Studies in Scottish Literature

Volume 42 | Issue 2 Article 6

11-30-2016

Topic Modeling and the Historical Geography of Scotland

Michael Gavin
University of South Carolina

Eric Gidal
University of Iowa

TOPIC MODELING AND THE HISTORICAL GEOGRAPHY OF SCOTLAND

Michael Gavin and Eric Gidal

With the recent public releases of large-scale digital archives, the history of language and culture can now be studied quantitatively in ways that were impossible just a few years ago.[1] To explore these new avenues of research is, broadly speaking, one of the foremost challenges facing the humanities today. As part of this larger research agenda, the "spatial humanities" adapt geographic information science (GIS) to historical and cultural scholarship.[2] Geospatial text analysis combines the techniques of GIS with those of corpus linguistics, identifying places mentioned in document collections and analyzing concepts associated with those places.[3] In this paper, we present some initial findings from a research

[1] This argument has been advanced most publicly by the designers of the Google N-gram Viewer, Erez Aidan and Jean-Baptiste Michel, in *Uncharted: Big Data as a Lens on Culture* (Riverhead Books, 2013) and within academia by Matthew Jockers in *Macroanalysis: Digital Methods and Literary History* (Southern Illinois University Press, 2013).

[2] The spatial humanities are covered in a series of recent scholarly collections. Among these are David Bodenhamer, John Corrigan, and Trevor Harris, eds., *Deep Maps and Spatial Narratives* (Indiana University Press, 2014), as well as their earlier collection, *The Spatial Humanities: GIS and the Future of Humanities Scholarship* (Indiana University Press, 2010). These collections were preceded by the somewhat more technical collection edited by Ian N. Gregory and Paul S. Ell, *Historical GIS: Technologies, Methodologies, and Scholarship* (Cambridge: Cambridge University Press, 2007). Most recently, GIS uses for literary history have been collected by David Cooper, Christopher Donaldson, and Patricia Murrieta-Flores in *Literary Mapping in the Digital Age* (Routledge, 2016).

[3] The methods presented here share much in common with those outlined by Ian N. Gregory and Andrew Hardie in their essay, "Visual GISting: Bringing

project we have begun using geospatial text analysis to study industrial and environmental history in Scotland. Over the coming years our goal will be to expand the archive, refine our analytical methods, and develop case studies that contribute to historical geography, literary and cultural studies. What we present here is simply a sampling of recent work.

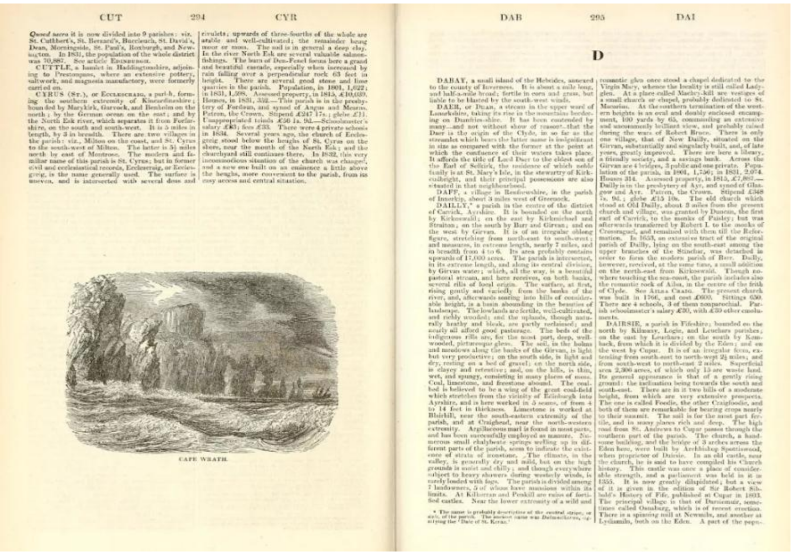
This project began as an engagement with spatial approaches to the *Poems of Ossian*. As Eric Gidal has detailed in his recent study *Ossianic Unconformities: Bardic Poetry in the Industrial Age*, for a select group of Scottish enthusiasts over the eighteenth and nineteenth centuries, the problem of the poems' authenticity became a problem of historical geography.[4] Seeking to connect James Macpherson's controversial "translations" of the 1760s with indigenous oral traditions of the Scottish Highlands, Hugh Blair, Sir John Sinclair, Henry Mackenzie, and an evolving network of correspondents combined statistical geography with antiquarian philology to map the poems onto Gaelic lands on either side of the North Channel. The "deep maps" of mythological, social, agricultural, and increasingly industrial change that they produced strike us as early and imaginative methodological forays into geocriticism and spatial analytics. Part of our early work, therefore, has been to update information arranged in publications such as the *Report of the Committee of the Highland Society of Scotland...into the Nature and Authenticity of the Poems of Ossian* (Edinburgh, 1805) into the protocols of modern GIS. This has enabled us to visualize a complex semantic footprint of Macpherson's epic composition that includes multiple generations of oral transmission, manuscript collection, and antiquarian inquiry, and their administrative integration into the pages of the Highland Society's *Report*. Our approach continues a sequence of translations and remediations initiated by Macpherson's original publications in new

Together Corpus Linguistics and Geographical Information Systems," *Literary & Linguistic Computing,* 26:3 (2011). Within the field of geography, this set of techniques shares much in common with "geospatial semantics," which is devoted to study the concepts used to describe and model spaces. For an overview of that field, see Werner Kuhn, "Geospatial Semantics," *Journal on Data Semantics III* (2005): 1-24. For a review of similar techniques used by geographers for this purpose, see Angela Schwering, "Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey," *Transactions in GIS,* 12:1 (2008): 5-29.

[4] Eric Gidal, *Ossianic Nonconformities: Bardic Poetry in the Industrial Age* (Charlottesville: University of Virginia Press, 2015).

technologies of data collection and display and speaks to analogous though greatly accelerated conditions of cultural and environmental change.

Our next task has been to move beyond the Ossianic literature to a wider archive of Scottish writings. Our current model builds on a corpus of nineteenth-century geographical publications, matching the places described in that corpus with their geographical locations, as recorded in modern GIS datasets published by the British Ordnance Survey.[5] Geographical writing played an important role in the Scotland's modernization, which included the rapid expansion of commerce, manufacturing, and agriculture, alongside an infrastructure of new roads, canals, ports, and planned villages. The creation of geographic knowledge went hand-in-hand with the invention of Scotland as a modern geographical entity. Both depended on particular uses of print: compiling descriptions of named locales, gathering testimonies of local cultural



CUT 294 CYR

Quoad sacra it is now divided into 9 parishes: viz. St. Cuthbert's, St. Bernard's, Buccleuch, St. David's, Dean, Morningside, St. Paul's, Roxburgh, and Newington. In 1831, the population of the whole district was 70,887. See article Edinburgh.

CUTTLE, a hamlet in Haddingtonshire, adjoining to Prestonpans, where an extensive pottery, saltwork, and magnesia manufactory, were formerly carried on.

CYRUS (St.), or Ecclesgraig, a parish, forming the southern extremity of Kincardineshire; bounded by Marykirk, Garvock, and Benholm on the north; by the German ocean on the east; and by the North Esk river, which separates it from Forfarshire, on the south and south-west. It is 5 miles in length, by 3 in breadth. There are two villages in the parish: viz., Milton on the coast, and St. Cyrus to the south-west of Milton. The latter is 5½ miles north by east of Montrose. The modern and familiar name of this parish is St. Cyrus; but in former civil and ecclesiastical records, Ecclesgraig, or Ecclesgreig, is the name generally used. The surface is uneven, and is intersected with several dens and rivulets; upwards of three-fourths of the whole are arable and well-cultivated; the remainder being moor or moss. The soil is in general a deep clay. In the river North Esk are several valuable salmon-fishings. The burn of Den-Fenel forms here a grand and beautiful cascade, especially when increased by rain falling over a perpendicular rock 63 feet in height. There are several good stone and lime quarries in the parish. Population, in 1801, 1,622; in 1831, 1,598. Assessed property, in 1815, £10,039. Houses, in 1831, 332.—This parish is in the presbytery of Fordoun, and synod of Angus and Mearns. Patron, the Crown. Stipend £247 17s.; glebe £11. Unappropriated teinds £56 1s. 9d.—Schoolmaster's salary £33; fees £33. There were 4 private schools in 1834. Several years ago, the church of Ecclesgreig stood below the heughs of St. Cyrus on the shore, near the mouth of the North Esk; and the churchyard still continues there. In 1832, this very incommodious situation of the church was changed, and a new one built on an eminence a little above the heughs, more convenient to the parish, from its easy access and central situation.

CAPE WRATH.

DAB 295 DAI

D

DABAY, a small island of the Hebrides, annexed to the county of Inverness. It is about a mile long, and half-a-mile broad; fertile in corn and grass, but liable to be blasted by the south-west winds.

DAER, or Dear, a stream in the upper ward of Lanarkshire, taking its rise in the mountains bordering on Dumfries-shire. It has been contended by many—and not without show of reason—that the Daer is the origin of the Clyde, in so far as the streamlet which bears the latter name is insignificant in size as compared with the former at the point at which the confluence of their waters takes place. It affords the title of Lord Daer to the eldest son of the Earl of Selkirk, the residence of which noble family is at St. Mary's Isle, in the stewartry of Kirkcudbright, and their principal possessions are also situated in that neighbourhood.

DAFF, a village in Renfrewshire, in the parish of Innerkip, about 3 miles west of Greenock.

DAILLY,* a parish in the centre of the district of Carrick, Ayrshire. It is bounded on the north by Kirkoswald; on the east by Kirkmichael and Straiton; on the south by Barr and Girvan; and on the west by Girvan. It is of an irregular oblong figure, stretching from north-east to south-west; and measures, in extreme length, nearly 7 miles, and in breadth from 4 to 6. Its area probably contains upwards of 17,000 acres. The parish is intersected, in its extreme length, and along its central division, by Girvan water; which, all the way, is a beautiful pastoral stream, and here receives, on both banks, several rills of local origin. The surface, at first, rising gently and variedly from the banks of the river, and, afterwards soaring into hills of considerable height, is a basin abounding in the beauties of landscape. The lowlands are fertile, well-cultivated, and richly wooded; and the uplands, though naturally heathy and bleak, are partly reclaimed; and nearly all afford good pasturage. The beds of the indigenous rills are, for the most part, deep, well-wooded, picturesque glens. The soil, in the holms and meadows along the banks of the Girvan, is light but very productive; on the south side, is light and dry, resting on a bed of gravel; on the north side, is clayey and retentive; and, on the hills, is thin, wet, and spongy, consisting in many places of moss. Coal, limestone, and freestone abound. The coal-bed is believed to be a wing of the great coal-field which stretches from the vicinity of Edinburgh into Ayrshire, and is here worked in 5 seams, of from 4 to 14 feet in thickness. Limestone is worked at Blairhill, near the south-eastern extremity of the parish, and at Craighead, near the north-western extremity. Argillaceous marl is found in most parts, and has been successfully employed as manure. Numerous small chalybeate springs welling up in different parts of the parish, seem to indicate the existence of strata of ironstone. The climate, in the valley, is generally dry and mild, but on the high grounds is moist and chilly; and though everywhere subject to heavy showers during westerly winds, is rarely loaded with fogs. The parish is divided among 7 landowners, 5 of whom have mansions within its limits. At Kilkerran and Penkill are ruins of fortified castles. Near the lower extremity of a wild and romantic glen once stood a chapel dedicated to the Virgin Mary, whence the locality is still called Ladyglen. At a place called Machry-kill are vestiges of a small church or chapel, probably dedicated to St. Macarius. At the southern termination of the western heights is an oval and doubly enclosed encampment, 100 yards by 65, commanding an extensive and uncommonly brilliant view, and probably raised during the wars of Robert Bruce. There is only one village, that of New Dailly, situated on the Girvan, substantially and singularly built, and, of late years, greatly improved. There are here a library, a friendly society, and a savings bank. Across the Girvan are 4 bridges, 3 public and one private. Population of the parish, in 1801, 1,756; in 1831, 2,074. Houses 314. Assessed property, in 1815, £7,887.—Dailly is in the presbytery of Ayr, and synod of Glasgow and Ayr. Patron, the Crown. Stipend £348 7s. 9d.; glebe £15 10s. The old church which stood at Old Dailly, about 3 miles from the present church and village, was granted by Duncan, the first earl of Carrick, to the monks of Paisley; but was afterwards transferred by Robert I. to the monks of Crossraguel, and remained with them till the Reformation. In 1653, an extensive tract of the original parish of Dailly, lying on the south-east among the upper branches of the Stinchar, was detached in order to form the modern parish of Barr. Dailly, however, received, at the same time, a small addition on the north-east from Kirkoswald. Though nowhere touching the sea-coast, the parish includes also the romantic rock of Ailsa, in the centre of the frith of Clyde. See Ailsa Craig. The present church was built in 1766, and cost £600. Sittings 650. There are 4 schools, 3 of them nonparochial. Parish schoolmaster's salary £30, with £30 other emoluments.

DAIRSIE, a parish in Fifeshire; bounded on the north by Kilmany, Logie, and Leuchars parishes; on the east by Leuchars; on the south by Kemback, from which it is divided by the Eden; and on the west by Cupar. It is of an irregular form, extending from south-east to north-west 2½ miles; and from south-west to north-east 2 miles. Superficial area 2,306 acres, of which only 15 are waste land. Its general appearance is that of a gently rising ground: the inclination being towards the south and south-east. There are in it two hills of a moderate height, from which are very extensive prospects. The one is called Foodie, the other Craigfoodie, and both of them are remarkable for bearing crops nearly to their summit. The soil is for the most part fertile, and in many places rich and deep. The high road from St. Andrews to Cupar passes through the southern part of the parish. The church, a handsome building, and the bridge of 3 arches across the Eden here, were built by Archbishop Spottiswood, when proprietor of Dairsie. In an old castle, near the church, he is said to have compiled his Church history. This castle was once a place of considerable strength, and a parliament was held in it in 1355. It is now greatly dilapidated; but a view of it is given in the edition of Sir Robert Sibbald's History of Fife, published at Cupar in 1803. The principal village is that of Dairsiemuir, sometimes called Osnaburg, which is of recent erection. There is a spinning mill at Newmills, and another at Lydiamills, both on the Eden. A part of the popu-

* The name is probably descriptive of the central stripe, or *dale*, of the parish. The ancient name was *Dalmakkeran*, signifying the 'Dale of St. Keran.'

Fig. 1: ***A Topographical Dictionary of Scotland*** **(1846).**

[5] Datasets from the Ordnance Survey can be downloaded from the web at https://www.ordnancesurvey.co.uk/business-and-government/products/opendata-products.html.

and environmental conditions, describing histories specific to those places, reconciling competing definitions of places, and articulating their sometimes shifting physical and linguistic boundaries. These activities left a textual record that is especially amenable to geospatial analysis. In their efforts to modernize their nation, nineteenth-century writers built an elaborate geospatial model of Scotland as such, one rich with historical detail, but still largely untapped by scholars in its digital form.[6]

The textual data used in this study is drawn from a medium-sized corpus (approximately seven million words) of nineteenth-century geographical writing about Scotland—the Old and New Statistical Accounts of Scotland, primarily, but also topographical dictionaries and gazetteers.[7] All the files used in our collection were taken from *The Internet Archive*, a public resource that stores page facsimiles and transcriptions of scanned books. (Each of the books we used were transcribed automatically using optical character recognition.) The corpus includes:

The Statistical Account of Scotland, 21 vols. (1791-99)
A Topographical Dictionary of Scotland, 2 vols. (1813)
Gazetteer of Scotland (1825)
A Topographical Dictionary of the United Kingdom (1826)
Descriptive Account of the Principal Towns in Scotland (1828)
The Gazetteer of Scotland, 2 vols. (1838)
The Topographical, Statistical, and Historical Gazetteer of Scotland, 2 vols. (1842)
The New Statistical Account of Scotland, 15 vols. (1834-45)
A Topographical Dictionary of Scotland, 2 vols. (1846)
The Imperial Gazetteer of Scotland, 2 vols. (1854)
Ordnance Gazetteer of Scotland, 6 vols. (1882-85)

As this list shows, more than half our corpus is drawn from the *Statistical Account of Scotland*, including both the “Old” (1791-99) and the “New”

[6] In their critique of GIS, Pater J. Taylor and Ronald J. Johnston call attention to its origins in the “quantitative revolution” that they pinpoint in the 1950s and 1960s, and they associate the compilation of statistics with the exercise of the power of the state. “Geographic Information Systems and Geography,” in *Ground Truth: The Social Implications of Geographic Information Systems*, ed. John Pickles (New York: The Guilford Press, 1995), 51-67. Our research underscores that this power dynamic reaches back yet further to the earliest attempts at systematically gathering and managing geographic information.

[7] The data and scripts used for this study are available for download at https://github.com/michaelgavin/mapping-ossian.

(1834-45) versions.[8] The bulk of the corpus is not dated beyond the 1850s, but we also included Francis Groome's *Ordnance Gazetteer of Scotland*, which was published in six volumes during the 1880s. One volume, called *A Topographical Dictionary of the United Kingdom* (1826), includes entries from the whole of the British Isles (including Ireland), but otherwise the focus is squarely on Scotland.

Using this corpus, we extracted descriptions of British places and matched them to places listed in the Ordnance Survey's datasets. Geographic dictionaries and encyclopedias are highly structured and so are particularly amenable to parsing and analysis (Fig. 1). The Statistical Accounts are laid out like encyclopedias with long entries describing each parish. Entries in gazetteers tend to be shorter, typically not more than a paragraph or a few sentences. But in all cases, the regular type-face patterns that distinguish entries on the page appear similarly in the transcriptions and so provide easily identifiable labels.

The next step was to match those labels against existing places in the Ordnance Survey's GIS datasets. In most cases, there was a direct match between the names listed in the nineteenth-century dictionaries and the modern GIS systems, but if no perfect match was found the computer would perform a "fuzzy search" to find names with slightly variant spelling. In cases where the Ordnance Survey listed more than one location for a given name, the algorithm searched through the place description, then chose the location nearest whatever places were mentioned therein. We tried to be conservative, and so for about 19% of entries no suitable match was found. Nonetheless, the remaining text was georeferentially copious: 17,047 distinct places were identified, and of those 60,951 geographical descriptions were extracted. Because our goal in this study was to see how geographical texts recorded change over time, we further targeted our analysis to individual sentences that included references to dates (of which there were just over 78,000), and then ran a topic model over those sentences.

Topic modeling should be familiar to most literary historians. It's a form of text analysis designed to identify themes or topics that occur throughout a collection of documents, even if those topics are not listed as "subjects" in a catalogue nor specified in titles. The technique was invented by David Blei, with important contributions from Andrew McCallum, David Mimno, and others; it has been used for literary

[8] Text in some volumes of Sinclair's original *Account* were of insufficient quality to use in the corpus; the dataset includes volumes 1-9, 11-13, and 16.

purposes by scholars like Matthew Jockers, Ted Underwood, and Andrew Goldstone, and in the social sciences by John Mohr.[9] Essentially, it's a mode of summary: researchers begin with a document collection and instruct the system to summarize that collection along a chosen number of axes, or "topics." For each topic, the model finds a cluster of words that tend to appear together and finds the documents in which they are likely to appear. If the researcher chooses a small number of topics—say, five topics over a corpus of a hundred articles—the model will find big themes that gather the collection into large categories. If the researcher chooses a higher number of topics, the model will find more finely grained patterns. For our analysis, we looked for 200 topics across the collection of 78,000 sentences. Our goal was to group sentences into themes that would span different locations while also being specific enough to differentiate from the whole. Of the 200 topics, several appear very widely across thousands of sentences. These tend to refer to sentences that describe basic geographical facts, as for example Topic 116 (*population, parish, resident, town*) or Topic 196 (*church, built, sittings, contains, parish*). However, most appear in smaller but still significant numbers. The smallest topics appear in only about 200 sentences, while most make up between 250 and 500 sentences. When measured geographically, this means that each topic appears across a few dozen to a few hundred locations, with varying levels of intensity. We'll discuss examples of these topics below.

[9] The fundamentals of topic modeling are explained in David Blei, "Introduction to Probabilistic Topic Models," *Communications of the ACM* 55, 4 (2012): 77-84. For an overview of their use in literary studies, see Ted Underwood and Andrew Goldstone, "The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us," *New Literary History,* 45:3 (Summer 2014): 359-384; Matthew Jockers, *Macroanalysis: Digital Methods & Literary History* (University of Illinois Press, 2013); and Lisa M. Rhody, "Topic Modeling and Figurative Language," *Journal of Digital Humanities,* 2:1 (Winter 2012). An overview of their potential use in the social sciences can be found in Mohr and Bogdanov, "Topic Models: What They Are and Why They Matter?," *Poetics,* 41:6 (December 2013): 545-69. For discussions of some common pitfalls in topic modeling, see Ben Schmidt, "Words Alone: Dismantling Topic Models in the Humanities," *Journal of Digital Humanities,* 2:1 (Winter 2012) and David Mimno et al., "Optimizing Semantic Coherence in Topic Models," in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (Edinburgh, 2011): 262-72.

Text mining reveals how words cluster together in documents. Geospatial modeling reveals how words cluster together in space. Geospatial text analysis combines these techniques. We begin from the assumption that similar places will tend to be described using similar words—port towns will tend to be described using a common language of boats, docks, and fisheries, while farming regions will share a common vocabulary of agriculture. To identify these regions, we use an analytic tool called "hot spot analysis."[10] Criminologists use this same method to identify areas of concentrated violence, and linguists use it to identify regional dialects.[11] For our purposes, the technique identifies regions with common geographical and historical traits.[12] Our goal is to see how past places are described, to uncover how historical events and spatial transformations are noted in different locales, and to trace how Industrial Era Scottish geographers understood their own modernity in relation to the past.

Many of the resultant hotspot maps correlate with established patterns in Scottish historical geography, specifically regarding modernization and industrialization. This alone is worth remarking, for it indicates how geospatial text analysis can identify significant topics from a large corpus

[10] The particular mathematical calculations used in hot spot analysis were designed by Arthur Getis and J. Keith Ord: see Getis and Ord, "The Analysis of Spatial Association by Use of Distance Statistics," *Geographical Analysis*, 24:3 (July 1992): 189-206, and Ord and Getis, "Local Spatial Autocorrelation Statistics: Distributional Issues and an Application," *Geographical Analysis* 27:4 (October 1995): 286-306.

[11] For an example of hot spot analysis applied to the study of crime, see John E. Eck, Spencer Chainey, James G. Cameron, Michael Leitner, and Ronald E. Wilson, *Mapping Crime: Understanding Hot Spots* (Washington D.C.: National Institute of Justice, 2005). For mapping regional dialects, see Yuan Huang, Diansheng Guo, Alice Kasakoff, and Jack Grieve, "Understanding U.S. Regional Linguistic Variation with Twitter Data Analysis," *Computers, Environment and Urban Systems* (December 2015): http://dx.doi.org/10.1016/j.compenvurbsys.2015.12.003.

[12] Hot spot analysis has the added virtue of filtering out errors in the data. Topic models are built using a machine-learning process that inevitably misreads some documents—two sentences may share several keywords even if they're about fundamentally different things. Similarly, the algorithm we used to match names of places with their corresponding descriptions necessarily entails a margin of error. However, these errors represent "noise" in the data that is evenly distributed spatially. By identifying clusters of hotspots where topics occur with unusually high density, this analysis filters out most such noise.

of primary materials and distribute them coherently across space and time. For example, Topic 136, which concerns the textile industry (keywords include *manufacture, linen, employed, looms, trade, cotton, weaving*), is concentrated in the west around Glasgow and the Clyde, and in the east in Tayside and Fife, the historical locations of the textile mills south and east of the Highland Line, with smaller concentrations in the lowland region of the Grampians, the Midlands, and, as a notable outlier, the area of Cullen (Banff and Buchan) where linen manufacture was introduced by the Earl of Findlater in 1748.[13]

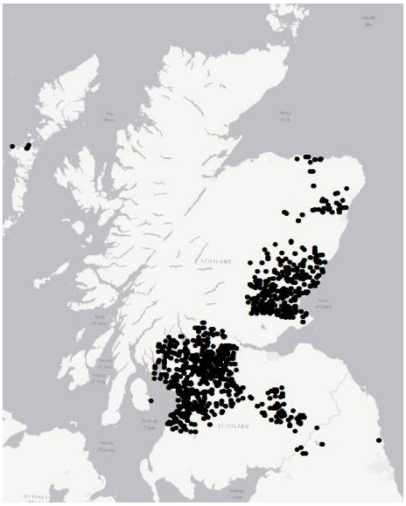
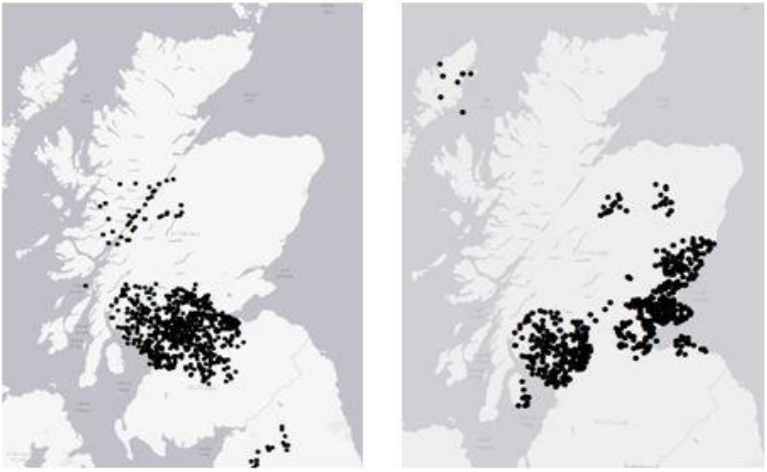


Fig. 2: textile industry keywords

Topic 83 concerns the canals around Glasgow as well as the Caledonian Canal (keywords including most of the central locations as well as *canal, between, river, port, navigation, opened, steam*, as well as *watt*, referencing James Watt who, among his many accomplishments, first surveyed the Caledonian Canal in the 1770s). It is located precisely in these areas around the major canals (the Union, the Forth and Clyde, the Crinan, the Caledonian). Turning outward to export trade and transport, Topic 107 (*tons, vessels, trade, port, foreign, tonnage, british, ships*), concentrates in the coastal areas of the Central Belt with some movement north along the eastern coast past St. Andrews and Dundee. Outlier concentrations in the Cairngorms and Lewis and Harris, harder to read at first sight, call our attention to potentially interesting connections in the literature. All three of these topics show an exponential increase in the nineteenth century; again, not at all surprising, but noteworthy for the

[13] These general patterns are outlined in David Turnock, *The Historical Geography of Scotland since 1707* (Cambridge: Cambridge University Press, 1982); for more detailed information on the textile industries in Scotland during the nineteenth century, see John Butt and Kenneth Ponting, eds., *Scottish Textile History* (Aberdeen: Aberdeen University Press, 1987).



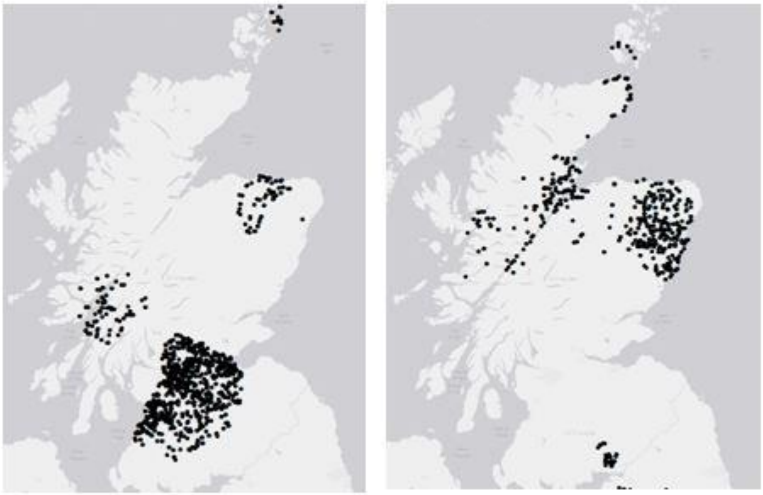
Figs. 3 and 4: keywords for canals (left) and export/shipping (right)

ability of these automated methods to discern and visualize such historical patterns.

Scottish industrial and environmental history is not uniform. Topic 184 concerns iron furnaces *iron, works, large, extensive, added, tons, cast, year, coal, furnaces*) and is concentrated predictably in the Central Belt where the iron industry developed alongside coal mines and canals, and in the vicinity of Loch Etive and Loch Awe, where some of the first iron furnaces were established in the eighteenth century. Its timeline helps us to position most of these locations in a dramatic concentration from the mid-eighteenth through the first two-thirds of the nineteenth centuries (when our corpus ends).[14] By contrast, Topic 140, which concerns agricultural improvement (*acres, land, arable, extent, pasture, waste, cultivation, soil, ground, parish, wood*), distributes mostly in Aberdeenshire, the Great Glen, and the northern Highlands and has a much less dramatic, though still noticeable, uptick during this same

[14] On industrial historical geography, in addition to Turnock, see, for example, Roy Campbell, *The Rise and Fall of Scottish Industry, 1707-1939* (Edinburgh: John Donald, 1980); Christopher A. Whately, *The Industrial Revolution in Scotland* (Cambridge: Cambridge University Press, 1997).

period, registering the longer and relatively more gradual history of agrarian reform.[15]

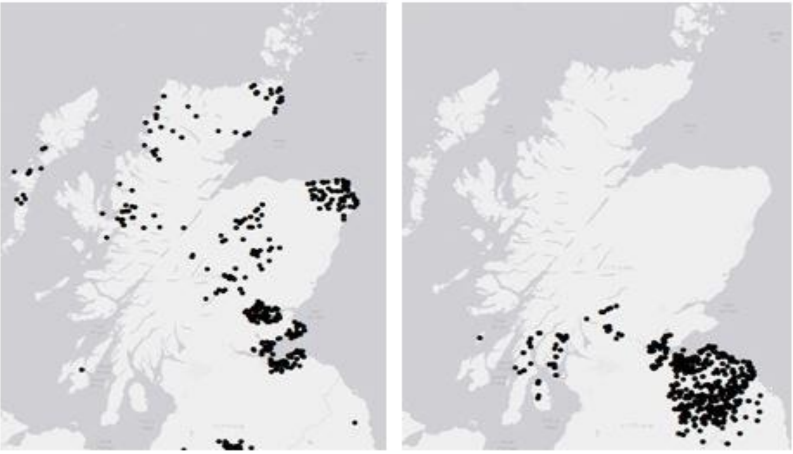


Figs. 5 and 6: keywords for ironworks (left) and arable (right)

These maps can also offer semantic footprints of cultural and literary topics. Topic 144 represents an antiquarian's wish list (*castle, ancient, seat, house, family, residence, ruins, old, once, built*). The map is hardly comprehensive as a guide to Scottish ruins, and hot spot analysis is designed to find regions of concentration, rather than specific points, but an antiquarian scholar or picturesque tourist could do worse than start here. The timeline for this topic contrasts sharply with the industrial and agricultural timelines, concentrating predictably in the Middle Ages. The related Topic 76 concerns William Wallace (*edward, english, castle, sir, england, scottish, wallace, during, army, battle, baliol, king*). It concentrates spatially in the southeast around the major sites of Wallace's

[15] On modern Scottish agricultural history, see T.M. Devine, *The Transformation of Rural Scotland: Social Change and the Agrarian Economy, 1660-1815* (Edinburgh: Edinburgh University Press, 1994) and *Clearance and Improvement: Land, Power and People in Scotland 1700-1900* (Edinburgh: Edinburgh University Press, 2006). See also Eric Richards, *A History of the Highland Clearances,* 2 vols. (London: Croom Helm, 1982) and David Turnock, *Patterns of Highland Development* (London: Macmillan, 1970).

battles, invasions, and defeats while its timeline includes a single spike in the late thirteenth century.



Figs. 7 and 8: keywords for antiquities (left) and William Wallace (right)

Topic 165 (Fig. 9) offers an intriguing though by no means simple map of print culture (*published, scotland, account, time, well, known, author, work, written, letter, learned, edition, map, country, curious, poem*) circulating in the Central-West, southern isles, Oban and the Isle of Mull, the Cairngorms, and the Lothians. Quite a number of topics concern ecclesiastical history, but these concentrate in different geographic and historical patterns than the earlier ones from economic or antiquarian topics. For example, Topic 4 is mostly drawn from sentences in the Statistical Accounts of Scotland (New and Old) that recount the tenures of parish ministers (*minister,* whole of Scotland, the map of its textual traces in the corpus concentrates almost exclusively along and just north of the Central Belt as well as Arran, Kintyre, Islay, and Jura. By contrast Topic 77 is dominated by sentences that tell of new congregations being formed, especially for the various secession kirks and the Disruption of 1843 (e*stablished, congregation, united, secession, house, built, meeting, stipend, relief, church, original, chapel*). It concentrates more around Glasgow and the inland regions of Argyll in the west, Caithness in the north, Angus in the east, and just south of the border in Northumberland. The first topic is distributed relatively evenly

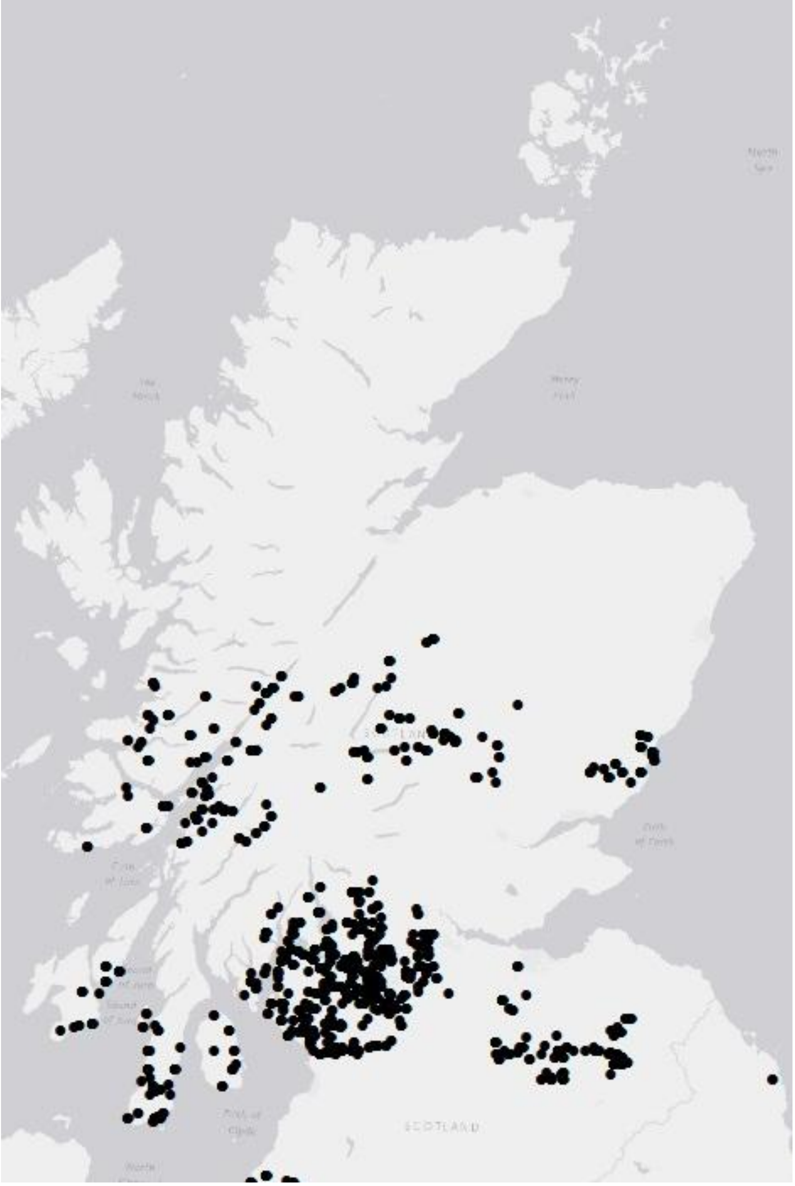
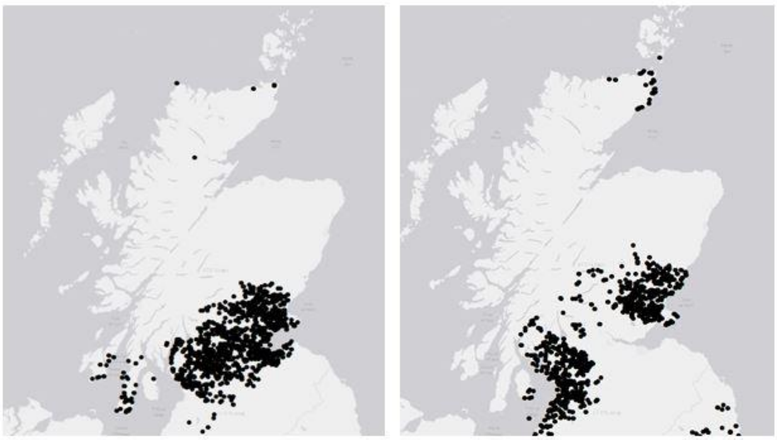


Fig. 9: mapping hotspots from keywords for print culture

across a 1500-1900 timeline, whereas the second follows the pattern of the industrial topics with a huge uptick in the nineteenth century.[16]



Figs. 10 and 11: parish ministers (left) and new congregations (right)

Taken individually, most of these topical distributions should be not be particularly surprising. The patterns they reveal, especially on the earlier topics, largely conform to our understanding of Scottish history and geography. However, we review the model in these broad outlines to emphasize its capaciousness and accuracy, as well as to invite readers to imagine with us how many different lines of inquiry could be pursued using this method. We hope to have indicated the ability of geospatial text analysis to discern verifiable historical and geographical patterns across large archives of texts. This method offers us the potential to reconceive those archives, be they in libraries, private collections, or digital databases, as records of historical geography. The patterns that are easiest to interpret in these archives will be those that conform to previous knowledge. Those less transparent may suggest irregularities in the corpus, but they may also lead to new discoveries—not only about material history, but also about its representation.

The ambition here is not simply to excerpt geospatial data from a body of literature, and so ground it in an ostensibly stable set of coordinates, but to explore how space has been produced according to

[16] On Scottish ecclesiastical history, see, e.g., Callum G. Brown, *Religion and Society in Scotland since 1707* (Edinburgh: Edinburgh University Press, 1997).

different print networks and protocols during periods of social and economic transformation. Once an archive has been mined and mapped in such a fashion, we can begin to ask questions not only about regional descriptions but about the cultural, economic, political, and environmental assumptions and imperatives that informed them. These maps help us to navigate through a wide body of material so as to combine close readings with broader patterns and to understand how place comes to be defined within textual media.

University of South Carolina
and *University of Iowa*