Cutting patterns in D.W. Griffith's Biographs: An experimental statistical study

Mike Baxter, 16 Lady Bay Road, West Bridgford, Nottingham, NG2 5BJ, U.K.
(e-mail: michaelj.baxter@btconnect.com)

1 Introduction

A number of recent studies have examined statistical methods for investigating cutting patterns within films, for the purposes of comparing patterns across films and/or for summarising 'average' patterns in a body of films. The present paper investigates how different ideas that have been proposed might be combined to identify subsets of similarly constructed films (i.e. exhibiting comparable cutting structures) within a larger body. The ideas explored are illustrated using a sample of 62 D.W Griffith Biograph one-reelers from the years 1909–1913.

Yuri Tsivian has suggested that 'all films are different as far as their SL structures; yet some are less different than others'. Barry Salt, with specific reference to the question of whether or not Griffith's Biographs 'have the same large scale variations in their shot lengths along the length of the film' says the 'answer to this is quite clearly, no'. This judgment is based on smooths of the data using seventh degree trendlines and the observation that these 'are nearly all quite different one from another, and too varied to allow any grouping that could be matched against, say, genre'[1]. While the basis for Salt's view is clear Tsivian's apparently opposing position that some films are 'less different than others' seems to me to be a reasonably incontestable sentiment.

It depends on how much you are prepared to simplify structure by smoothing in order to effect comparisons. You can push the idea close to a limit by fitting first degree trendlines (i.e. straight lines) through the data and removing scale effects due to differing average shot lengths. The smooths so obtained then differ only in the slopes of the lines and it is pefectly possible to then effect comparisons in terms of similarity. The present paper – and I would wish to stress the term 'experimental' in the title – pursues this idea without going to quite such an extreme of smoothing. The idea is to try and identify films with, more or less, similar smoothed structures, without regard to factors such as date or genre, and look for features similarly structured films might have in common only after some structural grouping has been achieved.

Nick Redfern, in a number of posts on his and the Cinemetrics websites, has explored the use of cumulative frequency distributions (CFDs) for comparing shot length (SL) distributions of films. These have been exploited, for example, to compare films of Chaplin and of Hitchcock. The comparisons have involved small numbers of films and are of a descriptive nature complemented by the use of statistical

[1] See Tsivian's paper first contribution to *Question 3: Looking for Lookalikes?* on the Cinemetrics website at http://www.cinemetrics.lv/dev/on_statistics.php, where Salt's paper is also located.

tests (the two-sample Kolmogorov-Smirnov test) to assess the significance of differences between pairs of films[2].

Keith Brisson and Yuri Tsivian have developed an idea for averaging cutting patterns across films to obtain a 'signature' pattern for a body of films, illustrated with reference to a sample of Griffith Biographs, similar to but not identical with those used here. The fundamental idea is a simple one; films are partitioned into an equal number of time intervals of equal duration; for a film 'localised' average shot lengths (ASLs) are computed for each interval based on the computation of fractional counts; and these are averaged across films for each interval to obtain an average pattern. Barry Salt, in commenting on Brisson's original post, noted that 'shot-density' and SL are inversely related and illustrated the use of localised and averaged densities to summarise patterns across films. The idea of smoothing the averaged pattern, which can be quite 'rough', was also illustrated[3].

My contribution to the discussion that contains Tsivian's paper investigated different methods that might be used for estimating averaged cutting patterns across films, among other things applying Salt's 'discrete density estimation' method to a sample of Griffith Biographs[4]. The point was made that it is possible, in principle, for no film in a sample to have a cutting profile that matches the average; by example it was illustrated that subsets of films, defined by genre and/or date, could have markedly different average patterns from each other and from the overall average. This is not especially surprising; it does not, however, exclude the possibilities either that films within a subset can differ noticeably from one another or that some may be similar to films in other subsets.

That is, any attempt to define a 'meaningful' average pattern in a body of films defined on filmic grounds alone – 'meaningful' in the sense that the average corresponds to the pattern exhibited by a reasonable number of individual films – is beset by the problem that no such correspondence may exist. This is an extreme possibility, though it can't be ignored. Tsivian's analysis of four crosscutting Griffith films in Figures 14–17 of his Cinemetrics paper is instructive in this respect. The graph of the average does well enough at capturing the acceleration in cutting that take place about half-way through three of the films, but hides some obvious differences and similarities between films, evident on comparing pairs of graphs [5].

The purpose of this paper is to explore how, in advance of any averaging procedure, subsets of structurally similar films (and 'outliers') might be identified. If this

[2]The use of the CFD with an application to two Hitchcock films is illustrated in Redfern's contribution to *Question 2: What Do Lines Tell?* on the Cinemetrics website at http://www.cinemetrics.lv/dev/on_statistics.php. For Chaplin see http://nickredfern.wordpress.com/2011/03/17/estimating-shot-length-distributions/ and http://nickredfern.files.wordpress.com/2010/04/nick-redfern-shot-length-distributions-in-the-early-films-of-charles-chaplin.pdf

[3]See http://www.cinemetrics.lv/topic.php?topic_ID=356 for Brisson's post with Salt's comments; Tsivian's paper is his contribution to *Question 3: Looking for Lookalikes?* on the Cinemetrics website at http://www.cinemetrics.lv/dev/on_statistics.php.

[4]At the time of writing that paper I had not implemented the Brisson approach; this has since been rectified and gives, as might be expected, similar results to the density-based methodology.

[5]Tsivian makes this point as well as expressing concerns about the validity of comparing films with rather different numbers of shots. Of the films in his example *The Prussian Spy* and *Behind the Scenes* have rather fewer shots, 11 and 14, than I'd be comfortable about using in a statistical analysis . The thought here is that with so few shot there is much less scope for variation to be exhibited than in films with more shote (and I hesitate to define 'more'). *The Prussian Spy* has the additional problem that the first half is dominated by a single shot which, with a sample of four, may unduly affect the averaged patten, though it exhibits the crosscutting in the second half well enough.

can be done at all convincingly it then becomes of interest to see if the groups so defined admit a 'filmic' or other interpretation. For example, are subsets temporally distinct; do crosscutting films group together, and if not what distinguishes them etc.? The approach to be illustrated is based on the following steps.

1. Cumulative frequency distributions (CFDs) are obtained of film duration against shot number. Duration is scaled to the interval [0, 100].

2. The CFDs are smoothed, during which process 'fitted' values at 100 equally spaced intervals are obtained.

3. The smoothed CFDs are grouped with the aid of standard multivariate statistical analysis methodology, based on an $n \times 100$ table of data, where n is the number of films in the sample, the 100 corresponding to the fitted values..

4. Within-group averages are obtained, and can be smoothed for comparative purposes.

Inspection of the original SLs for each film, plotted against cut-points may inform modification of any groups suggested by statistical analysis. A lot of data manipulation is involved in group definition, and groups can be formed that are visually unconvincing, containing members that clearly don't 'belong'.

The analysis of CFDs is applied, as far as I'm aware, to a much larger body of films than has previously been attempted. The smoothing is undertaken to obtain equal numbers of 'observations' for each film, needed for the subsequent multivariate analyses. The smoothing is accomplised using functions in the `fda` – functional data analysis (FDA) package of `R` and the exercise as a whole can be thought of as an application of FDA (Baxter, 2012b, Chapter 8). The statistical 'technology' used is of greater complexity than much statistical analysis in the cinemetric literature, and later exposition is confined to the closest I can get to 'ordinary language' explanations. The theory of FDA is available in Ramsay and Silverman (2005), with a series of case studies provided in Ramsay and Silverman (2002) and guidance about applications using `R` software in Ramsay *et al.* (2009). Examples in the last of these, with associated `R` scripts available via the `fda` package in `R`, were used as an initial template for the analyses conducted here.

Operating with CFDs estimated at an equal number of equally spaced intervals is analogous to the idea exploited by Brisson, whereby comparisons are effected by controlling both for unequal film durations and numbers of shots. Sampling a smoothed estimate at equally spaced intervals is an idea that has been exploited in Baxter(2012b) and Redfern (2013). In the final stage the Brisson methodology can be used to explore how the average SL distributions vary over film duration for different groups. This involves operating with the SLs rather than their CFDs and provides a check on whether any proposed grouping makes sense.

The data used are taken from two of the Cinemetric labs listing of D.W Griffith's Biograph films[6]. The larger of these contains 61 films, the smaller 19 'rescue' films. A few duplicates were removed from the combined samples, as were two–reelers and split–reelers (taken to be films of less than 900ft). This was done in order to have a

[6] At http://www.cinemetrics.lv/lab.php?ID=119 and http://www.cinemetrics.lv/lab.php?ID=127

sample consisting of films of roughly similar length. Films of 20 or fewer shots were also removed to minimise potential statistical issues arising with small numbers. The combined effect of these decisions was to eliminate films from 1908 (including several of the rescue films) leaving 62, dated 1909–1913, in total.

2 Introductory analyses

Some concrete illustrations of what is being attempted are provided below to illustrate some of the stages described above. Figure 1 shows the shot length distribution for *A Corner in Wheat* before and after applying the Brisson partitioning process. In the former case the time axis is scaled to lie between 1 and 100, with plotting positions defined by the cut-points after this rescaling[7]. By construction, plotting positions in the latter case are the integers in the range [1,100] and can be read as the proportion of a film's duration. The main point in showing both graphs is to make clear that there is little difference between them, any difference being most evident in intervals in the second plot where a cut occurs.

The procedure was originally devised to enable averaging across films, for which equal numbers of observations are needed, simultaneously controlling for the fact that films have different numbers of shots and duration. Scaling the vertical axis is a possibility illustrated later. The original data can be thought of as a ragged array of 62 rows with unequal numbers of observations; after partitioning, the data are in the form of a 62 × 100 table (or *data matrix*) of equally spaced values. Beyond the original motivation for transforming in this manner, considerable statistical advantages are gained in analysis if data can be coerced into this format, a motivation for the way in which CFDs have been treated in this paper.

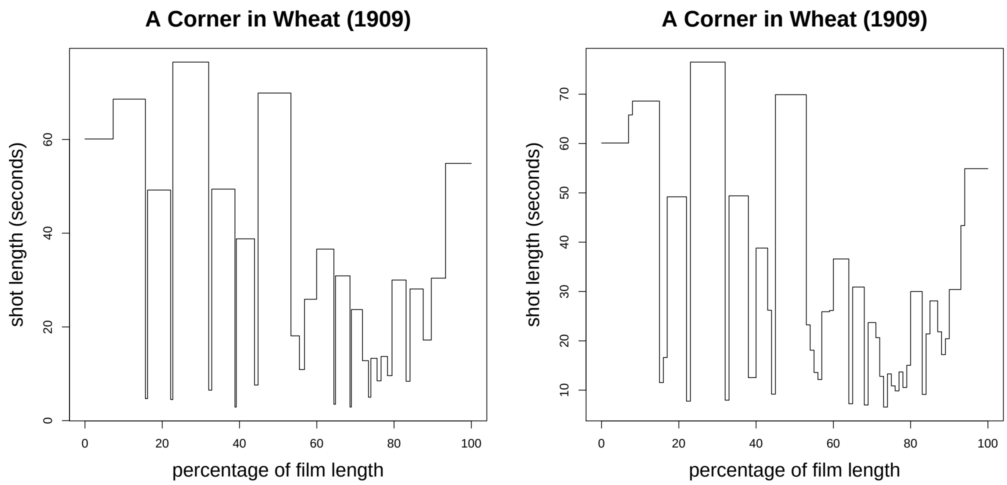

Figure 1: *The SL distribution of* A Corner in Wheat *before and after applying the Brisson method.*

Smoothed CFDs were obtained for each film using functions in the `fda` package in `R`, resulting in the left-hand panel of Figure 2[8]. It is obvious looking at the

[7]That is, divide SLs by their sum and multiply by 100.

[8]Readers are spared the details. The data are formally similar to the growth curve data introduced on pages

'envelope' of the plot that some films differ clearly from the majority and from each other; most are, however, not so easily differentiated, which is where statistics comes in.

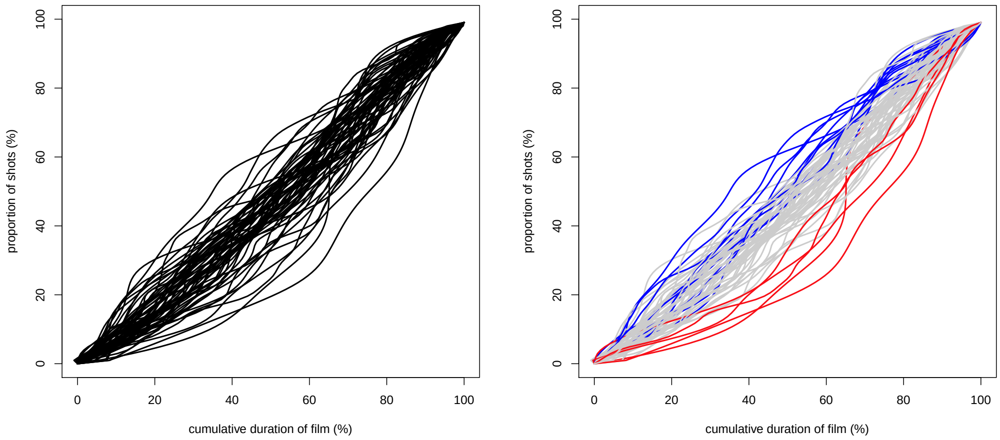

Figure 2: *Smoothed CFDs for 62 Griffith Biographs; the graph to the right is the same as that to the left with the color scheme explained in the text.*

For illustration, a sample of the more extreme films have been highlighted in the plot to the right. It is possible, more-or-less, to pick these out 'by eye' but they were, in fact, selected using statistical methods to be described in the next section. The films colored red and blue are at either end of a 'stylistic spectrum'. For further illustration a sample of plots for the original data is shown in Figure 3 – this kind of thing is advisable as a 'reality check' before getting too carried away.

The top row consists of three of the films colored red and are in the Cinemetrics lab of Griffith 'rescue' films. These films start 'slowly' in terms of numbers of shots, and get faster; for example the first 60% of a film involves only about 20-40% of the shots. This contrasts with the films that are coloured blue, a sample of which are shown in the rest of Figure 3. These films have more of a fast/slow trajectory. The 'red' and 'blue' films exhibit a reasonable similarity within their groups, and differences between them that might crudely be summarised as slow cutting in the first half and fast cutting in the second half for the 'red' films, with the reverse the case for the 'blue' films. Further discussion of issues involved in this kind of assessment are deferred until a more detailed analysis has been presented.

3 Further Analysis

It is obvious from Figure 2 that pairs of CFDs vary in the difference between them. The difference between any pair of CFDs can be thought of generally as 'distance'; mathematically, there are different ways of defining distance. Once, and however, this is done, a large number of applied multivariate statistical methods exist that

1–3 of Ramsay *et al.* (2009) and scripts they provide with the `fda` software to analyse these data in various parts of their book were used as templates.

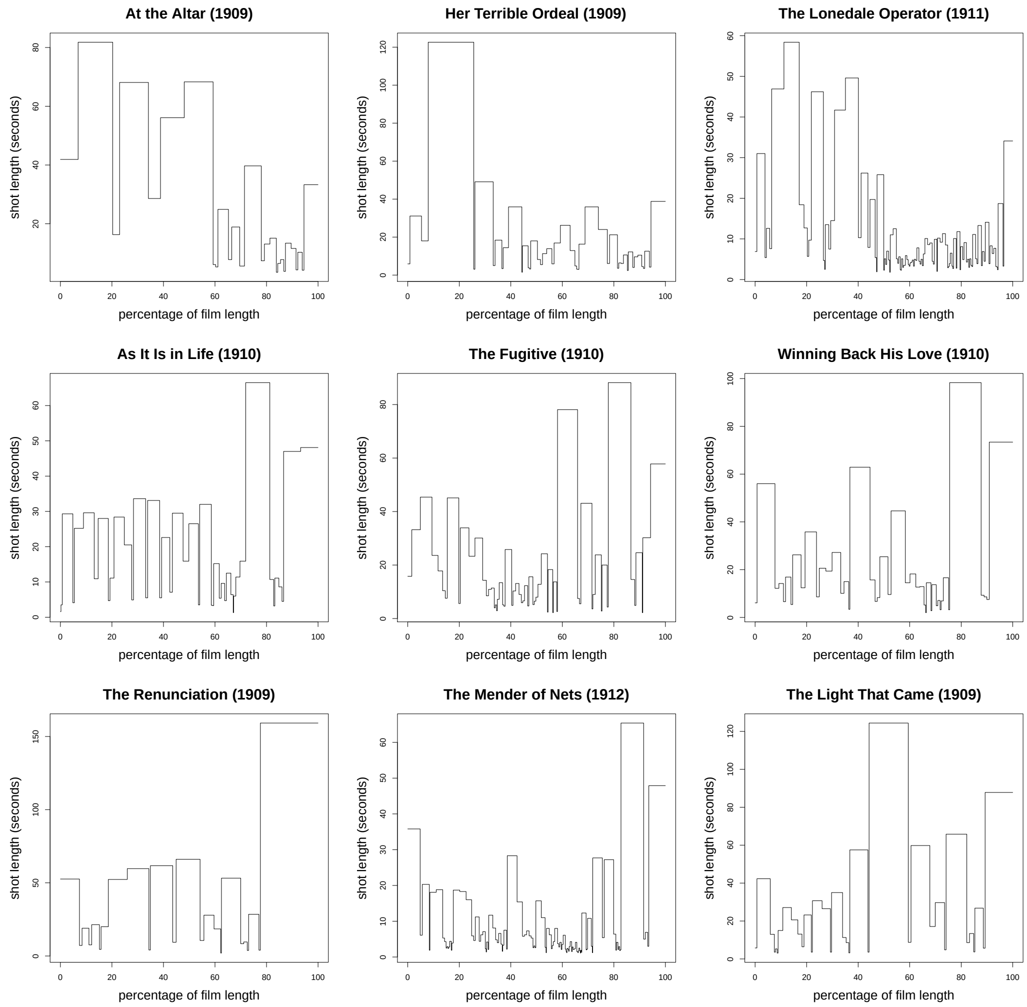

Figure 3: *A sample of plots for films highlighted to the right of Figure 2*

take all the pairwise distances between, in our example, the rows of a table of data (films) and convert them into a 'map'. This is done by constructing new variables from the original data, and plotting the most 'important' of these – often just the first two – to get a map. The idea is that the points on the map represent rows (films) and the real distances on the map *approximate* the actual distances used to construct the map. Issues can arise concerning the quality of approximation that are passed over here, but if it is reasonable the map can be inspected for potentially interpretable pattern – for example, clusters of films having a similar structure.

One of the most commonly used methods of multivariate analysis, and that employed here, is principal component analysis (PCA). A distinctive feature of FDA is that the rows of the data table are conceived of as samples along a curve; the curves (smoothed CFDs in our usage) are estimated as part of the analytical process using the `fda` package in `R`; and then PCA is used to obtain the desired map from the data matrix based on the CFDs, sampled at 100 equally spaced points. The practical guidance given in Sections 7.1–7.4 of Ramsay *et al.* (2009) is emulated

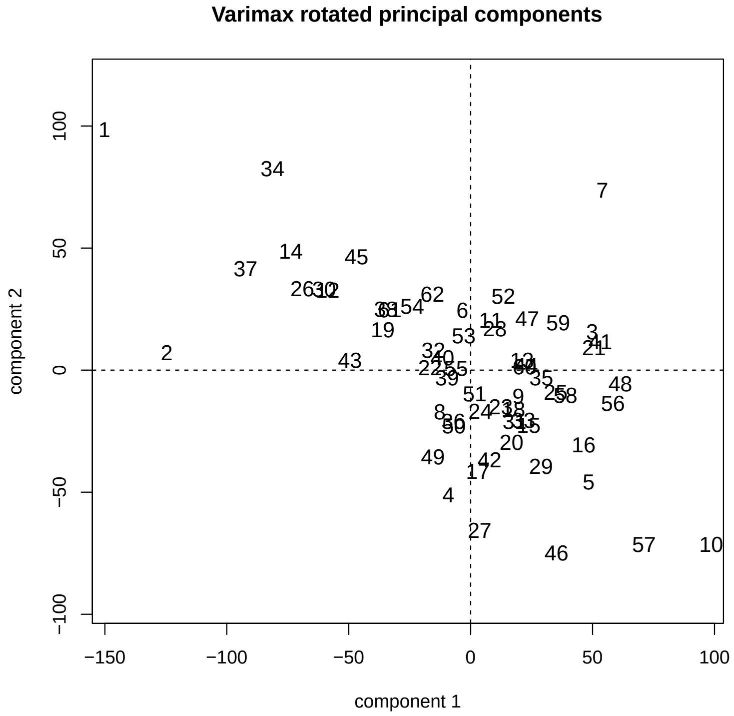

Figure 4: *A plot based on the first two principal components from a PCA based on the CFDs of the Griffith Biographs sample.*

here[9].

Applying this to the Griffith Biographs resulted in the map shown in Figure 4, with the concordance between films and numbers provided in Table 1 in Appendix 1. Interpretation of this plot is discussed in more detail in the next section but is unavoidably subjective when patterns are not obvious. As an *initial* aid Figure 5 shows the principal component plot with labelling suggested by cluster analysis[10].

That the map produced by the PCA is an approximation to the true distances between the CFDs, and that the results of cluster analysis depend on the method used and subsequent interpretation needs to be kept firmly in mind. The PCA shown can be thought of as a 'view' of the data from a particular position, and if you view it from another angle different features will emerge[11]. One point labelled as an outlier is not obviously so in the figure, but is revealed as such if you look from a different angle. Some of the points labelled red are obvious outliers but are

[9] Attention was focused on the first two components or 'harmonics'. The variation about the mean CFD 'explained' by these was 68.9%, 16.1% and 6.0% for the first three components. As recommended in Ramsay *et al.* (2009, p.104) varimax rotation of the components was used to produce Figure 4 to enhance interpretability.

[10] For a brief account of cluster analysis applied to SL data see Redfern (2013). I hope to provide a more thorough discussion of the detail in another paper. Briefly, the labelling is that suggested by an average-link cluster analysis of the unstandardised smoothed CFD estimates, but other methods such as Ward's method give similar results. It would make sense to treat some of the films labelled as red as outliers; they have been treated as part of a dispersed group as they are extremes that help in interpreting the pattern in the data.

[11] Think of how perception of a landscape varies according to whether you view it from the air, or from nearer ground level. The views in PCA are not determined arbitrarily; that presented is intended to be the most 'informative' in a mathematical sense I won't discuss here. Other views can be useful, and some of the commentary is based on inspection of plots using the third principal component, not shown here.

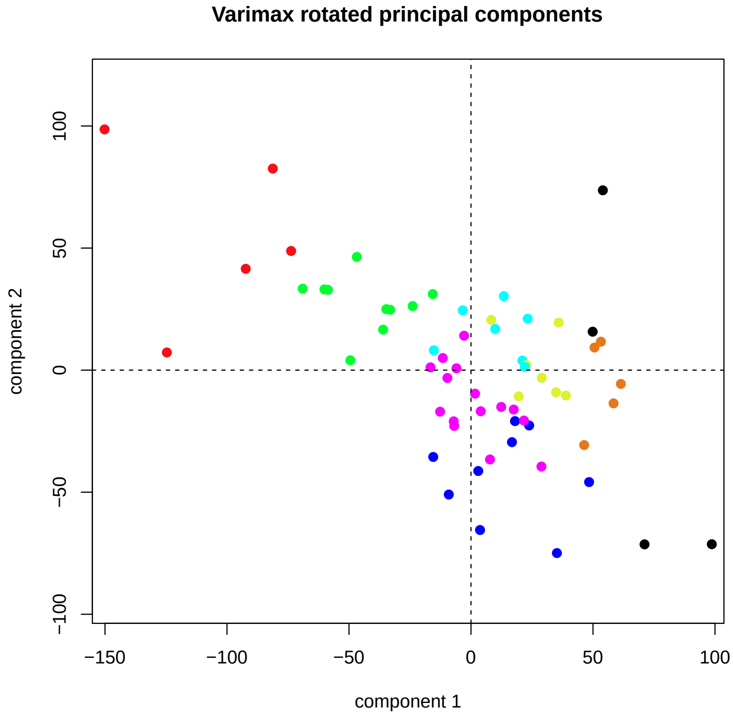

Figure 5: *Plots of the first two principal components labelled by groups suggested by cluster analysis. Points labelled black are outliers.*

treated as part of a group to aid interpretation.

Separation of the groups suggested by cluster analysis, viewed in the two-dimensional PCA plot, is not perfect. This is both a consequence of the fact that the plot is an approximation, and of 'algorithmic variation' in the results provided by cluster analysis and interpretation (choice of number of groups) placed on the results obtained. All this is by way of saying that there is nothing 'given' about the groups shown – even if they have been suggested by a statistical technique – they are the starting rather than end-point of analysis.

Often the first principal component will capture the most important aspect of variation in the data. In Figure 4 the films labelled 1, 14 and 37 are *At the Altar*, *Her Terrible Ordeal* and *The Lonedale Operator* for which graphs of the SL variation over the duration of the film are shown in Figure 3. They are among those coloured red in Figures 2 and 5 and have been characterised, a little crudely, as having a slow/fast editing pattern. Films 5, 10 and 49 in Figure 4, *The Renunciation*, *The Light That Came* and *The Mender of Nets* are also illustrated in Figure 3 and coloured blue in Figures 2. *The Light That Came* is in the bottom-right corner of Figure 5 and coloured as an outlier. These films have been characterised as having more of a fast/slow cutting pattern, and as a first attempt at interpretation the diagonal from the top-left to bottom-right of Figures 4 and 5 might be thought of as representing a 'stylistic' continuum running from a slow/fast, [SF], to fast/slow [FS] pattern.

This works up to a point but is too simplistic. If it is allowed that films like *Her Terrible Ordeal* can be described as having a reverse "J"–shaped cutting pattern and films like *The Renunciation* a "J"–shape, then it might be expected that some films between these extremes will have a "U'–shape. In fact, quite a few do, but there is a lot of complexity not captured by this reverse "J"–" "U"–"J" 'linear' characteristion of a 'cutting-style' continuum. It may be as well to clarify what is to be understood here when judgments about 'style' or 'shape' are being made. Appendix 2 shows plots similar to those in Figure 3 along with superimposed smooths of the data. It is these smooths that are being used as the basis for summary decriptions of style, and a selection is presented in Figure 6.

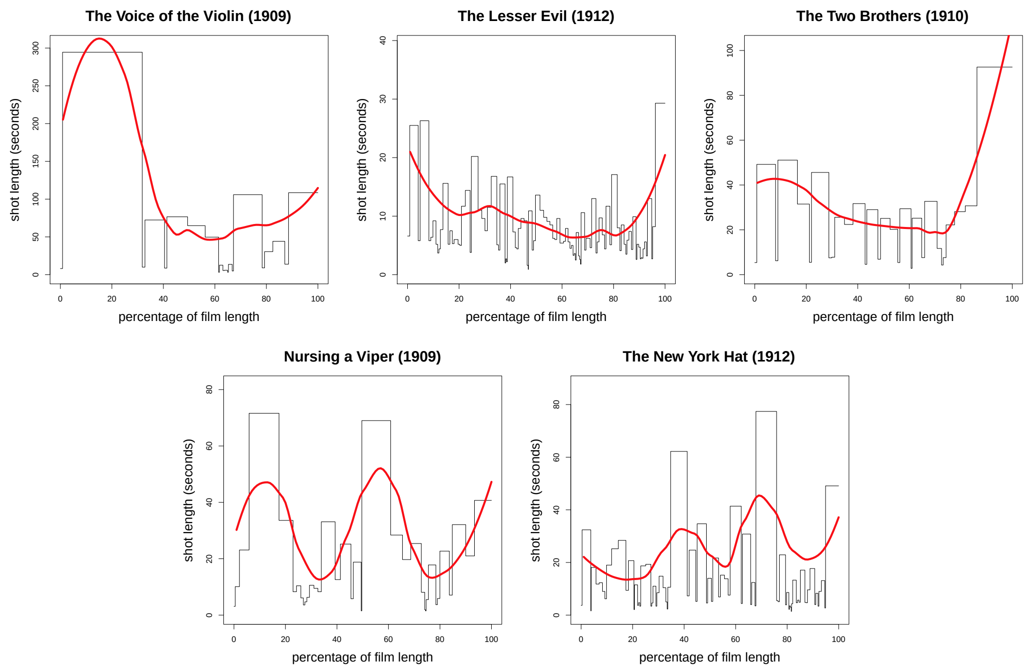

Figure 6: *Plots of SL against film duration, scaled to the interval [0, 100], with superimposed loess smooths based on partition SLs. The films have been chosen to illustrate a range of 'typical' or extreme shapes.*

Loess smooths with a span of 1/2, and otherwise the defaults in the `loess` function in `R` have been used. Caveats will shortly be entered about this choice, but initial discussion is on the basis that this choice is made. *The Voice of the Violin* is an example of what is being called reverse "J" shaped with an [SF] pattern. It isn't quite this, having a short opening title shot that produces what might be called an S[FS] smooth. This is fairly characteristic of many films where just one or two shots at the beginning or end of a film add an extra 'element' to the smoothed pattern. Later plots group films according to what I have judged to be the dominant pattern, within square brackets, so that plots can look 'messy' at the ends, where there is not a common pattern[12].

The *Lesser Evil* has what I am calling a "U" shaped [SFS] pattern. It has some

[12] I have wondered whether, for this kind of exercise, an analysis that concentrates on, say, the central 90% of a film might provide cleaner results, but haven't pursued this.

small bumps in the F section, and some further discussion of this is needed shortly. *The Two Brothers* might also be described as [SFS] but differs from the previous example in the prominence of the final S section. It might thus be labelled S[FS] to emphasise this; in adopting this way of describing a smoothed pattern one is attempting to impose discrete categories on a continuum, so distinctions between, for example, [SFS] and S[FS] are subjectively determined.

The bottom two films in Figure 6 are less simple to describe. Both have two clear peaks but those for *The New York Hat* are shifted to the right in comparison to those for *Nursing a Viper*. The patterns are sufficiently different, and repeated in other films that it seemed worth using them as a basis for classification, though the variable location of peaks and troughs, and different end behaviours complicates matters. *Nursing a Viper* has an [FSFSFS] pattern; *The New York Hat* an [SFSFSF]S pattern. A fuller discussion of the commonality and variety in films with these kind of patterns is provided in Appendix 2.

It is possible to disagree about interpretations of smooths in terms of [FS] stucture, but the more important caveat concerns the chosen level of smoothing. To my eye the smooths for four of the five films in Figure 6 seem reasonable, and close to what might be produced if 'freehand' smoothing was attempted. The smooth for *The Lesser Evil* seems less satisfactory and perhaps over-smooths. It's possible to read an [SFSFSFS] pattern in the data and a somewhat smaller span than 1/2 would be needed to show this. What is done in the remainder of the paper is to take the initial classification, based on the CFDs, as a starting position, and then to modify groups based on inspection of the smoothed SLs to try and identify films that show some resemblance in their cutting patterns to each other and one of the basic patterns suggested so far. Other than the extra effort involved there is no reason why different degress of smoothing shouldn't be applied to different films; however a span of 1/2 seemed reasonable for a majority of films and is that used for all of them[13].

That the analysis to date forms a reasonable basis for subsequent group modification can be seen in Figure 7 where, allowing for the compression of the plot and overprinting of lines, the groups identified by the different colours 'block' together. More detailed analysis, after group modification, is pursued below.

Temporal patterns in the groups, if any, are explored in more detail in Section 5, but one aspect of this is worth noting here. The 1909 films are numbers 1–14. These typically have fewer shots than films from later years; had films with 30 or fewer shots been excluded from analysis (rather than 20) 7/14 of these would have been omitted. If a bounding line is drawn through the extreme points of the plot in Figure 4 four of the 1909 films lie on it (1, 2, 7, 10), and another five (3, 4, 5, 6, 14) lie at or close to the edges of the point scatter when stripped of the more obvious extremes . That is, the 1909 films are a highly heterogeneous group, with a majority of the films differing noticeably from each other and from films from other years. This may be at least partly a function of the small number of shots in some films, with some exceptionally long individual shots in several of them, which result

[13]It would be possible to undertake analyses similar to those described, but based directly on smooths of the SLs against film duration. Applying this approach in the manner attempted here for CFDs, raises issues of detail that I hope to treat at greater length in a companion paper.

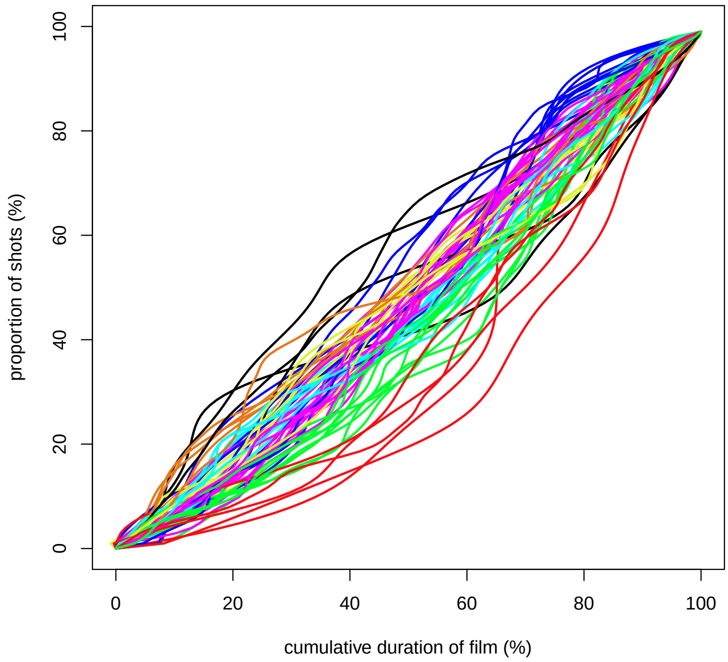

Figure 7: *Smoothed CFDs for 62 Griffith Biographs showing the groups suggested by average-link cluster analysis.*

in patterns much less likely to occur in later films with two to six times as many shots.

4 Aspects of interpretation

Figure 7 is a nice and compact way of illustrating group differences, but not that easy to read in terms of what the distinctions are between the proposed groups. For any proposed group Brisson's partitioning method can be used to obtain a group average curve, and the smoothed plots for the SLs of individual films can be compared with this to see how coherent the group is. Previous applications of the methodology are affected by the absolute values of the SLs involved, and this may be what's wanted, but if 'shape' rather than 'scale' differences are of interest rescaling is desirable.

Plots similar to those in Figure 6 were examined for all films and the groups, originally determined as in Figure 5, were modified in an attempt to define groups where the films involved more nearly matched the average profile for the group than was originally the case. Figure 8 is the result for some of the more interesting groups The thicker coloured lines are the group averages and the thinner black ones the individual films. This is a highly subjective exercise; some films better match the average profiles than others, and transferring some films from one group to another would not affect appearances much. The raw data is available in Appendix 2 for

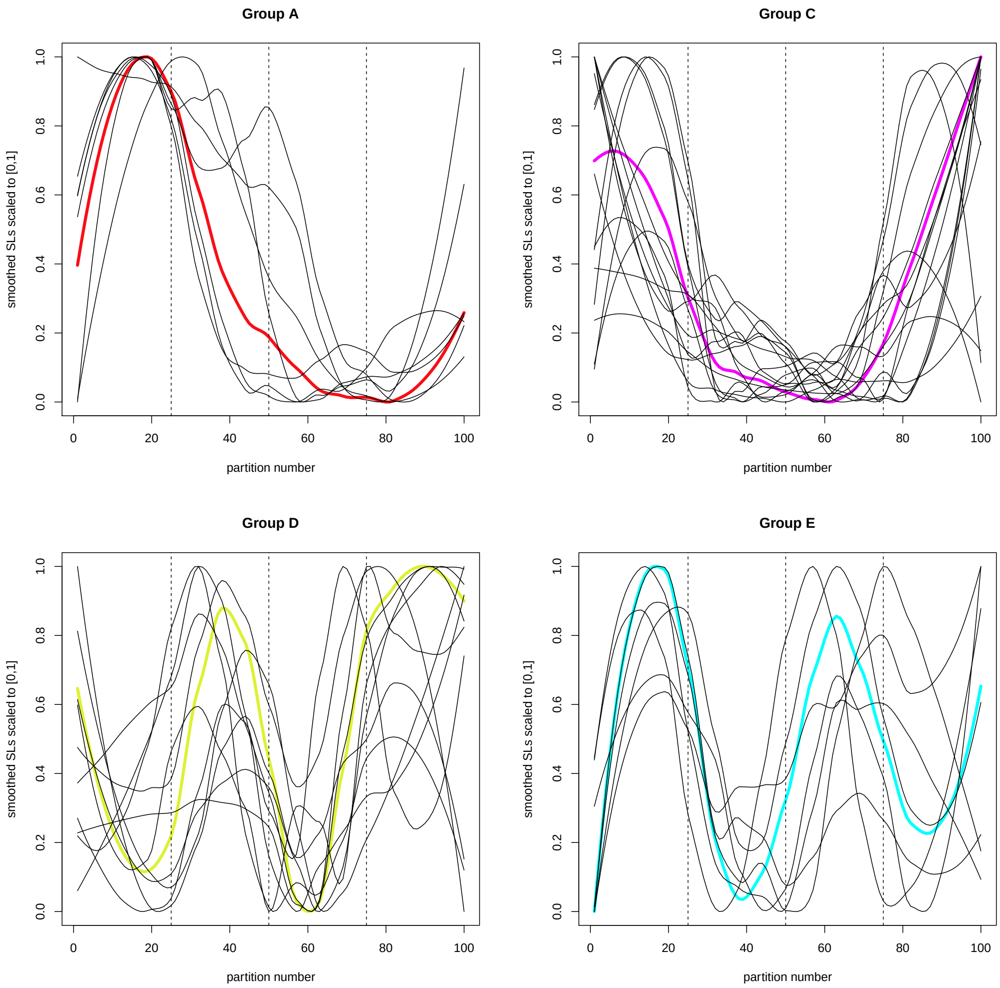

Figure 8: *Smoothed plots based on partition SLs for individual films for Groups A, C, D and E.*

those who wish to make their own judgment. The distributions have been scaled to lie in the interval [0, 1], in a similar way to Redfern (2013); another possibility would be to standardize the SLs to zero mean and unit variance[14].

As Yuri Tsivian observed in his introduction to the discussion *Question 3: Looking for Lookalikes?*, all films have different SL structures, but some are less different than others. It's easy enough to pick out films in the individual panels of Figure 8 that differ from each other, and from the average; but the more important point is that they should be more similar to each other than they are to films in other panels. One way of apprehending that there is pattern to discuss and compare is to notice that there are substantial areas of empty space in the different plots. If the collections of curves are surrounded by an imaginary 'envelope', noticeable areas of space occur for Group A below the envelope between the 0–40% marks, and above it between 60–100%. For Group C the 20–80% region stands out, with small spaces below the envelope at either end. The more complex Group D has regions of space

[14]Loess smoothing with a span of 1/2 has been used. This is in contrast to the occasional use of sixth-degree polynomials for the same purpose

above the envelope between the 0–20% and 40–60% marks, and below it between 20–40% and 70–100%. Group E is a partial inverse of Group D with space below the envelope between the 0–35% marks, and above it between 20–55%; the second half is messier, but with evidence of peaking between 55–75%. The areas of empty space define where the peaks and troughs occur for the different groups, and these clearly differ in both number and location[15].

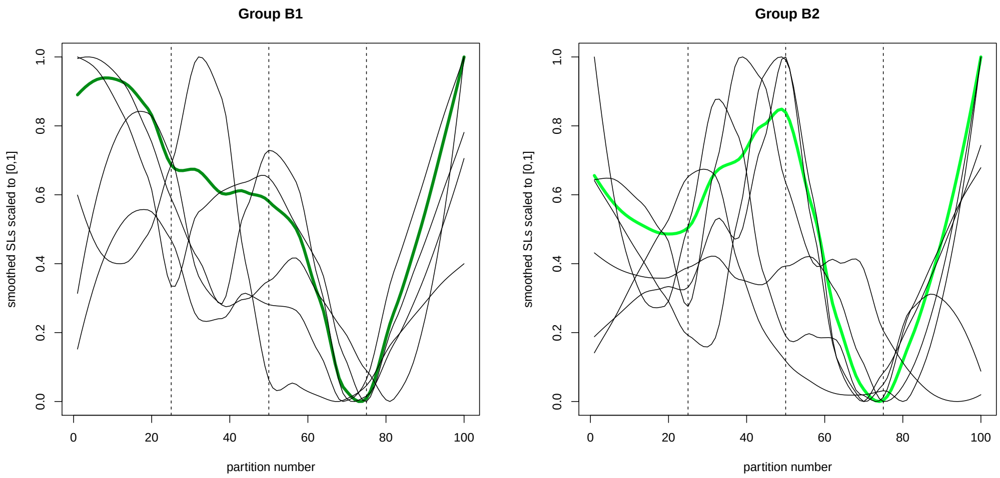

Figure 9: *Smoothed plots based on partition SLs for individual films for Groups B1 and B2.*

Figure 9 shows similar plots for what we shall call Groups B1 and B2. What they have in common, and the reason they have been grouped together, is a pronounced trough at about three-quarters of the way through. The first half of each group does not show an especially consistent pattern, but those in Group B1 tend, like those in Group A, to have slower cutting in the first 30% or so, while the peaks (or slower cutting), in Group B2 tends to be more evident about half way into the films. More so than in Group A there is a pronounced slowing down in the last 20% of the films, though the phenomenon is not absent from Group A films.

The smoothed average curve for the full sample of Biograph films using Brisson's methodology is shown as the thick black line in Figure 10, along with the curves for Groups A to E. Troughs correspond to faster cutting. The vertical dashed lines indicate the first and third quarters of film duration. Groups A, B1 and B2 have their fastest cutting at about 80%. Group C has a fairly flat region of faster cutting, but is reasonably fast at about the 60% point, which is also the position of some of the faster cutting for Group D. These groups account for nearly 70% of the sample and, along with some films from other groups, it their averaging that accounts for the distinct trough in the Griffith average at about 70%.

The differences between groups are clear, as is the fact that most of them differ in noticeable ways from the overall average curve. The relatively smooth decline in the average curve between about the 15% point and the trough at about 70% arises because some of the more obvious variation in the group averages between these

[15]The [0, 1] scaling 'magnifies' differences between peaks and troughs compared to the impression obtained when an 'absolute scale' is used for the smooths, as in Appendix 2 where further comment is provided.

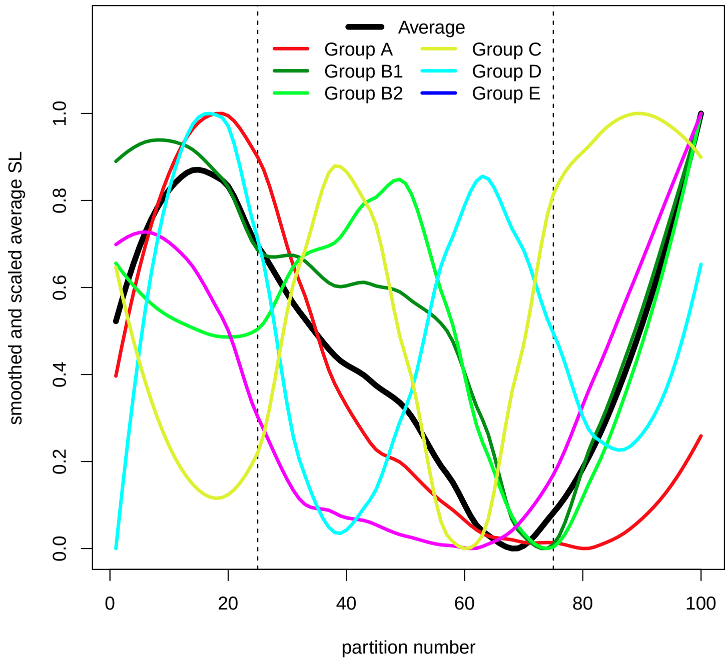

Figure 10: *Averaged localised ASL patterns, scaled to the interval [0, 1], across partitions for six groups proposed for the Griffith Biographs sample.*

points 'cancels out' (e.g., Groups D and E) and is overlaid on a pattern showing a general acceleration in cutting rate for other groups. A general point that can be made here is that the average curves sometimes only highlight what is common to the curves from which they are derived over part of the range (e.g., Groups B1 and B2).

Films so far not accounted for are shown as Groups F and G in Figure 11, though the term 'Group' is something of a misnomer as they lack any kind of obvious coherence. To take Group G first, it suggested one of the groups in the average-link cluster analysis, but I reassigned some films to other groups on the basis of the visual evidence of plots in Appendix 2. Essentially the films are outliers rather than a coherent group. They are being picked out as a group in the cluster analysis, as can happen, because they are rather different from everything else, rather than because they are similar to each other. Such features as some have in common are noted in discussion of the individual films in Appendix 2.

Group F was adopted wholesale from the cluster analysis.. The average pattern suggests a fast beginning and relatively slower middle or end, but the slowest cutting can be anywhere between 25–80% of the way through, and a more detailed discussion is deferred to Appendix 2.

To conclude this section Figure 12 shows a plot of the first two principal components, labelled according to the modified grouping described above. It may be compared with the original grouping suggested by the cluster analysis in Figure 5.

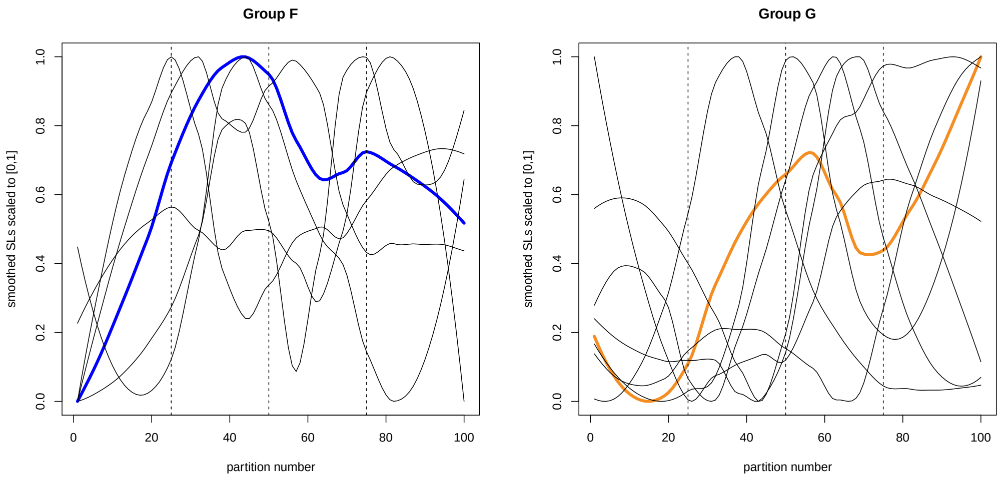

Figure 11: *Smoothed plots based on partition SLs for individual films for Groups F and G.*

Two films placed in Group A might equally well have been put into Group B1. Bearing in mind that Group G consists of outliers the grouping looks sensible – there is overlap between some of the groups, but this was also the case with those defined by the cluster analysis. Group B2 seems the least satisfactory but plots using the third principal component (not shown) separate it out better from those groups with which it is intermingled in Figure 12.

5 Exploiting other information

The statistical patterns in the data, particularly Figure 4, can be exploited to investigate variation using other ways of categorising films. The two examples below, using information on crosscutting and on date, are illustrative. Classifying by 'genre' would be another possibility. Brief explorations, not reported here, using the typology given in Henderson (1970), hasn't shown anything obvious to report, and neither are there obvious differences between the films made in California and those made elsewhere.

Crosscutting

Yuri Tsivian (pers. comm.) has drawn up a (partial) list of films that have have been suggested in the literature as 'crosscutting' films. These have been categorised according to the purposes to which crosscutting has been put. The most common category is 'suspense', often a 'rescue' film, and to these films have been added others that occur in the list of 'rescue' films in the relevant Cinemetrics lab. These are highlighted in red in the left-hand panel of Figure 13[16]. Those highlighted in blue are films where the use of crosscutting is for what Tsivian describes as

[16]Some 'classic' crosscutting films from Tsivian's list are omitted because they do not meet the crireia for duration and/or number of shots used here.

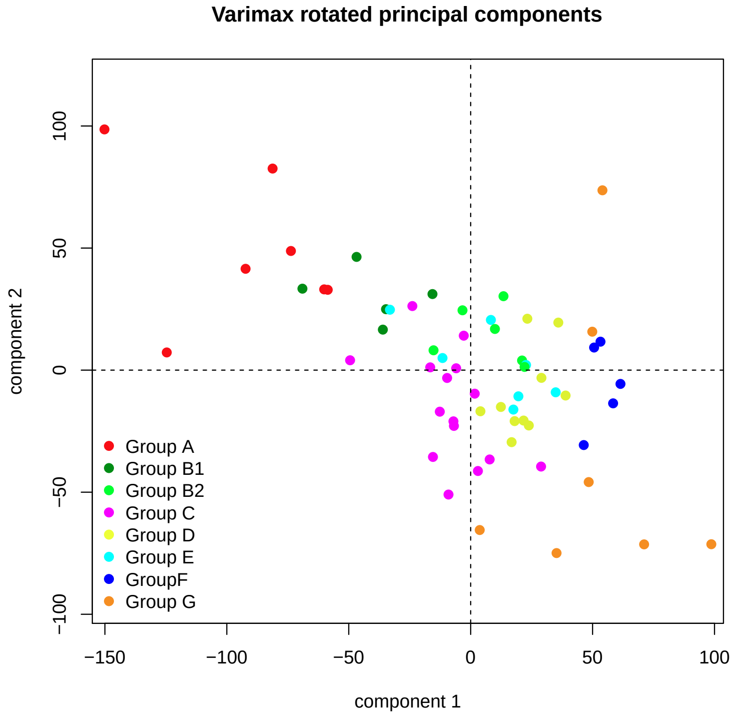

Figure 12: *Plots of the first two principal components with points labelled by their modified groups.*

'rhetorical' purposes (e.g., *A Corner in Wheat*). Isolated occurrences of three other crosscutting films, each differently classified, are shown as green.

The suspense/rescue films mostly occur, as one would hope, in the left-half of the plot associated with faster cutting in the second half of the film. There are three exceptions to this, just to the right. If a gross simplication is permitted here, the shape of a typical suspense/rescue film exhibits a reverse-"J" or skewed-"U"-shape, with the trough (on average) occurring about three-quarters of the way through (some a bit earlier). The three exceptions to this, *The Death Disc*, *A Child's Stratagem* and *The Modern Prodigal* have a passage of faster cutting in their second half, but also other sections with comparably faster cutting. The other crosscutting films to the left also have passages of fast cutting, often the dominant one, that occur earlier than is typical for the majority of suspense/rescue films. Given the provisional and incomplete nature of the listings currently available, and small number involved, not too much should be read into the analysis at this point, though the placement of the suspense/rescue films would seem to make sense. The question is raised, perhaps, of why films that plot close to crosscutting ones are not so classified, or whether, indeed, they might be regarded as such[17].

[17]For example, *The Voice of the Violin* (labelled 2 in Figure 4) plots where it does because the first third of the film is largely accounted for by a single shot, so it can't do much other than get 'faster' after that. It involves a 'rescue' as well.

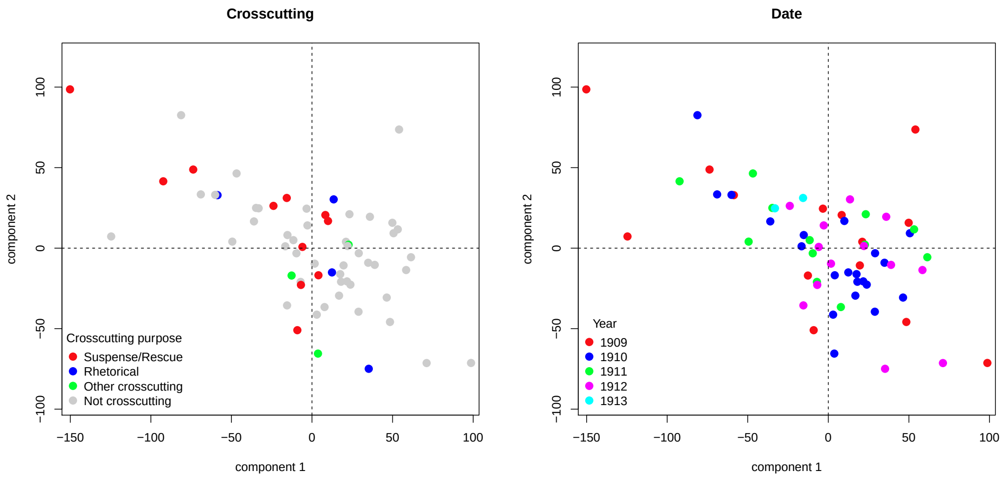

Figure 13: *Plots of the principal components, as in Figure 4 labelled by crosscutting type and date.*

Date

The right-hand panel of the figure shows films categorised by date[18]. That the earlier 1909 films are scattered around, with many at or near the periphery of the plot has already been noted. The two 1913 films plot close together, but not distinctly from films of other years. Those years that are well-represented are scattered throughout the plot. The south-east quadrant of the plot is dominated by two years but these, 1910 and 1912, are not temporally contiguous and perhaps not much can be read into this

The purpose of this paper can be viewed as an attempt to classify films by an aspect of 'style' that can be thought of as the 'shape' of a film as characterised by its cutting pattern. The diagonal band evident in Figure 4 might be interpreted as a 'stylistic gradient', though there is considerable scatter particularly in the right-half of the plot. Since films from most years occur in most regions along the gradient (or where they do not, the years represented are not contiguous) this might be interpreted as showing that, in terms of cutting patterns, Griffith's style did not evolve much over the period 1909-1913. The positive slant that could be put on this is that the cutting styles – as evidenced by pattern as opposed to content – that Griffith exploited were present from 1909.

Since one of the points of Griffith, and one of the reasons he has been studied so much, is his 'evolution', concluding that at least one aspect of style did not evolve might seem unsatisfactory. The most common quantitative measure of style that has been used in the cinemetrics literature is the average (or median) SL. In this respect Griffith clearly does evolve, the earlier films tending to have fewer shots and thus – since they are of about the same length – larger ASLs than the later films. Is this compatible with the argument that cutting style as evidenced by pattern does not evolve?

[18]These are dates of production, as given in Henderson (1970). Where these differ from other sources, including some on the Cinemetrics Measurement Database, this seems to depend on whether production or release dates (also give by Henderson) are being used. For present statistical purposes these differences are of no great moment.

For Figure 14, four pairs of films have been selected, (3, 57), (12, 37), (31, 58) and (29, 53), that plot close together and are well-separated in time, but show fairly similar smoothed cutting patterns. The numbering corresponds to their temporal sequencing within the sample of 62, according to the date of production. The number of shots, and consequently the ASLs since the films are of fairly similar duration, vary widely; the films, with date, number of shots, and ASL are *A Son's Return* (1909; 34; 19.7), *The Painted Lady* (1912; 77; 9.6), *A Corner in Wheat* (1909; 32; 25.7) and *The Lonedale Operator* (1911; 107; 9.5), *Winning Back His Love* (1910; 44; 18.3), *The Musketeers of Pig Alley* (1912; 88; 11.2), *The Golden Supper* (1910; 57; 15.7) and *The Lesser Evil* (1912; 105; 7.4).

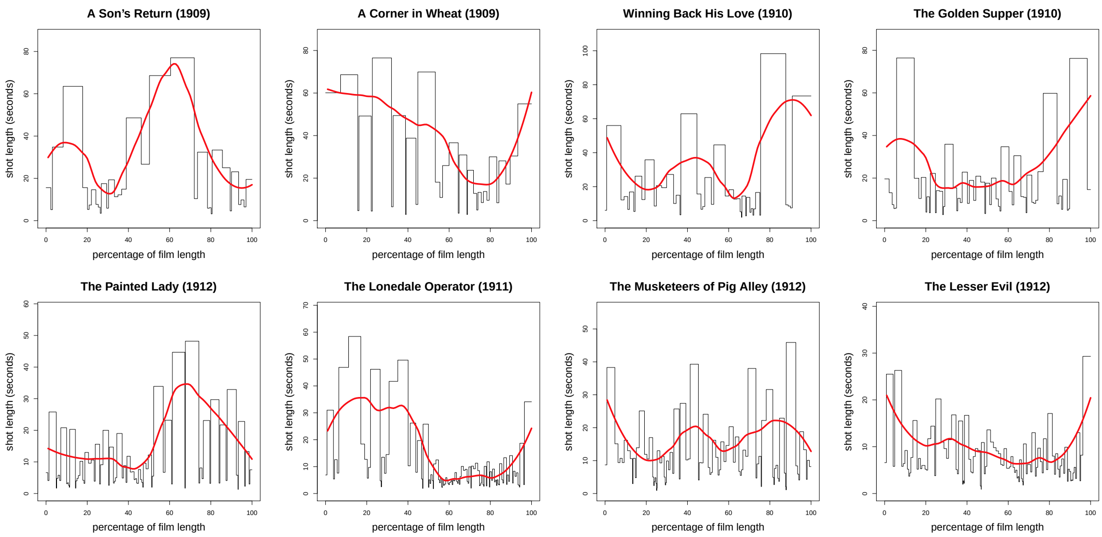

Figure 14: *Films from different years showing similar cutting styles within columns.*

It's possible to argue in at least two ways here. It has already been noted that the earlier films, and particularly those from 1909 in this sample, mostly have relatively few shots; that this is in some ways restrictive in the patterns that can be produced; and that, partially because of the use of shots of very long duration, many of these films show up as (near) extremes in Figure 4. It might be argued that the increased number of shots that tend to be used over the 1909–1913 period opens up more possibilities in the way that they can be deployed, and that 'stylistic' development in the cutting patterns used is inevitable. See the discussion of Group C in Appendix 2 for further comment on this.

What Figure 14 is intended to show, however, is that there is considerable variation in the broad patterns evident in the earlier films (the top row) that is replicated in later films with between two and three times as many shots and much smaller ASLs. All the groups proposed in Appendix 2 for which some kind of coherence is claimed contain films from at least three of the years 1909–1912, and there is no pattern to which the adjective 'early' or 'late' is readily applied.

That films from any one year are spread across the 'style continuum' – or equivalently, that 'style categories' contain films from several years – does not preclude the possibility of an association, possibly weak, between style and date. Investigating this statistically has its problems, not least that the sample available is not a random selection and probably not even 'representative'. Formally, it is possible to tabulate date by 'style group', but testing for association with a chi-squared test runs into problems with small expected values. The table was examined informally using correspondence analysis[19], but nothing worthy of note emerged.

6 Discussion

Introduction

It has, I hope, been made clear that this study is experimental; 'style groups' have been proposed as an aid to discussion but there is no pretence that these are 'definitive'. Appendix 2 shows plots of SLs for all the films in the sample by group, along with smooths of them, so readers can make up their own minds about how well the groups 'hold up'.

It will be clear that I find some more 'coherent' than others, and instances are noted in Appendix 2 where individual films don't really fit the more general pattern and/or might be reclassified. I have resisted (for the moment) an attempt to refine the classification proposed, though it could be done, and in more ways than one. One thing that I think is clear is that the average for Griffith's Biograph films, used to illustrate the Brisson methodology on the Cinemetrics site, has limitations as an overall summary. The trough at about 70% of the way through, associated with faster cutting, is a quantitative 'compromise' that captures a qualitative feature characterising many, but by no means all, of the films. That is, a lot of films do have a trough in the second half of the film, but there is a lot of variation about the 70% mark. Elsewhere, the curve is smoothing out a lot of interesting variation that differentiates films. The main emphasis of this paper is that it is desirable to separate films into groups with reasonably differentiated cutting patterns before employing averaging methodology. The main claim is that statistical methods can be used to provisionally define such groups with at least partial success.

The remainder of this discussion touches on issues that have occurred to me in writing this paper which might repay further study, but would lengthen the paper too much to pursue here in detail.

Comparisons involving small numbers of shots

The question of what to do with films with a 'small' number of shots, mostly from 1909, needs revisiting. Those with 20 or fewer shots were excluded from analysis, but I have had cause to notice and occasionally comment on several 1909 films with not much more than 20 shots that stand out for various reasons. The fundamental idea that underpins the analysis here, and which also motivates the Brisson/Tsivian analyses (they raised similar concerns), is that you can validly compare the *shape* of cutting patterns independently of the number of shots involved (the present study

[19]Correspondence analysis, like PCA, ia another widely used method of multivariate analysis for reducing a table of data to a picture, and there are close mathematical connections between the techniques. It is typically applied to tables of counts. Baxter (2012) has applied the methodology to tables of shot-scale categories by film, and Redfern (2012) to tables of film genre preference by age/gender categories.

attempts to control for duration, which is another potentially complicating factor). My current feeling is that you can, once you have a sufficiently 'large' number of shots, but I wouldn't care to define 'large' except that I wonder if some of the 1909 films used here qualify.

Some of the previous comment may appear contradictory. On the one hand, attention has been drawn to the unusual nature of patterns for many of the 1909 films. On the other hand, Figure 14 does provide examples of 1909 films which it is claimed have similar patterns to somewhat later films.

It is obvious that the more shots you have available the more variable the cutting patterns in films can be. The essence of the methodology used in this paper is to smooth out a lot of the detail, which may eliminate or reduce what may be interesting 'filmic' features (unusual beginnings/ends; deliberate exploitation of shots of unusual duration etc.) to reveal the pattern (whose existence is an implicit assumption) around which a film is structured. With a film of one shot there is no scope for variation in cutting pattern, since only one is possible; you can begin to talk about 'shape' with a film of two shots, but it isn't that interesting; at what point the number of shots allows the emergence of the range of patterns that you might reasonably look for in films with many more shots is the point at issue.

The terms 'fast' and 'slow' have been used rather loosely in earlier discussion, usually in connection with the appearance of troughs in smoothed graphs (but also the raw data) with 'slow' being passages of films that are clearly not fast, sometimes, but not necessarily, associated with a peak[20]. The terms are clearly being used in a relative sense here, and it can get quite complicated. For example, two passages in a film associated with troughs separated by a clear peak might be categorised as 'fast' but they may, or may not, exhibit very different cutting rates (that is, one 'fast' passage may be a lot slower than another).

That is, for the purposes of comparing pattern in the form of 'shape', 'size' factors associated with the fact that fast and slow cutting-rates can be thought of in absolute, as well as relative, terms are being deliberately ignored. Two films might show a similar pattern in terms of their shape, including the alternation and positioning of fast and slow passages, using these terms in a relative sense, but the cutting in one film may generally be much faster than that of the other. Whether this should matter in terms of what one might be attempting in comparing cutting styles is not, perhaps obvious. Is it reasonable, for example, to suggest that – ignoring variation at the beginning – the pairs *A Son's Return* and *The Painted Lady* in Figure 14 exhibit similar cutting styles, or that *A Corner in Wheat* and *The Lonedale Operator* do so?

Partitioning a continuum

A first attempt at partitioning the map shown in Figure 4 produced groups (and outliers) suggested by cluster analysis. Subsets of the films at either end stood out a little, but divisions in the central part of the plot were a little arbitrary. Inspection of *averaged* patterns in these groups (not shown) admitted a plausible

[20] I am using the term 'peak' here to refer to a noticeable bump in a plot. This excludes films with, for example, slow beginnings and ends where the pattern is one of descent or ascent from or to the end-points. The term 'peak' could be redefined to include such end-points as 'peaks' if you wished.

interpretation, but inspection of plots such as those in Figures 8 and 9 suggested that the placement of some films was unsatisfactory. Films were reallocated between groups on the basis of a subjective evaluation of their SL distributions and smooths to obtain the groups eventually arrived at.

My view on this is that some degree of subjectivity in group definition is not only inevitable, but also desirable. A consequence is that different analysts may arrive at different groups, and some might even conclude that the attempt to define useful groups is an unattainable aim. Multiple narratives do no-one any harm. The idea that statistical methodology, however complex, provides an 'objective' method of classification, the results of which everyone should be happy with, is not sustainable. If the effort is thought to be worth making in the first place, it provides a sensible starting point rather than an end.

Principal component plots, and other graphical displays such as the dendrograms associated with cluster analysis, can be thought of as Procrustean beds on which the films are forced to lie. Some (hopefully most) do so with little discomfort and sleep well; others, however, may be forced into close proximity to bed-fellows they are not comfortable with, and can, at the behest of the analyst, move to a more comfortable location, which may include splendid isolation.

Any partitioning of a continuum thus has an element of arbitrariness about it. Some kind of pattern undoubtedly exists in the films studied here, and it would be possible to discuss this without attempting to group them. Grouping does, however, provide a convenient aid to discussion, if only to dispute it or demonstrate why it doesn't work well. The evidence, in the form of the raw data and smoothed plots, is there for anyone to rearrange a suggested classification as they see fit.

Quantitative versus qualitative classification

For most of this paper a quantitative approach to exploring and categorising cutting patterns has been explored. It is not prescriptive; the numerical results have to be interpreted; any 'final' classification is the outcome of an iterative procedure; and, as the previous section suggests, judgments about similarities can not only be involved in this process but are essential. The statistical methodology is only a tool.

In the discussion of Figure 6 a more 'qualitative' approach to classifying films in terms of their fast/slow [FS] patterns was broached. This can be done using smoothed curves showing the SL pattern over time. Application of this kind of approach raises issues of both a 'filmic' and statistical nature. An obvious statistical one is the degree of smoothing to use. Under-smooth and there is not much point to it. Over-smooth and detail associated with, for example, short bursts of 'short' shots or isolated occurrences of very long shots may be smoothed out, possibly to the point of near-invisibility. It is arguable that this is the point of smoothing, but if the desire is for a 'stylistic comparison' to give weight to these kind of features then the methods explored here may not be especially suitable. It is, I think, a 'filmic' rather than statistical issue.

The same is true when it comes to the treatment of the start and end of a film. Smoothing methods are generally less reliable near the end points of a set of bounded data – smoothing takes place over contiguous sets of shots, and you run out of these at the ends. In the analyses of earlier sections, in making assessments about the similarities of cutting patterns I have on occasion – sometimes silently and sometimes not – mentally discounted variations in end behaviour. In statistical terms this can be justified as a focus on the more reliable parts of smoothed plots and might have been pursued more formally. It is possible that, in any case and from a 'filmic' point of view, one might wish to accord beginnings and ends separate treatment, perhaps using behavior in these regions as a basis for sub-dividing groups of films otherwise categorised as similar. In terms of FS patterns, for example, this might involve categorisation into different basic 'units' of pattern, or 'motifs', [SFS], [SFSF] etc. for, say, the central 80–90% of a film and then 'attaching' end behaviour where this varies from the basic pattern, F[SFS] of [SFSF]S for example.

This idea needs to be thought through in more detail than there is space for here, but one further and fairly obvious consideration can be mentioned. Thinking only in terms of films where the dominant feature is a trough or passage of fast cutting, call this an 'anchor point'. Films might have a similar basic pattern, [SFS] for example, but the 'anchoring' F can in principle appear early, late or centrally. It is thus unlikely that one would wish to regard all [SFS] patterns as the same, and an initial categorisation would need to be sub-divided using information on the positioning of the dominant features. This, as with the other issues raised, is left for the moment except that additional commentary in Appendix 2 some use is made of the idea with reference to individual groups.

References

Baxter, M. (2012a) Picturing the pictures. *Significance* **9**, 5-9.

Baxter, M. (2012b) *Cinemetrics Data Analysis.*
http://www.mikemetrics.com/#/cinemetrics-data-analysis/4569975605

Cox, D.R. and Donnelly, C.A. (2011) *Principles of Applied Statistics*, Cambridge: Cambridge University Press.

Henderson, R.M. (1970) *D.W. Griffith: The Years at Biograph.* New York: Farrar, Straus and Giroux.

Ramsay, J.O. and Silverman, B. (2002) *Applied Functional Data Analysis: Second Edition.* New York: Springer.

Ramsay, J.O. and Silverman, B. (2005) *Functional Data Analysis: Second Edition.* New York: Springer.

Ramsay, J.O., Hooker, G. and Graves, S. (2009), *Functional Data Analysis with* R *and* MATLAB. New York: Springer.

Redfern, N. (2012) Correspondence analysis of genre preferences in UK film audiences. *Participations* **9**, 45-55.

Redfern, N. (2013). Time series clustering and the analysis of film style.
http://www.cinemetrics.lv/dev/on_statistics.php

Appendix 1: The Sample

Index	Title	Date	Shots	ASL	Group
1	At the Altar	1909	28	21.8	A
2	Voice of the Violin, The	1909	22	43.3	A
3	Son's Return, A	1909	34	19.7	G
4	Country Doctor, The	1909	44	10.0	C
5	Renunciation, The	1909	24	29.6	G
6	Strange Meeting, A	1909	26	28.2	B2
7	Hessian Renegades, The	1909	33	18.5	G
8	Lines of White on a Sullen Sea	1909	26	24.3	C
9	Nursing a Viper	1909	36	17.1	E
10	Light that Came, The	1909	35	23.4	G
11	Death Disc, The	1909	25	29.0	E
12	Corner in Wheat, A	1909	32	25.7	A
13	Redman's View, The	1909	25	34.5	B2
14	Her Terrible Ordeal	1909	46	15.1	A
15	As It Is in Life	1910	44	16.2	D
16	Unchanging Sea, The	1910	44	18.3	F
17	Two Brothers, The	1910	30	22.5	C
18	Ramona	1910	68	14.3	E
19	In the Border States	1910	66	13.1	B1
20	Flash of Light, A	1910	45	19.1	D
21	Arcadian Maid, An	1910	47	15.9	F
22	House with the Closed Shutters, The	1910	66	15.0	C
23	Usurer, The	1910	61	17.4	D
24	Modern Prodigal, The	1910	55	14.3	D
25	Rose O'Salem Town	1910	52	16.6	E
26	Oath and the Man, The	1910	64	10.2	B1
27	Fugitive, The	1910	60	16.7	G
28	Child's Stratagem, A	1910	86	7.7	B2
29	Golden Supper, The	1910	57	15.7	C
30	Lesson, The	1910	64	13.1	A
31	Winning Back His Love	1910	44	18.3	D
32	His Trust	1910	58	14.5	B2
33	His Trust Fulfilled	1910	26	25.8	D
34	What Shall We Do With Our Old?	1910	32	25.2	A
35	Heartbeats of Long Ago	1910	77	11.7	D
36	His Daughter	1911	88	10.1	C
37	Lonedale Operator, The	1911	107	9.5	A
38	Indian Brothers, The	1911	55	15.0	B1
39	Fighting Blood	1911	101	6.4	C
40	Last Drop of Water	1911	72	10.8	E
41	Country Cupid, A	1911	69	12.9	F
42	Swords and Hearts	1911	114	8.5	C
43	Adventures of Billy, The	1911	106	8.3	C
44	Battle, The	1911	113	8.6	E
45	Miser's Heart, The	1911	82	11.7	B1
46	For His Son	1911	72	12.2	G
47	Billy's Stratagem	1911	111	7.4	D
48	Sunbeam, The	1911	83	10.6	F
49	Mender of Nets, The	1912	108	6.9	C
50	Girl and Her Trust, The	1912	135	6.6	C
51	Female of the Species,The	1912	73	11.3	C
52	One is Business, the Other Crime	1912	77	11.7	B2
53	Lesser Evil, The	1912	105	7.4	C
54	Beast at Bay, A	1912	105	6.3	C
55	Unseen Enemy, An	1912	130	6.9	C
56	Friends	1912	92	8.1	F
57	Painted Lady, The	1912	77	9.6	G
58	Musketeers of Pig Alley, The	1912	88	11.2	D
59	New York Hat, The	1912	76	12.7	D
60	Burglar's Dilemma, The	1912	86	10.6	B2
61	House of Darkness, The	1913	86	11.6	E
62	Death's Marathon	1913	122	7.2	B1

Table 1: *Films used in the analysis, with suggested 'group' classifications. The ordering of the films, the 'Index' used in labelling some plots, is chronological by production date, as given in Henderson (1970). Films 15–18, 36–40, 49–54 and 61–62 were produced in California.*

Appendix 2: Groups

A more detailed group-by-group analysis is presented below. Plots similar to those in Figure 8 are presented along with graphs of the raw SL data. The aim is to get some sense of how well films are represented by the average pattern for a group, shown in various figures in the main text. Some comment is provided on films that seem 'deviant' in some way, but an exhaustive discussion is not attempted.

The unsmoothed SLs are for the raw data; the smoothed plots have been obtained using partition SLs, which are much more satisfactory for application of this methodology for films with relatively few shots. The defaults in the `loess` function of `R` have been used, except that the default span of 2/3 has been replaced with one of 1/2. The idea is to produce a smooth capable of revealing variation in cutting patterns that does not get too 'bogged-down' with details. Comparing the smooths to the pattern in the raw data this degree of smoothing seems to work reasonably well – to my eye at least – in most cases, and the evidence is available for those who wish to disagree with this judgment.

The rather stern injunction of Cox and Donnelly (2011, p.84) that the interpretation of plots with substantial noise should not be compromised by imposing on them prominent smooth curves' might be borne in mind here. The data – and to my eye, again – are not what I'd regard as too noisy in many cases, and if you want to effect comparisons between graphs without the cue provided by smooths some form of 'mental' smoothing is, perhaps, needed. In Group B1 the pattern you might see for *In the Border States*, for example, might be much that provided by the smooth shown. *The Miser's Heart*, in the same group is, perhaps, a counter-example, the trough at about 50% being completely missed because of the effect of the long shots either side of it. This would be corrected if less smoothing was applied.

This may be the place for a reminder that the grouping proposed is based on (smoothed) *cumulative* distributions modified by inspection of raw data shown in the figures. Analysis based directly on smooths of the SLs rather than their cumulative distributions is a possible avenue for further exploration. This would raise issues about the degree of smoothing to use, touched on earlier, that may be more problematic than they are for CFDs.

Group A

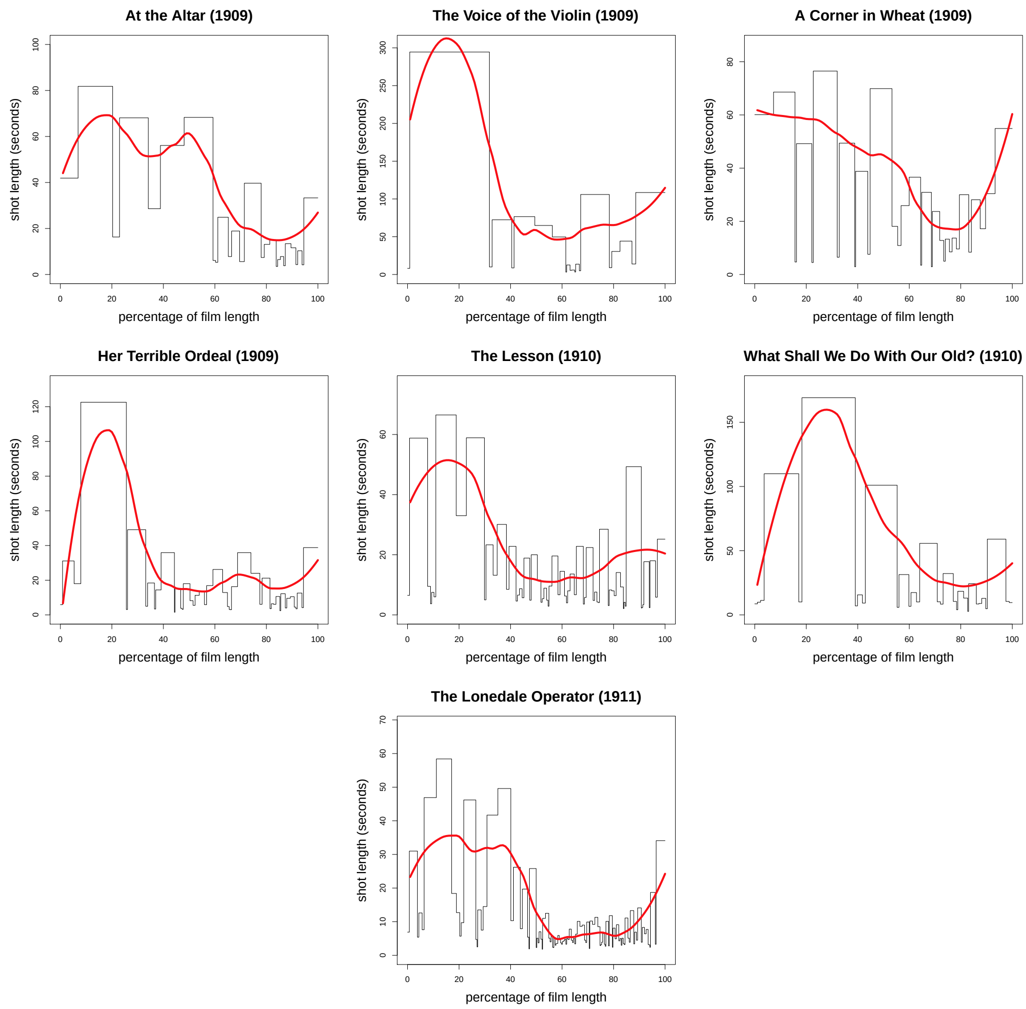

Figure 15: *Group A films* – At the Altar *(1),* The Voice of the Violin *(2),* A Corner in Wheat *(12),* Her Terrible Ordeal *(14),* The Lesson *(30),* What Shall We Do With Our Old? *(34),* The Lonedale Operator *(37)*

The plot of the principal components, Figure 4, shows that this is not a tightly defined group. *At the Altar* and *The Voice of the Violin* might be better treated as outliers, while *A Corner in Wheat* and *The Lesson* might be in Group B1. The dominant motif is [SF] with variants at either end often attributable to a very small number of shots. For example, the smooth for *The Voice of the Violin* could be described as F[SF]S, but the outer F is the product of a single title shot.

Group B1

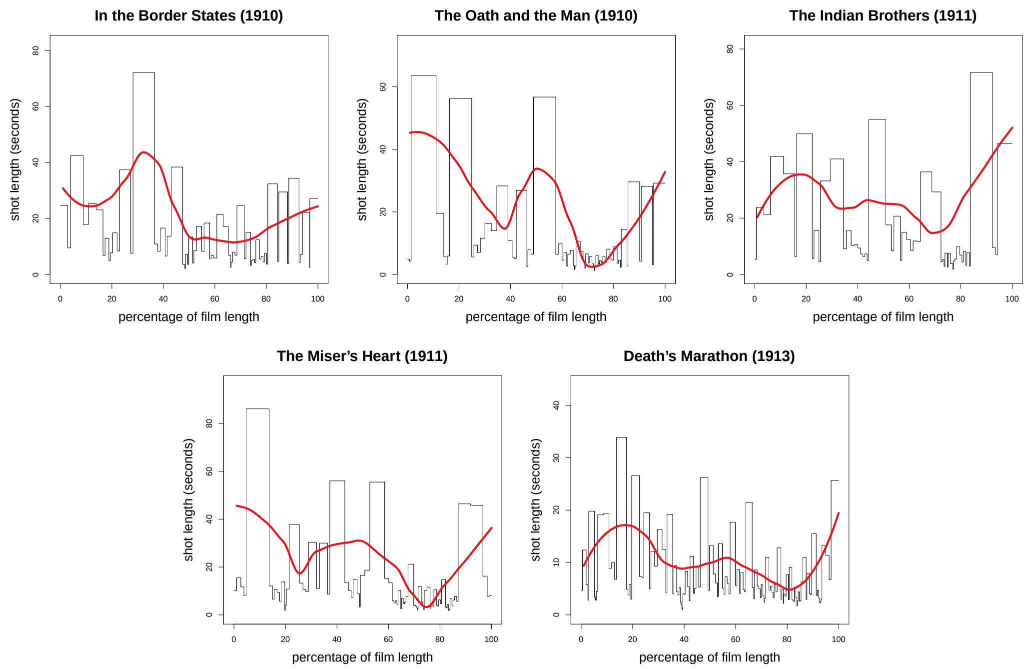

Figure 16: *Group B1 films* – In the Border States *(19)*, The Oath and the Man *(26)*, The Indian Brothers *(38)*, The Miser's Heart *(45)*, Death's Marathon *(62)*

These films have in common, after variable beginnings, an [FSFS] pattern with the final [SFS] mostly occuring in the second half of the film and responsible for the pronounced trough at about the three-quarters mark in Figure 9.

Group B2

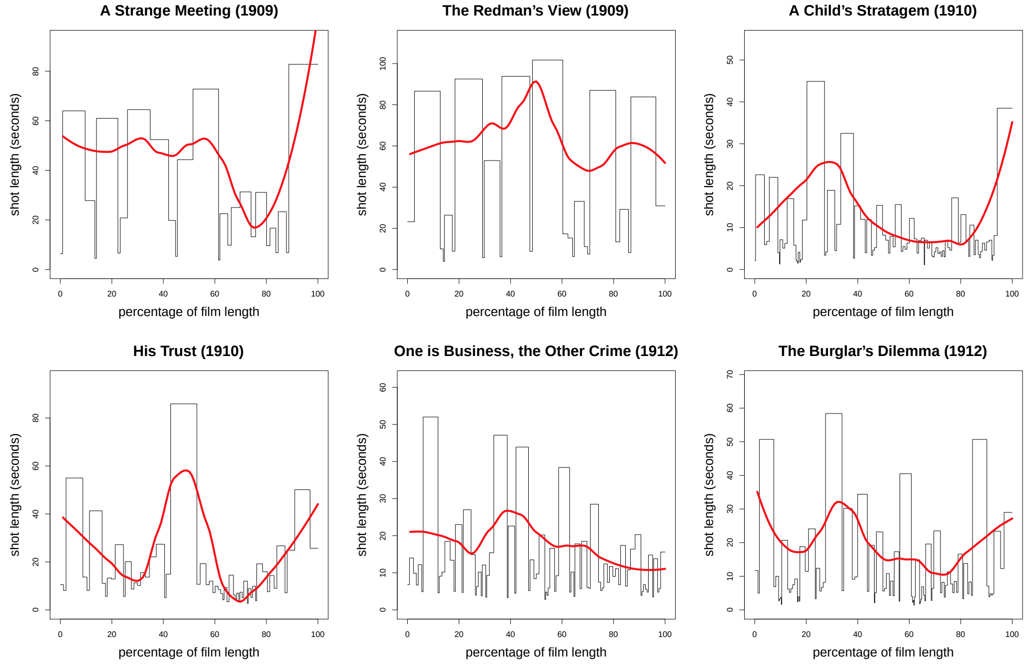

Figure 17: *Group B2 films* – A Strange Meeting *(6),* The Redman's View *(13),* A Child's Stratagem *(28),* His Trust *(32),* One is Business, the Other Crime *(52),* The Burglar's Dilemma *(60)*

With the exception of *One is Business, the Other Crime* an ending [SFS] pattern characterises the smooths, with *The Redman's View* having an additional S at the end. This gives rise to the pronounced trough at about the three-quarters mark in Figure 9. *The Redman's View* has only 25 shots and is an instance where I wonder if the methods used in the paper are particularly appropriate. *One is Business, the Other Crime* is also unconvincing, but presumably is picked up because of the [SFSF] pattern it has in common with other films in its first part.

Group C

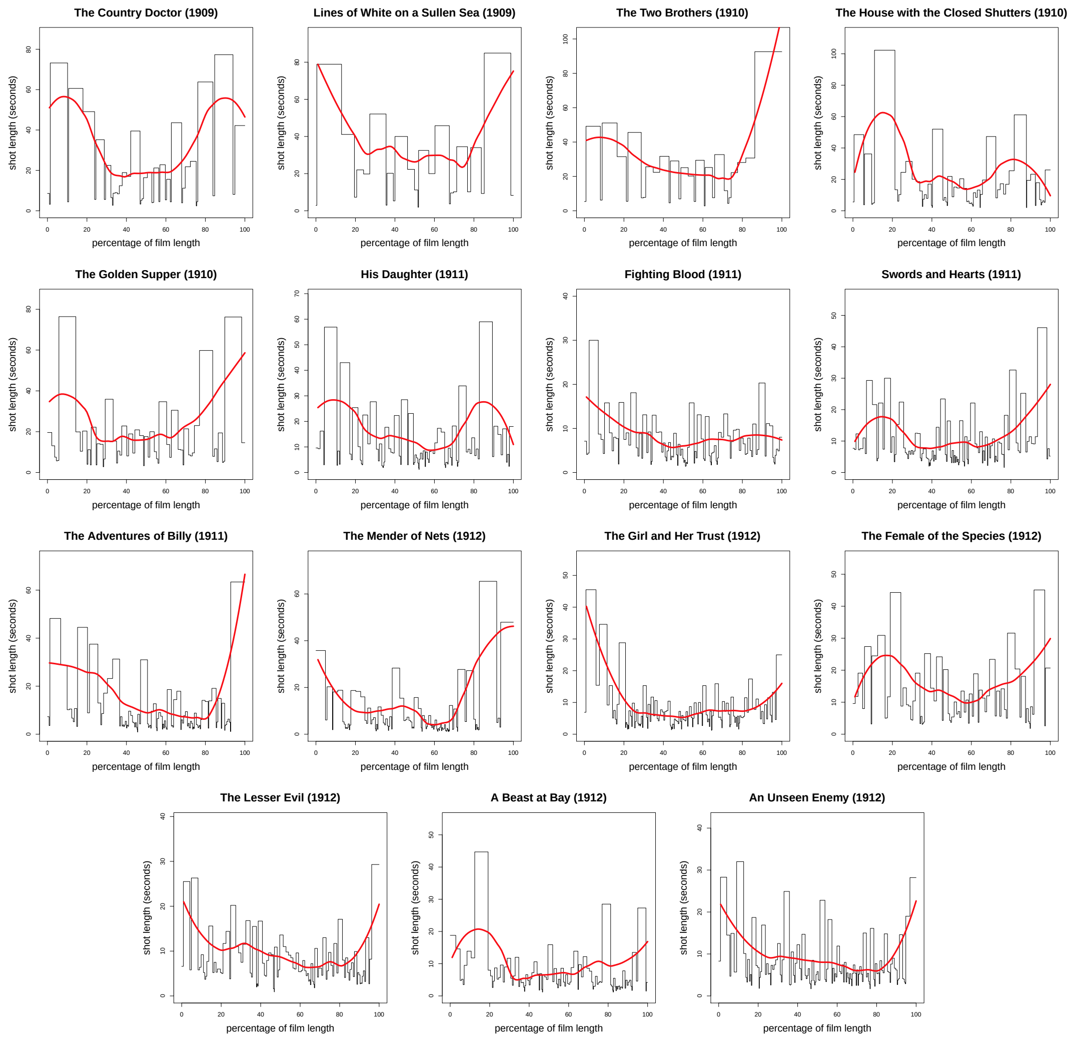

Figure 18: *Group C films* – A Country Doctor *(4),* Lines of White on a Sullen Sea *(8),* Two Brothers *(17),* The House with the Closed Shutters *(22),* The Golden Supper *(29),* His Daughter *(36),* Fighting Blood *(39),* Swords and Hearts *(42),* The Adventures of Billy *(43),* The Mender of Nets *(49),* The Girl and Her Trust *(50),* The Female of the Species *(51),* The Lesser Evil *(53),* A Beast at Bay *(54),* An Unseen Enemy *(55)*

The dominant motif for the smoothed data is of a "U" shaped [SFS] pattern, distorted in several cases. The F passage in many cases covers quite a wide range, embracing the middle of the film rather than being anchored at about the 70% mark as in the average for the full sample of 62. That is, the Griffith Biograph average that has been previously defined is not particularly representative of this, the largest of the groups that have been suggested. Allowing for end behaviour, several of the films can be described as F[SFS] or F[SFS]F.

The SLs seem over-smoothed for some films where, without the cue of the smooth, a more complex patterns looks plausible. For *Swords and Hearts*, for example, [FSFSFSFS]; for *An Unseen Enemy*, [SFSFSFS]; for *The Lesser Evil*, [SFSFS] etc. This raises the question, previously noted, of how much smoothing out of detail can be tolerated in the interests of simplification.

It could also be observed that 10/15 of the films in the group are 1911–1912, and the three examples just provided are among those with the greater number of shots in the sample. That is, later films with more shots are permitting stylistic variation not really seen in the earlier films. Further exploration of this is probably best explored by applying multivariate methods to the smoothed SLs, rather than their CFDs, and experimenting with different degrees of smoothing rather than imposing a fixed level. This is beyond the scope of the present paper.

Group D

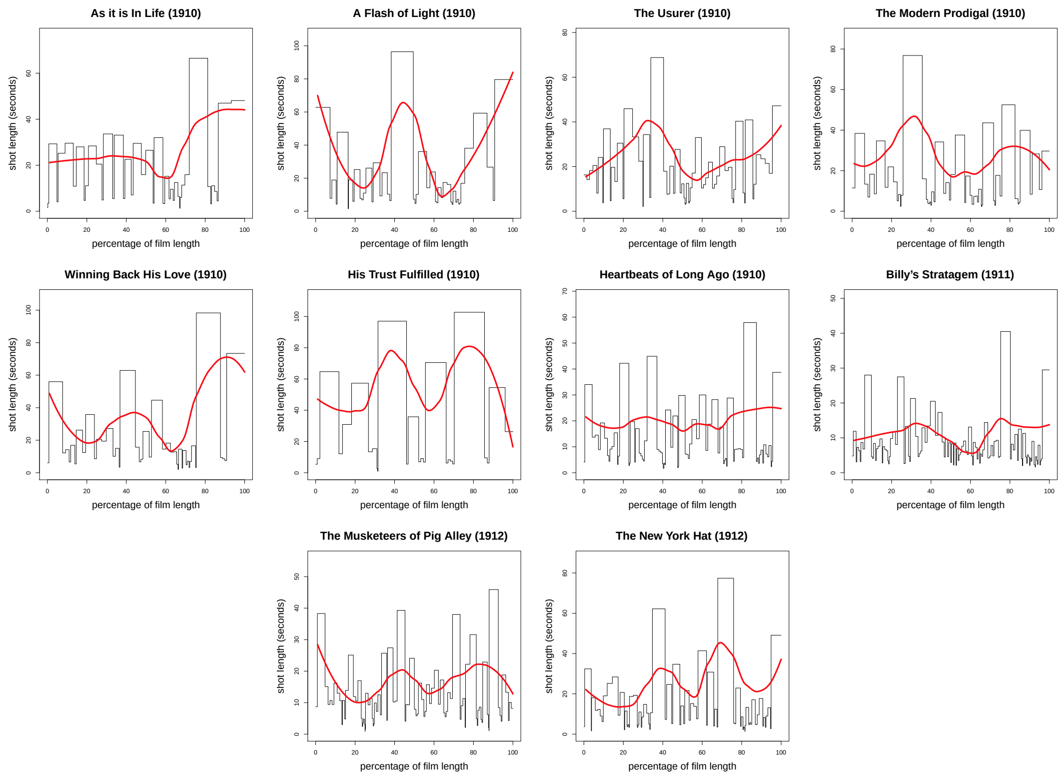

Figure 19: *Group D films* – As It is in Life *(15),* A Flash of Light *(20),* The Usurer *(23),* The Modern Prodigal *(24),* Winning Back His Love *(31),* His Trust Fulfilled *(33),* Heartbeats of Long Ago *(35),* Billy's Stratagem *(47),* The Musketeers of Pig Alley *(58),* The New York Hat *(59)*

The average pattern for the groups is fairly strongly [SFSFS] with the Fs occurring at roughly the 20% and 50% marks. *As It is in Life* isn't very convincing, and some films are missing the S start or have elements added on at the end – that is, the 'core' is [FSFS]. *Heartbeats of Long Ago* is an example of a type of film, mentioned earlier in the paper, where the not particularly noticeable absolute variation in the smoothed pattern is magnified by the scaling to [0, 1] in some of the figures.

Group E

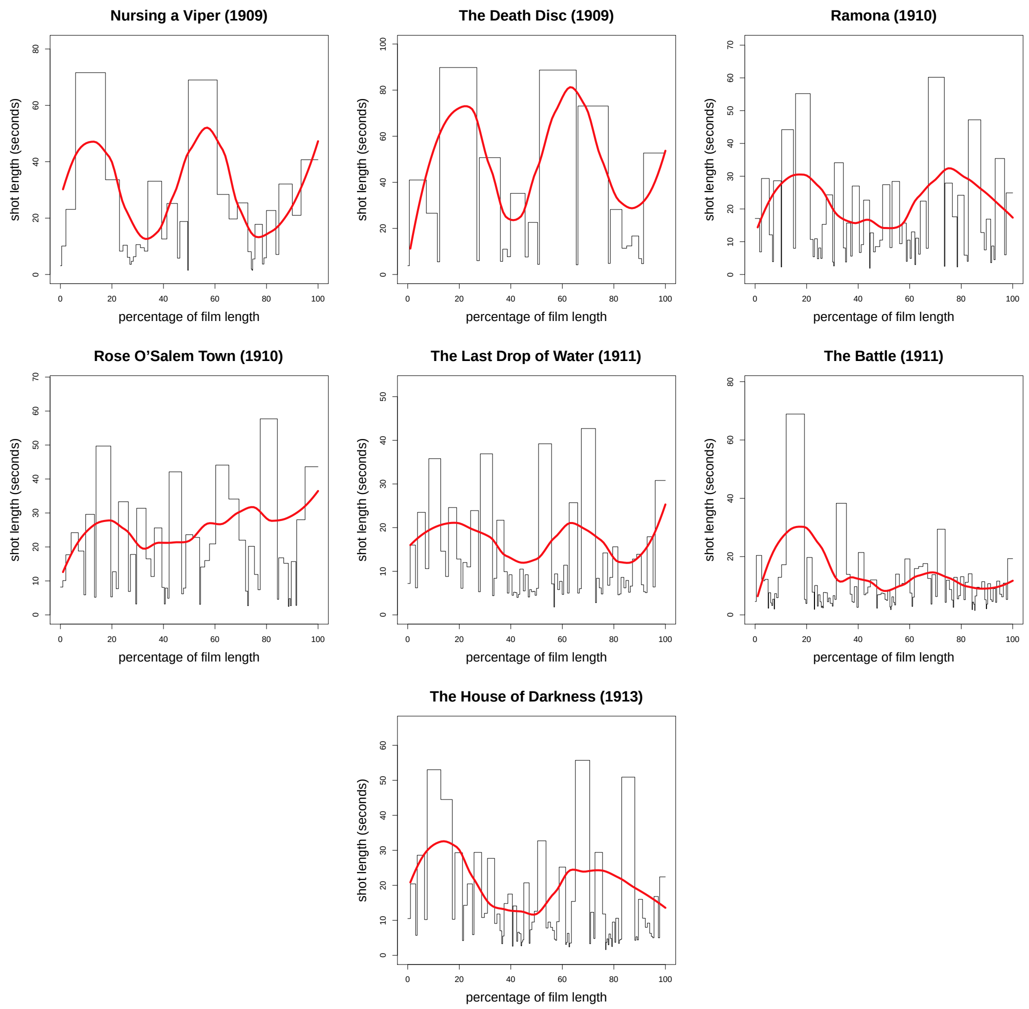

Figure 20: *Group E films* –Nursing a Viper *(9),* The Death Disc *(11),* Ramona *(18),* Rose O'Salem Town *(25),* The Last Drop of Water *(40),* The Battle *(44),* The House of Darkness *(55)*

The general pattern is [FSFSF] with an ending S in some films. *Rose O'Salem* town sits uncomfortably in the group and is there primarily because of what happens in the first half. The group is distinguished from Group D by the general positioning of the peaks and troughs.

Group F

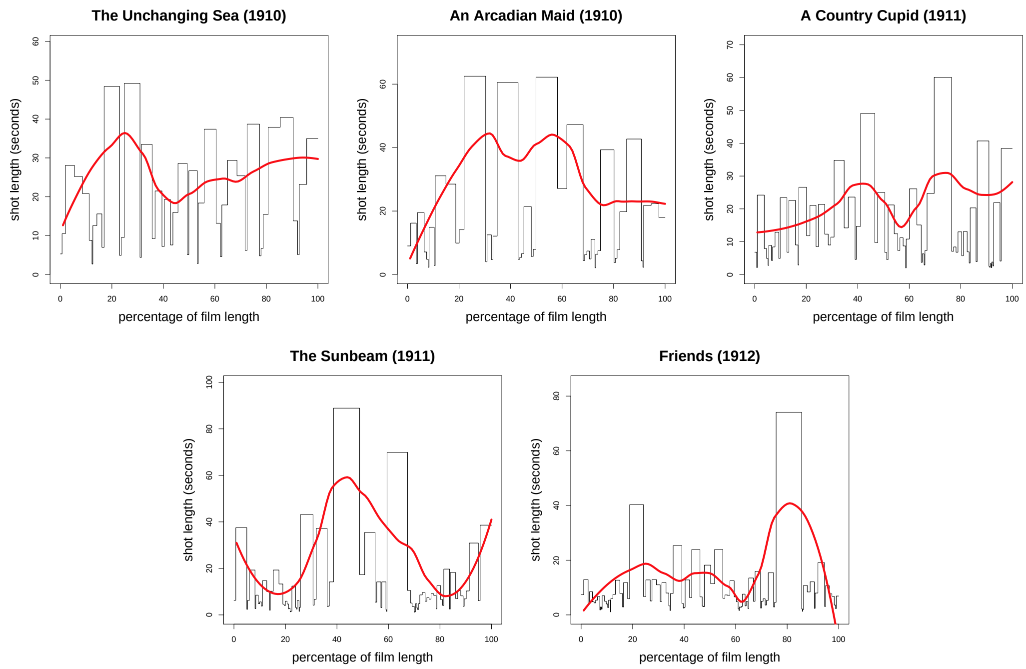

Figure 21: *Group F films* – The Unchanging Sea *(16),* An Arcadian Maid *(21),* A Country Cupid *(41),* The Sunbeam *(48),* Friends *(56))*

Group F features towards the extreme right of Figure 12 and, as these things go, plots as a fairly coherent cluster. In fact, as Figure 11 and the above plots show it is not really that. There is no obviously dominant motif, though they tend to start fast or, in the case of *The Sunbeam* have a fast early passage. The 'blocky' nature of the middle and/or ends, sometimes attributable to a small number of shots is perhaps what distinguishes them from earlier groups.

Group G

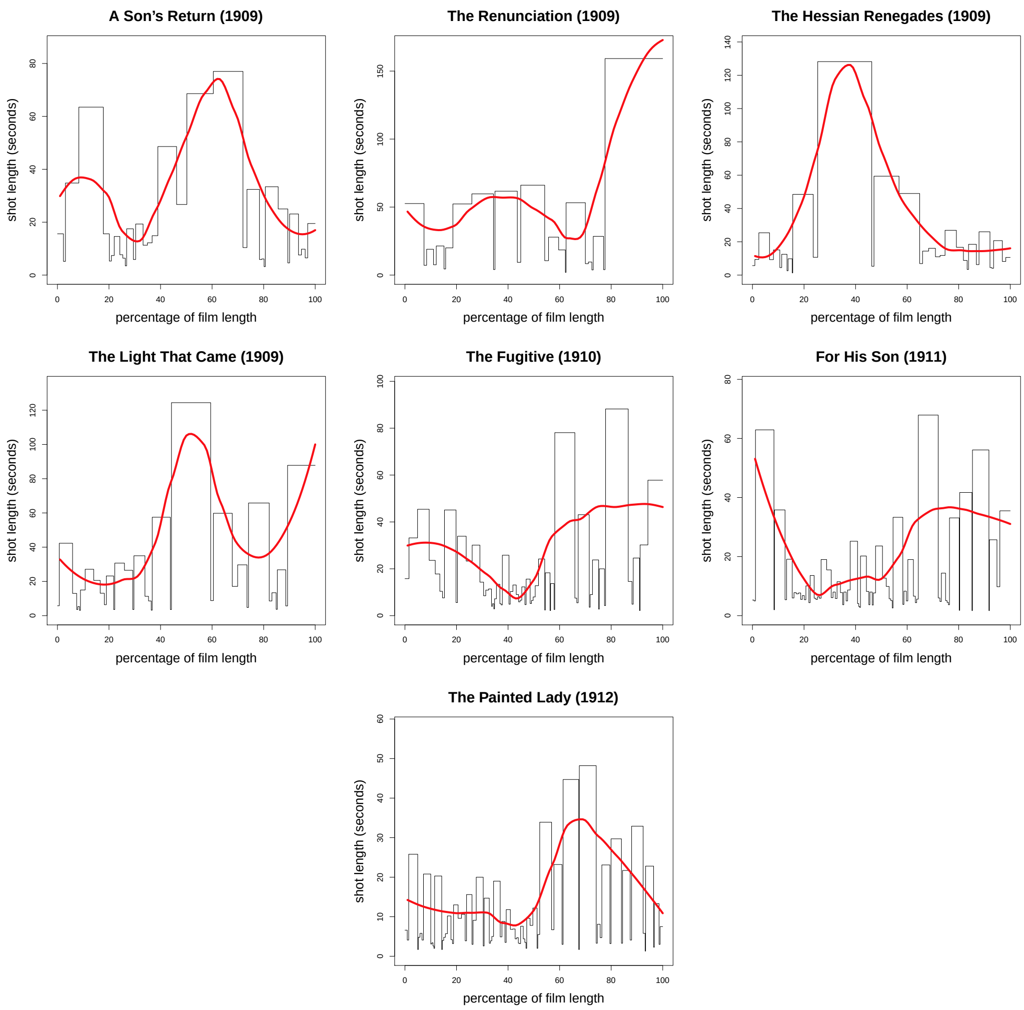

Figure 22: *Group G films* – A Son's Return *(3)*, The Renunciation *(5)*, The Hessian Renegades *(7)* , The Light that Came *(10)*, The Fugitive *(27)*,For His Son *(46)*, The Painted Lady *(57)*

As already noted this is not a coherent group, but rather a collection of outliers that cluster together more because of their differences from other films than because of their similarity to each other. Their position on the principal component plots is to the right end and bottom right, and some films have features in common with those in Group F that plot in the same general region. The films tend to be 'blocky' in the middle or end, and for the four 1909 films this is usually attributable to a small number of long shots, lending some support to earlier suggestions (to which there are exceptions) that many 1909 films are not suited to the kind of classification exercise attempted here. The two later films, *For His Son* and *The Painted Lady*, are distinguished by their relatively long slowly-cut endings, compared to faster cutting – other than the first two shots of *For His Son* – that occurs earlier.

Mike Baxter, May 2013