Hattrick analytics reloaded

The goal of the present report is to share some insight into the mechanics of the building-up and decay of the team spirit and self-confidence of national squads in the football manager game Hattrick.

The analysis has become necessary since the game developers have changed the base value for team spirit and self-confidence of national teams from 4.5 to 5.0 in summer 2021. Furthermore, the introduction of the simulating tool on the line-up page has provided new ways of estimating the team spirit and self-confidence of a squad one has access to. It needs, however, be stressed that the formulae proposed here are strictly speaking only valid for national squads since they have coaches with fixed properties, while in club teams, i) the rest value of self-confidence and team spirit, i.e. the value the team spirit and the self-confidence tends towards in absence of competition games has stayed at 4.5, and ii) the properties of the coach (leadership, coach level) and the assistants in the staff are affecting the mechanics as well. While some of the elements of the evolution of team spirit (TS) and self-confidence (SC) are probably the same for national squads and club teams, e.g. the SC increase due to a certain winning score, other elements may be distinct and need to be investigated. It is a first goal of this report to present a robust way to evaluate the current values of your squad in terms of TS and SC, in order to facilitate the sort of research that is needed.

In the present report we will therefore focus on the following aspects:

1. How can we measure the current level of self-confidence?
2. How can we measure the current level of team spirit?
3. How does the SC evolve with time in absence of competitive games?
4. How does the TS evolve with time in absence of competitive games?
5. How does a certain score of a competitive game affect the self-confidence?
6. How does the firing of a player affect the team spirit?

Some of these points have been investigated in the past and are more or less well known to the larger public in HT. The focus of the present report is hence also on how can we improve the precision of the quantitative analysis.

The results presented here can be implemented in a calculation program of team spirit and self-confidence in virtually any programming language. For those who are not familiar with programming environments I will provide a few simple-to-use Excel tools that allow you to predict expected team spirit evolution for your team (and to some extent also your adversary's team) and to go further from this point. Bear in mind that the author is an Engineer and not a Designer. Feel free to modify the layout of the Excel file to your liking.

Data source and data analysis methodology

Data source

The data used for this work stem essentially form national teams I had access to at some point in time, mainly the Swiss national squad from autumn 2020 to X-mas 2021 (thanks to Morelos and Florissimo), but also the German NT (thanks to K_Oz829) for parts of the self-confidence study, and the NT of Luxemburg (thanks to Grabner) for the limited analysis of the firing of players on team spirit. Having access to the line-up page of an NT squad allows determining self-confidence and team spirit within relatively tight bounds, which greatly improves the quality of the available data.

As complementary data, the evolution of visible self-confidence of all 128 national teams participating in WC32 has been tracked on a daily basis, together with changes in visible SC levels before and after competitive matches. This has been continued for the SC and TS levels of the 64 teams participating in the WC33, which is to the author's knowledge the only fairly complete data set available since the shut-down of the Uruguayan NT-tracking activity.

Methodology of data analysis

The methodology of data analysis is based on the hypothesis that a given line-up of a national team (including the form of each player, its health status etc., yet neglecting effects of weather) leads to a specific rating in the midfield and the attack sectors at the beginning of the game. The effective rating of the midfield sector will then depend on some modifiers:

- whether the coach selected PIC, PIN, or MOTS
- whether the match is played home (h) or away (a)
- what tactics were selected (normal (n), long shot (LS), counter attack (CA))
- what the team spirit (TS) of the team is.

For the attack ratings the modifiers will essentially be the position of the orientation slide bar (100def…100off), the use of long shot tactics, and the self-confidence level of the squad.

It is further hypothesized that the modifying factors (from PIC to PIN, from MOTS to PIN, CA/normal, Home/Away, TS8/TS5 etc. for the midfield, and e.g. 30def/0 for the attack ratings) are stable and do neither depend on the rating itself, nor on whether a side attack or a central attack rating is concerned. It is also assumed that there is no interdependence between the modifying factors, i.e. the combined effect from a (n,a,PIC)-setting to a (CA,h,MOTS)-setting is the product of the individual modifying factors $f(n \rightarrow CA) \cdot f(a \rightarrow h) \cdot f(PIC \rightarrow MOTS)$ and the latter, i.e. $f(PIC \rightarrow MOTS)$ is the product of $f(PIC \rightarrow PIN) \cdot f(PIN \rightarrow MOTS)$. These two hypotheses are widely accepted amongst HT managers trying to understand the mechanics of the game. It still is useful to be aware of these as you will see in what follows.

Determination of MF-modifying factors

If we accept these hypotheses, we can express the expected MF-rating for any condition of the match in terms of the MF-rating of a convenient reference state, which will be defined henceforth as the midfield rating for a match with no special tactics (normal), played away, PIN, and an arbitrary TS level, TS_0. In mathematical terms:

$$MF(t,p,o,TS) = MF(n,a,PIN,TS_0)f(n \rightarrow t)f(a \rightarrow p)f(PIN \rightarrow o)f(TS_0 \rightarrow TS) \quad (1)$$

where $MF(t,p,o,TS)$ is the indicated MF-rating for the setting one wants to calculate, $MF(n,a,PIN,TS_0)$ is the MF-rating of the reference setting, the factors $f(n \rightarrow t)$, $f(a \rightarrow p)$,

$f(PIN \rightarrow o)$, and $f(TS_0 \rightarrow TS)$ are the modifying factors to be applied for the chosen tactics (t) (with regard to normal), the place of play (p) (with regard to away), the team orientation (o) (with regard to PIN), and the level of team spirit (TS) (with regard to TS_0), respectively. From this description it becomes clear that precise knowledge of these modifying factors is very important. We take therefore a closer look at these modifiers.

In the past the approach has often been to take a specific line-up and to compare the indicated midfield ratings for settings that are identical but for one parameter, e.g. $MF(n,a,PIN,TS_a)$ and $MF(n,a,PIC,TS_a)$ (where TS_a is the actual or current level of team spirit) to find the modifying factor for the transition from *PIN* to *PIC* by dividing the latter by the former. Example: if one gets an indicated MF-rating of 10.75 for PIN and 10.00 for PIC the modifying factor would be

$$f\left(PIN \rightarrow PIC\right) = \frac{10}{10.75} = 0.930233\ldots$$

i.e. a close to 7% reduction in MF-rating. For one given line-up one can find 6 different values for all possible combinations of the midfield-modifying tactics (*n,CA,LS*) and places of play (*h,a*). This number can be further increased if one varies also the team spirit to various fixed levels (Cold war to PoE). With 11 levels to choose from one can get 66 values for $f\left(PIN \rightarrow PIC\right)$ for each line-up. Of course, similar combinations can be operated to determine the modifying factor between home and away matches or for the various tactics. For a few line-ups one can generate several hundred of values of a given modifying factor and take their average value.

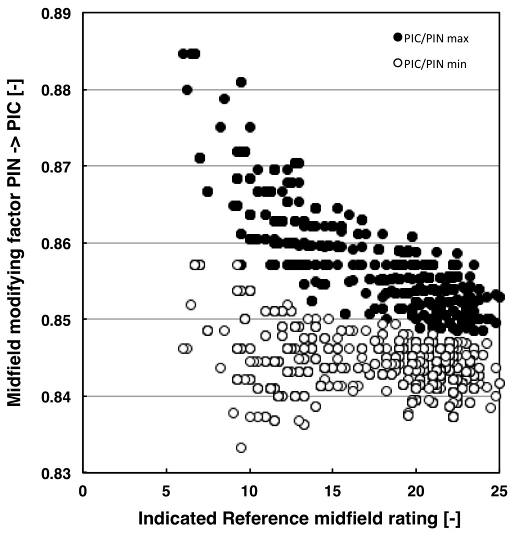
The problem with this approach is the following: It is generally accepted that the visible indication in the line-up predictor is the lower limit of the actual rating. A visible evaluation of 10.75 for the midfield means that the actual midfield rating is somewhere between 10.75 and 10.9999… or in mathematical terms in the interval [10.75,11) the parentheses indicating an open interval, i.e. not including the limit of the interval but getting infinitely close to it, while the brackets indicate a closed interval, i.e. one including the limiting value. This uncertainty in the actual evaluation obviously hampers the precise determination of the modifying factors. If we take the example from above, i.e. a MF-rating of 10.75 in PIN and 10.00 in PIC the actual value could be [10.75,11) for PIN and [10,10.25) for PIC. The modifying factor can then be anywhere between the two extreme cases, $f_{\max}\left(PIN \rightarrow PIC\right)$, $f_{\min}\left(PIN \rightarrow PIC\right)$:

$$f_{\max}\left(PIN \rightarrow PIC\right) = \frac{10.25}{10.75} = 0.953488\ldots$$

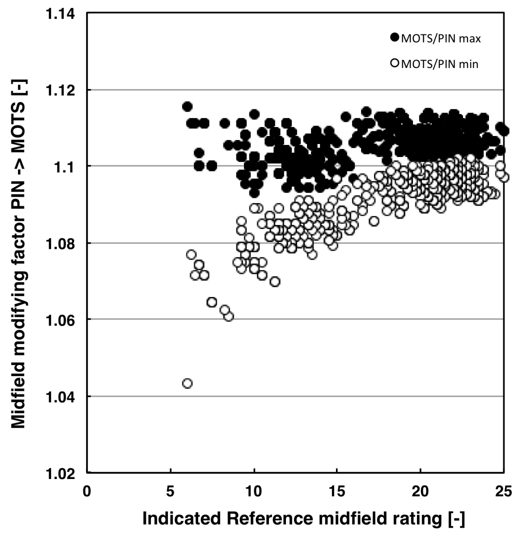
$$f_{\min}\left(PIN \rightarrow PIC\right) = \frac{10}{10.75} = 0.909091\ldots$$

If we do the analysis of a large number of setting combinations and line-ups we will find a very large number of those minimum and maximum values. The real value must then be in the range between the highest minimum value and the lowest maximum value in order to not create contradiction. This has been done for several thousands of different evaluations for each of these modifying factors. The result of the analysis is shown in Fig. 1 for the five factors $f\left(PIN \rightarrow PIC\right)$, $f\left(PIN \rightarrow MOTS\right)$, $f\left(a \rightarrow h\right)$, $f\left(n \rightarrow LS\right)$, and $f\left(n \rightarrow CA\right)$. Interestingly enough, in all evaluations the envelope, i.e. the imaginary line limiting the data points of $f_{\max}$, seems not to be a straight line but one that tends to lower values as the MF-rating goes up.

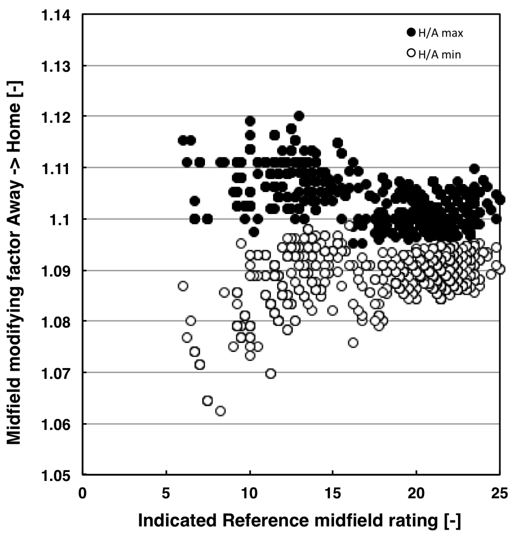
The same is true for the upper envelope of the f_{min} values. The really important point however is that the lowest f_{max} value is lower than the highest f_{min} value, meaning that there is no single value for the modifying factor that is consistent with all evaluations.



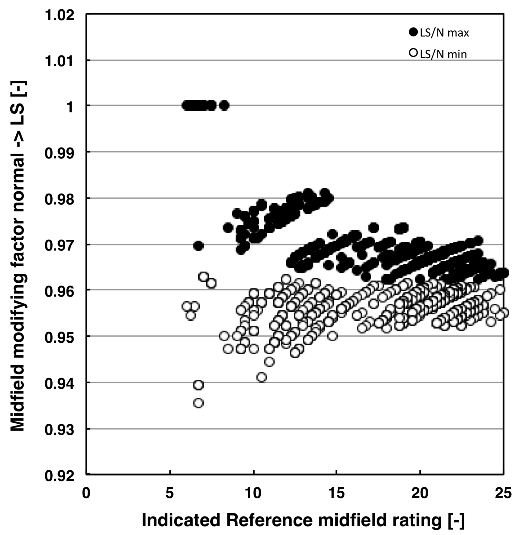
a)



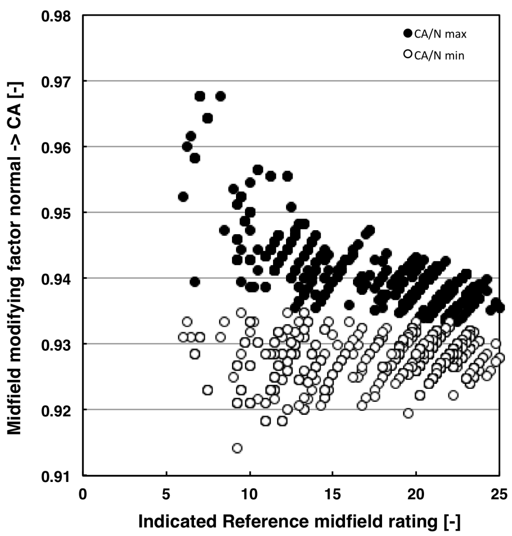
b)



c)



d)



e)

Fig. 1: Modifying factors for a) PIN to PIC, b) PIN to MOTS, c) Away to Home, d) normal to LS and e) normal to counter attack based on the indicated MF-ratings. The graphs show for all cases a clear separation between the f_{max} (full circles) and f_{min} (open circles) values. The value at which the separation happens changes as the MF-rating changes. Moreover, in all cases the lowest f_{max} value, which is an upper limit for a supposed unique f, is lower than the highest value of f_{min}, which is a lower limit for the supposed unique f. This leads to a contradiction.

The consequence of the way of evaluating the modifying factors, *f*, as discussed above is that depending on what typical MF-rating the evaluation is made one would end up with a different value for *f*. This may explain, why the indications for these modifying values have varied in the past depending on who has evaluated their values (and the typicial MF-rating they had at their disposal).

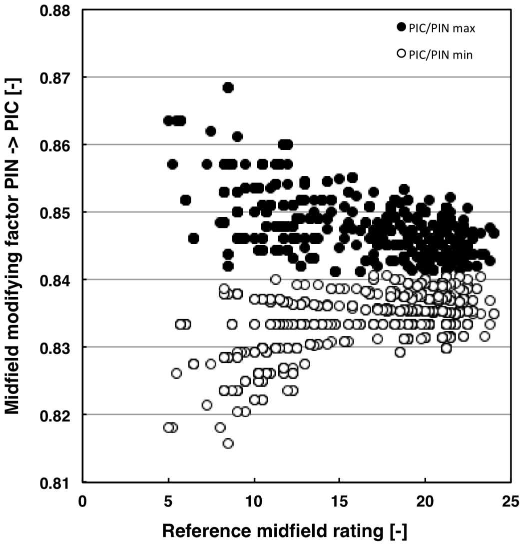
One could leave it at that or one could go and dig into it further. It is indeed well known that an empty line up has an indicated value of 0.75 and that this value switches to 1.00 as soon as a player with no valuable skill yet an experience of non-nil is placed on a position in the line-up simulator that has a contribution to midfield. Let me explain: the goalkeeper position has no contribution to midfield. So even if you put your best midfield player on that spot, the evaluation in the line-up predictor will give you 0.75 for the MF-rating. On the other hand even a player with all skills being "non existent" with a non-nil experience will lead to a midfield rating of 1.00 when placed on a spot that contributes to the midfield, i.e. all spots except the goal keeper spot, thanks to the contribution coming from his experience. This leads to the conclusion that the empty line-up ratings are below 1.00 but very close to 1.00. At last, it has been observed that probability to get a scoring chance based on the rating, i.e. the value that is shown on the match page is based on the midfield ratings of the two clubs but only if one subtracts one unit from the indicated rating (with the additional uncertainty that the true value is somewhat more than the indicated value, cf. the discussion of the interval above.) Therefore, it is conjectured that the rating value that is subject to the modifying factors is actually the difference to the empty line up value, and not the absolute value. Applying this logic to the evaluation of the upper and lower limits of modifying factors leads to the results shown in Fig. 2. Indeed in this evaluation there is a narrow value band between the lowest upper limit and the highest lower limit of that modifying factor. The expected midfield rating for any match setting can hence be calculated by

$$MF\left(t,p,o,TS\right)=1+\left(MF\left(n,a,PIN,TS_0\right)-1\right)f\left(n\rightarrow t\right)f\left(a\rightarrow p\right)f\left(PIN\rightarrow o\right)f\left(TS_0\rightarrow TS\right) \quad (2)$$

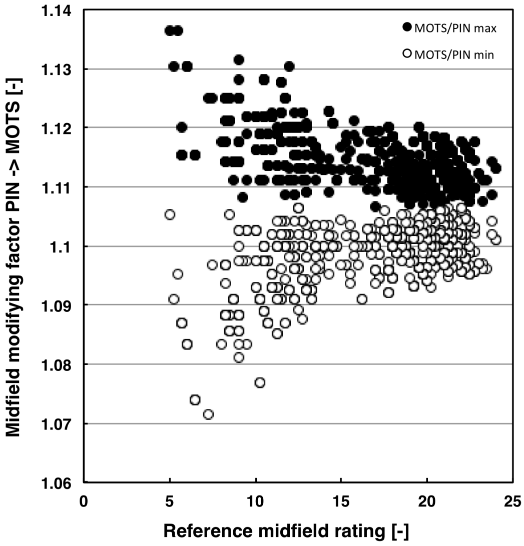
the terms having the same meaning as above. The values of the modifying factors that do no longer create any conflict with observed indications from the line up predictor on the HT-side are given in Tab. 1: it turns out that they are fairly close to what has been given in the unwritten manual and the values given by Schum (17404127.55), but this time with the specific information how to use them.

Table 1: Modifying factors for the conversion of midfield ratings. Values in the diagonal mean no change and have hence the value of 1. Empty cases mean that the combination of lines and rows is not mutually exclusive. The values are to be understood as the modifying factors corresponding to the transition from the setting in the row to that in the column. Example: the MF-rating by changing in a given lineup and for a fixed TS from a (*n,a,PIC*)-setting to a (*LS,h,MOTS*) setting will be MF(*LS,h,MOTS*) = 1 + (MF(*n,a,PIC*)-1) x 0.96 x 1.1 x 1.316.

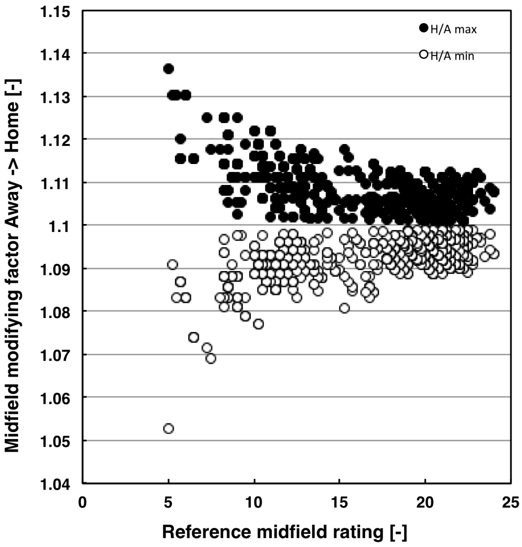
	PIC	PIN	MOTS	Home	Away	normal	LS	CA
PIC	1	1.1892	1.316					
PIN	0.8409	1	1.1082					
MOTS	0.7589	0.9024	1					
Home				1	0.90909			
Away				1.1	1			
normal						1	0.96	0.93
LS						1.04166	1	0.96875
CA						1.07527	1.0323	1



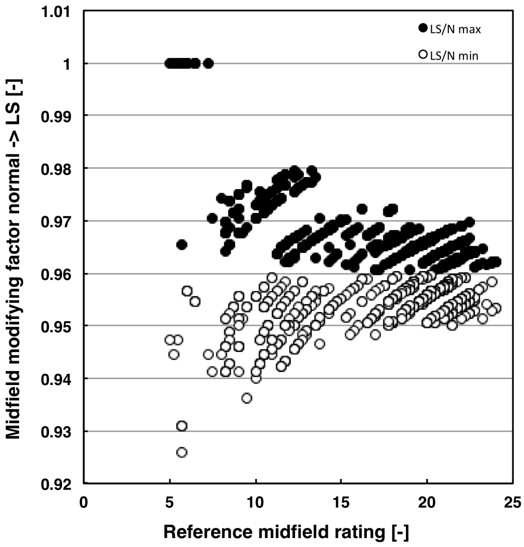
a)



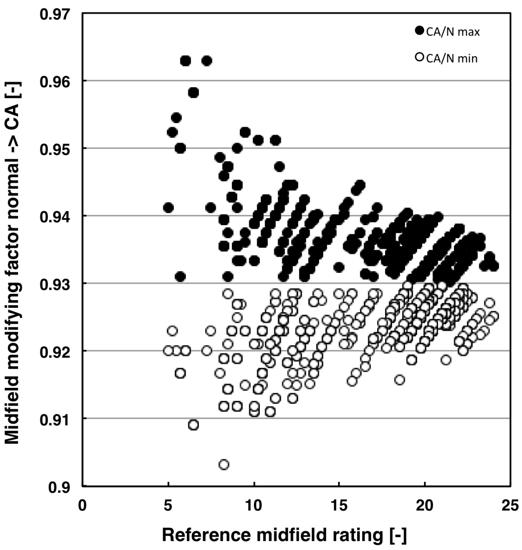
b)



c)



d)



e)

Fig. 2: Modifying factors for a) PIN to PIC, b) PIN to MOTS, c) Away to Home, d) normal to LS and e) normal to counter attack based on the evaluation of the modified ratings. The graphs show for all cases a clear separation between the f_{max} (full circles) and f_{min} (open circles) values. The value at which the separation happens no longer changes as the MF-rating changes. Moreover, in all cases the lowest f_{max} value, which is an upper limit for a supposed unique f, is higher than the highest value of f_{min}, which is a lower limit for the supposed unique f. The separation is typically smaller than ±0.001 around the average value given in Table 1.

Determination of the precise MF-rating

Once the question of the modifying factors is settled we can tackle the question of the precise MF-rating. We first note that, for a given lineup, there are 18 distinct MF-ratings, coming from the 18 possible combinations of (PIC,PIN,MOTS)x(h,a)x(n,LS,CA). The procedure is the following: since the MF-ratings for all possible settings can be calculated from the MF-rating for a given setting using Equation (2) and the factors collected in Table 1, we can find the *lowest possible MF-rating* for that reference setting that is coherent with all other 17 distinct indicated ratings in the lineup predictor. This gives us a *lower limit* of the MF-rating for the reference setting. We can along the same lines also find the *highest possible MF-rating* that leads to results consistent with the 17 other indicated MF-ratings. This leads then to the *upper limit* of MF-rating for the reference setting. The range of possible MF-ratings found by this procedure is typically within ±0.01 around their average. This has to be compared to the face value that is without applying the procedure described above within ±0.125 of the mid-point of the indicated MF-rating, e.g. for an indicated MF-rating of 16.75, the true MF-rating would be known only to 16.875±0.125. The procedure outlined above hence improves the precision of the true MF-rating by more than a factor of 10.

Extracting the current TS-level from precise MF-ratings

Besides the true MF-rating for the reference setting at the actual TS-level, we can also determine the true MF-rating for the reference setting for any TS-level by selecting that TS-level in the lineup simulator. From this point on, there is a multitude of ways extracting the actual TS-level from the information collected about the MF-rating for the reference stetting for the actual TS-level and the MF-rating of the reference setting at various fixed TS-levels. A few of these are discussed in what follows.

TS vs. MF-rating Master-curve

At the time when the line-up simulator was in its beta testing phase on stage, HT-Bodin had first programmed it to reflect selectable TS-levels as x.0. After some comments of beta-testers he changed that because people were used to having the values at x.5. Luckily enough, the community had access to both sets of values, which allows drawing a full master curve of MF rating vs. team spirit. In order to be a master-curve it needs to be normalized to be applicable to all sorts of line-ups. We choose to normalize it to its value at the highest available fixed TS-level, i.e. 9.5 corresponding to the fixed value "WoC" (in the test phase when the data were collected PoE was kept at 10.0 while now it is a 10.5). This normalized function has the following form:

$$MF\left(TS_{\text{act}}\right)/MF\left(TS=9.5\right)=A+B\cdot TS+C\cdot TS^2+D\cdot TS^3+E\cdot TS^4+F\cdot TS^5+G\cdot TS^6 \quad (3)$$

with the coefficients given in Table 2. To find the current level of team spirit, it suffices to calculate for which value of TS the equation gives the ratio of MF-ratings as determined by the procedure in the previous subsections.

Table 2: Coefficients of the MF vs. TS master-curve normalized to its value at TS = 9.5 (WoC). The notation "E-x" means that the value is to be multiplied by 10^{-x}.

Coefficient	A	B	C	D	E	F	G
Value	0.29533	0.12974	-6.9158E-3	-6.3908E-4	1.4072E-4	-6.433E-6	-3.7109E-9

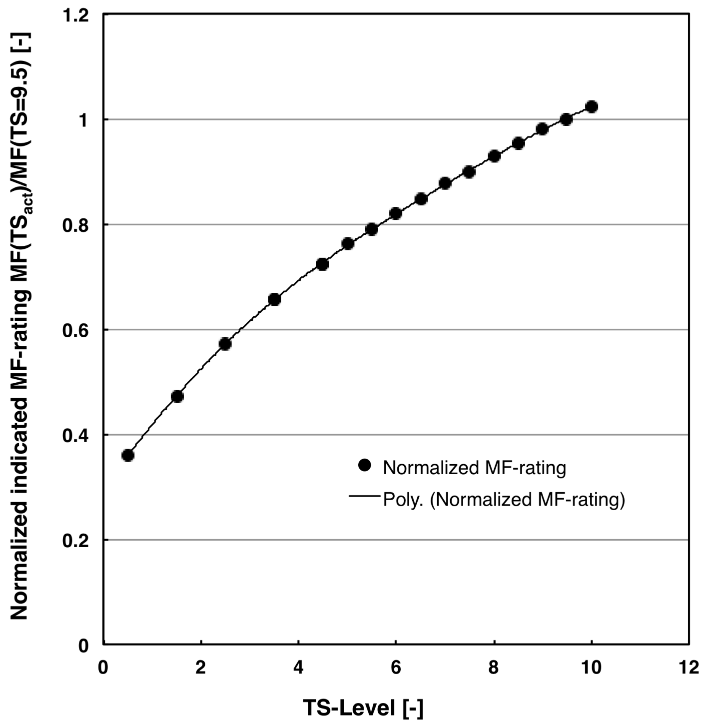


Fig. 3: Master curve of MF-rating vs. TS-level. The MF-ratings are normalized to the MF-ratings of the same setting at a team spirit of 9.5, i.e. "walking on clouds".

The graphical representation of that function is given in Fig. 3. It needs to be highlighted that the normalized MF-rating vs. TS is relatively flat in the high TS-range. Furthermore, the measurement of the MF-ratings needed to enter into Equation (3) has some uncertainty (on the order of 0.1%) and the descriptive capacity of the polynomial expression is on the same order of relative error. These two factors lead to a typical uncertainty of the evaluated TS using the master-curve approach that is on the order of ±0.05 to ±0.1 TS level. A further difficulty is that the polynomial expression to approximate the normalized MF-rating cannot be inverted and needs to be solved numerically. This can be done e.g. by using the goal seek function in Excel. All in all, the master curve approach is ok for people who would like to have a rough idea where they stand within a visible level of TS but is not sufficient to gain deeper insight in how the team spirit decays over time.

Extracting TS-level by 2-point or 3-point MF-rating interpolation

A second method of extracting TS-level information from refined measurement of the MF-rating is by 2- or 3-point interpolation. The 2-point interpolation is based on the idea that if we know the refined MF-rating for two precise TS-levels on either side of the actual TS-level, TS_1 and TS_2, we can linearly interpolate the actual TS-level by knowing the refined MF-rating for the same line-up and setting. The actual TS-level, TS_{act}, is then

$$TS_{act} = TS_1 + \frac{MF(TS_{act}) - MF(TS_1)}{MF(TS_2) - MF(TS_1)}(TS_2 - TS_1) \tag{4}$$

where $MF(TS_{act})$, $MF(TS_1)$, and $MF(TS_2)$, are the averages of the upper and lower limits determined according to the procedure described above. The advantage of this procedure compared to the master curve approach is that there is no need for solving a higher-order polynomial numerically. The disadvantage of the procedure is that we need to determine the refined MF-rating of the reference state three times, once for the current TS-level and once for each TS_1 and TS_2.

The MF-rating vs. TS curve is also only approximately linear and we may introduce uncertainty in the TS-determination by making the linearity assumption. A somewhat more adequate procedure is to make a 3-point interpolation, i.e. to measure the refined MF-rating for the visible level of TS and the ones above and below that visible level. To be explicit: if on the team page the indicated TS-level is "delirious", i.e. in the interval [8,9), one has to determine the refined MF-rating at "satisfied", i.e. 7.5, at "delirious", i.e. 8.5, and at "walking on clouds", i.e. 9.5, as well as for the "current" TS-level. The averages of the upper and lower limits of the refined MF-rating, cf. de procedure above, for TS 7.5, 8.5, and 9.5 are then the points of the "local master" curve that are to be modeled with a polynomial of second order, i.e. of the form:

$$MF\left(TS\right) = a + b \cdot TS + c \cdot TS^2$$

The factors *a*, *b*, and *c* are derived from the averages of the upper and lower limit of the refined MF-ratings at TS_1, TS_2, and TS_3, where TS_1 is that one below the visible level on the team page, TS_2 is the TS-level shown on the team page, and TS_3 is the TS-level above the visible TS-level on the team page. Explicitly, it holds:

$$c = \frac{MF\left(TS_3\right) + MF\left(TS_1\right)}{2} - MF\left(TS_2\right)$$

$$b = \frac{MF\left(TS_3\right)\left(1 - 2TS_2\right) - MF\left(TS_1\right)\left(1 + 2TS_2\right)}{2} + 2TS_2 MF\left(TS_2\right)$$

$$a = MF\left(TS_2\right) - b \cdot TS_2 - c \cdot TS_2^2$$

The level of current TS, TS_{act}, can then be found as

$$TS_{act} = \frac{-b + \sqrt{b^2 - 4c\left(a - MF\left(TS_{act}\right)\right)}}{2c} \tag{5}$$

This procedure has the advantage that it eliminates the linearity assumption of the 2-point interpolation and reflects the curved nature of the MF-rating vs. TS curve. It further does this in an explicit way, i.e. the calculation of the current TS-level can be automated, once the refined MF-ratings for TS_1, TS_2, and TS_3, as well as TS_{act} have been obtained.

In order to evaluate the maximum uncertainty in TS-level obtained by the interpolation procedure, we may, instead of the average values of the refined MF-rating (remember, there was an upper and a lower limit obtained by the procedure), take the *upper limit* values of the refined MF-rating at TS_1, TS_2, and TS_3, designated $MF^+(TS_1)$, $MF^+(TS_2)$, and $MF^+(TS_3)$ to establish the factors a^+, b^+, and c^+ of Equation (4) for the polynomial of the upper limit points and evaluate the current TS-level using the *lower limit* of the refined MF-rating at the actual TS, $MF^-(TS_{act})$. Inserting these values in Equation (5) will give us the lower limit of the current

TS-level, TS_{act}^{-}. Conversely, if we take the lower limits of the refined MF-ratings at TS_1, TS_2, and TS_3, designated $MF^{-}(TS_1)$, $MF^{-}(TS_2)$, and $MF^{-}(TS_3)$ to establish the factors a^{-}, b^{-}, and c^{-} of Equation (4) for the polynomial of the lower limit points and evaluate the current TS-level using the *upper limit* of the refined MF-rating at the actual TS, $MF^{+}(TS_{act})$, we will, after inserting in Equation (5), find the upper limit of the current team spirit, TS_{act}^{+}. This is illustrated by the schematic drawing in Fig. 4.

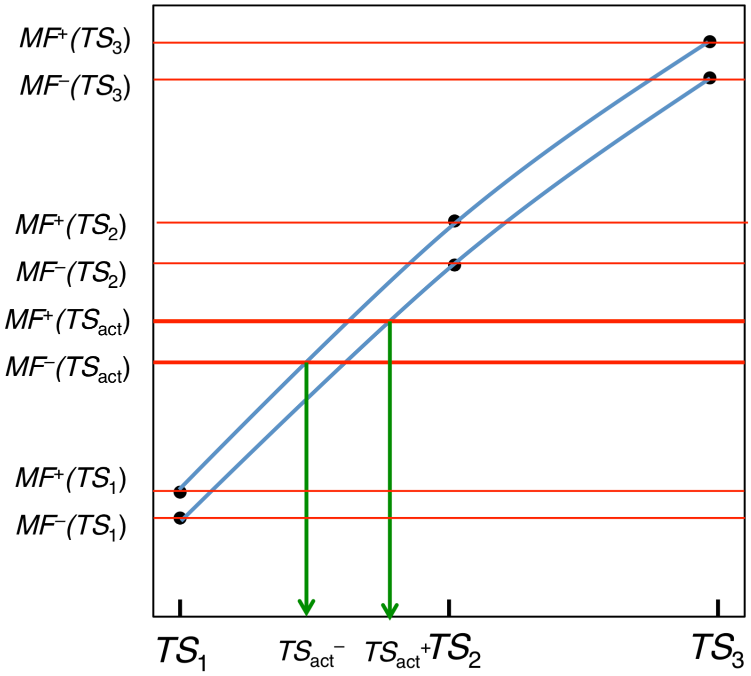


Fig. 4: Schematic drawing of the determination of the upper and lower limits of the current TS-level based on the crossed evaluation of the 3-point interpolation scheme. "Crossed" in this context means that the lower of the current TS-level is based on the *lower limit* of the refined actual MF-rating and the second order polynomial of the *upper limits* of the refined MF-ratings at the three fixed TS-levels and vice versa.

As an example let us consider a typical set of refined MF-ratings as obtained from the Swiss Hattrick NT[1] given in Table 3. The evaluation is done for the 3-point interpolation. For comparison, the same type of evaluation based on the 2-point interpolation would yield a current TS-level of 9.312 and the master curve approach would have led to 9.23. The current analysis shows that the current TS-level can be evaluated to typically well within ±0.02 by the 3-point interpolation scheme.

Table 3: Example of a data set of refined MF-rating at various fixed TS-levels and the actual refined MF-rating for a reference setting.

	$MF(8.5)$	$MF(9.5)$	$MF(10.5)$	$MF(TS_{act})$	a	b	c	TS_{act}
Upper limit	20.12733	21.23656	22.27273	21	7.7491	1.7668	-0.03653	9.302
Lower limit	20.12311	21.21247	22.26149	20.99099	9.2348	1.4524	-0.02017	9.273
Average	20.12522	21.22452	22.26711	20.9955	8.4915	1.6097	-0.02836	9.287

[1] Before you get impressed by the fairly high MF-ratings being achieved, be reassured that this is only possible in a 5-5-3 lineup with all players acting towards the midfield. Such a formation cannot be played but still can be used for evaluation of the TS-level. The reason to go for one of those high MF-ratings is that the higher the MF-rating, the smaller the relative uncertainty of the refined MF-rating is (its absolute uncertainty being typically on the order of ±0.02).

Last but not least we can further refine the 3-point interpolation by doing the same analysis for several lineups: when we change the lineup, the indicated MF-ratings will change for the various settings and so will the upper and lower limits for the MF-ratings at the actual and the fixed team spirit levels. This will give us several values for upper limits of the team spirit and several levels of lower limits, "several" standing for the number of lineups we are evaluating. The real level of actual team spirit will then be between in the range between the lowest value of the upper limit and the highest value of the lower limit. This procedure typically reduces the uncertainty of the actual team spirit value further. For three line-ups the range is typically reduced to the average ±0.01 or better.

Determination of the precise attack-rating

After this lengthy discussion concerning MF-rating and the team spirit, the same sort of consideration can be applied for the evaluation of the self-confidence. Contrarily to the team spirit, where we were varying the tactics, the place of match, and the orientation to obtain a refined information about the MF-rating for the reference setting, of all these factors none but the LS-tactic affects the attack ratings. Additionally, the aggressiveness slider (-100 to +100) can be used to produce a variation in attack ratings. As working hypothesis we apply again that the various corrections, i.e. those due to aggressiveness level and when choosing the long shot tactics as well as those based on self-confidence, are unique and mutually independent. That is we can write the attack rating, A_i, of attack side $i=l,m,r$ for left, middle and right, with regard to a reference state in terms of aggressiveness, self-confidence, and tactics as

$$A_i(t,agg,SC)=1+\left(A_i(n,0,SC_0)-1\right)f(n\to t)f(0\to agg)f(SC_0\to SC) \qquad (6)$$

with $f(n\to t), f(0\to agg), f(SC_0\to SC)$ being the factor for tactics other than normal tactics, the factor taking into account the aggressiveness, and the factor taking into account self-confidence, respectively. We note in passing that, similarly to what was found in the context of the precise MF-rating, the modifying factors are found to act on the player skills based attack rating, i.e. the indicated value minus 1. The factor $f(n\to t)$ is 1 for all but the long shot tactics for which it is actually $f(n\to LS)=0.96$. The factor for aggressiveness is different for positive and for negative aggressiveness levels:

$$f(n\to t)=(1+0.001175agg) \quad \text{for } agg<0$$
$$f(n\to t)=(1+0.000785agg) \quad \text{for } agg>0$$

For the factor due to self-confidence, it becomes numerically most simple when SC_0 = 4.5 (i.e. "decent"):

$$f(4.5\to SC)=\left(1+0.05(SC-4.5)\right)$$

Similarly to the development for the precise MF-rating determination, we can again look for the basic attack rating for a given level of SC, that for all 42 settings (21 levels of aggressiveness, two different tactics) leads to no conflict between the predictions by the line up simulator and the calculated values. As before this gives us an upper limit for the basic attack rating and a lower limit for the basic attack rating for a given side *i* of attack. It turns out that such procedure leads typically to basic attack ratings that are bounded to within 0.005 around the average value of the two limits. This again is to be compared with an uncertainty of ±0.125 around the middle level of the range indicated by the predictor.

It is clear that the determination of the current level of SC will be the more precise, the larger the value of the attack rating is. Therefore, if someone wants to track its SC level very precisely e.g. from one update to the next, or the change brought to SC by a given result in a competitive game, he/she is better off using a lineup with high attack ratings. For league teams this is most easily done by creating a lineup with eleven custom created players with all having divine level in scoring, passing, and wing and to place them in a 3-5-3 lineup with a lateral central defender towards wing, and the offensive IMs. The left and right IM can also be oriented towards the wing to increase the attack rating. Such a line up can obviously not be saved but will continue to work in the line up simulator as long as the window is not closed, and it will continue to use the updated SC levels of the team throughout. For such a lineup it is not unusual to achieve attack ratings in the 30 to 40 ballpark (divine +20!) or even 50 if one orients a forward to wing. This gives highest precision.

In the case of national teams it may be easier to keep either a saved, very offensive lineup as a permanent reference (with the need to keep track of changes due to training and changes in form or loss of skills), or to keep a lineup simulator window open. The latter has the advantage that also a lineup with more than 10 players on the field can be used for "measurement" purposes.

The extraction of the current self-confidence value from the values of the basic attack ratings, that is the attack ratings for the setting normal tactics and 0 aggressiveness is a priori somewhat simpler than for the team spirit determination, because the influence of self-confidence enters the attack ratings linearly. Due to the bounding procedure of the values for both the current and the fixed self-confidence values that can be set in the line-up simulator, the precision will typically be better for interpolation schemes than for direct calculation from the comparison of the attack ratings at the current SC level and the reference SC-level. Due to the simple linear influence of self-confidence, the most convenient way is to determine the upper and lower limits of the attack ratings at the current self-confidence level and compare them to the upper and lower limits of the corresponding attack ratings at the midpoint of the indicated self-confidence level.

As an example we consider the following case: Be $A^{+}_{i,\mathrm{curr}}$ and $A^{-}_{i,\mathrm{curr}}$ the upper and lower limit, respectively, of the basic attack rating in sector *i* at the current SC-level and $A^{+}_{i,\mathrm{ind}}$ and $A^{-}_{i,\mathrm{ind}}$ the upper and lower limit, respectively, of the basic attack rating in sector *i* at the SC-level indicated on the team page. The upper and lower limit of the current SC-level, SC^{+}_{curr} and SC^{-}_{curr}, respectively, can then be determined by

$$\begin{aligned} SC^{+}_{\mathrm{curr}} &= SC_{\mathrm{ind}} + \frac{A^{+}_{i,\mathrm{curr}} - A^{-}_{i,\mathrm{ind}}}{\alpha_{\mathrm{ind}}\left(A^{-}_{i,\mathrm{ind}} - 1\right)} \\ SC^{-}_{\mathrm{curr}} &= SC_{\mathrm{ind}} + \frac{A^{-}_{i,\mathrm{curr}} - A^{+}_{i,\mathrm{ind}}}{\alpha_{\mathrm{ind}}\left(A^{+}_{i,\mathrm{ind}} - 1\right)} \end{aligned} \qquad (7)$$

where α_{ind} is the proportionality factor associated with the self-confidence value indicated on the team page. α_{ind} is calculated as

$$\alpha_{\text{ind}} = \frac{0.05}{\left(1+0.05\left(SC_{\text{ind}}-4.5\right)\right)}$$

For your convenience the thus calculated factors are given in Table 4 for the various values of SC_{ind}.

Table 4: Collection of α_{ind}-values for the calculation of the current self-confidence level using Eq. (7).

SC_{ind}	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5
α_{ind}	0.0625	0.0588	0.0556	0.0526	0.05	0.0476	0.0455	0.0435	0.0417	0.04

Since every line up gives us three attack ratings, i.e. one for left, middle, and right attack, we will get three values for SC^{+}_{curr} and SC^{-}_{curr}, respectively. The real value of the current self-confidence level will then be between the lowest of the three SC^{+}_{curr}-values and the highest of the SC^{-}_{curr}-values. This leaves typically an uncertainty on the order of ±0.01.

One can of course choose any other means of interpolation as discussed in the context of the calculation of the TS values, to reach, mutatis mutandis, the current SC level.

Results

The result section will address the questions raised in the intro in the same order as given above. Questions 1 and 2 have been addressed in the previous section and are not further discussed here.

How does the SC evolve with time in absence of competitive games?

This question could of course be perfectly well addressed by doing regular measurements from one update to the next and thus produce a wealth of points. If collected by a large number of NT-managers one would probably quite rapidly get access to a fairly precise picture of the current mechanics. It turned out, though, that the mechanics of decay (from high values) and of build-up (from low values) as it had been measured at times when there were long periods of no competitive matches for national teams, i.e. in the formats of previous qualifiers and world cups, has been carried over to the new situation with the sole difference that the rest value has been adapted from 4.5 to 5.0 in August 2021. In essence, the change in team spirit operated by time update only can be written as (factors *a* to *h* below in Tab. 5):

$$\begin{aligned}\Delta SC &= a\left(SC_{\text{curr}}-5\right)^4 + b\left(SC_{\text{curr}}-5\right)^3 + c\left(SC_{\text{curr}}-5\right)^2 + d\left(SC_{\text{curr}}-5\right) \quad \text{for } SC > 5\\ \Delta SC &= e\left(SC_{\text{curr}}-5\right)^4 + f\left(SC_{\text{curr}}-5\right)^3 + g\left(SC_{\text{curr}}-5\right)^2 + h\left(SC_{\text{curr}}-5\right) \quad \text{for } SC < 5\end{aligned} \qquad (8)$$

Table 5: coefficients of the polynomials for change in self-confidence per update, ΔSC.

a	b	c	d	e	f	g	h
0.0001251	-0.0015432	0.007	-0.0426	0.002082	-0.0042388	-0.014876	-0.050678

How does the TS evolve with time in absence of competitive games?

For the evolution of the TS with number of updates, the data from the msn tables have been for a long time the point of reference. Thanks to the tools outlined above, we have now much cleaner access to the team spirit and can refine or even correct those tables where necessary. To this end the TS-decay per update has been tracked for the Swiss NT on a daily basis since spring 2021 where the rating simulator had first been introduced on stage and later to the larger public. By the nature of things, i.e. due to the fact that NTs try always to be in the high TS region, most data are in the region of TS between 8 and 12. In Fig. 5 the data are collected as drop in team spirit per update, dTS, vs. the difference to the rest value, ΔTS.

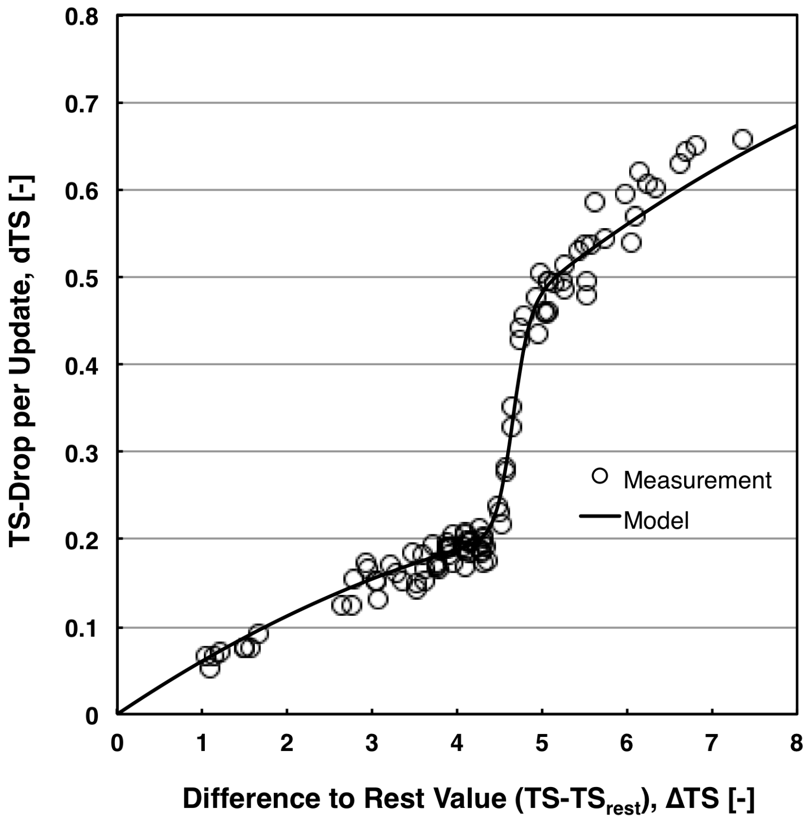


Fig. 5: Experimental measurements of TS-drops for a bit more than 100 updates for the Swiss NT. The scatter about the trend line is on the order of 0.02, which is comparable to the uncertainty of the measurement. The trend line designated by "Model" is obtained by plotting Equation 8.

The Data set shown in Fig. 5 has the unusual feature that the drop per update increases quite significantly for a difference to the rest value of 4.5 – 5, i.e. for TS-levels between 9.5 and 10 and then continues its evolution quite steadily. The data can be described by

$$\left(TS_{\text{old}} - TS_{\text{new}}\right) = dTS = -0.0046\underbrace{\left(TS_{\text{old}} - 5\right)}_{\Delta TS}^{2} + \Delta TS\left(0.0932 + 0.0278\tanh\left(\frac{\Delta TS - 4.65}{0.2}\right)\right) \quad (9)$$

Equation (9) has been applied to predict the TS evolution of the Swiss NT for more than 100 updates, i.e. essentially the whole campaign of WC 33. The prediction of the TS for the Swiss NT is compared to the measured evolution in Fig. 6. The spikes upwards are the results of the PIC choice for matches. Some manual correction had to be done because there were several players fired during the campaign. The manual correction was then by simply getting the value back on track at that point. Such intervention had to be done at four occasions. The downward spikes around update 75 and update 110 where the 80 and 60% TS-resets. Throughout the whole campaign the deviation was no more the ±0.05 in team spirit level.

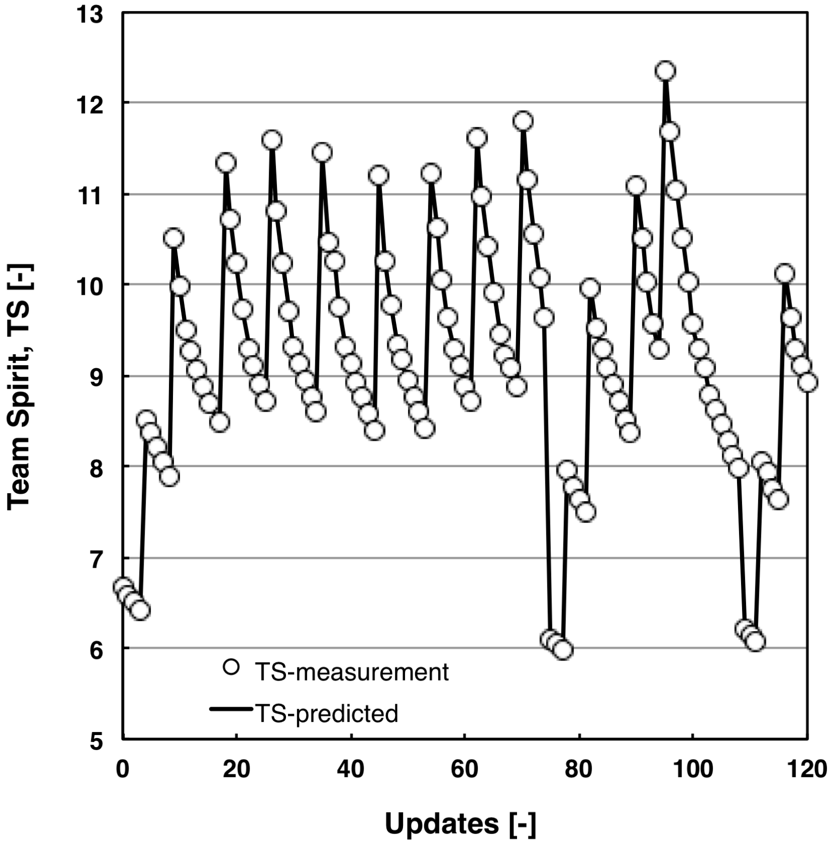


Fig. 6: Experimental measurements of TS-drops for a bit more than 100 updates for the Swiss NT. The scatter about the trend line is on the order of 0.02, which is comparable to the uncertainty of the measurement. The trend line designated by "Model" is obtained by plotting Equation 8.

It might be argued that the experimental data that are modelled here have also been the basis of the model and hence they ought not lead to discrepancies with the model. It needs however be borne in mind that i) the data being modelled are only the observed variations from one update to the other and not the full temporal evolution in which errors tend to accumulate, and ii) there is no reason why the TS evolution of another NT should differ from that observed in Switzerland (except of course due to choices concerning the PIC/PIN/MOTS strategy and the firing of players). Therefore, the practice to evaluate the quality of the prediction on the data set that allowed deducing the description is considered less critical.

For the other branch of the TS-curve, i.e. the recovery from very low values again the TS of the Swiss NT has been measured after firing of all players once the NT has been eliminated from the world cup. Again, the recovery the data are well described by a 4[th] order polynomial:

$$TS_{\text{new}} - TS_{\text{old}} = dTS = 0.0038\underbrace{\left(TS_{\text{rest}} - TS_{\text{old}}\right)}_{\Delta TS}^4 - 0.0246\Delta TS^3 + 0.0628\Delta TS^2 + 0.0202\Delta TS \quad (10)$$

Note that the definition of dTS and ΔTS is in this case different that what was used in Eq. (9).

How does a certain score of a competitive game affect the self-confidence?

The statements of the HTs, in particular by HT-Tasos, indicate that the effect of a result in a competition game is to multiply the self-confidence by a specific factor that depends solely on the result. With the general rule that a draw leads to a factor of unity, a win to a factor >1 and a loss to a factor <1. The problem reduces hence to find the factors corresponding to the various possible results. With the tools to measure the self-confidence presented above this question could, in principle, be tackled by measuring the changes in SC-levels for any result of a competitive match. Since some results are more frequent than others, it would require the combined effort of a number of managers to get those values for all possible match results. This would further also only work, if the presence of a psychologist and its level had no effect on these factors. Indeed, preliminary measurements would suggest that the psychologist has only an effect on the rate of decay and the rest value but not on the factors themselves. In the exploratory stage of this work, these factors have hence been determined by following closely the evolution of SC of the 128 national teams involved in the qualification and main stages of Hattrick WC 32. The result is summarized in Table 5 where the factors for all possible results are listed. By the nature of things not all possible results have appeared in that campaign and some values have simply be inferred by interpolation. Those are designated by grey shaded background in Table 6.

Table 5: Factors to be applied to the self-confidence based on the outcome of a competition match.

Goals scored

Goals conceded	0	1	2	3	4	5	6	7	8	9	10
0	1	1.096	1.142	1.191	1.23	1.272	1.315	1.345	1.37	1.39	1.41
1	0.905	1	1.067	1.109	1.165	1.2	1.233	1.27	1.31	1.33	1.36
2	0.836	0.935	1	1.058	1.102	1.145	1.18	1.2	1.235	1.27	1.3
3	0.76	0.875	0.941	1	1.052	1.097	1.12	1.16	1.18	1.21	1.24
4	0.7	0.82	0.887	0.946	1	1.047	1.07	1.12	1.135	1.165	1.18
5	0.634	0.76	0.83	0.9	0.955	1	1.042	1.08	1.1	1.12	1.14
6	0.58	0.7	0.78	0.85	0.91	0.957	1	1.04	1.067	1.085	1.1
7	0.52	0.64	0.73	0.8	0.865	0.914	0.96	1	1.038	1.055	1.075
8	0.46	0.59	0.68	0.75	0.82	0.871	0.92	0.963	1	1.032	1.05
9	0.4	0.54	0.63	0.7	0.775	0.828	0.88	0.926	0.965	1	1.025
10	0.34	0.5	0.58	0.65	0.725	0.78	0.84	0.889	0.93	0.97	1

It needs to be stated that the self-confidence has, at least for the club teams a hard cap at 9.5, i.e. completely exaggerated. Whether this is also the case for national teams is not sure. What can be said however, is that none of the 128 teams involved in the WC 32 ever reached a self-confidence of >9. The closest that was obtained was a level of ≈8.8 by Sweden. The factors of this table together with the equations (8) for decay and build up of self-confidence can now be used to calculate the self-confidence of an arbitrary team during a past, present of future WC campaign. This is done for the example of the self-confidence of Singapore during the WC in

Venezuela. The example of Singapore is chosen because they reached amongst the highest Self-confidence levels in that competition. The upward jumps in self-confidence at the beginning were due to two 4-0 wins. The further events are indicated in the graph by arrows, including a 60% reset.

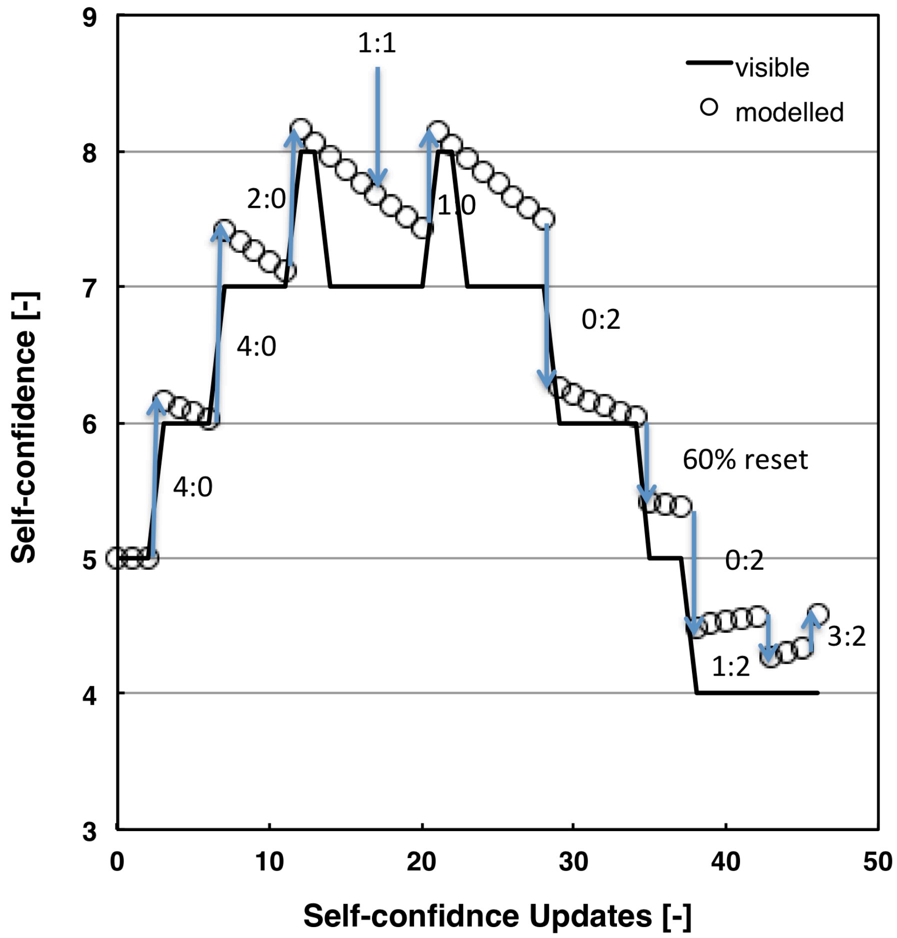


Fig. 7: Evolution of the visible and predicted self-confidence of Singapore's NT between September 29, 2021 and November 5, 2021.

Similar analyses can be done for all other teams involved in NT-competition. Typically, the predictions do not deviate by more than 0.05 SC-levels from the measureable level.

How does the firing of a player affect team spirit?

The answer to this question is in principle well known: depending on the character of the player there are known intervals of loss in team spirit that are associated with the firing of a player in NTs. The question actually was whether a given player would always lead to the same reduction in TS (within that range associated with his character), which could then be interpreted as the sub-skill of the visible character that affects where in the know range e.g. for pleasant guys the drop would lie. This idea has been put forward in various discussions but there has never really been the means to nail this. With the present tools to measure the actual team spirit quite precisely, one can indeed check the hypothesis by firing and reengaging the same player several times. Grabner, at the time NT manager of Letzebuerg was so kind to agree to destroy the team spirit from hi level down to nothing by doing exactly this for various players with various levels of visible character. The outcome was that the firing of a given player would lead to a very specific percentage of reduction in team spirit: for one player with pleasant character the drop was 9.1, 8.8, 9.0 per cent while for another player with pleasant character the drop was 10.2, 10.0, and 10.4 per cent. The variations in value are on the order of the uncertainty of the measurement. Hence, the present investigation supports the hypothesis that the drop in team spirit encountered upon firing of a player is indeed linked to a not indicated sub-skill of his character, or, to the very least, the present investigation gives no evidence for that the drop in team spirit caused by the firing of a player with a given visible character is a random value chosen ad hoc at the moment of firing within the range associated with the visible character.

Outlook

The unveiling of the mechanism of build-up and decay of team spirit and self-confidence and the tools that can be constructed based on this information allow to get much more precise information about how team settings, e.g. specialists, leadership level of the coach and other would change (or confirm) the mechanics as observed with the national teams. This would be achieved much more efficiently if many different managers with a range of different settings would do the analysis in parallel and then share the results.

January 2022, Luiteger Manager of Kickers Le Mont