A better way to rank pilots

by Douglas Lovell

The problem

There are two methods in use for ranking pilots at aerobatics contests. CIVA uses a system of number crunching manipulations called "Fair Play." IAC uses straight averaging.

CIVA finds straight averaging unsatisfactory because one judge can completely ruin the chances of a competitor by giving them very bad scores.

IAC finds "Fair Play" unsatisfactory because regional contest flights generally have a small number of pilots, and because the number crunching presents a black box that has to be taken pretty much on faith. Further, "Fair Play" has a property that some find highly unsatisfactory-- the flight of each pilot affects the rankings of every other pilot. If pilot A is ranked before pilot B, the flight of pilot C can alter the ranking to place pilot B before pilot A.

Examination of the merits of straight averaging and "Fair Play" is a topic for a different presentation. This article presents a new alternative that combines the best qualities of both:

1. It is readily verified by examination.

2. It is simple to understand how it really works.

3. It resists sabotage from a judge.

4. It makes sabotage so risky that there is very strong disincentive to try it.

Description. How it works.

We now present the method without rationale or justification. The idea is first to understand it, then to evaluate it.

We start with the scores from each judge for each pilot.

Pilot	J1	J2	J3	J4	J5
P1	2803.0	2701.5	2888.0	2864.5	2732.5
P2	2744.0	2604.0	2781.0	2686.5	2808.5
P3	2735.5	2578.0	2722.5	2634.5	2649.5
P4	2612.5	2585.0	2676.0	2563.5	2745.5
P5	2442.5	2632.5	2651.5	2601.5	2548.5
P6	2657.0	2429.0	2636.0	2527.0	2710.5
P7	2714.5	2517.5	2566.0	2488.0	2599.0
P8	2554.5	2446.0	2506.5	2504.0	2706.0
P9	2674.5	2405.5	2629.5	2488.0	2583.0
P10	2595.5	2395.5	2605.0	2610.0	2673.5
P11	2574.0	2499.5	2599.0	2344.5	2497.0
P12	2539.5	2446.5	2553.0	2513.0	2562.0
P13	1192.0	1137.0	911.5	954.0	1059.0

Scores from the judges

We convert the scores to ranks from each judge for each pilot. The rank is the number of pilots who get a better score from the judge, plus one. The highest scoring pilot from a judge has rank 0+1=1 for that judge. The second highest scoring pilot has rank two.

Pilot	J1	J2	J3	J4	J5
P1	2803.0	2701.5	2888.0	2864.5	2732.5
P2	2744.0	2604.0	2781.0	2686.5	2808.5
P3	2735.5	2578.0	2722.5	2634.5	2649.5
P4	2612.5	2585.0	2676.0	2563.5	2745.5
P5	2442.5	2632.5	2651.5	2601.5	2548.5
P6	2657.0	2429.0	2636.0	2527.0	2710.5
P7	2714.5	2517.5	2566.0	2488.0	2599.0
P8	2554.5	2446.0	2506.5	2504.0	2706.0
P9	2674.5	2405.5	2629.5	2488.0	2583.0
P10	2595.5	2395.5	2605.0	2610.0	2673.5
P11	2574.0	2499.5	2599.0	2344.5	2497.0
P12	2539.5	2446.5	2553.0	2513.0	2562.0
P13	1192.0	1137.0	911.5	954.0	1059.0

Scores from the judges

Pilot	J1	J2	J3	J4	J5
P1	1	1	1	1	3
P2	2	3	2	2	1
P3	3	5	3	3	7
P4	7	4	4	6	2
P5	12	2	5	5	11
P6	6	10	6	7	4
P7	4	6	10	10	8
P8	10	9	12	9	5
P9	5	11	7	10	9
P10	8	12	8	4	6
P11	9	7	9	12	12
P12	11	8	11	8	10
P13	13	13	13	13	13

Ranks from the judges

Next we find the overall ranking that has the fewest number of disagreements with the individual judge rankings. A disagreement occurs when the overall ranking places pilot A before pilot B while an individual judge places pilot B before pilot A. We call this ranking with the fewest disagreements the "consensus ranking."

Pilot	rank	J1	J2	J3	J4	J5
P1	1	1	1	1	1	3
P2	2	2	3	2	2	1
P3	3	3	5	3	3	7
P4	4	7	4	4	6	2
P5	5	12	2	5	5	11
P6	6	6	10	6	7	4
P7	7	4	6	10	10	8
P8	8	10	9	12	9	5
P9	8	5	11	7	10	9
P10	10	8	12	8	4	6
P11	11	9	7	9	12	12
P12	12	11	8	11	8	10
P13	13	13	13	13	13	13

The Consensus Ranking

Looking down the ranking table you can readily verify that each pilot is preferred over the next pilot by a majority of judges. If the number in a judge column is lower for a pilot than the number for the next pilot down, the judge in that column ranks the first pilot before the next.

We have now ranked the pilots for the flight.

Next we need to give each pilot a score. A score does two things for us:

1. It allows us to combine flights using the sum of scores to select a champion.

2. It tells us how far the first horse is ahead of the second horse.

We assign scores as follows:

The first place pilot gets every judge's best score.

Pilot	rank	J1	J2	J3	J4	J5
P1	1	1	1	1	1	3
P2	2	2	3	2	2	1
P3	3	3	5	3	3	7
P4	4	7	4	4	6	2
P5	5	12	2	5	5	11
P6	6	6	10	6	7	4
P7	7	4	6	10	10	8
P8	8	10	9	12	9	5
P9	8	5	11	7	10	9
P10	10	8	12	8	4	6
P11	11	9	7	9	12	12
P12	12	11	8	11	8	10
P13	13	13	13	13	13	13

The Consensus Ranking

rank	J1	J2	J3	J4	J5	score
1	2803.0	2701.5	2888.0	2864.5	2732.5	2813.10
2	2744.0	2604.0	2781.0	2686.5	2808.5	2717.90
3	2735.5	2578.0	2722.5	2634.5	2649.5	2685.80
4	2612.5	2585.0	2676.0	2563.5	2745.5	2659.20
5	2442.5	2632.5	2651.5	2601.5	2548.5	2618.30
6	2657.0	2429.0	2636.0	2527.0	2710.5	2609.50
7	2714.5	2517.5	2566.0	2488.0	2599.0	2583.60
8	2554.5	2446.0	2506.5	2504.0	2706.0	2546.50
8	2674.5	2405.5	2629.5	2488.0	2583.0	2546.50
10	2595.5	2395.5	2605.0	2610.0	2673.5	2519.90
11	2574.0	2499.5	2599.0	2344.5	2497.0	2509.60
12	2539.5	2446.5	2553.0	2513.0	2562.0	2437.20
13	1192.0	1137.0	911.5	954.0	1059.0	1050.70

Scores given the first place pilot

The second place pilot gets every judge's second best score.

Pilot	rank	J1	J2	J3	J4	J5
P1	1	1	1	1	1	3
P2	2	2	3	2	2	1
P3	3	3	5	3	3	7
P4	4	7	4	4	6	2
P5	5	12	2	5	5	11
P6	6	6	10	6	7	4
P7	7	4	6	10	10	8
P8	8	10	9	12	9	5
P9	8	5	11	7	10	9
P10	10	8	12	8	4	6
P11	11	9	7	9	12	12
P12	12	11	8	11	8	10
P13	13	13	13	13	13	13

The Consensus Ranking

rank	J1	J2	J3	J4	J5	score
1	2803.0	2701.5	2888.0	2864.5	2732.5	2813.10
2	2744.0	2604.0	2781.0	2686.5	2808.5	2717.90
3	2735.5	2578.0	2722.5	2634.5	2649.5	2685.80
4	2612.5	2585.0	2676.0	2563.5	2745.5	2659.20
5	2442.5	2632.5	2651.5	2601.5	2548.5	2618.30
6	2657.0	2429.0	2636.0	2527.0	2710.5	2609.50
7	2714.5	2517.5	2566.0	2488.0	2599.0	2583.60
8	2554.5	2446.0	2506.5	2504.0	2706.0	2546.50
8	2674.5	2405.5	2629.5	2488.0	2583.0	2546.50
10	2595.5	2395.5	2605.0	2610.0	2673.5	2519.90
11	2574.0	2499.5	2599.0	2344.5	2497.0	2509.60
12	2539.5	2446.5	2553.0	2513.0	2562.0	2437.20
13	1192.0	1137.0	911.5	954.0	1059.0	1050.70

Scores given the second place pilot

… and so forth on down.

That is the procedure.

Rationale

The rationale for using individual judge ranks, rather than scores to rank the pilots is this: The individual judge ranks eliminate differences in judge scoring styles or quirks. We ask judges to apply the judging criteria as they know them consistently. So long as they use consistent criteria they will fairly rank the pilots as they see the flights.

When we derive ranks from the scores, there is no longer such a thing as a judge who gives generally low scores, or a judge who gives generally high scores. There is no longer any concern with the range of scores given by the judge.

Using ranks derived from the scores is the simplest, most straightforward, most justifiable way to eliminate any differences in scoring styles.

The consensus ranking method, known as the Kemeny-Young method, was published in 1959 by John Kemeny. Kemeny was a mathematician and computer scientist who co-invented the BASIC programming language. He served as Chair of Mathematics and later as President at Dartmouth college. Hobart Payton Young, a game theorist with many pedigrees in England and the United States, added his name to the method by proving, in 1978, that it is the **best estimator of the true preference order**. Read that again. It is proven there is no better ranking of the pilots.

It also makes intuitive sense. Who is the better pilot? The pilot whose scores show that they are preferred by the majority of judges.

What remains is deciding a method for combining flights.

Justification for assigning scores

The practice of assigning scores to pilots is the hardest thing to accept in this method. Pilots think of the scores as their scores. It feels counter-intuitive to give the score to someone else. Note that the assignment of scores does not change the ranking or the outcome of the flight program.. It only enables us to combine flight programs into an overall winner.

There are three rationale for the assignment of scores after ranking the pilots:

First, winner takes all. The best ranked pilot gets the best from all of the judges. The second best gets second best.

Second, a judge's highest and second best scores are wholly and completely representative of what that judge awards a best and second best performance on the flight. You cannot concoct better first and second best scores for that judge. We treat the score given each pilot as representative of the score that each judge would give for a pilot of that rank.

Third, this method of assigning scores has benefits with regard to honesty of the judging, as we will explain.

Resistance judging bias or favorites

There are two aspects of the method that make it resistant to strategic grading. The first has to do with consensus ranking. The second has to do with the application of scores.

First refer to the example. In a straight average computation, if judge five had given pilot one the score they gave their seventh ranked pilot, then pilot one would have been knocked into second and the majority judge ranking would have been violated. Judge five could tank pilot one. A well designed report of the flight could make that behavior obvious. The pilot could rightly file a protest, but straight averaging cannot prevent the crime.

On the other hand, with the consensus ranking, a rank of seven given to pilot one by judge five does not change the result. Pilot one still wins. In fact, judge five can give their very worst ranking to pilot one. They can rank pilot one at thirteen. Pilot one will still win.

One judge cannot by themselves tank a pilot.

Second, suppose judge five intentionally gave their very best grade to pilot two and a poorer grade to pilot one. Where did that best grade from judge five go? It completely backfired because it went to

pilot one. There is absolutely no incentive for a judge to give their favored pilot an undeserved superior grade. That grade will go instead to a pilot higher in the consensus ranking. It will put their favored pilot not further ahead, but further back in the running.

The practice of giving each judge's best score to the pilot favored by the majority of the judges completely fouls any judge attempt to favor a pilot.

Conclusion

So what we have is a system for translating judge grades into pilot standings that everyone can follow, visually verify and accept; that has lots of nice qualities; and that is highly resistant to intentional or unintentional tampering from judges. The pilot most favored by the majority of judges will always win. Always.