Handicapping algorithms for measured performance HCing of keel boats.

A TopYacht discussion paper.

Introduction

There are a great variety of HC methodologies and philosophies used for keel boat racing. We recently investigated a few of these in an endeavour to find a simple but effective HCing algorithm that would provide high quality, computer generated HCs.

Before entering the discussion it is important to note that we are concerned in this paper about "Measured Performance HCing" (not "measured boat" HCing). That is HCing where a performance HC is provided as a result of studying the elapsed times of a number of competitors over a number of races. I deliberately say "competitors" because the skipper and crew have a significant impact on the performance of most boats.

The two continuum

A key question is "What is the aim of HCing?"
There are divergent thoughts on this question.
The answers that HCers give to this question could be placed on a continuum.
One end of this continuum says that the handicapping should ensure that only the "best" sailor(s) should win, the other says that it is important to keep club interest up so let us have a system that simply shares the honours around each week.

A second key question: "how often should the HCs change and by how much?".
At one end of another continuum are those who believe that the HC is virtually a fixed value that is applied to a competitor and that it should not be allowed to vary up or down except for say an annual review. At the other end of the spectrum are those who believe that HC value must be allowed to move significantly after each race and should reflect the current form/performance of the competitor.

The observations

Through our work on the TopYacht race results software we have had access to data from a number of keel boat clubs. We have been able observe the "performances" of boats over a Series and in some cases over several seasons. By "performance" I mean the relative performance of a competitor when compared to the HC corrected time of a reference boat for that race. This reference boat can be selected in various ways. For all our investigations the reference (boat) time was either from a boat at a fixed percentile down the fleet or the average time of a large group of boats in the middle of the fleet. This "performance" or "Back Calculated Handicap" is the handicap each competitor needed to have finished on an equal HC Corrected time to every other competitor in the (division in the) race.
Mathematically: BCH = Ref Time / Elapsed time of Competitor.

These performances are most easily studied by looking at performance graphs of individual competitors. That is a graph showing the competitor's BCHs for each consecutive race in a series (see Performance graph below).
Other graphs of interest include a plot of HC Corrected times for all competitors in a race against their placing in that race (see race data graph below)

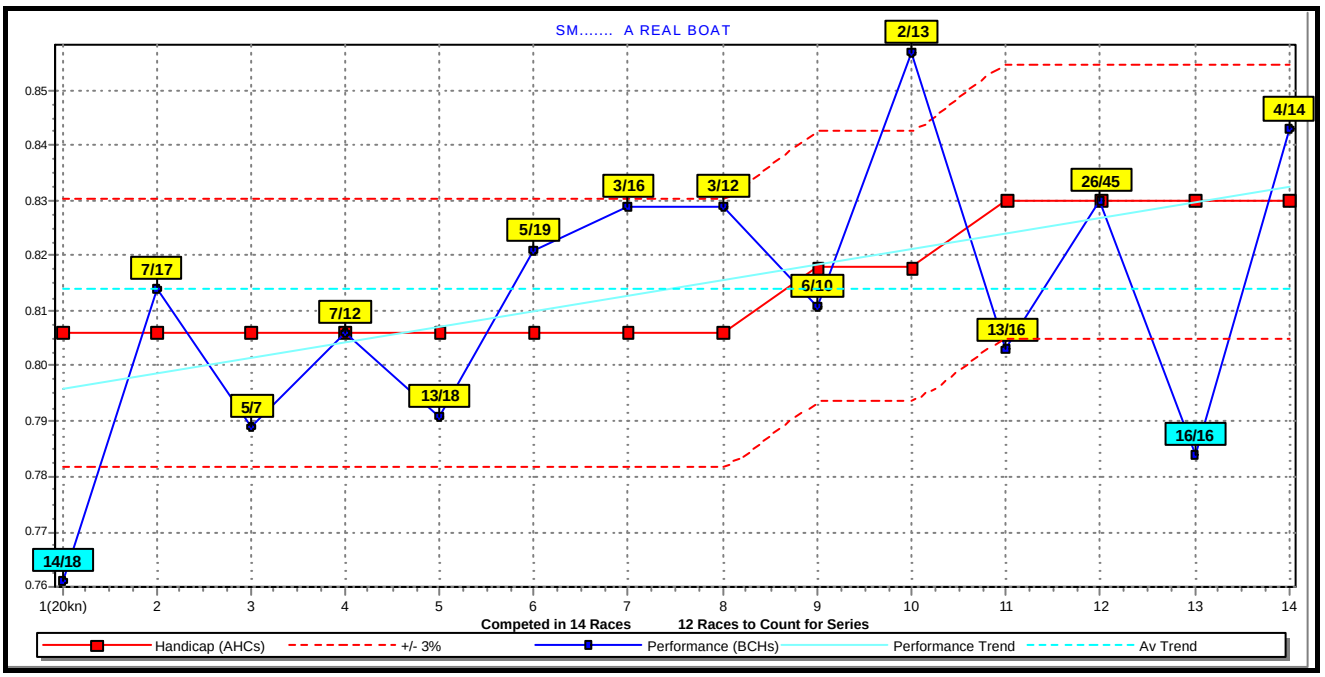
Purely from these observations a number of things stand out.
For any reasonable HC system .....

1) A competitor's performance varies up and down around a "mean" value.
2) Usually, most competitors performances remain with a +/- 3% window around their "mean" for a majority of races.
3) Some competitors fluctuate up and down considerably while other's performances moved up and down but not nearly so much.
4) It is extremely unusual for any competitor's performance to continue away from their "mean" for more than 3 consecutive races. Instances of 4 consecutive results each moving still further away from the "mean" were very, very rarely observed.
5) Over a series most competitor's performances had a neutral trend. That is, it did not gradually go up nor go down over the series. For those that did, the trend was usually rather gentle.
6) Most competitor's had one or two exceptionally good performances over say a 16 race series. For these races they performed well above their "mean". Values of 5% or higher were not uncommon.
7) Competitors occasionally performed well below their "mean"; values of -5% were not uncommon. Presumable when something went very wrong. Who moved that sandbank there?????
8) To win a race a competitor must perform well above his/her "mean". Those who perform at the "mean" being placed approximately mid fleet for that race.

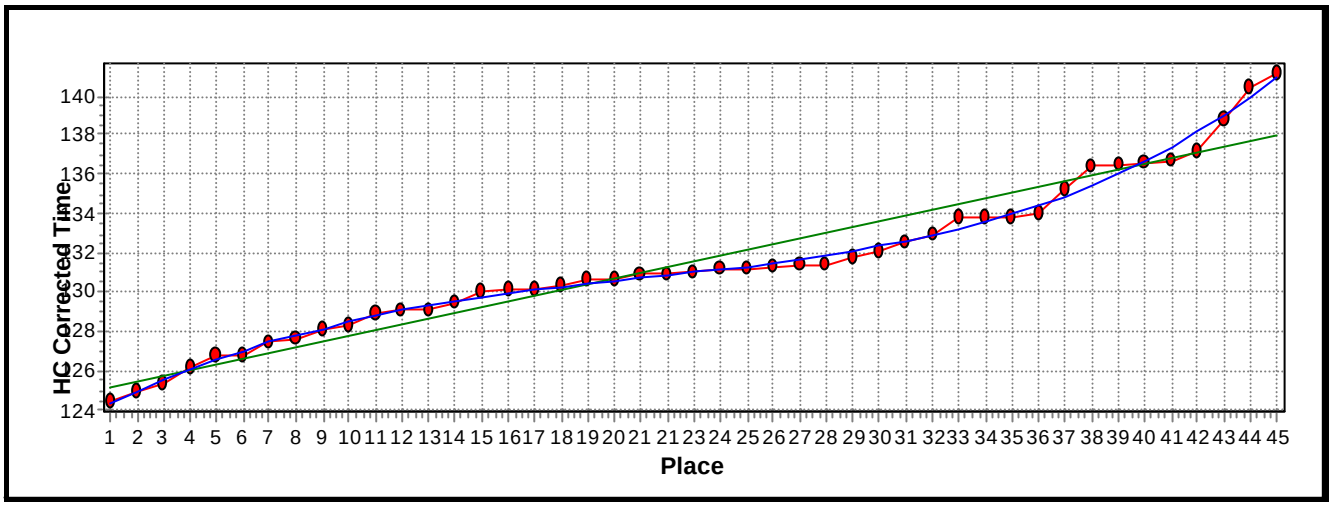
[It is a most interesting thing to study the performance graphs of the first 3 place getters in each division in a series.]

9a) Some series were won by competitor's who were entered into a series on far too generous a handicap and by the time it had been adjusted to a "fair" value, that competitor had sufficient placings near the top of the fleet to win the Series.
9b) Conversely others spent the entire series waiting for the HC to adjust down from the sky unlit it approached their actual performances.
10) Some Series are won by those whose performance genuinely improved through the series.

11) Studying individual races (see Race Graph, next page) it was not uncommon for the tail- ender(s) to be rather off the pace. It was surprisingly rare for a lead boat(s) to be a significantly ahead of the rest.



TopYacht Performance Graph



TopYacht Race Data

What are Tests for a "good" Handicap System?

Unfortunately such tests tend to reflect ones standing on the two continuum mentioned earlier.
So at this point I shall have to get off the fence and make a stand on those two continuum.

How do we set the criteria??
From talking to folks at a number of clubs there seems to be some consensus....

1 to provide HCs that give everyone a "fair" or even chance of winning, coming second etc.

2 BUT it is not deemed acceptable that an absolute novice should get up and win several races

3 the system should preclude anyone from deliberately "bending" their HC by putting in a few very poor performances

4 BUT the system should quickly react to someone who is repeatedly performing well above (or below?) Their "mean".

5 the HC should not just wander up and down on a weekly basis just as a reflection of normal performance fluctuations. (Remember most competitors move up and down less than 3% and rarely move away from the "mean" for many consecutive races before returning to it. So, say at most, 3 races going up to + 3% then 3 races back down to the "mean". An average difference from the "mean" of under 2% for most competitors. This is born out by the looking at the graphs.)

[Having said that; many competitors did fluctuate considerably within the +/-3% window and occasionally straying outside that window. Those competitors who ***constantly*** have wider fluctuations ***could*** be those whose boats favoured particular courses, wind directions, wind strengths, styles of racing, seas states or be those who ventured out in ***all*** types of weather and therefore experienced it all???]

The tests

If these criteria are accepted then possible tests might include.....

A) you would expect that a large number of different competitors should each have tasted a win in at least one race in the series. Likewise you would expect that the seconds, thirds etc were reasonably well spread out among the competitors over a series. See "Score Spread" table on the right.

[Notice the "split" scores of 5.5, 9.5 etc where the handicap corrected times provided ties for a place in the race. This is real data with no alterations.]

B) as a direct consequence of A) you would expect only a small spread of scores in the series or aggregate scores provided you ignore those whose scores were weighted down by DNCs DNFs etc.

TopYacht Software: Sandringham Yacht Club
Printed on : 11/04/00 23:14:04
Spread of Scores for :
SUMMER 1999-2000 VYC upto Race no 14
Handicap Parameters: Rods 25,80,c,4,4,0,0,l,y,y

Division	Score	No Diff Comps
1	1	11
1	2	13
1	3	10
1	4	9
1	5	8
1	5.5	2
1	6	11
1	7	11
1	8	10
1	9	10
1	9.5	2
1	10	9
1	11	10
1	12	8
1	13	6
1	14	6
1	15	6
1	16	7
1	17	4
1	18	3
2	1	9
2	2	8
2	3	9
2	4	11
2	5	9
2	6	9
2	6.5	4
2	7	9
2	8	8
2	9	8
2	10	8
2	11	6
2	12	8
2	13	8
2	14	5
2	15	4
2	16	3
2	17	2
2	19	2
3	1	5
3	2	7
3	3	5
3	4	5
3	5	6
3	6	2
3	7	4
3	8	1
3	9	1

TopYacht Score Spread

If the "mean" value or "handicap" value for each competitor is chosen to reflect their average performance and most competitors perform within +/- 3% of their HC then...

C) there should be a rather small spread of HC corrected times for the bulk of the middle of the fleet. This spread being in the order of +/- 3% of the average HC finish time. See Standard Deviation Normalised Corrected Times Table on the right..

[This table shows the Standard Deviation for the handicap corrected times for each race where the times have been normalised to provide a 100 minute average for each race. Notice that in fact the spreads were even closer than predicted. The predicted spread being +/- 3% whereas the real data is often closer to just 3%.]

If the HC system is working well then..

D) The average HC for a fleet should show little or no overall "drift" over a season even though it may well vary a little from week to week. See HC Drift table below. ['0's indicate Abandoned races.]

Printed on : 11/04/00 23:13:29
Average AHC for all races in Series:
SUMMER 1999-2000 VYC upto Race no 14
Handicap Parameters: Rods 25,80,c,4,4,0,0,I,y,y

Race No	DIV 1	DIV 2	DIV 3
1	0.902	0.765	0.669
2	0.907	0.765	0.000
3	0.915	0.766	0.674
4	0.930	0.766	0.696
5	0.920	0.769	0.686
6	0.915	0.768	0.687
7	0.898	0.767	0.675
8	0.897	0.767	0.687
9	0.904	0.000	0.000
10	0.876	0.783	0.711
11	0.908	0.000	0.000
12	0.890	0.772	0.687
13	0.916	0.769	0.000
14	0.908	0.781	0.673
15	0.000	0.000	0.000

TopYacht HC Drift

TopYacht Software: Sandringham Yacht Club
Printed on : 11/04/00 23:08:56
Standard Deviations of Normalised Corrected Times:
SUMMER 1999-2000 VYC upto Race no 14
Handicap Parameters: Rods 25,80,c,4,4,0,0,I,y,y

Division 1 with Creep Correction Factor of 1.000

Race No	Std Dev	No Comps	Av AHC Comps
1	6.451	18	0.902
2	3.638	17	0.907
3	3.266	7	0.915
4	3.229	11	0.930
5	3.428	17	0.920
6	2.574	18	0.915
7	2.459	16	0.898
8	2.756	11	0.897
9	2.688	10	0.904
10	3.508	8	0.876
11	6.854	16	0.908
12	2.243	13	0.890
13	2.77	16	0.916
14	2.806	14	0.908

Division with Creep Correction Factor of 1.000

Race No	Std Dev	No Comps	Av AHC Comps
Av	3.476		

Division 2 with Creep Correction Factor of 1.000

Race No	Std Dev	No Comps	Av AHC Comps
1	12.43	17	0.765
2	2.765	14	0.766
3	7.121	9	0.766
4	1.924	13	0.766
5	3.145	15	0.769
6	3.265	15	0.768
7	2.658	16	0.767
8	5.342	12	0.767
10	5.265	7	0.783
12	2.886	17	0.772
13	2.545	11	0.769
14	3.563	13	0.781

Division with Creep Correction Factor of 1.000

Race No	Std Dev	No Comps	Av AHC Comps
Av	3.779		

Division 3 with Creep Correction Factor of 1.000

Race No	Std Dev	No Comps	Av AHC Comps
1	7.034	9	0.669
3	6.233	8	0.674
4	3.399	5	0.696
5	4.69	6	0.686
6	3.041	7	0.687
7	3.736	8	0.675
8	2.456	5	0.687
10	0.928	3	0.711
12	11.907	5	0.687
14	1.929	5	0.673

Division with Creep Correction Factor of 1.000

Race No	Std Dev	No Comps	Av AHC Comps
Av	3.24		

TopYacht Standard Deviation Normalised Cor'td Times

If we accept these as fair criteria then we can start comparing the results of various HC systems.

The bulk of the comparisons were done on the Summer 1999-2000 results for Sandringham Yacht Club Victoria Australia.

The systems compared.

Systems that seem to provide a sound mathematically model were examined. Those that just used the "knock down because you came first" model were rejected from the observation that most boats have at least one inexplicable "blinder" in a series and if a single high is normal then it is not in need of being "punished".

1) Results based on the HC provided by very experienced HCers who basically adjusted the HCs **manually** in line with the VYC HC system

2) A computer model of the **VYC** system that was simplified a little in that it did not provide for HC adjustments for a win. It provided a HC as a running average of the last 8 races (raced anywhere against a known fleet) with no upper clamp, a 3% lower clamp and where the two lowest and one highest performance were discarded leaving 5 race performances to average.

3) An **exponential average** where the new HC = (4-1)/4 of the previous HC plus 1/4 of the most recent performance i.e. BCH. The number "4" is referred to as the "gain".

4) A running average of the last few race performances where there was a strong weighting to the most recent races. I call it the "**more recent weighted running average**". The formula applied was HC =

   For 3 race
   3/6 Latest Performance Value, 2/6 Next oldest, 1/6 oldest.
   For 4 races
   4/10Latest,3/10 next oldest, 2/10 almost oldest,1/10 oldest.

5) A **simple running average** but **weighted** as follows.....
   performance values were clamped at +/- 3% if they exceeded that window
   if 2 or more values exceed +5% then the second and consequent values are included in full
   if the performance was 5% below HC then it was discarded and the next oldest chosen.
   The HC value was not altered unless it altered by at least 1% (we call this the "Step Size" parameter)

What was the result of the comparisons??

Not surprisingly all the systems tried provided "reasonable" HCs. By which I mean they all faired reasonably well on the tests defined above. Some were better in some tests, others were better elsewhere. It really came down to your position on those two continuum.

Comparisons

The **manual** system and the **VYC** system produced very similar results on all tests. When compared to other systems their biggest draw back was how long they took to respond to a significant change in performance by one or more competitors. This slow rate of change could easily hand a Series to a competitor.

The **exponential average**. It can be argued that trained statisticians would examine boat racing's characteristics and apply this method. It gave rather similar test results to the manual and VYC systems. My biggest concern is its slow reaction time to fast changes and its inability to consider any history of the competitor. If I changed the "gain" from 4 to 5 the rate of changes was even slower as was expected.

The **more recent weighted running average**. This system appeared to have merit in that it should provide a quicker response to changing performances.
It provided the closest aggregate results when it was set to just consider the last 3 races. It was judged to be poor to average on the other tests.
I consider that it has several problems.
The graphic evidence suggests that 3 races do not give a fair "averaging" indication of a competitor's performance.
Secondly the HC values generated by this system fluctuate considerably from race to race and this I believe is largely reflecting the normal spread of performance rather than providing a useful tool.
Finally, with only 3 races being considered it is too easy for it to be "mis-trained" by one or two abnormal races.
I tried to remedy these perceived problems by applying various clamps and limitations as described above and below, but in each case this resulted in the aggregate results being more spread out than without the constraints.
I increased the number of races to 4. At this stage the system delivered results no better than the other alternatives on most tests.

The simple running average (weighted).
If the number of races considered was set to 4 then this system seems to deliver the best compromise while not providing quite the closeness of aggregate results of the previous contender. It was on a par with and often better than the manual and VYC systems on the various tests.
If the number of races was reduced to 3 then it provided closer Aggs scores than with 4 races but this again causes me the same concerns as noted above.

My choice?
I prefer **the simple running average (weighted)** with 4 races.
On all tests it is on a par with the **manual** and the **VYC** tests and is ahead on several tests.
While it does require 4 races, at least it only requires 4 races to give sensible adjustments.
It seems to fit the various observations made above.
It is a compromise in terms of a fast adjusting system but one that doesn't adjust too quickly.
It is not difficult to implement.
All the data shown above came from this system.

Other Issues
From the data examined there are two other potential problems.
1) when a new series starts, if this is say a winter series, then the HCs from the summer series may not fit too well as the boat may well be sailed by a different combination of crew and/or with a different intent eg. "We'll use winter to try some new ideas"

2) when a new (unknown) competitor first joins a series

Possible Solutions to those issues.
Provide a "fast Start" system that more quickly adjusts HCs for the first few races of a competitor in a series
OR
Allow a "fast start" system for those awarded a "P" for a provisional or new HC.

Run each series independent of other series, taking only a starting HC into the new series then providing a second algorithm that adjusts HC for the first few races.
This has several advantages

1) You don't need the computer to have access to the previous data for a competitor.
2) The previous race by race performance data may be from a series where the crew and/or attitude were different and does not really reflect heavily on how they will perform in this series. Further, if the HC system was working well in the previous series then the final HC should have encompassed the previous performances any way.

Measured Performance Handicapping limitations....
Like all 'one value only' HC systems that I have seen, MPHCing does not endeavour to take into account such factors as wind strength, sea conditions, amount of windward work nor the length of the course/race. But it does provide surely the simplest way of generating interesting keel boat racing at the local club level giving most competitors a reasonably equal chance of doing well and hence maintaining their interest in sailing. With just a little bit of cooperative effort it is also relatively easy to keep HCs in sync across different clubs provided there is some regular multi club events.

The future?
The data examined provides other interesting topics (several noted in the previous paragraph) to be explored. Several of these may further allow significant enhancements to this simple to administer form of HCing. But for now I hope that enough issues have already been raised to start some healthy discussion.