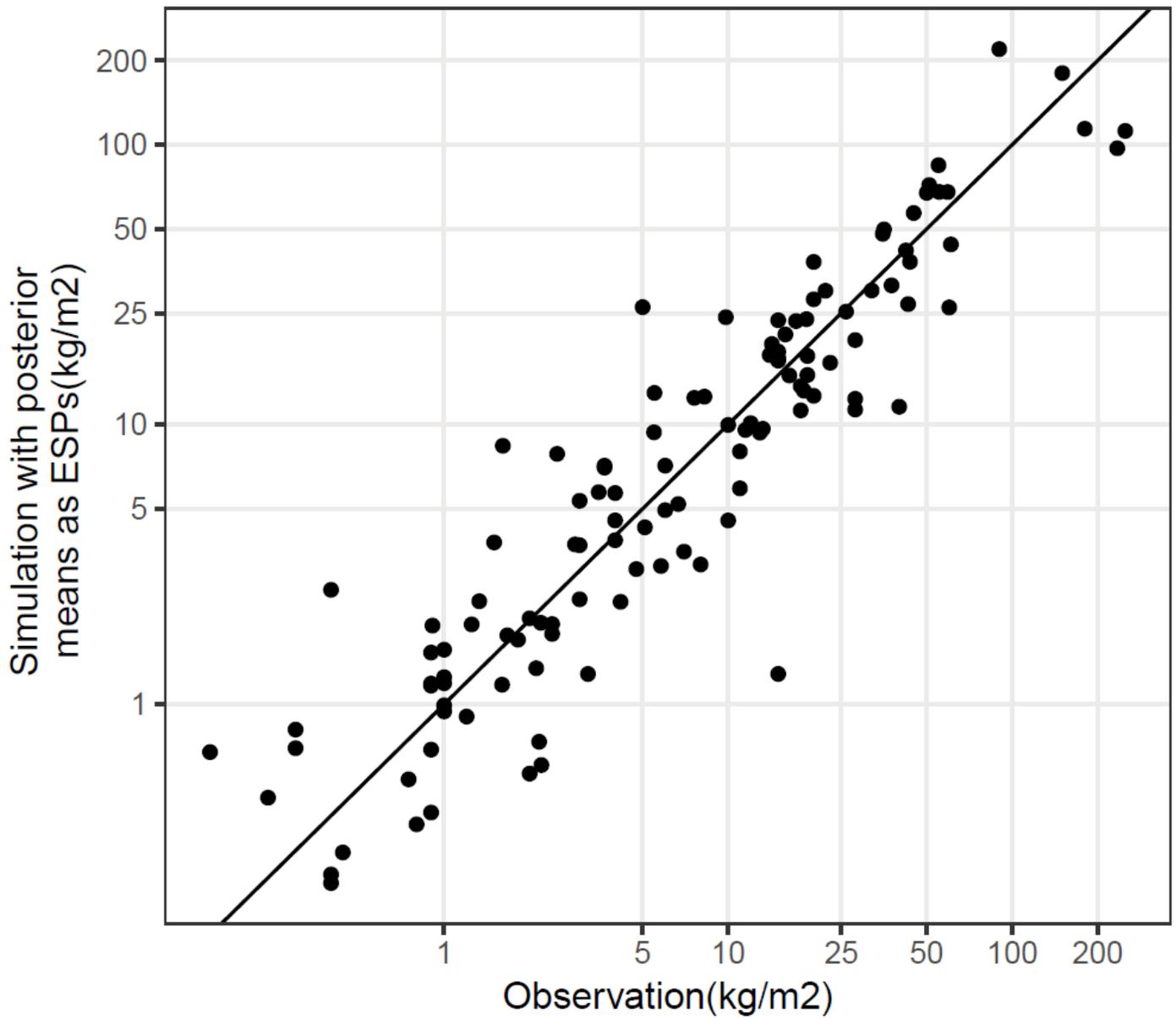
**Figure 3**

Selected sampled posterior distributions of column height and log-transformed total eruption mass in 2D. Dashed lines mark true values of column height and log-transformed total eruption mass. a: posterior distributions from Set 1 Experiments # 0 (red; reference experiment) and 4 (blue; experiment with incorrect priors); b: posterior distributions from Set 1 Experiments # 5 (red; experiment with 30 observations) and 9 (blue; experiment with 30 observations and incorrect priors). c and d display posterior distributions from Set 1 Experiments # 10 and 11 (experiments with different sample site locations), respectively.



**Figure 4**

a and b: observed and simulated tephra masses per unit area, respectively, of the 2011 Kirishima-Shinmoedake eruption dataset. In b, the ESPs and wind conditions are determined based on the posterior means. See Table 8 for their values. The corresponding isopachs are also shown in the four subfigures. Levels of these isopachs are 50, 25, 10, 5, 1 kg/m<sup>2</sup>; c: absolute difference (point size) and difference (point color) between simulated and observed masses per unit area. d: relative error, i.e., (simulation-observation)/observation for each sample. Points that have relative error with magnitude greater than 1 are marked as crosses.



**Figure 5**

Simulated data from Tephra2 using posterior means as ESPs and wind conditions plotted against observation data of the tephra deposit from the 2011 Kirishima-Shinmoedake eruption under log-scale.