The Professional Geographer

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rtpg20

Investigating Sherlock Holmes: Using Geographic Profiling to Analyze the Novels of Arthur Conan Doyle

Michael C. A. Stevens, Gillian Ray, Sally C. Faulkner & Steven C. Le Comber

To cite this article: Michael C. A. Stevens, Gillian Ray, Sally C. Faulkner & Steven C. Le Comber (2020) Investigating Sherlock Holmes: Using Geographic Profiling to Analyze the Novels of Arthur Conan Doyle, The Professional Geographer, 72:4, 566-574, DOI: 10.1080/00330124.2020.1758575

To link to this article: https://doi.org/10.1080/00330124.2020.1758575

OPEN ACCESS

Investigating Sherlock Holmes: Using Geographic Profiling to Analyze the Novels of Arthur Conan Doyle

Michael C. A. Stevens
Queen Mary University of London

Gillian Ray
The London Clinic

Sally C. Faulkner and Steven C. Le Comber
Queen Mary University of London

Central to spatial analysis of crime is the assumption that the offender has visited the locations analyzed. We show that even if this is incorrect, meaningful patterns can be identified. Using 390 locations from the Sherlock Holmes stories, we show that geographic profiling (GP) ranks Conan Doyle in the top 13 percent of 2,678 historical figures in London, despite the fact at the time of writing he had not visited all of the sites. Restricting the analysis to thirty authors contemporary with the Sherlock Holmes stories, Conan Doyle ranks first, with a hit score of 2.8 percent (above Holmes's address at 221b Baker Street). Finally, we show that GP prioritizes sites strongly associated with Conan Doyle (example.g., his home) compared to those more tangentially associated with him, even when sites are close together. Our analysis, although mostly for amusement, underlines the ability of GP to extract useful information from complex data. **Key Words: Bayesian statistics, criminology, mapping literature, point pattern analysis, spatial data.**

犯罪空间分析的关键是假设犯罪者到过某地理位置。我们证明，即使这个假设不正确，也能发现有益的模式。利用夏洛克·福尔摩斯小说中的390个地理位置，我们证实了，尽管作者在写作过程中没有到过所有的地点，伦敦2,678个历史人物的地理画像排名中柯南·道尔位居前13%。如果只考虑与夏洛克·福尔摩斯小说同时代的30位作者，柯南·道尔以2.8%的分数位列第一（福尔摩斯在贝克尔大街221-b号的地址）。最后，与弱相关地点（尽管空间位置相邻）相比，地理画像优先选择与柯南·道尔强相关的地点（例如，他的住宅）。尽管我们的分析主要是娱乐性质的，也强调了地理画像能够从复杂的数据中提取有益的信息。**关键词：贝叶斯统计，犯罪学，制图，点模式分析，空间数据.**

Un aspecto medular en el análisis del crimen es el supuesto de que el delincuente ha visitado las localizaciones analizadas. Indicamos que aun si tal cosa fuere incorrecta, con ello pueden identificarse patrones significativos. Con el uso de 390 locaciones de las historias de Sherlock Holmes, mostramos que ese trazado de perfil geográfico (GP) coloca a Conan Doyle dentro del grupo cimero del 13 por ciento de 2.678 figuras históricas de Londres, pese al hecho de que en la época en que escribió él no había visitado todos los sitios. Restringiendo el análisis a treinta autores contemporáneos con las historias de Sherlock Holmes, Conan Doyle para al primer lugar, con un puntaje del 2.8 por ciento (encima de la dirección de Holmes en 221b Baker Street). Por último, mostramos que el GP prioriza sitios fuertemente asociados con Conan Doyle (su casa, por ejemplo) en comparación con los que están asociados con él de manera más tangencial, incluso cuando los sitios son cercanos en conjunto. Nuestro análisis, así sea mayormente por diversión, subraya la capacidad del GP para extraer información útil de datos complejos. **Palabras clave: análisis de patrones de puntos, criminología, datos espaciales, estadística bayesiana, mapeo de literatura.**

> It must amuse you to see the vast and accurate knowledge of London I display. I worked it all out from a Post Office map.
>
> —Arthur Conan Doyle, in a letter to his publisher after the publication of *The Sign of Four* (Roden and Roden 1994; Miller 2009, 120)

Geographic profiling (GP) was originally developed in criminology to prioritize the large lists of suspects typical of investigations of serial crime such as murder, rape, and arson. It uses the spatial locations of linked crime sites to build a probability surface that is overlaid on the study area to form a geoprofile and that describes the probability that the offender's anchor point (usually a home or workplace) can be found at this point. Suspects are prioritized according to the heights of locations associated with them on the geoprofile (Rossmo 2000).

GP is highly successful in criminology and is used by law enforcement agencies worldwide (Rossmo 2012). It has recently been applied to biological data—for example, using locations at which animals

The Professional Geographer, 72(4) 2020, pages 566–574
Initial submission, December 2019; revised submissions, February and March 2020; final acceptance, March 2020.

have been observed foraging to identify nests, roosts, or dens or using the addresses of malaria patients to identify breeding sites of the mosquitoes that carry the parasite that causes malaria (Le Comber et al. 2011; Le Comber and Stevenson 2012; Verity et al. 2014; Faulkner et al. 2015). Simultaneously, Bayesian models such as those of O'Leary (2009, 2010) and Verity et al. (2014) have been developed. Whereas O'Leary's models explicitly assume a single source, the last of these is based on a Dirichlet process mixture (DPM) model that can be used even when the number of sources is large and unknown, as will often be the case with biological data.

Our data set consists of locations mentioned in the Sherlock Holmes stories by Arthur Conan Doyle, probably the most famous examples of detective fiction in literature. The principal interest of this data set lies in the fact that—rather than using spatial data associated with an individual to identify anchor points associated with that individual—it allows us to use spatial locations associated with a fictional character to find a real person; in this case, the author, Arthur Conan Doyle.

Until 1890, Doyle lived in Southsea and, in a letter to his publisher dated 6 March 1890 (the month after the publication of the second Sherlock Holmes novel, *The Sign of Four*), wrote, "It must amuse you to see the vast and accurate knowledge of London I display. I worked it all out from a Post Office map" (Roden and Roden 1994; Miller 2009, 120). In addition, the constraints of the genre are likely to have dictated the inclusion of some locations, whether the author was personally familiar with them or not. For example, Threadneedle Street, the home of the Bank of England and the financial heart of the City of London, occurs at least twice in the Holmes stories. A lack of direct evidence that Doyle visited all of the locations where he set the Holmes stories does not hinder the applicability of this study. In criminology there are several cases where a list of prospective crime sites produced by an offender has been subject to GP methods (Rossmo and Harries 2011; Rossmo 2018). It could be argued, then, that Doyle's choice of setting for the Holmes stories came from his own awareness space and understanding of London's geography (Gould 1975; P. J. Brantingham and Brantingham 1981; P. L. Brantingham and Brantingham 1984; Gould and White 1986).

Here, we use the DPM model to analyze the spatial locations of sites in and around London that feature in the Sherlock Holmes stories and ask whether, using these crime sites as input to the model, we can locate sites associated with the author, Arthur Conan Doyle. Specifically, we ask these questions: (1) Using as suspect sites the locations of 2,678 blue plaques (signs commemorating links with famous people, events, or former buildings on the site), how efficiently does the DPM model prioritize the three plaques associated with Conan Doyle? (2) Using the same crime sites but restricting the suspect list to thirty London-based authors (including Conan Doyle) contemporaneous with the original publication of the Sherlock Holmes stories, does the model correctly prioritize Conan Doyle ahead of the other suspects? Finally, using as suspect sites the locations of seventy-four London locations connected to Conan Doyle, we ask (3) whether the model assigns higher priority to areas most strongly associated with the author than areas more tangentially associated with him.

Method

Locations in the Sherlock Holmes Stories

Data were obtained from Wheeler (2018) and are available as a .kml file at https://www.google.com/maps/d/u/0/viewer?ll=51.51034565331305%2C-0.12769532132745098&z=14&mid=11hi6OwDoifyUI4kFsg7suBQm1t8. The list was accessed on 19 June 2019 and contains 390 unique locations.

Blue Plaque Data

Now run by English Heritage, the London blue plaques scheme was started in 1866 and is thought to be the oldest of its kind in the world. There are now more than 900 plaques across London commemorating writers, artists, musicians, and other notable figures. Their popularity has led other organizations—for example, the London County Council and the British Film Institute—to place similar plaques in and around London, and similar schemes now operate in other parts of the world (http://openplaques.org/).

A list of 2,812 blue plaques was downloaded from http://openplaques.org/ using data from April 2018. Three of these are associated with Conan Doyle, with the following addresses and inscriptions: (1) Twelve Tennison Road, South Norwood: "Sir Arthur Conan Doyle author 1859–1930 worked and wrote here 1891"; (2) Two Upper Wimpole Street, London: "Sir Arthur Conan Doyle 1859–1930 creator of Sherlock Holmes lived here 1891–1894"; and (3) Langham Hotel, Portland Place: "Oscar Wilde and Arthur Conan Doyle dined here with the publisher of *Lippincott's Magazine* on 30 August 1889, a meeting that led to *The Sign of Four* & *The Picture of Dorian Gray*." Excluding plaques outside of the study area (i.e., outside an area between –0.38 and 0.08 degrees longitude and between 51.36 and 51.61 degrees latitude) left a total of 2,678 for which hit score percentages were calculated (see the section "Hit Score Percentage" for more details). We also calculated the hit score percentage for 221b Baker Street, Sherlock Holmes's address in the stories.

Table 1 *Locations and hit score percentages for the London addresses of thirty authors, including Arthur Conan Doyle, whose homes feature blue plaques and who were alive during the period from 1887 to 1927, when the Sherlock Holmes stories were originally published*

Latitude	Longitude	Hit score (%)	Author	Author abbreviation
51.3953	–0.08074	2.8	Arthur Conan Doyle	ACD
51.49187	–0.210226	5.5	Henry Rider Haggard	HRH
51.528442	–0.113771	6.5	Amelia Edwards	AE
51.410877	–0.080794	8.2	Raymond Chandler	RC
51.360992	–0.169453	10.7	William Hale White	WHW
51.489797	–0.242602	11.5	AP Herbert	APH
51.448407	–0.124914	11.7	Arthur Mee	AM
51.484969	–.033362	11.7	Italo Svevo	IS
51.497799	–0.242539	21.8	Ouida	O
51.466417	–0.00439	22.1	Samuel Smiles	SS
51.456768	–0.096316	24.0	Sax Rohmer	SR
51.460294	–0.217008	26.4	Theodore Watts Dunton	TWD
51.465985	–0.147978	28.9	Natsume Soeseki	NS
51.475938	–0.164341	29.6	Norman Douglas	ND
51.462714	–0.143576	31.7	Graham Greene	GG
51.569869	–0.186111	33.7	Evelyn Waugh	EvW
51.441824	–0.165834	33.8	Thomas Hardy	TH
51.463757	–0.026288	38.8	Edgar Wallace	EdW
51.445352	–0.065289	40.9	CS Forrester	CSF
51.539783	–0.079529	44.4	Edmund Gosse	EG
51.532575	–0.042763	46.3	Israel Zangwill	IZ
51.489402	–0.267407	47.4	Patrick Hamilton	PH
51.492137	–0.267519	47.9	EM Forster	EMF
51.373387	–0.304103	50.8	Enid Blyton	EB
51.448806	0.062828	58.7	Richard Jeffries	RJ
51.426134	–0.203119	61.8	Georgette Heyer	GH
51.565256	–0.342379	64.1	RM Ballantyne	RMB
51.422477	–0.224531	67.7	Robert Graves	RG
51.578598	–0.004242	90.5	Solomon T Plaatje	STP
51.60012	0.012398	97.4	James Hilton	JH

Author List

A restricted suspect list of thirty individuals was generated by searching the list of English Heritage blue plaques in London for the words *creator*, *writer*, *novelist*, or *author* and removing any nonwriters (e.g., hits for songwriter); authors who fell outside of the study area were again excluded. Only authors whose life span overlapped the publication of the Sherlock Holmes stories (1887–1927) were included (Table 1). Postcodes were recorded and converted to latitude and longitude using https://www.doogal.co.uk/BatchGeocoding.php. The transformation from postcode to longitude and latitude can incur an error due to the shift from polygon to point. Given, however, the inverse relationship between postcode size and population density and that the analysis is based in London, we believe this error to be negligible.

Sites Associated with Conan Doyle

Sites in London associated with Conan Doyle ($n = 74$) were identified from Lycett (2007) and ranked as high ($n = 16$), medium ($n = 21$), or low ($n = 37$) priority (Table 2). High-priority sites are those with a very strong connection to Conan Doyle, such as home addresses; his ophthalmic practice; or clubs of which he was a member (e.g., The Athenaeum and the Reform Club), as opposed to clubs he simply visited; and hotels in which he stayed regularly (e.g., Morley's, where he wrote *The Hound of the Baskervilles*). Medium-priority sites are other sites that Conan Doyle is either known to have visited more than once or where this appears likely; for example, his bank at 125 Oxford Street, his friend EW Hornung's house, or theaters where his plays were staged. Finally, low-priority sites are those remaining locations for which there is no evidence that he visited more than once or twice. Hit score percentages for all three groups were compared using a linear model.

The Dirichlet Process Mixture Model

The DPM model of GP was first introduced in Verity et al. (2014) and combines advantages from Rossmo's (2000) criminal geographic targeting algorithm and O'Leary's (2010) simple Bayesian model. The DPM model provides a mathematically robust way of estimating source locations from the spatial locations of the observations. The DPM model is preferable over more conventional spatial clustering algorithms because it does not require a user to specify the number of sources the model must estimate prior to running the model. It will, in fact, estimate this parameter from the data, a huge advantage in cases where the data describe multiple unknown sources. The model splits the difficult problem of estimating the locations of multiple sources into two much simpler problems using a Gibbs sampler (Geman and

Table 2 *Locations and hit score percentages for seventy-four sites identified from Lycett (2007) and designated as high, medium, and low priority*

Latitude	Longitude	Hit score %	Description
High priority			
51.508361	–0.126214	0.0	Golden Cross Hotel
51.50831	–0.126909	0.0	Morley's Hotel
51.50676	–0.132742	0.2	Athenaeum
51.50699	–0.126017	0.2	Grand Hotel
51.50648	–0.133242	0.8	Reform Club
51.506207	–0.124249	1.0	Authors Club
51.520548	–0.149099	2.4	ACD's opthalmic practice
51.3953	–0.08074	2.8	ACD home
51.520344	–0.128141	3.5	ACD lodgings
51.506076	–0.134815	4.8	Royal Automobile Club
51.509362	–0.137475	6.7	ACD's church
51.505337	–0.137016	9.3	Pall Mall Club
51.499382	–0.130253	14.5	ACD's bookshop
51.511651	–0.113727	17.6	Idlers's Club
51.496734	–0.143732	20.4	ACD home
51.516501	–0.173064	21.1	ACD lodgings
Medium priority			
51.510207	–0.127084	0.1	Duke of York's Theatre (ACD play staged)
51.509784	–0.126803	0.1	London Coliseum (ACD play staged)
51.50349	–0.194225	0.1	EW Hornung's house
51.506468	–0.124359	1.0	Hotel Metropole
51.511264	–0.133553	2.0	Lyric Theatre (ACD play staged)
51.511361	–0.13337	2.0	Apollo Theatre (ACD play staged)
51.418688	–0.018876	3.8	Beckenham Golf Club
51.510024	–0.122853	4.0	Adelphi Theatre (ACD play staged)
51.509452	–0.136322	5.0	St. James' Hall
51.512563	–0.135607	7.6	AP Watt, ACD's agent
51.510312	–0.120918	7.9	Savoy Theatre (ACD play staged)
51.510556	–0.120833	8.4	Terry's Theatre (ACD play staged)
51.511575	–0.11999	11.0	Lyceum Theatre (ACD play staged)
51.515841	–0.135681	12.9	Capital and Counties Bank (ACD's bank)
51.513056	–0.1175	13.4	Gaiety Restaurant
51.499694	–0.130076	14.5	Imperial Theatre (ACD play staged)
51.499253	–0.127309	15.3	Westminster Abbey
51.494553	–0.177701	15.6	London Spiritualist Alliance
51.501378	–0.14189	18.0	Buckingham Palace
51.529418	–0.172792	20.3	Lord's cricket ground
51.535184	–0.173696	21.6	ACD attends seances
Low priority			
51.515758	–0.084037	0.3	Amateur Billiards Championship
51.510733	–0.133025	1.1	Trocadero restaurant
51.519163	–0.15653	1.2	Baker Street Bazaar
51.420971	–0.07037	2.4	Crystal Palace Park
51.463821	0.003123	4.6	Dines with Jean Leckie
51.517782	–0.143986	5.1	Langham Hotel
51.512596	–0.149254	5.3	Savile Club
51.513629	–0.129817	5.5	Royal English National Opera
51.513851	–0.077941	6.4	ACD tours Jack the Ripper sites
51.503312	–0.127612	6.8	Downing Street
51.518024	–0.142585	7.6	Queen's Hall
51.494079	–0.210414	7.6	Society for Psychical Research
51.514	0.0605482	9.0	ACD tours Jack the Ripper sites
51.512111	–0.144663	9.9	Aeolian Hall
51.515445	–0.119278	10.4	Kodak House
51.502783	–0.124245	10.9	Scotland Yard
51.523443	–0.13684	12.1	James Schoolbred, shop
51.509229	–0.139699	12.1	Burlington House
51.51194	–0.13956	12.5	Café Monico
51.515819	–0.101792	14.5	Old Bailey
51.499885	–0.126708	15.2	St. Margaret's Church, Westminster
51.501027	–0.177415	15.2	Royal Albert Hall
51.498521	–0.134961	15.4	Caxton Hall
51.517342	–0.105982	15.5	City Temple
51.498889	–0.137515	16.1	Wellington House
51.499911	–0.124605	16.7	House of Commons
51.49906	–0.176892	16.9	International Health Exhibition 1884
51.4984	–0.1401	17.7	St. Peter's and St. Edward's church
51.518376	–0.075403	18.8	ACD tours Jack the Ripper sites
51.505518	–0.145579	19.6	Royal Societies Club
51.511907	–0.106894	19.8	*Daily Mail* offices
51.495174	–0.143886	21.3	Victoria Station

(*Continued*)

Table 2 *(Continued).*

Latitude	Longitude	Hit score %	Description
51.520003	–0.060549	25.5	ACD tours Jack the Ripper sites
51.520315	–0.072661	26.1	ACD tours Jack the Ripper sites
51.503123	–0.110746	26.2	Union Jack Club
51.481484	–0.165399	36.1	Automatic Sculpture Company
51.6	–0.246	85.5	Hendon aerodrome

Note: ACD = Arthur Conan Doyle.

Geman 1984; Neal 2000). First, the model partitions the observations into clusters, where observations that are close together are more likely to be derived from the same source and there is some non-zero probability that an observation is associated with a new unseen source. Second, the model estimates the locations of these sources, conditional on the clustering in the first step. The model repeats these steps many thousands of times using standard Bayesian Markov chain Monte Carlo methods until it converges on the posterior distribution of interest. The output of the DPM model is a superposition of mixture densities where each component is a bivariate normal distribution centered on a source location. Faulkner et al. (2016) extended this process by allowing for the DPM model to estimate the standard deviation (referred to as *sigma*) of the bivariate normal distributions that make up the mixture.

Model Implementation

The DPM model described here was implemented in R (R Core Team 2019) using version 2.1.0 of the package *Rgeoprofile* introduced by Verity et al. (2014); this package is available at https://github.com/bobverity/Rgeoprofile, where model settings are explained in detail by the documentation. The model settings chosen for analyzing the locations associated with Sherlock Holmes stories were as follows. Our prior on *sigma* was governed by the settings: *sigma_mean* = 1, *sigma_var* = NULL, and *sigma_squared_shape* = 2. The combination of *sigma_var* = NULL and *sigma_squared_shape* = 2 corresponds to manually defining the shape parameter of the inverse gamma prior on sigma squared. Appendix 1 of Faulkner et al. (2016) describes this in full detail, but these settings correspond to a very diffuse prior on sigma, so that the model fits the standard deviation of the bivariate normal from the data. Furthermore, the settings for running our Markov chain Monte Carlo algorithm were as follows: *samples* = 100,000, *chains* = 10, and *burnin* = 1,000.

Hit Score Percentage

Typically, the model's performance is assessed by the hit score percentage. The suspect site's hit score is the proportion of the geoprofile that must be searched before that site is located. Suspects are ranked in order of their hit score percentage and investigated in this order; the lower the hit score, the higher priority the suspect. For a random, unprioritized search, the hit score percentage will be, on average, 50 percent. Thus, a hit score percentage of 5 percent describes a search strategy that is ten times more effective than an unprioritized search, because only 5 percent of the total area will have to be searched before locating the site of interest.

Results

Blue Plaque Data

The model fitted a value of sigma of approximately 680 m. This equates to a cluster size with a diameter of approximately 4 km, similar to that typical of data from humans in urban environments (Rossmo 2000). The DPM model estimated between twenty-one and forty source locations best described the data, with a mean of twenty-eight and a standard deviation of two. The three sites associated with Conan Doyle had hit score percentages of 1.6 percent, 5.1 percent, and 2.8 percent respectively, ranking 349, 615, and 438 out of 2,678 sites (Figure 1). Thus, these would be found after searching 13.0 percent, 23.0 percent, and 16.4 percent of the list of plaques—more effective than an unprioritized search by factors of approximately four, two, and three, respectively. The lowest hit score percentage was for the plaque for physician John Hughlings Jackson (1835–1911), at Three Manchester Square, Marylebone (0.0028 percent). Sherlock Holmes's address, 221b Baker Street, had a hit score percentage of 15.65 percent, which means that it would have ranked 1,337 out of 2,678 had it been included in this list.

Author List

When all authors were ranked by hit score percentages, Arthur Conan Doyle ranked first out of thirty, with a hit score percentage of 2.8 percent. Thus, the DPM model correctly allocates a higher priority to Conan Doyle than to the other suspects. Henry Rider Haggard ranked second, with 5.5 percent (Table 1, Figure 2). 221b Baker Street would have ranked ninth had it been included in this list.

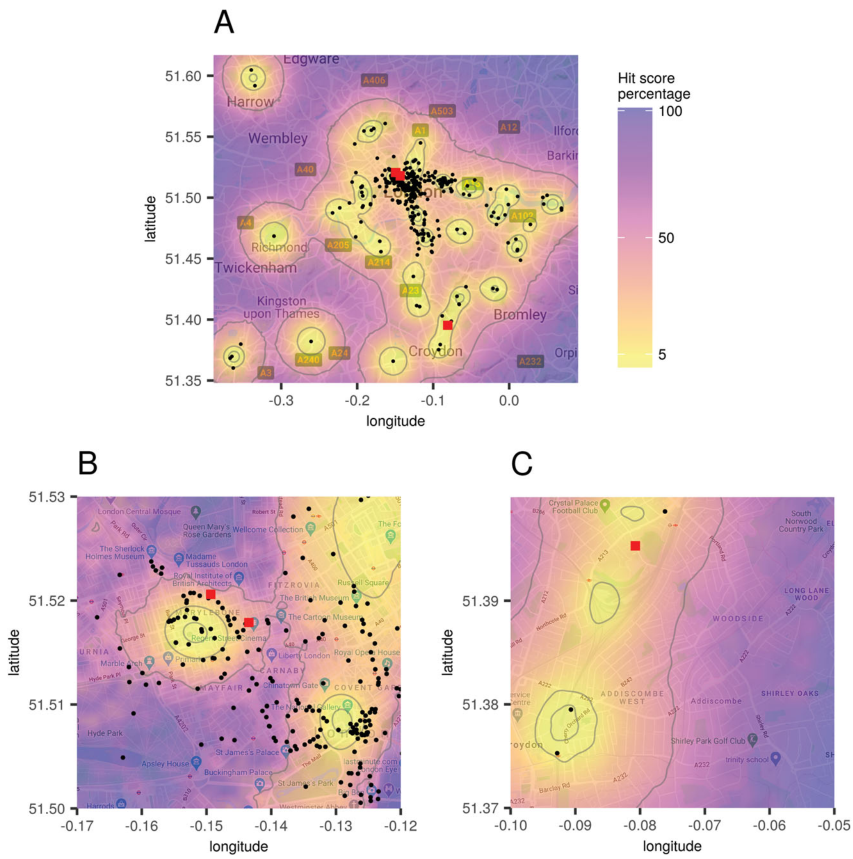

Figure 1 *(A) Full geoprofile, with zoomed views of the areas around plaques relating to Conan Doyle in (B) central London and (C) Norwood. In all three panels, locations mentioned in the stories are shown in black, with the locations of the three blue plaques relating to Conan Doyle in red. For clarity, the remaining 2,675 plaques are not shown and contour lines corresponding to the top 1, 5, 50, and 100 percent of the profile are visible.*

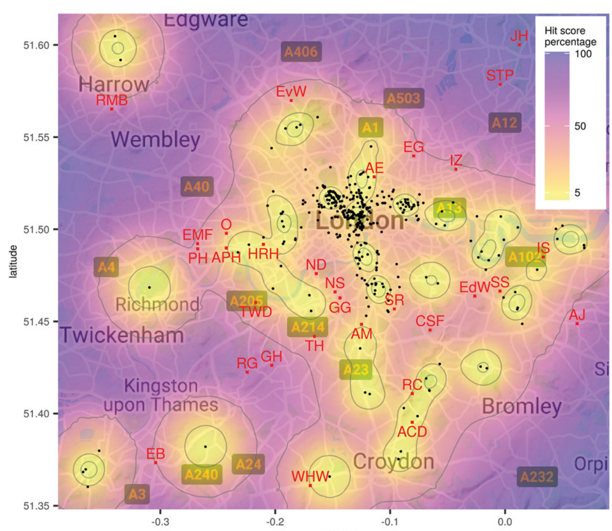

Figure 2 *Geoprofile from Figure 1, showing the locations of plaques associated with authors contemporary with the publication of the Sherlock Holmes stories (red squares). Author abbreviations are given in Table 2. Contour lines for the top 1, 5, 50, and 100 percent of the profile are shown.*

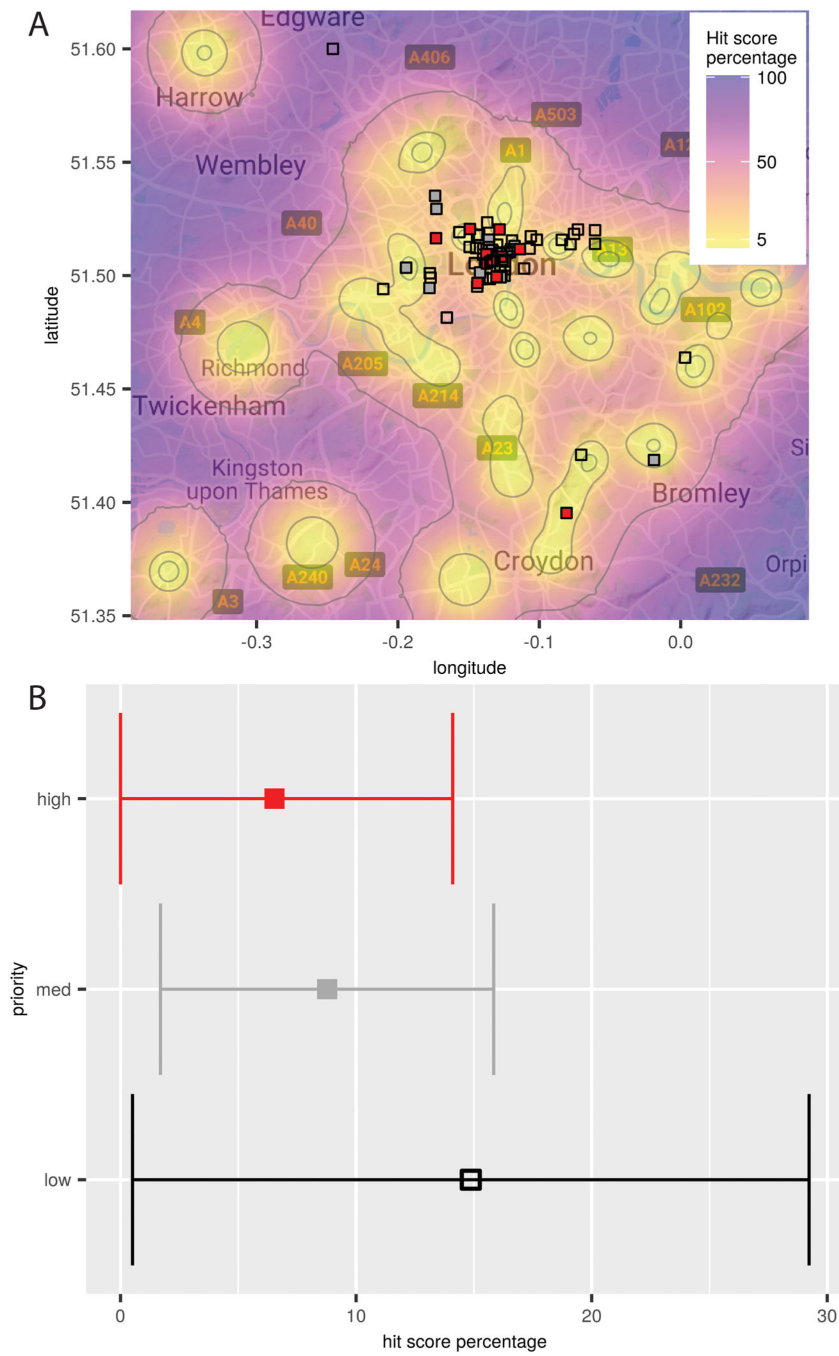

Figure 3 *(A) Geoprofile from Figure 1, showing low-priority (open squares), medium-priority (gray squares), and high-priority (red squares) sites associated with Conan Doyle. Sites and their priority are listed in Table 2. Contour lines for the top 1, 5, 50, and 100 percent of the profile are shown. (B) Hit score percentage means with standard deviations for the same three groups of sites.*

Sites Associated with Conan Doyle

The restricted nature of the study area meant that low-, medium-, and high-priority sites occurred in close proximity to each other. On average, low-priority sites were 1.7 km ($SD = 2.3$ km) from the nearest high-priority site, and medium-priority sites were 0.9 km ($SD = 1.2$) from the nearest high-priority site. In total, fifty-five of fifty-eight low- or medium-priority sites were within three sigma of a high-priority site and would thus overlap with a cluster around that site. Despite this, hit score percentages differed significantly between sites of different priority, linear model $F(2, 71) = 3.739$, $p = 0.03$. Low-priority sites had a hit score percentage of 14.9 percent ($t = 7.955$, $p = 2 \times 10^{-11}$), with medium-priority sites 6.15 percent lower, but not significantly so ($t = 1.970$, $p = 0.05$), and high-priority sites 8.35 percent lower ($t = 2.441$, $p = 0.02$; Figure 3).

Discussion

In this study, we present the first instance in which GP has successfully pinpointed locations of interest associated with a real individual when analyzing locations of a fictitious one. In the first analysis, the three plaques associated with Conan Doyle had hit score percentages from 1.6 to 5.1 percent, meaning that they ranked in the top quarter of the full suspect list, comprising 2,678 blue plaques. Each of these is thus found much more efficiently than would be the case for a random, unprioritized search. Similarly, in the second analysis, Conan Doyle ranked at the top of a suspect list of thirty authors, with a hit score percentage of 2.8 percent. The search strategy described by the DPM model is thus seventeen times more efficient than an unprioritized search.

Perhaps most impressive is the model's ability to differentiate between sites strongly associated with Conan Doyle and others that are more weakly linked to him. This is particularly striking given the close proximity of sites to each other. On average, medium- and low-priority sites were within 1.4 km of the nearest high-priority site. The fitted value of 680 m for sigma (the standard deviation of the bivariate normal distribution around a source) implies clusters with a diameter of 4 km, larger than the mean distance from low- or medium-priority sites to a high-priority site. Despite this, the model was able to distinguish between low- and high-priority sites, with the latter having lower hit score percentages. Although the model failed to distinguish between low- and medium-priority sites, this is perhaps not surprising because in this instance—as well as the issue of proximity between sites—designating sites as low or medium priority is in some cases a matter of judgment and might not always reflect reality. For example, Crystal Palace Park is mentioned only once in Lycett (2007; when Conan Doyle plays his first-ever first-class cricket match, for the Marylebone Cricket Club against WG Grace's London County team, scoring four runs) and is thus designated here as low priority. It was close to the house in Norwood, however, where Conan Doyle lived from 1891 to 1894, and it is likely that he visited the park more than once—as indeed its hit score percentage (2.4 percent) might suggest.

A related difficulty is that we have not attempted to take into account the fact that the bulk of the Sherlock Holmes canon—all four novels and forty-four of fifty-six short stories—were written by 1917. Obviously, sites where Conan Doyle's connection postdates this (including high-priority sites like his home in Buckingham Palace Mansions, Victoria) can have only limited impact on the locations the author chose to use in the stories. It is interesting to note, then, that this site had the second highest hit score percentage (20.4 percent) of any high-priority site except Conan Doyle's lodgings in Forty-Four Norfolk Square (21.1 percent), where he appears to have stayed for only for a very short period—perhaps even a matter of days (Pugh 2018). Similarly, Conan Doyle's Psychic Bookshop at Two Victoria Street is designated as high priority but was only set up in 1925, with just six short stories remaining to be published, and has a hit score percentage of 14.5 percent. Excepting these three sites, all but one of the remaining high-priority sites have hit score percentages under 10 percent, with more than half of them (ten out of sixteen) below 5 percent.

Conclusion

We show here that the DPM model of GP can perform well even when the data are loosely associated with an individual, when different suspect sites occur in close proximity, or when temporal issues confound the analysis—and suggest that our study underlines the power of spatial mapping in fields as diverse as criminology, ecology, and epidemiology. We note that Sherlock Holmes—probably the most famous detective in all of literature—appears himself to have been aware of the importance of spatial analysis: In one story, "The Adventure of the Priory School," Conan Doyle included a map of the school and its neighborhood presented as drawn by the great detective and which is crucial to the case. We would like to think that Holmes—or at least Arthur Conan Doyle—would approve of this study. ■

Supplemental Materials

Supplemental data for this article can be accessed on the publisher's site.

ORCID

Michael C. A. Stevens http://orcid.org/0000-0003-2480-0312
Sally C. Faulkner http://orcid.org/0000-0003-0342-4389
Steven C. Le Comber http://orcid.org/0000-0001-9027-8409

Literature Cited

Brantingham, P. J., and P. L. Brantingham. 1981. *Environmental criminology*. 39th ed. Beverly Hills, CA: Sage.

Brantingham, P. L., and P. J. Brantingham. 1984. *Patterns in crime*. New York: Macmillan.

Faulkner, S. C., M. D. Stevenson, R. Verity, A. H. Mustari, S. Semple, D. G. Tosh, and S. C. Le Comber. 2015. Using geographic profiling to locate elusive nocturnal animals: A case study with spectral tarsiers. *Journal of Zoology* 295 (4):261–68. doi: 10.1111/jzo.12203.

Faulkner, S. C., R. Verity, D. Roberts, S. S. Roy, P. A. Robertson, M. D. Stevenson, and S. C. Le Comber. 2016. Using geographic profiling to compare the value of sightings vs. trap data in a biological invasion. *Diversity and Distributions* 23 (1):104–12. doi: 10.1111/ddi.12498.

Geman, S., and D. Geman. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (6):721–41. doi: 10.1109/TPAMI.1984.4767596

Gould, P. 1975. Acquiring spatial information. *Economic Geography* 51 (2):87–99. doi: 10.2307/143066.

Gould, P., and R. White. 1986. *Mental maps*. 2nd ed. London and New York: Routledge.

Le Comber, S. C., D. K. Rossmo, A. N. Hassan, D. O. Fuller, and J. C. Beier. 2011. Geographic profiling as a novel spatial tool for targeting infectious disease control. *International Journal of Health Geographics* 10 (1):35–38. doi: 10.1186/1476-072X-10-35.

Le Comber, S. C., and M. D. Stevenson. 2012. From Jack the Ripper to epidemiology and ecology. *Trends in Ecology & Evolution* 27 (6):307–8.

Lycett, A. 2007. *Conan Doyle: The man who created Sherlock Holmes*. London: Weidenfeld & Nicholson.

Miller, R. 2009. *The adventures of Arthur Conan Doyle*. London: Pimlico.

Neal, R. M. 2000. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* 9 (2):249–65. doi: 10.2307/1390653.

O'Leary, M. 2009. The mathematics of geographic profiling. *Journal of Investigative Psychology and Offender Profiling* 6:253–65.

O'Leary, M. 2010. Implementing a Bayesian approach to criminal geographic profiling. Paper presented at the 1st International Conference and Exhibition on Computing for Geospatial Research & Application, 22 June, Washington, DC.

Pugh, B. W. 2018. *A chronology of the life of Sir Arthur Conan Doyle*. 2nd ed. London: MX.

R Core Team. 2019. *R: A language and environment for statistical computing*. Accessed January 30, 2019. https://www.r-project.org/

Roden, C., and B. Roden. 1994. The newsletter of the Arthur Conan Doyle Society, Number 11. *The Parish Magazine* 11:1–32.

Rossmo, D. K. 2000. *Geographic profiling*. Boca Raton, FL: CRC.

Rossmo, D. K. 2012. Recent developments in geographic profiling. *Policing* 6 (2):144–50. doi: 10.1093/police/par055.

Rossmo, D. K. 2018. Geoprofiling terrorism. In *The Oxford handbook of environmental criminology*, ed. G. J. N. Bruinsma and S. D. Johnson, 801–20. Oxford, UK: Oxford University Press.

Rossmo, D. K., and K. Harries. 2011. The geospatial structure of terrorist cells. *Justice Quarterly* 28 (2):221–48.

Verity, R., M. D. Stevenson, D. K. Rossmo, R. A. Nichols, and S. C. Le Comber. 2014. Spatial targeting of infectious disease control: Identifying multiple, unknown sources. *Methods in Ecology and Evolution* 5 (7):647–55. doi: 10.1111/2041-210X.12190.

Wheeler, T. B. 2018. *The mapped London of Sherlock Holmes*. London: Author.

MICHAEL C. A. STEVENS is a PhD Student at the School for Biological and Chemical Sciences at Queen Mary University of London, London E1 4DQ, UK. E-mail: M.Stevens@qmul.ac.uk. His research interests include developing geographic profiling to cater for different data types in ecology and epidemiology.

GILLIAN RAY is Secretary of the Head Physician to The Royal Household at The London Clinic, Marylebone W1G 6BW, UK. E-mail: Geckogill82@mail.com. She is a longtime Holmesian and geographic profiling enthusiast and graduate of Queen Mary University of London with interests in herpetology and apiology.

SALLY C. FAULKNER is a Lecturer at the School for Biological and Chemical Sciences at Queen Mary University of London, London E1 4DQ, UK. E-mail: S.C.Faulkner@qmul.ac.uk. She has developed the geographic profiling model for use with biological data, addressing conservation issues. Her interests lie primarily in wildlife crime and conservation.

STEVEN C. LE COMBER was a Lecturer at the School for Biological and Chemical Sciences at Queen Mary University of London, London E1 4DQ, UK. E-mail: S.C.Lecomber@qmul.ac.uk. His interests included evolutionary biology, spatial patterns within ecology, and mathematical models of molecular evolution.