Cities, Conflict, and Corridors*

Shuhei Kitamura† Nils-Petter Lagerlöf‡

Abstract

In this paper, we link European state fragmentation to geography, conflict, and the locations of capitals. First we document that military battles tend to occur close to the shortest-distance corridors between the capitals of the belligerent powers, *except* where that corridor is intercepted by certain types of geography, specifically seas, mountains, and marshes. Geography thus seems to have influenced the effective military distance between the belligerents' capitals. Then we explore similar corridors between a multitude of European cities, documenting two patterns: (1) capitals tend to be closer to each other when the geography between them is more separating, as measured by similar types of geography found to affect battle locations; (2) controlling for distance, the likelihood that any two cities are located in the same state decreases with the same types of geography between them. We present a model consistent with these patterns.

Keywords: Conflict, battles, state fragmentation, cities

JEL codes: F52, H56, N43, N90

*We are grateful for helpful comments from Amrita Kulka, Nathan Nunn, and Yang Xie, as well as participants at the following conferences, workshops, and invited seminars: APEN Brisbane, EEA-ESEM Milan, GRIPS, Hitotsubashi, Keio, Kyoto, OEIO, SIOE Toronto, SWET, The BSE Summer Forum, The Tokyo Labor Economics Workshop, and UEA London. Kitamura acknowledges financial support from JSPS (18K12768, 21K132840). Lagerlöf acknowledges financial support from the Social Sciences and Humanities Research Council of Canada (Grant No 435-2019-0701).

†CiDER, Osaka University. E-mail: kitamura@cider.osaka-u.ac.jp.

‡Department of Economics, York University. E-mail: lagerlof@yorku.ca.

1 Introduction

Two of Europe's historically most powerful states, France and Britain, were fierce competitors for many centuries and usually of comparable military strength. Their capitals, Paris and London, are relatively close: about 400 km as the crow flies. So why did neither of them ever dominate and/or absorb the other? And how come they so often ended up fighting each other far beyond their own state territories?

This paper proposes a new way to understand state fragmentation in Europe. Our starting point is not the states themselves, or their borders, but rather something more temporally and spatially stable: the states' political centers of power—i.e., their capitals—and the spaces between these. Specifically, we argue that terrain which slows down military incursions makes capitals more secure, and as a consequence allows them to locate closer to each other, thus giving rise to a less unified state structure.

To concretize this argument, we consider the Great Power era in Europe and utilize a novel dataset from Kitamura (2021) on geo-coded battles. We document that Great Power battles tend to occur close to the shortest-distance corridors between the capitals of the belligerent powers (i.e., the most direct routes connecting them). However, the battles tend to deviate from the shortest-distance corridor precisely where it is intercepted by certain types of geography, namely seas, mountains, and marshes. In other words, these types of geography seem to push battles "off the corridor." Because battles should occur along the paths where the belligerents advance or retreat, our interpretation is that these features of the geography tend to extend the effective military distance between capitals at a given geodesic (direct-route) distance.[1]

This is arguably relevant for understanding the location of capitals, and state structure more generally, because it is well known from military history that state security in preindustrial Europe depended in large part on staying out of reach of foreign armies. Treivish (2016) finds that capitals tend to be farther from state boundaries compared to similarly sized non-capital cities. Bosker (2022, p.3) writes that capital cities are often located "in

[1] This interpretation is broadly consistent with a large body of work in military history on how geography shaped warfare and campaign routes; see, e.g., Engels (1978) on Alexander the Great and Collins (1998, Ch.1) for examples from modern times.

places further away from a [country's] borders or coastline that are less vulnerable from attack by foreign powers." States have also on occasion moved their capitals in response to military threats, as when the Royal Government fled from London to Oxford during the English Civil War (Toynbee, 1970, Ch.6).[2]

To illustrate the point, consider the example that we started off with: the historical Great Powers of France and England (or Great Britain). The reason their capitals could be so close, in our narrative, is that the English Channel stood between them. In military contests they often ended up battling far away from either capital (e.g., in Ireland and Spain) precisely because of the sea that separated them. By contrast, when France fought Russia the battles took place much closer to the shortest-distance corridor. The invader (mostly France in that case) could advance on land and relatively directly. It is often argued that Russia has survived threats because of its size, or "strategic depth."[3] However, one can just as well say that, absent seas and mountains, Russia's long-term survival has necessitated longer distances between its capital and those of its potential enemies.

Having documented that geography seems to impact the effective distance between capitals, we argue that this has some interesting implications for the location of capitals, and state structure more broadly. We illustrate this in a simple spatial model, where states locate their capitals to maximize security from neighboring states. Terrain varies spatially in terms of "separatedness," meaning how difficult it is to cross. The model predicts that areas with more separating terrain have more and smaller states, with capitals closer to each other. Intuitively, in a more separating environment states need less (geodesic) distance between them to achieve a given effective distance.

To test the model's predictions, we use data on European city locations from Bosker et al. (2013) and look at pairs of such cities in 1800. In our first exercise, we look at capital cities and document that pairs of capitals tend to be closer to each other (in a geodesic sense) when the geography between them is more separating, as measured by similar

[2]While we consider a strictly European context, the underlying mechanism may also relate to how rugged terrain protected African societies from slave traders (Nunn and Puga, 2012).

[3]See, e.g., Friedman (2020) and Marshall (2015, p. 13). Spolaore and Wacziarg (2016, p. 13) writes that "[g]eodesic distance [...] limits the ability to project force."

types of geography as used in the battle analysis, in particular seas and marshes (with some caveats for mountains, as discussed later). This is consistent with the model's predictions.

We also find some indirect evidence of our proposed mechanism. For example, the results do not hold for pairs of non-capital cities, presumably because military security is a concern specific to governments, while most other cities benefit from being well connected, e.g., for trade reasons. Consistent with this interpretation, we also find that the results strengthen when dropping trade hubs. The results are not driven by capitals being larger than other cities either, which could otherwise be a confounding factor: the patterns we document hold for large capitals, but not large non-capitals.

In our second exercise, we look across all city pairs in 1800 (not only capitals) and find that the likelihood that both belonged to the same state 100 years later (i.e., in 1900, which is when Europe was the most unified, as measured by its number of countries) decreases with the same types of geography along the corridor between them. This also matches the model's predictions.

We also document similar same-state correlations when using a smaller sample of cities existing already in 800 CE, (mostly) prior to modern state formation and before any of these could be described as capitals of modern states, and when measuring outcomes at various points in time later.

Finally, we illustrate these findings by mapping locations predicted to be most likely to belong to the same state as different European Great Power capitals. These show a striking similarity to the actual states at the time, with some interesting deviations that we think illustrate variation in the degree of state capacity, and viability of state territories as they appeared in 1900.

The rest of this paper is organized as follows. Section 2 discusses some of the existing literature. Section 3 sets up a model to inform our empirical exploration. Section 4 presents the data we use. We then move on to the empirical analysis. Section 5 first presents battle data to support that effective distances do seem to depend on geography. Section 6 then tests specific model predictions about how geography affects geodesic distances between pairs of capitals and the likelihood that pairs of cities belong to the same state. Section 7 concludes.

2 Existing Literature

The topics discussed here relate broadly to research on the relationship between trade, war, borders, and political unification (e.g., Alesina and Spolaore, 2003; Rohner et al., 2013; Gancia et al., 2022; Spolaore, 2023). One contribution in relation to Alesina and Spolaore (2003) is that we study the location of capitals in relation to one another, rather than their own states' borders.

We also differ by focusing on the role of geography as a deep determinant of state formation, which connects us to an older debate about the link between Europe's specific geography and high degree of state fragmentation (see, e.g., Diamond, 1997; Jones, 2003; Hoffman, 2015; Weese, 2016; Ko et al., 2018; Scheidel, 2019; Kitamura and Lagerlöf, 2020; Allen, 2023; Fernández-Villaverde et al., 2023) To test this hypothesis, earlier studies have explored the correlation between border locations and geography (Kitamura and Lagerlöf, 2020), or simulated quantifiable models of state expansion with geography as an input (Fernández-Villaverde et al., 2023). Other papers discuss the role of geography for conflict, but not for state fragmentation or the location of capitals (see, e.g., Jia, 2014; Iyigun et al., 2017; Dincecco et al., 2021).

One novelty with our empirical approach compared to all these is that we measure geography, and its effect, not where borders are located, or where battles occur, but across corridors between capitals. Our motivation is that the locations of battles in European history have not always been along state borders, or in any particular types of geography, but rather reflected the feasible paths of hostile incursions aimed at a state's centre of power, and attempts by defenders to stop these.[4]

Our "corridor" approach may have something in common with work on how spatial proximity affects interstate conflict (e.g., Gleditsch and Singer, 1975; Bremer, 1992;

[4]However, it is possible that our findings are specific to the European Great Power era, and may not hold for, e.g., the Roman Empire or Imperial China. As discussed later, they do not seem to hold in Europe after the outbreak of WWI either.

Stinnett et al., 2002). More recently, Spolaore and Wacziarg (2016) explore other distance measures (in particular genetic distances), finding that geodesic distances are negatively correlated with interstate conflict, also with various other distance controls. However, none of these papers explores where conflict occurs spatially or interacts with other measures of geography, such as seas, mountains, or marshland.

A large literature examines how geography affects the locations of modern cities, and economic activity more generally. The specific types of geography considered vary but examples include coastlines (Rappaport and Sachs, 2003; Michaels and Rauch, 2018), portage sites (Bleakley and Lin, 2012), and land productivity (Henderson et al., 2018), as well as proximate historical factors that might fundamentally depend on geography, e.g., the early emergence of statehood (Cook, 2021) and agriculture (Dickens and Lagerlöf, 2023), historical population density (Maloney and Valencia Caicedo, 2016), and transportation and trade networks (Bosker and Buringh, 2017; Barjamovic et al., 2019; Bakker et al., 2021). Different from most of these studies we explore where capitals tend to locate, and how state territories form around them, reflecting the way geography affects military security.

Dincecco and Onorato (2016) study the effect of battles on city growth, but not what determines battle locations, state territories, or the location of capitals.

3 A Model

Consider a world where locations are represented by points on a unit-length circle. To facilitate the graphical illustrations below, we shall project that circle to the unit interval, letting locations 0 and 1 be the same (i.e., where the circle closes). Locations are indexed by $x \in [0, 1]$.

We let "separatedness" at location x be denoted $g(x)$, which is assumed to be differentiable as many times as we need it to be. Empirically, a high degree of separatedness at x, i.e., a high $g(x)$, corresponds to terrain that is more difficult to cross, i.e., more mountains, sea, and/or marshland.

There are N states indexed by $i \in \{1, 2, ..., N\}$. Each state i has a capital at location

λ_i. Thus, $\lambda_i - \lambda_{i-1}$ represents the *geodesic* distance between capitals i and $i-1$ (i.e., the distance "as the crow flies"). By contrast, we let $E_{i-1,i}$ denote the *effective* distance between capitals i and $i-1$, given by

$$E_{i-1,i} = \int_{\lambda_{i-1}}^{\lambda_i} g(x)dx = G(\lambda_i) - G(\lambda_{i-1}), \tag{1}$$

where $G(x) = \int_0^x g(z)dz$ and $G'(x) = g(x)$. In other words, for a given geodesic distance between two capitals the effective distance is greater when the geography between them is more separating. Figure 1 provides an illustration.

Since the space is circular, the neighbor to the left of state 1 is state N, and, vice versa, the neighbor to the right of state N is state 1. We return to these special cases below and focus first on states $i \in \{2, ..., N-1\}$.

3.1 Optimal Locations of Capitals

Each state is assumed to locate its capital to maximize the product of the effective distances to its neighboring capitals. Although we do not model conflict explicitly, the idea is that states want to keep capitals secure from incursions by hostile neighbors. We write the objective function as:

$$\pi_i = E_{i-1,i} \times E_{i,i+1}. \tag{2}$$

We postulate that state N's capital is located at point 1, which (recall) is the same as point 0, soon verified to be optimal in equilibrium. Consider next a state $i \in \{2, ..., N-1\}$, which sets λ_i to maximize (2), subject to (1), and (1) forwarded to $E_{i,i+1} = G(\lambda_{i+1}) - G(\lambda_i)$, taking as given the locations of the neighboring states' capitals, λ_{i-1} and λ_{i+1}.

The first-order condition can be seen to imply that the effective distances are equalized: $E_{i-1,i} = E_{i,i+1}$.[5] Using (1), this can be written

$$G(\lambda_i) - G(\lambda_{i-1}) = G(\lambda_{i+1}) - G(\lambda_i). \tag{3}$$

This hints at the main mechanism in this model: where $g(x)$ is high, and $G(x)$ steep, the geodesic distance between capitals is shorter, since a given distance between λ_i and λ_{i-1}

[5] The first-order condition can be written $G'(\lambda_i)/E_{i-1,i} = G'(\lambda_i)/E_{i,i+1}$, which simplifies to $E_{i-1,i} = E_{i,i+1}$.

is associated with a greater gap between $G(\lambda_i)$ and $G(\lambda_{i-1})$. That is, a more separating geography "affords" a shorter geodesic distance between capitals.

Note also that the effective distance between state 1 and state N (and vice versa) equals $G(\lambda_1)$: state 1's leftward neighbor is state N with its capital at location 0 (same as location 1), and $G(0) = 0$. Likewise, the optimality condition for state 1—corresponding to that in (3)—becomes $G(\lambda_1) = G(\lambda_2) - G(\lambda_1)$.

3.2 Equilibrium

As mentioned, we here assume, without loss of generality, that the capital of the Nth state is located exactly where the circle closes: at location 1, which (recall) is the same location as 0. This can be shown to give

$$G(\lambda_i) = \frac{iG(1)}{N}. \tag{4}$$

(See Section A.1 in the Online Appendix for details.) It is easy to verify from (4) that $\lambda_N = 1$. Moreover, because the effective distances between all capitals equalize in equilibrium, $\lambda_N = 1$ is optimal for state N. That is, the effective distance between 0 and λ_1 is the same as that between λ_{N-1} and 1.

State 1 also locates its capital optimally, since (4) satisfies $G(\lambda_1) = G(\lambda_2) - G(\lambda_1)$. Similarly for the remaining states $i \in \{2, ..., N-1\}$, the optimality condition in (3) is implied by (4).

3.3 Simulations

With a functional form for $g(x)$ we can determine the location of each capital on the circle. For example, applying (4) to the special case where $g(x) = x$ [and $G(x) = x^2/2$] gives $\lambda_i = \sqrt{i/N}$.

With richer functional forms for $g(x)$ it is easiest to use numerical illustrations, as shown in Figures 2 to 4 [where Figure 2 shows the case where $g(x) = x$].

Panel A of each figure shows the different shapes of $g(x)$ and the equilibrium location of the capitals, assuming $N = 15$ states. Borders between the states are also indicated, here assumed to be located (geodesically) halfway between the capitals. That is, the left border of state i is halfway between λ_i and λ_{i-1} at $(\lambda_i + \lambda_{i-1})/2$; the right border of the same state is located at $(\lambda_{i+1} + \lambda_i)/2$.

The patterns shown in **Panel B** are less obvious. There we consider all different pairs of capitals [$N(N-1)/2 = 105$ pairs in this case, with $N = 15$] and explore how the geodesic distances between the capitals (as measured on the circle and thus between 0 and 1/2) correlate with the average level of separatedness, $g(x)$, between the capitals (here normalized to fall between 0 and 1). All three figures show a negative relationship. The correlation coefficient is $-.43$ in Figure 2, $-.42$ in Figure 3 and $-.65$ in Figure 4. We can sum this up as follows:

Result 1. *The geodesic distance is shorter between pairs of capitals with more separating geography between them.*

Since state borders are located between capitals, it follows that a more separating geography also results in more state fragmentation. To illustrate this, **Panel C** considers multiple pairs of locations (not only capitals), all at the same fixed distance (here set to 0.1). The bar graph shows how separatedness between the two locations in each pair differs between those pairs which are located in the same state and those which are split between different states (i.e., located on different sides of a border, possibly more than one border). Separatedness thus tends to be lower for pairs of locations in the same state, compared to those in different states. Note that this holds when considering a fixed distance between the locations. We can sum this up as follows:

Result 2. *Holding the geodesic distance between two locations constant, a more separating terrain between them makes it more likely that the two locations belong to different states.*

Results 1 and 2 follow from three different numerical simulations. For the sake of brevity we here refrain from proving them analytically, but they can be seen to be qualitatively robust to any way we choose $g(x)$ (and N). This is not surprising, since the mechanisms are so intuitive.

We could make the model much more realistic by allowing for, e.g., state heterogeneity, or an endogenous number of states. However, this need not affect any of the specific mechanisms that we are after here.[6]

4 Data

4.1 The Battle Data

Our starting point for the empirical analysis is a new battle dataset compiled by Kitamura (2021). Most of it originates from Wikidata and Wikipedia.[7] This source material changes over time, but according to Kitamura (2021) edits to the information used here (i.e., years and locations) tend to be few and minor.

The full dataset contains information about, e.g., start and end years of battles, their geo-coordinates, and lists of belligerent powers on different sides of the battle.[8] Although it covers battles throughout human history and across the world, here we focus on Europe and an era in which regular Great Power (GP) conflicts shaped its political geography. To that end, we drop all battles with geo-coordinates outside a rectangle with its northwestern and southeastern corners in Reykjavík and Baghdad, respectively. We also restrict attention to battles with a start year from 1525 up to and including 1913. The starting point coincides with the birth of Prussia, and the end point is chosen to avoid World War I battles. The Online Appendix considers the period 1914-1945 and discusses how and why the results differ for this period.

We focus on battles involving the major historical GP states in Europe. Obviously, the identities, names, regimes, and territories of these powers have changed over time. For example, one GP has been known as England, Great Britain, and the United Kingdom (of Great Britain and Ireland) at different points in history. Germany and Prussia have

[6]For example, we could let states care about their territorial size, defined as the geodesic distance between the borders of each state, but that would not change anything. With borders located geodesically half-way between capitals, the territory of state i becomes $(\lambda_{i-1} + \lambda_{i+1})/2$, which does not depend on λ_i.

[7]There are other papers using Wikidata and Wikipedia for different applications (see, e.g., Iaouenan et al., 2021, who study notable people in human history), but to the best of our knowledge Kitamura (2021) is the first to compile data on battles using this source.

[8]The dataset also contains information on outcomes of battles (who won or lost, etc.), but we do not use that information here.

intertwined histories, the latter being (a dominant) part of the former when the German Empire was created in 1871.

Here we consider the following seven GPs: England/Great Britain; France; Russia; Prussia/Germany; Austria/Habsurg Empire/Austria-Hungary; Spain; and the Ottoman Empire. These are the ones discussed in most detail in the influential study of the European Great Power system by Levy (1983).[9] The matching of battles to GPs was done manually by Kitamura (2021), who provides further details on this process.

These GPs also had relatively stable capital locations, with two exceptions: Moscow was the Russian capital before 1712 and after 1917, and 1728-1730, and St. Petersburg otherwise; Königsberg (Kaliningrad) was the capital of Prussia before 1701 and Berlin after. We return to these changes in capitals below.

We ignore those battles where the same state (by our definition) was the only belligerent involved, i.e., on both sides of the battle. This drops many (or most) civil war battles, with the exception of those where another GP was involved on one side of the battle. These are primarily battles fought in the English Civil War and during the French Revolution.

We also drop battles with locations based on rivers and valleys, because exact geo-coordinates for those battles are not reported by the sources used by Kitamura (2021).

We include naval battles in the benchmark analysis, but the results are robust to dropping these (see Section A.2 of the Online Appendix). It arguably makes sense to include naval battles, since a negative effect of sea on the likelihood of battle might otherwise seem obvious.

The seven GPs can form 21 pairs in total, but some of these fought no, or very few, battles over the period considered. In our benchmark analysis, we drop those pairs which fought fewer than ten battles, leaving eleven pairs in total.

With this adjustment, our data contains no battles involving Russia or Prussia during the years when these had Moscow and Königsberg, respectively, as capitals. In effect,

[9] Three more European states that were defined as Great Powers by Levy (1983) are ignored here, namely Sweden, Italy, and the Netherlands. However, these were not GPs over nearly as long periods of time as the other seven; see Levy (1983, Table 2.1).

we can thus treat St. Petersburg and Berlin as the capitals of Russia or Germany/Prussia in our benchmark analysis.[10] More generally, even though their territories and regimes were often fluid, all seven GPs can be thought of as having fixed political centers.

The upshot is a set of 685 battles fought between these eleven different pairs of GPs.

4.1.1 Cell Data

The battle analysis is done at the cell level, allowing us to measure battle/non-battle outcomes. We divide the rectangular area considered (with corners in Reykjavík and Baghdad) into cells of equal size, with sides of one degree latitude and longitude.

We want our results not to be based on cells in the extreme periphery of Europe, where no battles are likely to be fought. To that end, we drop all cells north of the most northerly cell in which battles took place between any of the eleven pairs, and cells south of most southerly such cell, etc. This leaves us with 1,450 cells in total. For each cell, we can measure the number of battles fought between each of the eleven GP pairs.

All in all, this gives us a dataset with $11 \times 1,450 = 15,950$ observations, where the unit of observation is the combination of a GP pair and a cell. The outcome of interest in the battle analysis is an indicator for whether a cell had any battles, or not, involving any particular GP pair. Although there is no time variation, the data structure is panel-like, in the sense that it displays variation across both cells and GP pairs; for example, a cell could record battles between England and France, but not between France and Spain, or England and Spain.

Figure 5 shows a map of the precise battle locations and which cells are coded as battle cells for at least one GP pair.

[10] Berlin would probably have been a more important power center than Königsberg ever became. The choice between Moscow and St Petersburg might be less obvious; for example, Napoleon's invasion of Russia aimed for Moscow. However, assigning Moscow as the capital of Russia does not change the main results; see Section A.2 of the Online Appendix.

4.2 Geography and Shortest-Distance Corridor

The variable that we call the *Shortest-Distance corridor* (SD corridor, for short) is an indicator for cells intersected by a 50 km buffer zone around the shortest-distance line between the relevant pair of capitals. This line takes into account the curvature of the Earth, so it does not look like a straight line on a projected map.

Different segments of a SD corridor may of course have different access to roads, ports, and rest stops. However, such factors seem endogenous and probably changed over time (and across seasons); railways began to matter later in our study period. Moreover, troops need not necessarily follow roads but could often travel across open fields or frozen waterways. Rather than using road data, we here try to measure directly some dimensions of the underlying geography that may at a deeper level have hindered troop transports.

We consider three geography variables. Marshland data are from the Global Lakes and Wetlands Database maintained by the World Wildlife Foundation (linked to here; Level 3, Categories 4 and 5). We use a relatively broad definition, including freshwater marshes, floodplain and swamp forest, and flooded forest. The binary cell-level variable is an indicator of whether a cell is intersected by anyone of those types of marshes.

To define mountains we use elevation data from NOAA National Centers for Environmental Information (linked to here). A cell is defined as having a mountain when its mean elevation exceeds 800 meters, with alternative cutoffs explored in the Online Appendix.

We define sea as the absence of land, using data from GADM. The sea indicator equals one when a cell is intersected by sea, i.e., not fully covered by land.

Figure 6 illustrates the battle cells for six GP pairs, together with the associated shortest-distance corridors, and cells where each of the three geography variables are present.

4.3 City and State Data

The city data are from Bosker et al. (2013), who provide information on multiple European cities at the turns of the centuries from 800 CE to 1800 CE. City population is reported for city-years when they exceed 5,000. The dataset also contains geo-coordinates, as well as information about which cities were capitals at different points in time.

The spatial coverage is approximately Europe and surrounding areas, such as North Africa and parts of Near East.

Our benchmark analysis in Section 6 considers cities with a population above 5,000 in 1800 CE, with some robustness checks in the Online Appendix. We choose the year 1800 because it is the latest available in Bosker et al. (2013). The unit of analysis is a pair of (capital) cities, with geography measured across buffer zones around the shortest-distance line between cities (or capitals).

The sources for the geography variables are the same as for the cell-level data (see Section 4.2 above), except that we here measure sea using Natural Earth. Different from the cell-level analysis, where we constructed binary indicators, we here use the *fraction* of the relevant buffer zone covered by mountains, marshland, and sea. This makes more sense in this context, since the corridors are so much larger geographical areas than the cells.

Data on state borders are from Euratlas (Nüssli, 2010). These contain geospatial information on the borders of sovereign states in Europe and surrounding areas at the turn of the centuries from 1 CE to 2000 CE. We use these data to determine which pairs of cities belonged to the same sovereign state. The benchmark analysis considers state borders in 1900 based on Euratlas, while the Online Appendix explores other years and data sources.

5 Battle Data Analysis

For the battle-level analysis the unit of observation is a one-degree cell. We consider cells both with and without battles, thus allowing us to use information about locations that did not see any battles. For each cell we measure if there were any battles fought there during the period of interest and involving the GPs under consideration.

More precisely, our main outcome variable is an indicator variable denoted $B_{i,p}$, taking the value one if a battle between pair p occurred in cell i over the benchmark period (1525-1913), and zero otherwise. (Section A.2 of the Online Appendix considers an intensive-margin measure as the outcome variable, i.e., the number of battles rather than a battle indicator.)

Our independent variables of interest include three geography variables, all binary indicators. $H_{800,i}$ equals one if average elevation in cell i exceeds 800 meters above the sea, and zero otherwise. (We consider different heights in Section A.2.) M_i is an indicator for a marsh (or swamp) intersecting cell i. S_i indicates whether the cell is intersected by sea.

The remaining variable of interest is the shortest-distance corridor. Like the geography variables, this is also a binary indicator, and denoted by $D_{i,p}$. Note that $D_{i,p}$ varies both across cells and GP pairs.[11]

5.1 Direct Effects

We are going to present results from a few different regression specifications. Consider first this:

$$B_{i,p} = \alpha + \beta_{\mathrm{D}} D_{i,p} + \lambda_{\mathrm{S}} S_i + \lambda_{\mathrm{H}} H_{800,i} + \lambda_{\mathrm{M}} M_i + \omega_p + \varepsilon_{i,p}, \tag{5}$$

where ω_p is a GP pair fixed effect, and $\varepsilon_{i,p}$ is an error term. If $\widehat{\beta}_{\mathrm{D}} > 0$, then battles tend to happen more often in cells along the shortest-distance corridor than elsewhere.

The first three columns of Table 1 bear this out. In column (1) we consider a specification without any geography controls or fixed effects; column (2) adds geography controls; and column (3) adds both geography controls and pair fixed effects. Throughout $\widehat{\beta}_{\mathrm{D}}$ comes out as positive and significant. We also note that all three geography measures carry negative coefficients, suggesting that battles tend to occur on land, and in terrain that is not too mountainous or marshy. However, these direct effects are hard to interpret, since geography can vary with, e.g., distance from the corridor.

We can also add cell fixed effects to the formulation in (5), absorbing the geography controls, and giving us the following specification:

$$B_{i,p} = \beta_{\mathrm{D}} D_{i,p} + \omega_p + \gamma_i + \varepsilon_{i,p}, \tag{6}$$

[11] For example, if p refers to the pair England-France, and cell i intersects with the shortest-distance corridor between London and Paris, then $D_{i,p} = 1$, while $D_{j,p} = 0$ for cells $j \neq i$ off the London-Paris corridor, and $D_{i,q} = 0$ for all GP pairs $q \neq p$, whose corridors do not cover cell i.

where γ_i capture the cell fixed effects. This is estimated in column (4) of Table 1, again showing us $\widehat{\beta}_{\mathrm{D}} > 0$.

One possibility is that the positive coefficient on the shortest distance corridor merely captures an effect from cells far away from the belligerent states, in regions where they had no reason to fight. To address this, columns (5) and (6) of Table 1 consider the same specifications as in columns (3) and (4), but restrict the sample to cells within 300 km of the shortest-distance corridor. This shrinks the sample to about 10% of its original size. While the estimated coefficient of interest shrinks in magnitude, it remains positive and significant.

Finally, column (7) of Table 1 considers the same specification as in column (4), but allows standard errors to be clustered at the pair and cell level. The corridor indicator becomes slightly less precisely estimated, but remains significant at the 5% level.

5.2 Interaction Effects

So far we have documented that GPs tend to fight more battles along their shortest-distance corridors. Next we examine if our measures of geography tend to push battles off that corridor. To that end, we estimate the following regression equation:

$$
\begin{aligned}
B_{i,p} = \;& \beta_{\mathrm{D}} D_{i,p} \\
& + \beta_{\mathrm{S}} D_{i,p} S_i \\
& + \beta_{\mathrm{H},800} D_{i,p} H_{800,i} \\
& + \beta_{\mathrm{M}} D_{i,p} M_i \\
& + \omega_p + \gamma_i + \varepsilon_{i,p},
\end{aligned}
\tag{7}
$$

where, as before, ω_p and γ_i are fixed effects for GP-pair and cell, respectively, and $\varepsilon_{i,p}$ is an error term. As earlier, we expect $\widehat{\beta}_{\mathrm{D}} > 0$. Now we should also expect $\widehat{\beta}_{\mathrm{S}} < 0$, $\widehat{\beta}_{\mathrm{H},800} < 0$, and $\widehat{\beta}_{\mathrm{M}} < 0$. As discussed above, we might expect this geography effect to be present in all cells, not only along the corridor, but any such effects are absorbed by the cell fixed effects.

In other words, we expect seas, marshes, and mountains to make the hypothesized path of military advance deviate from the shortest route. If this is the case, it suggests that these geographical characteristics increase the effective military distance between the two GPs political centers, at given geodesic distance.

Table 2 considers a few different regressions involving these interaction effects. Columns (1)-(3) show the results from three separate regressions, where the independent variables include the indicator for cells on the shortest-distance corridor, and each of the three geography variables and their interactions with the shortest-distance corridor, entered one at a time. The interaction effects all come out as negative, although not significant for marshes. Column (4) enters them all together and now the coefficients on the interaction terms become precisely estimated, all three being significantly different from zero at the 5% level, or lower. This holds also when entering GP pair fixed effects in column (5), and with both pair and cell fixed effects in column (6); note that the direct geography effects are dropped in column (6), as they are absorbed by the cell fixed effects. Column (7) uses the same fixed-effects specification as in column (6), but allows standard errors to be clustered at the pair and cell level. This renders the coefficient on marshes insignificant, but seas and mountains still come out as significant at the 5% level.

Overall, this supports the idea that these types of geography tend to push battles off the shortest-distance corridor, on which battles would otherwise tend to be fought, the result being and increase in the effective distance between the capitals.

Figure 7 illustrates how the means of the different geography variables vary between observations (cell-GP pairs) with and without battles, both for the full sample and for observations on the shortest-distance corridor between the belligerents' capitals. This shows that geography indeed differs between observations with and without battles, in particular when we consider cell/pairs on the corridor. In other words, these types of geography do push battles off the corridor.

Section A.2 of the Online Appendix examines the robustness of the results in Table 2, e.g., by adding city interactions, dropping battles close to capitals, letting Moscow be the capital of Russia (instead of St. Petersburg), dropping sea battles, using the number of battles (rather than a battle dummy) as the dependent variable, and allowing for spatially correlated standard errors. None of these changes alters the results much, at least not in ways suggesting that the correlations of interest are spurious; in some cases the results rather strengthen.

One thing that does weaken the results is measuring battle outcomes over a later period, 1914-1945. However, this finding arguably makes sense, since advances in transport technology at some point should make geography less of an obstacle for advancing armies. It is also consistent with how new modes of transport, such as railroads and steam ships, affected the spatial distribution of economic activity (see, e.g., Delventhal, 2018; Nagy, 2020; Ellingsen, 2021).

6 City Data Analysis

The analysis so far suggests that certain types of geography tend to push battles off the shortest-distance corridor between the belligerents' capitals. This supports the main assumption underlying the model in Section 3, that geography affects the effective distance between the capitals.

Next we are going to explore the specific results of the model, as summarized by Results 1 and 2. To recap, Result 1 states that the geodesic distance between capitals should be shorter when the geography between them is more separating. Result 2 states that two locations, holding constant the geodesic distance between them, are more likely to belong to different states if the geography between them is more separating.

6.1 Geodesic Distances Between Capitals

To test Result 1 we use the dataset from Bosker et al. (2013), and look at pairs of capital cities in 1800. We also add the Russian capital of St. Petersburg, to get a little closer to our battle data, but results are not sensitive to this.

We then run a few regressions where the dependent variable is the geodesic distance between the capitals, or the length of the corridor, denoted $L_{i,j}$. The three independent variables of interest correspond to those used in our earlier battle analysis: the fraction mountain (with elevation above 800 m), $H_{800,i,j}$; the fraction sea $S_{i,j}$; and the fraction marsh, $M_{i,j}$. These are all measured as fractions across a corridor's total area (the 50 km buffer zones around then shortest-distance line). The regression equation can be written:

$$L_{i,j} = \delta_{\mathrm{S}} S_{i,j} + \delta_{\mathrm{H},800} H_{800,i,j} + \delta_{\mathrm{M}} M_{i,j} + \eta_i + \eta_j + \varepsilon_{i,j}, \tag{8}$$

where η_i and η_j denote city fixed effects, one for each of the capital cities in the pair.[12] These fixed effects absorb anything that directly affects distances for any particular capital and/or its location, and follows the approach of Spolaore and Wacziarg (2006) (see also Spolaore and Wacziarg, 2009, Footnote 42).

The estimates of the different δ's should all be negative according to Result 1. Columns (1)-(4) of Table 3 present results from a number of such regressions.

As seen in columns (1) and (3) of Table 3, larger fractions sea or marshland along the corridors are associated with a shorter geodesic distances, with the estimated coefficients being negative and highly significant. This is consistent with Result 1. That is, a more separating terrain tends to push the capitals closer to each other.

The coefficient on the fraction mountain in column (2) carries the wrong sign, and also comes out as highly significant. However, when all three geography variables enter together in column (4), the coefficient on the fraction mountain shrinks in absolute magnitude and becomes less precisely estimated, while the corresponding coefficients on the fractions sea and marshland become larger in absolute terms.

The correlations are thus broadly consistent with our theory, in which each capital locates to maximize its effective distances to its neighboring capitals. At the same time, all cities would presumably benefit from being close to their trading partners, and this may apply to at least some capital cities too. (Alternatively, trade hubs may be more likely to become capitals than other cities, e.g., due to their size or administrative skills.) One way to explore this is to identify which capitals are also trade hubs, and drop these from our sample. If our hypothesis is correct, then we should expect our results to be stronger in this restricted sample. In column (5) of Table 3, we drop all pairs where both of the capitals

[12]More precisely, let $\eta_i\phi_i + \eta_j\phi_j$ be two terms in the sum $\sum_{k=1}^{N} \eta_k\phi_k$, where N is the number of capitals (or number of cities), and η_k is the coefficient on the dummy variable for capital k, denoted ϕ_k. This dummy is such that $\phi_k = 1$ if $i = k$ or $j = k$, and $\phi_k = 0$ otherwise. The two terms (and the whole sum) thus equal $\eta_i\phi_i + \eta_j\phi_j = \eta_i + \eta_j$ for capital cities i and j.

are trade hubs, defined as either a port city or a city located on a Roman road.[13] This shrinks the sample from 435 to 248 capital pairs. The estimates of the coefficients on the sea and marshland variables remain negative and significant, as in column (4), while the coefficient on the fraction mountain now comes out as negative, although not significant. This suggests that our model is more applicable when military security matters more relative to trade, as we would expect.

Rather than looking at each type of geography in isolation we can consider a composite measure that we call a *Separatedness Index*, constructed as this weighted average of the three geography variables:

$$.162 \times S_{i,j} + .166 \times H_{800,i,j} + .13 \times M_{i,j}. \tag{9}$$

The weights are given by the estimated coefficients on the interaction terms in column (5) of Table 2. These capture to what degree the presence of each type of geography tends to push battles off the shortest distance corridor, thus extending the effective distance.[14] The coefficient on this index is negative and significant in column (6) of Table 3, consistent with our theory: capitals tend to be geodesically closer to each other when they have more separating geography between them.

Columns (7) and (8) present results based on the same specifications as in (4) and (6), but with standard errors clustered on pairs of 5×5 degree cells.[15] The estimates stay significant. Finally, columns (9) and (10) drop pairs where both of the capitals are trade hubs, as in column (5), here also with clustered standard errors. The results stay significant in both specifications. In column (10), the estimated coefficient on the Separatedness Index is in fact larger than in column (8), again suggestive of our proposed mechanism.[16]

[13] For the latter, we rely on Bosker et al. (2013), and include what they call "hubs" and "non-hubs" (i.e., cities on a single Roman road or at the intersection of at least two). To define port cities, we rely on manual coding and data from Natural Earth (see Section 4.3).

[14] Since the weights in (9) are similar in size an equal-weighted average produces similar results.

[15] That is, we divide the map into cells centered on degrees latitude and longitude divisible by 5. Each pair of capitals (or cities) belongs to one unique cell pair and we cluster the standard errors on such cell pairs.

[16] However, this is somewhat sensitive to how we define trade hubs; the estimate does not change much compared to column (8) if we instead define trade hubs as cities on a crossing of at least two roman roads,

In Section A.3 of the Online Appendix, we explore various alternative specifications based on the Separatedness Index. We restrict the sample to capitals of Great Powers only (trying two different definitions), finding negative, and mostly significant, correlations. We also run regressions based on pairs of non-capital cities, here defined as pairs of cities where at least one of them is not a capital. We find that the negative effects of a more separating geography on geodesic distances pertain to pairs of capitals specifically, rather than other city pairs.

In the Online Appendix we also restrict the sample to pairs of particularly large cities, as measured by population. This could be of interest because capitals tend to be larger than other cities, suggesting size itself might drive these patterns.[17] However, we find no significant relationship between distances and separatedness for large non-capitals, while we do find it for large capitals. Our interpretation, consistent with the discussion about trade hubs above, is that large non-capital cities tend to be large because they are commercial and/or transport hubs, for which trade and connectedness matter more than security.

All in all, the results presented in this section seem consistent with Result 1 of the model.

6.2 Same-State Outcomes

To test Result 2 we again use the city data from Bosker et al. (2013), and the year 1800 CE, but consider all cities with a population above 5,000 (i.e., not only capitals). We want to know if these were more likely to belong to different states if the geography between them was more separating controlling for the geodesic distance between them.

As in the analysis of capital pairs above, we use the geo-coordinates of cities to find the shortest-distance line between city pairs and measure the same types of geography as in our earlier analysis across 50 km buffer zones from the shortest-distance line. As

or port cities, and drop these from the sample.

[17] Larger cities may be more likely to become capitals, and capitals may also grow faster than other cities. For an example of the latter, see, e.g., Kulka and Smith (2023), who finds that US cities grow faster when becoming county seats.

before, we refer to these as corridors between cities.

Also in line with the capital-pairs analysis, we focus on 1800 CE, the latest year for which Bosker et al. (2013) report data. Same-state outcomes are based on Euratlas borders of independent states in 1900, a century after we measure cities. This is when the number of states in Europe was at its lowest (see, e.g., Gancia et al., 2022, Table 1), and also a point in time when the states that we considered in our battle analysis formally existed, in particular Italy and Germany.

Excluding cities outside Euratlas state territories in 1900, this gives us 241,860 city pairs.

Let the outcome variable be an indicator denoted $C_{i,j}$, taking the value one if the two cities i and j belonged to the same state (in 1900, the year when we measure outcomes), and zero otherwise. Using the same notation as earlier for the remaining variables, we can now write the regression equation as

$$C_{i,j} = \lambda_{\mathrm{L}} L_{i,j} + \lambda_{\mathrm{S}} S_{i,j} + \lambda_{\mathrm{H},800} H_{800,i,j} + \lambda_{\mathrm{M}} M_{i,j} + \eta_i + \eta_j + \varepsilon_{i,j}, \tag{10}$$

where the terms η_i and η_j represent the same type of fixed effects as in (8), although referring to all cities (not only capitals). These absorb anything that varies at the city level.

We are interested in the estimates of the different λ's, which we all expect to carry negative signs. That is, any two cities should be less likely to belong to the same state if they are farther from each other and if they are more separated by seas, mountains, or marshes. Put another way, they should be more likely to belong to the same state if they lie close to each other, with flat, non-marshy dry land between them.

Table 4 presents least-squares estimates from various specifications similar to that in (10), letting the different geography variables enter both one by one and together.

The signs come out the expected way, and highly significant, when all geography controls enter together in column (4). The same is true when entering the fraction sea or the fraction mountain separately in columns (1) and (2). The significant and positive effect when entering the fraction marshes separately in column (3) is an anomaly, but (as mentioned) this result reverses when entering all geography variables together in column (4). We also see that the inclusion of the fraction marshes increases the size of the estimated coefficients on the other two geography variables, suggesting that these variables capture different dimensions of the separating effects of geography. Notably, the marshes variable has a negative correlation with the other two, as swampy areas tend to be located on land and at low elevation.

Column (5) uses the Separatedness Index, defined in (9), in lieu of the three geography variables. The index comes out as negative and significant at the 1% level.

Columns (6) and (7) cluster the standard errors on cell pairs (same as in Table 3), with similar results as in columns (4) and (5), except that the fraction marshes comes out as insignificant only at the 10% level in column (6).

Section A.4 of the Online Appendix makes several robustness checks of the results in Table 4. First, we explore if the results hinge on using 1900 as the outcome year, a point in time when Europe was at its most unified. We find that they are not. Letting the same-state dummy be defined on state borders in later years than 1900 the results are almost identical.

We also consider pairs of cities that existed earlier than 1800. This could be important if we believe that some cities emerged simultaneously and/or endogenously with states. However, when using pairs of cities that existed in 800 CE already—preceding modern European state formation by a few centuries, and the earliest year for which Bosker et al. (2013) have data—the results are similar to those in Table 4.

6.2.1 Heat Maps

We can use the same-state regressions to make predictions about which city locations are most likely to lie within the state territories associated with each of the Great Power capitals used in the analysis of battles earlier. We here focus on London, Paris, Madrid, Berlin, Vienna and Istanbul. Figure 8 shows different so-called heat maps, indicating which other cities are predicted to be most likely to belong to the same states as each of these respective capital cities, or what we can label connectedness. The predictions are based on the regressions with the Separatedness Index in column (5) of Table 4 (but ignoring the city fixed effects when generating the predictions).

The maps in Figure 8 show a striking resemblance between the territories of the actual states in 1900 and the lighter colors of the heat maps. Most importantly, the contours are not circular, as they would be if the geodesic distance alone was used to predict same-state outcomes. For example, the territory associated with Vienna clearly takes a non-circle shape because of the Alps.

Some exceptions are also interesting to note. For example, the part of France closest to the Dover-Calais straight has a relatively high connectedness to London (i.e., high probability of belonging to the same state as England). These are areas where England displayed some early military presence in wars against France. Similarly, southern France has a relatively low connectedness to Paris, compared to northern France, consistent with the weaker and later spread of centralized state capacity to the south, where *Langue d'Oc* (or Occitan) languages were long spoken. The areas well connected to Istanbul reach deep into the Balkans and Europe, which does not match well with the map of the Ottoman Empire in 1900, but fits much better a century earlier (see Figure A.3 in the Online Appendix).

7 Conclusion

In this paper we use data on battles and cities in Europe and its surrounding areas to gain insights about the role of geography in determining the location of capitals and state structure more generally. The focus is on the Great Power era, here defined as 1525-1913.

The conceptual starting point is that geography matters because it affects what we call the effective (military) distance between different locations, in particular between capitals. We motivate this assumption with some novel data compiled by Kitamura (2021) on battle locations. We find that battles tend to occur within a 50 km buffer zone around the shortest-distance line between the capitals of the belligerent powers, what we call a shortest-distance corridor. However, battle locations deviate from that corridor where it is intercepted by certain types of geography, specifically seas, mountains, and marshes. This result is robust to various controls, sample restrictions, and econometric specifications. Because battles may be expected to occur close to where armies advance or retreat, our interpretation is that these types of geography tend to extend the effective military distances between capitals, holding constant the geodesic distance.

To help us think about the implications of this finding, we set up a model where states locate their capitals to maximize security from neighboring states, and where the terrain varies in how difficult it is to cross, what we call separatedness. The model predicts that areas with more separating terrain have more and smaller states, with capitals closer to each other.

To test these predictions we use data on the location of capitals and other cities from Bosker et al. (2013), and examine pairs of capitals and other cities, measuring the same three types of geography along corridors between these pairs. We find the model predictions to be broadly consistent with the data. Capitals tend to be closer in a geodesic sense when separated by more seas and marshes (although the results for mountains are more mixed). Similarly, pairs of cities (capitals and others) are more likely to belong to different states when the geography between them is more separating, as measured by these three types of geography.

To illustrate this last result, we also construct maps showing the most probable state territories predicted by our same-state regressions, based on the geo-coordinates of some of Europe's Great Power capitals. The maps show striking resemblance to the actual state territories, with a couple of exceptions that we argue are interesting in their own right.

References

Alesina, A. and Spolaore, E. (2003). *The Size of Nations*. MIT Press, Cambridge, Massachusetts.

Allen, T. (2023). The topography of nations. *NBER Working Paper*.

Bakker, J. D., Maurer, S., Pischke, J.-S., and Rauch, F. (2021). Of mice and merchants: connectedness and the location of economic activity in the Iron Age. *Review of Economics and Statistics*, 103(4):652–665.

Barjamovic, G., Chaney, T., Coşar, K., and Hortaçsu, A. (2019). Trade, merchants, and the lost cities of the Bronze Age. *Quarterly Journal of Economics*, 134(3):1455–1503.

Bleakley, H. and Lin, J. (2012). Portage and path dependence. *Quarterly Journal of Economics*, 127(2):587–644.

Bosker, M. (2022). City origins. *Regional Science and Urban Economics*, 94:103677.

Bosker, M. and Buringh, E. (2017). City seeds: Geography and the origins of the European city system. *Journal of Urban Economics*, 98:139–157.

Bosker, M., Buringh, E., and van Zanden, J. L. (2013). From Baghdad to London: unraveling urban development in Europe, the Middle East, and North Africa, 800-1800. *Review of Economics and Statistics*, 95(4):1418–1437.

Bremer, S. A. (1992). Dangerous dyads: Conditions affecting the likelihood of interstate war. *Journal of Conflict Resolution*, 36(2):309–341.

Colella, F., Lalive, R., Sakalli, S. O., and Thoenig, M. (2023). acreg: Arbitrary correlation regression. *The Stata Journal*, 23(1):119–147.

Collins, J. M. (1998). *Military Geography: For Professionals and the Public*. National Defense University Press, Washington, DC.

Cook, J. C. (2021). Subnational persistence of state history: evidence from geolocalized civilizations. *Mimeo, University of California, Merced*.

Delventhal, M. J. (2018). The globe as a network: Geography and the origins of the world income distribution. *Mimeo, Claremont McKenna College*.

Diamond, J. (1997). *Guns, Germs, and Steel: The Fates of Human Societies*. W.W. Norton & Company, New York.

Dickens, A. and Lagerlöf, N. P. (2023). The long-run agglomeration effects of early agriculture in Europe. *Economic Inquiry*, 61(3):629–651.

Dincecco, M., Fenske, J., and Menon, A. (2021). The Columbian exchange and conflict in Asia. *SSRN Electronic Journal*.

Dincecco, M. and Onorato, M. G. (2016). Military conflict and the rise of urban Europe. *Journal of Economic Growth*, 21(3):259–282.

Ellingsen, S. (2021). Long-distance trade and long-term persistence. *Mimeo, Universitat Pompeu Fabra and IPEG*.

Engels, D. W. (1978). *Alexander the Great and the Logistics of the Macedonian Army*. University of California Press, Berkeley and Los Angeles, California.

Fernández-Villaverde, J., Koyama, M., Lin, Y., and Sng, T.-H. (2023). The fractured-land hypothesis. *Quarterly Journal of Economics*, 138(2):1173–1231.

Friedman, G. (2020). Russia's search for strategic depth. *Geopolitical Futures*, November.

Gancia, G., Ponzetto, G. A., and Ventura, J. (2022). Globalization and political structure. *Journal of the European Economic Association*, 20(3):1276–1310.

Gleditsch, N. P. and Singer, J. D. (1975). Distance and international war, 1816–1965. In *Proceedings of the International Peace Research Association (IPRA) Fifth General Conference*, pages 481–506. IPRA, Oslo.

Henderson, J. V., Squires, T., Storeygard, A., and Weil, D. (2018). The global distribution of economic activity: nature, history, and the role of trade. *Quarterly Journal of Economics*, 133(1):357–406.

Hoffman, P. T. (2015). *Why Did Europe Conquer the World?* Princeton University Press, Princeton.

Iaouenan, M., Eyméoud, J.-B., Gergaud, O., Palaash, B., Plique, G., and Wasmer, E. (2021). A cross-verified database of notable people, 3500BC-2018AD. *CEPR Discussion Paper No. DP15852*.

Iyigun, M., Nunn, N., and Qian, N. (2017). The long-run effects of agricultural productivity on conflict, 1400-1900. *NBER Working Paper*.

Jia, R. (2014). Weather shocks, sweet potatoes and peasant revolts in historical China. *Economic Journal*, 124(575):92–118.

Jones, E. (2003). *The European Miracle Environments, Economies and Geopolitics in the History of Europe and Asia*. Cambridge University Press, Cambridge, 3rd edition.

Kitamura, S. (2021). World historical battles database. *Mimeo, Osaka University*.

Kitamura, S. and Lagerlöf, N.-P. (2020). Geography and state fragmentation. *Journal of the European Economic Association*, 18(4):1726–1769.

Ko, C. Y., Koyama, M., and Sng, T.-h. (2018). Unified China and divided Europe. *International Economic Review*, 59(1):285–327.

Kulka, A. and Smith, C. (2023). Agglomeration over the long run: evidence from county seat wars. *Mimeo, University of Warwick*.

Levy, J. S. (1983). *War in the Modern Great Power System, 1495-1975*. University Press of Kentucky, Lexington, Kentucky.

Maloney, W. F. and Valencia Caicedo, F. (2016). The persistence of (subnational) fortune. *Economic Journal*, 126(598):2363–2401.

Marshall, T. (2015). *Prisoners of Geography: Ten Maps that Explain Everything about the World (Volume 1)*. Scribner, New York.

Michaels, G. and Rauch, F. (2018). Resetting the urban network: 117-2012. *Economic Journal*, 128(608):378–412.

Nagy, D. K. (2020). Hinterlands, city formation and growth: Evidence from the U.S. westward expansion. *Mimeo, CREi, Universitat Pompeu Fabra and Barcelona GSE*.

Nunn, N. and Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. *Review of Economics and Statistics*, 94(1):20–36.

Nüssli, C. (2010). Euratlas georeferenced vector data description. *Mimeo, Euratlas-Nüssli, Concise, Switzerland*.

Rappaport, J. and Sachs, J. D. (2003). The United States as a coastal nation. *Journal of Economic Growth*, 8(1):5–46.

Rohner, D., Thoenig, M., and Zilibotti, F. (2013). War signals: A theory of trade, trust, and conflict. *Review of Economic Studies*, 80(3):1114–1147.

Scheidel, W. (2019). *Escape from Rome: The Failure of Empire and the Road to Prosperity*. Princeton University Press, Princeton.

Spolaore, E. (2023). The Economic Approach to Political Borders. In Wilson, T. M., editor, *Border Studies: A Multidisciplinary Approach*. Edward Elgar Publishing, Cheltenham and Northampton.

Spolaore, E. and Wacziarg, R. (2006). The diffusion of development. *NBER Working Paper*, 12153.

Spolaore, E. and Wacziarg, R. (2009). The diffusion of development. *Quarterly Journal of Economics*, 124(2):469–529.

Spolaore, E. and Wacziarg, R. (2016). War and relatedness. *Review of Economics and Statistics*, 98(5):925–939.

Stinnett, D. M., Tir, J., Diehl, P. F., Schafer, P., and Gochman, C. (2002). The Correlates of War (COW) Project: Direct contiguity data, version 3.0. Technical Report 2.

Toynbee, A. (1970). *Cities on the Move*. Oxford University Press, New York and London.

Treivish, A. I. (2016). Capital cities and state borders: spatial relationships and shifts. *Regional Research of Russia*, 6(3):244–257.

Weese, E. (2016). European Political Boundaries as the Outcome of a Self-Organizing Process. *Mimeo, The University of Tokyo*.

Tables and Figures

	Dependent variable is the Battle Indicator						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Shortest-Distance Corridor Indicator	0.171^{***}	0.170^{***}	0.176^{***}	0.146^{***}	0.079^{***}	0.069^{**}	0.146^{**}
	(0.019)	(0.019)	(0.019)	(0.017)	(0.020)	(0.032)	(0.049)
Sea Indicator		-0.011^{***}	-0.011^{***}		-0.039^{**}		
		(0.003)	(0.003)		(0.019)		
Mountain Indicator (800 m)		-0.006	-0.006		-0.026		
		(0.005)	(0.004)		(0.036)		
Marsh Indicator		-0.018^{***}	-0.018^{***}		-0.007		
		(0.004)	(0.004)		(0.032)		
R^2	0.03	0.03	0.05	0.21	0.19	0.61	0.21
Number of obs.	15950	15950	15950	15950	1379	1379	15950
Fixed effects	None	None	GP-pair	GP-pair, Cell	GP-pair	GP-pair, Cell	GP-pair, Cell
Sample	Full	Full	Full	Full	Cells < 300 km of SD Corr.	Cells < 300 km of SD Corr.	Full
Standard errors	Robust	Robust	Robust	Robust	Robust	Robust	Clustered

Notes: Ordinary least squares regressions. Robust standard errors are indicated in parentheses, except for column (7), which uses two-way clustering on pairs and cells. The unit of observation is a cell/Great-Power pair combination. The Shortest-Distance (SD) Corridor is an indicator variable for whether a cell is intersected by a 50 km buffer zone around the Shortest-Distance line between the relevant pair of capitals. Columns (5) and (6) restrict the sample to cells intersected by a 300 km buffer zone around the shortest-distance line between capitals. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table 1: Battle Locations and the Shortest-Distance Corridor.

	Dependent variable is the Battle Indicator						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Shortest-Distance Corridor Indicator	0.243***	0.181***	0.176***	0.286***	0.288***	0.248***	0.248***
	(0.031)	(0.021)	(0.020)	(0.037)	(0.036)	(0.034)	(0.071)
Sea Indicator	−0.004*			−0.006**	−0.006**		
	(0.002)			(0.003)	(0.003)		
Mountain Indicator (800 m)		0.002		−0.001	−0.002		
		(0.004)		(0.004)	(0.004)		
Marsh Indicator			−0.010***	−0.013***	−0.014***		
			(0.003)	(0.004)	(0.004)		
Sea × SD-Corridor	−0.137***			−0.168***	−0.162***	−0.141***	−0.141**
	(0.038)			(0.041)	(0.040)	(0.037)	(0.060)
Mountain × SD-Corridor		−0.103**		−0.171***	−0.166***	−0.177***	−0.177**
		(0.052)		(0.054)	(0.054)	(0.045)	(0.075)
Marsh × SD-Corridor			−0.066	−0.135**	−0.130**	−0.135**	−0.135
			(0.060)	(0.065)	(0.065)	(0.053)	(0.093)
R^2	0.04	0.03	0.03	0.04	0.05	0.22	0.22
Number of obs.	15950	15950	15950	15950	15950	15950	15950
Fixed effects	None	None	None	None	GP-pair	GP-pair, Cell	GP-pair, Cell
Standard errors	Robust	Robust	Robust	Robust	Robust	Robust	Clustered

Notes: Ordinary least squares regressions. Robust standard errors are indicated in parentheses, except for column (7), which uses two-way clustering on pairs and cells. The unit of observation is a Cell/Great-Power pair combination. The Shortest-Distance (SD) Corridor is an indicator variable for whether a cell is intersected by a 50 km buffer zone around the shortest-distance line between the relevant pair of capitals. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table 2: Battle Locations and the Shortest-Distance Corridor: Interactions with Geography.

	Dependent variable is the Geodesic Distance Between Capitals									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Fraction Sea	-1.267^{***}			-1.348^{***}	-1.746^{***}		-1.348^{***}		-1.746^{***}	
	(0.208)			(0.203)	(0.204)		(0.217)		(0.218)	
Fraction Mountain		1.209^{***}		0.577^{**}	-0.108		0.577		-0.108	
		(0.254)		(0.285)	(0.324)		(0.372)		(0.447)	
Fraction Marsh			-20.745^{***}	-27.307^{***}	-20.754^{***}		-27.307^{***}		-20.754^{***}	
			(4.678)	(4.975)	(7.069)		(5.670)		(7.460)	
Separatedness Index						-4.973^{***}		-4.973^{***}		-8.294^{***}
						(1.373)		(1.590)		(1.372)
R^2	0.56	0.51	0.51	0.61	0.72	0.52	0.61	0.52	0.72	0.68
Number of obs.	435	435	435	435	248	435	435	435	248	248
Sample	Full	Full	Full	Full	Dropping trade hubs	Full	Full	Full	Dropping trade hubs	Dropping trade hubs
Standard errors	Robust	Robust	Robust	Robust	Robust	Robust	Clustered	Clustered	Clustered	Clustered

Notes: Ordinary least squares regressions across city pairs made up by those cities which were capitals in 1800 according to Bosker et al. (2013), plus St Petersburg. Robust standard errors are indicated in parentheses, except for columns (7)-(10), which cluster standard errors on cell pairs. The dependent variable is the geodesic distance between the capitals, i.e., the length of the corridor. Columns (5), (9), and (10) drop pairs of capitals where both are trade hubs, defined as port cities or cities located on at least one Roman road. All specifications include city fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table 3: The Effects of Geography on the Geodesic Distance between Capitals.

	Dependent variable is the Same-State Indicator in 1900						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Length of Corridor	−0.291***	−0.292***	−0.294***	−0.283***	−0.283***	−0.283***	−0.283***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.018)	(0.018)
Fra Sea along Corr.	−0.352***			−0.532***		−0.532***	
	(0.003)			(0.004)		(0.047)	
Fra Mountain along Corr.		−0.207***		−0.669***		−0.669***	
		(0.007)		(0.008)		(0.133)	
Fra Marsh along Corr.			1.132***	−4.270***		−4.270*	
			(0.361)	(0.392)		(2.317)	
Separatedness Index					−3.289***		−3.289***
					(0.022)		(0.302)
R^2	0.30	0.27	0.27	0.33	0.33	0.33	0.33
Number of obs.	241860	241860	241860	241860	241860	241860	241860
Standard errors	Robust	Robust	Robust	Robust	Robust	Clustered	Clustered

Notes: Ordinary least squares regressions across city pairs made up by cities which has a population above 5,000 in 1800, plus St Petersburg. Robust standard errors are indicated in parentheses, except for columns (6) and (7), which cluster on cell pairs. The dependent variable is an indicator taking the value one if the two cities belonged to the same state in 1900, and zero otherwise. All specifications include city fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table 4: Same State Outcomes across City Pairs.

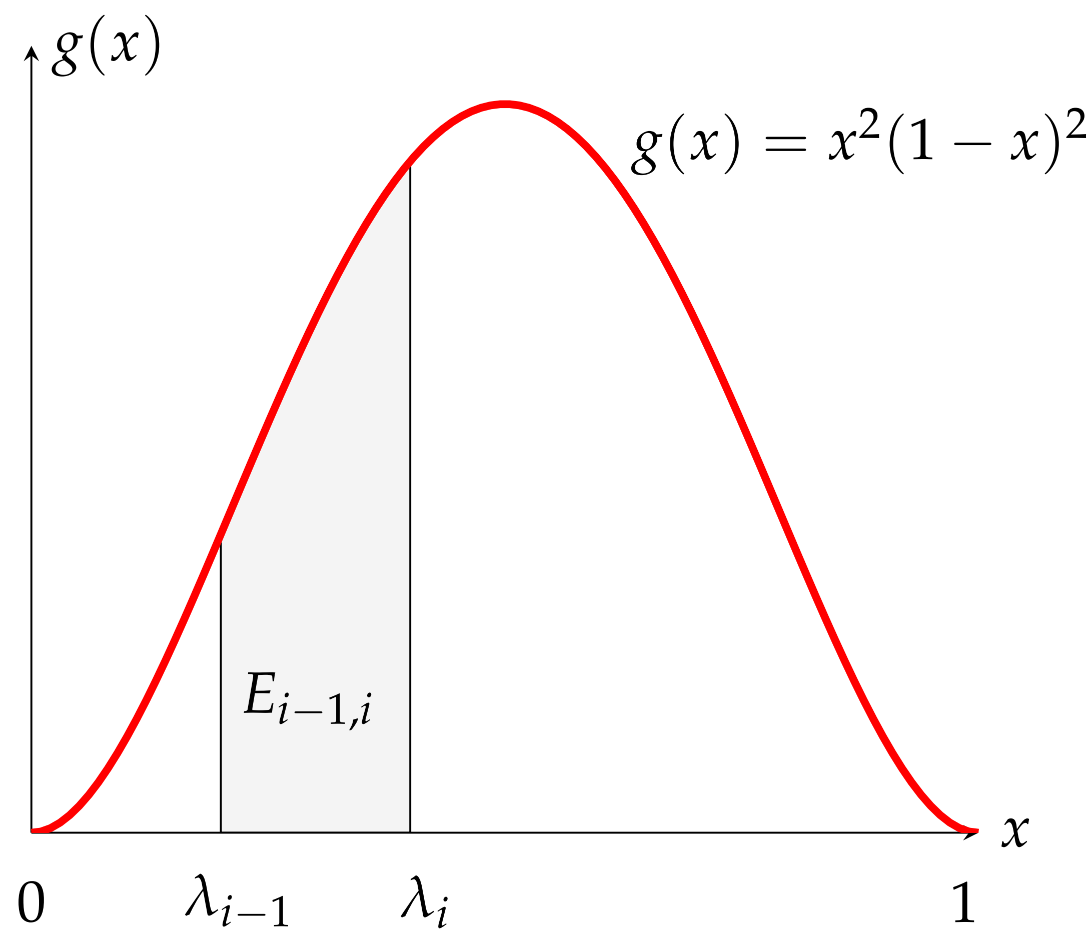


Figure 1: Illustration of geodesic distance ($\lambda_i - \lambda_{i-1}$) and effective distance ($E_{i-1,i}$) in a world with $g(x) = x^2(1-x)^2$.

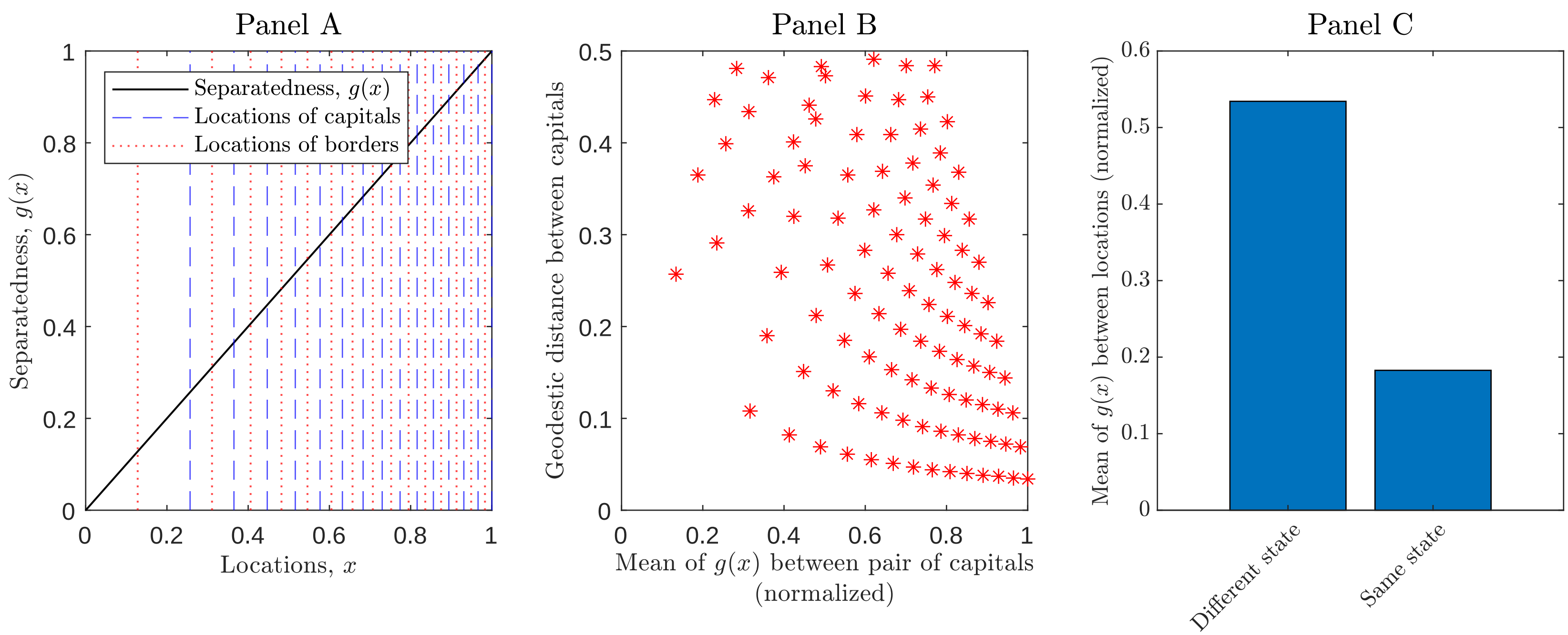


Figure 2: Model illustration with $g(x) = x$.

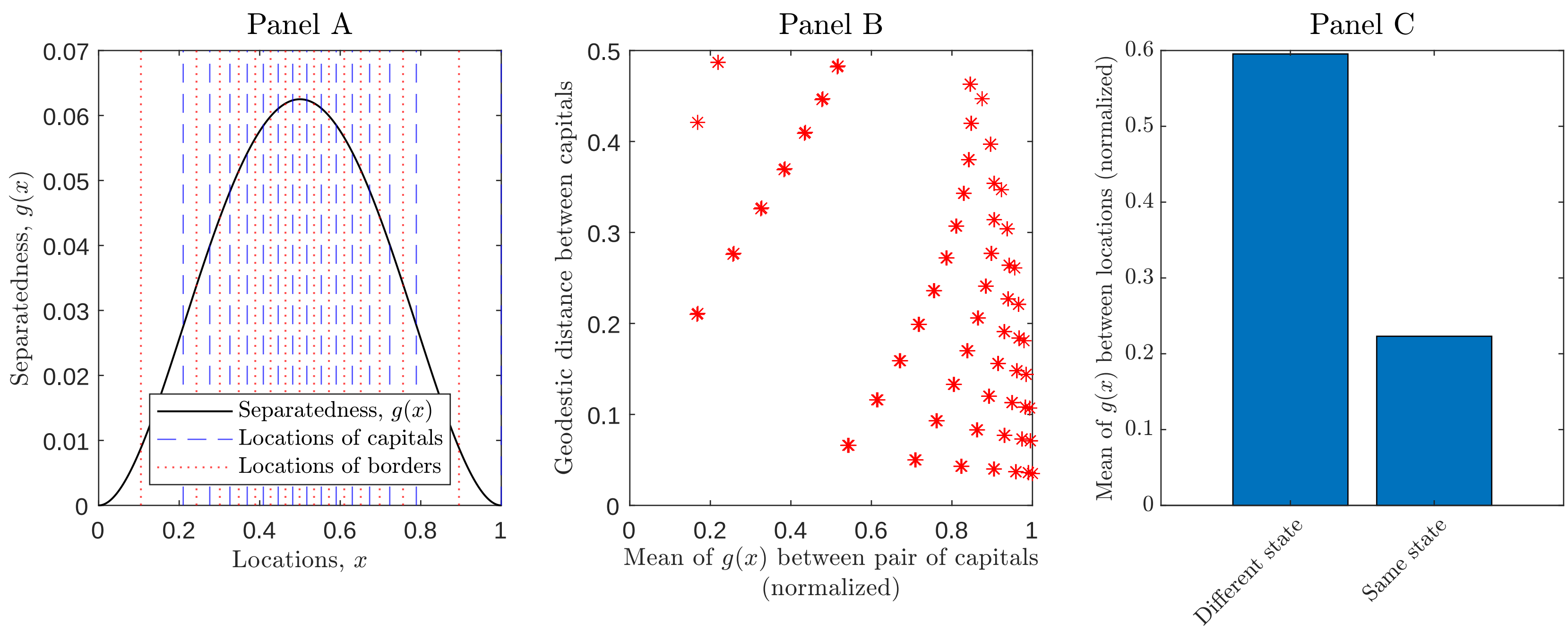


Figure 3: Model illustration with $g(x) = x^2(1-x)^2$.

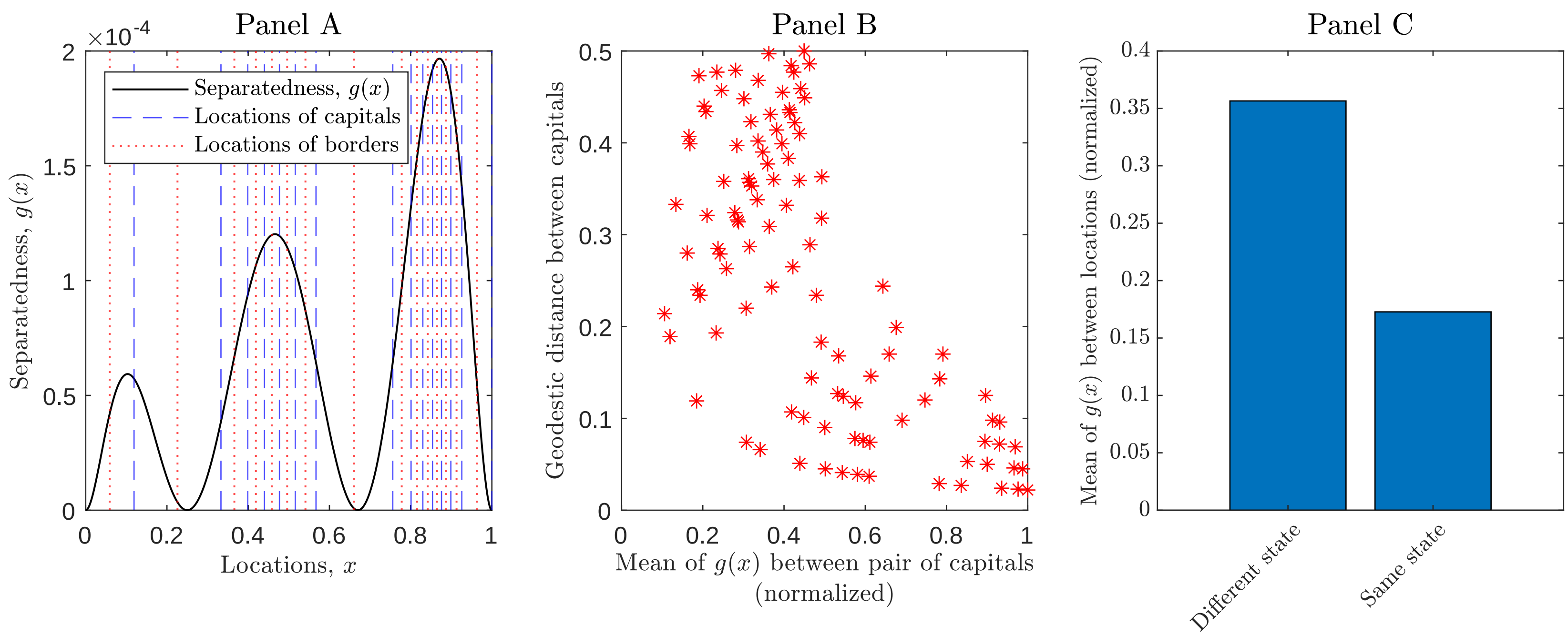


Figure 4: Model illustration with $g(x) = x^2(1-x)^2(x-1/4)^2(x-2/3)^2$.

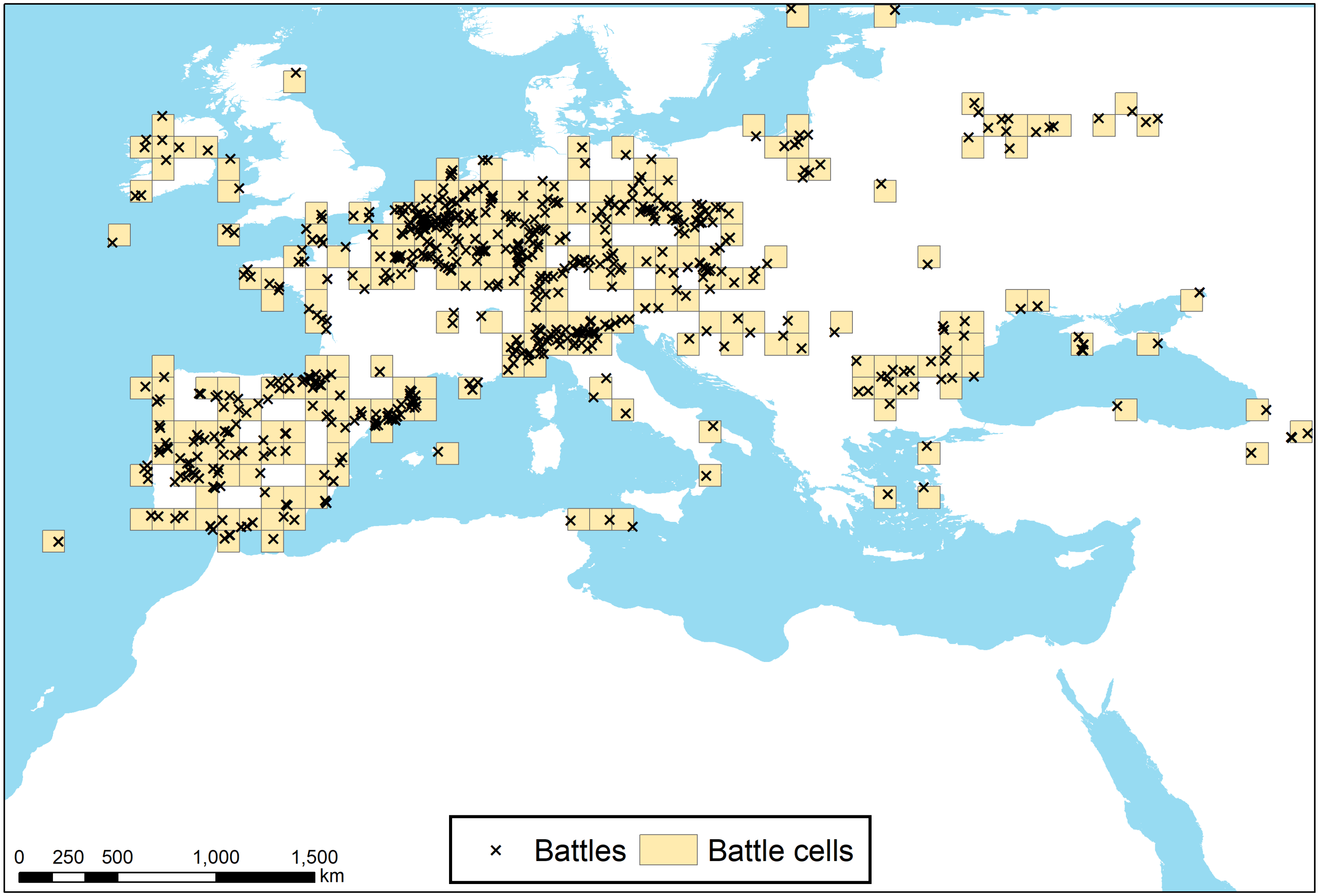
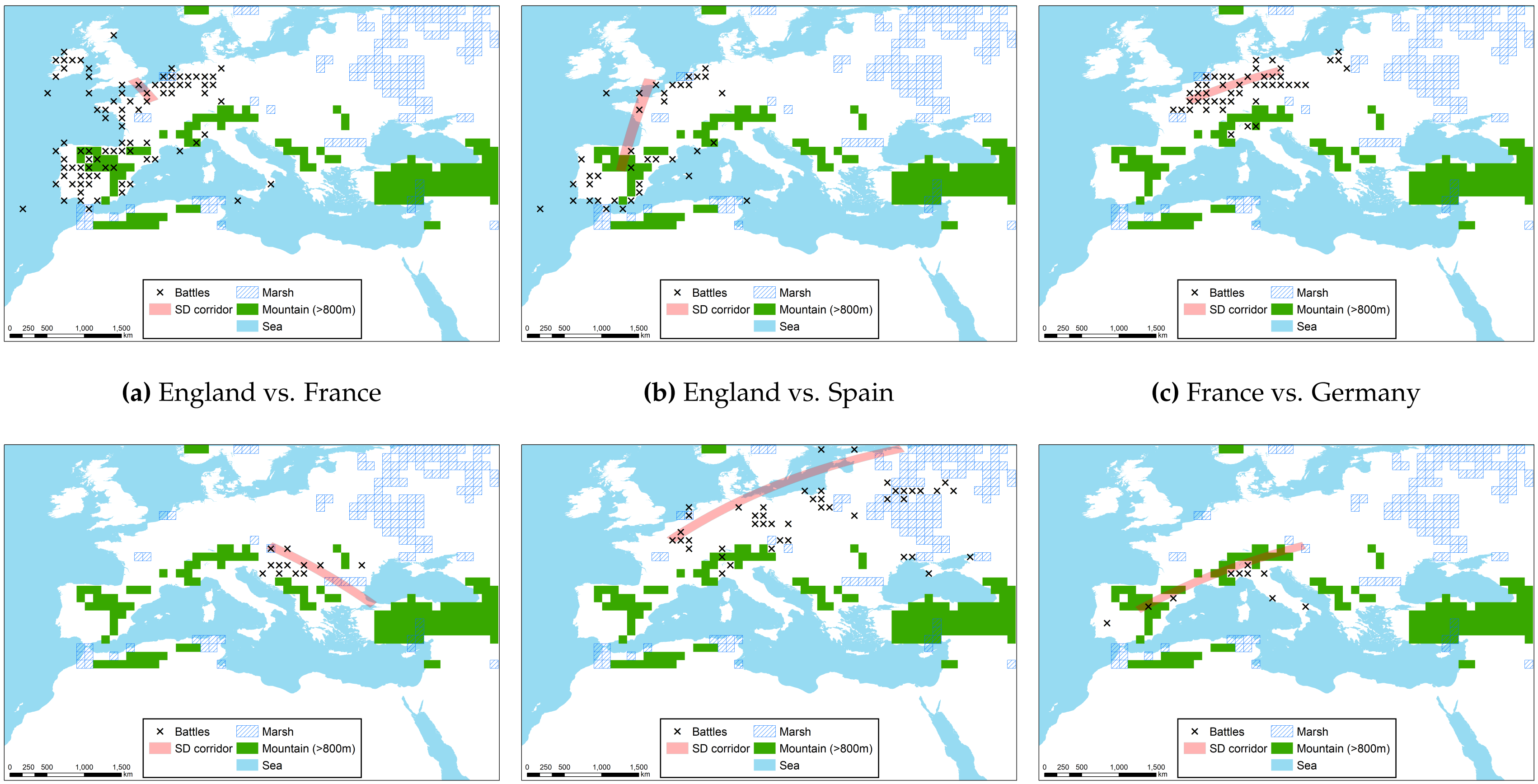


Figure 5: Map of all battles and battle cells.



(a) England vs. France

(b) England vs. Spain

(c) France vs. Germany

(d) Ottoman Empire vs. Austria

(e) France vs. Russia

(f) Austria vs. Spain

Figure 6: Maps of some Great Power battles.

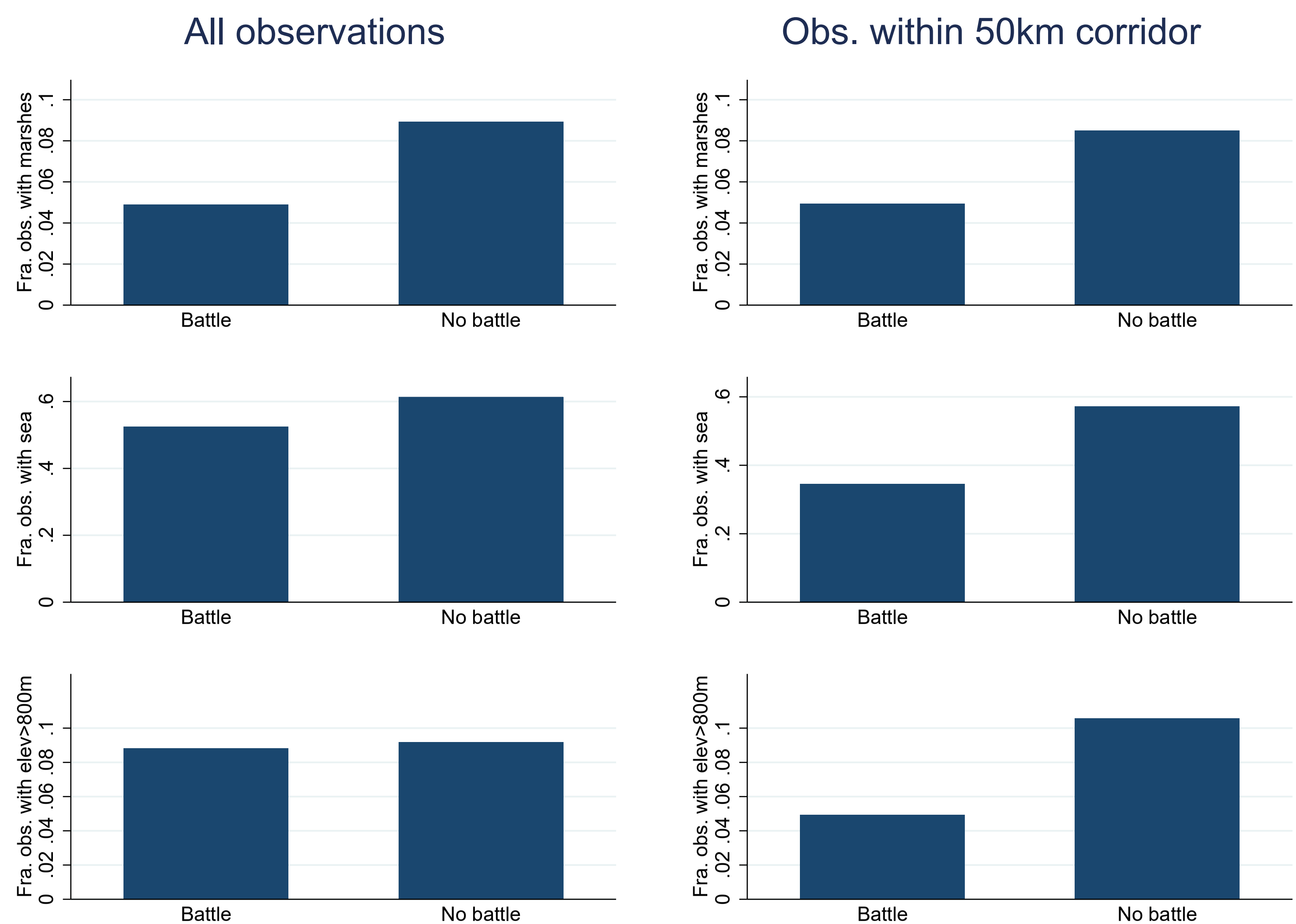
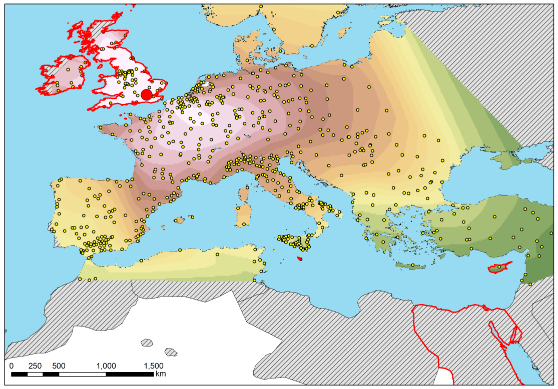
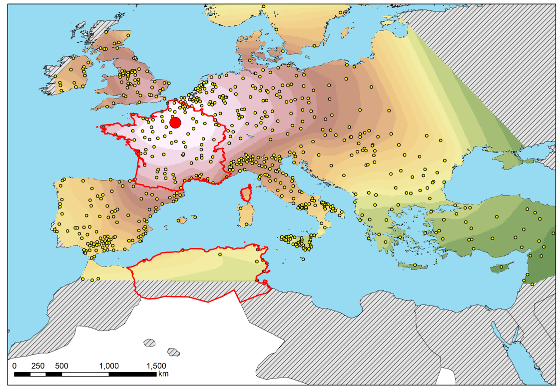


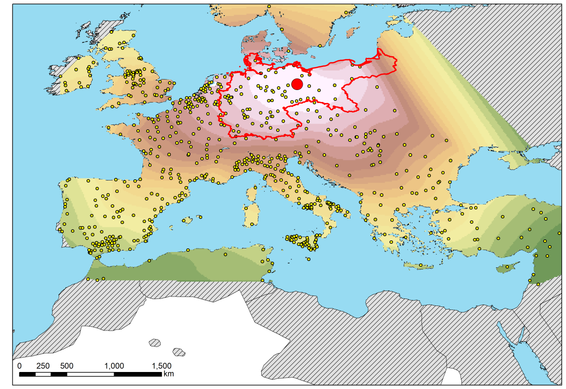
Figure 7: How geography differs between cell-pairs with and without battles, for the full sample and for cell-pairs close to the shortest-distance corridor between the belligerents' capitals.



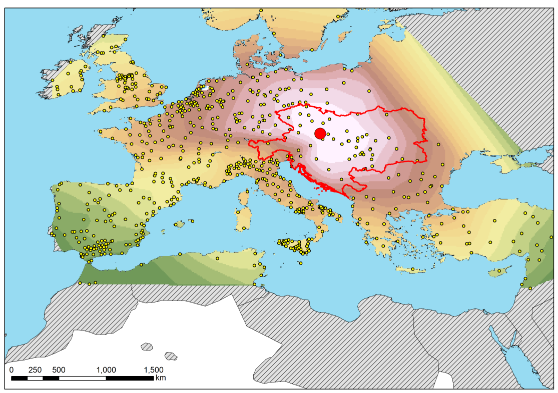
(a) Heat map for London.



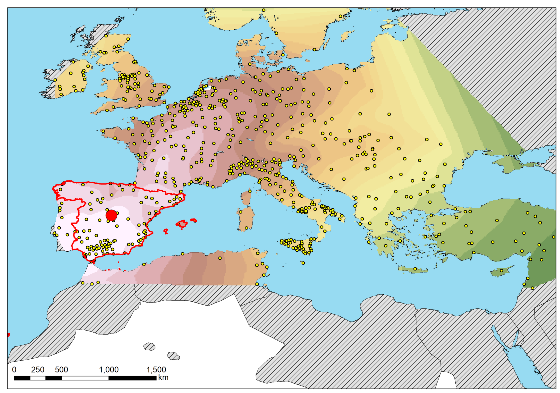
(b) Heat map for Paris.



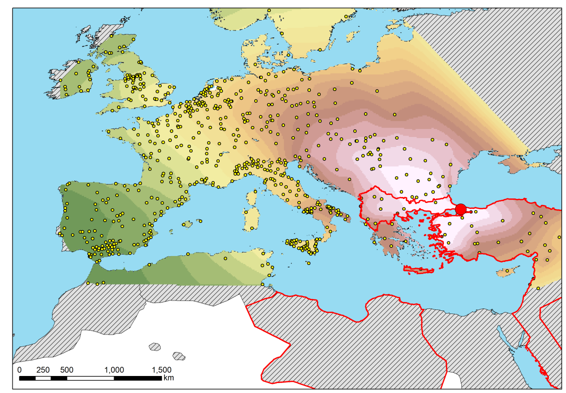
(c) Heat map for Berlin.



(d) Heat map for Vienna.



(e) Heat map for Madrid.



(f) Heat map for Istanbul.

Figure 8: Heat maps for simulated state territories for select capitals, with actual state borders in 1900 marked in red.

A Online Appendix

A.1 Equilibrium Distribution of Capitals

Define the effective distance between the capital of state i and the capital of its neighbor to the left as

$$\chi_i = G(\lambda_i) - G(\lambda_{i-1}), \tag{A.1}$$

with $\chi_1 = G(\lambda_1)$, since (recall) we postulate that state 1's neighbor to the left (state N) has its capital at 0 (same as 1). We know from (3) that χ_i equals the same constant for all states $i \in \{2, ..., N-1\}$. Call that constant $\overline{\chi}$.

For state 1, we also have $\chi_1 = G(\lambda_1) = \overline{\chi}$, since it equalizes the effective distance from its capital to both its neighbors' capitals, and the effective distance to its rightward neighbor's capital equals $G(\lambda_2) - G(\lambda_1) = \chi_2 = \overline{\chi}$. The same holds for state N. The effective distances to the capitals of its neighbors to the right and left must equalize for its capital location to be chosen optimally, so χ_N must equal the effective distance between the capitals of states 1 and N, i.e., $\chi_N = \chi_1 = \overline{\chi}$.

Thus, $\chi_i = \overline{\chi}$ for all states $i \in \{1, ..., N\}$, implying that $\sum_{i=1}^{N} \chi_i = \overline{\chi}N = G(1)$ [recalling that $G(0) = \int_0^0 g(z)dz = 0$], which gives

$$\overline{\chi} = \frac{G(1)}{N}. \tag{A.2}$$

We now see that $G(\lambda_1) = \chi_1 = \overline{\chi} = G(1)/N$. Then (A.1) says that $G(\lambda_2) = G(1)/N + G(\lambda_1) = 2G(1)/N$; $G(\lambda_3) = 3G(1)/N$; and so on, with $G(\lambda_N) = NG(1)/N = G(1)$. We can write this more succinctly as

$$G(\lambda_i) = \frac{iG(1)}{N}. \tag{A.3}$$

For example, in the case where $g(x) = x$, and $G(x) = x^2/2$, we see that $G(\lambda_i) = \lambda_i^2/2 = iG(1)/N = i/(2N)$. Disregarding the negative root (for obvious reasons) it follows that

$$\lambda_i = \sqrt{\frac{i}{N}}. \tag{A.4}$$

A.2 Battle Data Analysis: Robustness and Further Exploration

In the main analysis we defined cells with mean elevation above 800 meters as mountain cells. Table A.1 considers alternative definitions, using the same specifications as in column (6) of Table 2. The largest positive, and most significant, coefficients are found when using the 800-meter threshold. For very low levels of elevation the coefficients turn negative, which is due to cells at low elevation often having marshes, or being (fully or partially) covered by sea.

Table A.2 considers the same specifications as in Table 2, but uses only land battles when defining which cells are battle cells, dropping naval battles. This renders the negative interaction effect from sea cells more significant, for reasons that are rather obvious and not interesting. More importantly, the negative interaction effects for mountains and marshes stay robust.

The negative effect of the SD-distance corridor, and its interactions with geography, could be driven by battles happening close to the capitals between which the corridor spans. To explore this possibility, Table A.3 drops those cells that are closer than 200 km from any of the relevant capitals for each pair. The results are largely robust to this chance, with negative interaction effects throughout, slightly less significant for sea interactions and more significant for marshes, when compared to Table 2.

Not every military incursion was directly aimed at capturing the opponent's capital. The perhaps most well-known example is the French invasion of Russia in 1812. Even though the Russian capital was St. Petersburg at the time, Napoleon actually advanced towards Moscow. Table A.4 presents regressions results similar to those in Table 2 but based on a dataset where Moscow is treated as the Russian capital instead of St. Petersburg. The results change very little compared to those in Table 2.

Table A.5 presents the same regressions as in Table 2 but lets the dependent variable be the number of battles in the cell (between the relevant pair and from 1525 to 1913), rather than just a battle indicator. The results are robust to this change, and in fact strengthen for marshes in column (7).

Table A.6 allows for spatially adjusted standard errors and declining weights, applying the acreg command in Stata and the Bartlett option from Colella et al. (2023). The specifications are the same as in column (6) of Table 2, changing the distance cut-off within which standard errors are allowed to be correlated. The results are broadly consistent with the benchmark results, with slightly weaker results for marsh interactions, similar to when using two-way clustering in column (7) of Table 2.

One concern is that geography simply captures an effect of urbanization. For example, battles might not happen where the SD-corridor intersects mountains or marshes because those areas are uninhabited, which can make it hard to feed and service troops. To explore this, Table A.7 adds an interaction with cities along the SD-corridor to the specification in column (6) of Table 2. The variable we call City (or City Indicator) is equal to one for cells having a city with population above 5,000 in the year indicated for each column of Table A.7. Population data come from Bosker et al. (2013). For all years, there is a positive interaction effect between city presence and the corridor, meaning battles are more likely to happen on the SD-corridor where cities are located, i.e., in more populated areas. More importantly, the interaction effects with our three geography variables are robust to the inclusion of these city interactions.

The benchmark analysis is focused on battles fought between 1525 and 1913. Table A.8 runs the same regressions as in Table 2, but uses battles taking place 1914-1945, i.e., during the two world wars. This renders the interaction effects for seas and marshes insignificant, and weakens the interactions for mountains. One possibility is that advances in transport technology from the early 20th century started to make geography less of an obstacle for advancing armies.

A.3 Geodesic Distances: Robustness and Further Exploration

Table A.9 shows some variations on the regression in column (6) of Table 3, using the Separatedness Index.

Column (1) considers a much larger sample consisting of pairs of non-capital cities (241,425 pairs in total, i.e., the same sample as in Table 4), while column (2) replicates (6) of Table 3, using only capital city pairs. As mentioned earlier, we here define non-capital pairs as those where at least one city is not a capital, by the relevant definition, but the results are very similar if we use a sample where both cities in the pair are non-capitals.

As we have already discussed, the negative coefficient in column (2) implies that a more separating terrain is associated with pairs of capitals being located closer to each other. Interestingly, the relationship is the opposite when we consider non-capital cities. In other words, the separating effects of geography seems to be specific to capitals. Figure A.1 shows binscatter plots illustrating the different relationships in columns (1) and (2) of Table A.9; note that both the geodesic distance and the index are measured net of city fixed effects.

Our analysis so far has been based on capitals as defined by Bosker et al. (2013). There are of course different definitions of what constitutes a capital (and/or a sovereign state). It stands to reason that the mechanisms that we are after might easiest be found among states that are in regular conflict with each other, such as the Great Power nations of Europe. Column (4) considers the same Great Power capitals that we used in our battle analysis, here called a narrow definition of Great Powers, and column (3) adds Stockholm and Amsterdam, what we call a broad definition. These samples are much smaller than that made up by all capitals as defined by Bosker et al. (2013), but as seen in columns (3) and (4) we still find a negative relationship, although insignificant for the narrow definition. See Figure A.2 for the associated plots, where both the geodesic distance and the index are reported as residuals net of city fixed effects.

Columns (5)-(7) present the same regressions as in columns (2)-(4), but with standard errors clustered on cell pairs, similar to Table 3. The results are very similar.

Table A.10 analyses results for large cities, again looking at pairs of capitals and non-capitals separately. We restrict the samples to pairs where both cities have populations above the 50th, 75th, 90th, and 95th percentiles. Both population data and the defintion of capitals is from Bosker et al. (2013). [Note that columns (1) and (3) are identical, because the set of capital city pairs is the same when restricting populations to be above median as when restricting them to be in the 75th percentile.]

Throughout in Table A.10, we find that the negative relationship between distances and separatedness is negative and mostly significant for pairs of capitals, while insignificant and carrying inconsistent sign for non-capitals [The coefficient estimate is borderline significant at the 10% level in column (7).] This shows that the patterns we describe for capitals is likely not caused by their size, but rather something else that makes them unique. Since large non-capital cities are likely to be commercial centers, these patterns seem broadly consistent with the idea that seperatedness matters more for security, and connectedness more for trade. That is, being out of reach by one's enemy is important for cities that are centers of government, while for other large cities trade and communication might matter more.

A.4 Same-State Outcomes: Robustness and Further Exploration

This section considers variations on the regressions in Table 4. Table A.11 presents results with the same-state indicator measured in 2000 based on Euratlas data (same source as in Table 4), and Table A.12 shows the results when using modern country borders from the Global Administrative Boundaries (GADM) database (version 3.6, the most recent at the time these data were extracted). The results when using these modern state borders are qualitatively very similar to those in Table 4, which were based on 1900 borders.

The location of cities with populations above 5,000 may well be endogenous to how state territories form. As yet another complementary exercise, Table A.13 thus considers similar gravity regressions as those in Table 4, but here across pairs formed only by cities present in 800 CE and in the year for which we measure same-state outcomes, which we let vary from 800 CE to 1800 CE. Here all specifications include city fixed effects and standard errors are clustered on cell pairs.

While shrinking the sample considerably, dropping cities that emerged after 800 CE should mitigate some of these endogeneity concerns, since centralized statehood did not exist (or was at least not widespread) in Europe by then. The coefficient estimates in Table A.13 come out with roughly the expected negative signs: not all estimates are highly significant, but those that are carry the right (negative) sign.

Table A.14 presents results from the same regressions, but using the Separatedness Index instead of the three geography variables separately, which facilitates interpretation. The pattern is similar to Table A.13, with the most significant negative estimates around 1300-1500, and slightly less precise after 1600.

Online Appendix Tables and Figures

	Dependent variable is the Battle Indicator					
	(1)	(2)	(3)	(4)	(5)	(6)
Shortest-Distance Corridor Indicator	0.170***	0.201***	0.229***	0.248***	0.228***	0.220***
	(0.015)	(0.014)	(0.013)	(0.013)	(0.012)	(0.012)
Sea × SD-Corridor	−0.091***	−0.109***	−0.126***	−0.141***	−0.124***	−0.118***
	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Mountain × SD-Corridor	0.096***	0.052***	−0.039*	−0.177***	−0.120***	−0.136**
	(0.018)	(0.020)	(0.023)	(0.028)	(0.035)	(0.064)
Marsh × SD-Corridor	−0.084***	−0.099***	−0.119***	−0.135***	−0.119***	−0.112***
	(0.030)	(0.030)	(0.030)	(0.030)	(0.029)	(0.029)
R^2	0.22	0.22	0.22	0.22	0.22	0.22
Number of obs.	15950	15950	15950	15950	15950	15950
Threshold for Mountain Indicator	200 m	400 m	600 m	800 m	1000 m	1500 m

Notes: The same ordinary least squares regressions as in column (6) of Table 2 but using different thresholds for what defines a mountain. All specifications include fixed effects for cells and Great Power pairs. Column (4) of this table replicates column (6) of Table 2. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.1: Interactions with Geography: Different Height Thresholds for Mountain Indicator.

	Dependent variable is the Land Battle Indicator						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Shortest-Distance Corridor Indicator	0.659***	0.464***	0.450***	0.807***	0.809***	0.700***	0.700**
	(0.113)	(0.070)	(0.069)	(0.140)	(0.139)	(0.125)	(0.246)
Sea Indicator	−0.010			−0.016**	−0.017**		
	(0.006)			(0.007)	(0.007)		
Mountain Indicator (800 m)		0.002		−0.005	−0.006		
		(0.010)		(0.012)	(0.011)		
Marsh Indicator			−0.026***	−0.034***	−0.035***		
			(0.005)	(0.007)	(0.007)		
Sea × SD-Corridor	−0.433***			−0.540***	−0.524***	−0.482***	−0.482**
	(0.130)			(0.145)	(0.143)	(0.130)	(0.210)
Mountain × SD-Corridor		−0.355***		−0.578***	−0.566***	−0.533***	−0.533**
		(0.103)		(0.136)	(0.134)	(0.123)	(0.210)
Marsh × SD-Corridor			−0.253*	−0.477***	−0.464***	−0.443***	−0.443*
			(0.131)	(0.158)	(0.156)	(0.130)	(0.214)
R^2	0.03	0.03	0.03	0.04	0.05	0.22	0.22
Number of obs.	15950	15950	15950	15950	15950	15950	15950
Fixed effects	None	None	None	None	GP-pair	GP-pair, Cell	GP-pair, Cell
Standard errors	Robust	Robust	Robust	Robust	Robust	Robust	Clustered

Notes: The same ordinary least squares regressions as in Table 2 but using a Land Battle Indicator as dependent variable. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.2: Interactions with Geography: Only Land Battles.

	Dependent variable is the Battle Indicator						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Shortest-Distance Corridor Indicator	0.175***	0.133***	0.136***	0.225***	0.229***	0.190***	0.190**
	(0.036)	(0.022)	(0.022)	(0.043)	(0.043)	(0.040)	(0.068)
Sea Indicator	−0.003			−0.005*	−0.005*		
	(0.002)			(0.003)	(0.003)		
Mountain Indicator (800 m)		0.001		−0.001	−0.001		
		(0.004)		(0.004)	(0.004)		
Marsh Indicator			−0.009***	−0.011***	−0.011***		
			(0.003)	(0.004)	(0.004)		
Sea × SD-Corridor	−0.089**			−0.126***	−0.124***	−0.097**	−0.097
	(0.043)			(0.047)	(0.047)	(0.044)	(0.059)
Mountain × SD-Corridor		−0.108**		−0.148***	−0.141***	−0.146***	−0.146**
		(0.050)		(0.051)	(0.051)	(0.051)	(0.061)
Marsh × SD-Corridor			−0.148***	−0.220***	−0.219***	−0.203***	−0.203***
			(0.023)	(0.039)	(0.039)	(0.039)	(0.051)
R^2	0.01	0.01	0.01	0.02	0.03	0.19	0.19
Number of obs.	15626	15626	15626	15626	15626	15626	15626
Fixed effects	None	None	None	None	GP-pair	GP-pair, Cell	GP-pair, Cell
Standard errors	Robust	Robust	Robust	Robust	Robust	Robust	Clustered

Notes: The same ordinary least squares regressions as in Table 2 but dropping those observations that are closer than 200 km from anyone of the relevant capitals for each pair. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.3: Interactions with Geography: Dropping Observations Close to Capitals.

	Dependent variable is the Battle Indicator						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Shortest-Distance Corridor Indicator	0.230***	0.197***	0.193***	0.262***	0.265***	0.239***	0.239***
	(0.028)	(0.021)	(0.021)	(0.031)	(0.031)	(0.029)	(0.055)
Sea Indicator	−0.003			−0.005*	−0.005*		
	(0.002)			(0.003)	(0.003)		
Mountain Indicator (800 m)		0.002		−0.000	−0.000		
		(0.004)		(0.004)	(0.004)		
Marsh Indicator			−0.011***	−0.013***	−0.013***		
			(0.003)	(0.004)	(0.004)		
Sea × SD-Corridor	−0.104***			−0.121***	−0.115***	−0.120***	−0.120**
	(0.038)			(0.038)	(0.038)	(0.035)	(0.046)
Mountain × SD-Corridor		−0.119**		−0.158***	−0.154***	−0.173***	−0.173**
		(0.052)		(0.053)	(0.052)	(0.044)	(0.068)
Marsh × SD-Corridor			−0.079	−0.114**	−0.108*	−0.115**	−0.115
			(0.056)	(0.058)	(0.057)	(0.049)	(0.064)
R^2	0.04	0.04	0.04	0.04	0.06	0.22	0.22
Number of obs.	15950	15950	15950	15950	15950	15950	15950
Fixed effects	None	None	None	None	GP-pair	GP-pair, Cell	GP-pair, Cell
Standard errors	Robust	Robust	Robust	Robust	Robust	Robust	Clustered

Notes: The same ordinary least squares regressions as in Table 2 but based on a dataset where Moscow is treated as the Russian capital instead of St. Petersburg. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.4: Interactions with Geography: Moscow as Capital.

	Dependent variable is Number of Battles						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Shortest-Distance Corridor Indicator	0.654***	0.473***	0.460***	0.803***	0.807***	0.696***	0.696**
	(0.113)	(0.070)	(0.069)	(0.140)	(0.139)	(0.125)	(0.243)
Sea Indicator	−0.005			−0.011	−0.011		
	(0.006)			(0.007)	(0.007)		
Mountain Indicator (800 m)		−0.003		−0.008	−0.009		
		(0.010)		(0.011)	(0.011)		
Marsh Indicator			−0.025***	−0.030***	−0.030***		
			(0.007)	(0.008)	(0.008)		
Sea × SD-Corridor	−0.407***			−0.515***	−0.499***	−0.455***	−0.455*
	(0.131)			(0.145)	(0.144)	(0.130)	(0.211)
Mountain × SD-Corridor		−0.363***		−0.578***	−0.567***	−0.532***	−0.532**
		(0.104)		(0.135)	(0.134)	(0.122)	(0.207)
Marsh × SD-Corridor			−0.267**	−0.485***	−0.471***	−0.443***	−0.443*
			(0.131)	(0.158)	(0.156)	(0.130)	(0.212)
R^2	0.03	0.03	0.03	0.04	0.05	0.21	0.21
Number of obs.	15950	15950	15950	15950	15950	15950	15950
Fixed effects	None	None	None	None	GP-pair	GP-pair, Cell	GP-pair, Cell
Standard errors	Robust	Robust	Robust	Robust	Robust	Robust	Clustered

Notes: The same ordinary least squares regressions as in Table 2, but using the *number* of battles (in each cell and between the relevant pair) as the dependent variable. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.5: Interactions with Geography: Number of Battles as Dependent Variable.

	Dependent variable is the Battle Indicator					
	(1)	(2)	(3)	(4)	(5)	(6)
Shortest-Distance Corridor Indicator	0.248***	0.248***	0.248***	0.248***	0.248***	0.248***
	(0.034)	(0.043)	(0.050)	(0.053)	(0.056)	(0.058)
Sea × SD-Corridor	−0.141***	−0.141***	−0.141***	−0.141***	−0.141***	−0.141***
	(0.037)	(0.042)	(0.043)	(0.045)	(0.049)	(0.051)
Mountain × SD-Corridor	−0.177***	−0.177***	−0.177***	−0.177***	−0.177***	−0.177***
	(0.045)	(0.054)	(0.047)	(0.045)	(0.043)	(0.045)
Marsh × SD-Corridor	−0.135**	−0.135*	−0.135	−0.135	−0.135*	−0.135*
	(0.053)	(0.080)	(0.090)	(0.085)	(0.076)	(0.069)
R^2	0.22	0.22	0.22	0.22	0.22	0.22
Number of obs.	15950	15950	15950	15950	15950	15950
Distance cut-off	0 km	200 km	400 km	600 km	800 km	1000 km

Notes: Ordinary least squares regressions with spatially adjusted standard errors and declining weights (the Bartlett option in acreg). All specifications include Cell and Great Power pair fixed effects. Column (1) replicates the results in Table 2, column (6), with robust standard errors now adjusted for spatial correlation. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.6: Interactions with Geography: Spatially Adjusted Standard Errors.

	Dependent variable is the Battle Indicator					
	(1)	(2)	(3)	(4)	(5)	(6)
Shortest-Distance Corridor Indicator	0.200***	0.220***	0.217***	0.205***	0.203***	0.179***
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.036)
Sea × SD-Corridor	−0.119***	−0.128***	−0.127***	−0.124***	−0.124***	−0.115***
	(0.038)	(0.038)	(0.038)	(0.038)	(0.037)	(0.038)
Mountain × SD-Corridor	−0.163***	−0.169***	−0.175***	−0.172***	−0.175***	−0.162***
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
Marsh × SD-Corridor	−0.110**	−0.120**	−0.119**	−0.112**	−0.121**	−0.110**
	(0.052)	(0.052)	(0.052)	(0.052)	(0.051)	(0.052)
City × SD-Corridor	0.116***	0.081*	0.079*	0.103**	0.103***	0.116***
	(0.044)	(0.048)	(0.044)	(0.042)	(0.040)	(0.036)
R^2	0.22	0.22	0.22	0.22	0.22	0.22
Number of obs.	15950	15950	15950	15950	15950	15950
Year in which city presence measured	1300	1400	1500	1600	1700	1800

Notes: Ordinary least squares regressions with robust standard errors. The variable City × SD-Corridor is an interaction term defined as the product of a City Indicator and the SD-corridor Indicator. The City Indicator equals one for cells having a city with population above 5,000 in the year indicated. All specifications include both Cell and Great Power pair fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.7: Interactions with Geography: City Presence.

	Dependent variable is the Battle Indicator						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Shortest-Distance Corridor Indicator	0.077***	0.087***	0.090***	0.085***	0.085***	0.078***	0.078*
	(0.022)	(0.017)	(0.018)	(0.026)	(0.026)	(0.025)	(0.037)
Sea Indicator	−0.003			−0.002	−0.002		
	(0.003)			(0.003)	(0.003)		
Mountain Indicator (800 m)		−0.000		−0.001	−0.001		
		(0.004)		(0.005)	(0.005)		
Marsh Indicator			0.004	0.003	0.003		
			(0.005)	(0.005)	(0.005)		
Sea × SD-Corridor	0.018			0.013	0.017	−0.003	−0.003
	(0.034)			(0.036)	(0.036)	(0.035)	(0.037)
Mountain × SD-Corridor		−0.105***		−0.106***	−0.089***	−0.092*	−0.092*
		(0.018)		(0.022)	(0.021)	(0.047)	(0.043)
Marsh × SD-Corridor			−0.038	−0.037	−0.033	−0.037	−0.037
			(0.045)	(0.048)	(0.047)	(0.049)	(0.060)
R^2	0.01	0.01	0.01	0.01	0.02	0.22	0.22
Number of obs.	10290	10290	10290	10290	10290	10290	10290
Fixed effects	None	None	None	None	GP-pair	GP-pair, Cell	GP-pair, Cell
Standard errors	Robust	Robust	Robust	Robust	Robust	Robust	Clustered

Notes: The same ordinary least squares regressions as in Table 2, but using battles over the period 1914-1945, instead of 1525-1913. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.8: Interactions with Geography: Battles 1914-1945.

	Dependent variable is the Geodesic Distance Between Capitals or Non-Capitals						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Separatedness Index	1.738***	−4.973***	−8.459***	−2.480	−4.973***	−8.459***	−2.480
	(0.067)	(1.373)	(2.452)	(4.717)	(1.590)	(2.528)	(4.670)
R^2	0.46	0.52	0.72	0.70	0.52	0.72	0.70
Number of obs.	241425	435	45	21	435	45	21
Capitals/non-capitals	Non-capitals	Capitals	Capitals	Capitals	Capitals	Capitals	Capitals
Definition of capitals	Bosker	Bosker	Great Pwrs (broad)	Great Pwrs (narrow)	Bosker	Great Pwrs (broad)	Great Pwrs (narrow)
Standard errors	Robust	Robust	Robust	Robust	Clustered	Clustered	Clustered

Notes: Some variations on the regressions presented in Table 3, based on the Seperatedness Index. The samples labelled capitals and non-capitals refer to city pairs where both are capitals, or at least one is not a capital, respectively, according to the definition indicated. Robust standard errors reported in parentheses, except for columns (5)-(7), where we cluster standard errors on cell pairs. All specifications include city fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.9: Geodesic Distances: Variations Using the Separatedness Index.

	Dependent variable is the Geodesic Distance Between Capitals or Non-Capitals							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Separatedness Index	−4.068**	−0.663	−4.068**	−0.570	−4.439**	−0.003	−3.814	−1.839
	(1.677)	(1.075)	(1.677)	(1.024)	(1.943)	(1.467)	(2.315)	(1.671)
R^2	0.56	0.49	0.56	0.56	0.55	0.60	0.66	0.59
Number of obs.	378	45678	378	16827	276	2352	153	442
Capitals/non-capitals	Capitals	Non-capitals	Capitals	Non-capitals	Capitals	Non-capitals	Capitals	Non-capitals
Percentile pop. size	50%	50%	75%	75%	90%	90%	95%	95%

Notes: Regressions similar to those in Table 3, but considering capitals and non-capitals, and restricting the samples to observations where both cities in the pair have populations above the median (50th percentile), and in the 75th, 90th, and 95th percentiles of the full sample; the associated population levels (in thousands) are 15, 20, 40, and 76, respectively. The population data are from Bosker et al. (2013). The capitals/non-capitals samples are explained in Table A.9, here based on the definition in Bosker et al. (2013). Standard errors, reported in parentheses, are clustered on cell pairs throughout. All specifications also include city fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.10: Geodesic Distances Between Capitals and Non-Capitals: Large Cities.

	Dependent variable is the Same-State Indicator in 2000						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Length of Corridor	-0.255^{***}	-0.256^{***}	-0.258^{***}	-0.247^{***}	-0.248^{***}	-0.247^{***}	-0.248^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.018)	(0.018)
Fra Sea along Corr.	-0.336^{***}			-0.515^{***}		-0.515^{***}	
	(0.003)			(0.004)		(0.049)	
Fra Mountain along Corr.		-0.220^{***}		-0.667^{***}		-0.667^{***}	
		(0.007)		(0.008)		(0.132)	
Fra Marsh along Corr.			1.443^{***}	-3.855^{***}		-3.855^{*}	
			(0.336)	(0.366)		(2.210)	
Separatedness Index					-3.194^{***}		-3.194^{***}
					(0.022)		(0.310)
R^2	0.28	0.24	0.24	0.31	0.31	0.31	0.31
Number of obs.	241860	241860	241860	241860	241860	241860	241860
Standard errors	Robust	Robust	Robust	Robust	Robust	Clustered	Clustered

Notes: The same regressions as in Table 4, except that the dependent variable is an indicator for the pair of cities belonging to the same state in 2000, based on maps from Euratlas. All specifications include city fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.11: Same State Outcomes: Euratlas Borders for 2000.

	Dependent variable is the Same-State Indicator based on the most recent GADM borders						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Length of Corridor	-0.253^{***}	-0.254^{***}	-0.256^{***}	-0.245^{***}	-0.246^{***}	-0.245^{***}	-0.246^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.018)	(0.018)
Fra Sea along Corr.	-0.333^{***}			-0.510^{***}		-0.510^{***}	
	(0.003)			(0.004)		(0.049)	
Fra Mountain along Corr.		-0.219^{***}		-0.661^{***}		-0.661^{***}	
		(0.007)		(0.008)		(0.132)	
Fra Marsh along Corr.			1.653^{***}	-3.594^{***}		-3.594	
			(0.336)	(0.365)		(2.189)	
Separatedness Index					-3.167^{***}		-3.167^{***}
					(0.022)		(0.309)
R^2	0.27	0.24	0.24	0.31	0.31	0.31	0.31
Number of obs.	241860	241860	241860	241860	241860	241860	241860
Standard errors	Robust	Robust	Robust	Robust	Robust	Clustered	Clustered

Notes: The same regressions as in Table 4, except that the dependent variable is an indicator for the pair of cities belonging to the same state in modern times, based on the most recent borders from the Global Administrative Areas (GADM). All specifications include city fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.12: Same State Outcomes: State Borders from GADM.

	The dependent variable is the Same-State Indicator for different years									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Length of Corridor	−0.181***	−0.267***	−0.197***	−0.190***	−0.121***	−0.149***	−0.180***	−0.278***	−0.264***	−0.375***
	(0.032)	(0.029)	(0.027)	(0.025)	(0.018)	(0.028)	(0.032)	(0.025)	(0.026)	(0.030)
Fra Sea along Corr.	−0.263*	−0.176	−0.008	0.132	−0.254***	−0.240**	−0.304**	−0.045	0.068	0.095
	(0.141)	(0.153)	(0.130)	(0.129)	(0.094)	(0.115)	(0.136)	(0.159)	(0.182)	(0.152)
Fra Mountain along Corr.	−0.594***	−0.227	−0.696***	−0.336	−0.891***	−0.639***	−0.732***	−0.438**	−0.421*	−0.583**
	(0.221)	(0.227)	(0.209)	(0.228)	(0.140)	(0.207)	(0.232)	(0.201)	(0.232)	(0.240)
Fra Marsh along Corr.	−9.984*	1.606	−9.745*	−8.001*	−11.427***	−3.360	−4.762	6.269	8.127	0.063
	(5.513)	(6.408)	(5.005)	(4.333)	(3.662)	(6.864)	(6.945)	(7.283)	(7.092)	(6.513)
R^2	0.38	0.41	0.41	0.34	0.40	0.35	0.37	0.47	0.43	0.65
Number of obs.	946	1035	946	990	990	780	861	741	630	780
Year in which state borders are measured	900	1000	1100	1200	1300	1400	1500	1600	1700	1800

Notes: Ordinary least squares regressions with standard errors in parentheses, clustered on 5 × 5-degree cell pairs. The unit of observation is a pair of cities that existed both in 800 CE and in the year in which outcomes are measured. The dependent variable is an indicator for whether the two cities belonged to the same state in the years indicated. All specifications include city fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.13: Regressions across City Pairs by 800 CE: Same-State Outcomes by Century.

	The dependent variable is the Same-State Indicator for different years									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Length of Corridor	-0.171^{***}	-0.266^{***}	-0.180^{***}	-0.179^{***}	-0.110^{***}	-0.144^{***}	-0.175^{***}	-0.273^{***}	-0.260^{***}	-0.369^{***}
	(0.035)	(0.029)	(0.032)	(0.029)	(0.023)	(0.031)	(0.036)	(0.022)	(0.025)	(0.030)
Separatedness Index	-2.074^{**}	-1.163	-1.060	0.191	-2.424^{***}	-1.900^{**}	-2.351^{**}	-0.777	-0.065	-0.043
	(0.916)	(0.945)	(0.938)	(0.933)	(0.690)	(0.805)	(0.961)	(0.932)	(1.117)	(0.961)
R^2	0.36	0.41	0.35	0.31	0.32	0.32	0.35	0.46	0.41	0.62
Number of obs.	946	1035	946	990	990	780	861	741	630	780
Year in which state borders are measured	900	1000	1100	1200	1300	1400	1500	1600	1700	1800

Notes: The same ordinary least squares regressions as in Table A.13 but with the Separatedness Index as the independent variable of interest, instead of entering the three geography variables separately. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.14: City Pairs by 800 CE and Later Same-State Outcomes: Seperatedness Index Results.

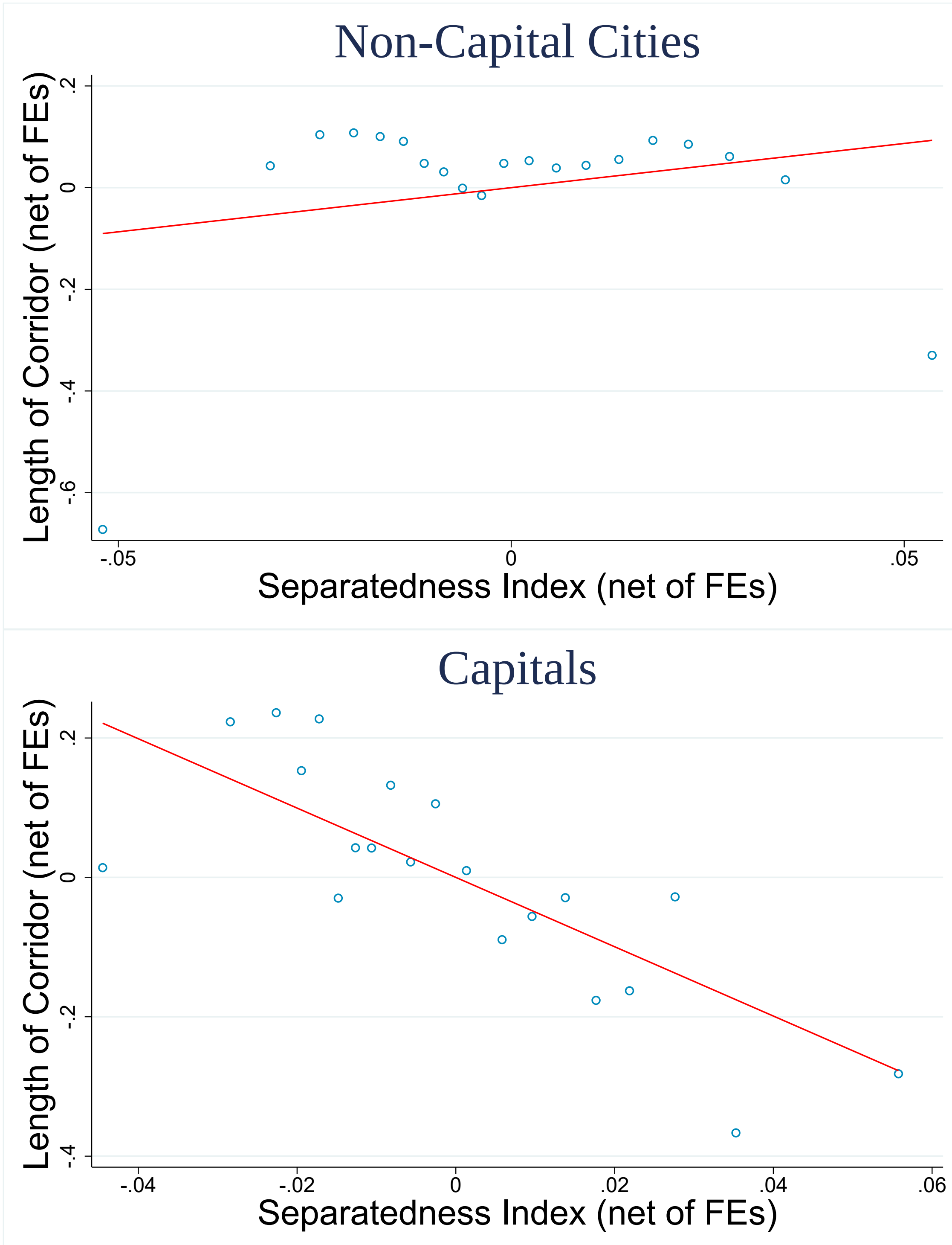


Figure A.1: Binscatter plots contrasting the different relationships between geodesic distance and separatedness for non-capital cities and capitals, based on the definition of capitals from Bosker et al. (2013).

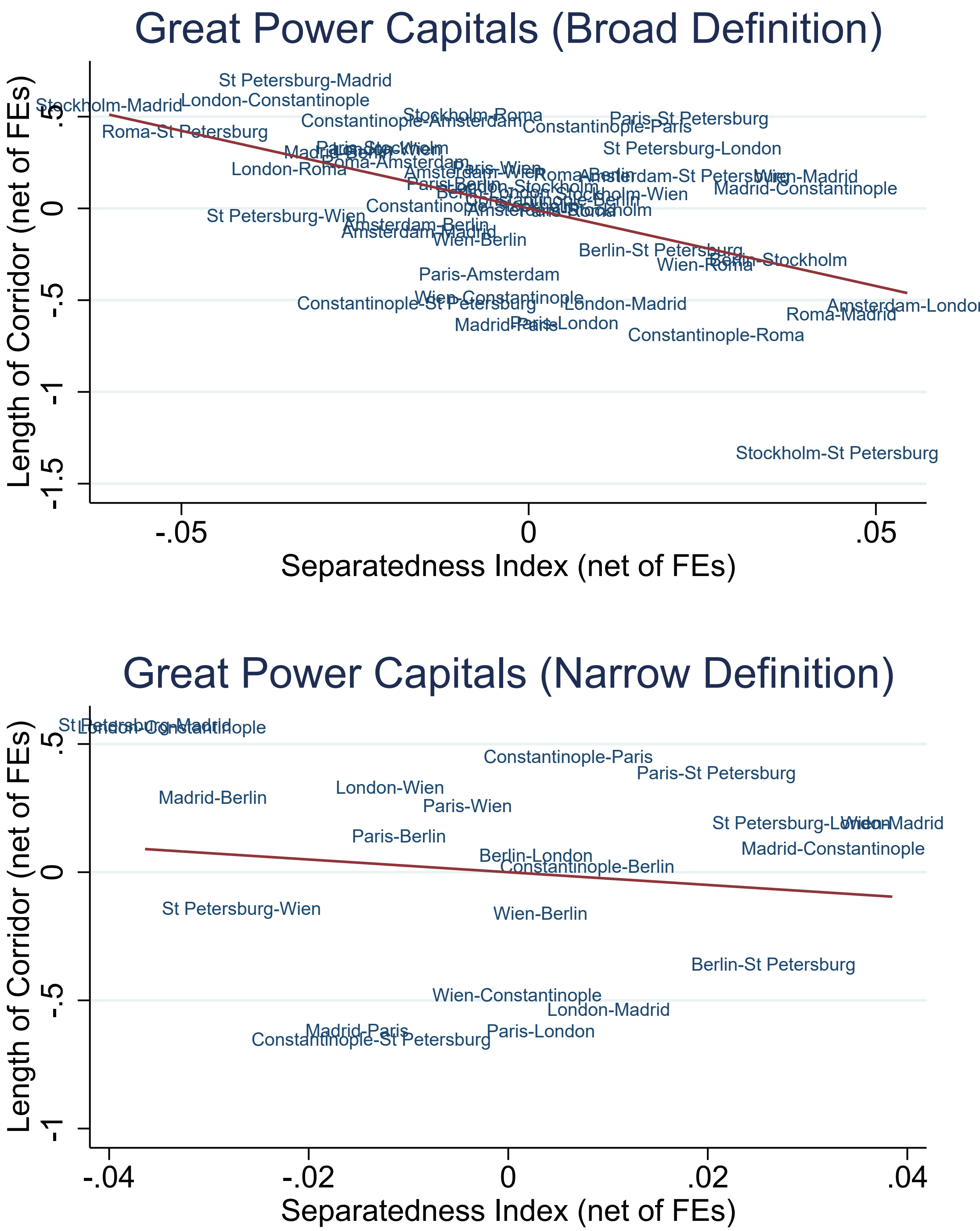


Figure A.2: Plots showing the negative relationship between geodesic distance and separatedness for Great Power capitals, based on both a narrow definition, using the same set of Great Power capitals as in the battle analysis, and a broader definition including Stockholm and Amsterdam.

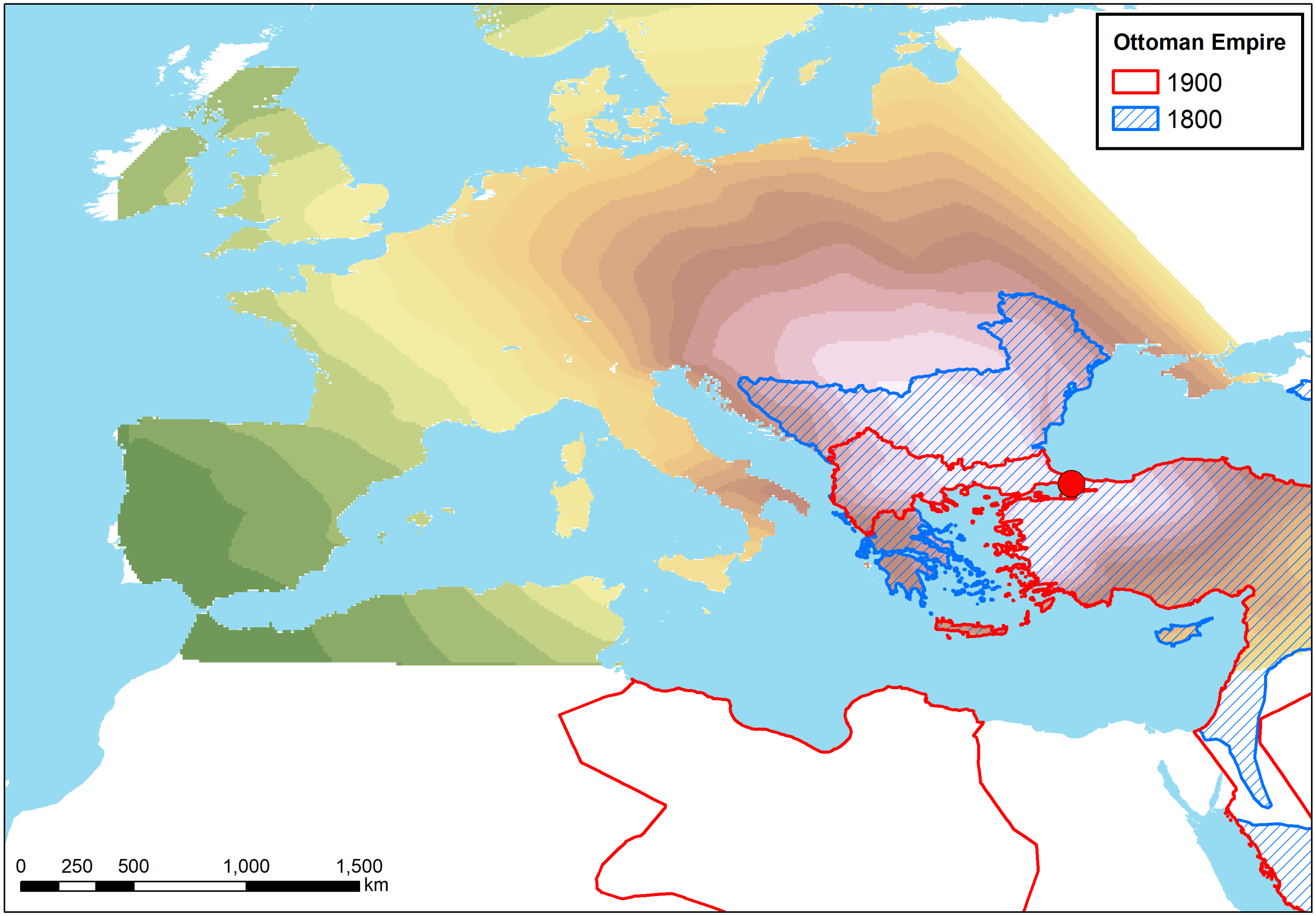


Figure A.3: Heat maps for simulated state territories around Istanbul, similar to Figure 8, but showing the actual borders of the Ottoman Empire in both 1800 and 1900.