

## Environmental Research Communications



## PAPER

## The effect of uncertainty in humidity and model parameters on the prediction of contrail energy forcing

## OPEN ACCESS

## RECEIVED

16 January 2024

## REVISED

31 July 2024

## ACCEPTED FOR PUBLICATION

13 August 2024

## PUBLISHED

17 September 2024

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



John C Platt<sup>1,\*</sup>, Marc L Shapiro<sup>2</sup>, Zebediah Engberg<sup>2</sup>, Kevin McCloskey<sup>1</sup>, Scott Geraedts<sup>1</sup>, Tharun Sankar<sup>1</sup>, Marc E J Stettler<sup>3</sup>, Roger Teoh<sup>3</sup>, Ulrich Schumann<sup>4</sup>, Susanne Rohs<sup>5</sup>, Erica Brand<sup>1</sup> and Christopher Van Arsdale<sup>1</sup>

<sup>1</sup> Google Research, United States of America

<sup>2</sup> Breakthrough Energy, United States of America

<sup>3</sup> Department of Civil and Environmental Engineering, Imperial College London, United Kingdom

<sup>4</sup> Institute of Atmospheric Physics, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Germany

<sup>5</sup> Institute of Energy and Climate Research 8—Troposphere, Forschungszentrum Jülich GmbH, Germany

\* Author to whom any correspondence should be addressed.

E-mail: [platt@google.com](mailto:platt@google.com), [marc.shapiro@breakthroughenergy.org](mailto:marc.shapiro@breakthroughenergy.org), [zeb.engberg@breakthroughenergy.org](mailto:zeb.engberg@breakthroughenergy.org), [mccloskey@google.com](mailto:mccloskey@google.com), [geraedts@google.com](mailto:geraedts@google.com), [tsankar@google.com](mailto:tsankar@google.com), [m.stettler@imperial.ac.uk](mailto:m.stettler@imperial.ac.uk), [roger.teoh15@imperial.ac.uk](mailto:roger.teoh15@imperial.ac.uk), [Ulrich.Schumann@dlr.de](mailto:Ulrich.Schumann@dlr.de), [s.rohs@fz-juelich.de](mailto:s.rohs@fz-juelich.de), [ebrand@google.com](mailto:ebrand@google.com) and [cvanarsdale@google.com](mailto:cvanarsdale@google.com)

**Keywords:** contrails, climate forcing, weather prediction

Supplementary material for this article is available [online](#)

## Abstract

Previous work has shown that while the net effect of aircraft condensation trails (contrails) on the climate is warming, the exact magnitude of the energy forcing per meter of contrail remains uncertain. In this paper, we explore the skill of a Lagrangian contrail model (CoCiP) in identifying flight segments with high contrail energy forcing. We find that skill is greater than climatological predictions alone, even accounting for uncertainty in weather fields and model parameters. We estimate the uncertainty due to humidity by using the ensemble ERA5 weather reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF) as Monte Carlo inputs to CoCiP. We unbiased and correct under-dispersion on the ERA5 humidity data by forcing a match to the distribution of *in situ* humidity measurements taken at cruising altitude. We take CoCiP energy forcing estimates calculated using one of the ensemble members as a proxy for ground truth, and report the skill of CoCiP in identifying segments with large positive proxy energy forcing. We further estimate the uncertainty due to model parameters in CoCiP by performing Monte Carlo simulations with CoCiP model parameters drawn from uncertainty distributions consistent with the literature. When CoCiP outputs are averaged over seasons to form climatological predictions, the skill in predicting the proxy is 44%, while the skill of per-flight CoCiP outputs is 84%. If these results carry over to the true (unknown) contrail EF, they indicate that per-flight energy forcing predictions can reduce the number of potential contrail avoidance route adjustments by 2x, hence reducing both the cost and fuel impact of contrail avoidance.

## 1. Introduction

When aircraft fly in the upper troposphere and lower stratosphere, they form condensation trails (contrails), line-shaped clouds made up of ice particles [1]. Under ice-supersaturated conditions, contrails may persist for several hours and, over this time, shear, spread and merge, transitioning to contrail-cirrus [2]. These persistent contrails make up a significant fraction of aviation's contribution to climate change [3], and have been highlighted as a potentially promising climate impact mitigation opportunity [4].

Contrail reduction efforts are complicated by uncertainties in the radiative forcing of contrails at different times, locations, and seasons. Contrail modeling studies report uncertainty from multiple sources, including the

use of parameterized approximations to physical processes, such as contrail evolution [5], optical properties [6, 7], and radiative transfer [8]. These approximations are used to reduce the computational burden of both microphysical models [9, 10] and global climate simulations [11]. Reducing the uncertainty of contrail forcing is an area of active research [12–16].

When using a microphysics model to estimate the forcing caused by contrails, there are two main sources of uncertainty: the relative humidity of a potential contrail location is not perfectly known, and the model reflects an incomplete representation of ice microphysics, plume evolution and radiative transfer. Uncertainty in relative humidity can be reduced by averaging over many samples of a weather distribution (i.e., the uncertainty is aleatoric). In contrast, the model uncertainty is epistemic: while we can average the outputs of a model over many Monte Carlo simulations with different parameter settings, we have no assurance that the average will converge to the true physical model.

There are two broad approaches to estimating the forcing of contrails with a microphysics model: climatological estimates (e.g. seasonal) or short-term estimates (e.g. per flight). For climatological estimates, the effects of humidity uncertainty are reduced, because we average over many instances of weather states that are drawn from a seasonal or annual average. However, the climatological averaging has an inherent limitation: it ignores the effect of current local weather on contrail forcing. Conversely, the short-term or per-flight predictions have much more weather uncertainty, but can exploit knowledge of current local weather. Previous work was pessimistic that per-flight predictions could ever have high skill, due to the high uncertainty in relative humidity [17].

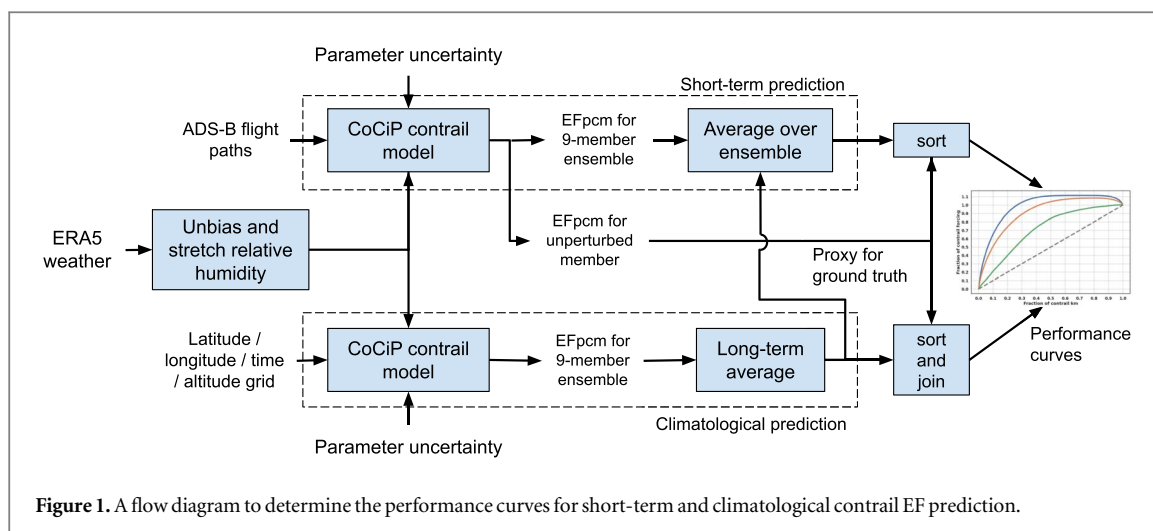
This paper explores a potential way that short-term predictions could have higher skill than climatological approaches, based on two observations. First, most of the energy forcing from contrails comes from a small number of ‘big hit’ contrails [18]. This implies that targeting a limited fraction of flight distance for contrail avoidance maneuvers can reduce the majority of contrail energy forcing as long as these can indeed be predicted. The possibility of high skill on a small fraction of flight distance motivates the current work. Second, relative humidity predictions by weather models tend to be poorly correlated with *in situ* measurements [17, 19]. The current study therefore introduces the use of a non-parametric quantile mapping technique [20] to correct biases in humidity predictions, and also corrects for the under-dispersion that occurs in the output of numerical weather humidity prediction [21].

There are multiple studies that have established that most contrail forcing is caused by a small fraction of flight distance. Explorations into the observed distributions of contrail lifespan using an automated contrail tracking algorithm [22] have estimated that a small fraction of contrails had the longest lifespans [5]. Similarly, reports of distributions of contrail energy forcing (EF) based on reanalysis weather data and parameterized radiative transfer calculations indicate a small fraction of contrails are responsible for a large fraction of forcing [5, 23]. In 2020, Teoh, *et al* [15] found 2.2% of flights contributed 80% of the EF in a study of Japanese airspace. A recent global study [24] finds again roughly 2% of all air traffic is responsible for 80% of the annual global contrail EF between 2019 and 2021. While these studies have explored uncertainty in global contrail EF due to humidity and model uncertainties, they did not evaluate how these uncertainties would affect short-term predictions.

Several studies have reported that numerical weather products, including the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) ensemble [25], have biases and can under-represent the variance of weather properties when compared to reality [26, 27]. In particular, the humidity fields provided by ERA5 are generally under-saturated [28]. The current study shows that using quantile mapping and under-dispersion correction can make ERA5 humidity statistics more aligned with *in situ* measurements. We then apply a Lagrangian contrail model, the Contrail Cirrus Prediction model (CoCiP) [10], to each instance of the 10 ensemble members of ERA5, to make both climatological and short-term predictions of contrail forcing.

This paper demonstrates a method with potentially high skill in predicting ‘big hit’ contrails per flight, but with a number of limitations:

- Most importantly, we do not know the true energy forcing caused by each contrail, so we cannot definitively state that our method has high skill. Instead, we use a proxy for the true energy forcing. We run CoCiP 100 times on every flight, each time choosing one of the 10 ensemble members as the weather. We select one of the ensemble members to approximate the truth. When CoCiP is run on that ensemble member, we treat that as a proxy to the ground truth forcing of the flight’s contrails, rather than as a prediction. We then estimate skill by comparing the CoCiP predictions averaged over 9 ensemble members compared to the CoCiP prediction run on the proxy.
- Second, we limit this study to predicting EF per contrail meter (EFpcm), assuming that a contrail has already formed. This study thus does not consider uncertainties in predicting contrail formation, which we leave to future work.



- Third, this study uses reanalysis weather, rather than a true forecast such as the ECMWF Integrated Forecast System (IFS). Reanalysis may be more accurate than a forecast product. We limit our analysis here to the 10 ensemble members of ERA5 due to data availability and because the skill of predicting forcing even from reanalysis data has been called into question [17, 29].
- Fourth, this paper only considers the short-term EF caused by contrails. Due to feedback effects in the atmosphere, ocean and surface, the actual impact of contrails on the climate has additional factors. Converting the instantaneous radiative forcing to the effective radiative forcing (ERF) introduces further uncertainty which we do not study here as it requires running a climate model [3]. In the Supplementary Material, we use the Monte Carlo CoCiP simulation to generate distributions over contrail cross-section and ice crystal radius, to facilitate further research into long-term climate impact of contrails.

## 2. Methods and materials

The goal of the experiment in this paper is to estimate the skill of predicting ‘big hit’ contrails, including both weather and physical parameter uncertainty, and using a proxy for the ground truth. We estimate this skill for both short-term and climatological predictions.

To estimate the skill of predicting ‘big hit’ contrails, we create a *performance curve*, analogous to a Receiver Operating Characteristic curve [30]. The performance curve for a contrail predictor shows the trade-off between the fraction of contrail-generating flight distance selected by a predictor and the fraction of total contrail EF contained in that distance (section 2.6, figure 5). The more accurate the prediction, the more forcing is concentrated in less distance. A random predictor would produce a diagonal line as the performance curve.

The workflow to produce these performance curves is shown in figure 1.

We first take ERA5 ensemble weather data and make them match the relative humidity observed by in-flight instruments, via unbiasing and correcting under-dispersion (described in section 2.1). We then run CoCiP (described in section 2.2) in two different ways. We apply CoCiP along known flight paths to create short-term EFpcm predictions (described in section 2.4). We also apply CoCiP on a four-dimensional grid to create climatological EFpcm predictions (described in section 2.3). For both of these cases, we perform Monte Carlo simulations of CoCiP that incorporate our lack of certainty of the physical parameters in CoCiP (described in section 2.5).

We create the climatological EFpcm predictions by averaging the grid values produced by CoCiP across three months (as described in section 2.3). We create the short-term EFpcm predictions by averaging the values produced by CoCiP for a waypoint, using the climatological EFpcm values as prior knowledge (as described in section 2.4).

Finally, we compute the short-term and climatological prediction performance curve by using the CoCiP predictions from one ERA5 ensemble member as a proxy for the ground truth. The ERA5 ensemble component consists of one ‘control’ member and nine ‘perturbed’ members; this unperturbed control member is the one selected as the proxy for ground truth, and is not used in any prediction computation. The ground truth proxy is sorted by the predicted EFpcm value and the performance curves are produced (as described in section 2.6).

## 2.1. Unbiasing and removing under-dispersion in ERA5

The first step to compute the performance curve is to correct the ERA5 ensemble member estimates of relative humidity with respect to ice (RH<sub>i</sub>). Like many numerical weather models, the ERA5 ensemble [25] has biases and can under-represent the variance of weather properties when compared to reality [26]. We address these issues by calibrating the ERA5 humidity against *in situ* humidity measurements from the European research infrastructure In-service Aircraft for a Global Observing System (IAGOS) [31, 32]. We choose to calibrate RH<sub>i</sub> because that is the most sensitive parameter for persistent contrail formation and EF, according to recent analysis that investigated the sensitivity of EF with respect to several CoCiP parameters [33]. We obtained the 10-member Ensemble of Data Assimilations ERA5 data from the ECMWF Copernicus Climate Data Store [34].

First, building on the methodology developed by Teoh, Maraun, and others [15, 33, 35], we match the cumulative distribution function of the ERA5 RH<sub>i</sub> values to the RH<sub>i</sub> values from the worldwide IAGOS RH<sub>i</sub> observations from 2019. This matching is done via quantile mapping, where each ERA5 RH<sub>i</sub> value is mapped to an RH<sub>i</sub> value that has the identical quantile in the global IAGOS dataset from the year 2019 (using linear interpolation between quantiles) [20]. We calculate RH<sub>i</sub> values from interpolated ERA5 temperature, pressure, and specific humidity, where the temperature and pressure are interpolated quadrilinearly and the specific humidity is interpolated via a monotonic cubic spline that is fit to the averaged ERA5 specific humidity lapse rate (`interpolation_q_method="cubic-spline"` of [36]). We found it necessary to apply a non-linear interpolation of specific humidity because vertical changes in specific humidity are frequently non-linear and a linear interpolation of specific humidity biased the resulting RH<sub>i</sub> values to be too high. Every ERA5 ensemble member has a distinct quantile map onto the IAGOS 2019 quantiles, spanning 0% to 165% RH<sub>i</sub> [36], and we use 801 quantiles: the smallest sufficient number that qualitatively avoids discretization artifacts in the quantile-mapped humidity values.

Second, weather ensemble members are frequently under-dispersed relative to weather observations [37]. We apply the ‘shift and stretch’ transformation proposed by Eckel *et al* [21] to ensure that the inter-ensemble-member variance of ERA5 RH<sub>i</sub> values (the denominator in (2)) matches the variance between the ERA5 ensemble mean and the IAGOS RH<sub>i</sub> values (the numerator in (2)), averaged over all IAGOS samples. Using the notation of [26], the bias (shift) coefficient  $a$  is given by

$$a = \frac{1}{n} \sum_{t=1}^n (\bar{x}_t - y_t), \quad (1)$$

where  $\bar{x}_t$  is the (histogram-matched) ERA5 ensemble mean RH<sub>i</sub> interpolated at a waypoint  $t$  (out of  $n$  waypoints), and the IAGOS RH<sub>i</sub> measurement is  $y_t$ . The scaling (stretch) coefficient  $c$  is computed as

$$c = \left[ \frac{\frac{m}{(m+1)n} \sum_{t=1}^n (\bar{x}_t - y_t)^2}{\frac{1}{(m-1)n} \sum_{t=1}^n \sum_{k=1}^m (x_{t,k} - \bar{x}_t)^2} \right]^{1/2}, \quad (2)$$

where  $x_{t,k}$  is the (histogram-matched) RH<sub>i</sub> at waypoint  $t$  from ERA5 ensemble member  $k$  (out of  $m$  total members,  $m = 10$ ). The fitted ‘shift’ and ‘stretch’ coefficients  $a$  and  $c$  are  $-6.26 \times 10^{-5}$  and 2.684 respectively (when RH<sub>i</sub> is a percentage ranging 0 to 165); the post-processed RH<sub>i</sub> used by CoCiP ( $y_{t,k}$ ) at a waypoint  $t$  is then given by:

$$y_{t,k} = \bar{x}_t + 0.000\,062\,6 + 2.684(x_{t,k} - \bar{x}_t). \quad (3)$$

We can see that applying histogram-matching already has led to a negligible ‘shift’ coefficient  $a$  as expected. The ‘stretch’ coefficient  $c$  being larger than 1 indicates the ERA5 ensemble members were under-dispersed prior to this transformation; without this preprocessing the uncertainty due to RH<sub>i</sub> in subsequent Monte Carlo analysis could have been under-estimated by a factor of 2.684.

Once the RH<sub>i</sub> of the ERA5 ensemble has been unbiased and stretched so that its distribution matches the distribution of IAGOS, we can use the inter-ensemble-member uncertainty as a proxy to our uncertainty of the actual weather state. Section 3.1 shows the results of the distribution matching.

## 2.2. CoCiP contrail model

Once the ERA5 RH<sub>i</sub> values are corrected, we can feed them to the CoCiP Lagrangian model of contrail energy forcing [10, 38]. We use an implementation of CoCiP published in the open-source pycontrails repository (v0.42.0) [36] that can also evaluate arbitrary grid waypoints independently and in parallel [39, 40].

For each input flight segment to CoCiP, we estimate fuel consumption, overall propulsion efficiency, and non-volatile particulate emissions (nvPM) using the same methods as Global Aviation Emissions Inventory (GAIA) described in [41]. Briefly, GAIA uses BADA 3.15 and 4.2 [42, 43] with regional passenger load factors to estimate aircraft mass, fuel consumption, and overall propulsion efficiency. GAIA uses the ICAO Aircraft Engine Emissions Databank (EDB) [44] to estimate engine nvPM emissions.

CoCiP considers a contrail generated at a segment to be persistent if ice particles nucleated in the exhaust stream survive the initial wake vortex phase. CoCiP evolves contrail points in time using a second-order two-step Runge-Kutta integration scheme [10], here using a time step of 10 minutes until their end of life, which is defined as when the ice number concentration is lower than the background ice nuclei ( $<10^3 \text{ m}^{-3}$ ), contrail optical depth ( $\tau_{\text{contrail}}$ ) is less than  $10^{-6}$ , or when the contrail lifetime exceeds 10 hours [24, 33]. Other CoCiP implementations set longer maximum contrail lifetime limits (12 hours [24, 33], 24 hours [10, 15]); here we select 10 hours to reduce computation time, which clips only a few ( $\approx 0.01\%$ ) of the highest forcing contrail segments [24].

To execute the CoCiP model, we use quadrilinear interpolation to estimate all required ERA5 weather variables (except specific humidity) within the  $0.25^\circ \times 0.25^\circ$  latitude longitude grid at pressure levels [350, 300, 250, 225, 200, 175, 150] hPa provided by ECMWF. As we do when calibrating against the IAGOS data, the specific humidity is interpolated with quadrilinear interpolation in latitude, longitude, time, and via monotonic cubic spline fit to the averaged ERA5 specific humidity lapse rate in pressure levels (`interpolation_q_method="cubic-spline"` of [36]).

### 2.3. Climatological energy forcing prediction

We run the CoCiP model in two different ways: in *grid mode* to compute the climatological predictions, and in *per-flight mode* to compute the short-term predictions. This subsection describes grid mode.

To create climatological EFpcm predictions, we need a consistent space-time basis to average local weather. We create a four-dimensional grid of latitude/longitude (0.5 degree spacing, from  $80^\circ\text{S}$  to  $80^\circ\text{N}$  latitude), altitude (3 pressure levels: 200, 250, and 300 hPa), and time (every 3 hours UTC). Averaging over a box on this grid produces a climatological EFpcm prediction that is independent of flight density. In this flight-independent model, each waypoint represents a flight segment of one meter length placed at the center of the grid cell. The contrail produced by this nominal flight segment evolves according to the dynamics of CoCiP, independent of other grid waypoints. We run the simulation for all time coordinates of the ERA5 EDA data for every fifth day of 2019 (January 3, January 8, January 13...), which produces approximately 405 million grid samples. Each sample was run through CoCiP a total of 100 times, each time with a randomly selected member of the ERA5 EDA.

The aircraft type is set to the default B737 for climatological EF predictions. We select B737 as the single representative aircraft type for climatological predictions because it is the most common aircraft family (B73X) in the individual trajectory dataset. We use BADA 3.15 and 4.2 [42, 43] to estimate nominal aircraft performance conditions at cruise and estimate the emissions profile for the meteorological conditions in each grid cell. The details of the performance and emissions modeling can be found in (Teoh, *et al*, 2023) [24]. Nominal emissions profiles are used to initialize the CoCiP simulation in each grid cell. The sensitivity of climatological predictions to aircraft type is an important area of future research, but is out of scope of the current study.

To make any EF prediction, we desire an EF per contrail-generating flight meter, so any averaging operation in the prediction must only take into account segments that have a non-zero EF. To compute EFpcm (energy forcing per contrail-generating flight meter), we compute

$$\text{EFpcm} = \frac{\sum_i \text{EF}_i}{\sum_i L_i \mathbb{1}(|\text{EF}_i| > 0)} \quad (4)$$

where  $\text{EF}_i$  is the energy forcing from the  $i$ th segment,  $L_i$  is the length of the  $i$ th segment, and  $\mathbb{1}$  is an indicator function that is 1 if its argument is true, 0 otherwise.

To compute the climatological EF, first we convert longitude and UTC time to a local time via  $(\text{UTC hour} + \text{longitude}/15) \bmod 24$ , averaging together points with the same latitude and local time. We then average all points across 3 month periods (Feb/Mar/Apr, May/Jun/Jul, Aug/Sep/Oct, Nov/Dec/Jan), average across 9 weather ensemble members, and average across all altitude levels. We then run a box filter: for every point, we average all points within  $\pm 1.5$  hours of local time and  $\pm 10$  degrees of latitude. We choose the 3 hour window because ERA5 is only available every 3 hours: any smaller window will not average all longitudes equally. We chose 10 degrees of latitude to compensate for the limited number of days of data available: fine zonal details likely reflect weather sampling rather than climatology. In the end, this averaging produces a map of EFpcm that shows the average contrail forcing for 2019, rather than daily weather. The result can be seen in figure 3.

### 2.4. Short-term energy forcing prediction

To estimate EFpcm for a particular flight, we execute CoCiP on ERA5 weather along the trajectory for that flight, in *per-flight mode*. This emulates a day-of-flight prediction system, though a real prediction system would rely on a forecast product rather than a reanalysis like ERA5.

We use flight trajectories provided by the 2019 GAIA dataset derived from Automatic Dependent Surveillance—Broadcast (ADS-B) signals that are provided by Spire Aviation [41]. The flight trajectory ADS-B data includes ICAO aircraft type and 4D positions (longitude, latitude, barometric altitude, and time) recorded



**Table 1.** Nominal values, spread, and uncertainty distribution for CoCiP model parameters.

Parameter	Purpose	Nominal	Spread	Distribution
Initial wake vortex depth	Scales the max contrail downward displacement after the wake vortex phase to set initial contrail depth. Denoted by $C_{D0}$ in equation (14) of [10].	0.5	[0.3, 0.7]	Triangular
Wind shear enhancement exponent	Scales the magnitude of the shear at contrail scale to account for subgrid velocity fluctuations. Denoted by $n$ in equation (39) [10].	0.5	[0, 1]	Triangular
Sedimentation impact factor	Tunable parameter to scale the terminal fall velocity and effective vertical diffusivity. Denoted by $f_i$ in equation (35) of [10]. Nominal values previously reported in [10] as 0.1, but now taken as 0.5 [15].	0.5	$\sigma = 0.1$	Normal

nominally every 5 minutes. We then resample waypoints to a uniform 1 minute resolution using a great circle interpolation in latitude / longitude and implement a linear vertical interpolation [41].

The CoCiP simulation is performed analogously to the climatological energy forcing prediction, except that the actual aircraft and engine type are used. If the engine type is unknown, the nvPM emissions index is set to a constant value of  $10^{15} \text{ kg}^{-1}$  as in (Schumann, *et al.*, 2015) [45] and (Teoh, *et al.*, 2020) [15]. For short-term EF prediction, we define a flight segment as the vector between two waypoints. Contrail segments are initialized when two adjacent waypoints satisfy the Schmidt-Appleman Criterion (SAC) [1, 46, 47].

We simulate a representative sample of flight segments by running Monte Carlo experiments on 1,000 randomly selected flights every third day of 2019 (January 1, January 4, January 7, ...). The 18,220,520 flight segments are run through CoCiP a total of 100 times, with each simulation using a random choice of one of the 10 ERA5 EDA ensemble members.

For each flight segment in the dataset, we averaged the EF predictions by CoCiP run on nine ERA5 ensemble members to form a per-segment EFpcm prediction. In addition to the nine ensemble members, a tenth value was provided to the average: the value for the waypoint from the climatological prediction described in section 2.3. The EFpcm is computed analogously to (4):

$$\text{EFpcm} = \frac{V + \sum_i \text{EF}_i}{L(1 + \sum_i \mathbb{1}(|\text{EF}_i| > 0))} \quad (5)$$

where  $\text{EF}_i$  is the energy forcing predicted by CoCiP on the  $i$ th ensemble member,  $V$  is the climatological prediction evaluated on the segment, and  $L$  is the length of the segment. This averaging reduces the variance of the per-segment prediction, and also provides a value if none of the ensembles produce a non-zero EF, which happens around 2% of the time.

## 2.5. Incorporating model uncertainty

Recall the goal is to estimate the skill of ‘big hit’ contrail prediction, given both humidity and model uncertainty. The humidity uncertainty is captured by the methods described in section 2.3 and section 2.4, but we also want to capture our uncertainty due to CoCiP model parameters. While CoCiP has been applied in several publications and compared to measurements and other model results [48], CoCiP is a highly parameterized model with inherent epistemic limitations. We aim to emulate prediction uncertainty by capturing the uncertainty range in CoCiP model parameters. To do so, every run of CoCiP has model parameters drawn from an uncertainty distribution.

CoCiP includes three critical model parameters with nominal values supported by physical and empirical evidence [10, 15, 33]. Table 1 presents model parameters with a short description of each parameter function and reference to the original publication variable [10, 15]. For each model parameter, we define a range and distribution of possible input values based on observations [48] and prior studies of model uncertainty [15, 33, 49]. We set the center of initial wake vortex depth, wind shear enhancement exponent, and sedimentation impact factor distributions based on the nominal values [10, 15]. The wind shear enhancement exponent is bounded by 0 for stably stratified flows and estimated as 2/3 for isotropic turbulence [10]. We allow the value to exceed 2/3 but place a hard limit at 1. The initial wake vortex depth has hard limits at 0 (since the vortex must be below the plane) and 1 (since it cant exceed the maximum depth). We limit the value to between 0.3 and 0.7 based on figure 7 in [10] and discussion with U. Schumann. Figure S5 and table S1 in the supplement show how parameter values correlate with the mean EF from all Monte Carlo samples run with a specific parameter vector.

**Table 2.** Nominal values, spread, and uncertainty distribution for additional scaling parameters.

Parameter	Purpose	Nominal	Spread	Distribution
RF <sub>SW</sub> scale factor	Scales shortwave radiative forcing (RF <sub>SW</sub> ) to account for the error in model fit. Based on table 2 from [38].	1	$\sigma = 0.106$	Normal
RF <sub>LW</sub> scale factor	Scales longwave radiative forcing (RF <sub>LW</sub> ) to account for the error in model fit. Based on table 2 from [38].	1	$\sigma = 0.071$	Normal
Habit weight mixtures	Accounts for the uncertainty in habit weight mixtures defined as a function of volumetric mean radius [10]. Modifies default values provided in table 2 of [10]	$G_i$	$\alpha_i = 0.5 + 96G_i$	Dirichlet
nvPM EI <sub>n</sub> scale factor	Scales flight nvPM emissions to account for uncertainty in modeling.	1	$\sigma = 0.15$	Log-normal

In addition to the standard CoCiP parameters defined in [10], we define four additional scaling parameters to capture uncertainty in other model parameterizations. Table 2 presents these additional model parameters with a short description of the parameter function and reference to the original publication [15, 38, 50]. We set the distribution of parameters values to be triangular when the range of potential values has clear limits and normal when the range is less well defined. The non-volatile particulate matter emissions index number (nvPM EI<sub>n</sub>) scale factor is defined with a log-normal distribution to account for uncertainties in the nvPM emissions [33, 51].

The scaling parameters for RF<sub>SW</sub> and RF<sub>LW</sub> capture the fit errors in the parametric contrail RF model described in table 1, figure 5 of [38]. The parametric RF model presumes a linear superposition of ice particle habits with habit mixtures defined as a discrete function of volumetric mean radius ( $r_{vol}$ ) when calculating the local contrail RF, the instantaneous change in energy flux per contrail area [10, 52].

While uncertainty in input  $r_{vol}$  (and hence  $r_{eff}$ ) is captured by the dispersion-corrected ERA5 ensemble members, in previous literature the habit mixtures in bands of  $r_{vol}$  are point estimates (see figure 14, [10]). We need to add variance to the point estimates to reflect our uncertainty about the habit mixtures. Statisticians use a Dirichlet distribution to model uncertainty in mixture weights [53]. Samples from a Dirichlet distribution are a vector of non-negative weights that always sum to 1. A Dirichlet distribution has a mean vector of weights, and some spread around that mean. The parameters of the Dirichlet distribution are  $\alpha_i$ , one for each member of the mixture. The mean of the Dirichlet is

$$p_i = \frac{\alpha_i}{\sum_i \alpha_i}, \quad (6)$$

while the variance decreases as  $\sum_i \alpha_i$  increases. To set  $\alpha_i$ , we thus need to set the mean from the  $G_i$ , the original habit fraction from [10], and choose a scaling factor for the  $\alpha_i$  to control how much noise to add to the mean. We set

$$\alpha_i = 0.5 + 96G_i, \quad (7)$$

where 96 was chosen to add approximately  $\pm 10\%$  noise to the dominant habit mixture weights.

For both the short-term and climatological prediction, the CoCiP model is run a total of 100 times for each flight or grid. For each simulation, we randomly select one model parameter vector from a collection of 500 pre-constructed model parameter vectors. This sampling is in addition to the random selection of ERA5 ensemble member. We assume that uncertainties in the CoCiP model parameters are independent. We construct the parameter vector collection ahead of time by randomly sampling the parameter distributions. The CoCiP outputs with the additional parameter variation flow through the prediction system as shown in figure 1. Performing all the CoCiP simulations from sections 2.3-2.4 with the different model parameter vectors used 5,408 CPU hours. Halving the number of CoCiP simulations by reducing the number of flights or the number of parameter vectors changes our results by much less than the error bars reported in this work, suggesting that further increasing the number of CoCiP simulations would not lead to a meaningful change in these results.

## 2.6. Testing the predictions

Once we have short-term and climatological predictions that incorporate both humidity and model uncertainties, we can estimate their respective skills. To estimate the skill of a prediction, we first need to create a performance curve which can vary between no skill and perfect skill.

To create such a curve, we first sort flight segments in descending order based on their predicted EFpcm. We start at 0 contrail kilometers and 0 contrail forcing avoided, with an avoidance threshold of positive infinity. For each segment, if we lower the avoidance threshold just below the predicted EFpcm, we have to avoid the segment

lengths worth of flight, but we also get to avoid the EF. The performance curve would then go to the right by the segment length and up (or down) by the EF of the segment. We loop through all segments, moving right and up/down, until we reach the end of the sorted segments. The result is a performance curve that shows the performance for all possible avoidance thresholds of EFpcm, which can be seen in figure 5. Algorithm 1 presents the pseudocode to generate such a curve.

---

**Algorithm 1.** Pseudocode for generating performance curves

---

```
SortedKeys ← ArgumentSort (PredictedForcing, order = descending)
SortedForcing ← []
SortedLength ← []
for Key in SortedKeys do
  SortedForcing.append (ProxyForcing[Key])
  SortedLength.append (ContrailLength[Key])
CumulativeForcing ← [0]
for Forcing in SortedForcing do
  CumulativeForcing.append(CumulativeForcing[len(CumulativeForcing) - 1]
  + Forcing)
CumulativeLength ← [0]
for Length in SortedLength do
  CumulativeLength.append (CumulativeLength[len(CumulativeLength) - 1] +
  Length)
Plot (x = CumulativeLength/CumulativeLength [len(CumulativeLength) - 1],
y = CumulativeForcing/CumulativeForcing [len(CumulativeForcing) - 1])
```

---

We must establish ‘no skill’ and ‘perfect skill’ performance curves in order to compute a skill quantity. If we feed random values into **PredictedForcing** for algorithm 1, on average it will produce a curve that is a diagonal line with a slope of 1, because the fraction of forcing avoided would equal the fraction of distance avoided. If we set **PredictedForcing** to be **ProxyForcing** in algorithm 1, it will be a perfect predictor whose skill cannot be exceeded, since it has perfect knowledge.

We thus define the skill of a predictor to be the area under the performance curve, scaled so that the diagonal line has 0% skill and the perfect knowledge curve has 100% skill:

$$\text{skill} = \frac{\text{Area under performance curve} - 0.5}{\text{Area under perfect knowledge curve} - 0.5} \quad (8)$$

### 3. Results

We now present results of the experiments: tests of how well we can calibrate the ERA5 humidities versus IAGOS, maps illustrating the climatological prediction, and performance curves for both short-term and climatological predictions.

#### 3.1. Intra-ensemble dispersion of ERA5 relative humidity

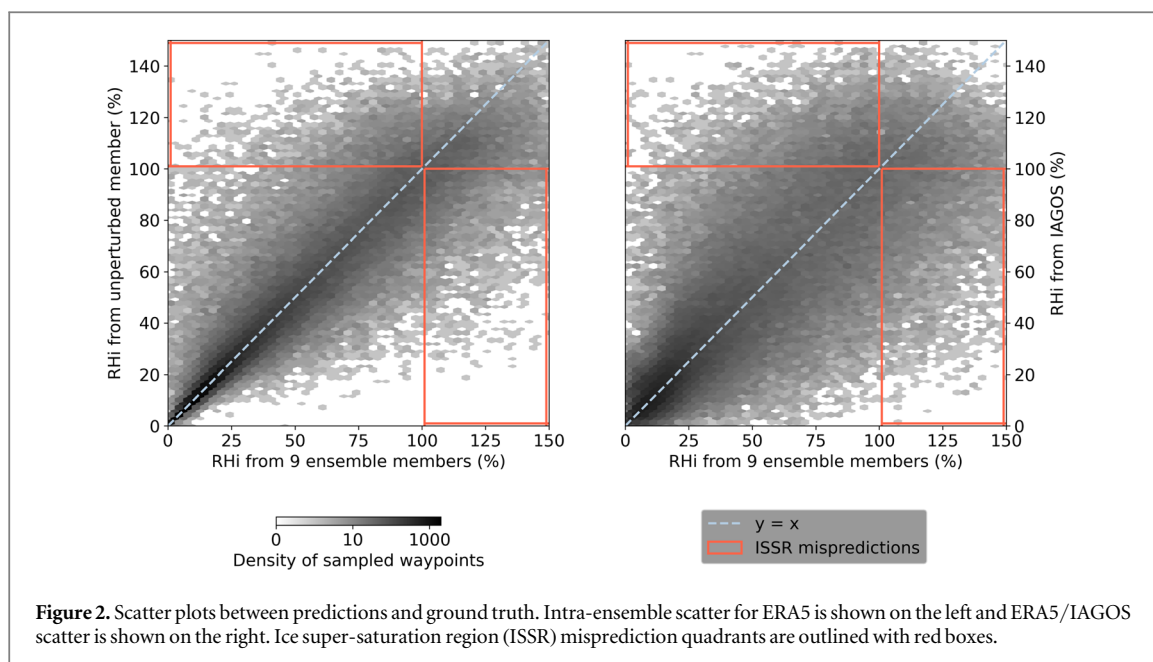
The goal of calibrating ERA5 is to make the unperturbed ERA5 ensemble member RH<sub>i</sub> be a good proxy for the IAGOS RH<sub>i</sub>, which is only known on a sparse set of points. This section performs three tests on proxy quality.

The first test is the Kolmogorov-Smirnov (KS) test comparing the distribution of the unperturbed ERA5 RH<sub>i</sub> to the IAGOS RH<sub>i</sub>. The KS test measures the maximum gap between two cumulative distribution functions (cdfs), which is 0.0203. That is, the cdf of the unperturbed ERA5 RH<sub>i</sub> and the IAGOS RH<sub>i</sub> differ by no more than 2.03%.

The second test is to show that the joint probability density  $p(\text{Unperturbed ERA5 RH}_i, \text{Perturbed ERA5 RH}_i)$  is similar to  $p(\text{IAGOS RH}_i, \text{Perturbed ERA5 RH}_i)$ . This test can be performed qualitatively by creating a density plot of both of these joints, as can be seen in figure 2. Each point in the density plot reflects one sample from the RH<sub>i</sub> ensemble of nine versus a corresponding ground truth (IAGOS) or proxy ground truth (ERA5 unperturbed member).

The joint densities appear qualitatively different at low relative humidity, but are better matching at high relative humidity. Given that we compute EF per contrail meter and assume that we know whether a contrail is persistent, our analysis is only sensitive to calibration at relative humidity above 100%. Table 3 shows a





**Table 3.** Comparison of predicting ice supersaturation in IAGOS and in the ERA5 unperturbed member. To serve as a proxy for IAGOS, the ERA5 prediction should have a similar equitable threat score (ETS) and similar error rates.

Preprocessing	Assumed truth	True negatives	False negatives	False positives	True positives	ETS
None	IAGOS	88.3%	7.3%	2.0%	2.5%	0.179
None	ERA5	93.3%	2.2%	1.5%	3.0%	0.425
Unbias only	IAGOS	85.5%	4.8%	4.8%	4.9%	0.291
Unbias only	ERA5	88.2%	2.1%	2.1%	7.6%	0.609
Unbias + stretch	IAGOS	84.0%	5.1%	6.3%	4.7%	<b>0.241</b>
Unbias + stretch	ERA5	84.8%	4.3%	4.9%	6.0%	<b>0.346</b>

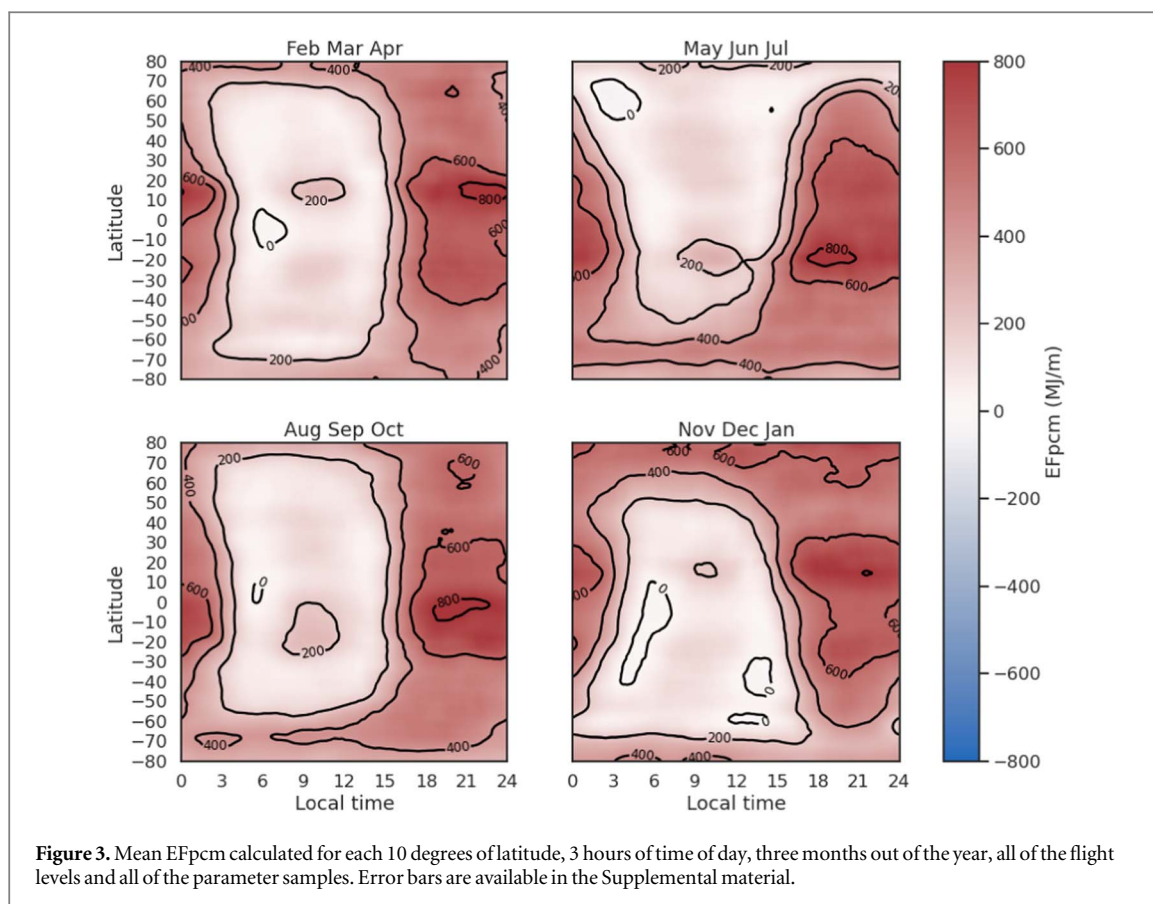
quantitative analysis of the two joint densities. We assess the error rates of using an ensemble member RHi at a space-time point to predict whether a proxy (ERA5 unperturbed) and the ground truth (IAGOS) are ice-supersaturated. There are four possible outcomes to this prediction, which corresponds to the four quadrants of a density plot in figure 2. The lower left is a true negative, the upper right is a true positive, the upper left is a false negative, and the lower right is a false positive. We can compute these four rates by integrating the density in each of the four quadrants. Following [17], we also compute the equitable threat score (ETS) for the ice-supersaturation prediction.

When we do not use any calibration of ERA5, or only use quantile-mapping calibration, ERA5 ensemble members predict the ice supersaturation of the unperturbed ERA5 member with much higher skill than IAGOS, which implies that the weather uncertainty is not yet captured. The ETS without calibration is 0.179, which is comparable to the values reported by [17]. When we use the stretching transformation of (3), the ETS gap between predicting unperturbed ERA5 and predicting IAGOS drops to 0.105, and the largest gap in the rates is only 1.4% in the false positive rates. The matching of the rates and ETS gives us confidence that using the unbiased and stretched ERA5 is a reasonable proxy for predicting the true weather. Even though our intra-ERA5 ETS has a low value of 0.346, we show in the next two sections that the outputs of CoCiP can adequately find ‘big hit’ contrails with highly-positive EFpcm.

The third test of the proxy is to compute calibration curves for the RHi distribution around the proxy. These calibration curves are shown in the Supplementary Material. We find that the P-P plots of the calibration improve when both unbiasing and (3) are used.

### 3.2. Climatological predictions of energy forcing.

Figure 3 shows the EFpcm as a function of local time of day, latitude, and season. The mean EFpcm in our individual simulation is  $309 \pm 38$  MJ/contrail meter, which overlaps the estimate of 331 MJ/contrail meter from [41]. There is significant variation in EFpcm over space and time. Roughly speaking, we find that contrails that form between 3 hours before sunset to 3 hours before sunrise have on average large positive EF, usually  $>400$  MJ/m, while contrails at other times are usually  $<100$  MJ/m. The ‘local time’ axis of figure 3 refers to the



time the aircraft created the contrail. Since contrails last several hours the boundary between the strongly forcing and less forcing contrails happens a few hours before sunset/sunrise.

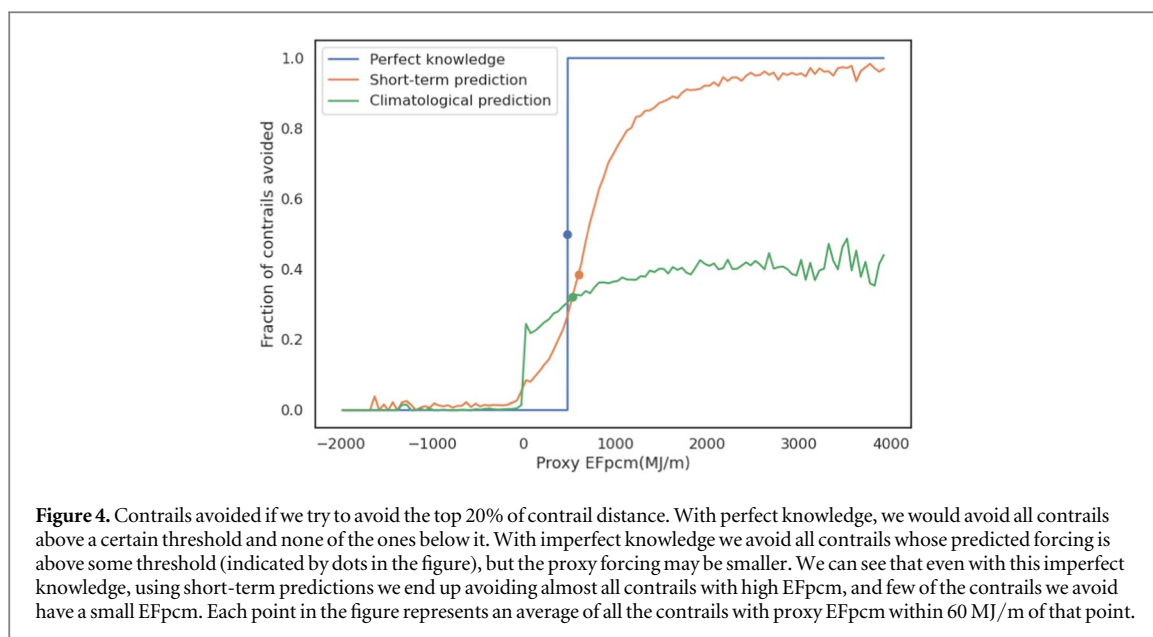
By using a large amount of averaging, we can make a low-variance estimate of contrail forcing. Note however that the values in figure 3 are averages, and an individual flight may deviate from the average value. For all the contrails in our dataset, the root mean squared difference (RMSD) between the proxy contrail forcing (computed from unperturbed ERA5 data) and the values in figure 3 was 598 MJ/m. While this RMSD is almost twice the average EFpcm, in the section below we show that the climatological predictions still have skill in predicting the ‘big hit’ contrails.

### 3.3. Short-term predictions

In the previous section, we computed EFpcm for different latitudes, times of day, and seasons by averaging over ERA5 ensemble members and days. Because of the large variation in the EFpcm of individual flights, we examine whether predicting a short-term EFpcm is a better method of finding ‘big hit’ contrails instead of simply relying on climatological EFpcm. The RMSD between the nominal contrail forcing and the resulting short-term EFpcm prediction is 425 MJ/m, 29% lower than the climatological prediction value from the previous subsection (598 MJ/m). This is because the short-term prediction can use the daily weather to predict almost all of the ‘big hit’ contrails (as we will see in figure 5).

We now investigate how short-term EF predictions can still have skill despite the remaining uncertainty. We start by imagining an intervention with the goal of preventing as much forcing as possible while avoiding 20% of the contrail distance. If we knew the true weather, we could do this by finding the 20% of contrails with largest EF and only avoiding those. In figure 4, we represent this intervention by the ‘Perfect knowledge’ curve, which shows how we avoid no contrails below some threshold and every contrail above it. In the dataset studied in this work, these 20% contrails with largest EF account for 90% of all contrail forcing.

We can also see in figure 4 the impact of having less than perfect knowledge. If we instead choose the 20% of flights with the largest predicted EFpcm, we get a less effective intervention because some flights predicted to have a large forcing actually do not (and vice versa). For example, using our short-term predictions would avoid the 20% of contrails with a predicted EFpcm of  $>595$  MJ/m (the orange dot in the figure). Because our predictions are imperfect, we wind up failing to avoid some contrails whose proxy EFpcm is above this threshold, and we also end up avoiding some contrails whose proxy EFpcm is below this threshold, even though we might have preferred to avoid more strongly forcing contrails instead. Note, however, that the fraction of



highly-forcing contrails avoided (proxy EFpcm >2000 MJ/m) is still quite high. These contrails have so much forcing that even our imperfect short-term prediction is usually correct about whether to avoid them. Similarly, the fraction of contrails with negative forcing (<0 MJ/m) that we avoid is very small. We find that the contrails avoided by an intervention using short-term predictions account for 75% of the total contrail forcing. Though this is not as good as the 90% of forcing avoided with perfect knowledge, if we had no knowledge of which contrails were the contrails with the largest EF we could only avoid 20% of the forcing by avoiding 20% of the contrails. Our short-term predictions enable us to do much better than that.

The climatological prediction is less skillful. Compared to the short term predictions, it avoids less highly-forcing contrails and more slightly-forcing contrails, though it also almost never avoids cooling contrails. The contrails avoided account for 40% of all contrail forcing, which is much less than what we would get with perfect knowledge but still twice as much forcing as we would have prevented with no knowledge.

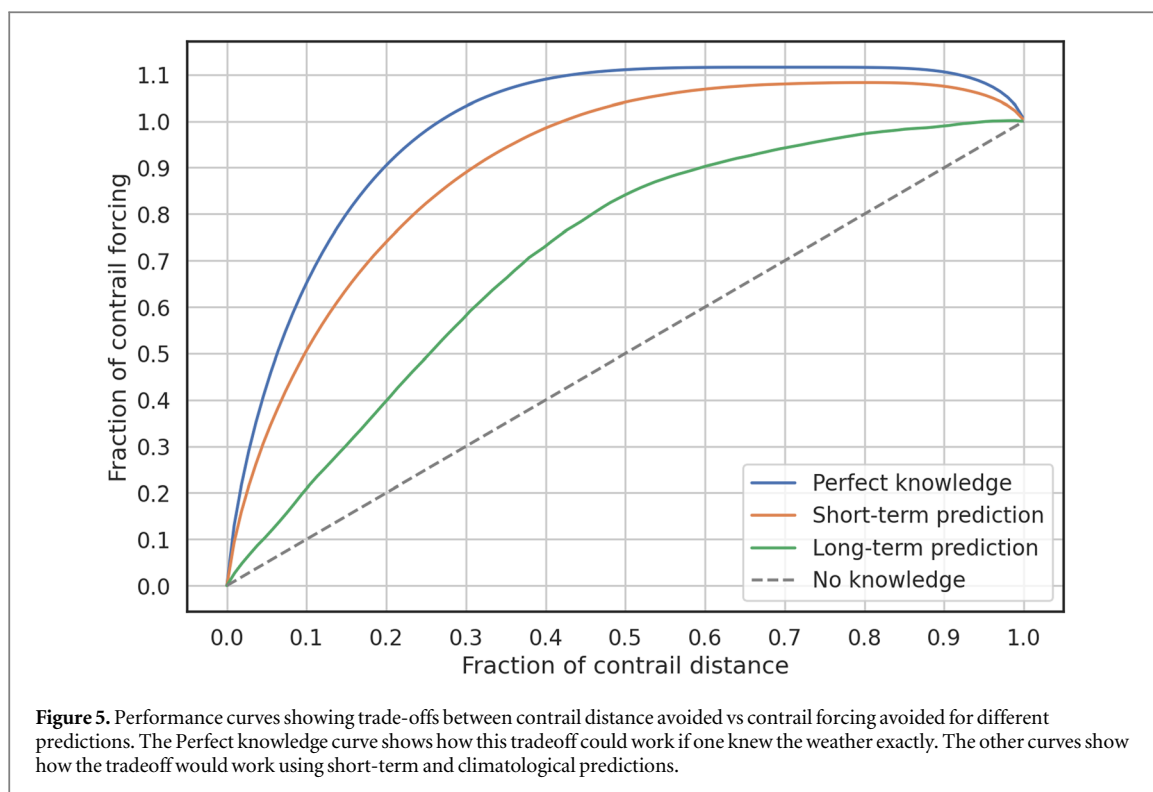
We just described in detail an intervention to avoid 20% of contrails, but of course other interventions are also possible. We can describe how our different prediction methods handle different interventions using the performance curves described in section 2.6. We show such a curve in figure 5. The case where the fraction of contrail distance=0.2 represents the intervention described above, but the curve also maps out all other possible fractions of avoided contrails.

For the perfect knowledge curve, we see that the fraction of contrail forcing avoided initially rises quickly, reflecting the small number of ‘big hit’ contrails that contribute a large amount of forcing. In contrast, if we had no knowledge of the EF of each contrail, the fraction of contrail forcing avoided would be equal to the fraction of contrails avoided (see dashed line in figure 5). Note that these results are roughly consistent with previous work [15, 24, 33] if we account for the fact that around 5% of the total flight distance makes contrails, although since the results are for different locations and times they cannot be directly compared. We also show the skill of the short-term and climatological predictions in figure 5. If a prediction curve matches the perfect knowledge curve, it would have 100% skill. If a prediction curve matches the no knowledge curve, it would have 0% skill.

In particular, we can see that the short-term prediction has high skill in predicting the proxy, which implies that most of the forcing we could avoid with perfect knowledge can also be avoided using short-term predictions. Note that the perfect knowledge curve also assumes perfect knowledge of the CoCiP parameters, while the short-term prediction and climatological prediction curves include uncertainty in the CoCiP parameters.

The perfect knowledge and short-term prediction curves in figure 5 indicate that certain operating points can avoid more than 100% of the net contrail EF, because there are some contrails that have negative EF. These operating points occur when the slopes of the curves are negative. If the contrails that have negative EF are predicted correctly, they get sorted to the right side of the curve and move the curve downwards. In theory, with short-term prediction, 100% of the net forcing from contrails can be avoided by route changes for the 42% contrail kilometers with the highest positive forcing.

Finally, we can compute the skill in predicting the proxy by integrating the area under the short-term or climatological curves and comparing them to the ‘no knowledge’ curve (0% skill) and the ‘perfect knowledge’ curve (100% skill). This yields a short-term skill of 84% and a climatological skill of 44%.



#### 4. Discussion

Despite high intra-weather-ensemble variance of EF estimates, when averaged over nine reanalysis ensemble members, the mean EFpcm produced by CoCiP has skill in predicting ‘big hit’ contrail conditions, using a proxy to the ground truth. This suggests EFpcm could be useful in a contrail prediction system, by combining it with a persistent contrail forecast to create maps of zones which are likely to generate highly-forcing contrails. Flight planners and pilots could then leverage these maps to maximize the amount of contrail forcing avoided while minimizing the cost of contrail avoidance maneuvers. Figure 5 gives guidance for how a per-flight contrail prediction system could work in practice. The results show that the uncertainty in EF across ERA5 ensemble members would not prevent predictions of EF from being useful when deciding how to target mitigation efforts inside these simulations. Imagine one has a goal of mitigating 65% of contrail EF. Without EF predictions, one would need to avoid creating approximately 65% of contrails to achieve this goal. With EF prediction in these simulations, even though the predictions are imperfect, one can achieve the same goal by avoiding only  $\approx 15\%$  of contrails, an improvement of more than 4x. Figure 5 does not specifically inform trade-offs, including added fuel burn and operational considerations, when implementing contrail avoidance in practice. One might decide to try to target 80% of contrail forcing, or try to target the most highly-forcing 10% of contrails. But in all cases in these simulations both the short-term and climatological predictions allow the fraction of forcing prevented to be much larger than the fraction of contrails avoided. Alternatively, the distance of contrail avoidance can be substantially reduced, e.g. if 75% forcing reduction is targeted, then per-flight prediction requires avoidance of only 20% of the contrail distance, rather than 40% from climatological prediction. If this result holds in operational contrail avoidance systems, it would reduce operational burdens and potentially accelerate contrail avoidance as a climate mitigation strategy. Figure 5 also does not cover details that are important to contrail avoidance in practice. The figure only considers a threshold on contrail EF, and does not include uncertainty in predicting whether a persistent contrail will form, nor does it cover other factors such as limitations on contrail avoidance due to air traffic congestion. ‘Big hit’ contrails may be a small fraction of contrails avoided in practice. If there is any reason that ‘big hit’ contrails are systematically ignored by an avoidance process, then the fraction of EF avoided will decrease.

The climatological estimates can be applied beyond the current study to improve baseline contrail forcing estimates when missing specific meteorological conditions or observations. For example, current guidance for company reporting of aviation non-CO<sub>2</sub> emissions suggests multiplying total CO<sub>2</sub> emissions by a factor of 1.9 [54] regardless of time and location of flight. The climatological estimates could form the foundation of a basic algorithm to account for average contrail forcing based on origin–destination pair and time of departure. The

same system could be used as a fallback option for monitoring, reporting, and verification systems while higher fidelity mechanisms are in development and implementation.

This study is one step towards an end-to-end live contrail management system. More research is needed to create a production system:

- Though we used ERA5 ensemble members for both prediction and a proxy to ground truth, a more realistic setup would be to use a forecast product (such as ECMWF IFS) to make predictions. An interesting future research direction would be to evaluate how the prediction uncertainties increase with the lead time of the short-term forecasts.
- Using a CoCiP calculation on one ERA5 ensemble member as a proxy for ground truth as in this analysis may result in mis-estimating the true performance curve, to the degree that CoCiP failing to accurately model contrail phenomena substantially changes the ordering of which flight segments actually caused the largest contrail energy forcing. In particular, CoCiP does not take into account long-term climate feedback effects like a climate simulation can.
- Predicting contrail formation is currently a difficult problem if based purely on the humidity fields provided by numerical weather predictions [17, 19]. Improvements by incorporating other data, such as contrail detections [55, 56] or other numerical weather fields [57] into the predictions using machine learning may be possible.
- Validating the properties of contrails after the fact is useful for both scientific and regulatory purposes. It is currently often possible to observationally validate whether a given flight has made a persistent contrail [58, 59], however the total EF produced by a contrail is difficult to estimate from observations.
- Observations could be used to measure a ground truth of energy forcing in future work. Current work in observational estimation of EF has high uncertainties for multiple reasons, among them contrail cirrus morphology being very similar to natural cirrus, and small sample sizes relative to the total population [60].

This paper probes the uncertainty of the EF of contrails using CoCiP that arise from humidity and model parameter uncertainties. The paper does not assess the uncertainty of the CoCiP model itself: given the uncertainties in, e.g. ice habit distribution, future work may improve upon CoCiP, reduce its uncertainties, or even replace it with a different model. Previous work in estimating effective radiative forcing (ERF) from RF uses global flight patterns [61–63]. More research is required to understand the ERF impact of mitigating only ‘big hit’ contrails.

## 5. Conclusions

We have shown that when calibrating the RH<sub>i</sub> predictions of the ERA5 ensemble members, the resulting ensemble predictions for short-term contrail energy forcing obtained by use of CoCiP — while accounting for uncertainty in model parameters — are more skillful than climatological predictions alone, under the assumption that one of the ensemble CoCiP predictions corresponds to the ground truth. To our knowledge, this is the first exploration of global Monte Carlo contrail simulations propagating both extensive parameter uncertainty and uncertainty from (bias- and dispersion-corrected) relative humidity meteorological inputs.

We measure the skill of the predictions against the proxy by constructing a performance curve of fraction of total contrail forcing versus fraction of contrail kilometers for all possible decision thresholds. Skill is the area under the performance curve, scaled so that random decisions are 0% and perfect performance is 100%. Under this setup, the skill of the climatological predictions is 44% while the skill of the per-flight predictions is 84%.

## Acknowledgments

This work was entirely performed with no external funding.






Flight-level and map-level experimental data are available on Zenodo at <https://doi.org/10.5281/zenodo.12746362>. This includes the 500 random parameter vectors used in the Monte Carlo simulation, and the 500 samples of ice crystal sizes and contrail cross-sectional area (described in the Supplementary Material), to enable further contrail climate modeling. **Author contributions:** JCP, MLS, ZE designed and performed experiments and wrote the paper. KM, SG contributed algorithms, made figures, and edited the paper. TS did experiments. MEJS, RT, US contributed ideas and edited the paper. SR contributed data to the experiments. EB and CVA organized the work and gave feedback on the paper.



## Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://doi.org/10.5281/zenodo.12746362>.

## ORCID iDs

John C Platt  <https://orcid.org/0000-0002-5652-5303>  
Marc L Shapiro  <https://orcid.org/0000-0003-0864-6394>  
Kevin McCloskey  <https://orcid.org/0000-0001-9967-4117>  
Scott Geraedts  <https://orcid.org/0009-0001-6496-0153>  
Tharun Sankar  <https://orcid.org/0000-0002-0557-0972>  
Marc E J Stettler  <https://orcid.org/0000-0002-2066-9380>  
Ulrich Schumann  <https://orcid.org/0000-0001-5255-6869>  
Susanne Rohs  <https://orcid.org/0000-0001-5473-2934>  
Erica Brand  <https://orcid.org/0009-0003-8173-9914>  
Christopher Van Arsdale  <https://orcid.org/0000-0001-8473-6725>

## References

- [1] Schumann U 1996 On conditions for contrail formation from aircraft exhausts *Meteorol. Z.* **4**–23
- [2] Haywood J M, Allan R P, Bornemann J, Forster P M, Francis P N, Milton S, Rädcl G, Rap A, Shine K P and Thorpe R 2009 A case study of the radiative forcing of persistent contrails evolving into contrail-induced cirrus *J. Geophys. Res. D: Atmos.* **114** D24201
- [3] Lee D S *et al* 2021 The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018 *Atmos. Environ.* **244** 117834
- [4] Caldeira K and McKay I 2021 Contrails: tweaking flight altitude could be a climate win *Nature* **593** 341
- [5] Gierens K M and Vazquez-Navarro M 2018 Statistical analysis of contrail lifetimes from a satellite perspective *Meteorol. Z.*
- [6] Markowicz K M and Witek M 2011 Sensitivity study of global contrail radiative forcing due to particle shape *Journal of Geophysical Research: Atmospheres* **116** D23203
- [7] Myhre G and Stordal F 2001 On the tradeoff of the solar and thermal infrared radiative impact of contrails *Geophys. Res. Lett.* **28** 3119–22
- [8] Myhre G *et al* 2009 Intercomparison of radiative forcing calculations of stratospheric water vapour and contrails *Meteorol. Z.* **18**
- [9] Fritz T M, Eastham S D, Speth R L and Barrett S R 2020 The role of plume-scale processes in long-term impacts of aircraft emissions *Atmos. Chem. Phys.* **20** 5697–727
- [10] Schumann U 2012 A contrail cirrus prediction model *Geoscientific Model Development* **5** 543–80
- [11] Kärcher B 2018 Formation and radiative forcing of contrail cirrus *Nat. Commun.* **9** 1–17
- [12] Bräuer T *et al* 2021 Reduced ice number concentrations in contrails from low aromatic biofuel blends *Atmos. Chem. Phys.* (<https://doi.org/10.5194/acp-21-16817-2021>)
- [13] Märkl R S *et al* 2023 Powering aircraft with 100% sustainable aviation fuel reduces ice crystals in contrails *Atmospheric Chemistry and Physics* **24** 3813–37
- [14] Sridhar B, Ng H K and Chen N Y 2012 Aircraft trajectory optimization and contrails avoidance in the presence of winds *J. Guid. Control Dyn.*
- [15] Teoh R, Schumann U, Majumdar A and Stettler M E J 2020 Mitigating the climate forcing of aircraft contrails by small-scale diversions and technology adoption *Environ. Sci. Technol.* **54** 2941–50
- [16] Voigt C *et al* 2021 Cleaner burning aviation fuels can reduce contrail cloudiness *Communications Earth & Environment* **2** 1–10
- [17] Gierens K, Matthes S and Rohs S 2020 How well can persistent contrails be predicted? *Aerospace* **7** 169
- [18] Royal Aeronautical Society. Greener by design; annual report 2018–2019. Technical report, London, UK, 2019. (<https://aerosociety.com/media/12007/greener-by-design-report-2018-2019.pdf>)
- [19] Agarwal A, Meijer V R, Eastham S D, Speth R L and Barrett S R H 2022 Reanalysis-driven simulations may overestimate persistent contrail formation by 100%–250% *Environ. Res. Lett.* **17** 014045
- [20] Panofsky H A and Brier G W 1968 Some applications of statistics to meteorology *Earth and Mineral Sciences Continuing Education, College of Earth and Mineral Sciences*
- [21] Eckel F A, Allen M S and Sittel M C 2012 Estimation of ambiguity in ensemble forecasts *Weather Forecast.* **27** 50–69
- [22] Vazquez-Navarro M, Mannstein H and Mayer B 2010 An automatic contrail tracking algorithm. *Atmospheric Meas. Tech.* **3** 1089–101
- [23] Wilhelm L, Gierens K and Rohs S 2022 Meteorological conditions that promote persistent contrails *NATO Adv. Sci. Inst. Ser. E Appl. Sci.* **12** 4450
- [24] Teoh R, Engberg Z, Schumann U, Voigt C, Shapiro M, Rohs S and Stettler M 2023 Global aviation contrail climate effects from 2019 to 2021 *EGU sphere* **2023** 1–32
- [25] Hersbach H *et al* 2020 The ERA5 global reanalysis *Quart. J. Roy. Meteor. Soc.* **146** 1999–2049
- [26] Wilks D S 2018 Univariate ensemble postprocessing *Statistical Postprocessing of Ensemble Forecasts* (Elsevier) pp 49–89
- [27] Wolf K, Bellouin N, Boucher O, Rohs S and Li Y 2023 Correction of temperature and relative humidity biases in ERA5 by bivariate quantile mapping: Implications for contrail classification (<https://doi.org/10.5194/egusphere-2023-2356>)
- [28] Reutter P, Neis P, Rohs S and Sauvage B 2020 Ice supersaturated regions: properties and validation of ERA-Interim reanalysis with IAGOS in situ water vapour measurements *Atmos. Chem. Phys.* **20** 787–804
- [29] Hofer S, Gierens K and Rohs S 2024 How well can persistent contrails be predicted? an update *Atmos. Chem. Phys.* **24** 7911–25
- [30] Marzban C 2004 The ROC curve and the area under it as performance measures *Weather and Forecasting* **19** 1106–14
- [31] Boulanger D, Blot R, Bundke U, Gerbig C, Hermann M, Nédélec P, Rohs S and Ziereis H 2018 IAGOS final quality controlled observational data L2—time series (<https://iagos.aeris-data.fr/landing-page/?uuid=575882c0-64ce-4648-bb19-00030d5d63af>)

- [32] Petzold A *et al* 2020 Ice-supersaturated air masses in the northern mid-latitudes from regular in situ observations by passenger aircraft: vertical distribution, seasonality and tropospheric fingerprint *Atmos. Chem. Phys.* **20** 8157–79
- [33] Teoh R, Schumann U, Gryspeerdt E, Shapiro M, Molloy J, Koudis G, Voigt C and Stettler M E J 2022 Aviation contrail climate effects in the North Atlantic from 2016 to 2021 *Atmos. Chem. Phys.* **22** 10919–35
- [34] ECMWF. Copernicus climate data store. <https://cds.climate.copernicus.eu/#!/home>. Accessed: 2023-12-27.
- [35] Maraun D 2016 Bias correcting climate change simulations—a critical review *Current Climate Change Reports* **2** 11–20
- [36] Shapiro M, Engberg Z, Teoh R, Stettler M and Dean T 2023 pycontrails: Python library for modeling aviation climate impacts
- [37] Raftery A E, Gneiting T, Balabdaoui F and Polakowski M 2005 Using Bayesian model averaging to calibrate forecast ensembles *Mon. Weather Rev.* **133** 1155–74
- [38] Schumann U, Mayer B, Graf K and Mannstein H 2012 A parametric radiative forcing model for contrail cirrus *J. Appl. Meteorol. Climatol.* **51** 1391–406
- [39] Engberg Z, Teoh R, Abbott T, Dean T, Stettler M E J and Shapiro M L 2024 Forecasting contrail climate forcing for flight planning and air traffic management applications: the CocipGrid model in pycontrails 0.51.0 *EGUsphere* **2024** 1–45
- [40] Shapiro M L, Engberg Z, Zugic B, Teoh R, Stettler M E J, Schumann U and McKay I 2022 Forecasting contrail climate forcing for flight planning and air traffic management applications *The V International Conference on Transport, Atmosphere and Climate (TAC-5)* ([https://pa.op.dlr.de/tac/2022/archiv/web/presentations/29\\_Wed\\_PM/29-18Shapiro\\_Wednesday\\_oral.pdf](https://pa.op.dlr.de/tac/2022/archiv/web/presentations/29_Wed_PM/29-18Shapiro_Wednesday_oral.pdf))
- [41] Teoh R, Engberg Z, Shapiro M, Dray L and Stettler M 2024 A high-resolution global aviation emissions inventory based on ADS-B (GAIA) for 2019–2021 *Atmospheric Chemistry and Physics* **24** 725–44
- [42] Baulleret P 1998 *User Manual for the Base of Aircraft Data (BADA) - Revision 3.1*
- [43] Nuic A, Poles D and Mouillet V 2010 BADA: an advanced aircraft performance model for present and future ATM systems *Int. J. Adapt Control Signal Process.* **24** 850–66
- [44] EASA 2021 ICAO aircraft engine emissions databank URL: (<https://easa.europa.eu/domains/environment/icao-aircraft-engineemissions-databank>)
- [45] Schumann U, Penner J E, Chen Y, Zhou C and Graf K 2015 Dehydration effects from contrails in a coupled contrail-climate model *Atmos. Chem. Phys.* **15** 11179–99
- [46] Appleman H 1953 The formation of exhaust condensation trails by jet aircraft *Bull. Am. Meteorol. Soc.* **34** 14–20
- [47] Schmidt E 1941 Die entstehung von eisnebel aus den auspuffgasen von flugmotoren *Schriften der Deutschen Akademie der Luftfahrtforschung, Verlag R. Oldenbourg, München, Heft 44* vol 5 (Verlag R. Oldenbourg, München) pp 1–15
- [48] Schumann U *et al* 2017 Properties of individual contrails: a compilation of observations and some comparisons *Atmos. Chem. Phys.* **17** 403–38
- [49] Schumann U and Graf K 2013 Aviation-induced cirrus and radiation changes at diurnal timescales *J. Geophys. Res. D: Atmos.* **118** 2404–21
- [50] Schumann U, Mayer B, Gierens K, Unterstrasser S, Jessberger P, Petzold A, Voigt C and Gayet J-F 2011 Effective radius of ice particles in cirrus and contrails *J. Atmos. Sci.* **68** 300–21
- [51] Teoh R, Stettler M E J, Majumdar A, Schumann U, Graves B and Boies A M 2019 A methodology to relate black carbon particle number and mass emissions *J. Aerosol Sci.* **132** 44–59
- [52] Baum B A, Heymsfield A J, Yang P and Bedka S T 2005 Bulk scattering properties for the remote sensing of ice clouds. part i: Microphysical data and models *J. Appl. Meteorol. Climatol.* **44** 1885–95
- [53] Murphy K P 2022 *Probabilistic Machine Learning: An Introduction* (MIT Press)
- [54] Thistlethwaite G *et al* 2022 *Government greenhouse gas conversion factors for company reporting: Methodology paper* Department of Business, Energy, and Industrial Strategy, Government of the UK (<https://assets.publishing.service.gov.uk/media/62aee1f90e0765d523ca33/2022-ghg-cf-methodology-paper.pdf>)
- [55] Meijer V R, Kulik L, Eastham S D, Allroggen F, Speth R L, Karaman S and Barrett S R H 2022 Contrail coverage over the United States before and during the COVID-19 pandemic *Environ. Res. Lett.* **17** 034039
- [56] Ng J Y-H, McCloskey K, Cui J, Meijer V R, Brand E, Sarna A, Goyal N, Van Arsdale C and Geraedts S 2024 Contrail detection on GOES-16 ABI with the OpenContrails dataset *IEEE Transactions on Geoscience and Remote Sensing* **62** 1–14
- [57] Wang Z, Bugliaro L, Gierens K, Hegglin M I, Rohs S, Petzold A, Kaufmann S and Voigt C 2024 Machine learning for improvement of upper tropospheric relative humidity in ERA5 weather model data *EGUsphere* **2024** 1–28
- [58] Chevallier R, Shapiro M, Engberg Z, Soler M and Delahaye D 2023 Linear contrails detection, tracking and matching with aircraft using geostationary satellite and air traffic data *Aerospace* **10** 578
- [59] Geraedts S *et al* 2024 A scalable system to measure contrail formation on a per-flight basis *Environ. Res. Commun.* **6** 015008
- [60] Vázquez-Navarro M, Mannstein H and Kox S 2015 Contrail life cycle and properties from 1 year of MSG/SEVIRI rapid-scan images *Atmos. Chem. Phys.* **15** 8739–49
- [61] Bickel M, Ponater M, Bock L, Burkhardt U and Reineke S 2020 Estimating the effective radiative forcing of contrail cirrus *J. Clim.* **33** 1991–2005
- [62] Ponater M, Marquart S, Sausen R and Schumann U 2005 On contrail climate sensitivity *Geophys. Res. Lett.* **32**
- [63] Rap A, Forster P M, Haywood J M, Jones A and Boucher O 2010 Estimating the climate impact of linear contrails using the UK met office climate model *Geophys. Res. Lett.* **37** L20703