Summary of an Analysis of Scoring Systems for UKAC Contests

Introduction

A team of volunteers from both inside and outside the RSGB VHF Contest Committee have been working to analyse scoring systems for the Tuesday evening UK Activity Contests. This report gives the output from that study and is background information to feed into the RSGB VHF Contest Committee Rules Consultation for 2017. That consultation, which will follow in the next week, will provide a formal means for providing feedback.

The report is broken into two pieces - this shorter summary of the analysis and conclusions, and a much more detailed report for those people who are interested in looking at the subject in more depth.

Details of Study

The objective we set ourselves were to come up with a scoring system which

- Will be recognised as fair by the vast majority of entrants across a wide geography
- Is straightforward enough to be easily understood and explained (and coded)
- Works for all bands from 6m through SHF
- Works for all sections of the contest (Open, Restricted and Low Power)

The primary rule in taking part in this analysis was to put any pre-conceptions behind us, and to use the real data from the newly published open logs of 2016 to tell us what was actually happening, and what would happen if we tried different schemes.

As a result of this work, we believe that we have come up with a proposal for a fairer scoring system for these contests, but in the process we have been surprised by much of what we uncovered, and by how inaccurate some of the speculation on the behaviour of different scoring systems was.

What constitutes fairness has been a critical element of the work. We have adopted the concept of a Potential Score for a square. From the data in real logs showing what stations were active in the contest, the relative score that a station in each of the UK¹ squares could achieve if they had equivalent skills, equipment and take-off is calculated. A correction is applied for different types of scoring systems (bonuses or multipliers), and then the closer that the resulting scores are across the whole UK for a particular scoring system, the fairer it is.

We haven't tried to compensate for differences in station's equipment, operating skill, or take-off. We don't think that we *should* compensate for equipment and skill – these are hard won factors which contesters develop through experience. It would be nice to compensate for take-off, but we haven't been able to determine a sufficiently reliable way of doing this with public domain data.

What we *have* tried to compensate for is the differences in the numbers of QSO partners which are accessible to a particular station to work in the contest based on their location and the actual way that activity is spread around the country.

To make sure that the theoretical analysis of this new scoring system doesn't create unexpected results, we also rescored the real contest logs against each scoring system. This is to make sure that

¹ Throughout this document, where UK is used – read UK & Crown Dependencies

we're not changing the basic structure of these highly successful and popular contests, but at the same time we are helping the most disadvantaged stations.

We repeated this analysis for each band from 6m to 23cm, and for Open, Restricted and Low Power, and in some cases for multiple months of logs.

The detailed analysis is described in detail in the main report, but a brief summary is provided here.

- The amount of activity outside the UK, and the impact on scores is much less than commonly assumed, even from the East Coast. UK activity represents 96% of the QSOs made across the UK as a whole on 2m.
- M5 is broadly more favourable to stations in the North of the country, while M7 is more favourable to stations in the South, but both schemes exhibit equivalent levels of bias when measured across the country as a whole.
- Multiplier systems generally make scoring less fair than 1 pt/km or bonus based scoring because they exaggerate the differences between high and low activity areas.
- The people who are most disadvantaged by both M5 and M7 (relative to 1 pt/km) in these contests are those in Central and Northern Scotland. By the time that you venture that far North, activity is very thin indeed.
- Bonus based scoring systems provide a means to reward the most disadvantaged stations better than multiplier based schemes.

A New Proposal – B2

Based on this analysis, we are proposing a new bonus based scheme (B2) with the first contact in each square adding a bonus of 500, 1000, or 2000 points as shown in Figure 1 below.



- We believe that this scheme results in a significantly fairer scoring system relative to either M5 or M7 across the whole of the UK
- It significantly helps the most disadvantaged stations uplifting them by a few places in the results table

- It provides some incentive for stations to take the necessary time to complete aircraft scatter or meteor scatter contacts with the most remote stations. This won't work for everyone, but could be a useful technique for mid-table stations to enhance their score, at the same time making the contest more interesting and rewarding for the stations on the far edges of activity.
- It does not result in disruptive changes to results across the majority of stations and so keeps the popular character of the contest
- It generates scores which are closer together across the table. We have received feedback that some entrants (particularly newcomers) are put-off by the huge numerical gap between their score and that of the leaders. This scoring system will help that.

An Alternative Proposal – M8

We have also received some feedback that entrants would like to retain a multiplier based scheme, but not M5 or M7. We have spent some time looking for a fairer system than M5 or M7 across the whole country, and propose a new option called M8 as shown below in Figure 2.



Figure 2 - M8 Multiplier Structure

This scheme is significantly fairer across the UK as a whole than M5 and M7, and, compared to B2, achieves similar but smaller impacts to positions of the most disadvantaged stations.

Computer Logging

One of the challenges of variable multiplier or bonus scoring systems like this is that they are not currently scored accurately during the contest by any logging packages except for MINOS. Not having the score calculated correctly during the contest doesn't stop you taking part, it just means you won't know your multiplier or bonus score accurately during the contest. After the contest, when you submit your log to the contest entry web pages, our entry system will re-calculate those multipliers or bonuses correctly. You can then use this score to populate the Claimed Score system if

you like. However, knowing your correct score during the contest could help you optimise your strategy as you go along, although this can be achieved quite effectively just by making paper notes.

The results of a recent mini-survey suggest that this is not an issue for the majority of our entrants. We will discuss how to get wider software support with the authors.

We think these scoring systems work well to at least 23cm, but not at SHF with very low QSO totals and short ranges. We are recommending 1 point/km scoring for the SHF UKACs.

Conclusions

We believe that the B2 proposal outlined here is a very strong candidate for a new UKAC scoring system from 6m through to 23cm. It appears to provide a fairer scoring system across the whole UK than either the existing M5 or M7 mechanisms, without changing the fundamental character of a set of remarkably successful contests. It should also make contesting more enjoyable for those in remote areas. This should result in maintaining and encouraging higher VHF activity levels in these areas.

We have also offered an alternative multiplier based scheme called M8. This is less effective at achieving the goals of uplifting the positions of the most disadvantaged stations although it achieves this to some degree, but is significantly more even handed across the UK as a whole than M5 or M7.

We don't believe that the lack of support for real-time scoring for B2 or M8 in all logging software should be a block to adoption of this scoring system and will be approaching software authors to provide support for these scoring systems.

Detailed Analysis of Scoring Systems for UKAC Contests

Background

Discussions about the fairness of scoring systems in contests have probably been raging for as long as contests have existed. Over the past few years there has been a particularly fierce debate over the scoring systems used within the regular UK Activity Contests which take place between 6m and 10 GHz on Tuesday evenings. These have been a tremendously successful set of contests involving very large numbers of entrants – with over 600 in a year at 2m, over 400 on 6m and 70cm, and even over 200 on 23cm and 4m. Nevertheless, there continues to be a passionate debate on how they should be scored.

The activity which we are reporting on has attempted to providing a solid, evidence based analysis of a number of scoring systems, with the aim of identifying the fairest across the whole of the UK and was prompted by the RSGB Presidential Review of contesting which took place in 2015, and subsequent discussion and debate.

The team of volunteers carrying out this analysis have been drawn from both within and outside the RSGB VHF Contest Committee, and consists of

Andy Cook, G4PIQ (JO02) - Chair Pete Lindsay, G4CLA (IO92) Mike Goodey, G0GJV (IO91) Ian Pawson, G0FCT (IO91) Roger Dixon, G4BVY (IO82) Richard Baker, GD8EXI (IO74) John Quarmby, G3XDY (JO02) Martin Hall, GM8IEM (IO78) Steve Hambleton, G0EAK (IO93)

The objectives that we set ourselves was to come up with a scoring system which

- Will be recognised as fair by the vast majority of entrants across a wide geography
- Is straightforward enough to be easily understood and explained (and coded)
- Works for all bands from 6m through SHF
- Works for all sections of the contest (Open, Restricted and Low Power)

The primary rule in taking part in this analysis was to put any pre-conceptions behind us, and to use the real data from the newly published open logs of 2016 to tell us what was actually happening, and what would happen if we tried different schemes.

As a result of this work, we believe that we have come up with a proposal for a fairer scoring system for these contests, but in the process we have been surprised by much of what we uncovered, and by how inaccurate some of the speculation on the behaviour of different scoring systems was.

But - what does fair mean?

This was one of the most important questions that we had to answer. We decided that there were some factors that we either shouldn't or couldn't compensate for. These included operator skill and station capability and optimisation. Learning how to operate efficiently, accurately and tactically, and learning how to build the most competitive station are all skills which we develop in order to be successful in contesting. We didn't think that we should be compensating for these hard learnt traits.

The other factor that we felt that we could not compensate for was a station's take-off. There is no doubt that this is a major factor affecting success in VHF contests, but the effects can be tremendously localised - moving just 10 metres can turn a great site into a relatively poor one if this moves you behind an obstruction, be it topography or a building. Apart from being fiendishly complex to devise and calculate, publicly available topography data with its 50m lateral resolution and 10m height resolution does not look granular enough to provide a highly reliable handicap system. While it might work in many situations to provide a good estimate of the capability of a site, there will be a significant number of situations where it is inaccurate – and generally in the direction that a station will have a much poorer take-off than estimated. For these reasons, we have not attempted to compensate for take-off.

But – the biggest source of complaint which we have received over the years is that some stations are a long way from the centre of activity and so are naturally handicapped by their location. This was an area that we felt we could tackle, and the document outlines the approach we have taken.

What have we done?

We've generated the concept of a 'potential score' from a particular location. This is the score a notional station with a good 360 degree take-off could achieve from that location with good operating and good equipment. We generate the potential score for a square by taking the following steps

- 1. Take the real logs from a specific UKAC event from 2016 and record the location of every logged station
- 2. For every large locator square (e.g. JO01) in the UK, identify a location approximately in the centre of activity
- 3. Calibrate the model according to the performance of the leading station on that band and section. Do this by having a sliding scale of the probability of working a particular active station based upon distance. (E.g. 40% chance of working stations up to 200km, 30% for stations from 200 to 300km, 20% for 300 400km, and so on up to 800 km). These percentages vary by band and section and are set such that the model predicts a similar mix of QSOs, score, multiplier total and best DX for the leading stations to that achieved in practice.

- 4. For every square in the UK, calculate a potential maximum distance based score.
- 5. Based on the performance of the leading station, set a radius inside which all multipliers / bonuses are worked, and calculate the multiplier / bonus score.
- 6. Calculate the overall potential score for every UK locator square for each band and section with each scoring system under consideration.

We have used standard deviation by population (st dev) of those potential scores across all UK squares as a means of identifying relative fairness of a scoring system. The results did not have a normal distribution about the mean so this is not the classical use of st dev. It was also necessary to take the root of the multiplier based scores so they could be compared with bonus based scoring systems. However we did check we end up similar types of distributions and so we believe the st dev was indicative of spread of results between squares.

We then went on to re-run the actual entrants' logs for each of these contests and sections against each of the new scoring schemes to ensure that the effect on the real contest results is as we would expect it to be, and that we are not generating an obviously skewed set of results.

What scoring systems have we looked at?

We started by looking at some of the recent and common scoring systems, in particular

- 1 point per kilometre worked. No multipliers
- M5 1 point / km + a multiplier of x1 for all UK locator squares worked
- M7 1 point / km + a multiplier of x2 for all UK locator squares worked, and x1 for all others worked

The results of this initial analysis are shown below in Figure 3.



Figure 3- 2m Open Potential Score Distribution - Jan 2016 Data

It takes the highest scoring square for a particular scoring system and marks that as 100%. It then shows the potential percentage of that score which could be achieved by an equivalent station / operator / take-off in all the other squares based on the analysis described above. The chart also shows a line with the average (mean) score by square for each scoring system, so squares above the line are doing better than the UK average on that scheme, and those below the line worse. The square by square impact of different scoring systems is more easily visualised below in Figure 4.



Figure 4 - Square by Square Potential Score against Mean Score - 2m Open Jan 16 Data

We are using the 'tightness' of the distribution of scores across the whole country for a particular scoring system as a measurement of fairness – the narrower the distribution, the fairer that particular scoring algorithm is. We have measured that tightness of distribution by taking the standard deviation of the scores in all squares. We took this analysis one stage further and looked at the tightness of the distribution across both all UK squares, and those squares with at high activity (those squares which have had 20 or more active stations in 2m UKAC from Jan – Aug 2016). This gives the 'Most Active Squares' distribution shown below in Table 1.

	All Squares	Most Active Squares
1pt/km	0.186	0.112
M5	0.208	0.116
M7	0.211	0.089

Table 1 - Standard Deviation for Existing Scoring Systems for 2m Open Jan 2016 Data

So – what does this tell us, and why do we think this happens?

- When viewed across the UK as a whole, M5 and M7 are similarly fair. When viewed across the squares with significant activity, M7 is fairer than M5.
- M5 broadly favours stations in the North of the country, while M7 favours stations in the South.
- The differences are much smaller than many expected because the commonly held assumption of high activity levels in Europe is in-correct. The reason for this is that the UKAC events have become so successful that, even at 2m, less than 4% of the total QSOs are from outside the UK, and even in the JO squares, this figure only rises to just over 7%. Activity is completely dominated from the UK in particular from activity in IO83, IO91, IO92 and IO93.
- The people who are most disadvantaged by both M5 and M7 (relative to 1 pt/km) in these contests are those in Central and Northern Scotland. By the time that you venture that far North, activity is very thin indeed.

Activity level by square in the 2m UKACs between Jan and Aug 2016 is shown below in Figure 5. For a station to count as active they must have worked 4 or more stations – this helps us remove busted calls from the data.



We have repeated this analysis for all bands up to 23cm, and for all sections. This data is included in the Appendix A, but the patterns within the results are consistent across the bands and sections.

A proposal had been made previously for a new M6 mechanism (UK squares + DXCC Countries). This was found to be not significantly more or less fair than either M5 or M7 and so was dismissed from further analysis.

The original multiplier scheme of M2 where all squares counted equally was also briefly tested, but likewise found to be less fair than both M5 and M7 and so was also dismissed from further analysis.

How do we test if this is a valid model?

Models are all well and good, but they are just that – models of the real world. We need to have some confidence that the model is not at odds with reality. This output of this model suggests that some of the best places to operate the UKAC contests from are in the IOx4 and IOx5 lines of squares.

Table 2 below shows the distribution of top 3 placings in the 2m Open section of the contest from January 2016 to September 2016. While, because of variations in station capabilities and take-offs, this in an imperfect check, there is a good selection of leading stations (66%) in the IOx4 and IOx5 line of squares. This analysis feels valid for the Open Section where there is reasonable representation of competitive stations in many of the major squares. That situation does not exist in the Restricted Section for example.

Locator	1064	1074	1085	1092	1094	JO01	JO02
No of position #1,2,3 for	5	9	1	7	3	1	1
2016 2m Open section							

Table 2 - Number of 1st, 2nd and 3rd places in 2m Open Section 2016 to September by Square

An addition measure of validity of this type of model was to compare the actual points per QSO of the leading stations across the country with the predicted results from the model as shown below in Figure 6. This gave a correlation coefficient $R^2 = 0.81$ which implies a good fit between model and reality.



Figure 6 - Relationship between Actual Pts/QSO and Predicted pts/QSO

There has been a previous analysis based on a model developed by Martin GM8IEM during the 2015 Presidential Review of Contesting, which was presented to and accepted by the review workshop. It was subsequently scrutinised in detail by the Contest Committee and used in the development of B1 prior to the 2015 Supplemental Consultation. This work was based on the best information available at the time, i.e. population/amateur density both within the UK and Europe rather than the actual activity density which is used in this analysis. It is only this year's launch by the VHFCC of Open Logs, allowing everyone to see what is actually being worked, which has enabled us to move from population density to this much better informed analysis based on real activity distribution. The previous model exaggerates the relative advantage enjoyed by stations in the south-east of England because of the low activity levels within Europe relative to the UK described earlier, and also because there is a very high population density peak in London which has never

been reflected in VHF SSB activity. Nevertheless, despite its limitations, the general conclusion that Scottish stations are disadvantaged by the M7 scoring rule has been confirmed by our most recent analysis., which also demonstrated for the first time that stations in the south of England were disadvantaged to a similar level by M5.

A New Proposal – B2

Having established a baseline for the existing scoring systems, we set about looking at alternatives which would be fairer, particularly to the most disadvantaged areas, while retaining good fairness across the whole of the UK. We established earlier that multiplier based schemes inherently magnified the differences between successful and less successful areas, so we also looked at bonus schemes. In these scoring systems, rather than the 1 pt/km score being multiplied by a factor based on number of squares worked, an additional number of bonus points are awarded for each square worked. This approach is used in Scandinavia for their Nordic Activity Contests which have a very similar format to the UKACs.

Last year, VHFCC consulted on a bonus scheme which was called 'B1', and which was fairly narrowly rejected as a scoring system for 2016. We have looked hard at the best bonus options, looking at many different options – some simple, some complex. However, in the end we have found that B1 was a very good starting point for a fair scoring system, and by making a few small alterations to it, we are now proposing a system called 'B2' which is simpler and the fairest option that we have found to date. The bonus points per square for B2 are shown in the map below in Figure 7. Note that all non-UK squares count for 1000 bonus points.



The distribution of southerly UK squares at 500 points and setting non-UK squares at 1000 points may appear odd at first sight, but this mix results in the narrowest distribution of scores across the UK.

Another alternative – M8

We have also received some feedback that entrants would like to retain a multiplier based scheme, but not M5 or M7. We have spent some time looking for a fairer system than M5 or M7 across the whole country, and propose a new option called M8 which takes the same grouping of squares, but assigns

them multipliers of x1, x2 and x3 as shown below in Figure 8. Again, this structure gives the narrowest distribution of scores across the UK that we have found



This scheme is fairer across the UK as a whole than both M5 and M7, and achieves similar but a smaller impact to positions of the most disadvantaged stations.

The effect of these scoring system relative to the existing scoring system of M7 is shown below in Figure 9 & Figure 10Error! Reference source not found., again using the 2m Open section as an example.



Figure 9 - M8, M7 and B2 Potential Scores for Jan 2016 2m Open Data



Figure 10 - Square by Square Potential Score against Mean Score - 2m Open Jan 16 Data

	All	Most
	Squares	Active
		Squares
M8	0.190	0.071
M7	0.211	0.065
B2	0.154	0.053

Table 3 - Standard Deviation for M8, M7 and B2 Scoring Systems for 2m Open Jan 2016 Data

From this analysis we learnt that

- B2 exhibits a higher degree of fairness than any existing scoring scheme, both across the UK as a whole, and in the higher activity areas. It is clear from the graphs that B2 has a narrower potential score distribution than M7 or M8 even after correcting for the multiplicative structure of M7 and M8.
- M8 has a narrower distribution than M7 and so is fairer across the whole UK, and provides some correction to the most disadvantaged stations.
- B2 delivers a significant enhancement in score for the most badly disadvantaged stations uplifting them by a few places in the results table
- Neither B2 or M8 result in disruptive change to results across the majority of stations.
- Both B2 and M8 provides some incentive for stations to take the necessary time to complete aircraft scatter or meteor scatter contacts with the most remote stations. This won't work for everyone, but could be a useful technique for mid-table stations to enhance their score, at the same time making the contest more interesting and rewarding for the stations on the far edges of activity.

One of the challenges of a variable bonus based scoring system like this is that it is not currently scored accurately during the contest by any logging packages except for MINOS. Not having the score calculated correctly during the contest doesn't stop you taking part, it just means you won't know your multiplier or bonus score accurately during the contest. After the contest, when you submit your log to the contest entry web pages, our entry system will re-calculate those multipliers or bonuses correctly. You can then use this score to populate the Claimed Score system if you like. However, knowing your correct score during the contest could help you optimise your strategy as you go along, although this can be achieved quite effectively just by making paper notes.

We recently ran a mini-survey to establish the importance of having a real-time display of score to entrants. MINOS (which will score system B2 correctly) is used by approximately 50% of entrants. About another 20% currently use loggers or techniques which do not provide an accurate real-time view even with the current M7 system, and then of the remaining 30% of entrants, there was a mild preference towards accurate real-time scoring being important to them.

For these reasons, we don't believe that the current lack of full scoring support in some logging software should prevent us proposing this option. We will discuss gaining support for this mechanism with the other major logging software authors.

Our testing shows that this scoring system works well to at least 23cm, but can deliver skewed results where QSO totals are very low. For this reason, we are not recommending the B2 scheme for the SHF (2.3 GHz – 10 GHz) UKACs and will be providing an alternative recommendation for these contests.

Impact on Actual Results

We have also reviewed the impact of B2 and M8 on actual contest results. In general, the impact on position in the tables has been relatively small to the majority of entrants. The biggest impact is on stations in the more remote parts of the UK (IOx6 line and North in particular), moving them usefully up the tables.

An example of the overall impact on position is shown below in Figure 11. This graph shows the positions of the entrants relative to M7 for M8 and B2. An entrant dropping below the trend line moves up the table, one moving above the trend line moves down it. You can see from this that this scoring system does not undermine the successful overall structure of the contests, but B2 in particular generally helps to promote the disadvantaged.



Figure 11 - Positional Moves by Square or May 2m Open

What's different about this year's analysis?

During last year's review of scoring options, there were a number of points made against the original B1 proposal which we have tried to address in this study. Specifically

- Analysis flawed based on population data.
 The analysis from 2015 was based on population, this analysis is based on actual activity levels
- The playing field cannot be levelled. It's very true that we can't completely level a playing field, but it's not unreasonable to make it better where possible – that is what this proposal seeks to do
- B1 too complex.

Both B1 and B2 are more complex than some other scoring systems, but they are not out of line with scoring systems used worldwide and are not actually difficult to keep track of - even manually

- Logging Programs won't support it This has been covered in some detail in the paper. Software support is by no means essential. MINOS will support these options, and we will approach other software authors for support.
- Selection of Bonuses Arbitrary

A wide variety of different bonus structures and levels were analysed during this work. The B2 proposal is the one which delivers the tightest distribution that we have found to date.

- B1 is non-reciprocal (i.e. points accrue to the station working the higher bonus square) This is true, but the approach here should encourage stations to make the effort to work the more distant stations which should result in higher scores for them.
- No objective stated for change / no way to measure success This is quite straightforward. We measure success through the number of entrants and the geographical distribution of those entrants. It is hoped that this proposal will allow the contest to continue to grow, especially away from the main population areas.
- Contest should promote activity within the UK rather than between the UK and the Continent This series of contests is actually about encouraging activity inside the UK, whether those contacts are to Europe, or within the UK. As shown earlier in this paper, these contests have been so successful within the UK that the QSO totals and score contribution from overseas QSOs are very small, even for the East Coast stations, and 96% of stations worked – even on 2m are in the UK.

Conclusions

We believe that the analysis carried out in this study is new, in that it takes a purely objective view of how the distribution of available stations impact a stations ability to score well in the UKAC contests. We believe that the B2 proposal outlined here is a very strong candidate for a new UKAC scoring system from 6m through to 23cm. It provides a fairer scoring system across the whole UK than either the existing M5 or M7 mechanisms, without changing the fundamental character of a set of remarkably successful contests. It should also make contesting more enjoyable for those in remote areas. This should result in maintaining and encouraging higher VHF activity levels in these remote areas of the country.

We don't believe that the lack of support for real-time scoring for B2 in all logging software should be a block to the adoption of this scoring system and will be approaching software authors to provide support for this scoring system.

We commend the B2 mechanism as the fairest system which we have found to date, however, based on feedback from some entrants, we have also proposed an alternative multiplier based approach, M8, which improves in fairness on both of the existing M5 and M7 schemes, but is still less fair than this B2 proposal.



Appendix A – Potential Score Distributions & Corrected Standard Deviations for all bands and sections

	All	Most
	Squares	Active
		Squares
M8	0.211	0.090
M7	0.210	0.069
B2	0.156	0.064



6m Restricted





	All	Most
	Squares	Active
		Squares
M8	0.314	0.077
M7	0.299	0.083
B2	0.213	0.081

4m Open



4m Restricted

B2

0.205

0.060



4m Low Power







	0.150	0.07 1
M7	0.211	0.065
B2	0.154	0.053

2m Restricted

B2

0.162

0.054



2m Low Power

70cm Open

	Squares	Active Squares
M8	0.269	0.112
M7	0.294	0.082
B2	0.215	0.093

	All	Most
	Squares	Active
		Squares
M8	0.291	0.117
M7	0.336	0.099
B2	0.251	0.091

23cm Open

	All	Most
	Squares	Active
		Squares
M8	0.271	0.111
M7	0.280	0.083
B2	0.225	0.080

B2

0.227

0.071

23cm Low Power

