



Using multiple contrail models for decision making could increase efficacy of warming-based contrail avoidance strategies

Caleb Akhtar Martínez^{1,2,3}, Jerome P. Jarrett³, Sebastian D. Eastham^{1,2}

¹Brahmal Vasudevan Institute for Sustainable Aviation, Imperial College London, London, SW7 2AZ, United Kingdom

5 ²Department of Aeronautics, Imperial College London, London, SW7 2AZ, United Kingdom

³Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, United Kingdom

Correspondence to: Caleb Akhtar Martínez (c.akhtarmartinez@imperial.ac.uk)

Abstract.

Contrail avoidance through flight deviations is thought to be the fastest way to reduce aviation-related warming. Nevertheless, uncertainty in contrail model accuracy makes it difficult to prioritize contrails for avoidance according to their predicted warming. To minimise the effect of model disagreement on contrail avoidance decisions, we reformulate the selection of candidate contrails for avoidance as an aggregation problem – a family of problems enabling joint decisions by several agents (contrail models) with different preferences. We compare several aggregation rules to determine whether using multiple contrail models can increase savings in the lifetime total extinction when compared to just using the most popular model, CoCiP. In the absence of a ground truth, we use the lifetime total extinction from a high-fidelity large eddy simulation (LES) as a proxy for the real warming. We test these aggregation strategies with two real models, and synthetic contrail model results (produced by applying a chosen functional bias and random noise to the reference results). For interpretability, we describe a contrail model as “moderately correlated” if its Pearson correlation coefficient relative to the reference lies between 0.50 and 0.60. We find that two separate aggregation strategies using two moderately correlated contrail models is expected to avoid 20 % more lifetime total extinction relative to using CoCiP by itself. This improvement rises to 33 % when using three moderately correlated contrail models. Since two aggregation strategies with different working principles achieve similar reductions in the lifetime total extinction, they may improve the efficacy of warming-based contrail avoidance.

1 Introduction

Given the ongoing climate emergency and the short remaining window to mitigate its impact, it is imperative to take rapid action to achieve a timely reduction of anthropogenic warming. However, due to increasing demand and stagnating reductions in emissions, the aviation industry increasingly burdens the climate. The approximately equal climate contribution of aviation carbon dioxide and contrails (Lee et al., 2021) makes both decarbonisation and contrail mitigation crucial. However, unlike decarbonisation, contrail mitigation through in-flight rerouting does not require changes to the aircraft configuration or airport infrastructure, making it one of the fastest ways to reduce the climate impact of aviation (Smith et al., 2026).



30 Contrail avoidance strategies can be organised into one of two categories: randomised strategies and warming-based strategies. Randomised strategies randomly select contrails to avoid, and so they are agnostic to the estimated contrail warming (e.g. Avila et al., 2019; Sonabend-W. et al., 2024). Warming-based strategies determine which contrails to avoid according to the simulated warming of potential contrails (e.g. Teoh et al., 2020a; Teoh et al., 2020b; Dean et al., 2025; Elmourad, 2023; Martín Frías et al., 2024).

35 Despite the necessity to eliminate contrails, both types of avoidance strategies are subject to risks of exacerbating global warming. First, it is possible that any contrail avoidance strategy results in a warming increase if the fuel burn penalties from avoidance moves cumulatively outweigh the warming reduction from avoiding the contrail. In a recent study by Smith et al. (2026) it was estimated that this risk is low (a 9 % chance for an increase in the temperature by the year 2100 even in the most pessimistic model sets assuming a fleet-wide fuel burn penalty of 10 %), so it is excluded from further discussion. The second

40 type of risk only afflicts warming-based contrail avoidance. Consider a contrail model that has a contrarian bias, such that cooling contrails are predicted to be warming and vice-versa. In such a case, it is possible for cooling contrails to be avoided systematically, resulting in a net detriment to the climate. The randomised contrail avoidance strategy circumvents this by targeting the same proportion of all types of contrails. Hence, although warming-based strategies have the potential to be more effective than the randomised strategy, their safe use requires good understanding of the behaviour and limitations of the

45 contrail models employed in them, including performance evaluations against reference validation data. Obtaining validation data for contrail models is challenging due to the difficulties associated with observing the full contrail lifetime. Contrails begin as narrow, line-shaped clouds and, if they persist for long enough, they can become widespread, amorphous, and dispersed, making them hard to distinguish from natural cirrus (Kärcher, 2018). Considering the practical limitations of ground-based sensing techniques (e.g. Schumann et al., 2013) and remote sensing techniques (Euchenhofer et

50 al., 2025; Driver et al., 2025), no single contrail observation technique can track the contrail properties from inception to demise. Furthermore, meteorological uncertainties likely place a fundamental limit to the accuracy of the warming predictions made through modelling, even when modelling uncertainties are accounted for (Xu, 2024) or eliminated (Wilhelm et al., 2021). Thus, time-evolving data of different contrails (referred to as *ground truth*), and not just single snapshots in time, are necessary to validate the evolution of modelled contrails by constraining their uncertainty ranges when evaluated on the reference cases.

55 Although it is not yet possible to obtain a ground truth dataset from current observational techniques, contrail model intercomparisons are starting to emerge in the literature. Akhtar Martínez et al. (2025), showed that the open-source contrail models CoCiP and APCEMM disagreed in their predictions of lifetime total extinction (a proxy for climate impact) by a factor of 3.8 under equivalent meteorology. A further intercomparison by Akhtar Martínez et al. (2026b; preprint) indicates that there is disagreement between contrails simulated in APCEMM and equivalent contrails simulated in the large eddy simulation

60 (LES) model EULAG-LCM. Overall, these intercomparisons reveal that there may be irreconcilable differences between contrail plume models of different fidelities and scales.

Contrail model disagreement may have adverse effects on contrail avoidance. At worst, contrail model disagreement may act as a barrier to the deployment of regulatory measures necessary to perform contrail avoidance (due to the low apparent level

of scientific understanding). At best, contrail model disagreement could limit the efficacy of contrail avoidance. Circumventing this dilemma by discarding all but a single contrail model is not advisable because this only hides the modelling uncertainties instead of reducing them.

As far as we know, no solutions have been proposed in the literature to make warming-based avoidance decisions under model disagreement. We aim to address this by considering the entire set of candidate contrails for avoidance simultaneously (avoidance on an ensemble basis) instead of considering them individually (avoidance on a case-by-case basis). This allows for the normalisation of the solution space for each model, after which the candidate contrails can be compared across models. By then treating the contrail models as decision-making agents with preferences, group decisions may be reached by using aggregation strategies. The main objective of this study is therefore to determine whether contrail avoidance decisions can be improved by using aggregation strategies when compared to contrail avoidance relying on a single contrail model.

1.1 Contrail avoidance and decision-making under uncertainty

Consider a scenario in which the air traffic management (ATM) workload remains high. In it, avoidance through flight deviation would only be performed for a specific proportion (α) of flight segments predicted to lie in contrail-likely zones (CLZs). We refer to contrail avoidance on an *ensemble basis* as the set of warming-based decision-making strategies in which all flight segments identified to lie within CLZs are considered for re-routing simultaneously, and only α are selected.

Individually, each of the models will produce a different subset of flights recommended for avoidance according to their simulated warming independently of the diagnostic probabilities associated with contrail avoidance (see Meijer, 2024). Hence, an aggregation strategy that can robustly select the most impactful flights from the predicted warming (as evaluated by all the models) is required. This is an example of an aggregation problem.

Aggregation problems are those problems in which group recommendations must be issued by considering the individual recommendations from numerous stakeholders, sensors, or models (hereafter agents), and they are relevant to a variety of disciplines (e.g. economics, philosophy, political science, and computer science; Grossi and Pigozzi, 2014). *Judgement aggregation* involves making a collective judgement from individual judgements produced by each agent (Grossi and Pigozzi, 2014). Similarly, *preference aggregation* involves creating a list of preferred candidates based on the preferences of the individual agents (Grossi and Pigozzi, 2014). In simple terms, aggregation strategies can be thought to work by attempting to find regions of the evaluation space (the vector space containing all possible evaluations of the available proposals) in which the different agents agree, discarding some potential outliers, hence reducing epistemic uncertainties.

Aggregation strategies have been used in engineering contexts. For example, voting has been used in sensor fusion to produce error-resilient collective outcomes given incomplete and imperfect information from multiple sensors (e.g. Urken, 2005; Montijano et al., 2011). This example reflects how aggregation strategies can be used to mitigate epistemic uncertainties in a computationally efficient manner when multiple agents are available – making it particularly relevant to the problem of selecting contrails to avoid under model disagreement. Although aggregation strategies have seen success in the engineering literature, they are not necessarily guaranteed to produce an improvement in the collective recommendations (see Section 5.2).



Nevertheless, their low computational cost, traceability, and simplicity are sufficient to motivate their analysis in the context of contrail avoidance.

Due to the difficulties involved in obtaining ground truth data, we consider a future model set in which whole-lifetime validation data are available for a small number of contrails. Hence, deep learning approaches for decision-making under uncertainty are not considered due to their poor suitability for training with sparse data. Another reason to avoid deep learning methods is that they often behave as logic black boxes. This makes traceability of the decisions at best difficult and at worst impossible. Aggregation strategies do not have these drawbacks, justifying their use for the mitigation of epistemic uncertainties in contrail avoidance.

Increasingly, approaches in the contrail literature attempt to capture variability in meteorological model outputs by considering the spread of ensembles produced by weather models and climate models. For example, Hanst et al. (2025) define the ice supersaturated region (ISSR) forecast probability as the fraction of the ICON model ensemble members predicting an ISSR. Hanst et al. (2025) also show that, when creating an ISSR mask from the ensemble, decision-making using probabilistic forecasts have a higher Matthews correlation coefficient (a metric which is higher with better classification performance) than decision-making using deterministic forecasts. However, there are two setbacks associated with using probabilistic forecasts for deciding areas to perform contrail avoidance in. First, such approaches do not account for meteorological scenarios not included in any of the ensembles. Second, it is not possible to decouple the potential impact of a contrail from the meteorology due to the epistemic uncertainties associated with modelling. Hence, aggregation techniques used in this study can be used alongside probabilistic forecast approaches to account for both meteorological and modelling uncertainty. This makes exploring aggregation strategies worthwhile.

1.2 Types of aggregation strategies

1.2.1 Threshold-based aggregation

If models could vote for which contrails to avoid (given a limited budget), selecting an appropriate set of contrails from the votes would require the use of judgement aggregation. Here, we consider threshold-based rules (Grossi and Pigozzi, 2014), a subset of judgement aggregation rules that rely on a threshold number of votes being reached for the selection of a particular contrail to avoid.

Given a potential contrail s containing all simulations evaluated on the contrail models of interest in a set of all potential contrails S , threshold-based rules follow the relationship (Eq. 2.3 of Grossi and Pigozzi, 2014):

$$f_t(S; V^*) = \{s \in S \mid V_s \geq V^*\}, \quad (1)$$

where f_t is the set of contrails selected for avoidance according to the threshold rule, V^* is the threshold number of votes, and V_s is the number of votes for each potential contrail.

Two notable threshold-based rules are the *consensus* and the *majority* rules. In the consensus rule, the threshold is equal to the number of models. In our definition of the majority rule, at least half of the contrail models must have voted for a particular



130 contrail for it to be avoided. We do not use the strict majority rule because it collapses to the consensus rule when using two agents, which is a relevant case considered in this study. These rules are commonly used for day-to-day group decision-making.

1.2.2 Weighted aggregation

We now move away from aggregation using votes from each contrail model and instead consider the predicted contrail properties directly. For an example of a weighted aggregation rule applicable to contrail models, we first define the loss function f_{loss} of a set of n contrails to be avoided ($\Omega \subseteq S$) to be the weighted sum of a predicted quantity M from each model, summed over all potential contrails in Ω :

$$f_{\text{loss}}(\Omega) = \sum_{s \in \Omega} \sum_k w_k \cdot M_k(s), \quad (2)$$

140 where k is the index of the contrail model, and w_k is the weight associated with the k -th contrail model. In the case of contrail avoidance, M is equal to the negative of a warming metric (such as total extinction) so that the most warming contrails are avoided when f_{loss} is minimised. The weighted aggregation rule minimises the value of the loss function by varying Ω :

$$\Omega_w(S; n) = \arg \min_{\Omega \subseteq S} f_{\text{loss}}(\Omega), \text{ s.t. } |\Omega| = n, \quad (3)$$

145 where Ω_w is the set of contrails selected for avoidance according to the weighted rule, and n is the number of contrails to avoid. The weights represent the relative importance of the k -th contrail model in the calculation of the loss function. Selection of the weights is, to an extent, arbitrary since there is no single way to prescribe them. For example, some use cases may require using equal weights, whereas use cases that value a particular feature (e.g. a multimodal particle size distribution) may want to use increased weights for models with that feature. Weighted sums have been used to analyse model outputs from the Coupled Model Intercomparison Project, phase 6 (CMIP6) (e.g. Brunner et al., 2020; Massoud et al., 2023) – a project that informed the IPCC AR6 report. Furthermore, weighted loss functions are common in optimisation problems, like those solved in machine learning. Given their relevance in policy and in optimisation, the use of a weighted aggregation strategy is justified. We select the loss function and weights for this study in Sect. 2.2.3.

150 1.2.3 Preference aggregation: classic Pareto and Multiple Dominance Relations ranking

The predicted contrail properties can also be used to determine the order of prioritisation (preference) for avoidance according to the models of interest. This requires the use of preference aggregation. Although preference aggregation can use the ranked preference from each model as the basis for aggregation, we instead use a metric from the simulated contrails directly. This follows the design ranking algorithms used in engineering multi-objective optimisation, where there are several objectives to optimise for simultaneously (Phillips and Jarrett, 2020a, 2020b; Flanagan, 2022; Akhtar Martínez and Jarrett, 2023). Hence, 155 in this section we explore two of these ranking algorithms for preference aggregation: the well-known Pareto criterion and the more recent development known as Multiple Dominance Relations (MDR), first introduced by Cook et al. (2020).



Consider two potential contrails $s_a \in S$ and $s_b \in S$ containing simulation results from different models. We denote the binary relation $s_a \preccurlyeq s_b$ to indicate that s_a dominates s_b . In other words, $s_a \preccurlyeq s_b$ indicates that it is preferable to avoid contrail s_a than contrail s_b .

As in Phillips and Jarrett (2020a, 2020b), Flanagan (2022), and Akhtar Martínez and Jarrett (2023), we define a set $\min(S, \preccurlyeq)$ whose elements are not dominated by any other potential contrail according to the $\preccurlyeq$ binary relation:

$$\min(S, \preccurlyeq) := \{s_a \in S \mid \nexists s_b \in S \text{ for which } s_b \preccurlyeq s_a \cap s_a \not\preccurlyeq s_b\} \quad (4)$$

In classic multi-objective optimisation, s_a Pareto dominates s_b if all the objective function evaluations in s_a are smaller than in s_b (Phillips and Jarrett, 2020a, 2020b; Flanagan, 2022; Akhtar Martínez and Jarrett, 2023). Similarly, we define the Pareto dominance relation $\preccurlyeq_{\text{Pareto}}$ to indicate that all model simulations of s_a predict a lower value of the feature to be minimised M_k than those simulations of s_b :

$$s_a \preccurlyeq_{\text{Pareto}} s_b \Leftrightarrow M_k(s_a) < M_k(s_b), \forall k \in 1, \dots, n_m, \quad (5)$$

where n_m is the number of contrail models. It follows that the set forming the Pareto front is given by $S_{\text{Pareto}} = \min(S, \preccurlyeq_{\text{Pareto}})$.

Pareto fronts are known to be increasingly populated when the number of objectives (contrail models) considered increases. To address this, instead of considering all the contrail models simultaneously, MDR nests different dominance relations such that the optimised set of contrails S_{MDR} satisfies the following relation (Cook et al., 2020):

$$S_{\text{MDR}} = \min(\dots(\min(\min(S, \preccurlyeq_1), \preccurlyeq_2) \dots), \preccurlyeq_{n_r}), \quad (6)$$

where n_r is the number of dominance relations, and where the nested binary relations denote Pareto dominance for their corresponding layer. The greater the number of nested relations, the lower the number of remaining candidates.

The Pareto and MDR filters can be adapted to become ranking functions instead. For the classic Pareto ranking, consider a ranking function $R(s, S, \preccurlyeq)$ that returns the number of potential contrails in S that dominate s (here we use Algorithm 2 from Cook et al., 2020). The Pareto front consists of the set containing the zero-rank candidates. When using MDR, there are two ranks: the layer rank R_D (the layer depth that a potential contrail reaches), and the Pareto rank within that layer.

Aggregation is therefore straightforward. Using classic Pareto ranking, assuming n contrails need to be avoided, we select the first n contrails of the set S after it has been sorted according to the Pareto rank. Using MDR ranking, this is repeated for each layer (starting from the deepest layer) until n unique candidates have been selected.

MDR has been used by few studies in the literature, but it has consistently been reported that it provides benefits to the computational time and designer understanding when used for multi-objective optimisation (Phillips and Jarrett, 2020a, 2020b). Phillips and Jarrett (2020a) noted that the order of the dominance relations in MDR affects the outcome in a more traceable way than when using other optimisation (and hence aggregation) methods. MDR has also seen modifications in the work by Flanagan (2022), but these have not seen further use in the literature yet. The use of MDR as a ranking function for preference aggregation outside the context of multi-objective optimisation is therefore unprecedented and is one of the main contributions from this study.



1.3 Summary and outline of this study

In Sect. 1 we have identified that it is not currently possible to validate contrail models with available observational datasets. Hence, in the presence of a growing number of available contrail models, model uncertainties remain a potential barrier to the implementation of contrail avoidance strategies which prioritize the most warming contrails. To address this, this study aims to evaluate the efficacy of aggregation strategies as a way to mitigate the effects of model uncertainties. Section 2 describes the methodology, including a description of how synthetic contrail model results are constructed to evaluate future model sets where more contrail models are available than at present. The results are analysed in Sect. 3, and their implications are discussed in Sect. 4. Finally, the limitations of this study are raised in Sect. 5, and the conclusions are presented in Sect. 6.

2 Methods

Fig. 1 provides a flow chart overviewing the experimental process for this study. Each meteorological scenario is simulated in different contrail models, including the Contrail Cirrus Prediction model (CoCiP; Schumann, 2012), the Aircraft Plume Chemistry, Emissions, and Microphysics Model (APCEMM; Fritz et al., 2020), and the large eddy simulation (LES) EULAG-LCM, which consists of the computational fluid dynamics solver EULAG (Smolarkiewicz et al., 2014) fully coupled to the particle-based ice microphysics module LCM (EULAG-LCM; Sölch and Kärcher, 2010). Since there are few available contrail models, we artificially construct results of contrail models which do not exist (referred to as synthetic contrail models; see Section 2.1). This is done because we aim to evaluate the performance of the aggregation strategies in future scenarios where more contrail models are available. For CoCiP, APCEMM, and EULAG-LCM we use the results from Akhtar Martínez et al. (2026b). More details on the meteorology and the versions of the models used for these simulations are available in Appendix A.

To benchmark the performance of the aggregation strategies, we use the results from the 48 EULAG-LCM simulations as a substitute for the ground truth. Since the dataset is limited, we perform 1000 Monte Carlo simulations of all aggregation strategies to estimate their statistical properties.

At each Monte Carlo evaluation, the available contrails are randomly split into a training set and an evaluation set. The random splitting results in the sampling of different combinations of contrails, which allows us to estimate the uncertainties associated with each aggregation strategy. We then evaluate the contrail models over the training set and perform the aggregation strategies over the evaluation set. After 1000 simulations (see the convergence analysis in Appendix B), we analyse the ensemble results to determine the mean efficacy and the corresponding 2.5th to 97.5th percentile range across the Monte Carlo evaluations of each aggregation strategy (referred to as the 95 % uncertainty interval). This process is repeated for several contrail model sets, some of which are synthetic.

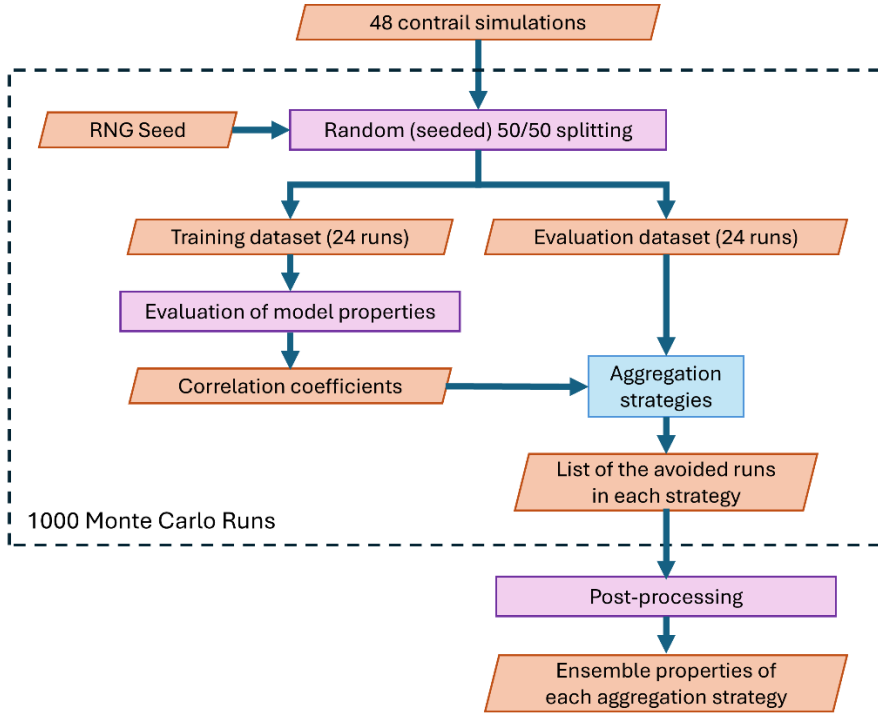


Figure 1: Flow chart showing the experimental setup for the evaluation of the aggregation strategies. The blue rectangle indicates the aggregation step, orange rhombi indicate outputs, and purple rectangles indicate data processing units.

2.1 Construction of synthetic contrail model results

Since we only have access to the results from three real contrail models (CoCiP, APCEMM, and EULAG-LCM), we construct synthetic contrail model results to evaluate the aggregation strategies across a range of scenarios, including future scenarios in which more contrail models are available. As detailed in Akhtar Martínez et al. (2026b), when comparing the normalised lifetime total extinction $\hat{E}^*$ (defined in Appendix A as a proxy metric for warming) from simulations to the corresponding ground truth quantity on a parity plot, a contrail model is effectively reduced to two elements: a functional bias and random noise (captured by the Pearson correlation coefficient β). Therefore, by choosing a functional bias and iterating on the noise level, it is possible to create synthetic contrail models with the desired Pearson correlation coefficient. This process is illustrated in Fig. 2.

In summary, we evaluate the bias function (which can be linear, exponential, etc.) at the normalised $\hat{E}^*$ of the EULAG-LCM simulations. We then apply noise to each of the evaluations of the bias function. For the TAN and LOG model families, this is a multiplicative noise with distribution $\zeta \sim U[1 - l, 1 + l]$, where $l \in (0, 1]$ (referred to as Type A noise in Table 1). For the GRAD model families, this is defined to be $\zeta \sim U\left[\frac{1}{l}, l\right]$, where $l \in [1, 5]$ (referred to as Type B noise in Table 1). Next, by performing a grid search in the noise level l we achieve the desired Pearson correlation coefficient β . Throughout the grid



search, β is only evaluated over a pre-determined set containing half of the available contrails. This reflects the split between training and evaluation sets used in the Monte Carlo simulations. Finally, we normalise the results by the median. Since the evaluation set is pre-determined and constant, this process only needs to be done once per synthetic contrail model before the Monte Carlo simulations. The results from the synthetic models are then treated like the results from CoCiP or APCEMM (the real models that are not used for benchmarking).

To test the robustness of the aggregation strategy to bad predictions, we also create two *poisoned* models: one made by randomising the EULAG-LCM results, and one made by sampling a uniform distribution (so the points are completely random). Table 1 summarises the properties of all available models (real, synthetic, and poisoned) on the pre-determined simulation set used for the grid search. An example of a synthetic contrail model is available in Fig. 3.

This way of creating synthetic model results assumes a degree of independence between evaluations. For instance, the parity plots in Fig. 10 of (Akhtar Martínez et al., 2026b) contains evaluations that are clustered (close to each other). Such pairs are not easily produced when assuming that the multiplicative noise factor is independent for each run. Although they may arise due to the perturbation of a variable to which the models are not very sensitive to (e.g. when changing the updraught velocity from 0.02 to 0.05 m s⁻¹ – see Fig. 11 of Akhtar Martínez et al., 2026b), we cannot discard the possibility that there are deeper sources of dependencies that give rise to these evaluation clusters. Given the purpose of the study being to provide a proof-of-concept for the use of aggregation strategies for contrail avoidance, we exclude the characterisation of dependencies from this work. This is justified because we do not produce actionable results when using these synthetic contrail models, nor do we make generalised statements based on our findings due to further limitations of our sandbox approach (see Sect. 5).

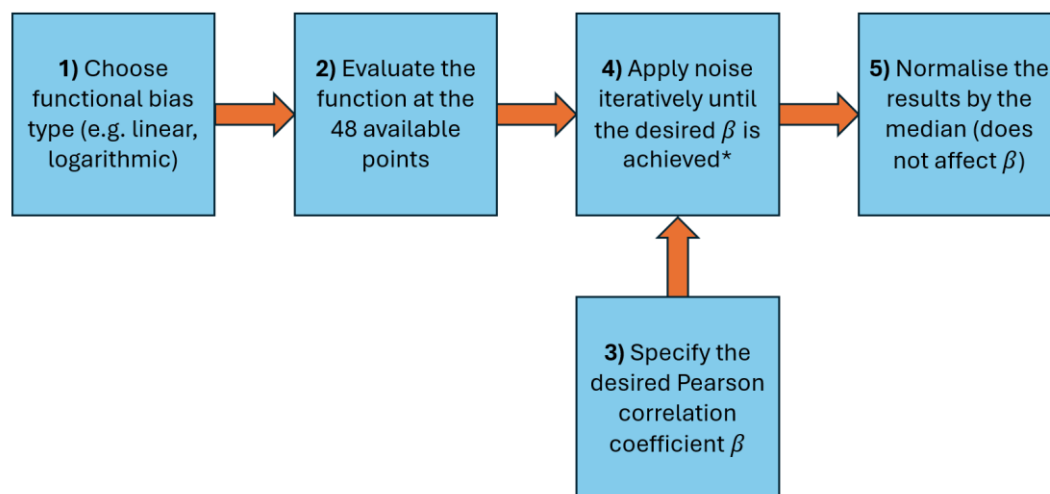
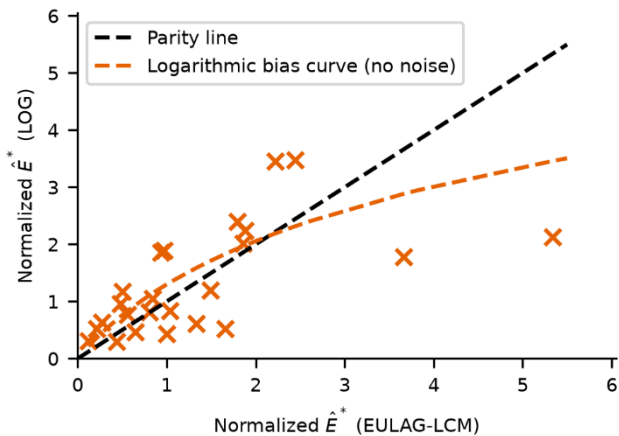


Figure 2: Flow chart showing the experimental setup for the generation of synthetic model results. The Pearson correlation coefficient β is only evaluated over a pre-determined set containing half of all available simulations.



260 **Figure 3: The results of a synthetic contrail model with a logarithmic bias and a Pearson correlation coefficient of 0.60.**

Table 1: Properties of the contrail models evaluated on the pre-determined simulation set used for the grid search. The models have been sorted by the Pearson correlation coefficient (lowest first).

<i>Model name</i>	<i>Functional bias type</i>	<i>Noise type</i>	<i>Linear fit gradient</i>	<i>Pearson correlation coefficient</i>	<i>Classification</i>
<i>PSN-B</i>	Random sorting	N/A	0.42	-0.17	Uncorrelated
<i>PSN-A</i>	Random sampling (uniform dist.)	N/A	0.64	0.08	
<i>GRAD</i>	Linear	Type B	0.84	0.50	Moderately correlated
<i>CoCiP*</i>	N/A	N/A	0.60	0.56	
<i>TAN-MID</i>	Tangential	Type A	0.90	0.58	
<i>LOG</i>	Logarithmic	Type A	0.76	0.60	
<i>TAN</i>	Tangential	Type A	1.26	0.70	Highly correlated
<i>LOG-HI</i>	Logarithmic	Type A	0.70	0.80	
<i>GRAD-HI</i>	Linear	Type B	0.93	0.90	
<i>APCEMM*</i>	N/A	N/A	0.79	0.95	

2.2 Aggregation strategies and tie-breaking

The following aggregation strategies are considered in this study:

- 265
- Aggregation by consensus
 - Aggregation by majority



- Weight-based aggregation
- Pareto-based aggregation
- MDR-based aggregation

270 We also consider the efficacy of using a single model (CoCiP or APCEMM) to select the contrails to avoid. This may be thought of as either any aggregation strategy applied on only one model, or alternatively as the aggregation strategy which only considers the preference of one model.

Given that different models have different qualities (as given by their Pearson correlation coefficient), strategies will rely on the quality of the models to inform the weights, the dominance relation hierarchy, or the tie-breaking procedures. Furthermore, 275 some of these strategies may need to be adapted to satisfy a quota n . Although it is an artificial simplifying constraint, this study uses a quota because it can be related to air traffic management (ATM) workload, air traffic conflict rate, or limitations to fuel burn penalty. A more realistic application of aggregation strategies would be used to find regions in space where many contrail models indicate a high likelihood of highly warming contrails, but this is beyond the scope of the current study.

We now consider the practicalities of each strategy individually, as well as the standard tie-breaking procedure.

280 2.2.1 Tie-breaking procedure

If tie-breaking is required, the simulations with the highest normalised $\hat{E}^*$ (according to the model with the highest Pearson correlation coefficient) are selected. This tie-breaking procedure is chosen to ensure that the selection of contrails to be avoided is deterministic and explainable. An alternative tie-breaking procedure is to choose contrails randomly. Although this alternative procedure is unbiased, it is not deterministic given any set of candidate contrails, making it unsuitable for 285 commercial aviation applications.

2.2.2 Voting for threshold-based aggregation

For the two threshold-based aggregation strategies (aggregation by consensus and majority), a voting framework is needed that is aware of the quota. In this investigation, assuming n contrails are to be selected for avoidance, contrail models will vote for the n simulations producing the highest median-normalised $\hat{E}^*$. If fewer than n simulations meet the required total votes 290 for selection, another round of voting with the remaining simulations is performed. This is repeated until the quota is met, or no more simulations are selected. If no more simulations are selected without having met the quota, the simulations with the highest number of votes are selected. In this case, if there are more simulations of the lowest relevant vote than the quota allows for, the tie-breaking procedure is performed.

2.2.3 Selection of weights for the weighted aggregation strategies

295 We perform two runs with weighted aggregation: one using equal weights, and another using quality-dependent weights. We define the quantity M from Eq. 2 to be equal to the median-normalised $\hat{E}^*$ times -1 (to ensure the avoided contrails have the



highest median-normalised $\hat{E}^*$). For the latter, the weights in Eq. 2 are chosen to be equal to the Pearson correlation coefficient (truncated at 0 to prevent negative weights).

2.2.4 Aggregation with classic Pareto and MDR ranking

300 As in Sect. 2.2.3, we define the quantity M from Eq. 5 to be equal to the median-normalised $\hat{E}^*$ times -1 (to ensure the avoided contrails have the highest median-normalised $\hat{E}^*$). For Pareto aggregation, simulations with the lowest Pareto rank are selected for avoidance. For MDR to function properly, it requires the specification of the dominance relations and their hierarchy. As explored by Phillips and Jarrett (2020b), the importance of the dominance relations decreases with increasing layer depth. Hence, the contrail models are sorted according to decreasing Pearson correlation coefficient and are paired in order. For the
 305 model sets containing an odd number of contrail models, we choose to pair the first model with the last model in the final layer. This choice gives the highest correlated model discriminatory power in the absence of a pair of models not used yet, resembling the tie-breaking procedure (see Sect. 2.2.1). Alternative arrangements are possible, but they have not yet been explored in the literature.

We now provide an example of how the models are split into layers for the MDR strategy, and how the dominance relations
 310 function. Consider the list of models, sorted in descending order according to their Pearson correlation coefficient β : [modelA, modelB, modelC, modelD, modelE]. The dominance layers will comprise the following models:

1. Layer 1: [modelA, modelB]
2. Layer 2: [modelC, modelD]
3. Layer 3: [modelE, modelA]

315 Candidate contrails can only move from the higher layers (i.e. the layers with the lowest number) to deeper layers if they are members of the Pareto front of the negative of the median-normalised $\hat{E}^*$ predicted by the pair of contrail models of the current layer (see Fig. 4 for a visualization). When selecting contrails to avoid, the MDR aggregation strategy selects the contrails in the deepest layer first, progressively moving upwards if the quota is not met. All members of the Pareto front of a layer are treated equally by MDR, so the tie-breaking protocol is used if there are more contrails at the current layer than the remaining
 320 avoidance slots.

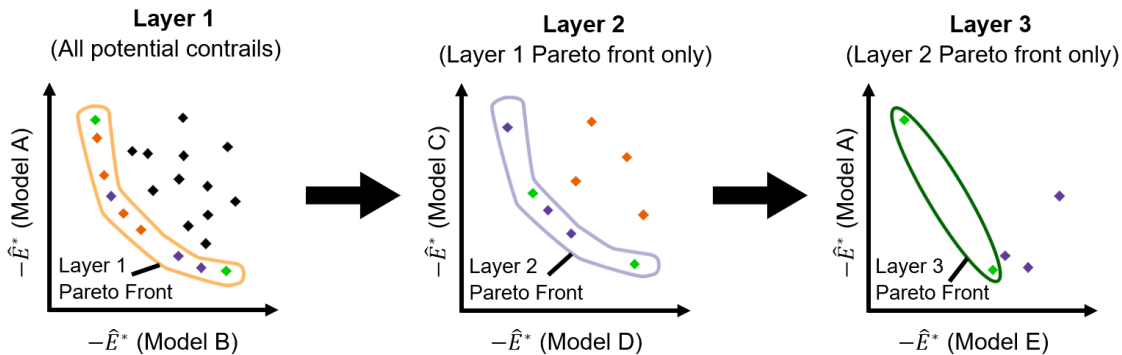




Figure 4: Visualization of the Multiple Dominance Relations (MDR) recursive filtering technique. This graph uses fabricated data for illustration purposes.

2.3 Efficacy of the aggregation strategies

325 To determine the performance of each aggregation strategy we now define some metrics. Consider the proportion of avoided lifetime total extinction λ when targeting a % of potential contrails for avoidance through the q -th aggregation strategy:

$$\lambda(a; q) = 100 \cdot \frac{\sum_{\Omega(a; q)} \hat{E}_{\text{ref}}^*}{\sum_S \hat{E}_{\text{ref}}^*}, \quad (7)$$

where the subscript ref indicates that $\hat{E}^*$ is the quantity predicted by EULAG-LCM (our proxy for the ground truth), where $\Omega(a; q)$ is the set of avoided contrails as determined by the q -th aggregation strategy, and where S is the candidate set of
 330 contrails for avoidance such that $\Omega \subseteq S$. In plain language, λ is the proportion of the total $\hat{E}_{\text{ref}}^*$ across the pool of candidate contrails that the q -th aggregation strategy can avoid at a specific budget a .

We identify that the maximum possible value λ_{max} and the value for the random-choice avoidance strategy λ_{rand} are special cases:

$$\lambda_{\text{max}}(a) = 100 \cdot \max_{\Omega \subseteq S} \frac{\sum_{\Omega(a)} \hat{E}_{\text{ref}}^*}{\sum_S \hat{E}_{\text{ref}}^*} \quad (8)$$

$$\lambda_{\text{rand}}(a) = a. \quad (9)$$

With the above, it is now possible to define the efficacy of the aggregation strategy Ψ :

$$\Psi(b; q) = \frac{\int_0^b \lambda(a; q) - \lambda_{\text{rand}}(a) da}{\int_0^b \lambda_{\text{max}}(a) - \lambda_{\text{rand}}(a) da} \quad (10)$$

where b is the upper limit of the possible proportion of contrails that can be avoided. The efficacy of the aggregation strategy has the following mathematical properties:

- 340
- $\Psi = 1$ if the strategy achieves the maximum possible reduction in $\hat{E}_{\text{ref}}^*$.
 - $\Psi > 0$ if the strategy achieves a larger reduction in $\hat{E}_{\text{ref}}^*$ than the mean effect of the random-selection strategy.
 - $\Psi = 0$ if the strategy achieves the same reduction in $\hat{E}_{\text{ref}}^*$ as the mean effect of the random-selection strategy.
 - $\Psi < 0$ if the strategy achieves a smaller reduction in $\hat{E}_{\text{ref}}^*$ than the mean effect of the random-selection strategy.

Unless otherwise specified, we provide the mean of the λ and Ψ metrics across all Monte Carlo simulations in Sect. 3.

345 2.4 Model sets evaluated

As time passes, it is likely that more contrail models of varying quality will become available. Hence, we evaluate the aggregation strategy performance for 6 different model sets containing more than one model, and 2 model sets containing only one model (see Table 2). Each of the prescribed model sets containing more than one model also has a corresponding *poisoned* version, in which the PSN-A and PSN-B synthetic models are included in addition to the models already present in each model
 350 set. The analysis conducted in Section 3 hence considers the poisoned model sets separately as part of the robustness analysis for each aggregation strategy.



355 selected strategies to the maximum avoidance budget b in Section 3.2.

Table 2: Model sets considered for evaluating the aggregation strategies.

<i>Model set name</i>	<i>Description</i>	<i>Models</i>	<i>Symbol</i>	<i>Poisoned version available?</i>
<i>Model set A</i>	Two moderately correlated models	CoCiP, GRAD	●	Yes
<i>Model set B</i>	Three moderately correlated models	CoCiP, GRAD, TAN-MID	■	Yes
<i>Model set C</i>	Four moderately correlated models	CoCiP, GRAD, TAN-MID, LOG	◆	Yes
<i>Model set D</i>	Three highly correlated models	APCEMM, GRAD-HI, LOG-HI	▲	Yes
<i>Model set E</i>	Moderately correlated and highly correlated models	CoCiP, LOG, TAN, GRAD-HI, LOG-HI	▼	Yes
<i>Model set F</i>	Real models	CoCiP, APCEMM	+	Yes
<i>Single model A</i>	Real moderately correlated model	CoCiP	N/A	No
<i>Single model B</i>	Real highly correlated model	APCEMM	N/A	No

3 Results

3.1 Performance evaluation of the aggregation strategies

360 In this section we evaluate the performance of the aggregation strategies by first considering the achievable λ at a budget of 21 % (Fig. 5). We then consider their efficacy Ψ at a maximum budget of 21 % (Fig. 6). Finally, we select two aggregation strategies that operate on different principles for further analysis.

Before poisoning the model selection, the two weighted aggregation strategies yield the highest average λ across all model sets at a budget of 21 % ($\lambda = 43$ %), see the bars in Fig. 5). After poisoning the model selection, the two aggregation strategies with the highest average λ at a budget of 21 % are the β -weighted ($\lambda = 43$ %) and MDR ($\lambda = 41$ %) aggregation strategies.



365 Both the β -weighted and MDR strategies experience no appreciable decrease in average λ when the model selection is poisoned, indicating resilience (not immunity) to model error for the model sets considered.

When considering the average efficacy across all model setups to a maximum budget of 21 %, both weighted loss function strategies have the maximum average efficacy (0.78) prior to poisoning (see the bars in Fig. 6), whilst MDR has an average efficacy of 0.74. After poisoning, the β -weighted strategy retains its average efficacy of 0.78, which is followed by the MDR strategy with an average efficacy of 0.73. This corresponds to a negligible decrease in efficacy for the β -weighted strategy, and a decrease in efficacy of 0.01 for the MDR strategy. Given that both λ and the efficacy are highest for the β -weighted and MDR aggregation strategies when poisoned, we select them for further analysis.

When compared to the strategy that just uses one model (CoCiP) to determine the contrails to avoid at a maximum budget of 21 %, the β -weighted and MDR aggregation strategies have a higher average efficacy after poisoning across all model sets (0.78 and 0.73 respectively) than just using CoCiP (0.40) (Fig. 6). At a budget of 21 %, this corresponds to average λ values across all model sets of 43 %, 41 %, and 30 % for the β -weighted, MDR, and single model aggregation strategies respectively (Fig. 5). In other words, using the β -weighted aggregation strategy leads to an expected increase in the $\hat{E}_{ref}^*$ savings of 43 % relative to the savings made when only using CoCiP (calculated from the ratio of average λ values). This quantity is limited to 37 % when using the MDR aggregation strategy. Similarly, the expected increase in the $\hat{E}_{ref}^*$ savings across all model sets is 105 % for the β -weighted strategy, 95 % for the MDR strategy, and 43 % for the CoCiP strategy, relative to the random selection strategy. We now consider the model sets individually in Figs. 5 and 6.

The model set that consistently yields the lowest efficacy and λ is model set A – consisting of two moderately correlated models (see Table 2). Conversely, the model set that most frequently yields the highest efficacy and λ is model set D – consisting of three APCEMM-like models. This is expected since the models in model set D have a higher Pearson correlation coefficient than those in model set A (see Tables 1-2). However, all model sets have a mean efficacy greater than 0, indicating that they all perform better than random selection when there are no cooling contrails.

When considering the β -weighted and the MDR aggregation strategies, all model sets have a higher efficacy and λ than the CoCiP strategy (both before and after the poisoning). The other strategies that achieve this are the Pareto and equal weight strategy. Furthermore, when using three moderately correlated models (model set B) at an avoidance budget of 21 % both the β -weighted and MDR aggregation strategies have a λ value of 41 % and 40 % respectively, whereas using CoCiP by itself yields a λ value 30 %, and the random avoidance strategy yields a λ value of 21 %. This corresponds to an expected increase in savings of $\hat{E}_{ref}^*$ of at least 33 % relative to the savings from CoCiP and at least 90 % relative to the savings from the random strategy when using either the β -weighted or the MDR aggregation strategies (calculated from the ratio of average λ values). When only considering two moderately correlated models (model set A), the β -weighted and MDR aggregation strategies have a λ value of ~ 40 % and ~ 36 % respectively. Taking the lowest λ out of the two strategies, this corresponds to an expected increase in savings of $\hat{E}_{ref}^* \sim 20$ % relative to the savings from CoCiP and ~ 71 % relative to the savings from the random strategy.

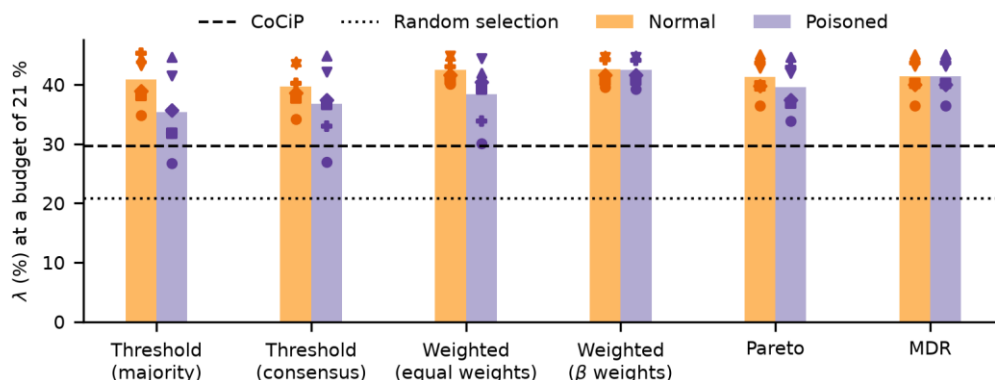


Figure 5: Bar chart showing the mean avoided lifetime total extinction λ for each aggregation strategy as computed by EULAG-LCM at an avoidance budget of 21 %. The markings correspond to the Monte Carlo mean for each model set (see Table 2 for the legend), and the bars correspond to the average of the Monte Carlo mean of all model sets per aggregation strategy.

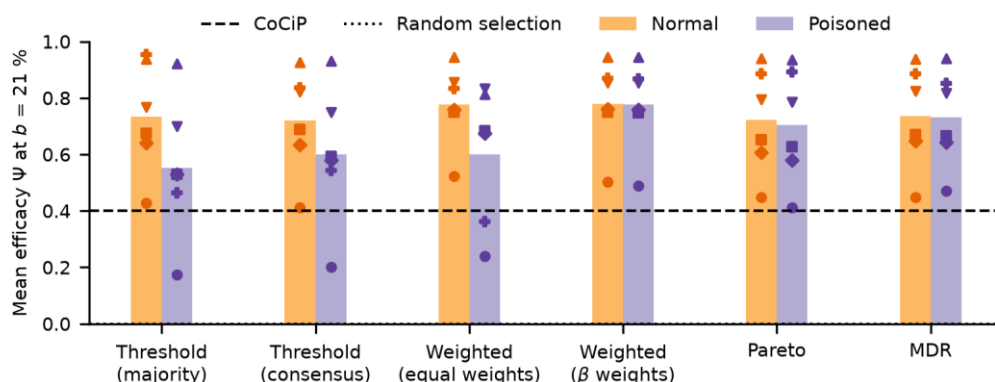


Figure 6: Bar chart showing the mean aggregation strategy efficacy Ψ at a maximum avoidance budget of 21 %. The markings correspond to the Monte Carlo mean for each model set (see Table 2 for the legend), and the bars correspond to the average of the Monte Carlo mean of all model sets per aggregation strategy.

3.2 Sensitivity analysis of the selected strategies

In this section we analyse the sensitivities of the β -weighted and MDR aggregation strategies to changes in the upper bound of integration (Fig. 7) and the model number (Figs. 7 and 8).

Both the β -weighted and MDR aggregation strategies have a higher mean efficacy than just using CoCiP for all values of b (the upper limit of the possible proportion of contrails that can be avoided) for model sets B and C – regardless of whether the model setup has been poisoned or not (Fig. 7). This indicates that, in the case that there are no cooling contrails, both the aforementioned aggregation strategies perform better than CoCiP for model sets B and C. For model set A, however, the aggregation strategies perform worse than CoCiP when $b < \sim 15\text{--}20\%$. Since there are only two contrail models being considered in model set A, this lower-than-CoCiP performance is expected because the group decision is more sensitive to



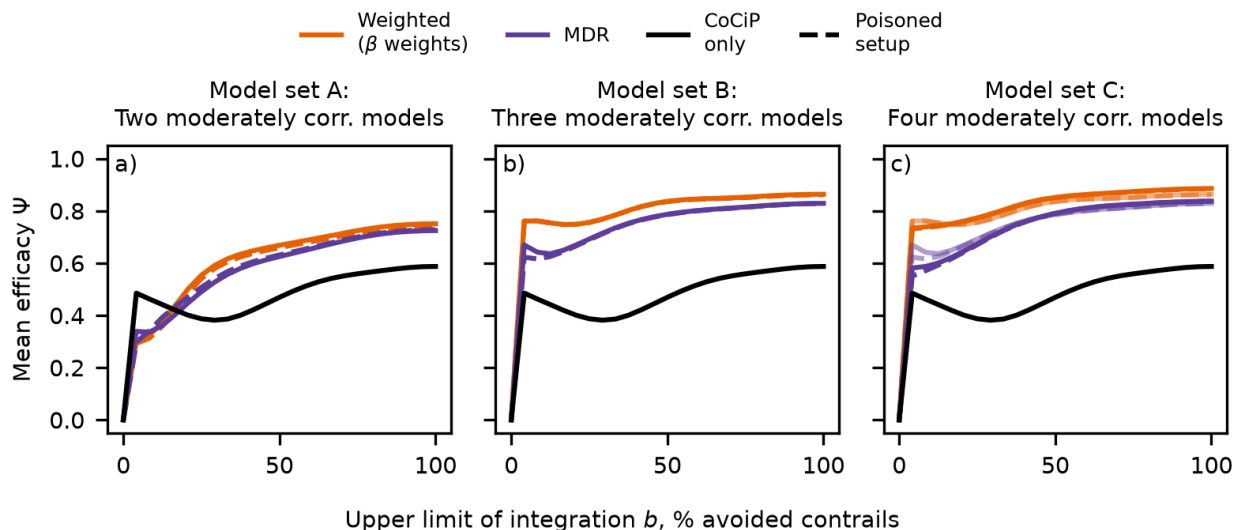
415 disagreement between models when the budget is extremely low. Nevertheless, all the discussed efficacies are above zero, indicating that even CoCiP may yield more effective contrail avoidance than random selection for the case in which no cooling contrails are considered. The implications of including cooling contrails in this analysis are discussed in Section 5.

When moving from using three models (model set B) to four models (model set C), we see limited improvements in the mean efficacy. Furthermore, at certain values of b ($b < \sim 15\%$ and $b < \sim 45\%$ for the β -weighted and MDR aggregation strategies
 420 respectively), the mean efficacy when using three models is higher than when using four models. This indicates one of two things: there are limited benefits of adding additional models – or there may occasionally be performance drops due to decreased relative importance of more correlated models. This is likely because, given a small avoidance budget, it is difficult for models with a low Pearson correlation coefficient to agree on which contrail is the most impactful. Alternatively, with a larger number of contrail models to consider, the relative importance of the more correlated models is reduced, causing worse
 425 performance than when there are fewer models. Both effects may be magnified in the current investigation because a 21 % budget only corresponds to 5 contrails to avoid due to the training/testing data split. With a higher number of contrails to avoid, we expect this effect to be reduced. Due to this, model set C has been excluded from the potential technology timeline in Fig. 8. The limitations of the aggregation strategies and their potential to worsen performance are explored in Sect. 5.2.

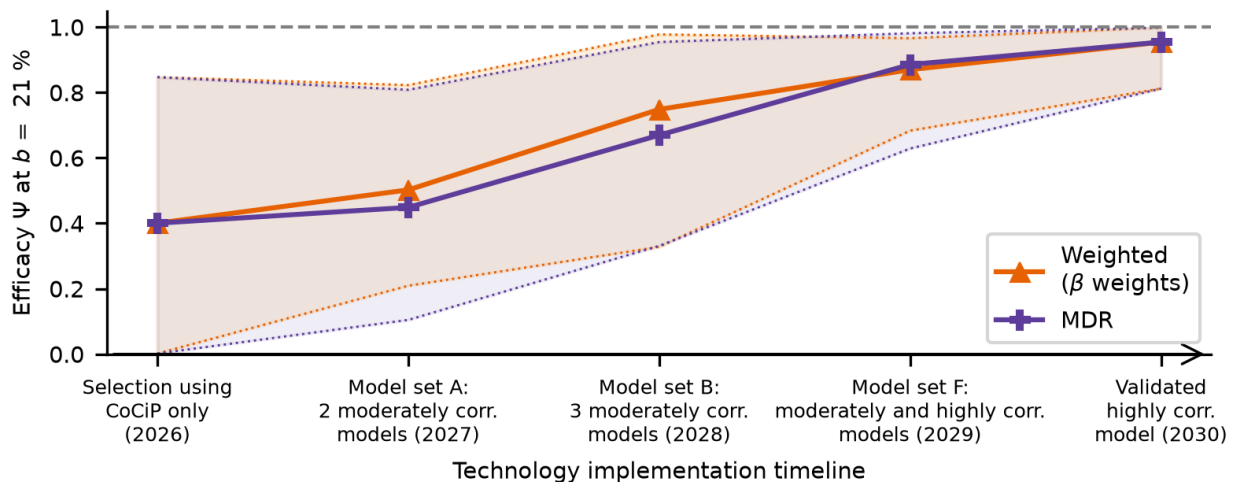
Poisoning of the model sets also has a negligible effect on the β -weighted aggregation strategy for all model sets, but it has a
 430 noticeable effect on the MDR aggregation strategy when $b < 20\%$ for model set B (Fig. 7b). Moreover, without considering poisoning, the β -weighted aggregation strategy has a higher mean efficacy than the MDR strategy at all values of b , with the maximum difference between them being 0.12 and 0.15 at $b = 8\%$ and $b = 4\%$ for model sets B and C respectively (Fig. 7b and Fig. 7c). This indicates that the β -weighted aggregation strategy has a more robust performance than the MDR strategy. Furthermore, for model set A, MDR performs better when exposed to the poisoned setup than to the standard setup (Fig. 7a).
 435 This is expected because model set A contains two models in the standard setup, meaning that the highest correlated model (CoCiP in model set A) is repeated in the lowest layer when the poisoned model is added. The additional weight of the highest correlated model also causes visible performance improvements in the MDR column of Fig. 6 for model sets A and D.

We now consider the mean and 95 % uncertainty intervals of the efficacy of the β -weighted and MDR aggregation strategies when exposed to more model sets, presented as a possible implementation timeline in Fig. 8. Qualitatively, the efficacy
 440 increases for both strategies as more models (or models with higher correlation coefficients) are introduced for the cases considered. Simultaneously, the 95 % uncertainty intervals shrink for both strategies. Furthermore, the size of the efficacy uncertainty interval for the β -weighted strategy is smaller than or equal to that from the MDR strategy. Similarly, the efficacy of the β -weighted strategy is greater than or equal to the efficacy of the MDR strategy across all model sets. Note that, as discussed in Sect. 5.2, improvements in the performance and uncertainty intervals are not guaranteed to take place when using
 445 aggregation strategies.

Finally, for the contrail selection using only one model (CoCiP), the lower end of its 95 % uncertainty interval for efficacy is close to zero, indicating a worst-case performance comparable to random selection of contrails for the experimental setup considered.



450 **Figure 7:** Graphs showing the effect of the integration limit b on the efficacy for the β -weighted and MDR aggregation strategies for model set A (panel a), model set B (panel b), and model set C (panel c). The less saturated lines in panel c indicate the results of model set B, superimposed for convenience of comparison.



455 **Figure 8:** Timeline showing a potential progression of the aggregation strategy as new hypothetical contrail models are developed. The shaded areas correspond to the efficacy 95 % uncertainty interval range for the β -weighted (orange) and the MDR (purple) aggregation strategies. The dashed horizontal line indicates the theoretical limit of the efficacy.

4 Discussion

The two strategies identified as the most suitable for aggregation – β -weighted and MDR – are those with explicit penalisation for models with low Pearson correlation coefficients. The β -weighted strategy performs better than the MDR strategy in terms of robustness to poisoning, robustness to changes in the model number, robustness to changes in the maximum budget, and in

460



its 95 % uncertainty interval. Purely from a performance perspective, we recommend the β -weighted strategy based on the results in this study.

The lower relative performance of the MDR strategy may be due to its formulation. Once a set of contrails reaches a lower dominance layer, the models in higher layers no longer have any input as to which runs to choose (except when there are an odd number of models, and the most correlated model forms part of the deepest layer). For optimisation strategies, this may be favourable because layers are composed of either optimisation objectives or desirable features (Phillips and Jarrett, 2020a). However, this property of MDR may not be as desirable when dealing with an aggregation problem because it may allow for lower-correlated models to have a stronger influence on the results than is warranted. Hence, we believe that MDR may be able to perform better if a different approach to nesting than the one employed in this study (and in the broader literature, e.g. Phillips and Jarrett, 2020a) is used. Nevertheless, MDR still performed second best under the influence of poisoned model sets – showing that it possesses the robustness required for the task.

MDR has other strengths that the β -weighted strategy lacks. For example, the nesting order in MDR is easier for an engineer to justify (for example, in a multi-objective optimisation problem; Phillips and Jarrett, 2020a) when compared to justifying the weights for any strategy employing a weighted sum. Furthermore, the logic flow used in MDR is straightforward to follow, whereas the weighted function does not necessarily follow any obvious logic. In the context of aviation, traceability and explainability are important – especially when faced with the possibility that mistakes in the aggregation strategy could end up systematically causing further climate warming. When considering these practicalities, MDR emerges as the most favourable strategy.

Nevertheless, the confidence in the results is increased by the fact that two fundamentally different aggregation strategies robustly improved the performance of contrail avoidance in most of the circumstances analysed.

Finally, Fig. 8 also shows that one high-fidelity validated model may outperform all aggregation strategies. Hence, the findings obtained in this study may best serve the times in which such a model is not available but lower-fidelity models are. We also note that, if two or more validated models are available, aggregation strategies may still yield more robust results than using one model by itself. This is because models will inevitably have epistemic uncertainties due to biases or errors. Using an aggregation strategy may help mitigate the epistemic uncertainties, yielding more robust results, as seen in this investigation. However, it is important to note that such improvements are case-by-case dependent and are therefore not guaranteed. We now discuss these caveats in detail.

5 Limitations

In this section, we first consider the limitations of the experimental setup in Section 5.1, including the implications of the important omission of cooling contrails. We then consider the limitations of the aggregation strategies in Section 5.2, including a note regarding why results from any analysis employing aggregation strategies may not be generalisable.



5.1 Limitations of the experimental setup

First, in this study we ignore the probabilities related to forecasting the location and persistence of contrails. We recognise that a comprehensive approach including formation probabilities and spatial correlations may be needed in practice to better exploit patterns in the forecast and observed meteorologies. Neglecting this may limit the effectiveness of the aggregation approach for decision-making in the context of contrail avoidance.

Second, although we have presented results that imply that the β -weighted aggregation strategy, MDR aggregation strategy, and using CoCiP all perform on average better than random selection of contrails for avoidance, these results are only valid in the absence of cooling contrails. This is because aggregation strategies that rely on modelling may inadvertently and systematically target cooling contrails instead of the intended warming contrails, as detailed in Section 1. As illustrated in Fig. 9, although the random selection strategy yields a straight line between 0 to 100 % regardless of whether there are cooling contrails, the worst-case may cause negative savings in ERF. This risk of warming the planet demands the re-evaluation of the performance of the aggregation strategies when the full functional relations between contrail model inputs and climate impact – not just the contrail radiative impact – can be properly estimated. This is to ensure that we do not cause a net climate detriment when deploying warming-based avoidance strategies.

Another limitation in this study is the assumption that the random noise used to generate the synthetic contrail model results is independent. Results from previous studies (Akhtar Martínez et al., 2025; Akhtar Martínez et al., 2026b) show some model simulations that are related to each other. Although this could be explained by the perturbation of parameters with minor effects on the contrail evolution, we do not discard the possibility of deeper dependencies existing.

The results in this study are also limited because of the low number of runs available, and because EULAG-LCM simulations are used as ground truth. We acknowledge that this limits the generalisability of our findings. However, our study should simply be taken as a proof of concept that aggregation strategies might be worth exploring in the future, when a ground truth dataset is available and when more contrail models that can be deployed globally are ready.

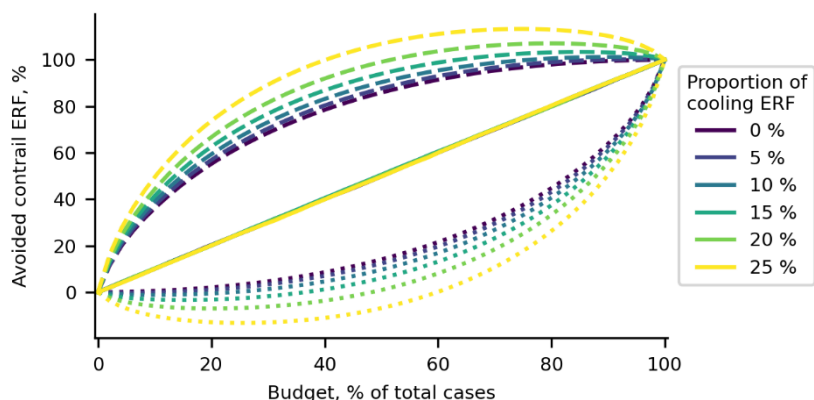


Figure 9: Avoided contrail ERF against the avoidance budget a for the random selection strategy (solid lines), best-case scenario (dashed lines), and the worst-case scenario (dotted lines) at different proportions of cooling ERF. This graph uses synthetic ERF



results in 1000 Monte Carlo simulations, and it is provided for illustration purposes only. For the random selection, the results approach the line of parity, so they are valid even if made with synthetic data.

5.2 Limitations of the aggregation strategies

520 Although two of the aggregation strategies in this study have robustly improved the performance of contrail avoidance given a limited number of avoidance moves, there may be cases in which any aggregation strategy could worsen the performance instead of improving it.

Consider the vector space containing all possible simulations (represented as a 2D plane in Fig. 10). Since aggregation strategies consider the agreement of multiple contrail models, it is likely that they will select contrails in the regions of the simulation space where contrail models agree (represented by the intersecting areas in Fig. 10). However, Fig. 10b shows that there may be situations in which simulations which agree with the ground truth are not produced in the regions where contrail models agree. In such situations, aggregation strategies could worsen the performance (since less of the ground truth intersects with the regions of model agreement in Fig. 10b). Therefore, the improvements of the aggregation strategies performance when increasing the number of models are not generalisable, and are model dependent. This reinforces the need to conduct a performance analysis when adding a new model to be considered in the aggregation strategies.

530

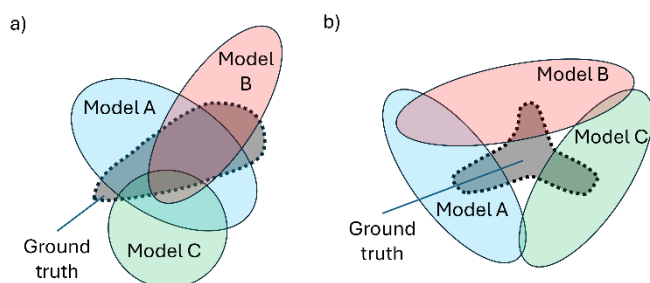


Figure 10: Venn diagrams visualising contrail model agreement within the vector space containing all possible simulations. Panel a shows a situation in which the contrail models approach the ground truth when all contrail models agree. Panel b shows a situation in which agreement with the ground truth happens when the contrail models do not agree.

535 6 Conclusions

Although previous studies (Akhtar Martínez et al., 2025; Akhtar Martínez et al., 2026b) indicated that contrail model predictions may be irreconcilable on a case-by-case basis regardless of the model fidelity, this study showed that an ensemble approach to selecting contrails for avoidance may mitigate the epistemic uncertainties at hand. The use of normalised metrics and aggregation strategies permits this.

540 When considering two moderately correlated contrail models (as determined by their Pearson correlation coefficient) at an avoidance budget of 21 %, we found that using aggregation strategies yielded an expected increase in $\hat{E}_{ref}^*$ savings of at least 20 % relative to the savings obtained when just using CoCiP to determine the contrails to avoid. When considering three moderately correlated models, the relative expected increase in savings reached 33 % compared to CoCiP. Both CoCiP and

the aggregation strategies outperformed random selection. The two strategies responsible for this are the weighted aggregation strategy (using truncated Pearson correlation coefficients as weights) and the MDR aggregation strategy.

We found that, for the model sets considered in this study, the above aggregation strategies outperform CoCiP when the number of models available increases from two to three models, even when subsequently introducing models providing random information. We also found that the aggregation strategies had smaller 95 % uncertainty interval ranges than selection using CoCiP for the model sets consisting of two and three moderately correlated models. This indicates that, even if contrail models have a correlation coefficient smaller than 1 relative to a set of reference simulations, an aggregation strategy (that accounts for their correlation coefficients against a reference set of data) using more than one model may be able to mitigate the epistemic uncertainties exposed by model disagreement. This is the case in this study, indicated by the fact that two independent aggregation strategies achieve similar performance improvements. Despite aggregation strategies improving the performance of contrail avoidance in this study, these findings are not necessarily generalisable – hence requiring another performance analysis when new models are considered.

Nevertheless, the results presented in this study are subject to important limitations in the experimental setup. These include the critical scenarios where using warming-based selection avoids not only the warming contrails, but also some cooling contrails. In such scenarios, not only is the benefit of warming-based selection limited, but it could also cause a climate detriment. Meanwhile, the random selection strategy targets an equal proportion of cooling and warming contrails, achieving a proportional net climate impact benefit equal to the targeted proportion of contrails. For this reason, we only recommend the use of warming-based targeting for contrail avoidance when the relationship between contrail optical properties and climate impact is better understood, and ideally with a set of ground truth data to support it.

Appendix A: Contrail simulation setup for CoCiP, APCEMM, and EULAG-LCM

This Appendix summarises Section 2 from Akhtar Martínez et al. (2026b). Reused with permission – the manuscript will be published as a preprint under a Creative Commons BY 4.0 license.

Akhtar Martínez et al. (2026b) perform equivalent simulations using the Contrail Cirrus Prediction model (CoCiP; Schumann, 2012), the Aircraft Plume Chemistry, Emissions, and Microphysics Model (APCEMM; Fritz et al., 2020), and the CFD large eddy simulation code EULAG (Smolarkiewicz et al., 2014) coupled to the Lagrangian ice microphysics module LCM (EULAG-LCM; Sölch and Kärcher, 2010), thus making EULAG-LCM. To minimise the variability from the jet and vortex regime parametrisations, the authors initialise the simulations directly at the diffusion regime with equivalent initial conditions. For these simulations Akhtar Martínez et al. (2026b) use an idealized representation of an ice-supersaturated region consisting of a trapezoidal RH_i profile (see Fig. A1), fully described by 5 meteorological parameters.

To estimate the magnitude of the potential climate implications of each simulated contrail, Akhtar Martínez et al. (2026b) use the lifetime total extinction $\hat{E}$ as a proxy metric for the energy forcing of an individual contrail. It is defined as follows (see also Lottermoser and Unterstrasser, 2026):



$$E(t) = \int \chi(x, z) dx dz = \tau B, \quad (\text{A1})$$

$$\hat{E}^* = \hat{E}(t = 8 \text{ h}) = \int_0^{t=8 \text{ h}} E(t) dt, \quad (\text{A2})$$

where x is the horizontal direction perpendicular to the flight direction, z is the vertical direction, $E(t)$ is the total extinction (a proxy for instantaneous radiative forcing), τ_z is the vertical optical depth at a location, τ is the optical depth averaged across the contrail width (perpendicular to the flight direction), and B is the contrail width. Akhtar Martínez et al. (2026b) only use the width-averaged expression to calculate $E(t)$ from CoCiP only since its plumes have a prescribed shape.

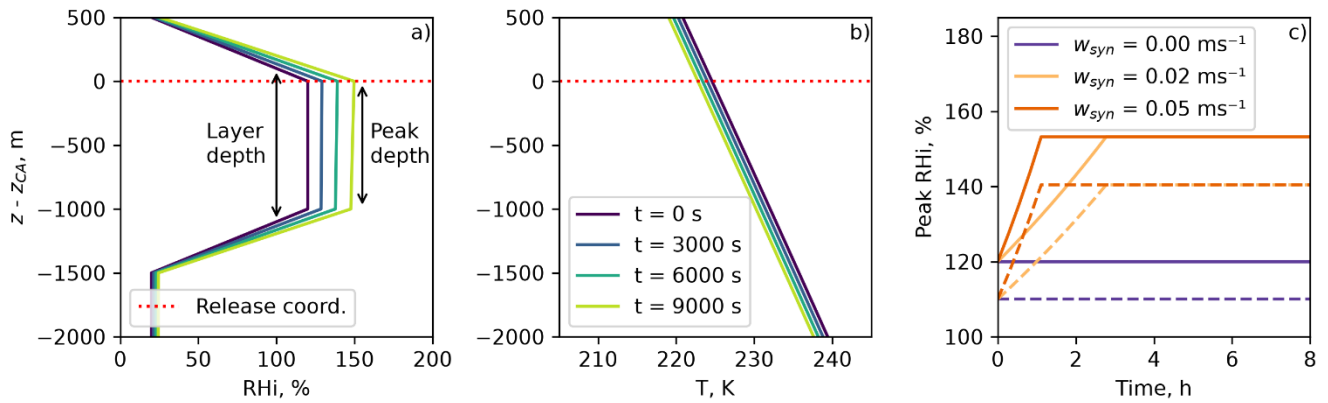


Figure A1: Panel a: evolution of the layer relative humidity profile used in Akhtar Martínez et al. (2026b). Panel b shows the temperature profile, and panel c shows the time evolution of the peak RH_i for an example case. This figure has been taken directly from Figure 4 in Akhtar Martínez et al. (2026b).

Table A1: Values of the meteorological parameters used in Akhtar Martínez et al. (2026b). This table has been taken directly from Table 3 in Akhtar Martínez et al. (2026b).

Parameter	Units	Low value	High value
Initial peak relative humidity, RH_{i0}	%	110	120*
Updraught velocity*, w_{syn}	cm s^{-1}	2	5
Initial release temperature, $T_{CA,0}$	K	217*	225
Peak supersaturation depth, d	m	500	1000*
Vertical wind shear, s	s^{-1}	0.002*	0.006

Appendix B: Convergence analysis

To ensure convergence of the results presented we define robust convergence criteria. Consider a set of $Q_{\max}$ Monte Carlo simulations per model set. Convergence is achieved after Q^* runs if no runs between Q^* and $Q_{\max}$ (where $Q^* < Q_{\max}$) have a

difference of more than 5 % in the quantities of interest relative to $Q_{\max}$. For this study, the quantities of interest are the mean and standard deviation of the proportion of saved $\hat{E}_{ref}^*$ (evaluated over the first Q runs). The example in Fig B1 illustrates this. The convergence criteria are evaluated over each contrail model, over each aggregation strategy and budget. Through this, we found that the highest number of simulations to converge the mean is 160, and 893 to converge the standard deviation. Since both convergence thresholds are well below 1000 simulations, the use of 1000 Monte Carlo simulations in this study ensures that all presented quantities have converged.

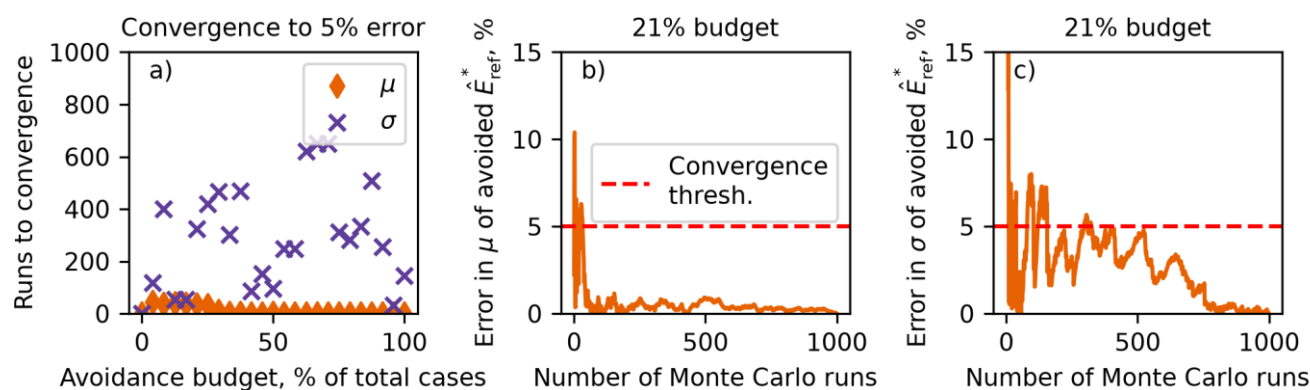


Figure B1: Convergence analysis for the strategy selecting according to the $\hat{E}_{ref}^*$ predicted by CoCiP. This is only an example.

600 Author contributions

All authors provided extensive technical and scientific contributions throughout the investigation. CAM set the overall scope of the manuscript and produced the code. All authors contributed to the structure of the experimental design. CAM prepared the manuscript with reviews and edits from all co-authors. JPJ and SDE provided research guidance and reviewed the quality, importance, and validity of the arguments raised.

605 Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The authors would like to thank Dr. Annemarie Lottermoser and Dr. Simon Unterstrasser for providing access to EULAG-LCM simulation results, for providing feedback on early versions of this work, and for participating in insightful conversations about contrail modelling and avoidance. This work was supported by the UK Engineering and Physical Sciences Research



Council [grant number EP/W524633/1]. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising.

Code availability

The scripts required to generate the graphs of this manuscript (including the algorithms for the aggregation strategies) are
615 available at (Akhtar Martínez et al., 2026a).

Data availability

The data required to generate the graphs of this manuscript are available at (Akhtar Martínez et al., 2026a).

Use of artificial intelligence

CAM made use of “Claude Sonnet 4.5”, “Claude Sonnet 4.6”, “Claude Sonnet 5”, and “Claude Opus 4.6” by Anthropic through
620 GitHub Copilot in a purely assistive manner. Example use cases include retrieving code scaffolding of common patterns
(analogous to how tutorial websites and help forums are used), asking questions about functions (akin to using the code
documentation), assistance in the navigation of existing code bases, assistance with understanding some error messages,
checking for bugs in new code before compilation and running, refactoring existing code, and commenting existing code. Any
artificial intelligence (AI) output was thoroughly checked for correctness and, if presented with any unfamiliar functions, the
625 official documentation was consulted and understood. The code produced with AI assistance would be able to be produced by
CAM without the AI if given access to help forums, tutorial websites for unfamiliar packages, and code documentation. CAM
also used artificial intelligence for English language and typesetting assistance, but not for generation of new content in the
manuscript. This mainly consisted of identifying spelling, grammar, and typographical errors. CAM reviewed and refined the
language suggestions made by these tools prior to their inclusion in the manuscript.

630 References

Akhtar Martínez, C., Eastham, S. D., and Jarrett, J. P.: Zero-dimensional contrail models could underpredict lifetime optical
depth, *Atmos. Chem. Phys.*, 25, 12 875–12 891, <https://doi.org/10.5194/acp-25-12875-2025>, 2025.

Akhtar Martínez, C. and Jarrett, J. P.: Configuration Optimization of a Hydrogen-Kerosene Hybrid Combustion Aircraft
635 Retrofit, in: AIAA SCITECH 2023 Forum, <https://doi.org/10.2514/6.2023-1161>, 2023.



Akhtar Martínez, C., Jarrett, J. P., and Eastham, S. D.: Scripts and Data for "Using multiple contrail models for decision making could increase efficacy of warming-based contrail avoidance strategies", [Computer software], Zenodo, <https://doi.org/10.5281/zenodo.23592477>, 2026a.

640

Akhtar Martínez, C., Lottermoser, A., Unterstrasser, S., Eastham, S. D., and Jarrett, J. P.: Simplifying assumptions made by contrail–cirrus models lead to irreconcilable differences in geometric and optical contrail properties, EGU sphere [preprint], <https://doi.org/10.5194/egusphere-2026-5576>, 2026b.

645 Avila, D., Sherry, L., and Thompson, T.: Reducing global warming by airline contrail avoidance: A case study of annual benefits for the contiguous United States, *Transportation Research Interdisciplinary Perspectives*, 2, 100 033, <https://doi.org/10.1016/j.trip.2019.100033>, 2019.

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., and Knutti, R.: Reduced global warming from
 650 CMIP6 projections when weighting models by performance and independence, *Earth Syst. Dynam.*, 11, 995–1012, <https://doi.org/10.5194/esd-11-995-2020>, 2020.

Cook, L. W., Willcox, K. E., and Jarrett, J. P.: Design optimization using multiple dominance relations, *Int. J. Numer. Meth. Eng.*, 121, 2481–2502, <https://doi.org/10.1002/nme.6316>, 2020.

655

Dean, T. R., Abbott, T. H., Engberg, Z., Masson, N., Teoh, R., Itcovitz, J. P., Stettler, M. E. J., and Shapiro, M. L.: Impact of forecast stability on navigational contrail avoidance, *Environ. Res.: Infrastruct. Sustain.*, 5, 045 008, <https://doi.org/10.1088/2634-4505/ae1da5>, 2025.

660 Driver, O. G. A., Stettler, M. E. J., and Gryspeerdt, E.: Factors limiting contrail detection in satellite imagery, *Atmos. Meas. Tech.*, 18, 1115–1134, <https://doi.org/10.5194/amt-18-1115-2025>, 2025.

Elmourad, J. A.: Evaluating Fuel–Climate Tradeoffs in Contrail Avoidance, Master’s thesis, Massachusetts Institute of Technology, <https://dspace.mit.edu/handle/1721.1/150282>, online. Last accessed 1 February 2026, 2023.

665

Euchenhofer, M. V., Prashanth, P., Parke, S. A., Eastham, S. D., and Waitz, I. A.: Contrail Observation Limitations Using Geostationary Satellites, *Geophys. Res. Lett.*, 52, e2025GL118 386, <https://doi.org/10.1029/2025GL118386>, 2025.

Flanagan, F.: Fundamental Configuration Optimisation of Aircraft with Electric Propulsion, Ph.D. thesis, University of
 670 Cambridge, <https://doi.org/10.17863/CAM.96496>, 2022.



Fritz, T. M., Eastham, S. D., Speth, R. L., and Barrett, S. R. H.: The role of plume-scale processes in long-term impacts of aircraft emissions, *Atmos. Chem. Phys.*, 20, 5697–5727, <https://doi.org/10.5194/acp-20-5697-2020>, 2020.

675 Grossi, D. and Pigozzi, G.: Judgment Aggregation: A Primer, vol. 8 of Synthesis Lectures on Artificial Intelligence and Machine Learning, Synthesis Lectures on Artificial Intelligence and Machine Learning, Springer, <https://doi.org/10.1007/978-3-031-01568-7>, 2014.

Hanst, M., Köhler, C. G., Seifert, A., and Schlemmer, L.: Predicting ice supersaturation for contrail avoidance: ensemble
 680 forecasting using ICON with two-moment ice microphysics, *Atmos. Chem. Phys.*, 25, 17253–17274, <https://doi.org/10.5194/acp-25-17253-2025>, 2025.

Kärcher, B.: Formation and radiative forcing of contrail cirrus, *Nat. Commun.*, 9, 1824, <https://doi.org/10.1038/s41467-018-04068-0>, 2018.

685 Lee, D., Fahey, D., Skowron, A., Allen, M., Burkhardt, U., Chen, Q., Doherty, S., Freeman, S., Forster, P., Fuglestad, J., Gettelman, A., De León, R., Lim, L., Lund, M., Millar, R., Owen, B., Penner, J., Pitari, G., Prather, M., Sausen, R., and Wilcox, L.: The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018, *Atmospheric Environment*, 244, 117 834, <https://doi.org/10.1016/j.atmosenv.2020.117834>, 2021.

690 Lottermoser, A. and Unterstrasser, S.: Modeling the Impact of Alternative Fuels and Hydrogen Propulsion on Contrail-Cirrus: A Parameter Study, *J. Geophys. Res.-Atmos.*, 131, e2025JD044 604, <https://doi.org/10.1029/2025JD044604>, 2026.

Martín Frías, A., Shapiro, M. L., Engberg, Z., Zopp, R., Soler, M., and Stettler, M. E. J.: Feasibility of contrail avoidance in a
 695 commercial flight planning system: an operational analysis, *Environ. Res.: Infrastruct. Sustain.*, 4, 015 013, <https://doi.org/10.1088/2634-4505/ad310c>, 2024.

Massoud, E. C., Lee, H. K., Terando, A., and Wehner, M.: Bayesian weighting of climate models based on climate sensitivity, *Commun. Earth. Environ.*, 4, 365, <https://doi.org/10.1038/s43247-023-01009-8>, 2023.

700 Meijer, V. R.: Satellite-based Analysis and Forecast Evaluation of Aviation Contrails, Ph.D. thesis, Massachusetts Institute of Technology, <https://dspace.mit.edu/handle/1721.1/155350>, online. Last accessed 6 February 2026, 2024.



- Montijano, E., Martínez, S., and Sagues, C.: Distributed robust data fusion based on dynamic voting, in: 2011 IEEE
 705 International Conference on Robotics and Automation, pp. 5893–5898, <https://doi.org/10.1109/ICRA.2011.5980096>, 2011.
- Phillips, S. and Jarrett, J. P.: Assessing the Significance of Nesting Order in Optimization Using Multiple Dominance
 Relations, in: AIAA AVIATION 2020 FORUM, AIAA, <https://doi.org/10.2514/6.2020-3134>, 2020a.
- 710 Phillips, S. and Jarrett, J. P.: Enhancing Designer Understanding by Combining Multiple Dominance Relations and Tabu
 Search, in: AIAA SCITECH 2020 Forum, AIAA, <https://doi.org/10.2514/6.2020-0161>, 2020b.
- Schumann, U.: A contrail cirrus prediction model, *Geosci. Model Dev.*, 5, 543–580, <https://doi.org/10.5194/gmd-5-543-2012>,
 2012.
- 715 Schumann, U., Hempel, R., Flentje, H., Garhammer, M., Graf, K., Kox, S., Lösslein, H., and Mayer, B.: Contrail study with
 ground-based cameras, *Atmos. Meas. Tech.*, 6, 3597–3612, <https://doi.org/10.5194/amt-6-3597-2013>, 2013.
- Smith, J. R., Grobler, C., Hodgson, P. J., Mukhopadhaya, J., Shapiro, M. L., Mirolo, M., Stettler, M. E. J., Eastham, S. D., and
 720 Barrett, S. R. H.: The climate opportunities and risks of contrail avoidance, *Nat. Commun.*, <https://doi.org/10.1038/s41467-026-68784-8>, 2026.
- Smolarkiewicz, P., Kühnlein, C., and Wedi, N.: A consistent framework for discrete integrations of soundproof and
 compressible PDEs of atmospheric dynamics, *J. Comput. Phys.*, 263, 185–205, <https://doi.org/10.1016/j.jcp.2014.01.031>,
 725 2014.
- Sonabend-W, A., Elkin, C., Dean, T., Dudley, J., Ali, N., Blickstein, J., Brand, E., Broshears, B., Chen, S., Engberg, Z., Galyen,
 M., Geraedts, S., Goyal, N., Grenham, R., Hager, U., Hecker, D., Jany, M., McCloskey, K., Ng, J., Norris, B., Opel, F.,
 Rothenberg, J., Sankar, T., Sanekommu, D., Sarna, A., Schütt, O., Shapiro, M., Soh, R., Van Arsdale, C., and Platt, J. C.:
 730 Feasibility test of per-flight contrail avoidance in commercial aviation, *Comms. Eng.*, 3, 184, <https://doi.org/10.1038/s44172-024-00329-7>, 2024.
- Sölch, I. and Kärcher, B.: A large-eddy model for cirrus clouds with explicit aerosol and ice microphysics and Lagrangian ice
 particle tracking, *Q. J. Roy. Meteor. Soc.*, 136, 2074–2093, <https://doi.org/10.1002/qj.689>, 2010.

735



Teoh, R., Schumann, U., Majumdar, A., and Stettler, M. E. J.: Mitigating the Climate Forcing of Aircraft Contrails by Small-Scale Diversions and Technology Adoption, *Environ. Sci. Technol.*, 54, 2941–2950, <https://doi.org/10.1021/acs.est.9b05608>, 2020a.

740 Teoh, R., Schumann, U., and Stettler, M. E. J.: Beyond Contrail Avoidance: Efficacy of Flight Altitude Changes to Minimise Contrail Climate Forcing, *Aerospace*, 7, <https://doi.org/10.3390/aerospace7090121>, 2020b.

Urken, A.: Using collective decision system support to manage error in wireless sensor fusion, in: 2005 7th International Conference on Information Fusion, vol. 2, pp. 8 pp.–, <https://doi.org/10.1109/ICIF.2005.1592047>, 2005.

745

Wilhelm, L., Gierens, K., and Rohs, S.: Weather Variability Induced Uncertainty of Contrail Radiative Forcing, *Aerospace*, 8, <https://doi.org/10.3390/aerospace8110332>, 2021.

Xu, M.: Development and Evaluation of Contrail Models, Master's thesis, Massachusetts Institute of Technology,
750 <https://hdl.handle.net/1721.1/155484>, online. Last accessed 15 January 2026, 2024.