TBL: Saint or Villian?

Many competition fliers, judges, officials, and administrators are still not sure of the benefits that the TBL statistical averaging system provides to radio control aerobatics, because they have not taken the trouble to understand the process. Some have even gone to the extreme by suggesting that it may be manipulated by scoring personnel and competition organisers to favour particular fliers or nations. Because they do not understand the system, it is considered ‘evil’.

The arithmetical calculations (algorithm/algorism) can not be altered, and score manipulation is simply not possible. It is true however that this algorithm is not usually made public, since it would allow any layman to have an attempt at writing scoring software, with disastrous results. It is also a fact that not all people are mathematically inclined, and may not have knowledge of intricate mathematical calculations, and the algorithm will not mean anything to them. The following explanation is taken from the IAC rule book and will hopefully give a layman’s guide to what it is that TBL does, and does not do.

Competition aerobatics is scored by a panel(s) of judges who grade each manoeuvre in each flight on the basis of their perception of how well the competitor flies against the criteria of perfection, smoothness and gracefulness, positioning, and size. Judges, just like the competitors they grade, are not infallible. Whether caused by bias, inaccuracy, inconsistency, experience level, fatigue, or other factors, there will always be some variation in the scores. Different judges also have different “styles”. Some score generously, others score miserly. Certain judges use the full range of scores, while the scores of others are may be clustered around a single value (maybe 7, or 8). It is because of these human factors that we use a panel of five judges rather than a single judge.

The TBL (Messrs Tarasov, Bauer, Long) system bases its foundations on the premise, that by combining the scores from several judges in a proven statistical way, we can eliminate many of those undesirable effects.

Even if a single judge viewed an identical flight a number of times, slightly different scores would be recorded for each flight. There is simply an element of randomness in the scoring process, no matter how competent, accurate, consistent, objective, experienced, or unbiased the judge may be. There is however, a low probability that this randomness will be identical for all the judges on a panel. This is why statistical methods give a more reliable result than simple averaging, or discarding of high and low scores, as was used in the past.

The use of straight averages for scoring will result in the judges who use a wide range of scores having far more influence on the final standings than the judges who confine their scores to a narrow range. Discarding the high and low scores does not remove the second or third out-of-range, or biased judges, and often removes judges who are the most rigorous with their scores.

The TBL statistical averaging system normalises the scores, and removes the biased and out-of-range scores from the final result, even if there were more than one biased or out-of-range score for a single flight. In operation, TBL first normalises the individual judge’s scores to remove the distortion caused by different judging styles. This is done by:

- raising the scores of the low-scoring judges
- lowering the scores of the high-scoring judges
- compressing the scores of the wide-range scoring judges
- expanding the scores of the narrow-range scoring judges.

This normalisation process results is a mean (average) score, and a scatter (distance from the mean) for each judge. In case you are concerned that this process corrupts the ranking of the

competitors: when the scores are normalised in this manner to remove stylistic differences between judges, the relative standing of each competitor is not affected. An individual competitor's score may be bumped up or down, but the relative finish order, as seen by each judge, does not change.

If each judge's score is compared to the mean, it will be seen that some scores are greater than the mean, and some lesser than the mean (this difference is termed "scatter"). This means, the scores do not fall in a straight line, but within a window, with the mean at the centre line.

How wide should that window be? That depends on how confident we want to be that a given score is valid. Based on many trials and analysis on contest data from the full-size aerobatic fraternity, it was established that that one "standard deviation" (a bit less than 70% of the scores population) was good enough. To increase the population would increase the risk of passing biased, or incompetent scores. To lower the population would discard too many valid scores.

With the width of the window set, TBL removes any scores that are outside the window, and averages the remaining scores to produce a final score for the competitor. The removal of those out-lying scores is not done on a black and white basis. The window in TBL has "fuzzy" edges. Instead of a score being simply in, or out of the window, the score is examined to determine exactly how far outside the window it is. If it is only slightly out, the score is still used, but is given less weight (importance) when the final average is calculated. This means that TBL progressively removes scores near the edge of the acceptance window from the final results, rather than using an all-or-nothing approach.

When a judge's score does fall completely outside the "fuzzy" edge of the window, the score is totally discarded, and not taken into consideration in the final calculation of the score for a competitor.

TBL can "fix" inconsistent, biased, and inaccurate judging, but it can never turn a bad judge into a good one. All R/C aerobatic judges are urged to study, know, and apply the Sporting Code and the Judges' Guide, and to obtain as much practical experience as possible. Then there will be no need for the TBL system to discard your scores, even if the scores are modified to more accurately reflect the performance of the competitors. Competitors are urged to study the TBL process, so that they may understand why a TBL score if often different from a raw score, but that the relative placing and ranking of an individual competitor does not change.

Any TBL scoring system should only be applied where there are at least five judges and at least twenty competitors, since there would then be enough data to work with.