

Flyers can identify the most-warming flights when booking

Steven Davis

sjdavis@stanford.edu

Stanford University <https://orcid.org/0000-0002-9338-0844>

Silas Whiteson

Colorado College

Xavier Bonnemaizon

Laboratoire des Sciences du Climat et de l'Environnement, Paris, France

Ken Caldeira

Carnegie Institution for Science <https://orcid.org/0000-0002-4591-643X>

Roger Teoh

Department of Civil and Environmental Engineering, Imperial College London

Marc Shapiro

Contrails.org

Physical Sciences - Article

Keywords:

Posted Date: October 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11285103/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

Aircraft contrails and the cirrus clouds they seed may be warming the planet about as much as all the CO₂ aviation has ever emitted. But whereas CO₂ warming persists for centuries, contrails dissipate in hours; to the extent we can avoid making them, the warming they cause will fade within years. And a small fraction of flights produce a disproportionately large share of the contrail warming. Researchers have previously demonstrated the potential for airlines to adjust flight scheduling and routes to avoid the most-warming contrails, but here we show that it is also possible to predict many of the most-warming flights weeks in advance using only information available when booking (and selectively avoid them). Our prediction model, trained on flight-specific, process-based simulations of contrail energy forcing (hereinafter “contrail forcing”) from 52 million commercial flights in 2019 and 2021, uses only features of flights knowable from a published flight schedule (e.g., origin, destination, scheduled time, aircraft type, and great-circle geometry but no atmospheric data) to correctly identify the 10% of flights responsible for 68% of total contrail forcing, and a 5% subset responsible for 47%. Applied to real annual flight logs from two corporations, avoiding the predicted worst 10% in favor of same-route alternatives departing within ± 16 hours of the original flight could avoid 14–29% of contrail forcing, and this share increases to 42–54% if any same-route flight is acceptable (regardless of timing). Our findings show that booking-time information could differentiate flights by expected contrail forcing and potentially shift demand toward lower-warming schedules and routes.

Introduction

The climate impact of aviation includes long-lived CO₂ emissions from fossil kerosene (jet fuel) combustion as well as short-lived but intense radiative forcing primarily related to contrail cirrus and aviation-induced cloudiness^{1,2}. Integrated over 100 years, the mean CO₂ and non-CO₂ warming contributions of aviation are estimated to be of similar magnitude despite the latter acting on far shorter (hours) time scales¹. Yet climate accounting frameworks have historically omitted contrail forcing³, perhaps because contrail forcing is complex with large uncertainties: contrail formation requires temperature and humidity in the cruise layer to cross specific thresholds, and the resulting cirrus clouds can warm or cool the surface depending on time of day, aircraft-engine emissions (more fuel burn typically corresponds to more particles), and background radiation fields⁴⁻⁸.

Despite this complexity, guidance from the Greenhouse Gas (GHG) Protocol and U.K. Department of Environment, Food, and Rural Affairs has for many years given corporations the option to estimate the CO₂-equivalence of non-CO₂ forcing (including contrails) by applying an average multiplier to the CO₂ emissions of their business flights^{9,10}. But recent process-based simulations have underscored the extreme heterogeneity of contrail forcing, suggesting that roughly 2% of flights account for 80% of contrail forcing¹¹. Such concentration of forcing points to the need for flight-specific contrail modeling, which the EU Emissions Trading System began requiring for flights within Europe in 2025, and will extend to all flights in 2027¹².

Previous simulations have shown that airlines could re-route flights in real time based on atmospheric forecasts, whereby contrail forcing could be reduced by 55-75% at a cost of <\$1 per ton of CO₂e forcing¹³⁻¹⁶. This research has led to several active efforts by airlines, air navigation service providers, and policy makers to test such interventions¹⁶⁻²². These trials involve coordination of dispatchers, controllers, flight planners, and pilots based on short-term meteorological forecasts (e.g., minutes to hours). But there has been little effort to empower climate-conscious flyers when buying tickets; steering their demand to low-contrail flights could reward and encourage the operational investments airlines are exploring.

Here, therefore, we investigate the extent to which flights' contrail forcing can be predicted at the time of booking (weeks-to-months ahead of departure), which would allow consumers and corporate purchasers to avoid the highest-warming flights and selectively choose alternatives. Whether schedule-level features such as origin, destination, departure time, aircraft type carry sufficient predictive signal to identify a meaningful subset of warming flights, or whether the variability of upper-troposphere humidity²³ confounds any structural pattern, has not been quantified.

Details of our analytic approach are in the Methods. In summary, we train a machine learning model using per-flight contrail forcing (i.e. the energy forcing of a flight's contrails integrated over their lifetime and spatial extent, in joules; summed over many flights it is a net quantity that includes the small negative contribution of contrails that cool) for 52 million commercial flights in 2019 and 2021 computed using a Contrail Cirrus Prediction (CoCiP) process model²⁴ and European Centre for Medium-Range Weather Forecast (ECMWF) reanalysis weather²⁵, whereby the actual flight trajectories are provided by the global aviation emissions inventory based on automatic dependent surveillance – broadcast (GAIA)²⁶. Our regression target is the signed log of per-km contrail energy forcing, expressed in CO₂-equivalent units, so that a single model encompasses cooling, neutral and warming flights. We restrict model features to those knowable from a published flight schedule alone: coordinates of origin and destination airports, great-circle distance, day of year and scheduled departure hour, aircraft ICAO type, sun-altitude-weighted night fraction along the great-circle route (computed from solar-position only), and the fraction of the route over land. The resulting model is further validated against independently-published values for concentration of energy forcing across flights, the share of flights forming net-warming contrails, diurnal and seasonal asymmetries, and regional structure (Supplementary Table S1), and by comparing predictions for 273 flights from four seasonal days with forcing re-simulated with the open-source pycontrails (v0.63) implementation of CoCiP²⁷. Finally, we apply our schedule-only model to two anonymized corporate flight logs from 2021 to demonstrate potential reductions as a function of assumed booking flexibility.

Results

Geographical and temporal patterns of contrail forcing

Whether a given contrail warms or cools depends on the radiation it intercepts. Contrails always trap a fraction of outgoing terrestrial (longwave) radiation, a warming effect (wavy red arrows in Fig. 1a). In daylight, they also reflect part of the incoming solar radiation back to space, which cools relative to the no-contrail counterfactual (yellow arrows in Fig. 1a). As a result, daytime contrails can have weak or even negative contrail forcing, but at night contrails only trap heat and are always warming. At high latitudes, contrail forcing tends to be stronger in winter when solar reflection is weak or absent, whereas greater incoming solar radiation in summer can increase the potential for daytime cooling more strongly with season. Mapped to flown great-circle waypoints, the highest mean contrail forcing per passenger-kilometer occurs over Europe, the North Atlantic, and northern Asia—major mid-latitude air corridors where cold, ice-supersaturated upper-tropospheric air is encountered most frequently (Figs. 1b and 1c).

These geographic and temporal patterns mean that per-flight contrail forcing is heterogeneous and concentrated. Across all 52.5 million flights in our 2019 and 2021 datasets, the worst 5% of flights by fully-modeled (i.e. CoCiP with weather data) contrail forcing account for roughly 95% of all contrail forcing, and the worst 10% for 103% (more than 100% because the total is offset by cooling flights; Fig. 2a). The remaining flights have much weaker forcing, including some cooling, but just 6.4% of flights account for all of the contrail forcing (Fig. 2a). Splitting these flights into different subsets illustrates that long-haul flights (23% of all flights, here defined as >2000 km) account for 71% of energy forcing and twice the contrail forcing per km of shorter flights, both because they remain at (contrail-prone) cruise altitudes for a greater share of their duration and because they are mostly flown by widebody aircraft that emit more soot particles per km and so form contrails of smaller, longer-lived ice crystals (widebodies are 13% of flights but 44% of net forcing)²⁸. However, warming is more concentrated among shorter flights (3.6% of flights <2000 km account for 100% of their contrail forcing, compared to 16% of long-haul flights; Fig. 2b).

Afternoon and evening departures similarly dominate warming because their contrails form late in the day and may persist into night (when they do not reflect sunlight). Morning departures contribute little contrail forcing because their contrails mostly live out their lives in daylight; what warming they do contribute is disproportionately in the long-haul minority whose contrails outlast the day (1.6% of morning flights account for 100% of their contrail forcing; Fig. 2c). Seasonally, winter and autumn flights (here categorized by boreal calendar since >90% of flights originate and arrive in Northern hemisphere) contribute disproportionately to contrail forcing since cold-season cruise-level air is more often ice-supersaturated and shorter days limit reflection of sunlight, but concentration among flights is similar in every season (5–8% of each season's flights account for all of their contrail forcing; Fig. 2c).

Flights that at any point reach latitudes >40° (roughly half of all flights) account for roughly 3x the contrail forcing of flights that do not (because of colder troposphere and less sunlight to reflect), but warming among low-latitude flights is especially concentrated in long-haul, afternoon-and-evening departures that spend time cruising in the dark in areas where air is more often ice-supersaturated (i.e. the humid deep tropics and the subtropical margin; Fig. 2e). Regionally, North America's large share of contrail forcing (36%) roughly matches its share of air traffic (40% of flights), whereas Europe (20% of

flights; 30% of contrail forcing) and especially Russia (2% of flights, 9% of contrail forcing) are disproportionately warming, often flying in the high latitude corridors (Fig. 2f). In contrast, Asia flights are disproportionately less warming, with what warming there is quite concentrated (3.0% of Asia flights account for 100% of their contrail forcing), consistent with the drier, subtropical cruise altitudes of much of that traffic (Fig. 2f).

Scheduling model skill and key features

Using 14 flight features that are knowable at the time of booking, the flights our scheduling model flags as most-warming account for most of the contrail forcing in the withheld test data estimated by the CoCiP model (which uses reanalysis meteorological data that are not knowable that far ahead, and whose results we treat as true throughout). For example, the 10% of flights our scheduling model predicts to be most-warming represent 68% of the total contrail forcing compared to 109% (more than 100% because the total is offset by cooling flights) when flights are re-ranked by the meteorology-aware model (red “model” curve versus blue “perfect” curve in Fig. 3a). In contrast, simply ranking the 10% of flights with the least mean insolation (i.e. solar radiation) along their predicted routes captures only 20% of contrail forcing (green curve in Fig. 3a). The scheduling model is equally skilled when 2019 and 2021 are separated: the 10% predicted to be most-warming represent 67.4% and 68.1% of contrail forcing, respectively), suggesting that the model captures persistent rather than year-specific structure. Thus, although the scheduling model only explains 15% of overall variance in the test data ($R^2=0.149$ in signed-log space), its predicted worst decile captures about 60% of the warming that perfect foresight could achieve. We also test the scheduling model’s skill in predicting the most-cooling flights. The 10% of flights with the most cooling contrails offset 12% of total contrail forcing, but the 10% of flights our model predicts to be least-warming offset just 3%, and the 10% with the most direct sunlight along their routes offset none (blue, red, and green curves in Fig. 3b, respectively).

We further evaluate the relative importance of different features to the scheduling model by randomly shuffling the values of each feature in turn, thereby severing that feature’s relationship to the outcome while leaving everything else intact, and observing the degradation of the model’s predictive skill (R^2 in signed-log space, on test data). By this measure, loss of information about flight distance matters most ($\Delta R^2 = 0.079$), followed by aircraft type (0.065), day of year (0.061), mean insolation (0.053), origin latitude (0.043), and departure hour (0.040; see Fig. 3c), consistent with exposure time at cruise altitude, engine and airframe characteristics, the seasonal cycle of ice-supersaturation and daylight, the absence of solar radiation to reflect, and the latitudinal climatology of contrail-forming conditions. Aircraft type matters through engine and airframe characteristics rather than age per se: holding route and departure time fixed, the model predicts roughly an order of magnitude less forcing for aircraft with low-soot lean-burn engines (787s and 747-8s; 737 MAX and many A320neos) than for other types, including the new A350 (Supplementary Fig. S3). These differences reflect some combination of soot (nvPM) emissions, typical cruise altitude, and the routes each type flies, so they should be read as type-level associations rather than isolated engine effects. The overland fraction of the great-circle is the least important feature (0.007).

Model application

Figure 4 shows the potential for booking decisions to reduce contrail forcing as a function of the share of flights rescheduled (ranked worst-first by predicted or actual forcing, x axes) and the degree of flyers' scheduling flexibility (the maximum difference in departure time between the avoided flight and its replacement, y axes). The contours in panels a and b show the percentage of global contrail forcing that could have been avoided if a given share of flights had been rebooked onto same-route alternatives not in the predicted worst decile and within the acceptable window (if no such alternative flights existed, no change). Across all 22.3 million flights in 2021, contrail forcing is reduced by 60-70% if the predicted worst 10-25% of flights are rebooked and there is ± 2 hours of flexibility (Fig. 4a), versus $>90\%$ reduction under perfect foresight (Fig. 4b). Indeed, the potential reductions are insensitive to travel time flexibility, reflecting the fact that the busy corridors that produce most contrail forcing are also densely scheduled, so alternative flights almost always depart nearby in time (often a different aircraft or operator at virtually the same hour), meaning most fleet-wide avoidance requires little schedule change at all. Fleet-wide potential is therefore limited by the accuracy of the ranking rather than the availability of alternatives.

However, alternatives may be much more limited for individual flyers, whose itineraries may be served by only a few flights per day. We further demonstrate the potential of the scheduling model to avoid contrail forcing from such specific itineraries by applying it to two anonymized corporate flight logs from 2021. The meteorology-aware model estimates that the 565 flights taken by employees of Company 1 and 1,154 flights taken by employees of Company 2 caused per-passenger contrail forcing equivalent to 76.1 and 309.3 tCO₂-eq, respectively. If we counterfactually replace the worst 10% of flights predicted by our scheduling with the real, same-route alternatives that had the lowest-predicted warming (regardless of timing), the total attributed contrail forcing would decrease by 54% for Company 1 and 42% for Company 2 (versus 82% and 75% reductions, respectively, with perfect-foresight under the same procedure). With a more practical rebooking window of ± 16 -hour different from the original flights' departure time, the total attributed contrail forcing would have been reduced by 29% and 14% for Companies 1 and 2, respectively (white stars in Figs. 4c and 4e). But potential reductions decline sharply as flexibility contracts: if the flexibility is $< \pm 4$ hours, the potential for avoided contrail forcing nearly disappears (Fig. 4c–f; Supplementary Fig. 2). The companies in our case study differ mostly with respect to the availability of acceptable alternatives: within ± 16 hours, 79% of Company 1's worst-flagged flights have a preferable alternative versus only 39% for Company 2, reflecting Company 1's trunk-route-heavy mix against Company 2's longer-haul, lower-frequency network.

Discussion and conclusions

Our results show that contrail forcing is partially predictable from flight features that flyers can know when they book flights, and that this information could substantially reduce per-flight warming (by 14-29% for the two companies) by preferring available same-route alternatives within ± 16 -hours. Such

weeks-ahead booking information could complement operational re-routing on minutes-to-hours horizons.

The predictive skill of our scheduling model is dominated by four features: flight distance, aircraft type, mean insolation (i.e. daylight) along the expected route, and origin latitude. These translate to useful heuristics for climate-conscious flyers: when possible, prefer daytime departures and lower-forcing aircraft types (e.g., 787s rather than 747s or 777-200s on long-haul routes), and recognize that long-haul, high-latitude and wintertime trips carry the largest contrail burden, so flexibility on those trips is most valuable. Yet as we saw with the application to corporate flight logs, acceptable same-route alternatives may be more or less available depending on the locations involved. Realizing potential reductions will thus depend on educating flyers of the heuristics and/or making flight-specific predictions readily available when and where flights are booked. As a start, we have developed a prototype of an online tool that will allow flyers to assess the view our scheduling model predictions for flights they are considering taking and view lower-warming alternatives within a selected flexibility window, as well as a web browser extension “Contrail check for Google Flights” that makes flight-specific predictions directly visible on Google Flights. Google itself offers a metric of anticipated contrail forcing, but it is highly qualitative (low, medium, high), not exposed without clicking, and lacks any methodological details. Published schedules may differ from operated flights through delays and equipment changes, but since our features are coarse (departure hour, aircraft type), such differences perturb inputs only marginally, and the tool's booking-time inputs face the same weather-realization uncertainty already reflected in the reported skill.

Our results are subject to several important limitations and caveats. First, the scheduling model's performance decreases when applied out of distribution. For example, when trained on 2019 only and tested on 2021 flights, model predictions of the worst 10% capture 57% of contrail forcing as compared to 75% when the model trained and tested on 2021 data alone. Although the model thus retains most of its ranking skill across even the exceptional 2019-to-2021 disruption in air traffic and fleet composition, the decrease underscores that there are year-specific weather patterns that a schedule-only model trained in one year to predict warming in another year may miss. Relatedly, because year-to-year changes include evolution of aircraft fleets and fuel composition (the latter of which may be increasingly important as sustainable aviation fuels scale up), predictions might be improved by supplanting the proxy of “aircraft type” with more directly relevant engine characteristics such as soot production, typical cruise altitude, and bypass ratio.

Second, our model is trained on the output of a contrail process model rather than atmospheric observations. For this reason, we further evaluate our results against independent previous studies as well as the results of a different process-model. The distribution of warming represented in our training data reproduces independently published findings: the worst 2% of flights account for 72% of contrail forcing in our data (compared with roughly 2% of flights causing 80% of global energy forcing in 2019¹²); 18% of flights are net-warming (compared to ~14%¹²); flights whose great-circle routes lie mostly in darkness contribute 96% of contrail forcing; and winter flights at northern mid-latitudes produce 3.2x the mean forcing of summer flights (consistent with the diurnal and seasonal asymmetries established by

Stuber et al.¹³; Supplementary Table S1). Re-simulating a stratified sample of 273 flights (four seasonal days in 2021, spanning all deciles of predicted forcing) with an independently-implemented process-modelling approach (the open-source pycontrails implementation of CoCiP²⁷ driven by ERA5 reanalysis and the Poll–Schumann aircraft-performance model; see Methods). Of the 273 flights, 59 (22%) formed persistent contrails. Among flights our model predicts in the top quintile of contrail forcing, 45% formed persistent contrails, 36% produced net warming, and 16% were in the top tercile of simulated warming intensity, and this top quintile alone accounted for 76% of all simulated positive forcing (Supplementary Fig. 3a). In contrast, none of the flights in the bottom quintile of predicted contrail forcing were in the top tercile of simulated warming, and flights in that quintile accounted for only 2% of simulated positive forcing (Supplementary Fig. 3a). The rank correlation between predicted and re-simulated per-km forcing is $\rho = +0.37$, and the comparable correlation between the independent process models (our validation simulations and training data) is $\rho = +0.60$. The +0.37 correlation between our model and the validation simulations suggests that the learned patterns are not specific to the process-model implementation used for training. The gap between +0.37 and +0.60 reflects atmospheric uncertainties that schedule-only predictions cannot overcome, and the gap between +0.60 and 1 reflects differences between the two contrail process models themselves (Supplementary Fig. 3b; cf. ref. ¹).

Finally, the CO₂-equivalence which researchers, companies, and climate accountants commonly use to report contrail impacts invites flyers to treat CO₂ emissions and contrails as fungible when they are not. Radiative forcing from CO₂ emissions accumulates and persists for centuries, but contrail cirrus forcing lasts for hours and disappears when the activity stops. Any reported equivalence between the two thus embeds a time-horizon choice and obscures a difference that matters greatly for policy: avoided contrail forcing cools immediately but only for as long as the behavior persists, whereas avoided CO₂ creates lasting benefits. Conversely, no amount of reduced contrail forcing can offset the cumulative commitment of emitted CO₂. Separately reporting the impacts of CO₂ and contrails, or presenting a given flight's contrail forcing as a percentile ranking, may better serve decisions and tools like our booking support prototype (which currently shows both percentile and kg CO₂e).

But despite these methodological caveats, our results have some important implications for climate accounting and policy. First, static multipliers that estimate contrail-related warming uniformly across all flights drastically understate and overstate the climate impact of specific flights and offer no incentive for preferring less-warming flights. Even imperfect schedule-only predictions provide flight-specific information that could be deployed in booking platforms today. Second, demand-side interventions can be complementary to operational re-routing because the time horizons and decision makers are distinct. Flyers and corporate travel managers choose flights months ahead whereas routing changes depend on coordination among airlines, dispatchers, pilots, and air-traffic controllers based on forecasts generated hours or minutes ahead. Moreover, unlike operational re-routing, booking-time predictions are potentially market-shaping: showing flight-level predictions at the point of sale could raise awareness and allow climate-conscious demand to shift toward lower-forcing departures. In turn, because the underlying levers (daytime departures, lower-forcing aircraft types, less contrail-prone routes) are largely within the

control of airlines, sustained demand for low-contrail itineraries could incentivize the schedule and equipment choices that are needed to reduce warming contrails. Contrail warming is not an unavoidable cost of flying but a property of particular flights, and flyers who can identify those flights can create demand for the ongoing operational efforts to reduce aviation's non-CO₂ climate effects.

Methods

Process-model labels. Per-flight contrail energy forcing was provided by the GAIA global aviation emissions inventory¹¹ using a CoCiP-family contrail process model²⁴. For each flight the dataset provides per-waypoint contrail formation, persistence, and energy forcing along the actual flown trajectory, integrated with ECMWF reanalysis humidity and temperature fields and parameterized radiation-budget shortwave/longwave background²⁹. The output we use is the per-flight scalar total contrail energy forcing in joules, which we convert to CO₂-equivalent mass as

$$\text{kgCO}_2\text{e} = E / (\text{AGWP}_{100} \cdot A_{\text{Earth}} \cdot s_{\text{yr}}),$$

where $\text{AGWP}_{100} = 8.25 \times 10^{-14} \text{ W m}^{-2} \text{ yr kgCO}_2^{-1}$ (ref. ¹), $A_{\text{Earth}} = 5.101 \times 10^{14} \text{ m}^2$, and $s_{\text{yr}} = 3.154 \times 10^7 \text{ s}$. Equivalently, 1 kg CO₂-e per 1.33 GJ of contrail energy forcing. Per-km contrail forcing is then the CO₂-equivalent forcing divided by flight distance in km, and the regression target is the signed log of that per-km contrail forcing.

Flight database and filtering. We use a database of per-flight outputs from 2019 and 2021²⁶. 2020 was excluded as anomalous due to COVID-related traffic collapse. We filtered rows with missing features or flight distances <0.1 km (taxi/data artifacts), and retained 47 weekly parquet shards per year (94 total) covering all months. The resulting dataset includes 30.1M flights for 2019 and 22.3M for 2021, totaling 52.5M flights.

Schedule-only feature constraint. Because we want to predict forcing weeks to months ahead at booking time, we exclude features derived from atmospheric state and include only the kinds of quantities a published flight schedule provides: (i) origin/destination latitude (linear) and longitude (cyclic-encoded as sin/cos of $2\pi \cdot \text{lon}/360$); (ii) day-of-year and UTC departure hour, both cyclic-encoded; (iii) great-circle total flight distance from origin–destination airport coordinates; (iv) aircraft type (categorical; 257 distinct types in the dataset); (v) the share of the great-circle trajectory over land, normalized by total trajectory length (pure geometry; no atmospheric input); and (vi) insolation, evaluated as the sun-altitude-weighted fraction of the great-circle in darkness, computed with the pyephem solar-ephemeris library at each of 30 interpolated waypoints (no atmospheric input). In the training data, departure time and aircraft type are taken from the operated flight — the first waypoint time of the flown ADS-B track and the transponder registry's airframe type — which we treat as proxies for their published-schedule counterparts; because both features are coarse (hour of day; aircraft type), the differences introduced by delays and equipment substitutions are small. Cyclic encoding maps each periodic quantity (longitude, day of year, departure hour) to a sine/cosine pair on the unit circle, so that values

adjacent across the wrap point (the antimeridian, New Year, midnight) remain numerically adjacent to the model. We explicitly exclude features derived from the flown trajectory's specifics (mean aircraft mass, total fuel burn, passengers, load factors, and the number or location of tracked waypoints) or from atmospheric state (temperature and humidity multipliers).

Model and training. We use XGBoost regression³⁰ (version 3.2) tuned using a maximum tree depth of 11, a learning rate of 0.05, 348 boosting rounds, a minimum child weight of 50, and 85% row subsampling per tree, with histogram-based tree construction and native categorical handling for aircraft type (parameters `max_depth`, `learning_rate`, `n_estimators`, `min_child_weight`, `subsample`, `tree_method="hist"`, `enable_categorical`, respectively). Categorical features (aircraft type) are handled natively by XGBoost. We pool 2019 and 2021 flights, split them 2:1 train:test stratified by year, and report year-as-out-of-distribution results as a robustness check.

Hyperparameter selection. We tuned hyperparameters with a structured grid search over the settings that most affect gradient-boosted trees at this scale: maximum tree depth (5, 7, 9, or 11), learning rate (0.05, 0.1, or 0.2), and minimum child weight (1 or 50), thus 24 configurations in all. Row subsampling was fixed at 85% and column subsampling was disabled, the latter because the model has only 14 features. To keep the search tractable, each configuration was trained on a stratified subsample of 1.6 million flights and evaluated on a held-out slice of 400,000, with the number of boosting rounds set by early stopping (training halts once validation loss fails to improve for 30 consecutive rounds, to a maximum of 1,500 trees). We selected the configuration that concentrated the most forcing in its predicted most-warming decile of the testing slice and then retrained the model on the full 35-million-flight training set.

Evaluation metrics. We report (i) R^2 in signed-log space on the held-out test set; (ii) top-k% forcing capture (i.e. the cumulative actual forcing of the model's predicted worst k% of flights as a share of the test set's total contrail forcing, net of cooling flights); and (iii) permutation importance (i.e. the decrease in held-out R^2 when a feature's values are randomly permuted, averaged over five permutations on a 50,000-flight test sample, with cyclically encoded variables permuted jointly²⁷). Capture at every threshold from 0–100% are plotted by the cumulative curves in Figs. 3a and 3b.

Company applications. The two corporate flight logs record per-flight identifiers in the same YYMMDD–NNNN–CCCN format as the contrails-team database. Matching these identifiers to the 2021 process-model output yields 565 analyzable flights for Company 1 (of 576 unique flights in its log) and 1,154 for Company 2 (of 1,246). We attribute per-passenger shares of each flight's contrail forcing by dividing by the relevant aircraft's seat capacity and sum to estimate 76.1 tCO₂e and 309.3 tCO₂e for Company 1 and 2, respectively. We then run the scheduling model on the matched flights, sort them by predicted forcing, and compute each company's remaining forcing under each worst-K% avoidance rule by replacing flagged flights with the best-available same-route alternative within a specified time window (see below). Because avoided and replacement flights are apportioned identically, the percentage reductions reported in Fig. 4 are independent of the apportionment convention.

Independent validation. We re-simulate a stratified sample of 273 flights (eight per decile of predicted forcing on each of four seasonal days: 15 January, April, July, and October 2021; restricted to aircraft types supported by the Poll–Schumann performance model) using the open-source pycontrails (v0.63) implementation of CoCiP²⁷, driven by ERA5 reanalysis²⁵ from the ARCO archive (pressure-level fields at 175–350 hPa, 0.5°, 3-hourly; hourly top-of-atmosphere radiation), with histogram-matching humidity scaling and a 12-hour maximum contrail age. We use pycontrails' default microphysics, with engine-specific soot (nvPM) number emissions from its emissions model and without volatile-particle (vPM) activation or wake-vortex ice-survival corrections. Since these processes mainly increase ice-crystal numbers for low-soot engines at cold temperatures, omitting them would, if anything, understate forcing for the newest engine types and in very cold regions (e.g., over Russia and the Arctic); this affects the magnitude of simulated forcing more than the ranking we evaluate. Trajectories were synthesized as great-circle cruise paths at fixed altitude (10,050 m for flights under 1,500 km; 11,280 m otherwise) with 25-km waypoint spacing, using each flight's observed departure time and duration. We compare independently simulated per-km energy forcing against our schedule-only predictions and against the process-model labels using Spearman rank correlation — over all flights and over the contrail-forming subset, since 78% of flights form no persistent contrail — and by overlap of the top-decile sets.

Realistic alternative-flight replacement. For each of the two corporate flight logs from 2021 we flag the predicted worst $k\%$ of flights and search for replacements among all 2021 flights on the same route (identical origin and destination airports) whose (1) own predicted forcing falls below the 90th percentile of all 22.3 million 2021 flights and (2) scheduled departure time is within $\pm W$ hours of the original flight's scheduled departure time. For the corporate applications, under the scheduling-model rule, each flagged flight is replaced by the acceptable alternative with the lowest predicted forcing in the window, and the forcing simulated by the process model for the replacement is accounted in its place. Under the perfect-foresight rule, the flights flagged are the truly worst $k\%$ (ranked by actual per-km forcing), and the replacement is the alternative with the lowest forcing simulated by the process model (and swaps that would not reduce forcing are declined). Under either rule, flagged flights with no acceptable alternative in their window keep their original forcing. For the all-flights analysis, we instead assign each avoided flight its route's mean forcing among acceptable alternatives within the window, rather than selected the best alternative. Figure 4 shows the results of these replacements when sweeping the flagged share quasi-continuously from 1% to 100% (45 values) and the rebooking window from ± 1 to ± 48 hours (40 log-spaced values), an avoided flight is rebooked at its route's average acceptable alternative rather than the best one, provided at least one acceptable alternative departs within the window. Supplementary Figure S2 shows the same results at representative operating points ($k = 1, 5, 10$, or 20% and $W = 2, 4, 8$, or 16 hours or unlimited).

Declarations

Acknowledgements

S.W. was supported by a gift from Gates Ventures. X.B. was supported by UC MRPI grant M23PL6005. We thank the two anonymous corporations that shared their flight logs from 2021 for use as test cases.

Author contributions

S.J.D. and S.W. designed the study. S.W. and X.B. led the early feature engineering and ML experiments on the Contrails.org/GAIA dataset. M.S. and R.T. provided the per-flight process-model outputs and engaged on physical interpretation. S.J.D. led the writing with input from all authors.

Competing interests

The authors declare no competing interests.

Data availability

The cleaned per-flight feature parquets used for training, excluding the proprietary corporate flight logs (i.e. Company 1 and 2), has been deposited on Zenodo at: <https://doi.org/10.5281/zenodo.23194382>. The underlying process-model outputs were generated by Contrails.org and are described therein. The web-based tool is available at: <https://sustainablesolutions.stanford.edu/tools/contrails> and the Chrome browser extension is at: <https://chromewebstore.google.com/detail/contrail-check-for-google/jgoiipmnojcdecidalejcadljdcjacgl>

Code availability

All code is available at <https://github.com/Sustainable-Solutions-Lab/predicting-contrails>, including the canonical training pipeline, hyperparameter search, figure-generation scripts, and customer-application procedure. Analyses were conducted in Python 3.11 with XGBoost 3.2, scikit-learn 1.3, pandas 2.3, and pyproj/geopandas/shapely for the great-circle geometry.

References

1. Lee, D. S. *et al.* The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018. *Atmospheric Environment* **244**, doi:10.1016/j.atmosenv.2020.117834 (2021).
2. Bergero, C. *et al.* Pathways to net-zero emissions from aviation. *Nature Sustainability* **6**, 404-414 (2022).
3. IPCC. *Guidelines for National Greenhouse Gas Inventories*. Vol. 4 (IPCC WGI Technical Support Unit, 2006).
4. Burkhardt, U. & Kärcher, B. Global radiative forcing from contrail cirrus. *Nature Climate Change* **1**, 54-58 (2011).
5. Kärcher, B. Formation and radiative forcing of contrail cirrus. *Nature Communications* **9**, 1824 (2018).
6. Thomas, T. M., Duan, L., Bala, G. & Caldeira, K. A Stylized Study of the Climate Response to Longwave and Shortwave Forcing at the Altitude of Aviation-Induced Cirrus. *Earth's Future* **13**,

e2025EF006201 (2025).

7. Schumann, U., Mayer, B., Graf, K. & Mannstein, H. A Parametric Radiative Forcing Model for Contrail Cirrus. *Journal of Applied Meteorology and Climatology* **51**, 1391-1406, doi:10.1175/JAMC-D-11-0242.1 (2012).
8. Teoh, R. et al. Aviation contrail climate effects in the North Atlantic from 2016 to 2021. *Atmospheric Chemistry and Physics* **22**, 10919-10935 (2022).
9. DEFRA. Greenhouse gas reporting: conversion factors 2024. (UK Department for Environment, Food & Rural Affairs, 2024).
10. Technical Guidance for Estimating Scope 3 Emissions. (GHG Protocol, 2013).
11. Teoh, R. et al. Global aviation contrail climate effects from 2019 to 2021. *Atmospheric Chemistry and Physics* **24**, 6071-6093, doi:10.5194/acp-24-6071-2024 (2024).
12. in *OJ L*, 2024/2493 (European Commission, 2024).
13. Teoh, R., Schumann, U., Majumdar, A. & Stettler, M. E. J. Mitigating the climate forcing of aircraft contrails by small-scale diversions and technology adoption. *Environmental Science & Technology* **54**, 2941-2950 (2020).
14. Smith, J. R. et al. The climate opportunities and risks of contrail avoidance. *Nature Communications* **17**, 2092, doi:10.1038/s41467-026-68784-8 (2026).
15. Frias, A. M. et al. Feasibility of contrail avoidance in a commercial flight planning system: an operational analysis. *Environmental Research: Infrastructure and Sustainability* **4**, 015013 (2024).
16. Sonabend-W, A. et al. Feasibility test of per-flight contrail avoidance in commercial aviation. *Communications Engineering* **3**, 184 (2024).
17. *Contrail Trials*, <<https://www.blue-lines.org/trials>> (
18. Hodgson, P. & Langermann, C. Vol. 2026 (Google, 2026).
19. Vol. 2026 (American Airlines, 2023).
20. Matthes, S. et al. D-KULT: data and tools for routine eco-efficient flight operations. *Journal of Environmentally Compatible Air Transport System* (2026).
21. Sausen, R. et al. Can we successfully avoid persistent contrails by small altitude adjustments of flights in the real world? *Meteorologische Zeitschrift* **33**, 83-98 (2024).
22. Sankar, T. et al. Efficacy of Scalable Airline-led Contrail Avoidance. *arXiv* (2026).
23. Gierens, K., Matthes, S. & Rohs, S. How well can persistent contrails be predicted? *Aerospace* **7**, 169 (2020).
24. Schumann, U. A contrail cirrus prediction model. *Geoscientific Model Development* **5**, 543-580 (2012).
25. Hersbach, H. et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999-2049, doi:10.1002/qj.3803Digital Object Identifier (DOI) (2020).
26. Teoh, R., Engberg, Z., Shapiro, M., Dray, L. & Stettler, M. E. J. The high-resolution Global Aviation emissions Inventory based on ADS-B (GAIA) for 2019–2021. *Atmospheric Chemistry and Physics* **24**,

725-744 (2024).

27. Shapiro, M., Engberg, Z., Teoh, R., Stettler, M. E. J. & Dean, T. (Zenodo, 2026).
28. Teoh, R. *et al.* Aviation contrail climate effects in the North Atlantic from 2016 to 2021. *Atmospheric Chemistry and Physics* **22**, 10919-10935 (2022).
29. Loeb, N. G. *et al.* Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product. *Journal of Climate* **31**, 895-918, doi:10.1175/JCLI-D-17-0208.1 (2018).
30. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794, doi:10.1145/2939672.2939785 (2016).

Figures

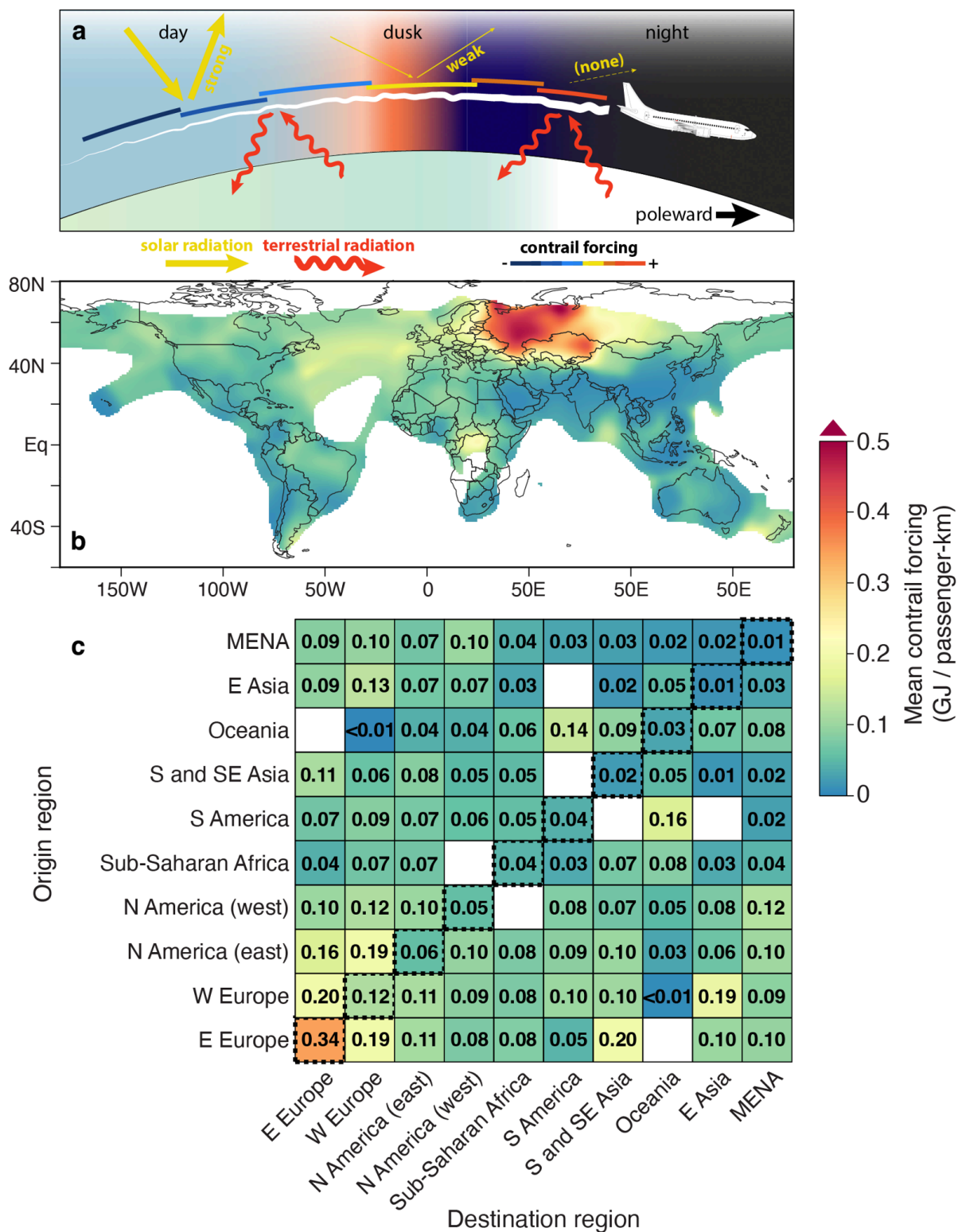


Figure 1

Spatial and temporal heterogeneity of contrail climate forcing. (a) Whether a contrail warms or cools depends on when and where it forms: at night contrails only trap outgoing heat, while by day they also reflect sunlight, which there tends to be less of near the poles (schematic). (b) Mean contrail forcing per passenger-kilometer along flown great-circle routes, 2019 and 2021 (GJ per passenger-km; white = no data). (c) Mean forcing for flights within and between world regions (same color scale).

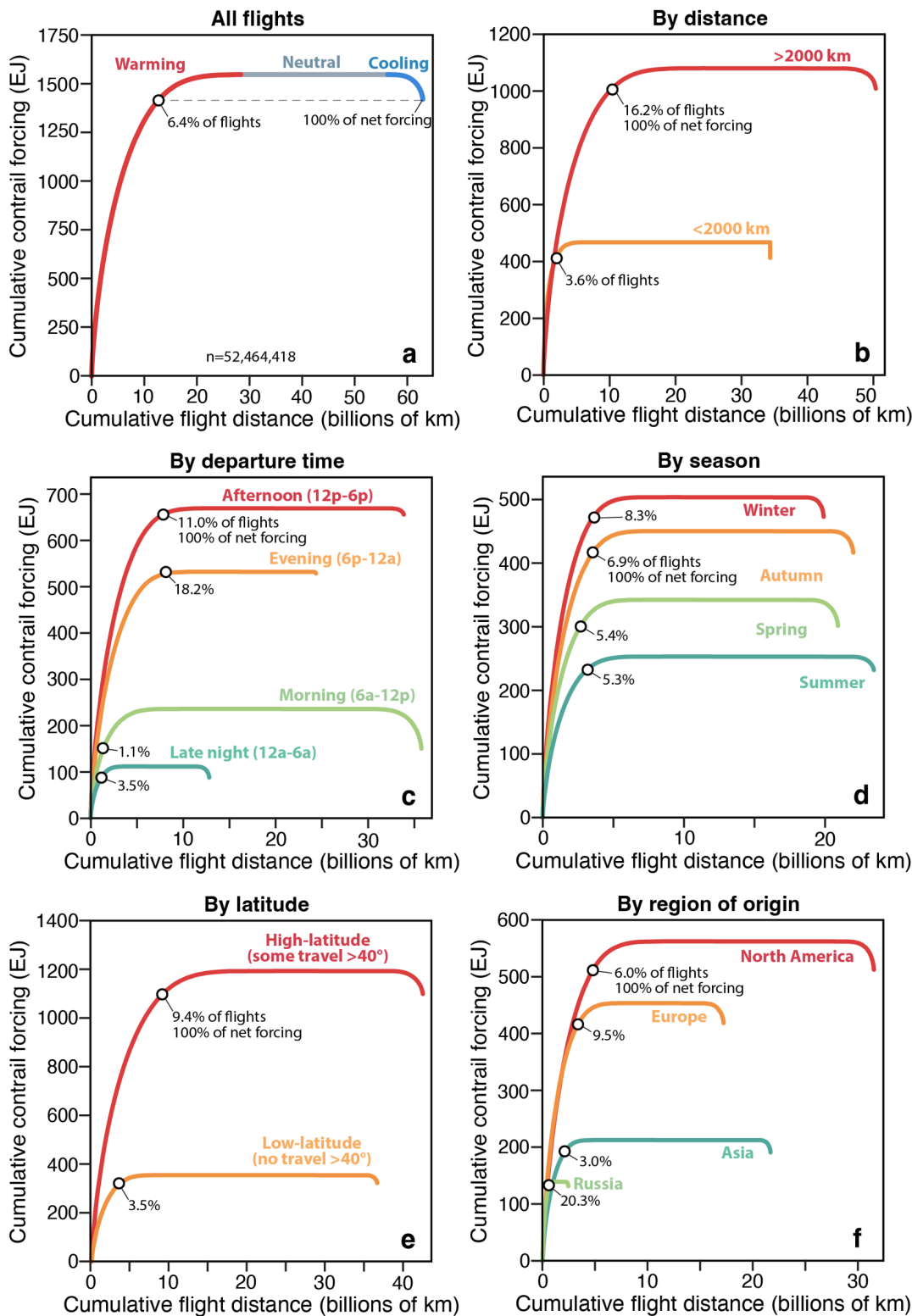


Figure 2

Cumulative contrail forcing versus distance flown. Cumulative contrail forcing (EJ) versus distance flown and ranked most- to least-warming per km, for (a) all 52M flights and disaggregated by (b) flight distance, (c) local solar departure time, (d) season, (e) latitude, and (f) origin region. Curves peak where flights cease to be net-warming; white circles indicate the smallest set of worst flights whose forcing equals the entire grouping's contrail forcing.

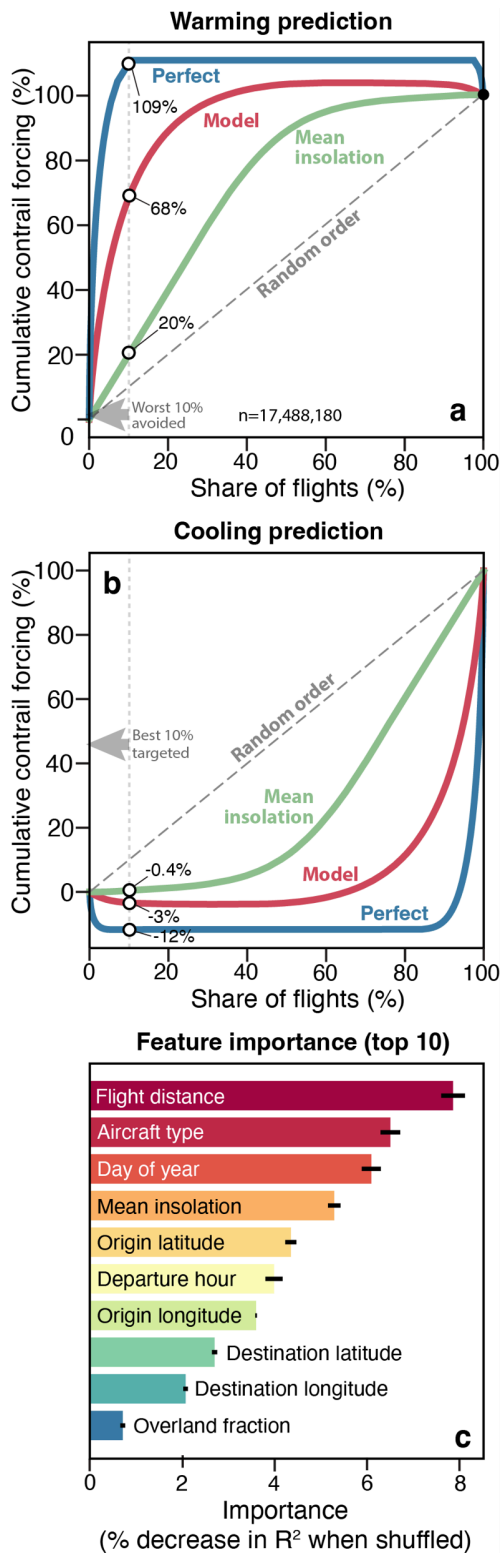


Figure 3

Schedule-only predictions capture most of the warming. (a) Share of contrail forcing as a function of the share of flights as ranked by perfect foresight (blue), our scheduling model (red), or route-mean insolation alone (green). The 10% of flights predicted to be most-warming by our model capture 68% of contrail forcing, in comparison to 109% by perfect foresight (because some flights are net-cooling), and just 20% by mean insolation. (b) In contrast, the model predictions of 10% least-warming (i.e. cooling)

captures only -3% in comparison to -12% with perfect foresight. **(c)** Permutation importance (ΔR^2 when shuffled) reveals the relative importance of features to the model results: distance, aircraft type, route-mean insolation, and geography dominate. Bars show the mean decrease in held-out R^2 over five random permutations of each feature (error bars: ± 1 s.d. across permutations).

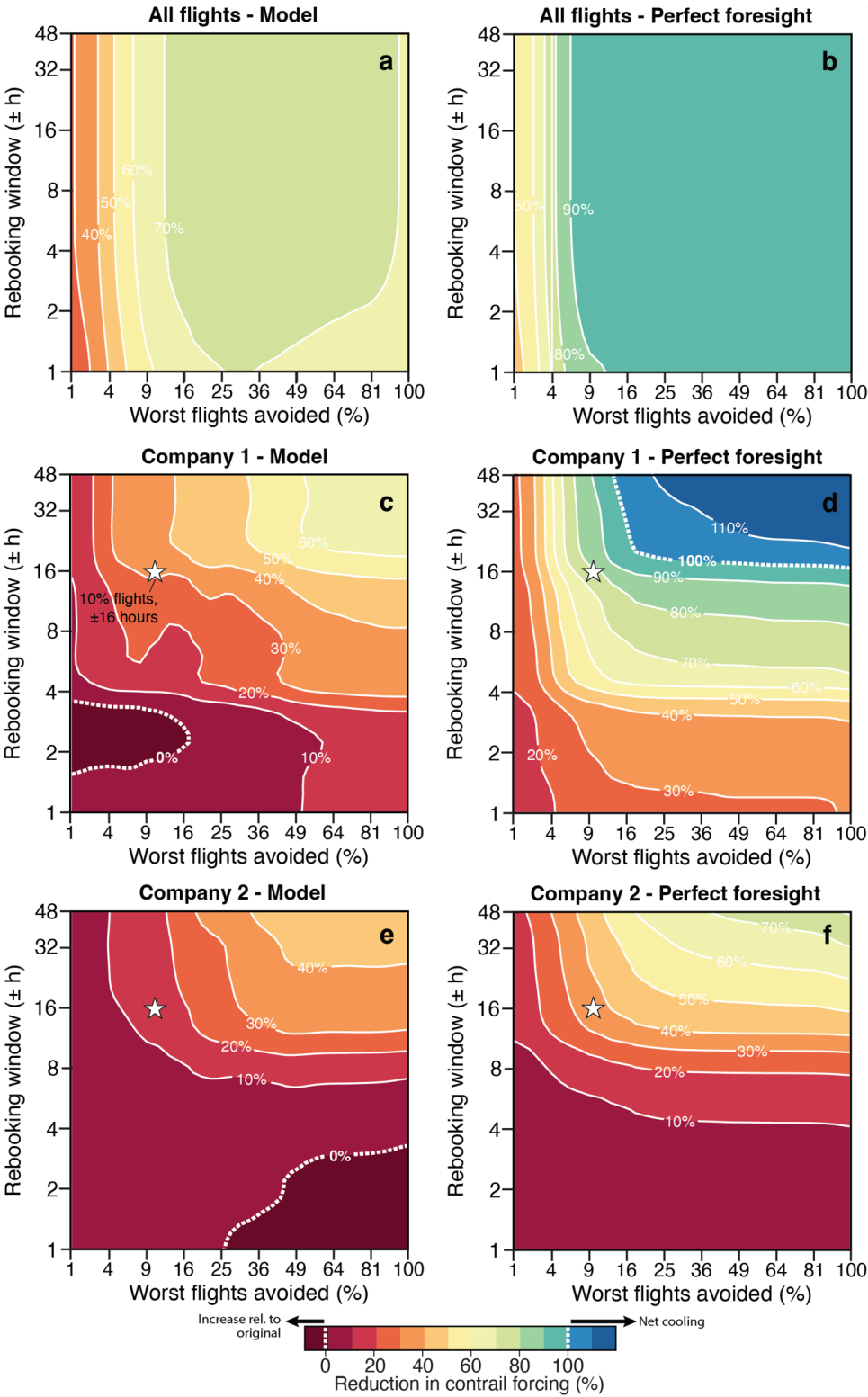


Figure 4

Contrail forcing avoided as a function of ambition and flexibility. Reduction in contrail forcing (% color and contours) from avoiding the worst share of flights (horizontal, square-root scale; per-km ranked) given same-route rebooking within $\pm W$ hours (vertical, log scale), for all 22.3M 2021 flights (**a, b**), flights taken by Company 1 employees (**c, d**), and flights taken by Company 2 employees (**e, f**), under the scheduling model (left) and perfect foresight (right). Company panels rebook each avoided flight to the best acceptable same-route departure in the window; all-flights panels use the route-average acceptable alternative (Methods). Dashed lines mark net-zero (beyond which changes induce net-cooling; blue) and increased forcing (beyond which predictions backfire and warming increases; dark red).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [WorkingSI6Oct26Contrails.docx](#)