MPRA
Munich Personal RePEc Archive

A resolution of emissions-estimate confusion for informing flight choice

Kaivanto, Kim and Zhang, Peng

Lancaster University

22 April 2016

Online at https://mpra.ub.uni-muenchen.de/70923/
MPRA Paper No. 70923, posted 26 Apr 2016 07:57 UTC

A Resolution of Emissions-Estimate Confusion for Informing Flight Choice

Kim Kaivanto* and Peng Zhang

Department of Economics, Lancaster University, Lancaster LA1 4YX, UK

this version: April 22, 2016

Abstract

Air transport Greenhouse Gas (GHG) emissions estimates differ greatly, depending on the calculation method employed. Among the IPCC, ICAO, DEFRA, and BrighterPlanet calculation methods, the largest estimate may be up to 4.5 times larger than the smallest. Such heterogeneity – and ambiguity over the true estimate – confuses the consumer, undermining the credibility of emissions estimates in general. Consequently, GHG emissions estimates do not currently appear on the front page of flight search-engine results. Even where there are differences between alternative flights' emissions, this information is unavailable to consumers at the point of choice. When external considerations rule out alternative travel-modes, the *relative ranking* of flight options' GHG emissions is sufficient to inform consumers' decision making. Whereas widespread agreement on a gold standard remains elusive, the present study shows that the principal GHG emissions calculation methods produce consistent rankings *within specific route-structure classes*. Hence, for many consumers, the question of which calculation method to employ is largely irrelevant. But unless GHG emissions information is displayed at the point of decision, it cannot enter into consumers' decision making. A credible and ambiguity-free alternative would thus be to display GHG *ranking* information on the front page of flight search-engine results.

Keywords: greenhouse gas emissions, carbon footprint computation, scheduled passenger air transport, informed choice, decision making, behavior, policy

JEL classification: Q54, D62, D03

*Corresponding author: tel +44(0)1524594030; fax +44(0)1524594244; e-mail k.kaivanto@lancaster.ac.uk

1 Introduction

At the point of decision, consumers are generally not provided with the GHG emissions information required to make an environmentally responsible choice between different flight options. Flight search engines typically display GHG information, if at all, on CO_{2e}-offset pages which are reached after a specific flight has been selected for purchase. Given differences between emissions calculation methods and differing views on Radiative Forcing (RF), any one CO_{2e} calculator confers advantage to some flights and disadvantage to others in a manner that, to an airline company, may appear arbitrary and unwelcome. For competitive reasons, airlines have been reluctant to elevate GHG emissions to be front-page product-defining attributes alongside price and convenience. Multiple-methods, multiple-estimates ambiguity is a major impediment to the promotion of CO_{2e} to the front page of flight search-engine results, where it must appear in order to have an impact upon consumers' choices.

In this study we show that even though there is considerable heterogeneity in the estimated emissions *levels* calculated with different methods, there is consistency and cordance among the *rankings* generated with different methods. We show this both through the calculation methods' formulae as well as through statistical analysis of flight data. The empirical analysis also reveals that the strong linear relationship between different methods does not generalize well across route-structure classes, where each class collects together flights that have the same number of segments (legs) and share similar relative segment lengths and passenger load factors. Front-page, point-of-choice display of within-class rankings is implementable, valid, and avoids multiple-methods, multiple-estimates ambiguity. For consumers without a viable alternative to air travel, GHG emissions rankings are informationally sufficient to allow identification of the least-harmful flight from among those within the route-structure class. But where the travel mode is not restricted by external considerations, there is no substitute for direct comparison of emissions levels – with all the attendant ambiguity and credibility problems.

2 Previous research

Civil aviation GHG emissions are estimated as global-level inventories (Olivier, 1995; IPCC, 1999; Wilkerson et al., 2010; Simone et al., 2013), national-level inventories (Pejovic et al., 2008; GAO, 2009), airport-level inventories (Hudda et al., 2014; Sherry, 2015), airline-level inventories (van Dorland et al., 2009; Zou et al., 2012), and as performance statistics for individual engines or aircraft (Green, 2002). At each of these levels there is a well-defined clientele for these emissions-estimate data, and hence there is a substantial volume of research being continually conducted and published on all of these levels. For numerous reasons – participation in a regional emissions trading scheme (e.g. the EU Emissions Trading Scheme), meeting carbon reduction targets, and the study of transport-mode choice for informing transport policy – GHG emissions have become the subject of a vast volume of transport research.

In stark contrast, research on flight-level GHG emissions-calculation methods – for informing consumers at the point-of-choice – is exceptionally sparse. Moreover, this research remains within the grey literature (institutional reports), despite point-of-choice GHG-emissions calculation being the necessary input for making environmentally responsible decisions when purchasing flights through flight-search engines, and despite the transport-research field recognizing point-of-choice influence as being an environmentally consequential and legitimate research question in its own right (see e.g. Avineri and Waygood, 2013, in this journal).

These grey-literature studies originate from the Stockholm Environment Institute (Kolmuss and Lane, 2008; Kolmuss and Myers Crimmins, 2009), the Oxford Environmental Change Institute (Jardine, 2009), and the Breda Centre for Sustainable Tourism & Transport (Eijgelaar et al., 2013). Common to these studies is the aim to identify the best calculation method, if such an optimal method exists and is discernible as such. The calculators – and the studies at least in part – are motivated by GHG-emissions off-setting systems' requirements for such estimates, by GHG-emissions accounting and reporting requirements, and by teleological reasoning fixed on the objective of having travel-mode choices respond to flight-specific GHG emissions.

These studies document great differences between the flight-specific estimates provided by the existing GHG-emissions calculation methods. They conclude that, whereas different calculation methods have different advantages and strengths,[1] ultimately all of the calculation methods are imperfect and involve strong compromises.

Whereas we discuss the calculation methods in detail below (Section 3), here we note two sources of uncertainty emphasized in the grey-literature studies. First, aside from any inac-

[1]for instance some utilize extensive detailed information about the flight, whereas at the other extreme some methods employ simple, robust calculations with low informational requirements

curacies in the raw input information regarding plane type and its engines,[2] actual emissions will deviate from calculated emissions because of (i) variation in climatic conditions, such as headwinds or tailwinds, (ii) variation in flight distances and paths, due e.g. to weather-related routing, (iii) variation in time spent in the holding-pattern 'stack', and (iv) variation in the mass of the aircraft from one flight to the next (Jardine, 2009). These sources of irreducible variation entail that there are limits to the precision with which realized GHG emissions may be estimated ex ante. Second, there are numerous metrics with which to adjust airliner CO_2 estimates to account for non-CO_2 effects[3]: Radiative Forcing (RF),[4] Radiative Forcing Index (RFI), Integrated Radiative Forcing (IRF), Global Warming Potential (GWP), Global Temperature Change Potential (GTP), and Integrated Change in Temperature over Time (ICTT) (Kolmuss and Myers Crimmins, 2009). The most commonly used adjustment metric is RFI, which is defined as the ratio of total RF to RF from CO_2 emissions alone. The RFI value used in some calculators may be as high as 4 (Jardine, 2009). However most pre-2005 implementations employed the IPCC (1999) report's central estimate of 2.7, and most post-2005 implementations employ Sausen et al.'s (2005) updating of the original IPCC estimate to 1.9.

Nevertheless in a strict technical sense RFI is a fundamentally flawed metric for gauging the impact of individual flights in the future, as RFI (i) is based on the cumulative effect of past emissions, (ii) is not independent of background atmospheric conditions, and (iii) is not able to account for the different latency periods of different forcing agents (Jardine, 2009; Kolmuss and Myers Crimmins, 2009; Peeters and Williams, 2009; Eijgelaar et al., 2013).

Despite the improvements that GWP, GTP and ICTT bring over RFI, considerable uncertainty and lack of precision remains. In part, this is due to "the relatively low understanding of the impact of contrails and contrail-induced cirrus on radiative forcing and the feedback loops between climate change and the local and global occurrence of these aviation-related impacts" (Peeters et al., 2007). Furthermore, each adjustment-metric implementation involves making an assumption regarding the horizon over which it is to be calculated. Hence each metric can generate a range of different estimates, depending on the horizon assumed in calculation. This

[2]Due to operational exigencies, airlines will occasionally substitute one aircraft with another that is not a precise match down to airframe model and variant, engine type, vintage and efficiency.

[3]emission of water vapour (H_2O), nitrogen oxides (NO_x), particulates (sulfates and soot aerosols), and the formation of contrails and cirrus clouds

[4]The RF of a forcing agent (a gas) is the difference between incoming solar radiation and outgoing infrared radiation, expressed in Watts per square meter (W/m^2).

in turn needs to be determined to match the duration(s) of the forcing agent(s) that the analyst wishes to capture. For instance Kolmuss and Myers Crimmins (2009) advocate a short time horizon of e.g. 20 years to capture the more short-lived effects. These authors acknowledge that "this is a value-based choice... ...in order to best estimate [sic] the footprint of an individual or company due to their current air travel" (Kolmuss and Myers Crimmins, 2009).

While acknowledging the fundamentally flawed nature of RFI for gauging the prospective climate-change impact of individual flights, the grey-literature studies concur that a multiplier greater than 1 should be employed to account for non-CO_2 effects (Kolmuss and Myers Crimmins, 2009; Eijgelaar et al., 2013). The value recommended for this multiplier is in the 1.9–2 range (Kolmuss and Myers Crimmins, 2009). Implicitly, therefore, RFI is accepted as a pragmatic compromise solution. Subsequent studies have adopted RFI as a compromise solution between reliability and usefulness (Eijgelaar et al., 2013, p. 70).

Whereas the existing literature acknowledges the dispersion in estimates calculated with different methods, it has not attempted to address the associated problems of credibility and confusion that impede their adoption for front-page, point-of-choice display. The purpose of the present study is to resolve these problems of credibility and confusion for point-of-choice display.

3 Calculation methods

According to general guidance issued by the Intergovernmental Panel on Climate Change (IPCC), the GHG emissions of scheduled commercial passenger air transport depends on a variety of factors including the type and efficiency of the aircraft and its engines, fuel consumption, distance flown, the composition of the flight in terms of take-off, climb, cruise, decent & landing operating phases, the engine power settings in these operating phases, and flight altitude, among others.

There are several methods for calculating the carbon emissions of a specific flight. In practice the choice of method depends primarily on data availability and the accuracy required. Here we introduce four of the most commonly cited methods. The principal distinction is between methods that employ a fuel-based approach and those that employ a distance-based approach.

A The UK Department for Environment, Food and Rural Affairs (DEFRA) method employs a linear distance-based approach, where the impacts of CO_2, CH_4 and N_2O are included (DEFRA, 2011). DEFRA recommends a 1.9 multiplier if the impact of water vapor, contrails and NO_x are to be included. The distance of the flight is multiplied by an emission factor as follows to obtain an estimate of the CO_2-equivalent (CO_{2e}) emissions:

$$\text{Emissions } (CO_{2e}) = \text{Distance (km)} \times \text{Emission Factor} \tag{1}$$

This emission factor incorporates a 9% uplift to account for delays/circling and route deviations from the Great Circle Distance (GCD) lines between destinations, following IPCC recommendations. In the 2011 DEFRA guidelines the average value of the emission factor is set to be 164.84g/km for UK domestic flights, 96.84g/km for European flights (or distances up to 3700km), and 111.48g/km for all other international flights (or distances above 3700km) (DEFRA, 2011). These values are calculated based on a flight length from the EMEP/EEA Guidebook of 463km, 1108 km and 6482km respectively (EMEP/EEA, 2009).

B The International Civil Aviation Organization (ICAO) adopts a more sophisticated distance-based approach, where each calculation is based on the origin and final destination airports for direct flights, and on the chain of airport pairs for indirect flights (ICAO, 2012). Then published scheduled flights data are used to obtain the aircraft type, which is mapped onto one of fifty 'equivalent type' classes for the purpose of calculating fuel consumption. When a unique aircraft type cannot be identified and scheduled flight data identify a set of possible aircraft types, fuel consumption is estimated as the frequency-weighted average of the fuel consumptions over this set of potential aircraft types. The GCD method is used to calculate the flight distance. ICAO collects passenger load factor data and passenger-to-cargo-ratio data These variables are used to calculate the average fuel consumption per economy class passenger. The fuel consumption is then multiplied by 3.157 (representing the number of tonnes of CO_2 produced by burning one tonne of aviation fuel) to obtain the average CO_2 footprint per economy class passenger.

C The European Environment Agency (EEA) and the IPCC both follow a 3-Tier approach to the calculation of carbon emissions for scheduled commercial air transport (EMEP/EEA, 2009; IPCC, 2006). The approach is conceptual and general. As long as the appropriate emission factor is used, the approach can be applied to direct flights as well as to the individual legs of a multi-segment flight. Equally, it can and be used for calculating CO_2 emissions or CO_{2e} emissions.

Tier 1 is a purely fuel-based approach, where total emissions are obtained by multiplying fuel consumption with an emission factor.

$$\text{Emissions } (CO_2 \text{ or } CO_{2e}) = \text{Fuel Consumption} \times \text{Emission Factor} \tag{2}$$

Tier 2 , also a fuel-based approach, accounts for the fact that jet engines have higher emissions during the Landing/Take-Off (LTO) phase than during the cruise phase of flight. In this approach, total emissions are the sum of LTO emissions and cruise emissions, which are calculated respectively by multiplying LTO fuel consumption with the LTO emission factor and multiplying cruise fuel consumption with the cruise emission factor.

$$\begin{aligned}
\text{Total emissions} &= \text{LTO Emissions} + \text{Cruise Emissions} && (3)\\
\text{LTO Emissions} &= \text{LTO Fuel Consumption} \times \text{Emission Factor LTO} && (4)\\
\text{Cruise Emissions} &= \text{Cruise Fuel Consumption} \times \text{Emission Factor Cruise} && (5)\\
\text{Cruise Fuel Consumption} &= \text{Total Fuel Consumption} - \text{LTO Fuel Consumption} && (6)\\
\text{LTO Fuel Consumption} &= \text{Number of LTOs} \times \text{Fuel Consumption per LTO} && (7)
\end{aligned}$$

Tier 3 follows a distance-based approach where origin- and destination-airport information is utilized. The specific aircraft type is then found from an appropriate database. Again the LTO fuel consumption and cruise fuel consumption are calculated for the specific engine and total emissions are the sum of LTO emissions and cruise emissions.

D BrighterPlanet characterizes flights by origin airport, destination airport, distance, airline, aircraft, seat class, load factor and round-trip versus one-way, and other variables (Kling and Hough, 2010). The BrighterPlanet methodology is also distance-based. The main innovation of this method is that it allows additional parameters (seat class, round-trip versus one-way, etc.) instead of relying on averages over these parameters. For indirect flights the method can be applied to each segment separately.

With the origin-destination parameters for any flight, the passenger-specific emissions are calculated as follows

$$\text{Emissions} = \frac{\text{Total Fuel}}{\text{Passengers}} * \text{Seat Class Multiplier} * (1 - \text{Freight Share}) * \text{EF} * \text{RFI} \quad (8)$$

where the emission factor (EF) is taken from the EIA and a radiative forcing index (RFI) of 2 is used to incorporate the effects of high-altitude emissions and contrails. Seat Class Multiplier is calculated using data from SeatExpert and SeatGuru. Freight Share and Passengers are calculated as passenger-weighted averages of coresponding values of all flights matching this origin-destination pair. Total Fuel is calculated as

$$\text{Total Fuel} = \text{Fuel per Segment} * \text{Segments} * \text{Trips} \quad (9)$$

$$\text{Fuel per Segment} = b + (m_1 * d_s) + (m_2 * d_s^2) + (m_3 * d_s^3) \quad (10)$$

$$d_s = 1.07 * \frac{\text{Distance}}{\text{Segments}} * 1.25^{(\text{Segments}-1)} \quad (11)$$

where coefficients b, m_1, m_2, m_3 are calculated by fitting a third-order polynomial equation to the EEA fuel-consumption data.

4 Analysis

The above-documented diversity in calculation methods leads to large differences in the levels of estimated CO_{2e}. In Case 1 examined below, the largest estimate is 3.5 times the smallest. In Case 3 the largest estimates are 4.5 times the smallest (excluding a small number of outliers, the largest of which is 14.1). This is larger than the factor-of-three maximum deviation previously cited in Kolmuss and Lane (2008).

The diversity in calculation methods nevertheless accommodates high between-method correlations. For instance, the Tier 1 method from the EEA and the IPCC is equivalent to the DEFRA method, assuming that fuel consumption is piece-wise linear in flight distance; the Tier 3 method from EEA and IPCC adopts a similar principle to that of the ICAO method in using more accurate data for the specific jet engine to calculate fuel consumption, although the ICAO method does not distinguish between LTO and cruise phases but rather assumes that fuel consumption is determined by flight distance; the Tier 3 method from the EEA and IPPC also shares some common elements with BrighterPlanet's method.

Merely because the EEA and IPCC methods distinguish between the LTO and cruise phases does not necessarily preclude a high correlation between these emissions figures and – for instance – that calculated with the ICAO method. To see this, note that the IPCC regularly cites Olivier's finding that about 10% of fuel is burnt in the LTO cycles of civil aviation world-wide (Olivier, 1995). The IPCC adopts this fraction as an assumption in several calculations when actual data are not available. Compared to ICAO's fuel consumption calculation, the assumption that 10% of fuel is burnt during LTOs leads to a proportionately higher total fuel-burn estimate for the same flight. However, holding everything else constant, this difference in levels will nevertheless go hand-in-hand with a high correlation between the total emissions figures.

More generally, the fuel-based approach and the distance-based approach share the common principle that fuel consumption should be multiplied and/or divided by a number of factors to generate the emissions estimate. These factors include: the emission factor, the passenger load factor, and the passenger-to-cargo ratio, among others. The main difference lies only in the fact that the distance-based approach requires additional information – e.g. on the origin and destination airports and aircraft type – in order to estimate fuel consumption more precisely. If fuel consumption is indeed approximately a linear function of flight distance, an equivalency can be established between the fuel-based approach and the distance-based approach, with the qualification that the latter requires more information and leads to more accurate estimates.

BrighterPlanet estimates are consistently higher than those generated with the DEFRA and ICAO methods. Whereas the DEFRA and ICAO methods are restricted to direct emissions, the BrighterPlanet method also comprehensively incorporates factors contributing to *indirect* CO_{2e} emissions. These include the effects of methane, nitrous oxide, sulfur hexafluoride, hydrofluorocarbons, perfluorocarbons, land-use change, radiative forcing (a multiplier of 2 rather than 1.9), and *indirect supply-chain emissions from the production of goods and services used as inputs in the provision of air transport services.* These indirect – a.k.a. embodied – emissions often account for a large fraction of total emissions. (Kling and Hough, 2010)

Case 1: MAN-NY Consider the task of choosing among different flight options from Manchester UK to New York. We collect data for 27 different route-time-airline combinations departing between 9am and 12am on May 1st 2013. For present purposes, only flights of no more than 24 hours duration are considered. Flights that involve two or more transfers are excluded. Flights with prices above 1000GBP are excluded.

The price (on the horizontal axis) and carbon emissions (on the vertical axis) pairs for all 27 flights are plotted in Figures (1a) to (1c), where the emissions are calculated by DEFRA, ICAO and BrighterPlanet methods respectively.

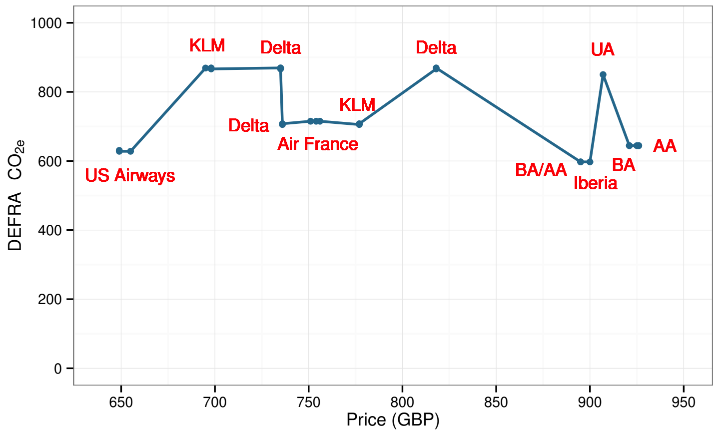
The estimates differ by factors ranging from 1.5 to 3.5. Nevertheless the estimates are nearly perfectly correlated with each other. All pairwise correlation coefficients are above 0.95. Regressing the carbon emissions estimates on each other reveals nearly perfect linear relationships. For instance, regressing the ICAO estimate on the DEFRA estimate and a constant yields an adjusted coefficient of determination of 0.956 (see Table 1). Although statistically the intercept coefficient does not differ from zero, we retain this here and in subsequent regressions for consistency and comparability.

Table 1: MAN-NY estimation results from regressing ICAO emissions on DEFRA emissions.

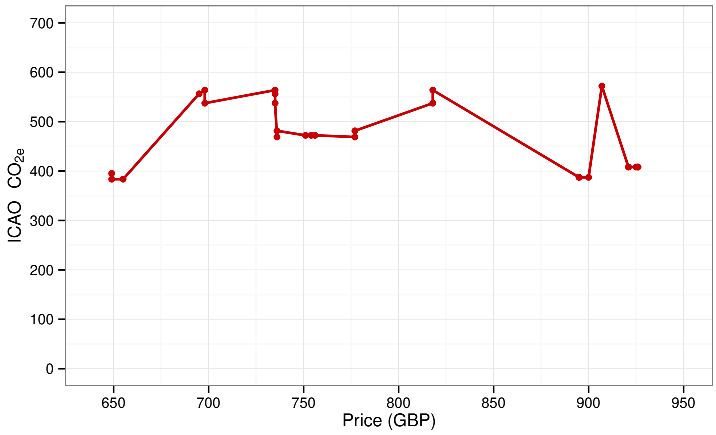
Regressand: $\mathrm{CO}_{2e}^{\mathrm{ICAO}}$					
Regressor	Coefficient	S.E.	t	$P(>\lvert t\rvert)$	(95% C.I.)
Constant	-0.185	19.99	-0.009	0.993	(-41.4, 41.0)
$\mathrm{CO}_{2e}^{\mathrm{DEFRA}}$	0.644	0.0272	23.7	0.000	(0.588, 0.700)
$R^2_{\mathrm{adj}} = 0.956$, $F_{1,25} = 562.6$, $p < 0.001$					

The $\mathrm{CO}_{2e}^{\mathrm{DEFRA}} - \widehat{\mathrm{CO}_{2e}^{\mathrm{DEFRA}}}$ residuals are small, ranging from -21.05 to 25.46. The regression coefficient on $\mathrm{CO}_{2e}^{\mathrm{DEFRA}}$ is statistically significantly different from both zero and one at the $\alpha = 0.05$ level (i.e. the 95% confidence interval excludes both zero and one). The significant F-test value reveals that the mean-only model is rejected in favor of the linear model. Regressions between the remaining method-pairs provide qualitatively comparable conclusions.

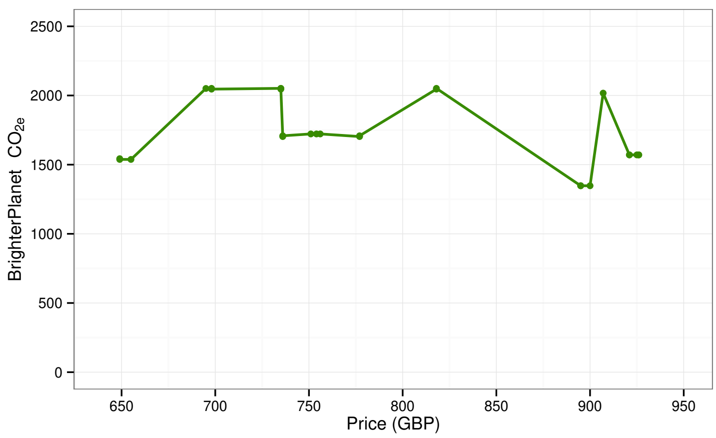
This nearly perfect collinearity implies that emissions calculated with these three methods contain essentially the same information and are indicative of each other. It suggests that whenever the absolute levels of emissions are not of primary interest – for instance when other considerations rule out non-air-travel modes – comparison of different flight options' GHG emissions is likely to lead to be the same ranking with any one of the methods.



(a)



(b)



(c)

Figure 1: Manchester, UK, to New York (any airport); CO_2 on the vertical axis, price (£) on the horizontal axis.

The consistency of the rankings generated by the ICAO, DEFRA, and BrighterPlanet methods may be investigated further using Kendall's coefficient of concordance W, which takes values between 0 and 1 and is related via simple transformation to both (i) the average of all bivariate Spearman rank-correlation coefficients ($\overline{r}$), and (ii) Friedman's non-parametric rank-sum test statistic. Table 2 presents both global and *a posteriori* tests of concordance between the ICAO, DEFRA, and BrighterPlanet methods using the sample of 27 MAN-NY flights.[5] Globally, the three methods are almost-perfectly concordant ($W = 0.9819$). The null hypothesis that the three methods' rankings are independent is rejected by both the F test as well as the permutation-based χ^2 probability $p = 0.001$, which is guaranteed to be of correct size also on small samples (Legendre, 2005; Bonnini et al., 2014). Under the alternative hypothesis of the global test, at least one method is concordant with one or both of the remaining methods in this sample.

Table 2: Tests of concordance on MAN-NY sample.

Global test				
Kendall's W	F	p-value	χ^2	p-value
0.9819	108.3	0.000	76.59	0.001

***A posteriori* tests**					
j	Method	$\overline{r}_j$	W_j	p_j	p_j^H
1.	CO_{2e}^{DEFRA}	0.9796	0.9864	0.0010	0.0030
2.	CO_{2e}^{ICAO}	0.9592	0.9728	0.0010	0.0030
3.	CO_{2e}^{BrPl}	0.9796	0.9864	0.0010	0.0030

The *a posteriori* tests identify the individual contributions of the j methods to the global concordance statistic. Each $\overline{r}_j$ represents the mean of the bivariate Spearman rank-correlation coefficients between method j and the remaining methods. Accordingly, each partial concordance coefficient W_j represents the contribution of method j to the global concordance W. The final column reports p-values adjusted for multiple comparisons using the Holm method (p_j^H). In each *a posteriori* test, the null hypothesis is that method j is independent of the *all the remaining methods*. Hence finding that $p_j^H < \alpha = 0.05$ entails rejection of the null hypothesis, and we say that method j is concordant with all the remaining methods. In Table 2 all three Holm-corrected p-values are smaller than $\alpha = 0.05$, and thus each method is concordant with the other methods.

[5]See the kendall.global and kendall.post functions in the R package 'vegan', available from CRAN.

Case 2: MAN-KUL The results from different emissions calculation methods are not always highly correlated. For instance, the same analysis can be done on different flight options from Manchester, UK (MAN) to Kuala Lumpur, Malaysia (KUL). We restrict attention to economy flights of less than 24 hours duration departing between 7am and 10pm local time on Mayst 2013. Flights that involve more than three segments are excluded. Flights with prices above 1000GBP are excluded as well. For each flight option we record price, DEFRA carbon emissions estimate, BrighterPlanet carbon emissions estimate, ICAO carbon emissions estimate, flight duration, flight distance, fuel consumption estimate, airline company and transfer airport(s).

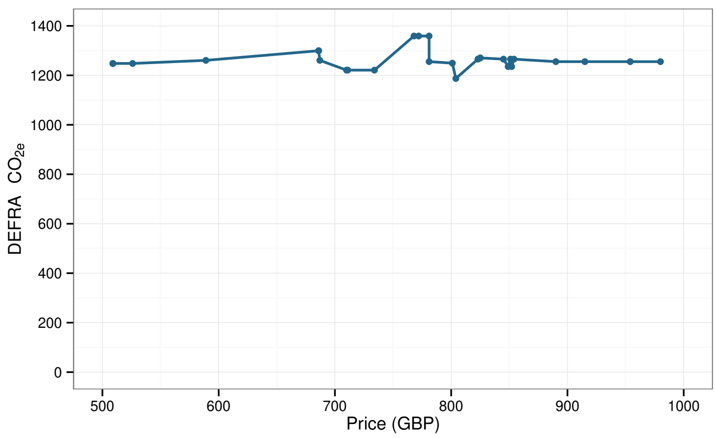
The price (on the horizontal axis) and carbon emissions (on the vertical axis) pairs for all the 29 available flights are plotted in Figures (2a) to (2c), where the emissions are calculated by DEFRA, ICAO and BrighterPlanet methods respectively.

It is clear that the between-method correlations are much weaker here than in the Manchester–NY sample.

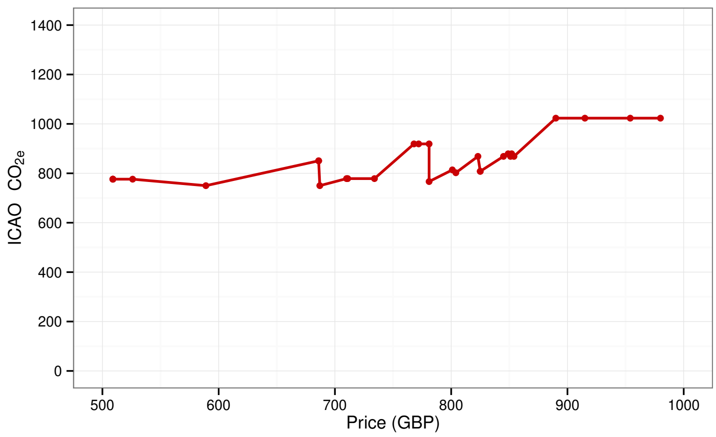
Table 3: MAN-KUL estimation results from regressing ICAO emissions on DEFRA emissions

Regressand: CO_{2e}^{ICAO}					
Regressor	Coefficient	S.E.	t	$P(>\|t\|)$	(95% C.I.)
Constant	-34.66	498.53	-0.07	0.945	(-1058, 988.2)
CO_{2e}^{DEFRA}	0.702	0.394	1.78	0.0863	(-0.1071, 1.511)
$R^2_{\text{adj}} = 0.0719$, $F_{1,27} = 3.17$, $p < 0.0863$					

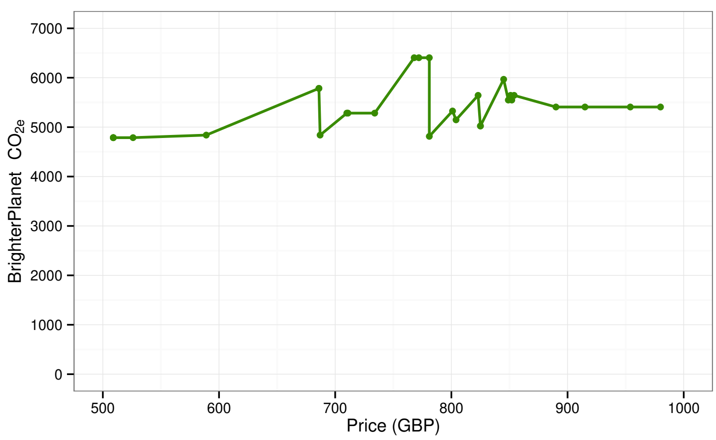
Both the intercept and slope coefficients in Table 3 are statistically insignificant, and the coefficient of variation indicates very low explanatory power ($R^2_{\text{adj}} = 0.0719$). The result is largely determined by the four flights with the highest prices, which show up as outliers in both residuals-vs.-fitted and Normal Q-Q plots. For the four flights in this subset, the ICAO method generates high emissions estimates whilst both the DEFRA and BrighterPlanet methods yield only moderate emissions estimates. The ICAO result is influenced by a particularly low passenger load factor from Manchester to Helsinki and thus particularly high average passenger emissions on that leg. As a result, these four flights stand as outliers in the regression analysis and contaminate the linear regression result. (In what follows, we will accommodate this type of characteristic in the definition of a 'route-structure class'.) Excluding these four observations markedly improves



(a)



(b)



(c)

Figure 2: Manchester, UK, to Kuala Lumpur; CO_2 on the vertical axis, price (£) on the horizontal axis.

both variance explained and statistical significance.

However, in Table 4 the linear relationship between the ICAO estimate and the DEFRA esti-

Table 4: MAN-KUL estimation results from regressing ICAO emissions on DEFRA emissions, excluding flights via Helsinki.

Regressand: CO_{2e}^{ICAO}					
Regressor	Coefficient	S.E.	t	$P(>\|t\|)$	(95% C.I.)
Constant	-252.9	254.7	-0.993	0.331	(-779.8, 274.1)
CO_{2e}^{DEFRA}	0.852	0.201	4.23	0.000	(0.436, 1.27)
$R^2_{\text{adj}} = 0.414,\ F_{1,23} = 17.93,\ p < 0.001$					

mate is still much weaker than that found in the previous section for MAN-NY data. The reason may be that the flights from Manchester to New York fall into the same route-structure class: most share similar route-structure characteristics, consisting of a long-haul segment followed by a short-haul segment. Meanwhile, the flights from Manchester to Kuala Lumpur belong to different route-structure classes – they include a mixture of one-stop and two-stop flight options, and the structures of the flights differ considerably depending on the transfer airports. For instance, Manchester–London–Kuala Lumpur has very different characteristics from Manchester–London–Hong Kong–Kuala Lumpur, as the latter transfers once more than the former and hence consists of one long-haul segment between two short-haul segments. The Manchester–London–Kuala Lumpur route also has different characteristics from Manchester–Dubai–Kuala Lumpur, as the former has two unbalanced segments and the latter has two balanced segments.

Table 5: Tests of concordance on MAN-KUL sample, excluding flights via Helsinki.

Global test				
Kendall's W	F	p-value	χ^2	p-value
0.7657	6.535	0.000	55.13	0.001

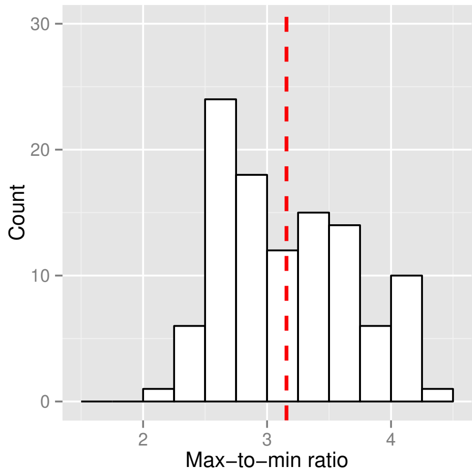
***A posteriori* tests**					
j	Method	$\overline{r}_j$	W_j	p_j	p_j^H
1.	CO_{2e}^{DEFRA}	0.5281	0.6854	0.0030	0.0030
2.	CO_{2e}^{ICAO}	0.6876	0.7917	0.0010	0.0030
3.	CO_{2e}^{BrPl}	0.7297	0.8198	0.0010	0.0030

Although the three methods' global concordance is weaker on the MAN-KUL sample ($W = 0.7657$) than on the MAN-NY sample ($W = 0.9819$), the hypothesis that the three methods' rankings are independent is rejected ($p = 0.001$). And on the *a posteriori* tests, $p_j^H < 0.05 \; \forall \, j = (1, 2, 3)$ so H_0 is rejected, whereby we infer that each method is concordant with the remaining methods.

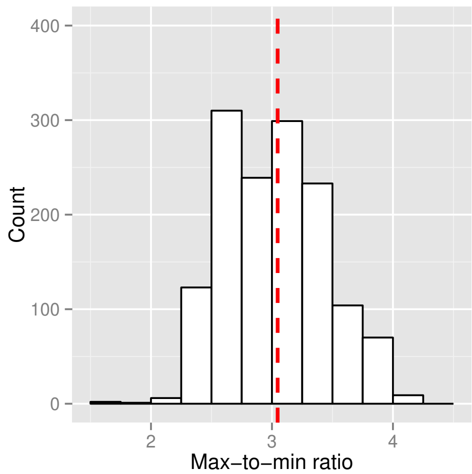
Since the complexity among various flight options will have different impacts on different emissions-calculation methods, in the following section we build two subsamples, each of which pools together flights that have different destinations but share common route-structure characteristics. Specifically, we propose the concept of a 'route-structure class' as a categorization criterion that takes into account the number of segments, the relative length of each segment (balanced/unbalanced) and passenger load factors. We test this concept below.

Case 3: WATS-based sample We seek a broader, representative global sample for testing the rank equivalence of the emissions-calculation methods. Ideally this would entail compiling a frequency distribution of True-Origin–True-Destination (TO-TD) pairs and then sampling from this distribution. As it is based on ticketing data, IATA's PaxIS database allows the construction of such a TO-TD frequency distribution. However direct access to IATA's PaxIS database, even for purely academic research purposes, is prohibitively costly.[6] Hence we proceed to compile a sample that is broadly representative of the highest-volume TO-TD pairs using IATA's World Air Transport Statistics (WATS) digest of the PaxIS Plus data. The properties of the sample data and the procedures used to construct the sample are documented in Appendix A.

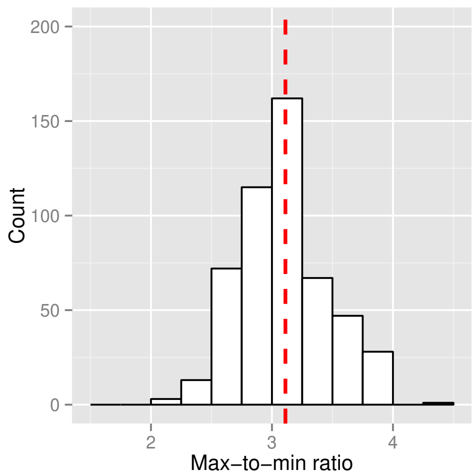
[6] The authors were quoted prices in range USD 50,000–60,000.



(a) One-segment flights ($n = 107$).



(b) Two-segment flights ($n = 1,403$). Seven outliers fall outside plot range.



(c) Three-segment flights ($n = 512$). Four outliers fall outside plot range.

Figure 3: Frequency distributions of max-to-min ratios.

We focus on how subsampling will affect the result of regressing the ICAO emissions estimate on the DEFRA emissions estimate, and examine whether a clearer conclusion emerges. The first subsample contains flights with one segment, the second subsample contains flights with two segments, and the third subsample contains flights with three segments. Within the one-segment subsample ($n = 107$) the max-to-min ratio ranges from 2.23 to 4.40. That is, the largest estimate (invariably BrighterPlanet) ranges between 223% and 440% of the smallest estimate. Within the two-segment subsample ($n = 1,403$) there are 7 outliers with a max-to-min ratio greater than 4.5. The remaining 99.5% of the two-segment subsample's max-to-min ratio falls between 1.55 and 4.5. And in the three-segment subsample the max-to-min ratio ranges from 1.43 to 4.5, with the largest of 4 outliers being 5.80. In all subsamples, both the mean and the median of the max-to-min ratio falls within the interval between 3 and 3.2. The associated frequency distributions are illustrated in Figure 3.

Table 6: Estimation results for the one-segment route-structure class subsample (107 distinct flights).

(a) Regression.

Regressand: CO_{2e}^{ICAO}					
Regressor	Coefficient	S.E.	t	$P(>\|t\|)$	(95% C.I.)
Constant	47.1	3.05	15.4	0.000	(41.07, 53.17)
CO_{2e}^{DEFRA}	0.591	0.00654	90.4	0.000	(0.5781, 0.6041)
$R^2_{adj} = 0.9872$, $F_{1,105} = 8,177$, $p < 0.001$					

(b) Tests of concordance.

Global test				
Kendall's W	F	p-value	χ^2	p-value
0.991	213.4	0.000	315.0	0.001

***A posteriori* tests**				
j Method	$\overline{r}_j$	W_j	p_j	p_j^H
1. CO_{2e}^{DEFRA}	0.9885	0.9924	0.0010	0.0030
2. CO_{2e}^{ICAO}	0.9803	0.9869	0.0010	0.0030
3. CO_{2e}^{BrPl}	0.9893	0.9929	0.0010	0.0030

In the regression of CO_{2e}^{ICAO} on CO_{2e}^{DEFRA} for the one-segment route-structure class, both the intercept and slope coefficients are statistically significant at the 0.1% level (see Table 6a). The 95% confidence interval for the slope coefficient excludes both zero and one; the slope is statistically significantly different from both one and zero. The linear relationship is very strong; the linear model as a whole performs statistically significantly better than the mean-only model ($F_{1,105} = 8177$, $p < 0.001$).

Table 6b presents the results of concordance tests on the two-segment route-structure class. In this subsample the three methods' global concordance is strong ($W = 0.991$), and the hypothesis that the three methods' rankings are independent is rejected ($p = 0.001$). On the *a posteriori* tests, $p_j^H < 0.05 \ \forall j = (1, 2, 3)$ whereby H_0 is rejected. We infer that each method is concordant with the remaining methods.

Table 7: Estimation results for the two-segment route-structure class subsample (1,403 distinct flights).

(a) Regression.

Regressand: CO_{2e}^{ICAO}					
Regressor	Coefficient	S.E.	t	$P(>\|t\|)$	(95% C.I.)
Constant	96.2	1.72	55.9	0.000	(92.86, 99.61)
CO_{2e}^{DEFRA}	0.586	0.00215	273	0.000	(0.5816, 0.5901)
$R^2_{\text{adj}} = 0.9815$, $F_{1,1401} = 74,320$, $p < 0.001$					

(b) Tests of concordance.

Global test				
Kendall's W	F	p-value	χ^2	p-value
0.993	279.1	0.000	4176	0.001

***A posteriori* tests**				
j Method	$\overline{r}_j$	W_j	p_j	p_j^H
1. CO_{2e}^{DEFRA}	0.9915	0.9944	0.0010	0.0030
2. CO_{2e}^{ICAO}	0.9870	0.9912	0.0010	0.0030
3. CO_{2e}^{BrPl}	0.9896	0.9931	0.0010	0.0030

Precisely the same inferences emerge from the analyses of two-segment (Table 7) and three-segment (Table 8) subsample data. Thus, within each route-structure subsample, both linear association and concordance of rankings are strong and statistically significant.

Table 8: Estimation results for the three-segment route-structure class subsample (512 distinct flights).

(a) Regression.

Regressand: CO_{2e}^{ICAO}					
Regressor	Coefficient	S.E.	t	$P(>\|t\|)$	(95% C.I.)
Constant	115.4	4.09	28.2	0.000	(107.4, 123.5)
CO_{2e}^{DEFRA}	0.603	0.00352	171	0.000	(0.5960, 0.6098)
$R^2_{\text{adj}} = 0.9829, \ F_{1,510} = 29,370, \ p < 0.001$					

(b) Tests of concordance.

Global test				
Kendall's W	F	p-value	χ^2	p-value
0.992	257.1	0.000	1521	0.001

***A posteriori* tests**					
j	Method	$\overline{r}_j$	W_j	p_j	p_j^H
1.	CO_{2e}^{DEFRA}	0.9906	0.9937	0.0010	0.0030
2.	CO_{2e}^{ICAO}	0.9875	0.9917	0.0010	0.0030
3.	CO_{2e}^{BrPl}	0.9872	0.9914	0.0010	0.0030

5 Conclusion

To summarize, we have found high correlations between different GHG emission-calculation methods when the flights included in the sample belong to a common route-structure class defined by the number of segments, the relative length of each segment (balanced/unbalanced) and the passenger load factor. Pooling flights from different route-structure classes weakens the correlation structure.

As long as the relevant set of flight options falls within the same route-structure class, analysis based on any one of the emissions-calculation methods will generally lead to the same GHG-footprint ranking – which is precisely the information needed to make environmentally responsible decisions at the point of purchase.

This partial resolution of emissions-estimate confusion suggests relative rankings as an informationally sufficient and pragmatically implementable way forward. However the precise manner in which this relative ranking information should be presented remains an open question. The study of choice architecture is rapidly growing due to ample evidence – amassed within psychology, behavioral economics and interface design – that information *presentation format* impacts upon choice behavior as much, if not more, than the mere presentation of information itself (Johnson et al., 2012). In transport choice, the framing of information affects choice behavior (Avineri and Waygood, 2013), as do social comparisons (Gaker et al., 2010). Accordingly, the present results suggest the incorporation of relative CO_{2e}-ranking information into flight-choice interface design and into the supporting choice-architecture research.

Acknowledgements

This research was supported by Research Councils UK under research grant number EP/I00033X/1. The authors wish to thank Adrian Friday, Paul Peeters, the Editor in Chief, and two anonymous referees, for constructive comments. The usual disclaimer applies.

References

Avineri, E., Waygood, E.O.D., 2013. Applying valence framing to enhance the effect of information on transport-related carbon dioxide emissions. Transp. Res. Part A 48, 31–48.

Bonnini, S., Corain, L., Marozzi, M., Salmaso, L., 2014. Nonparametric Hypothesis Testing: Rank and Permutation Methods with Applications in R. John Wiley & Sons, Chichester, UK.

DEFRA, 2011. Guidelines to Greenhouse Gas (GHG) Conversion Gactors for Company Reporting. Department for Environment, Food and Rural Affairs, UK.

van Dorland, N., van der Zwan, F., Ghijs, S., Santema, S., Curran, R., 2009. Aviation and the environment: Rating airlines on their CO_2 efficiency. Papers Am. Inst. Aer. Astron. 3, 1449–1464.

Eijgelaar, E., Miedema, W., Bongaerts, R., Peeters, P., 2013. Overview of carbon calculators and the calculation of greenhouse gas emissions in tourism. CARMATOP Research Note to Working Package 1, NHTV Breda University of Applied Sciences, Breda, The Netherlands.

EMEP/EEA, 2009. Air Pollutant Emission Inventory Guidebook. European Environment Agency, Luxembourg.

Gaker, D., Zheng, Y., Walker, J., 2010. Experimental economics in transportation: A focus on social influences and the provision of information. Transp. Res. Record 2156, 47–55.

GAO, 2009. Aviation and Climate Change. United States Government Accountability Office, Washington, DC.

Green, J.E., 2002. Greener by design – the technology challenge. Aer. J. 106 (1056), 57–113.

Hudda, N., Gould, T., Hartin, K., Larson, T.V., Fruin, S.A., 2014. Emissions from an international airport increase particle number concentrations 4-fold at 10 km downwind. Environ. Sci. Technol. 48 (12), 6628–6635.

IATA, 2014. World Air Transport Statistics, 58th edition. International Air Transport Association, Montréal, Canada.

ICAO, 2012. Carbon Emissions Calculator Document, Version 5. International Civil Aviation Organization, Montréal, Canada.

IPCC, 1999. Aviation and the Global Atmosphere. Intergovernmental Panel on Climate Change, Geneva, Switzerland.

IPCC, 2006. Guidelines for National Greenhouse Gas Inventories. Intergovernmental Panel on Climate Change, Geneva, Switzerland.

Jardine, C.N., 2009. Calculating the carbon dioxide emissions of flights. Environmental Change Institute, University of Oxford, Oxford.

Johnson, E.J., Shu, S.B., Dellaert, B.G.C., Fox, C., Goldstein, D.G., Haubl, G., Larrick, R.P., Payne, J.W., Peters, E., Schkade, D., Wansink, B., Weber, E.U., 2012. Beyond nudges: Tools of a choice architecture. Marketing Lett. 23, 487–504.

Kling, M., Hough, I., 2010. Brighter Planet Carbon Modelling Methodology. `http://brighterplanet.com/`

Kollmuss, A., Lane, J., 2008. Carbon offsetting & air travel; Part 1: CO_2 emissions calculations. SEI Working Paper, Stockholm Environment Institute, Stockholm.

Kollmuss, A., Myers Crimmins, A., 2009. Carbon offsetting & air travel; Part 2: Non-CO_2 emissions calculations. SEI Discussion Paper, Stockholm Environment Institute, Stockholm.

Legendre, P., 2005 Species associations: The Kendall coefficient of concordance revisited. J. Agr. Biol. Env. Stat. 10 (2), 226–245.

Olivier, J.G.J., 1995. Scenarios for Global Emissions from Air Traffic. The Development of Regional and Gridded ($5° \times 5°$) Emissions Scenarios for Aircraft and for Surface Sources, Based on CPB Scenarios and Existing Emission Inventories for Aircraft and Surface Sources. RIVM-773002003, Rijksinstituut voor Volksgezondheid en Milieu, Bilthoven, The Netherlands.

Peeters, P., Williams, V., Gössling, S., 2007. Air transport greenhouse gas emissions. Tourism and Climate Change Mitigation: Methods, Greenhouse Gas Reductions and Policies. NHTV Academic Studies No. 6, NHTV Breda University of Applied Sciences, Breda, The Netherlands.

Peeters, P., Williams, V., 2009. Calculating emissions and radiative forcing. In Gössling, S., Upham, P. (Eds.), Aviation Climate Change: Issues, Challenges and Solutions. Earthscan, London, pp. 69–87.

Pejovic, T., Noland, R.B., Williams, V., Toumi, R., 2008. Estimates of UK CO_2 emissions from aviation using air traffic data. Climatic Change 88 (3–4), 367–384.

Sausen, R., Isaksen, I., Grewe, V., Hauglustaine, D., Lee, D.S., Myhre, G., Kohler, M.O., Pitari, G., Schumann, U., Stordal, F., Zerefos, C., 2005. Aviation radiative forcing in 2000: An update on IPCC (1999). Meteorol. Zeitschrift 14 (4), 555–561.

Sherry, L., 2015. Improving accuracy of airport emissions inventories using disparate data sets. IIE Transactions 47, forthcoming.

Simone, N.W., Stettler, M.E.J., Barrett, S.R.H., 2013. Rapid estimation of global civil aircraft emissions with uncertainty quantification. Transp. Res. Part D 25, 33–41.

Wilkerson, J.T., Jacobson, M.Z., Malwitz, A., Balasubramanian, S., Wayson, R., Fleming, G., Naiman, A.D., Lele, S.K., 2010. Analysis of emission data from global commercial aviation: 2004 and 2006. Atmos. Chem. Phys. 10 (13), 6391–6408.

Zou, B., Elke, M., Hansen, M., 2012. Evaluating air carrier fuel efficiency and CO_2 emissions in the U.S. airline industry. National Center of Excellence for Aviation Operations Research (NEXTOR), University of California, Berkeley, CA.

Appendices

A Construction of the WATS-based sample

Step 1. In constructing the sample of flights for analysis in Case 3 we make use of IATA's 'World Air Transport Statistics' (WATS) report (IATA, 2014). The WATS report describes its coverage and inclusion as follows.

> The top city-pair rankings presented in this section have been sourced from IATA's PaxIS Plus. Passenger volumes are based on the origin and destination of the passenger's itineraries, which may have included intermediate connections. The coverage is total scheduled traffic, including IATA members as well as non-members. All city-pair rankings in this section consider inbound and outbound passengers, and consolidate commercial airports with a metropolitan area. (IATA, 2014)

Overall, we obtain a sample of 54 top passenger city pairs. For each of the following route areas, we select the three top passenger city pairs: Asia – Southwest Pacific, Europe – Far East, Europe – Middle East, within the Far East, the Mid and South Atlantic, Middle East – Far East, North Atlantic, North America – Latin America/Caribbean, North and Mid Pacific, Europe – Southern Africa, Within Europe, China, Japan and USA. One top passenger city pair is chosen for each of 12 route areas including Australia, Brazil, India, Korea, Russia, Africa – Middle East, Europe – Northern Africa, Europe – Southwest Pacific, within the Middle East, within North America, within South America and within the Southwest Pacific.

Step 2. Based on these 54 city-pairs, we collect all flight route-options utilizing Skyscanner's consolidated search engine.[7] For each of the city-paris, it is assumed that a passenger is to arrange a single flight from the origin city to the destination city on 01/09/2015. To follow the WATS report, we also consolidate commercial airports with a metropolitan area. All flight route options with 3 or fewer legs are recorded for each city-pair, including the information of departing airport, arrival airport and transfer airport(s) for flight routes with 2 or 3 legs.

Step 3. Then all the flight route are pooled together and divided into three subsamples: flight routes with 1 leg, 2 legs and 3 legs respectively.

[7] `www.skyscanner.net`

Step 4. For each flight-route option, we query the ICAO carbon emission calculator[8] for flight distance (GCD) and emissions for each leg.[9] DEFRA emissions for each leg can be calculated by applying a multiplier to the flight distance. BrighterPlanet emissions for each leg can be collected from the BrighterPlanet API.[10].

[8] `http://www.icao.int/environmental-protection/CarbonOffset/Pages/default.aspx`

[9] Flights for which the emissions figure is missing in the ICAO database are subsequently excluded before the analysis is conducted.

[10] `www.brighterplanet.com`

Highlights

- GHG emissions estimates vary greatly, depending on the calculation method employed
- the strictest method's estimate is up to 4.5 times larger than the most conservative
- this heterogeneity prevents industry from displaying CO_{2e} at the point of choice
- we show that different calculation methods yield consistent *rank* orderings
- point-of-choice display of flights' CO_{2e} ranking suffices for emissions-minimizing choice