Chapter 1

LITERARY STUDIES IN THE DIGITAL AGE

[T]here is no method, however well adapted to a given science, that literary history can transplant and apply to its own researches. The illusion that this is possible is responsible for much poor and childish work: statistics and charts, evolution of species, and quantitative analysis are processes, methods, and hypotheses excellent in their place, but their place is not literary history.¹

In the last decade, and especially in the last four or five years, the insistence in this epigraph – that quantitative methods have no place in literary history – has been repeated many times. The fact that this particular passage comes from a book first published in 1922, and intended as a guide for graduate students, should demonstrate that both the application of such methods, and the resistance to them, are of considerably longer standing in debates about literary history than is generally acknowledged. Nonetheless, discussion of quantitative methods has almost certainly never been as heated or as widespread – or as apparent to the majority of literary scholars – as it is today. While there are a number of quantitative approaches to literature,² the current debate focuses on Franco Moretti’s work in literary history. As Priya Joshi says, literary scholars have for a long time ‘regarded quantitative analysis with suspicion bordering on contempt’.³ But in the response to the publication in 2000 of Moretti’s ‘Conjectures on World Literature’, and in 2005 of his book Graphs, Maps, Trees,⁴ this contempt has escalated – especially in the American humanities – to an intense stand-off.⁵

The controversy surrounding Moretti’s work is, to a significant extent, specific to it. But this debate also presents important criticisms of quantitative methods that need to be engaged with if such studies are to make a productive contribution to literary history and humanities scholarship generally. This chapter considers three closely related criticisms that have been levelled at quantitative literary research (predominantly at Moretti’s ‘experiments’ in literary history): first, that such approaches reduce the inherent complexity and multiplicity of literature and language to uniform data; second, that quantitative methods make false claims to authoritative and objective knowledge; and finally, that such studies resonate, in problematic and complicit ways, with contemporary institutional discourses, especially neoliberal or economic rationalist managerial practices.

I am not proposing that such criticisms are never applicable to quantitative approaches; like all research practices, these can be applied in varying ways. Nor is this chapter a defence of Moretti’s scholarship. Although his centrality to the debate makes an engagement with his arguments and methods unavoidable – and while I find his work well worth the engagement – some aspects of Moretti’s research justify some of the
criticisms that have been made. However, I will argue that reductionism, absolutism and acquiescence to neoliberalism are not intrinsic to quantitative methods. In this chapter I discuss work in book history and the digital humanities that I have found useful in developing my approach to literary historical data. Specifically, I argue that an approach based on book history’s methodological pragmatism regarding the nature and use of data, and the digital humanities’ method of modelling, offers a productive way of integrating empirical data with the paradigm of humanities knowledge as a critical, analytic and speculative process of inquiry. This approach maintains what Donna Haraway calls ‘a no-nonsense commitment to a faithful account of the “real” world’, while preserving, in George Levine’s words, ‘a tentativeness that keeps all aspirations to knowledge from becoming aspirations to power as well’.

**I Quantitative Method and its Critics**

As the criticisms of quantitative approaches to literature are largely directed at Moretti’s work, I will begin with a brief summary of his arguments: both against conventional approaches to literary history and for quantitative methods. For decades, Moretti has argued that a literary history based only on the texts that make up the canon offers no insight into the vast ‘mass’ of literature, and no basis for understanding the causes and processes of literary change. In 1983 he wrote that:

> At present, our knowledge of literary history closely resembles the maps of Africa of a century and a half ago: the coastal strips are familiar but an entire continent is unknown. Dazzled by the great estuaries of mythical rivers, when it comes to pinpointing the source we still trust too often to bizarre hypotheses or even to legends.

More recently, Moretti has refined this critique into a specific challenge to the reliance of literary history on detailed textual analysis or ‘close reading’ as the source of historical evidence. He identifies ‘close reading’ – where the ‘representative individual’ defines the ‘whole’, or the ‘one per cent of the canon’ signifies ‘the lost 99 per cent of the archive’ – as a form of ‘topographical thinking’. The main problem with this approach, and the source of what Moretti considers as irrationality, lies in the fact that the ‘rare and… exceptional’ works of the canon are by definition not representative. In taking the canon as its object, literary history fails to consider the ‘banal, everyday, normal’ operations of the literary field and the wider context in which literary change occurs.

For Moretti, the means of overcoming this unrepresentative focus cannot be more reading. The size of the archive renders this potential solution impossible to achieve: even ‘a novel a day every day of the year would take a century or so’ to cover nineteenth-century British fiction. As well as a matter of scale, close reading gives no insight into the workings of the literary system:

> At present, our knowledge of literary history closely resembles the maps of Africa of a century and a half ago: the coastal strips are familiar but an entire continent is unknown. Dazzled by the great estuaries of mythical rivers, when it comes to pinpointing the source we still trust too often to bizarre hypotheses or even to legends.

This chapter has been published in the volume “Reading by Numbers: Recalibrating the Literary Field” by Katherine Bode. London: Anthem Press, 2012. ISBN: 9780857284549.
To this end, Moretti offers a paradigm of ‘distant reading’ that deliberately abstracts both the material and textual features of literary works to provide new accounts of literary history based on ‘a specific form of knowledge: fewer elements, hence a sharper sense of their overall interconnection. Shapes, relations, structures. Forms. Models.’ Moretti’s work provides an important statement of the contribution quantitative methods can make to literary history: namely, their potential to represent historical trends and, in so doing, enable a form of analysis that moves beyond the handful of exceptional texts and authors that are repeatedly discussed in literary history. However, and although he is often perceived as such, Moretti is not the only scholar to make these arguments: both his challenge to established practices in literary history and his rationale for quantitative analyses align closely with ideas in book history.

Since the emergence of this interdisciplinary field in the 1980s, book historians have – like Moretti – rejected a canonical approach to literary history and challenged that discipline’s reliance on theory, insufficiently grounded in empirical, historical evidence (what Moretti calls literary history’s basis in ‘bizarre hypotheses’ and ‘legends’). Robert Darnton, for instance, describes the canonical approach to – or ‘great-man, great-book variety’ of – literary history as an artifice, pieced together over many generations, shortened here and lengthened there, worn thin in some places, patched over in others, and laced through everywhere with anachronism. It bears little relation to the actual experience of literature in the past.

As a substitute for this canonical focus, Darnton recommends that literary scholars ‘work through theoretical issues by incorporating them more thoroughly in more research of a concrete, empirical character’. Darnton’s focus on reception – the ‘experience of literature in the past’ – is characteristic of most work in book history. This signals another connection between such scholarship and Moretti’s analyses in Graphs, many of which treat the reading community as the catalyst for literary development. Other connections include the identification of social history – especially the Annales school – as an important historical and intellectual antecedent, and a focus on literature as a system (or, as book historians tend to call it, a ‘communications circuit’) rather than a collection of individual texts. Finally, and most importantly for the purposes of this chapter, like Moretti, many book historians use quantitative methods to explore this system or circuit. This aspect of book history includes studies that focus on the operations of the publishing industry and the reception of literature in history, as well as an emerging body of work that uses historical data – as this book does – to explore changes and developments in particular literary forms.

In presenting these connections, I am not aiming to minimise the innovation and uniqueness of Moretti’s work. His application of quantitative methods extends well beyond any other work in book history, especially in his use of what might be called textual as well as material or historical data. While the first chapter of Graphs is (as Moretti acknowledges) essentially an exercise in quantitative book history – drawing on historical data to explore trends in book publication, authorship and genre across a range of national fields – the other two chapters are based on datasets created from elements
within particular literary texts, such as character, setting, plot and device. From this textual dataset, Moretti produces visual representations (in his words, ‘abstract models’) of what is occurring within the pages of books. In more recent work, Moretti employs quantitative methods to analyse language patterns in much larger groups of texts. These other studies have their own antecedents: Literary Darwinists use textual data, and Moretti’s analyses of language patterns draw on methods developed in linguistics and digital humanities (or humanities computing). But no one else incorporates this range of approaches, or combines them in ways as original and provocative as Moretti. My intention in establishing these connections between Moretti’s work and other quantitative approaches to literature, particularly those of book history, is to signal the relevance of ongoing methodological discussions in these other fields to the current debate about Moretti’s quantitative ‘experiments’ and, in particular, to the criticisms these experiments have received.

The most general of these criticisms is that, in reducing aspects of the literary field to data, quantitative approaches provide and privilege a simplistic view of literature, one that fails to understand – or more pointedly, dismisses and violates – such things as aesthetic value and literary complexity. Discussing Moretti’s analysis of British book titles, Katie Trumpener describes

> the designation of a novel as a novel, a poetry volume as poems…[as] alienating, reducing books to mere commodities – a box of salt with the generic label ‘Salt’, a bag of flour announcing itself ‘Flour’ – as if the book’s content (and the irreducibility of authorial style) was virtually irrelevant.

According to Trumpener, such designations – and Moretti’s ‘statistically driven model of literary history’ more broadly – ‘violate the individuality of the text’. Similarly, in a review of *Graphs*, Robert Tally argues that, in relying on data, the ‘literary historian will overlook, or deliberately elide, the particulars that make the study of literature critical. The practice leads to, and even encourages, generalisations that critics would normally eschew.’ Referring specifically to analyses that use words as data, but also discussing Moretti’s *Graphs* as a whole, Michael Rothberg asks whether ‘quantitative cultural historians [can] prevent the massification of word-based data from performing a reduction of the inherent polysemy or aporetic nature of the signifier?… Can we quantify without losing the disruptive detail and splitting significations to which we have learned to attend?’

This perception that quantitative analysis will replace complexity with simplistic explanation underpins another criticism of such methods, also made primarily in relation to Moretti’s work: that quantitative approaches make a false claim to absolute knowledge and objective truth. In similar terms to Rothberg, Gayatri Spivak perceives in Moretti’s quantitative experiments an attempt to control the inherent ‘undecidability’ of literary culture by creating ‘authoritative totalizing patterns’ that reduce the complexity of the literary field to simplistic models. However, she identifies the ‘real problem’ with distant reading as its ‘claim to scopic vision’. Such vision – described by Haraway as a ‘god trick’, claiming to see everything ‘from everywhere and nowhere equally
and fully — asserts a form of knowledge that is transcendent, central, total and true. This charge of false objectivity is probably the most developed aspect of the critiques of Moretti’s work, and I will return to it in detail later in this chapter.

The association of quantitative research with objective knowledge is seen as having major implications for power relations between literary scholars and within the institution of the university. Jonathan Arac describes Moretti’s model of ‘distant reading’ as ‘covert imperialism’ due to the hierarchical difference it creates between ‘readers’ and the ‘global synthesizer, who becomes the maestro di color che sanno (“master of those who know”). Similarly, though focusing on the different national locations and languages of these readers, Spivak criticises Moretti’s use of ‘native informants’, predominantly from non-Anglophone literary cultures, to provide ‘close reading[s] from the periphery’ that are amassed at the Anglophone centre. These critics are responding, specifically, to Moretti’s proposal for world literary studies and the implicit hegemony of the English language (and for Spivak, of American nationalism) they perceive in his framing of this agenda. However, similar claims regarding the inequalities between readers and synthesisers could be made of all projects that use the scholarship of others – as I do with the bibliographical work in AustLit – to identify trends in literary history.

More broadly in terms of power inequalities, there is the view that quantitative analyses resonate, and are complicit, with other paradigms that foreground numerical measures, especially the neoliberal or economic rationalist ideology underpinning managerial practices in today’s universities and in capitalist societies generally. Referring to the American academy, James English describes the ascendancy of a ‘naive or cynical quantitative paradigm that has become the doxa of higher-educational management’. This ‘hegemony of numbers’ favours the social and natural sciences – disciplines that also deploy statistics. Under these conditions, ‘antagonism toward counting has begun to feel like an urgent struggle for survival’ for literary studies. Susan Lever makes a similar argument in relation to the Australian university system, arguing that literary criticism – a practice which requires ‘time rather than money’ – falls between the gaps in terms of gaining funding in an institutional context that values research based on a ‘science model’. Projects that require ‘research assistants, travel, even equipment’, speak to this model in ways that marginalise traditional humanities research: ‘[t]hat’s one reason’, Lever proposes, ‘why cultural history, media studies, “distant reading” are now the fashion’ for literary studies in Australia. The idea that quantitative methods support and institute power inequalities between disciplines relates to a wider argument regarding the oppressive consequences – for society generally – of forms of knowledge based on statistical evidence, numerical data and averages.

There can be no doubt that numbers and statistics are imbued with significant power in modern society, and that much of this power comes from the rhetoric of objectivity and truth surrounding such measures or, as Sally Engle Merry puts it, ‘the magic of numbers and the appearance of certainty and objectivity that they convey’. I strongly agree with many of the scholars above that this rhetoric is employed in contemporary universities to channel and control research, and that this configuring of knowledge is having major negative consequences for the humanities. These institutional factors are perhaps one of the main reasons why Moretti’s work has received so much attention and...
he stands, in the American academy, as a symbol of broader changes that are only beginning to be articulated by humanities scholars. It is much easier to criticise an individual than the system as a whole, especially as this is a system that humanities scholars are ensconced within and reliant upon.

With computers and computation embedded in the same rhetoric of objectivity and truth that surrounds quantitative approaches, there is also significant potential for the integration of digital methods in humanities scholarship to reinforce and compound institutional trends. In a review of *A Companion to Digital Literary Studies*, Scott Hermanson echoes many of the concerns above, but locates them specifically in relation to digital research. Hermanson worries that ‘data-driven scholarship’ will be ‘misconstrued as more valuable or more legitimate because it relies on hard numbers’. In a social and institutional context where the humanities ‘have lately struggled…adequately [to] explain themselves to outside viewers’, Hermanson is concerned that ‘this type of data-driven research becomes elevated above others because it is easy to sell, quantifiable, and a product of exact numbers. The danger exists’, he says, returning us to the issue of reductionism discussed above, ‘in privileging the 1 and the 0 and obscuring the infinite gradations in between’. The strategies for visualising data that modern technology make available arguably represent the epitome of this rhetoric of truth and objectivity. In particular, the slippage between seeing and knowing functions to accord graphs and charts – what are, in essence, arguments made using visual rhetoric – the status of self-evident fact or, precisely, ‘scopic vision’. As a result, and in an extension of the division that Spivak and Arac describe between close reading’s ‘native informants’ and their quantitative masters, there is a clear possibility – with universities privileging paradigms of knowledge and funding that produce measurable outcomes and productivity – that a divide will open up between humanities scholars with the opportunities and technological abilities to frame their arguments in terms of quantitative evidence, and those without.

However, recognising the implication of quantitative and computational methods in complex and challenging power dynamics does not constitute an argument for excluding such approaches from literary historical research. If we avoided all methods implicated in difficult power relations, literary scholars would long ago have abandoned language. As poststructuralist theory emphasises, language is a form of knowledge and a means of representation that carries, in its structure, values that privilege some voices and attempt to silence others. Instead of abandoning language, literary scholars have sought to understand the ways it works and to challenge and critique the relations of power it perpetuates. We need to do the same with numbers: to recognise them as a form of representation and, as such, to explore how they operate and the ways in which numbers accrue authenticity and authority. Like language, numbers provide an imperfect and mediated way of accessing the world; but in the absence of any perfect or unmediated access, they are tools we can use in our attempts to understand and investigate the literary field.

While this sense of numbers as an imperfect and mediated representation might not be the exact way they are discussed in the sciences, no scientist approaches statistics as neutral, true and infallible. Awareness of the way scientists interrogate – rather than simply accept or promote – statistical measures is often lacking in current humanities’
debate about quantitative approaches and their ideological resonances. As part of the contemporary, corporate university system, the sciences – like the humanities – are implicated in its managerial strategies and neoliberal or economic rationalist political ideology. This does not mean that scientists adopt the positivistic approach to quantification that prevails among ‘the doxa of higher education management’. Instead, significant effort is devoted in scientific studies to addressing the ‘problems of categorisation, bias, rhetorical presentation and distortion’. (Pat Hudson provides this list, her point being that these precise problems afflict ‘detailed description or narrative’ as much as ‘quantitative approaches’.)

On one level, developing a more sophisticated and theorised understanding of numbers and statistical analysis is necessary for the humanities whether or not such approaches are employed directly. This is simply because of the status of quantification and, increasingly, computation as central regimes of knowledge and forms of power in contemporary society. In this context, understanding the way social relations are organised and institutions function requires a framework for engaging critically – rather than contemptuously or fearfully – with quantitative forms of representation. On another level, such engagement is worthwhile because – as I will discuss in more detail in what follows, and as the case studies in this book aim to demonstrate – quantitative methods allow us to explore aspects of the literary field, especially trends and patterns, broad developments and directions, that would otherwise remain unrepresented and unrepresentable.

For both purposes – and particularly because, rather than in spite, of the potential collusions between quantitative analyses and various forms of neoliberal ideology – literary scholars need to reconceptualise data and computation not as inevitably reductive and absolute regimes of power, but as products of theoretical processes and decisions, and as means of argumentation and theorisation. Although there has been little direct conversation between the two fields, the methodological underpinnings of both book history and the digital humanities signal important directions for developing a critical and theoretically aware approach to working with data – one that has significant potential for quantitative literary history, and for the humanities generally. Importantly, the methodological frameworks of both fields either anticipate and avoid, or answers, the criticism of quantitative approaches I have outlined above, while at the same time circumventing blind spots that emerge in the defence of close reading.

II Critical Quantification: Book History and the Digital Humanities

Many quantitative book histories begin with the deceptively simple statement that, like all cultural fields, the literary one includes features that ‘cannot be “counted”’, such as ‘the reading experiences of an individual’ and the ‘quality’ of a literary work. But it also contains elements that ‘can be quantified…for example, the number of books printed; the number of books sold, the quantities of books exported’. In such accounts, instead of qualitative and quantitative methods being inherently divided and opposed practices, they become, in Darnton’s words, ‘a matter of perspective’, their use suited to the investigation of different aspects of the literary field. Close readings can reveal
information that numbers cannot just as ‘[s]tatistics can reveal configurations and proportions that escape other kinds of observation’. As Joshi writes, ‘rather than forcing a divide between quantitative method and literary study, between statistics and cultural understanding’, quantitative book historians acknowledge the potential of each approach to ‘enhance the other’.

These book historians do not deny that quantitative methods suppress the ‘particularities’ and ‘singularities’ of individual literary works, as critics of such methods contend. However, they do argue that the trends and patterns – the ‘generalisations’ – that emerge from this process justify the loss of detail. As Jonathan Zwicker says:

[N]umbers flatten out the peculiarities and individuality of their object, but this is also part of their value, they ‘simplify the better to come to grips with their subject’ and so make accessible – through patterns and series – solutions to problems that are virtually inaccessible through the methods of traditional literary history.

In similar terms, while acknowledging the vital role of detailed case studies for publishing history, Simon Eliot insists that:

Any number of individual studies would not be sufficient, because you could never be certain that you had assembled a reliable sample that did justice to the particular period or area you were studying. Also the individual studies need a context to confer on their details a proper significance.

Quantitative methods, in other words, do not tell us everything about the literary field; but they provide a way of exploring aspects of that field that could not be investigated by other means. From this perspective, identifying one approach as inherently better simply results in ‘an impoverished understanding of a [complex] phenomenon’.

It may seem almost bizarre for book historians to devote so much energy to insisting on the presence of both quantitative and qualitative features of the literary field and, hence, the value of both qualitative and quantitative forms of understanding. Indeed, Literary Darwinist Jonathan Gottschall simply dismisses the dichotomisation of methodological debate:

To argue for the superiority of quantitative over qualitative approaches (or vice versa) would be as vacuous as arguing that hammers are better than drills. As the carpenter requires a collection of widely varied and subtle tools for effectively confronting widely varied challenges, so too does the scholar.

Yet in the broader controversy about quantitative methods, it is often difficult to determine what value, if any, Moretti and his critics accord to the methodology other than the one they champion. For instance, in her critique of Moretti’s work, Trumpener identifies book history as a potential ‘middle ground’ between statistical methods and close reading, and describes as ‘brilliant’ the ‘statistical work of bibliographers and book historians like Peter Garside and James Raven’. However, the type of book history she
subsequently delineates and advocates is one that replicates the individualised focus of close reading. Thus, she valorises the book historian who endeavours ‘to figure out, book for book, who determined’ – in this case – ‘each novel’s title: author, publisher or publicist’. As Trumpener puts it, this approach involves ‘real footwork – and…commitment to specific novels’.\(^{58}\) Trumpener thus appears to endorse, or at least allow space for, quantitative approaches to literature (admittedly in a book history realm that she separates from literary history). However, the moral associations she draws between the study of particular texts – signifying commitment and hard work – and quantitative approaches – which ‘violate’ the integrity or ‘individuality’ of the object of study – denigrates any form of history not based on direct acquaintance with each literary object.\(^{59}\)

Moretti’s views on the appropriate fate of close reading are similarly opaque. Sometimes, like the book historians above, he appears to accord value to qualitative as well as quantitative approaches. In ‘Conjectures on World Literature’, he notes, ‘[r]eaders “more” is always a good thing’,\(^{60}\) and describes ‘distant reading’ as ‘a little pact with the devil’.\(^{61}\) In the introduction to Graphs, he claims, ‘for me, abstraction is not an end in itself, but a way to widen the domain of the literary historian, and enrich its internal problematic’.\(^{62}\) ‘Distant reading’, from this perspective, is a way of addressing particular questions. Rachel Serlen, who has written a detailed critique of Moretti’s oeuvre, notes that such instances – where Moretti ‘appears to say that both methods can peacefully coexist’ – are the ones ‘[h]is most sympathetic critics’ seize upon.\(^{63}\) Timothy Burke, for instance, contends: ‘There is no requirement to purchase the entire methodological inventory [Moretti] makes available, or to throw overboard close reading’.\(^{64}\)

But, Serlen continues, ‘In more recent restatements of the problem…Moretti takes the more radical stance that the distant turn he advocates entails the rejection of interpretation’.\(^{65}\) In one of his many responses to the commentary his work has evoked, Moretti asserts: ‘Between interpretation (that tends to make a close reading of a single text) and explanation (that works with abstract models on a large groups [sic] of texts) I see an antithesis. Not just difference, but an either/or choice’. He makes this proclamation even while acknowledging that, ‘[i]t may be tactically silly for me to say so now, given that the general consensus is that what I do could be interesting, as long as it doesn’t want to get rid of current procedures’.\(^{66}\) As Moretti elsewhere describes ‘close reading’ as less ‘rational’ than distant reading,\(^{67}\) this ‘either/or choice’ between interpretation and explanation – like Trumpener’s association of close reading with commitment and hard work – raises methodology from modus operandi to moral imperative. Where Trumpener’s moral/methodological framework suggests the Protestant work ethic, Moretti’s invokes the Kantian view of rationality as the basis of morality. In light of this morally loaded bifurcation of debate about qualitative and quantitative methods, the careful and pragmatic insistence by book historians that different questions – and different features of the literary field – require different approaches appears, far from simple-minded or bizarre, astute and necessary.

As I have said, critics such as Spivak have attacked what they see as Moretti’s false claim to objective truth and totalising knowledge. To my mind, this is the most pertinent criticism of Moretti’s approach, and I want to spend some time detailing its elements, and adding my own criticism, before describing how book history’s approach to data...
avoids this charge. While Spivak broadly criticises Moretti’s assumption of scopic vision,68 other commentators focus on two aspects of his work that they argue underpin this assertion of absolute knowledge: his definition and use of data. John Frow criticises Moretti’s understanding of data, arguing that he takes ‘genres or forms as given and then derive[s] structures from large data sets based on them in such a way that literary history can be conceived as an objective account of patterns and trends’. In this process, Moretti ‘ignore[s] the crucial point that these morphological categories he takes as his base units are not pre-given but are constituted in an interpretive encounter by means of an interpretive decision’.69 In other words, although Moretti’s data – which include such abstract concepts as clues in detective fiction, free indirect style,70 and formal compromise71 – are the results of the subjective process of reading, his analysis disregards ‘that moment of interpretive constitution of the categories of analysis’ to produce a historical approach that is uncritically positivist.

Related to this, Serlen highlights Moretti’s ‘ad hoc’ categorisations of data, and proposes that he ‘runs the risk of identifying genres and devices whose totality is as artificial as that of the individual texts he is trying to displace’.72 I would add that, even when Moretti uses data constructed by others, he is inclined to overplay its accuracy. For instance, in the first chapter of Graphs he claims that: ‘[q]uantitative data can tell us when Britain produced one new novel per month or week or day, or hour for that matter’.73 Moretti may intend this claim to refer to average levels of production; or to the possibilities of future datasets, more complete than current records of eighteenth- and nineteenth-century British book publication. Yet in not making these distinctions, and instead asserting an impossible level of precision – it is certain that not even the titles of all British novels are recorded, let alone the hour of their production – Moretti accords to his results an accuracy and objectivity they cannot (and probably can never) possess.

This quote regarding the quantitative data on British book publishing (which continues, ‘where the significant turning points lie along the continuum – and why – is something that must be decided on a different basis’) is one that Serlen uses as an example of where Moretti acknowledges the subjective nature of his interpretations.74 Quantitative research provides, he argues in this case, ‘data, not interpretation’.75 However, she sees this example as exceptional, arguing that Moretti generally describes his method in ways that downplay or occlude the subjective and interpretive aspects of data analysis:

While Moretti’s own work shows interpretation to be as important to distant reading as the accumulation of data…interpretation is curiously elided in his descriptions of distant reading as a method…made to seem easy, natural – the inevitable result of the accumulated data.76

This minimisation of interpretation is another way Moretti presents his arguments as objective descriptions or – the word he prefers – explanations of the literary field.

Significantly, and resonating with Spivak’s charge of ‘scopic vision’,77 Moretti’s elision of the subjective process of interpretation is particularly apparent in respect to the visualisation of data. Quantitative analysis and the visual models it enables are presented
not as representations and arguments but as essentially transparent windows into the literary
text or the field more broadly. At the beginning of *Graphs*, Moretti announces: ‘graphs, maps,
and trees place the literary field literally in front of our eyes – and show us how little we
still know about it’. Elsewhere he describes these ‘three models’ as ‘three snapshots of the
literary field’, implying the capacity of a camera’s lens to capture the view (with the speed
with which these snapshots are taken further minimising the process of interpretation). If
these visualisations present the literary field itself, then seeing what is there – and explaining
it – can be substituted for the subjective processes involved in collecting and constructing
the data for these graphs, and in deciding how the data will be visualised and what it means.
As Serlen remarks, ‘[h]ow the data get interpreted is replaced by how the data are seen,
which minimizes the explanatory work involved in distant reading’.

Another consequence of what Spivak calls Moretti’s ‘claim to scopic vision’ is the
transition this assumption supports from context-specific, socio-historical accounts
of literary data to explanations that are ‘global rather than local’. Michael Friedman
describes a global focus as characteristic of ‘scientific explanation’: what is explained in
such accounts is ‘a general regularity or pattern of behavior – a law if you like – i.e. that
water turns to steam when heated’. Moretti aims in *Graphs* to reveal the ‘hidden thread
of literary history’: its cycles. As ‘[v]ariations in a conflict that remains constant’, these
cycles operate separately from and beyond the social world and events in it. Moretti ties
this focus on cycles to the *Annales* school of history – an approach, as I said earlier, that is
also foundational to book history. However, where the *Annaliste* historians investigate the
persistence of cultural ideas (or *mentalities*) over long stretches of time (as well as the effects
of geographical and geological realities on the social world), Moretti aims to discover
general regularities or patterns that, like global scientific ‘laws’, exist beneath or beyond
social or historical context.

This understanding of cycles as literary ‘laws’ leads Moretti, in some cases, to neglect
obvious differences in context when comparing literary fields in different places and times
and, in other cases, to propose explanations that relate changes in the literary field solely
to features internal to that field. In relation to this first tendency, Moretti attributes the
‘pattern’ that emerges in the first graph in *Graphs* (the much-discussed rise of the novel in
five separate countries) to a general theory of

the horizon of novel-reading… As long as only a handful of texts are published each
year…the novel is an unreliable commodity: it disappears for long stretches of time, and
cannot really command the loyalty of the reading public; it resembles a fashion, more
than a literary genre. With a new text every week, however, the novel becomes that great
modern oxymoron of the *regular novelty*: the unexpected that consumers expect so often
and eagerly that they can no longer do without it.

Leaving aside the issue of whether five instances constitutes a pattern – and disregarding
the important fact that novels are published and read in many forms besides the book – this
hypothesis might have relevance to understanding growth in the production of British novels
in the early eighteenth century. However, in identifying the same ‘horizon of novel-reading’
in late nineteenth-century Nigeria, for instance, Moretti overlooks the vital contextual
point that, while Britain may have had a relatively isolated book market, there were novels circulating in Nigeria before the 1960s and the emergence of the Nigerian novel. Thus, the same relationship between the publication of novels by Nigerian authors and the reading habits of those in that country – Moretti’s ‘horizon of novel-reading’ – is not inevitable, and cannot be assumed.

Elsewhere, Moretti’s search for literary cycles leads him to propose explanations that are, as Serlen puts it, ‘purely internal to the formal structure of the literary object’: what she calls ‘depoliticized form’. For instance, in Moretti’s explanation of gender trends in authorship, the novel – and other, unspecified, literary forms – become not only active, but the only participants in this literary cycle. Noting the various rises and falls in the number of novels by British men and women in the eighteenth and nineteenth centuries, Moretti asserts:

[I]f the conflict remains constant, then the point is not who prevails in this or that skirmish, but exactly the opposite: no victory is ever definitive, neither men nor women writers ‘occupy’ the British novel once and for all, and the form keeps oscillating back and forth between the two groups...allow[ing] the novel to use a double pool of talents and of forms, thereby boosting its productivity, and giving it an edge over its many competitors.

I agree with Moretti that gender trends in authorship tend to oscillate. Certainly, the Australian data shows this same ‘back and forth’ movement between a predominance of novels by men and women. However, in abstracting this phenomenon from the social world, Moretti’s explanation adds little to our understanding of the operations of gender in the literary field. While it seems clear that, throughout history, writing by men and women has been constructed and received differently, Moretti presents this gendered distinction and dichotomisation as itself a law: men and women have different ‘talents’ and produce contrasting ‘forms’ across space and time. The British novel is personified, and plays men and women off against each other to gain what Moretti proposes as an evolutionary advantage over other literary forms (‘its many competitors’). Although Moretti defines ‘[f]orms...as the abstract of social relationships’, such that ‘formal analysis is in its own modest way an analysis of power’, in his explanations of gender trends Moretti substitutes a consideration of social power relations for a seemingly uncritical mixture of literary statistics and evolutionary and economic theory.

Such explanations, as Serlen argues, make it ‘unclear what weight historical forces can have if their results cannot change’, or, we might add, if they change randomly. Moretti has responded to criticisms of his use of scientific paradigms to explain literature by asking: ‘why on earth should I drop a perfectly plausible explanation? Because it sounds politically wrong? My problem with these particular explanations is not that they are politically incorrect, but that they miss important features of, or do not add to our understanding of, literary change. This step that Moretti takes in abstracting literature from its social and historical context occurs because, as critics like Frow and Serlen argue, Moretti imagines his data as being objective, in the sense of being separate from the social world. Admittedly, Moretti moves very quickly through a number of
different historical and literary trends, and is, as he says, more interested in ‘opening new conceptual possibilities…than justifying these in every detail’. 90 However, in some cases, the conceptual possibilities that are offered do not seem to me to attain the level of understanding and insight that should be the aim of humanities scholarship. In this book I offer an alternative approach to both of the issues I have just addressed in relation to Moretti’s work: that is, a form of ‘distant reading’ that attends to the complex historical, social, geographical, political and economic factors involved in the rise of the novel and its relationship to reading communities, and in the shifts or oscillations in gender trends in authorship.

These criticisms of Moretti’s ‘claim to scopic vision’ describe an association between a conception of data as a direct reflection of what is in the literary field, and a form of analysis that proposes to tell and, indeed, to show the truth of that field. Methodological discussions in quantitative book history work against both tendencies, acknowledging that literary statistics are mediated and limited, and that the understanding or knowledge gained from such studies is necessarily partial and qualified. Eliot provides an apt summary of the understanding of data in book history when he asserts:

The past has left us some data, but they were not produced in laboratory conditions; they were not designed to answer our questions; they were not collected as a representative sample – and they rarely used a classification system that we might find at all helpful.

However, they are all we have got and we must work with them. 91

Such acknowledgements of the limitations of data are often ‘pounce[d] upon’, Joshi says, by ‘quantitative history’s detractors…as “further” evidence of the dubious value of statistical methods’. 92 However, in book history, this understanding simply demonstrates the need for a clear and detailed account of the origins, biases and limitations of literary historical data. This is a process that Eliot describes as providing the ‘biography of the [data] source’:

If we are to use our sources well (that is, exploit them to the full without asking them to bear a weight of interpretation that they are not strong enough to carry) we need to know our sources well: who compiled them, why they were compiled, and how they were compiled. 93

All archives are the outcome of what Frow terms ‘interpretive encounter[s] by means of…interpretive decision[s]’. 94 The biography of an archive – data-based or otherwise; historical or modern – attempts to identify the ideas, values, definitions and meanings, the theories and biases, that underpin and produce the collection, so as to enable a more critical and astute reading of the information it contains.

As I said in the introduction, the new history of the Australian novel I present in this book draws predominantly on the AustLit database, a non-profit, electronic archive that it announces on its website, ‘aim[s] to be the definitive virtual research environment and information resource for Australian literary, print, and narrative culture scholars, students, and the public’. 95 Created in 2000, AustLit merged a number of existing specialist...
databases and bibliographies, and has subsequently involved well over a hundred individual researchers – from multiple Australian universities and the National Library of Australia – in an effort to ‘correct unevenness and gaps in bibliographical coverage’ and continually update the collection. The database has received significant government and institutional funding and support, and includes bibliographical details on hundreds of thousands of works and authors.

AustLit is well suited for quantitative analysis. The database has a high degree of comprehensiveness, due to the substantial and longstanding investment of money and scholarly energy in its creation, and because of the relatively recent origins of ‘Australian literature’. Its construction according to established bibliographical standards and fields, which give the data a high degree of consistency across the collection, also facilitates quantitative and computational approaches. However, none of this implies that the data in AustLit is complete or perfect. I focus on novels predominantly for historiographical reasons: they are the fictional form most directly tied, in general as well as academic discussion, to the state and status of Australian literary culture. But there is also an important pragmatic reason for this focus: novels are the most comprehensively recorded fictional form in AustLit. Even so, not every Australian novel is included. AustLit notes that its ‘coverage of some popular fiction genres such as westerns and romances, and of self-published works, is representative rather than full’. Likewise, although over a thousand different periodicals are indexed, this coverage is not comprehensive.

My datasets are also not identical with AustLit’s. This is partly because there are instances where I have chosen to exclude titles that AustLit includes, such as entries for ‘Non-AustLit Novels’ and novels by overseas authors included because they were banned in Australia. There is also a considerable amount of data in AustLit that my study does not explore. In particular, while many titles are reprinted multiple times, due to the complexity of this dataset and the certainty of substantial gaps in coverage, I only consider the first publication in book and/or serial form. Most significantly, as an online rather than a print bibliography, AustLit is updated continually as Australian authors write more fiction and as historical authors and works are included or excluded. I updated my datasets approximately every six months during the four years of this project. This process enhanced my awareness of the fluid nature of the dataset and, in particular, of the adjustments – including retrospective ones – that occur as AustLit modifies its parameters for inclusion or its interpretation of them. But it also confirmed for me the general stability of this collection, in that these adjustments did not change the overall trends.

Even leaving aside the authors and works not discovered by AustLit, and the differences between this dataset and my own, in a fundamental and important way, it is impossible for any bibliographical record of the Australian novel to be complete. As AustLit acknowledges, ‘[t]he definition of “Australian” and “Literature” moves according to current debates and changing reading, teaching and research patterns’. AustLit, in other words, is engaged in an ongoing process of representing and constructing the category of Australian literature, including the Australian novel. The complexity of this process of construction comes to the fore in relation to the question of who or what is an Australian author. Drawing on a set of parameters for defining an Australian author – including
such considerations as where they were born, where they spent their formative years, and
the content of their fiction – decisions about which texts and authors to incorporate
are made on an individual basis. The effects of these decisions are particularly apparent
for nineteenth-century records. In certain cases, some of an author’s titles are included
while others are not (as with novels by Fergus Hume and B. L. Farjeon, for example). However, Australian literature’s status as a constructed category is true for all periods: as Maryanne Dever says of the 1920s and 1930s, ‘the concept of an “Australian Author”… was by no means a fixed and fully-constituted category’. More broadly, the notion of Australian literature as it currently stands is a product – like the AustLit database itself – of a particular period and paradigm of cultural nationalist research into, and funding for, Australian literature. Indeed, a peculiar irony of this book is that, while I criticise aspects of this cultural nationalist ideology (especially in Chapter 3), this new history of the Australian novel is only possible because of the research and infrastructural outcomes of that paradigm.

In book historical accounts, awareness of the inevitably constructed or mediated and limited nature of any cultural data strongly tempers the type of knowledge or understanding scholars claim for their quantitative analyses. This point is often made through metaphors that relate the results of quantitative studies to other forms of partial representation. Darnton, for instance, compares the ‘general picture of literary culture’ provided by book historical data ‘to the early maps of the New World, which showed the contours of continents, even though they did not correspond very well to the actual landscape’. While acknowledging that Moretti’s analogy of literary history with ‘the maps of Africa of a century and a half ago’ – quoted at the start of this chapter – is from his earlier work, comparing it with Darnton’s map provides a salutary demonstration of the different epistemological claims made for quantitative analysis by the two historians. Both refer to historical maps, but whereas for Darnton the ‘general picture’ offered by statistics is the point – because literary data is inevitably ‘flawed or distorted’ – Moretti’s analogy implies that quantitative analyses will provide literary scholars with the framework to fill in the map and ‘pinpoint’ the source of changes in the ‘coastal’ canon (in effect, regarding quantitative analyses as GPS technology).

Similar metaphors to Darnton’s appear in many other descriptions of the potential and limitations of quantitative book history. Joshi builds on Darnton’s analogy of statistics and maps of the new world, arguing:

[Q]uantitative methods expand literary history and make all sorts of discoveries possible, much the way that early maps did in the dissemination of knowledge about ‘new’ worlds. Statistics, like maps, are indeed lies to some extent…but they are lies that tell a truth that would not otherwise be evident.

Discussing ‘[l]iterary statistics from a poorly documented book culture’ – such as the Nigerian novels she explores – Wendy Griswold argues that quantitative analyses are ‘like a very rough sketch: some of the lines may be off, but a picture emerges anyway’. These metaphors in book history serve to acknowledge the limitations of literary data and present the results of quantitative analyses as indications, rather than proof, of historical
trends. At the same time, in emphasising the wider perspective such studies enable, these metaphors maintain the importance of quantitative perspectives. The refusal of an objective standpoint does not, in other words, slide into a claim of equivalence between qualitative and quantitative studies. Any reading of statistics is, like any reading of a text, a subjective process of selection and decision-making; and in both cases, there are readings that are more accurate and enlightening than others. But for identifying literary trends over time, quantitative analyses enable a broad, historical and comparative perspective not achievable based on studies of particular texts or publishers’ records.

These discussions in book history – regarding the mediated nature of data and the form of knowledge that quantitative analyses make possible – provide an important critical framework for data-rich literary history. But they have a significant blind-spot: although the majority (if not all) recent quantitative book histories use computers and computational techniques, especially in collating and visualising data, their methodological framework does not acknowledge the adoption of this technology, let alone reflect on its methodological and epistemological implications. Accordingly, while these book historical studies show, for instance, precisely why the archive is not an unmediated repository of information, computational processes are rendered entirely transparent. The nature of the metaphors employed to describe the visualisation of data makes this assumed lack of mediation apparent: both quill on parchment (to produce the historical maps) and pencil on paper (for Griswold’s ‘rough sketch’) signify forms of representation where there is no apparent intermediary between input and output. As well as determinedly analogue, these metaphors present the visualisation of data as a final (and singular) end product, whereas computer visualisations – like the data on which they are based – can exist in temporary and transitional, as well as multiple and transferable, forms.

A failure to acknowledge – let alone interrogate – the implications of working with computers is not unique to book history, but occurs throughout the humanities. While the established view is of humanities scholars and technology as ‘virtual strangers’, the ready acceptance of the computer as simply a tool – a ‘system to deliver results’, entirely separate from analysis and no more worthy of mention than the use of a word processing program to prepare an article for publication – suggests that ‘we are all too comfortably at home in the digital’. As Rothberg continues, lacking the critical distance to question and ‘defamiliarize powerful technological framings’, there is a significant risk that humanities scholars will not perceive computation for what it is: a new set of representational and epistemological practices and processes, whose adoption has profound consequences for humanities scholarship, and requires careful consideration.

This perception of computers, to use Willard McCarty’s formulation, as ‘knowledge jukeboxes’ – that simply play whatever is loaded into them – has two major implications. First, it reinforces the same rhetoric of objectivity and certainty for computational approaches that book historians have worked so hard to challenge in relation to quantitative analyses. Given that computational approaches always involve quantification, and that quantitative studies are increasingly carried out with the computer, this uncritical view of computational analyses has the potential to cancel out the important methodological insights of quantitative book history. This uncritical understanding of computers as providing objective and certain information – rather than

This chapter has been published in the volume "Reading by Numbers: Recalibrating the Literary Field" by Katherine Bode. London: Anthem Press, 2012. ISBN: 9780857284549.
a way of knowing (or ‘prosthetic extensions...for critical reflection’, as Jerome McGann puts it)\textsuperscript{120} – also unconsciously replicates the understanding of knowledge production privileged in contemporary university management practices. As McCarty writes, the ‘knowledge jukebox’ view of computers ‘harkens to the commodification of knowledge as something that can be packaged in units, stored somewhere else and delivered to a consumer or dispensed from a machine on demand’.\textsuperscript{121} Second, in viewing computers as simply tools, humanities scholars risk not taking full advantage of the critical possibilities computers can enable, nor of playing an active part in developing such approaches ‘in ways we wish to develop them’.\textsuperscript{122}

Although the theorised nature of data does not receive the same focused attention in the digital humanities as it does in book history, both fields share a basic view of data as shaped by ideas and values circulating in the world. For instance, Cathy Davidson’s description of ‘data collection...[as] data selection’, as well as her contention that ‘data transform theory; theory, stated or assumed, transforms data into interpretation’,\textsuperscript{123} clearly resonate with many of the claims I have discussed in relation to quantitative book history. In integrating what might be called the analogue conception of quantitative analysis in book history, with an understanding of how working with a computer informs and remediates my approach to data, I have found McCarty’s notion of modelling particularly useful. Although McCarty develops this approach for use with language, his description of modelling – as an exploratory and experimental practice, aimed not at producing final and definitive answers but at enabling a process of investigation and speculation – can be adapted for quantitative analysis of literary historical data, and its visualisation in particular.

Like quantitative book historians, McCarty acknowledges that his method ‘obfuscates difference’: a ‘model of something’ is, of necessity, ‘an abstraction or simple representation of a more complex real phenomenon’. And like book historians, he justifies this process of reduction because it facilitates the development of a form of understanding and knowledge that would not be possible by other means. Indeed, McCarty argues that models are necessary precisely because the object of study – such as, in my case, historical trends in the production and reception of Australian novels – is otherwise ‘inaccessible or intractable’.\textsuperscript{124} While this aspect of modelling resembles the approach to data and quantitative analysis in book history, what I find particularly enabling about McCarty’s methodology is the emphasis it places on the analysis and representation of data as a process of knowledge production and experimentation, enabled by the computer.

McCarty emphasises the status of models as ‘pragmatic instruments of investigation’.\textsuperscript{125} Where Moretti’s ‘abstract models’ present transparent windows into, or snapshots of, the literary field – and where book historical metaphors emphasise partial but completed indications of historical trends – McCarty describes models as ‘experimental device[s]’, and as constructs or stages in a ‘process of coming to know’.\textsuperscript{126} To ground this notion in an example from my own research, the publishing data for a particular period in the history of the Australian novel might lead me to suppose a particular influence on the field: for instance, that a particular government funding model, supporting a particular group of local publishers, enabled the rise of a particular Australian novel genre. The manipulability of a digital representation, and the fluid nature of the computing environment more
broadly, encourages and enables modification of, or experimentation with, data – for instance, subtracting particular publishers or genres – as a way of testing such hypotheses. This process might fulfil my expectations – thus strengthening my original hypothesis – or challenge them, bringing that hypothesis into question and compelling me to seek an alternative account of the data. Although the codex form does not allow me to display the literally hundreds of models I developed in exploring the history of the Australian novel, this practice of trying and testing a range of hypotheses underpins all the graphs, and informs all the arguments, that I present in the pages that follow. I have also made the datasets used for this book available, so that others can explore and experiment with them, and in doing so, check, extend or challenge my categorisations, visualisations and interpretations.

Modelling as McCarty theorises it, then, is directed at ‘making new knowledge’ using quantitative information and a digital research environment. However, the emphasis on manipulation also transforms the meaning of these representations. Models are built to be modified, and this process of modification emphasises their status as fictions: ‘not only by being a representation, and so not the thing itself, but also by selective omission and perhaps by distortion or inclusion as well’. As a result, what is emphasised in modelling is not knowledge as an end product but the development of knowledge as an ongoing process. As McCarty elaborates:

The drastically reduced investment in an obviously temporary product, plus the means at hand to alter it immediately, mean that one is much less likely to mistake this product for a true or final representation, indeed unlikely to think that any such product would ever reach perfection.

No matter how complete the graphs in this book might appear, as McCarty says of his models, they ‘are better understood as temporary states in a process of coming to know rather than fixed structures of knowledge’.

This description of modelling resonates productively with AustLit’s status as an online and, hence, fluid archive. While the established comprehensiveness of AustLit remains important in this context – as it means the data used for modelling will not change radically and render the process of hypothesis testing so abstract as to be redundant – McCarty provides a framework wherein potential shifts in the data do not disallow quantitative analysis. To put this another way, where the perception of a graph as a final product would prevent – or at least, significantly curtail – the critical potential of mining, modelling and visualising an online (and hence changeable) database, modelling enables a relationship between data and argumentation that is, explicitly, an ongoing and evolving one. Data is contingent: inherently due to its constructed nature and, as is increasingly the case, because of the online environment in which it is presented. Understanding data representation as a process of research rather than an end-product signifies a quantitative approach that resonates with the humanities process of interpretation: an approach that is explicitly contingent and speculative while remaining critical and committed to scholarly rigour.

***

This chapter has been published in the volume “Reading by Numbers: Recalibrating the Literary Field” by Katherine Bode. London: Anthem Press, 2012. ISBN: 9780857284549.
As I have proposed in this chapter, there are some who view all quantitative approaches to literature as pointlessly reductive, falsely totalising in their knowledge claims, and inherently complicit with the neoliberal or economic rationalist managerial practices increasingly prominent in today’s corporate university. But any claim to knowledge (including the defence of close reading) that does not admit its partiality – and thus, inherently, the value of other ways of knowing – is bound to transform the will to knowledge into the will to power, and to enable only an impoverished understanding of any complex phenomenon. Combined, the book history and digital humanities approaches I have discussed offer a critical way of working and thinking with literary and digital data that does not fall into this trap. Quantitative book histories provide a framework for acknowledging the limitations of data while upholding the importance of analyses based on empirical evidence. As Eliot says, such studies are not ‘exercise[s] in justifying the use of any figures at any time in any context’. Rather, they propose that, ‘interpreted cautiously and used intelligently’, literary data indicates trends in the literary field that cannot be investigated otherwise.\textsuperscript{131} The digital humanities – and McCarty’s method of modelling in particular – extends the possibilities of quantitative analysis by outlining a speculative and experimental approach to computation and data visualisation. The new history of the Australian novel I offer in the pages that follow draws on both sets of approaches to provide an account that is, of necessity, partial. However, due to the perspective enabled by quantitative representation and analysis, it is also an account that enables insights into the literary field: insights that challenge established interpretations and offer new understandings of the history of the Australian novel.