Please cite the Published Version

Thomas Gray, Samuel Taylor Coleridge and Geographical Information Systems: A Literary GIS of Two Lake District Tours

Ian N. Gregory and David Cooper

Abstract:

There have been growing calls to develop the use of Geographical Information Systems (GIS) across the humanities. For this shift to take place, two things must be demonstrated: firstly, that it is technically possible to create a useful GIS of textual material, the main medium through which humanities research is conducted; and, secondly that such a database can be used to enhance our understanding of disciplines within the humanities. This paper reports on a pilot project that created a GIS of two textual accounts of tours of the Lake District: Thomas Gray's 1769 excursion and Samuel Taylor Coleridge's self-styled 'circumcursion' of 1802. It describes how these accounts were converted into a GIS and explores various methodologies that can then be used to explore the spatialities embedded within their respective tours. The resulting material has been placed on the internet at http://www.lancs.ac.uk/mappingthelakes. The pilot demonstrates that it is both possible, and conceptually fruitful, to move GIS beyond the quantitative arena in which it currently resides and into more qualitative areas of humanities research.

Bio: Ian Gregory is a Senior Lecturer in Digital Humanities at Lancaster University. His main area of expertise is the use of GIS in the humanities, the subject of his recent book *Historical GIS: Techniques, methodologies and scholarship* (CUP, 2007 with P.S. Ell), as well as numerous journal articles.

David Cooper's research focuses on the literature of place and space, with a particular emphasis on the spatial history of the Lake District. His current projects include work on the

intersections between literature, cartography and the environmental note-taking of Samuel Taylor Coleridge. He teaches English literature at Lancaster University.

Introduction

In recent years the use of Geographical Information Systems (GIS) has become increasingly widespread in historical studies and, as a result, Historical GIS is now a widely accepted field of scholarly research.[1] This field has well-established technical and methodological underpinnings and, critically, is making significant contributions to knowledge across a wide range of topics, including environmental history,[2] economic history,[3] historical demography,[4] medieval history[5] and urban history[6]. Whilst this represents a major achievement for a nascent scholarly field, the impact of GIS, to date, has been significantly circumscribed by the fact that its databases are best suited to handling information in tabular form. This represents a significant limitation as most of the sources used by historians are unstructured text. Over recent years, however, there have been increasing calls for the development of a Humanities GIS: a development which requires the use of GIS to spread from quantitative to qualitative historical research and even literary studies. Clearly, for this shift to take place, there is a need for GIS to operate with textual data. The first challenge, then, is technical; but, alongside this, Humanities GIS must reach out to a wider scholarly constituency by demonstrating that the use of GIS can make a significant contribution to knowledge.

This paper describes a pilot project that explored the potential for a GIS in literary studies. It focuses on two historical accounts of two tours of the English Lake District. The aims are two-fold: to demonstrate that it is technically feasible to build a GIS of unstructured text; and to demonstrate the techniques and approaches that open up critical and conceptual potentiality for future mappings of literary texts. The principal aim of the pilot project was to establish an overarching methodology for literary GIS; and, through the mapping of texts by

Thomas Gray and Samuel Taylor Coleridge, it is hoped that we have provided the technical foundations for further theory and praxis.

Challenges for a literary GIS

Although most literary texts are set in particular geographical locations at specific historical moments, there has been a tendency for literary criticism to privilege the representation of temporality over issues of spatiality. This privileging of time over space has been recognised by P. E. Wegner, who argues that: 'Characters are fundamentally temporal constructs that unfold in a space, or "setting", which, once established, seems to remain constant.' [7] In other words, geographical setting is seen to be of secondary importance to the 'unfolding' of individual characters over time. The particular object of Wenger's critical attention is the novel; but his argument can be applied to the criticism of all literary genres, as evidenced by the fact that New Historicism remains the dominant theoretical mode for critical responses to Romantic poetry. [8]

As Wenger points out, however, literary criticism has fed off, and back into, the so-called 'spatial turn' across the humanities: 'an emerging interdisciplinary formation centred on the problematics of "space", "place" and "cultural geography"'. [9] This conceptual preoccupation with spatial experience, and, by extension, its textual representation, has been manifested in numerous ways, many of which have emerged directly out of mid-twentieth century phenomenological thinking. A key sub-strand within this field of 'spatial criticism' has involved the mapping of geo-specific literature: a cartographical mode of criticism which can be located in M. Bradbury's *The atlas of literature* and which was developed further by F. Moretti in his seminal work, *Atlas of the European novel 1800-1900*. [10] For both Bradbury and Moretti, maps are used to chart the movement of characters through places and spaces; maps are used to reveal the patterns and trajectories which are embedded within the textual representation of space and everyday spatial practice. Although both Bradbury and Moretti

are primarily interested in the novel, a cartographical approach to literary texts clearly offers particular potential for the critic interested in travel writing: the textual documentation of the actual movement through named topographies.

In a more recent monograph, *Graphs, maps, trees*, Moretti develops his interdisciplinary approach to critical practice through the articulation of a model of 'distant reading': a mode of critical engagement 'in which the reality of the text undergoes a process of deliberate reduction and abstraction'; a mode in which the critic is able to develop an understanding of overarching patterns and connections across a large corpus of primary texts. [11] Moretti, in developing this 'abstract' methodology, draws upon the representational tools of scholars working in other disciplines: 'graphs from quantitative history, maps from geography, and trees from evolutionary theory'. [12] Graphs are used, for example, to represent the shifting popularity of different fictional forms (the epistolary, gothic and historical novels) between 1760 and 1850; whilst a diagrammatic tree is constructed to facilitate further understanding of the generic evolution of detective fiction. Both graphs and trees seem to open up critical and conceptual potentiality through the visualisation of broad trends and literary connections; they provide a means of concretising the abstract.

Moretti's maps, however, offer a less convincing visualisation of his critical mode of 'distant reading'. Maps are used, for instance, to illustrate the representation of space in Mary Mitford's *Our village*: *sketches of rural life and scenery*: a series of sketches of spatial everydayness which was published between 1824 and 1832. The mapping of the detailed geo-specificity of Mitford's first volume, though, is necessarily predicated upon close critical engagement with the primary text; and, as a result, the practice of 'close reading' seems to be privileged over 'distant reading'. A second problem with the literary mappings of Mitford's writings – a problem in common with all conceptual research based upon traditional

cartographical practice – is that such maps are time-consuming to produce and, once completed, are not easily altered. As a result, the map is presented as a fixed, static representation of space and the opportunities for cartographical refinement and layering remain limited.

GIS has the potential to resolve both of these problems: it allows users to produce maps with relative speed and thereby facilitates the drive towards the identification of large-scale spatial patterns; and it grants users the opportunity to revisit and to revise maps throughout the research process. That said, the use of GIS also brings a number of problems of its own. The first of these is how to convert a text into a GIS quickly and effectively. This will not be answered in this paper but possible approaches which are being developed will be discussed in the conclusion. The second is that a GIS models the world by having a table of data, referred to as *attribute data*, in which each row is linked to a location on the map usually represented as a point, line, or polygon.[13] This locational information is called the *spatial data*. This data model is not directly applicable to literary studies where the ambition is to be able to link unstructured text to spatial locations. If these two problems are solved, however, a literary GIS has the potential to be a research tool that can be used to visualise the spatial patterns embedded in geo-specific texts and, by extension, to open up further conceptual thinking with regards to the literature of place and space.

The literary Lake District and the choice of texts

[Figure 1]

The Lake District, shown in figure 1, is a small but mountainous area in the north-west of England. It incorporates the only three mountains in England to reach over 3,000 feet and around ten significant lakes, the largest of which, Windermere is over ten miles long. The

area, which was designated a National Park in 1951, historically included tracts of land from the old counties of Cumberland, Westmorland and north Lancashire; but, following regional boundary changes in 1974, the Lake District National Park was absorbed into the new administrative territory of Cumbria. Despite its small size - the National Park covers an area of less than 900 square miles - the Lake District has an incredibly rich literary tradition that includes writers such as William Wordsworth and the Lake Poets, Beatrix Potter, Arthur Ransome, Alfred Wainwright, and Norman Nicholson.[14] Clearly, this imaginative literature has been influenced by the singularity of the regional topography; but, at the same time, this writing has helped to shape the ways in which this landscape has come to be perceived within the national, and even international, consciousness.

This intersection of geographical and cultural factors made the Lake District an ideal site for a pilot study on literary GIS. The selection of primary texts, however, was more problematic. As the primary aim was the exploration of potentiality, rather than the construction of a complete system, the decision was made to focus on two relatively short textual accounts of moving through the Lake District: Thomas Gray's account of a tour in the autumn of 1769; and Samuel Taylor Coleridge's nine-day walking expedition of August 1802. From a technical perspective, these texts have distinct advantages: both accounts are relatively short (approximately 10,000 words); and, as written documentations of actual tours, the respective texts contain a significant amount of geographical information on spatial data. [15] The two tours also have particular literary significance within the cultural history of the Lake District. Gray's account can be read as a proto-Picturesque text; that is to say, his representation of spatial experience contains many of the characteristics which would come to be associated with that touristic way of perceiving the Lake District landscape. Coleridge's account, on the other hand, represents a self-conscious rejection of this eighteenth-century aestheticised mode

of landscape engagement; the Romantic writer privileges the articulation of embodied experience over the position of physical detachment from landscape favoured by the Picturesque tourist.

Thomas Gray's first attempt to tour the Lake District in 1767 was abandoned when his companion, Dr Wharton, was taken ill with asthma. On the second attempt, in the autumn of 1769, Wharton again became ill; this time, however, Gray continued alone and spent 12 days touring the region. He started his tour by travelling west from Brough to Penrith, and then on to Keswick where he spent a number of days and nights. Next, he headed south crossing the central Lake District to visit Ambleside and Kendal, continuing on to Lancaster. [16] Gray, who is best known as the author of 'Elegy Written in a Country Church Yard' (1751), kept a detailed written account for the benefit of his absent friend. This account, known as Gray's 'Journal of his Tour in the Lake District', was published posthumously in 1775 and became one of the eighteenth-century's most popular guides to the Lakes.

In 1798, Samuel Taylor Coleridge and William Wordsworth published *Lyrical Ballads*: a volume which brought together poems by the two writers and which would come to be recognised as perhaps the foundational text in British Romanticism. In 1799, Wordsworth, and his sister, Dorothy, moved to Dove Cottage in Grasmere, Westmorland; and, in the following year, Coleridge relocated his own family to Greta Hall, Keswick, Cumberland. Coleridge, however, was a troubled figure: he felt suffocated by domestic life; he was increasingly addicted to opium; and his relationship with the Wordsworths was in decline. It was within this biographical context, then, that, Coleridge set out alone from Greta Hall on a walking tour of the western fringes of the Lake District in August 1802. Coleridge's experience of moving through the region was documented in his Notebooks; alongside this,

his tour was recorded in a series of letters sent to Sara Hutchinson, the woman with whom he was in love.

Building the GIS

Constructing the GIS of these tours effectively fell into four stages: digitising the texts; extracting place names from these digital texts; adding co-ordinates to the place names; and then converting this spatial data into GIS form. The first of these was relatively simple. Given the short lengths of the texts the easiest approach was simply to type them up by hand. Texts were sourced from the definitive scholarly editions: P. Toynbee and L. Whibley's *Correspondence of Thomas Gray*;[17] E.L. Griggs's *Collected Letters of Samuel Taylor Coleridge*;[18] and K. Coburn's *The Notebooks of Samuel Taylor Coleridge*.[19] They were typed up using eXtensible Markup Language (XML) tags in order to structure the texts. Tags were particularly used to identify place names. In addition, place names were flagged in an attempt to identify whether the writer actually visited a particular location – which we have termed a 'visit' – or whether he was referring to a geographical site from a physical and imaginative distance. By extension, place names that are tagged as 'visits' in one part of the text can also be tagged as 'not visited' in another part. Additional tags were used to identify the day/date of the tour, people encountered/mentioned and so on.

Having the place names tagged in XML meant that it was easy to extract this information and to insert it into a database. From here they were joined to the Ordnance Survey's (OS) 1:50,000 gazetteer, a database table that gives the names of each of the 260,000 places marked on the OS's Landranger series of maps of Britain together with a National Grid reference.[20] These grid references give the co-ordinates of the bottom-left hand corner of the one kilometre grid square in which the name of the place appears on the OS map. Adding

500m to these in both x and y gives the grid reference of the centre of the square which, in theory, means that the co-ordinate is accurate to within 500m in both x and y.

Joining the list of place names that appears in the texts to the gazetteer represented a first step towards the creation of a list of co-ordinates for every place name that occurs in the accounts. Further work then had to be carried out to deal with spelling mis-matches; and a small amount of disambiguation was required where a place could refer to more than one location. This process provided a database table that gave the name of the place as provided in the text, a standardised version taken from the OS gazetteer, and a northing and an easting also from the gazetteer. In total we identified co-ordinates for 280 places mentioned by Gray and 384 in Coleridge's account.

The final stage was to convert this data into a GIS. This is a relatively straightforward process as a database table with x and y fields can be easily converted into a GIS through software which uses the co-ordinates to create its spatial data. This provides us with a basic GIS in which the spatial data are points representing the places to which the writers refer, and the attributes include the dates on which they mention places, whether they visited those sites or not, and so on. One problem that immediately comes to light when this is mapped is the accuracy of the locations of place names. While towns and villages, mountains, and many other features can logically be given a point location based on the one kilometre grid square in which the name is centred on an OS map, rivers and long thin lakes such as Ullswater and Windermere cause problems. On the first day of his tour Gray states the following: '[P]assed Kirby-thore, Sr W: Dalston's house at Acorn-bank, Whinfield-park, Hawthorn-oaks, Countess-pillar, Brougham-Castle, … crossed the Eden & the Eimot (pronounce *Eeman*) with its green vale, & at 3 o'clock dined with Mrs Buchanan, at Penrith on trout & partridge.' [21]

This part of his account is almost a list of place names starting at Kirby-thore and finishing with the River Eamont which he calls the Eimot.

[Figure 2: The river Eden]

Figure 2 shows these places mapped in their raw form; and, in turn, this figure illustrates a number of problems with the GIS. To start with, the sequence works well with 'Kirby-thore' (Kirkby Thore in modern spelling), 'Acorn-bank' (Acorn Bank), 'Whinfield-park' (Whinfell Park), 'Countess-pillar' (Countess' Pillar) and 'Brougham-Castle' (Brougham Castle) all being located in logical places; the only difficulty with this particular mapping is that no location can be found for Hawthorn-oaks in the gazetteer. With the rivers, however, two problems come to light. The map makes it clear that Gray's textual documentation of crossing these rivers does not seem to match the sequentiality of his geographical experience. That is to say, he refers to the Eden after listing the series of place names given above, but he must have crossed that particular river between 'Acorn-bank' and 'Whinfield-park'. This is only a relatively minor problem and, crucially, it serves to raise salient questions regarding the ways in which written records of spatial practices do not necessarily replicate the linearity of the experiences themselves. The second problem, however, is more significant. While minor errors in identifying the precise locations of most topographical features do not create significant representational inaccuracies, the mis-mapping of linear features – such as rivers – is more problematic. [22] As mentioned above, the gazetteer gives the location of the named place on OS 1:50,000 maps rather than the physical site itself. Fortunately, the Eamont is named in close proximity to the correct location: the point location, given by the gazetteer, lies approximately two kilometres east of where Gray is likely to have crossed the river. The Eden, however, is considerably misplaced in that it is at least ten kilometres south-east of where the writer must have crossed.

The combined impact of these two problems means that if we join the places in sequence it appears as if Gray has back-tracked on a detour of around twelve kilometres between Brougham Castle and the Eamont, before progressing on to Penrith. A solution to this would clearly be to move the points to a more probable location for Gray's movement through space. This process would raise further issues, however, and would heighten the speculative nature of the literary map. There is an argument, therefore, that the locations taken from the gazetteer provide an objective representation of the sequentiality of Gray's naming of places. As a result, we have usually left the points provided by the gazetteer in place; they may be inaccurate but their existence raises important methodological and conceptual issues for the literary GIS. In cases of obvious mis-mappings, though, these cartographical anomalies have been omitted from subsequent analyses.

Exploring the texts using the literary GIS

[Figure 3]

Figure 3 shows a point GIS of the places referred to by Gray: distinctions are made between places that he claims to have physically visited and places he viewed from a distance or experienced only imaginatively; we have also acknowledged that it remains intrinsically unclear as to whether Gray visited a number of places to which he refers. The places that he visited have been joined by a straight line in order of the visits; but major errors, such as the River Eden, described above, have been omitted. The route of the tour is clear from the maps. On day 1, Saturday, 30 September 1769, Gray leaves Brough and follows the Eden Valley north-west to Penrith. He spends two nights in Penrith, travelling down to Ullswater on the second day, before moving on to Keswick. He spends a number of days here travelling out to places such as Bassenthwaite Lake and Derwent Water. Then he heads south in a long day's travel, passing though Ambleside and on to Kendal, where he spends two nights before

travelling further south to Lancaster. Two things become clear from this mapping of the text: first of all, Gray devotes much of his writing to documenting the time he spends in and around the town of Keswick; secondly, Gray's spatial practice is frequently carried out (perhaps unsurprisingly given his touristic status) alongside the region's lakes and is shaped by the area's major routes of travel.

[Figure 4]

Figure 4 shows the corresponding map for Coleridge's tour. He departs from his home in Keswick and heads south-west, spending three nights in the area near the coastal village of St Bees. It is clear from the maps that his practice of place naming is not at its most intense in this particular area. On the fourth day of his tour he turns south-east, back into the fells and the current National Park area, travelling into the valley of Wasdale. He spends the next two days in the area around Scafell, including ascending the mountain itself, and during this time writes extensively about this particularly precipitous landscape. He then travels east to Coniston, the northern end of Windermere, before heading north back to Greta Hall, travelling over Dunmail Raise: the pass which lies between Keswick and Grasmere and which separated the old county boundaries of Cumberland and Westmorland. On these later days he again writes far less about his environmental experiences; indeed, on the final day of his tour, the only place that he explicitly mentions visiting is Dove Cottage, the Wordsworths' home at Grasmere.

While point maps represent the places mentioned in the text in a simple form, the human eye is not particularly adept at distinguishing point patterns. Additionally, if an author repeatedly writes about the same place, that location will only appear once on the map as the multiple symbols will be superimposed on top of one other. To make point patterns more comprehensible, then, a technique called *density smoothing* can be used. [23] This approach,

pioneered in the spatial analysis field to analyse point patterns such as those found in epidemics or crime maps, aims to measure the density of events around each location. The challenge is to change from precise and discrete points to the intensity of events as it changes continuously over space. In density smoothing a point layer is converted into a raster surface which consists of a fine mesh of square pixels. A simplistic way to do this would be to set an arbitrary distance, such as one kilometre around each pixel, and say that events that occur within this radius are 'near' the point and everything else is not. The density of events around this pixel is then:

$$\lambda_i = \frac{n_{ri}}{2\pi r^2} \quad (1)$$

where λ_i is the density of events at pixel i and n_{ri} is the number of events within radius r of pixel i. The problem with this is that the radius r is highly arbitrary and an event either lies within it or without. Density smoothing techniques avoid this problem by allowing the impact that an event has on the density to decline with distance, so that near events have a larger impact than those further away. How quickly the impact declines is set using a bandwidth h which is also arbitrary but which has less impact than a radius r. A very small bandwidth will simply reflect the actual point patterns; whilst a very large one will smooth out almost all of the variation in density. A good choice, therefore, will show the general patterns but limit the amount of complexity. Equations for density smoothing are complex but a typical one is:

$$\lambda_i = \sum_{dk<=h} \frac{3}{\pi h^2} \left(1 - \frac{d_k^2}{h^2}\right)^2 \quad (2)$$

where d_k is the distance from pixel i to event k.[24]

[Figure 5]

[Figure 6]

Fortunately, density smoothing is incorporated into many GIS software packages such as ArcGIS so implementing it is not difficult. Figures 5 and 6 show the results of applying density smoothing to the two texts. The two maps simplify the patterns and stress that much of the writing focuses on small geographical areas and on relatively short parts of the tour: in Gray's case this means the area around Keswick on days 4 to 8 of his tour; in Coleridge's case it is the area around Scafell on day 5. What is more, it becomes apparent that the two writers concentrated on different parts of the Lake District and, as a result, their tours do not spatially overlap except to a limited extent on the road between Ambleside and Keswick.

[Figure 7]

Figure 7 shows a density smoothed surface of the combined place name mentions from the two texts. This GIS suggests that both writers offer comprehensive accounts of their spatial experiences of the old county of Cumberland, but seem to move through Westmorland and north Lancashire with relative little attention to environmental detail. The layering of the modern boundary of the Lake District National Park also illustrates that the writers spent significant parts of their respective tours outside this privileged space: of the 12 nights of Gray's tour that we have mapped, six were spent beyond the boundaries of what is now the National Park (two in Penrith, two in Keswick, and one each in Brough and Lancaster); whilst Coleridge spent much of his second and third days to the west of the National Park in the western fringes of the region.

Integrating other data

One of the key advantages of GIS is that it allows apparently disparate datasets to be integrated using their spatial data. Data on altitude, for example, is available in the form of

Digital Elevation Models (DEMs) which are raster surfaces in which every pixel, usually with sides of less that 100m, is given a height value. The point layers representing the two tours can be overlaid onto a DEM to give a height for every point. Obviously there are issues of accuracy concerned with doing this given that there will be positional error in the points, as discussed above, and in the DEM. Nevertheless this offers a relatively accurate representation of the heights of the places to which each writer refers. It might be assumed that Gray's proto-Picturesque approach to landscape experience would necessarily dictate that he spent most of his tour moving through the Lake District at low levels and that, as a result, his account would concentrate on the experiences of describing the region's lakes. Coleridge, by contrast, might be expected to move through the topography at higher levels as he endeavoured to experience, and to document, the area's more vertiginous territories.

[Figure 8]

Figure 8 shows frequency distributions for the heights of the places mentioned by both writers. It seems to suggest that there is some validity to the hypothesis advanced above. Gray's tour is strongly based on places below 200m (around 600 feet) with almost 90 per cent of mentions of visits being below this height. The only three places that he mentions visiting that lie above 300m are all rivers and are probably positional errors. This suggests that the highest point on Gray's tour was his ascent of Beacon Hill (285m on the DEM) - near Penrith and well outside the National Park - after his lunch on trout and partridge on the first day of his tour. The next highest points were his crossing of the passes of Penruddock between Penrith and Keswick on day 3 and Dunmail Raise between Keswick and Ambleside on day 9. Both of these would have taken him to around 250m. Places to which he refers, but does not definitely visit, show a more complex distribution. While the majority, 66 per cent, are again below 200m, he also talks about a considerable number of places above 600m (approximately

15

2,000 feet). These constitute 17 per cent of places that he mentions while visiting. Intermediate height places, those between 300m and 600m (around 1,000 to 2,000 feet), only constitute 7 per cent of mentions of this type.

Coleridge's distribution is far less pronounced. As with Gray, the majority of places he mentions visiting are under 200m, but in Coleridge's case these represent two-thirds of the places he visits rather than 90 per cent. 11 per cent of places he visits are over 600m although these are all found on and around Scafell. Apart from this ascent, however, the highest places Coleridge describes visiting are again on passes, although in his case these are slightly higher and more remote than Gray's. These are found between Keswick and Ennerdale on the first day of his tour, and between Eskdale and Ulpha on the sixth day; both are around 300m in height. The distribution of the places that Coleridge mentions without actually being situated at the named location is similar to Gray's but less pronounced. Only 39 per cent of these are less than 200m compared to 66 per cent for Gray. 17 per cent of the places he mentions without being there are over 600m, almost identical to Gray's. The major difference between the two distributions is that 32 per cent of the places Coleridge mentions without being there are between 300 and 600m, an altitude that Gray largely ignores.

[Figure 9]

[Figure 10]

Figure 9 shows the average heights by day of the tour for both authors. The pattern for Gray is consistent with the idea that he spent most of his tour at low levels commenting on the peaks above him. The exceptions to this are day 5, 4 October, in which he gives a relatively short account of an excursion south from Keswick; and the last few days of his tour, once he reaches Kendal on day 10, when he has begun to move away from the craggy landscape of

the central fells. Coleridge's pattern is intriguing in its similarities and differences to Gray's. On day 3, around St Bees and Egremont, he only presents a very short account and the places he mentions are mainly low-lying. On day 5 he climbs Scafell and thus the averages of both his visits and the other places he mentions are very high. On the other days, however, the pattern is not very different to Gray's. He visits places that are on average at between 100m and 200m, and talks about other places that, at 300m to 400m on average, are only slightly higher than Gray's. Figure 10 illustrates this by showing the impact of the Scafell ascent on Coleridge's tour. While the average height of a place that Coleridge says he visited was 234m, nearly twice Gray's 124m, if the day of Coleridge's ascent of Scafell is removed the average drops to only 155m, only a little higher than Gray. For places mentioned without being there, once day 5 is removed Coleridge's average of 184m is actually noticeably lower than Gray's 221m. This is possibly because Gray, the touristic outsider, concentrates on referring to those low-level landscapes through which he moves and to the high fells towards which his gaze is attracted; whilst Coleridge draws upon his familiarity with the regional landscape to referring to sites located at all heights within and across the Cumbrian terrain.

Linking texts and Google Earth

As we have shown, both textual accounts are littered with place names to the extent that even a reader familiar with the region may have difficulty locating a particular geo-specific reference. In an attempt to facilitate further spatial understanding of the primary texts, then, there is a need to create a virtual space in which the user is allowed to read the text and, simultaneously, to follow the place name references on a cartographical representation of regional space. What is required is a tool that allows the reader to read the text while at the same time following the places mentioned on a map. Additionally, the spatial understanding of the texts would be enriched through access to historical maps produced at the time of the

tours; visual representations of the landscape from the eighteenth- and nineteenth-centuries; and even photographs of the landscape today. This can be achieved using an HTML version of the text and maps in Google Earth. [25] Google Earth is effectively a spatial data viewer designed to work over the internet. It provides aerial photographs of the Earth's surface represented in three dimensions to show the topography. Users can add their own data to it using Keyhole Markup Language (KML), an XML implementation. [26] Google Earth provides basic viewing functionality that allows the user to pan and zoom, and to query spatial features. Hyperlinks can be included in the attribute data to provide a link to web pages which can be viewed either within Google Earth or an external browser.

[Figure 11]

To create a Google Earth version of the Lake District GIS the texts were converted into HTML, and the points representing mentions and the lines providing the outline of the tour converted into KML. The data were structured in such a way that the reader could look at all of the tour or explore it day by day. Clicking on a point representing a place mentioned provides basic attribute data on this location such as the standardised version of its spelling and the date the author visited it. It also provides a link to the day of the appropriate tour which, when followed, opens either within Google Earth or an external browser. Figure 11 shows a screenshot of part of the first day of Coleridge's tour as he sets of south-west from Keswick. The attribute data for Kescadale is shown. Clicking on the link to 'Day 1: 1/8/1802 (Sun)' opens the text window at the bottom of the screen. It provides the text for this part of Coleridge's account with the place names highlighted in red. Following the other link to 'Photos (external link)' opens the photographic website Flickr and selects all photographs near this location that are tagged as 'landscape' images. [27] This works because in Google Earth we have a latitude and longitude for the point that in this case represents Kescadale.

Flickr allows users to 'geotag' their photographs with a latitude and longitude so that they also have a location. This allows a URL to pass the coordinate from Google Earth to Flickr along with a zoom level that determines how near to the point the photograph has to be to be included and the tag 'landscape.' Flickr then selects all of the images that match this criteria. For Kescadale this returns 480 images showing the richness of content in the area. [28]

In this way the reader can see contemporary views of the Lake District. There is also a contemporary map of the Lake District dating from 1815 that can be superimosed over Google Earth's landscape. While we have not included historical images of the Lakes, as yet, there is no technical reason why this can not be done. It would also be relatively simple to link from the spatial data directly to the place name in the text rather than the start of the day in which the geo-specific reference is located. Through the use of GIS, therefore, the literary critic is able to produce fluid maps which allow the user to move freely between textual, cartographical and visual representations of place and space.

Conclusions

This article has demonstrated that a textual GIS can be a useful tool for research within literary studies. The creation of a literary GIS can be achieved either through the use of basic mapping techniques, such as dot maps, or through more complex approaches, adapted from other disciplines, such as density smoothing. Having the data in GIS form also allows us to integrate different texts together to facilitate comparative analysis and also to integrate other sources of data such as the border of the National Park or height data from the DEM. As a result, the paper has been able to show that while Gray and Coleridge both toured the Lake District, there was very little overlap in the routes that they took or the places that they documented in their respective texts. The paper has also illustrated that both writers focused on the old county of Cumberland, as well as places and spaces which lie beyond the

boundaries of what is now the National Park; and it has indicated that, apart from Coleridge's ascent of Scafell, the heights at which the two writers spent much of their tours were broadly similar. A major distinction, however, is that when Gray describes places at which he is not currently located, he tends to focus either on low landscapes, below 200m, or high terrains, over 600m; Coleridge, by contrast is far less discriminatory in his place naming. Finally, the paper has been able to show how GIS, through Google Earth, can be used to explore the accounts in detail by providing an interactive map which links directly to the texts, as well as other forms of data such as historical maps and contemporary photographs. The GIS has also provided much of the material for a website devoted to the two texts[29] which has received attention in the local media including the *Westmorland Gazette* and BBC Radio Cumbria. [30]

The article, therefore, has illustrated that the abilities of GIS to summarise spatial patterns, to integrate data from different sources, and to disseminate information to a broad audience can all be applied to textual material as well as the more traditional quantitative sources. GIS has one other major ability, namely the capacity to handle large volumes of material effectively. This paper has not examined this potentiality as its principal aim has been to describe a pilot project which has mapped just two short texts. Clearly, however, the techniques described here could be used on much larger datasets: for example, all writing on the Lake District could be incorporated into this framework, or the techniques used could be applied to a much broader geographical area, potentially all of Britain, Europe, or even the whole world. Such mapping would clearly facilitate Moretti's proposed critical mode of 'distant reading'.

Assuming that the texts are in digital form, the major obstacle to the production of such literary GIS is the problematic process of allocating grid references to place names. This involves first extracting the names and then joining them to a gazetteer. In this paper the first part of this was performed manually and the Ordnance Survey gazetteer was used for the

second. To scale this up to cope with far larger datasets, potentially corpora consisting of millions of words, requires further work but is far from impossible. Corpus linguistics techniques can be used to automatically extract proper nouns from texts and this would provide a reasonable first pass on possible place names. Linking to a gazetteer requires improved gazetteers to be built, particularly historical gazetteers capable of handling variant spellings as they changed over time. It may also require improved methodologies to handle ambiguous place names such as Newcastle. In these cases a certain degree of user interaction is likely to remain inevitable but accuracy could be improved by better techniques. It is hoped that developments will be made in this area in the near future as this paper has demonstrated that GIS the potential to be an important tool in literary studies and, more particularly, the interdisciplinary field of spatial criticism.

Acknowledgements:

Figure 1: An outline of the Lake District.
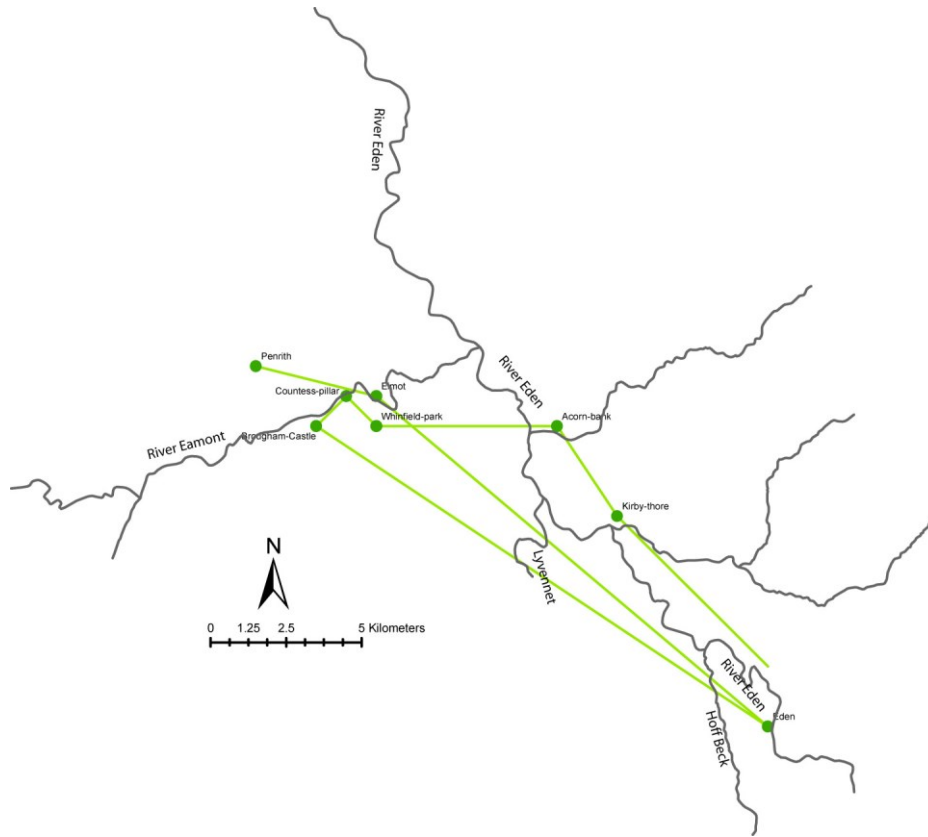
Figure 2: Places mentioned by Gray near the River Eden. The line shows the order in which the places are mentioned starting at Kirby-thore.
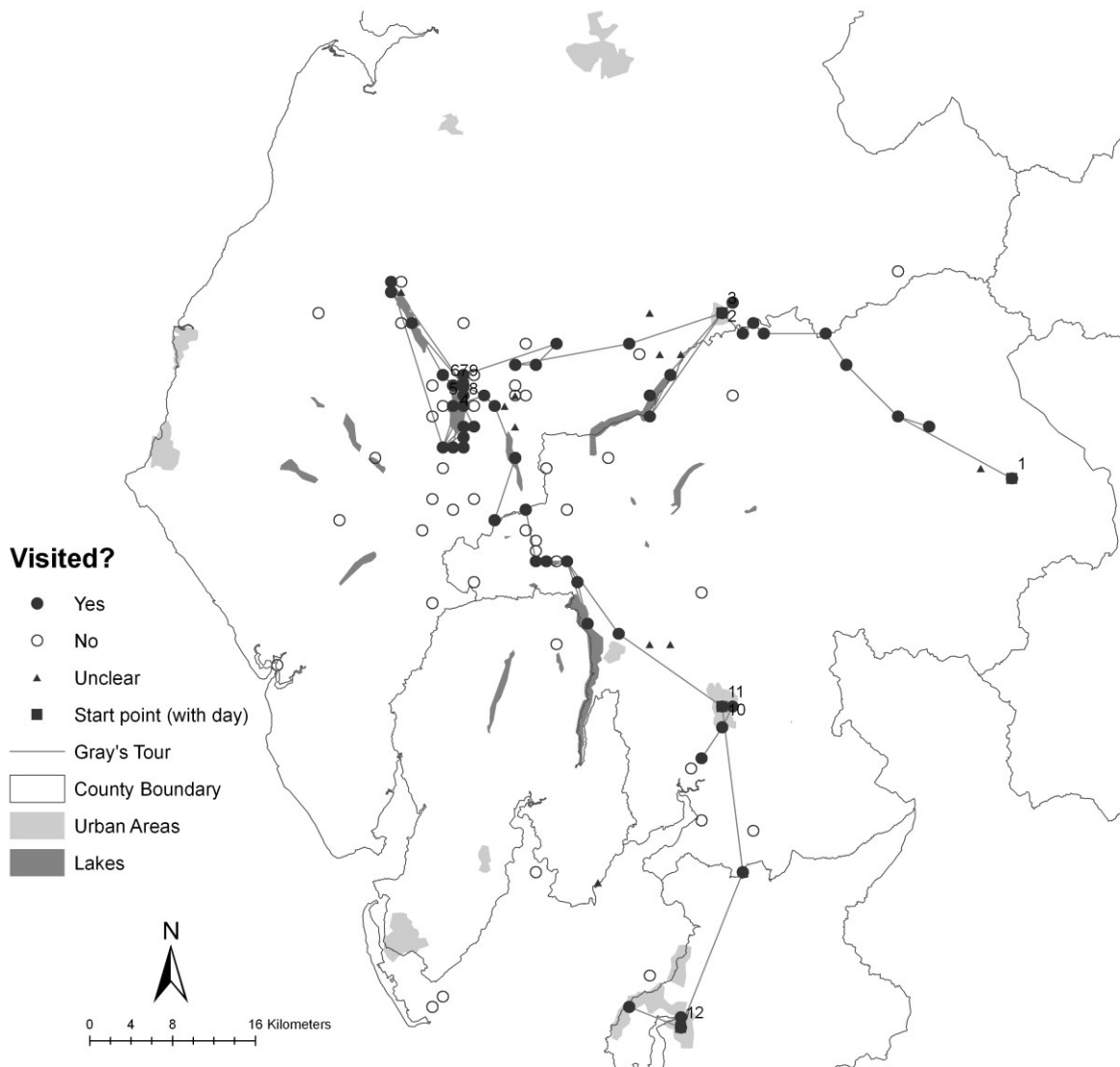
Figure 3: A point GIS of Gray's tour. The line joins places that he visited in the order he mentions them. Obvious errors such as the River Eden discussed in Figure 2 have been omitted. His itinerary was: 1. Brough to Penrith; 2. Penrith; 3. Penrith to Keswick; 4-8. Keswick; 9. Keswick to Kendal; 10. Kendal; 11. Kendal to Lancaster; 12. Lancaster.
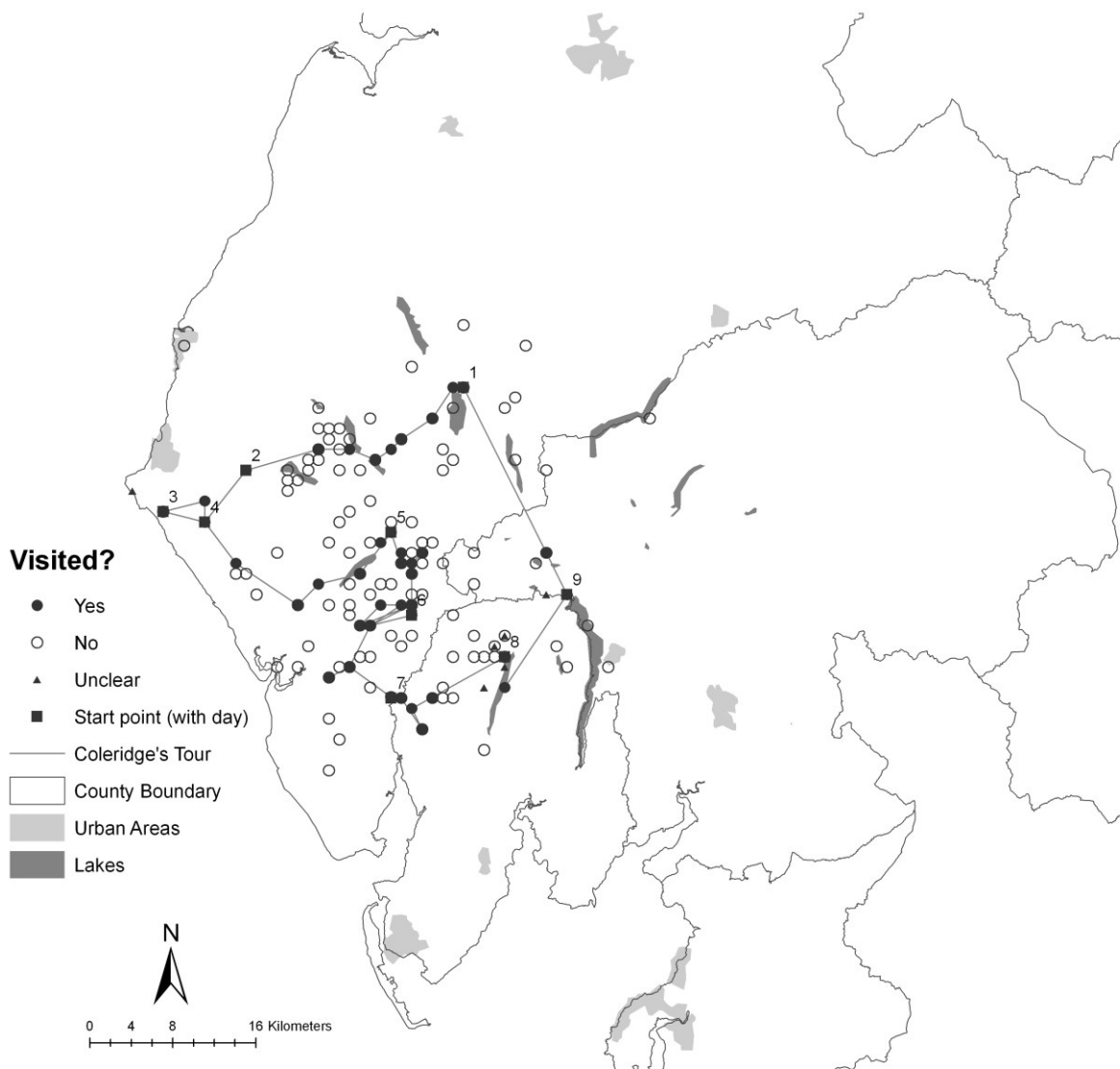
Figure 4: A point GIS of Coleridge's tour. His itinerary was: 1. Keswick to Ennerdale; 2. Ennerdale to St. Bees; 3. St. Bees to Egremont; 4. Egremont to Wasdale; 5. Wasdale to Eskdale; 6. Eskdale to Ulpha; 7. Ulpha to Coniston; 8. Coniston to Clappersgate; 9. Clappersgate to Keswick
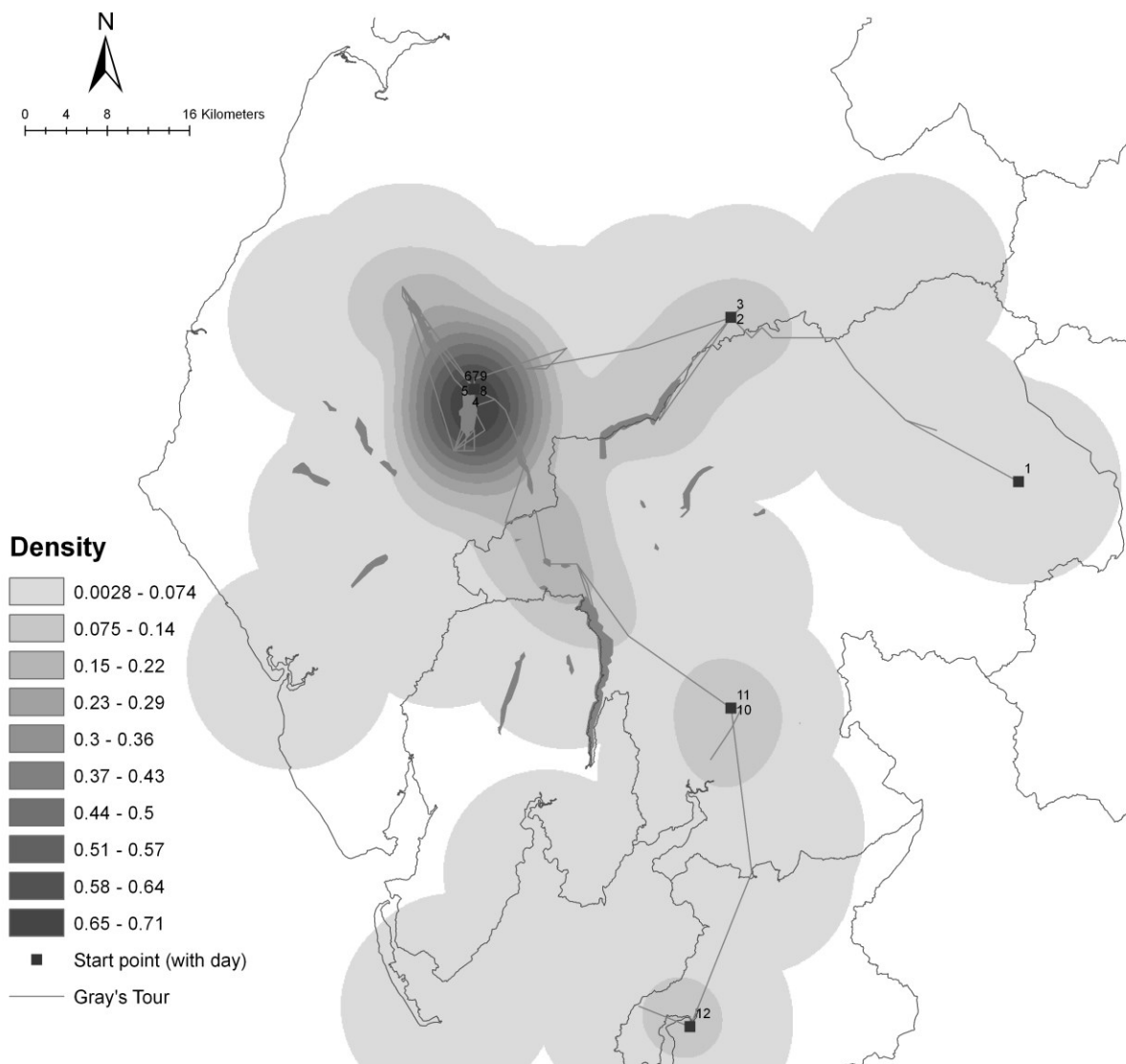
Figure 5: A density smoothed map of Gray's tour. All places mentioned are included whether he visits them or not. Legend uses equal intervals.
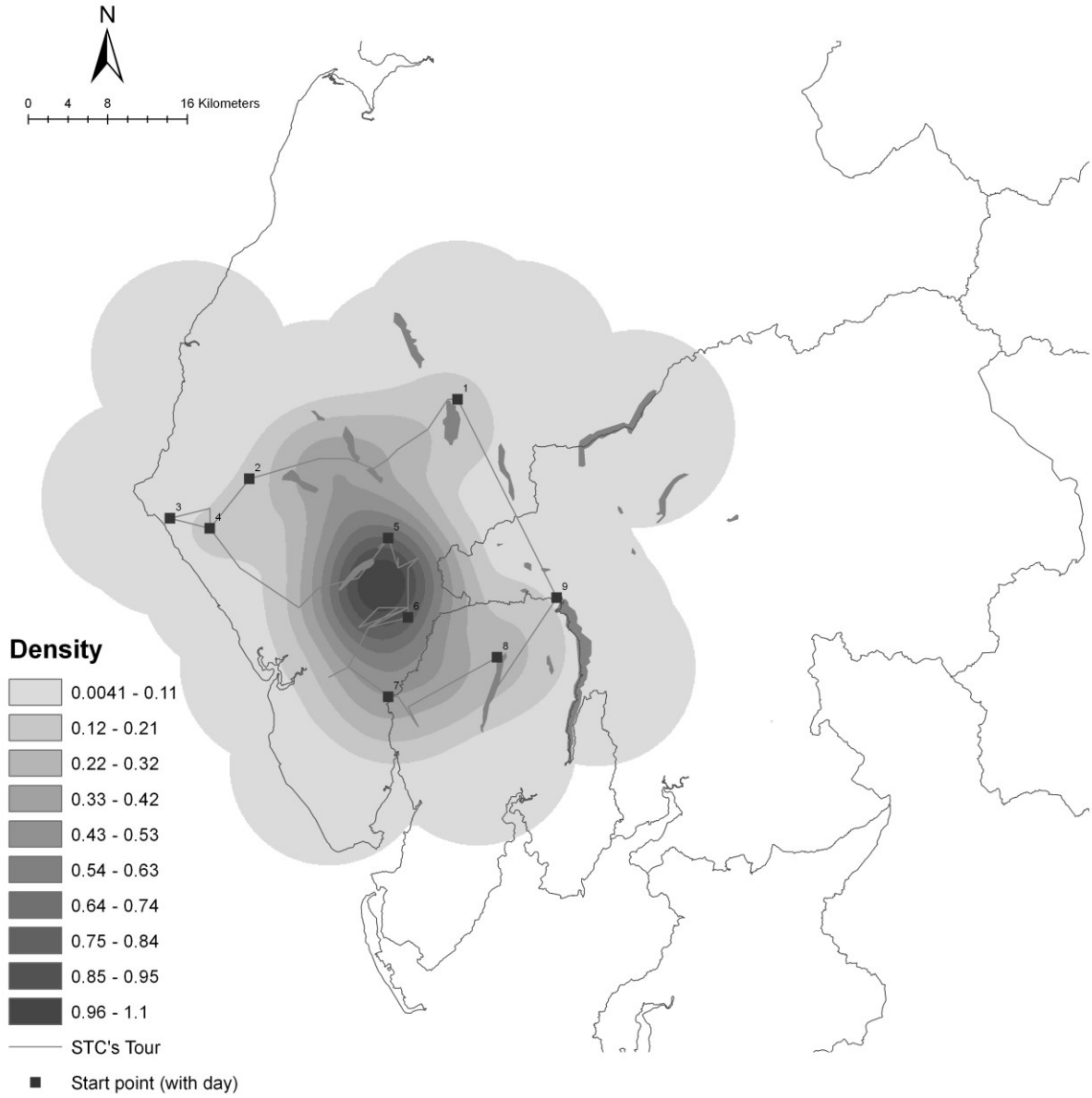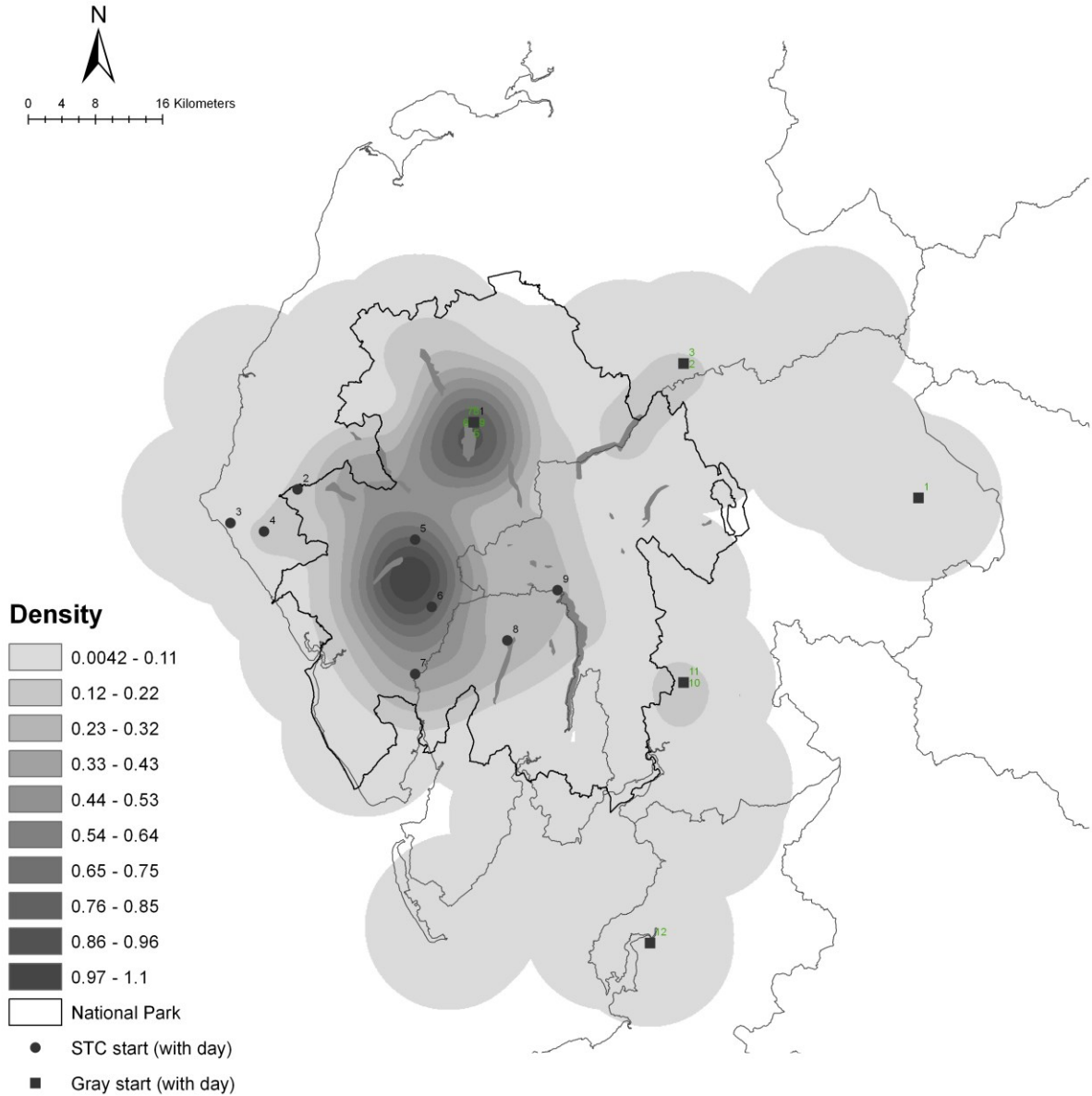
Figure 6: A density smoothed map of Coleridge's tour. All places mentioned are included whether he visits them or not. Legend uses equal intervals.
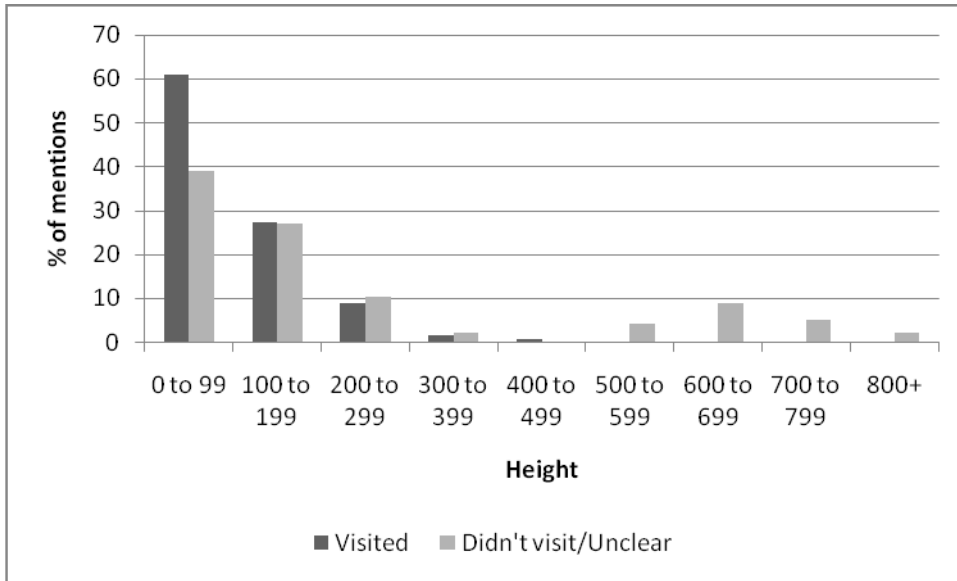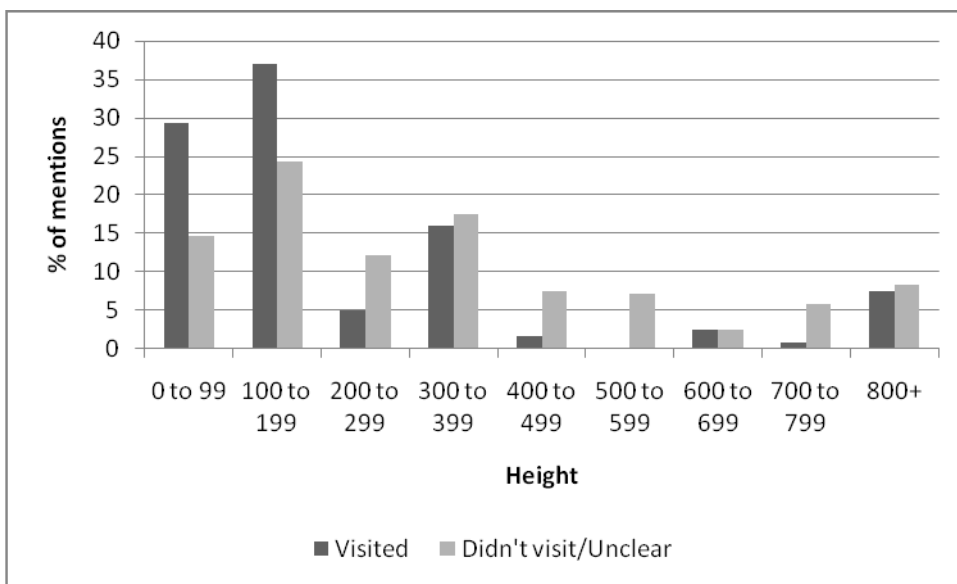
Figure 7: A density smoothed map of both tours. All places mentioned are included whether they were visited or not. Legend uses equal intervals. The modern National Park boundary is also shown.
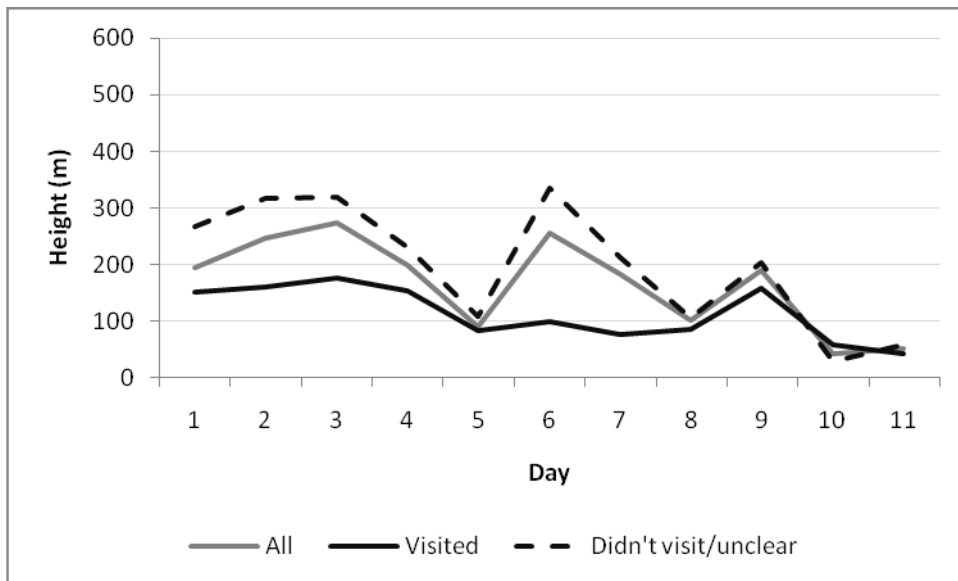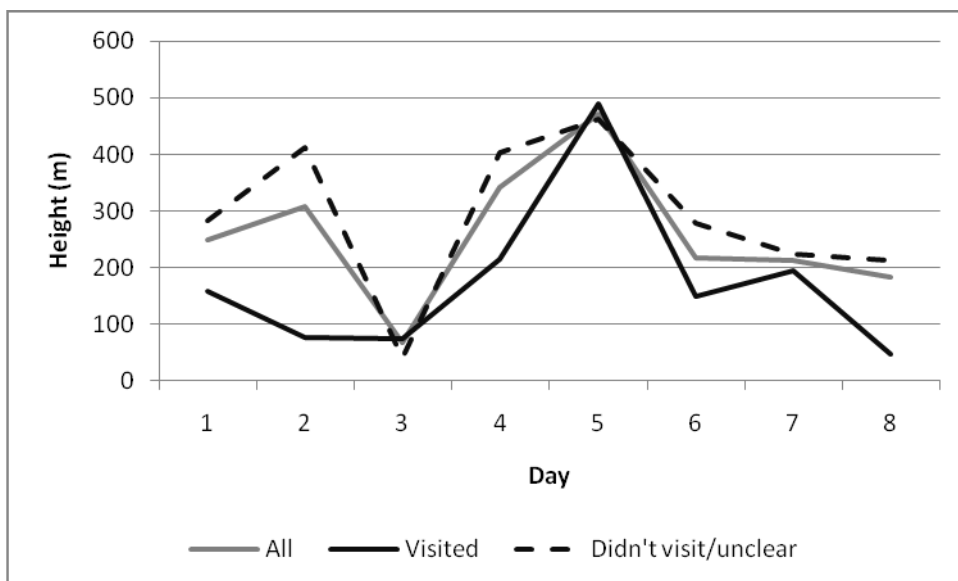
a. Gray

b. Coleridge

Figure 8: Frequency distribution of heights mentioned by the two authors. Heights are in meters. The percentage of mentions is by type for each author.

a. Gray

b. Coleridge

Figure 9: The average hieght of the places mentioned by the writers by day. The three lines distinguish between all mentions for the day: places that the writer unambiguously visited; places that they did not visit; or places where it is fundamentally unclear. In both cases the final day of the tour has been missed off because of a lack of environmental observations within the respective texts.
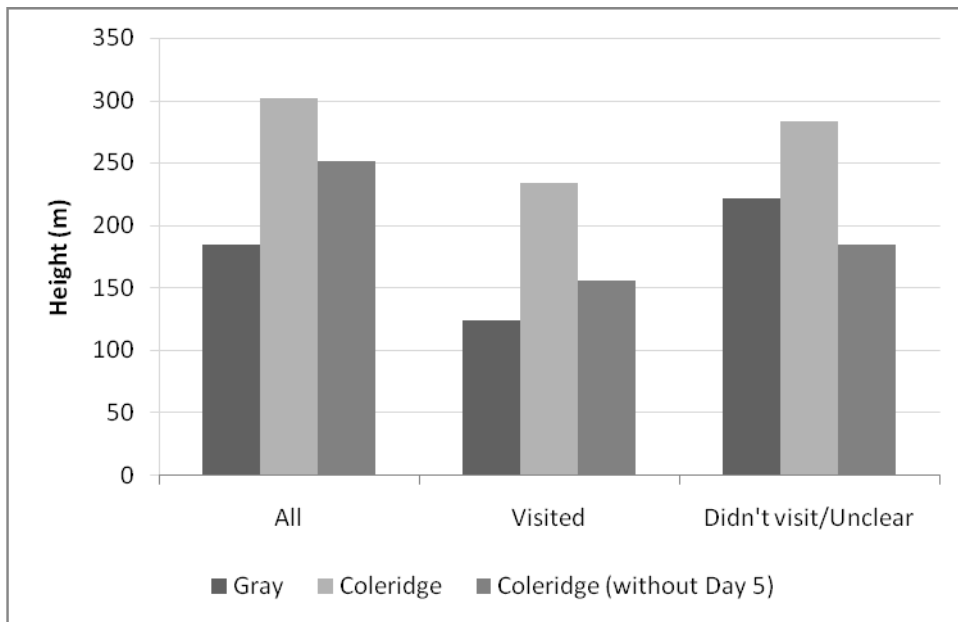
Figure 10: A comparison of the average heights of places mentioned by Gray and Coleridge. The graphs show the average heights of all places mentioned by both writers. In addition, the graph shows the average height of Coleridge's tour, excluding the day on which he climed Scafell (Day 5).
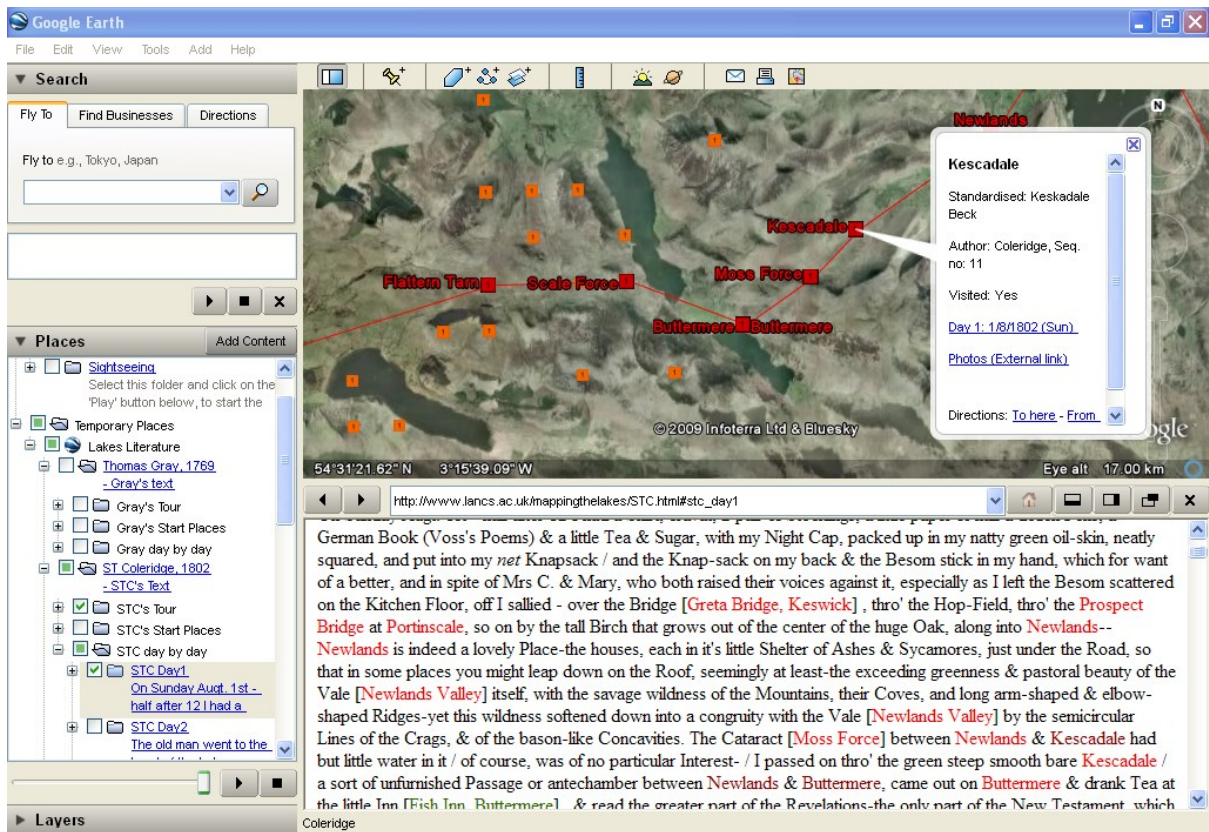
Figure 11: Screenshot of the Lake District GIS in Google Earth. Part of day 1 of Coleridge's tour is shown. Kescadale is being queried.

[1] See: A.R.H Baker, *Geography and History: bridging the divide* (Cambridge, 2003); D.W. Holdsworth, 'Historical geography: New ways of imaging and seeing the past', *Progress in Human Geography*, 27 (2003), 486-493; I.N. Gregory and P.S. Ell, *Historical GIS: technologies, methodologies and scholarship* (Cambridge, 2007); I.N. Gregory and R.G. Healey, 'Historical GIS: Structuring, mapping and analysing geographies of the past', *Progress in Human Geography*, 31 (2007), 638-653.

[2] G. Cunfer, *On the Great Plains: agriculture and environment*. (College Station, TX, 2005).

[3] A.K. Knowles and R.G. Healey, 'Geography, timing, and technology: a GIS-based analysis of Pennsylvania's iron industry, 1825-1875', *Journal of Economic History*, 66 (2006), 608-634.

[4] I.N. Gregory, 'Different places, different stories: Infant mortality decline in England & Wales, 1851-1911', *Annals of the Association of American Geographers*, 98 (2008), 773-794.

[5] B.M.S. Campbell and K. Bartley, *England on the eve of the Black Death: an atlas of lay lordship, land and wealth, 1300-49* (Manchester, 2006).

[6] C. Gordon, *Mapping decline: St Louis and the fate of the American city* (Philadelphia, 2008).

[7] P. E. Wegner, 'Spatial criticism: critical geography, space, place and textuality', in J. Wolfreys, ed., *Introducing criticism at the twenty-first century* (Edinburgh, 2002), 179-201. Cited here at 180.

[8] 'New Historicism' is a branch of literary theory which was developed during the 1980s and which continues to interrogate the relationship between history and literature.

[9] Wegner, 'Spatial criticism', 180.

[10] M. Bradbury, ed., *The atlas of literature* (London, 1996); F. Moretti, *Atlas of the European novel 1800-1900* (London, 1998).

[11] F. Moretti, *Graphs, maps, trees* (London, 2005), 1.

[12] Moretti, *Graphs, maps, trees*, 1-2.

[13] A polygon is the structure that a GIS uses to represent an area or zone.

[14] G. Lindop, *A literary guide to the Lake District* (London, 1993).

[15] Both accounts have complex textual histories which need to be explored in detail and beyond the parameters of the present article.

[16] Gray's tour of the North of England continued into Yorkshire; but, to facilitate the process comparative mapping, our digitisation of Gray's text ended up with his stay in Lancaster.

[17] P. Toynbee and L. Whibley, eds, revised by H.W. Starr, *Correspondence of Thomas Gray. Volume III, 1766-1771* (Oxford, 1971).

[18] E.L. Briggs, ed, *Collected letters of Samuel Taylor Coleridge. Volume II, 1801-1806.* (Oxford, 1956).

[19] K. Coburn, ed., *The Notebooks of Samuel Taylor Coleridge. Volume I, 1794-1804.* (London, 1957).

[20] This was downloaded from the Digimap service at EDINA, see <http://digimap.edina.ac.uk/gaz-download/GazetteerFullDownload.jsp > [15 January 2009].

[21] Toynbee and Whibley, 1076.

[22] In GIS terminology an *error* is defined as the difference between reality and the feature as represented in the database. It does not imply a mistake. An error is usually the result of minor inaccuracies in locational information. See D. Unwin, 'Geographic Information Systems and the problem of "error and uncertainty"' *Progress in Human Geography*, 19 (1995), 549-558.

[23] See T.C. Bailey and A.C. Gatrell, *Interactive spatial data analysis* (Harlow, Essex, 1995); or A.S. Fotheringham, C. Brunsdon and M. Charlton, *Quantitative geography: perspectives on spatial data analysis* (London, 2000).

[24] Modified from C. Lloyd, *Local models for spatial analysis* (Boca Raton, FL, 2007).

[25] Google Earth can be freely downloaded from <http://earth.google.com> [15 January 2009].

[26] For a description of KML see: <http://code.google.com/apis/kml/documentation> [15 January 2009].

[27]< http://www.flickr.com> [15 January 2009].

[28] This was returned on the 6th Feb. 09. It will change over time.

[29]< http://www.lancs.ac.uk/mappingthelakes> [15 January 2009].

[30] 'Virtual tour of literary Lakes.' *The Westmorland Gazette*, 6 August 2008. See: <http://www.thewestmorlandgazette.co.uk/news/3573999.Virtual_tour_of_literary_Lakes> [15 January 2009].