**Takeoff Mass Estimates based on the PS Performance Model and Data-derived Corrections**.

as of 27 Oct 2024

Ulrich Schumann, DLR Oberpfaffenhofen, Germany, email: ulrich.schumann@dlr.de

Ian Poll, Univ Cranfield, UK, email: d.i.a.poll@cranfield.ac.uk

and Roger Teoh, Imperial College, London, UK, email: roger.teoh15@imperial.ac.uk

## 1. Introduction

The take-off weight (TOW) of aircraft determines its fuel consumption, flight profile and flight range, but generally is not well known. The EUROCONTROL Performance Review Committee (PRC) set up a competition call in order to find the "best", or most accurate, practically applicable method which provides this information. For this purpose, the OpenSky Network and its community provided flight data including the observable flight trajectory. We participate in this contest based on our previous work on the PS performance model (Poll, 2018; Poll and Schumann, 2021a; b; Poll and Schumann, 2022; 2024a; b; c) and on practical experiences from investigating the climate impact of aviation, in particular in respect to contrails with traffic data from EUROCONTROL, the CoCiP contrail model, and the PS model (Schumann et al., 2021), and a new global traffic data set GAIA using SPIRE data and BADA3 or BADA4 for performance analyses (Teoh et al., 2024a; Teoh et al., 2024b).

On 2 August 2024, the "EUROCONTROL PRC Data Challenge" was announced by email. The related internet page provided Challenge data with information for **369013** flights (the "C-flights") flown all over Europe in 2022, including long-distance flights to and from outside Europe. As summarized in **Table 1**, the Flight data include the origin and destination airport names, the ICAO aircraft type, the off-block and arrival times, together with estimated or "given" takeoff weights (TOW). In addition, trajectory (FLIGHT-data) were provided. These data come at (max) 1 s time resolution, in the form of around 200 GB parquet-format files (about 1.2 TB in CSV format), including ADSI-B-observed flight positions (time, longitude, latitude, altitude), together with information on groundspeed, flight track angle relative to ground, vertical rate of climb (CLR) and meteorological data (horizontal wind components u and v, temperature T, and humidity q) for each time instance. Some of these data have gaps (e.g., for flights over the North Atlantic), are partially unsorted in time, and come with some jumps and otherwise obviously erroneous data. These data were given together with Submission data (S-data) comprising initially 105959 flights in similar format but without the TOW value. Since 17 October 2024, the S-data were extended to the final submission data including a total of **158149** S-flights. The challenge participants are asked to provide their TOW estimates for these S-flights. The root-mean square (rms) of the errors $\varepsilon$ ($\varepsilon$ = ( TOW_estimate - TOW_given) / TOW_given), RMSe= square root of $\Sigma\ \varepsilon^2$ over all flights divided by the number of flights) were announced to be used to rank the submission results as one input for the selection of the prize-winning team, besides some information on the method used.

**Table 1: Fortran-90 Types describing the FLIGHT, FFLISTC and FLLISTS challenge input data files**

| FLIGHT data (more than 600000), each up to about 30000 elements long | FFLISTC (369013) and FFLISTS (158149) data |
|---|---|
| Type TFLIGHT<br><br>integer No<br>integer flight_id<br>character*10 timestamp<br>character*14 timestamp2<br>real latitude<br>real longitude<br>real altitude<br>real groundspeed<br>real track<br>real vertical_rate<br>integer icao24<br>real u_component_of_wind<br>real v_component_of_wind<br>real temperature<br>real specific_humidity<br><br>end type TFLIGHT | type TFLIGHTLISTC<br><br>integer flight_id<br>character*10 cdatex<br>character*32 callsign<br>character*4 adep<br>character*40 name_adep<br>character*2 country_code_adep<br>character*4 ades<br>character*40 name_ades<br>character*2 country_code_ades<br>character*20 actual_offblock_time<br>character*20 arrival_time<br>character*4 aircraft_type<br>character*1 wtc ! weightclass<br>character*32 Airline<br>Integer Flight_duration<br>integer Taxiout_time<br>integer Flown_distance<br>real TOW<br><br>end type TFLIGHTLISTC<br><br>type TFLIGHTLISTS<br><br>same as above but without TOW<br><br>end type TFLIGHTLISTS |

## 2. Method Overview

First, we had to download the data from the internet (the large files took more than 30 h of time to download to DLR) and then to convert parquet format files to csv format files (taking about 6 h of laptop computing time). Actually, we had to redo this 3 times because of data changes by the PRC team. The last conversion was completed 5 days before the final submission date, by 24 UTC 22 October 2024. We uploaded our final submission values during the last day before the deadline 00 UTC 28 October 2024.

We solve this challenge by executing a sequence of computing tasks as explained in **Table 2**. Each task is implemented in a separate Fortran f90 code. The larger data sets are kept on a memory disk (total 3 TB), smaller ones are first kept in the laptop memory. The latter is saved also on the disk "F:"

**Table 2: Tasks performed, Input and Output data**

| Task name, purpose, computing time | Input | Output |
|---|---|---|
| R1<br>copies FLIGHT data from csv format<br>and combines FFLISTC and FLLISTS data and with airport coordinates.<br><br>Computing time on a laptop about 30 h. | csv format FLIGHT input data which were downloaded from the PRC files in parquet format and converted to csv and kept on a disk in file F:\EU_parquet;<br>also FFLISTC and FFLISTS in csv format.<br>In addition, a file (from Roger Teoh) "global_airports.csv " listing the latitude/ longitude/ altitude of 21208 airports identified by 4-character string ICAO symbols. | The combination of FLIGHT data with the flight lists in FFLISTC (with TOW) or FLISTS (without TOW) data, with times converted from character strings to integer times and with airport information included, are output to F:\FLLISTC\'//cdate, as unformatted and hence quickly readable files.<br>Here cdate is a string like "2022-01-01" referring to the day of departure of the flights. |
| R2<br>for first TOW mass analysis.<br><br>Computing time on a laptop about 1 or 2 min. | FFLISTC (with given TOW) and FLISTS (without given TOW) | sets up a first estimate of load factors for each flight in FLLISTX, and outputs daily mean values for the 365 days of 2022 in 'git/LFYEARs3.txt' |

| | | |
|---|---|---|
| | | and “MassLF3s.txt”based on FFLISTC and in a submission file (early version v1, later version v2) based on FFLISTS |
| R3<br>Performs a detailed trajectory analysis, derives mass estimates with methods 1, 2, 3, and 5 (based on the PS model, see below),<br>and computes the necessary input for estimating the mass from the initial climb rate (method 4, see below). R3 computes performance for each flight segment using the PS model from Poll and Schumann (2024)<br><br>Computing time of order 9 h. Most time is used for input/output | FFLISTC,<br>FLISTS,<br>and<br>F:\FLLISTC\'//cdate containing the FLIGHT trajectory data in updated form from R2, including airport information.<br><br>and<br>'git/LFYEARs3.txt' | Output of one line for each flight into<br>'F:\FLLISTCoutnew\'<br>//'outmass_'//cdate//'.txt',<br>containing:<br>flight_id,ICAO,IC,IPS, Jairline, imassvalid1, imassvalid2, imassvalid3, imassvalid4, imassvalid5, massresult1/kg, massresult2/kg, massresult3/kg, massresult4/kg, massresult5/kg, clr/(m/s), nclrsteps, given_TOW/kg, err1, err2, err3, err4,err5, FLmax/feet, PMALTmax/feet, flownairdist/m, flown_grounddistance/m<br><br>In addition, output of climb rate statistics on to<br>'F:\FLLISTCout2/climbdata2 with mean climb rates per aircraft type and per airline in the format: real clrm, a12, a22, b1, b2, x1, x2 and INTEGER nclr |
| R4<br>runs in 5 “test steps”.<br>Completes the climb data and searches for mean errors and best approximations and computes a final submission data set version v2 for given aircraft type, given airline id, and ground distance value from FFLISTS | Output from R3:<br><br>'F:\FLLISTCoutnew\'<br>//'outmass_'//cdate//'.txt'<br><br>and<br><br>'F:\FLLISTCout2/’<br>//’climbdata2' | 'git/R4_result.txt',<br>'F:\FLLISTCout2/’<br>//’climbdata2',<br>'F:\FLLISTCout2/’<br>//’meanerrorac',<br>'Results2/R6_bestof.txt',<br>and a submission file |

| | | |
|---|---|---|
| and the load factor analysis of Ian Poll with corresponding PS model data input.<br><br>Computing time order 3 min | and from R2 | |
| R5<br>makes the same analysis as R6 (see below) but for the FLLISTC data with known TOW, and computes the mean rms error for this method.<br><br>About 4 h computing time | FFLISTC and FLIGHT<br>and<br>output from R3 and R2:<br>'F:\FLLISTCoutnew\'<br>//'outmass_'//cdate//'.txt'<br><br>and<br>'F:\FLLISTCout2/'<br>//'climbdata2' | mean RMS error estimate for same method as used below in R6 |
| R6<br>Generates the final submission file, version v3. Performs the analysis of the FLLISTC for data with unknown TOW.<br>About 2 h computing time | same as above for R5 but FFLISTS instead of FLLISTC | final submission data file version v3 |
| R7<br>Takes the same basic method as in R6, but restricts the final selection of best fitting mass estimates to those based on load factors computed with air distance.<br>About 2 h computing time | same as above for R6 | final submission data file version v4 |

## 3. Some methodological details

We use 5 methods, They are described here as coded and  in the sequence as developed, not necessarily in the scientifically most obvious sequence.

### 3.1 Mass estimates from load factors (in the codes identified by a string LF)

The take off mass can be estimated (Poll, 2011; Poll, 2018) from

$$TOM \approx \frac{LF\ MZFM\ + (1-LF)OEM}{\left(EXP\left(-\left(0.014+1.015\left(\frac{g(R_t)_{air}}{(\eta_o L/D)_o LCV}\right)\right)\right)-0.05\right)} \qquad (1).$$

Here appears the yet unknow load factor LF, which can in principle vary between 0 and 1. The maximum zero fuel mass *MZFM*, the operation empty mass *OEM*, the maximum landing mass *MLM* and the performance parameter $(\eta_o L/D)_o$ (product of overall propulsion efficiency $\eta_o$ times lift to drag ratio L/D under optimum flight conditions; index o), are characteristics of the aircraft and $(R_t)_{air}$ (the total distance travelled through the air) depends upon the ground distance between destination and departure points and the wind field. The *MZFM* and *MLM* are "certified" mass values which can be found in the aircraft's Type Certificate and aircraft manufacturer documents for the given aircraft type.

For given TOW, this equation can be inverted to determine the load factor LF.

That is what is done first in task R2. It uses the ground distance as first estimate for $R_t$ and the given TOW from the FLLISTC, and uses values for MZFM, OEM, and $(\eta_o L/D)_o$ as given in the PS model tables (and g=9.80665 m/s$^2$; LCV=42 MJ/kg), and computes the load factor LF for each flight.

The results are averaged into daily mean values. The results exhibit a weekly cycle and clear indications of reduced traffic in the first part of 2022 because of the COVID counteractions. The maximum is reached in August to September, but the values are lower than what we would have expected (order 0.6) but show reasonable time trends.. Then we assume that the mean load factor applies also to the flights in the submission table FFLISTS. So, we compute an estimate for the TOW for given daily load factors from this equation in task R2.

In tasks R4 to R7, we compute the air distance from the trajectories, and compute one estimate for the TOW, besides by other methods (see below), from the above equation with air flight distance $(R_t)_{air}$ instead of ground distance $R_t$.

Note, we do not use the MTOW in this computation, except that it is used to limit the results from above. The MTOW is highly variable even for constant ICAO code. However, the same is not true of the maximum landing weight, MLM. This is almost the same for all weight variants of a given aircraft type. It is also a certified weight and so we have high confidence in our values.

### 3.2 TOW estimate from lift for given Mach number immediately after takeoff (CL)

The second method is based on lift at takeoff. The lift coefficient is

$$C_L = \frac{L}{(\gamma/2)p_\infty M_\infty^2 S_{ref}}$$

and depends mainly on the Mach number. The lift must be large enough to carry the aircraft weight. Hence assuming weight = lift, we can estimate the TOW for given lift coefficient, aircraft wing area $S_{ref}$, static pressure $p_\infty$(known for given flight level) and an estimate for the lift coefficient at take-off. This value is estimated starting with the lift coefficients for the "clean" aircraft from the PS model as first estimate and corrected empirically for take-off conditions, possibly still with flaps and undercarriage not yet fully retracted, such that it fits the given flight conditions in the least squares sense.

The derived mean linear errors e are used to correct the first estimate by multiplying them with 1/(1+e). We do this for all 5 methods described here. We considered also to separate for different distance categories, but the project time did not allow to implement this yet.

The linear errors e are computed in task R3, averaged in task R4, and then used later (in task R5 to R7) to correct the computed values.

### 3.3 TOW estimate assuming given Flight Level (FL) at cruise close to "optimum" flight (CLO)

The third method estimates the aircraft weight at cruise assuming the given cruise Mach number and flight level (FL) are close to those for minimum fuel consumption, i.e. maximum value of $\eta_0 L/D$. This is one of the key tasks solved by the PS model, see the references cited above. For details we refer to them. We then add to the resultant mass estimate the mass of consumed fuel since take off. This fuel mass is estimated using the full PS performance model along the given (sorted, completed and partially corrected) trajectory. From the C-data we compute the TOW errors which we obtain when taking nominal PS model parameters. The trajectory uncertainties cause part of the TOW estimate errors

### 3.4 TOW estimate from climb rate during upper part of ascent (CLR)

The idea for this method is quite obvious and can be found with only few details in a thesis paper by Lovegren and Hansman (2011). We assume that an aircraft climbs the quicker the lighter it is. So, we take the flight trajectory data with given TOW values and gather values on the climb-rate CLR in m/s from flight altitudes of 10 000 feet to 21 000 feet (or smaller limits when the upper limit is close to the maximum permitted FL). The CLR is simply the $\Delta z/\Delta t$, the altitude difference dived by the time it needs to overcome this altitude range. We correct the time difference for any delay in intermediate constant flight level trajectory parts. The CLR and TOW values are gathered for each flight in task R2, and mean values over all flights of given aircraft type and airline are passed on to task R3. There we compute mean values and least square fit coefficients of a linear approximation, x1 and x2, such that TOW= x1+ x2*CLR in the least squares sense over all flights.

### 3.5 TOW estimate from thrust required at cruise (CT)

The fifth method estimates the aircraft weight at cruise from the fact that the required thrust at cruise must be deliverable with some maneuver reserve by the given aircraft engines. So, we compute the thrust coefficient, relate it to the lift coefficient using available PS equations and then derive the actual mass from this lift coefficient. The result cannot be exact and needs calibration. This is done by computing the mean approximation errors in tasks R2 and R3 and then correcting the result to a likely zero mean error by multiplying the first estimate with 1/(1-e), as explained before.

## 4. Assessment of the validity of the given TOW values

We have indications that the given TOW values cannot be right in all cases, and we have prepared a few ppt slides with figure (see git file), which illustrate such indications. The arguments given there could possibly be extended:

1. The data show for example, that for some aircraft (e.g. B773) the max (TOW) exceeds at last sometimes given MTOW values. Others stay far below the MTOW value (A310). For further ones (e.g., A310), the minimum of the given TOW values reaches below the OEM.
2. For some aircraft (e.g. A319), the TOW is often close to constant specific discrete values from day to day, constantly all over the year. Others (e.g., the B738) do show in contrast far more reasonable variability and seasonal dependency.
3. For some aircraft (e.g., B772 and B778), the TOW is mostly very low, all over the year. For other aircraft the TOW is far more frequently close to the maximum MTOW value.
4. For most of the aircraft types, our TOW estimates based on distance flow, optimum eta(L/D) and fitted load factors is within a few percent agreeing with the given TOW values. For two aircraft (B772, B778), the TOW derived from the load factor equation shows very larger errors (about 25%). Further, the errors usually change significantly when the used eta(L/D) value is changed. However, for the B772 and B778, the deviations get reduced from about 25 % (a very large rms) to a minim value of 22 % (still very large) when eta(L/D) is increased from our best estimate value (see PS model) by 13 % upwards, which is a large change. So, the estimate errors are not dependent on the selected performance parameter, which in principle should be accurate to about 5 or at worst 10 %. This indicates clearly that the TOW values given for these aircraft do not depend on the flight distance. This cannot be correct.
5. In addition, we note that the passengers are not weighed - their weight is estimated, i.e. the payload is not known accurately. This error is then compounded by the taxi fuel. Provided none of the estimates result in a legally binding maximum mass being exceeded, the aircraft is safe to fly. However, the TOW value given is only an estimate of the true TOW. Also, the taxi fuel is not well known. Therefore, much of the time the recorded TOW may be an underestimate.

There may be methods which give high agreement with the given TOW values, but possibly with large deviations from the true values.

Our approach intends to return best estimates for the true values, but are not so good in fitting the values given by the operators. As a consequence, it is to be expected, that our rms mean values cannot be the lowest.

So, we asked PRC whether their goal is to get the best fit to operator estimates or whether their goal is to get the best fit to the true values? And, how is this dilemma considered in the contest rating?

The answer was: “We have no guarantee on the quality of the data, I.e. QAR/FMD for each flight...we have been given some TOWs but we have no way to assess the quality (for this year's challenge at least...). There are no "true values" as what could come from aircraft direct measurement.”

Finally, we were informed that “The leaderboard ranking is a discriminating factor to get in the list of candidates for our rating together with the description of the rationale/ heuristics/ algorithms used to build the model and its (commented) code.”

So, we hope for some consideration of these aspects.

## References


Lovegren, J. A., and R. J. Hansman (2011). Estimation of potential aircraft fuel burn reduction in cruise via speed and altitude optimization strategies*Rep. Report No. ICAT-2011-03*, MIT.

Poll, D. I. A. (2011). A first order method for the determination of the leading mass charactertics of civil transport aircraft. *Aero. J.*, *115*(May, No. 1167), 257-272.

Poll, D. I. A. (2018). On the relationship between non-optimum operations and fuel requirement for large civil transport aircraft, with reference to environmental impact and contrail avoidance strategy. *Aero. J.*, *122*(1258 ), 1827-1870, doi: 10.1017/aer.2018.121

Poll, D. I. A., and U. Schumann (2021a). An estimation method for the fuel burn and other performance characteristics of civil transport aircraft during cruise: Part 2, determining the aircraft’s characteristic parameters. *Aero. J.*, *125*(1284), 257-295, doi: 10.1017/aer.2020.124.

Poll, D. I. A., and U. Schumann (2021b). An estimation method for the fuel burn and other performance characteristics of civil transport aircraft in the cruise. Part 1: fundamental quantities and governing relations for a general atmosphere. *Aero. J.*, *125*(1284), 296-340, doi: 10.1017/aer.2020.62.

Poll, D. I. A., and U. Schumann (2022). An estimation method for the fuel burn and other performance characteristics of civil transport aircraft. Part 3 Generalisation to cover climb, descent and holding. *Aero. J.*, *submitted*.

Poll, D. I. A., and U. Schumann (2024a). A simple model for the estimation of turbofan engine performance in all airborne phases of flight. *Aero. J.*, 1-29, https://doi.org/10.1017/aer.2024.92.

Poll, D. I. A., and U. Schumann (2024b). On the conditions for absolute minimum fuel burn for turbofan powered, civil transport aircraft and a simple model for wave drag. *Aero. J.*, *online*, https://doi.org/10.1017/aer.2024.10.

Poll, D. I. A., and U. Schumann (2024c). An estimation method for the fuel burn and other performance characteristics of civil transport aircraft. Part 3 Full flight profile when the trajectory is specified. *submitted*.
Schumann, U., I. Poll, R. Teoh, R. Koelle, E. Spinielli, J. Molloy, G. S. Koudis, R. Baumann, L. Bugliaro, M. Stettler, and C. Voigt (2021). Air traffic and contrail changes over Europe during COVID-19: A model study. *Atmos. Chem. Phys.*, *21*(10), 7429–7450, doi: 10.5194/acp-21-7429-2021.
Teoh, R., Z. Engberg, U. Schumann, C. Voigt, M. Shapiro, S. Rohs, and M. E. J. Stettler (2024a). Global aviation contrail climate effects from 2019 to 2021. *Atmos. Chem. Phys.*, *24*, 6071-6093, https://acp.copernicus.org/articles/24/6071/2024/.
Teoh, R., Z. Engberg, M. Shapiro, L. Dray, and M. E. J. Stettler (2024b). The high-resolution Global Aviation emissions Inventory based on ADS-B (GAIA) for 2019–2021. *Atmos. Chem. Phys.*, *24*, 725-744, https://acp.copernicus.org/articles/24/725/2024/.