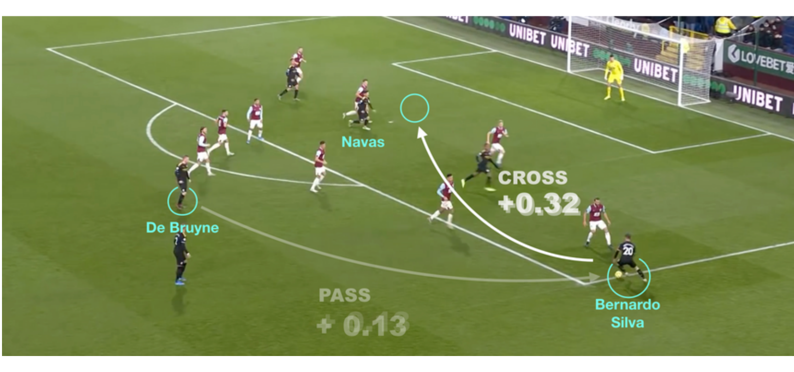
Expected Possession Value - EPV

1. Introduction

So what is EPV? EPV is the expected value of a possession, the probability that the attacking team will score given the current situation on the pitch. This includes many factors, notably the ball position, player positions and the gameplay scenario. This can be looked at mathematically as:

EPV = probability (GOAL | current situation), where current situation = a multitude of factors

So whilst a value such as expected goals is more widely understood as a measure of valuing shots, EPV answers how do we value possession of the ball in any given scenario on the pitch? Leading into how do we value different options on the ball?



The potential uses of EPV are extensive. It paves the way for new methods to analyse gameplay. As an example, EPV surfaces can be plotted of all the possession build up for opponent shots and goals, to evaluate strengths and weaknesses in their build up play. Additionally, the optimal pass during a possession can be evaluated, to determine how efficient a midfielder being scouted, is at creating dangerous opportunities. There are numerous more applications, some of which will be shown further in tihs notebook. Firstly, an EPV model is presented.

2. EPV Modelling

The first step in this process is to think of every possession as a 'state', where each state is a situation of play. For example, this can be a set piece situation, open play ball possession in a coordinate of the pitch, a penalty etc. These are all states, where the next state is determined by the current state.

Given the current state, what is the probability that the ball moves to another state

This is what is known as a Markov process, a stochastic modelling methodology. Of course the key here is that there are only 2 end states in a football match:

- 1. Loss of possession
- 1. Scoring a goal

Thus, we can use this methodology to quantify the probability of scoring or losing possession in the current state by formulating a transition matrix:

$$T = \begin{array}{c|ccccccc} & \multicolumn{7}{c}{\text{Current state}} \\ & S_1 & S_2 & S_3 & \dots & S_N & Goal & Loss \\ \hline S_1 & P(S_1 \to S_1) & P(S_2 \to S_1) & & & P(S_N \to S_1) & 0 & 0 \\ S_2 & P(S_1 \to S_2) & P(S_2 \to S_2) & & & P(S_N \to S_2) & 0 & 0 \\ S_3 & P(S_1 \to S_3) & P(S_2 \to S_3) & & & P(S_N \to S_3) & 0 & 0 \\ \dots & & & & & & 0 & 0 \\ S_N & P(S_1 \to S_N) & P(S_2 \to S_N) & & & P(S_N \to S_N) & 0 & 0 \\ Goal & P(S_1 \to G) & P(S_2 \to G) & & & P(S_N \to G) & 1 & 0 \\ Loss & P(S_1 \to L) & P(S_2 \to L) & & & P(S_N \to L) & 0 & 1 \end{array} \left.\vphantom{\begin{array}{c}1\\1\\1\\1\\1\\1\\1\end{array}}\right\} \text{New state}$$

The transition matrix is filled out using large quantities of event data, and is then used to calculate the probability of a goal given the current state. For example the probability of moving from one state to another (t+1) is:

Transition Matrix Current State*

And, the probability of moving from one state to a state in 2 transitions time (t+2), i.e. A throughball (t+1), and then obtaining a penalty (t+2):

Transition Matrix Transition Matrix *Current State*

Therefore, the probability from any current state to the end state being a goal can be calculated, thus giving a valuation of the possession value of the current state.

Note that this is of course a brief explanantion of the EPV calculation, as the full mathematics is beyond the remit of this notebook. An EPV grid is kindly provided via the Friends of Tracking initiative for public use, along with some helpful functions shared on GitHub by Laurie Shaw. The following section delves into the applications of EPV to advance the analysis and understanding of football.

3. EPV Applications

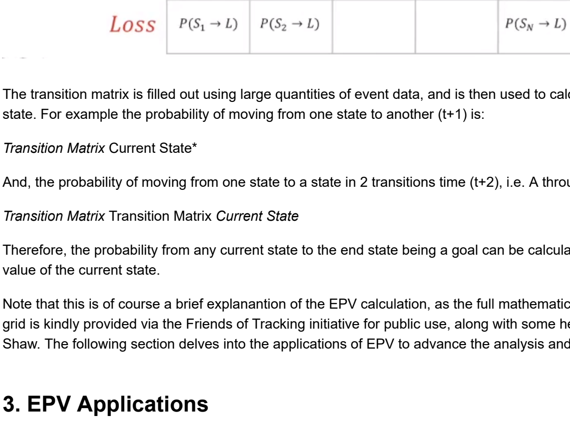
Firslty, certain functions are executed to set up the data for analysis such as reading in the Metrica datasets, converting the Metrica pitch coordinates to meters and setting up single playing direction to account for the change in halves.

In [16]:
```
import Metrica_IO as mio
import Metrica_Viz as mviz
import Metrica_Velocities as mvel
import Metrica_EPV as mepv

# set up initial path to data
DATADIR = r'C:\Users\steff\Documents\Football Analytics\FOT Tracking Data - Laurie\sample-data-master\d
ata'

game_id = 2 # let's look at sample match 2

# read in the event data
events = mio.read_event_data(DATADIR,game_id)

# read in tracking data
tracking_home = mio.tracking_data(DATADIR,game_id,'Home')
tracking_away = mio.tracking_data(DATADIR,game_id,'Away')

# Convert positions from metrica units to meters (note change in Metrica's coordinate system since the
 last lesson)
tracking_home = mio.to_metric_coordinates(tracking_home)
tracking_away = mio.to_metric_coordinates(tracking_away)
events = mio.to_metric_coordinates(events)

# reverse direction of play in the second half so that home team is always attacking from right->left
tracking_home,tracking_away,events = mio.to_single_playing_direction(tracking_home,tracking_away,events
)
```

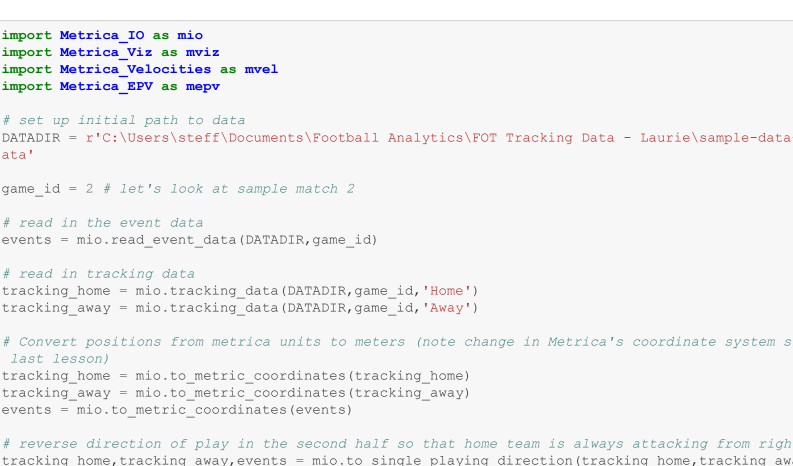
```
Reading team: home
Reading team: away
```

Once this is set up, the aforementioned EPV grid that has been calculated can be plotted to show the probability of a goal, given the possession of the ball at any location on the pitch.

In [17]:
```
""" *** GET EPV SURFACE **** """
home_attack_direction = mio.find_playing_direction(tracking_home,'Home') # 1 if shooting left-right, el
se -1
EPV = mepv.load_EPV_grid(r'C:\Users\steff\Documents\Football Analytics\FOT Tracking Data - Laurie\Lauri
eOnTracking-master\EPV_grid.csv')
# plot the EPV surface
mviz.plot_EPV(EPV,field_dimen=(106.0,68),attack_direction=home_attack_direction)
```



The key insight here is that the value added of a pass or dribble can be calculated.



For example, in this image above, the pass highilighted in green would give a value added of:

EPV ADDED = EPV2 - EPV1 = 0.047 - 0.018 = 0.029

This is a simple example to demonstrate the basic element of EPV, however, of course the influence of ball and player positions completely redefines the value of every possession. Therefore, it is a combination of the EPV and the probability that a teammate actually recieves the ball, that defines the real value. This means that EPV has to be combined with the pitch control model. Please review the pitch control notebook if necessary before continuing with the following analysis.

COMBINING EPV WITH A PITCH CONTROL MODEL TO EVAULATE EXPECTED VALUE ADDED

Given that EPV is the probability that a possession leads to a goal, and that pitch control is the probability that a team controls the ball if it arrives at any position on the pitch, the expected value added is calculated as the product:

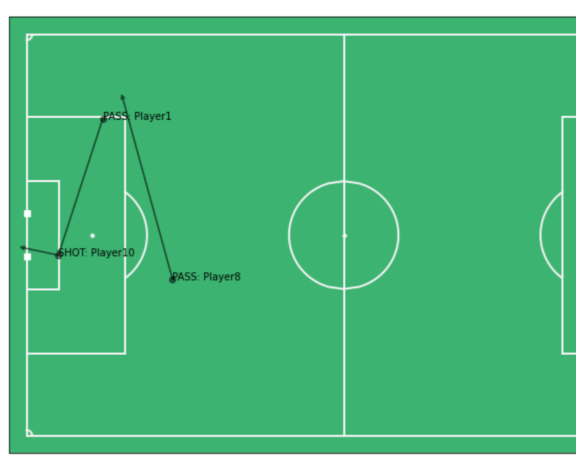
EXPECTED VALUE ADDED = (EPV2 PC2) - (EPV1 *PC1*)

Consequently, the entire pitch can be mapped out in an incredibly useful manner, giving the expected value added for every single option available to a player in possession of the ball.

To demonstrate this, firstly, the event data is looked at to select the first goal scored in this match as an example scenario to analyse. Note this is the same event used in the pitch control notebook for consistency. The following figure below is the pass netwok for the 2 passes leading to the goal.

In [18]:
```
# plot event leading up to first away team goal
mviz.plot_events( events.loc[196:198], color='k', indicators = ['Marker','Arrow'], annotate=True )
```

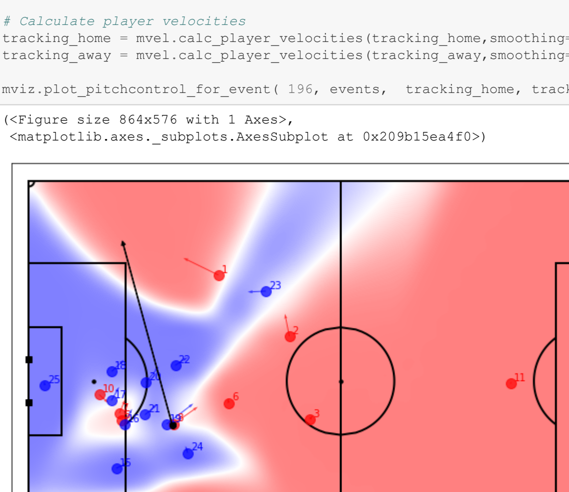
Out[18]:
```
(<Figure size 864x576 with 1 Axes>,
 <matplotlib.axes._subplots.AxesSubplot at 0x209b152cac0>)
```



The player velocities are then calculated to add to the tracking data in order for the pitch control calculations to be made. Again, please view the pitch control notebook for any clarification on how this visualization is formulated.

In [19]:
```
import Metrica_PitchControl as mpc

# first get pitch control model parameters
params = mpc.default_model_params()
# find goalkeepers for offside calculation
GK_numbers = [mio.find_goalkeeper(tracking_home),mio.find_goalkeeper(tracking_away)]

# Calculate player velocities
tracking_home = mvel.calc_player_velocities(tracking_home,smoothing=True,filter_='moving_average')
tracking_away = mvel.calc_player_velocities(tracking_away,smoothing=True,filter_='moving_average')

mviz.plot_pitchcontrol_for_event( 196, events,  tracking_home, tracking_away, PPCF, annotate=True)
```

Out[19]:
```
(<Figure size 864x576 with 1 Axes>,
 <matplotlib.axes._subplots.AxesSubplot at 0x209b15ea4f0>)
```

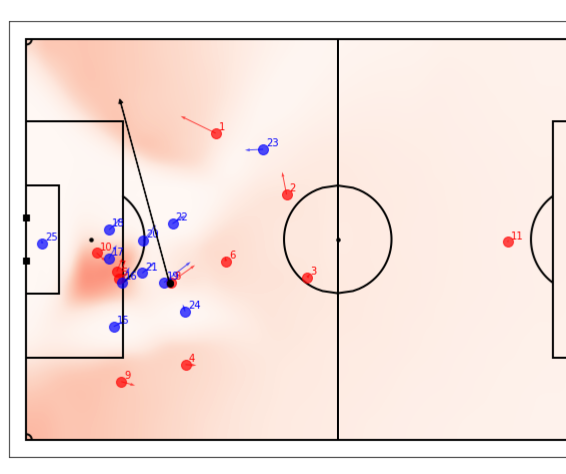


Here we see that the pass played is into a very good area in that it advances the attack into a more dangerous position as well as having a high probability of being a successful. Now, we can advance this visualization by combining it with the EPV model as EPV * Pitch Control, to create the entire pitch value surface shown in the figure below.

In [20]:
```
# Calculate value-added for the penultimate pass, and plot expected value surface
event_number = 196 # away team first goal
EEPV_added, EPV_diff = mepv.calculate_epv_added( event_number, events, tracking_home, tracking_away, GK
_numbers, EPV, params)
PPCF,xgrid,ygrid = mpc.generate_pitch_control_for_event(event_number, events, tracking_home, tracking_a
way, params, GK_numbers, field_dimen = (106.,68.,), n_grid_cells_x = 50, offsides=True)
fig,ax = mviz.plot_EPV_for_event( event_number, events,  tracking_home, tracking_away, PPCF, EPV, annot
ate=True, autoscale=True )
fig.suptitle('Pass EPV added with pitch control: %1.3f' % EEPV_added, y=0.95 )
```

Out[20]:
```
Text(0.5, 0.95, 'Pass EPV added with pitch control: 0.021')
```

Pass EPV added with pitch control: 0.021



Here we can see this is an extremely valuable pass played from the central position out wide to the player in space. EPV adds the extra information that is not visualized by pitch control alone.

Some other points to consider:

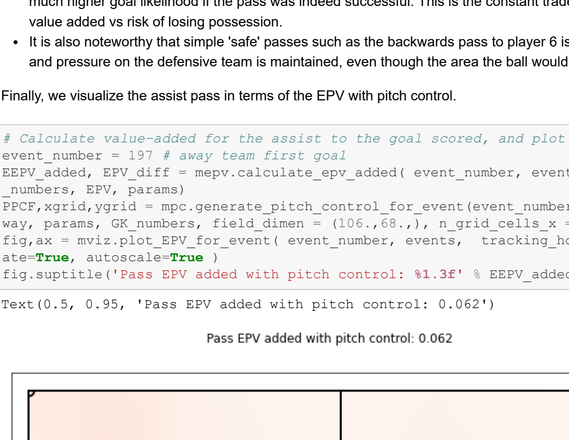
- The area in the box around player 10 is valued even higher by the model than the pass that was made, which is a new insight not observed from the pitch control alone. Whilst it is a very 50/50 area in terms of making a successful pass, the value is higher due to a much higher goal likelihood if the pass was indeed successful. This is the constant tradeoff between attempting high value passes: value added vs risk of losing possession.
- It is also noteworthy that simple 'safe' passes such as the backwards pass to player 6 is also a valued pass as possession is retained and pressure on the defensive team is maintained, even though the area the ball would be recieved in is a less dangerous position.

Finally, we visualize the assist pass in terms of the EPV with pitch control.

In [21]:
```
# Calculate value-added for the assist to the goal scored, and plot expected value surface
event_number = 197 # away team first goal
EEPV_added, EPV_diff = mepv.calculate_epv_added( event_number, events, tracking_home, tracking_away, GK
_numbers, EPV, params)
PPCF,xgrid,ygrid = mpc.generate_pitch_control_for_event(event_number, events, tracking_home, tracking_a
way, params, GK_numbers, field_dimen = (106.,68.,), n_grid_cells_x = 50, offsides=True)
fig,ax = mviz.plot_EPV_for_event( event_number, events,  tracking_home, tracking_away, PPCF, EPV, annot
ate=True, autoscale=True )
fig.suptitle('Pass EPV added with pitch control: %1.3f' % EEPV_added, y=0.95 )
```

Out[21]:
```
Text(0.5, 0.95, 'Pass EPV added with pitch control: 0.062')
```

Pass EPV added with pitch control: 0.062



We see the assist in this move was the decisive pass adding even more value than the pass recieved (penultimate pass = 0.021 value added, assist = 0.062 value added), giving the team a high chance of scoring. In this scenario the high value pass was indeed converted into a goal.

Though football can be a cruel game in that luck and missed opportunities can ultimately be decisive in a fixture, creating high value opportunities in the long run will undoubtedly yield results. In Jurgen Klopp's final year at Dortmund for example, which was a 'disaster', his team actually remained very industrious in terms of creating opportunities and dominating games. Analysts at Liverpool were well aware of this, and did not simply view final results that year to evaluate him as a manager. They signed him that very summer after a failed season in the Germany, 5 years later... look now!

4. Concluding Remarks

Summary

This notebook has been an introduction into evaluating possession in football. A Markov process model has been explained as a way of finding the probability of a goal resulting from the current possession. Combined with a pitch control model to leverage the tracking data of the 22 players, this methodology yields an extremely powerful tool to analyse gameplay.

Future Applications

The applications of what has been introduced here are huge. The ability to quantitavely evaluate the value added of every potential pass available to any player provides insight that has been impossible until now. Whilst coaches in the past have of course been able to view video and pass judgment on where a player has made optimal decicions or not, never have they had a tool that could be run on software that has the capacity to evaluate every pass a player has made over an entire season. In what positions do they play their most dangerous passes? Where has the opposition left high value areas exposed in their new system? How do two players being scouted compare in their distribution of safe passes vs high value passes?

Computing power gives rise to the ability to break the game down into new metrics, whilst analysing them millions of times quicker than the brain could dream of, watching video alone. The journey now in football analytics is about finding the very best metrics and learning to best communicate with coaches to work together and advance the game.