Cities, Conflict, and Corridors

Shuhei Kitamura[†] Nils-Petter Lagerlöf[‡]

October 2021

Abstract

In this paper we propose that state structure in European history is linked to how geography affects the effective distance between state 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) state 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.

[†]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 that the location of state capitals and a region's overall state structure (i.e., the territorial size of states) both depend on geography. Specifically, we argue that terrain that slows down military incursions makes state capitals more secure, in effect allowing them to locate closer to each other, and also gives rise to a less unified state structure, i.e., many small states rather than one single unified state.

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 these 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.

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. Bosker (2021, p.3) writes that capital cities are often located "in places further away from a [country's] borders or coastline that are less vulnerable from attack by foreign powers."

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 neither of them ever dominated and/or absorbed the other, 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."[1] 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 then explore the implications for the location of capitals, and state structure more broadly. We consider a simple model where two competing states care about security and territorial size. Territories are split halfway between the two capitals. A state's security is taken to be synonymous with the effective distance to the other state's capital, which depends both on the geodesic distance and on geography. By moving its capital closer to that of the other state, each state can increase its territory, but at the cost of less security. In equilibrium, a more separating geography is found to result in the capitals being closer to each other, and (under plausible assumptions, when allowing for endogenous state unification) also make a two-state outcome more likely.

To test these model predictions, we use data on European city locations from Bosker et al. (2013) and look at pairs of such cities in 1800. First we focus on 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 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.

Then we look across all city pairs in 1800 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.

[1] 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."

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

Part of our narrative relates to an old insight articulated by Alesina and Spolaore (2003, p. 34), who write: "From just a glance at a world map, it is apparent that the capital cities, which is where many public goods are located and public services decided, tend to be around the middle of the country." This follows from many models where rulers care about the average distance from the capital to locations within the state. We rather study distances between different state capitals, and how these are linked to geography and state fragmentation.

This topic also relates broadly to research on the relationship between trade, war, and political unification (e.g., Gancia et al., 2021; Rohner et al., 2013). However, that literature does not focus on the role of geography or distances between state capitals.

In that regard, this paper connects more to an old debate about whether Europe's high degree of state fragmentation compared to other world regions was caused by its geography (see, e.g., Diamond, 1997; Fernández-Villaverde et al., 2020; Hoffman, 2015; Jones, 2003; Kitamura and Lagerlöf, 2020; Ko et al., 2018; Scheidel, 2019). 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., 2020). We are probably the first to use battle and city data, and measures of geography across corridors between cities, to explore the same topic.

Researchers using historical battle and conflict data often take an interest in the link between agricultural productivity and conflict (see, e.g., Dincecco et al., 2021; Iyigun et al., 2017; Jia, 2014). We rather examine how battle locations depend on geography and the capitals of the belligerents.

It is well known that spatial proximity and interstate conflict have a positive association (e.g., Bremer, 1992; Gleditsch and Singer, 1975; 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.

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 (Michaels and Rauch, 2018; Rappaport and Sachs, 2003), 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, 2021), historical population density (Maloney and Valencia Caicedo, 2016), and transportation and trade networks (Bakker et al., 2020; Barjamovic et al., 2019; Bosker and Buringh, 2017). Different from most of these studies we explore which cities are more likely to become state capitals, and how state territories form around those capitals, 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, or state territories, or what makes some cities more suitable to become state capitals.

3 A Model

Consider a Hotelling-type world, where two states, L and R, choose to locate their capitals on a unit interval. The locations of the capitals are denoted λ_L and λ_R, where we assume (without loss of generality) that

$$0 \leq \lambda_L \leq \frac{1}{2} \leq \lambda_R \leq 1. \tag{1}$$

The states' territories depend on where they locate their capitals. All land to the left of λ_L belongs to L (since it is closer to L), and half of the land between λ_L and λ_R belongs to L. The reminder belongs to R. Letting the size of the territories be X_L and X_R, it can be seen that

$$\begin{aligned} X_L &= \lambda_L + \tfrac{\lambda_R - \lambda_L}{2}, \\ X_R &= 1 - \lambda_R + \tfrac{\lambda_R - \lambda_L}{2}. \end{aligned} \tag{2}$$

The states (or the agents in the states making decisions about where to place the capitals) care about the total amount of state territory and the "effective" distance to the other capital, explained below. In short, they want as much land as possible, but also not be too close to the other state.

We assume the following payoff function:

$$\pi_i = E_i^{\beta} X_i, \tag{3}$$

where $i \in \{L, R\}$, E_i denotes the effective distance between L's to R's capitals, and β is an exogenous parameter capturing the relative weight on the effective distance in the payoff function.

We can interpret $1 - E_i^{\beta}$ as the probability of war in a setting where war is fully destructive, meaning each state's payoff is zero in the event of war. The fits with the empirical observation that wars are more likely to be fought between states that are closer to each other (see Section 2). The more general case where war is only partially destructive generates qualitatively identical results; see Section A.1 of the Online Appendix.

The effective distance is assumed to be increasing with the geodesic distance between the capitals, which equals just $\lambda_R - \lambda_L$, and what we call "geography" for short, denoted g, and interpreted such that a larger g means a more separating environment, such as more mountains, seas, or marshes. Specifically, we assume this functional form for the effective distance:

$$E_i = (\lambda_R - \lambda_L)^{1-g}, \tag{4}$$

where $i \in \{L, R\}$. It is easy to see that E_i is increasing with $\lambda_R - \lambda_L$ (the geodesic distance) and with g. Intuitively, whatever a high g captures (mountains, seas or marshes) is assumed to increase the effective distance between the capitals for a given geodesic distance. One interpretation could be that it slows down any potential enemy, although we do not explicitly model conflict between the states.

Note that geography is here a fixed parameter that does not vary across space. This is for tractability and serves to avoid having the geography itself depend on the endogenous capital locations, which would make the first-order conditions hard to interpret.

3.1 Equilibrium

We can now look at a Nash equilibrium, where L chooses λ_L to maximize π_L, subject to (4) and $X_L = \lambda_L + \frac{\lambda_R - \lambda_L}{2}$, taking as given λ_R (R's action). Similarly, R chooses λ_R to maximize π_R, subject to (4) and $X_R = 1 - \lambda_R + \frac{\lambda_R - \lambda_L}{2}$, taking as given λ_L (L's action).

The first-order conditions and some algebra reveal that in equilibrium the capitals are located at

$$\begin{aligned} \lambda_L &= \tfrac{1-\beta(1-g)}{2}, \\ \lambda_R &= \tfrac{1+\beta(1-g)}{2}, \end{aligned} \tag{5}$$

implying that

$$\lambda_R - \lambda_L = \beta(1-g). \tag{6}$$

We now see the following:

Proposition 1. *The geodesic distance between the capitals, $\lambda_R - \lambda_L$, is decreasing in g.*

All proofs are in Section A.1 of the Online Appendix

Thus, a flat geography, with few obstacles, implying a low g, is associated with a longer geodesic distance between the capitals, with λ_L being closer to 0 and λ_R closer to 1.

3.2 Endogenous state structure

We can extend the setting to allow for endogenous state structure. Suppose the two states (or the elites ruling them) have the option to unify and share the territory equally, meaning $X_L = X_R = 1/2$. Since there exist no external states with capitals when the territory is unified, we assume that the variable capturing the effective distance equals its maximum value, $E_i = 1$. This is what the effective distance would be if there were two states with capitals at $\lambda_L = 0$ and $\lambda_R = 1$. Alternatively, we can interpret $1 - E_i^{\beta}$ as the probability of (fully destructive) war between the two states, which is zero under unification; cf. Section A.1 in the Online Appendix.

Finally, in order for unification to not always dominate a two-state structure, we assume that this option carries a utility cost of $S > 0$. This could capture an intrinsic value to the elites of ruling their own sovereign states. The payoffs under unification then become

$$\pi_i^{\text{unif}} = \frac{1}{2} - S, \tag{7}$$

for $i \in \{L, R\}$.

To find the payoff to not unifying, what we call a two-state structure, we use (3), (4), (5) and (6) to find that

$$\pi_i^{\text{frag}} = \frac{1}{2} \left[\beta(1-g)\right]^{\beta(1-g)}, \tag{8}$$

for $i \in \{L, R\}$.

To get the results we are after, we impose a lower bound for g, denoted $\underline{g} \in (0, 1)$, such that

$$g \in (\underline{g}, 1], \tag{9}$$

and assume that

$$-\ln\left[\beta(1-\underline{g})\right] > 1. \tag{10}$$

Now we can note that:

Lemma 1. *Under the assumptions in (9) and (10), π_i^{frag} is increasing in g.*

Intuitively, there are two different effects on π_i^{frag} from an increase in g, both working through the effective distance. A direct effect works through the exponent in (4), and raises the effective distance holding the geodesic distance (the base) constant. An indirect effect works in the opposite direction by shrinking the geodesic distance in equilibrium. Generally, π_i^{frag} may be increasing or decreasing in g, but under the assumptions made here it is increasing, meaning the direct effect dominates, so that a more separating geography (a higher g) implies a higher payoff to a two-state structure. Put another way, under these parametric assumptions a more separating geography implies a longer effective military distance in equilibrium, which arguably seems more plausible than the opposite.

Finally, define

$$S^{\max} = \frac{1 - \left[\beta(1-\underline{g})\right]^{\beta(1-\underline{g})}}{2} > 0. \tag{11}$$

We can now state the following:

Proposition 2. *Suppose $S < S^{max}$. Then there exists a $g^* \in (\underline{g}, 1]$, such that a two-state structure is preferred over unification ($\pi_i^{frag} > \pi_i^{unif}$) when $g > g^*$, and unification is preferred ($\pi_i^{frag} < \pi_i^{unif}$) when $g < g^*$.*

The above proposition thus says that a more separating environment (a higher g) makes a two-state structure more likely. This follows from the payoff to having two states being increasing in g, which in turn hinges on the way in which g affects the effective distance between the capitals in equilibrium.

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.[2] 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.[3] 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-cordinates 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 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 GP's: England/Great Britain; France; Russia; Prussia/Germany; Austria/Habsurg Empire/Austria-Hungary; Spain; and the Ottoman

[2]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.

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

Empire. These are the ones discussed in most detail in the influential study of the European Great Power system by Levy (1983).[4] The matching of battles to GP's was done manually by Kitamura (2021), who provides further details on this process.

These GP's 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; Koeningsberg (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 geocoordinates 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 GP's 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 Koeningsberg, respectively, as capitals. In effect, we can thus treat St. Petersburg and Berlin as the capitals of Russia or Germany/Prussia in our benchmark analysis. (Letting Moscow be the capital of Russia does not change the main results; see Section A.2 of the Online Appendix.) More generally, even though their territories and regimes were often fluid, all seven GP's can be thought of as having fixed political centers.

[4] 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 GP's over nearly as long periods of time as the other seven; see Levy (1983, Table 2.1).

The upshot is a set of 685 battles fought between these eleven different pairs of GP's.

4.1.1 Cell Data

Much of 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 panel dataset with $11 \times 1,450 = 15,950$ GP-pair/cell combinations. Figure 1 shows a map of the precise battle locations and which cells are coded as battle cells.

4.2 Geography and Shortest-Distance Corridor

In the cell-level analysis, the variable that we call the *shortest-distance corridor* 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.

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 2 illustrates the battle cells for six GP pairs, together with cells making up the associated shortest-distance corridor as well as the three geography variables.

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. 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

5.1 Distance Correlations

We start by exploring the relationship between geography and battle locations for one pair of GP's at a time, and plot the distances between the location of their battles to the two GP's capitals; see Figure 3. The shortest distance line between the belligerents' capitals is represented by a line with slope minus one and intercepts equal to the full distance between the two capitals. Note that no battles can fall below the shortest-distance line since the sum of the two distances is minimized when the battle is on the line.

If all battles take place exactly on the shortest-distance line between the two capitals, then the distances from the battle location to each the respective capitals should have a correlation of minus one. Figure 3 considers two GP pairs: England and France, where the correlation is positive, and the distance between the capitals relatively short; and Russia and France, where the correlation is negative and the distance between the capitals long. This fits with the idea that battles tend to be fought away from the shortest-distance line when it is intercepted by certain types of geography, in the case of England-France by the English Channel, in turn allowing the capitals to be closer. By contrast, Russia and France have a less separating geography between them, and thus fight their battles closer to the shortest-distance line.[5]

Obviously, this example considers just two pairs, and appeals to one particular type of geography (seas). There are limits to how systematic we can be with just 11 pairs, but we pursue some further investigation in Section A.2 of the Online Appendix.

5.2 Cell-Level Analysis

For the remainder of this section we use one-degree cells, 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 GP's under consider-

[5]We here use distances to St. Petersburg, which was formally the Russian capital when these battles were fought, although Napoleon's 1812 invasion advanced towards Moscow. We return to this discussion later.

ation.

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. 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 $D_{j,p} = D_{i,q} = 0$ for $j \neq i$ (i.e, cells off the London-Paris corridor) and $q \neq p$ (i.e., other GP pairs).

5.2.1 Direct Effects

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

$$B_{i,p} = \alpha + \beta_{\mathrm{D}} D_{i,p} + \lambda_{\mathrm{S}} S_i + \lambda_{\mathrm{H}} H_{800,i} + \lambda_{\mathrm{M}} M_i + \omega_p + \varepsilon_{i,p}, \tag{12}$$

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 (12), absorbing the geography controls, and giving us the following specification:

$$B_{i,p} = \beta_{\mathrm{D}} D_{i,p} + \omega_p + \gamma_i + \varepsilon_{i,p}, \tag{13}$$

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.

5.2.2 Interaction Effects

So far we have documented that GP's 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:

$$\begin{aligned} B_{i,p} = \; & \beta_{\mathrm{D}} D_{i,p} \\ & + \beta_{\mathrm{S}} D_{i,p} S_i \\ & + \beta_{\mathrm{H},800} D_{i,p} H_{800,i} \\ & + \beta_{\mathrm{M}} D_{i,p} M_i \\ & + \omega_p + \gamma_i + \varepsilon_{i,p}, \end{aligned} \tag{14}$$

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 effect is 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 GP's 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 4 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., 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; Ellingsen, 2021; Nagy, 2020).

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 predictions of that model, as summarized by Propositions 1 and 2. To recap, Proposition 1 states that the geodesic distance between capitals should be shorter when the geography is more separating, and Proposition 2 states that a two-state equilibrium is more likely in a more separating geography.

6.1 Geodesic Distances Between Capitals

To test Proposition 1 we use the dataset from Bosker et al. (2013), and look at pairs of capital cities in 1800. 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).

We can now write the regression equation as

$$L_{i,j} = \delta_{\mathrm{S}} S_{i,j} + \delta_{\mathrm{H},800} H_{800,i,j} + \delta_{\mathrm{M}} M_{i,j} + \chi_i + \rho_j + \eta_{i,j}, \tag{15}$$

where $\eta_{i,j}$ is the error term. The estimates of the different δ's should all be negative according to Proposition 1. Table 3 presents results from a number of such regressions. All specifications include fixed effects for both origin and target capitals, i.e., χ_i and ρ_j in (15).

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 Proposition 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 of 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.

Moreover, the negative estimate of the mountain coefficient is not robust. Column (5) drops those pairs where both cities were capitals only in 1800, and became non-capital cities in either 1900, 2000, or in recent modern times (according to Euratlas and GADM, respectively; see below). These cities are Firenze, Genoa, Milano and Naples in modern Italy and Munich in modern Germany; note that Italy and Germany did not exist as states in 1800. As seen in column (5), when dropping these pairs, thus shrinking the sample by about 4%, the coefficient on the fraction mountain is no longer significantly different from zero.

We also note that the positive sign on the fraction mountain comes out as insignificant in column (6), which reverts to the full sample and allows standard errors to be clustered on both origin an target city, while the signs on the other geography variables stays negative and significant.

In summary, while we tend to find insignificant results for the fraction mountain, the two coefficient estimates that come out as consistently significant—the fractions sea and marshland—carry negative signs, as we would expect from Proposition 1.

6.2 Same-State Outcomes

To test Proposition 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 (corresponding to a high g in the model), controlling for the geodesic distance between them (corresponding to $\lambda_R - \lambda_L$ in the model).

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 a 50 km buffer zones from the shortest-distance line. As before, we refer to these as corridors between cities.

In the benchmark 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, 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 215,495 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

$$C_{i,j} = \lambda_{\mathrm{L}} L_{i,j} + \lambda_{\mathrm{S}} S_{i,j} + \lambda_{\mathrm{H},800} H_{800,i,j} + \lambda_{\mathrm{M}} M_{i,j} + \kappa_i + \theta_j + \eta_{i,j}, \tag{16}$$

where 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 (16), letting the different geography variables enter both one by one and together. All specifications include fixed effects for both cities in the pair, i.e., κ_i and θ_j in (16).

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) allows standard errors to be clustered on origin and target city with very similar results, except that the fraction marshes comes out as insignificant.

In sum, with a slight caveat regarding the results for marshes, the overall pattern suggests that geography affected the shape of states in ways that are consistent with Proposition 2, when we measure geography in the same way as in our battle-level analysis.

Section A.3 of the Online Appendix explores a few robustness checks of the results in Table 4, measuring same-state outcomes in later years than 1900, and city pairs based on cities that existed in 800 CE already. The results are broadly 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 the capitals for which we have data from Bosker et al. (2013). These are London, Paris, Madrid, Berlin, Vienna and Istanbul, but not St. Petersburg. Figure 5 shows different so-called heat maps, indicating which other cities are predicted to be most likely to belong to each of these respective capital cities, or what we can label connectedness. The predictions are based on the regression in column (4) of Table 4 (but ignoring the city fixed effects when generating the predictions).

The maps in Figure 5 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.2 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 state 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 state 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 tend to deviate from that corridor where it is intercepted by certain types of geography, specifically seas, mountains, and marshes. This result is robust to several sorts of controls, sample restrictions, and econometric specifications. Because battles would 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 the leaders of two states care about the effective military distances to the other state's capital, as well as state territories, both of which depend on where capitals are located. From this model we derive two predictions: (1) the geodesic distance between state capitals should be shorter when the geography between them is more separating (i.e., when it is intercepted by the same geography found to affect battle locations); and (2) a two-state outcome (as opposed to state unification) should be more likely when geography is more separating.

To test these predictions we use data on the location of state 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. State 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 intersected by more seas, mountains, and marshes, implying that these types of geography make unification less likely.

To illustrate this last result, we also present various maps showing the most probable state territories, as predicted by our same-state regressions, and based on the geo-coordinates of the Great Power capitals from the battle analysis (London, Paris, Madrid, Berlin, Vienna, and Istanbul; city data is missing for Russia). The maps show some striking resemblance to the the actual state territories, with a couple of exceptions that we argue are interesting in their own right.

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)
Fra Sea along Corr.	−1.103***			−1.163***	−1.302***	−1.163***
	(0.230)			(0.213)	(0.226)	(0.289)
Fra Mountain along Corr.		1.205***		0.689**	0.279	0.689
		(0.254)		(0.283)	(0.344)	(0.441)
Fra Marsh along Corr.			−21.139***	−27.230***	−27.927***	−27.230***
			(5.380)	(5.990)	(5.895)	(5.875)
R^2	0.65	0.62	0.62	0.68	0.68	0.68
Number of obs.	405	405	405	405	389	405
Sample	Full	Full	Full	Full	Dropping pairs within countries in 1900 or today	Full
Standard errors	Robust	Robust	Robust	Robust	Robust	Clustered

Notes: Ordinary least squares regressions across city pairs made up only by those cities which were state capitals in 1800 according to Bosker et al. (2013). Robust standard errors are indicated in parentheses, except for column (6), which uses two-way clustering on origin and target capital city. The dependent variable is the geodesic distance between the capitals, i.e., the length of the corridor. Column (5) drops those pairs of capitals where both are located within the same country in either 1900 (according to Euratlas), or today (as defined by GADM); all those pairs include a city in Italy or Germany. Column (6) uses the same specification as in column (4), but reports standards errors clustered both on origin and target city. All specifications include fixed effects for origin and target city. * 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)
Length of Corridor	-0.281^{***}	-0.283^{***}	-0.285^{***}	-0.274^{***}	-0.274^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.012)
Fra Sea along Corr.	-0.301^{***}			-0.427^{***}	-0.427^{***}
	(0.003)			(0.002)	(0.022)
Fra Mountain along Corr.		-0.106^{***}		-0.470^{***}	-0.470^{***}
		(0.007)		(0.006)	(0.056)
Fra Marsh along Corr.			1.853^{***}	-2.171^{***}	-2.171
			(0.332)	(0.365)	(1.450)
R^2	0.47	0.45	0.44	0.48	0.48
Number of obs.	215495	215495	215495	215495	215495
Standard errors	Robust	Robust	Robust	Robust	Clustered

Notes: Ordinary least squares regressions across city pairs made up by cities which has a population above 5,000 in 1800. Robust standard errors are indicated in parentheses, except for column (5), which uses two-way clustering on origin and target city. 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 fixed effects for origin and target city. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table 4: Same State Outcomes across City Pairs.

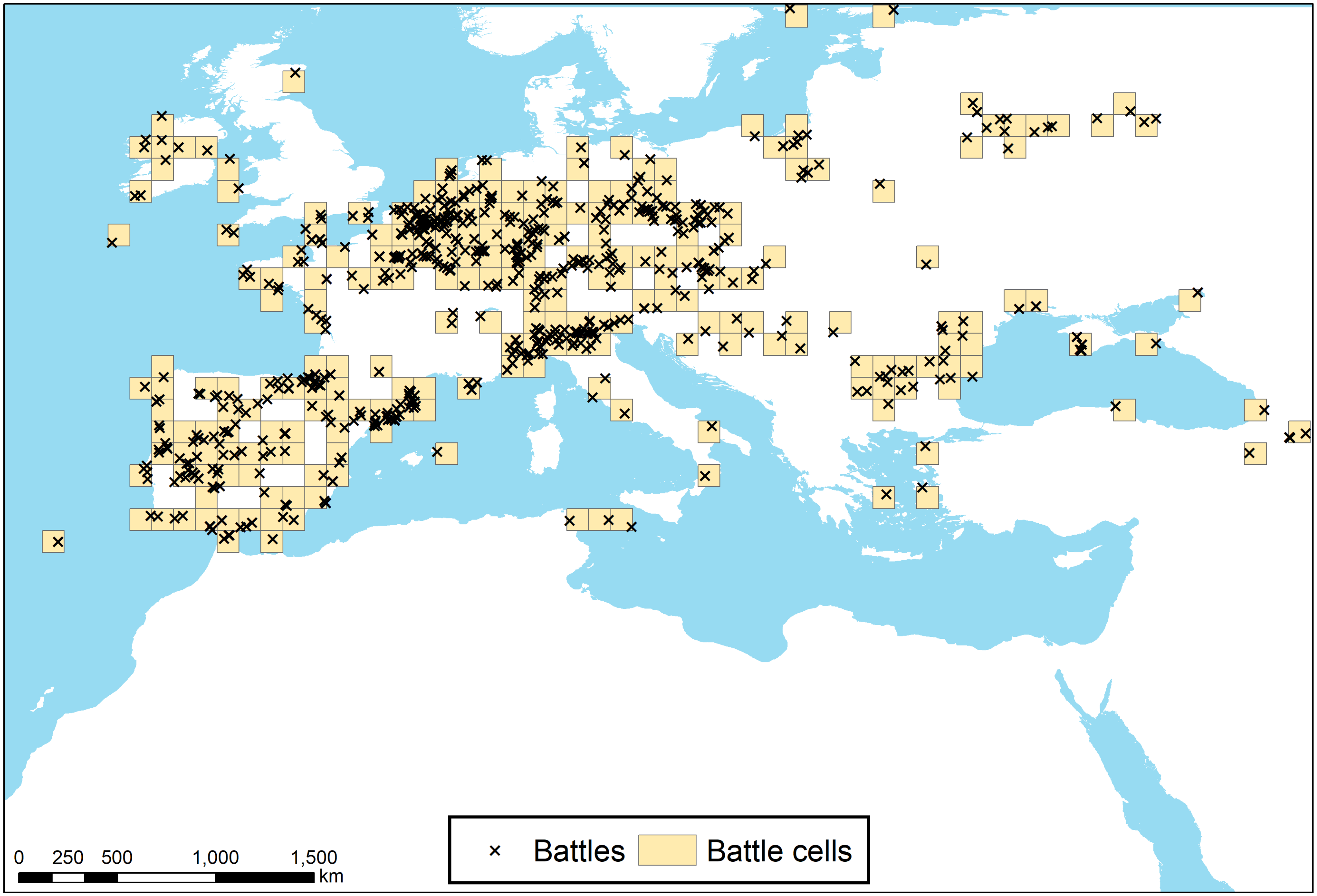
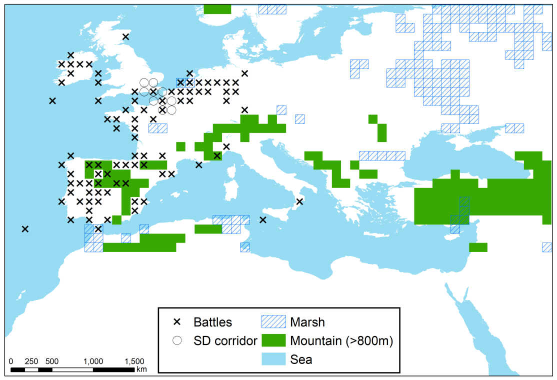
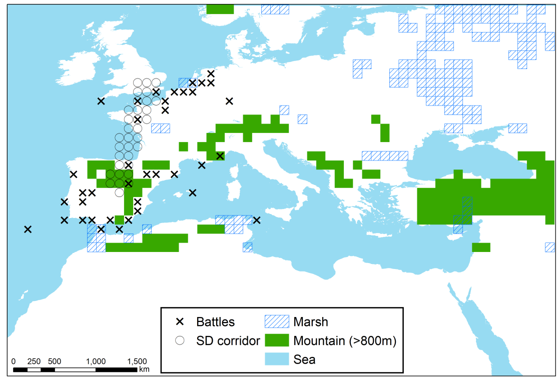


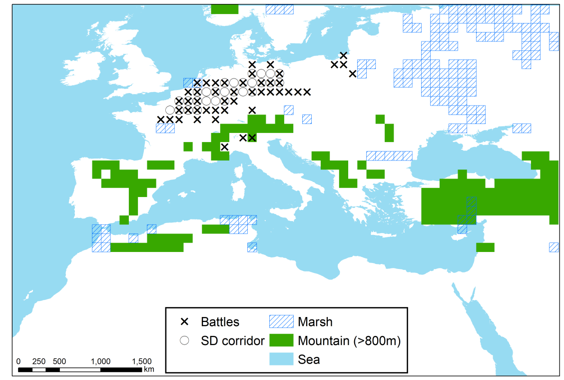
Figure 1: Map of all battles and battle cells.



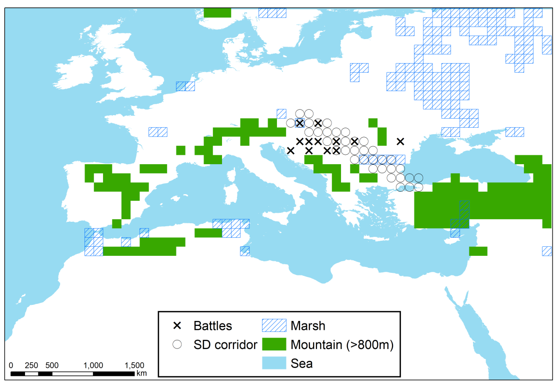
(a) England vs. France



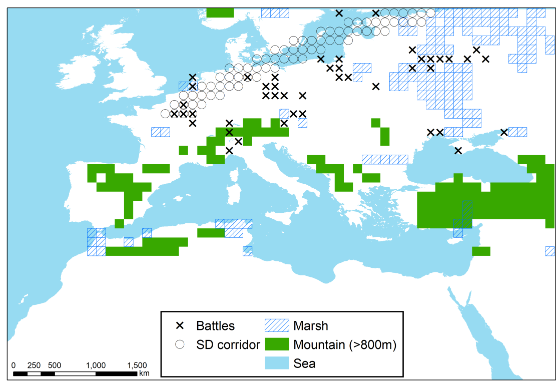
(b) England vs. Spain



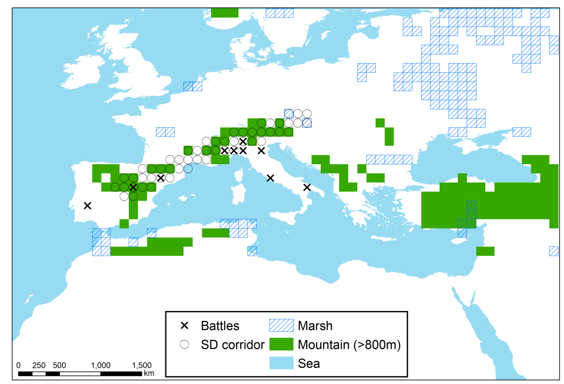
(c) France vs. Germany



(d) Ottoman Empire vs. Austria



(e) France vs. Russia



(f) Austria vs. Spain

Figure 2: Maps of some Great Power battles.

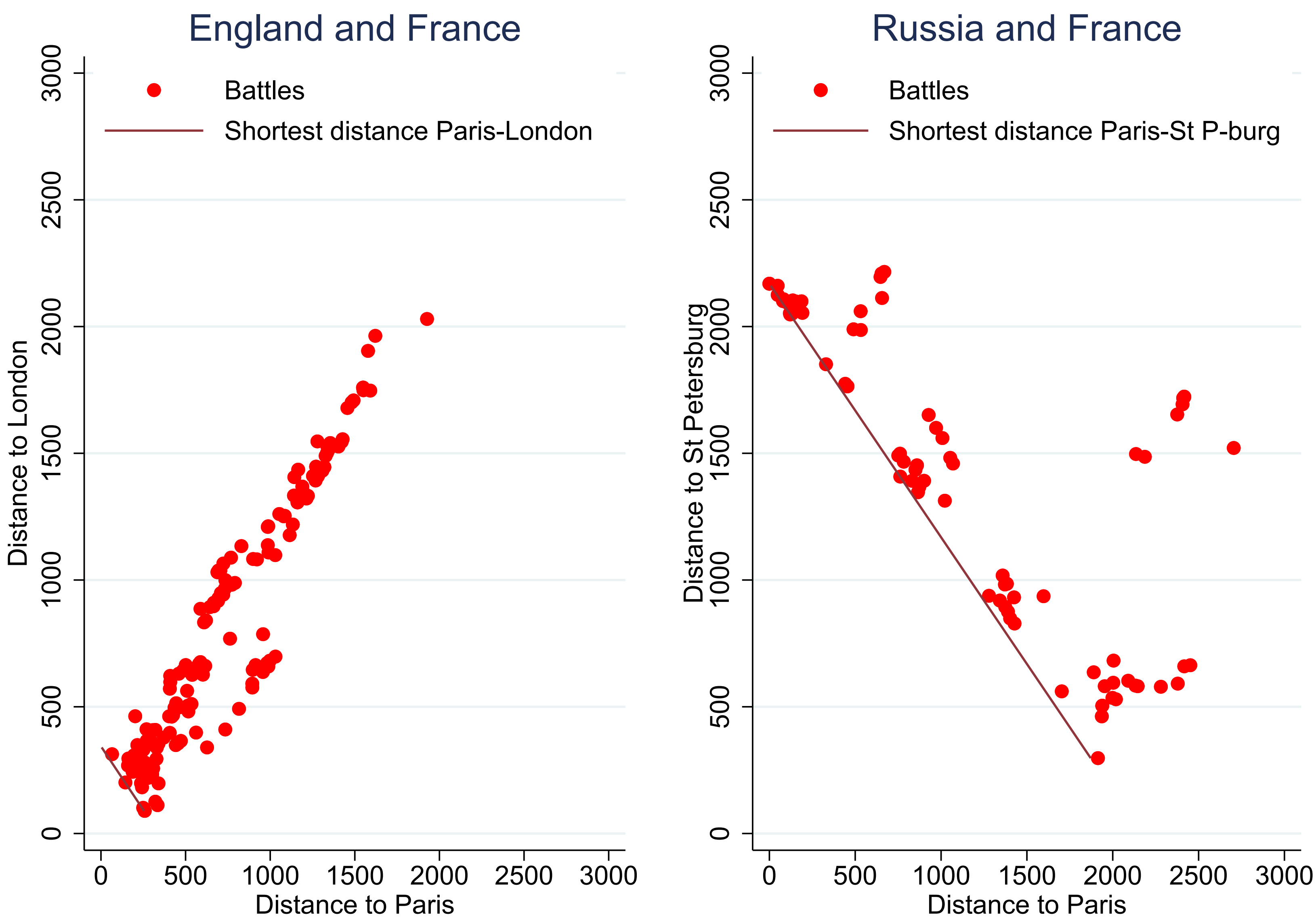


Figure 3: Examples of a positive and negative relationships between distances to capitals.

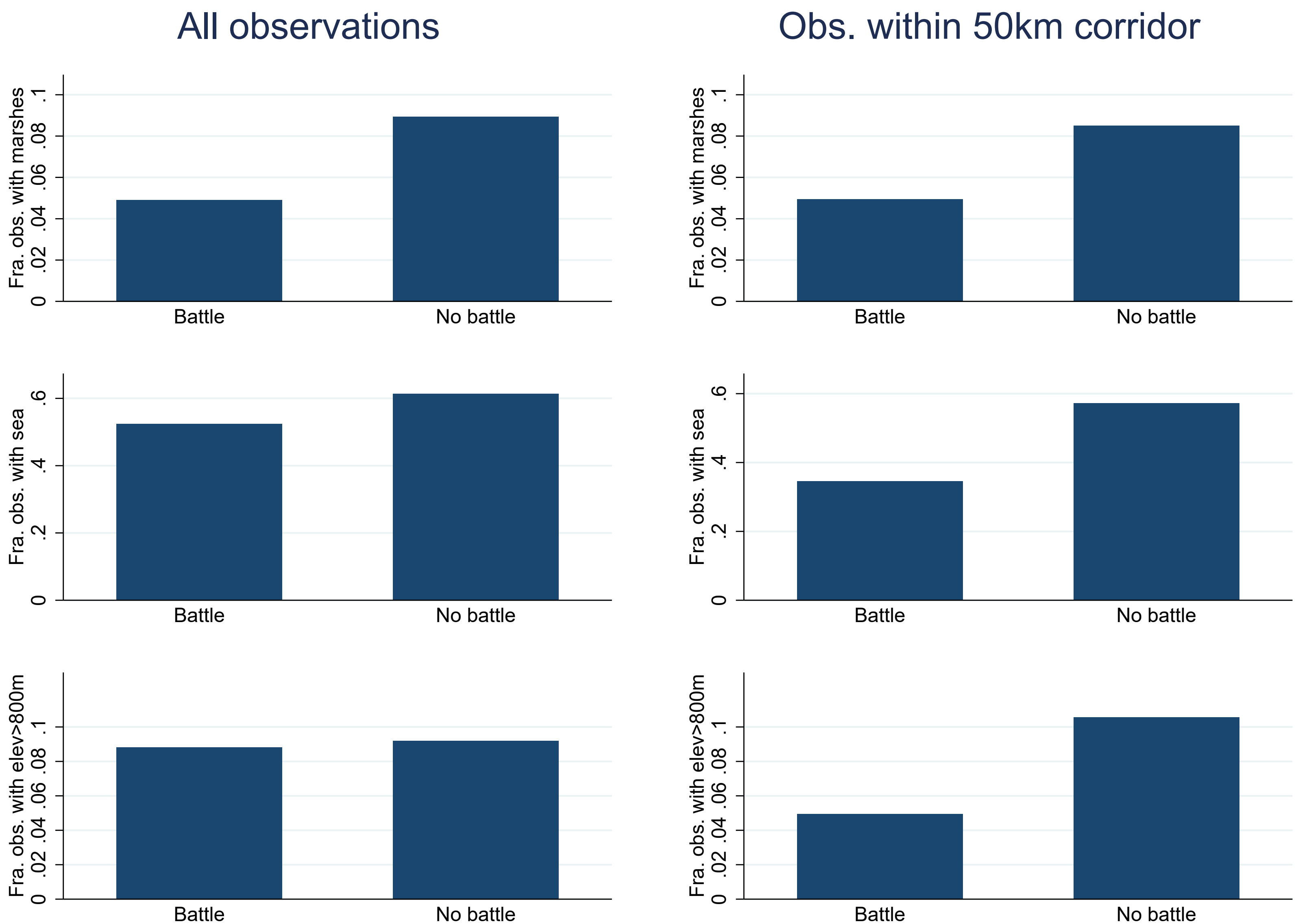
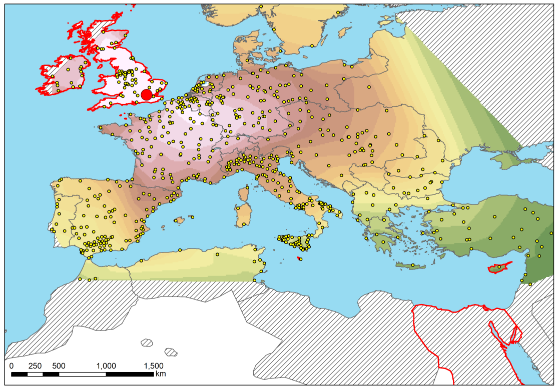
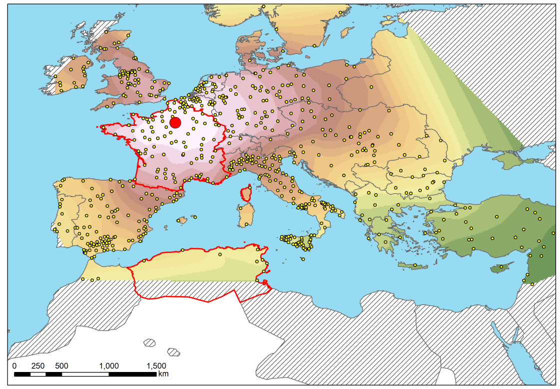


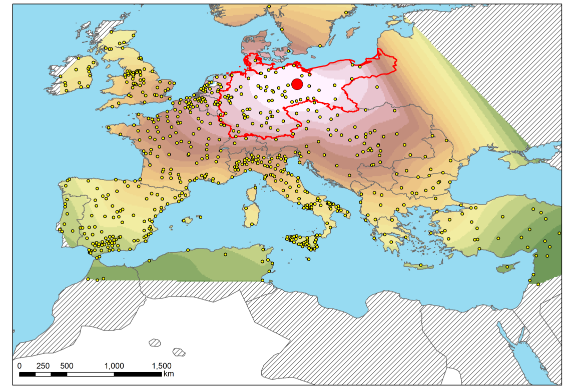
Figure 4: How geography differs between cell-pairs with and without battles, for the full sample and for cell-pairs close to the shortest-distance corridor between the belligerents' capitals.



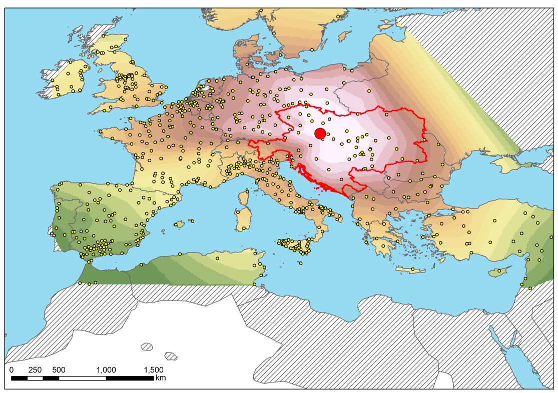
(a) Heat map for London.



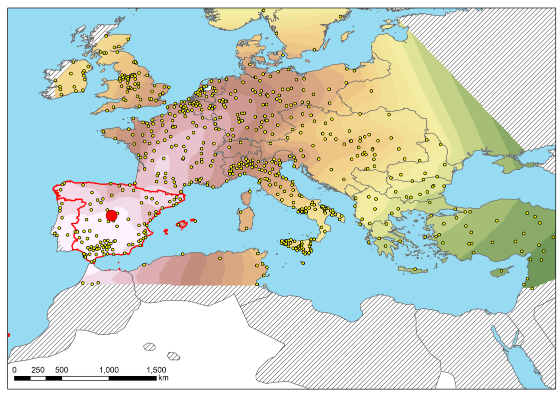
(b) Heat map for Paris.



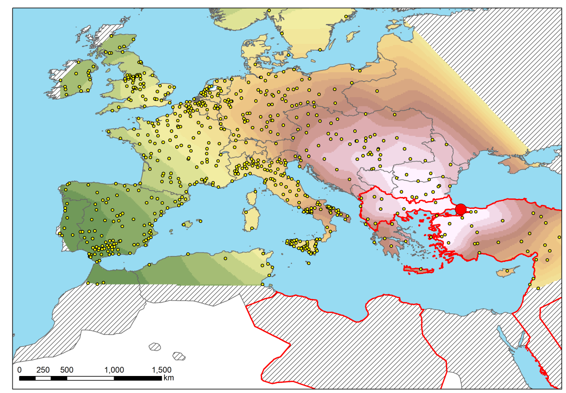
(c) Heat map for Berlin.



(d) Heat map for Vienna.



(e) Heat map for Madrid.



(f) Heat map for Istanbul.

Figure 5: Heat maps for simulated state territories for select capitals, with actual state borders in 1900 marked in red.

References

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

Bakker, J. D., Maurer, S., Pischke, J.-S., and Rauch, F. (2020). Of mice and merchants: connectedness and the location of economic activity in the Iron Age. *Review of Economics and Statistics, forthcoming*.

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. (2021). City origins. *Regional Science and Urban Economics, forthcoming*.

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.

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. (2021). The long-run effects of early agriculture: evidence from carbon dated archaeological sites. *Mimeo, York University and Brock University.*

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*.

Fernández-Villaverde, J., Koyama, M., Lin, Y., and Sng, T.-H. (2020). The fractured-land hypothesis. *Mimeo, University of Pennsylvania, GMU, NUS, and Guangdong University.*

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

Gancia, G., Ponzetto, G. A., and Ventura, J. (2021). Globalization and political structure. *Mimeo, QMU, UPF and Barcelona GSE*.

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(February):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 Euope. *International Economic Review*, 59(1):285–327.

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*.

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. 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.

A Online Appendix

A.1 The Model

A.1.1 Interpreting the Effective Distance in Terms of Conflict Probabilities

This section explores a setting where the effective distance between the two states' capitals determines the probability of war, and where expressions similar to those in (3) in the main text can be derived as expected payoffs. More precisely, let the probability of war be $1 - E_i^\beta$, meaning a shorter effective distance makes the outbreak of war more likely. This seems empirically plausible, since it is well documented that states which are closer to each other are more likely to engage in military conflict, and since E_i is increasing in the geodesic distance between capitals.

The outbreak of war is here an exogenous event and not a choice by any of the players.

If peace prevails, each state gets a territory given by (2). If war breaks out, each state wins the conflict with equal probability $1/2$. The winner seizes the full unit interval (including that of the other state), minus some loss $\delta \in (0, 1]$, while the loser gets a territorial payoff of zero. That is, there is an aggregate loss of war (here captured by the territorial term in the payoff function, but could just as well be interpreted as loss of lives and expenditure). The winner gets a net (territorial) payoff of $1 - \delta$, and the loser zero; in peace the aggregate territorial payoff equals one, split according to (2).

For notational convenience, we temporarily let $\omega = 1 - \delta$ be the winner's net territorial gain in war, and define $\alpha = \beta(1 - g)$, so that $E_i^\beta = (\lambda_R - \lambda_L)^\alpha$. The expected payoff for state i can now be written

$$\begin{aligned} \pi_i &= E_i^\beta X_i + (1 - E_i^\beta)\left(\tfrac{1}{2}\right)(1 - \delta) \\ &= E_i^\beta X_i + (1 - E_i^\beta)\tfrac{\omega}{2} \\ &= \tfrac{\omega}{2} + E_i^\beta\left(X_i - \tfrac{\omega}{2}\right) \\ &= \tfrac{\omega}{2} + (\lambda_R - \lambda_L)^\alpha\left(X_i - \tfrac{\omega}{2}\right). \end{aligned} \tag{17}$$

The main text consider the case where war is fully destructive (meaning $\delta = 1$ and $\omega = 0$), making the payoff in case of war zero for both players.

Maximizing π_i in (17), subject to the expression for X_i in (2), the first-order conditions for each state's capital can be written

$$\begin{aligned}
\frac{\partial \pi_L}{\partial \lambda_L} &= -\alpha\left(\lambda_R-\lambda_L\right)^{\alpha-1}\left(\tfrac{\lambda_R+\lambda_L-\omega}{2}\right)+\left(\lambda_R-\lambda_L\right)^{\alpha}\tfrac{1}{2} &&= 0, \\
\frac{\partial \pi_R}{\partial \lambda_R} &= \alpha\left(\lambda_R-\lambda_L\right)^{\alpha-1}\left(1-\tfrac{\lambda_R+\lambda_L+\omega}{2}\right)-\left(\lambda_R-\lambda_L\right)^{\alpha}\tfrac{1}{2} &&= 0.
\end{aligned} \tag{18}$$

It is straightforward to see that $\lambda_R + \lambda_L = 1$ holds in equilibrium. Substituting this back into the first-order conditions above some algebra shows that

$$\begin{aligned}
\lambda_L &= \tfrac{1-(1-\omega)\alpha}{2} = \tfrac{1-\beta\delta(1-g)}{2}, \\
\lambda_R &= \tfrac{1+(1-\omega)\alpha}{2} = \tfrac{1+\beta\delta(1-g)}{2}.
\end{aligned} \tag{19}$$

where we have switched back to the original notation, $\omega = 1-\delta$ and $\alpha = \beta(1-g)$. Letting war be fully destructive (setting $\delta = 1$) brings us back to 5 in the main text.

A.1.2 Proofs of Propositions

Proof of Proposition 1. The claim follows directly from 6. ■

Proof of Lemma 1. Define $f(g) = (1-g)\ln[\beta(1-g)]$. We see from (8) that, if $f'(g) > 0$, then π_i^{frag} is increasing in g. First we see that $f''(g) > 0$. Next some algebra shows that $f'(\underline{g}) > 0$ follows from (10). Now, $f'(\underline{g}) > 0$ and $f''(g) > 0$ together imply that $f'(g) > 0$ for all $g \in (\underline{g}, 1]$. ■

Proof of Proposition 2. From (7) and (8) it follows that a two-state structure is preferred when

$$S > \left[1-[\beta(1-g)]^{\beta(1-g)}\right]/2. \tag{20}$$

The right-hand side of (20) is monotonically decreasing in g and goes from $S^{\max}$ to 0 as g goes from $\underline{g}$ to 1; cf. the proof of Lemma 1. Since $S \in (0, S^{\max})$ it follows that there exists a threshold $g^* \in (\underline{g}, 1]$ with the properties described in the proposition. One can see this by drawing a graph with g on the horizontal axis, and the right-hand side of (20) on the vertical axis. ■

A.2 Battle Data Analysis: Robustness and Further Exploration

A.2.1 Distance Correlations

This section looks at battle-level correlations between distances from battles to capitals. We let the unit of observation be a battle and regress the distance to one capital on the distance to the other capital, both measured from the site of the battle. We do this for each of the eleven GP pairs separately; note that a battle can involve several pairs. The results are shown in the eleven columns of Table A.1. We use the following GP acronyms (in parentheses): England/Great Britain/United Kingdom (EN); France (FR); Russia (RU); Prussia/Germany (GE); Austria/Habsurg Empire (AT); Spain (SP); and the Ottoman Empire (OE).

Starting with battles between England and France in column (1) we find a positive coefficient. This tells us that battle locations far away from London were also far from Paris, and those closer to London were also closer to Paris. In other words, these battles were *not* clustered along the shortest-distance corridor between Paris and London; cf. Figure 3.

The same positive relationship holds for battles between England and Russia in column (2).

By contrast, battles between England and Spain show a negative distance association in column (3), and for most of the remaining columns we also find negative relationships. That is, these battles tended to take place close to the shortest-distance corridor. Figure 3 illustrates the examples if France-England and France-Russia, corresponding to columns (1) and (4) in Table A.1. Curiously, while all these coefficients vary in sign and magnitude, most are relatively precisely estimated, despite a very small sample size in some cases.

This heterogeneity can be understood from the geography along the corridors between the capitals, as measured by seas, marshes, and mountains (elevation exceeding 800 meters). With some simplification, the more of the corridor was covered by sea, marshland, and mountains, the more positive the distance correlation tends to be. Intuitively, these features of the geography tend to push battles away from the shortest-distance corridor between the capitals, thus making distances to both capitals longer.

The length of the corridor also matters. The longer is the corridor the more likely it is that battles are fought along it, all else equal, implying a more negative coefficient in Table A.1.

The third factor is more ad-hoc: battles between Spain and England tend to be located along the corridor between Madrid and London, even though it is mostly covered by sea; most of these were naval battles. The reason these two Great Powers fought so much by sea is a separate story, but if we accept the frequency of naval battles as "explanation," then we end up with a relatively complete account of the variation in regression outcomes seen in Table A.1.

Figure A.1 shows how these five factors relate to the estimated coefficients in Table A.1 across these eleven pairs. Not all partial relationships carry the expected sign. However, a meta analysis in Table A.2 explores what happens when these are entered jointly in a regression, with the distance coefficient from Table A.1 as the dependent variable, and each of the above factors as independent variables, i.e., the three geography measures, the length of the corridor, and the fraction naval battles. All five come out as significant at 10% level or lower, and mostly with the correct signs, even though the dataset consists of only eleven observations.

The bottom right panel of Figure A.1 shows the relationship between the actual values for the estimated coefficients in Table A.1 and those predicted from column (6) of Table A.2, showing a very good fit. In particular, the five variables can jointly account for the three positive distance coefficients in Table A.1, referring to the pairs of Ottoman Empire-Russia, England-France, and England-Russia.

A.2.2 Cell-Level Analysis

In the main analysis we defined cells with mean elevation above 800 meters as mountain cells. Table A.3 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.4 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.5 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.6 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.7 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.8 allows for spatially adjusted standard errors and declining weights (applying the acreg command in Stata and the Bartlett option). 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.9 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 5000 in the year indicated for each column of Table A.9. 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.10 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 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 the same Euratlas data as for 1900), and Table A.12 shows the results when using modern country borders from the Global Administrative Boundaries (GADM) database (version 3.6). These results are qualitatively similar to those in Table 4.

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. 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, strongest around 1500-1600 and slightly weaker by 1800.

Online Appendix Tables and Figures

	Dependent variable is distance to:										
	London			Paris				Berlin	Madrid	Istanbul	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Dist. to Paris	1.095***										
	(0.019)										
Dist. to St. P-burg		0.799***		−1.015***							0.417*
		(0.157)		(0.066)							(0.234)
Dist. to Madrid			−0.991***		−0.647***						
			(0.113)		(0.065)						
Dist. to Vienna						−0.567***		−0.810***	−0.810***	−0.899***	
						(0.186)		(0.219)	(0.161)	(0.027)	
Dist. to Berlin							−1.015***				
							(0.061)				
R^2	0.90	0.83	0.56	0.54	0.50	0.38	0.56	0.58	0.75	0.92	0.15
Number of obs.	167	12	62	83	136	191	101	38	12	16	38
Pair of belligerents	EN-FR	EN-RU	EN-SP	FR-RU	FR-SP	FR-AT	FR-GE	GE-AT	SP-AT	OE-AT	OE-RU

Notes: Ordinary least squares regressions with robust standard errors in parentheses. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.1: Distance regressions for 11 Great Power pairs.

	Dependent variable is the distance coefficient from Table A.1.					
	(1)	(2)	(3)	(4)	(5)	(6)
Fraction Sea	2.447*			2.516*	4.016*	6.870***
	(1.285)			(1.282)	(1.660)	(1.224)
Fraction Mountain (>800m)		−2.954		−0.567	0.624	3.377*
		(1.852)		(1.606)	(2.129)	(1.657)
Fraction Marsh			2.953***	3.227***	5.617**	9.779***
			(0.816)	(0.920)	(1.688)	(1.506)
Length of Corridor					−0.009	−0.019***
					(0.006)	(0.004)
Fraction Naval Battles						−3.415**
						(1.001)
R^2	0.27	0.16	0.12	0.45	0.61	0.81
Number of obs.	11	11	11	11	11	11

Notes: Ordinary least squares regressions with robust standard errors in parentheses. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.2: Explaining variation in the distance regression coefficients in Table A.1.

	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.3: 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.4: 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.5: 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.6: 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.7: 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	0.245^{***}	0.245^{***}	0.245^{***}	0.245^{***}	0.245^{***}	0.245^{***}
	(0.033)	(0.042)	(0.049)	(0.053)	(0.055)	(0.058)
Sea × SD-Corridor	-0.121^{***}	-0.121^{***}	-0.121^{***}	-0.121^{***}	-0.121^{**}	-0.121^{**}
	(0.036)	(0.040)	(0.041)	(0.043)	(0.047)	(0.049)
Mountain × SD-Corridor	-0.150^{***}	-0.150^{***}	-0.150^{***}	-0.150^{***}	-0.150^{***}	-0.150^{***}
	(0.031)	(0.039)	(0.041)	(0.044)	(0.045)	(0.046)
Marsh × SD-Corridor	-0.136^{***}	-0.136^{*}	-0.136	-0.136	-0.136^{*}	-0.136^{**}
	(0.052)	(0.080)	(0.089)	(0.085)	(0.075)	(0.068)
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.8: 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 5000 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.9: 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 1525-1914. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.10: Interactions with Geography: Battles 1914-1945.

	Dependent variable is the Same-State Indicator in 2000				
	(1)	(2)	(3)	(4)	(5)
Length of Corridor	-0.243^{***}	-0.245^{***}	-0.246^{***}	-0.237^{***}	-0.237^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.011)
Fra Sea along Corr.	-0.291^{***}			-0.407^{***}	-0.407^{***}
	(0.002)			(0.002)	(0.022)
Fra Mountain along Corr.		-0.088^{***}		-0.433^{***}	-0.433^{***}
		(0.006)		(0.006)	(0.056)
Fra Marsh along Corr.			2.420^{***}	-1.353^{***}	-1.353
			(0.293)	(0.325)	(1.187)
R^2	0.51	0.49	0.49	0.52	0.52
Number of obs.	215495	215495	215495	215495	215495
Standard errors	Robust	Robust	Robust	Robust	Clustered

Notes: The same ordinary least squares regressions across city pairs, but letting the dependent variable be an indicator of the two cities belonging to the same state in 2000, rather than 1900. All specifications include fixed effects for origin and target city. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.11: Same State Outcomes across City Pairs: Euratlas Data in 2000.

	Dependent variable is the Same-State Indicator Today				
	(1)	(2)	(3)	(4)	(5)
Length of Corridor	-0.244^{***}	-0.245^{***}	-0.246^{***}	-0.238^{***}	-0.238^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.012)
Fra Sea along Corr.	-0.290^{***}			-0.401^{***}	-0.401^{***}
	(0.002)			(0.002)	(0.021)
Fra Mountain along Corr.		-0.086^{***}		-0.425^{***}	-0.425^{***}
		(0.006)		(0.006)	(0.056)
Fra Marsh along Corr.			2.537^{***}	-1.239^{***}	-1.239
			(0.289)	(0.317)	(1.182)
R^2	0.51	0.48	0.48	0.52	0.52
Number of obs.	223445	223445	223445	223445	223445
Standard errors	Robust	Robust	Robust	Robust	Clustered

Notes: The same ordinary least squares regressions across city pairs, but letting the dependent variable be an indicator of the two cities belonging to the same state according to modern borders from GADM. All specifications include fixed effects for origin and target city. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.12: Same State Outcomes across City Pairs: GADM Data.

	The dependent variable is the Same-State Indicator for different years									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Length of Corridor	−0.143***	−0.253***	−0.156***	−0.196***	−0.108***	−0.110***	−0.139***	−0.102***	−0.093***	−0.213***
	(0.023)	(0.023)	(0.024)	(0.022)	(0.017)	(0.022)	(0.021)	(0.021)	(0.024)	(0.019)
Fra Sea along Corr.	−0.310***	−0.195*	−0.175	−0.062	−0.305***	−0.284***	−0.371***	−0.264***	−0.280***	−0.338***
	(0.099)	(0.101)	(0.108)	(0.113)	(0.079)	(0.094)	(0.089)	(0.089)	(0.096)	(0.085)
Fra Mountain along Corr.	−0.167	0.270*	−0.399***	−0.335**	−0.887***	−0.299*	−0.429***	−0.232*	−0.267*	0.043
	(0.147)	(0.141)	(0.140)	(0.133)	(0.097)	(0.159)	(0.158)	(0.123)	(0.144)	(0.110)
Fra Marsh along Corr.	−13.269***	−0.539	−12.526***	−9.783**	−15.254***	−11.790***	−14.653***	−7.017***	−8.227***	−2.736
	(3.780)	(4.220)	(4.094)	(4.207)	(3.542)	(3.551)	(3.666)	(2.448)	(2.801)	(2.093)
R^2	0.49	0.56	0.53	0.53	0.55	0.57	0.60	0.78	0.77	0.87
Number of obs.	861	946	861	903	903	703	780	666	561	703
Year in which state borders are measured	900	1000	1100	1200	1300	1400	1500	1600	1700	1800

Notes: Ordinary least squares regressions with robust standard errors in parentheses. The unit of observation is a pair of cities that existed both in 800 CE and in the year for which state borders are measured. The dependent variable is an indicator for whether the two cities belonged to the same state in the years indicated. The independent variables of interest measure the fraction cells with sea, mountains (above 800m), or marshes along a 300km buffer zone around the shortest distance line between the pair of cities. All specifications include fixed effects for origin and target city. * 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.

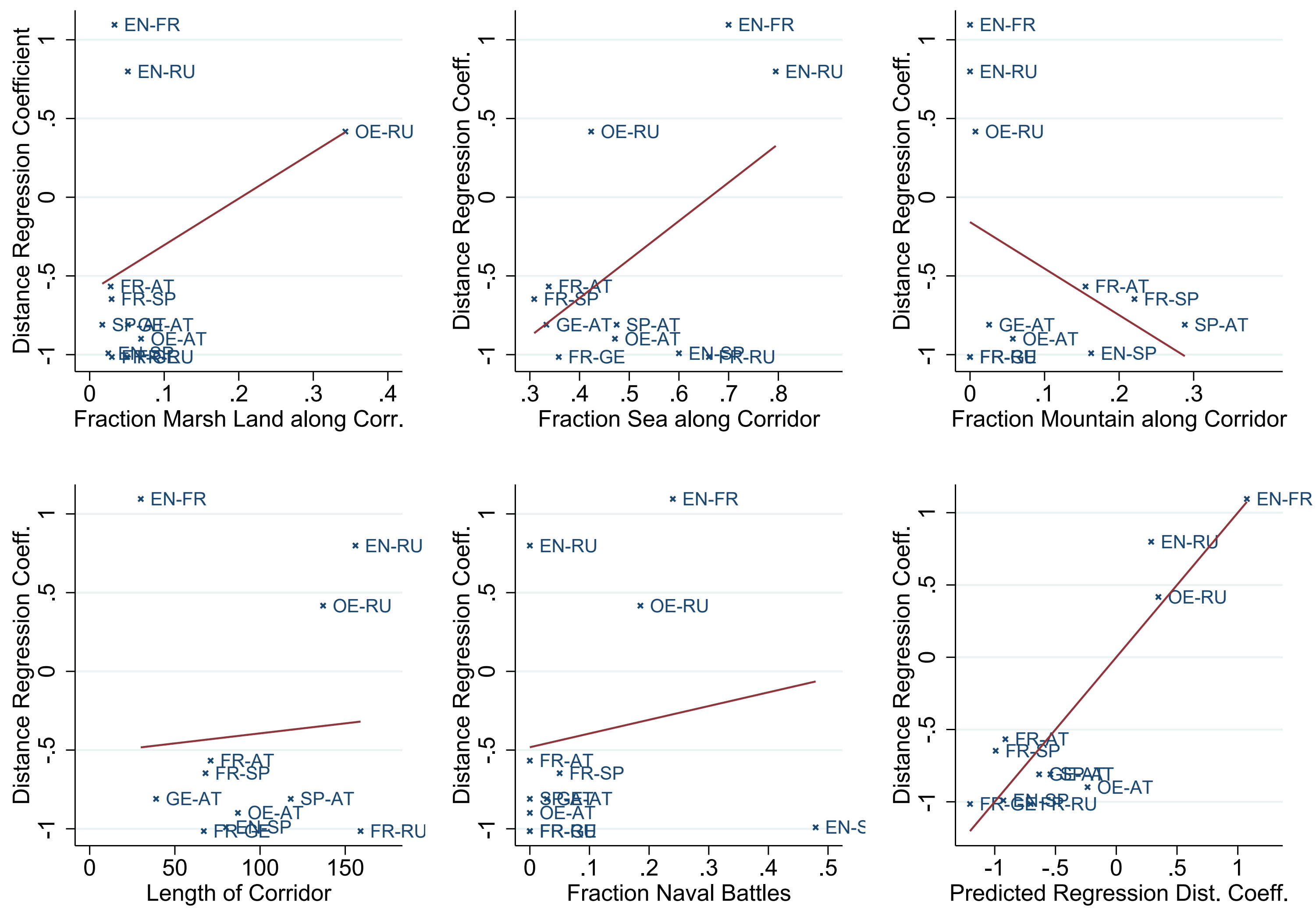


Figure A.1: The coefficients in Table A.1 and geography.

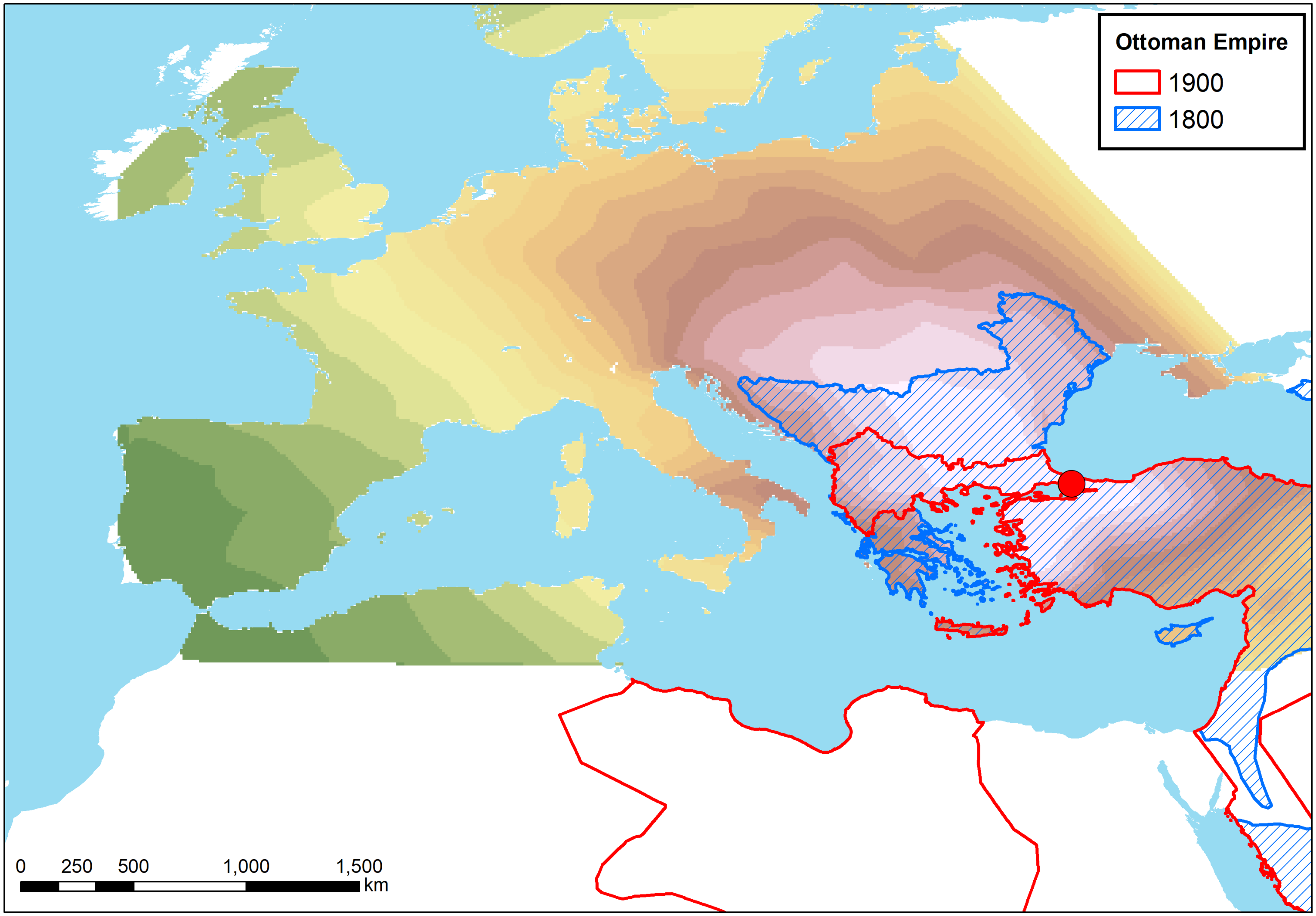


Figure A.2: Heat maps for simulated state territories around Istanbul, similar to Figure 5, but showing the actual borders of the Ottoman Empire in both 1800 and 1900.