The CERDI-seadistance database

Simone Bertoli, Michaël Goujon, Olivier Santoni

CENTRE D'ETUDES
ET DE RECHERCHES
SUR LE DEVELOPPEMENT
INTERNATIONAL

SÉRIE ÉTUDES ET DOCUMENTS

The CERDI-seadistance database

Simone Bertoli

Michaël Goujon

Olivier Santoni

Études et Documents n° 7

March 2016

To cite this document:

Bertoli S., Goujon M., Santoni O. (2016) “ The CERDI-seadistance database”, *Études et Documents,* n° 7, CERDI. http://cerdi.org/production/show/id/1791/type_production_id/1

CERDI
65 BD. F. MITTERRAND
63000 CLERMONT FERRAND – FRANCE
TEL. + 33 4 73 17 74 00
FAX + 33 4 73 17 74 28
www.cerdi.org

The authors

Simone Bertoli
Associate Professor
CERDI – Clermont Université, Université d'Auvergne, UMR CNRS 6587, 63009 Clermont-Ferrand, France.
E-mail: simone.bertoli@udamail.fr

Michaël Goujon
Associate Professor
CERDI – Clermont Université, Université d'Auvergne, UMR CNRS 6587, 63009 Clermont-Ferrand, France.
E-mail: michael.goujon@udamail.fr

Olivier Santoni
Research Assistant
CERDI – Clermont Université, Université d'Auvergne, UMR CNRS 6587, 63009 Clermont-Ferrand, France.
E-mail: olivier.santoni@udamail.fr

Corresponding author: Olivier Santoni

This work was supported by the LABEX IDGM+ (ANR-10-LABX-14-01) within the program "Investissements d'Avenir" operated by the French National Research Agency (ANR).

Études et Documents are available online at: http://www.cerdi.org/ed

Director of Publication: Vianney Dequiedt
Editor: Catherine Araujo Bonjean
Publisher: Mariannick Cornec
ISSN: 2114 - 7957

Abstract

We describe the publicly available CERDI-seadistance database, which contains bilateral maritime distances between 227 countries and territories. The relevant port(s) for countries with access to the sea is defined as the coastal cell of a country that contains the highest number of shipping lines, and each landlocked country is associated to the (foreign) port with the shortest road distance to its capital city. The length of the existing shortest sea route between the two ports is then computed; this represents the bilateral maritime distance included in our database, which also contains a few additional ancillary variables.

Keywords

International trade; Bilateral maritime distances; Road distances.

JEL code

F19

Acknowledgment

The authors are grateful to Martine Bouchut for the technical help she provided for the treatment of the data; the geolocalized databases program led by the CERDI is supported by the Agence nationale de la recherche of the French Government through the program `Investissements d'Avenir' (ANR-10-LABX-14-01), and the IDGM+ initiative conducted by the FERDI (Fondation pour les études et recherches sur le développement international); the usual disclaimers apply.

1 Introduction

Since the seminal contribution of Tinbergen (1962), trade economists have relied on the great-circle or as the crow flies distance, which represents the length of the shortest path between two points on the surface of a sphere, to measure the distance between two trading countries.[1] This distance is included in widely used datasets of bilateral and dyadic variables (see, for instance, Mayer and Zignago, 2011). The great-circle distance represents a very good proxy of the air distance between two countries, but a substantial–albeit declining–share of world trade still goes through the sea, and maritime and air distances can greatly differ for pairs of countries that are separated by huge land masses; for instance, the sea distance between Bremen (Germany) and Tokyo (Japan) via the Suez Canal stands at 11,399 nautical miles, while the great circle distance between the two cities stands at 4,898 nautical miles.

The CERDI-seadistance database provides a matrix of bilateral sea distances for 227 countries and territories.[2] The construction of the dataset has required to: (*i*) identify the location of the relevant port(s) for each country, and (*ii*) compute of the sea distance between the relevant ports of each one of the 227×226 country pairs. We also computed and included in the dataset (*iii*) the road distance between the capital of each country and its relevant port(s). For countries with access to the sea, the relevant port is located in the coastal cell of its territory that contains the highest number of shipping lines among the cells that include a major road.[3] For a landlocked country, the relevant port is defined as the (foreign) relevant port with the shortest road distance to its capital city. We also include in the dataset (*iv*) the road distance between the capital cities of two countries when this is shorter than the sum of the road distance connecting each of them to the corresponding relevant port.

The CERDI-seadistance database is closely related to Feyrer (2009), but it differentiates from it in several respects; specifically, we describe explicitly which is the criterion used to identify the relevant port for each country, including also landlocked countries, and we rely on existing maritime routes rather than on an algorithm that finds the shortest path between two ports to define maritime distances. Last but not least, our database is made freely publicly available. Another closely related database is represented by Fugazza *et al.* (2013),

[1] Feyrer (2009) represents a notable exception here, as discussed below.

[2] The list of countries and territories comes from Gridded Population of the World Database by the SEDAC-CIESIN, Columbia University.

[3] For some large countries, notably the United States, Canada and Russia, we have identified two relevant ports, one of the Eastern and one on the Western coast, as in Feyrer (2009).

who measure the shortest maritime linear shipping routes between any pair of countries for a sample of 178 countries over 2006-2012. The maritime distances between main container ports in each country are computed, and information about the need of container transshipments when countries are not connected directly, which could induce higher transportation costs.To the best of our knowledge, this database is not publicly available.

The paper is structured as follows: Section 2 describes the various analytical choices and the different steps that we followed for the construction of the database, while Section 3 describes the structure and content of the Excel and Stata data files that contain the CERDI-seadistance database.

2 Construction of the database

We provide here a detailed description of the different phases that have led to the construction of the CERDI-seadistance database.

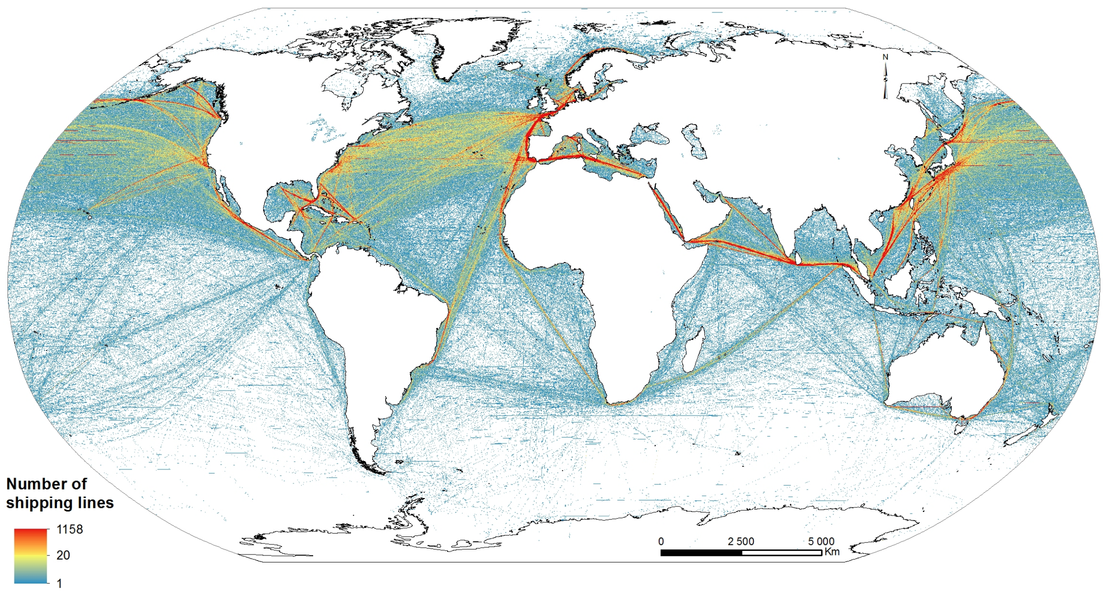
2.1 Identification of the relevant port(s)

The identification of the relevant port(s) for each country goes through a series of intermediate steps. First, we divided the area of each country or territory into a grid of cells of 100 squared kilometers, and selected the coastal cells. Then, we relied on Halpern *et al.* (2008), who provide information about the number of shipping lines that went through each squared kilometer of the sea in 2005 (see Figure 1), to count the number of shipping lines that go through each coastal cell using the Spatial Analyst Zonal statistics tool.[4] Finally, we relied on the RWDB2 database by the FAO to identify the coastal cells that include major land roads, and selected from this subset of cells the one with the highest number of shipping lines. This last step is meant to avoid identifying as the relevant port a coastal cell that is located in an unpopulated area, as it would have happened, for instance, in the case of Namibia, where the coastal cell with the highest number of shipping lines is within the Skeleton Coast National Park.

We have identified two relevant ports for three countries, namely Canada, Russia and

[4]Specifically, we attribute to each coastal cell the highest number of shipping lines that go through the set of maritime cells that it contains.

Figure 1: Shipping lines



Source: Authors' elaboration on Halpern *et al.* (2008)

the United States, following Feyrer (2009). The justification for this choice is related to that the bilateral sea distances of these countries with the other 226 countries in the database would be otherwise greatly sensitive with respect to the choice to locate the relevant port on their Western or Eastern coast. Specifically, the average distance between Canada and the relevant ports of the other countries with access to the sea stands at 8,126 kilometers, and the average of the absolute value of the difference in the distances measured from its relevant port on the Eastern or the Western coast stands at 7,333 kilometers, i.e., more than 90 percent of the average distance.[5] Canada is clearly not the only country that have ports located both on the Atlantic and on the Pacific Ocean; what differentiates, say, Canada from Colombia is that in the latter case the two ports are *not* separated by a huge land mass, as shipping lines that go through the Panama Canal, and this is why we have decided not to select two relevant ports for other countries.

42 out of 227 countries do not have access to the sea;[6] for each one of them, we selected the

[5]The corresponding figures for Russia and the United States stand at 51 and 120 percent respectively.

[6]The three countries bordering the Caspian Sea, namely Azerbaijan, Kazakhstan and Turkmenistan are considered as being landlocked, but we also computed the direct maritime distances between their (domestic)

(foreign) relevant port with the minimal road distance to its capital city.[7,8] We also computed the road distance between the capital city of each country and its relevant port(s), adopting the adjustment for the slope of the terrain described above.[9] Figure 2 provides an example of our use of the (adjusted) data on the road network to identify the relevant port for Mali, which happens to be located in Guinea.

2.2 Computation of the sea distance between relevant ports

Distances between relevant ports have been calculated as the shortest path between them, using the raster dataset in Halpern *et al.* (2008) and the Mollweide projection. From each relevant port, a raster distance map is created using the tool Coast Distance from Spatial Analyst. Then, with Spatial Analyst's Zonal statistics we get the distance from the home port to all other relevant ports, following actual shipping lines from Halpern *et al.* (2008). The computation is limited to 11,000 kilometers from the central meridian to avoid distortions due to map projection far away from it; to calculate all distances, the central meridian is moved eight times by 45° (Figure 3 provides an example).

Notice that the bilateral maritime distance can be equal to zero for two landlocked countries if they are both connected to the same (foreign) relevant port. For instance, this is the case for the bilateral distance between Austria, Hungary, Slovakia and Liechtenstein, as the relevant port for these four countries is the Slovenian one, or for Afghanistan, Kyrgyzstan, Tajikistan and Uzbekistan, which are all associated to a relevant port located in Pakistan.

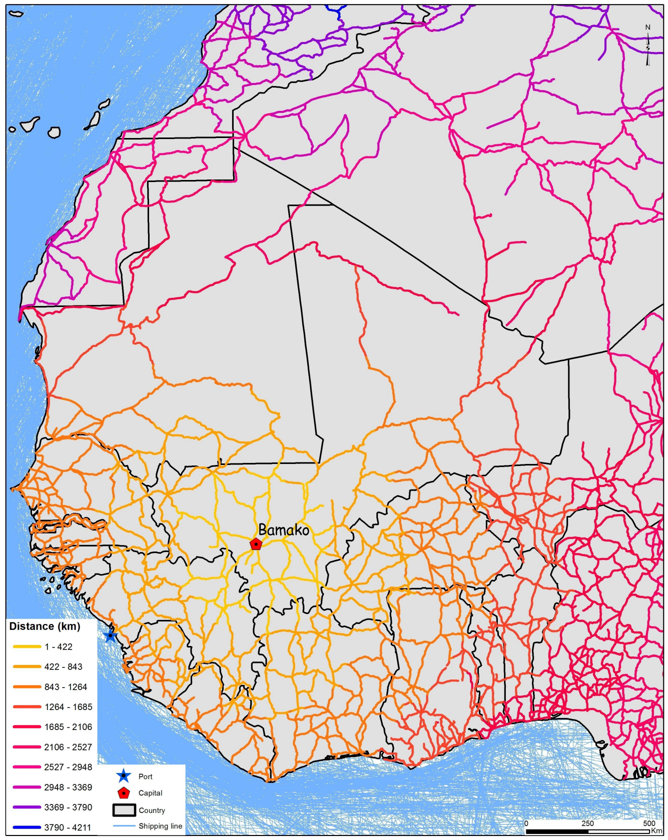
relevant ports.

[7]The data source for the road network is the GeoNetwork database by the FAO (http://www.fao.org/geonetwork/srv/en/); we adjusted the road distance through a function of its slope using the digital terrain model provided by ISciences-Terraviva (http://www.TerraViva.NET/DataSolutions/projects_elevation.html), by multiplying the length of the road by a coefficient that is an increasing function of the slope of the terrain (notice that 86 percent of the cells have a zero slope, so that this adjustment has actually a limited empirical relevance).

[8]We considered, but did not opt for, the option of taking two relevant ports also for some landlocked countries, e.g., Chad, for which the road distances between its capital city and two relevant ports located on two different coasts were very close to each other.

[9]This distance was not calculated for small islands, small coastal territories, such as Macau, and for the Democratic Republic of the Congo and some islands such as Cuba for which information on the road network was not available.

Figure 2: Identification of the (foreign) relevant port for Mali



Source: Authors' elaboration on FAO GeoNetwork and ISciences-Terraviva

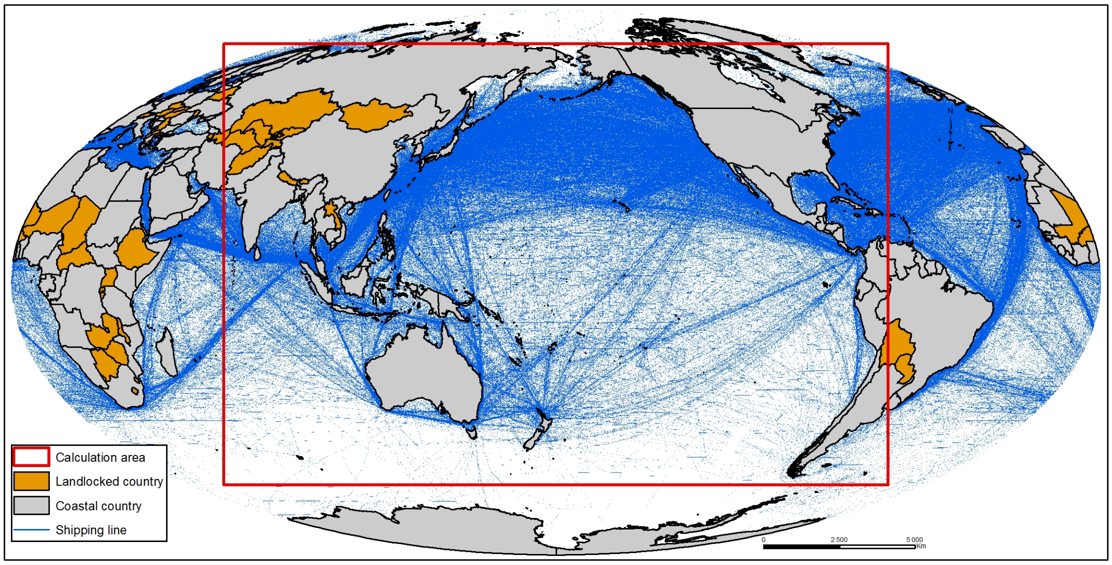
2.3 Computation of the road distance for landlocked countries

For some pairs of landlocked countries, the road distance between their capitals is shorter than the sum of the road distances between them and the two relevant ports (for instance, Mali and Burkina Faso), so that merchandise trade is unlikely to go through the sea in this case. The database reports the direct road distance, and it includes a dummy when this is smaller than the sum of the distances to the two ports.

3 The CERDI-seadistance database in Excel and Stata

The CERDI-seadistance database, whcih is publicly available on Zenodo at the address http://dx.doi.org/10.5281/zenodo.46822, contains the following seven variables:

Figure 3: Area calculation with the central meridian at 180°



Source: Authors' elaboration on Halpern *et al.* (2008)

1. `iso1`: the alpha-3 ISO country code of the first country (string).
2. `iso2`: the alpha-3 ISO country code of the second country (string).
3. `seadistance`: bilateral sea distance in kilometers (numeric).
4. `capitalport1`: road distance between the capital of the first country and the relevant port in kilometers (numeric).
5. `capitalport2`: road distance between the capital of the second country and the relevant port in kilometers (numeric).
6. `roaddistance`: bilateral road distance between two capitals in kilometers (numeric).
7. `short`: a dummy variable which is equal to 1 if `roaddistance` is lower than `capitalport1+capitalport2`, and 0 otherwise (numeric).

Notice that the variables `capitalport1` and `capitalport2` are bilateral rather than country-specific as for six countries (Canada, Russia, the United States, plus Azerbaijan, Kazakhstan and Turkmenistan) we use two relevant ports.

Users should cite the CERDI-seadistance database as follows:

BERTOLI, S., M. GOUJON, AND O. SANTONI (2016): "The CERDI-seadistance database," Études et Documents No. 7, CERDI, Clermont-Ferrand.

References

FEYRER, J. (2009): "Trade and Income - Exploiting Time Series in Geography," NBER Working Paper No. 14910, Cambridge MA.

FUGAZZA, M., J. HOFFMANN, AND R. RAZAFINOMBANA (2013): *Building A Dataset For Bilateral Maritime Connectivity*. New York: United Nations.

HALPERN, B. S., S. WALBRIDGE, K. A. SELKOE, C. V. KAPPEL, F. MICHELI, C. D'AGROSA, J. F. BRUNO, K. S. CASEY, C. EBERT, AND H. E. FOX (2008): "A global map of human impact on marine ecosystems," *Science*, 319(5865), 948–952.

MAYER, T., AND S. ZIGNAGO (2011): "Notes on CEPII's distances measures: the GeoDist Database," CEPII Working Paper 2011-25, Paris.

TINBERGEN, J. (1962): *Shaping the world economy: suggestions for an international economic policy.*