1 Introduction

Volcanic eruptions that spread out ash over large areas can have tremendous economic consequences, although they are relatively rare compared to other high-impact natural hazards (such as tropical cyclones, storms, etc). For instance, the eruption of Eyjafjallajökull in 2010 forced the cancellation of about 100 000 flights and generated a 1.4 billion Euro loss to the airline operators (IATA, 2010). Considering the different levels of risks to safety when an aircraft encounters ash, including failure of aircraft turbines during operation (Guffanti et al., 2010; Alexander, 2013), flight cancellations and re-routings out of the ash-contaminated areas were the most common decision during this eruption.

In order to mitigate the consequences of such types of volcanic eruptions on aviation, operational centres continuously watch and issue warnings about ash dispersion in the atmosphere, that support the decisions in the frame of predefined procedures (Bolić and Sivčev, 2011). This watching and warning role worldwide has been the duty of Volcanic Ash Advisory Centres (VAACs). Following the consequences of the Eyjafjallajökull eruption in Europe, the London and Toulouse VAACs procedures have changed and they now also provide concentration charts, in the Flight Level (FL) bands FL000-200, FL200-350, FL350-550, for three contamination levels: ≤ 2 mg m^{-3} (low contamination), >2 mg m^{-3} and <4 mg m^{-3} (medium contamination), and ≥ 4 mg m^{-3} (high contamination) (ICAO, 2016). These volcanic ash contamination charts (up to 18 hours ahead) indicate hazardous zones and hazard levels, which can be used by authorities for flight safety.

Volcanic ash warnings and charts are based on the outputs from numerical prediction models. As a consequence, they are usually prone to large errors and uncertainties (Kristiansen et al., 2012; Dacre et al., 2016), which arise from the uncertainties in the ash source term, in the modelling of transport including meteorology, and the parameterisation of physical processes. Although these components have been improving thanks to active research (Beckett et al., 2020), it is highly probable that predictions with sufficient accuracy cannot be reached in a near future. The ash source term, i.e., the temporal evolution of volcanic ash mass emitted by the volcano at every vertical level and distributed over aerosol size groups, cannot be fully observed, and even after inversion of satellite observations some error and uncertainty remain (Kristiansen et al., 2012). Aerosol processes (sedimentation, wash and rain out, aggregation in the presence of liquid or solid water) and aerosol transport depend on the aerosols' representation in the model, and also on the meteorological conditions. The meteorological forecasts, for which error grows inevitably with time (Dacre et al., 2016) and which are essential input information for ash dispersion forecasts, also contribute significantly to uncertainties in ash forecasts. In order to account for the uncertainty of volcanic ash forecasts, some studies have already shown the added value of ensembles (Kristiansen et al., 2012). They recommended to use probabilistic forecasts in the decision process (Prata et al., 2019), in a similar manner as meteorological probabilistic weather forecasts, which have shown to have a large benefit compared to deterministic forecasts (Richardson, 2000; Osinski and Bouttier, 2018; Fundel et al., 2019).

Proof-of concept studies for flight planning are in progress (Steinheimer et al., 2016), to use probabilistic meteorological forecasts to support aviation safety, capacity, and cost-efficiency. Regarding volcanic ash hazard, interest for ensemble prediction is also raising among the meteorological operational community (such as VAACs) for ensemble prediction. Beyond the important safety aspect, one potential application of probabilistic ash forecasts is related to ash dosage (i.e., the accumulated mass of ash encountered by the aircraft along its track), which is an important parameter to characterize the impact of ash on aircraft engines (Clarkson et al., 2016). While medium ash concentration over long time can also be a safety issue, low ash dose can lead to longer term damage, which becomes an issue for engine maintenance as shorter intervals are needed in order to prevent performance loss (Clarkson et al., 2016). Contaminated regions can be avoided by flight rerouting but this increases costs due to delays and additional fuel. As a consequence, a cost/loss rationale (Richardson, 2000) can be considered in regions where ash concentrations remain below the safety margins, and in that sense, probabilistic ash forecasts (Prata et al., 2019) could be used to make optimal decisions for air traffic management (ATM). Such applications require good estimates of ash concentrations in 4 dimensions (3D space and time) as well as their uncertainty. Cost-optimized consideration of such hazards could result in a significantly reduced impact on flight cancellations, rerouting, and traffic flow congestion during volcanic ash events (Rokitansky et al., 2019).

In the European Natural Airborne Disaster Information and Coordination System for Aviation (EUNADICS-AV) project, a multi-model approach has been developed and assessed on several test cases. The outputs from several models were collected to build a mini-ensemble and probabilistic charts of ash concentrations. This ensemble has a $0.1°$ horizontal resolution on a large Euro-Atlantic domain, and provides information on 13 vertical FLs. The integration of these data into a flight-planning software and their relevance for ATM were shown during an exercise simulating a fictitious crisis situation (Hirtl et al., 2020).

The purpose of the present article is to provide more precise understanding of the performance and benefit of the multi-model multi-source-term ensemble approach developed during the EUNADICS-AV project, using measurements as reference for comparison. The performance of individual model runs performed with four models using four different source terms each, is also evaluated to better understand ensemble characteristics. The uncertainty of the meteorological conditions was not taken into account, i.e., meteorological analyses were concatenated with short-term forecasts.

The study focuses on a particular period of the Eyjafjallajökull eruption from 13 to 20 May 2010, when ash spread across the North Sea and the Atlantic Ocean, and then over continental Europe. During this period, the amount of measurements was particularly high compared to other phases of the eruption. Dacre et al. (2016) pointed out a low predictability of the dispersion of ash during this period and they studied how the error grew with time. Their study emphasized the need for ensemble approaches in order to deal with uncertainty. For the same phase of the eruption, Kristiansen et al. (2012) compared two different models with different source terms. They showed overall a good agreement between models, and that the ensemble obtained as the mean ash concentration of the different models (a mean-ensemble) usually, but not generally, outperforms any of the models.

The outline of the present article is as follows: Section 2 gives a description of the models, the source terms, and the reference data. Section 3 compares the model outputs and evaluates them against reference observations. Section 4 presents how mini-ensembles are built and compares them to reference observations as well. Section 5 discusses how such an ensemble can be used in flight planning for future eruptions. Conclusions are drawn in Section 6.