2. Towards a 'Humanistic Cinemetrics'?

Christian Gosvig Olesen

In recent years, film scholars have increasingly developed quantitative methodologies to produce data visualizations for a historical analysis of film style. Through methods of Cultural Analytics, Cinemetrics (2005) and ACTION (2014), which can be described as cinemetric or stylometric, scholars measure, quantify and visualize stylistic patterns in, for instance, editing, light or sound. These are used to investigate historical developments in film aesthetics and narration and to produce statistical profiles of films, directors or national cinemas.

Characterized by scientific rigour, deduction and hypothesis-testing, Cinemetrics' quantitative framework has been perceived as introducing a new empiricism or positivism in film historical research (Christie 2008). A significant reason for this is that the tool resuscitates a scientific paradigm of historical style analysis initiated in the 1970s to produce evidence for research on film editing and shot types (*ibid.*). The recent quantitative, statistical approaches of Cultural Analytics and ACTION, which build on this paradigm as a conceptual departure point, have equally been associated with a positivist epistemology (Manovich 2012a; Casey & Williams 2014). Yet, as I will argue in this article, their emergence is engendering an inductive, exploratory form of Cinemetrics which necessitates a change in the perception of cinemetric methodology as being primarily scientist. As I shall discuss further on, their practices, in the words of Eef Masson, suggest 'prob[ing] data in an exploratory manner' and highlight how data visualization defamiliarizes our objects of study by foregrounding their constructed nature (see the chapter by Mason in this publication, p. 33). In these aspects it seems they qualify as humanistic data research but that they still need to be fully recognized and distinguished as such. By attending to the development and deployments of Cultural Analytics and ACTION in comparison to Cinemetrics, my article takes further steps in this direction.[1] This, I argue, may open a critical path for contemporary, statistical style analysis and contribute to increased methodological pluralism in data-driven film historical research.

1 The developers of Cultural Analytics and ACTION, respectively, have emphasized their practices as more exploratory than Cinemetrics through blog posts and conference papers. It is these steps that I wish to acknowledge and theorize further by situating them within the discussion of scientism vs. hermeneutics in the present volume.

To introduce such a distinction is important because film studies, unlike history, sociology or literary studies, have tended to develop digital tools at a slow pace, arguably because digitized films for many years remained too data-heavy objects for automated analysis.[2] Consequently, data-driven research on film is characterized by less methodological pluralism than other disciplines and in particular lacks critical, quantitative methodologies. It seems however, that there is a desire for the latter among film historians. As Yuri Tsivian, co-founder of Cinemetrics, has observed regarding the tool's potential users, '[…] not every student of film history is ready or eager to masquerade as a scientist' (2008: 765). Furthermore, as discussed by Masson in Chapter 1 in this publication, many humanities scholars broadly speaking tend to wish to reflect their critical legacy and interpretative frameworks in their conceptualization of digital tools. In this regard, distinguishing and theorizing Cultural Analytics and ACTION as 'humanistic cinemetrics' based on what I see as their exploratory, critical *modus operandi* may provide a fruitful departure in this direction and make them more widely compelling to film scholars.

To present this argument, I draw on the theory of history of Michel de Certeau. De Certeau, responding to enthusiastic claims that had surrounded quantitative, computational history throughout the 1970s, reminded his colleagues that these methods remain technically and socially biased towards the specific traditions and institutions they emanate from (1986). When historians adopt the computer's scientific procedures, he argued, they essentially express and respond to their contemporary concerns and historical fictions, producing a 'science fiction' which combines scientific and poetic gestures of interpretation (*ibid.*: 215). Therefore he argued that historians should challenge the 'neutral' aura of computational history by remaining acutely aware of its institutional processes of knowledge production and limitations, and seek to reflect the latter in their methods (*ibid.*). Through this theoretical lens I analyse how the underlying social and technical processes of Cinemetrics, Cultural Analytics and ACTION attribute meanings to data visualizations as epistemic images in order to elicit the tools' differences.

My article is divided into three parts. First, I provide a historical discussion of statistical style analysis' epistemology by tracing cinemetric practices back to its foundational ideas in film studies in the 1970s. Second, I focus

2 However, it should be noted that film scholars have engaged extensively with and adopted digital methods of text editions deriving from literary studies to develop scholarly DVD presentations of films and with GIS technologies from socio-economic history to study film distribution, exhibition and reception. Yet only little work has been done which intervenes analytically in digitized archival films to detect formal patterns.

on the software Cinemetrics as an emblematic development of computer-based style analysis which introduces scientific, visual analytics to produce evidence for stylistic history. Finally, I discuss how cinemetric theory and techniques have developed beyond this approach in Cultural Analytics and ACTION to highlight the poetic aspects of data visualization. In this section I argue that they gesture towards a humanistic data analysis which may open a critical, methodological avenue for film scholars.

Statistical Style Analysis and Representation of Filmic Structure

To understand where cinemetric tools come from it is necessary to go back to the 1970s. In this period, film historians began developing systematic, quantitative methods to study film style, as film studies was institutionalizing as an academic discipline (Bordwell 1997). As Harvard professor of film studies at the time Vlada Petric contended, the film histories used in academic curricula had described especially film editing and style's developments haphazardly without a firm, empirical basis (1975).[3] According to Petric, to reliably account for film editing's 'historical evolution', film historians should scrutinize archival films as 'primary documents' to produce extensive and precise analytical documentation of editing patterns from canonical films, genres and periods and disseminate them in 'visual/analytical' representations (*ibid.*: 23-24).

Concurrently, film scholar Barry Salt questioned contemporary style analysis which he perceived as relying too much on hermeneutics rather than systematized, scientific procedures (1974).[4] Salt instead envisioned a film history which would achieve a more objective, scientific foundation by embracing statistical methods and the natural sciences' attitude (1983). He suggested a form of Scientific Realism, which would observe and measure stylistic features such as cutting rates, camera movements and shot scales as real phenomena, to verify or disprove hypotheses about film editing's historical development.[5] In doing so, Salt aspired to discover if aesthetic or

3 Petric in particular addressed the film histories written by Georges Sadoul, Rachael Low, Lewis Jacobs and Lotte Eisner.

4 Specifically, Salt addressed Andrew Sarris's classic work *The American Cinema: Directors and Directions 1929-1968* (New York: E.P. Dutton, 1968). In his key work *Film Style & Technology. History & Analysis* (Starword, 1983) which also explains his method in greater depth than his 1970s articles, Salt developed this approach with attention to classic mainstream cinema.

5 Salt mentioned that he found inspiration for his Scientific Realism in the work of philosophers of science such as Karl Popper, Thomas Kuhn and Imre Lakatos, without however providing a detailed discussion of how exactly they informed his work.

narrative qualities followed recurrent or exceptional patterns (for example, slow versus fast editing) and to facilitate a comparative, historical analysis following uniform, scientific principles (*ibid.*).

Attending primarily to films' cutting rates, what Salt dubbed Average Shot Lengths (ASL), he developed statistical, reduced forms of representation to express his results. In doing so, he aspired to yield more objective insights into film editing's evolution and norms, to facilitate the comparative analysis of films (*ibid.*). For this, Salt relied in particular on the widely used method of lognormal distribution to create histograms of ASLs which visualized patterns in film directors' *oeuvres* or norms in films from specific periods (Salt 2006a).[6] Using lognormal distribution for film style analysis implies that shots are grouped into class intervals or bins out of their sequential order to establish normal distributions of shot lengths. This creates a histogram displaying the film structure as a curve with a simple shape which, according to Salt, is ideal for visualizing and comparing film structure and discerning patterns, or for instance identifying outliers – meaning shots of potential analytical interest.[7]

While widely known by now and influential in film studies, the rigour of Salt's method was hotly debated in the 1970s and 1980s. Film scholars like Kristin Thompson and David Bordwell pointed to several inconsistencies, such as the circumstance that Salt initially calculated ASLs using 30-minute samples, which he regarded as representative, and not entire films (Bordwell & Thompson 1985). This, they argued, led Salt to provide inaccurate data himself.[8] Furthermore, they felt Salt's approach exaggerated the general applicability of quantitative approaches and was essentially positivist, in spite of presenting itself as a softer, scientific method.[9]

6 Lognormal distribution analysis emerged in the late nineteenth century developed by British scientist Francis Galton as a response to contemporary probability statistics and has since been refined into several variants. A somewhat simplified explanation of its scope is that it calculates the probability of a phenomenon's occurrence from a given data set with the aim of predicting its future development. It is widely used for instance to predict price developments, the occurrence of illnesses or for weather forecasts. See: Theodore M. Porter, *The Rise of Statistical Thinking, 1820-1900* (Princeton: Princeton University Press, 1986), 139 and J. Aitchinson and J.A.C. Brown's classic *The Lognormal Distribution, with Special Reference to its Uses in Economics* (Cambridge University Press, 1957), the latter of which guided Salt in the conception of his method.

7 The terms 'bin' and 'interval' can be used interchangeably and in statistical style analysis refer to the different categories of shot lengths. Illustrative examples of Salt's histogram visualizations can be seen in his article 'The Metrics in Cinemetrics' which is accessible online. See: www.cinemetrics.lv/metrics_in_cinemetrics.php, last accessed 6 April 2016.

8 Bordwell calculated the ASLs of several entire films to compare them to Salt's results based on 30-minute samples, to support this criticism.

9 As they remarked, Salt seemed to suggest 'that science's strongest certainties are those which can be reduced to numbers' (*ibid.*: 225).

Nonetheless, Bordwell and Thompson endorsed Salt's method more broadly in light of how stylistic history had hitherto been produced. As they wrote:

> His demand for precision of description, including statistical representation, comes as a welcome alternative to the practices of a generation of historians who relied upon memory, reviews, and gossip for their evidence. (*ibid.*: 234)

This remark may be taken as a concise characteristic of the direction in which film historical research was pushed by Petric and Salt in the 1970s. Responding to a scholarly need to develop scientific approaches for film style's history, they developed statistical methods which have increasingly appealed to film historians drawn to empirical research in the following decades (Buckland 2008).

Having sketched the emergence of statistical style analysis in the 1970s, I shall now turn to a discussion of Cinemetrics, which develops Salt's concept of ASL into a more fully fledged scientific method.

Style Analysis as Scientific Data Research: Cinemetrics

Conceived by University of Chicago professor Yuri Tsivian together with computer scientist Gunars Cijvans, Cinemetrics was launched in 2005 as 'an open-access interactive website to collect, store and process digital data related to film editing' (Tsivian 2008: 766). It shares statistical style analysis' assumption that film editing is a key distinguishing feature of film art and places it within an even broader theoretical reference frame to underline how scholars, also long before the 1970s, studied film editing quantitatively.[10] Highlighting how great directors throughout film history have measured segments at the editing table to achieve the pinnacle of

10 See 'Cinemetrics Predecessors': www.cinemetrics.lv/topic.php?topic_ID=38. Last accessed 28 July 2015. Tsivian has recurrently pointed to early film theorist Hugo Münsterberg's measurements of cutting rates in the mid-1910s for studies of spectatorship and psychology. Kristin Thompson finds inspiration in German film critic Georg Otto Stindt's article 'Bildschnitt' (1926) which compared shot lengths in US and German fiction films. And film historian Frank Kessler highlights German film historian Herbert Birett's foundational work on film statistics initiated in the 1960s. For a representative example of Herbert Birett's statistical style analysis see: Herbert Birett, 'Alte Filme: Filmalter und Filmstil', *Diskurs Film. Münchner Beiträge zur Filmphilologie* 2 (1988): 69-87.

their art through variations in shot length, the website's presentation calls for a metric, computational approach to film style (Tsivian 2008).[11] Currently, Cinemetrics' database counts approximately 15,000 titles uploaded by more than a thousand users.[12] These uploads do not follow a unifying selection criterion nor apply standard details on provenance and technical specificities. They constitute a heterogeneous data mass which facilitates comparison between primarily limited corpora with uniform, technical standards rather than providing evidence for a universal, evolutionary film history as in the 1970s (*ibid.*).[13] Popular among scholars propagating statistical style analysis in the 1970s and 80s such as Salt, Thompson and Bordwell, as well as newcomers, Cinemetrics users seek to refine style analysis into a more scientifically sound theoretical approach based on computer-generated data visualizations. In the following, I shall attend to key aspects of this development through a discussion of the underlying processes of its data visualization, the Cinemetrics graph.

Though not the only visualization format used by cinemetricians, the Cinemetrics graph is the primary 'inscription device' and evidentiary image used for summarizing and distributing editing data among the site's community (Latour 1987: 68).[14] As a standard representation, it consists of a custom-made red graph plotted onto a grid of horizontal lines, using

11 In particular, Tsivian highlights the formally dense works of the avant-garde directors Abel Gance, Dziga Vertov and, perhaps most emblematically, Peter Kubelka's 'metric' cinema.

12 See: www.cinemetrics.lv/database.php, last accessed 11 June 2014. However, it should be noted that this number includes a fair amount of television programmes, music videos and film excerpts as well. This aspect has however prompted discussions among academic Cinemetrics users about the data's reliability and the possibility of introducing rankings of user data to ensure cleaner data. See discussion thread 'Data Ranking and Verification': www.cinemetrics.lv/topic.php?topic_ID=355, last accessed 30 July 2015.

13 In this aspect, Cinemetrics nurtures a piecemeal approach. In general, Cinemetrics is critical of the teleological, universalizing accounts which 1970s style analysis supported. Tsivian argues at length how the earlier teleological film histories' account of cinema becoming an accomplished art form only in the late silent era, obfuscates an understanding of early cinema's distinct modes of expression.

14 Some users, such as Mike Baxter and Nick Redfern, explore alternative visualization formats, using the open-source software *R*. British scholar Nick Redfern for instance finds order structure matrices to structure the data in such a way that it allows for easier identification of clusterings of shots in sequences within films and shifts between segments. See Mike Baxter, *Notes on Cinemetric Data Analysis* (Nottingham, self-published, 2014), 46. See: www.cinemetrics.lv/dev/Cinemetrics_Book_Baxter.pdf, and Nick Redfern, 'An introduction to using graphical displays for analysing the editing of motion pictures', p. 22, 24, www.cinemetrics.lv/dev/redfern_q2_opt.pdf, last accessed 11 December 2015. For background information on *R* see: http://cran.r-project.org/.

a classic statistical format as has existed for centuries (Drucker 2014).[15] Numbered shots appear as white bars from above in sequential order. The x- and y-axes represent the variables of time code and shot duration, respectively, and users can annotate and comment on each shot/bar.[16] While the principle of showing a film's shot lengths as bars, combined through a curve resembles in fundamental aspects Salt's idea to use a histogram for comparative analysis, there remain significant differences.

First of all, the Cinemetrics graph reflects adversary, scholarly positions on the key parameter ASL. Tsivian considers Salt's ASL problematic because it only offers a single datum per film as a basis for comparison (Tsivian 2013). A single datum does not convey how cutting rates shift in relation to depicted events or motifs giving little insight into internal film dynamics. Therefore, Cinemetrics represents shots sequentially and is also designed to reflect a wider array of parameters such as *cutting swing*, which measures how the cutting rate shifts throughout a film's segments and diverges from its overall ASL. It also shows a film's *cutting range*, which is the difference between its shortest and longest shots (*ibid.*). In this regard, Cinemetrics offers a way for scholars to identify and link editing statistics on specific shots to film narration in greater detail.

In addition to Tsivian, British media scholar Nick Redfern has challenged the ASL concept by suggesting Median Shot Length (MSL) as an alternative (2011). ASL represents a mean value and is calculated by dividing a film's duration with its number of shots to find its average. MSL, on the other hand, locates the middle value of the cutting range to define it as a film's norm. In practice, this means that MSL performs outlier correction of the film's longest and shortest shots, producing different values.[17] Redfern has argued that MSL gives a more accurate impression of the typical shot length one may expect to see in a film because it is less sensitive to extreme outliers. Opposing MSL, Salt contends that MSL alters the data to an undesirable degree in cases where outliers may be relevant,

15 As Drucker points out, 'before the seventeenth century, the number of statistical graphs – that is, visual expressions of variables charted against each other as abstract quantities – was extremely small', but flourished in the following centuries with René Descartes' work in analytical geometry.

16 For a representative example of the Cinemetrics graph made by Yuri Tsivian, see for instance his visualization of Alfred Hitchcock's Rear Window (US, 1954), added to the database 23 May 2009: http://cinemetrics.lv/movie.php?movie_ID=3166, last accessed 6 April 2016.

17 Redfern gives examples of two Josef von Sternberg films, *The Lights of New York* (USA, 1928) and *Scarlet Empress* (1934). For the former the ASL is 9.9 seconds and MSL 5.1 seconds. For the latter the ASL is 9.9 and MSL 6.5.

stressing that ASL is also more widely accepted among cinemetricians (Salt 2013).[18] This discussion is reflected in the graph where MSL is included as an alternative to the ASL to enable comparison in each visualization. In this way, the graph accommodates internal adversary propositions on the function and value of the visual evidence's graphical properties (Amann & Knorr-Cetina 1990).

In different ways, these discussions of ASL negotiate the relation between data, representation and analytical purpose to determine an ideal 'analysability' of Cinemetrics' scientific image (Amann & Knorr-Cetina 1990: 107). This process can be seen as reminiscent of the way scientists debate how to fix their evidence in a representational form according to a shared set of assumptions and best practices. In this respect, Cinemetrics leans towards a scientist form of data research, following the natural sciences' attitudes to data visualization in the lineage of Salt's approach.

A second aspect of Cinemetrics that instantiates a scientist position is its emphasis on producing accurate data by eliminating potential inaccuracies caused by human reaction time. The tool's first 'classic' version launched in 2005 is semi-automatic and requires full user participation throughout a film's playback. During playback the user runs Cinemetrics in a separate window, clicking a 'Shot Change' button for every new shot to calculate the ASL and generate a graph. With Cinemetrics second version, Frame Accurate Cinemetrics Tool (FACT), which has currently only been released in a beta-version under testing, shot boundary detection has become more fine-grained and accurate by allowing users to pause and rewind so as to perform the shot segmentation with greater exactitude. Furthermore, while this is not integrated into FACT, users have expressed the overall ambition and projected as a future development – as also stated by Yuri Tsivian already in 2006 – to automate shot boundary detection in Cinemetrics in order to eliminate potential human inaccuracies, or to simply make the process of data collection quicker. However, there are different stances towards automatisation among cinemetricians and on whether human or computational annotation is most accurate or desirable. For instance, inspired by the key parameters of Cinemetrics' underlying theory, and to complement Tsivian's initiative, the related software Shot Logger – created by media scholar Jeremy Butler – goes a step further by offering automatic shot boundary detection developed in the PHP scripting language. However,

18 As Salt dryly remarks, 'Such an idea seems reminiscent of the Catholic church continuing its ban on the discussion of the idea of the earth going round the sun, even after the concept was in wide use'.

while FACT still only exists in a beta version which does not offer automatic shot boundary detection, it projects a new relation between the scholar and the viewing equipment previously used by creating a closer approximation of filmic structure through visual analytics of cutting rates in digital source material.

To conclude, Cinemetrics recasts statistical style analysis' methodology as articulated in the mid-1970s in central aspects. Salt's ASL analysis initially favoured comparison between films, representing each by a single datum. Cinemetrics, on the other hand, privileges a microscopic perspective on films, displaying text-internal, hitherto imperceptible dynamics as, in Tsivian's words, 'hard facts', from a wider range of perspectives, such as the MSL (Tsivian 2013). Cinemetrics enables scholars to closely study text-internal, micro-perspectives of single films and to switch to a macro-perspective to raise questions on film editing's historical development by comparing groups of films. One of Cinemetrics' great affordances is especially its micro-perspective, which was less prominent, if not absent, in 1970s sample-based style analysis, because it allows for a fine-grained analysis of the dynamic relationship between shot lengths and depicted events.

The display of internal dynamics and the FACT acronym's bold proposition that its procedures yield more accurate empirical data arguably advance style analysis' realism by assuming a closer approximation between film editing as a real-life phenomenon and its description (Salt 2006b). While Cinemetrics has its clear advantages for the study of film style based on editing data, it can however also be said to embody a scientist conception of data research, which humanities scholars more broadly would feel uncomfortable engaging with. In particular, the observer-independence which Cinemetrics' graph seems to imply and champion contradicts many humanities scholars' consideration of visual evidence as inherently ambiguous and contingent. Therefore, as I argued in my introduction, it is crucial to develop more critical, cinemetric approaches for film historians who do not regard scientific images as observer-independent and wish to reflect the ambiguity of their research methods in their results.

Cultural Analytics and ACTION – Gesturing Towards Humanistic Cinemetrics?

Cinemetric analysis has developed beyond Tsivian's initiative, in a variety of conceptually related, quantitative software applications. Some of these, ShotLogger and Edit2000 for instance, as Cinemetrics, analyse ASL but produce

differently styled graphs.[19] Others, such as Cultural Analytics or ACTION, draw on cinemetric theory to focus on different moving image features such as light, sound or colour. While sharing Cinemetrics' conceptual departure point in statistical style analysis their attitude towards data visualization differs fundamentally. They proceed inductively, without a preconceived theoretical framework and are less bound to tradition and established methodological operations. In this last section I would like to attend to these applications as practices that can be recognized as a form of 'Humanistic Cinemetrics' in their deployment of scientific data visualizations for style analysis.

Cultural Analytics is a research program which develops toolkits for visual analytics of cultural patterns in large image sets created within media theorist Lev Manovich's Software Studies Initiative (Yamaoka, Manovich, Douglass & Kuester 2011). Suggesting a middle way between scientist and hermeneutic approaches to visual analytics, it departs from the question, 'What will happen when humanists start using interactive visualizations as a standard tool in their work, the way many scientists already do?'[20] Its core application is ImagePlot, an extension of the open-source scientific visualization software ImageJ, first known as NIH Image, developed by the US National Institute of Mental Health.[21] Conceived by programmer Wayne S. Rasband in 1987 it advanced the combination of modern computation techniques with microscopy and gained widespread success in the natural sciences, because of its later translation into Java-programming (Schneider, Rasband & Eliceiri 2012).[22]

The software has always been open source, enabling users to tweak it and resulting in around 500 plug-ins by May 2012 (Schneider, Rasband & Eliceiri 2012). ImagePlot added four of these.[23] Manovich has developed ImagePlot's visual analytics approach within the 'Cultural Analytics' research program and initially conceived it for analysing the digital age's big data image sets, especially of amateur image sites such as Flickr and Instagram. In this regard, Manovich considered ImageJ capable of providing adequate 'super-visualization technologies' to match these sets' scale and discover patterns in them.[24]

19 See www.data2000.no/EDIT2000/ and www.shotlogger.org/, last accessed 10 April 2015.

20 See the project introduction 'Cultural Analytics': http://lab.softwarestudies.com/p/cultural-analytics.html, last accessed 27 September 2015.

21 'About NIH Image', see: http://rsb.info.nih.gov/nih-image/about.html, last accessed 22 April 2015. ImageJ followed the NIH Image software on the basis of which it was created.

22 According to Schneider, Rasband and Eliceiri, in the late 1990s Java programming became considered an 'operating system-agnostic' language, compatible between Macintosh and PC. The 'J' in ImageJ stands for Java. (pp. 671-672).

23 See: http://lab.softwarestudies.com/p/imageplot.html#features1, last accessed 11 May 2015.

24 'Cultural Analytics', op. cit.

Cultural Analytics' scope quickly expanded to digitized heritage collections, art history and moving images.[25] Regarding the latter, Cultural Analytics thinks along the lines of Cinemetrics by evoking how formally dense and complex film works necessitate statistical approaches for analysing filmic structures (Manovich 2013).[26] However, its distinguishing feature as a method is that it processes entire films as image sets instead of extracting metadata to produce reduced, statistical representations. It breaks down video files into sequences of separate images and seriates or layers them according to various image features in different visualization types. The ImageJ Montage visualization, for example, orders frames onto a grid according to their sequential order, from left to right, enabling a quick, comprehensive overview of movements between shots (*ibid.*).[27] Figurative in comparison to Cinemetrics' graph, it seems close to early scientific cinematography, such as Etienne-Jules Marey's and Eadweard Muybridge's sequential photography, in particular the latter's famous *The Horse in Motion* (1878) (Tosi 2005). It has been applied to films by Soviet avant-garde director Dziga Vertov to grasp his film's complex structures and for understanding his reuse of footage within different films. With another ImageJ visualization type, the Summary image, one can layer image sequences to visualize median values of colours in films.[28] Subsequently, with ImagePlot one may plot these visualizations or entire image sets on a y- and x-axis with different values.

Closer to Cinemetrics' reduced visual analytics, the recent project Audio-Visual Cinematic Toolkit for Interaction, Organization and Navigation

25 For examples of the wider array of visualization formats, see: http://lab.softwarestudies.com/p/research_14.html, last accessed 11 May 2015.

26 As in Cinemetrics Dziga Vertov's documentary theory is used to conceptualize the potential of digital tools, aligning their analytical potential with Vertov's conception of cinema as a machinic vision which unveils hidden structures of life to the human eye. Yet in contrast to Cinemetrics, Manovich has also prominently invoked Vertov's documentary theory to regard new media as dynamic and as privileging multiple viewpoints rather than positivism, by analogy to Vertov's staging of editing. See: Lev Manovich, *The Language of New Media* (Cambridge, MA: The MIT Press, 2001), 199.

27 See: http://lab.softwarestudies.com/2013/01/visualizing-vertov-new-article-by-lev.html, last accessed 6 April 2015. The visualization of Vertov's *The Eleventh Year* can be seen via the following link: www.flickr.com/photos/culturevis/3988919869/in/album-72157632441192048/, last accessed 6 April 2016.

28 For reasons of space I do not include a discussion of the Summary visualization here. For examples and an interesting recent application, I refer to film scholar Kevin L. Ferguson's use of Summary for studying the Western. See: Kevin L. Ferguson, 'What Does the Western Really Look Like?', https://medium.com/the-outtake/what-does-the-western-look-like-545981d93ae8, last accessed 27 September 2015.

(ACTION) developed by Michael Casey, Mark Williams and Tom Stoll at Dartmouth College also analyses patterns in film style. Using the open-source software Matplotlib and Python to visualize 'latent stylistic patterns' of colour, sound and movement it creates auteur and film profiles from a sample of 120 films.[29] Though not focused on film editing, it extends cinemetric theory to comprise other stylistic features and develops it by putting greater emphasis on machine learning processes in the hope of producing more precise, clean data (Casey & Williams 2014).[30] Using algorithms to extract for instance mean values of colour and sound, it charts the results onto order structure matrices or tabular diagrams where single data of mean values represent *auteur* profiles to enable comparison, much in the vein of Salt's original approach. In the latter format, directors are represented by their initials, AH for Alfred Hitchcock and JLG for Jean-Luc Godard for instance, and are classified according to their mean values of colour.

While conceptually related to Cinemetrics, these uses of visual analytics can be said to take different measures to distance themselves from its scientism. Manovich evokes statistician John Tukey's tradition of Exploratory Data Analysis (EDA) as an inductive approach, to underline that ImagePlot does not depart from a clearly defined hypothesis but uses visualizations for exploratory purposes as a stepping stone to new research questions (Manovich 2012a).[31] According to Manovich, this produces open answers rather than finite, hard scientific explanations and encourages multiple interpretations. Furthermore, Cultural Analytics also nods to literary scholar Franco Moretti's quantitative, historical approach as middle way between scientist methodological rigour and hermeneutics' 'free play' of subjectivity (Manovich & Douglass 2009; Moretti 2008). Arguably, this 'free play' manifests itself in an attitude towards data visualizations which does not regard them as hard evidence but equally contemplates their graphic features to highlight their abstract and constructed nature. Manovich, for example, underscores the limits of ImagePlot visualizations when he associates its graphic properties with the characteristic compositions of Soviet photographer Alexander Rodchenko's avant-garde photography (Manovich 2012b). In doing so, he stresses how ImagePlot's visualizations may also be taken to render reality more unfamiliar to us rather than serving solely a revelatory, scientific function.

29 See https://sites.dartmouth.edu/mediaecology/content-partners/campus-partners/action/, last accessed 10 November 2015.

30 The project's white paper is available online and contains the list of the 120 films analysed within the project as well as the visualizations which I refer to here. See: https://securegrants.neh.gov/PublicQuery/main.aspx?f=1&gn=HD-51394-11, last accessed 6 April 2016.

31 See also Bernhard Rieder's and Theo Röhle's discussion of Tukey in this anthology.

Through a similar attitude, ACTION also seems to locate itself midway between the scientific and aesthetic contemplation of data visualizations to emphasize its contingencies (Casey 2014). This can be seen in the appropriation work *One Million Seconds* (US, 2014) which Casey produced using sound classifications of film samples analysed within ACTION.[32] Where Manovich associatively muses on ImagePlot's visualization in relation to Rodchenko, Casey uses Glenn Gould's famous recording of Bach's *Goldberg Variations* (1981) as a template from which film excerpts are retrieved based on their audio similarities with Gould's recording. Thus, Casey creates a frenetic video piece where glimpses of barely recognizable film excerpts replace each other in rapid succession based on their audio similarity to Gould's *Goldberg Variations*, in which both the films' and Gould's recording are audible.

While tentative, experimental gestures, both Manovich's and Casey's appropriations can be said to point to the uncertainty in their analytic and representational practices, inviting us to think critically about the meanings we assign to data visualizations. In this regard ImagePlot and ACTION differ from Cinemetrics because they do not strive towards best practices following positivist aspirations nor idealize data visualization for stylistic analysis. Whereas Cinemetrics' graph is perceived as a strong evidentiary image among its practitioners, ACTION and ImagePlot embrace the analytical potential of computational stylometry while stressing how data visualizations can also be perceived as abstract and, as discussed by Masson, defamiliarize our objects of study (Masson in this volume). In doing so, they arguably appreciate scientific, graphical expressions within a historically long-standing intersection of science and art to open for less formalized exploratory methodologies.[33] Consequently, Cultural Analytics and ACTION come across as more self-reflexive towards data's visual shapes and may be seen as congruent with a humanistic approach which, as Johanna Drucker defines it, 'calls to attention its madeness – and by extension, the constructedness of knowledge, its interpretative dimensions' (2014: 178).

Bearing in mind this observation, I would conclude by suggesting that Cultural Analytics and ACTION can also productively be considered

32 Michael Casey's video appropriation and description of the work can be accessed via the following link: https://vimeo.com/105909439, last accessed 14 May 2015.

33 On this subject see Caroline A. Jones and Peter Galison, 'Introduction' in Caroline A. Jones & Peter Galison (eds.), *Picturing Science, Producing Art* (Abingdon, New York: Routledge, 1998) and Monique Sicard, *La fabrique du regard. Images de science et appareils de vision* (*Xve-XXe siècle*) (Paris: Editions Odile Jacob, 1998).

consistent with de Certeau's notion of computational history as a 'science fiction' referred to in my introduction, by foregrounding both the scientific and poetic dimensions of its making (de Certeau 1986). Concretely, I take them to suggest, lending a characterization of de Certeau's historiography from Jeremy Ahearne, 'a method [which] is alternately scientific and anti-scientific. It oscillates between interpretation and something like anti-interpretation' (Ahaerne 1995: 35). Thinking along these lines when visualizing data, I believe, may especially enable film historians with reservations about style analysis' scientific realism to move in a new, critical direction which restores one of the fundamental tasks of the historian on their terms, namely to emphasize the ambiguity of the relationship between past and present and its construction (de Certeau 1986). In doing so, we may to a greater degree underline the enigmatic enterprise of (film) history making, while embracing computational methods in fruitful new ways to study filmic structures and directorial styles and review the way we understand the inner workings of films new and old.

References

Ahaerne, Jeremy. 1995. *Michel de Certeau. Interpretation and its Other*. Cambridge: Polity Press.

Amann, K. & K. Knorr Cetina. 1990. "The fixation of (visual) evidence." In *Representation in Scientific Practice,* ed. Michael Lynch & Steve Woolgar. Cambridge, MA: The MIT Press.

Baxter, Mike. 2014. *Notes on Cinemetric Data Analysis*. Nottingham, self-published. Available online: www.cinemetrics.lv/dev/Cinemetrics_Book_Baxter.pdf.

Birett, Herbert. 1988. "Alte Filme: Filmalter und Filmstil." *Diskurs Film. Münchner Beiträge zur Filmphilologie* 2.

Bordwell, David & Kristin Thompson. 1985. "Towards a Scientific Film History?" *Quarterly Review of Film Studies* 10 (3): 224-237.

Bordwell, David. 1997. *On the History of Film Style*. Cambridge, MA, and London, England: Harvard University Press.

Buckland, Warren. 2008. "What Does the Statistical Style Analysis of Film Involve? A Review of Moving into Pictures. More on Film History, Style, and Analysis." *Literary and Linguistic Computing* 23 (2): 219-230.

Casey, Michael & Mark Williams. 2014. *White Paper. ACTION: Audio-visual Cinematics Toolbox for Interaction, Organization, and Navigation of Film*. Hanover: Dartmouth College.

Casey, Michael. 2014. "Investigating Film Authorship with the ACTION Toolbox." Conference paper given at the 55 Annual conference of the Society for Cinema and Media Studies, 22 March.

Christie, Ian. 2006. "'Just the Facts, Ma'am?' A Short History of Ambivalence towards Empiricism in Cinema Studies." *Tijdschrift voor Mediageschiedenis* 9 (2): 65-73.

Daston, Lorraine & Peter Galison. 1992. "The Image of Objectivity." In *Representations*, 40.

De Certeau, Michel. 1986. "History: Science and Fiction." In Michel de Certeau, transl. Brian Massumi, *Heterologies. Discourse on the Other*. Minneapolis: Minnesota University Press. 217-218.

Drucker, Johanna. 2014. *Graphesis. Visual Forms of Knowledge Production*. Cambridge, MA, and London, England: Harvard University Press.

Chávez Heras, Daniel. 2012. "The Malleable Computer: Software and the Study of the Moving Image." *Frames Cinema Journal* 1 (1). n.p.

Jones, Caroline A. & Peter Galison. 1998. "Introduction." In *Picturing Science, Producing Art*, ed. Caroline A. Jones & Peter Galison. Abingdon, New York: Routledge.

Ladurie, Emmanuel Le Roy. 1979 [1973]. *The Territory of the Historian*, transl. Ben and Siân Reynolds. Sussex: The Harvester Press.

Latour, Bruno. 1988. *Science in Action. How to Follow Scientists and Engineers Through Society.* Cambridge, MA: Harvard University Press.

Manovich, Lev. 2001. *The Language of New Media*. Cambridge, MA: The MIT Press.

—. 2012a. "The Meaning of Statistics and Digital Humanities." *Software Studies* blog entry 27 November: http://lab.softwarestudies.com/2012/11/the-meaning-of-statistics-and-digital.html.

—. 2012b. "How to Compare One Million Images?" In *Understanding Digital Humanities*, ed. David M. Berry. New York: Palgrave Macmillan. 249-278.

—. 2013. "Visualizing Vertov." Available online: http://softwarestudies.com/cultural_analytics/Manovich.Visualizing_Vertov.2013.pdf.

Manovich, Lev and Jeremy Douglass. 2009. "Visualizing Temporal Patterns in Visual Media." Available online: http://softwarestudies.com/cultural_analytics/visualizing_temporal_patterns.pdf.

Moretti, Franco. 2008. *Graphs, Maps, Trees. Abstract Models for Literary History.* London: Verso.

Petric, Vladimir. 1974. "A Visual/Analytic History of the Silent Cinema (1895-1930)." Paper presented to the 30th Congress of the International Federation of Film Archives, May 25-27. Available online at: http://files.eric.ed.gov/fulltext/ED098639.pdf.

—. 1975. "From a Written Film History to a Visual Film History." *Cinema Journal* 14 (2), 20-24.

Porter, Theodore. 1986. *The Rise of Statistical Thinking, 1820-1900*. Princeton: Princeton University Press.

Redfern, Nick. "An introduction to using graphical displays for analysing the editing of motion pictures." Available online: www.cinemetrics.lv/dev/redfern_q2_opt.pdf.

—. "Films and Statistics – Give and Take. The average shot length as a statistic of film style". Available online: http://cinemetrics.lv/dev/fsgt_q1b.php.

Salt, Barry. 1974. "Statistical Style Analysis of Motion Pictures." *Film Quarterly* 28 (1).

—. 1983. *Film Style and Technology: History and Analysis*. London: Starword.

—. 2006a. *Moving into Pictures. More on Film History, Style, and Analysis*. London: Starword.

—. "Statistical Film Analysis (Basic Concepts and Practical Details)." Available online: www.cinemetrics.lv/articles.php.

—. 2006b. "The Metrics in Cinemetrics." Available online: www.cinemetrics.lv/metrics_in_cinemetrics.php.

—. "Films and Statistics – Give and Take. Graphs and Numbers." Available online: http://cinemetrics.lv/dev/fsgt_q1d.php.

Schneider, Caroline A., Wayne S. Rasband & Kevin W. Eliceiri. 2012. "NIH Image to ImageJ: 25 Years of Image Analysis." *Nature Methods* 9 (7): 671-675.

Sicard, Monique. 1998. *La fabrique du regard. Images de science et appareils de vision* (*XVe-XXe siècle*). Paris: Editions Odile Jacob.

Tosi, Virgilio. 2005. *Cinema Before Cinema. The Origins of Scientific Cinematography.* London: British Universities Film & Video Council.

Tsivian, Yuri. 2008. "What Is Cinema? An Agnostic Answer." *Critical Inquiry* 34 (4), 754-776.

—. "Films and Statistics – Give and Take. Question 1: Median or Mean." Available online: http://www.cinemetrics.lv/dev/fsgt_q1a.php.

—. "Taking Cinemetrics into the Digital Age (2005-now)." Available online: www.cinemetrics.lv/dev/tsivian_2.php.

Yamaoka, So, Lev Manovich, Jeremy Douglass & Falko Kuester. 2011. "Cultural Analytics in Large-Scale Visualization Environments." *Computer* 44 (12), 39-48.