https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

Long-Term Ash Dispersal Dataset of the Sakurajima Taisho Eruption for Ashfall Disaster Countermeasure

Haris Rahadianto[1, 2], Hirokazu Tatano[2], Masato Iguchi[3], Hiroshi L. Tanaka[4], Tetsuya Takemi[2], Sudip Roy[5]

[1]Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, Japan
[2]Disaster Prevention Research Institute, Kyoto University, Uji, 611-0011, Japan
[3]Sakurajima Volcano Research Center, Disaster Prevention Research Institute, Kyoto University, Sakurajima, 851-1419, Japan
[4]Center for Computational Sciences, Division of Global Environmental Science, University of Tsukuba, Ibaraki, 305-8577, Japan
[5]Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, 247-667, India

Correspondence to: Haris Rahadianto (haris.rahadianto.88n@st.kyoto-u.ac.jp)

Abstract. We present the ashfall deposit and airborne ash concentration dataset from ash dispersal simulation of a large-scale explosive volcanic eruption as a reference for ashfall disaster countermeasure. We select the Taisho (1914) eruption in Sakurajima volcano, regarded as the strongest eruption in Japan in the last century, as our case study to provide a baseline for the worst-case scenario. We employ one eruption scenario approach by replicating the actual eruption under various extended weather conditions to show how it would affect contemporary Japan. Accumulated ashfall has devastating impacts on both surrounding areas of the volcano and other regions, affecting airline transportation, socio-economics activities, and human health. Therefore, it is crucial to discover places with a high probability of exposure to ashfall deposition. This knowledge can help assess the additional risk in the infrastructures, human lives, and economic impacts to make a better volcanic eruption response plan. We generate the ash dispersal dataset by simulating the ash transport of the Taisho eruption scenario with a volcanic ash dispersal model and meteorological reanalysis data for 64 years (1958-2021). We explain the dataset production process and provide the dataset in multiple formats for broader audiences. We also clarify the validity of the dataset with its limitations and uncertainties. The dataset is available at the DesignSafe-CI Data Depot: https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-2848v2 or through the DOI: https://www.doi.org/10.17603/ds2-vw5f-t920 by selecting Version 2 (Rahadianto and Tatano, 2020).

1 Introduction

A volcanic eruption is one of the events that emits several dangerous pollutants affecting human lives. Large eruptions eject enormous tephra and other eruptive materials into the air. In the form of very fine ash, tephra hamper the residents near the volcano and citizens in the farther area (Bonadonna et al., 2021).

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

When a large eruption happens, volcanic ash could travel far away from the volcano, disrupting socio-economic activities in many different ways, damaging critical infrastructures, buildings, and causing health problems among the people (Wilson et al., 2012; Zuccaro et al., 2013). Settled volcanic ashes (i.e., ashfall) are directly affecting human health and livelihoods; destroying vegetation, crops, and pastures; causing physical damages to infrastructure, such as clogging drainage systems and contaminating water supplies; also disrupting the traffic and damaging the vehicles on roads (Barsotti et al., 2010; Ayris and Delmelle, 2012; Damby et al., 2013; Zuccaro et al., 2013; Poulidis et al., 2018). The cumulative weight of ashfall on the roof could collapse buildings and cause short-circuit electricity inside them (Zuccaro et al., 2013; Hampton et al., 2015). In addition, ash plumes from the volcanoes are a safety hazard that can severely damage commercial airlines (Folch et al., 2012; Peng and Peterson, 2012; Tanaka and Iguchi, 2019). The 2010 eruption of Eyjafjallajökull amplifies the evidence. Volcanic ashes flew to thousands of kilometres away, forcing almost entire airports in 25 countries to close for a week-long and costing a loss of billions of dollars revenue for airlines (Folch et al., 2012; Peng and Peterson, 2012). Apart from the eruption scale, the weather condition, mainly winds, primarily contribute to escalating ashfall catastrophe to a broader range. Ergo, ashfall hazards have uncertainty based on (1) its eruption magnitude and intensity and (2) the wind condition at the time of eruption that can bring ashes to distant places. Based on the evidence that ashfall hazards have dynamically changing exposures toward socio-economic activities, we believe estimating the risk for finding effective countermeasures is critical.

The present densely populated and modernised cities require comprehensive volcanic risk reduction strategies (Miyaji et al., 2011). One of the necessary actions for lessening the impacts of ashfall hazards is to assess the risk in the infrastructures, human lives, and economic impacts to develop a better response plan. Accordingly, the hazards and risk assessment for ashfall mostly rely on quantifying accumulated ashfall on the ground, then extending the analysis by combining social data based on the subject of interest (population data, building data, and many more). Many researchers use such a typical method to give expected impacts probability for future eruption events (Jenkins et al., 2014, 2015, 2018; Wilson et al., 2012; Biass et al., 2017). A comprehensive ashfall risk assessment requires both long-term weather and simulations for statistical discussion purposes, such as selecting typical conditions and analysing circumstances on exceptional conditions (Hattori et al., 2013, 2016). An important component to conducting such a study is to have long-term ashfall deposit data encompassing vast impact areas during an extended period.

This paper presents the dataset generation of the ash dispersal products over a vast region for 64 years (1958-2021) to provide vital input to develop a more comprehensive ashfall risk assessment and emergency management. We select Sakurajima volcano as an area of interest due to its high explosivity and its potential to have a large-scale explosive eruption within the following 20 to 25 years (Yamasato et al., 2013; Hickey et al., 2016; Biass et al., 2017; Poulidis et al., 2018). As supporting evidence, three large eruptions in Sakurajima volcano have occurred within the last six centuries: the Bunmei eruption in 1471, the An'ei eruption in 1779, and the Taisho eruption 1914. The latter is considered the biggest and the strongest eruption in Japan during the last century (Todde et al., 2017; Poulidis et al., 2018). Furthermore, Hickey et al. (2016) show that similar phenomena of magma supply rates that led to the Taisho eruption are currently happening, suggesting 130 years return period from the last eruption to the next one in the future.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

Consequently, the Sakurajima volcano tends to have another large-scale explosive eruption in the following multiple decades. Addressing such an urgent issue, Kagoshima city municipal government has been preparing the risk management and evacuation plan by providing an updated Sakurajima volcano hazard map. The hazard map contains necessary information about the impacts of historical eruptions, volcanic alerts and warnings, with a guideline on evacuation directions and procedures (Kagoshima City, 2010). Furthermore, the Osumi Office of the Rivers and National Highways produced maps for the potential ash deposition from a large eruption case in municipalities surrounding the Sakurajima volcano (Kyushu Regional Development Bureau, 2017). Lately, some research has assessed the impacts of both continuous Vulcanian activities and scenarios for an explosive eruption of Sakurajima volcano, providing insights on how volcanic ashes would disperse and affect livelihoods (Biass et al., 2017; Poulidis et al., 2018). However, those researches and precautions only discussed the proximal impacts in short to medium terms, with no further analysis available for the impacts to the distal locations. Furthermore, the previous large-scale explosive eruptions brought calamities across almost entire Japan. Thus, it is crucial to assess and develop a comprehensive risk analysis to a broader range as a larger population will be exposed when the ashfall arrives in major socio-economic centres of Japan, such as Fukuoka, Osaka, and Tokyo.

Using the Taisho eruption case, we simulate volcanic ash transport and dispersion model (VATDM) to demonstrate an ash dispersal process in large areas (entire Japan). We approach this issue by adapting one eruption scenario (OES) strategy, specifically as a baseline for the worst-case scenario (Bonadonna, 2006; Barsotti et al., 2018). We conducted this simulation in an extended period (1958-2021, 64 years) with the long-term reanalysis data to compensate for the incomplete wind profiles at the time of the eruption and to further validate the dataset. It is an attempt to characterise ashfall impacts from the Taisho eruption to contemporary Japan and understand the variety of ash dispersal characteristics of a large-scale explosive eruption in Sakurajima volcano. Furthermore, it is important to recognise the behaviour of ash transport, from discovering the condition of reproducing the actual ashfall distribution to finding the condition that led to much worse consequences. We produce this dataset for various purposes and usage, such as estimating the volcanic ashfall risk by analysing both the dispersal and deposition pattern volcanic ashfall over a long period; developing hazard maps for the disaster risk management process; and seeking a better comprehension of the tendency from ashfall distribution from a large eruption over a large region. Our contributions in this paper are as follow:

1. We produce the dataset consisting of:
    - The daily record of ashfall deposit for an extended period (1958-2021), covering Japan.
    - The daily record of ashfall deposit for an extended period (1958-2021), covering the southern Kyushu region, focusing mainly on Kagoshima prefecture as the closest urban area from the volcano.
    - Time-series record of airborne ash concentration, recorded at every one and a half hours after each simulation for an extended period (1958-2021).
2. We validate the accuracy of the dataset by using a novel method to find similar weather conditions to the day of the eruption using a machine learning technique.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

This paper is structured as follows: Section 1 introduces the general background and motivation of this study, followed by Section 2, which presents the process and impacts of the ashfall dispersal from the Taisho eruption in 1914, derived from previous studies and historical reports of the eruption. Section 3 provides the methodology for producing the dataset, and Section 4 describes the format of the dataset. Section 5 thoroughly addresses the validation method with the limitations of the dataset, and finally, Section 6 concludes the paper. Unless specified otherwise, all time dimensions in this paper use the Japan Standard Time (JST, +09UTC)

2 Ashfall Dispersal during the Taisho Eruption

The Taisho eruption is one of the eruption events that extensively studied by researchers all over the world, mainly due to:

1. The availability of detailed historical reports in English, which compiled shortly after the eruption (Omori, 1914; Koto, 1916).
2. The concurrent unrest activities of Sakurajima volcano (Iguchi, 2016; Iguchi et al., 2020; Poulidis et al., 2017, 2018).
3. The high possibility of a large-scale explosive eruption in the following decades probably resembles the past event (Hickey et al., 2016).

This paper will focus on how volcanic ashes travelled during the eruption process, especially to distal locations spread across Japan. For reference in regards to the complete evidence and chronology of the Taisho eruption, please refer to the chronicles mentioned above and the studies done by Kobayashi (1982); Yasui et al. (2006, 2007); and Todde et al. (2017). This section will explain the ashfall dispersal process of the Taisho eruption with references, unless specified otherwise, from both historical chronicles and the latest comprehensive study by Todde et al. (2017).

The Taisho eruption is a large-scale explosive eruption with Volcanic Explosivity Index (Newhall and Self, 1982) (VEI) 4. The Taisho eruption ejected enormous eruptive materials from two main active vents during its explosive phases. The explosive eruption started at 10 am on 12 January 1914 and lasted 48 hours. The tephra mostly ejected from the western vent (Nabeyama) mixed with the ones from the eastern vent (Yokoyama), producing ~~gigantic~~ plumes estimated to be between 10 km and 18 km high. The westerlies in the upper altitude and the surface winds influence the ashfall dispersal pattern. Both prevailing winds and ~~considerable~~ plume height are the significant drivers of the vast ash deposition process over a very wide area. Most tephra dispersed eastward, leaving Kagoshima City with only minor ash deposits on the first several days of the eruption.

In contrast, the location close to the eastern vent had four meters deep ashfall deposit. Three days after the eruption started, the ashfall deposits were reported at most of Kyushu, Shikoku, western Japan, and Sendai in northern Japan. Village offices, tobacco plantations, and local meteorological observatories kept a record of the exact time of ash arrival and sighting, showing that volcanic ashes reached the major cities of Fukuoka at 08 am on 13 January, Osaka by midnight of 13 January, Tokyo in the morning of 14 January 1914, and finally Sendai in the afternoon of the same day.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Another important notion is the report of heavy ash fallout (~1.5 × 10^6 km^2) at Ogasawara (Bonin Islands), around 1222 km from the vent. Todde et al. (2017) present a simple and straightforward isochrones map derived from the chronicles in Fig 1. Before reaching the climactic phase, the dispersal of very fine ashes during the first 13 hours of the eruption perhaps caused such a phenomenon. Meanwhile, the activity in the following days was largely minor in intensity, and ashfall kept occurring within the 200 km range from the volcano until 19 January 1914. The average velocity of the winds carrying tephra was found to be around nine m/s. Unfortunately, although there was an abundance of ashfall reports in entire Japan, there is no concrete measurement that can give clear and definitive clarification on the extent of ash dispersal. As most of the recent research only focused on the proximal impacts near the volcano, there is still a lack of complete analysis on the effects in distal locations. Based on the collected reports, there is only one attempt to draw all ashfall affected areas completely. The wide exposure map, shown in Fig. 2, did not get the same interest for modernisation compared to the dispersal maps at proximal regions that have been updated recently (see supplementary materials by Todde et al., (2017) and the latest report conducted by Mita et al., (2018)). The ash dispersal map showing the eruption impacts in the southern Kyushu region contains complete information about the ground measurement of ash deposits. However, the map of ash dispersal for the entire Japan did not have an exact ash measurement at all. Though limited, the information provided in Fig. 2 is vital as the extent of distance the ashfall travelled is known. The fast-travelling ash endangers a larger population and greatly disrupts airline transportation. Pieces of evidence from the latest large eruption events (e.g., St. Helens in 1980, Pinatubo in 1991, Eyjafjallajökull in 2010) demonstrated the cataclysmic consequences of ashfall on wider areas (Tilling et al., 1990; Fero et al., 2008, 2009; Folch et al., 2012; Peng and Peterson, 2012).

Economic losses amounted to billions due to obstructed air transportation is one of the more significant repercussions that did not realise by the authorities during the Taisho eruption event that occurred more than a century ago. Moreover, the Japanese airspace would suffer worse implications if a similar explosive eruption occurred shortly. At least 40% of the air traffic in Japanese airspace will be affected if Sakurajima volcano has a large-scale explosive eruption in the present day. The number of possible affected passengers is almost four times higher than the number of people affected by devastating Typhoon Jebi back in 2018 (Takebayashi et al., 2021). The current built environment and infrastructures vulnerable to ashfall hazards require better comprehension of the upcoming volcanic risk for proximal and distal locations from the volcano. It is urgent to learn how large the catastrophic would affect contemporary Japan if the same event occurred in recent times. One of the main aims we produce the dataset here is to provide the necessary information, which is vital in modern times. By having an extended period dataset on ash dispersal in wider areas for a worst-case scenario, hopefully, comprehensive countermeasure strategies can be derived to reduce the ashfall risk further.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

3 Data Generation Methods

3.1 Simulation on Selected Eruption Scenario

We use a fixed volcanological scenario over an extended contemporary time recorded in the meteorological reanalysis dataset to capture extended daily variability of ash dispersal patterns from the Taisho eruption. We run the PUFF model for cases of 23376 days from 1 January 1958 to 31 December 2021. The PUFF model is an ash tracking model developed during the Redoubt volcano eruption in 1989, which is famous for its high resolution and high accuracy results (Tanaka, 1994; Searcy et al., 1998; Scollo et al., 2011; Folch, 2012). This model considers the ash as a collection of a finite number of virtual particles and computes the motion of the particles with time (Searcy et al., 1998; Tanaka, 1994). The PUFF model is a Lagrangian-based model that has several advantages compared to another approach (Eulerian), such as: provides trajectory information, has physical realism, can describe non-diffusive near-field to sources, has numerical stability, lacks numerical diffusion, has conservation properties, and can resolve sub-grid scale variability (Lin et al., 2012). Among all the nine Volcanic Ash Advisory Centers (VAACs) worldwide, eight VAACs, including Tokyo VAAC, use and operate models with the Lagrangian approach (Folch, 2012; Lin et al., 2012). The recent example of the usage of the PUFF model for the simulation for explosive eruptions in Sakurajima volcano shows a satisfactory outcome. This model applied to cases of explosive eruption in 2017 and 2018 to assess the utility of forecasting real-time ash fallout. Both simulations agreed well with the measurement recorded by instruments installed at several points nearby the volcano (Tanaka and Iguchi, 2019).

Here, we construct the model using a large number of random variables $r_i(t)$, where $i = 1 \sim M$ and M is the total number of particles. $r_i(t) = (x, y, z)$ represents the position vector for an i^{th} particle at time t with its origin at the volcano vent. Using discrete-time increment $\Delta t = 300$ s, the governing equation is written as:

$$\begin{cases} r_i(0) = S_i, & i = 1 \sim M, \text{ for } t = 0, \\ r_i(t + \Delta t) = r_i(t) + V\Delta t + Z\Delta t + G\Delta t, & i = 1 \sim M, \text{ for } t > 0 \end{cases} \quad (1)$$

where S_i is the initial location of all the particles at the vent, $V = (u, v)$ is a vector for the wind velocity moving the particles and $Z = (c_h, c_h, c_v)$ is a vector for the diffusion velocity containing diffusion speeds generated by Gaussian random numbers. $G = (0,0, -w_t)$ is the gravitational fallout velocity by approximating the extended Stokes Law for diverse particle sizes (Tanaka, 1994; Tanaka and Yamamoto, 2002). The movement of particles steers the diffusion Z direction, and the size of particles affects the gravitational fallout G (Searcy et al., 1998; Tanaka, 1994). The diffusion speed c (c_h or c_v) obtained by the random walk process related to the diffusion coefficient K as $c = \sqrt{2K/\Delta t}$ (Tanaka et al., 2016; Tanaka and Iguchi, 2019). Tanaka and Yamamoto (2002) conducted several diffusion tests with various values of K, and compared the results with satellite images of actual dispersals from several volcanic eruptions in the past. Based on such research, we assign the suitable horizontal diffusion coefficient $K_h = 150$ m^2s^{-1} and vertical diffusion coefficient $K_v = 1.5$ m^2s^{-1} as mentioned in Table 1.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

These diffusion coefficients are consistent with *in situ* observation documented by Eliasson et al. (2014) and adjusted for Sakurajima volcano. The values used in this study follow previous studies by Tanaka et al. (2016, 2019) but are smaller than the study conducted by Fero et al. (2008, 2009) and Kratzmann et al. (2010). In order to further investigate the ash dispersal process to distal locations in entire Japan, we extend the simulation period of the PUFF model to 96 hours after the eruption started. Following Eq. (1), by modelling the source location of volcano vent as $S(x, y, z)$, we can adjust the number of particles released at each time step to draw the optimal statistical information from the model. For each time step, we assign a Gaussian random number for each of the particles (M_0), scale it according to the emission rate, making the value of the total number of particles (M) changes linearly with the emission rate (Tanaka et al., 2016). We assign the initial number of particles to 5000 due to constraints in the available computational power.

The size of volcanic ash varies from fine ash to boulders, as large particles tend to settle out shortly and then become smaller as time goes. Due to this process, each ash particle will have a various grain size (Bonadonna et al., 1998), and we assume each particle has an initial grain size of a logarithmic Gaussian distribution with a standard deviation of 2.0 centred at -3.0. Thus, the average particle size is approximately 1.0 mm on a log scale, with 68% of the particles are within the range of 10 μm to 10 cm, as stated in Table 1. Given the initial vertical velocity of the emission with specified damping (e-folding) time τ_0, the particles are distributed randomly in a vertical manner from the vent z_1 to the plume's peak z_2 continually. Then, by using time integration for the vertical velocity in the momentum equation, we obtained the final form as $z(0) = z_1$ and $z(t) = z_2$ during the time step Δt (Tanaka et al., 2016).

3.2 Estimation on Mass Eruption Rate based on Plume Heights Transition

In this study, we use the estimated emission rate from the calculation done by Iguchi (2014), based on the previous study (Kobayashi, 1982). As shown in Fig. 3, we consider the maximum estimated plume height at 17,890 m around midnight of 13 January 1914, close to the range of plume heights in the chronicles explained in Sect. 2. The temporal change of the plume height allows us to obtain the value of mass eruption rate (ε) and the mean eruption mass (see Table 1). The values we obtained agree well with the recent observation (Todde et al., 2017). The pressure of the reservoir that ejects gas through a nozzle ($P_{(t)}$) at time t decays exponentially with time (Nishimura, 1998) following Eq. (2):

$$P_{(t)} = \Delta P_0 \left[1 + \frac{\gamma-1}{2}\left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{2(\gamma-1)}} v_0 \frac{A}{Vol} t\right]^{\frac{-2\gamma}{\gamma-1}} \quad (2)$$

Where Vol is the volume of the reservoir, A is the cross-sectional area of the nozzle, v_0 is the initial ejection velocity, P_0 is the initial pressure, and γ is the specific heat ratio (values within the range 1.01 – 1.4). Since the Taisho eruption is a Plinian eruption, the plume height (H) is proportional to a quarter of the power of heat discharge rate ($\frac{dQ}{dt}$) (Morton et al., 1956). Assuming the volume of the reservoir is much larger than the ejected gas volume, the mass eruption rate (ε) is equivalent to decay rate of pressure ($\frac{dP}{dt}$), which can be estimated from the changes in plume height as follows:

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

$$H \propto \left(\frac{dQ}{dt}\right)^{1/4} \propto (\varepsilon)^{1/4} = k\left(\frac{dP}{dt}\right)^{1/4} \quad (3)$$

3.3 Long-Term Wind Field

In order to conduct the long-term ash transport simulation, the long-term weather data from the meteorological reanalysis values are necessary, as the wind characteristics strongly affect the sedimentary properties of the volcanic ashfall (Hattori et al., 2013, 2016; Poulidis et al., 2018). One of the available long-period historical data is the JRA-55 reanalysis dataset (Kobayashi et al., 2015). This data contain the global meteorological reanalysis dataset from 1958 to date, developed by Japan Meteorological Agency (JMA) to study a long-time variation in atmospheric and climate phenomena. The JRA-55 employs a reduced Gaussian horizontal grid system and applies a vertically conservative semi-Lagrangian advection scheme. We utilise the JRA-55 data for the PUFF model, considering the horizontal wind grid as 1.25° × 1.25°. We employ 16 vertical layers from 1000 hPa to 10 hPa atmospheric pressure, and we use this data of four different times (00, 06, 12, 18 UTC) daily. The 3D wind field consists of zonal wind (*U*), meridional wind (*V*), and geo-potential height (*gph*), which are interpolated to the position of each ash particle using the cubic spline method in space and time. In Fig. 4, each wind rose corresponding to the probability of the provenance identified by sectors and sequential colours mark the speed gradients.

As discussed in Sect. 2, the ashfall dispersal from the Taisho eruption controlled by both westerlies (250 hPa) and surface winds (1000 hPa), we only portray the wind conditions in those pressure levels, with the first row indicating the upper altitude winds and the second row is the latter. The columns, from the second column onward, reflect seasonal variations: spring (March to May), summer (June to August), autumn (September to November), and winter (December to February), with the first column displaying entire dates selected for the simulation (1958-2021). As noted by previous research concerning the climatological conditions in the Sakurajima area, wind conditions change heavily based on the season. Thus, the ash deposition process largely depends on the time of the eruption (Biass et al., 2017; Poulidis and Takemi, 2017; Poulidis et al., 2018)

3.4 Recording Ashfall Location and Thickness

This study considers the ashfall deposits on the ground as particles with non-positive altitude, with their location marked as a longitude-latitude pair. For each simulation, as the computation process goes, the particles with a negative value in the altitude dimension were subset to other files before being compiled when the simulation finished. During such a process, we captured the rest of the flowing particles as airborne ash concentrations, copied them to different files, and saved them using the time-marking for every 90 minutes. We separate the measurement of ashfall deposit and the airborne ash concentrations into two different files using this mechanism. Further, we can measure the ashfall thickness (ashfall depth) by dividing the surface area with the ashfall density according to its particle size. First, we allocate the ashfall particles to a grid according to their location. Then, we set all ashfall particles in that grid as the average value considering all the adjacent grids and the pre-assigned mass values multiplied by the number of particles to obtain the total mass of the ashfall deposit.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

It is possible to get the mass of each particle based on the particle size. However, since we deliberately designate the total initial number of particles as a constant number and smaller than the actual number in the real case, the total mass of all particles is less than the actual, and the difference may increase as the scale of the eruption increases. Therefore, considering the particle size under the total mass conservation, giving each particle a virtual mass is necessary (Shimbori et al., 2009). We assume that one ash particle contained 450 tons of ash mass when plume height reached its peak, and by giving the virtual mass to each particle, we can alleviate the uncertainty that appeared when assigning the small number of particles in the simulation (see Scollo et al., 2011). After assigning the virtual mass to each particle in one grid, the thickness of the ashfall χ_{ij} at grid i, j obtained from the following equation:

$$\chi_{ij} = \sum_n \frac{\tilde{m}(D_n)}{\Delta x \Delta y} \quad (4)$$

where $\tilde{m}$ is the total virtual mass for the total n particles within the grid i, j of area $\Delta x \Delta y$ corresponding to the particle size D_n (Shimbori et al., 2009). All computation from running the simulation, processing wind fields, and computing the total ashfall accumulation on the ground done with a small workstation with 32GB RAM, dual 16-cores Intel Xeon E5-2620V4 2.10 GHz processors, and 4GB NVIDIA Quadro P1000 GPU.

4. Dataset Overview and Availability

Each ash dispersal simulation produces three data:

1. Ashfall deposit measurement from 128.5° E to 148.6° E, and 30.0° N to 45.9° N in 0.1°×0.1° grid (~10 km^2).
2. Ashfall deposit measurement from 129° E to 132° E, and 30.5° N to 32.5° N in 0.01°×0.01° grid (~1 km^2).
3. Time series collection of total airborne ash concentrations in the air recorded for every 90 minutes.

In order to ensure maximum availability for the diverse users of the dataset, we prepare the dataset in two formats. We stored the dataset as a space-separated ASCII value table in comma-separated value (CSV) format (see Table 2) and a multi-dimensional array structure in Network Common Data Format (NetCDF). NetCDF is useful for supporting access to diverse kinds of scientific data, and its files are self-describing, network-transparent, directly accessible, and extendible (Unidata, 2021). We develop the NetCDF files using the *xarray* library in Python using the NETCDF4 package (Hoyer and Hamman, 2017). Due to the high dimensionality of the airborne ash concentrations, we only provide this data in the NetCDF format. Meanwhile, the ashfall deposit data are available in CSV and NetCDF formats. We set the name of each dataset file as the simulation date in ISO-8601 format (*YYYYMMDD*<<*.csv, .nc*>>), stored inside a directory named by year (*YYYY*) within each respective region (Japan and Kagoshima). In addition, each data contain an ordered set of location marking (longitude, latitude) and ashfall depth (in cm), stored in two decimal floating-point formats. The set of location marking is the coordinates in the NetCDF data format of ashfall deposit data. Contrasting to the deposit data, the collection of airborne ash concentrations provide the total number of particles at a specific location and time. The tracking period of the ash particles start at 90 minutes after the simulation start and finish after 96 hours.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

As the three-dimensional shape of the affected area and the total number of particles change throughout the time dimension, we assign the *date_time* as the primary coordinate, as portrayed in Fig. 6. The dataset directory consists of two main folders based on the data observation: *Airborne* and *Deposit*. The *Deposit* folder contains two child directories based on the data format and observation range, then breaking down to every year starting from 1958 and ending in 2021. The *Airborne* dataset has a straightforward directory (year, *YYYY*) since it only consists of one data format, and there is no separate observation range. The dataset is available on the DesignSafe-CI Data Depot (https://www.designsafe-ci.org) hosted in the Texas Advanced Computing Center (TACC). This data depot is provided by the Natural Hazards Engineering Research Infrastructure (NHERI), which provides the natural hazards engineering community and researchers with comprehensive state-of-the-art cyberinfrastructure (Rathje et al., 2017). The dataset can be accessed directly at https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-2848v2 or through the DOI: https://www.doi.org/10.17603/ds2-vw5f-t920 (Rahadianto and Tatano, 2020). Users can access the dataset directly without any prior registration and can choose any data to download. The data can be downloaded as a single file or as a collection by selecting multiple files at once. Please select the Version 2 of the dataset for the latest update.

5 Dataset Validation and Limitations

5.1 Common Validations Techniques and Limited Information on Ground Truth of the Taisho Eruption

Traditionally, the primary way to confirm the accuracy of the simulation result of volcanic ash transport models is to compare with the ground data directly. Another method to validate the simulation result of such models is through satellite imagery for both qualitative and quantitative measurement. Apart from the space-based remote sensing, combining multiple ground and aeroplane-based remote sensing techniques can validate the models, although it is not as common as the primary method (Webley and Mastin, 2009; Folch, 2012). For example, Fero et al. (2009) measured the model's accuracy by comparing the model output with satellite data and the deposit report for the 1991 Mount Pinatubo (Webley and Mastin, 2009). Scollo et al. (2011) presented a novel mechanism to validate results from PUFF model simulations by evaluating the number of particles necessary to improve its results for the tephra deposit value in the ground and the ash concentration in the air. The study claimed that the PUFF model requires up to 33 and 220 million particles set in the simulation parameter to match the field data for ash deposits and their concentration in the air, respectively. This finding further confirms that simply increasing the particle number set in the model input improves model accuracy (Scollo et al., 2011; Bursik et al., 2012). However, increasing the number of particles in the simulation means heightening computational requirements (Peterson and Dean, 2002; Scollo et al., 2011; Madankan et al., 2014). Furthermore, increasing the number of particles in the simulation means increasing the computational time exponentially, requiring a trade-off between desired accuracy and computational cost (Scollo et al., 2011; Madankan et al., 2014). In this study, we found that validating our simulation results using the aforementioned conventional ways is challenging due to:

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

1. The lack of ground truth. Section 2 outlined the available reports gathered in cities and tobacco plantations across Japan only mentioned subjective ash deposit observations without a quantitative measurement, e.g., "frost-like ash", "house roof become white", and "slight deposit".
2. The complete wind data at the time of the eruption is absent. Only surface observation is available, and it is impossible to completely replicate the ash transport process, although we are using near-precise ESP. Therefore, finding the correct number of particles, following the study by (Scollo et al., 2011), is also not feasible due to the unavailability of sufficient computing resources and best-fit value for the simulation settings.

Despite all those complications, we try to utilise an innovative approach to verify the dataset presented here using all available information. We focus on validating the ash dispersal results to entire Japan as it is the main data produced here and one of the observations that previous studies have never touched before. The most conclusive information that can help us validate the dataset is the ash dispersal map presented in Fig. 2 and the weather condition that asserts control on how the ashes disperse all over Japan. Therefore, we refer to that information to proceed with the validation process. In order to facilitate the validation mechanism, we digitalised the ash dispersal map in Fig. 2 using ArcGIS software with WGS 84 projection (ESRI, 2017), as portrayed in Figure 7. In addition, we added the current airports' location to signify the impacts of the very fine ashfall on entire Japan. First, we conducted the map transformation by putting binary marking to indicate affected regions, i.e., 1 for red-stripped grids and 0 for the rest. All the binary values have a pair of longitude-latitude values for representing their position in the map. This procedure is necessary, as now we have validation data in the same area dimension with the model output. Second, we further observe the weather condition during the eruption that appears crucial. This feature of the Taisho eruption is often left out, although it explains why the ashfall from the eruption moves erratically to distal locations. Several prior researchers did not advance their analysis about this abnormality, which is proven to be influential concerning the impacts in distal municipalities (e.g., Biass et al., (2017), among many others). Understanding the weather component at the time of eruption is crucial to determine the behaviour of ash fallout trajectory.

The Taisho eruption occurred in winter when mid-latitude cyclones were active. With the stationary planetary waves generated by the land-sea contrasts and the flow over the Himalayas, these phenomena reinforce the tropospheric jet stream and the baroclinic wave to be the strongest. In Japan, the reinforced jet streams cause the westerlies at 250 hPa pressure levels to exceed 70 m/s. The winds in the tropospheric jet stream blow from the west throughout the year; they are strongest during winter and weakest during summer (Wallace and Hobbs, 2006). The Meteorological Research Institute of the JMA (JMA-MRI) finding further clarifies how the westerly wind traverses above Sakurajima volcano all year long from west to east direction. The westerlies reach their peak velocity in January, and then it keeps weakening until their lowest speed in summer, before rising again (Maeda et al., 2012). This cycle runs for the whole year and becomes one of the deciding factors of the ash dispersal pattern. Figure 8 portrays a time-latitude cross-section of the east-west wind at 250-hPa level (about 12 km from the surface), averaged at 130°E to 140°E based on the 30-year survey (Maeda et al., 2012; Mita et al., 2018). The position of Sakurajima volcano and its vicinity are at around 31° N (red dashed square), including Kagoshima city. Therefore, the westerlies would bring the most of the ashfall from the eruption to the east.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

Instead, the chronicles reported ashfall traversed to the north, west, and southwest (Omori, 1914; Todde et al., 2017). This pattern somehow contrasts the precedent eruption (An'ei eruption – November 1779) that follows the pattern of the westerlies (Tsukui, 2011). Research investigating the past ash dispersal pattern from Sakurajima volcano also reaffirm the differences due to the seasonal features, agreeing that during winter, tephra should mostly go eastward, which most of it would fall upon the Pacific (Biass et al., 2017; Poulidis et al., 2018; Mita et al., 2018). When the eruption occurred, the vicinity of Kagoshima city and Sakurajima volcano were experiencing a maximum atmospheric pressure in four days, providing a unique characteristic of the Taisho eruption. From 9 January 1914 to the morning of 13 January 1914, atmospheric pressure in Japan consistently recorded an unusual high, resulting in stable sea waves and low-velocity winds indicating calm and clear weather at most of the places (Omori, 1914). Unexpectedly, the weak winds were helpful to allow many people living near the volcano to evacuate by boat (Kitagawa, 2015). The eruption seemed to play a significant role in the increasing temperature in Kagoshima, but apparently, it turned out to be the opposite (Omori, 1914). The leading cause of the high temperature at the eruption time was a sudden increase of southerly winds, as observed in Okinawa, 373 km from the vent (Omori, 1914). Further, the chronicles explained tri-daily general weather conditions at the eruption and found that similar weather phenomena also occurred on 17 January 1914 (Omori, 1914). This vital information helped us derive a method to validate our simulation results.

Following such reports, we concluded that the clear, sunny weather with weak winds played a big part in transporting ashfall to entire Japan, and also such conditions, although rare, can be repeated. From here, we derive the assumption that on days with identical weather characteristics as irregular as the day of the Taisho eruption, our simulation results will agree quite well with the transformed ash dispersal map in Fig. 7. Furthermore, due to seasonal differences, we simplify the validation process by validating our simulation results in winter, specifically on days with similar phenomena. Using this assumption, we will get wind data on various atmospheric pressure heights matching the wind data on the eruption. Therefore, the simulation results can produce similar ashfall distribution. In summary, we show the complete validation process in Fig. 9.

5.2 Finding the Identical Weather Characteristics

The validation method, as shown in Fig. 9, consists of two parts:

1. Find the day(s) with weather conditions similar to the eruption day (12 January 1914). The search process will only focus on winter days (December, January, February, and March) from 1 January 1958 to 31 December 2021.
2. Then, evaluate the ashfall distribution by comparing the transformed map (Omori's map) and each simulation result on simulation date(s) found in the first part.

This subsection will focus only on the first part, and the next subsection will explain the rest. The important main features to query are clear, sunny days, weak winds with high atmospheric pressure (anti-cyclonic) and greater temperature, which are quite distinct and rarely occur in Japanese winter. Those features reiterate the peculiarity of the Taisho eruption with its diverging ash distribution. We conduct weather comparisons using available weather information from the chronicles and historical weather information provided by the JMA on their website.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

However, the past meteorological data are limited since the Japanese authority only kept the surface weather condition (temperature, pressure, humidity, precipitation, wind speed) in a major meteorological observatory located in the capital of selected prefectures. These past data consist of daily and hourly observations, and apart from surface observation, there is no other available recorded measurement in different atmospheric pressure levels. Other available information is the surface weather chart on the day of the eruption (12 January 1914, Fig. 10) and entire simulation dates. Therefore, we decided to utilise a surface weather chart instead of surface observation, as it gives a clearer description of a weather phenomenon on a particular. A weather chart, also known as a weather map, usually consists of symbols and features that describe a particular meteorological pattern in a specific space-time dimension. Weather charts are created by plotting the relevant measurements such as mean sea level pressure, wind barbs, and cloud cover. This plotting can help find synoptic scale features such as weather fronts. A meteorologist usually performs an isobaric analysis on these maps, which involves the construction of lines of equal mean sea level pressure (Wallace and Hobbs, 2006). Figure 10 portrays the weather chart at 06 JST on 12 January 1914, with the modern convention of pressure measurements. This weather chart displays how the high-pressure condition covering almost entire Japan centred in the central Honshu. This local maxima in the pressure field, denoted by a big "**H**" letter, can simply explain the common weather characteristics in that particular region. The wind will become stronger when the pressure gradient between high-pressure and low-pressure systems increases. Land frictions weaken the wind coming out from high-pressure systems. Hence, high-pressure systems usually result in weaker winds with clear skies and relatively warm weather (Wallace and Hobbs, 2006). These implications correctly corroborate the weather reports during the eruption day. Accordingly, using similarity on the surface weather chart is appropriate for finding identical weather conditions.

5.2.1 Weather Report Commentary on the Surface Weather Charts

Our initial attempt to search identical weather by using weather chart similarity is by employing an image similarity method to compare features drawn in a pair of weather charts. However, this attempt turned out to be too difficult as the weather chart on the eruption day is very different from the modern weather chart currently used in Japan. The measurement time and scale of barometric pressure, the dimension and projection of the map, and the area included in the map are all different between the current daily weather chart and the weather chart on 12 January 1914. The format of daily weather charts in Japan has changed several times after World War II. Starting from 1883, the JMA produced daily surface weather charts for the Asia Pacific region, consisting of Japan and Japanese-occupied areas pre-World War II. Since August 1958, the JMA then provided daily weather charts in various upper levels (500 hPa, 700 hPa, and 850 hPa) for the Asia Pacific and 500 hPa for the Northern Hemisphere region. Then, starting in March 1996, the JMA added the daily surface weather chart for Northern Hemisphere and 300 hPa level for the Asia Pacific region. Finally, in February 1999, the JMA updated all the weather charts and added a specific surface weather chart for the Japan region. Those changes also came with changes in weather chart format, and the current daily weather charts are following the 1999 format.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Thus, due to the ever-changing formats between past weather charts and the modern ones, doing image similarity is not feasible and is a heavy time-consuming task since we need to have multiple training data for every format to decide whether the two images are similar. We moved to another option by employing other information available in the daily weather chart. Another piece of information available inside the weather chart is the general weather report commentary (*天気概況*). The experts in meteorology developed this commentary as a guide to weather conditions during a particular day. Figure 11a portrays the weather report commentary on 12 January 1914, located in the bottom left of the second page. All historical daily weather charts, from 1883 up to March 1938, have a commentary written on the weather chart explaining the weather condition. However, from April 1938 until December 1995, this feature was omitted. The JMA resumed adding a general weather report commentary feature in the daily weather chart from January 1996 until today, under the weather chart images in the monthly compilation of historical daily weather charts (Database of Weather Charts for Hundred Years, 2022). Figure 12b portrays the first page of the daily weather charts compilation for December 2020 from the JMA for reference (Daily Weather Chart, 2022).

The weather report commentary usually includes the important weather characteristic at that day, such as the existence of high- or low-pressure centres or fronts; the description of extreme weather phenomena such as typhoons, heavy rains, or heavy snow; and other special events on the day, such as disaster, and many more. This description allows us to find similar weather between the dates used in simulation and the day of the eruption, even though further decreasing the number of dates used in the searching process (1996-2021, 3120 days). However, before proceeding to the searching process, we need to do several conversions since there is a minor distinction between the old commentary and the modern ones, which are:

1. The old commentary on the weather chart of the day of the eruption was using the old Japanese syntax.
2. The old commentary addressed both weathers in Japan and weather in a Japanese-occupied area.

We solve those slight differences by simply converting the old commentary to contemporary Japanese syntax and removing all the observation areas outside the modern Japan region. The first column on Table 3 denotes the original form written in the Old Japanese syntax. The second column shows the transformation from the Old Japanese to the contemporary one. The old commentary consists of three parts: explaining the location of high- or low-pressure centres, the weather on main islands of Japan (Honshu, Kyushu, and Hokkaido), and lastly on other parts in the Asia Pacific, coincidentally, at that time all regions mentioned are under Japanese occupation. This commentary agrees with the meteorological report stated in the chronicles (Omori, 1914). Therefore, we omitted the last phrases and focused only on the weather in the main islands of Japan, then further converting it to have a similar form with the commentary in the current daily weather charts. The new phrases in the last column give more brief comments that match with the contemporary style in the modern weather charts while still maintaining the important features (high-pressure all over Japan, clear/sunny in main islands, Kyushu is cloudy, and Hokkaido is either clear or cloudy with strong south-westerly winds). We then compared these phrases in contemporary Japanese with all the commentaries found in the modern weather chart from 1 January 1996 to 31 March 2021, provided by the JMA in winter from January 1996 to March 2021 from the historical archive collected by the National Institute of Informatics (Database of Weather Charts for Hundred Years, 2022).

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

These commentaries explain the weather condition on a particular day in random sequence, i.e., not in similar order to the processed commentary phrases from 12 January 1914. Finally, we utilise the natural language processing (NLP) method to find semantic textual similarity between all commentaries with sentence embedding using SentenceBERT (Reimers and Gurevych, 2019).

5.2.2 Semantic Textual Similarity on Weather Report Commentary using SentenceBERT (SBERT)

In order to find similar weather, both commentaries should have an equivalent meaning that is semantically similar, regardless of sentence structures and words used to form the sentence. The vast improvements of machine learning (ML) and artificial intelligence (AI), especially on NLP tasks, allow us to find the similarity between two sentences using semantic textual similarity (STS) by embedding all sentences first. Word embedding is a process to represent a word in the form of a real-valued vector. This process encodes the meaning of the word such that the words that are closer in the vector space will be similar in meaning. (Jurafsky and James, 2000). A sentence embedding is just the sum of the individual word embedding (Cer et al., 2017). STS scores semantic similarity in varying degrees (e.g., a vehicle and a car are more similar than a wave and a car) instead of binary (e.g., a vehicle and a car are the same, a wave and a car are not the same). STS provides a unified framework that allows for an extrinsic evaluation of multiple semantic components. This framework includes word sense disambiguation and induction, lexical substitution, semantic role labelling, multi-word expression detection and handling, anaphora and co-reference resolution, time and date resolution, named-entity handling, under-specification, hedging, semantic scoping and discourse analysis (Agirre et al., 2012). Here, we can consider a pair of sentences is similar (in this case, the weather condition) if both are located close to each other, as both sentences mapped to a dense vector space (Reimers and Gurevych, 2019, 2020). Google leads the current state-of-the-art NLP development with its technology, Bidirectional Encoder Representations from Transformers (BERT), which can handle various NLP tasks accurately, including question answering, sentence classification, and sentence-pair regression (Devlin et al., 2019).

However, BERT is not optimal to handle several tasks such as large-scale semantic similarity comparison, clustering, and information retrieval via semantic search (Reimers and Gurevych, 2020). To alleviate this issue, Reimers and Gurevych (2019) developed SentenceBERT (SBERT), a modification of the BERT network using Siamese and Triplet networks that can derive semantically meaningful sentence embedding. The Siamese network architecture enables input sentences as fixed-sized vectors. Thus, we use SBERT to find a similar weather condition on 12 January 1914 (query) from all the corpora available (general weather report commentary on daily weather chart 1996-2021). We will not explain in detail the architecture of SBERT and how they can solve various NLP tasks with accurate performances since it is beyond the scope of this paper. For further reference, please visit the official site of SBERT at https://www.sbert.net/index.html. Figure 12 explains all steps to find similar weather conditions through semantic search based on STS between the old commentary and the modern ones. We define the corpora as a collection of a commentary, which may contain one or more sentences explaining the weather condition in a non-uniform sequence. SBERT embeds each sentence on both query and corpus with pre-trained models, specifically to handle the Japanese language.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

The result of the sentence embedding is vectors with 768 values length, specific to each sentence, representing the sentence in the vector space (Reimers and Gurevych, 2019). We employ pre-trained models available for the Japanese language from the Hugging Face model hub (Hugging Face, 2022). However, sentence embedding resources for the Japanese language are limited, and we only found one appropriate pre-trained model exclusively developed for Japanese sentences, *sbert-base-ja* (Abe, 2021). This pre-trained model is from Colorful Scoop, a non-profit entity developing a conversational AI (Colorful Scoop, 2022). This company produced their own BERT Japanese model, trained using the Japanese Wikipedia corpus, released in 2021 under CC-BY-SA 3.0 license. The *sbert-base-ja* pre-trained model is an extension from their original model and trained using Japanese translation of Stanford Natural Language Inference (JSNLI) dataset released by Kyoto University under CC-BY-SA 4.0 license (Yoshikushi et al., 2020), with 535,000 sentence pairs, utilising AdamW optimiser with a $2x10^{-5}$ learning rate (Abe, 2021).

Once we encode both queries and corpora into the vector space, we then conduct a semantic search to find the most similar commentary between query and corpora. The benefit of using semantic search is that we can find the sentences that are synonyms, although constructed using completely different words and order, as the search process focuses on understanding the query's content. Specifically, we did a symmetric semantic search as both query and corpus have similar lengths and amounts of sentences (Reimers and Gurevych, 2019). We use cosine similarity to compute the similarity scores between query and all corpus in the corpora. Cosine similarity is a popular similarity metric used to identify a pair of sentences with an overlapping meaning, as those sentences are likely located next to each other in the vector space (Cer et al., 2017; Reimers and Gurevych, 2019; Arsov et al., 2019).

$$\mathrm{sim}_{\mathrm{v}}(q,c) = \frac{\mathrm{v}(q)\cdot\mathrm{v}(c)}{\|\mathrm{v}(q)\|_2\|\mathrm{v}(c)\|_2} \quad (4)$$

Given a pair of sentence vectors $\mathrm{v}(q)$ and $\mathrm{v}(c)$, cosine similarity $\mathrm{sim}_{\mathrm{v}}$ is computed by the dot product and a magnitude between two vectors, where q and c are components of vector $\mathrm{v}(q)$ and vector $\mathrm{v}(c)$, respectively. Here, q is each sentence in the queries, which is the feature we want to find, and c is each sentence in all commentaries collected from 1996 to 2021. The results of the comparison are between $\{0\}$ for most dissimilar, and $\{1\}$ if a pair of sentences are very similar (Singhal, 2001). The last step is to find similar weather conditions is to rank all the corpus based on its respective similarity score with the query. Ideally, the date(s) with similar weather conditions will have a high similarity score. To keep it brief, we will select only the top two results from all the results obtained from the searching process and describe the weather condition on both dates. Table 4 describes the general weather report commentary of the top two results of the semantic search. The commentary in English is a direct translation using DeepL translation software (DeepL Translator, 2022), with sentences in bold indicating similar weather conditions to the Taisho eruption.

5.2.3 Weather Condition on the Selected Dates

Both dates with the highest STS score agree well with the weather feature mentioned in the chronicles. The weather on selected dates was generally clear, sunny due to the high-pressure covering Japan's main islands.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

However, the commentary on both dates mentions some rain that the chronicles did not mention, proving that we may not observe the exact similar weather completely. This slight difference can slightly affect the ashfall distribution produced by the simulation. Figure 13 depicts the weather chart on those dates, illustrating more apparent features analogous to the weather chart published at the eruption. Both dates contradict each other regarding high-pressure centres but exhibit the same weather characteristics. On the first selected date (5 March 1997), the high-pressure centre moved from the south and then covered entire Japan at night. Meanwhile, on the second date (24 February 2006), the high-pressure centre was in the opposite direction near Hokkaido and advancing eastward. Another difference is the appearance of fronts in the eastern part of the Japanese archipelago coming from the low-pressure centre further in the north. As explained in the commentary, this front brought some rains on the Pacific side of the Kanto region (Tokyo and surrounding areas). Both weather charts presented in Fig. 13 further substantiate the continuous format transition in surface weather analysis in Japan. The general weather report commentary accompanied by establishing a daily weather chart is proven to be critical information that clarifies the weather condition on the respective date.

Another feature not mentioned in the commentary is the wind condition during the selected dates. The chronicles briefly revealed that the wind was weak, making it possible to evacuate using boats from the island to escape the eruption (Omori, 1914; Kitagawa, 2015). In order to further analyse the weather similarities between the selected dates and 12 January 1914, we draw the condition of the wind using the wind rose graph as depicted in Fig. 14. The graph consists of winds observation for every 6 hours from the start of the simulation (00UTC on the selected date) until the simulation process finished (00UTC four days after the eruption start). We choose the winds observation both at the 1000 hPa pressure level (equivalent to the surface pressure) and 250 hPa pressure level (equivalent to the location of the westerlies) based on the findings explained in Sect. 2. The surface winds characteristics differ in terms of directions but have a comparable velocity between the two selected dates. The surface winds on the eruption are more akin to March 1997 than February 2006, judging from the dominant south-westerly winds portrayed in Fig. 14a. In contrast, the surface winds on February 2006 mostly comprise stronger north-easterly winds that conflict with the finding in chronicles despite having identical weather features. Meanwhile, there is a little difference in higher altitude as the upper winds at March 1997 had a stronger velocity but mostly came from the west, compared to the latter having slower north-eastward direction. This disparity in wind characteristics between the two dates considered to have very similar weather will certainly affect how the ash distribution formed on the simulation.

5.2 Simulation Results Performance on Dates with Identical Weather Characteristics

Once we found the date(s) with similar weather conditions from the previous section, we evaluated the dataset performance using the wind data on the selected date(s). Here, due to the unavailability of the concrete ground measurement, we use a binary contingency table between the ash dispersal map from the simulation results and the original one (Figure 6). This method is appropriate since we intend to measure how the ash dispersal map produced from the simulation on selected dates will equally cover an equivalent region of interest to the original ash dispersal map.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Therefore, the main aim of this validation method is to assess the performance of the selected simulation result enclosing all distal regions covered by ashfall deposit in the Taisho eruption. Table 5 briefly explains the contingency table used to evaluate the simulation results. This table denotes two dimensions of affected and unaffected grid points. Then, from matching the binary contingency table, we can quantify the hit rate (sensitivity) of the simulation result as follows:

$$Hit = \frac{TP}{TP+FN} \quad (5)$$

TP denote the true positive, which means the same grid points affected in both the original ash dispersal map and the simulation results, and vice versa. Meanwhile, *FN* denotes the false negative, which particular grid points marked as unaffected in the simulation results but appear to be affected in the original ash dispersal map. Therefore, hit rate score *Hit* is ashes from the simulation result covering the same grid points with the original ash dispersal map divided by all affected grid points from the original ash dispersal map. Here, we use hit rate score *Hit* as we only involved the region of interest in the original ash dispersal map (red-striped area on Fig. 6) due to bias presented in the wind data. It should be noted that the simulation result may, in reality, depicts worse impacts on the map.

Nevertheless, we argue that it is still a valid result as long as the ashfall deposit reflected from the simulation covers most of the same regions as the original ash dispersal map. Therefore, using the hit rate is justifiable in this particular case. Table 6 provides the satisfactory performance of each selected simulation result reaching a high *Hit* score. This result means that both selected dates with similar weather conditions to the Taisho eruption have a matching ash dispersal pattern. Figure 15 further portrays the evidence of the similarities in the ash dispersal map for both simulation results. The first selected date (1997/03/05) with a higher similarity score has less *Hit* than the second selected date (1996/02/24). However, the maps presented in Fig. 15 resemble more similarities. The second selected date produced a more erratic and scattered distribution, implying far worse impacts than the Taisho eruption.

For reference, we also portray the ash dispersal map from the supplementary data focusing on Kagoshima covering the southern Kyushu region of each respective date on Fig 16. The sequential colour gradient indicates differences in the thickness of the ashfall deposit in m. The isolines of complete ground measurement (Mita et al., 2018) overlayed in the map. Both maps produce a more erratic ash dispersal pattern than the Taisho eruption, with some regions having a thicker ashfall deposit in the ground. The proximal ashfall deposit of the simulation on 24 February 2006 more resembling the original than its counterpart. Though the simulation produces a somewhat agreeable deposit distribution with the report, it should be noted that having a matching ashfall distribution in distal locations does not necessarily mean having a matching ash distribution at a proximal location. Section 2 mentioned that different winds govern the ashfall distribution from Sakurajima volcano on the Taisho eruption. In summary, we successfully demonstrate the quality of the dataset under limited validation data with a novel method incorporating a machine learning technique.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

575 **5.4 Dataset Limitations**

Generally, the ash deposition results from the simulation of volcanic ash dispersal and transport models rely on how the parameters value of both eruption source parameters and the wind condition during the eruption are set. These parameters are the most critical parts determining how ashfall impacts proximal and distal locations (Bonadonna et al., 2012; Folch, 2012; Macedonio et al., 2016; Mastin et al., 2009; Webley and Mastin, 2009). This study uses the deterministic value of
580 plume height, erupted mass and mass eruption rate as the eruption source parameters. We utilised the calculation by Iguchi (2014) incorporating the study by Morton et al. (1956) for the relationship between observed plume height in the historical reports and studies (Kobayashi, 1982; Koto, 1916; Omori, 1914; Todde et al., 2017) to produce mass eruption rate. There is a positive correlation between the change of plume height over time to both total eruption mass and mass eruption rate. However, a researcher must be careful when using this method as it introduces a significant bias in the mean height value.
585 The eruption rate is roughly proportional to the fourth power of height; error in assigning the precise height will heavily affect its conversion to eruption rate (Folch et al., 2012; Mastin et al., 2009). We assign the parameters of last eruptions as a baseline of the largest eruption size or plume height as a conservative safety factor, though we recognise wide variability that could change the significance of the implied hazard. The values we used in this study agreed well with the past observations, and the bias presented in the eruption source parameter should be small and would not heavily affect the ash dispersal
590 mechanism. However, variations on those parameters could result in different eruption rates and total mass, resulting in different ashfall footprints.

In addition, future vent location, eruption style and size, and eruption duration significantly affect the impacts of the ashfall for both proximal and distal areas (Bonadonna et al., 2012; Mastin et al., 2009; Selva et al., 2018). Furthermore, depending on the models, the parameters which need to be set arbitrarily, such as the diffusion coefficient, settling particle velocity law
595 (see Table 1), orographic effects, and the lack of topographical data, could cause discrepancy if different variations were used (Macedonio et al., 2016; Poulidis et al., 2017, 2018; Scollo et al., 2011). Finally, the distribution of total grain size, the particle density and its shape, and the chosen aggregation method also affect the accuracy of the model outputs (Bonadonna et al., 2012; Folch et al., 2010, 2012; Mastin et al., 2009; Selva et al., 2018). Here, we approach the simulation with the main purpose to characterise the Taisho eruption, which eliminates the variations on those variables since we use the fixed
600 scenario as a worst-case. We acknowledged the limitations of this study that used the fixed values for the eruption source parameters to precisely replicate the transition in the Taisho eruption, which strongly restrict the various possibility of future events. Following the result mentioned earlier by Hickey et al. (2016) that illustrates how the current uplift situation is similar to the preceding eruption event, our results could provide valuable input to both authorities and stakeholders. The dataset presented here will be a guideline to provide an insightful understanding of how such a large explosive eruption
605 would affect contemporary Japan. Moreover, inaccuracies can occur from the effect of the winds and atmospheric humidity, among other things (Folch et al., 2012; Mastin et al., 2009). The meteorological conditions heavily determine the ashfall dispersion trajectory and ashfall deposit location (Macedonio et al., 2016).

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Therefore, reanalysis datasets are favourable to use since they have better accuracy than the forecast ones (Folch et al., 2012). It should be noted that the variability in the different meteorological databases seems relatively insignificant (Selva et al., 2018). As intended to provide guidance for a worst-case scenario, the dataset provided can remain helpful to extend the present study toward the ashfall risk analysis and its decision-making process. We argue that the dataset here is vital to support future researchers and emergency managers in analysing and planning crisis responses over different conditions. It is important to note that these assignments are rough estimates of one of the possibilities about the most likely future eruption size and type at Sakurajima volcano. We understood that there might be inconsistency in the dataset that likely introduces a bias in the ashfall hazard and risk analyses. Future users should consider all these sources of uncertainty.

6 Conclusion

The ash dispersal products from large eruptions brought devastating impacts to various sectors both in a nearby areas from the volcano and distal locations. The emergency managers and crisis response planners need to acknowledge the compounded effect of ashfall deposits on the infrastructures to manage the loss appropriately. This paper presents the dataset useful for more extensive research and planning, focusing on the context of the entire Japan. The paper describes the complete generation process of a dataset based on the recent large eruption of Sakurajima volcano (the Taisho eruption) by replicating the eruption process over an extended period from 1958 to 2021. Furthermore, we conduct the validation procedure by identifying the similarity in the weather pattern at the time of the eruption and proving that similar weather conditions will bring identical ashfall distribution, thus verifying the dataset. Finally, we acknowledged the limitation and uncertainties the dataset could bring. Although it contains some degree of inaccuracies, the dataset can still encourage further studies in the ashfall risk analysis and decision-making process as limited as it is. Such advancement will be crucial to support future researchers and emergency managers in devising disaster responses over different conditions.

Author Contributions

HR did the most work to produce the dataset by doing the data processing and simulations, analysing the results, and writing this manuscript. In addition, HT conceptualised the project and supervised both the simulations and the result analysis. MI produced the Taisho eruption data used as eruption source parameters used in the model to generate the dataset. HLT is the creator of the PUFF model, provided the original FORTRAN codes of the model and the default value of the parameters used in the simulation. TT verified the weather similarity process incorporating the general weather report commentary inside the weather chart. SR assisted in developing the semantic search methodology to find the semantic similarity through the commentary.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgements

We acknowledge the works of Sakurajima Volcano Research Center (SVRC), Disaster Prevention Research Institute, Kyoto University in Kagoshima, Japan, that provided the Taisho eruption data used in this work. In addition, we thank the JMA for supplying the JRA-55 Reanalysis data from 1958 to 2020. The JRA-55 data were obtained from the archive of the Center for Computational Sciences (CCS), University of Tsukuba, Ibaraki, Japan. Furthermore, the authors were grateful for the help from Si Ha, Ryo Hirako, and Toraemon Matsumoto, who assisted in the map digitalising process of the original ash dispersal map and the translation process for the old weather charts.

References

Abe, N.: Sentence BERT base Japanese model, Python, Colorful Scoop, 2021.

Agirre, E., Cer, D., Diab, M., and Gonzalez-Agirre, A.: SemEval-2012 Task 6: A Pilot on Semantic Textual Similarity, in: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012), SEM 2012: The First Joint Conference on Lexical and Computational Semantics, Montréal, Canada, 385–393, 2012.

Arsov, N., Dukovski, M., Evkoski, B., and Cvetkovski, S.: A Measure of Similarity in Textual Data Using Spearman's Rank Correlation Coefficient, arXiv [eprint], arXiv:1911.11750, 2019.

Ayris, P. M. and Delmelle, P.: The immediate environmental effects of tephra emission, B. Volcanol., 74, 1905–1936, https://doi.org/10.1007/s00445-012-0654-5, 2012.

Barsotti, S., Andronico, D., Neri, A., Del Carlo, P., Baxter, P. J., Aspinall, W. P., and Hincks, T.: Quantitative assessment of volcanic ash hazards for health and infrastructure at Mt. Etna (Italy) by numerical simulation, J. Volcanol. Geoth. Res., 192, 85–96, https://doi.org/10.1016/j.jvolgeores.2010.02.011, 2010.

Barsotti, S., Di Rienzo, D. I., Thordarson, T., Björnsson, B. B., and Karlsdóttir, S.: Assessing Impact to Infrastructures Due to Tephra Fallout From Öræfajökull Volcano (Iceland) by Using a Scenario-Based Approach and a Numerical Model, Front. Earth Sci., 6, 196, https://doi.org/10.3389/feart.2018.00196, 2018.

Biass, S., Todde, A., Cioni, R., Pistolesi, M., Geshi, N., and Bonadonna, C.: Potential impacts of tephra fallout from a large-scale explosive eruption at Sakurajima volcano, Japan, , B. Volcanol., 79, https://doi.org/10.1007/s00445-017-1153-5, 2017.

Bonadonna, C.: Probabilistic modelling of tephra dispersion, in: Statistics in Volcanology, Special Publications of IAVCEI, 1, edited by: Mader, H. M., Coles, S. G., Connor, C. B., and Connor, L. J. (eds), The Geological Society of London on behalf of The International Association of Volcanology and Chemistry of the Earth's Interior, 243–259, https://doi.org/10.1144/IAVCEI001.19, 2006.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Bonadonna, C., Ernst, G. G. J., and Sparks, R. S. J.: Thickness variations and volume estimates of tephra fall deposits: the importance of particle Reynolds number, J. Volcanol. Geoth. Res., 81, 173–187, https://doi.org/10.1016/S0377-0273(98)00007-9, 1998.

Bonadonna, C., Folch, A., Loughlin, S., and Puempel, H.: Future developments in modelling and monitoring of volcanic ash clouds: outcomes from the first IAVCEI-WMO workshop on Ash Dispersal Forecast and Civil Aviation, B. Volcanol., 74, 1–10, https://doi.org/10.1007/s00445-011-0508-6, 2012.

Bonadonna, C., Biass, S., Menoni, S., and Gregg, C. E.: Assessment of risk associated with tephra-related hazards, in: Forecasting and Planning for Volcanic Hazards, Risks, and Disasters, Volume 2, edited by: Schroeder, J. F. (series editor), Papale, P. (volume editor), Elsevier, 329–378, https://doi.org/10.1016/B978-0-12-818082-2.00008-1, 2021.

Bursik, M., Jones, M., Carn, S., Dean, K., Patra, A., Pavolonis, M., Pitman, E. B., Singh, T., Singla, P., Webley, P., Bjornsson, H., and Ripepe, M.: Estimation and propagation of volcanic source parameter uncertainty in an ash transport and dispersal model: application to the Eyjafjallajokull plume of 14–16 April 2010, B. Volcanol., 74, 2321–2338, https://doi.org/10.1007/s00445-012-0665-2, 2012.

Cer, D., Diab, M., Agirre, E., Lopez-Gazpio, I., and Specia, L.: SemEval-2017 Task 1: Semantic Textual Similarity - Multilingual and Cross-lingual Focused Evaluation, in: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 11th International Workshop on Semantic Evaluation (SemEval-2017), Vancouver, Canada, 1–14, https://doi.org/10.18653/v1/S17-2001, 2017.

Colorful Scoop: https://colorfulscoop.com/, last access: 27 January 2022.

Daily Weather Chart: https://www.data.jma.go.jp/fcd/yoho/hibiten/, last access: 3 January 2022.

Damby, D. E., Horwell, C. J., Baxter, P. J., Delmelle, P., Donaldson, K., Dunster, C., Fubini, B., Murphy, F. A., Nattrass, C., Sweeney, S., Tetley, T. D., and Tomatis, M.: The respiratory health hazard of tephra from the 2010 Centennial eruption of Merapi with implications for occupational mining of deposits, J. Volcanol. Geoth. Res., 261, 376–387, https://doi.org/10.1016/j.jvolgeores.2012.09.001, 2013.

Database of Weather Charts for Hundred Years: http://agora.ex.nii.ac.jp/digital-typhoon/weather-chart/, last access: 3 January 2022.

ESRI: ArcGIS Pro 2.7, 2017.

DeepL Translator: https://www.deepl.com/translator/, last access: 27 January 2022.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv [eprint], arXiv:1810.04805, 2019.

Eliasson, J., Yoshitani, J., Weber, K., Yasuda, N., Iguchi, M., and Vogel, A.: Airborne Measurement in the Ash Plume from Mount Sakurajima: Analysis of Gravitational Effects on Dispersion and Fallout, International Journal of Atmospheric Sciences, 1–16, https://doi.org/10.1155/2014/372135, 2014.

Fero, J., Carey, S. N., and Merrill, J. T.: Simulation of the 1980 eruption of Mount St. Helens using the ash-tracking model PUFF, J. Volcanol. Geoth. Res., 175, 355–366, https://doi.org/10.1016/j.jvolgeores.2008.03.029, 2008.

© Author(s) 2022. CC BY 4.0 License.

Fero, J., Carey, S. N., and Merrill, J. T.: Simulating the dispersal of tephra from the 1991 Pinatubo eruption: Implications for the formation of widespread ash layers, J. Volcanol. Geoth. Res., 186, 120–131, https://doi.org/10.1016/j.jvolgeores.2009.03.011, 2009.

Folch, A.: A review of tephra transport and dispersal models: Evolution, current status, and future perspectives, J. Volcanol. Geoth. Res., 235–236, 96–115, https://doi.org/10.1016/j.jvolgeores.2012.05.020, 2012.

Folch, A., Costa, A., Durant, A., and Macedonio, G.: A model for wet aggregation of ash particles in volcanic plumes and clouds: 2. Model application, J. Geophys. Res., 115, B09202, https://doi.org/10.1029/2009JB007176, 2010.

Folch, A., Costa, A., and Basart, S.: Validation of the FALL3D ash dispersion model using observations of the 2010 Eyjafjallajökull volcanic ash clouds, Atmos. Environ., 48, 165–183, https://doi.org/10.1016/j.atmosenv.2011.06.072, 2012.

Hampton, S. J., Cole, J. W., Wilson, G., Wilson, T. M., and Broom, S.: Volcanic ashfall accumulation and loading on gutters and pitched roofs from laboratory empirical experiments: Implications for risk assessment, J. Volcanol. Geoth. Res., 304, 237–252, https://doi.org/10.1016/j.jvolgeores.2015.08.012, 2015.

Hattori Y., Suto H., Toshida K., and Hirakuchi H.: Development of estimation method for tephra transport and dispersal characteristics with numerical simulation technique (part1) - meteorological effects on ash fallout with Shinmoe-dake eruption (in Japanese), Central Research Institute of Electric Power Industry (CRIEPI), Japan, 2013.

Hattori Y., Suto H., Toshida K., and Hirakuchi H.: Development of estimation method for tephra transport and dispersal characteristics with numerical simulation technique (part 2) - A method of selecting meteorological conditions and the effects on ash deposition and concentration in air for Kanto-area (in Japanese), Central Research Institute of Electric Power Industry (CRIEPI), Japan, 2016.

Hickey, J., Gottsmann, J., Nakamichi, H., and Iguchi, M.: Thermomechanical controls on magma supply and volcanic deformation: application to Aira caldera, Japan, Sci. Rep., 6, 32691, https://doi.org/10.1038/srep32691, 2016.

Hoyer, S. and Hamman, J. J.: xarray: N-D labeled Arrays and Datasets in Python, Journal of Open Research Software, 5, https://doi.org/10.5334/jors.148, 2017.

Hugging Face: https://huggingface.co/, last access: 27 January 2022.

Iguchi, M.: Sakurajima Taisho Eruption, 100th Anniversary of the Disaster Prevention in Sakurajima, Kyoto University, Kyoto, Japan, 2014.

Iguchi, M.: Method for Real-Time Evaluation of Discharge Rate of Volcanic Ash – Case Study on Intermittent Eruptions at the Sakurajima Volcano, Japan, Journal of Disaster Research, 11, 4–14, https://doi.org/10.20965/jdr.2016.p0004, 2016.

Iguchi, M., Nakamichi, H., and Tameguri, T.: Integrated Study on Forecasting Volcanic Hazards of Sakurajima Volcano, Japan, Journal of Disaster Research, 15, 174–186, https://doi.org/10.20965/jdr.2020.p0174, 2020.

Jenkins, S. F., Spence, R. J. S., Fonseca, J. F. B. D., Solidum, R. U., and Wilson, T. M.: Volcanic risk assessment: Quantifying physical vulnerability in the built environment, J. Volcanol. Geoth. Res., 276, 105–120, https://doi.org/10.1016/j.jvolgeores.2014.03.002, 2014.

Jenkins, S. F., Wilson, T. M., Magill, C., Miller, V., Stewart, C., Blong, R., Marzocchi, W., Boulton, M., Bonadonna, C., and Costa, A.: Volcanic ash fall hazard and risk, in: Global Volcanic Hazards and Risk, edited by: Loughlin, S. C., Sparks, S.,

© Author(s) 2022. CC BY 4.0 License.

Brown, S. K., Jenkins, S. F., and Vye-Brown, C., Cambridge University Press, Cambridge, 173–222, https://doi.org/10.1017/CBO9781316276273.005, 2015.

Jenkins, S. F., Magill, C. R., and Blong, R. J.: Evaluating relative tephra fall hazard and risk in the Asia-Pacific region, Geosphere, 14, 492–509, https://doi.org/10.1130/GES01549.1, 2018.

Jurafsky, D. and James, M. H.: Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition, Prentice Hall, Upper Saddle River, New Jersey, 2000.

Kagoshima City, Crisis Management Division: Sakurajima Volcano Hazard Map, Kagoshima Municipal Government, Kagoshima, Japan, 2010.

Kitagawa, K.: Living with an Active Volcano: Informal and Community Learning for Preparedness in South of Japan, in: Observing the Volcano World, edited by: Fearnley, C. J., Bird, D. K., Haynes, K., McGuire, W. J., and Jolly, G., Springer International Publishing, Cham, 677–689, https://doi.org/10.1007/11157_2015_12, 2015.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, Journal of the Meteorological Society of Japan, 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.

Kobayashi, T.: Geology of Sakurajima Volcano: A Review, Bulletin of Volcanology Society of Japan, 27, 277–292, 1982.

Koto, B.: The Great Eruption of Sakura-Jima, Journal of the College of Science, Imperial University of Tokyo, 38, 1–237, 1916.

Kratzmann, D. J., Carey, S. N., Fero, J., Scasso, R. A., and Naranjo, J.-A.: Simulations of tephra dispersal from the 1991 explosive eruptions of Hudson volcano, Chile, 190, 337–352, https://doi.org/10.1016/j.jvolgeores.2009.11.021, 2010.

Kyushu Regional Development Bureau, Osumi Office of the Rivers and National Highways: Sakurajima Volcano Wide Area Disaster Prevention Map (in Japanese), Ministry of Land, Infrastructure, Transport, and Tourism, Sakurajima, Kagoshima, Japan, 2017.

Lin, J., Brunner, D., Gerbig, C., Stohl, A., Luhar, A., and Webley, P.: Lagrangian Modeling of the Atmosphere: An Introduction, in: Lagrangian Modeling of the Atmosphere, American Geophysical Union, Washington, DC, USA, 1–15, 2012.

Macedonio, G., Costa, A., Scollo, S., and Neri, A.: Effects of eruption source parameter variation and meteorological dataset on tephra fallout hazard assessment: example from Vesuvius (Italy), Journal of Applied Volcanology, 5, 5, https://doi.org/10.1186/s13617-016-0045-2, 2016.

Madankan, R., Pouget, S., Singla, P., Bursik, M., Dehn, J., Jones, M., Patra, A., Pavolonis, M., Pitman, E. B., Singh, T., and Webley, P.: Computation of probabilistic hazard maps and source parameter estimation for volcanic ash transport and dispersion, J. Comput. Phys., 271, 39–59, https://doi.org/10.1016/j.jcp.2013.11.032, 2014.

Maeda, S., Hirahara, S., Hagiya, S., Murai, H., and Oikawa, Y.: Forecasting Changes in Seasonal Progress in Future Climate (based on the changes of Westerly Wind) (in Japanese), Climate Information Division, Global Environment and Ocean Department, Japan Meteorological Agency, 05 March 2012, http://wind.gp.tohoku.ac.jp/yamase/reports/meeting5.html, 2012.

© Author(s) 2022. CC BY 4.0 License.

Mastin, L. G., Guffanti, M., Servranckx, R., Webley, P., Barsotti, S., Dean, K., Durant, A., Ewert, J. W., Neri, A., Rose, W. I., Schneider, D., Siebert, L., Stunder, B., Swanson, G., Tupper, A., Volentik, A., and Waythomas, C. F.: A multidisciplinary effort to assign realistic source parameters to models of volcanic ash-cloud transport and dispersion during eruptions, J. Volcanol. Geoth. Res., 186, 10–21, https://doi.org/10.1016/j.jvolgeores.2009.01.008, 2009.

Mita, K., Ishimine, Y., Iwamatsu, A., Kamikozuru, H., Koga, M., Takeda, N., Nakamura, T., Matsumoto, S., Iguchi, M., Baba, M., Yamaguchi, K., Shimokawa, E., and Maki, M.: Tentative Reports on the Traffic Countermeasures for a Large-Scale Eruption of Sakurajima Volcano. Case Study: Ash Deposit on Kagoshima City (in Japanese), Earthquake and Volcanic Disaster Prevention Center, Kagoshima University, 2018.

Miyaji, N., Kan'no, A., Kanamaru, T., and Mannen, K.: High-resolution reconstruction of the Hoei eruption (AD 1707) of Fuji volcano, Japan, J. Volcanol. Geoth. Res., 207, 113–129, https://doi.org/10.1016/j.jvolgeores.2011.06.013, 2011.

Morton, B. R., Taylor, S. G., and Turner, J. S.: Turbulent gravitational convection from maintained and instantaneous sources, Proc. R. Soc. Lond. A Mat., 234, 1–23, https://doi.org/10.1098/rspa.1956.0011, 1956.

Newhall, C. G. and Self, S.: The volcanic explosivity index (VEI) an estimate of explosive magnitude for historical volcanism, J. Geophys. Res., 87, 1231, https://doi.org/10.1029/JC087iC02p01231, 1982.

Nishimura, T.: Source mechanisms of volcanic explosion earthquakes: single force and implosive sources, J. Volcanol. Geoth. Res., 86, 97–106, https://doi.org/10.1016/S0377-0273(98)00088-2, 1998.

Omori, F.: The Sakura-Jima Eruption and Earthquakes, Bulletin of the Imperial Earthquake Investigation Committee, The University of Tokyo, 8, 1–630, 1914.

Peng, J. and Peterson, R.: Attracting structures in volcanic ash transport, Atmos. Environ., 48, 230–239, https://doi.org/10.1016/j.atmosenv.2011.05.053, 2012.

Peterson, R. and Dean, K.: Analysis and Optimization of a Lagrangian Volcanic Ash Particle Tracking Model called Puff, AGU Fall Meeting Abstracts, 2002.

Poulidis, A. P. and Takemi, T.: A 1998-2013 climatology of Kyushu, Japan: seasonal variations of stability and rainfall, Int. J. Climatol., 37, 1843–1858, https://doi.org/10.1002/joc.4817, 2017.

Poulidis, A. P., Takemi, T., Iguchi, M., and Renfrew, I. A.: Orographic effects on the transport and deposition of volcanic ash: A case study of Mount Sakurajima, Japan, J. Geophys. Res-Atmos., 122, 9332–9350, https://doi.org/10.1002/2017JD026595, 2017.

Poulidis, A. P., Takemi, T., Shimizu, A., Iguchi, M., and Jenkins, S. F.: Statistical analysis of dispersal and deposition patterns of volcanic emissions from Mt. Sakurajima, Japan, Atmos. Environ., 179, 305–320, https://doi.org/10.1016/j.atmosenv.2018.02.021, 2018.

Rahadianto, H. and Tatano, H.: 62 Years Simulated Sakurajima Taisho Eruption Ashfall Deposit Data (1958-2019), Disaster Prevention Research Institute, Kyoto University, Japan [dataset], https://doi.org/10.17603/ds2-vw5f-t920., 2020.

Rathje, E. M., Dawson, C., Padgett, J. E., Pinelli, J.-P., Stanzione, D., Adair, A., Arduino, P., Brandenberg, S. J., Cockerill, T., Dey, C., Esteva, M., Haan, F. L., Hanlon, M., Kareem, A., Lowes, L., Mock, S., and Mosqueda, G.: DesignSafe: New Cyberinfrastructure for Natural Hazards Engineering, Nat. Hazards Rev., 18, 06017001, https://doi.org/10.1061/(ASCE)NH.1527-6996.0000246, 2017.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Reimers, N. and Gurevych, I.: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, arXiv [eprint], arXiv:1908.10084, 2019.

Reimers, N. and Gurevych, I.: Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation, arXiv [eprint], arXiv:2004.09813, 2020.

Scollo, S., Prestifilippo, M., Coltelli, M., Peterson, R. A., and Spata, G.: A statistical approach to evaluate the tephra deposit and ash concentration from PUFF model forecasts, J. Volcanol. Geoth. Res., 200, 129–142, https://doi.org/10.1016/j.jvolgeores.2010.12.004, 2011.

Searcy, C., Dean, K., and Stringer, W.: PUFF: A High-Resolution Volcanic Ash Tracking Model, J. Volcanol. Geoth. Res., 80, 1–16, https://doi.org/10.1016/S0377-0273(97)00037-1, 1998.

Selva, J., Costa, A., De Natale, G., Di Vito, M. A., Isaia, R., and Macedonio, G.: Sensitivity test and ensemble hazard assessment for tephra fallout at Campi Flegrei, Italy, J. Volcanol. Geoth. Res., 351, 1–28, https://doi.org/10.1016/j.jvolgeores.2017.11.024, 2018.

Shimbori T., Aikawa Y., Fukui K., Hashimoto A., Seino N., and Yamasato H.: Quantitative Tephra Fall Prediction with the JMA Mesoscale Tracer Transport Model for Volcanic Ash: A Case Study of the Eruption at Asama Volcano in 2009 (in Japanese), Pap. Meteorol. Geophys., 61, 13–29, https://doi.org/10.2467/mripapers.61.13, 2009.

Singhal, A.: Modern Information Retrieval: A Brief Overview, 9, 2001.

Takebayashi, M., Onishi, M., and Iguchi, M.: Large volcanic eruptions and their influence on air transport: The case of Japan, J. Air Transp. Manag., 97, 102136, https://doi.org/10.1016/j.jairtraman.2021.102136, 2021.

Tanaka, H. L.: Development of A Prediction Scheme for Volcanic Ash Fall from Redoubt Volcano, Alaska, in: Proceedings of the First International Symposium on Volcanic Ash and Aviation Safety, First International Symposium on Volcanic Ash and Aviation Safety, 283–291, https://doi.org/10.3133/b2047, 1994.

Tanaka, H. L. and Iguchi, M.: Numerical Simulations of Volcanic Ash Plume Dispersal for Sakura-Jima Using Real-Time Emission Rate Estimation, Journal of Disaster Research, 14, 160–172, https://doi.org/10.20965/jdr.2019.p0160, 2019.

Tanaka, H. L. and Yamamoto, K.: Numerical simulation of volcanic plume dispersal from Usu volcano in Japan on 31 March 2000 using PUFF model, Earth, Planets and Space, 54, 743–752, https://doi.org/10.1186/BF03351727, 2002.

Tanaka, H. L., Iguchi, M., and Nakada, S.: Numerical Simulations of Volcanic Ash Plume Dispersal from Kelud Volcano in Indonesia on February 13, 2014, Journal of Disaster Research, 11, 31–42, https://doi.org/10.20965/jdr.2016.p0031, 2016.

Tilling, R. I., Topinka, L. J., Swanson, D. A., and Survey (U.S.), G.: Eruptions of Mount St. Helens: Past, Present, and Future, U.S. Department of the Interior, Geological Survey, 1990.

Todde, A., Cioni, R., Pistolesi, M., Geshi, N., and Bonadonna, C.: The 1914 Taisho eruption of Sakurajima volcano: stratigraphy and dynamics of the largest explosive event in Japan during the twentieth century, B. Volcanol., 79, 72, https://doi.org/10.1007/s00445-017-1154-4, 2017.

Tsukui M.: Ash-Fall Distribution of 1779 An'ei Eruption, Sakurajima Volcano : Revealed by Historical Documents (in Japanese), Bulletin of The Volcanological Society of Japan, 56, 89–94, https://doi.org/10.18940/kazan.56.2-3_89, 2011.

Unidata: NetCDF, UCAR/Unidata, https://doi.org/10.5065/D6H70CW6, 2021.

© Author(s) 2022. CC BY 4.0 License.

Wallace, J. M. and Hobbs, P. V.: Atmospheric science: an introductory survey, 2nd ed., Academic press, Amsterdam Paris, 2006.

Webley, P. and Mastin, L.: Improved prediction and tracking of volcanic ash clouds, J. Volcanol. Geoth. Res., 186, 1–9, https://doi.org/10.1016/j.jvolgeores.2008.10.022, 2009.

Wilson, T. M., Stewart, C., Sword-Daniels, V., Leonard, G. S., Johnston, D. M., Cole, J. W., Wardman, J., Wilson, G., and Barnard, S. T.: Volcanic ash impacts on critical infrastructure, Phys. Chem. Earth, Parts A/B/C, 45–46, 5–23, https://doi.org/10.1016/j.pce.2011.06.006, 2012.

Yamasato, H., Funasaki, J., and Takagi, Y.: The Japan Meteorological Agency's Volcanic Disaster Mitigation Initiatives, National Research Institute for Earth Science and Disaster Prevention, Japan, 2013.

Yasui, M., Takahashi, M., Ishihara, K., and Miki, D.: Records on the 1914-1915 eruption of Sakurajima volcano, Japan (in Japanese), Bulletin of The Volcanological Society of Japan, 41, 75–107, 2006.

Yasui M., Takahashi M., Ishihara K., and Miki D.: Eruptive Style and its Temporal Variation through the 1914-1915 Eruption of Sakurajima Volcano, Southern Kyushu, Japan (in Japanese), Bulletin of The Volcanological Society of Japan, 52, 161–186, https://doi.org/10.18940/kazan.52.3_161, 2007.

Yoshikushi, T., Kawahara, D., and Kurohara, S.: Multilingualization of a Natural Language Inference Dataset Using Machine Translation, Kyoto University, 2020.

Zuccaro, G., Leone, M. F., Del Cogliano, D., and Sgroi, A.: Economic impact of explosive volcanic eruptions: A simulation-based assessment model applied to Campania region volcanoes, J. Volcanol. Geoth. Res., 266, 1–15, https://doi.org/10.1016/j.jvolgeores.2013.09.002, 2013.

Figures Caption

Figure 1: The arrival time of ashfall during the Taisho eruption event throughout Japan. The Blue dashed line indicates the chronology of ash sighting in each respective location. Inset, the graph of ash observation time from the eruption source (Todde et al., 2017).

Figure 2: The ash dispersal map of the Taisho eruption on entire Japan (Omori, 1914).

Figure 3: Plume height progression during the Taisho eruption on 12-13 January 1914 with information on both subsequent events and materials ejected at particular times (Takebayashi et al., 2021).

Figure 4: The ten wind roses illustrate the wind conditions inferred from the JRA-55 Reanalysis dataset (Kobayashi et al., 2015) for both surface winds (first row) and upper altitude winds at approximately 10 km asl (second row). The first column portrays all wind conditions for the entire 1958-2020, while the rest shows the influence of seasonality. Roses correspond to the probability of the provenance and sequential colours indicating the speed gradients.

Figure 5: The structures of the sample deposit data for both regions in the NetCDF format.

Figure 6: The structures of the sample airborne data in the NetCDF format.

Figure 7: Transformed ash dispersal map (Figure 2) from ArcGIS software with airports icon as impacts reference.

Figure 8: Seasonal progression of the westerlies over Japan at 12 km from the surface (Maeda et al., 2012). The red dashed square indicates the region covering Sakurajima volcano and Kagoshima city. The vertical axis shows the latitude degree of the wind

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

position. The horizontal axis shows monthly changes of the westerlies velocity for the left image and the longitude of the winds on the right image. The straight blue lines indicate all regions in Japan affected by the westerlies. The small number inside the left image is the wind contour.

Figure 9: The flowchart of the validation method to judge the performance of the dataset produced in this paper.

Figure 10: The weather chart on the Taisho eruption with additional markings of pressure measurement converted to the latest convention (mmHg → hPa). Modified from Omori (1914).

Figure 11: The general weather report commentary on the weather condition inside the weather chart. (a) The old weather chart put the commentary on the bottom of the second page on the day of the Taisho eruption (Database of Weather Charts for Hundred Years, 2022). (b) The modern weather chart mentions the commentary on the bottom of the monthly compilation of daily weather charts (Japan Meteorological Agency, 2022).

Figure 12: The flowchart for finding the date with a similar weather report commentary to the day of eruption using SBERT.

Figure 13: (a) Surface weather chart for Asia Pacific region at 00 UTC 5 March 1997, and (b) Surface weather chart for Asia Pacific region at 00 UTC 24 February 2006. Both weather chart images were obtained from the JMA (Database of Weather Charts for Hundred Years, 2022; Daily Weather Chart, 2022).

Figure 14: The wind rose diagrams show the six-hourly wind direction corresponding to which from the winds blow at (a) 1000 hPa altitude (~100 m asl.) on 5 March to 10 March 1995, (b) 1000 hPa (~100 m asl.) altitude on 24 February to 1 March 2006; (c) 250 hPa altitude (~10 km asl.) on 5 March to 10 March 1995, (d) 250 hPa altitude (~10 km asl.) on 24 February to 1 March 2006.

Figure 15: The comparison of the ash distribution coverage between the original ash dispersal map and (a) the simulation results on 5 March 1997, and (b) the simulation results on 24 February 2006. The grey-stripped area indicates the original areas affected by the Taisho eruption, and the dark grey colour represents the ash dispersal from the simulation.

Figure 16: The comparison of the ash distribution coverage between the original ash dispersal map and (a) the simulation results on 5 March 1997, and (b) the simulation results on 24 February 2006. The grey-stripped area indicates the original areas affected by the Taisho eruption, and the dark grey colour represents the ash dispersal from the simulation.

Tables

Eruption Source Parameters	Value
Mean Estimated Eruption Mass ($\times 10^{10}$kg)	1.1
Mass Eruption Rate, $\boldsymbol{\varepsilon}$ (kg/h)	5.8×10^{6} (Min) - 2.7×10^{10} (Max)
Horizontal Diffusion Coefficient, $\boldsymbol{K_h}$ (m/s)	150
Vertical Diffusion Coefficient, $\boldsymbol{K_v}$ (m/s)	1.5
Log-Scale Mean Grain Size (mm)	1.0

Table 1: Input values for the simulations used in this study.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

Deposit/CSV/Japan/<YYYY>/<YYYYMMDD.csv>			***Deposit/CSV/Kagoshima/<YYYY>/<YYYYMMDD.csv>***		
latitude	**longitude**	**deposit**	**latitude**	**longitude**	**deposit**
30.00	128.50	0.00	30.00	129.00	0.00
30.00	128.60	0.10	30.00	129.01	0.10
…	…	…	…	…	…
30.10	128.50	0.20	30.01	129.00	0.20
30.10	128.60	0.30	30.01	129.01	0.30
%.2f	*%.2f*	*%.2f*	*%.2f*	*%.2f*	*%.2f*

Table 2: The tabular view of the sample deposit data for both regions. Each column is separated by a comma, corresponding to the CSV format.

Original Form	Contemporary Form	Final Form
高気圧東に移り、中心本邦中部に在りて、773mm を示し。揚子江低気圧は目下漢口付近に存す。本邦にては、天気おおむね良好なれとも、九州にてはすでに曇天となり、北海道にては南西風強く晴曇相半す。満州および朝鮮にてはおおむね曇天となれり、琉球および台湾にてはこれ反して快晴なり。	*高気圧は東に移動し、日本の中心にあり、気圧計の値は 773mm を示すしている。長江にある低気圧は、ちょうど武漢付近に存在する。日本では、天気は晴れているが、九州地方では既に曇りとなっており、北海道では南西の風が強く晴れたり曇ったりしている。満州と朝鮮はおおむね曇り、琉球(沖縄)と台湾では、反して快晴。。*	*日本付近は移動性高気圧に覆われている。日本列島わ晴れ。一方、九州は曇り。北海道は南西風が吹き、晴れや、曇り。*

Table 3: The conversion of the general weather commentary in the weather chart on 12 January 1914.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Date	STS Score	Commentary (ja)	Commentary (en)
1997/03/05	0.963	*全国的に平年より5~10℃高い。日本付近は移動性高気圧に覆われるが、大陸から気圧の谷が進んでくる。北海道で曇りの他は似中晴れ。夜になると日本海側は曇り。山陰や九州北部では雨。*	*5~10℃ above normal nationwide.* ***A mobile high-pressure system will cover the area around Japan****, but a trough of pressure will move in from the continent.* ***Mostly cloudy in Hokkaido and sunny throughout****. Cloudy over the Sea of Japan at night. Rain in the San-in region and northern Kyushu.*
2006/02/24	0.957	*日本海から高気圧に覆われる。北海道東部~西日本の日本海側は晴れ。その他は曇り。関東~西の太平洋側は所々雨。*	***High pressure will cover the area from the Sea of Japan. Clearing on the Sea of Japan side from eastern Hokkaido to western Japan. Cloudy in other areas.*** *Some rain on the Pacific side from Kanto to the west.*

Table 4: The top two results of the semantic search process with its weather description from the commentary.

910

Actual Ash Precipitation of the *Taisho* Eruption	**Daily Simulation Results**	
	Affected	*Unaffected*
Affected	**TP**	**FN**
Unaffected	**FP**	**TN**

Table 5. The binary contingency table for the validation process.

Date	STS Score	*Hit* Score
1997/03/05	0.964	0.963
2006/02/24	0.957	0.999

Table 6. The performance of the simulation results from the selected dates.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

Figures

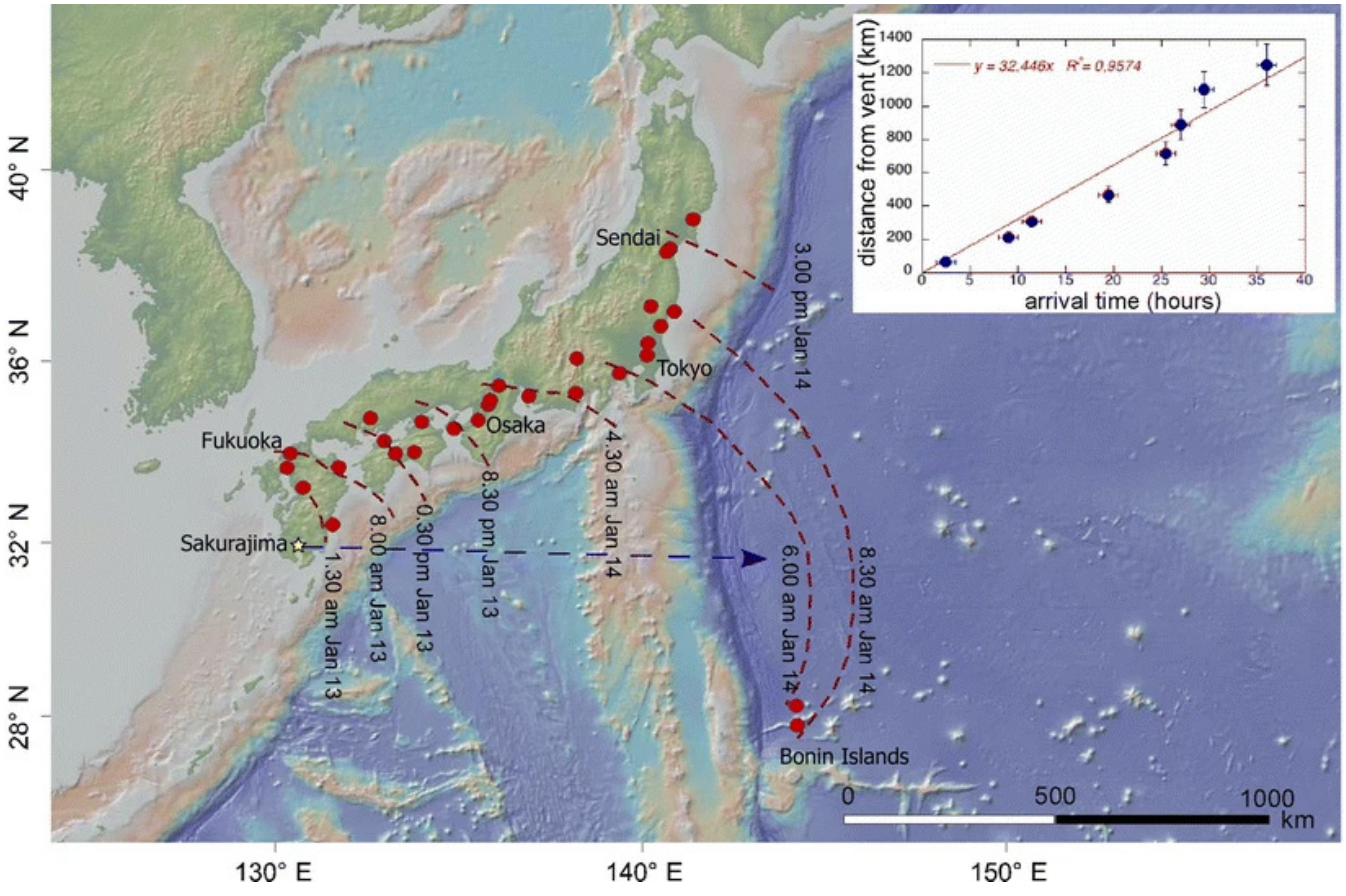
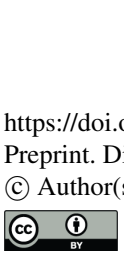


Figure 1.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.



Open Access Earth System Science Data Discussions

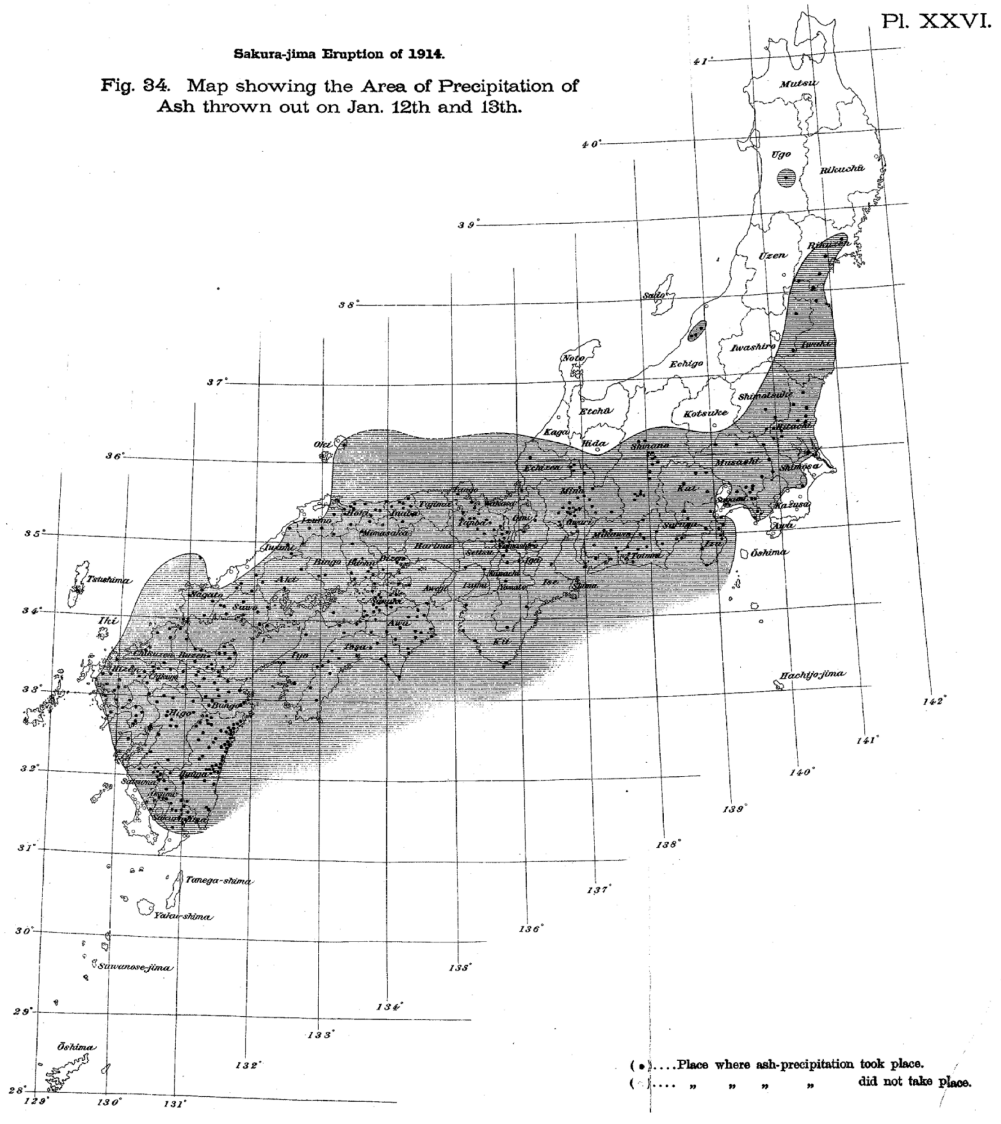


Figure 2.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

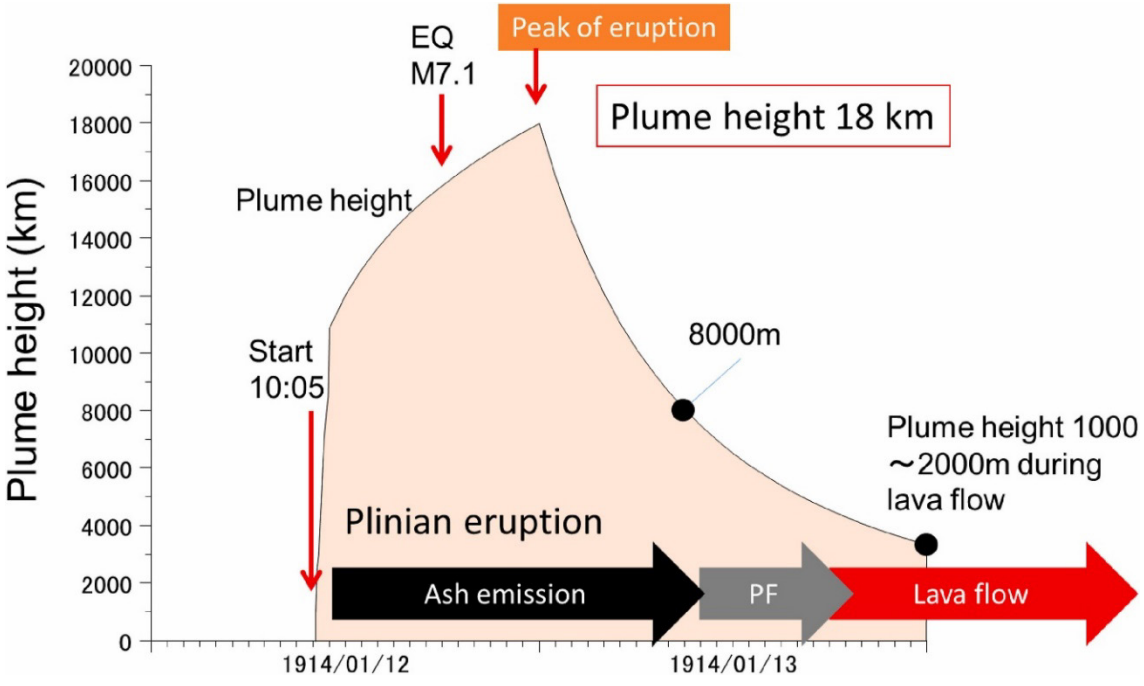


Figure 3.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

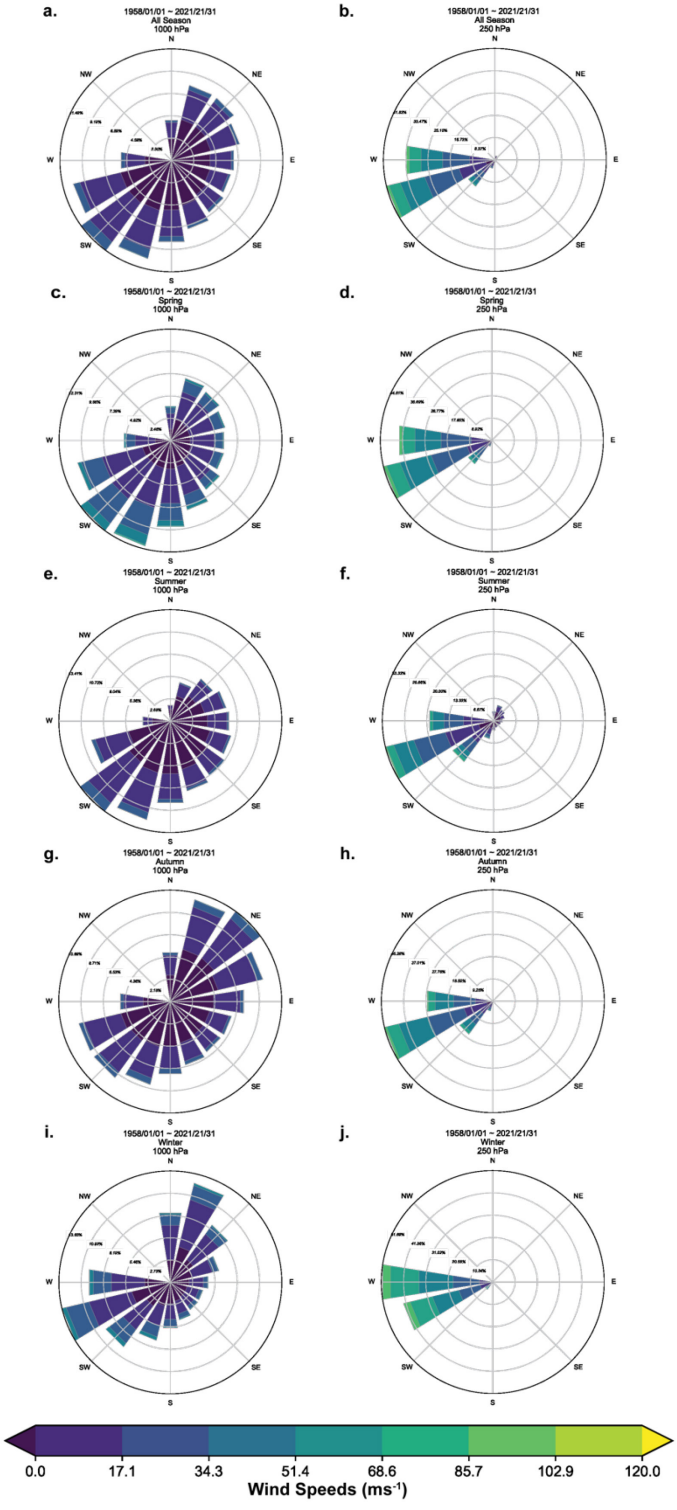


Figure 4.

925

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

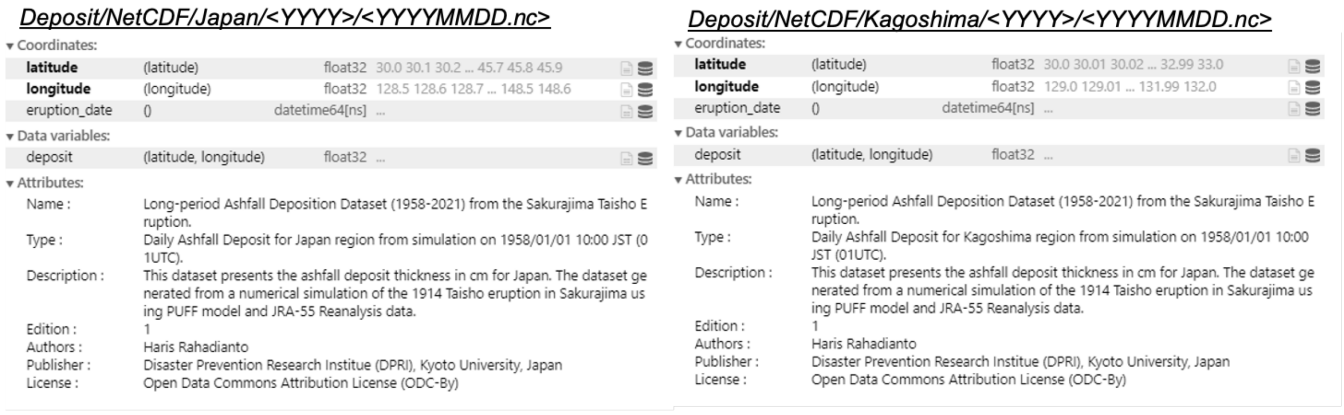


Figure 5.

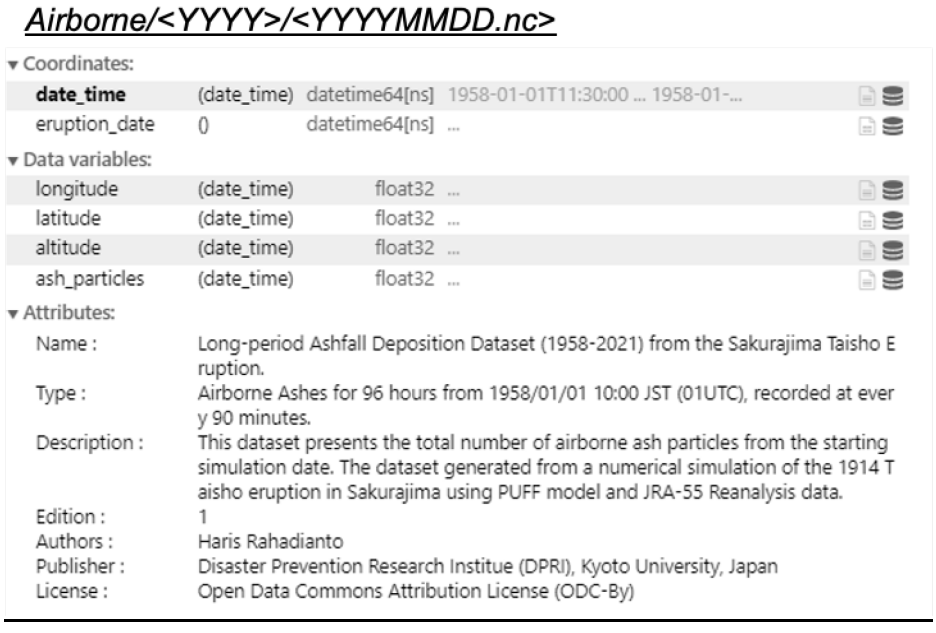
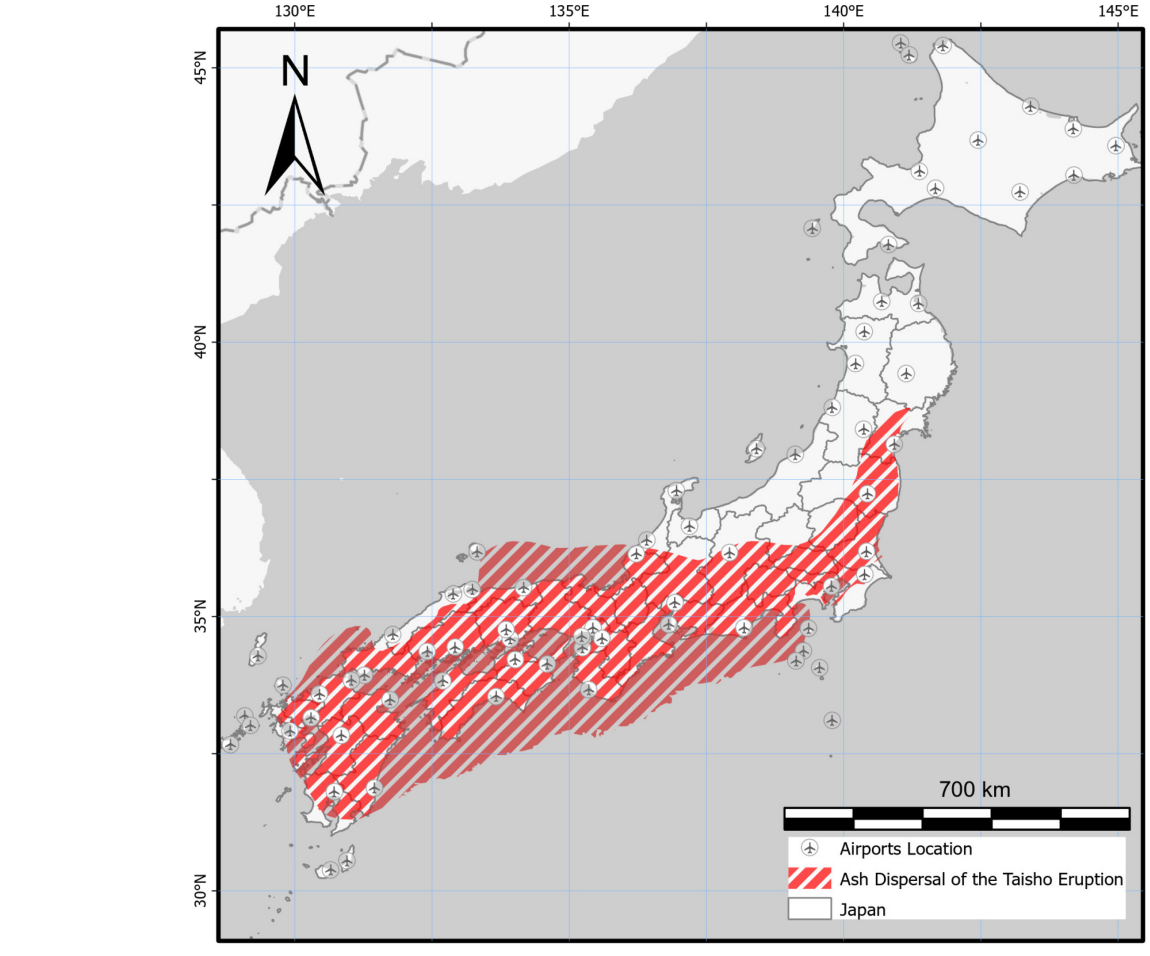


Figure 6.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.



930

Figure 7.

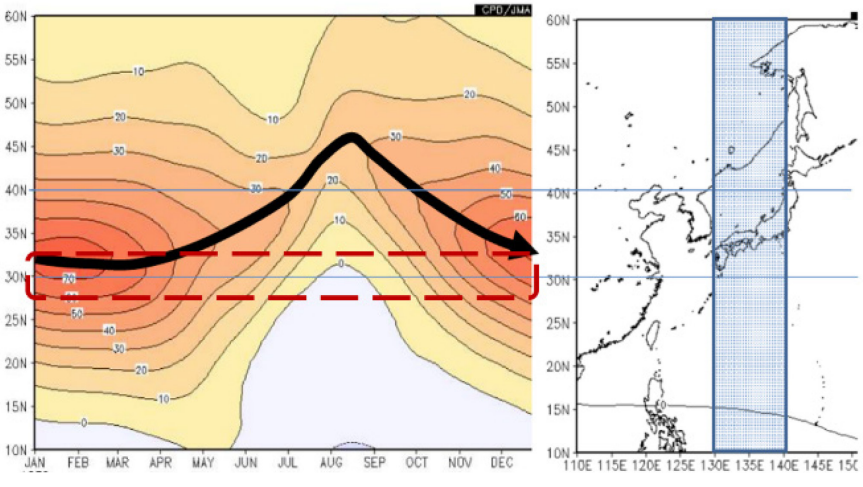
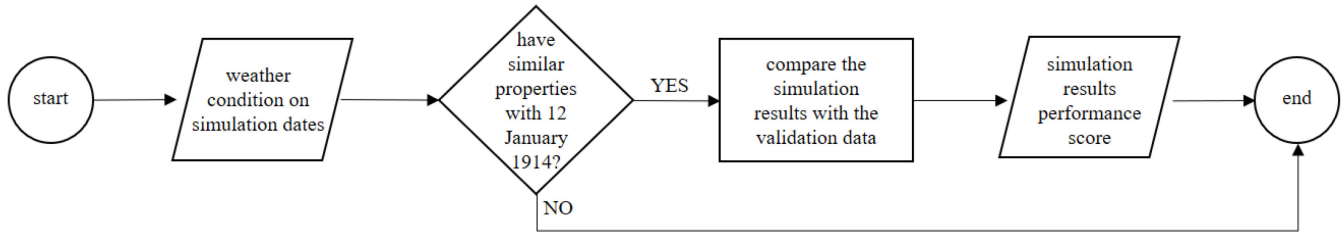


Figure 8.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.



935 **Figure 9.**

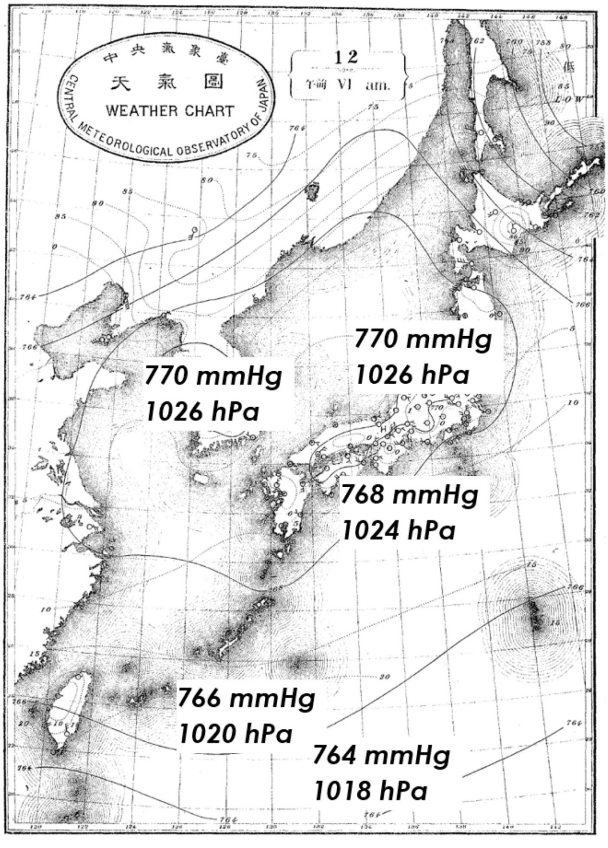


Figure 10.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

940

Figure 11.

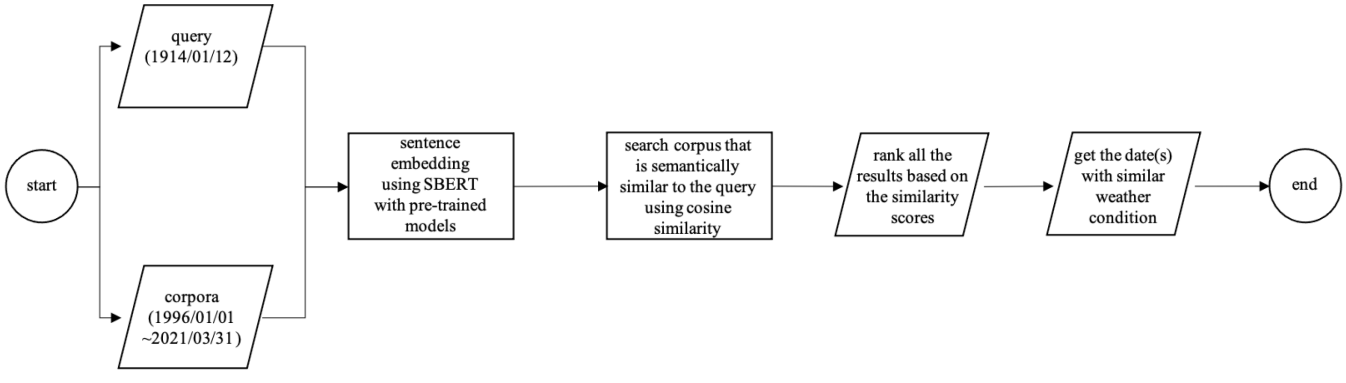
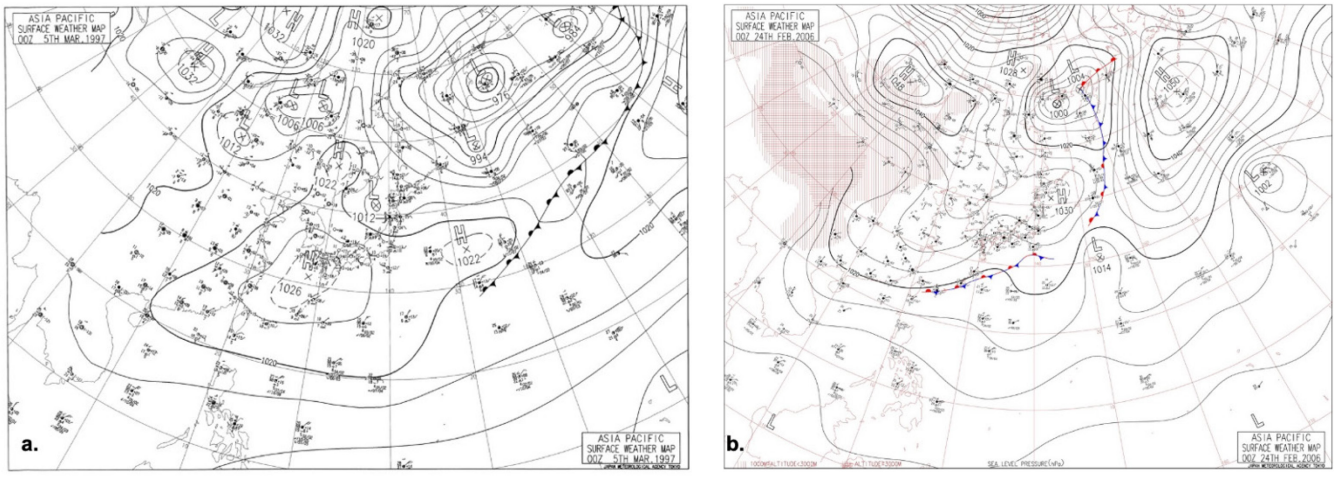


Figure 12.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.



945

Figure 13.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.

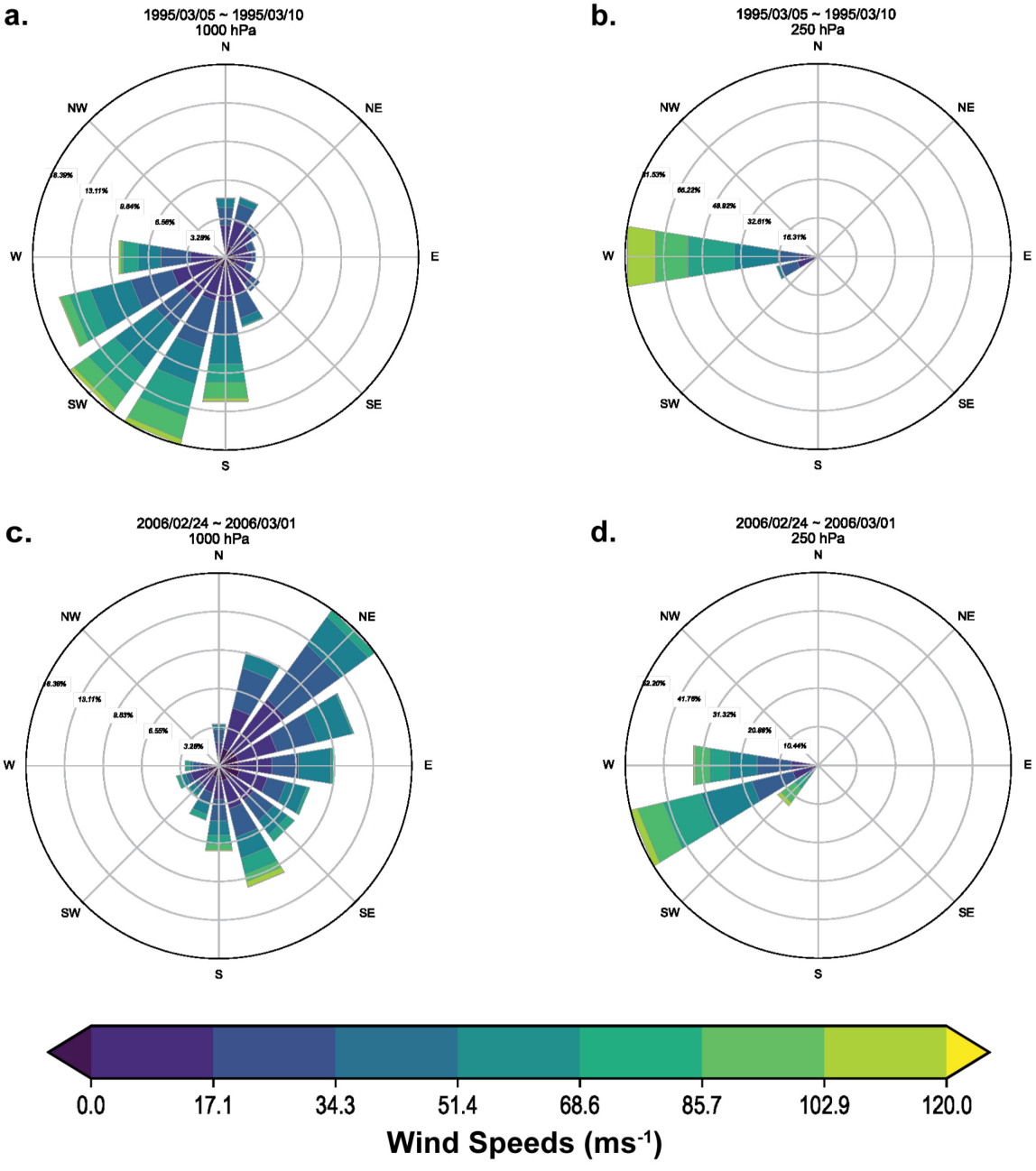
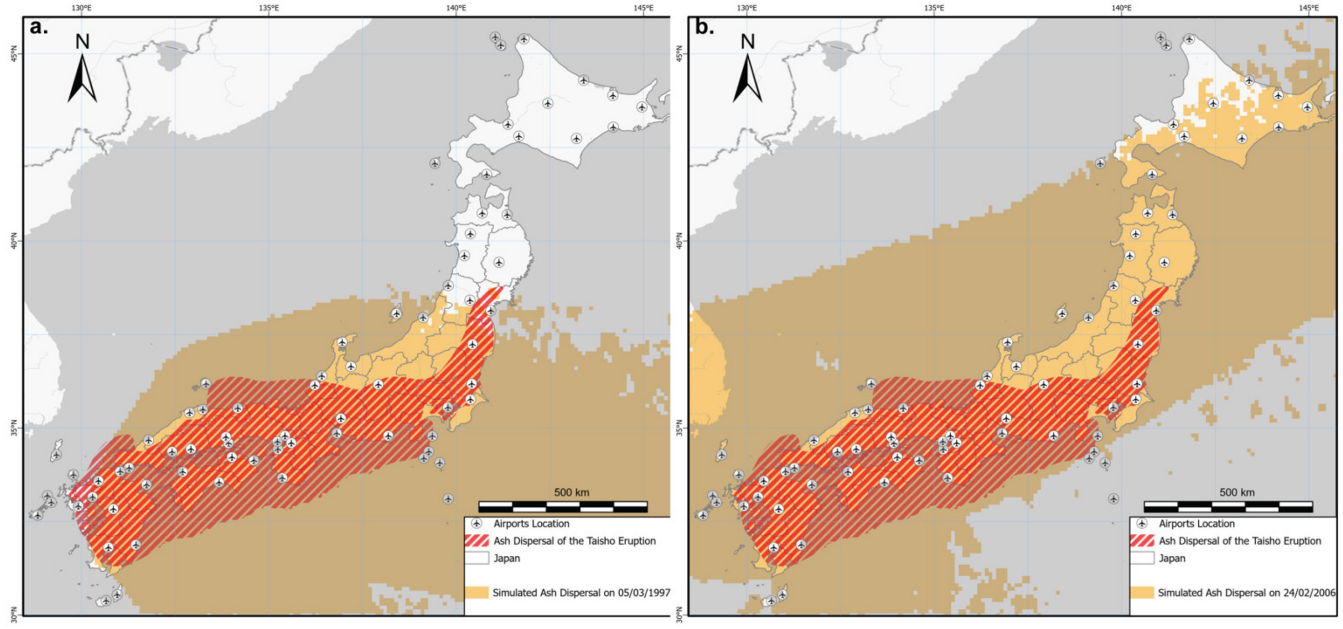


Figure 14.

https://doi.org/10.5194/essd-2022-42
Preprint. Discussion started: 7 June 2022
© Author(s) 2022. CC BY 4.0 License.



950 **Figure 15.**

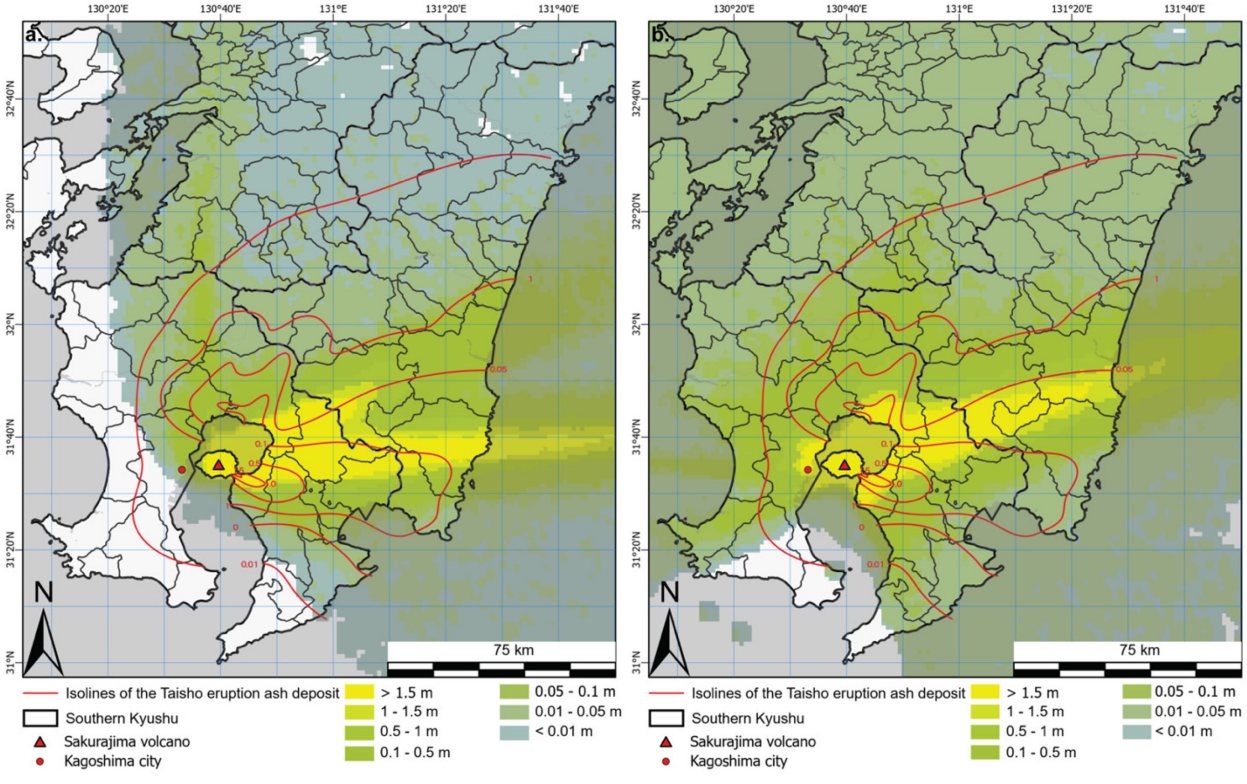


Figure 16.