

# **Fly Quiet & Green Forensics**

## **Introduction**

This paper discusses the results of Heathrow's Fly Quiet & Green programme from 2017 to date, examines their feasibility in the light of the published methodology and compares them to results produced by AvGen (using exclusively Heathrow's published inputs and methodology).

## **Background**

In July 2017, Heathrow published the first of a new series of quarterly Fly Quiet & Green statistics, ranking airlines on several different aspects of their environmental performance.

Shortly after publication, AvGen drew Heathrow's attention to apparent anomalies in the calculation of the results. The published "league table" did not appear to be reproducible, despite using exactly the methodology and inputs published by Heathrow.

AvGen requested some worked examples of the methodology from Heathrow in an attempt to identify where the discrepancies lay. Heathrow did not respond to this request. A subsequent request by AvGen elicited the response from Heathrow that it was satisfied with the accuracy of its results, but still without providing any examples to substantiate them.

AvGen then provided Heathrow with a paper highlighting (with specific examples) areas where it believed Heathrow's analysis was flawed. No response was received to this.

A further request from AvGen to Heathrow centred on the different results from the respective parties' analyses regarding which airline (BA or Aer Lingus) rightfully merited the Number 1 position in Heathrow's "league table" for Q1 2017. Heathrow responded that there were other unspecified and undocumented "inputs" (not mentioned in the published methodology) that would have the effect of altering the results, but no information was supplied about what those inputs could be or how they might work.

AvGen then concluded that further requests for clarification were pointless, given Heathrow's repeated failure to provide any evidence that substantiated the published statistics.

### **Basic Methodology – Selection of Airlines**

The Fly Quiet & Green quarterly statistics aim to compare the environmental performance of the 50 "busiest" airlines, i.e. those with the most flights into Heathrow during the quarter in question. Because of the changes to airline routes, schedules and frequencies, this list of busiest airlines will typically change slightly from quarter to quarter.

In addition, airlines that fly significant numbers of flights with both narrow-body and wide-body aircraft may have each type of operation considered separately. In differentiating between the two subfleets of aircraft, Heathrow uses the terms "Short-haul" and "Long-haul", respectively, even if the routes and distances flown by both subfleets are the same. In order for an airline to be split in this way, at least 20% of its flights must be operated by aircraft in each of the two subfleets and, where an airline is split, only a subfleet that qualifies as one of the 50 busiest airlines in its own right is included in the results.

### **Basic Methodology – Individual Metrics**

Heathrow measures seven aspects of airlines' environmental performance for the Fly Quiet & Green programme. The selected airlines and subfleets (as above) are assessed and their comparative performance relative to the other operators for each metric is translated into a table showing all the airlines surveyed, ranked from the best performer to the worst.

No quantitative information is published to indicate how much better (or worse) any airline's performance is compared to any other airline for individual metrics, though this information is not required for the next stage of the analysis, which (according to the published methodology) is based solely on the ranking positions.

## **Basic Methodology – Aggregation of Metrics**

The position rankings for the seven individual metrics are used to derive an aggregate score for each airline. These aggregate scores, in turn, enable an overall "league table" to be produced, showing the relative position and aggregate points score for every carrier.

**The aggregation process, according to the published methodology description, is based solely on the position rankings for the individual metrics (adjusted by a published weighting schema) and does not use the actual performance figures for each metric (which are neither required nor published).**

## **Airline Rankings for Individual Metrics**

The first part of Heathrow's calculation, deciding in which order the airlines are ranked for each metric, relies on the unpublished results from measurements and data relating to the characteristics of the aircraft and flights involved. While performance against those metrics could be independently assessed from sources such as noise certification data, flight monitoring, etc, a comparison with Heathrow's unpublished performance figures is clearly not possible.

In view of the above, AvGen accepts that Heathrow's position rankings for **individual metrics** probably constitute a true comparison of airlines' relative, albeit unquantified, performance.

## **Aggregate Points Deriving from Metric Rankings**

The published methodology states, unambiguously, that

*"The final score for each airline is calculated by adding up combinations of an airline's ranking position for each individual metric and the weighting set for the given metric"*

It goes on to specify that

*"This means that in order to get a 'perfect' overall score of 1000 an airline would have to be ranked #1 in all metrics across the board; conversely to score a 0 an airline would have to be ranked #50 in all metrics in the programme"*

It follows from the above that if the maximum available points contributions for each metric are added together, the total is 1,000.

Heathrow does not explicitly state what the maximum achievable points values are for each individual metric's contribution to the 1,000 points, but this can readily be determined from the published weighting schema, where the relative weighting of each metric's contribution to the aggregate score is specified:

<i>Noise quota/seat</i>	50%
<i>Chapter number (noise certification)</i>	50%
<i>NOx emissions/seat</i>	50%
<i>CAEP standard (engine emissions certification)</i>	50%
<i>Continuous Descent Approach (CDA) violations</i>	150%
<i>Track keeping (TK) violations</i>	150%
<i>Early or late movements between 23:30 and 04:30</i>	60%

Clearly, if 1,000 points is the total of the maximum contributions available from all the metrics, the individual maximum contribution for each metric can be calculated by dividing 1,000 points in the ratio 50:50:50:50:150:150:60. This gives:

<i>Noise quota/seat</i>	89.3 points
<i>Chapter number (noise certification)</i>	89.3 points
<i>NOx emissions/seat</i>	89.3 points
<i>CAEP standard (engine emissions certification)</i>	89.3 points
<i>Continuous Descent Approach (CDA) violations</i>	267.9 points
<i>Track keeping (TK) violations</i>	267.9 points
<i>Early or late movements between 23:30-04:30</i>	107.1 points
<i>(totals rounded to the nearest 0.1 points)</i>	

Having calculated the maximum points available for each metric (as awarded to the airline ranked in 1st position) and given that the airline in last (50th) position is awarded zero points for that metric, the number of incremental points gained or lost for each place up or down in the ranking for each metric can easily be derived:

<i>Noise quota/seat</i>	1.82 points
<i>Chapter number (noise certification)</i>	1.82 points
<i>NOx emissions/seat</i>	1.82 points
<i>CAEP standard (engine emissions certification)</i>	1.82 points
<i>Continuous Descent Approach (CDA) violations</i>	5.47 points
<i>Track keeping (TK) violations</i>	5.47 points
<i>Early or late movements between 23:30-04:30</i>	2.19 points
<i>(increments rounded to the nearest 0.01 points)</i>	

It follows from the published methodology that an individual airline's score for any given metric can be readily calculated by

multiplying the number of places up from the bottom of the ranking by the incremental points per place.

So an airline in 46th position (i.e. four places from bottom) for the CAEP metric (the position held by BA Short-haul in the Q4 2018 results) would score 7.3 points ( $4 \times 1.82$ ) for that metric, out of a possible 89.3 points. Adding the results of the corresponding calculation for all seven metrics would give the aggregate points that determine the airline's position in the overall league table.

**N.B. In the above instance from Q4 2018, BA loses 82 points (89.3 maximum minus 7.3 achieved) from the CAEP metric's contribution to the airline's overall aggregate score. It follows that even if BA had achieved maximum points for the other six metrics (it didn't) then it would have scored, at best, 918 (1,000 minus 82) aggregate points. Instead, it was awarded 946 points, which is impossible.**

### **A Simple Mathematical Sanity Check of Heathrow's Results**

If all 50 ranking positions were occupied for a given metric (i.e. there were no tied positions) then the points scores for the 50 airlines would clearly be evenly distributed from 0 to the maximum points value available for that metric, since the increment between each successive rung on the ladder is a constant for any given metric.

It follows that the average points score (both the arithmetic mean and the median) that any metric contributes towards the sum of the aggregate scores would equal half the maximum available score for that metric.

Equally, it follows that the average aggregate score in the final league table (again, assuming no tied places for any of the metrics) would be half of the maximum (1,000) available points for all metrics, i.e. an arithmetic mean and median of exactly 500 points.

If tied places exist (as can reasonably be expected) for any individual metrics then the average score for that metric will increase and therefore so will its contribution to the overall average aggregate score (explained in more detail below).

Heathrow's average aggregate score for Q4 2018 is **813** points per airline, some **313** points (63%) higher than the "no-tie" average of 500 points, a difference far too great to be accounted for by the relatively small number of tied places for some of the metrics.

It was a similarly obvious discrepancy in the Q1 2017 results that first alerted AvGen to potential errors in Heathrow's calculation of the results and league table. When asked about the discrepancy, Heathrow declined to comment on it.

### **Effect of Tied Places on the Average Scores**

For several of the metrics in each quarter's results, there are typically some airlines shown as tied, i.e. two or more airlines jointly occupy the same position in a metric ranking. It can therefore be assumed that they have achieved the same (unpublished) performance rating for that metric.

If, for example, six airlines tied for a single position (say 1st place) for the Track-keeping metric (as occurred in Q4), then instead of there being airlines in 2nd, 3rd, 4th, 5th and 6th positions, those positions would be vacant and the six airlines would all jointly qualify for the (higher) points score awarded to 1st place. This would add an amount equal to 5.47 (the increment per place for the Track-keeping metric) multiplied by 15 (the total number of places gained, i.e.  $1+2+3+4+5$ ), resulting in an extra 82 points being added to the sum of the aggregate scores, i.e. an increase of 1.64 points in the average airline score.

The most marked example of this in the Q4 results is for the "Early or late movements" metric, where 30 airlines (presumably all with zero infringements of the Night Quota) share the Number 1 slot. However even this adds only 1017 points to the total awarded across **all the airlines** for that metric, i.e. the average airline aggregate score increases by 20.3 points, far short of the increase of approximately **313** points **per airline** on average (15,643 points in total) that Heathrow has awarded for Q4 2018, compared to the "no-tie" average.

It is not possible to say whether this large discrepancy is a result of simple arithmetical errors, or whether a different methodology from that published has been used to calculate the anomalous results.

As discussed above, requests to Heathrow to clarify the processes that they have followed, and for comment on the discrepancies compared to results obtained by AvGen (using Heathrow's own published methodology and data) have remained unanswered.

## March 2019 Update

Following two years of unsuccessful attempts to persuade Heathrow to explain how and why its results differed from those obtained independently using the airport's published rankings and stated methodology, AvGen resorted to writing direct to Heathrow's CEO, repeating its concerns that the statistics did not appear to be realistic representations of an airline's environmental performance and asking for a breakdown of the most recent statistics that would substantiate (or otherwise) the published results.

Heathrow responded, describing AvGen's concerns as "bizarre", but at the same time declining to provide any substantiating data to refute those concerns. The airport disingenuously argued that other, undocumented factors had been applied to the results to produce the published scores, effectively confirming AvGen's contention that the stated methodology had not been used and that the entire process was therefore far from transparent.

## Limitations of the Fly Quiet & Green Programme

Leaving aside issues around whether the results have been correctly calculated, a fundamental limitation of Heathrow's approach is that it does not allow any meaningful comparisons to be made between an individual airline's performance over successive quarters.

For example, Airline A may be ranked higher for, say, NO<sub>x</sub> emissions in Q2 than in Q1 (and therefore gain more points to contribute to its aggregate Fly Quiet & Green score) even if its NO<sub>x</sub> performance figure has actually worsened in absolute terms.

It follows, therefore, that neither the rankings for individual metrics, nor the aggregate points awarded to an airline which determine its position in the published "league table" will necessarily be evidence of whether the airline in question has improved, or worsened, over time. An airline can improve its performance and still score fewer aggregate points than it did in the previous quarter; conversely an airline can climb up the league table even if has performed more poorly on some, or even all, of the metrics.

Furthermore, there have been instances in the results where Airline A has been ranked equal to or higher than Airline B for **every** metric and yet has been awarded a lower Fly Quiet & Green score overall (for example Air France vs Alitalia in Q1/Q2 2018).

## **Recommendations**

AvGen strongly urges Heathrow to:

- a) clarify how and why the methodology it has used differs from that published
- b) provide worked examples to demonstrate clearly to stakeholders exactly how the results are derived
- c) identify and document what other "inputs", over and above the metric rankings, are affecting the results
- d) render the Fly Quiet & Green programme transparent by publishing figures that allow stakeholders to see if a given airline's performance on a particular metric has improved or deteriorated over time