

Insights from signal within the noise of football tracking data

Christian Gilson

Mathematical Modelling of Football

BEYOND EVENTS

Events data provides a blow-by-blow of actions throughout a football match, often enabling the calculation of more robust measures of performance than traditional statistics like goals and assists, whilst also provoking questions through data stories that materialise from the events analysis. Often those questions can't be answered using events data alone, as off-the-ball context is essential to building a more complete picture of the game.

Tracking data serves as part of the toolkit to help provide that fuller picture. As an initial discovery analysis, this report will focus on player position, speed, and acceleration (as well as metrics derived from those quantities) to highlight the insights and limitations of tracking data for a Hammarby data analyst.

20 years ago in the Premier League, the likes of Teddy Sheringham and Dennis Bergkamp weren't the quickest but were often lauded as doing the first five yards in their heads. In today's game, player athleticism is deemed essential. Usain Bolt nearly switched his spikes for a pair of boots in a trial with Australian club Central Coast Mariners in 2018 – *scoring two goals against professional opposition!* – highlighting the perceived importance of applying speed in the right areas of the pitch.

HAMMARBY

We start by looking at a handful of key moments from three matches that we have tracking data for: Hammarby home games against IF Elfsborg, Malmö FF, and Örebro. The tracking data provides 25 frames of player x, y position every second, allowing us to calculate the rate of change in position per $\frac{1}{25}$ second time step to produce player x, y velocities (the magnitude of which is the player speed).

FIGs 1, 2 and 3 show the moments just before the final ball was delivered to score Hammarby's first goals against the three teams, with arrow direction and length representing player velocity direction and magnitude. In both scenarios against Elfsborg and Malmö:

- Djurdjic (#40) crossed from the left of the penalty area;
- Kacaniklic (#20) sprinted towards the penalty spot;
- Khalili (#7) ran towards the six-yard box.

Against Elfsborg, Djurdjic squeezed the ball past two defenders and the goalkeeper to provide unmarked Khalili with a tap in. Defensive dynamics were different against Malmö, with defender Safari (#4) matching Khalili stride for stride to block the passing lane across the face of goal. Djurdjic cut the ball back towards the penalty spot before Kacaniklic had even entered the area, where Kacaniklic's velocity arrow in FIG. 2 indicates his likely arrival at the penalty spot one second later, just in time to sweep the ball past the keeper without breaking stride. Great split-second decision making from Djurdjic.

At 1-0 down against Örebro, Hammarby switched the play to the right wing after struggling to penetrate down the left flank. Kacaniklic again ran towards the penalty spot, but it was Rodric' (#11) superior pace against 35 year old Lorentzon (#15) that made the difference, heading home at the near post before going on to score a hat-trick.

Whilst Hammarby demonstrate impressive wide play in attack, FIG. 4 highlights an instance of failing to defend a late run into the

area from a deep cross. Rather than heading towards a well-defended penalty spot, ex-Macclesfield Town midfielder Kasim (#16) sprinted towards the open space at the back post, getting the run on Hammarby's Widgren (#3) and easily beating him in the air to head the ball back across goal for Prodel (#9) to finish acrobatically.

LIMITATIONS OF THE DATA

Additional kinematic quantities beyond player velocity are straightforward to derive. Similar to the calculation of x, y velocity components from the rate of change in x, y position, x, y acceleration components are derived from the rate of change in x, y velocities (as acceleration is the second time derivative of position). A player's overall acceleration is then calculated as the magnitude of the x, y components.

Knowing which half of the pitch a team's goalkeeper is positioned (on average) enables a data-driven inference of shooting direction from the tracking data, which in turn enables the distance to goal to be calculated for every player on the pitch as the opponent's goal position is easily inferred given the team's shooting direction.

Visualising these kinematic quantities provides an intuitive way of spotting systematic issues with the tracking data quality. Anytime entities are involved in quantitative analysis that need to be linked through a sequence of events – whether those entities are company stocks interacting through mergers and acquisitions or football players interacting during a match – entity mapping issues will introduce spurious relationships within the data that can be tricky to unpick.

A striking example of this within the tracking data can be seen in FIG. 5, where Khalili and Fällman erroneously flip positions on the pitch and then flip back again a few seconds later after the tracking technology correctly resolves the player jersey numbers again. Outlier signatures of these kinds of mapping errors can be seen in the speed and acceleration plots: a speed of 600 ms^{-1} is 50 times faster than Usain Bolt's top speed of 12 ms^{-1} . These by-eye sanity checks provide a practical way of ruling out the impossible before weight is added to the story the data is telling.

Once these errors have been systematically highlighted, they can either be fixed or masked. Rather than going through each mapping issue and manually resolving player identities, this analysis masks tracking frames where a player's derived speed is above 12 ms^{-1} . Set pieces were the most common systematic cause of these errors where the bunching of players inevitably causes jersey numbers to be obfuscated.

Another data limitation is related to the tracking of the ball rather than players. The Signality cameras clearly struggle to track the ball once it leaves the ground. Whilst the dataset provides a z component of ball position in the data model, in reality the sum of every z ball position from the three matches is zero. It's as though the ball is glued to the pitch. This causes particular problems when tracking the path of a high, curving ball and relating that path to player movement.

Ball tracking issues are unfortunately not limited to whilst it's in the air as often the ball's position can vanish for handfuls of frames at a time when being played along the ground. To fill in those gaps, this analysis interpolates ball position linearly to help with the interpretability of play (with the clear limitation for curved passes).

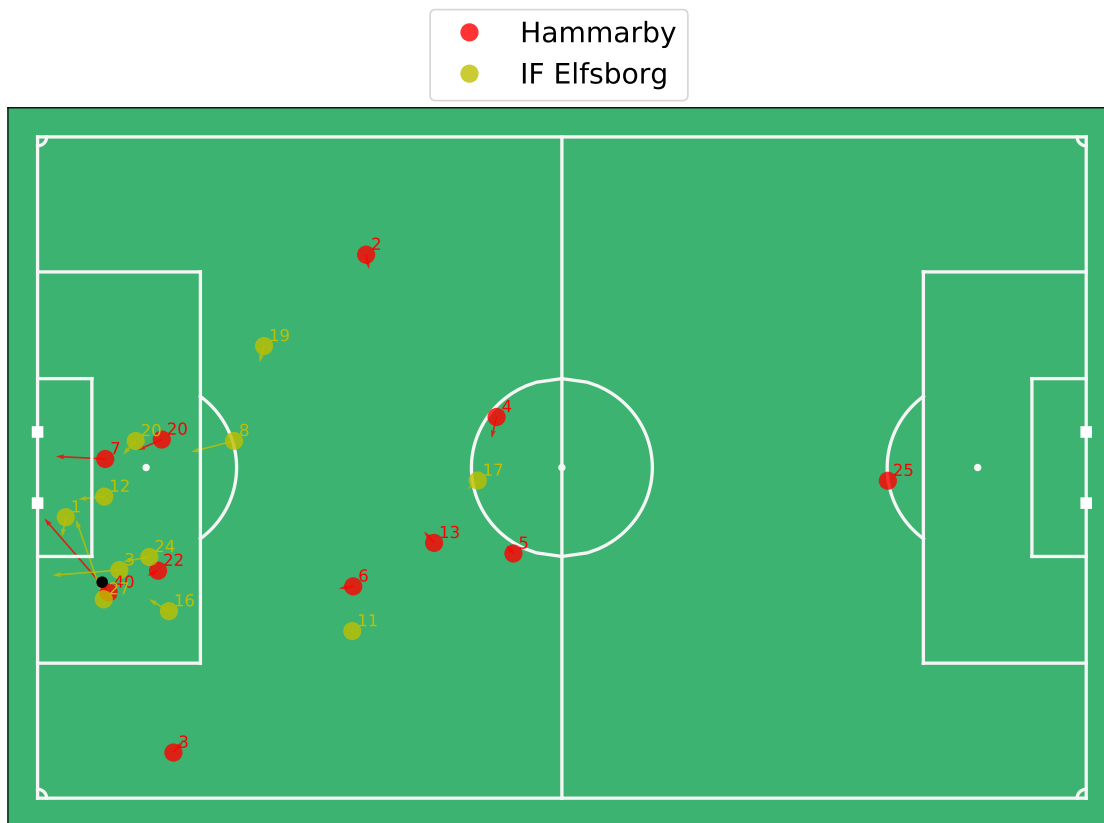


FIG. 1: Opening goal for Hammarby against Elfsborg with Khalili (#7) finishing Djurdjic's (#40) low ball across the face of goal.

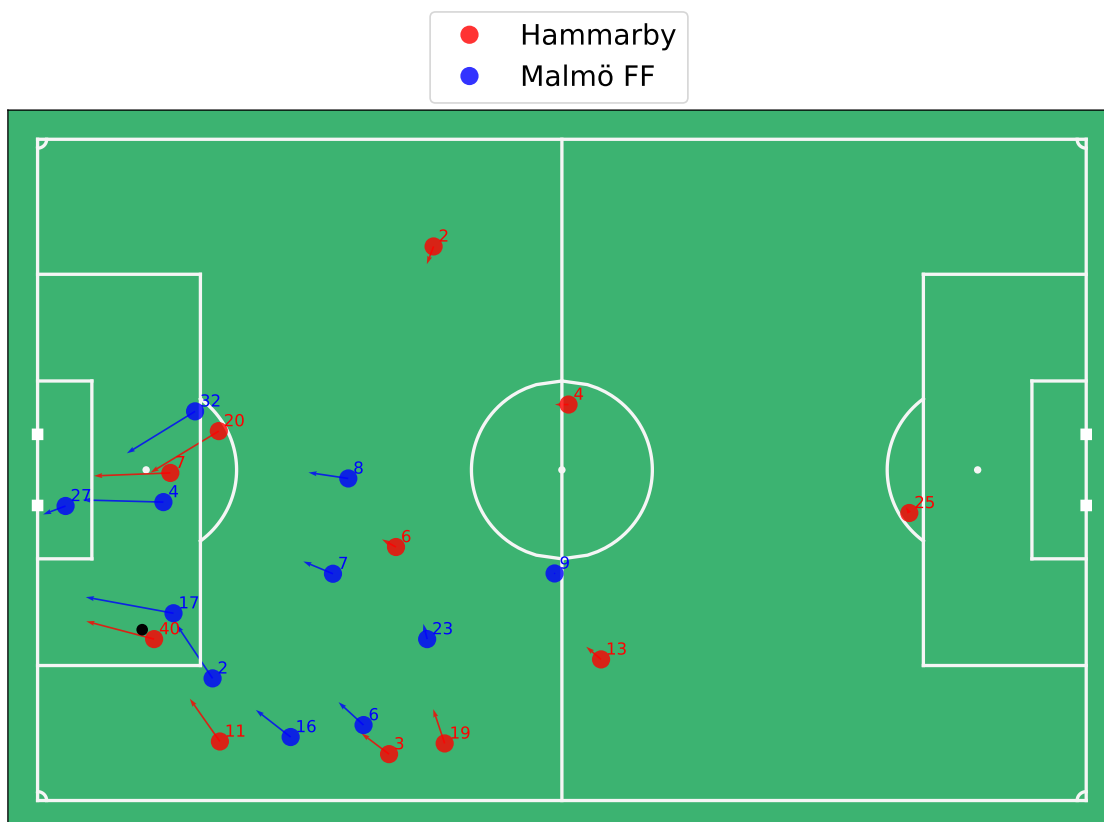


FIG. 2: Another assist for Djurdjic (#40) as Kacaniklic (#20) opens the scoring against Malmö.

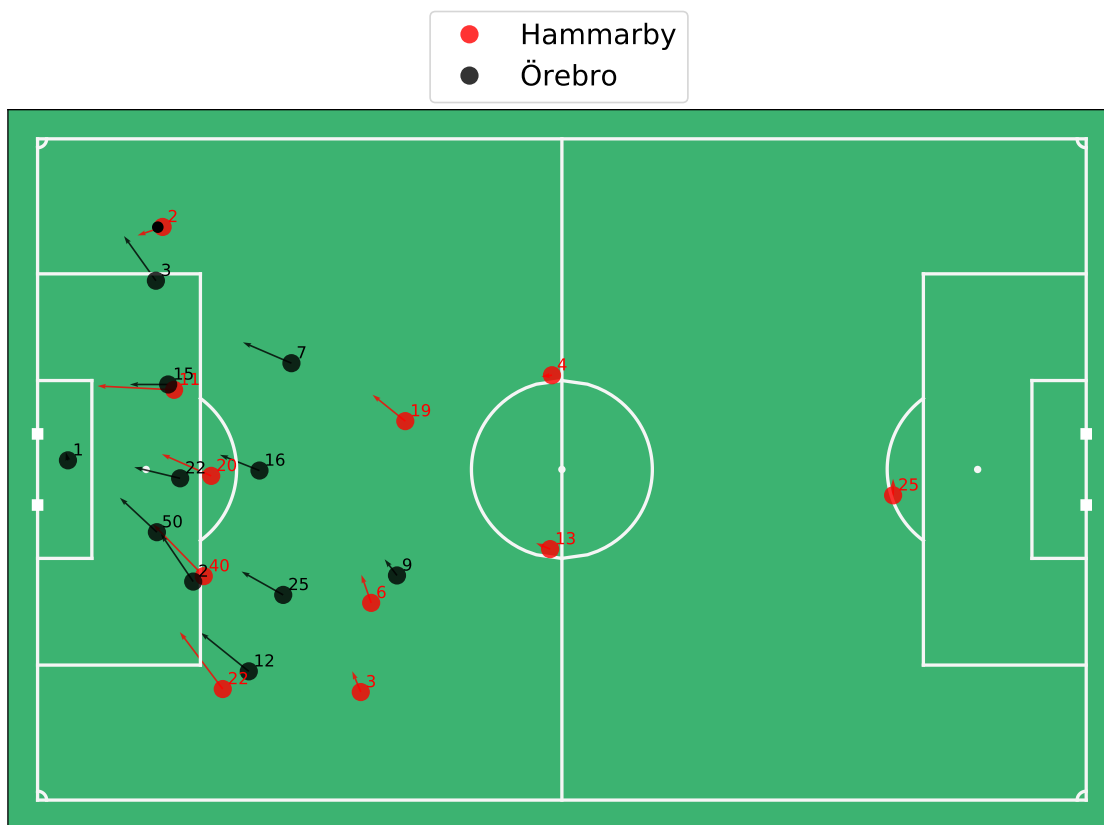


FIG. 3: Rodić (#11) heads in to equalise 1-1 against Örebro.

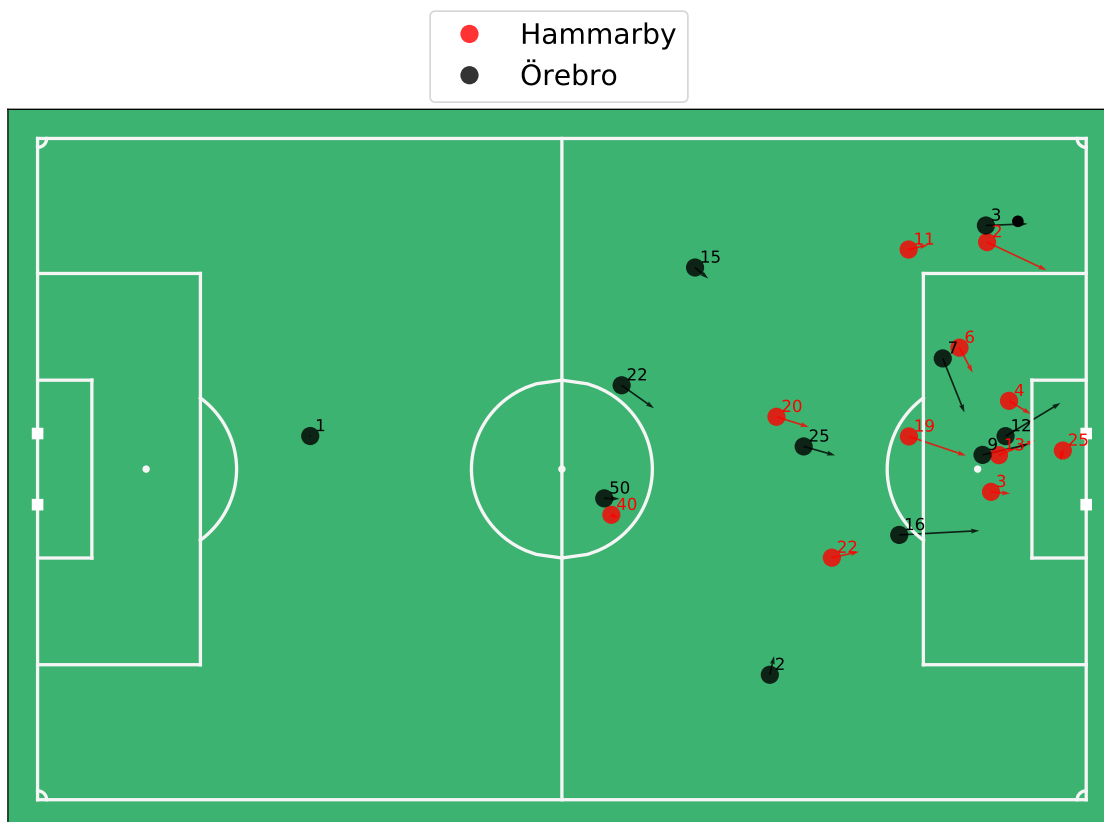


FIG. 4: Early goal conceded by Hammarby, with Prodell (#9) scoring a bicycle kick after late-arriving Kasim (#16) nodded a deep cross back toward goal.

A DEEPER LOOK AT KHALILI'S SECOND GOAL VS ELFSBORG

Hammarby's third goal against Elfsborg was a masterclass in off-the-ball movement that is captured in FIG. 6. Midfielder Bojanic started the sequence by dribbling the ball from the middle of the pitch towards the final third. Djurdjic, ever the creator, moved *away* from the ball from the left striker position towards the left wing, creating space by dragging away *both* the opponent right back, Hoist, *and* right centre half, Gregersen. It's barely noticeable on the broadcast footage, but clear in a tracking data motion sequence as the dragging is detected from the velocity directions of the defenders syncing with Djurdjic' as he made his move. Tankovic made the opposite movement, darting from the left flank in behind the out of position Gregersen, which forced the left centre half, Portillo, to cover and lose track of Khalili on his shoulder. Khalili had started accelerating towards goal *before* the ball had even arrived at Tankovic's feet, creating separation between himself and his marker Portillo (final plot of FIG. 6). Tankovic bisected the centre halves with a killer pass into Khalili's path – with 6 metres of space between himself and the nearest defender – who finished with a dink over keeper, Ellegaard.

Two sharp accelerations injected one-pass-ahead cut through the Elfsborg defence and killed the game before halftime. The only time Khalili stopped moving during the sequence was during his goal celebration, and the closest that keeper Ellegaard could get to him was after the ball had already hit the back of the net.

The opposing runs of Tankovic and Djurdjic in the plane of the defensive line had a *shearing effect* on the defenders, creating space for Khalili to score.

Djurdjic didn't take himself out of play entirely either; with all eyes on Khalili as the final ball was played, Djurdjic darted back towards goal – the closest Hammarby player to Khalili as the shot was made (penultimate plot of FIG. 6) – to follow up if the chance was missed. A valuable contribution that may have gone unnoticed without the tracking data considering he didn't kick the ball once!

Player distance to the nearest teammate / opponent was calculated per frame by measuring the distance between a player and all teammates / opponents and keeping track of the jersey number and distance of the closest.

SHEARING RUNS

Since the side-to-side *shear* running of Tankovic and Djurdjic seemed to open up the Elfsborg defence to enable Khalili's forward running to have maximum impact, we first inspect the directionality of runs made in the opponent's half in the three games that we have tracking data for, normalising per 90 minutes. We define running as sustained movement of over 4ms^{-1} for at least a second.

Rodic tops TABLE I in terms of forward running percentage in the opponent's half, closely followed by Khalili. Full backs Sandberg and Widgren also make up the top five, highlighting Hammarby's intent to attack through wide areas; a strategy showcased in each of the opening goals against IF Elfsborg, Malmö FF, and Örebro, with Sandberg providing the assist for Rodic in FIG. 3 from a cross.

It's interesting to see a different profile of runner emerge in TABLE II when looking at players whose primary axis of movement in the opponent's half is side-to-side. Three of the four players involved in Khalili's second goal against Elfsborg feature here, with Djurdjic – a creator of space off the ball and assists on it – topping the list.

A hypothesis that starts to emerge from this discovery analysis of a handful of Hammarby games is that there's far more to effective movement than running towards goal as fast as humanly possible. At

Player	# Forward	# Backward	# Left	# Right	% Forward
Vladimir Rodic	68	24	20	16	53.3
Imad Khalili	58	14	19	30	48.1
Simon Sandberg	33	20	7	10	47.0
Muamer Tankovic	50	26	17	17	45.4
Dennis Widgren	34	23	2	18	44.8

TABLE I: Top five run makers in a forward direction per 90 minutes made in the opponents half.

Player	# Forward	# Backward	# Left	# Right	% Side to Side
Nikola Djurdjic	49	19	26	66	57.4
Darijan Bojanic	17	15	14	28	56.5
Mads Fenger Nielsen	12	14	7.5	13	44.2
Alexander Kacaniklic	67	27	26	45	42.8
Imad Khalili	58	14	19	30	40.4

TABLE II: Top five side-to-side runners per 90 minutes made in the opponents half.

32 years old, Imad Khalili does not possess the kind of blistering pace that would traditionally strike fear into defenders. The synchronous movement of teammates perpendicular to goal enabled Khalili to find himself in 6 metres of space in a one-on-one situation with the goal-keeper, *and he didn't even break into a sprint*.

Runner #1	Runner #2	Shear Run Count	Shear Run Time [s]
Tankovic	Djurdjic	5	3.96
Bojanic	Kacaniklic	2	2.36
Kacaniklic	Tankovic	2	3.44
Rodic	Kacaniklic	2	1.96
Khalili	Tankovic	2	2.12

TABLE III: Pairs of players producing shear runs: synchronised sideways movement in opposite directions in the opponent's half.

As the final part of the discovery analysis we produce the *shear run count* metric: the number of side-to-side runs in opposite directions at the same time in the opponent's half by a pair of Hammarby players. Tankovic and Djurdjic – the pair that produced the shearing run for Khalili's goal – are firmly at the top. It would be interesting to learn whether these runs were tactical or opportunistic. All five of their shearing runs came in the game against Elfsborg, hinting at tactical play. Another possibility is that Tankovic – primarily a forward runner yet occupying three spots in the shear run list – has a knack for making complementary runs with respect to his teammates.

DISCOVERY ANALYSIS: FINAL THOUGHTS

This discovery analysis of tracking data has shown that, as with any dataset, the features and limitations must first be understood before insights can be trusted. For open-play sequences where the tracking data really shines, it's clear that Hammarby are proficient at building attacks through wide areas.

A central theme of this work is that a single player's movement in isolation isn't necessarily the key to opening up the opposition, but instead, it's a symphony of movement that imparts stress on the opponent's shape to create space to play through them. And that Sheringham and Bergkamp's teammates perhaps deserved a bit more of the credit...

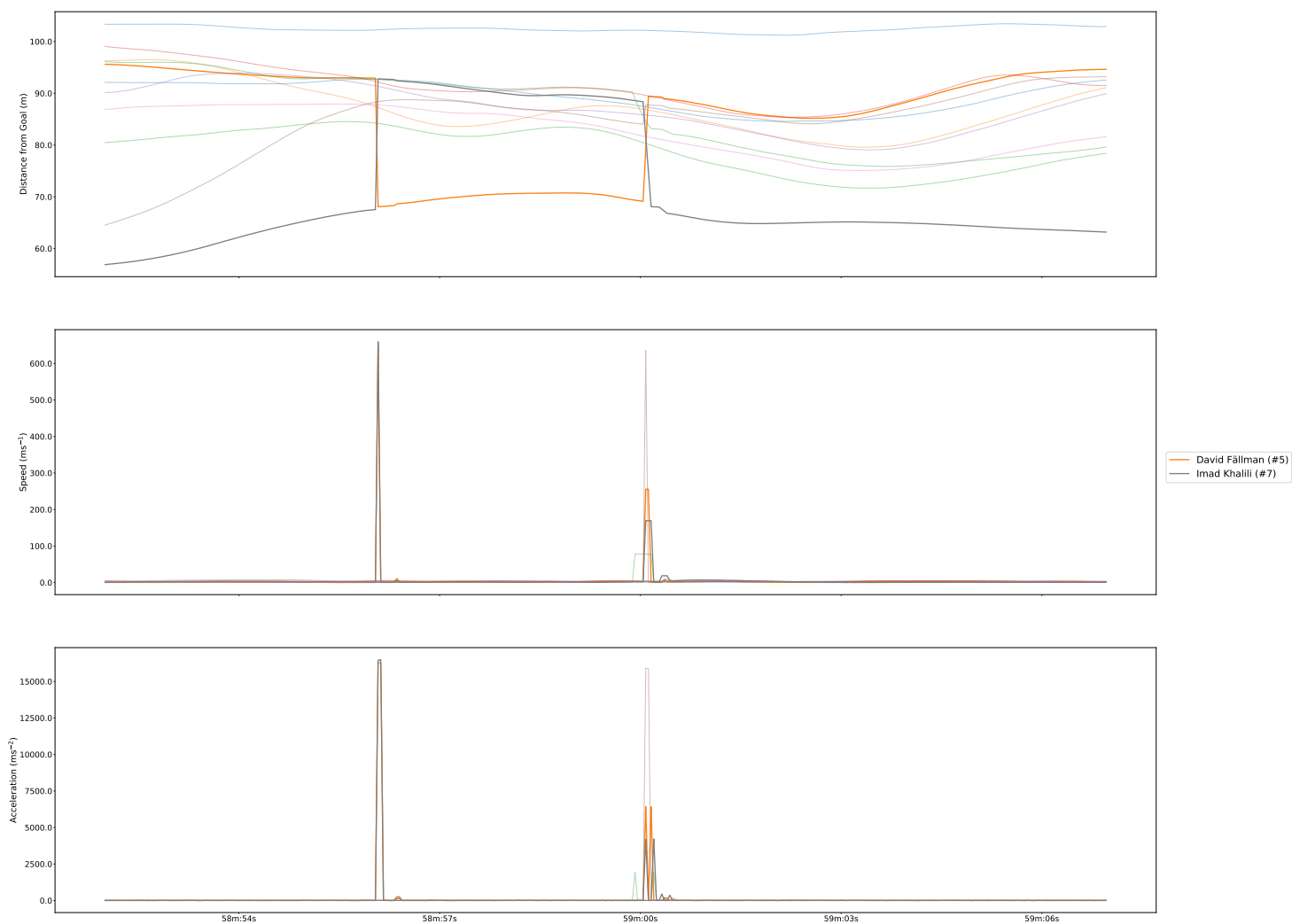


FIG. 5: **Player mapping errors.** Top panel illustrates the erroneous flipping of position between Khalili and Fällman, with the following panels displaying spikes in speed and acceleration as a result of the mapping error.

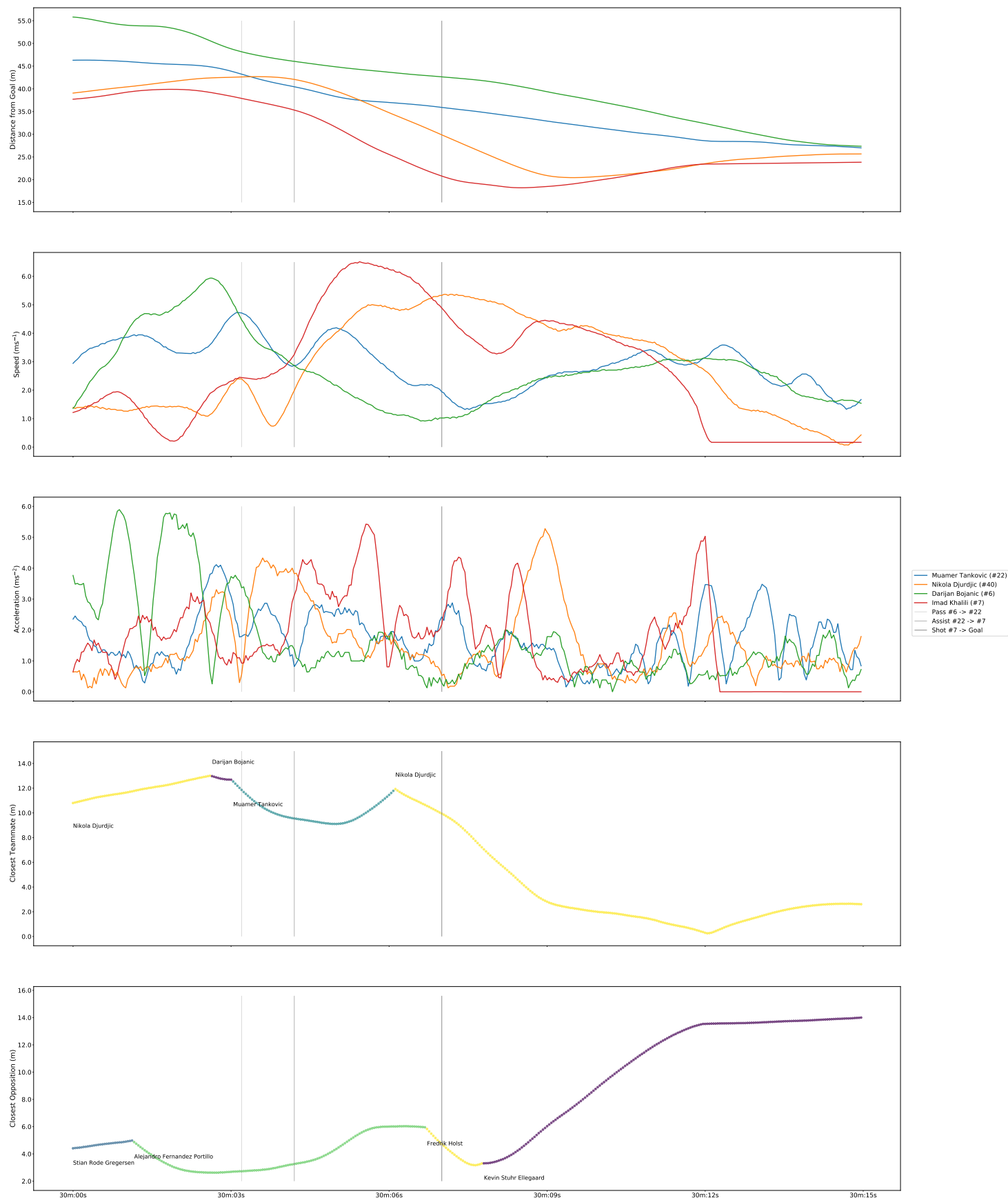


FIG. 6: Off-the-ball movement. Distance from goal, speed, and acceleration panels (as in FIG. 5) for the four main players involved in Hammarby's third goal Vs IF Elfsborg. Final two panels show goal scorer Khalili's closest teammate and opposition player throughout the sequence. Vertical lines indicate the pass, assist, and shot actions during the sequence.