

# CIVA FairPlay System Working Group Proposal – 2017



The development work discussed here was carried out by the FPS Working Group as a response to French proposal NP #2017 introduced at the CIVA 2016 plenary conference and Matthieu Roulet's paper "[MRoulet Rebalancing CIVA Scoring Method](#)", which together questioned whether a scoring system could be developed so that the competitors scores obtained in a programme would not depend upon the scoring of other competitors.

## Summary

Though primarily concerned with the FairPlay statistical system the working group has reviewed many of the simpler methods used by other major sporting regimes such as swimming, gymnastics, high-board diving etc. which deliver fixed or stable individual scores from judges opinions, in case any could simplify or render obsolete the more sophisticated techniques employed by FPS to provide the fairest possible result taking into account all relevant data. In every case however by setting a "fixed" result as the primary criterion the balance and coordination of the underlying judgements are handled with less regard for bias and fairness, and for CIVA the outcome would very likely be unacceptable. Doug Lovell's IAC publication "[Throwing Away High and Low](#)" illustrates many examples from this minefield, though IAC continue with raw averaging at all US events except their annual Nationals, where FPS and ACRO have been employed throughout for several years.

By contrast a fundamental requirement of FPS methodology is that all stored scoring data within each programme should be assessed every time results are calculated in order to provide the most robust, fair and unbiased result possible, and therefore individual pilot score and rank refinements are inevitable as more data becomes available. It follows that if the existing FPS algorithms can be developed to stabilise these minor variations this would present a more appropriate solution than moving to a radically different system offering an inherently inferior level of fairness and accuracy.

## FPS developments

Using existing championship contest files we have found that a significantly improved level of stability between successive results calculations as more pilots are added into the database can be achieved by replacing the existing Go/No-Go method of data rejection with immediate substitution of Fitted Values by one that provides smooth and proportional Fitted Value substitution over a range of uncertainty values. It can also be argued that this revised approach more closely mimics the normal human response to matters of judgement, in which personal deliberations normally move through a gradual transition from sure to unsure.

In the FairPlay process other items that can cause instability between successive calculations are:

- a) The FPS "60% Rule" stipulates that if more than 60% of judges' marks for any pilot/figure combination fail their acceptance test then all must be rejected and replaced by their Fitted Value. This assessment is influenced by all other FPS considerations within each group, and on the margin can occasionally slip from one state to the other. The logic for this item is not

at fault, and in this area too a more flexible and proportionate solution has been found that provides improved stability.

- b) When more than 30 pilots have flown a re-calculation is triggered if any scores are below 60% (in Knowns) or 50% (in Unknowns) that temporarily excludes the lower performing pilots and thereby avoids any influence that they might have on pilots with scores above these levels. Minor variations in scores can on occasion change the selection of pilots in the above/below sections, and inevitably there is some effect on the results.
  - Revising the 30 value to 11 for example would be more logical and also encompass smaller (e.g. glider) events as well as the larger power championships.
  - Discarding this system entirely is also a supportable move, as in reality the judges' opinions are influenced by everything they see during each programme and thus arbitrarily excluding some data for some of the time is not logically secure.

Further work to resolve these features are still being assessed to provide a recommended solution. Whatever is decided most appropriate will not however affect the underlying statistical evaluations that FPS uses to detect and address mark and score anomalies.

The W/G therefore proposes that the 'proportional' FPS method (FPS2) should be adopted by CIVA for use from 2018 forwards.

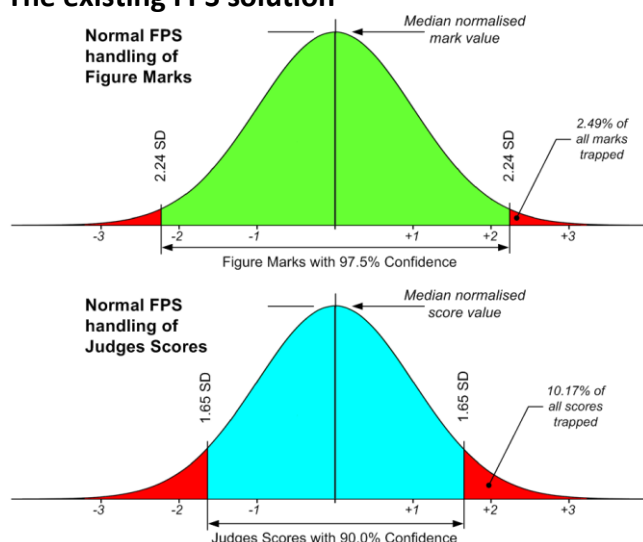
### **An explanation of the FPS judges mark and score handling process**

The judges' marks are initially collated into groups within which the figures have notable similarities so that as far as possible 'like' is compared to 'like' and logical data comparisons can be established.

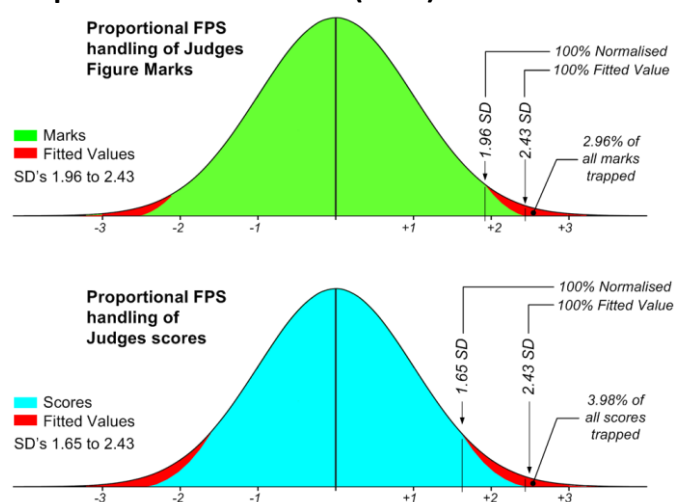
After the judge styles have been normalised (balanced) within each group, each data point is assessed by its Standard Deviation (SD) to measure its "fit" within the group with respect to each pilot and judge. In the existing FPS solution marks with SD greater than 2.24 (97.5% confidence) or scores with SD greater than 1.65 (90.0% confidence) are deemed unacceptable; these are replaced by a Fitted Value that is calculated after the rejected grade(s) have been removed, to match the judge's style.

In the proposed FPS2 solution however the transition from normalised mark to Fitted Value is made gradually, and in each case the final mark is the result of a smooth change from one to the other. Marks and scores are still subject to careful analysis throughout, but successive results iterations are significantly more stable and we have found that as more scoring data is added into the system the stability of pilots' scores is considerably improved and hence their rank positions are less likely to be affected. Whereas in the current FPS approach some 2.5% of all marks and 10.2% of all scores are likely to be rejected and the Fitted Value abruptly substituted, in the proportional FPS2 version with the SD's pairs that we have found to be most suitable just 1.7% of marks and 5.5% of scores transition smoothly from their normalised or 'balanced' value to their Fitted Value, and with these settings we have found the influence on results stability at its most effective.

## The existing FPS solution



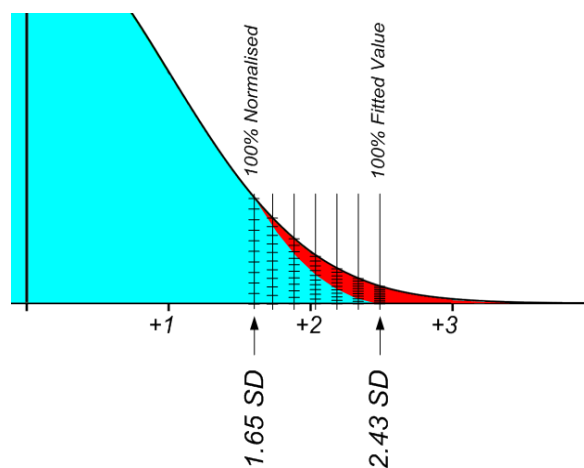
## Proportional FPS solution (FPS2)



In our development of the proportional FPS2 method, from many trials we have found the following SD ranges to offer the most stable and repeatable advantages –

	SD low	SD high	Confidence %
Figures:	1.96	2.43	95.0 to 98.5
Scores:	1.65	2.43	90.0 to 98.5

The broadened scope of this method is such that at each successive calculation of the results the additional judging data is unlikely to cause the marks and score SD's to vary significantly from their prior value and thus their influence on score (and hence rank) changes is of reduced concern.



## Appendix

On the next pages are some graphic illustrations of the ranking output as produced by a modified version of the ACRO scoring software, using typical data from programmes at recent glider and power championships. They show the variations in scores and hence the rank positions of each flown pilot at each possible stage of the programme, starting with just 3 flights completed (the minimum that FPS can handle) right through to the last. Pilots have been added according to the Flying Order established at the event, so that each one appears in the results in the correct order.

When using any system that reviews the value of all available data every time it is run to create, as far as possible, a balanced and bias-free overall result, it is inevitable that successive re-evaluations of the data as the pool content increases will reach similar but slightly different conclusions. The Working Group is satisfied however that this modified process provides a robust improvement over the existing system, which has remained virtually unchanged since its first use in 2005.

In each case here it is very clear that the use of the proportional FPS2 method makes a positive contribution to the stability of calculated successive rank orders by comparison with the established

FPS system, and markedly reduces instances of rank exchanges between pilots with similar scores. It can also be seen that the final scores provided by the proportional system are very similar to those from current FPS methodology.

We have also concluded that the proportional FairPlay system described here is likely to provide a marginally more realistic assessment of all the available data, and can therefore be accepted by CIVA as a beneficial change that implies no impropriety toward prior established results.

Note that when the revised logic of FPS2 is incorporated into CIVA's scoring systems it will be necessary that the software retains the original FPS process so that this can be used for legacy contest data files to accurately repeat championship results from years prior to 2018.

Nick Buckenham  
CIVA FPS Working Group chairman

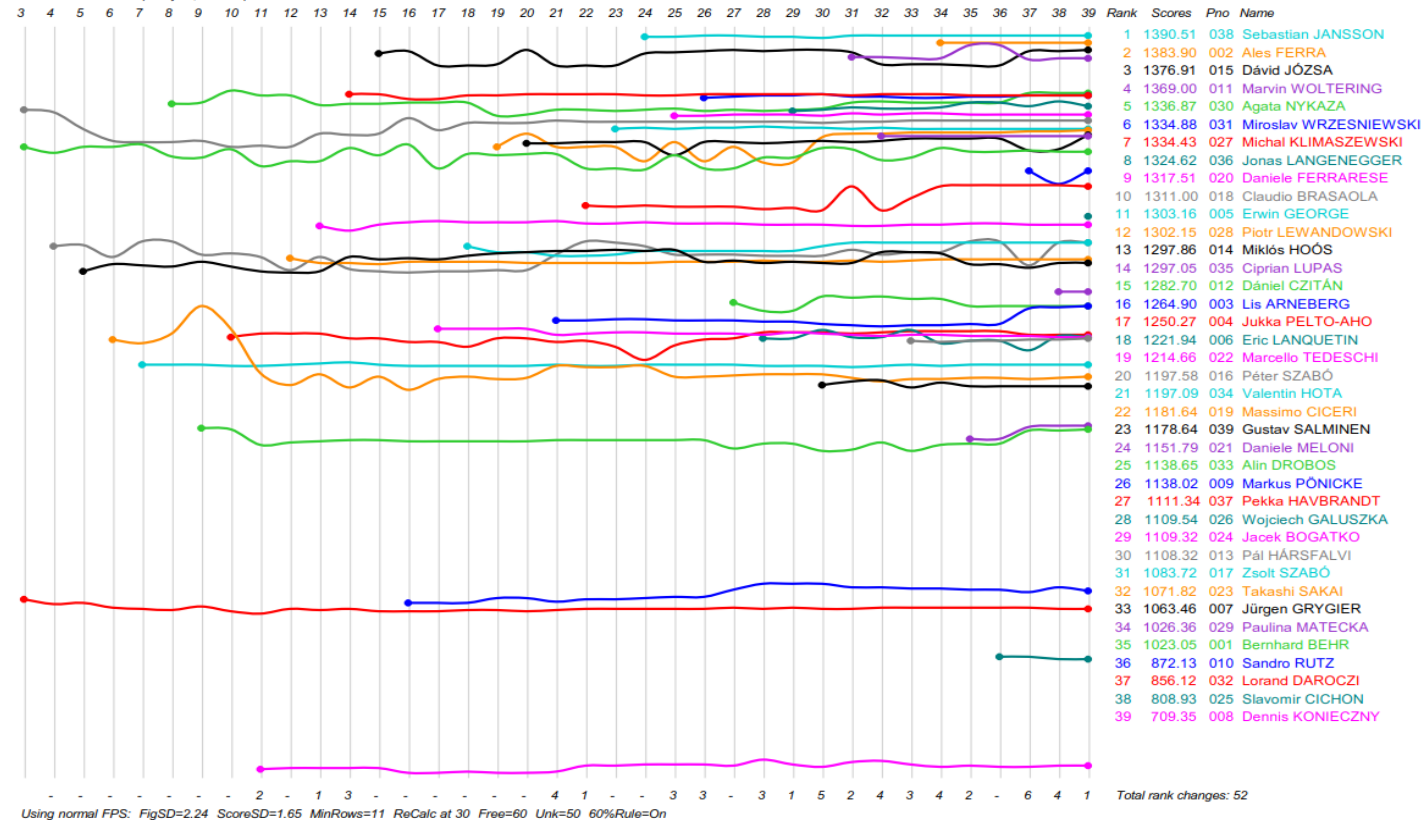
FPS W/G Members:

Gilles Guillemard	FRA
Doug Lovell	USA
Mikhail Mamistov	RUS
Vladimir Machula	CZE

# Progression of Pilot Scores Report: Free Unknown, 7th WAGAC

Data is in scores order per cycle, F/O steps 3 to 39 of 39

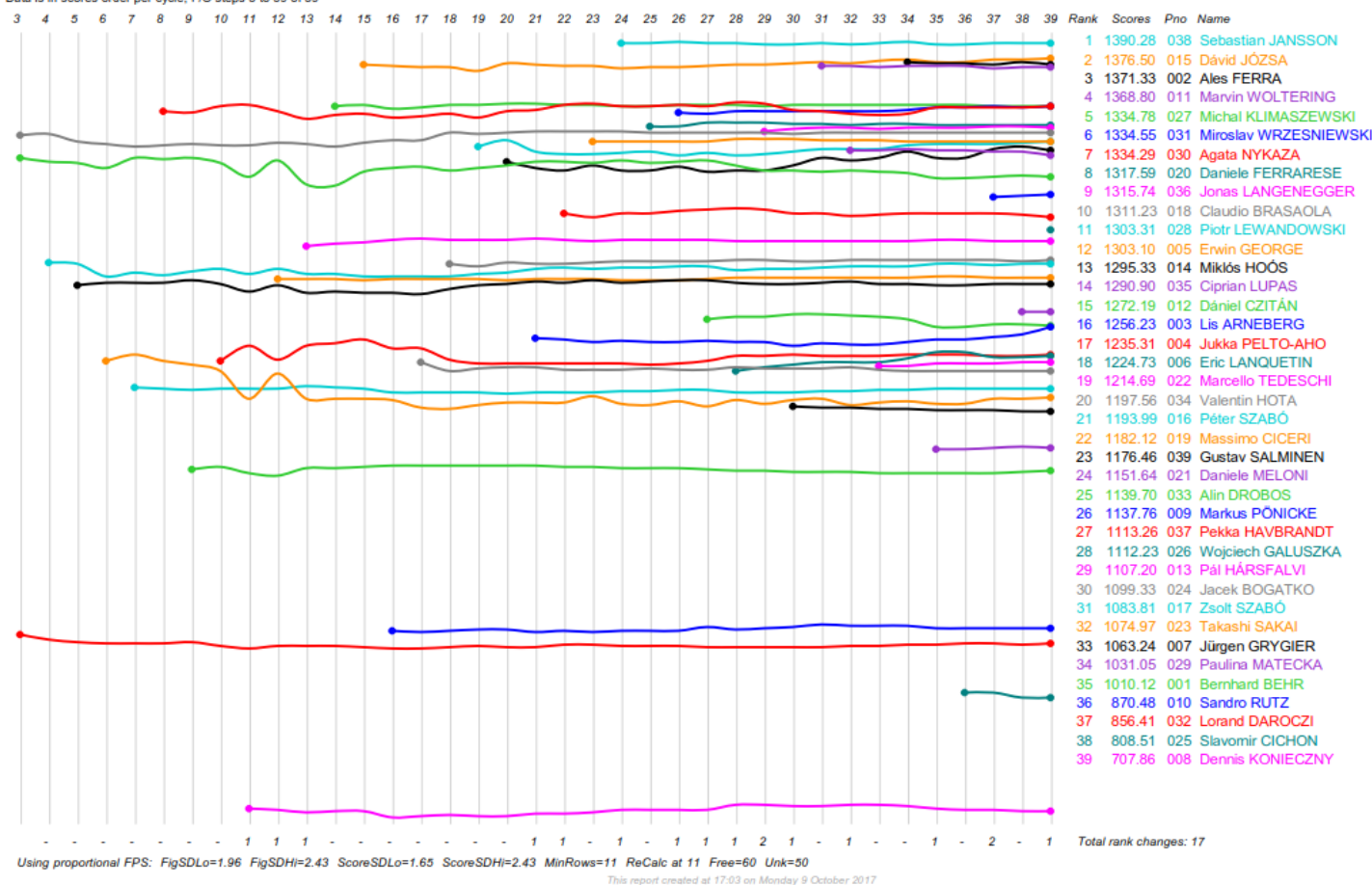
FPS: Normal 2.24 & 1.65, Re-Calc @ 30 per CIVA



# Progression of Pilot Scores Report: Free Unknown, 7th WAGAC

Data is in scores order per cycle, F/O steps 3 to 39 of 39

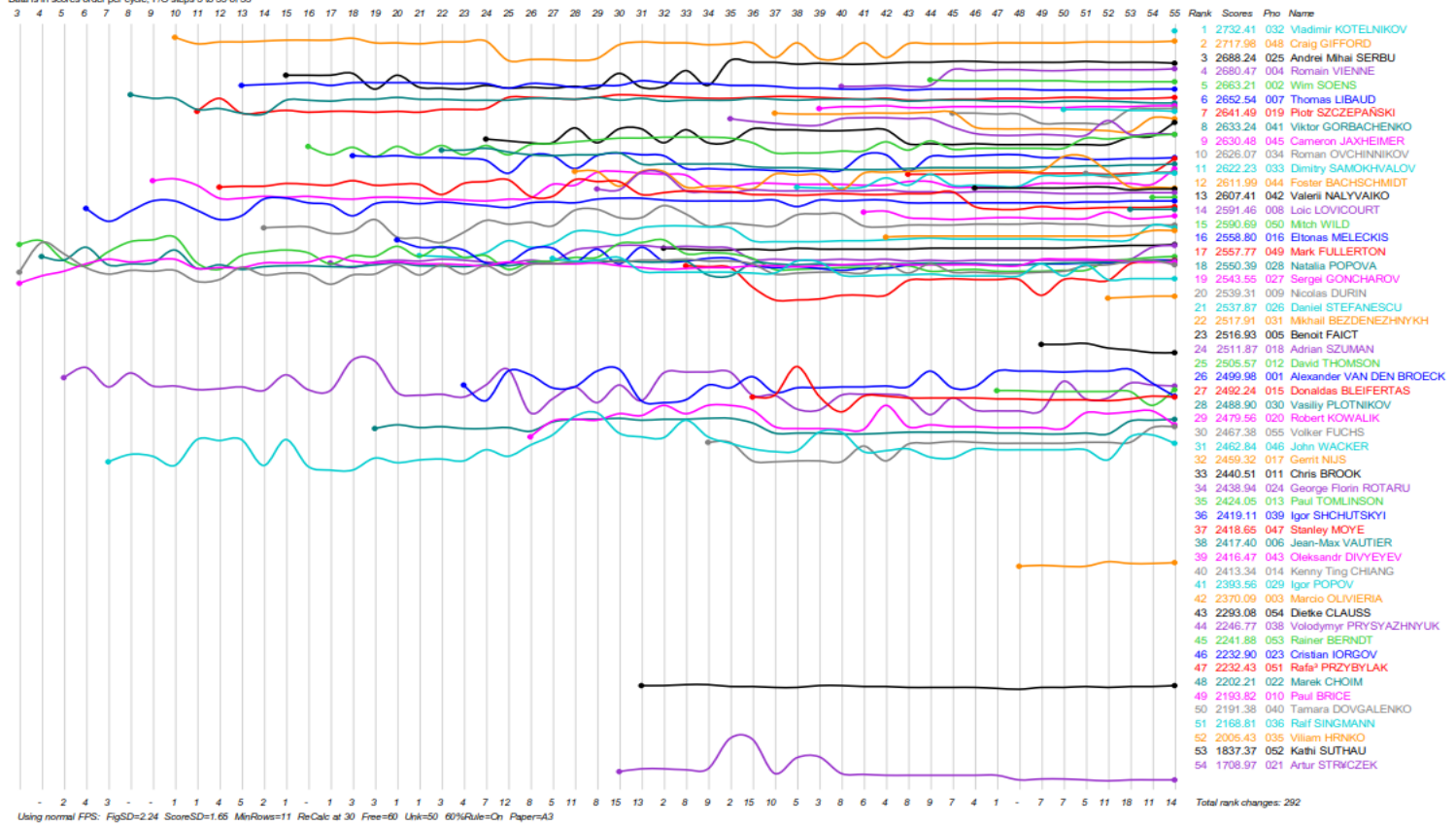
FPS2: Figs 1.96 to 2.43, Scores 1.65 to 2.43, Re-Calc @ 11



# Progression of Pilot Scores Report: Programme 1 - Free Known, 12-th FAI WAAC 2016

Data is in scores order per cycle, FIO steps 3 to 55 of 55

FPS: Normal 2.24 & 1.65 per CIVA, Re-Calc @ 30



# Progression of Pilot Scores Report: Programme 1 - Free Known, 12-th FAI WAAC 2016

Data is in scores order per cycle, FIO steps 3 to 55 of 55

FPS2: Figs 1.96 to 2.43, Scores 1.65 to 2.43, Re-Calc @ 11

