Geolocation of Business Register data:

Results from a prototype database

Paper presented to the
Meeting of the Group of Experts on Business Registers
organized jointly by UNECE, Eurostat and OECD
Brussels, 21-23 September 2015

Authors:

Alessandro Alasia and Alexander Davies

Centre for Special Business Projects

Contact email: alessandro.alasia@statcan.gc.ca

Statistics Canada Statistique Canada

Canada

Table of Contents

Summary and key findings

This paper presents a prototype Resource Area Database containing a set of measures of proximity between communities (here represented by census subdivisions - CSDs) and resource areas. The database is computed by using Business Register data combined with travel distance data generated from the Google Maps API. Five resource industrial sectors are considered (based on groupings of NAICS codes): agriculture, fishing, forestry, energy (oil and gas), and mining.

Using BR micro data, a classification of six community types is developed to allow an understanding of the characteristics of resource areas. This classification accounts for the absolute and relative size of each resource sector in the community, and for the presence of dominant enterprises in the sector. A global proximity index is computed for each of the five resource sectors. The global index is then decomposed into sub-indices that measure proximity to each of the six community types.

For each CSD, the proximity indices provide information on the relevance of resource sector activities for the community by accounting for the presence of these activities within the community as well as in the surrounding region. The indices are used to analyze the distribution and co-distribution patterns of resource sector activities at a small geographic scale, illustrating the information potential resulting from the geocoding of BR data.

Overall approach and methods

- A classification of six community types was developed to allow an understanding of the characteristics of resource areas. This classification accounts for the absolute and relative size of each resource sector in the community and for the presence of dominant enterprises in each resource sector.
- A global proximity index was computed based on a gravity model that uses a measure of agglomeration of the resource sector at the CSD level and a matrix of proximity (travel time) between CSDs. Travel time between CSDs is generated using the Google Maps API. The value of the proximity index for a community is the sum of "agglomeration divided by distance" to each CSD within two-and-a-half hours' travel time.
- Two measures of agglomeration were tested; both are generated from the Business Register. This resulted in an **employment-based** index in which the agglomeration measure is the CSD's employment in the sector, and a **revenue-based** index in which agglomeration is captured by the total revenue of businesses in the sector in the CSD. Where otherwise mentioned, employment and revenue data comes from the January 2014 Business Register, which gives information for the twelve-month period ending in roughly mid-2013.
- Each global index (employment-based and revenue-based) can be decomposed by proximity to each type of community (described in the bullet point after next). The relevance of this decomposition for analysis is that it provides information on the type of resource areas near each CSD. While the global proximity measure can be identical for two communities, the decomposition of the index might show that the nearby resource areas differ in terms of the size (according to revenue or employment), industrial concentration, or enterprise concentration (as measured by the revenue of the top three enterprises).
- Each proximity index (global, and its six components) can be presented in its original value, or in a transformed form (logarithmic or linearly rescaled form of the logarithm). The original and transformed forms of these indices are all suitable for comparing CSDs' exposure to resource sector activity.

Types of resource communities

- Six types of communities (resource areas) were defined using thresholds of absolute and relative size of the resource sector and size of the three largest resource sector businesses in the CSD. For the purposes of the prototype database created for this report, the same threshold values are applied to each resource sector and the classification is based only on revenue-based thresholds. The six types are named: (1) major-project resource-reliant; (2) cluster resource-reliant; (3) major-project diversified; (4) cluster diversified; (5) small resource-reliant; (6) small diversified.

- The distribution of resource community types shows some major differences between resource sectors. As could be anticipated, agriculture and fishing have a significant presence of small reliant CSDs. A sizable share of forestry, oil and gas, and mining, on the other hand, are major-project diversified CSDs.

Global proximity indices

- A global proximity index is computed for each of the five resource sectors. The index is computed for each CSD within commuting distance of a CSD containing a resource sector business. These global proximity indices are scaled so that the minimum value is 0 and the largest value is 1. Overall, revenue-based indices and employment-based indices yield similar results.
- Overall, these results show that proximity measures are generally higher (>0.6 for the rescaled index) for proximity to agriculture resource areas. The global mining proximity index tends to be lower (<0.5) for the majority of the CSDs; only a relatively small group of CSDs record a mining proximity index over 0.6.
- The distribution of the global proximity index for forestry is similar to that of agriculture, with a relatively large number of CSDs in some degree of proximity to forestry resource areas. This is not surprising, given the relevance of land as factor of production for these two sectors.
- Fishing is the resource sector that shows the most pronounced bi-modal distribution of the global proximity index; while most CSDs fall below the 0.6 threshold, a second peak of the distribution is found around the 0.8 value, indicating CSDs in close proximity to fishing resource areas.
- A preliminary look at the co-distribution patterns of the indices shows that the strongest correlation between the five revenue-based proximity indices (mining, agriculture, fishing, energy and forestry) is between mining and agriculture, indicating that communities close to the mining sector tend also to be close to the agriculture sector. Almost half of the variation in the mining proximity index is explained by variation in the agriculture proximity index (r=0.66).
- The weakest association between the five proximity indices is between forestry and energy, which in fact have a negative correlation, indicating that communities close to forestry have a slight tendency to be far from energy and vice versa.

Decomposition of global indices

- The same computational model used for the global index is implemented to decompose the index by proximity to each community type. In this way, for each CSD and resource sector, the sum of the original untransformed indices of the six components (one for each of the six community types) is equal to the global index.
- This decomposition shows that, for the average CSD in Canada, proximity to both the energy sector and the mining sector is largely due to proximity to major-project diversified communities. This is least true for the agriculture and fishing sectors, for which proximity to cluster (reliant or diversified) communities constitute 28% and 19%, respectively, of the average global proximity index value.
- Average proximity to community types can be used to benchmark and profile individual CSDs. An example is provided in the report.

CSD-level indicators for a monitoring framework

- CSD-level indicators were derived from the Business Register, the 2011 Census and the 2011 National Household Survey. These indicators included one-year CSD employment growth and one-year CSD revenue growth.

1. Introduction

To date, the potential of conducting economic analysis of small geographic areas by using Business Register (BR) microdata has not been fully examined. This paper is intended to be a preliminary exploration in that direction. It is also intended to illustrate how geocoding of BR microdata can enhance our understanding of the microdata themselves.

This paper outlines a framework for the identification and analysis of “resource areas” and a prototype set of measures of proximity to resource areas. These measures are generated by using Business Register data geocoded at the census subdivision level, in combination with travel distance data generated from the Google Maps API. Five major resource industries were identified: energy (specifically oil and gas), mining, forestry, fisheries and aquaculture, and agriculture.

The analysis has three main objectives:

- The specification of a conceptual and operational definition of “resource areas”, resulting in a method for the localization of those areas;
- The identification and measurement of the proximity of communities to resource areas, as calculated through a set of industry-specific indices which can be combined into a single measurement;
- The creation of a set of proposed indicators to be used to analyse the impact of resource areas on economic and social change.

The paper is organized as follows: after this introduction, the second section presents a short overview of the key definitions, concepts and data sources which are used in the analysis. The third section presents the methodology used for the proposed definition of resource areas. The fourth section presents the proximity indicator results and analysis. A short conclusion wraps up the paper.

2. Definitions and concepts and data sources

The key geographic concept used in this analysis is the census subdivision (CSD), which, in the 2011 version used in this project, divides the entire land area of Canada into 5,253 discrete areas. The CSD is an administrative geographical region that is defined using population characteristics as well as municipal and reserve boundaries. The “area” in the concept of “resource area” will be identified by the boundaries of the census subdivision (CSD) in which each project is located.

Geographic proximity between CSDs is measured by using the road distance from a representative point of one CSD to the representative point of the other CSD. The general rule was to locate the representative point in the most populated area of the CSD, always near a major road.

The travel distance and the travel time between CSDs are calculated by the Google Maps Distance Matrix API. Google Maps maintains a road network database for Canada that can be used as the basis for calculations of distance and travel time between an origin and a destination. This road network database contains information on road type, speed limit and type of surface. The database also includes ferries with regular service, which permits the inclusions of communities on islands within the main network. The travel time which was obtained from the Google Maps Distance Matrix API represents the key input parameter for the computation of proximity used in this analysis. Travel time is a more appropriate measure of proximity than kilometric distance because it accounts for differences in road quality and for ferries.

The identification of resource sectors is based on the aggregation of business data at the CSD level from the Business Register (BR). The BR is a central data source that allows access to a consistent and well-maintained national inventory of businesses. In the BR, the spatial coding of business units is based on Statistics Canada's standard geographical classification: the BR gives the census subdivision identifier for each unit. Business units in the BR are coded based on the concept of major business activity in a manner consistent with the approach outlined by the North American Industry Classification System (NAICS).[1] Custom groups of NAICS codes have been constructed from this classification in order to create five major resource sector groups as specified below. The term "sectors" in this report refers to these five groups.

To allow linkage with an existing database of CSD-to-CSD distances, data from the BR were aggregated at the CSD level; that is, the employment and revenue of business units in the same industry and CSD was summed. To implement the computations, a database was created, containing:

- BR employment data, derived from payroll deduction (PD7) files,
- BR revenue data, derived from the General Index of Financial Information, and
- the six-digit North American Industrial Classification System (NAICS) code from the Business Register.

In this prototype version of the database, the specific BR fields used were the *PD7 Number of Employees* and *Income Tax Revenue*. These quantities were obtained from the BR at the enterprise level. Where otherwise mentioned, employment and revenue data in this report comes from the January 2014 Business Register, which gives information for the twelve-month period ending in roughly mid-2013.

For simple enterprises, which by definition have only one location and which are the vast majority of enterprises, the CSD of the enterprise is, by definition, the CSD of the location. For approximately 3% of enterprises, known as complex enterprises, the CSD of each location may not actually be the same as the CSD of the enterprise. In these cases, using the *Profiled Number of Employees* and the *Profiled Revenue* instead would have provided proximity scores with greater geographical precision, although complications arising from missing data and mixed sources would have had to have been resolved. Alternatively, data could have been obtained directly from the PD7 file and directly from the General Index of Financial Information, which would be somewhat less aggregated than the enterprise-level data on the BR, but would require record linkage approval from Statistics Canada's Executive Management Board because the CSD and NAICS would need to linked to these files.

When computing total revenue and total employment, all NAICS, except code 91 (public administration), are considered in-scope for this analysis. The exclusion of public administration is due to the fact that it is both difficult to measure and to locate.

Table 1 shows the NAICS codes used to identify the five resource sectors. Agriculture includes one three-digit subsector code and seven four-digit industry groups associated with agriculture. The higher-level subsectors 112 and 115 contain agriculture and either aquaculture or forestry, so four digits must be used to extract only agriculture business activity. Fisheries and aquaculture includes two four-digit industry groups associated with fishing and aquaculture. To extract fisheries and aquaculture business activity only, four-digit detail is necessary, because the higher-level subsectors 112 and 114 contain agriculture and trapping also. Forestry includes the three-digit subsector associated with forestry and the four-digit industry group code which provides support to primary production. To extract forestry business activity only, four-digit detail is necessary because the higher-level subsector 115 contains agriculture also. Oil and gas (energy) includes the three-digit subsectors associated with oil and gas and two six-digit Canadian industry codes which provide services to primary production. To extract energy business activity only, six-digit detail is necessary because the higher, less detailed, levels (213, 2131 and 21311) contain mining also. Mining includes the three-digit subsectors associated with mining and two six-digit Canadian industry codes which provide services to primary production. To extract mining business activity only, six-digit detail is necessary because the higher, less detailed levels (213, 2131 and 21311) contain energy also.

[1] See: http://www.statcan.gc.ca/subjects-sujets/standard-norme/naics-scian/2012/index-indexe-eng.htm.

Table 1: NAICS codes used to assess the presence of resource sector activities

NAICS	Description
Agriculture	
111	Crop production
1121	Cattle ranching and farming
1122	Hog and pig farming
1123	Poultry and egg production
1124	Sheep and goat farming
1129	Other animal production
1151	Support activities for crop production
1152	Support activities for animal production
Fisheries and aquaculture	
1125	Aquaculture
1141	Fishing
Forestry	
113	Forestry and logging
1153	Support activities for forestry
Energy	
211	Oil and gas extraction
213111	Oil and gas contract drilling
213118	Services to oil and gas extraction CAN
Mining	
212	Mining and quarrying (except oil and gas)
213117	Contract drilling (except oil and gas) CAN
213119	Other support activities for mining CAN

2.1. Resource area types: a proposed synthesis

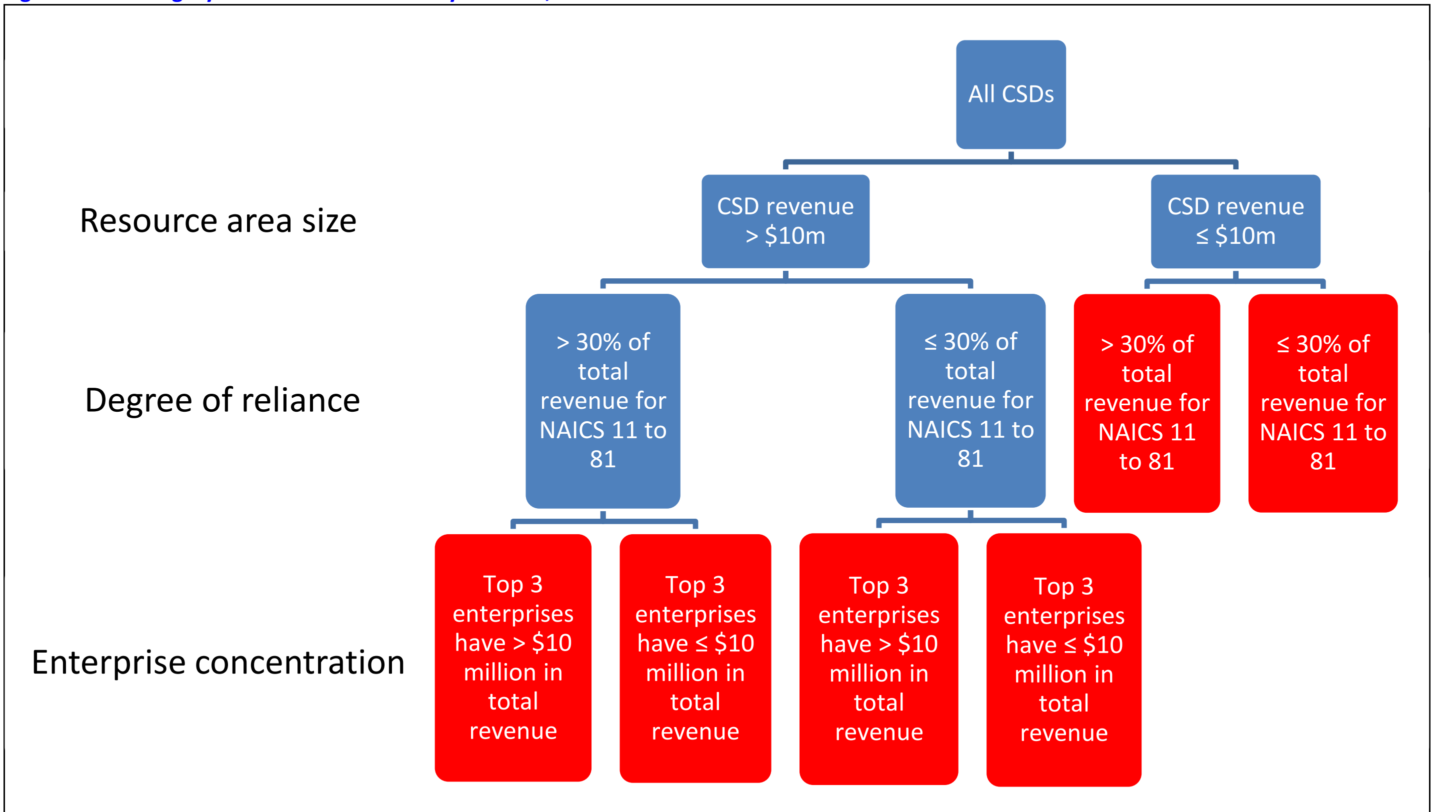
The various ways in which “resource area” is usually defined can be reduced to two main approaches.The first type of approach focuses on some concept of clustering of activities in a given geographic area, without taking into consideration any specific measure of industrial organization of activities in the areas other than possibly counts of businesses operating in the areas. A geographic unit would be considered a resource area if the area presents a certain level of agglomeration (in absolute size) or concentration (in relative terms) of activity in the resource sector. For instance, a resource area could be defined as a geographic unit in which more than 30% of the total employment is directly related to the resource sector.

A second approach would focus on areas in which “major projects” are located. That is, it considers a resource area to exist only if a large enterprise of a certain size is located there which, in itself, meets some given threshold in order to be considered one of the major undertakings of its type or industry.

The two approaches are not necessarily mutually exclusive. In fact, they both highlight relevant dimensions for policy analysis and formulations. These complementary dimensions can be preserved by combining a classification system of resource areas with a composite index of proximity.

For the purpose of this project, the conceptual framework for a classification of resource areas is presented in Figure 1. In this figure, the six red boxes identify the six types of communities defined by the classification system.

Figure 1: Six-category classification of CSDs by revenue, reliance and concentration



CSDs are classified as belonging to one of a set of mutually exclusive groups using a set of threshold criteria. For each type of group, a specific methodology is used to compute a proximity index, which can be used either as a single proximity value or decomposed by community type into a set of values representing the contribution to the total proximity value which was made by each of the various types of resource areas.

As outlined in Figure 1, the first classification criterion is the **size** of the aggregate measure of the resource sector in the community. The classification scheme of Figure 1 uses revenue to measure size, but some other measure such as employment could conceivably be used instead. A given threshold value of *x* is given to distinguish between those communities which have attained a certain level of agglomeration within the CSD and CSDs which do not meet that agglomeration threshold.

The second classification criterion captures the degree to which activity in a resource sector contributes to the total economy of a community. A distinction is made between CSDs in which the resource sector represents more than some level of the total CSD economy for that specific measure under consideration. This dimension is intended to capture the **degree of reliance** of the community on the resource sector. For example, CSDs with more than 30% of total revenue in resource sector A would be considered to be **reliant** on resource sector A.

The third classification criterion is intended to provide information on the **enterprise concentration** of the sector in the community. A distinction is made between CSDs which contain large businesses in the resource sector and CSDs which contain only smaller businesses. In the current specification, the threshold is set in such a way so that the threshold value of the top three largest businesses has to be at least equal to *x* (the first classification criterion). This rule, *de facto*, sets the minimum size of the "major project" for that resource sector in the community. Moreover, the way this criterion is set ensures that if a community is not a "resource area" according to the first criterion, it will automatically be excluded from any further breakdown on the third criterion (as shown in Figure 1).

This classification scheme results in the creation of 6 mutually exclusive types of CSDs, as will be further discussed below.

A global proximity measure to each resource sector (agriculture, fishing, forestry, energy and mining) regardless of community type is generated. For details on the methodology used to compute the proximity measures, see Section 3.1 below. The global proximity measure can be decomposed into components, each measuring proximity to a community, where the community types are the mutually exclusive classes resulting from the classification rules described above. What should be emphasized is that this decomposition into additive parts identifies the contribution of each type of resource area to the global proximity index of a CSD. See Section 3.2 for more details.

This approach distinguishes between types of resource communities and measures the proximity to each and every type of resource community. This offers more information than one single proximity measure that does not permit users to gain an understanding of the regional context.

The classification implemented in this analysis identifies each type of community based on the resource sector existing in the community. The proximity index provides: (1) a measure of geographical proximity between each community and economic activity in the resource sector; and (2) a measure of geographical proximity to each type of resource community, so that the sum of proximity measures to each type of community is equal to the overall proximity index.

3. Methodology

Measuring proximity to resource areas sets a framework for monitoring the potential impact of the employment and spending of resource industries. For the purposes of this project, measures were implemented first for each individual type of resource sector and then consolidated into a single index which encapsulates the accessibility of all types of resource areas.

Several variant indices using different ways of measuring the proximity to a resource sector were created. For instance, both revenue and employment were used to quantify the economic activity in a resource sector.

3.1. Specification of the general index

Each measure of proximity to resource activity is based on a model which follows the general specification of Alasia et al. (2012). Following the idea of a gravity model, the proximity of some community (CSD) i to resource areas can be described as the summation, over all CSDs within a travel time of 2.5 hours from community i, of the ratios between CSD-level resource activity and distance (between each resource area CSD and CSD i in minutes of travel time) (Equation 1).

In other words, for each community of interest, the contribution to the proximity measure of each resource-enterprise-containing CSD within 2.5 hours is divided by the distance to this CSD:

Equation 1:

$$A_i = \sum_{j=1}^{n} \left(\frac{E_{k,j}}{D_{i,j}}\right)$$

where A_i is the value of the proximity for a given CSD i, n is the number of CSDs within the threshold radius of 2.5 hours, $E_{k,j}$ is the measure of size (total revenue or employment, for example) for a given resource industry k in each resource area CSD j within 2.5 hours, and $D_{i,j}$ is the distance between CSD i and the resource area CSD j (measured in travel time). The index can then be transformed in logarithmic form and rescaled.

If CSD i has no access to CSD j, or if the travel time between the two CSDs is at more than 2.5 hours, then $E_{k,j}$ / $D_{i,j}$ takes a value of zero (Alasia et al. 2012). If the travel time between the two CSDs is under 3.75 minutes, the travel time is assumed to be 3.75 minutes. The travel time from a CSD to itself is also considered to be 3.75 minutes. This imputation ensures that the revenue or employment available within the CSD for which the index is computed is taken into consideration in the equation.

The size or intensity of operations in the resource area is weighted by the reciprocal of the distance to that area through the use of a distance matrix, in which all distances are expressed as travel time. If everything else is equal, a community one hour from a large mining operation, for example, will obtain a higher value of the indicator than a community one hour from a small mining operation.

3.2. Decomposition of the general index

The global index can be decomposed into the contributions to the global proximity measure made by each of the community types described in Figure 1; this provides insights into the characteristics of the resource areas which are in proximity to the community of reference. The computational equation that is applied to each type of area is the same; therefore, the decomposition equation can be expressed as follows:

Equation 2: $A_i = A_i^1 + A_i^2 + A_i^3 + A_i^4 + A_i^5 + A_i^6,$ where

Equation 3: $A_i^t = \sum_{j=1}^{n} (\frac{E_{k,j}^t}{D_{i,j}})$

Equation 2 states that the global index A_i for community i is the sum of the six components, A_i^t, where the superscript t indicates the type of area (from type one to type six) as described in Section 2.1 above, and the rest is as indicated in the description of Equation 1 above. Equation 3 states that the measure of proximity to community type t is computed using the same method as the global index, with the only difference being that the measure of size E for the resource sector k refers only to the specific community type t. As in the global index, each term in the summation is the agglomeration weighted by the inverse of $D_{i,j}$, the distance between the CSD of reference i and each within-range (2.5 hours) CSD j. Since the classification of CSDs into community types is exhaustive and mutually exclusive, the summation of all components, A_i^t, yields the value of the global index.

The indices can be analyzed in their original forms, without any transformation or rescaling, or they can be rescaled into various forms. Each form provides specific information and is valuable for different types of analysis. For this reason, multiple forms of the index are provided in the database.

In the remaining sections of this report, the term original form index will be used to refer to the index generated by Equation 1 or Equation 3. In this original form, the index can be described in plain language as a measurement of the total access of a community of reference to resource sector economic activity. The access to economic activity is considered to diminish proportionately with travel time from the community, and economic activity at a distance of 2.5 hours or greater is considered to be inaccessible. The computation accounts for the fact that, other conditions being the same, the community of reference will usually have less access to resource areas located further away.[2]

An advantage of using the original form index (without any transformation) is that this value provides an understanding of the size of the resource area (total revenue or employment) in proximity to the community of reference. The values of different indices (using the same measure of agglomeration) are comparable to each other across sectors, or across components for each sector, or across components for different sectors. For other analytical purposes, a transformed or rescaled index might provide a more straightforward interpretation. The database includes two sets of transformed variables: a logarithmically transformed variable, and a variable to which a second transformation is applied to the log-transformed values to normalize all indices in a range between zero and one.

4. *Proximity indices: results and analysis*

The presentation of results and analysis of the proximity indices is organized as follows. After this overview, the next section describes the distribution of CSDs by the six community types, as defined in Figure 1. This is followed

[2] Clearly, there are specific exception to this, such as for instance, communities in Newfoundland and Labrador with strong labour mobility ties with resource areas of Alberta. However, the intent of the index is to capture general proximity to economic activities and it appears fair to say that the direct economic effect generated by resource development is generally likely to be stronger (or offer more opportunities) for communities near these areas.

by an analysis of the global proximity indices, an analysis of the decomposition of the indices for each resource sector, and then a comparison of aboriginal and non-aboriginal communities' proximity indices. The penultimate section compares the distributions of selected proximity indices and presents other descriptive statistics on the proximity measures, while the final section discusses the relevance of population density.

All the results presented in this section are generated using the prototype Resource Area Database compiled for this study.

4.1. Classification of community types

For the purpose of this analysis, the following thresholds values were used for the classification of CSDs into community types. It is important to emphasize that only revenue is used in these thresholds. This classification may be made for any one of the five resource sectors; in the description below, this sector is called "resource sector A".

Table 2: Threshold values used for the classification of community types

Dimension	Revenue thresholds
Resource area size	Total revenue of businesses in resource sector A in the CSD is at least $10 million
Degree of reliance	Total revenue of businesses in resource sector A represents 30% or more of the total business revenue in the CSD of businesses in NAICS 11 to 81 (excluding NAICS 91 — public administration)
Industrial organization	Total revenue of the three largest businesses in resource sector A in the CSD is at least $10 million

As a result of this classification, the six mutually-exclusive types that are generated following the scheme outlined in Figure 1 are explicitly defined as follows.

(1) A **major-project resource-reliant** community is a CSD with $10 million or more in aggregate total revenue for resource sector A, representing 30% or more of total business revenue in the community, and with the three largest businesses of resource sector A capturing $10 million or more of the total revenue of the resource sector.

(2) A **cluster resource-reliant** community is a CSD with $10 million or more in aggregate total revenue for resource sector A, representing 30% or more of total business revenue in the community, with the largest three businesses of resource sector A capturing less $10 million of the total revenue of the resource sector.

(3) A **major-project diversified** community is a CSD with $10 million or more in aggregate total revenue for resource sector A, representing less than 30% of the total business revenue in the community, and with the three largest businesses of resource sector A capturing $10 million or more of the total revenue of the resource sector.

(4) A **cluster diversified** community is a CSD with $10 million or more in aggregate total revenue for resource sector A, representing less than 30% of the total business revenue in the community, with the three largest businesses of resource sector A capturing less than $10 million of the total revenue of the resource sector.

(5) A **small resource-reliant** community is a CSD with less than $10 million in aggregate total revenue for resource sector A, but representing 30% or more of total business revenue in the community. By definition, the three largest businesses of resource sector A represent less than $10 million in total revenue for the resource sector, so all of these communities are under the industrial organization threshold.

(6) A **small diversified** community is a CSD with less than $10 million in aggregate total revenue for resource sector A, representing less than 30% of the total business revenue in the community. Therefore, as in (5), the three largest businesses of resource sector A represent less than $10 million in total revenue for the resource sector.

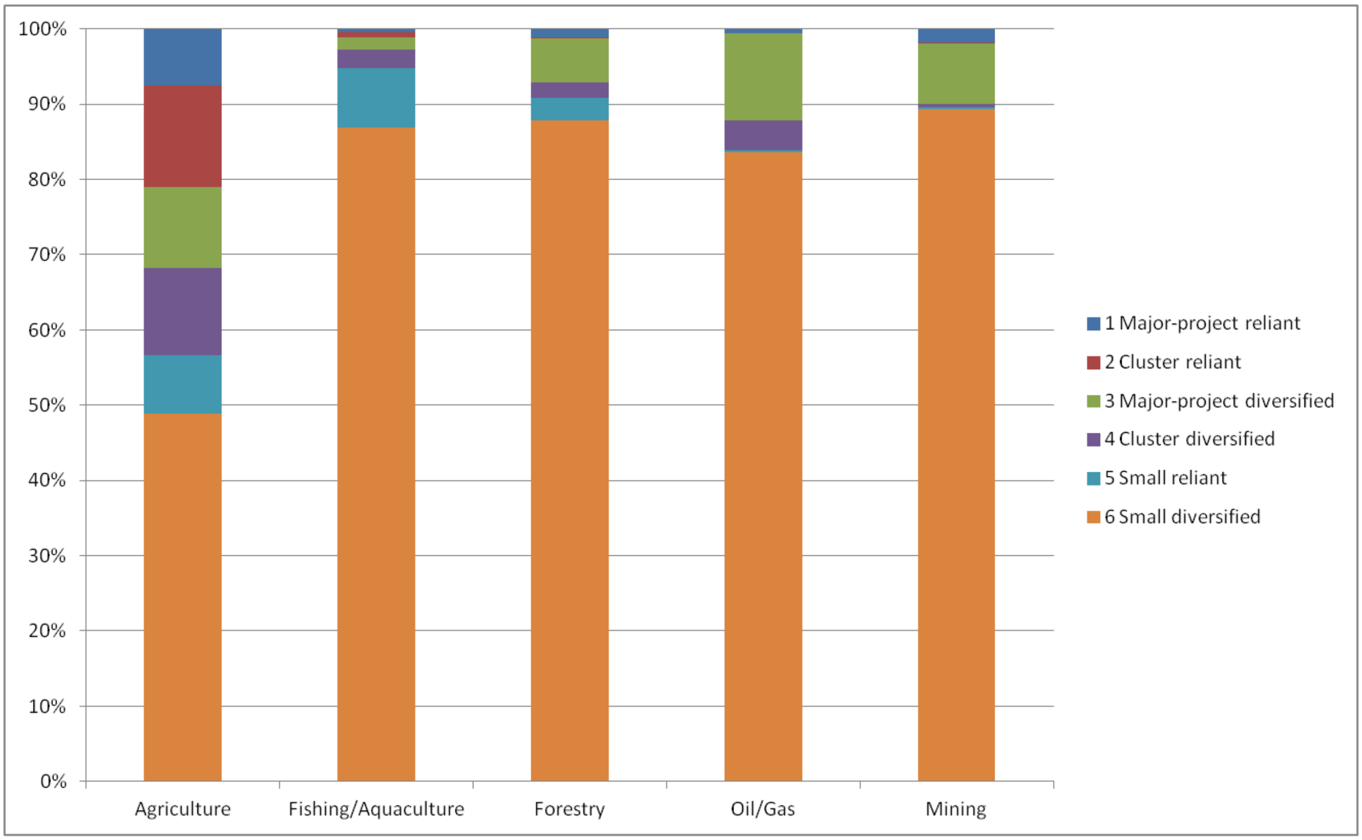
This description highlights the fact that proximity to each type of these resource areas might have different implications in terms of potential options for development and engagement in the sector. For instance, proximity to a resource sector that is entirely due to proximity to major-project resource-reliant communities is likely to have different implications for local economic development than the same proximity value due only to proximity to small diversified CSDs.

Figure 2 displays the proportional distribution of the community types across the five resource sectors for all those CSDs which report some revenue activity in the given sector. These distributions highlight some major differences between resource sectors.

For all sectors except agriculture, communities are most likely to fall into the small diversified category, indicating that they have less than $10 million in revenue for that resource sector and that the community is not reliant on the revenue generated from that sector.

As could be anticipated, agriculture and fishing have a significant presence (in both absolute and relative terms) of small reliant CSDs; in contrast, forestry, oil and gas and mining have a sizable share of major-project diversified CSDs.

Figure 2: Distribution of community types across resource sectors



Note: For each sector, proportions are calculated only for those CSDs that reported some revenue in the applicable sector.
Source: Authors' computations.

Table 3 shows the same information as Figure 2 but in tabular form. It should be noted that the totals refer to communities reporting some revenue in that resource sector. Thus, there are 3,167 CSDs in which a primarily agricultural enterprise (generating some revenue) is located; however, there are only 939 CSDs in which an oil & gas enterprise is located.

Table 3: Distribution of community types across resource sectors

Community type	Agriculture		Fishing		Forestry		Oil/Gas		Mining	
	Count	%	Count	%	Count	%	Count	%	Count	%
Major-project reliant	239	8	6	0	25	1	6	1	25	2
Cluster reliant	427	13	8	1	4	0	0	0	2	0
Major-project diversified	339	11	22	2	139	6	108	12	109	8
Cluster diversified	367	12	31	2	49	2	37	4	6	0
Small reliant	249	8	102	8	69	3	3	0	3	0
Small diversified	1,546	49	1,124	87	2,075	88	785	84	1,214	89
Total	**3,167**	**100**	**1,293**	**100**	**2,361**	**100**	**939**	**100**	**1,359**	**100**

Note: For each sector, counts and percentages are calculated only for those CSDs that reported some revenue in the applicable sector.
Source: Authors' computations.

More detailed analysis of the contribution of each sector and type of resource area can be undertaken. For instance, Table 4 shows the proportion of forestry sector revenue from each of the six forestry sector community types. The majority of revenue (77.9%) is located in CSDs generating $10 million or more. Amongst these six types, the major-project diversified type covers the majority (63.7%) of revenue.

Table 4: Detailed classification breakdown, forestry sector

Type of community	CSD Count	Revenue %
Total	**2,361**	**100.0%**
Small: Resource area size below $10M	**2,144**	**22.1%**
Diversified	2,075	19.7%
Reliant	69	2.4%
Resource area size $10M or more	**217**	**77.9%**
Diversified	188	69.1%
Clustered	49	5.4%
Major project	139	63.7%
Reliant	29	8.9%
Clustered	4	0.5%
Major project	25	8.3%

Source: Authors' computations.

It should be noted that the parameters used to define the community types in the prototype database could be adjusted and adapted for each specific resource sector. These parameters include the threshold values used to define the minimum size of a resource area, the share defining the degree of reliance, and the number or minimum size of the top businesses used to define industrial organization.

4.2. Industrial structure of the resource sector at the community level

The Resource Area Database permits analysis of the industrial structure of the resource sector at the community level. An example is shown below.

Table 5 below shows the number of CSDs having an agricultural enterprise over a certain revenue threshold. This table uses two thresholds - $5 million in revenue and $10 million in revenue. In this table, it is apparent that CSDs containing such enterprises are most likely to have a single enterprise over the threshold.

Table 5: Number CSDs with agricultural enterprises over revenue threshold

Number of agricultural enterprises with revenue over threshold	$5m revenue threshold			$10m revenue threshold		
	Number of CSDs	Percentage of all CSDs	Percentage of 1+ CSDs	Number of CSDs	Percentage of all CSDs	Percentage of 1+ CSDs
Total	**3,171**	**100.0%**		**3,171**	**100.0%**	
0	2,583	81.5%		2,901	91.5%	
1+	588	18.5%	100.0%	270	8.5%	45.9%
1	340	10.7%	57.8%	190	6.0%	32.3%
2	112	3.5%	19.0%	39	1.2%	6.6%
3	53	1.7%	9.0%	10	0.3%	1.7%
4	29	0.9%	4.9%	15	0.5%	2.6%
5	10	0.3%	1.7%	4	0.1%	0.7%
6+	44	1.4%	7.5%	12	0.4%	2.0%

Source: Authors' computations.

Table 6 and Table 7 look at the number of CSDs with at least one, three or five agricultural enterprises over a revenue threshold of $5 million, and the agricultural revenue share of the top one, top three and top five enterprises, respectively, averaged over all CSDs. From these tables it is apparent that, although these business may be the largest in their CSD, they do not dominate the agricultural revenue in their CSD.

Table 6: CSDs with agricultural enterprises over $5m in revenue, by province

Province	CSDs with at least one $5m+ agricultural enterprise		CSDs with three or more $5m+ agricultural enterprises		CSDs with five or more $5m+ agricultural enterprises	
	Number of CSDs	Average Top 1 Share of Revenue	Number of CSDs	Average Top 3 Share of Revenue	Number of CSDs	Average Top 5 Share of Revenue
Canada	**588**	**26%**	**136**	**33%**	**54**	**14%**
Nfld. & Lab.	5	75%	0	..	0	..
P.E.I.	11	35%	2	43%	0	..
N.S.	12	28%	2	43%	1	15%
N.B.	11	59%	0	..	0	..
Que.	119	37%	19	54%	4	26%
Ont.	139	24%	43	27%	22	12%
Man.	54	20%	16	38%	2	29%
Sask.	84	22%	7	22%	3	9%
Alta.	117	16%	38	30%	16	14%
B.C.	36	37%	9	30%	6	11%

Source: Authors' computations.

Table 7: CSDs with agricultural enterprises over $10m in revenue, by province

Province	CSDs with at least one $10m+ agricultural enterprise		CSDs with three or more $10m+ agricultural enterprises		CSDs with five or more $10m+ agricultural enterprises	
	Number of CSDs	Average Top 1 Share of Revenue	Number of CSDs	Average Top 3 Share of Revenue	Number of CSDs	Average Top 5 Share of Revenue
Canada	**270**	**31%**	**41**	**36%**	**16**	**15%**
Nfld. & Lab.	2	80%	0	..	0	..
P.E.I.	3	28%	0	..	0	..
N.S.	4	37%	0	..	0	..
N.B.	3	82%	0	..	0	..
Que.	55	44%	5	57%	1	44%
Ont.	82	28%	14	26%	4	16%
Man.	18	26%	3	61%	2	29%
Sask.	27	29%	2	25%	2	10%
Alta.	58	21%	13	40%	4	11%
B.C.	18	35%	4	18%	3	6%

Source: Authors' computations.

4.3. Global proximity measures

For each resource sector, a global proximity index was computed taking into account proximity to any type of community reporting employment or revenue for that given resource sector.

It should be emphasized that the number of CSDs with a non-zero value for a particular resource sector's proximity index is different from the number of CSDs reporting employment or revenue in that resource sector. This is because the proximity index yields non-zero values for the CSDs that have no resource sector activities but are, indeed, in proximity of CSDs reporting resource sector activities.

It should also be recalled that resource sector activities occurring within the CSD of reference are captured by using an imputed value of 3.75 minutes of travel time as the distance, so that the presence of these business can enter into the gravity model computations.

The number of communities (CSDs) with a value of each global revenue-based proximity index is shown in Figure 3 in the column headed "revenue-based index". The "employment-based index" column gives the number of communities (CSDs) with a value of each global employment-based proximity index. There are more CSDs with a revenue-based index value because the some CSDs are within 2.5 hours of revenue-generating enterprises but are not within 2.5 hours of employer-enterprises.

Figure 3: Number of CSDs for which each raw (untransformed) proximity index has a positive value

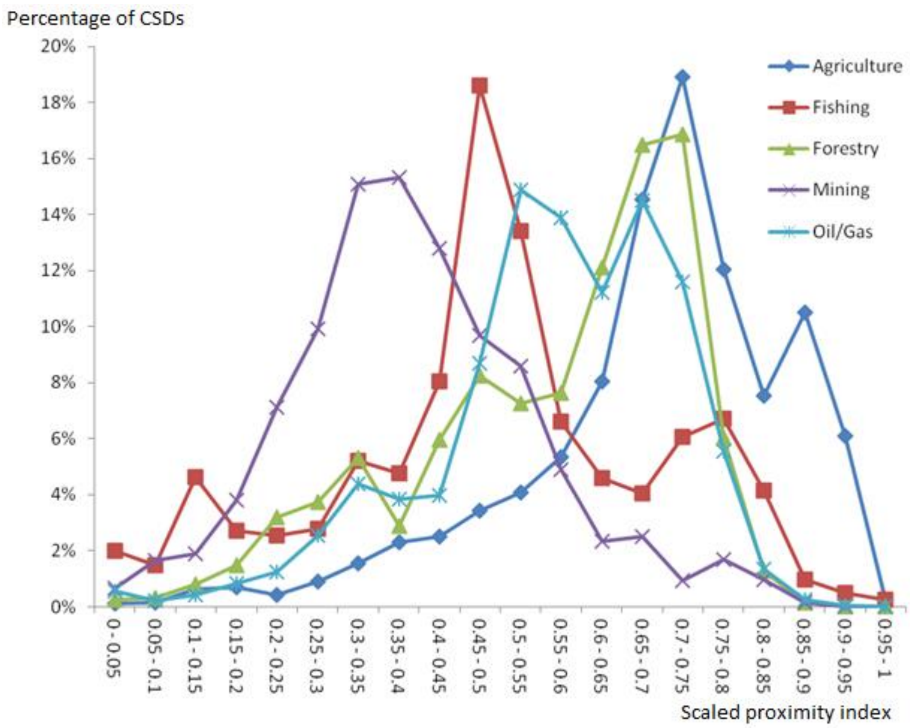
Index sector	Number of CSDs	
	Revenue-based index	Employment-based index
Agriculture	4,994	4,882
Fishing	4,929	3,561
Forestry	4,992	4,810
Oil & gas	4,597	4,405
Mining	4,922	4,741
All-resources	5,071	4,994

Source: Author's computations.

Figure 4 shows the distribution of the global proximity indices for the five resource sectors. Overall, proximity values are generally higher (>0.6 for the normalized index) for proximity to agriculture resource areas. On the contrary, the global mining proximity index tends to be lower (<0.5) for the majority of the CSDs; only a relatively small group of CSDs record a mining proximity index greater than 0.6. Since the scores are rescaled to a range between 0 and 1, this means that there is a low outlier in agriculture and a high outlier in mining.

The distribution of the global proximity index for forestry is similar to that of agriculture, with a relatively large number of CSDs in some degree of proximity to forestry resource areas. This is not surprising given the relevance of land as factor of production for these two sectors. Fishing is the resource sector that shows the most pronounced bi-modal distribution of the global proximity index; while most CSDs fall below the 0.6 threshold, a second peak of the distribution is found around the 0.8 value (indicating a group of CSDs in close proximity to fishing resource areas).

Figure 4: Histograms showing distributions of global indices of employment proximity for five resource sectors, all CSDs



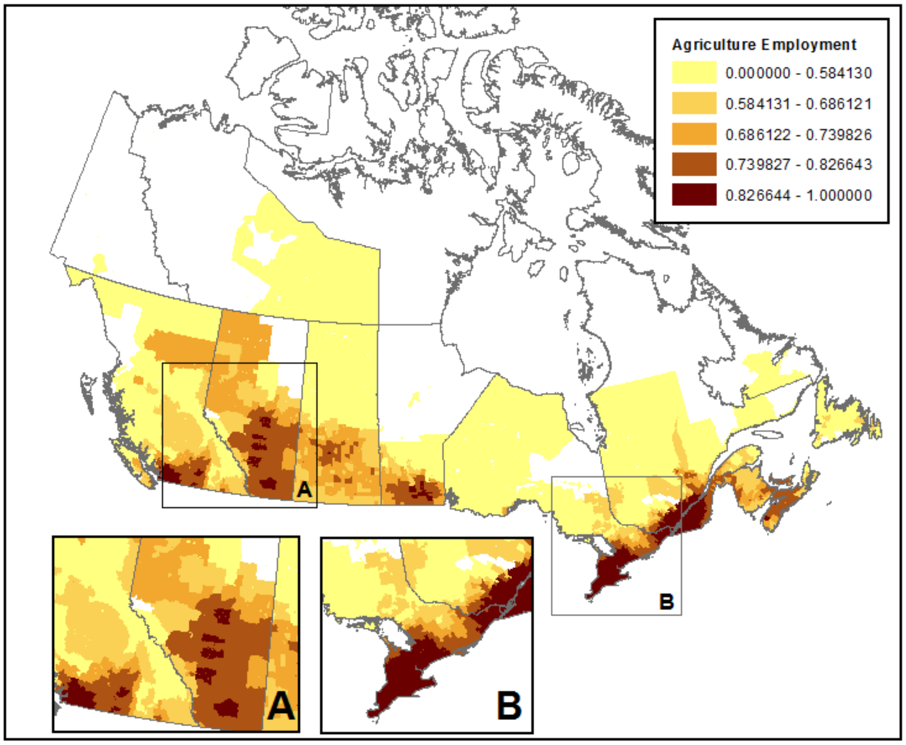
Note: Smaller values indicate lower proximity to the indicated resource sector. Larger values indicate higher proximity.
Source: Authors' computations.

Map 1 shows the geographic distribution of these scores for all CSDs in Canada. This index uses the same employment-based metric with the value of the index of proximity increasing as the colours move from light shades to darker colours. These five shades represent quintiles based on the index of proximity, each containing an equal number of CSDs. Unshaded areas indicate communities which did not fall within the proximity thresholds (2.5 hours' travel time) of any resource areas. Map 2 shows the revenue-based agricultural proximity, in which the Prairie region is the largest top-quintile (dark brown) region.

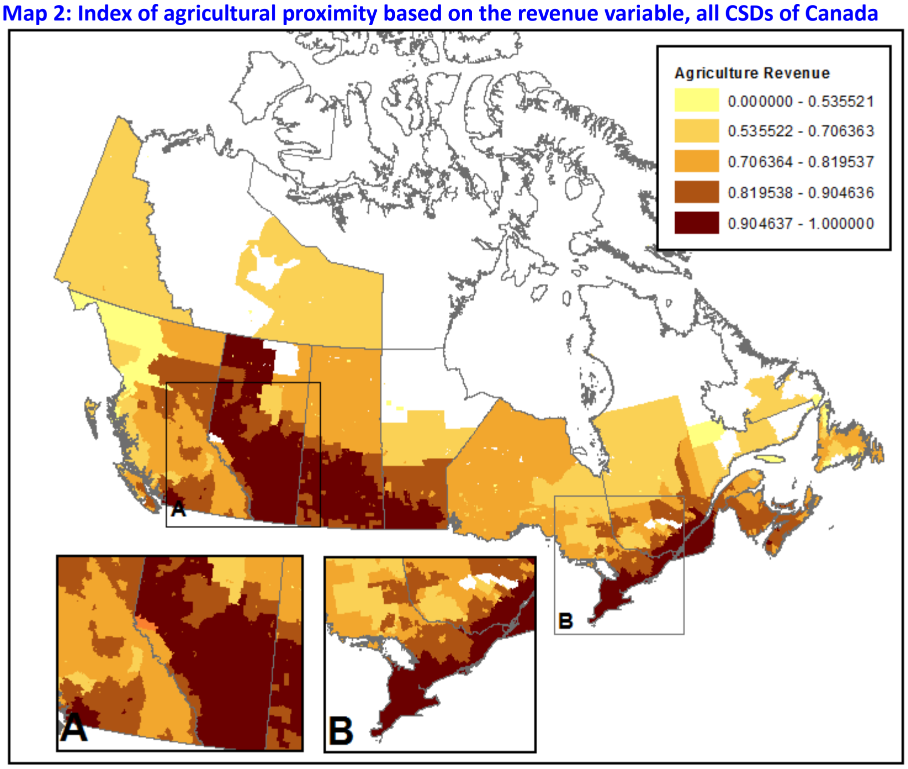
The spatial pattern reflects the distribution of farming areas of Canada. The lowest values of the index are recorded for the most northern regions, where the climate is less conducive to agriculture, whereas the geographically largest areas of high proximity to agriculture are found in southern Ontario and Quebec. Note that the index is employment-based, so it is not surprising that some areas in the prairies fall into lighter shades.

The maps showing other proximity scores (employment-based and revenue-based) for all resource sectors combined and for the other four resource sectors are reported in Appendix 1.

Map 1: Index of agricultural proximity based on the employment variable, all CSDs of Canada



Map 2: Index of agricultural proximity based on the revenue variable, all CSDs of Canada



4.4. Decomposition of the global index value

For each resource sector, the global index of proximity was decomposed into six sub-indices of proximity. Each sub-index measures proximity to one of the types of resource area in the classification scheme described in Figure 1. When using an index in its original value (not scaled or log transformed), the sub-indices add up to the global proximity index (in its non-rescaled form).

Table 8 presents summary results for all CSDs of Canada reporting some value of the global proximity index for the corresponding resource sector. The index values are in the original form, while the percentage values represent the contribution of each component (proximity to type of resource areas) to the global proximity measure. For instance, the average value of the global proximity index for the agriculture sector is $48.8 million (per minute of travel time). When this index value is decomposed by type of area, the result is that, on average, the largest contribution is due to proximity to “major-project diversified” communities. (These are communities with: over $10 million in total revenue generated by agriculture; revenue of agricultural businesses less than 30% of the total business revenue of that resource community; and size of the largest three agricultural businesses more than $10 million in revenue).

The energy (oil and gas) index stands out, with almost all (98%) of this index coming from major-project diversified CSDs.

Table 8: Decomposition of the proximity index: average values of each global index and component sub-indices for CSDs having a positive value of the global index

	Agriculture		Forestry		Mining		Fishing		Oil/Gas	
	Index ('000 $/min)	**%**	**Index ('000 $/min)**	**%**	**Index ('000 $/min)**	**%**	**Index ('000 $/min)**	**%**	**Index ('000 $/min)**	**%**
All Types	**48,787**	**100%**	**4,439**	**100%**	**27,742**	**100%**	**1,195**	**100%**	**39,734**	**100%**
Major-project reliant	12,052	25%	225	5%	2,739	10%	57	5%	235	1%
Cluster reliant	7,319	15%	21	0%	104	0%	31	3%	-	0%
Major-project diversified	20,053	41%	2,796	63%	24,276	88%	498	42%	38,870	98%
Cluster diversified	6,562	13%	253	6%	29	0%	196	16%	235	1%
Small reliant	766	2%	76	2%	3	0%	34	3%	9	0%
Small diversified	2,035	4%	1,069	24%	590	2%	378	32%	385	1%

Source: Authors’ computations.

The values presented in Table 8 can be used to benchmark different groups of CSDs. An example is provided by Table 9, which shows the value of each component of the proximity indices for the CSD of Prince George, BC. Not surprisingly, the largest original value of the indices is reported for the forestry sector ($75.6 million per minute of travel time), and the vast majority of this value (97%) is due to proximity to major-project diversified communities. (Note that the index value includes proximity to Prince George itself). Proximity to mining and agriculture also play a sizable role (being roughly between $7 million/min and $8 million/min) and also in this case they are almost entirely due to proximity to major-project diversified resource areas.

Table 9: Decomposition of the proximity index: component values for a specific community (Prince George)

Resource sector	Community type	Proximity measure (original value)	%
Forestry	All Types	75,572,056	100%
	Major-project reliant	1,732,733	2%
	Cluster reliant	-	0%
	Major-project diversified	73,105,771	97%
	Cluster diversified	-	0%
	Small reliant	35,718	0%
	Small diversified	697,834	1%
Mining	All Types	8,542,419	100%
	Major-project reliant	-	0%
	Cluster reliant	-	0%
	Major-project diversified	8,466,517	99%
	Cluster diversified	-	0%
	Small reliant	-	0%
	Small diversified	75,902	1%
Agriculture	All Types	7,227,094	100%
	Major-project reliant	-	0%
	Cluster reliant	-	0%
	Major-project diversified	6,212,407	86%
	Cluster diversified	240,257	3%
	Small reliant	20,816	0%
	Small diversified	753,613	10%
Fishing	All Types	25,287	100%
	Major-project reliant	-	0%
	Cluster reliant	-	0%
	Major-project diversified	-	0%
	Cluster diversified	-	0%
	Small reliant	-	0%
	Small diversified	25,287	100%
Oil/Gas	All Types	642,291	100%
	Major-project reliant	-	0%
	Cluster reliant	-	0%
	Major-project diversified	173,781	27%
	Cluster diversified	-	0%
	Small reliant	-	0%
	Small diversified	468,510	73%

Source: Authors’ computations.

4.5. Comparisons and relationships between indices

Table 10 shows simple statistics for the employment-based proximity indices. Statistics for imputed versions of these indices, in which missing (not calculated) values of the index are replaced by zero, are shown in the grey-shaded rows. Notably, amongst the non-imputed indices, the fisheries and aquaculture index was calculated for the fewest (3,561) CSDs. While eliminating the cut-off time of 2.5 travel hours would have increased the number of CSDs with an index value, this would have been inconsistent with previous CSBP work.

Table 10: Simple statistics, resource sector scaled employment-based proximity scores of CSDs

Variable	N	Mean	Minimum	Maximum
Agriculture, employment-based, no imputation	4,882	0.68987	0	1.0
Agriculture, employment-based, zero-imputed	5,253	0.64114	0	1.0
Fishing, employment-based, no imputation	3,561	0.49260	0	1.0
Fishing, employment-based, zero-imputed	5,253	0.33393	0	1.0
Forestry, employment-based, no imputation	4,810	0.56362	0	1.0
Forestry, employment-based, zero-imputed	5,253	0.51609	0	1.0
Oil & gas, employment-based, no imputation	4,405	0.39518	0	1.0
Oil & gas, employment-based, zero-imputed	5,253	0.33138	0	1.0
Mining, employment-based, no imputation	4,741	0.56598	0	1.0
Mining, employment-based, zero-imputed	5,253	0.51082	0	1.0

Source: Authors' computations.

Table 11 below, a matrix of Pearson correlation coefficients, shows the strength of the linear relationship between each pair of proximity indices. This matrix shows that the agriculture proximity score and the mining proximity score have the strongest linear relationship, with a Pearson correlation coefficient of 0.625.

The forestry proximity score and the energy (oil & gas) proximity score have the strongest negative correlation, meaning that communities close to forestry sector tend to be far from energy sector and vice versa. The correlation between each pair of indices is statistically significant, with the one exception that fisheries is not significantly correlated with mining, meaning that the proximity of CSDs to fisheries enterprises is not related to their proximity to mining enterprises. Note that each CSD was weighted equally in this analysis; different results would be obtained if CSDs were weighted differently, such as by revenue, employment or population.

Table 11: Correlations between scaled employment-based proximity scores of CSDs

Pearson Correlation Coefficients
Prob > |r| under H_0: Rho=0
Number of Observations

	Agriculture	Fisheries	Forestry	Energy	Mining
Agriculture	1	0.13058	0.37384	0.28578	0.62512
		<.0001	<.0001	<.0001	<.0001
	4,882	3,513	4,733	4,384	4,692
Fisheries	0.13058	1	0.07523	-0.16283	0.00607
	<.0001		<.0001	<.0001	0.7229
	3,513	3,561	3,498	3,165	3,415
Forestry	0.37384	0.07523	1	-0.29080	0.23094
	<.0001	<.0001		<.0001	<.0001
	4,733	3,498	4,810	4,305	4,601
Energy	0.28578	-0.16283	-0.29080	1	0.29500
	<.0001	<.0001	<.0001		<.0001
	4,384	3,165	4,305	4,405	4,369
Mining	0.62512	0.00607	0.23094	0.29500	1
	<.0001	0.7229	<.0001	<.0001	
	4,692	3,415	4,601	4,369	4,741

Source: Authors' computations.

Table 12: Correlations between imputed scaled employment-based proximity scores of CSDs

Pearson Correlation Coefficients, N = 5,253
Prob > |r| under H_0: Rho=0

	Agriculture	Fisheries	Forestry	Energy	Mining
Agriculture	1	0.36920	0.62918	0.64341	0.80756
		<.0001	<.0001	<.0001	<.0001
Fisheries	0.36920	1	0.48308	-0.01936	0.20906
	<.0001		<.0001	0.1607	<.0001
Forestry	0.62918	0.48308	1	0.17962	0.54555
	<.0001	<.0001		<.0001	<.0001
Energy	0.64341	-0.01936	0.17962	1	0.62324
	<.0001	0.1607	<.0001		<.0001
Mining	0.80756	0.20906	0.54555	0.62324	1
	<.0001	<.0001	<.0001	<.0001	

Source: Authors' computations.

Table 13 shows that the non-imputed revenue-based proximity score was calculated for more CSDs than the employment-based score (Table 10). This is because some enterprises are non-employers but still have revenue. For comparison, simple statistics for zero-imputed versions of the variables are also given, in grey-shaded rows.

Table 13: Simple statistics: revenue-based proximity indices (scaled), all CSDs

Variable	N	Mean	Minimum	Maximum
Agriculture, revenue-based, no imputation	4,994	0.85764	0	1.0
Agriculture, revenue-based, zero-imputed	5,253	0.81535	0	1.0
Fishing, revenue-based, no imputation	4,929	0.71097	0	1.0
Fishing, revenue-based, zero-imputed	5,253	0.66712	0	1.0
Forestry, revenue-based, no imputation	4,992	0.74803	0	1.0
Forestry, revenue-based, zero-imputed	5,253	0.71087	0	1.0
Oil & gas, revenue-based, no imputation	4,597	0.64933	0	1.0
Oil & gas, revenue-based, zero-imputed	5,253	0.56824	0	1.0
Mining, revenue-based, no imputation	4,922	0.73174	0	1.0
Mining, revenue-based, zero-imputed	5,253	0.68563	0	1.0

Source: Authors' computations.

Table 14 below is a correlation matrix of the revenue-based resource area proximity scores. As in Table 11, the highest pair-wise correlation is between the Mining proximity index and the Agriculture proximity index. Again, the lowest correlation is between the forestry proximity index and the energy (oil & gas) index. As before, the correlation is slightly negative, indicating that CSDs that are close to the forestry sector tend to be far from the energy sector.

Table 14: Correlations between scaled revenue-based proximity scores of CSDs

Pearson Correlation Coefficients
Prob > |r| under H_0: Rho=0
Number of Observations

	Agriculture	Fisheries	Forestry	Energy	Mining
Agriculture	1	-0.01404	0.32626	0.62041	0.66399
		0.3267	<.0001	<.0001	<.0001
	4,994	4,882	4,961	4,587	4,903
Fisheries	-0.01404	1	0.28027	-0.00261	0.05145
	0.3267		<.0001	0.8599	0.0003
	4,882	4,929	4,877	4,567	4,837
Forestry	0.32626	0.28027	1	-0.12299	0.32398
	<.0001	<.0001		<.0001	<.0001
	4,961	4,877	4,992	4,582	4,901
Energy	0.62041	-0.00261	-0.12299	1	0.52310
	<.0001	0.8599	<.0001		<.0001
	4,587	4,567	4,582	4,597	4,585
Mining	0.66399	0.05145	0.32398	0.52310	1
	<.0001	0.0003	<.0001	<.0001	
	4,903	4,837	4,901	4,585	4,922

Source: Authors' computations.

Table 15: Correlations between imputed scaled revenue-based proximity scores of CSDs

Pearson Correlation Coefficients, N = 5,253
Prob > |r| under H_0: Rho=0

	Agriculture	Fisheries	Forestry	Energy	Mining
Agriculture	1	0.59158 <.0001	0.78530 <.0001	0.76370 <.0001	0.86080 <.0001
Fisheries	0.59158 <.0001	1	0.64008 <.0001	0.46233 <.0001	0.56456 <.0001
Forestry	0.78530 <.0001	0.64008 <.0001	1	0.53081 <.0001	0.73413 <.0001
Energy	0.76370 <.0001	0.46233 <.0001	0.53081 <.0001	1	0.75014 <.0001
Mining	0.86080 <.0001	0.56456 <.0001	0.73413 <.0001	0.75014 <.0001	1

Source: Authors' computations.

Figure 5 below is a bubble plot comparing the agriculture proximity score of each CSD with its mining proximity score. This size of each bubble in this plot is proportional to the population estimate of the CSD according to the 2011 Census. As with the top-10 tables (Table 16 and Table 17), this plot hints that adjusting the unscaled log scores by subtracting a similarly-calculated log index of proximity to population might be useful.

The proximity scores as shown in ***Error! Reference source not found.*** in certainly measure the absolute level of resource activity; however, for the purpose of measuring the per-capita impact of resource areas on residents of Canada, it appears that populous CSDs are over-represented amongst the highest unadjusted scores. If, for instance, the population surrounding a CSD is in the hundreds of thousands, then a medium level of resource sector employment probably would be relatively insignificant in terms of its effect per capita. On the other hand, the same level of resource sector employment would be very significant for a small isolated community.

Figure 5: Comparison of agriculture and mining revenue-based proximity scores

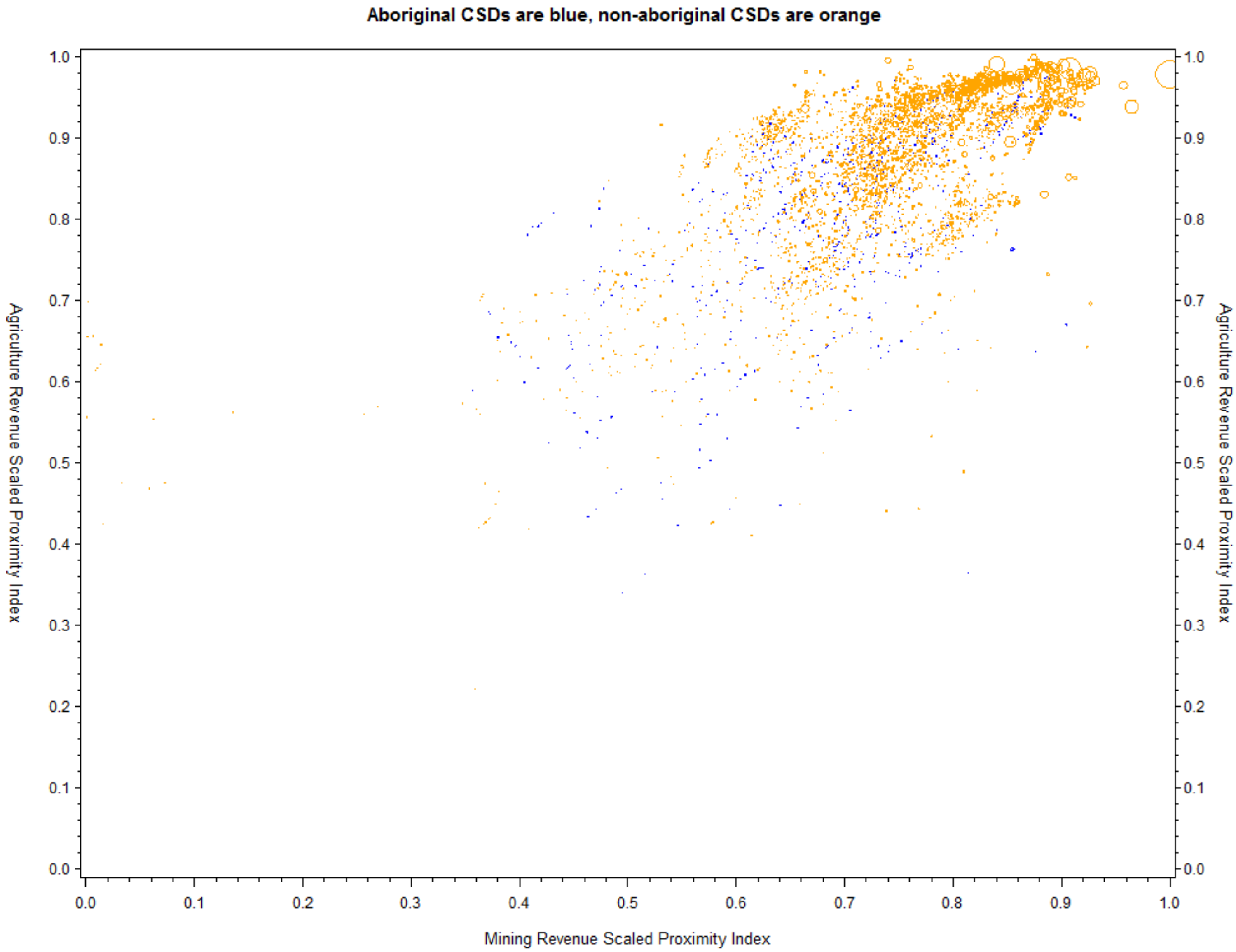


Figure 6 below shows the zero-imputed versions of the two variables in Figure 5. The only difference is the inclusion of the no-access CSDs along the axes (at the value of 0).

Figure 5: Comparison of agriculture and mining zero-imputed revenue-based proximity scores

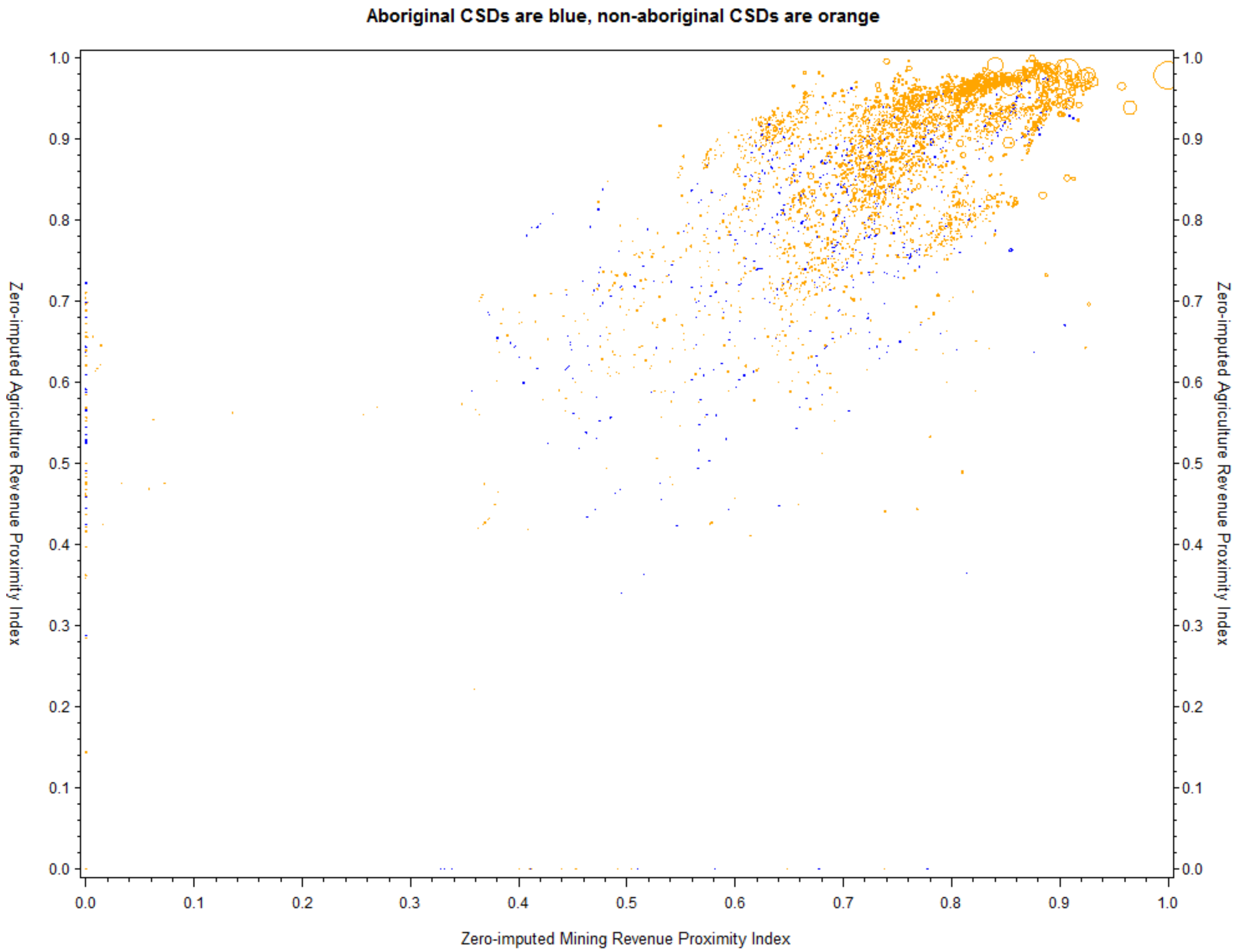


Figure 6 below is a bubble plot similar to Figure 5, but for forestry and energy (oil and gas), which are actually slightly negatively correlated.

Figure 6: Comparison of forestry and energy revenue proximity scores

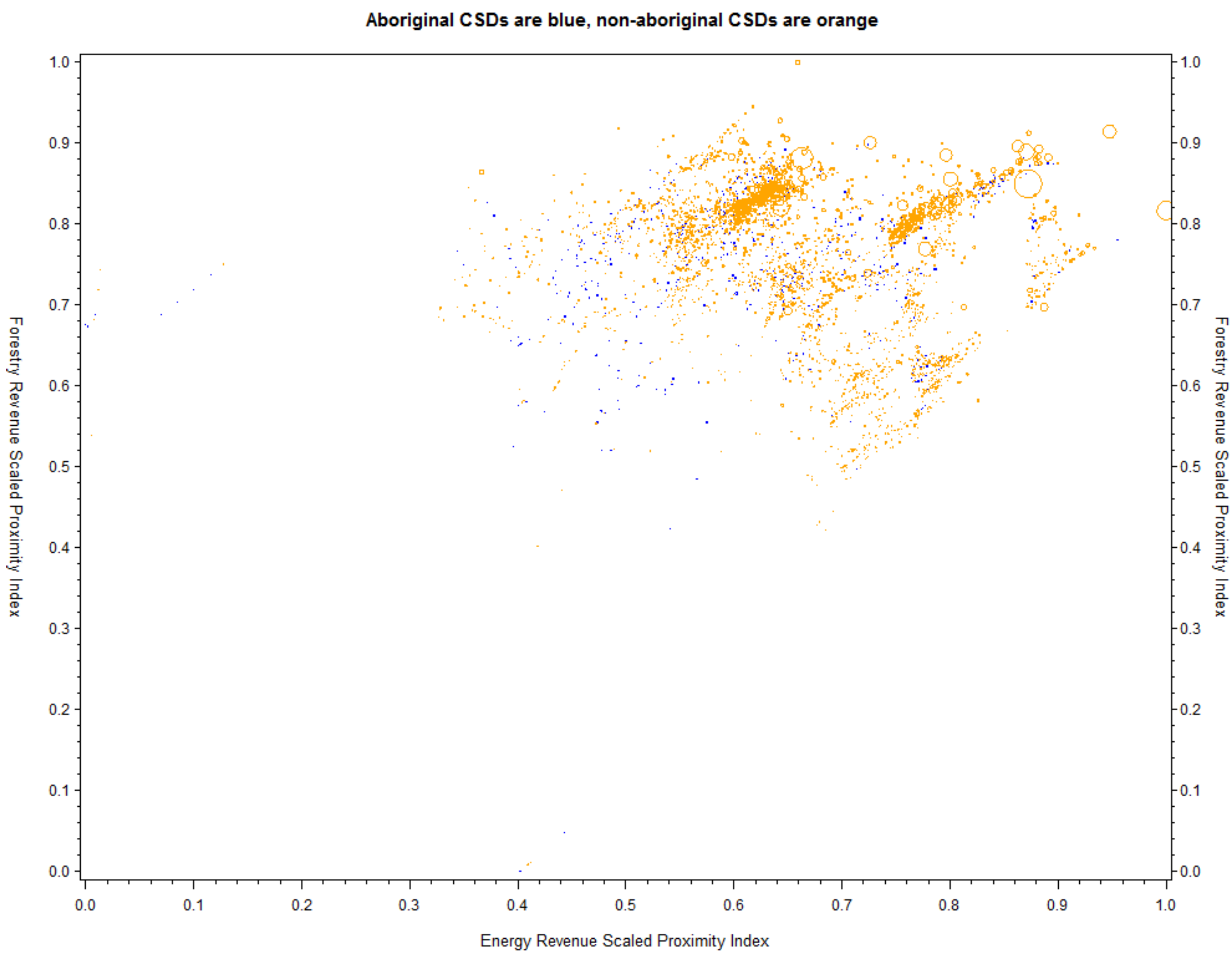
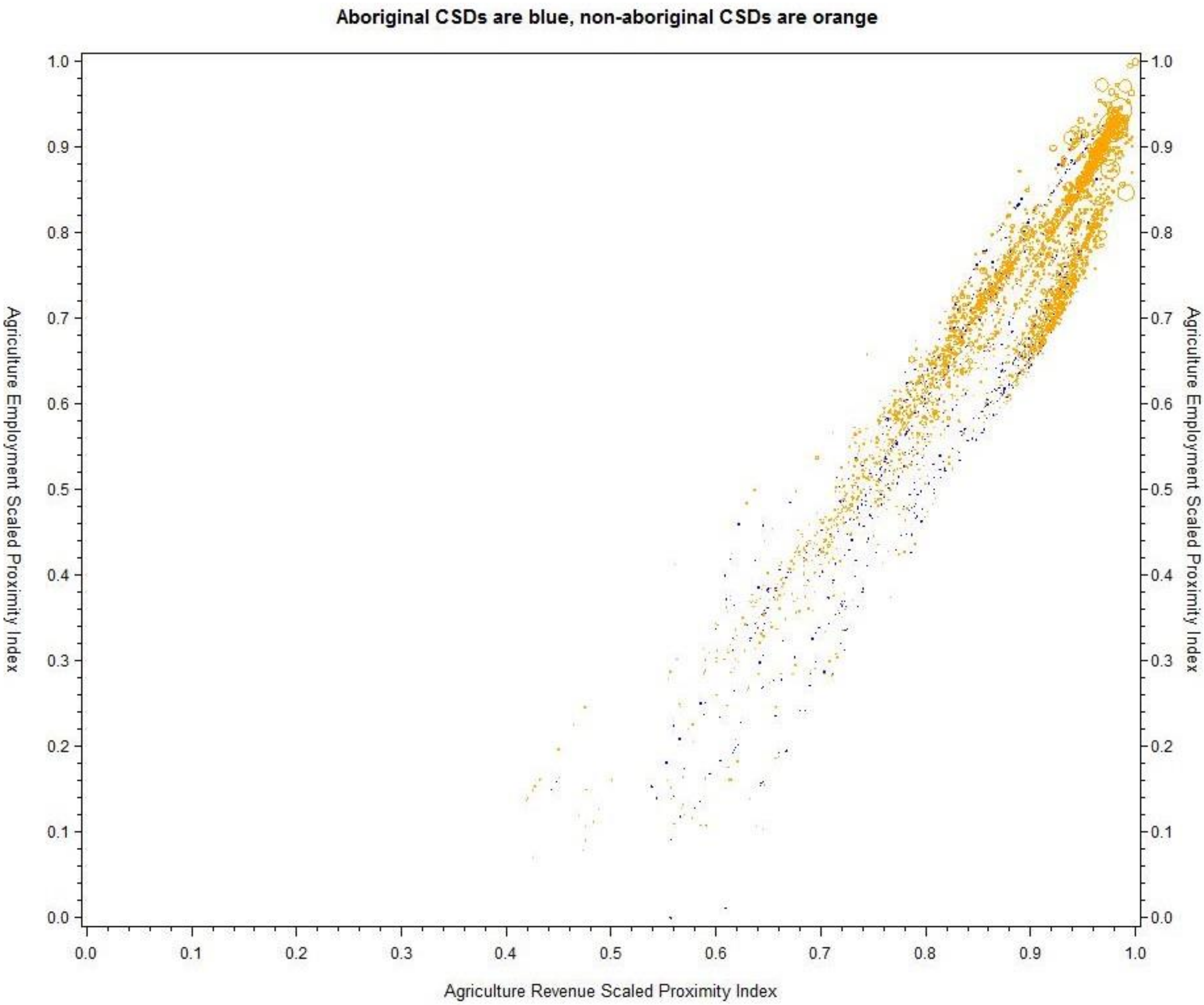


Figure 7 below is a bubble plot comparing the agriculture revenue-based proximity score with the agriculture employment-based proximity score. The size of each bubble in this plot represents the population estimate of the CSD according to the 2011 Census. The correlation between the scores is quite high, as expected: it doesn't make much difference whether resource sector activity is measured by employment or total revenue (gross income). As in Figure 5 and Figure 7, the large bubbles in the top right corner represent populous CSDs.

Figure 7: Comparison of agriculture revenue-based and employment-based proximity scores



An analysis of indices for selected CSDs can also be undertaken. As an illustrative example, Table 16 shows the top 10 CSDs according to the agriculture scaled proximity index based on employment. This table sheds some light on one of the main challenges in developing and interpreting this database. For instance, urbanized CSDs could have high values if nearby CSDs have high agricultural employment. As an example, Hamilton is surrounded by CSDs with high agricultural employment (Lincoln, Niagara-on-the-Lake, and St. Catharines are among these CSDs). However, since the source for this index gives employment at the enterprise level, certain other urbanized CSDs might have higher values than one would expect when a large enterprise, with locations in other CSDs, is headquartered there (although this is not the case for Hamilton, which itself contains no large agricultural enterprise).

Table 16: Top 10 CSDs according to agriculture scaled proximity index based on employment

Name of CSD	Agriculture employment scaled proximity index
Abbotsford	1.000
Norfolk County	0.995
Leamington	0.973
Surrey	0.973
Hamilton (City)	0.971
Langley	0.965
Chatham-Kent	0.963
Matsqui Main 2	0.963
Lincoln	0.960
Kingsville	0.955

Source: Authors' computations.

Table 17 below is similar to Table 16 but shows the revenue-based version of the index instead. This index would appear more accurate (based on expectations of geographic distribution of economic activities); urban areas like Hamilton drop out of the top 10 CSDs most proximate to agriculture sector activities (although they still report high values of the score). This might also be due to the fact that many farms do not pay payroll tax (since the workers are family members or self-employed), but still report revenue, so they would contribute to the revenue-based index but not the employment-based index.

Table 17: Top 10 CSDs according to agriculture scaled proximity index based on revenue

Name of CSD	Agriculture revenue scaled proximity index
Abbotsford	1.000
Huron East	0.997
Lethbridge County	0.997
West Perth	0.996
Chatham-Kent	0.996
Norfolk County	0.995
North Perth	0.995
Zorra	0.993
Brant	0.993
South-West Oxford	0.993

Source: Authors' computations.

5. CSD-level indicators

A test was undertaken to develop a set of indicators to assess the impact of resource areas on economic and social change. Two main indicators were created in this domain: (1) One-year CSD employment growth; and (2) One-year CSD revenue growth.

These indicators were calculated using BR microdata, for agriculture, fishing, forestry, mining, energy, all-resources, non-resources, and all-industries (excluding public administration). Two time periods were used:

- January 2013 Business Register (measuring the state of affairs in the twelve-month period ending roughly mid-2012) to January 2014 Business Register (measuring the state of affairs in the twelve-month period ending roughly mid-2013)
- January 2014 Business Register (measuring the state of affairs in the twelve-month period ending roughly mid-2013) to January 2015 Business Register (measuring the state of affairs in the twelve-month period ending roughly mid-2014)

Because of the possibility of releasing sensitive data, these indicators were converted to equally-sized quintiles. Although this conversion to quintiles loses some information, the quintiles have the advantage that outliers do not distort the results. Although these variables may have potential in future, there were implemented only for expoloratory purposes and to generated a first understanding of data patterns.

A preliminary analysis of these indicators is presented in (Table 18). As can be seen in the grey-shaded leading diagonal of the correlation matrix in Table 18, the January 2014 employment growth quintile is negatively correlated with the January 2015 employment growth quintile for all five resource sectors, although the correlation is not statistically significant for mining and energy.

Table 18: Correlations between January 2014 and January 2015 BR employment growth rates, CSD level, selected resource subsectors

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	Agriculture, January 2013 BR to January 2014 BR	Fisheries, January 2013 BR to January 2014 BR	Forestry, January 2013 BR to January 2014 BR	Mining, January 2013 BR to January 2014 BR	Energy, January 2013 BR to January 2014 BR
Agriculture, January 2014 BR to January 2015 BR	-0.13260 <.0001 2,322	0.05495 0.3349 310	0.01879 0.5478 1,026	-0.01206 0.7966 459	0.08355 0.0818 435
Fisheries, January 2014 BR to January 2015 BR	-0.09377 0.0967 315	-0.15930 0.0010 424	0.11856 0.0602 252	-0.06390 0.5132 107	0.01100 0.9313 64
Forestry, January 2014 BR to January 2015 BR	0.06153 0.0485 1,029	0.05478 0.3875 251	-0.10945 0.0002 1,181	-0.03730 0.5022 326	-0.07152 0.3082 205
Mining, January 2014 BR to January 2015 BR	0.03839 0.3838 517	-0.04064 0.6487 128	-0.01571 0.7641 367	-0.00294 0.9485 486	0.10908 0.1091 217
Energy, January 2014 BR to January 2015 BR	0.10371 0.0291 443	-0.17148 0.1720 65	-0.06229 0.3749 205	0.00198 0.9786 185	-0.00212 0.9650 432

Source: Authors' computations.

6. Conclusions

The analysis presented in this paper shows the potential of geocoding Business Register microdata, as well as some of the challanges that remains to be solved in further developing the prototype database on resource areas.

Geographic areas in close proximity to economic activity of resource sector enterprises can be identified using a gravity model that BR data aggregated at the CSD level with an inter-CSD distance matrix developed at Statistics

Canada. This model assumes that the economic activity of an enterprise occurs at the "representative point" of its operating address' CSD. This is a good approximation for most enterprises, but to date remains a limitation of the analysis. However, the detailed mapping of proximity scores may facilitate the identification of geographic areas for which further verification and validation would be needed.

The resulting proximity measure takes into account all economic activity within a distance of 2.5 hours by travel time, weighted by the inverse of the travel time. If no distance matrix had been available, the method that would have been employed would likely have taken into account only economic activity within the CSD, which is more orthodox, but less informative.

Creating this distance-derived measure allows the comparison of the proximity for selected types of communities (for instance, Aboriginal vs. Non-Aboriginal) to different resource sectors. Further, data from the 2011 Census and National Household Survey can easily be merged with these proximity measures. In this preliminary analysis, one-year CSD revenue growth and employment growth were also added to the database (as 5-level categorical variables) for two consecutive time periods, and were found to show indications of mean reversion (cyclical behaviour). A longer time period, however, might lead to results of greater significance.

There are numerous refiniments that could implemented to improve and further develope this prototype database. First, an adjustment of proximity indices population density could be implemented. As mentioned above in the discussion of Figure 5, it may be worthwhile to adjust the proximity log indices by subtracting an appropriately scaled "population" proximity log index. This would be equivalent to dividing the raw (before log) index by an appropriately scaled index of population proximity. This adjusted index would show each person's travel-time-corrected share of resource sector business activity: it would be a "per capita" version of the unadjusted index that would show the CSDs where there is an excess or deficit of business activity, which could identify communities undergoing labour shortages or economic downturns. Second, alternative metrics to the three dimensional classification of CSDs could be explored. Along these lines, alternative definition of sector reliance could be considered. For instance, the degree of sector reliance could be measured by the raw (not scaled) revenue-based proximity index for the sector divided by the all-industries index. This alternative definition would take into account all CSDs when calculating the reliance of a CSD on a particular industrial sector.

7. References

Aboriginal Affairs and Northern Development Canada (2009). Federal Framework for Aboriginal Economic Development. https://www.aadnc-aandc.gc.ca/eng/1100100033501/1100100033522.

Aboriginal Affairs and Northern Development Canada (2011). Annual Report: Indian Oil and Gas Report 2010-2011. http://www.pgic-iogc.gc.ca/DAM/DAM-INTER-IOGC/STAGING/texte-text/pubs_ar1011_1336398682735_eng.pdf.

Agriculture and Agri-Food Canada (2014). Aboriginal Agriculture Initiative, Program Terms and Conditions for Proposal-Based Projects for Agriculture and Agri-Food Canada. http://www.absn.ca/TermsandConditionsoftheAAISPI.pdf.

Alasia, Alessandro, Frédéric Bédard and Gordon Reichert (2010). Remoteness and accessibility: a review of the literature and potential data sources for computation. Prepared for Aboriginal Affairs and Northern Development Canada.

Alasia, Alessandro, Frédéric Bédard and Julie Bélanger (2013). Index of remoteness of First Nations and Inuit communities: proposed methodology for discussion. Prepared for Aboriginal Affairs and Northern Development Canada.

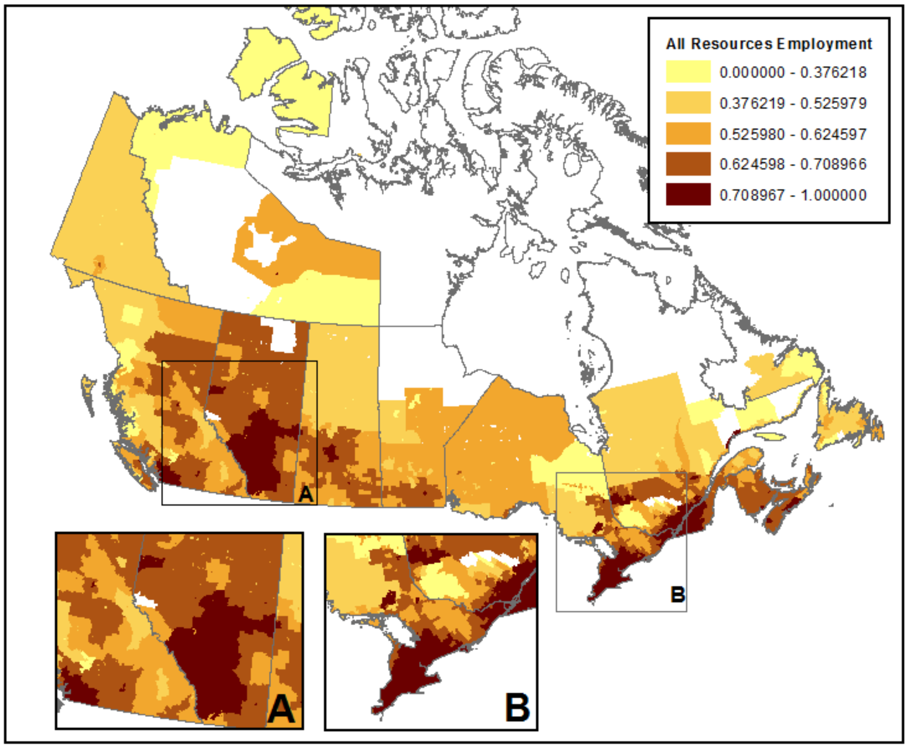
Natural Resources Canada (2012). Aboriginal participation in mining.
http://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/files/mineralsmetals/files/pdf/abor-auto/mining_infosheet_eng.pdf.

Natural Resources Canada (2014). Aboriginal Forestry Initiative.
http://www.nrcan.gc.ca/forests/federal-programs/13125.

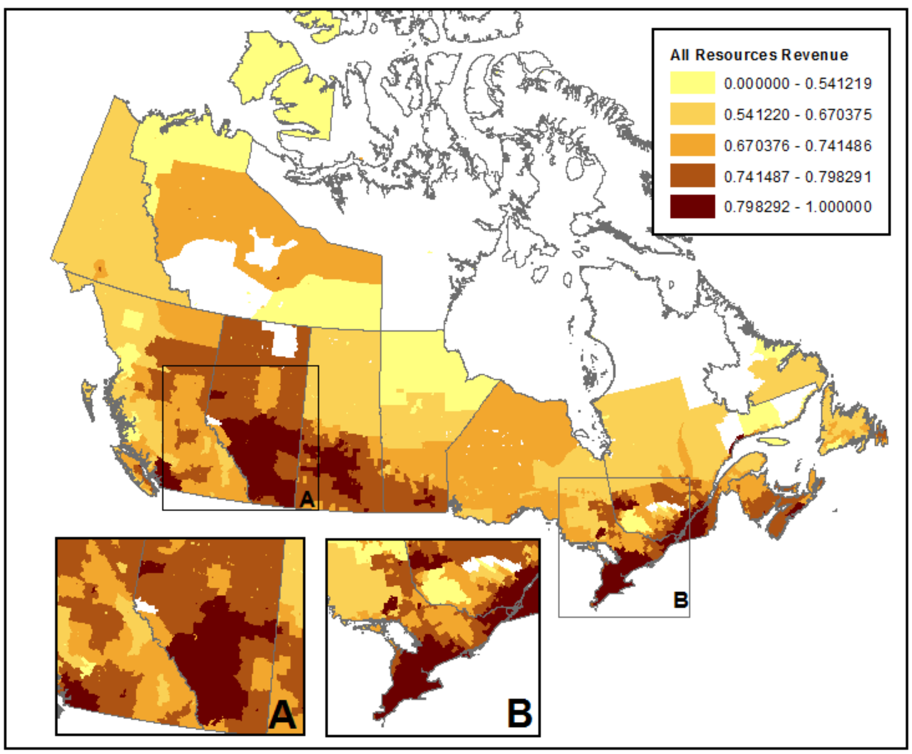
Natural Resources Canada (2014b). Interactive Map of Aboriginal Mining Agreements.
http://www2.nrcan.gc.ca/mms/map-carte/MiningProjects_cartovista-eng.html.

Appendix 1. Maps of global proximity scores

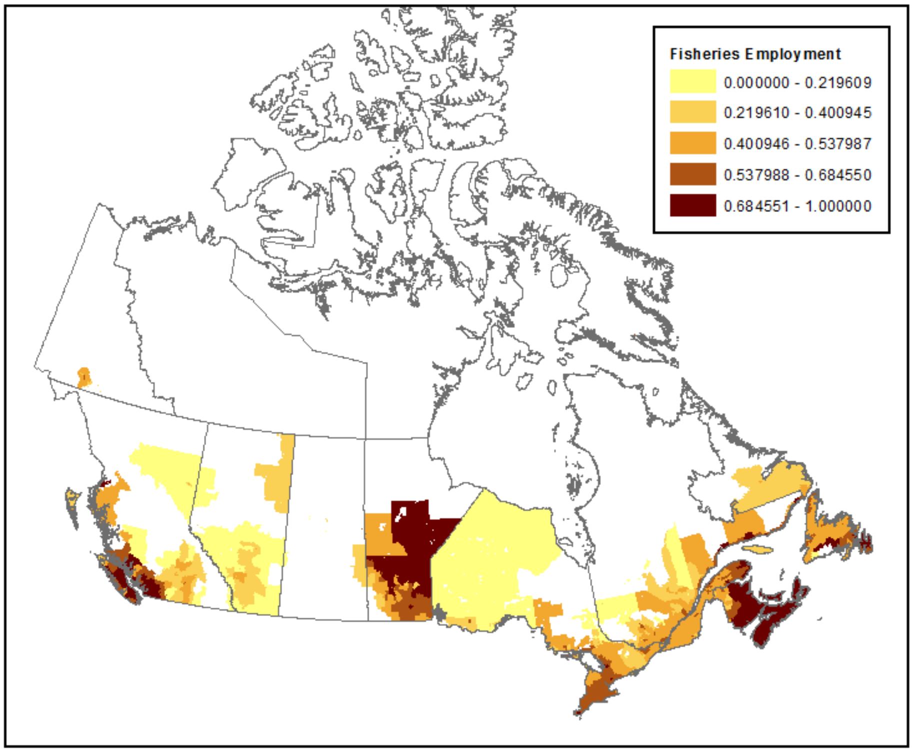
Map 3: Index of all-resources proximity based on the employment variable, all CSDs of Canada



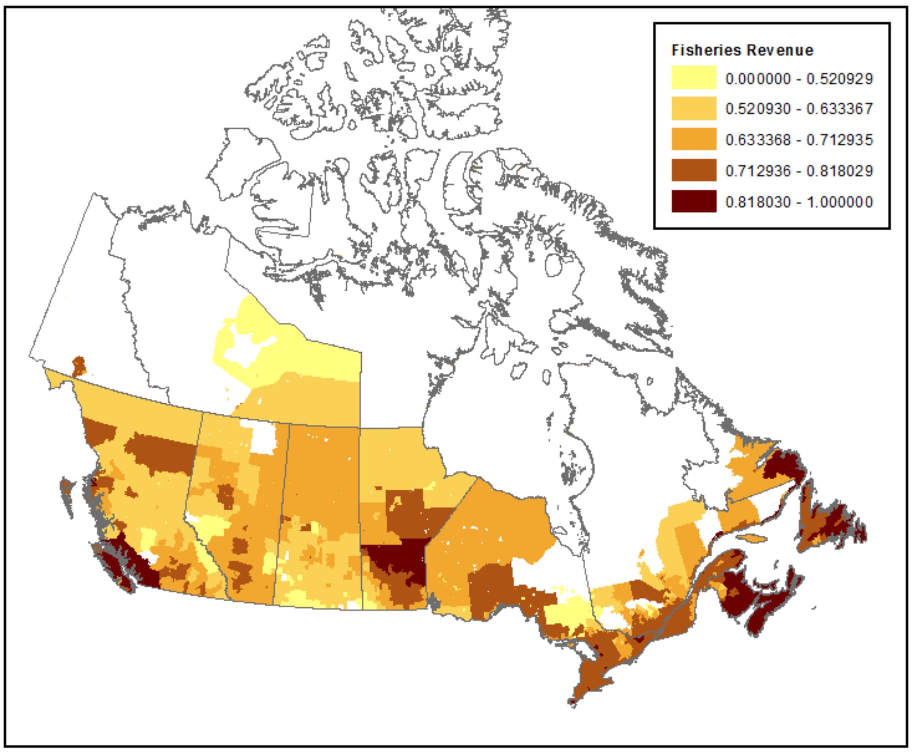
Map 4: Index of all-resources proximity based on the revenue variable, all CSDs of Canada



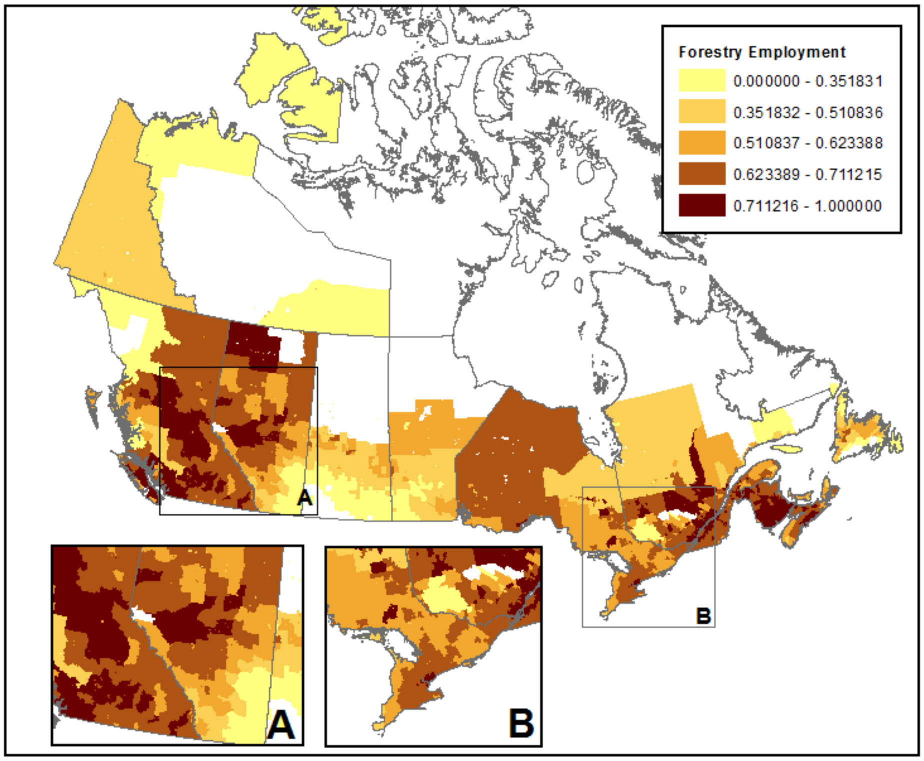
Map 5: Index of fisheries & aquaculture proximity based on the employment variable, all CSDs of Canada



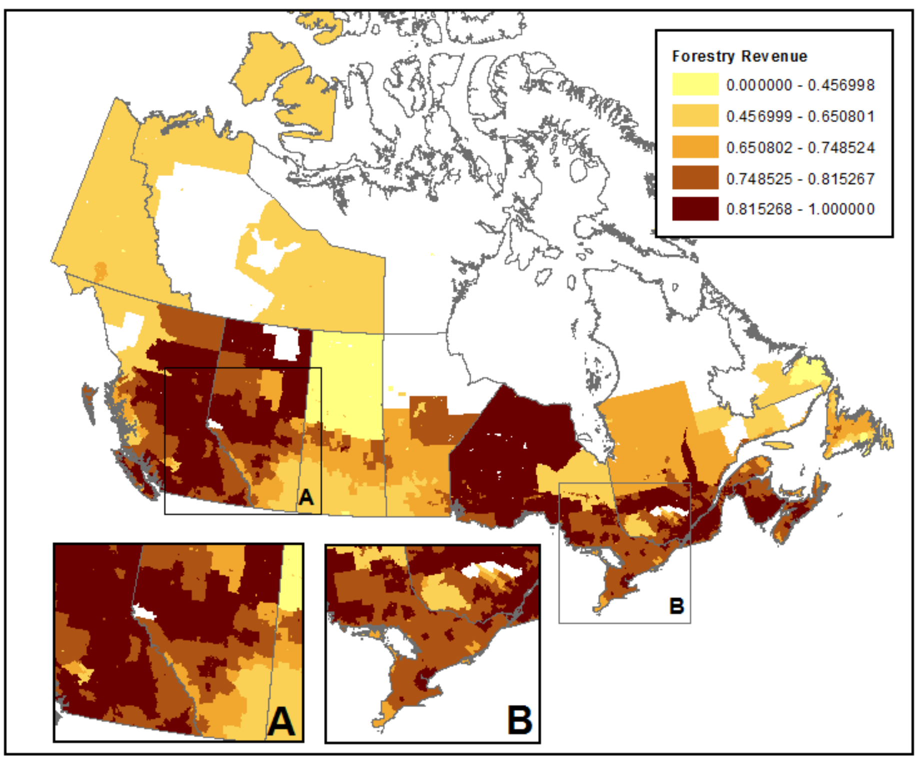
Map 6: Index of fisheries & aquaculture proximity based on the revenue variable, all CSDs of Canada



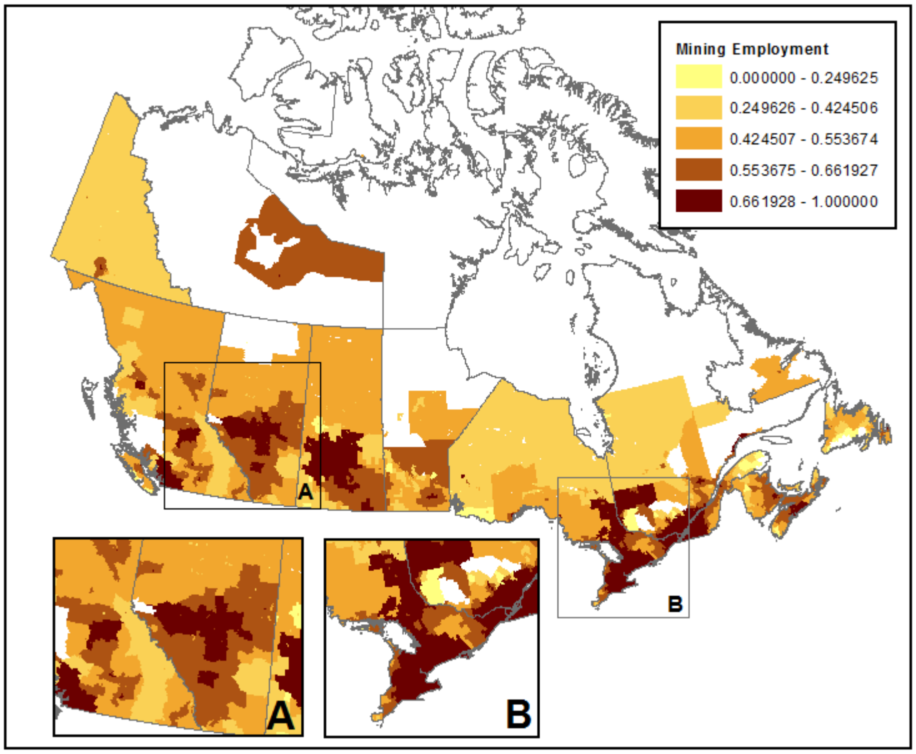
Map 7: Index of forestry proximity based on the employment variable, all CSDs of Canada



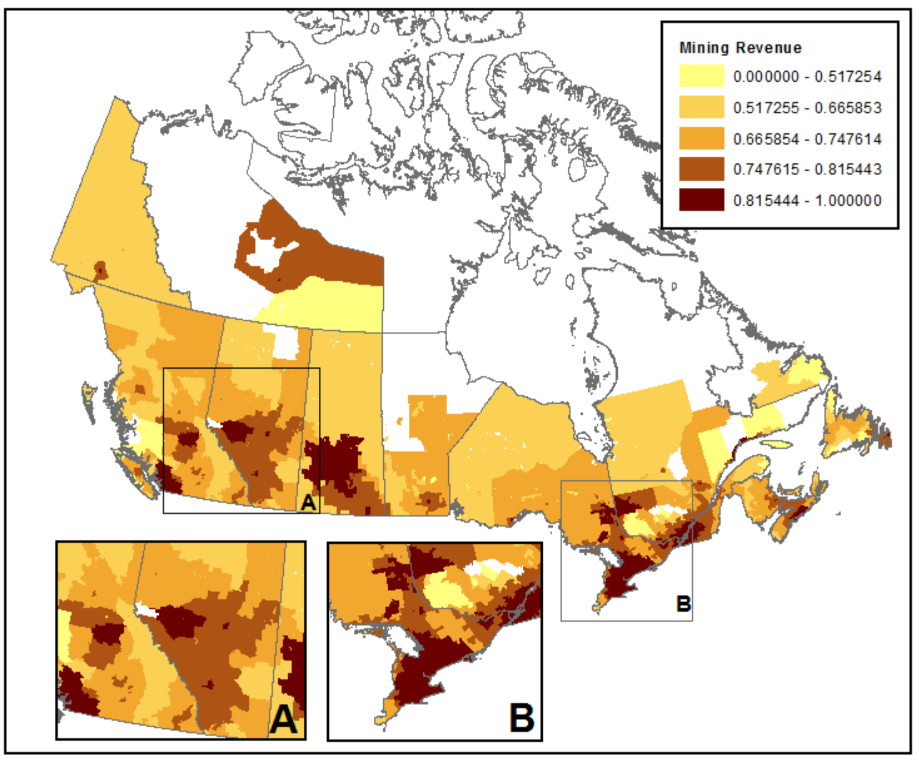
Map 8: Index of forestry proximity based on the revenue variable, all CSDs of Canada



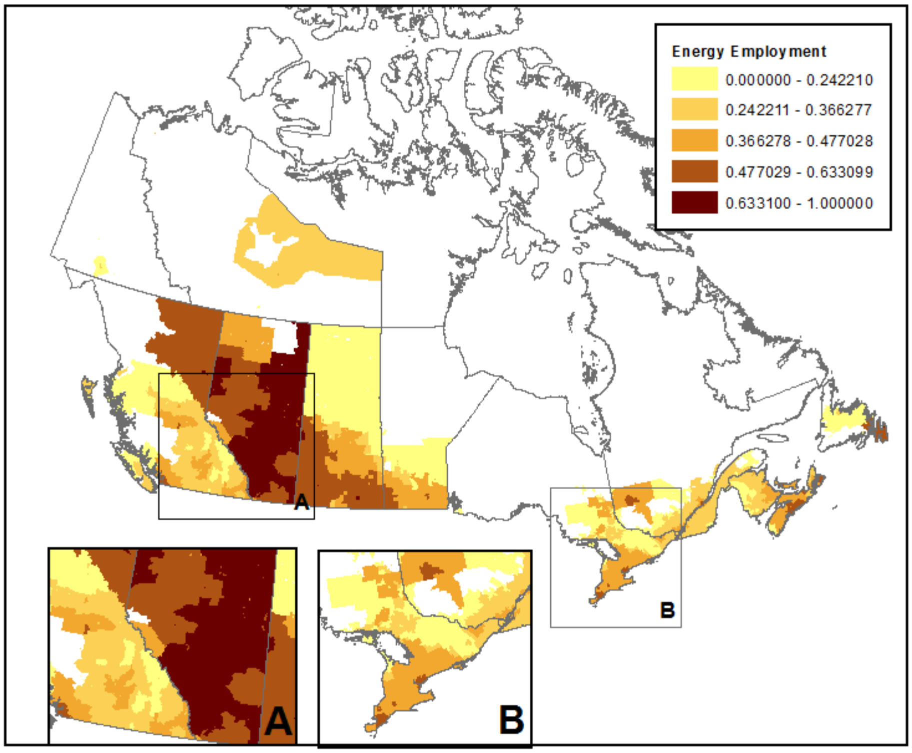
Map 9: Index of mining proximity based on the employment variable, all CSDs of Canada



Map 10: Index of mining proximity based on the revenue variable, all CSDs of Canada



Map 11: Index of energy proximity based on the employment variable, all CSDs of Canada



Map 12: Index of energy proximity based on the revenue variable, all CSDs of Canada