**Measuring Space: A Minimum-Distance Database and Applications to International Studies**

Kristian S. Gleditsch; Michael D. Ward

*Journal of Peace Research*, Vol. 38, No. 6. (Nov., 2001), pp. 739-758.

Stable URL:
http://links.jstor.org/sici?sici=0022-3433%28200111%2938%3A6%3C739%3AMSAMDA%3E2.0.CO%3B2-A

*Journal of Peace Research* is currently published by Sage Publications, Ltd..

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/sageltd.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org
Thu Jun 7 17:07:25 2007

journal of PEACE RESEARCH

© 2001 Journal of Peace Research,
vol. 38, no. 6, 2001, pp. 739–758
Sage Publications (London, Thousand Oaks, CA and New Delhi)
[0022-3433(200111)38:6; 739–758; 020475]

SPECIAL DATA FEATURE

# Measuring Space: A Minimum-Distance Database and Applications to International Studies*

KRISTIAN S. GLEDITSCH
*Department of Political Science, University of California San Diego*

MICHAEL D. WARD
*Department of Political Science, University of Washington & Espace Europe, University of Pierre Mendès France*

In this overview of a new database and approach to measuring distance among historical and contemporary independent nation-states, we review the utility of space to theory and empirical research in international studies. We identify weaknesses in existing empirical data on distances and contiguity among nations. Categorical data on distance treat proximity as an either–or issue and do not permit identifying degree of proximity among states. Continuous measures of distances between midpoints, such as capital cities, often overstate the actual distances between state borders and suffer for large states and irregular territories. We outline a new alternative approach, based on measuring the minimum distance for pairs of polities in the international system, which remedies some of these shortcomings. The current implementation of the minimum-distance database includes the minimum distances for all polities within 950 km of each other from 1875 to the present. We demonstrate the enhanced flexibility of the new minimum-distance approach relative to existing alternatives. Moreover, we illustrate how variables constructed from distance measures, combined with spatial statistical techniques, can contribute substantively to international relations and cross-national comparative research. We demonstrate the importance of dependence among geographical neighbors by examining the link between levels of economic wealth and prospects for democracy in the context of regional interdependence among states.

## Introduction: Distance in World Politics

Distance is widely acknowledged to be a primary force shaping the opportunity for interaction among states in the international system. The opportunity to cooperate, exchange, and impose sanctions will generally decline the greater the distance between countries. Similarly, the sensitivity and vulnerabilities of countries to other actors and institutions in the global arena are also affected by closeness, as proximity shapes actors' incentives to engage in various conflictual and cooperative behaviors.

* This article builds on research presented at the 1999 Annual Convention of the International Studies Association in Washington, DC, and the 1999 Annual Meeting of the Mid-West Political Science Association in Chicago, IL. We are grateful for comments from Sherry Bennett, Peter Brecke, Steve Chan, Andrew Enterline, Scott Gates, Nils Petter Gleditsch, Gary King, Mohan Penubarti, Richard Tucker, Shira Yoffe, and two anonymous reviewers. This research was supported by a grant from the Center for Statistics and Social Science at the University of Washington, Seattle. Gleditsch is grateful for support from the Human Security Project at Harvard University. The Perl code and data described in this article can be downloaded from http://weber.ucsd.edu/~kgledits/ and from http://faculty.washington.edu/mdw/replicate/.

Some examples may help underscore the theoretical importance of distance and proximity in different areas of contemporary world politics. Recent events in Central Africa indicate how distance is critical for understanding how internationalized conflicts at the intersection of interstate and civil wars evolve and diffuse. In July 1994, the aid of neighboring Uganda played a critical role in the defeat of the Hutu-dominated Rwandan government by the Tutsi rebel group Rwandan Patriotic Front (RPF). The civil war and genocide preceding the takeover made approximately 2 million Rwandan refugees flee to neighboring Burundi, Tanzania, Uganda, and Zaire. The new Rwandan regime change provided critical support facilitating the defeat of Zairian President Sese Seko Mobuto in 1997 by Laurent Kabila. Rwanda and Uganda later withdrew their support for Kabila, and a number of other neighboring states are currently involved in the civil war in the former Zaire, now the Democratic Republic of Congo. Although often considered a 'civil war', the Congo conflict has obvious international dimensions. States are sensitive to what takes place in neighboring states, and vulnerable to the actions of others. Although civil wars are often analyzed as interactions between governments and insurgents alone, any analysis of these events must also take into account the interaction *among* states and how proximity and distance shape their incentives, resources, and constraints (see McNulty, 1999; *The Economist*, 1996).

In August 1989, the decision of the new governments in Hungary and Czechoslovakia to permit East German nationals to leave to the West through their territories generated a mass exodus from the German Democratic Republic. These events gave a considerable boost to the New Forum opposition movement and the large anti-government demonstrations in Leipzig challenging the socialist state in September. Eventually, the magnitude and scale of the popular protests, along with pressure from Gorbachev, forced the resignation of the head of the East German Politburo, Erich Honecker, on 18 October. Although the process of democratization in Eastern Europe is the outcome of a complex interplay of domestic and international processes that are not necessarily spatially determined, distance and proximity nonetheless exert a strong influence on how these processes unfold (see Gleditsch, 2000; Kuran, 1991; Lohmann, 1994; O'Loughlin et al., 1998; Randle, 1991).

Research on economic performance has shown that there appear to be fairly marked differences in growth rates between geographical regions. Whereas countries in East Asia have enjoyed annual growth rates exceeding 5%, the average real GDP per capita of Africa did not grow at all over the 1965–90 period. In empirical work, researchers have addressed this by inserting various regional dummy variables. But even though the Africa/East Asia coefficient estimates are significantly negative/positive in most studies (e.g. Barro, 1991), such regional fixed effects models tell us little about *why* Africa is lagging behind or *why* East Asia does better than other areas (Collier & Gunning, 1999). Ades & Chua (1997) suggest that such concentrations stem from systematic contagion across national boundaries, where the favorable or unfavorable characteristics of neighbors can exert a strong influence on a country's long-run growth rate. Easterly & Levine (1998) show that the fixed Africa effect disappears once the growth rates of a state's neighbors are taken into account, indicating that bad economic policies may create important externalities and that unfortunate outcomes may spill over into neighboring states. Lopez-Baso et al. (1999) similarly find strong evidence for economic convergence in the European Union playing out as a spatial process.

The increasing theoretical attention to

distance and proximity increases our demand on data for empirical work. We contend that distance and distance-based measures of interaction networks and interaction opportunities are important for advancing empirical research in international relations and comparative politics. In this article, we present an overview of the strengths and weaknesses of extant empirical data on distance among countries, and introduce a new database on the minimum distances between states in the modern international system, a database that remedies some of the principal problems with existing data.

Much existing empirical research incorporates distance essentially as a control variable. Although our database certainly can provide better control variables than many existing alternatives, we also believe that research in international relations and comparative politics can benefit from greater attention to investigations of spatial processes (see Anselin, 1988; Cressie, 1991; Ripley, 1988). Even though world politics by definition *is* the study of interaction and dependence between actors, much empirical research considers only very limited forms of interdependence. Despite the widely held belief that transnational processes increasingly permeate the boundaries between the international and domestic arena, most analytical and empirical perspectives in comparative cross-national research and international relations still effectively treat states and entities as if these were completely independent of one another.

## Empirical Data on Distance

Several efforts to derive measures of distance between states in the international system already exist. The existing data can vary considerably in their organization as well as in level of measurement or granularity of the individual data entries. Conceptually, the most fundamental distinction is whether the approach is based on measuring (a) proximity or distance between borders and edges of countries or (b) distances between the midpoints of entities. Accordingly, we categorize existing measures as (a) *measures of contiguity* or (b) *continuous distance measures*. In this section, we describe the defining characteristics of these two approaches and briefly discuss their primary strengths and weaknesses.

### *Contiguity or Measures of Proximity*

The concept of *contiguity* is commonly used in international relations, and is sometimes treated as it if were an unambiguous characteristic of a dyad or pair of states. The term can be interpreted in two slightly different ways. *Strict contiguity* implies that entities are physically contiguous in the sense that they share some form of land border. In principle, it is straightforward to establish whether two entities are physically contiguous in terms of land borders. However, contiguity is at the same time often employed in a looser sense to denote that two entities are *proximate* or somehow 'close' to each other, even if not strictly physical contiguous through a shared land border.

Although the strict interpretation of contiguity has the virtue of being reasonably clear and easy to measure, very few researchers in practice restrict contiguity exclusively to direct physical contiguity.[1] Many states that are geographically proximate, and have high levels of interaction, may not have a shared border (e.g. Luxembourg and The Netherlands). Furthermore,

[1] The research project on the diffusion of democracy at the Institute of Behavioral Science at the University of Colorado (O'Loughlin et al., 1998) initially relied on a strict algorithm that checked whether there were physical land boundaries among countries. These data were generated from computerized maps by an algorithm that examined intersections of polygons on a world map as displayed in ARC/info. A modified version of these data, with a fair bit of manual tweaking added, is available online at http://www.colorado.edu/IBS/GAD/spacetime/data/Cmats.html.

many states are separated merely by minor stretches of water, as is the case for example with the USA and Cuba, El Salvador and Nicaragua, or Sweden and Denmark. In addition, many countries have disjoint enclaves (e.g. Cabinda and Kaliningrad) or dependencies separated from the core territory of the state. The strict concept of physical contiguity seems an overly restrictive approach to assessing proximity, and does not correspond to the theoretical concepts of closeness or connectedness that researchers generally are interested in. A strict interpretation of contiguity is in practice rarely intersubjective either, as different researchers differ in their opinions over whether certain pieces of territory should be considered part of the core territory of the state or separate possessions.

Most data on contiguity relax the requirement that the entities must be strictly physically contiguous. The problem of assessing proximity beyond physical contiguity is often tackled by coding contiguity on the basis of a summary evaluation of individual cases or pairs of states (see, for example, Gibson & Ward, 1992; Gleditsch, 1996). However, in the absence of clear criteria and operationalization of the concept of contiguity, such approaches often have an arbitrary slant, as they are quite dependent on subjective judgments. Moreover, a coder's decision as to whether countries should be considered contiguous may be influenced by the actual amount of interaction between states rather than the geographical distance per se between entities. Coding contiguity on the basis of perceived interaction, however, makes it more difficult to assess empirically how distance in and of itself is related to opportunities for interaction. Furthermore, without clear, intersubjective criteria, it can become difficult to account for how the data are generated, and it may be near impossible to replicate the end result. Although these problems can be partly remedied by explicit documentation, an approach based on subjective judgments, without clear intersubjective criteria, thus seems inherently unsatisfactory.[2]

The best-known approach to measuring contiguity is undoubtedly the database generated by the Correlates of War project (we will refer to these data as COW contiguity). In addition to direct or physical contiguity, this database identifies several types of contiguities by water at distances of 12, 24, 150, and 400 miles or less by water.[3] These data suffer from many of the problems discussed previously. Unfortunately, the data have never been extensively documented (see Gochman, 1991, for the most complete summary in print). It is not clear exactly how these have been generated, and whether they have been sufficiently checked for consistency and problematic cases and omission. At the time of writing, the data are not publicly available and have not yet been updated beyond 1993. Several versions of the data seem to be circulating.[4] Paul F. Diehl of the University of Illinois is renovating and updating this database. However, improved documentation will not remedy the structural problems inherent in contiguity data of this type.

Contiguity data are by construction discrete or categorical, typically binary. A principal problem with all binary measures of distance is the essentially dichotomized form, where distance and proximity are rendered either/or questions. States are considered

[2] The diffusion of democracy project fixed such issues by manual intervention on an ad hoc basis as problems became apparent. Even though the algorithm approach initially seemed an attractive approach to systematic coding, the advantage over other efforts became considerably weaker as the end product was increasingly dependent on subjective judgments.

[3] The COW contiguity data reportedly also include an indirect contiguity category encompassing shared borders through colonies and other non-contiguous territories. This category is missing from many of the versions of the data currently circulating, and has not yet been updated (personal communication, Paul Diehl, October 2000).

[4] No version of the COW contiguity data is publicly available at the time of writing, pending the completion of the update. Older copies may be purchased from the Correlates of War project.

either contiguous or not, and the structure of the data does not permit differentiation between *degree* of proximity or closeness between entities. Categorized measures may provide some information about degree of proximity beyond dichotomous measures. Yet, given the relatively coarse and ordinal nature of the categories in existing data – see the five categories in the COW data – much of the useful information about degrees of proximity has already been aggregated and collapsed in the process. As such, fully continuous alternative data on distances hold a strong attraction.

### *Continuous Measures of Distance*

Whereas measures of contiguity focus on whether the borders of countries are 'close', the continuous alternatives calculate the distance between some midpoints for two entities. It is relatively easy to calculate the exact distance between entities given data on the latitude and longitude of the two midpoints. If one assumes that the Earth is a perfect sphere, with no surface irregularities (i.e. disregarding mountains and valleys), the great circle distance can be calculated using the Haversine Formula, where the distance between two locations $x$ and $y$ is given by

$$R \times 2 \times \arcsin\left\{ \min\left( 1, \sqrt{\left( \sin\frac{(A_y - A_x)}{2} \right)^2 + \cos(A_y) \times \cos(A_x) \times \left( \sin\frac{(O_y - O_x)}{2} \right)^2} \right) \right\} \quad (1)$$

where $R$ is the number of distance units per Radian (i.e. R = 6,367 km = 3,956 miles), $A$ is the latitude, and $O$ is the longitude for two entities subscripted $x$ and $y$, respectively.[5]

Researchers have proposed various suitable candidates for the center or midpoints of entities. One common choice of midpoint is the capital city. The capital city of a state can be considered a 'center' in the sense that it usually tends to be the most important city of a country. Measures of the capital city also have the advantage of being relatively well defined, and the exact coordinates can easily be located from the index of any map or available database.[6]

A slightly more complicated alternative is to derive a measure of the *centroid* or the center-most point within an entity considered as a polygon. Lewis F. Richardson (1942) suggested a variety of different approaches to derive measures of the midpoint of differently shaped polygons. Such approaches are widely employed in geography and regional science, but have not seen much use in international relations (though see Vanzo, 1999).

Another attractive alternative midpoint measure for many purposes would be the so-called population center based on the population distribution of a state. This is calculated annually for the USA. The US Statistical Yearbook includes a map of the population center, which indicates that the population center of the USA has gradually shifted further west. In principle, it would be possible to calculate such measures for other countries. A research group at the National Center for Geographical Analysis (Tobler et al., 1995) has created a 'gridded' population dataset in which latitude/longitude quadrilaterals are used as the units of observation for population information.[7] Such information is currently not available for historical polities and national borders, however, and at present

[5] The assumption that the Earth is a perfect sphere is accurate enough for most applications. Corrections using an oblate ellipsoid allow precision of 1% in 800 km. Ellipsoid software can be found at http://kai.er.usgs.gov/ftp/PROJ.4/proj.html. Spherical earth computations generally underestimate distances measured in the direction toward the equator and overestimate those measured in the direction of the poles. Although care must be used in implementing these calculations to ensure that enough precision is employed to allow discrimination between small degrees, this is simple to calculate.

[6] Data on capital city locations and distances between them are available from http://weber.ucsd.edu/~kgledits/. A World Cities database with latitude and longitude of more than 2,500 cities around the world may be purchased from ESRI (http://www.esri.com/data/online/).

[7] These data are available at http://www.ciesin.org/datasets/gpw/globldem.doc.html.

it is not feasible to calculate the population center for states back to the 19th century.

Although these approaches allow for more finely grained measures of proximity or closeness between states, continuous measures of distance based on midpoints are not devoid of ambiguities for international relations research. Both the capital city and centroid approaches suffer for large states. A midpoint based on the capital city could be located quite far from the boundaries of entities. The sizeable distance between Washington, DC, and Mexico City clearly understates the long shared border between the USA and Mexico. Furthermore, some countries (e.g. Bolivia) have more than one capital city. In addition, capital cities may change over time – Kazakhstan, for example, moved its capital from Almaty to Astana/Akmola, 1,200 km further north in 1997. Measures of other types of midpoint, such as centroids, do not solve these problems for large countries. Fort Hays, KS is the centroid of the USA, but still some distance from the centroid of Mexico or Canada. In addition, the centroid approach suffers for strangely shaped countries, where the centroid may fall outside the boundaries of the country, as is the case for the contemporary configuration of Croatia, Israel, Norway, and Somalia.

### *Hybrid Measures of Distance and Contiguity*

Some researchers have proposed measures combining information on physical contiguity and distance between midpoints. The best-known example is probably the data available from the EUgene program developed by D. Scott Bennett and Allan Stam.[8] EUgene generates measures of bilateral distances between the capital cities of states, replaced with distances of 0 if dyads are deemed 'contiguous' in the COW project data. Although the idea of combining the advantages of contiguity and distance approaches in a single database is attractive, we believe that the solution provided in EUgene ultimately suffers from many of the problems discussed above. The implied difference in distance between countries that are strictly contiguous and those that are separated by minor distances between borders can be analytically misleading. Even though the USA and Cuba are not strictly contiguous, the 1,800 km between Havana and Washington, DC, understates the fact that Cuba is only 110 km off the coast of Florida.

[8] EUgene is available at http://www.eugenesoftware.org.

## The Minimum-Distance Data

We have developed a different approach to measuring distance systematically, which we believe remedies some of the principal shortcomings of existing approaches. Briefly stated, we have opted for the basic principle of measuring the shortest distance between the two closest physical locations for every pair of independent polities between 1875 and 1996. The database records the shortest distance in statute kilometers between points on the outer boundaries for two polities, regardless of whether the states are separated by land or sea, given the borders in place in a particular time period. Actual shared borders are coded as having distances of zero. In our view, this database provides an integrated framework that combines the advantages of binary contiguity data and more disaggregated continuous distance data.

In the longer run, we will create a code and software that will calculate the minimum distances between all polities since 1875 to the present using digitized maps. A full implementation of this project, however, is complicated by a number of factors. Digitized global coverage is required that takes into account all the changes in the borders of polities over history. At this point, we have historical coverage for the period after World War II, created in the context of the diffusion of democracy project

(O'Loughlin et al., 1998). However, digitizing historical maps for 1875 and 1945 would require a substantial amount of additional work, given the myriad of territorial changes over the period. In addition, the full implementation of such a project would require funds to cover costs such as licenses for expensive GIS software that we do not currently possess. We hope to undertake this effort in the future, but for the time being the full implementation remains beyond our reach.

In the interim, however, we have settled for a simplified feasible approach that nonetheless improves upon existing alternatives. We record the minimum distances for all proximate pair states within a relatively comprehensive threshold of 950 km within the time period 1875–1998.[9] More specifically, we have calculated the minimum distances by manual inspection for each pair of countries by determining the two closest points on their outer boundaries. If the distance between these points is 950 (statute) kilometers or less, we record the number of statute kilometers. Actual shared borders are coded as having distances of zero.

Ideally, we would want to know the minimum distances or distance between closest points for all pairs of polities. However, the task of measuring all pairs of states by manual inspection, regardless of distance, quickly expands to become nearly insurmountable. There are about 171 independent states in the current international system, excluding those that have fewer than 250,000 inhabitants (Gleditsch & Ward, 1999). This implies that we would have to determine the minimum distances for $\frac{171 \times (171 - 1)}{2} = 14,535$ pairs of states to generate data for all the possible pairings. The composition of states in the international system and the borders of states have also changed considerably from 1875 to the present. As a result, to derive minimum distances over time we would have to make new comparisons for a number of historical maps. Moreover, for states that are relatively far apart, the search processes to determine which two points actually are the closest can be cumbersome. The risk of measurement error by manual inspection increases rapidly with the complexity of the measures.

For most practical applications, however, it is not necessary to know the minimum distances between all polities. States that are more than 1,000 km apart can hardly be considered geographically close. By limiting ourselves to comparisons of pairs of states that are 950 km or less apart, we greatly reduce both the number and the complexity of the required comparisons. At the same time, having continuous minimum distances, instead of discrete data, allows us to vary the threshold for proximity from a relatively broad notion of region (i.e. a high threshold such as 950 km) to strict physical contiguity (i.e. 0 km). We return to the advantages and added flexibility of the minimum-distance approach over other approaches in more detail in the next section.

Beyond the issue of what to measure, we also had to make a number of other substantive decisions. We included most reasonably proximate enclaves disjoint from the core territory of the state, such as Alaska, Cabinda, Kaliningrad, and West Berlin. However, we did not include remote dependencies far from the core territory of the state, even if these are considered fully integrated parts of the possessing state. Accordingly, we do not consider Saint Pierre and Miquelon in considering distances between France and Canada or the USA.[10]

[9] These measures were carried out in Microsoft® Encarta® Virtual Globe (1998 and 1999 editions).

[10] A number of quite tricky questions remain here. We have opted for not including Algeria as part of France, even though this was considered a regular French department up to 1962, and included as part of NATO on maps in old editions of the *NATO Handbook*. Future versions of software to generate the data from the base information, however, will allow the user to modify such decisions as they find appropriate.

Furthermore, the composition of entities in the system and their boundaries are not constant, but change over time. We have outlined the set of independent countries since 1816 in an earlier article (Gleditsch & Ward, 1999).[11] In the context of the minimum-distance database, we have dealt with the problem of changing maps by measuring the distances between the actual pieces of territory, or polygons in GIS terminology, that have shifted between states. The distances between historical and contemporary polities at any given point of time can thereby be calculated using the appropriate unions of such polygons, given by the physical borders of the actors in the system at this particular time. The following list is a summary of the most important border changes that have taken place since 1875.[12]

- Up to 1878, the *Austro-Hungarian Empire* included, in addition to the territories of contemporary Austria and Hungary, also present-day Slovenia, Croatia, the Czech and Slovak republics, as well as parts of Italy and Poland. Croatia, Slovenia, and Bosnia were part of the Austro-Hungarian Empire between 1878 and 1919.
- *Germany* experiences various border changes after 1918. After its defeat in World War II, a considerable share of its territory was annexed by Poland and the Soviet Union, and the remainder of the country itself was split into the German Federal Republic and the German Democratic Republic. The German Democratic Republic merged with the Federal Republic in 1991.
- *Pakistan*, until 1971, encompassed both present-day Pakistan and the territory of East Pakistan, which became an independent polity under the name of Bangladesh in 1972.
- *Romania* between 1887 and 1945 included present-day Moldova.
- *Russia/Soviet Union* has undergone several major border changes since 1875. Imperial Russia before 1917 included the territories of Armenia, Azerbaijan, Belarus, Estonia, Finland, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Tajikistan, Turkmenistan, Ukraine, and Uzbekistan, in addition to the present-day Russian Federation. Finland, Estonia, Latvia, and Lithuania acquired independence from the Soviet Union after the revolution in 1917. The latter three Baltic republics were incorporated in the USSR in 1940, as were Moldova and former parts of Germany, Poland, and Japan after World War II. In 1991, Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Tajikistan, Turkmenistan, Ukraine, and Uzbekistan all acquired independence from the Soviet Union.
- *Turkey* (and its predecessor, the *Ottoman Empire*) has varied considerably in extension since 1875. Before 1878, the Ottoman Empire included parts of present-day Romania, Macedonia, Moldova, Albania, Bosnia, Bulgaria, Serbia, and Cyprus, in addition to the territory of contemporary Turkey. Whereas many of these states acquired independence in 1878, present-day Albania remained part of Turkey until 1913.
- *Vietnam* was partitioned into the Northern People's Republic of Vietnam and the Southern Republic of Vietnam after independence from France. The Republic of Vietnam then became part of the

[11] The list of independent states since 1816, and additional documentation, is available on http://weber.ucsd.edu/~kgledits/.

[12] We recognize that some earlier transfers of territory and border changes may not have been taken into account in these data at this point. Although most of these would not affect the direct contiguities between actors, they could in many cases change the minimum distances to other noncontiguous states. Over time, we hope that new versions of this database will encompass all pertinent changes.

Democratic Republic of Vietnam following its defeat in 1975.

- In the case of *Yugoslavia*, the precursor Kingdom of Serbia included present-day Macedonia before 1878. Between 1919 and 1991, Bosnia and Herzegovina, Croatia and Slovenia all became part of the Kingdom of Yugoslavia and subsequent Republic of Yugoslavia. The former Yugoslav republics of Bosnia and Herzegovina, Croatia, Macedonia, and Slovenia proclaimed independence from Yugoslavia in 1991 and 1992.

The current version of the data is distributed through a Perl program that allows the user to customize the format and scope of the data.[13] The current options allow users to generate binary contiguity data or continuous minimum-distance data in a dyadic or matrix format, as well as set the time period and some output format options.

## Using the Minimum-Distance Database

In this section, we demonstrate how the minimum-distance data may be used in international relations and cross-national research, and illustrate some of its advantages over existing data. Our data can improve upon many of the alternative measures of distance or contiguity used as control variables or measure of 'opportunities' for interaction. However, if we believe distance shapes or underlies substantively important processes, or that the behavior or properties of proximate entities influence a country's behavior and realm of feasibility, we should try not only to control, but also to incorporate, the spatial aspects of processes of interest directly into variables or statistical models. In this section, we indicate how international relations research may go 'beyond geographical contiguity' and incorporate the substantive implications of distance more systematically with measures based on the minimum-distance data.

### *Varying Relevance Thresholds by Distance*

The main advantage of the minimum-distance database is that it can be used to derive continuous data on distances as well as to generate binary or categorized contiguity data. From the minimum-distance data, we can derive various inverse weights to model forms of distance decay. As minimum distances less than 950 km are recorded by the actual distance, the analyst can vary the cutoff points for what is to be considered contiguous and weight entries in proportion to their degree of closeness. Thus, unlike other fixed contiguity data, the minimum-distance database allows generating different sets of contiguity data with different minimum-distance threshold criteria. The relevant distance or metric for a given problem is typically not known with certainty in advance, and may conceivably differ quite considerably between research questions and over time. The flexibility of the minimum-distance data allows various respecification tests to probe the most appropriate criteria for a given problem.

Consider the information in Table I, which details the connectives for a number of polities in the Middle East. Scrolling the columns of this table – from left to right – for each row or state indicates the additional entities that would be adjudged connected as we expand beyond the strict definition of land borders in the second column to increasingly more inclusive criteria for proximity or closeness. As can be seen, there are many pairs of states that are not connected by a strict criterion of direct contiguity that would be considered connected with a

[13] The current distribution program requires Perl, a free programming language available for virtually any operating system and platform. Perl can be found at http://www.perl.com.

slightly more inclusive threshold. Although Israel, for example, does not have a land border with Iraq and Saudi Arabia, expanding the threshold to a higher cutoff value (e.g. 50 km or 475 km) makes these pairs connected. Different distance thresholds will be appropriate for different purposes. Unlike other fixed alternatives, the minimum-distance data allow researchers to set the threshold for relevance and test the sensitivity to differences in operationalization.

Do the new minimum distance data yield much new information compared to existing alternatives? The most common source of data on distance is the COW contiguity data. Merely a cursory comparison reveals that much of the information in the minimum-distance data is simply absent from the COW contiguity data. In the case of Israel, for example, the COW contiguity data merely indicate that Israel has a land border with Egypt, Lebanon, Jordan, and Syria, is separated by 12 miles of sea or less from Saudi Arabia, separated by somewhere between 12 and 150 miles of sea from Cyprus, and somewhere between 150 and 400 miles of sea from Turkey. The information on distance across sea is provided in fixed categories, and researchers cannot modify these categories and thresholds, as is the case for our minimum-distance data. More importantly, the COW data do not contain any form of information about proximity over land. In the case of Israel, it completely disregards the seemingly important fact that the closest point in Iraq is no further than 215 miles from Israel. That Iraq is not within 400 miles across sea from Israel poses no major obstacles for launching a Scud or an Exocet missile; nor has it deterred Israeli military and civil defense programs. Intermediate countries by construction disconnect many countries in the COW contiguity data that commonly would be considered 'close'. Landlocked countries such as Uzbekistan can by construction never be connected with other relatively close states such as Iran (minimum distance 244 miles).

The only alternative provided by the COW project for a researcher interested in data on distance not limited to strict land contiguity, or the four categories of connectedness by sea, is the so-called regions classification. This is essentially a binary indicator of membership in regions as mutually exclusive categories. At most, these data tell us whether certain states are in some predefined geographical region such as 'the Middle East' or not. There is inevitably some arbitrariness in delineating regions where states are classified as either members or nonmembers. In the COW data, the Middle East category also encompasses all states in North Africa, but excludes states in the Caucasus and Central Asia. Many countries, however, may be actors in more than one of these 'regions'. From Table I, it is evident that Turkey is close to states in the Middle East, the Caucasus, Europe, and Central Asia. We submit that a broader distance span around each individual country corresponds better to what we think of as its regional context or interaction environment than strict physical contiguities or proper name classifications of regions that force 'Europe' or the regional context of European states to stop at the Urals and the Strait of Bosporus (Lake & Morgan, 1997). The minimum-distance data allow identifying neighbors or connected states relative to each country rather than joint membership in geographical regions. Moreover, these data permit rendering variation across regions a variable that can be compared between units, rather than treating regional differences by proper names (Przeworski & Teune, 1970). In the next sections, we outline how spatial statistics can contribute to create variables reflecting differences in spatial context, and how this may be used in empirical analyses.

Table I. Connectivities for Middle Eastern States: States Connected to a Reference Country, Cumulative

| *Reference country* | *Land border* | *50 km* | *475 km* | *950 km* |
|---|---|---|---|---|
| Bahrain | Saudi Arabia | Qatar | Iran, Iraq, Kuwait, Oman, United Arab Emirates | Yemen |
| Egypt | Israel, Libya, Sudan | Saudi Arabia, Jordan | Lebanon, Syria, Cyprus, Greece, Chad | Iraq, Turkey, Eritrea, Ethiopia |
| Iran | Afghanistan, Armenia, Azerbaijan, Iraq, Pakistan, Turkmenistan, Turkey | Kuwait | Bahrain, Georgia, Kazakhstan, Kuwait, Oman, Qatar, Russia, Syria, Saudi Arabia, United Arab Emirates, Uzbekistan | Cyprus, India, Jordan, Lebanon, Ukraine, Tajikistan |
| Iraq | Iran, Jordan, Kuwait, Saudi Arabia, Syria, Turkey | | Armenia, Azerbaijan, Bahrain, Georgia, Israel, Lebanon | Cyprus, Egypt, Iraq, Kazakhstan, Oman, Qatar, Russia, Turkmenistan, United Arab Emirates |
| Israel | Egypt, Jordan, Lebanon, Syria | Saudi Arabia | Cyprus, Iraq, Turkey | Greece, Sudan, Libya |
| Jordan | Iraq, Israel, Lebanon, Saudi Arabia | Egypt | Cyprus, Turkey | Armenia, Greece, Iran, Kuwait, Sudan |
| Kuwait | Iraq, Saudi Arabia | Iran | Bahrain, Qatar | Jordan, Oman, Syria, Turkey, United Arab Emirates |
| Lebanon | Israel, Jordan, Syria | | Cyprus, Egypt, Iraq, Saudi Arabia, Turkey | Greece, Iran |
| Oman | Saudi Arabia, United Arab Emirates, Yemen | | Pakistan, Iran, Qatar | Afghanistan, Bahrain, India, Iraq, Kuwait, Somalia |
| Qatar | Saudi Arabia | Bahrain | Iran, Kuwait, Oman, United Arab Emirates | Iraq |
| Saudi Arabia | Bahrain, Iraq, Kuwait, Jordan, Oman, Qatar, United Arab Emirates, Yemen | Egypt, Israel | Djibouti, Ethiopia, Eritrea, Iran, Lebanon, Sudan, Syria | Armenia, Azerbaijan, Cyprus, Pakistan, Somalia, Turkey |
| Syria | Iraq, Israel, Jordan, Lebanon, Turkey | | Armenia, Cyprus, Egypt, Iran, Saudi Arabia | Azerbaijan, Georgia, Kuwait, Russia, Ukraine |
| Turkey | Armenia, Azerbaijan, Bulgaria, Greece, Georgia, Iran, Iraq, Syria | | Albania, Cyprus, Israel, Lebanon, Jordan, Macedonia, Moldova, Romania, Russia, Saudi Arabia, Yugoslavia | Bosnia, Czech Republic, Croatia, Hungary, Egypt, Italy, Kazakhstan, Kuwait, Libya, Poland, Slovenia, Turkmenistan |
| United Arab Emirates | Oman, Saudi Arabia | | Iran, Bahrain, Yemen, Qatar | Afghanistan, Iraq, Kuwait, Pakistan |
| Yemen | Oman, Saudi Arabia | | Ethiopia, Qatar, United Arab Emirates | Bahrain, Sudan |

### *A Distance-Based Network Representation*

Geography and regional science have developed methods for spatial statistical analysis with data on geographical units. Data on distance can be converted to a matrix representation of the relationships between entities that can be used in statistical analysis. For simplicity, in this example we consider the 15 states in the left column as a separate subsystem, ignoring for the time being linkages that these actors have to other states.[14] Moreover, we base our discussion on a binary matrix, even though applications could also conceivably be based on matrices with differing weights.

The information in the minimum-distance data as displayed in Table I can be represented numerically in the form of a spatial connectivity matrix that allows a series of useful operations.[15] Generally stated, for a set $S = \{1, \ldots, n\}$ of N spatial units, a connectivity matrix is a $n \times n$ matrix $\mathbf{W}$ whose entries $w_{i,j}$ acquire non-zero values if units $i$ and $j$ are somehow considered to be connected, adjacent, or associated. The binary connectivity matrix $\mathbf{W}$ is similar to the idea of an 'adjacency matrix' in graph theory (e.g. Harary, Norman & Cartright, 1965). Bavaud (1998) stresses the resemblance of connectivity matrices and Markov chains,

$$\mathbf{W} = \begin{array}{c} \begin{array}{ccccccccccccccc} \text{Bahrain} & \text{Egypt} & \text{Iran} & \text{Iraq} & \text{Israel} & \text{Jordan} & \text{Kuwait} & \text{Lebanon} & \text{Qatar} & \text{Oman} & \text{Saudi Arabia} & \text{Syria} & \text{Turkey} & \text{United Arab Emirates} & \text{Yemen} \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{array}{l} \text{Bahrain} \\ \text{Egypt} \\ \text{Iran} \\ \text{Iraq} \\ \text{Israel} \\ \text{Jordan} \\ \text{Kuwait} \\ \text{Lebanon} \\ \text{Qatar} \\ \text{Oman} \\ \text{Saudi Arabia} \\ \text{Syria} \\ \text{Turkey} \\ \text{United Arab Emirates} \\ \text{Yemen} \end{array} \end{array} \quad (2)$$

[14] This is for purposes of illustration only. In an actual analysis, we would obviously want to include linkages to other entities. This example is adapted from Gleditsch (forthcoming).

[15] This is sometimes referred to as a 'contiguity' or 'spatial weight' matrix (e.g. Anselin, 1988). Since the entries need not refer to contiguities, space, or weights, we prefer the more general term 'connectivity matrix'.

and sets out some important proofs on their properties.

Equation (2) is an example of a binary spatial connectivity matrix **W** for the Middle East based on land borders or a threshold for relevance of 0 km.[16] The binary spatial connectivity matrix **W** in (2) is a network representation of the linkages among all the 15 polities in the Middle Eastern subsystem. It not only describes the first-order contiguities, but also permits derivation of *nth*-order contiguities or the 'neighbors of neighbors' by powering the matrix.[17] In the next section, we outline how such a connectivity matrix can be used in empirical analyses.

### *Spatial Context as a Variable*

Many of the existing applications of spatial statistics in international relations have an exploratory character, and are based largely on descriptive statistics of geographical clustering at a local or global level of some attribute.[18] Examining geographical clustering in the distribution of a single variable may be interesting as a first cut at understanding the role of spatial clustering and diffusion. Ultimately, however, researchers usually want to evaluate the role of various geographical and non-geographical factors jointly. The role of geographical factors can be examined more systematically by rendering clustering and regional differences as a variable that can be used as either an explanatory variable or a response variable where we seek to account for the variation on the variable by other explanatory variables.

[16] From the measures of the distance between units in the minimum-distance data, we can generate a series of different sets of connectivity matrices at different distance thresholds. Though the choice of any single distance threshold may seem arbitrary, the minimum-distance data allow the researcher to vary the definition of context and conduct specification tests.

[17] More specifically, the entries in a $\mathbf{W}^n$ matrix indicate the number of $n$-order linkages between $i$ and $j$ (see Harary, Norman & Cartright, 1965).

[18] We present such measures in some detail in Gleditsch & Ward (2000).

The most common type of a spatially constructed variable for a unit $i$ is the local mean or the average of some input variable $x$ for all units $J$ deemed to be connected with $i$.[19] The resulting variable can be interpreted as a summary measure of the regional context of the variable $x$ for each unit $i$. To create such a regional weighted average, the binary spatial connectivity matrix is first *row-standardized* by dividing each row by the sum of the entries in the row. This yields a so-called row-standardized connectivity matrix $\tilde{\mathbf{W}}$, where all the entries of each row sum to 1. The mean or average over the regional context of the variable $x$, denoted $x^R$, can then be found as the product of $\tilde{\mathbf{W}}$ and the vector of the values of $x$ over the set $S = \{1, \ldots, n\}$ of the N spatial units.

Standardization of the binary connectivity matrix **W** has the advantage that the values of spatial variable $x^R$ will have the same metric as the original variable, and that each observation $x_i^R$ can be interpreted as a spatial mean or weighted average of all entities within the given unit $i$'s regional context. Although this is by far the most common form to derive spatially based variables, merely reverting to average values based on standardized matrices with zeros on the diagonal entries may not be warranted as a standard solution for all purposes. Other approaches can make more sense theoretically in particular settings. In many applications, the aspect of interest is not so much the *average* values of a variable among states in a region, but rather the total spill-in or sum of the activities reflected by the variable. In such cases, a spatial variable with the sum of $x$ for all entities connected to $i$ (i.e. $\mathbf{W}x^T$) might be a more appropriate measure. Enterline & Gleditsch (2000: 33–35), for example, use a variable indicating the sum or total spill-in of regional conflict for each

[19] Borrowing from the terminology of proximity for time-series analysis, this is commonly referred to as 'the spatial lag' of a variable $x$.

state as a measure of threat. Murdoch, Sandler & Sargent (1997: 283–285) examine how differences in the total sum of emissions from geographic neighboring countries vary the extent of collective action problems in achieving emission reductions among European states.

Although no research or applications, to our knowledge, have explored such alternatives as of yet, it might in many settings also make sense to devise spatial variable constructions based on moments other than the mean and the total sum. More generally, the appropriate spatial variable type depends upon the causal mechanism that we believe operates across space. The public goods literature has explored various alternative supply technologies for public goods or externalities that may be helpful for thinking systematically about choices for construction of spatial influence or spill-in variables (see, e.g., Mueller, 1989: 22–25). If the amount of an externality is determined by the individual units' total consumption within some geographical bounds, as in the 'tragedy of the commons', an additive form would be appropriate (see Cornes & Sandler, 1996: 492–494, for a more general discussion). In the case of a 'weakest-link'/'best-shot' technology, where the externality is determined by the smallest/largest contribution of any of the individual units' provision, useful spatial variables could be constructed from the minimum or maxi-mum of the values among contiguous entities. If we are interested in the difference or variation among entities within a region, we could use measures based on moments such as the range or variance of the values. Clearly, the range of opportunities for innovative research is wide open.

In some applications, it might be helpful to create variables based on matrices where the diagonal entries $i$ are assigned non-zero values.[20] Gleditsch & Ward (2000) included the values of the $G_i^*$ statistic of local clustering for war and democracy as right-hand-side variables indicating the strength of regional clustering in these variables around each country $i$. Gleditsch (2000) used a measure of the strength of the geographical clustering in successive years without conflict involvement on the core territory of states as a measure of the relative degree of regional peace.

### *Spatial Variables in Regression Analysis: A Simple Example*

A wealth of recent research in international relations focuses on the consequences of political institutions for international interactions. The most prominent example is undoubtedly the so-called democratic peace, or the finding that democracies do not appear to have gone to war against each other (see Chan, 1997; Russett, 1993). This research primarily assesses the international consequences of differences in political institutions, and takes the actual distribution of these political institutions itself as predetermined or given. Most existing theoretical and empirical work on democracy and democratization tends to emphasize domestic factors, such as social or economic 'requisites' (Lipset, 1960), or transitions as political processes (O'Donnell, Schmitter & Whitehead, 1986) playing out within each state in isolation (see Vanhanen, 1990, for a comprehensive review).

Much recent work, however, argues that international factors and diffusion processes among states can influence political institutions as well, and that many of the observed changes in political institutions cannot be

[20] Whereas the convention is not to include a positive entry on the diagonal entries or 'connect countries to themselves', nothing in the definition or mathematical properties of the connectivity matrix itself prohibits non-zero diagonal values. Note, however, that the analogy to Markov Chains will no longer hold for many such matrices.

Table II. Regression of GDP Per Capita on Level of Democracy, 1992

| | *Model 1: Ordinary Least Squares Estimator* | | *Model 2: Spatial Autoregressive Estimator* | |
|---|---|---|---|---|
| *Variable* | *Coefficient estimate* | *Standard error* | *Coefficient estimate* | *Standard error* |
| Constant | –0.3798 | 0.8814 | –0.5877 | 0.7954 |
| Regional democracy | – | – | 0.4077 | 0.1126 |
| GDP per capita | 0.0008 | 0.0001 | 0.0005 | 0.0001 |
| | | $n = 82$ | | $n = 82$ |
| | $F(1,80) = 43.95$, LR $\chi^2 = 34.66$. | | $F(2,79) = 31.36$, LR $\chi^2 = 47.48$ | |

attributed purely to domestic processes alone (Gleditsch, forthcoming; Huntington, 1991; O'Loughlin et al., 1998). Space limitations preclude a fuller discussion of the role of international factors in democratization processes. Our main concern here is to illustrate that if such diffusion processes operate among proximate countries, ignoring spatial dependence and treating the distribution of democracy as independent among countries might severely undermine our inferences about the effects of domestic factors.

Consider a simple example where a country's GDP per capita is regressed on its level of democracy, as shown in Table II.[21] The column labeled Model 1 gives the results of a conventional linear regression of GDP per capita on democracy. This model is similar to most empirical analyses of the relationship between democracy and economic development (e.g. Burkhart & Lewis-Beck, 1994), and the results suggest a positive relationship between wealth and democracy. The expected democracy in each country is assumed to depend only on the country's GDP per capita, and to be independent of the characteristics of neighboring states. The latter part of this assumption will be violated if regional diffusion processes play out among neighboring states.

Model 2, by contrast, incorporates the average level of democracy among neighboring states as an additional right-hand-side variable.[22] The coefficient estimate for regional democracy indicates a strong correspondence between the regime characteristics of a state and those of its neighbors. This suggests that democracies and autocracies tend to cluster geographically in different zones. Using the common threshold of a POLITY score of 6 or above as a threshold to distinguish between democracies and non-democracies, we find that over 70% of the polities are located in regional contexts similar to their own.[23] The association seems both substantively strong and highly statistically

[21] Our measure of democracy here is the 21-point institutionalized democracy scale in the POLITY data (Jaggers & Gurr, 1995). The data on GDP per capita are taken from the Penn World Tables (Summers & Heston, 1991).

[22] The spatial autoregressive model here is given by $y = \rho \tilde{\mathbf{W}} y + \mathbf{X}\beta + u$, where $y$ is a $n \times 1$ vector of levels of democracy, $\tilde{\mathbf{W}}$ is an $n \times n$ row-standardized spatial connectivity matrix whose entries $\tilde{w}_{i,j}$ acquire a non-zero value if units $i$ and $j$ are within 950 km of each other, and $\mathbf{X}$ is an $n \times 2$ matrix of the other right-hand-side variables (i.e. a column of 1s for the constant term and a column of levels of real GDP per capita). OLS is inconsistent, given the presence of $\tilde{\mathbf{W}}y$ on the right-hand side, and the model must be estimated by maximum likelihood.

[23] Of a total of 82 countries in the sample, 26 are democracies located among other democracies, 32 non-democracies located among non-democracies, 3 non-democracies in a democratic regional context, and 21 democracies located in regions where states on average are non-democracies.

significant. A formal test that the coefficient for the regional context of democracy is zero yields an F = 9.78, and we can clearly reject the restriction in Model 1.[24]

In addition, we can see that the estimated coefficient for the impact of per capita GDP on democracy changes quite a bit once the spatial dependence among countries is taken into account. More specifically, although the coefficient estimate of GDP per capita remains positive and significant, the estimate is reduced to about two-thirds of its original size in Model 1 when the regional context of democracy is taken into account.[25]

Dependence between successive observations in space will give rise to serially correlated disturbance, with implications similar to dependence between observations over time, and the OLS standard errors will be incorrect in the presence of spatially correlated disturbances. Although the OLS estimator formally remains unbiased under serial correlation, the variance of the coefficient estimates will be higher than it would be if the errors were random, and the standard errors of the coefficient estimates will not be correct. In a time-series setting, including a lagged dependent variable may suffice to get unbiased standard error estimates (Beck & Katz, 1995). Introducing a spatial variable to emulate the spatial dependence between observations, however, does not resolve all the problems with the standard error estimates, and special estimators should be used for these kinds of models. A didactic summary of the statistical issues of research recognizing the spatial dependence is beyond the scope of this article.[26] As the example above shows, however, the analytical costs of disregarding the structure of dependence between entities can be substantial in social science applications. The minimum-distance data facilitate taking spatial dependence into account in empirical work.

### *The Mediation of Distance*

In international studies, a number of issues other than distance itself mediate the structure of dependence between states and actors. Factors such as resources, technological capabilities, and ideological disposition may also influence the opportunity and willingness of a state to engage in certain types of behavior, such as waging conflict. States with a large resource base and advanced technological capabilities have greater abilities to act far beyond their borders (e.g. Boulding, 1963), as witnessed in conflicts such as the Falklands and Gulf wars. States also differ in their international orientation and ambitions. Whereas some states, such as Cuba, place considerable emphasis on active involvement in international affairs far beyond their borders (as witnessed by the involvement of Cuban troops in many African conflicts), states such as Albania under Enver Hoxha and contemporary Myanmar seek to minimize any relations with neighboring states (e.g. Braumoeller, 1997). Geographical distance as a barrier or constraint is also mediated by factors such as type of terrain and existing infrastructure (e.g. Lemke, 1995; Starr,

[24] Raftery (1995) advocates evaluating two models corresponding to competing hypotheses through Bayes factors, or the ratio of the posterior odds for one of the models against the other. He develops a Bayesian Information Criterion (BIC) approximation to the Bayes factors that has gained widespread use. For some model $M_k$, we can find an approximation $BIC'_k = -\chi^2_{k0} + p_k \log n$, where $\chi^2_{k0}$ is the likelihood ratio statistic for testing the null model $M_0$ against $M_k$. The Bayes factor approximation is simply the difference between the BIC for the two models. In this case, $BIC'_1 = -25.85$ and $BIC'_2 = -34.26$. The difference (8.42) is in the 6–10 range that Raftery (1995) characterizes as providing 'strong' evidence for the superiority of the model with the larger *BIC*. This test thus clearly indicates that superiority of the spatial model over the non-spatial version.

[25] One reviewer expressed concerns that the differences between models 1 and 2 might be due to the estimator rather than the addition of the spatial variable. The coefficient estimates are substantially similar when model 2 is estimated by OLS; 0.1066 + 0.6270 *(regional democracy) + 0.0004 * (GDP per capita).

[26] We refer readers to Anselin (1988), Cressie (1991), and Ripley (1988) for general overviews as well as the applied articles cited earlier.

2000). Traversing flat plains and steppes is generally easier than dense rain-forest or mountain ridges. Troops and supplies can be moved much faster in areas with existing roads, airfields, and railroads in working order.

These considerations are all important theoretical questions about the role of distance for a particular issue area such as the study of international conflict. Many researchers have sought to create measures that take into account the influence on distance mediated by other factors. Examples include various sampling strategies for politically relevant dyads – where it is deemed war could possibly occur (e.g. Maoz & Russett, 1993) – or the politically relevant interaction environment of states (e.g. Maoz, 1998), measures of relevance taking into account terrain, technology, and infrastructure (e.g. Lemke, 1995; Starr, 2000), or measures of ability to project power based on exponential decay with distance (e.g. Bueno de Mesquita, 1981).

More specialized data on relevant distance, mediated by other factors, can provide useful additional information beyond raw geographical distance appropriate for many settings. More generally, however, the factors mediating distance will differ between research questions. In our view, it is essential to ensure that data collection on geographical distance remains disaggregated from data on other factors hypothesized to mediate the relevance of distance. Mixing distance and other factors in the data collected makes it difficult to separate analytically any effect of geography and distance in and of itself, as well as to examine how the implications of distance may be modified by other factors for some research question. Not one set of considerations applies universally, and imposing the expected influence of some factor on distance in collecting data on geographical distance reduces the general applicability of the data. Improved data on geographical distance, however, will provide a better basis for other data-collection efforts combining information on distance and other factors. Our overview has focused on spatial statistics in the context of the minimum-distance data and how this may be applied. However, the concepts and methods of spatial statistical analysis need not be limited to geographic or Euclidean distance. The structure of connectedness between states may be specified on the basis of factors such as power or observed indicators of interaction such as trade (see Ward, Heagerty & Gleditsch, 2001).

## Space, a Final Frontier?

In this article, we have presented our new data on minimum distances. We have briefly shown how these can allow us to go beyond distance as a control variable and consider dependence among observations explicitly. We recognize that the existing minimum-distance data by no means are perfect. Neither will they be useful in every plausible research question. Nonetheless, we hope the new database will be useful to a wide range of researchers. Finally, we hope to have convinced our readers that measures of distance can substantially and substantively contribute to the study of regional context in international relations and comparative politics. To conquer the final frontier requires advances in our attention to distance and space in theory, as well as improved empirical measures and methods. Most theories of international and comparative politics recognize that political actions, and indeed actors, are *interdependent*. It is time our empirical analyses reflected this basic fact of politics rather than ignored it.

## References

Ades, Alberto & Hak B. Chua, 1997. 'Thy Neighbor's Curse: Regional Instability and Economic

Growth', *Journal of Economic Growth* 2(3): 279–304.

Anselin, Luc, 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.

Barro, Robert, 1991. 'Economic Growth in a Cross Section of Countries', *Quarterly Journal of Economics* 106(2): 407–443.

Bavaud, Francois, 1998. 'Models for Spatial Weights: A Systematic Look', *Geographical Analysis* 30(1): 153–171.

Beck, Nathaniel & Jonathan N. Katz, 1995. 'What To Do (and Not To Do) with Time-Series Cross-Section Data', *American Political Science Review* 89(3): 634–647.

Boulding, Kenneth E., 1963. *Conflict and Defense: A General Theory*. New York: Harper and Row.

Braumoeller, Bear F., 1997. 'Isolationism in International Relations', paper presented at the 38th Annual Convention of the International Studies Association, Toronto, 18–22 March.

Bueno de Mesquita, Bruce, 1981. *The War Trap*. New Haven, CT: Yale University Press.

Burkhart, Ross & Michael Lewis-Beck, 1994. 'Comparative Democracy: The Economic Development Thesis', *American Political Science Review* 88(4): 903–910.

Chan, Steven, 1997. 'In Search of Democratic Peace: Problems and Promise', *Mershon International Studies Review* 41(1): 59–92.

Collier, Paul & Jan Willem Gunning, 1999. 'Why Has Africa Grown Slowly?', *Journal of Economic Perspectives* 13(3): 3–22.

Cornes, Richard & Todd Sandler, 1996. *The Theory of Externalities, Public Goods, and Club Goods*, 2nd edn. Cambridge: Cambridge University Press.

Cressie, Noel, 1991. *Statistics for Spatial Data*. New York: Wiley.

Easterly, William & Ross Levine, 1998. 'Troubles with the Neighbours: Africa's Problem, Africa's Opportunity', *Journal of African Economies* 7(1): 120–142.

Enterline, Andrew J. & Kristian S. Gleditsch, 2000. 'Threats, Opportunity, and Force: Externalization of Domestic Pressure, 1946–82', *International Interactions* 26(1): 1–53.

Gibson, Martha & Michael D. Ward, 1992. 'Export Orientation: Pathway or Artifact?', *International Studies Quarterly* 36(3): 331–344.

Gleditsch, Kristian S., 1996. *Aspects of Democratization: Economic Development, Spatial Diffusion, and Persistence in Time*. MA thesis, Department of Political Science, University of Colorado.

Gleditsch, Kristian S., 2000. 'International Dimensions of Democratization', paper presented at the ECPR joint session meeting, Copenhagen, 14–19 April.

Gleditsch, Kristian S., forthcoming. *All International Politics Is Local: The Diffusion of Conflict, Integration, and Democratization*. Ann Arbor, MI: University of Michigan Press.

Gleditsch, Kristian S. & Michael D. Ward, 1999. 'A Revised List of Independent States Since 1816', *International Interactions* 25(4): 393–413.

Gleditsch, Kristian S. & Michael D. Ward, 2000. 'War and Peace in Time and Space: The Role of Democratization', *International Studies Quarterly* 44(1): 1–29.

Gochman, Charles, 1991. 'Interstate Metrics since the Congress of Vienna', *International Interactions* 17(1): 93–112.

Harary, Frank; Robert Z. Norman & Dorwin Cartright, 1965. *Structural Models: An Introduction to the Theory of Directed Graphs*. New York: Wiley.

Huntington, Samuel P., 1991. *The Third Wave: Democratization in the Late Twentieth Century*. Norman, OK: Oklahoma University Press.

Jaggers, Keith & Ted R. Gurr, 1995. 'Tracking Democracy's Third Wave with the Polity III Data', *Journal of Peace Research* 32(4): 469–482.

Kuran, Timur, 1991. 'The East European Revolution of 1989: Is It Surprising That We Were Surprised?', *American Economic Review* 81(2): 121–125.

Lake, David A. & Patrick Morgan, eds, 1997. *Regional Orders: Building Security in a New World*. State College, PA: Pennsylvania State University Press.

Lemke, Douglas, 1995. 'The Tyranny of Distance: Redefining Relevant Dyads', *International Interactions* 21(1): 23–38.

Lipset, Seymour M., 1960. *Political Man: The Social Bases of Politics*. Garden City, NY: Anchor.

Lohmann, Susanne, 1994. 'The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989–91', *World Politics* 47(1): 42–101.

Lopez-Bazo, Enrique; Ester Vayá, Antonio J. Mora & Jordi Suriñach, 1999. 'Regional Economic Dynamics and Convergence in the European Union', *Annals of Regional Science* 33(3): 343–370.

McNulty, Mel, 1999. 'The Collapse of Zaire: Implosion, Revolution or External Sabotage?', *Journal of Modern African Studies* 37(1): 53–82.

Maoz, Zeev, 1998. 'Democratic Networks', typescript, Jaffe Center for Strategic Studies, University of Haifa.

Maoz, Zeev & Bruce M. Russett, 1993. 'Normative and Structural Causes of the Democratic Peace, 1945–1986', *American Political Science Review* 87(3): 624–638.

Mueller, Dennis, 1989. *Public Choice II*. Cambridge: Cambridge University Press.

Murdoch, James C.; Todd Sandler & Keith Sargent, 1997. 'A Tale of Two Collectives: Sulfur versus Nitrogen Oxides Emission Reduction in Europe', *Economica* 64(2): 281–301.

O'Donnell, Guillermo; Philippe C. Schmitter & L. Whitehead, 1986. *Transitions from Authoritarian Rule*. Baltimore, MD: Johns Hopkins University Press.

O'Loughlin, John; Michael D. Ward, Corey L. Lofdahl, Jordin S. Cohen, David S. Brown, David Reilly, Kristian S. Gleditsch & Michael Shin, 1998. 'The Diffusion of Democracy, 1946–1994', *Annals of the Association of American Geographers* 88(4): 545–574.

Przeworski, Adam & Henry Teune, 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley.

Raftery, Adrian E., 1995. 'Bayesian Model Selection in Social Research (with Discussion)', in Adrian E. Raftery, ed., *Sociological Methodology 1995*. San Francisco, CA: Jossey-Bass (111–196).

Randle, Michael, 1991. *People Power: The Building of a New European Home*. Stroud: Hawthorn.

Richardson, Lewis Fry, 1942. 'The Problem of Contiguity: An Appendix to Statistics of Deadly Quarrels', reprinted in *General Systems: Yearbook of the Society for General Systems Research*, Volume VI, 1961 (139–187).

Ripley, Brian D., 1988. *Statistical Inference for Spatial Processes*. Cambridge: Cambridge University Press.

Russett, Bruce M., 1993. *Grasping the Democratic Peace: Principles for the Post Cold War World*. Princeton, NJ: Princeton University Press.

Starr, Harvey, 2000. 'Opportunity, Willingness, and Geographic Information Systems (GIS): Reconceptualizing Borders in International Relations', paper presented at the Conference on New Methodologies for the Social Sciences: The Development and Application of Spatial Analysis for Political Methodology, Boulder, CO, 10–12 March.

Summers, Robert & Alan Heston, 1991. 'The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–1988', *Quarterly Journal of Economics* 106(2): 327–368.

*The Economist*, 1996. 'Someone else's doing, someone else's problem (fighting in Zaire between Tutsis and Zaire's army; Hutu refugees may be real target)', *The Economist* 341: 45–46.

Tobler, Waldo; Uwe Deichmann, Jon Gottsegen & Kelly Maloy, 1995. 'The Global Demography Project', working paper 95–6, National Center for Geographic Information and Analysis, Department of Geography, University of California Santa Barbara.

Vanhanen, Tatu, 1990. *The Process of Democratization: A Comparative Study of 147 States, 1980–88*. New York: Crane Russak.

Vanzo, John P., 1999. 'Border Configuration and Conflict: Geographical Compactness as a Territorial Ambition of States', in Paul F. Diehl, ed., *A Road Map to War: Territorial Dimensions of International Conflict*. Nashville, TN: Vanderbilt University Press (73–112).

Ward, Michael D.; Patrick Heagerty & Kristian S. Gleditsch, 2001. 'The Political Geography of Conflict and Trade', typescript, Center for Statistics in the Social Science, University of Washington.

KRISTIAN S. GLEDITSCH, b. 1971, PhD (University of Colorado, Boulder, 1999); Assistant Professor of Political Science, University of California San Diego (2001– ). This research was completed while he was lecturer in social science methodology in the Faculty of Social Sciences, University of Glasgow (1999–2001). Current research interests include international conflict and cooperation, democratization, and applications of spatial statistics.

MICHAEL D. WARD, b. 1948, PhD (Northwestern University, 1977); Professor of Political Science and member of the Center for Statistics in the Social Sciences, University of Washington, Seattle (1997– ); Professor of Economics, University of Pierre Mendès France, Grenoble. Current research deals with using Bayesian statistical methods to study the spread of democratic and economic institutions over time and especially space.

http://www.jstor.org

# LINKED CITATIONS

*- Page 1 of 3 -*

*You have printed the following article:*

**Measuring Space: A Minimum-Distance Database and Applications to International Studies**
Kristian S. Gleditsch; Michael D. Ward
*Journal of Peace Research*, Vol. 38, No. 6. (Nov., 2001), pp. 739-758.
Stable URL:
http://links.jstor.org/sici?sici=0022-3433%28200111%2938%3A6%3C739%3AMSAMDA%3E2.0.CO%3B2-A

---

*This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.*

## [Footnotes]

[21] **Tracking Democracy's Third Wave with the Polity III Data**
Keith Jaggers; Ted Robert Gurr
*Journal of Peace Research*, Vol. 32, No. 4. (Nov., 1995), pp. 469-482.
Stable URL:
http://links.jstor.org/sici?sici=0022-3433%28199511%2932%3A4%3C469%3ATDTWWT%3E2.0.CO%3B2-Q

[21] **The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988**
Robert Summers; Alan Heston
*The Quarterly Journal of Economics*, Vol. 106, No. 2. (May, 1991), pp. 327-368.
Stable URL:
http://links.jstor.org/sici?sici=0033-5533%28199105%29106%3A2%3C327%3ATPWT%285%3E2.0.CO%3B2-D

## References

**Economic Growth in a Cross Section of Countries**
Robert J. Barro
*The Quarterly Journal of Economics*, Vol. 106, No. 2. (May, 1991), pp. 407-443.
Stable URL:
http://links.jstor.org/sici?sici=0033-5533%28199105%29106%3A2%3C407%3AEGIACS%3E2.0.CO%3B2-C

**NOTE:** *The reference numbering from the original has been maintained in this citation list.*

http://www.jstor.org

# LINKED CITATIONS



**What to do (and not to do) with Time-Series Cross-Section Data**
Nathaniel Beck; Jonathan N. Katz
*The American Political Science Review*, Vol. 89, No. 3. (Sep., 1995), pp. 634-647.
Stable URL:
http://links.jstor.org/sici?sici=0003-0554%28199509%2989%3A3%3C634%3AWTD%28NT%3E2.0.CO%3B2-U

**In Search of Democratic Peace: Problems and Promise**
Steve Chan
*Mershon International Studies Review*, Vol. 41, No. 1. (May, 1997), pp. 59-91.
Stable URL:
http://links.jstor.org/sici?sici=1079-1760%28199705%2941%3A1%3C59%3AISODPP%3E2.0.CO%3B2-N

**Why Has Africa Grown Slowly?**
Paul Collier; Jan Willem Gunning
*The Journal of Economic Perspectives*, Vol. 13, No. 3. (Summer, 1999), pp. 3-22.
Stable URL:
http://links.jstor.org/sici?sici=0895-3309%28199922%2913%3A3%3C3%3AWHAGS%3E2.0.CO%3B2-Q

**Export Orientation: Pathway or Artifact?**
Martha Liebler Gibson; Michael D. Ward
*International Studies Quarterly*, Vol. 36, No. 3. (Sep., 1992), pp. 331-343.
Stable URL:
http://links.jstor.org/sici?sici=0020-8833%28199209%2936%3A3%3C331%3AEOPOA%3E2.0.CO%3B2-I

**War and Peace in Space and Time: The Role of Democratization**
Kristian S. Gleditsch; Michael D. Ward
*International Studies Quarterly*, Vol. 44, No. 1. (Mar., 2000), pp. 1-29.
Stable URL:
http://links.jstor.org/sici?sici=0020-8833%28200003%2944%3A1%3C1%3AWAPISA%3E2.0.CO%3B2-E

**Tracking Democracy's Third Wave with the Polity III Data**
Keith Jaggers; Ted Robert Gurr
*Journal of Peace Research*, Vol. 32, No. 4. (Nov., 1995), pp. 469-482.
Stable URL:
http://links.jstor.org/sici?sici=0022-3433%28199511%2932%3A4%3C469%3ATDTWWT%3E2.0.CO%3B2-Q

**NOTE:** *The reference numbering from the original has been maintained in this citation list.*

http://www.jstor.org

# LINKED CITATIONS



**The East European Revolution of 1989: Is it Surprising that We Were Surprised?**
Timur Kuran
*The American Economic Review*, Vol. 81, No. 2, Papers and Proceedings of the Hundred and Third Annual Meeting of the American Economic Association. (May, 1991), pp. 121-125.
Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28199105%2981%3A2%3C121%3ATEERO1%3E2.0.CO%3B2-L

**The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989-91**
Susanne Lohmann
*World Politics*, Vol. 47, No. 1. (Oct., 1994), pp. 42-101.
Stable URL:
http://links.jstor.org/sici?sici=0043-8871%28199410%2947%3A1%3C42%3ATDOICT%3E2.0.CO%3B2-7

**The Collapse of Zaïre: Implosion, Revolution or External Sabotage?**
Mel McNulty
*The Journal of Modern African Studies*, Vol. 37, No. 1. (Mar., 1999), pp. 53-82.
Stable URL:
http://links.jstor.org/sici?sici=0022-278X%28199903%2937%3A1%3C53%3ATCOZIR%3E2.0.CO%3B2-I

**Normative and Structural Causes of Democratic Peace, 1946-1986**
Zeev Maoz; Bruce Russett
*The American Political Science Review*, Vol. 87, No. 3. (Sep., 1993), pp. 624-638.
Stable URL:
http://links.jstor.org/sici?sici=0003-0554%28199309%2987%3A3%3C624%3ANASCOD%3E2.0.CO%3B2-Q

**A Tale of Two Collectives: Sulphur versus Nitrogen Oxides Emission Reduction in Europe**
James C. Murdoch; Todd Sandler; Keith Sargent
*Economica*, New Series, Vol. 64, No. 254. (May, 1997), pp. 281-301.
Stable URL:
http://links.jstor.org/sici?sici=0013-0427%28199705%292%3A64%3A254%3C281%3AATOTCS%3E2.0.CO%3B2-V

**The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988**
Robert Summers; Alan Heston
*The Quarterly Journal of Economics*, Vol. 106, No. 2. (May, 1991), pp. 327-368.
Stable URL:
http://links.jstor.org/sici?sici=0033-5533%28199105%29106%3A2%3C327%3ATPWT%285%3E2.0.CO%3B2-D

**NOTE:** *The reference numbering from the original has been maintained in this citation list.*