Received: 25 August 2020 | Accepted: 24 January 2022

DOI: 10.1111/rssa.12833

ORIGINAL ARTICLE

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Royal Statistical Society Series A Statistics in Society

A modelling strategy to estimate conditional probabilities of African origins: The collapse of the Oyo Empire and the transatlantic slave trade, 1817–1836

Ashton Wiens[1] | **Henry B. Lovejoy**[2] | **Zachary Mullen**[3] | **Eric A. Vance**[4]

[1]Department of Applied Mathematics, University of Colorado Boulder, Boulder, Colorado, USA

[2]Department of History, University of Colorado Boulder, Boulder, Colorado, USA

[3]Department of Computer Science, University of Colorado Boulder, Boulder, Colorado, USA

[4]Laboratory for Interdisciplinary Statistical Analysis (LISA), University of Colorado Boulder, Boulder, Colorado, USA

Correspondence
Eric A. Vance, Laboratory for Interdisciplinary Statistical Analysis (LISA), 331 ECOT, 1111 Engineering Drive Box 526, University of Colorado Boulder, Boulder, Colorado 80309-0526, USA.
Email: Eric.Vance@Colorado.EDU

Abstract

Intra-African conflicts during the collapse of the kingdom of Oyo from 1817 to 1836 resulted in the enslavement of an estimated 121,000 people who were then transported to coastal ports via complex trade networks and loaded onto slave ships destined for the Americas. Historians have a good record of where these people went across the Atlantic, but little is known about where individuals were from or enslaved *within* Africa. In this work, we develop a novel statistical modelling strategy to describe the enslavement of people given documented violent conflict, the transport of enslaved peoples from their location of capture to their port of departure, and—given an enslaved individual's location of departure—that person's probability of origin. We combine spatial prediction of conflict density via kriging with a Markov decision process characterising intra-African transportation. The results of this model can be visualised using an interactive web application to plot—for the first time—estimated conditional probabilities of historical origins during the African diaspora. Understanding the likely origins within Africa of enslaved people may have ramifications for the history

 © 2022 Royal Statistical Society

of the Atlantic world, whereby the ocean *connects*, rather than disconnects, Africa, the Americas, and Europe.

KEYWORDS

African diaspora, digital humanities, Gaussian process, kriging, Markov decision process, migration

1 | INTRODUCTION

The transatlantic slave trade consisted of over 12.5 million people boarding slave ships along the coast of Africa between 1514 and 1866 (Eltis, 2008; Manning & Liu, 2020). One of the egregious after-effects was the erasure of the identities of millions of Africans; children, women and men removed from their homes, forced into slavery, their names changed, their birth places and family ties obliterated. African historians have had great trouble understanding the origins of enslaved African people on a macro-scale during the entire era of the transatlantic slave trade because of the scarcity of primary sources in the pre-colonial period (Lovejoy, 2011). The size of inland populations during this period and the inland migrations that took place are unknown and remain a major debate within the field. This paper explores the internal African origins of enslaved people by focusing on a single quadrant of West Africa during the collapse of the kingdom of Oyo and expansion of the Sokoto Caliphate (see Figure 1) from 1817–1836. Oyo, a former slave trading state and major supplier to the transatlantic slave trade, was once located in what is now the Yoruba-speaking areas of southwestern Nigeria, as well as parts of Benin and Togo.

According to Eltis in *Voyages: The Trans-Atlantic Slave Trade Database* (2008), an estimated 121,000 individuals from this region and time period involuntarily boarded slave ships, which primarily went to three destinations: Brazil (42%), Cuba (40%), the British and French Caribbean (<1%), and due to British abolition efforts after 1807, Freetown, Sierra Leone (17%). Detailed shipping records sketch when and where people boarded slave ships at the coast, and when and where they went in diaspora (Eltis, 2008; Manning & Liu, 2020). These data allow for an

(a) (b)

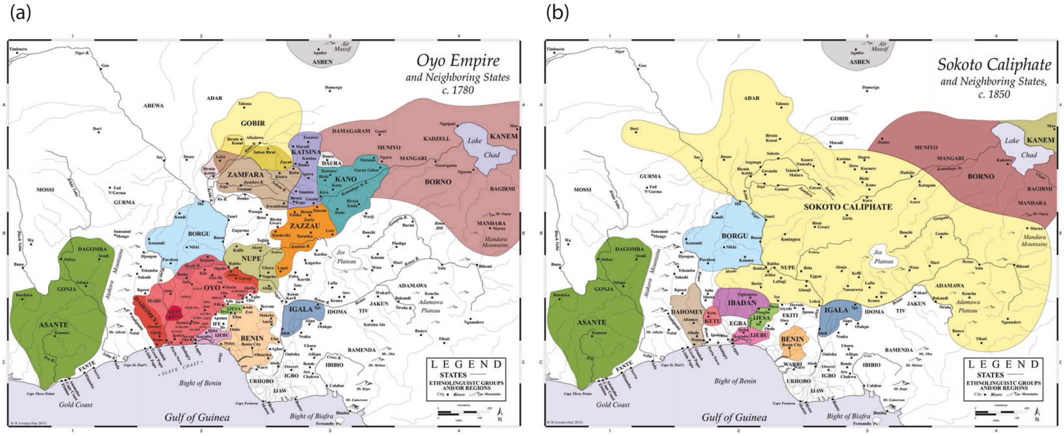

FIGURE 1 (a) Map of region from 1780, and (b) Map of region from 1850 (Lovejoy, 2019)

excellent understanding of when and where more than half of the enslaved people from Oyo went by way of the Atlantic, but they contain many gaps and inconsistencies (Manning et al., 2015), and they tell none of the story for the inland points of origin of these people. Genetic studies have observed a concordance between ancestry based on genetic composition and the existing historical documentation of slave ship voyages (Micheletti et al., 2020). However, this inference is done at large spatial scales, while the work developed here is applied at much smaller areas and periods of interest.

Oyo provided a motivating case study to model the origins of enslaved people during the transatlantic slave trade. Due to early diplomatic and missionary activity, Oyo's collapse is relatively well-documented compared with other places and periods in Africa. Transient European explorers witnessed and documented warfare in Oyo, though without fully understanding the local or regional context (Bruce Lockhart & Lovejoy, 2005; Lander, 1830; Lander & Lander, 1832). Missionaries recorded narratives of enslaved Africans and oral accounts, but usually several years after they boarded slave ships and were liberated in British abolition efforts in Sierra Leone (Ajayi, 1967; Curtin, 1967; Irving, 1856; Johnson, 1921; Lloyd, 1967). Recently, H.B. Lovejoy compiled annual maps of instances of conflict surrounding the turbulent era of Oyo's collapse, documenting the highest intensities of places of enslavement stemming from warfare in the Bight of Benin hinterland between 1817 and 1836 (2019). This recent breakthrough in mapping intra-African conflict has unlocked the possibility of (a) modelling how conflict can generate enslaved individuals who are (b) transported through trade routes to ports of sale. In this paper, we describe these models and then use them to (c) determine likely origins of Africans transported involuntarily to the Americas or elsewhere within Africa.

Our modelling strategy reconstructs the internal slave trade during Oyo's collapse by synthesising spatial statistical models from points of historical conflicts and conjoining them with models for decision processes governing the inland transit of enslaved individuals. Our model describes where individuals may have been captured and enslaved during conflict in a given time period and then predicts the port of departure accounting for the possible routes that could be taken out of the region. As an analogy, our methodology operates like a watershed model in geology that predicts where precipitation collects and drains into a body of water (Nelson et al., 1994); or like catchment areas in human geography that describe the geographical boundaries of populations comprising, as an example, medical practices (Jenkins & Campbell, 1996). Ultimately, we use our model to estimate the inland origin of an enslaved person from Oyo *given* that person's port of departure. The focus of this paper is on Oyo inland migrations to the coast, although our model could be applied to understand internal population movements into newly founded cities in Oyo or elsewhere within Africa.

This novel approach answers three questions relevant for historians and researchers of contemporary or historical forced migrations, which often involve highly uncertain or incomplete datasets:

1. How can one model the enslavement of people or the creation of refugees given historical documentation of war and violent conflict?
2. How can one model the transport of enslaved individuals, refugees, or migrants from their location of origin to their location of departure from the region? Or more generally, how can one model the migration of people using a sequential decision process?
3. Given an enslaved individual, refugee, or migrant's location of departure from a region, the conflicts in that region, and the possible transportation routes, what is that person's likely origin?

This paper presents a first step in an iterative process to infer origins of inland population movements, specifically for our case study of the transatlantic slave trade during the collapse of the Oyo Empire (1817–1836). As such, we believe it will help link together existing open-source data, such as the *Voyages* database (Eltis, 2008) and recently published results of genetic analyses of the African diaspora in the Americas (Micheletti et al., 2020). In addition, we hope that this paper will motivate the interdisciplinary community of researchers of the African diaspora to compile more data or uncover new sources of historical data, specifically from Oyo and West Africa. With more data from surrounding regions and periods, we could refine our modelling strategy to validate the model and tune its parameters with out-of-sample data, resulting in more confidently answering historical questions of interest. Knowing more about internal population movements and periods of conflict leading to enslavement could inform Africans and their descendants about their ancestry and heritage; thus providing a better understanding of the largest, forced transoceanic migration in human history.

In Section 2 we describe the historical data on conflict, trade routes, and slave ships in Oyo from 1817–1836. In Section 3 we describe the methods we use for creating conflict density functions, simulating enslavement from these functions, modelling the transit of enslaved individuals to points of sale, and creating maps of the conditional probabilities of origin given the point-of-sale. We present some illustrative results in Section 4.1 and describe our interactive web application for exploring the data and models in Section 4.2. Finally, we discuss in Section 5 how the models could be applied by historians; their limitations, potential extensions, and improvements; and their relevance for other applications in the study of forced migrations. Section 6 concludes this paper.

2 | HISTORICAL CONFLICT, TRADE ROUTES, AND SLAVE SHIP DATA

This section describes several geopolitical data sets used in the work curated by historian H.B. Lovejoy describing the enslavement, conflicts, and trade routes that existed during Oyo's demise from 1817–1836. First, we have annually recorded conflicts during the period, which may last one or more years. We have a trade network comprised of over 250 populated places that characterises the movement of people in the region, and this network is assumed valid during the entire period. We also have estimates of the total number of enslaved people departing the region as a whole and departing from specific trading ports, some with references to the years the journey took place. Finally, the visualisations produced in this work utilise shapefile data of prominent geographical features that existed in the region during 1817–1836. In particular, we include bodies of water in maps that are relevant to identifying the boundaries of the various states. These data were downloaded from http://www.diva-gis.org/ (Hijmans et al., 2001). Several bodies of water created by damming long after the historical period under analysis were removed from the data set.

While we have confidence in the quality of our data, we recognise that much historical data remains uncertain; cities may have been destroyed by conflict some years before or after we have indicated and trade routes used early in Oyo's collapse may have been abandoned decades later. We hope that this paper inspires historians and other researchers to debate and refine the data underlying our models.

2.1 | Historical conflict in Oyo, 1817–1836

Leading into the 19th century, Oyo was a major West African slave trading state that supplied tens of thousands of enslaved Africans to European slave traders at the coast. Prior to 1817, the kingdom had remained relatively stable. In 1817 this stability was disrupted by a Muslim slave uprising at Ilorin. Following a series of internal crises, jihad, and foreign invasions from virtually every direction between 1817 and 1836, Oyo gradually disintegrated. Cities, towns, and villages were attacked, some were destroyed and others were founded. When populated places (referred to hereafter as 'cities') were involved in conflict, those conquered were routinely captured, enslaved, and transported in slave caravans along existing trading routes to ports on the Atlantic coast as well as inland into internal slave markets, especially northward into the Sokoto Caliphate.

As Cameron has argued based on archaeological and bio-archaeological evidence, the enslavement of people around the world usually began following a violent act, most especially through warfare, raiding, or kidnapping (Cameron, 2016). Beyond human motivations, climate change also impacted enslavement, as drought and desertification would often precipitate conflict and increase vulnerabilities of people across the region (Webb, 1995).

H.B. Lovejoy compiled a list of cities in Oyo involved in conflict from 1817–1836, their corresponding spatial coordinates, start and end dates surrounding conflict at each city, political affiliations of the city, and an extensive list of primary or secondary sources from which the data are derived (2019). A conflict intensity scale is encoded as a categorical variable with two levels: 2 indicates a city was attacked and 3 represents a city that was destroyed. Cities could be attacked or destroyed over several years. Cities were often rebuilt in the same or slightly different locations and attacked or destroyed again in subsequent years. New cities were also founded, which is represented with 9 in the data set. The data are available in our data repository at https://osf.io/h6upw. For an annual overview of political and conflict maps, slave ship departures, available documentation, and imagery (including photographs) of the people involved in Oyo's collapse see http://www.yorubadiaspora.org (Lovejoy, 2021b).

Conflict and the transportation of enslaved people generally occurred during the dry season (December to April) since it was likely harder to move armies and caravans in the rainy season (Law, 1977; Nunn & Puga, 2012). This temporal framework immediately raises issues as to whether or not the Roman calendar is an appropriate method for dating conflict data on an annual basis (Lovejoy, 2019). However, to align with the slave voyage data based on the Roman calendar, we treat conflict data as annual, discrete time observations. As historians continue refining their understanding of conflict in Oyo, additional degrees or levels of conflict could be added to the data set. In Section 3.1 we describe how we create maps of conflict density based on this data set.

2.2 | Historical trade routes

Slavery in Oyo was highly visible and people were often sold more than once via various middlemen who facilitated transportation through different states. Heavily-armed caravans of large groups of enslaved people would have been a common result of intense periods of conflict. Since Oyo had been a major supplier of enslaved people to the Americas before its gradual collapse in the 19th century, slave routes and networks of slave traders had been well established across the region for centuries (Folayan, 1980; Law, 1977; Lovejoy, 2019; Nunn, 2008).

H.B. Lovejoy (2019) also describes how trade routes were compiled from lists and locations of cities in Oyo from 1817–1836. To record the trade network among these cities, we constructed

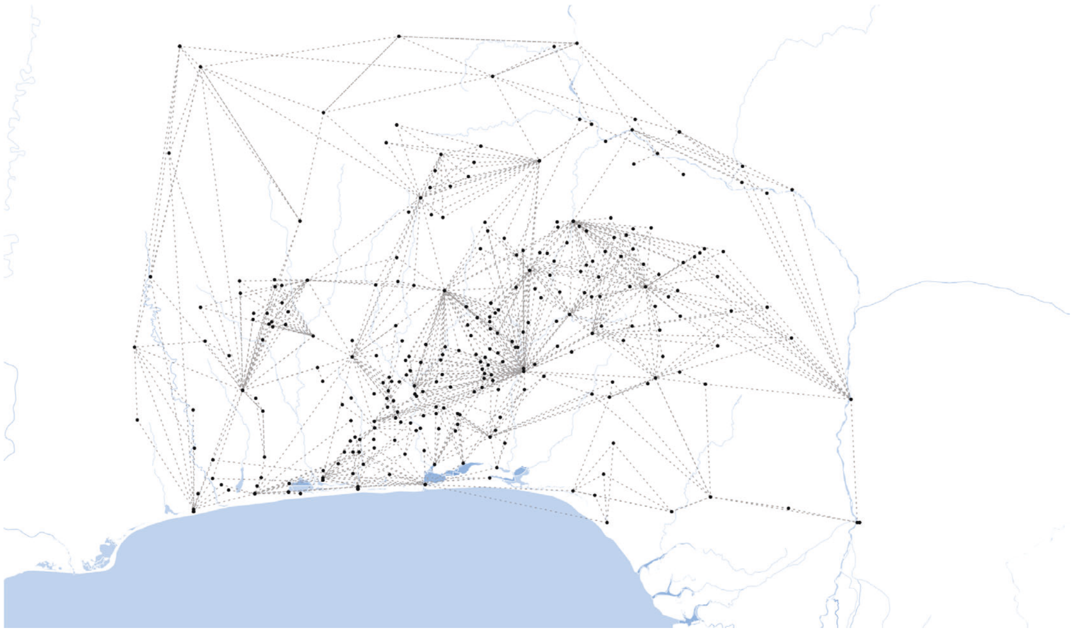

FIGURE 2 Map of trade in Oyo, 1817–1836, which is represented by the adjacency matrix for the Markov decision process

an adjacency matrix of the historical connections between cities, all of which are connected in some way to potential coastal and inland ports-of-sale. These data are also available at https://osf.io/h6upw. Figure 2 illustrates the possible routes in the trade network. One assumption of our model is that this trade network is constant over time. As Oyo lost control of the region over time, the historical trade routes may have shifted due to arising conflicts and the British Royal Navy's blockades of slave trading ports. This transition is not recorded in these data but could be incorporated into future modelling frameworks.

We encoded the relationships (edges) between the cities (nodes) of this graph into an $n \times n$ adjacency matrix A for a set of n locations $s_1, \dots, s_n$. An entry A_{ij} is equal to 1 if there is a connection starting at s_i and ending at s_j and 0 otherwise. Since trading caravans travelling from s_i to s_j could also travel in the reverse direction, the adjacency matrix describing the historical trade routes is symmetric, reflecting the fact that if it is safe and convenient to travel a trade route in one direction, the same holds in the opposite direction. We use this adjacency matrix to define the set of available actions in a Markov decision process, described in Section 3.2.

2.3 | Slave ships and passenger logs

Voyages: The Trans-Atlantic Slave Trade Database contains data for 243 documented slave ships departing several ports in the Bight of Benin carrying approximately 75,700 people out of a total, estimated 121,000 people departing the Atlantic coast between 1817 and 1836 (Eltis, 2008). However, many documented voyages have missing data for the port, year of departure or arrival, or the number of enslaved individuals embarking on the ships. Some researchers have applied statistical techniques such as sampling using Markov chain Monte Carlo methods to estimate total embarkations by decade and region of Africa for slave ships with missing data (Manning & Liu, 2020; Manning et al., 2015). Since these imputations are by decade and region (not year or specific port), we use the estimated annual departures data from Eltis (2008).

While specific documentation for who was on board slave ships do not exist for most ships, starting in 1807 the British Royal Navy captured many slave ships, escorted them into Vice Admiralty Courts and bilateral courts of mixed commission in Freetown, Rio de Janeiro, and Havana, and compiled detailed lists of the enslaved people forced on board. These detailed passenger lists included transliterations of African names which could be interpreted to determine a name's language (whether Yoruba, Fon, Hausa, among others) for use in future analyses surrounding the unique ethnolinguistic composition of individual slave ships, as contemplated in Nwokeji and Eltis (2002). In the future, linguistic data for transliterated African names might be aligned on an annual basis to the internal zones of conflict and probabilistic origins we are attempting to identify in this paper. Currently, scholars are examining over 30,000 transliterated African names of people leaving the Bight of Benin documented in ship registers made in Sierra Leona and Havana (Eltis & Misevich, 2009). Ojo has effectively shown how registers of Liberated Africans can help determine more accurately where people came from inland. In the case of the slave ship *Manuelita*, which departed Lagos in 1833 and where 477 enslaved people were registered in Havana, Ojo interpreted the transliterated African names and was able to determine that these names were 'common to central Yorubaland and specifically to Ife and Owu [Yoruba sub-groups]' (2017, p. 371). Based on these data, Ojo further demonstrates that the majority of people from the *Manuelita* almost certainly came from the town of Oko Oso/Ecomosho and 'were victims of either the Gbanamu or Erunmu war' (2017, p. 371).

3 | MODELS FOR MIGRATION DUE TO CONFLICT

The models we use to describe historical migration due to conflict are comprised of two main parts. First, we create a conflict surface from documented points of conflict and then use this surface as a density function representing points of capture and enslavement. Then, the simulated locations of the captured and enslaved individuals are used as input to a model describing forced migration through Oyo. The models used to create the conflict surface and internal migrations are discussed in the following sections. Ultimately, we use these models to determine the conditional probability of origin of an enslaved person given that they departed Oyo from a specific port.

3.1 | Mapping historical conflict density

Building on the historiography, H.B. Lovejoy (2019) has recorded major events surrounding the fall of Oyo, such as the kingdom's borders collapsing inwards, Dahomey achieving independence to the west, the founding of the emirate of Ilorin in jihad to the east, and lost territory to a coalition of neighbouring Yoruba-speaking kingdoms to the southeast. While the historical borders of the resulting states are relatively well known, answering the question of from where enslaved people originated requires analysing the conflicts themselves to determine the regions within the greater Oyo area most impacted by each conflict. Within our region and time period of study, we have geo- and time-referenced data on instances of conflict and the destruction of cities (with examples of protracted warfare recorded as less intense than the complete destruction of cities), as described in Section 2.1. However, a conflict does not unfold over just the sites of the major battles and events:

armies mobilise, advance, raid, retreat, occupy and abandon cities—thereby generating enslaved individuals—throughout contested regions (Cameron, 2016). To account for this behavior based on highly uncertain historical data, we created a continuous spatial map of the regions of conflict given the discrete time and place data available.

3.1.1 | A Gaussian process for generating slave capture locations

The data consist of a vector of conflict intensity measures $\boldsymbol{Y}$ observed at 2-D spatial locations $\mathbf{s}_1, \mathbf{s}_2, \ldots \mathbf{s}_n$. The statistical process model is

$$Y(\mathbf{s}) = Z(\mathbf{s}) + \varepsilon(\mathbf{s}),$$

where $\varepsilon(\cdot)$ is assumed to be an independent mean zero Gaussian white noise process with nugget variance τ^2, representing measurement error or microscale variation. $Z(\cdot)$ is modelled as a mean zero spatially correlated Gaussian process with Matérn covariance function. In particular, the Matérn covariance function is

$$\mathrm{Cov}(Z(\mathbf{s}), Z(\mathbf{s}')) = k(\mathbf{s}, \mathbf{s}') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{d}{\kappa}\right)^{\nu} K_\nu \left(\frac{d}{\kappa}\right), \tag{1}$$

where $d = \|\mathbf{s} - \mathbf{s}'\|$ is the Euclidean distance between points, σ^2 is the marginal variance (or sill), $\kappa > 0$ is the spatial range parameter, and $\nu > 0$ is the smoothness. $\Gamma(\cdot)$ is the gamma function, and K_ν is the modified Bessel function of the second kind of order ν. The Matérn covariance function is a popular choice in spatial statistics because it is a flexible model with interpretable parameters (Stein, 1999).

The Gaussian process considers the data $\boldsymbol{Y}$ to be a single draw from a multivariate normal on $\mathbb{R}^n$. This yields a log likelihood proportional to

$$\ell(\boldsymbol{Y}) \propto -\log \det \left(\Sigma + \tau^2 I\right) - \boldsymbol{Y}^T \left(\Sigma + \tau^2 I\right)^{-1} \boldsymbol{Y},$$

where I is an $n \times n$ identity matrix, and Σ is given by the Matérn covariance function in Equation (1) with the $(i,j)^{\text{th}}$ entry defined by $\Sigma_{i,j} = k(\mathbf{s}_i, \mathbf{s}_j)$.

The formal kriging estimator fills in a map of the Oyo region with a conflict intensity measure at chosen resolution by taking each desired location $\mathbf{s}_0$ and computing the estimated conflict intensity

$$\hat{\boldsymbol{Y}}(\mathbf{s}_0) = k(\mathbf{s}_0, \mathcal{S}) \left(\Sigma + \tau^2 I\right)^{-1} \boldsymbol{Y},$$

where $k(\mathbf{s}_0, \mathcal{S}) = (k(\mathbf{s}_0, \mathbf{s}_1), k(\mathbf{s}_0, \mathbf{s}_2), \ldots, k(\mathbf{s}_0, \mathbf{s}_n))$ is a $1 \times n$ vector with entries defined by the Matérn covariance function evaluated pairwise at $\mathbf{s}_0$ and every point in $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots \mathbf{s}_n\}$.

3.1.2 | Estimating kriging model parameters

We identified covariance parameters governing spatial distribution of conflict that correspond to our intuition about the the historical narrative using Cressie weights (Cressie, 1985) with the variogram based on the conflict data from the single year 1832, which had the highest number of observations of conflict (45) for any year in the period 1817–1836. The range was determined to be $\kappa = 0.13$ degrees latitude/longitude, the smoothness $\nu = 3$, the sill $\sigma^2 = 0.4$, and the nugget $\tau^2 = 0.1$.

While this is smooth in the context of dense spatial data, our observations are quite sparse, and higher smoothness helps ensure that our maps of conflict intensity can include ridge-like structures between separated conflicts, which mimic shifting borders resulting from the ebb and flow of warfare. Smaller values for the smoothness and range parameters would enforce a more rapid decay to zero away from observations and would push our model closer to one found from kernel density estimation with small bandwidth because it would result in conflicts being modelled as small, radially symmetric, disconnected, additive kernels around the observed locations. Note that the covariance parameters are treated as fixed, meaning that uncertainty in these parameters is not propagated through the model, as could be done with a more sophisticated model. Also, we did not assess the sensitivity of our model results as a consequence of using these specific kriging parameters.

We did not include a mean part of the statistical model so that the conflict intensity surface decays to zero away from conflict. Thus, these four covariance parameters are all that are required to perform spatial kriging at any desired location.

The surface created by kriging exists in the units of the $\boldsymbol{Y}$, which is the intensity of conflict at a location. A key assumption of our modelling strategy is that the estimated density of conflict is proportional to the probability of capture and enslavement. People near clusters of conflict are assumed to have a higher probability of enslavement than people far away from conflict. By using the kriging estimator to fill in a high-resolution surface over the Oyo region for each year or any set of years under investigation, we can create a probability density function of enslavement by normalising by the sum total of all predictions on the grid as follows. Formally, we predict $\hat{Y}(\mathbf{g}_i)$ on a regular rectangular grid $\mathbf{g}_i$ for $i = 1, \ldots, n_g$, and then normalise these predictions to create a probability density function

$$\tilde{Y}(\mathbf{g}_i) = \hat{Y}(\mathbf{g}_i) \bigg/ \left(\sum_{i=1}^{n_g} \hat{Y}(\mathbf{g}_i) \right). \tag{2}$$

Since the numerical computation is performed on a discrete regular grid, $\tilde{Y}$ is a probability mass function and can be used to simulate the origin locations of people enslaved as a result of conflict in the region for any given year or set of years.

3.2 | Modelling the transit of enslaved people

The map in Figure 2 displays cities within and around Oyo connected by the most likely trade routes at the time (Lovejoy, 2019). Such a depiction naturally translates to a graph-based approach to capturing the economics of the region. As described in Section 2.2, the map is derived from an adjacency matrix of valid city-to-city movements. In addition, we classify a handful of cities to be absorbing states where enslaved people historically departed the region externally via Atlantic ports (Lagos, Ouidah, Little Popo, Jakin, Badagry, and Porto Novo) or internally within slave markets (denoted Off Map NE, SE, and NW). Rather than directly assign probabilities to the flow of enslaved people in the region, we apply a finite horizon Markov decision process to model the decisions of slave traders during this period since this family of models can be constructed to emulate sequential decision making (Boucherie & Van Dijk, 2017).

3.2.1 | The historical narrative

To model the historical transit of enslaved people through the Oyo region, we must consider three historical realities and express them in statistical terms in our transit model. First, while historical trade routes have been pieced together by historians, the exact routes from points of capture to points of sale remain largely unknown and anecdotal (Ojo, 2017). Historically, slave trading routes changed over time, shifting in response to conflict, the economics of the region, and perhaps the ethnicity and personal preferences of the slave traders. These factors almost certainly resulted in a variety of plausible paths throughout the region. Consequently, our transit model requires sufficient variation to allow for enslaved people captured in similar locations to deviate to different ports of departure simply by chance, rather than always following the simplest path with the shortest route.

Second, much of the slave trade had to pass through pre-determined cities for the collection of taxes or tributes, as well as protection. Additionally, areas of conflict could ensnare slave traders causing them to be captured and enslaved. As a result, slave traders were incentivised to avoid areas of conflict (Ajayi, 1967; Curtin, 1967; Lloyd, 1967). Statistically, this means that we require a model for the transit of enslaved people to be able to downweight probabilities of transit based on conflict intensities as well as distance travelled.

Finally, the traffic of enslaved individuals at the ports of sale shifted over time, with a generally eastward preference among slave traders over time from Ouidah to Lagos. This shift was largely in response to British anti-slavery blockades becoming more prominent in front of Ouidah (Eltis, 2004). Therefore, to align our model with the historical narrative, our method must allow for some sale locations to be preferred to others a priori, whether that preference is informed by volume of trade, the price of slaves, or the gradual west-to-east blockade of West African slave ports by the British Royal Navy's African squadron during suppression efforts of the transatlantic slave trade.

3.2.2 | A Markov decision process

A Markov decision process (MDP) describes the partially deterministic and partially stochastic movement of an agent through a network in discrete time. The agent's actions at each state are chosen based on the rewards and costs associated with reaching future states in the network, and the actual event that takes place can be probabilistic. Formally, an MDP consists of a 5-tuple (S, A, P_a, R_a, γ).

1. S is a finite set of states in a network, often spatially located. These are the n_y cities $S_y = \{s_1, \dots, s_{n_y}\}$ in the trade network in year y, as well as the added classifications of some cities as terminal or absorbing states.
2. A is the set the actions an agent can take from any given state $s \in S$. These are the edges or roads connecting cities, or the action 'sell' that exists only in the absorbing states. A_y is an adjacency matrix of the trade network in year y such that $a_{ij} = 1$ if and only if cities with locations s_i and s_j are directly connected, that is, a slave trader could travel from s_i to s_j without passing through a third city s_k.

3. For each possible action $a \in A$ taken in state s at time t, we define the probability $P(s'_{t+1}|s_t, a)$ of reaching state s' for all states in S. In general, this probability of result-given-decision can cover a wide range of unexpected outcomes such as the deviation of a caravan from a planned path or a delay in which a caravan could remain in place for a number of iterations, which opens avenues for non-stationarity in time or estimating the duration of transit that we do not further explore. Instead we allow for deterministic results-given-actions so that $P(s'_{t+1}|s_t, a = s') = 1$.
4. For each possible transition s_t to s'_{t+1} we define a reward $R(s'_{t+1}|s_t, a)$. This includes two major components: a positive reward R_T associated with the 'sale' action available in absorbing states and the negative reward R_M associated with the cost of the movement from city s_t to s'_{t+1}.
5. The discount factor $\gamma \in [0, 1]$ multiplicatively downweights rewards incurred in later time steps. We fixed the discount factor at $\gamma = 1$ because we wanted the cost of movement to increase relative to the distance travelled and conflict traversed rather than by the raw count of the number of cities visited.

The trade networks defined by S_y and A_y are kept constant from year to year in our study period, but could change for future work as more historical data become available. The goal of the MDP is the optimal policy $\pi^*(s)$ for each state s, specifying the best choice among the actions a available to the agent at s. This choice is found by maximising the expected future rewards or utility $U(s)$ given by making only optimal choices beginning from state s. $U(s)$ is given by the utilities of the neighbouring states to s, so that the policy π or action $a \in A(s)$ is scored by

$$E\left[U^{\pi}(s)\right] = E\left[\sum_{s'} \gamma P(s'|s, a)\left(R(s'|s, a) + U(s')\right)\right]$$

and simplifies in our case of $\gamma = 1$ and $P(s'|s, a = s') = 1$ to

$$E\left[U^{\pi}(s)\right] = R(s'|s, a) + E[U(s')].$$

The resulting MDP policy is given by

$$\pi^*(s) = \underset{a \in A(s)}{\operatorname{argmax}}\, E\left[U^{\pi}(s)\right],$$

which—given our adjacency matrix—is solely dependent on the cost-benefit reward structure $R(s'|s, a)$.

We include two factors to describe the rewards $R = R_T - R_M$. Negative values ($-R_M$) represent cost of movement through conflict and positive values (R_T) correspond to the decision to sell an enslaved person at a point-of-sale absorbing state. Movement along each edge in the network incurred a cost proportional to $D * (1 + \lambda C)$, where D is the length of that edge, C is the maximum of the conflict kriging estimate $\tilde{Y}$ along that edge, and λ is a scaling factor for conflict. Formally,

$$R_M = R(s'_{t+1}|s_t, a) = \frac{D_{ss'}(1 + \lambda C_{ss'})}{C_{max}} \times \mathbb{1}_{t+1}(s \to s'),$$

where $\mathbb{1}_{t+1}(s \to s')$ is the indicator function of the agent moving from state s to s' at time $t + 1$ and C_{max} is the maximum value of $\tilde{Y}$ over the entire grid $\{\mathbf{g}_i\}_{i=1}^{n_g}$. The cost of movement could be generalised to include any other factors that might apply to the historical context, including disincentives to cross borders, venture through certain terrains, or deviate from historical trade routes.

Positive rewards R_T can be incurred in the MDP by reaching an absorbing state. Movement into these states represents a sale, gaining a (positive) terminal reward. If the historical sale price of slaves at each port were known, these could be used to set the terminal rewards R_T. Absent such data, we set terminal rewards R_T for the set of n_A absorbing states to be normally distributed with a common mean and a variance σ_R^2 reflecting the slave traders' (random) personal preferences and imperfect information, that is, $R_T \sim N(\boldsymbol{\mu}_R, \sigma_R^2 \boldsymbol{I})$, where $\boldsymbol{\mu}_R$ is a vector of length n_A, and I is an $n_A \times n_A$ identity matrix.

We optimised the total reward criterion using the policy iteration algorithm implemented in the `R` package `MDPtoolbox` (Chadès et al., 2017) to find the optimal policy $\pi^*(s)$. Given a draw from the distribution of R_T, the end result of the policy iteration algorithm is a 'best route' for a slave trader to reach a point-of-sale from any given origin location, which is determined by the trader's personal preferences and/or their imperfect information about the actual terminal sale rewards as well as the cost of moving through conflict. Due to the random deviations we encoded in the reward R_T, this route may be potentially different for each enslaved individual, even for those originating in the same location.

3.2.3 | Aligning the MDP with the historical narrative

Modelling the transit of enslaved people with this MDP formulation allows for considerable flexibility in meeting our criteria for a transit model that aligns with the historical narrative to allow for variability in paths taken from point of capture to point-of-sale, avoiding transit through conflict areas, and non-uniform rewards at points of sale.

Randomness in the terminal rewards R_T can account for both individual slave trader preferences and the broader temporal shift from the western port of Ouidah to the eastern port of Lagos. Setting all point-of-sale rewards to be equal asks the question, 'What is the least resistance route to *any* point-of-sale', whereas varying the reward vector allows for individual slave traders to balance preferred or higher revenue sale locations with the implied costs of a longer journey or a journey through regions of conflict.

Once the conflict intensity surface $\hat{Y}$ is created via kriging, $\tilde{Y}$ (see Equation 2) can be repeatedly sampled to generate an arbitrarily large sample of simulated people enslaved due to conflict. Each individual can be passed to the MDP with corresponding reward vector R_T, modelling the probabilistic movement of an individual through the trade network to an absorbing city. These simulations enable us to compute conditional probabilities of capture locations given final points-of-sale, which address the ultimate goal of this study: from which locations did enslaved people originate inland before departing specific ports? We further discuss the use of unequal rewards to add an appropriate amount of stochasticity to our simulation results in Sections 3.2.4 and 3.3.

3.2.4 | MDP parameter tuning

Due to the lack of historical information on the price of slaves within Africa as a function of point-of-sale, we choose as a baseline an MDP with equal expected rewards $\boldsymbol{\mu}_R = 10 \cdot \mathbf{1}_{n_A}$ with variance $\sigma_R^2 = 0.1$ for each absorbing state, where $\mathbf{1}_{n_A}$ is a unit vector of length n_A. σ_R^2

was chosen heuristically, but could be tuned in the future to out-of-sample data. One could envision including a random effect to induce correlation between the reward vector and the simulation location. However, we chose to capture this correlation through the cost of distance and conflict, which causes simulated individuals enslaved near each other to often reach the same ports.

Varying the conflict scaling factor λ results in conflict contributing more or less to the cost of travel and decision making in the MDP, resulting in different proportions of enslaved individuals arriving at terminal nodes. To validate the output of our model against historical data, we estimated λ by aggregating the conflict and ship totals data by Atlantic port for all years and minimising the difference between the expected totals at each port E_i calculated by our model and the observed totals at each port O_i. The expression

$$\sum_{i=1}^{n_P} \frac{(E_i - O_i)^2}{E_i},$$

where n_P is the number of absorbing states on the Atlantic coast, has the form of a $\chi^2_{n_P-1}$ statistic. Minimising this expression yielded an optimal value of $\lambda = 1.55$, which we used to produce conditional probability maps of origins for each year.

3.3 | Mapping conditional probabilities of origin

3.3.1 | Annual large-scale simulation

To gain origin and departure information from the models in Sections 3.1 and 3.2, for each year we generate 1000 enslavements of individuals from the conflict density function $\tilde{Y}$. Then, for each such enslaved individual, we generate a random terminal reward vector R_T. We fit an MDP for each enslaved individual and reward vector pair, resulting in an optimal path of travel to a port-of-exit. For each port, the origin locations of individuals arriving there are smoothed to form a map of probable origins, as is described in the next subsection. The simulations are generated in large number to be able to represent the variability in origins induced by the variability of $\tilde{Y}$ and R_T. Any number of simulations could be generated, and 1,000 was chosen to fill in the entire density function by the single set of simulations without excessive computation.

3.3.2 | Kernel density smoothing for maps

To integrate these simulations—at this point a large collection of origin points encoded by their port-of-exit—into more cleanly interpreted spatial maps, we use a simple kernel density estimator, which creates a small, radially decaying kernel function at each simulated slave capture location leaving from a specific port. The aggregation of each kernel function for every enslaved individual leaving a given port results in a heat map for enslavement origin given the port of departure. As historians often know the port of departure of enslaved people, we can use the repeated samples of simulated data to estimate probabilities of origins for the enslaved individuals who left a given port in a given year.

Formally, the kernel density estimate takes a radially symmetric function $K(r)$ and estimates the regional heat map $\hat{f}$ via the weighted sum

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{|x - x_i|}{h}\right),$$

where $x_1, x_2, \ldots x_n$ are the n capture locations for the enslaved people departing from the port in question. We choose the multivariate normal as the radial function K, as is often convention (Terrell & Scott, 1992). In general, a kernel density estimate requires only tuning one parameter: the bandwidth h that determines the distance/width of the kernel function centered on each simulated slave capture location. We used the `R` function `kde2d` in package `MASS` (Venables & Ripley, 2002). While a kernel density function can be sensitive to the number n of points employed, our simulation-based model allows us to simulate any arbitrary number of spatial samples and construct the resulting kernel density estimator to desired precision. In the web application described in Section 4.2 we allow h to vary from 0.25 to 6 degrees latitude/longitude for a sample of 1,000 simulated enslaved individuals per year and find that this range provides visually acceptable maps. A larger simulated sample would in turn allow for smaller bandwidths.

The sum of kernels corresponding to all simulated individuals approximates the conflict density function as more individuals are simulated. The density can be decomposed based on which port-of-exit the simulated individual ended in via the MDP. To generate a proper probability density function that integrates to 1, we normalise the sum of the kernels for any subset of the exit ports. Numerically, for each port(s) of exit, the KDE is predicted on a regular grid, with predictions in grid cells that lie in the ocean set to 0. Then the predicted values are normalised by the sum of all KDE predictions on the entire grid. This outcome results in a probability estimate, summing to 1, conditional on the individuals ending up in the selected absorbing state(s).

3.4 | Model summary

Our method for using historical data to estimate probabilities of origins of enslaved people during the collapse of Oyo from 1817–1836 works in four steps:

1. Using space-time locations for historical conflict, for each year, create a map estimating the conflict density $\hat{Y}$ that represents the shifting borders of the wars involved. Normalise $\hat{Y}$ into a density function $\tilde{Y}$ indicating the probability of enslavement, then sample enslavement locations from $\tilde{Y}$.
2. Create a trade map based on historical roads, cities, and common trading routes at the time. Specify the cities known to be historical points-of-sale on this map.
3. For each sample generated in step (a), create an MDP using the adjacency matrix in step (b) by pairing it with a randomised terminal reward vector R_T. Record each enslaved individual's point of capture, route, and point of departure. Our implementation of the MDP includes two flexible parameters: the cost-of-movement through conflict scaling factor λ and the variance of the terminal rewards σ_R^2.
4. For each port of departure, create a conditional heat map of enslavement origins via kernel density estimation.

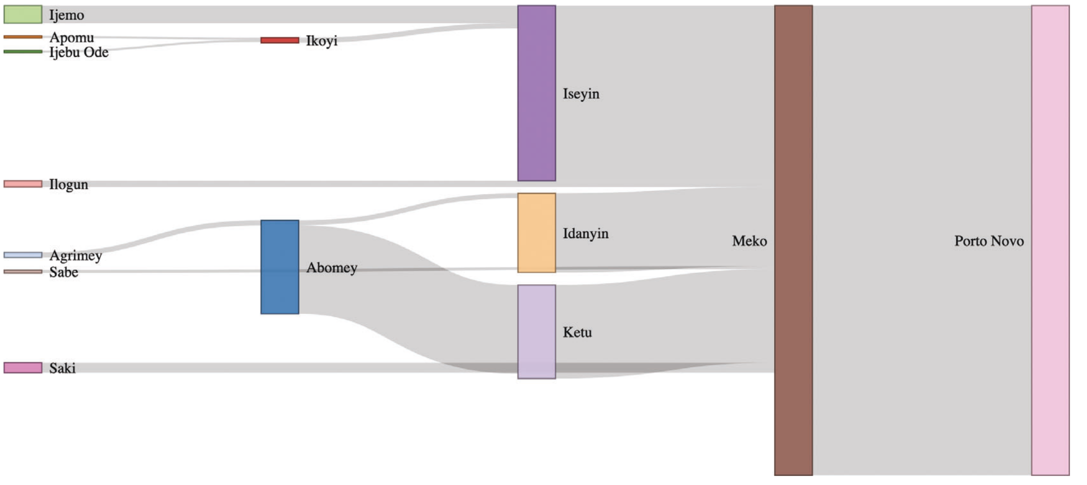

FIGURE 3 Sankey diagram showing the paths of all individuals arriving in Porto Novo over all years 1817–1836

4 | RESULTS AND INTERACTIVE WEB APPLICATION

4.1 | Results

Results from our modelling strategy can inform the possible origins and paths of enslaved individuals, as well as relative numbers of individuals arriving in each port of departure.

We present a Sankey diagram in Figure 3, which illustrates the probabilistic origins and paths of enslaved individuals who departed Oyo from the absorbing state Porto Novo based on our model aggregated over all years. The width of the line connecting cities indicates how many individuals travelled between these two cities. We do not display absolute numbers, and the size of the lines are to be interpreted relative to each other because the figure is based on how many individual simulations were produced in our model. The diagram shows that those who arrived at Porto Novo likely came through major centres of trade such as Abomey, Iseyin, Idanyin, and Ketu, while a much smaller number of individuals originated from places such as Ijebu Ode or Ijemo.

Another way to visualise the results of the model are through the conditional probability plots described in Section 3.3. For each year from 1817–1836 and for each point-of-sale, we estimated the conditional probabilities of origin for an enslaved person leaving that port. Below we present eight conditional probability maps: the first set illustrate the conditional probabilities of origin from four different ports from our model for the year 1824, and the second quartet illustrate the conditional probabilities of origin for an enslaved individual leaving any coastal port in the years 1828, 1829, 1831, and 1832.

Figure 4 shows the conditional probabilities of the origins of enslaved people departing from Badagry, Porto Novo, Off Map NE, and Off Map SE panelled clockwise starting in the top left. Off Map NE can be thought of as an enslaved individual leaving Oyo and entering into the Sokoto Caliphate.

These four maps demonstrate the likely origins of enslaved people involved in Oyo's collapse on a port-by-port basis. By providing these visualisations from our model, it is immediately

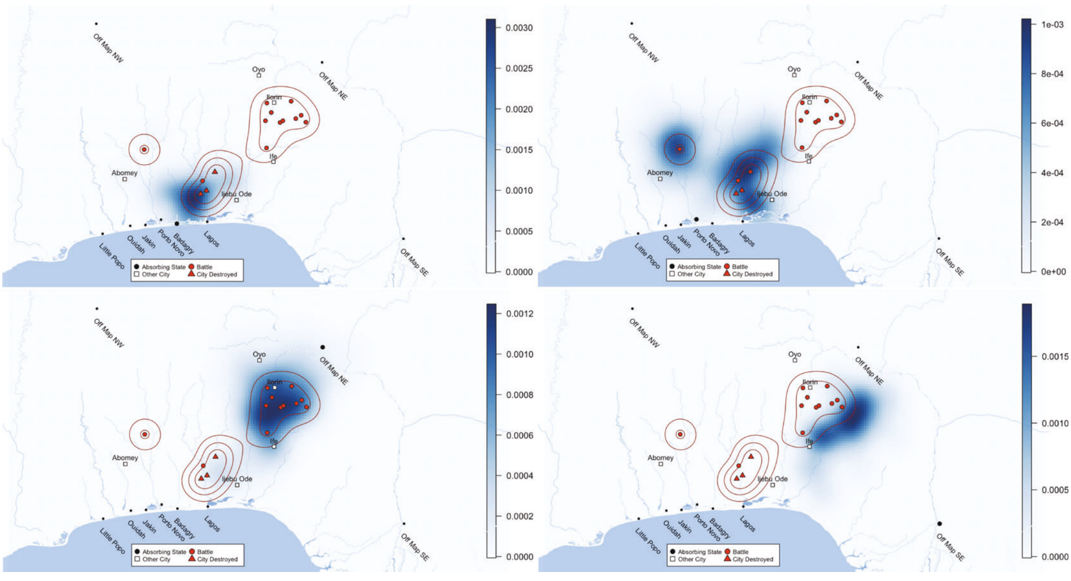

FIGURE 4 Conditional probabilities for 1824. The top left panel shows the probability of origin given the enslaved person departed Badagry. The top right shows the same for Porto Novo; bottom left is given they departed Off Map NE into the Sokoto Caliphate; bottom right for departed Off Map SE. The contours of conflict are displayed along with battles (circles) and cities destroyed (triangles)

possible for historians of the African diaspora to begin to assess from which inland cities people originated before boarding slave ships for the Americas. The red triangles represent cities being destroyed, and the red circles represent cities involved in protracted warfare. To summarise the selection of results in Figure 4, an enslaved person captured during Oyo's collapse in the year 1824 leaving through coastal ports most likely came from conflicts in the south, whereas an individual leaving via the Off Map NE or Off Map SE absorbing states most probably came from conflicts in the north.

This model output can inform more specific historical commentary. The top left panel shows that people leaving Badagry almost certainly arrived due to conflict at cities in the Egba, Awori, Egbado, and Anago regions. The top right panel illustrates the arrival of people embroiled in conflict due to an Ijebu, Ife, and Ijesha alliance along the Osun river. Much of this conflict would have also involved southern Oyo migrations due to the expansion of Ilorin in jihad as an emirate within the Sokoto Caliphate. The bottom left panel shows how people engaged in conflict also stemming from Ilorin were likely absorbed into the slave trade of the Sokoto Caliphate in the northeast. The final panel illustrates the likely areas for enslaved people being take off map to the southeast whereby small numbers of people likely went into the kingdom of Benin or into trade networks along the Niger River. Our model and application can produce similar results for every port for every year between 1817–1836.

Figure 5 illustrates another sample of results from our model, showing the conditional probabilities of origins for enslaved people departing Oyo from any coastal port for the years 1828, 1829, 1831, and 1832. These maps highlight one of the primary features of our model: the probable origins of enslaved persons leaving the coastal ports (or any absorbing state) change throughout the years based on shifting zones of conflict due to the ongoing pressure on Oyo from all fronts.

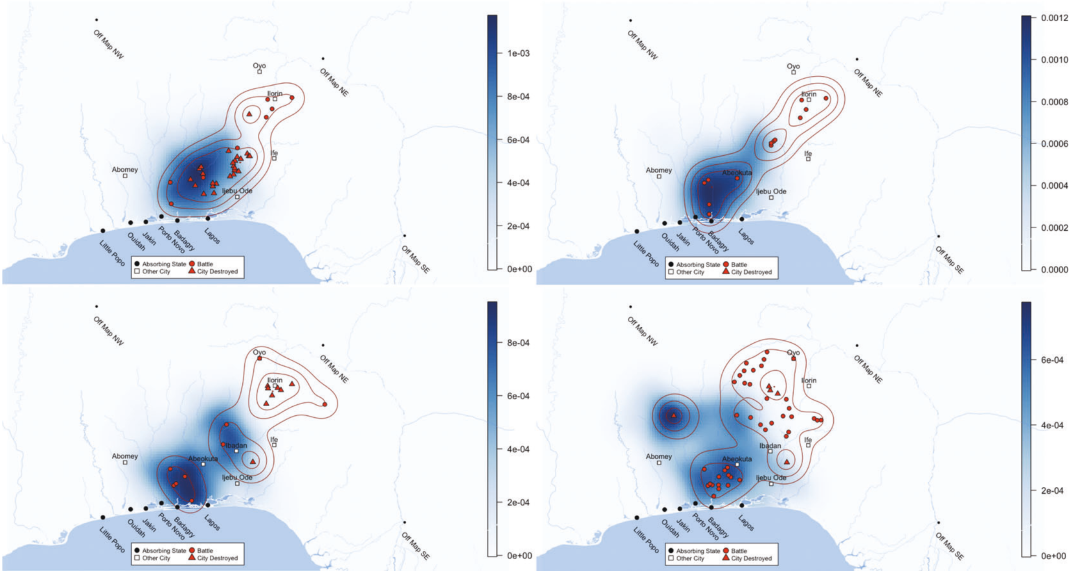

FIGURE 5 Conditional probabilities of origin of enslaved people departing Oyo from any coastal port into the transatlantic slave trade. The four panels represent conditional probabilities of origins for 1828, 1829, 1831, and 1832 clockwise starting in the top left

We can connect historical events to these shifting zones of origin probability. In 1828 (the map on the top left), the Ijebu Ife alliance continued to attack the Egba and Egbado regions, while Ilorin forced Oyo refugees further south destroying a series of cities. Meanwhile, Dahomey continued to push eastward into the Egbado, Anago, and Awori regions. The following year (top right), Ilorin slave raiding was widespread throughout Oyo territory and the city of Ibadan was founded as a refugee centre in the south. Initially, Ife armies dominated the new city, but as Oyo and Owu refugees arrived, they became strong enough to expel the Ife chiefs. In 1831 (bottom left), several key Oyo cities were abandoned in part due to the drought of 1830 and the consolidation of another refugee centre of predominately Egba and Owu at Abeokuta. Meanwhile, Ibadan—as a new city-state—began to wage war, although Abeokuta repulsed those attacks. Dahomey in the west continued its systematic campaign in the Egbado and Awori regions. In 1832 (bottom right), most of the Oyo cities came under attack by Ilorin in the north. In the south, the Owiwi war involved an unsuccessful attempt by Ijebu and support from Ibadan to disrupt trade routes to Abeokuta. In central Yorubaland, it is likely the Gbanamu or Erunmu wars occurred at this time, which resulted in the departure and registration of 477 people from Oko Oso/Ecomosho, who debarked in Lagos and arrived in Cuba on board the *Manuelita* in 1833 (Ojo, 2017). Meanwhile, Dahomey attacked from the west and destroyed the Mahi city of Kpaloko in the north. Eltis (2008) estimates that nearly 7,800 enslaved people departed these Oyo coastal ports for the Americas and Sierra Leone in 1832. Our model shows that their likely origins were mostly from the Egba, Egbado, Awori, and Anago regions, and from the southern fringes of Oyo around Ibadan. Others would have included those from Ijebu and Ife, among other groups from the forest, who were captured in conflict by their enemies. Our model indicates that a similar number of people may have been transported to the north off map into Sokoto.

4.2 | Interactive web application

To make this research more widely available to a general audience, we created an interactive web application using the `Shiny` package in the `R` programming language (Chang et al., 2019). Our web application is easy to use and freely hosted at https://walkwithweb.shinyapps.io/LA_Rshiny_app1/. It enables a general audience to interactively explore the history of the West African slave trade by visualising the data and models used in this paper. The user can select a year and one or more points-of-sale, and the application generates and displays a conditional probability map showing the most likely region of capture based on our model. The app can also display the yearly conflict data as discrete points or a contour plot of the conflict density surface. Furthermore, the trade network informing the MDP and annual approximate state borders (Lovejoy, 2019) can be overlaid. The maps produced by our app can be compared to the more detailed maps of Yoruba Diaspora (http://slaveryimages.org/), which is a cartographically based interactive digital archive currently under development by H.B. Lovejoy, Aswathanarayana, Brumfield Labs, and Chadha (Lovejoy, 2021a).

The data visualised in the application come from running our model for each year from 1817–1836. For each year, we generate 1,000 capture locations and record their spatial coordinates, the initial location in the trade network, and the point-of-sale. For any set of ports we produce an annual conditional probability surface using the methods of Section 3. Note that we do not claim that these maps display the historical truth, but rather the results from a model which provide an approximation of the truth. As historians and other researchers learn more about the West African slave trade, improved data can lead to more accurate results from our model.

5 | DISCUSSION

In this work, we developed a method for using historical data describing conflicts and trade routes in Oyo from 1817–1836 to estimate the uncertain origins of people enslaved in this region during this period. We used kriging to produce annual conflict density surfaces, from which we simulated the capture of individuals. We modelled the transport of enslaved people to ports of sale using an MDP. Then given an enslaved individual's location of departure, we created maps of that person's probability of origin. Our `Shiny` web application can produce such maps for any subset of nine departure points in the region during any of the years between 1817 and 1836. Note: interpretation of the model outputs is limited by the relatively sparse and uncertain historical data used as inputs.

5.1 | Contributions to historical literature

The first innovation of this work was to create continuous maps of conflict via kriging in pre-colonial Africa during the era of the slave trade, which may have an immediate and long-lasting impact on the field of African and African diaspora history. Historians can quickly access annual maps to visualise slave ship departures juxtaposed with the occurrence of conflict, which is the typical origin of slavery.

The second innovation of this work was to combine conflict data with historical trade routes to estimate likely routes enslaved people travelled through Oyo and maps of their likely origins. Such maps can spur future research and innovation. As an example, our work may help to classify origins of individuals in growing DNA and genealogical databases, such as 23andMe.com, Ancestry.com, or FamilySearch.org, which appear only able to provide broad regional origins (Bryc et al., 2015; Durand et al., 2014; Micheletti et al., 2020). If a person today knows that they descend from someone who was transported on a specific ship leaving the Oyo region, our methods and maps can provide a much richer understanding of their ancestral origins.

Historians can add and refine data to generate a broader and deeper understanding of how inland origins of enslaved people changed alongside the constant ebb and flow of conflict within Africa. In the case of the *Manuelita* (Ojo, 2017), our model not only visualises the inland origins of the 477 people who arrived in Cuba in 1833, but also provides a broader historical context to illustrate more instances of warfare across the wider region rather than a single isolated event of conflict.

Our model could also be used to highlight discrepancies by port between predicted embarkations based on conflict in the region and actual or estimated embarkations from historical data. Such discrepancies could be used to determine the preferences of slave traders or the prices of slaves by port and year and relate these to the factors outside Africa known to affect prices.

Our results indicate that many individuals in Oyo were enslaved due to conflicts around Ilorin and then likely transported north into the Sokoto Caliphate. P.E. Lovejoy has argued that 'the internal trade of [enslaved individuals in] West Africa appears to have been on a scale that was comparable to that of the transatlantic slave trade', which suggests that our model could be applied to re-evaluate the scale of demographic change within Africa during the precolonial period (Lovejoy, 2006, p. 104). If our model were able to withstand critique and validation against out-of-sample data for its treatment of the transport of enslaved individuals from Oyo to coastal areas, it could be applied to new data in the Sokoto Caliphate to help provide better estimates for the number of people involved in historical migrations from Oyo into the Central Sudan.

Estimating the probabilities of origin for the rest of pre-colonial Africa will require much more support in terms of extracting and collating geo-political data, which must be accomplished region-by-region, city-by-city and year-over-year. Input from African scholars and universities will be required, especially as new archaeological discoveries emerge. Ancient cities are being unearthed, many of which no longer have a place name or reflect the present-day geography. Ogundiran has, for example, identified remains of cities that are currently not located on modern-day or historical secondary source maps, but could have been places destroyed in the conflicts surrounding Oyo's collapse (Ogundiran, 2003, 2005, 2007a,b). Planning for this long process of data collection and assimilation has begun, whereby a team of historians have re-regionalised Africa with a vocabulary that expressly implements a more neutral terminology to avoid terms associated with European slave traders, colonial states, and modern-day countries, which often confuse the representation of inland Africa before 1900 (Lovejoy et al., 2019, 2021).

In this case study, the focus was on Oyo's collapse, but the methods can be applied to other contexts. Mapping conflict and origin locations of migrants in larger regions over longer periods in pre-colonial Africa will undoubtedly provide monumental results capable of making deeper connections to the complex heritage of African ancestry around the world by linking together existing and emerging open-source data sets, specifically those that have a biographic focus on enslaved Africans and their descendants. Due to increased activity in digital humanities practices, large data sets are in the ongoing process of being curated and expanded; and their themes are wide-ranging with projects centered on abolitionism, baptisms, self-liberating slaves, slave owners, slave narratives, and other land or maritime migrations beyond the Atlantic World (Eltis & Misevich, 2009; Hall, 2001; Landers, 2006; Lovejoy, 2004; Lovejoy, 2021a; forthcoming; Lovejoy et al., forthcoming). Connecting these data to our model will provide richer analysis for the intersection of people and historical sources at different moments in time. Filling the void of information surrounding pre-colonial Africa's shifting geo-political landscapes and migrations will revolutionise how African history is taught and understood, while providing a more empowered voice to the enslaved person's experience.

5.2 | Assumptions, limitations, and alternative models

The Gaussian process model is used only to convert the discrete spatial observations (sites of conflict) into a smooth spatial surface, and therefore there is no need to assume that the data generating process is Gaussian. We used kriging not in an attempt to fill in missing observations, but rather to model exposure to conflict (and therefore capture and enslavement) using a probability density function informed by discrete conflict locations. This provided flexibility in the shape of the distance decay function, enabling the model to be generalised to other historical contexts. Several alternative models could complete this task, such as representing conflict with a smoothing spline, a common variation of which is equivalent to the Gaussian process method we chose (Wahba, 1990). Other options include kernel density estimates or other radially decaying functions, which would yield mechanistically similar outcomes but may enforce unwanted symmetry. An entirely different modelling approach might treat the discrete observations as realisations of a point process model.

The MDP employed in this work represents a parsimonious approach to a migration model capturing costs and rewards possibly weighed by historical slave traders. This model is more complex than using a simple Markov chain and could be extended further using hierarchical statistical models, neural networks, or some combination of these. The MDP is constrained by the covariates available (i.e. distance and conflict intensity) to inform the probabilities of movement between cities, and could be improved to more accurately reflect history by including, among other covariates, a measure of the ruggedness of the terrain (Nunn & Puga, 2012).

We consider the methods described in this paper to be a first-order approximation of the historical process of the origins of slavery that will undoubtedly improve as we and the research community further explore the mechanisms of conflict, enslavement, and transport as they relate to the historical record. More data would enable us to tune model parameters and propagate uncertainty in the model through to the results.

5.3 | Potential extensions and future research

Future extensions of our methods could include estimating a population density surface for our region of study, and—instead of drawing a captured individual from the conflict density map—we could simulate enslavements from the product of the population surface and the conflict density surface, thereby reflecting the reality that more people would be captured from densely populated areas of conflict than from sparsely populated conflict areas. Such historical population density estimates do not currently exist, but historians could create such estimates in the future, similar to those in Manning (2010).

One of the advantages of our model is its flexibility to incorporate better sources of data. For example, as more slave ship logs are transcribed and the names categorised into linguistic groups with regional origins, more data will become available to tune the model parameters and validate it using out-of-sample data. Additionally, genetic databases and genealogical tracking have become much more powerful in recent years (Gouveia et al., 2020; Larmuseau et al., 2015;

Mathias et al., 2016; Primativo et al., 2017), and we look forward to an increase in the availability of such data. In particular, if descendants of passengers of any known ships were to compare their genome to the current genetic mapping of the Bight of Benin inland areas, we could improve the model validation. Our model could also be adapted to account for the time variance in the process of transporting captured individuals to ports-of-sale and the lack of precision in recorded conflict dates. One option would be adding positive spatio-temporal correlation from one year's conflict map to the next. Another option would be to add a time delay to each step in the MDP, allowing for recalculations of optimal routes as the conflicts shift each year. Future research could incorporate physical geographical features such as ruggedness of terrain (Nunn & Puga, 2012) to better model the travel of slave caravans.

A final source of tuning and validation would be to fit the model developed here to similar historical situations with better availability of data. Mapping continuous conflict borders from discrete city observations could be done for nearly any conventional war fought in the 19th or 20th centuries. Forced transit during the Holocaust did not originate from conventional battles as in Oyo, but have considerably better data due to its relative recency, and could be used to better tune the decision process and exit location models.

5.4 | Additional applications

Beyond the specific application of understanding the origins of enslaved individuals involuntarily involved in the transatlantic slave trade, our methods would translate well to other migrations of marginalised groups including the depopulation of indigenous groups in the Americas, the Jewish diaspora, contemporary refugees, as well as others associated with genocide and modern-day human trafficking. Specifically, our methods could help map the origins of refugees displaced during the contemporary Syrian Civil War, during which cities are known to have been attacked or destroyed, causing the migration of hundreds of thousands of individuals (Williams & Carlson, 2020). Given the known networks of roads and first-hand accounts of paths taken by Syrian migrants, an MDP could be created to model the migration of these displaced individuals into neighbouring countries. Given a specific refugee camp or border crossing, a map of probabilities of the origins of these migrants could be created to provide an estimation of from where in Syria these migrants/refugees originated. Such a model could be validated by known demographics and places of origin for a sample of refugees (Adalı & Türkyılmaz, 2020; Özden, 2013). Similarly, Olsson and Siba (2013) have compiled a list of 530 towns attacked in the Sudanese region of Darfur, and our methods could be used to better understand the resulting migrations of Darfuris. Ambrosius and Leblang (2019) argue that gang violence in El Salvador increases demand to emigrate to the United States. Our methods could be helpful to generate evidence for or against this claim.

6 | CONCLUSION

Our primary focus was to apply a novel modelling strategy for using historical data on conflicts and trade routes to estimate conditional probabilities of origin for those enslaved during the collapse of the Oyo Empire from 1817–1836. We believe that the results from our models represent the first step in an iterative process to better understand historical migrations. We hope that this paper will inspire historians and other researchers of the transatlantic slave trade and the African Diaspora to uncover new sources of data and collaborate with statisticians and data scientists to analyse these data with open-source models to gain valuable historical insight (Vance & Love, 2021). With more data—including out-of-sample data—we could more accurately tune model parameters and validate the models' performance, thus generating higher confidence in our results. Fully understanding the likely origins within Africa of enslaved people will have major ramifications for the history of the Atlantic world, whereby the ocean *connects*, rather than disconnects, Africa, the Americas, and Europe.

ACKNOWLEDGEMENTS

The authors thank the support of the Andrew W. Mellon Foundation; the National Science Foundation under Grant No. 1955109 for the project, 'IGE: Transforming the Education and Training of Interdisciplinary Data Scientists (TETRIDS)'; the United States Agency for International Development under Cooperative Agreement Number 7200AA18CA00022 for the project, 'LISA 2020: Creating Institutional Statistical Analysis and Data Science Capacity to Transform Evidence to Action'; and the Graduate School and the Center to Advance Research and Teaching in the Social Sciences at the University of Colorado Boulder. The authors also thank Vachan Daffedar Aswathanarayana and Kartikay Chadha for helping us mount the R Shiny App onto servers.

ORCID

Ashton Wiens https://orcid.org/0000-0002-7030-0602
Eric A. Vance https://orcid.org/0000-0001-5545-1878

REFERENCES

Adalı, T. & Türkyılmaz, A.S. (2020) Demographic profile of Syrians in Turkey. In: Carlson, E.D. & Williams, N.E. (Eds.) *Comparative demography of the Syrian diaspora: European and Middle Eastern destinations*, European Studies of Population, Cham: Springer International Publishing, pp. 57–91.

Ajayi, J.A. (1967) Samuel Ajayi Crowther of Oyo. In: Curtin, P.D. (Ed.) *Africa remembered: narratives by West Africans from the era of the slave trade*, Madison: University of Wisconsin Press, pp. 289–316.

Ambrosius, C. & Leblang, D.A. (2019) Immigration demand and the boomerang of deportation policies. *SSRN Scholarly Paper ID 3491522*, Social Science Research Network, Rochester, NY.

Boucherie, R.J. & Van Dijk, N.M. (2017) *Markov decision processes in practice*. Berlin: Springer.

Bruce Lockhart, J. & Lovejoy, P.E. (eds.) (2005) *Hugh Clapperton into the interior of Africa: Records of the second expedition, 1825–1827*, vol. 2. Leiden: Koninklijke Brill N. V.

Bryc, K., Durand, E., Macpherson, J., Reich, D. & Mountain, J. (2015) The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics*, 96, 37–53.

Cameron, C.M. (2016) *Captives: How stolen people changed the world*. Lincoln: University of Nebraska Press.

Chadès, I., Chapron, G., Cros, M.-J., Garcia, F. & Sabbadin, R. (2017) MDPtoolbox: Markov decision processes toolbox. Accessed 2020.

Chang, W., Cheng, J., Allaire, J.J., Xie, Y. & McPherson, J. (2019) Shiny: web application framework for R.

Cressie, N. (1985) Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*, 17, 563–586.

Curtin, P.D. (1967) Joseph wright of the Egba. In: Curtin, P.D. (Ed.) *Africa remembered: narratives by West Africans from the era of the slave trade*, Madison: University of Wisconsin Press, pp. 318–320.

Durand, E.Y., Do, C.B., Mountain, J.L. & Macpherson, J.M. (2014) Ancestry composition: a novel, efficient pipeline for ancestry deconvolution. *White paper 23-16*.

Eltis, D. (2004) The diaspora of Yoruba speakers, 1650–1865: dimensions and implications. In: Falola, T. and Childs, M. (Eds.) *The Yoruba diaspora in the Atlantic world*, Bloomington: Indiana University Press, pp. 19–39.

Eltis, D. (2008) Voyages: the trans-Atlantic slave trade database. Available from: https://www.slavevoyages.org/. Accessed 2019.

Eltis, D. & Misevich, P. (2009) African origins. Available from: http://www.african-origins.org. Accessed 2020.

Folayan, K. (1980) Trade routes in Egbado in the 19th century. In: Akinjogbin, I.A. & Osoba, S.O. (Eds.) *Topics on nigerian economic and social history*, Ife: University of Ife Press, 81–83, 100–102.

Gouveia, M.H., Borda, V., Leal, T.P., Moreira, R.G., Bergen, A.W., Kehdy, F.S.G. et al. (2020) Origins, admixture dynamics, and homogenization of the African gene pool in the Americas. *Molecular Biology and Evolution*, 37, 1647–1656.

Hall, G.M. (2001) Afro-Louisiana history and genealogy, 1719–1820. Available from: http://www.ibiblio.org/. Accessed 2020.

Hijmans, R.J., Guarino, L., Cruz, M. & Rojas, E. (2001) Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genetic Resources Newsletter*, (127), 15–19.

Irving, E.G. (1856) The Ijebu country. *Church Missionary Intelligencer*, 7, 65–72.

Jenkins, C. & Campbell, J. (1996) Catchment areas in general practice and their relation to size and quality of practice and deprivation: a descriptive study in one London borough. *BMJ: British Medical Journal*, 313, 1189–1192.

Johnson, S. (1921) *The history of the Yorubas: from the earliest times to the beginning of the British protectorate*. London: Routledge & Sons Limited.

Lander, R. (1830) *Records of Captain Clapperton's last expedition to Africa*. 2 vols. London: Colburn and Bentley.

Lander, R. & Lander, J. (1832) *Journal of an expedition to explore the course and termination of the Niger*. 2 vols. London: John Murray.

Landers, J. (2006) Slaves societies digital archive. Available from: https://www.slavesocieties.org/. Accessed 2020.

Larmuseau, M.H.D., Vessi, A., Jobling, M.A., Geystelen, A.V., Primativo, G., Biondi, G. et al. (2015) The paternal landscape along the Bight of Benin – Testing regional representativeness of West-African population samples using Y-chromosomal markers. *PLOS ONE*, 10, e0141510.

Law, R. (1977) *The Oyo Empire, c. 1600-c. 1836: a West African imperialism in the era of the Atlantic slave trade*. Oxford, UK: Oxford University Press.

Lloyd, P.C. (1967) Osifekunde of Ijebu. In: Curtin, P.D. (Ed.) *Africa remembered: Narratives by West Africans from the era of the slave trade*, Madison: University of Wisconsin Press, pp. 217–288.

Lovejoy, P.E. (2004) Studies in the history of the African diaspora-documents. Available from: https://lovejoy1943.wordpress.com/shadd-studies-studies-in-the-history-of-the-african-diaspora-documents/. Accessed 2020.

Lovejoy, P.E. (2006) *Jihad in West Africa during the age of revolutions*. Athens: Ohio University Press.

Lovejoy, P.E. (2011) *Transformations in slavery: a history of slavery in Africa*, vol. 117. Cambridge: Cambridge University Press.

Lovejoy, H.B. (2019) Mapping uncertainty: the collapse of Oyo and the trans-Atlantic slave trade, 1816–1836. *Journal of Global Slavery*, 4, 127–161.

Lovejoy, H.B. (2021a) Slavery images: a visual record of the African slave trade and slave life in the early African diaspora. Available from: http://slaveryimages.org/. Accessed 2020.

Lovejoy, H.B. (2021b) Yoruba Diaspora. Available from: http://yorubadiaspora.org

Lovejoy, P.E. (forthcoming) Freedom narratives. Available from: http://freedomnarratives.dev.matrix.msu.edu/. Accessed 2020.

Lovejoy, H.B., Lovejoy, P.E., Hawthorne, W., Alpers, E.A., Candido, M. & Hopper, M.S. (2019) Redefining African regions for linking open-source data. *History in Africa*, 46, 5–36. Available from: https://doi.org/10.1017/hia.2019.8

Lovejoy, H.B., Lovejoy, P.E., Hawthorne, W., Alpers, E.A., Candido, M., Hopper, M.S. et al. (2021) Defining regions of pre- colonial Africa: a controlled vocabulary for linking open-source data in digital history projects. *History in Africa*, 1–26. Available from: https://doi.org/10.1017/hia.2020.17

Lovejoy, P.E., Cottias, M., Glunec, J.-P.L. & Kelley, S. (forthcoming) Documenting Africans in trans-Atlantic slavery. Available from: https://documentingafricans.blogspot.com/

Manning, P. (2010) African population: Projections, 1850–1960. In: Ittmann, K., Cordell, D.D. & Maddox, G.H. (Eds.) *The demographics of Empire: the colonial order and the creation of knowledge*, chap. 10, Athens: Ohio University Press, pp. 245–276.

Manning, P. & Liu, Y. (2020) Research note on captive Atlantic flows: Estimating missing data by slave-voyage routes. *Journal of World-Systems Research*, 26, 103–125.

Manning, P., Zhang, Y. & Yi, B. (2015) Volume and direction of the Atlantic slave trade, 1650–1870: estimates by Markov chain Carlo analysis. *Journal of World-Historical Information*, 2–3(2), 127–149. https://doi.org/10.5195/jwhi.2015.31

Mathias, R.A., Taub, M.A., Gignoux, C.R., Fu, W., Musharoff, S., O'Connor, T.D. et al. (2016) A continuum of admixture in the Western Hemisphere revealed by the African diaspora genome. *Nature Communications*, 7, 12522.

Micheletti, S.J., Bryc, K., Esselmann, S.G.A., Freyman, W.A., Moreno, M.E., Poznik, G.D. et al. (2020) Genetic consequences of the transatlantic slave trade in the Americas. *Journal of Human Genetics*, 107, 1–13.

Nelson, E.J., Jones, N.L. & Miller, A.W. (1994) Algorithm for precise drainage-basin delineation. *Journal of Hydraulic Engineering*, 120, 298–312.

Nunn, N. (2008) The long-term effects of Africa's slave trades. *The Quarterly Journal of Economics*, 123, 139–176. Available from: https://doi.org/10.1162/qjec.2008.123.1.139

Nunn, N. & Puga, D. (2012) Ruggedness: the blessing of bad geography in Africa. *The Review of Economics and Statistics*, 94, 20–36. Available from: https://doi.org/10.1162/REST_a_00161

Nwokeji, G.U. & Eltis, D. (2002) Characteristics of captives leaving the Cameroons for the Americas, 1822–37. *The Journal of African History*, 43, 191–210.

Ogundiran, A. (2003) Archaeology and history in Ilare district, 1200–1900. British Archaeological Reports.

Ogundiran, A. (2005) *Precolonial nigeria*. Trenton, NJ: Africa World Press.

Ogundiran, A. (2007a) *Archaeology of atlantic Africa and the African diaspora*. Bloomington: Indiana University Press.

Ogundiran, A. (2007b) Living in the shadow of the Atlantic world: History and material life in a Yoruba-Edo hinterland, CA. 1600–1750. In: Ogundiran, A. & Falola, T. (Eds.) *Archaeology of Atlantic Africa and the African diaspora*, Bloomington: Indiana University Press, pp. 77–99.

Ojo, O. (2017) The slave ship manuelita and the story of a Yoruba community, 1833–1834. *Tempo*, 23, 360–382. Available from: https://www.scielo.br/j/tem/a/zffb6ZF7RzYZpHFcLVHLNNJ=lang=en

Olsson, O. & Siba, E. (2013) Ethnic cleansing or resource struggle in Darfur? An empirical analysis. *Journal of Development Economics*, 103, 299–312.

Özden, S. (2013) Syrian refugees in Turkey. *Tech. rep.*, MPC-Migration Policy Centre.

Primativo, G., Ottoni, C., Biondi, G., Serafino, S., Martínez-Labarga, C., Larmuseau, M.H.D. et al. (2017) Bight of Benin: a maternal perspective of four Beninese populations and their genetic implications on the American populations of African ancestry. *Annals of Human Genetics*, 81, 78–90.

Stein, M.L. (1999) *Interpolation of spatial data: some theory for kriging*. Springer Series in Statistics. New York: Springer-Verlag.

Terrell, G.R. & Scott, D.W. (1992) Variable kernel density estimation. *The Annals of Statistics*, 20, 1236–1265.

Vance, E.A., & Love, K. (2021) Building statistics and data science capacity for development. *Chance*, 34(3), 38–46. https://doi.org/10.1080/09332480.2021.1979810

Venables, W.N. & Ripley, B.D. (2002) *Modern applied statistics with S* 4th edn. New York: Springer.

Wahba, G. (1990) *Spline models for observational data*. Philadelphia: Society for Industrial and Applied Mathematics.

Webb, J.L.A. (1995) *Desert frontier: ecological and economic change along the Western Sahel, 1600–1850*. Madison, Wisconsin: University of Wisconsin Press. Google-Books-ID: 9YiEn0j5HBEC.

Williams, N.E. & Carlson, E.D. (2020) Conceptualizing the Syrian refugee crisis and migration during armed conflict. In: Carlson, E.D. & Williams, N.E. (Eds.) *Comparative demography of the Syrian diaspora: European and Middle Eastern destinations*, European Studies of Population, Cham: Springer International Publishing, pp. 3–12.

How to cite this article: Wiens, A., Lovejoy, H.B., Mullen, Z. & Vance, E.A. (2022) A modelling strategy to estimate conditional probabilities of African origins: The collapse of the Oyo Empire and the transatlantic slave trade, 1817–1836. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 1–24. Available from: https://doi.org/10.1111/rssa.12833