Air transport CO_2 emissions methodology update

Matthew De Queljoe (Matthew.DEQUELJOE@oecd.org), OECD Statistics and Data Directorate
June 2024

Introduction

The Air transport CO_2 emissions data available at OECD Data Explorer • Air transport CO2 emissions) is compiled from 2019-Q1 onwards using Automatic Dependent Surveillance-Broadcast (ADS-B) flight data provided by the International Civil Aviation Organisation (ICAO). The methodology is described in the OECD Statistics Working Paper CO_2 Emissions from air transport: A near-real-time global database for policy analysis, published in March 2022.

The Air transport CO_2 emissions 2024-Q1 data release includes updates to the methodology used to produce the estimates. The time series from 2019-Q1 onwards has been revised to reflect these updates. This note details the improvements made and identifies further work that is planned for future data releases. It should be noted that because of this ongoing process of improvements, this database is being re-labelled as "experimental".

Methodology improvements

This release contains two improvements to the existing methodology:

1. Adding an additional layer of ADS-B flight data validation
2. Resolving inconsistencies relating to missing residence information

Beyond these specific updates, the database is also subject to minor updates that occur on a continuous basis. These may involve ad-hoc corrections, such as checking that the re-use of airline carrier codes is accounted for in the flight data, or periodic updates to aircraft reference tables that are used as inputs to the emissions calculations.

Flight data validation

The CO_2 emissions results from 2019-Q1 onwards are based on ADS-B flight data from the ICAO. The ICAO processes and cleans the raw flight data producing good overall data quality. However, the OECD's database of Air transport CO_2 emissions contains detailed breakdowns of emissions from aviation, and any remaining issues in the flight data may potentially impact results. Therefore, it is important to add an additional layer of validation to the flight data to improve statistical accuracy. Table 1 summarizes the main issues that may be observed in the ICAO data.

Table 1. Flight data quality issues

Type	Description
Implausible route distances	Recorded distance outliers
	Recorded distance much longer than the great circle distance (GCD) of the route
	Recorded distance less than the GCD of the route
Anomalies in recording of aircraft and airports	Inclusion of helicopters (which are out of scope)
	Large aircraft arriving at/departing from small airports
	Implausible aircraft type - GCD of a route exceeds the aircraft's max range
Unusual new domestic routes	Domestic routes that were not flown in 2019 with mostly unknown records

Implausible route distances can be identified in two ways:

1. Standard outlier detection methods reveal invalid distances when looking at all the flight records for a given route. This works well when most route records are valid. The outlying distances are replaced by the great circle distance (GCD) multiplied by 1.1. In general, flight distances may exceed the GCD by up to 10 percent, so this multiplier is set at the upper bound.
2. The reported distances can be verified by calculating the GCD between the origin and destination for each route. Distances that are either much greater than or strictly less than the GCD are identified. The latter check is included as it is not possible for a flight distance to be less than the GCD. In both cases the invalid distances are replaced by the average valid flight distance for a given route.

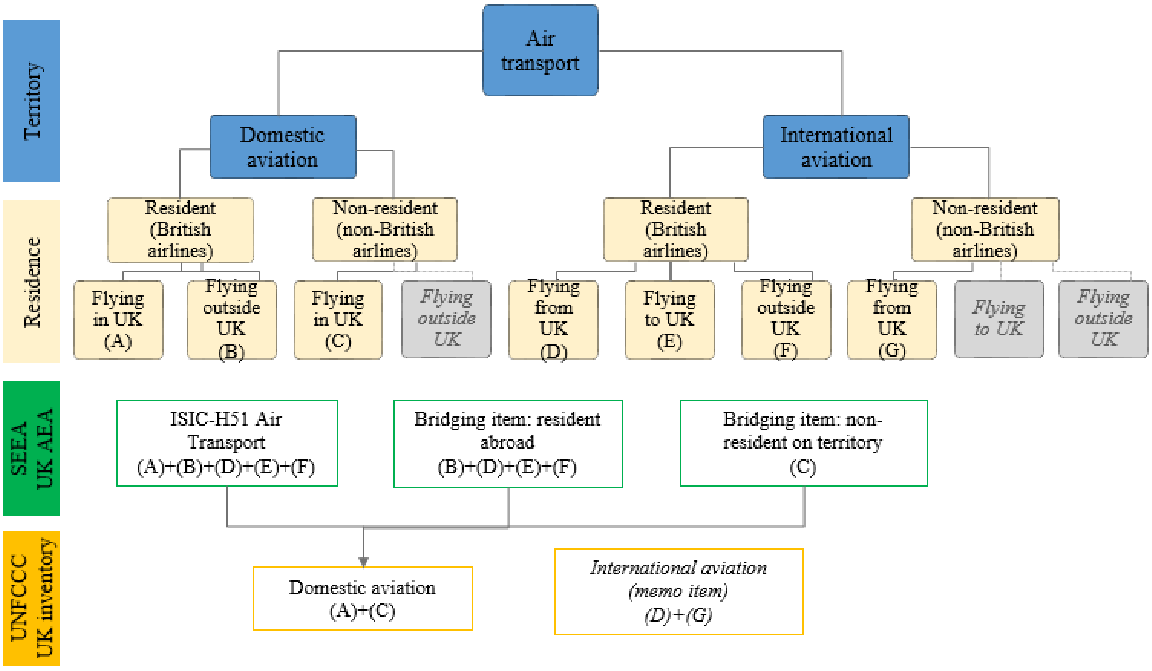
It is also possible to validate records by examining anomalies between the type of aircraft and the origin and/or destination airports. First, flights that depart from or arrive at heliports and hospitals are discarded because helicopters currently fall outside the scope of this database. This also removes cases of helicopter flights that are mis-recorded as another type of aircraft. Furthermore, by combining data on airport runway lengths with the takeoff and landing requirements for aircraft, it is possible to identify cases where the aircraft could not have operated at a given airport. For some small airports, data on runway lengths may be missing. Therefore, as an additional check, keywords in the recorded airport names such as "airfield" and "aerodrome" are checked against the incoming and outgoing aircraft. Large aircraft that operate at these airports are deemed to be implausible. Lastly, it is possible to check whether an aircraft type is plausible by comparing its maximum flying range to the GCD of the route. If the GCD is greater than the aircraft's max range, the aircraft is re-assigned to the most common valid aircraft type. For these checks, the flight record is re-coded using the most common valid aircraft type for the given route if this is available; otherwise it is discarded.

As a last validation, domestic routes from 2020 onwards are checked against the routes flown in 2019. A new route is kept if the airline carriers are known, but the records are discarded if most carriers are coded as "unknown" as the data quality is less certain for these cases. This check is possible since the flight data for 2019 are of high quality. Running this validation at the end catches cases that slip through the previous validations. For instance, the GCD calculation is not possible for some small airports due to unavailable coordinates data and therefore cannot be checked.

Missing residence information

The database contains both residence- and territory-based measures of CO_2 emissions. The March 2022 Working Paper on CO_2 emissions from air transport, which describes the methodology for this work, shows how these measures are compiled using a UK example (Figure 1).

Figure 1: From territory to residence: Allocation of flights from a UK perspective



Source: [CO_2 Emissions from air transport - A near-real-time global database for policy analysis] Clarke, D; Flachenecker, F; Guidetti, E and Pionnier P-A, March 2022.
Note: Bridging items in this Figure link CO_2 emissions related to domestic and international aviation on a residence basis (SEEA air emission accounts) to CO_2 emissions related to domestic aviation on a territory basis (UNFCCC inventories, excluding the international aviation memo item). Any additional flight category in the *Residence* section that is not relevant to the UK air emission accounts or inventories is shown in grey (e.g. a domestic flight outside the United Kingdom by a non-resident airline is neither accounted for in the UK air emission accounts nor in the UK inventories). Categories (B) and (F) correspond to flights operated by airlines that are resident in the United Kingdom but taking place outside of the United Kingdom. For example, a British Airways' flight between Paris and Nice would be recorded under category (B) and a British Airways' flight between Paris and Rome would be recorded under category (F).
Source: OECD own elaboration.

The territory-based measures are calculated using the origin and destination countries which are known for each flight record. The residence-based measures are calculated using the carrier's residence. These are then used to calculate the estimates for the System of Environmental-Economic Accounting (SEEA) Air Emission Accounts (AEAs), shown in green in Figure 1, and the UN Framework Convention on Climate Change (UNFCCC) inventories, shown in orange. If the residence information is missing for a given flight record it cannot be included in the calculation and therefore an inconsistency will arise. This will be the case when the airline operator is coded as "unknown" or if the flight was private.

For domestic flights, the previous methodology assigned all cases where the residence of the carrier was unknown to the reference country (in the Figure 1 example, the UK). This meant that domestic aviation emissions by resident operators were over-reported and emissions by non-resident operators were under-reported. Nevertheless, the accounting identity for domestic aviation measures for UNFCCC inventories (A+C) held since all records were included in both calculations.

However, for international flights no assignment rule was applied and the flight records where the residence of the carrier was unknown were excluded from the calculation. This led to potential inconsistencies in the international aviation (memo item) estimates of the UNFCCC inventories measures since both resident (D) and non-resident (G) carrier flights were under-reported.

The SEEA AEA residence-based CO_2 emissions measure for ISIC H51[1] Air transport (A+B+D+E+F) was also affected by the under-reporting of resident carrier flights.

To fully resolve this issue would require determining the country of residence for a given flight record when it is missing, which for commercial flights would mean first identifying the carrier. As this is a challenging task that requires further study it has been reserved as an improvement for a future methodology update.

In the meantime, even if the country of residence of the carrier is unknown, the problem can be simplified by classifying records into "resident" and "non-resident" using the information available from records where the residency is known. This ensures that the level of resident (D) and non-resident (G) emissions are no longer underreported by the reference country in its UNFCCC inventories international aviation memo item. (A) and (C) have also been adjusted so that they are good estimates in their own right, rather than (A) being overreported and (C) being underreported.

The classification of unknown records is done by first grouping the data by individual flight routes. Each unknown record is compared against the known records for the same route across all time periods. The method employed depends on the number of records for which the residence of the carrier is known, as shown in Table 2.

Table 2. Selection of method for classifying unknown records

Number of known records per flight route	Method
< 10	probability flight is resident
10 <= n < 50	match closest records using Euclidean distances
> 50	k-nearest neighbours classification model (k = 5)

If fewer than 10 known records are available, the classification is done using the probability that a flight is resident given the information available for the route. When 10 to 49 records are available, records can be matched to the closest known record using Euclidean distances; and when 50 or more records are available using a K-Nearest Neighbors classification model. Information is used on the aircraft, number of flights, the distance flown, and the average number of seats per flight to determine whether a flight was resident or non-resident.

However, this only partially resolves the issue. If an unknown record is classified as a non-resident from the perspective of the reference country, then the emissions from that flight would also need to be included in the UNFCCC inventories of the unknown country under the item *domestic flights outside of reference country, operated by resident airlines of reference country* - (B) in Figure 1 - or *international flights outside of reference country, operated by resident airlines of reference country* - (F) in Figure 1 – depending on the flight. As the country of the carrier is not known, this is not possible to do.

Table 3 provides an example for domestic flight emissions in France using the UNFCCC inventories (territory) measure. The accounting identity for domestic aviation in France is resolved by the

[1] International Standard Industrial Classification of All Economic Activities (ISIC) Revision 4, Section H Division 51: Air transport.

corrections described above: the total of A+C is unchanged. However, from the perspective of the unknown countries, (B) remains under-reported.

Table 3. Accounting for domestic flight emissions on UNFCCC inventories basis in France

Code	Emissions measure	France	Unknown countries
A	Domestic flights in reference country, operated by resident airlines	resolved	not applicable
B	Domestic flights outside of reference country, operated by resident airlines	not applicable	unresolved
C	Domestic flights in reference country, operated by airlines that are non-residents	resolved	not applicable
A + C	Domestic aviation	unchanged	unchanged

A key conclusion is that the SEEA AEA residence-based CO_2 emissions estimates for ISIC H51 Air transport (A+B+D+E+F) will remain underreported until we are able to identify the country of residence for all flight records.

Impacts on database

Overall, this update leads to notable downward revisions to CO_2 emissions for ISIC H51 - Air transport, which is the reporting on a SEEA AEA (residence) basis. Figure 1 displays the previous database series alongside the new series for the OECD aggregate. The largest revisions can be observed starting from 2020.

Figure 1. OECD CO_2 emissions for AEAs (residence basis)
Million tons, non-seasonally adjusted

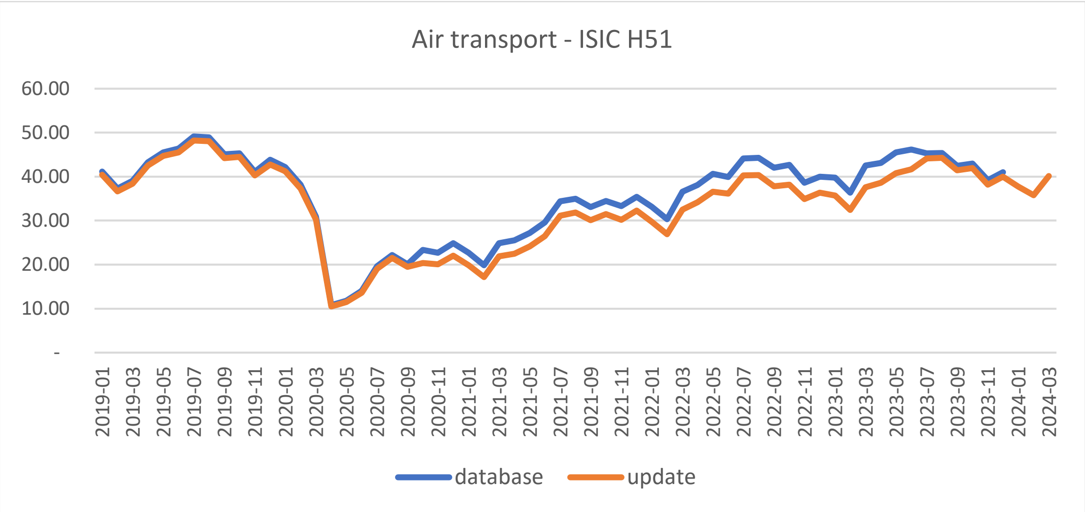
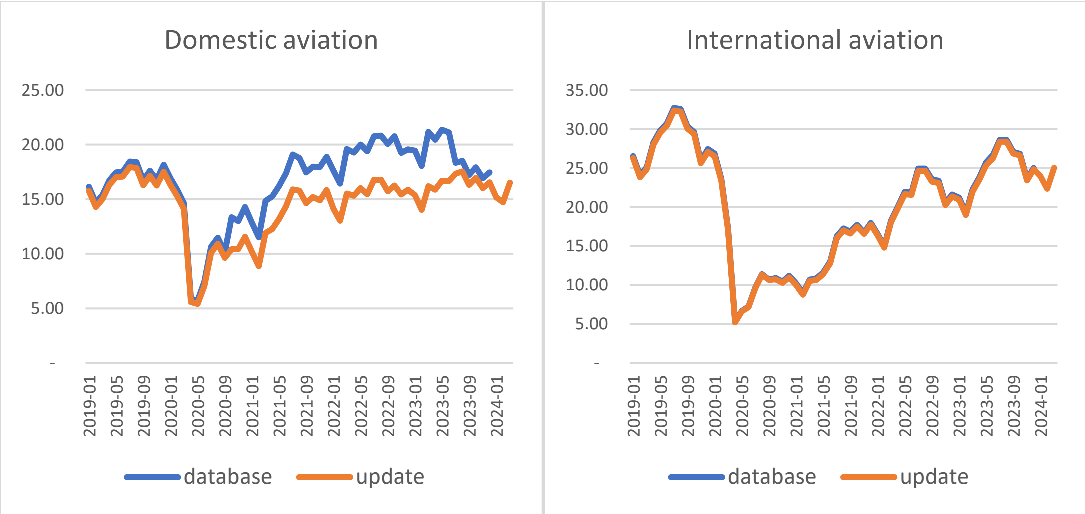


Figure 2 shows the information on a UNFCCC inventories (territory) basis for domestic aviation and international aviation (as a memo item). The main impact of the improvements can be seen in relation to domestic aviation.

Figure 2. OECD CO_2 emissions for inventories (territory basis)
Million tons, non-seasonally adjusted



The downward revisions are due to the improvements described in this paper. Resolving the inconsistencies relating to missing residence information was found to have only a minor impact on AEA ISIC H51 (residence-based) measures of emissions and no impact the two inventories (territory-based) measures. It is the improvements relating to flight data validation (distance checks to correct for implausible route distances and aircraft/airport checks) that have most impact.

Table 4 shows the contribution of each correction to the revision of the OECD total for the AEA measure. For example, the annual estimates for 2021 are revised downward by 10.5%, of which 6% is due to the correction of implausible distances and 4.3% due to the set of aircraft and airport checks. The corrections made so far for missing residence information had little impact, although there may be more impact in future as a result of the work still to be done.

Table 4. Contributions to revision of the OECD total for the AEA measure by correction
% change

Corrections	**2019**	**2020**	**2021**	**2022**	**2023**
Distance checks	0.46	-1.44	-6.07	-6.69	-3.26
Aircraft/airport checks	-2.33	-3.50	-4.33	-3.77	-3.49
Residence information	0.00	-0.07	-0.06	0.34	0.20
Total % change	**-1.87**	**-5.01**	**-10.46**	**-10.12**	**-6.55**

At country level, the extent of the revisions varies. Figure 3 displays the summary statistics for the revisions observed for OECD countries. Of the three main emissions measures, domestic aviation (inventories/territory basis) shows both the largest variation, ranging from -98% to +2.3%, and the largest revisions, with the median revision reaching -33% in 2021 and 2022. For international aviation (inventories/territory basis), the updated series generally remain close to the previous series with median revisions ranging from -1.1% to -0.2% across years. Lastly, for the AEAs (residence basis) the median revisions range from -4.4% to -0.4%. The revisions observed for this series are mainly due to revisions in the resident domestic aviation series.

Figure 3. Summary of annual data revisions for OECD countries
(% difference)

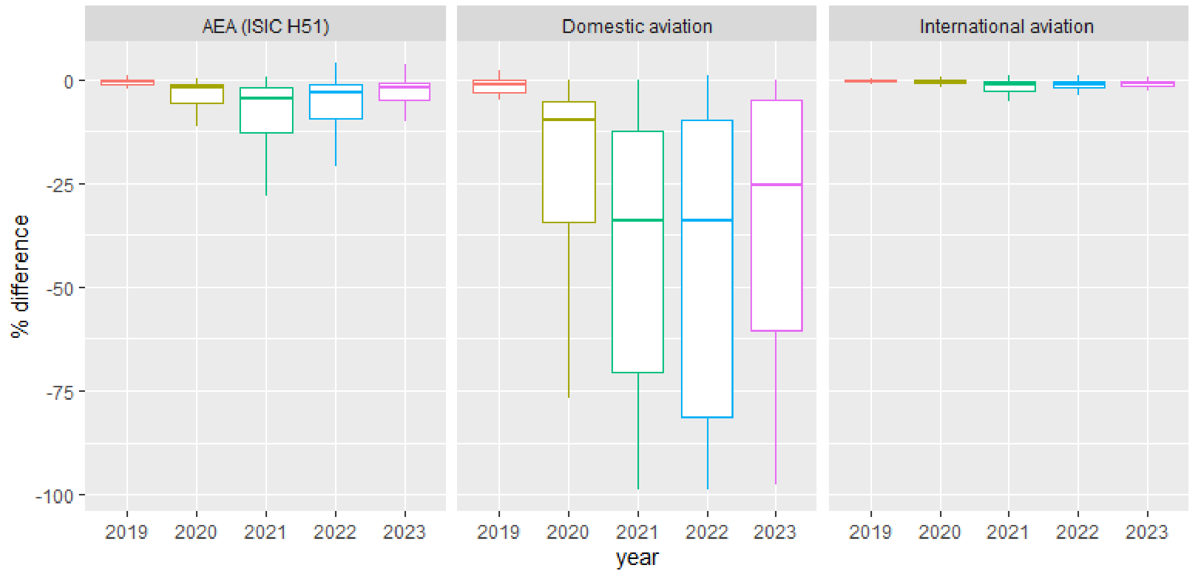


Table 4 and Figure 3 also show that the size of the revisions increased considerably during and after the Covid-19 pandemic (from 2020 onwards). This may be related to changes in the aviation industry as it responded to the challenges of the pandemic, probably affecting the quality of the data.

Planned improvements for future releases

Looking ahead to future releases, two further improvements are planned to be included by the end of 2024:

1. Identifying the country of residence of unknown airline carriers
2. Calculation of emissions from other pollutants (in addition to CO_2 emissions)

As discussed in this article, the first point would further improve the statistical accuracy of the database. The addition of other pollutants that are relevant to air transport emissions would provide further insight into the aviation industry's contributions to climate change. It is envisaged to include nitrogen oxide, water vapor, sulfur oxide, and particulate matter using data available from the European Organisation for the Safety of Air Navigation (EUROCONTROL).