Visualizing volcanic ash forecasts: scientist and stakeholder decisions using different graphical representations and conflicting forecasts

Article

Accepted Version

Mulder, K. J., Lickiss, M., Harvey, N. ORCID: https://orcid.org/0000-0003-0973-5794, Black, A., Charlton-Perez, A. ORCID: https://orcid.org/0000-0001-8179-6220, Dacre, H. ORCID: https://orcid.org/0000-0003-4328-9126 and McCloy, R. ORCID: https://orcid.org/0000-0003-2333-9640 (2017) Visualizing volcanic ash forecasts: scientist and stakeholder decisions using different graphical representations and conflicting forecasts. Weather, Climate and Society, 9 (3). pp. 333-348. ISSN 1948-8327 doi: 10.1175/WCAS-D-16-0062.1 Available at https://centaur.reading.ac.uk/68953/

To link to this article DOI: http://dx.doi.org/10.1175/WCAS-D-16-0062.1

Publisher: American Meteorological Society

Visualizing Volcanic Ash Forecasts: Scientist and Stakeholder Decisions using Different Graphical Representations and Conflicting Forecasts

Kelsey J. Mulder*

Department of Meteorology, University of Reading, United Kingdom

Matthew Lickiss

Department of Typography and Graphic Communication, University of Reading

Natalie Harvey

Department of Meteorology, University of Reading

Alison Black

Department of Typography and Graphic Communication, University of Reading

Andrew Charlton-Perez, Helen Dacre

Department of Meteorology, University of Reading

Rachel McCloy

Department of Psychology, University of Reading

**Corresponding author address:* Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading, RG6 6BB, United Kingdom.

E-mail: k.mulder@reading.ac.uk

ABSTRACT

During volcanic eruptions, Volcanic Ash Advisory Centres issue ash advisories for aviation showing the forecasted outermost extent of the ash cloud. During the 2010 Icelandic volcano Eyjafjallajökull eruption, the UK Met Office produced supplementary forecasts of quantitative ash concentration, due to demand from airlines. Additionally, satellite retrievals of estimated volcanic ash concentration are now available. To test how these additional graphical representations of volcanic ash affect flight decisions, whether users infer uncertainty in graphical forecasts of volcanic ash, and how decisions are made when given conflicting forecasts, a survey was conducted of 25 delegates representing UK research and airline operations dealing with volcanic ash. Respondents were more risk-seeking with safer flight paths and risk-averse with riskier flight paths when given location and concentration forecasts compared to when given only the outermost extent of the ash. Respondents representing operations were more risk-seeking than respondents representing research. Additionally, most respondents' hand-drawn no-fly zones were larger than the areas of unsafe ash concentrations in the forecasts. This conservatism implies that respondents inferred uncertainty from the volcanic ash concentration forecasts. When given conflicting forecasts, respondents became more conservative than when given a single forecast. The respondents were also more risk-seeking with high-risk flight paths and more risk-averse with low-risk flight paths when given conflicting forecasts than when given a single forecast. The results show that concentration forecasts seem to reduce flight cancellations while maintaining safety. Open discussion with the respondents suggested that definitions of "uncertainty" may differ between research and operations.

1. Introduction

a. Background

Volcanic ash is a significant hazard to aviation. For example, volcanic ash contains silica particles, which melt when ingested into airplane engines. This can cause temporary engine failure and permanent engine damage. Although avoiding flying through volcanic ash reduces risk of engine damage or failure, it also disrupts air traffic, resulting in substantial financial losses for the aviation industry. For example, the 2010 eruption of Icelandic volcano Eyjafjallajökull disrupted airspace over Europe for 13 days with over 95,000 flights grounded. This cost an estimated €3.3 billion in losses to the airline industry (Mazzocchi et al. 2010). One reason the event was so disruptive was that it occurred in the highly congested European airspace: 879 million people traveled by air in the European Union in 2014 (European Commission 2016). The 2010 eruption was not necessarily a rare event: a study of historic eruptions in Iceland over the past 1,100 years shows that volcanic eruptions occur 20–25 times every 100 years, with approximately three-quarters of these eruptions being explosive (Thordarson and Larsen 2007). Some of these eruptions can release much more ash into the atmosphere and erupt for longer (months to years) than the 2010 Eyjafjallajökull eruption (Thordarson and Larsen 2007). Globally, volcanic eruptions occur nearly daily.

The decision to fly or not during volcanic eruptions is solely the responsibility of the airline operator, not the Civil Aviation Authority (CAA, Safety and Airspace Regulation Group 2014). However, the CAA does require a safety risk assessment to be conducted before the operator is allowed to fly in airspace contaminated by volcanic ash. The safety risk assessment ensures that the operator has a safety management system, has a proven safety record, has the ability to evaluate volcanic ash risk, has documented procedures (such as how to avoid ash en-route), has received training in unusual circumstances and emergencies, and understands the impact of volcanic ash on the aircraft. The safety risk assessment must then be approved by the CAA (Safety and Airspace Regulation Group 2014).

The other requirement the CAA places on flights in airspace affected by volcanic ash is that operators are required to use Volcanic Ash Advisory Centre (VAAC) advisories, which are produced both graphically and in a text format. The London VAAC, based at the UK Met Office, is responsible for issuing volcanic ash advisories for the United Kingdom, Republic of Ireland, Iceland, and Scandinavia. The volcanic ash advisories, approved by the International Civil Aviation Organization, forecast the furthest extent of the ash cloud on three pre-approved flying altitudes.

b. Past literature

Questions of decision making in natural hazards have been widely studied, involving participants who are both experts and non-experts. Experts may behave differently from non-experts because of their familiarity with the hazard, data presentation, and the types of decisions that are made in the face of these hazards. Indeed, experts have been shown to have different risk perceptions than non-experts in hazards such as flash flooding (Morss et al. 2016) and therefore may be expected to behave differently in decision tasks.

However, similar to non-experts, experts can succumb to cognitive biases such as positive versus negative framing (e.g., Taylor et al. 1997) and anchoring (e.g., Whyte and Sebenius 1997; Englich et al. 2006). Additionally, some studies suggest that experts may not behave differently in decision-tasks than non-experts. In a study of decision making with different types of wind forecasts, both expert and novice forecasters had similar results: they performed most accurately when using a box plot, succumbed to anchoring when the worst-case scenario forecast was presented, and chose a box plot as easiest to use as a forecast aid (Nadav-Greenberg et al. 2008a).

Other studies suggest that the classification of a participant as "expert" may not be as important as other factors. In a decision-task study of military personnel, the amount of direct experience in a Combat Operations Center significantly affected decisions whereas rank and years of service did not (St. John et al. 2000). In another decision task, numeracy (which can vary widely across expert groups) predicted how well participants performed when given probabilistic information (Peters et al. 2006). Because there is not necessarily a distinction between how experts and non-experts perform in decision tasks, literature using both groups as participants have guided our research questions (discussed in section 1c).

Both experts and non-experts are able to process and use forecast information that is inherently uncertain to make decisions. For example, a non-expert student sample was able to understand basic hurricane track information (Wu et al. 2014). Additionally, evidence suggests that experts (e.g., St. John et al. 2000; Aerts et al. 2003; Riveiro et al. 2014) and non-experts (e.g., Morss et al. 2010; Correll and Gleicher 2014) interpret probabilities well enough to inform decisions when given uncertainty information on topics such as military tactics, land use, air traffic control, voter preference, snowfall predictions, and payout expected by a fund. Even with unfamiliar hazards or information, risk judgments can improve when training is provided (e.g., McCloy et al. 2007).

Although experts and non-experts can understand and use natural hazard information in decision making, their decisions may change based on how the information is presented. For example, for flood risk, a sample of non-experts indicated that, "within 40 years, there's a 33% probability of a flood" was riskier than "each year, there's a 1% probability of a flood," even though they represent the same likelihood of flooding (Keller et al. 2006). Similarly, the way information is shown for other hazards such as wind, hurricanes, snow, and precipitation has been shown to affect decisions in experts (e.g., Nadav-Greenberg et al. 2008b; Cox et al. 2013) and non-experts (e.g., Ibrekk and Morgan 1987; Abraham et al. 2015; Ruginski et al. 2015). However, in one study on hurricanes, non-experts perceived no significant difference in the likelihood of a hurricane striking a location when the hurricane forecasts showed the forecast track only, uncertainty cone only, or forecast track with an uncertainty cone (Wu et al. 2014). These studies suggest that further research needs to be conducted on the effect of information design on decision making.

One subset of research on decision making investigates whether giving more detailed information about a natural hazard affects respondents' decisions. Providing probabilistic forecast information rather than deterministic forecast information has been shown to encourage more economically rational decisions for both experts (e.g., Kirschenbaum and Arruda 1994; St. John et al. 2000; Nadav-Greenberg et al. 2008b; Riveiro et al. 2014) and non-experts (e.g., Joslyn et al. 2007; Nadav-Greenberg and Joslyn 2009; Roulston and Kaplan 2009; Joslyn and LeClerc 2012). Elaboration of the impact of a hazard also affects decisions: more serious volcanic eruption impacts encouraged more members of a community to take protective action (Ekker et al. 1988).

Increasing the resolution of the hazard information has also been tested. In the United States, reducing the size of tornado warnings from county-level to be polygons within and between counties had an effect on protective action, with more non-experts choosing to take protective action when given smaller warning polygons proximate to their locations (Nagele and Trainor 2012). When testing between deterministic and probabilistic tornado warning graphics, Ash et al. (2014) found that probabilistic forecasts encouraged non-expert protective action in the highest risk areas. In addition, non-experts indicated a non-zero probability of a tornado occurring just outside the warning areas, whereas with the deterministic polygon, the risk was perceived as localised to within the polygon (Ash et al. 2014). Providing airline pilots more information about the predicted future location of nearby aircraft encouraged safer decisions to prevent collisions (Wickens et al. 2000). These studies suggest that providing more information about a hazard encourages safer and more economically rational decisions.

Another important aspect of decision research is how users interpret deterministic forecasts when no uncertainty is provided. When given a deterministic forecast and decision task for either managing reservoir levels given a rain forecast or protecting crops given a temperature forecast, non-experts took protective action even when the forecast was on the safe side of the given threshold, inferring there was uncertainty in the forecast (Morss et al. 2010). In another study, when non-experts were only given a deterministic windspeed or temperature forecast, they forecasted much lower values than those given in the forecast, indicating they adjusted the forecast, perhaps based on the amount of uncertainty they perceived in the forecast (Joslyn et al. 2011). Non-experts also inferred additional uncertainty information into a probability of an event occurring in a one-week period, suggesting that the event was more likely toward the end than the beginning of the week (Doyle et al. 2014). These studies indicate that experts and non-experts infer uncertainty into text-based deterministic forecasts when it is not explicitly stated.

Uncertainty can also be inferred in graphical forecasts. For example, non-experts tend to infer a normal distribution of probabilities into a deterministic forecast, with a higher probability in the middle of a graphically defined area and lower probabilities toward the outside in both temperature forecasts (e.g., Tak et al. 2015) and tornado warnings (e.g., Sherman-Morris and Brown 2012; Ash et al. 2014; Lindell et al. 2016). However, in some circumstances, such as with tornadoes, the highest risk areas are at the edges, not in the middle of the polygon (Ash et al. 2014). Another way in which inferred uncertainty is evident is in the perception of risk just outside the warning or forecast area. Some studies have shown that non-experts acknowledge a low, but non-zero tornado probability just outside of tornado warning areas (e.g., Nagele and Trainor 2012; Lindell et al. 2016) and the hurricane cone of uncertainty graphic (e.g., Wu et al. 2014). However, other studies on the hurricane cone of uncertainty graphic suggest that non-experts gain little understanding of the uncertainty in hurricane track forecasts from the polygon graphic either because they are too focused on the forecast track line (e.g. Broad et al. 2007) or because they only interpret the direction of hurricane motion from the graphic (e.g., Wu et al. 2015). When inferring uncertainty into deterministic graphical forecasts, users may be inferring uncertainty incorrectly, which may lead to unsafe decisions.

c. Research questions

The combination of previous decision-based research and the 2010 Eyjafjallajökull eruption brought up three questions, which are the focus of this paper. First, during the 2010 Eyjafjallajökull eruption, the UK Met Office began producing supplementary forecasts of ash concentration in addition to the official VAAC forecasts showing the furthest extent of the ash cloud (Webster et al. 2012). Additionally, satellite retrievals of volcanic ash concentrations are becoming available. These changes in availability of graphical representations raised the question: how are different representations of ash concentration interpreted and used to make decisions by the aviation industry as well as the researchers who created these graphics?

Past research suggests that increasing the amount of information given about hazards leads to decisions that are safer and more economically rational. Therefore, the responses to the UK Met Office supplementary volcanic ash concentration forecasts and satellite retrievals of volcanic ash concentration may encourage safer decisions while still reducing the number of unnecessary flight cancellations. However, previous research has not tested how including more information graphically affects decision making in a volcanic ash context.

The second research question addressed in this article is: without uncertainty (e.g., uncertainty in 3D location or concentration of volcanic ash) being explicitly represented graphically in volcanic ash forecasts, how much uncertainty are users inferring from the forecasts? Does inferring uncertainty result in risky or over-conservative flight decisions? Past research suggests that users may infer uncertainty into both text-based and graphical deterministic forecasts, but they may make different inferences for volcanic ash. Therefore, it is important to understand how users make inferences about uncertainty from volcanic ash forecasts.

During the 2010 Eyjafjallajökull eruption, more than one VAAC provided volcanic ash forecasts, which were sometimes slightly different due to differences in the model being used and assumptions made about the state of the volcano. This problem inspired our third research question: how are operational decisions made when experts are given conflicting forecasts? Little research has been conducted on this topic, although it has been shown previously that experts do seek multiple sources of information to confirm their decisions (e.g., Morss et al. 2015).

To answer these questions, a survey was conducted at the National Environmental Research Council (NERC) Volcanic Ash Workshop in London on 22 February 2016. The workshop brought together 25 delegates representing research and airline operations (including pilots, engine manufacturers, airline representatives, and the Civil Aviation Authority) to discuss recent advances in volcanic ash forecasting and observations, ongoing challenges, and visualizations.

2. Methods

a. Participants

The Volcanic Ash Workshop was a one-day meeting in London, funded by the National Environmental Research Council (NERC) on 22 February 2016, designed to encourage discussion about volcanic ash from both academic and private sectors. The participants invited to the workshop were a mixture of airline operators, policymakers, and researchers (both academic and embedded in the aviation industry). Invitations to the workshop were extended to colleagues the co-authors had worked with previously on the topic of volcanic ash with further invitations being extended by the recommendations of those invited. Out of 78 individuals invited to the Volcanic Ash Workshop, 25 attended (excluding the co-authors and organizers). The final survey was completed by all 25 delegates. All attendees of the Volcanic Ash Workshop, except for the co-authors of this paper, agreed to participate in the survey.

Of the 25 respondents, 16 represented research (the majority of researchers were working at a university, but some were researchers embedded in institutions such as the UK Met Office) and 9 represented operations (including flight planners, airline manufacturers, airline representatives, pilots, and employees of the CAA). The level of job experience ranged from 2–18 years with a mean of 7 years. The respondents ranged in age from 28–69 with a mean age of 46. Most (80%) respondents were male. Although the 25-respondent sample size for this decision-making survey is small, expert groups are naturally smaller than public samples.

Because the sample size was small, comparing responses between other variables, such as age and gender, was not possible either because the sample size would be too small for one group or because no meaningful divisions between participants could be made. Comparisons between job experience were tested between those with less than or equal to five years of job experience and those with more than five years of job experience. The responses for these two groups were not significantly different.

b. Materials

This study was given favorable ethical opinion for conduct by the University Research Ethics Committee. The survey used in this study was piloted with five PhD students from the University of Reading Meteorology Department. The survey was distributed once the delegates arrived. The delegates were informed that participation was entirely voluntary, however every delegate participated. Respondents were given approximately 45 minutes to complete the survey. After they had completed the survey, there were a series of presentations from operations specialists and researchers discussing current challenges and recent advances in volcanic ash forecasting and observations. At the end of the day, there was an open group discussion about forecasting and communicating uncertainty of volcanic ash in aviation.

The survey consisted of four sections: low-, medium-, and high-risk flight decisions across three different graphic types; drawing no-fly zones onto four volcanic ash forecasts; four flight decisions given conflicting information; and sociodemographic information. The four sections were presented in the same order for each respondent, however the order of the graphics or forecasts were randomized within each section.

In the first section, respondents were given four flight paths overlaid onto a volcanic ash forecast (Fig. **??**a). The respondents then determined if they would approve the flight paths. The four flight paths were high risk (flight path A, going through the center of the volcanic ash plume), medium–high risk (flight path B, going through the polygon and going just outside the high levels of concentration in the filled contour and satellite graphics, described further below), medium–low risk (flight path C, going through the polygon and going just inside medium levels of concentration in the filled contour and satellite graphics), and low risk (flight path D, skimming the outside of the volcanic ash plume). Respondents were given the same flight paths and forecasts for three different graphic types: polygon, filled contour, and satellite.

The polygon graphic was similar to the official VAAC forecasts, showing the outermost extent of volcanic ash. The VAAC graphic is created by forecasters using an atmospheric dispersion model, local observations, reports from pilots, and satellite data (described below) (Millington et al. 2012). Operationally, these forecasts are presented in both graphical and text format so they can be transmitted to pilots mid-flight. Due to character limits in the text forecasts, the VAAC official polygons have limited complexity.

The filled contour graphic was similar to the forecast distributed by the UK Met Office since the 2010 Eyjafjallajökull eruption and showed both ash location and concentration. Similar to the polygon graphic, the filled contour graphic is created by forecasters using an atmospheric dispersion model, local observations, reports from pilots, and satellite data (Millington et al. 2012). Concentration levels for the filled contour graphic were shown in three bands: 200–2000, 2000–4000, and $> 4000\ \mu g\ m^{-3}$, similar to what is used operationally.

The satellite graphic simulated satellite ash retrievals. To produce this graphic operationally, difference in brightness temperature from satellite observations are used at three different wavelengths. Then, using data from a numerical weather prediction model and a radiative transfer model, ash column loading (the sum of all volcanic ash in a column), ash cloud height, and ash particle size are modeled. These quantities are dependent not only on the numerical weather prediction and a radiative transfer models, but also on the assumed refractive index of the ash. The satellite representation in the survey was artificially created and had six levels of concentration (500, 1000, 2000, 3000, 4000, and 5000 $\mu g\ m^{-3}$), rather than three for the filled contour representation.

It is of note that the level of ash concentration that was safe to fly through was debated as the 2010 Eyjafjallajökull eruption continued (for more information on the timeline of events, please see http://www.caa.co.uk/Safety-initiatives-and-resources/Safety-projects/Volcanic-ash/A-history-of-ash-and-aviation/). Further research has since been conducted on the effects of volcanic ash on airplane engines to further clarify what amount of volcanic ash is considered safe (e.g., Clarkson et al. 2016).

The purpose of the first section of the survey was twofold. First, by comparing decisions for different levels of risk for the same graphic, we could determine the risk appetite for each respondent. Second, by comparing the same flight path across different graphical representations, we determined how different graphical representations affected decision making. The responses were checked for consistency. Responses of one respondent, who appeared to misunderstand the task, were removed from the numerical analysis of this section only because their flight decisions shifted towards approval as ash concentrations increased. The respondent's qualitative feedback in this section and quantitative and qualitative responses from the other sections were included in this paper.

To establish context for responses from the first section, respondents were asked their familiarity with, trust in, and preferences for the three representations: polygon, filled contour, and satellite. Familiarity and trust were measured by rulers on 10-cm visual analogue scales ranging from "Never seen before" (0 cm) to "Have seen frequently" (10 cm) for familiarity and "Not at all trustworthy" (0 cm) to "Extremely trustworthy" (10 cm) for trust. Preference was measured as a multiple choice question.

The second section tested how much uncertainty respondents perceived in the filled contour and satellite graphical representations as well as whether including a gap in the forecast ash concentration influenced their perception of uncertainty. In the second section, respondents were given four different forecasts and were asked to draw no-fly zones directly on the forecast. The forecasts were shown for two different graphical representations (filled contour and satellite) and two different shapes of volcanic ash plume. The two shapes of volcanic ash plume were a "solid" ash plume with concentric concentration levels and a "gap" ash plume with two areas of high volcanic ash concentration and lower concentrations between them (Fig. **??**b).

To measure the perception of uncertainty in the second section, each no-fly zone map was scanned into Adobe Illustrator (a vector graphics software package). The boundary edge of the no-fly zones drawn by each participant were then traced as vector paths and sorted into individual layers. With all of the no-fly zones digitized as vectors, their areas were calculated and the no-fly zones were overlaid and compared visually in grouped layers.

The purpose of the third section of the survey was to investigate the impact of conflicting forecast information on decision making by analyzing the respondents' flight decisions and confidence levels. Respondents were given the same flight path overlaid onto two different filled contour forecasts, described as being issued simultaneously, and were asked whether they would approve the flight path. The forecasts were coded based on what color contours the flight paths went through: blue–blue, grey–grey, red–blue, and red–grey (Fig. **??**c). Additionally, respondents were asked what further information would help them make a decision to fly or not fly given conflicting forecasts.

For all the flight decisions, respondents were told that the forecast was issued three hours ago and valid now, when flights would take off. They were also told they had permission to fly through medium concentrations of volcanic ash (2000–4000 $\mu g\ m^{-3}$) corresponding to the blue and grey areas in the filled contour representation and the green, yellow, and orange areas in the satellite representation (Fig. **??**). This information was important because the safe level of ash concentration varies according to each airline's safety assessment, required by the CAA. None of the representations explicitly showed uncertainty, even though uncertainty was inherent in all three representations. For all flight decisions, respondents were also asked how confident they were in their decision, which was marked on a 10-cm visual analogue scale ranging from "Not at all confident" (0 cm) to "Extremely confident," (10 cm) and was measured using a ruler. All decision questions were also followed by an open-ended question asking what information influenced their decision.

The fourth section gathered respondents' job title (used to determine if the respondent worked in research or operations), length of time in current job, age, and gender.

3. Results

a. How do graphical representations of volcanic ash affect operational decisions?

Comparing flight decisions between graphical representations, fewer respondents approved high-risk flight paths (Fig. **??**a) and more respondents approved low- and medium–low-risk flight paths (Fig. **??**c and d) for the filled contour and satellite representations than the polygon representation. In the high-risk flight path, 17% of respondents approved the flight when given the polygon representation compared with 0% for the filled contour and 4% for the satellite representations (Fig. **??**a). In the low-risk flight path, 71% of respondents approved the flight when given the polygon representation compared with 83% for the filled contour and 83% for the satellite representations (Fig. **??**d). In other words, given concentration and location information, the respondents were more risk-averse for the riskier flight paths and risk-seeking for the safer flight paths. The exception was the medium–high-risk flight path, where both the polygon and filled contour representations encouraged risk-seeking decisions and the satellite representation encouraged risk-averse decisions (29% approved the flight path for both polygon and filled contour representations compared with 21% for satellite representation) (Fig. **??**b).

The filled contour and satellite representations also increased confidence in the respondents' decisions (Fig. **??**e–h). The mean confidence across all flight paths, was 6.3 for the polygon, 7.1 for the filled contour, and 7.2 for the satellite. Across all flight paths, there was a significant difference at the 5% level between mean confidence ratings across the different types of graphical representation (ANOVA, $F = 3.2$, $p = 0.04$).

In an open-ended question about what information influenced their flight decisions, over 50% of respondents indicated they needed more information to help them make a decision when given the polygon representation, compared with 20% for the filled contour and 16% for the satellite representation.

Respondents were asked in an open-ended format what further information they would need from each graphical representation to be more confident in their decisions. The responses varied widely and included altitude information, observations, past model performance, meteorological information, higher resolution, and uncertainty information. Interestingly, nine of the sixteen respondents representing research mentioned needing uncertainty, probability, accuracy, or confidence information whereas no respondents representing operations mentioned any of the above.

Separating the flight decisions by occupation, respondents working in operations (n =9) were more risk-seeking than those in research (n =15), with a higher percentage of respondents choosing to approve the flight path for all levels of risk (52% of the decisions of respondents in operations compared with 38% of the decisions of respondents in research, Fig. **??**a). This relationship was not statistically significant, perhaps because of the small sample size (t-test, $t = 1.4$, $p = 0.18$). Respondents representing operations were more confident in their decisions across all flight paths (means 7.4–9.0) than those in research (means 5.1–7.4, Fig. **??**b). The difference in mean confidence between respondents in operations and research was significant at the 5% level (t-test, $t = 4.6$, $p < 0.001$).

The respondents were most familiar with the filled contour (mean 6.7) and polygon (mean 6.1) representations and least familiar with the satellite representation (mean 5.3, Fig. **??**a). However, the respondents trusted the satellite representation (mean 6.6) more than the polygon (mean 5.4) and filled contour (mean 4.8) representations (Fig. **??**b). Respondents in operations (n =9) and research (n =16) had different familiarity in the graphical representations. Respondents in research were most familiar with the filled contour representation (mean 6.0), followed by the satellite (mean 5.1) and polygon (mean 4.9) representations, compared with those in operations who were most familiar with the polygon (mean 8.3), followed by filled contour (mean 7.9), and satellite representations (mean 5.6, Fig. **??**a).

Respondents trusted the satellite graphical representation the most (mean 6.6) followed by the polygon (mean 5.4) and filled contour (mean 4.8) representations. Respondents in operations trusted all graphical representations (mean 6.4) more than those in research (mean 5.1, Fig. **??**b). The difference in mean trust between operations and research was not statistically significant at the 5% level (t-test, $t = 1.3$, $p = 0.22$).

Product preferences varied among the respondents and whether they represented research or operations. Respondents in research preferred filled contour (45%) and satellite representations (45%) while respondents in operations preferred the satellite representation (42%, Fig. **??**c). Only respondents representing operations preferred "other" representations and specified graphics showing ash column loading and observational data. Ash column loading, which shows the sum of all the volcanic ash in a column, is similar to the satellite representation given in the survey, which showed the peak concentration in the column.

b. Do users infer uncertainty in graphical forecasts of volcanic ash?

When given a single volcanic ash forecast and four flight paths of differing risk (section 1 of the survey, Fig. **??**a), the respondents were conservative in their decisions. Only 79% of the low-risk flight paths (Path D), which travelled through safe concentrations of volcanic ash across all graphical representations, were approved (Fig. **??**d). This conservatism suggests that respondents infer uncertainty in the forecasts, otherwise 100% of respondents would approve the low-risk flight paths.

Respondents were asked to draw a no-fly zone around two different shapes of volcanic ash forecasts, one showing a gap between high concentrations of volcanic ash (simulating potential error in satellite retrieval of volcanic ash concentrations, as described in section 2) and one with a single area of high volcanic ash concentration (section 2 of the survey, Fig. **??**b). Six of the twenty-four respondents drew their no-fly zones to allow flights through the gap between the two areas of high volcanic ash concentrations, shown by overlaying the no-fly zones (Fig. **??**). Four of these six respondents were in operations.

To quantify the differences in the perception of forecast uncertainty for the gap and solid forecasts, the areas of the no-fly zones drawn by respondents were calculated. Respondents tended to draw no-fly zones with larger areas for the gap (mean 1214 mm^2) than for the solid (mean 1013 mm^2) forecasts (Fig. **??**b). However, the difference in means between solid and gap forecasts were not significantly different at the 95% level (t-test, $t = -1.0$ $p = 0.32$). The areas drawn in the different conditions may have been influenced by the larger area red zone in the gap (357 mm^2) than the solid (241 mm^2) forecast.

Most of the respondents' no-fly zone areas were larger than the areas of the red zones on the forecasts, again suggesting that the respondents inferred uncertainty from the forecasts. For the gap forecasts, the hand-drawn no-fly zones were between 7% smaller and 866% larger than the red zone (mean size was 231% larger than the red zone). For the solid forecasts, the hand-drawn no-fly zones were between 31% smaller and 1,182% larger than the red zone (mean size was 305% larger than the red zone). Only two respondents drew no-fly zones that were within 10% of the size of the red zone for the gap forecast and only three for the solid forecast. Because so few respondents drew no-fly zones that were within 10% of the size of the red zones for both forecasts, we assume that the no-fly zones were intentionally drawn larger than the red zone.

There was little difference in confidence in no-fly zones for the gap than the solid forecasts (Fig. **??**c). The mean confidence in the no-fly zone for the gap forecasts was 5.1 compared with 5.3 for the solid forecasts. Again, the mean confidence was not significantly different between the solid and gap no-fly zones (t-test, $t = 0.3$, $p = 0.73$)

The combination of being conservative in decisions and drawing larger no-fly zones suggests that respondents infer uncertainty in forecasts. This will be discussed further in section 4.

c. How do users make decisions when given conflicting forecasts?

In the third section of the survey, respondents were given conflicting forecasts for the same flight path and asked if they would approve the flight path (Fig. **??**c). Recall that respondents were informed they could fly through blue and grey regions on the map, but not red regions.

When given conflicting forecasts, respondents were overall more risk-averse for the lower-risk decisions (neither forecast indicates the flight path travels through unsafe concentrations, blue–blue and grey–grey) and risk-seeking for the higher-risk decisions (one forecast indicates the flight path travels through unsafe concentrations, red–blue and red–grey, Fig. **??**a) compared with the single forecast decisions from section 1 of the survey. For the lower-risk forecasts, only 52% of respondents would approve the flights in the blue–blue forecast and 52% would approve the flights in the grey–grey forecast (Fig. **??**a), more conservative than when given a single forecast (79% and 61% would approve the low- and medium–low-risk forecasts, Fig. **??**d and c, respectively). For the higher-risk forecasts, 16% would approve the flights in the red–blue forecast and 20% would approve flights in the red–grey forecasts (Fig. **??**a) (26% and 7% would approve the medium–high- and high-risk single forecasts, Fig. **??**a and b, respectively).

Respondents representing operations were more risk-seeking than those in research. For the higher-risk forecasts, 22% of respondents representing operations would approve the flights in the red–blue forecast and 22% in the red–grey forecasts compared with 13% and 19% of those in research, respectively. In the lower-risk forecasts, 67% of respondents in operations would approve the grey–grey forecast compared with 44% of those in research. The only exception was the lowest-risk decision (blue–blue), where 44% of respondents representing operations approved the flight path compared with 56% of those in research (Fig. **??**a), as was the case for single forecasts (see section 3a). The difference in mean decision was not statistically significant at the 5% level (t-test, $t = 0.4$, $p = 0.71$).

Confidence levels were lower for decisions given conflicting forecast information than for those with a single forecast (Fig. **??**e–h compared with Fig. **??**b). For all respondents, mean confidence levels for decisions given conflicting forecasts ranged from 5.7–6.3, compared with 6.2–7.8 for single forecasts. This relationship was not statistically significant, perhaps due to small sample size (t-test, $t = -1.6$, $p = 0.12$). Respondents in operations were more confident in their decisions when given multiple forecasts (mean 6.7–8.4) than those in research (mean 4.6–5.9, Fig. **??**b), as was the case for single forecasts (see section 3a). The difference in mean confidence between operations and research was significant at the 5% level (t-test, $t = 2.7$, $p = 0.01$).

After each conflicting forecast, respondents were asked, in an open-ended format, what information influenced their decisions. When given conflicting forecasts, 64% of respondents indicated they needed more information compared with 52% of respondents in the single forecast decisions. Respondents were then asked what further information they needed to help them make decisions given conflicting forecasts. There were a wide range of suggestions, including observational data, past model performance, meteorological information including wind speed and direction, information on model input, more model ensemble members, information about damage to engines, and uncertainty information. Of the ten respondents requesting uncertainty information for conflicting forecasts, nine represented research.

4. Discussion

The survey results suggested that giving volcanic ash concentration information in addition to the location of the outermost extent of the volcanic ash made the respondents more risk-averse in high-risk decisions and more risk-seeking in low-risk decisions. In an open-ended follow up question, respondents asked for further information more often when only given the outermost extent of the volcanic ash than when provided with ash concentration information. One of the main concerns respondents representing operations raised during the discussion at the end of the workshop was airline traffic disruption due to volcanic ash eruptions. Airlines want to maintain their high levels of safety while reducing the number of flight cancellations and diversions. In that context, our results suggest that providing volcanic ash concentration information is useful to operations because it encourages decisions to fly through safe volcanic ash concentrations and discourages decisions to fly through higher, potentially dangerous volcanic ash concentrations for aircraft. Providing more forecast information (specifically, providing probabilistic forecast information) had similar effects in other decision-making studies (e.g., Joslyn et al. 2007; Nadav-Greenberg and Joslyn 2009; Roulston and Kaplan 2009; Joslyn and LeClerc 2012). Similarly, providing probabilistic contours graphically in tornado warnings increased protective action in the highest probability areas when compared with a polygon only (Ash et al. 2014).

Although ash concentration information seemed to improve the respondents' decisions, providing conflicting volcanic ash concentration forecasts, which can be the case in operations when multiple VAACs are producing forecasts on the same eruption, had the opposite effect. Given two conflicting forecasts, respondents' decisions were more risk-seeking in high-risk situations compared with high-risk decisions given a single forecast. Respondents were also less confident in their decisions when given conflicting forecasts and asked for more information more often than when given a single forecast. However, during the discussion at the end of the workshop, one respondent representing operations said their company uses both the official VAAC forecasts as well as proprietary volcanic ash forecasts. If these two forecasts differ, the respondent said they would only ever increase their no-fly zones, never decrease them. This comment is not supported by the quantitative results from the survey. The action of seeking multiple sources to confirm decisions occurred with stakeholders in flash flooding as well (Morss et al. 2015). Seeking multiple sources, however, puts decision makers at risk of confirmation bias (preferring information that supports their previously held beliefs, e.g., Jonas et al. 2001). Further research into decision making given conflicting information is necessary, especially since stakeholders facing different hazards similarly seek multiple forecasts.

The question of what further information would help decision making given conflicting forecasts yielded a wide range of responses for a small sample of respondents, meaning there is no one-size-fits-all approach to providing volcanic ash information. Thompson et al. (2015), who studied preferences of volcanic hazard map representations for stakeholders in New Zealand, also found that user needs varied widely and one map could not meet all needs. Instead, Thompson et al. (2015) suggest that multiple maps be used that communicate a consistent message in different ways to suit all users' needs.

An additional concern was that respondents were least familiar with, but most trusting in, the satellite graphical representation. The concern with respondents trusting an unfamiliar graphical representation is a lack of knowledge in the ways in which the representation is unreliable. For example, the satellite retrievals are not direct observations; they have been produced by using brightness temperatures from satellite observations and data from numerical weather prediction and radiative transfer models (e.g., Francis et al. 2012). Satellite retrievals may be affected by errors in the models, meteorological cloud, or the angle at which the satellite is viewing the cloud

(Millington et al. 2012). Additionally, satellite retrievals are not forecasts, but instead suggest where ash was located in the past. These locations, of course, can change over time, which is not currently represented in the satellite representation. Lack of understanding of the limitations of the satellite graphic or any other graphical representation may result in poor decision making. Therefore, further training on the limitations of forecasts and the satellite graphic and its shortcomings could be provided to end-users. Providing training on information has previously been shown to help risk judgements (e.g., McCloy et al. 2007).

One suggestion of changes to volcanic ash forecasts, especially from respondents representing research, was to include uncertainty information. Results from the survey indicated that respondents made their own adjustments for uncertainty in the volcanic ash forecasts. For example, the respondents were conservative overall in their decision making, with one-fifth of respondents not approving flight paths through safe levels of volcanic ash concentrations, perhaps inferring uncertainty in the location and concentration of volcanic ash. Additionally, when asked to draw no-fly zones around forecasts, the areas of most respondents' no-fly zones were larger than the areas of unsafe ash concentrations. Similarly, when a non-expert sample made decisions given deterministic rain and temperature forecasts, some took protective action even when the forecast was on the safe side of the threshold, again inferring uncertainty (Morss et al. 2010). Although respondents were told in the survey instructions the levels of ash concentrations considered safe, respondents may have inferred more uncertainty due to the debate over what concentration of volcanic ash was safe during the 2010 Eyjafjallajökull eruption and ongoing research into the effects of volcanic ash on airplane engines (e.g., Clarkson et al. 2016). Respondents may also have inferred uncertainty due to other reasons, such as not trusting the forecast.

One problem with users inferring uncertainty is that there may actually be more or less uncertainty in the forecast depending on the conditions that day than the respondents are assuming. For example, the wide range of sizes of no-fly zones implies there is no universally assumed amount of uncertainty in the forecasts, which could inhibit decision making. This is one explanation for the fact that respondents representing operations were more risk-seeking and confident than those representing research, approving flight paths closer to the center of the ash plume and through higher concentrations of volcanic ash and being more likely to allow flights through the gap between high concentrations of volcanic ash. If the respondents representing operations inferred less uncertainty in the forecast, they would make decisions to fly closer to high concentrations of volcanic ash and be more confident of the boundaries shown in the forecasts. One way to investigate this issue is to test decision making given graphical representations including uncertainty information and to train users on how to interpret such information. Past research suggests that including probabilistic information in forecasts helps decision making (e.g., Roulston and Kaplan 2009; Joslyn and LeClerc 2012; Ash et al. 2014).

Although research indicates that uncertainty information in forecasts helps decision making, respondents in operations stated they do not want uncertainty information. During the discussion at the end of the workshop, respondents were asked if uncertainty information would be useful if provided in volcanic ash forecasts. Respondents in research were keen to provide uncertainty information, which could be possible using ensemble forecasts or emulators of ash dispersion models (Harvey et al. 2016). However, respondents in operations said they preferred deterministic forecasts. One respondent in operations said, “I have a fundamental problem using forecast uncertainty. If the best people in the world (VAACs) are not confident, are you really going to take the risk?” This was verified by an open-ended survey question where all of the respondents who specifically stated that uncertainty information would make them more confident in their decisions were in research. In a separate open-ended question, nine of the ten respondents who stated uncertainty information would help them make decisions given conflicting forecasts were in research.

Because there are so many operational decisions to be made in a short time during a volcanic eruption, respondents in operations were concerned that digesting uncertainty information would take too much time.

Experts in volcanic ash are not the only community to prefer deterministic forecasts. Nobert et al. (2010) found that flood managers also preferred deterministic forecasts, stating that they were not convinced probabilistic information could be made useful. Perhaps providing examples of graphics with uncertainty and practicing implementing them in training on real eruptions in Southeast Asia and Alaska would provide a better understanding of whether uncertainty information would be useful in forecasts and also provide opportunities for verification.

Interestingly, there seemed to be a difference in definition of “uncertainty information” between the respondents in operations and research. When the respondent in operations mentioned that their company paid for a proprietary volcanic ash forecast to compare with the official VAAC forecasts, the researchers in the room interpreted this action as one way to represent uncertainty: by providing multiple outputs for comparison. The operators did not interpret this action as seeking uncertainty information. This suggests there needs to be more conversation and perhaps a different choice of vocabulary when discussing uncertainty between operations and research. Terms such as “probabilistic forecasts”, “multiple model outputs”, and “confidence” might elicit different, more meaningful conversations between the groups than the vague umbrella term, “uncertainty.”

It is important to note that each airline operator is responsible for decision making in volcanic ash eruptions. These decisions are based on the safety risk assessment submitted to the CAA (Safety and Airspace Regulation Group 2014). Any changes to official graphics would require new safety assessments to be conducted by each airline. Therefore, it would take a long time to implement volcanic ash forecasts including uncertainty for the aviation community. This makes volcanic ash and its impact on aviation different from most industries, where communication and decision-making practices can change more quickly.

5. Conclusions

To discuss issues in forecasts and observations of volcanic ash and its effect on aviation, a group of 25 respondents from the United Kingdom representing operations and research were invited to a workshop in London. During the workshop, the respondents completed a survey consisting of numerous decisions given different representations of volcanic ash forecasts. The survey was designed to determine how different graphical representations of volcanic ash forecast affect flight planning decisions, if users infer uncertainty in graphical volcanic ash forecasts, and how flight decisions are made given conflicting volcanic ash forecasts.

When given forecasts containing ash concentration information in addition to the predicted location of the outermost extent of volcanic ash cloud, respondents became more risk-seeking in flight paths further from the center of the ash plume and more risk-averse in flight paths closer to the center of the ash plume. Additionally, fewer respondents mentioned they needed more information to help make their decisions when given the volcanic ash concentration forecasts. Therefore, our results indicated providing ash concentration information seems to encourage better decision making by reducing the number of flight cancellations, delays, and diversions when it is safe to fly. However, the respondents were most trusting in and least familiar with the satellite data, indicating more training is needed on the uses and shortcomings of the satellite representation.

Overall, the respondents were conservative in their decision making, with only 80% of flights through safe concentrations approved given a single forecast and only 50% of flights through safe concentrations approved given conflicting forecasts. In addition, the respondents drew no-fly zones that were larger than the areas of unsafe ash concentrations (no-fly zones drawn by users had means of 243 and 331% larger than the gap and solid forecast unsafe concentration zones, respectively). This implied that the respondents inferred uncertainty in the deterministic volcanic ash forecasts. Respondents representing operations were more risk-seeking and confident than those representing research in their flight decisions, perhaps because the two groups inferred different levels of uncertainty in the forecasts.

When given two conflicting forecasts, respondents became more conservative, being less likely to approve flight paths. However, respondents were more risk-seeking in high-risk flight paths (when one forecast suggested the flight would travel through unsafe concentrations) and more risk-averse in low-risk flight paths (when neither forecast suggested the flight would travel through unsafe concentrations) when given conflicting forecasts compared with single forecasts. Despite this observation, during the discussion following the survey, respondents indicated that when given conflicting information, they only ever increase their no-fly zone or become more risk-averse. This anecdotal evidence contradicts the findings from the survey and indicates inaccurate perception of the process amongst users. Because conflicting forecasts can be present in many natural hazards, further research in decision making given conflicting information is warranted.

There was no one-size-fits-all approach to volcanic ash forecasts, with many different suggestions for additional information to include in the forecasts. When discussing including uncertainty in graphical representations of volcanic ash forecasts, respondents representing operations stated that they only wanted deterministic information, not uncertainty information. However, there seemed to be a difference in the definition of "uncertainty" between the researchers and operations, warranting further conversation and collaboration between the operations and research communities. Continuing this collaboration and encouraging similar collaborations across hazards and user groups will help develop meaningful ways to convert environmental data into information useful to decision makers.

Acknowledgments. We thank two anonymous reviewers for their thoughtful comments that improved this article. We thank the five PhD students from the University of Reading for pretesting the survey, the UK Met Office for comments on the survey, the organizers of the Volcanic Ash Workshop, and the workshop delegates for participating. This research is funded by the National Environmental Research Council (NERC) under the Probability, Uncertainty and Risk in the Environment (PURE) Programme (NE/J017221/1). Data created during the research reported in this article are openly available from the University of Reading Research Data Archive at http://dx.doi.org/10.17864/1947.62.

References

Abraham, S., R. Bartlett, M. Standage, A. Black, A. Charlton-Perez, and R. McCloy, 2015: Do location-specific forecasts pose a new challenge for communicating uncertainty? *Met. App.*, **22**, 554–562, doi:10.1002/met.1487.

Aerts, J. C. J. H., K. C. Clarke, and A. D. Keuper, 2003: Testing popular visualization techniques for representing model uncertainty. *Cartography and Geographic Information Science*, **30**, 249–261, doi:10.1559/152304003100011 180.

Ash, K. D., R. L. Schumann III, and G. C. Bowser, 2014: Tornado warning trade-offs: Evaluating choices for visually communicating risk. *Weather, Climate, and Society*, **6**, 104–118, doi:10.1175/WCAS–D–13–00 021.1.

Broad, K., A. Leiserowitz, J. Weinkle, and M. Steketee, 2007: Misinterpretations of the "cone of uncertainty" in Florida during the 2004 hurricane season. *Bull. Amer. Meteor. Soc.*, **88**, 651–667, doi:10.1175/BAMS–88–5–651.

Clarkson, R. J., E. J. E. Majewicz, and P. Mack, 2016: A re-evaluation of the 2010 quantitative understanding of the effects volcanic ash has on gas turbine engines. *J. Aerospace Engineering*, **In Press**, doi: 10.1177/0954410015623 372.

Correll, M., and M. Gleicher, 2014: Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions of Visualization and Computer Graphics*, **20**, 2142–2151.

Cox, J., D. House, and M. Lindell, 2013: Visualizing uncertainty in predicted hurricane tracks. *International Journal for Uncertainty Quantification*, **3**, 143–156.

Doyle, E. E. H., J. McClure, D. M. Johnston, and D. Paton, 2014: Communicating likelihoods and probabilities in forecasts of volcanic eruptions. *Journal of Volcanology and Geothermal Research*, **272**, 1–15, doi:10.1016/j.jvolgeores.2013.12.006.

Ekker, K., G. Gifford, S. A. Leik, and R. K. Leik, 1988: Using microcomputer game-simulation experiements to study family response to the Mt. St. Helens eruptions. *Social Science Computer Review*, **6**, 90–105.

Englich, B., T. Mussweiler, and F. Strack, 2006: Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, **32**, 188–200, doi:10.1177/0146167205282 152.

European Commission, 2016: Eurostat: Air passenger transport by reporting country. Tech. rep., Accessible online: http://appsso.eurostat.ec.europa.eu/nui/show.do.

Francis, P. N., M. C. Cooke, and R. W. Saunders, 2012: Retrieval of physical properties of volcanic ash using Meteosat: A case study from the 2010 Eyjafjallajökull eruption. *J. Geophys. Res.*, **117**, 1–14, doi:10.1029/2011JD016 788.

Harvey, N., N. Huntley, H. Dacre, M. Goldstein, D. Thomson, and H. Webster, 2016: Multilevel emulation of a volcanic ash transport and dispersion model to quantify sensitivity to uncertain parameters. *Nat. Haz. Earth. Syst. Sci.*, doi:10.5194/nhess–2016–288.

Ibrekk, H., and M. G. Morgan, 1987: Graphical communication of uncertain quantities to non-technical people. *Risk Anal.*, **7**, 519–529.

Jonas, E., S. Schultz-Hardt, D. Frey, and N. Thelen, 2001: Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *J. Person. and Soc. Psych.*, **80**, 557–571, doi:10.103700 022–3514.80.4.557.

Joslyn, S., K. Pak, D. Jones, J. Pyles, and E. Hunt, 2007: The effect of probabilistic information on threshold forecasts. *Wea. Forecasting*, **22**, 804–812, doi:10.1175/WAF1020.1s.

Joslyn, S., S. Savelli, and L. Nadav-Greenberg, 2011: Reducing probabilistic weather forecasts to the worst-case scenario: Anchoring effects. *J. Experim. Psychol. Applied*, **17**, 342–353, doi:10.1037/a0025 901.

Joslyn, S. L., and J. E. LeClerc, 2012: Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *J. Experimental Psychology: Applied*, **18**, 126–140, doi:10.1037/a0025 185.

Keller, C., M. Siegrist, and H. Gutscher, 2006: The role of the affect and availability heuristics in risk communication. *Risk Analysis*, **26**, 631–639, doi:10.1111/j.1539–6924.2006.00 773.x.

Kirschenbaum, S. S., and J. E. Arruda, 1994: Effect of graphic and verbal probability information on command decision making. *Human Factors*, **36**, 406–418.

Lindell, M. K., S. Huang, H. Wei, and C. D. Samuelson, 2016: Perceptions and expected immediate reactions to tornado warning polygons. *Natural Hazards*, **80**, 683–707, doi: 10.1007/s11 069–015–1990–5.

Mazzocchi, M., F. Hansstein, and M. Ragona, 2010: The volcanic ash cloud and its financial impact on the European airline industry. *CESifo Forum*, **2**, 92–100.

McCloy, R., C. P. Beaman, B. Morgan, and R. Speed, 2007: Training conditional and cumulative risk judgments: The role of frequencies, problem-structure, and Einstellung. *Appl. Cognit. Psych.*, **21**, 325–344, doi:10.1002/acp.1273.

Millington, S. C., R. W. Saunders, P. N. Francis, and H. N. Webster, 2012: Simulated volcanic ash imagery: A method to compare NAME ash concentration forecasts with SEVIRI imagery for the Eyjafjallajökull eruption in 2010. *J. Geophys. Res.*, **117**, 1–14, doi:10.1029/2011JD016 770.

Morss, R. E., J. L. Demuth, A. Bostrom, J. K. Lazo, and H. Lazrus, 2015: Flash flood risks and warning decisions: A mental models study of forecasters, public officials, and media broadcasters in Boulder, Colorado. *Risk Analysis*, **35**, 1–20, doi: 10.1111/risa.12 403.

Morss, R. E., J. K. Lazo, and J. L. Demuth, 2010: Examining the use of weather forecasts in decision scenarios: Results from a US survey with implications for uncertainty communication. *Met. App.*, **17**, 149–162, doi:10.1002/met.196.

Morss, R. E., K. J. Mulder, J. K. Lazo, and J. L. Demuth, 2016: How do people perceive, understand, and anticipate responding to flash flood risks and warnings? Results from a pubilc survey in Boulder, Colorado, USA. *J. Hydrology*, **541**, 649–664, doi:10.1016/j.jhydrol.2015.11.047.

Nadav-Greenberg, L., and S. L. Joslyn, 2009: Uncertainty forecasts improve decision making among nonexperts. *J. Cog. Eng. and Dec. Making*, **3**, 209–227, doi: 0.1518/155534 309X474 460.

Nadav-Greenberg, L., S. L. Joslyn, and M. U. Taing, 2008a: The effect of uncertainty visualizations on decision making in weather forecasting. *Journal of Cognitive Engineering and Decision Making*, **2**, 24–47, doi:10.1518/155534 308X284 354.

Nadav-Greenberg, L., S. L. Joslyn, and M. U. Taing, 2008b: The effect of uncertainty visualizations on decision making in weather forecasting. *J. Cog. Eng. and Dec. Making*, **2**, 24–47, doi:10.1518/155534 308X284 354.

Nagele, D. E., and J. E. Trainor, 2012: Geographic specificity, tornadoes, and protective action. *Weather, Climate, and Society*, **4**, 145–155, doi:10.1175/WCAS–D–11–00 047.1.

Nobert, S., D. Demeritt, and H. Cloke, 2010: Informing operational flood management with ensemble predictions: Lessons from Sweden. *J. Flood Risk Management*, **3**, 72–79, doi:10.1111/j.1753–318X.2009.01 056.x.

Peters, E., D. Västfjäll, P. Slovic, C. K. Mertz, K. Mazzocco, and S. Dickert, 2006: Numeracy and decision making. *Psych. Sci.*, **17**, 407–413.

Riveiro, M., T. Helldin, G. Falkman, and M. Libram, 2014: Effects of visualizing uncertainty on decision-making in a target identification scenario. *Computers and Graphics*, **41**, 84–98, doi:10.1016/j.cag.2014.02.006.

Roulston, M. S., and T. R. Kaplan, 2009: A laboratory-based study of understanding of uncertainty in 5-day site-specific temperature forecasts. *Met. App.*, **16**, 237–244, doi:10.1002/met.113.

Ruginski, I. T., and Coauthors, 2015: Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition & Computation*, **16**, 154–172.

Safety and Airspace Regulation Group, 2014: Guidance regarding flight operation in the vicinity of volcanic ash. Tech. rep., Civil Aviation Authority, Accessible online: https://publicapps.caa.co.uk/docs/33/CAP%201236%20final%20version.pdf.

Sherman-Morris, K., and M. E. Brown, 2012: Experiences of Smithville, Mississippi residence with the 27 April 2011 tornado. *National Weather Digest*, **36**, 93–101.

St. John, M., J. Callan, S. Proctor, and S. T. Holste, 2000: Tactical decision-making under uncertainty: Experiments I and II. Tech. Rep. 1821, Space and Naval Warfare Systems Center, San Diego, CA, USA.

Tak, S., A. Toet, and van Erp J., 2015: Public understanding of visual representations of uncertainty in temperature forecasts. *J. Cog. Eng. and Dec. Making*, **9**, 241–262, doi:10.1177/1555343415591 275.

Taylor, H. A., C. E. Renshaw, and M. D. Jensen, 1997: Effects of computer-based role-playing on decision making skills. *Journal of Educational Computing Research*, **17**, 147–164.

Thompson, M. A., J. M. Lindsay, and J. C. Gaillard, 2015: The influence of probabilistic volcanic hazard map properties on hazard communication. *J. Applied Volcan.*, **4**, 1–24, doi:10.1186/s13 617–015–0023–0.

Thordarson, T., and G. Larsen, 2007: Volcanism in Iceland in historical time: Volcano types, eruption styles and eruptive history. *J. Geodynamics*, **43**, 118–152.

Webster, H. N., and Coauthors, 2012: Operational prediction of ash concentrations in the distal volcanic ash cloud from the 2010 Eyjafjallajökull eruption. *J. Geophys. Res.*, **117**, 1–17, doi:10.1029/2011JD016 790.

Whyte, G., and J. K. Sebenius, 1997: The effect of multiple anchors on anchoring in individual and group judgment. *Organizational Behavior and Human Decision Processes*, **69**, 75–85.

Wickens, C. D., K. Gempler, and M. E. Morphew, 2000: Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors*, **2**, 99–126.

Wu, H., M. K. Lindell, and C. S. Prater, 2015: Process tracing analysis of hurricane information displays. *Risk Anal.*, **35**, 2202–2220, doi:10.1111/risa.12 423.

Wu, H., M. K. Lindell, C. S. Prater, and C. D. Samuelson, 2014: Effects of track and threat information on judgments of hurricane strike probability. *Risk Anal.*, **34**, 1025–1039, doi:10.1111/risa.12 128.

LIST OF FIGURES

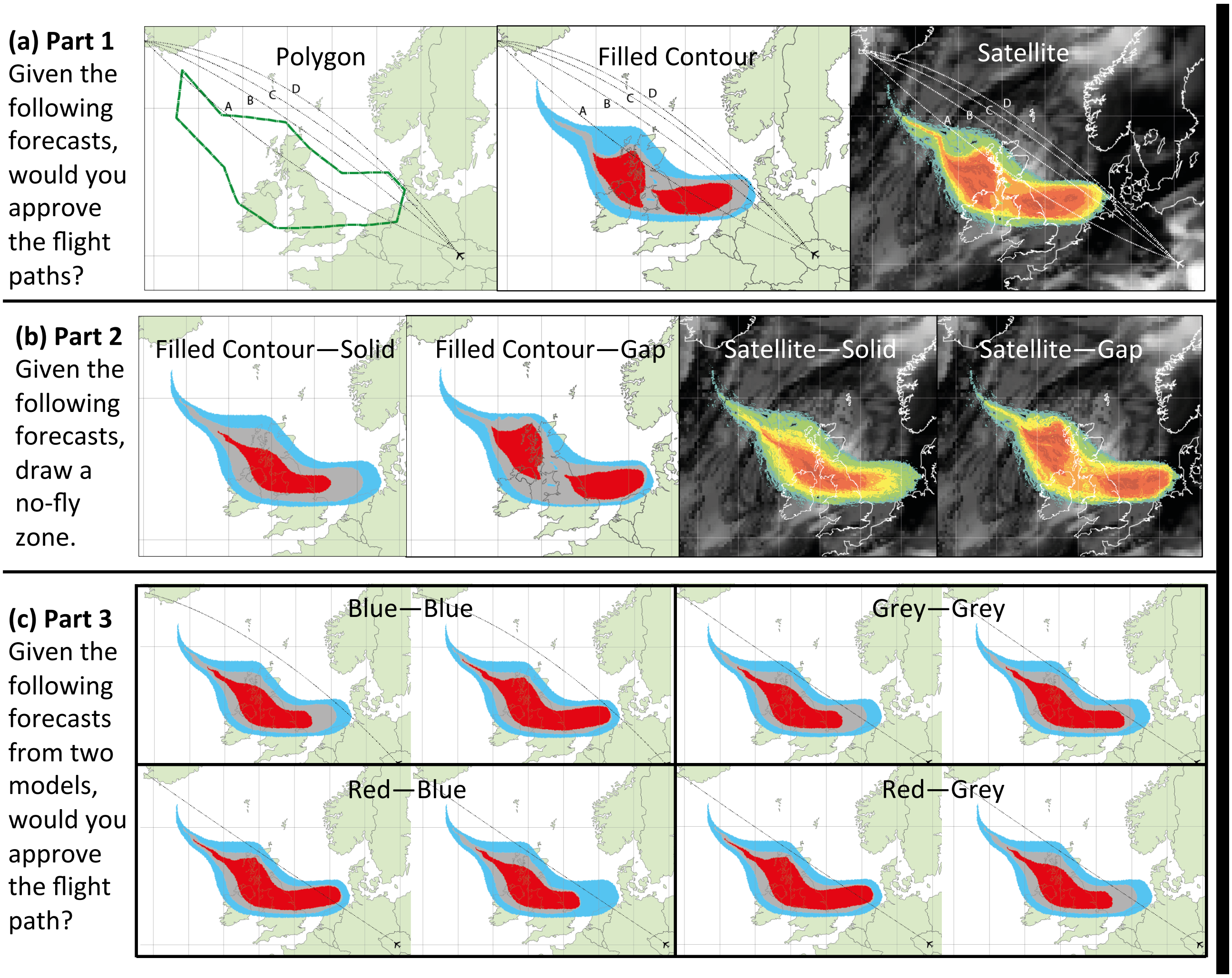


FIG. 1. Survey questions and graphical representations used for decision making for (a) part 1, (b) part 2, and (c) part 3 of the survey. (a) The same four flight paths were overlaid onto the polygon, filled contour, and satellite representations of the same volcanic ash forecast. Respondents were asked if they would approve each forecast and their confidence in their decisions. (b) Two forecasts (solid and gap) were represented in two ways (filled contour and satellite). Respondents were asked to draw a no-fly zone on the forecasts and their confidence in their no-fly zones. (c) Respondents were given conflicting forecasts for the same flight path and were asked if they would approve each forecast and their confidence in their decisions. The flight paths went through the following colored concentration contours: blue–blue, grey–grey, red–blue, or red–grey. For all figures, respondents were told it was safe to fly through medium concentrations of volcanic ash (2000–4000 $\mu g\ m^{-3}$) corresponding to the blue and grey areas in the filled contour representation and the green, yellow, and orange areas in the satellite representation

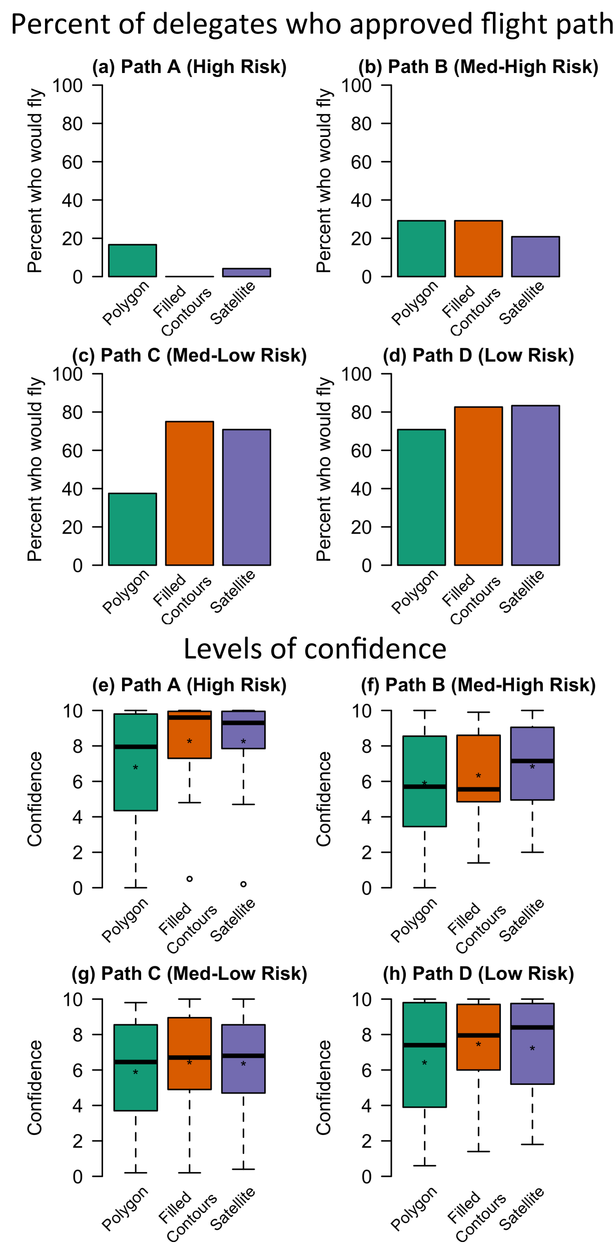
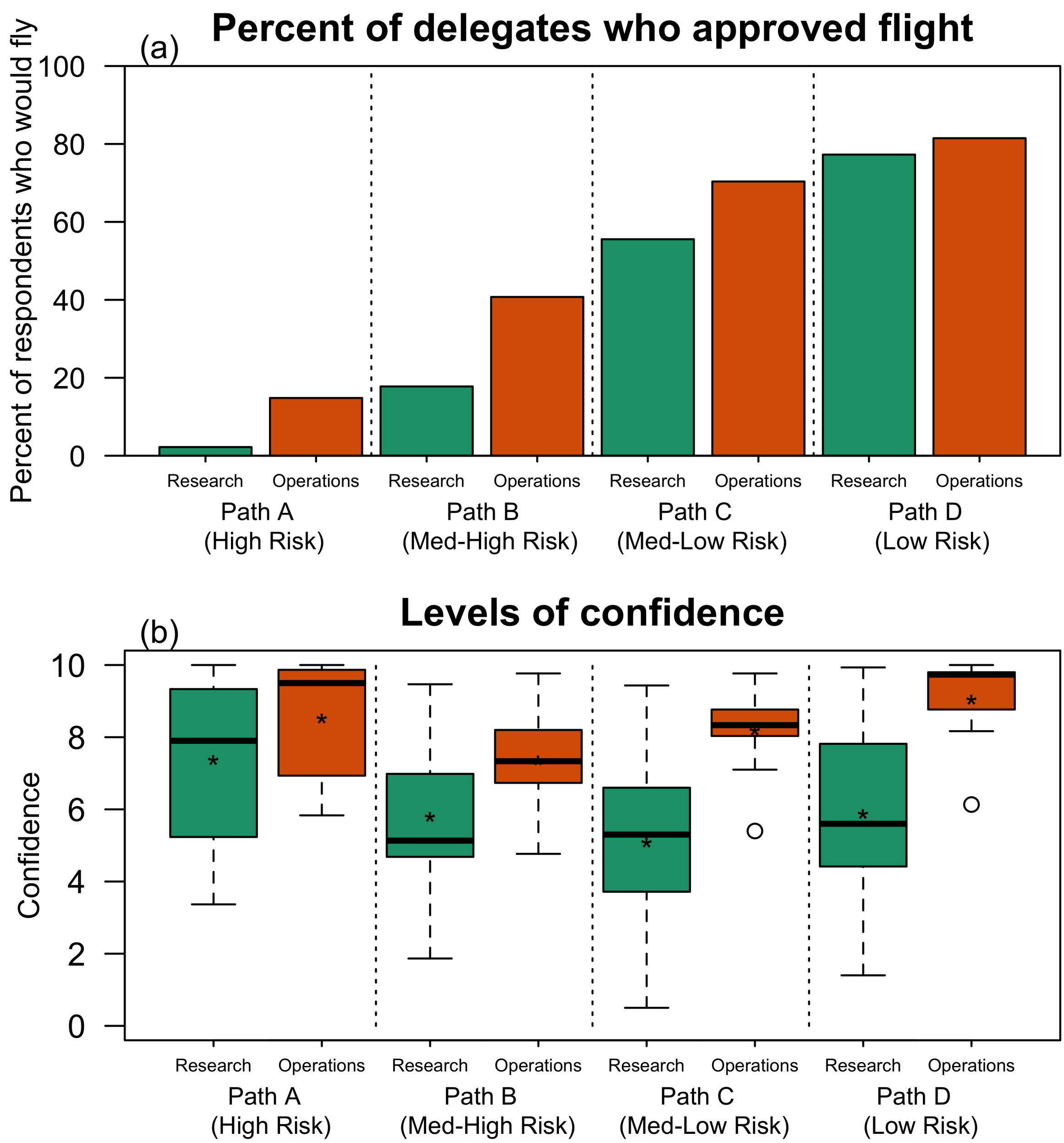


FIG. 2. Percent of respondents who approved flight [(a), (b), (c), and (d)] and levels of confidence [(e), (f), (g), (h)] for different flight paths by graphical representation. The polygon, filled contour, and satellite graphical representations are shown as green, red, and indigo, respectively. Path A (High Risk) is shown in (a) and (e); Path B (Medium–High Risk) is shown in (b) and (f); Path C (Medium–Low Risk) is shown in (c) and (g); Path D (Low Risk) is shown in (d) and (h). Graphical representations used for this section of the survey are shown in Fig. **??**a. Levels of confidence are rated on a scale from 0 (“Not at all confident”) to 10 (“Extremely confident”). The upper and lower whiskers represent the maximum and minimum values, respectively. The top and bottom of the box represent the 75th and 25th percentiles, respectively. The bar in the box represents the median. The star represents the mean. Circles on either side of the whiskers are outliers.



782 FIG. 3. (a) Percent of respondents who approved flight and (b) levels of confidence for different flight paths by occupation in either research (green) or operations (red). Graphical representations used for this section of the survey are shown in Fig. **??**a. Levels of confidence are rated on a scale from 0 ("Not at all confident") to 10 ("Extremely confident"). The box plot (b) is formatted as in Fig. **??**.

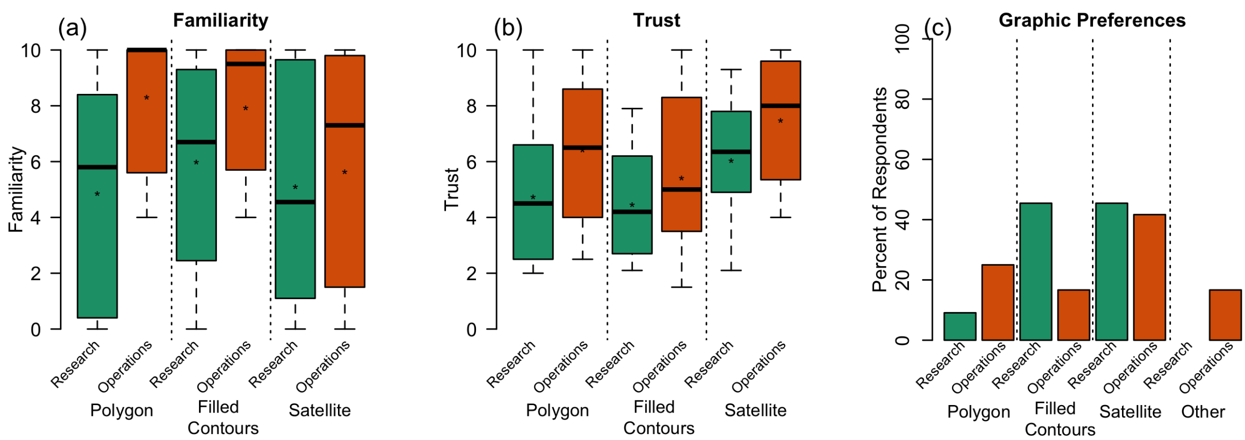


FIG. 4. (a) Familiarity with, (b) trust in, and (c) preferences of graphical representations by occupation in either research (green) or operations (red). Levels of familiarity and trust are rated on a scale from 0 ('Never seen before" or "Not at all trustworthy") to 10 ("Have seen frequently" or "Extremely trustworthy"). The box plots are formatted as in Fig. **??**.

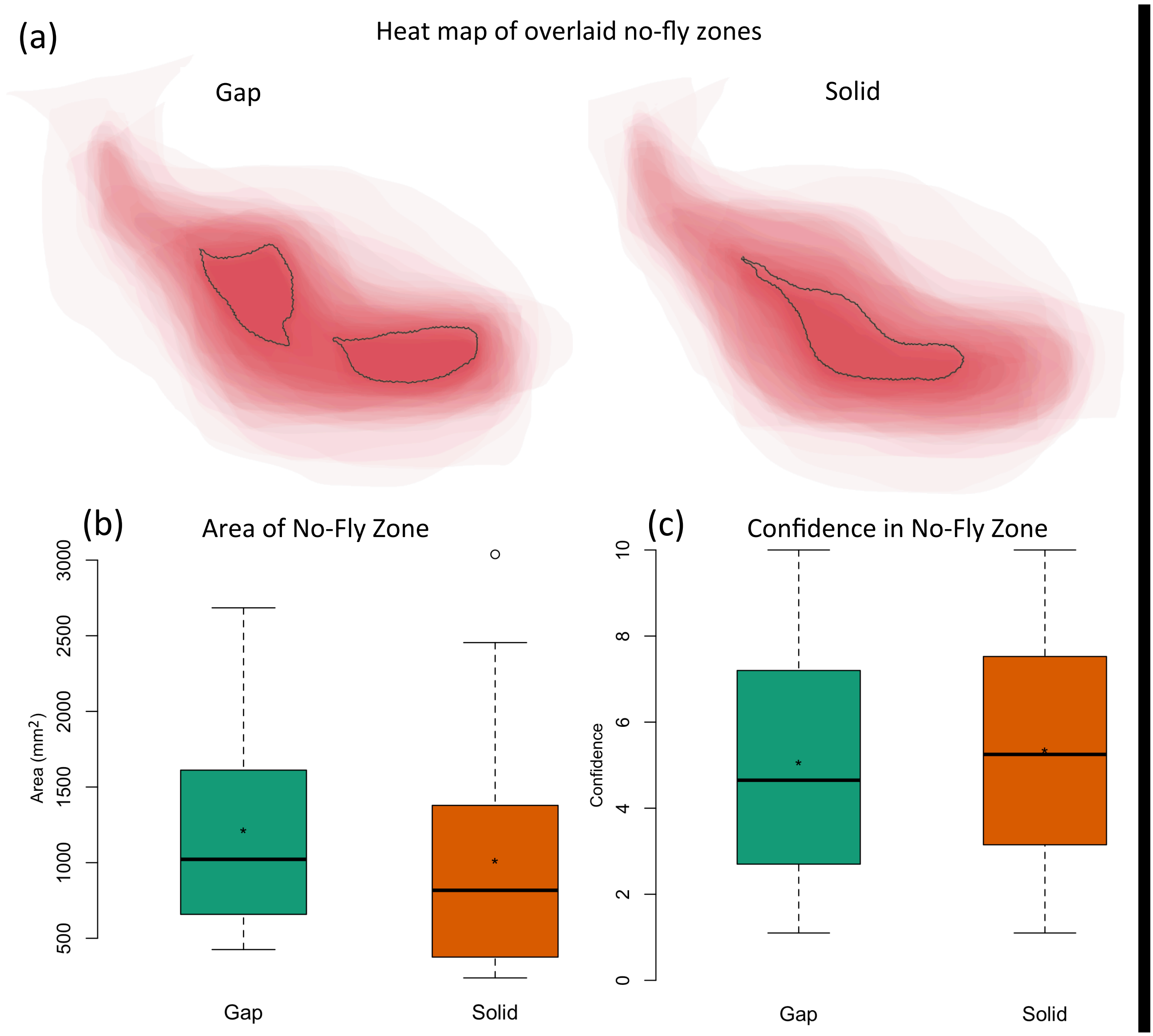


FIG. 5. (a) Heat map showing overlaid no-fly zones drawn by the respondents for the gap and solid forecasts. Darker colors indicate more respondents drawing a no-fly zone over that area. The black outlines show where respondents were told it was unsafe to fly. (b) Calculated areas of the no-fly zones in square millimeters by forecast type of either gap (green) or solid (red) (c) Levels of confidence in the no-fly zones by forecast type of either gap (green) or solid (red). Graphical representations used for this section of the survey are shown in Fig. **??**b. The box plots are formatted as in Fig. **??**.

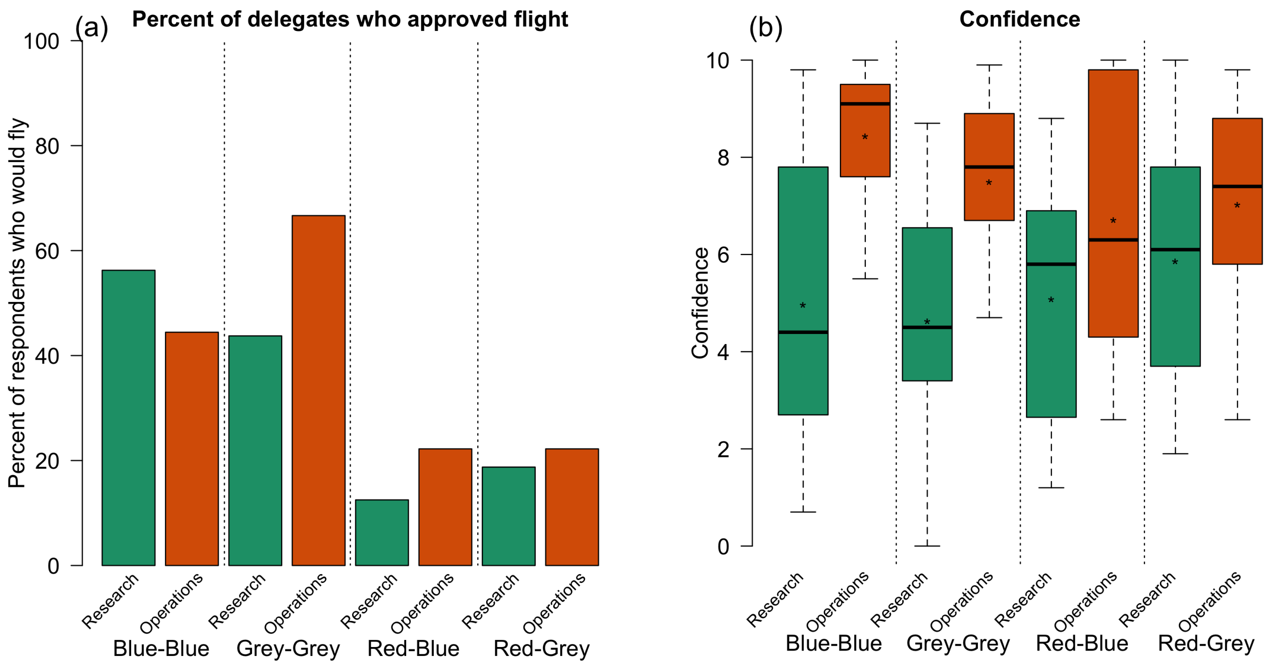


FIG. 6. (a) Percent of respondents who approved flight and (b) levels of confidence for flight paths given conflicting forecasts by occupation in either research (green) or operations (red). Graphical representations used for this section of the survey are shown in Fig. **??**c. Levels of confidence are rated on a scale from 0 ("Not at all confident") to 10 ("Extremely confident"). The box plot (b) is formatted as in Fig. **??**.