

# The Chilean Internal Migration (CHIM) database: A Temporally Consistent Spatial Data Framework for the Analysis of Human Mobility

Francisco Rowe<sup>1</sup>

<sup>1</sup> University of Liverpool, Liverpool, UK (email: [F.Rowe-Gonzalez@liverpool.ac.uk](mailto:F.Rowe-Gonzalez@liverpool.ac.uk))

Received: 20 July 2017/Accepted: 22 July 2017

**Abstract.** Changes in zonal boundaries and coding schemes severely compromise temporal comparison of data. In Chile, the Population and Housing census is the only comprehensive source of internal migration data, but municipal boundaries and occupation and industry sector coding schemes have undergone various changes which hamper the temporal comparability of census data. This paper presents the Chilean Internal Migration database which provides a temporally consistent framework for the analysis of internal migration over a period covering twenty-five years from 1977 to 2002. Specifically, it offers a hierarchical system of 304 municipalities, 51 provinces, and 13 regions, with 10 occupational groups and 11 industrial sectors which are temporally consistent over the 1977-82, 1987-92 and 1997-2002 census intervals. The database can be downloaded from: <https://s3.amazonaws.com/geoda/data/CHIM.zip>.

## 1 Background

Spatio-temporal analysis presumes temporally consistent geographical and data coding frameworks. However, data sources are commonly affected by changes in geographical boundaries and by changes in the way information is registered. In Chile, the Population and Housing census represents the most comprehensive source of internal migration and socio-economic data. The Chilean censuses conducted in 1982, 1992 and 2002 were all affected by changes in municipal boundaries and various schemes have been used to classify information on education, industry and occupation.

### 1.1 Changes to census data

Internal migration is inherently a spatial process so changes in municipal boundaries present a challenge for its temporal analysis. In Chile, municipalities represent the smallest geographical unit to which the place of usual residence is recorded and internal migration data can be derived. Any alteration to boundaries directly compromises the comparability of migration statistics over time. Various changes affected municipal boundaries over the three censuses (Rowe 2013): between the 1982 census and 1992 census, 82 of out 335 municipalities underwent boundary shifts, affecting 2.2% of the total population and 247 thousand people; while between 1992 and 2002 censuses, only ten municipalities were affected – but these changes involved a much larger number of people (321 thousand), increasing the total number of municipalities from 335 in 1992 to 342 in 2002 due to the division of existing areas and creation of new municipalities.

Additionally, census records also underwent changes in coding classification schemes. Different classification systems have been employed to organize information on occupations, industry sectors, and education categories for each census. Occupations were recorded using the Americas Classification of Occupations (COTA-70-developed by the Statistical Institute of American States (IASI)) at the 1982 census but this system was replaced for the International Standard Classification of Occupations (ISCO-88 -developed by the International Labour Organization (ILO)) at the 1992 and 2002 censuses. To classify data on industry sectors, the International Standard Industrial Classification of All Economic Activities (ISIC) Revision 2 was used at the 1982 census – which was then replaced for the ISIC Revision 3 for the 1992 and 2002 censuses. To record information on education, different national classification systems designed by the Chilean National Statistical Institute (INE) have been used for each of the three censuses.

### 1.2 Harmonization of census data

All these changes hinder spatio-temporal analysis of census data. To enable this, [Rowe, Bell \(2013\)](#) built a database, labelled Chilean Internal Migration (CHIM) database, based on census micro-data by harmonizing municipal boundaries and using consistent coding schemes. Census micro-data were obtained from the INE. To harmonize municipal boundary changes, [Rowe, Bell \(2013\)](#) used two procedures. For boundary shifts that occurred between 1982 and 1992, they implemented a ‘pseudo construct designer zones’ approach. It involves the construction of purpose-built zones from smaller building-bricks to harmonize the zonal system from different census periods based on a common set of boundaries. For boundary changes that occurred between 1992 and 2002, [Rowe, Bell \(2013\)](#) used a ‘freeze history’ approach. This approach consists of freezing the zonal system at a certain point in time and systematically tracks subsequent alterations in geographical boundaries. In doing so, later observations can be adjusted back to the original geography. For details on these procedures, see [Rowe, Bell \(2013\)](#).

To harmonize education, occupation and industry sector data, the United Nations’ international standard classification systems were used. Education data were recoded into five categories based on the International Standard Classification of Education, ISCED-1997 developed by The United Nations Educational Scientific and Cultural Organization ([UNESCO 1997](#)). To classify industry and occupation data, the ISIC Rev. 3 and ISCO-88 (respectively), which were used for the 1992 and 2002 censuses, were employed. Data from the 1982 census were mapped onto these schemes and industry and occupation information for all three censuses was then organized into 11 industry sectors and 10 occupational groups.

### 1.3 CHIM database

The outcome from the above procedures is a temporally consistent geography comprising 304 municipalities, 51 provinces, and 13 regions ([Figure 1](#)). The resulting municipal areas were defined as Temporal Municipalities (TMs) and the resulting hierarchical zonal system was classified using the INE’s current geographical classification system. Regions were numbered using the conventional order from 01 to 13 from Tarapacá to the Metropolitan Region. Provinces were numbered incrementally within each region using a three-digit identifier. So, for example, provinces within Tarapacá were numbered 011, 012 and 013, and within Los Lagos 101, 102, 103, 104 and 105. TMs were also numbered incrementally within provinces using a five-digit identifier. So, for example, TMs within the province of Iquique were numbered from 01101, 01102, and 01103 up to 01106, and within Llanquihue from 10101 up to 10108. The original municipality names at the 1992 census were preserved, and in cases where municipalities were amalgamated, the name of the municipality with the largest population in 1992 was adopted.

The CHIM database covers the full population at each census year, 1982, 1992 and 2002, and is a composite structure of multiple data sets: a set of micro-data, and a set of four aggregate data files<sup>1</sup>. [Figure 2](#) shows the aggregate files which consist of (1)

---

<sup>1</sup>Data from the 2012 Chilean census are not included in the CHIM database. This census has been questioned due to problems of imputation of unobserved housing units, with no identification of imputed

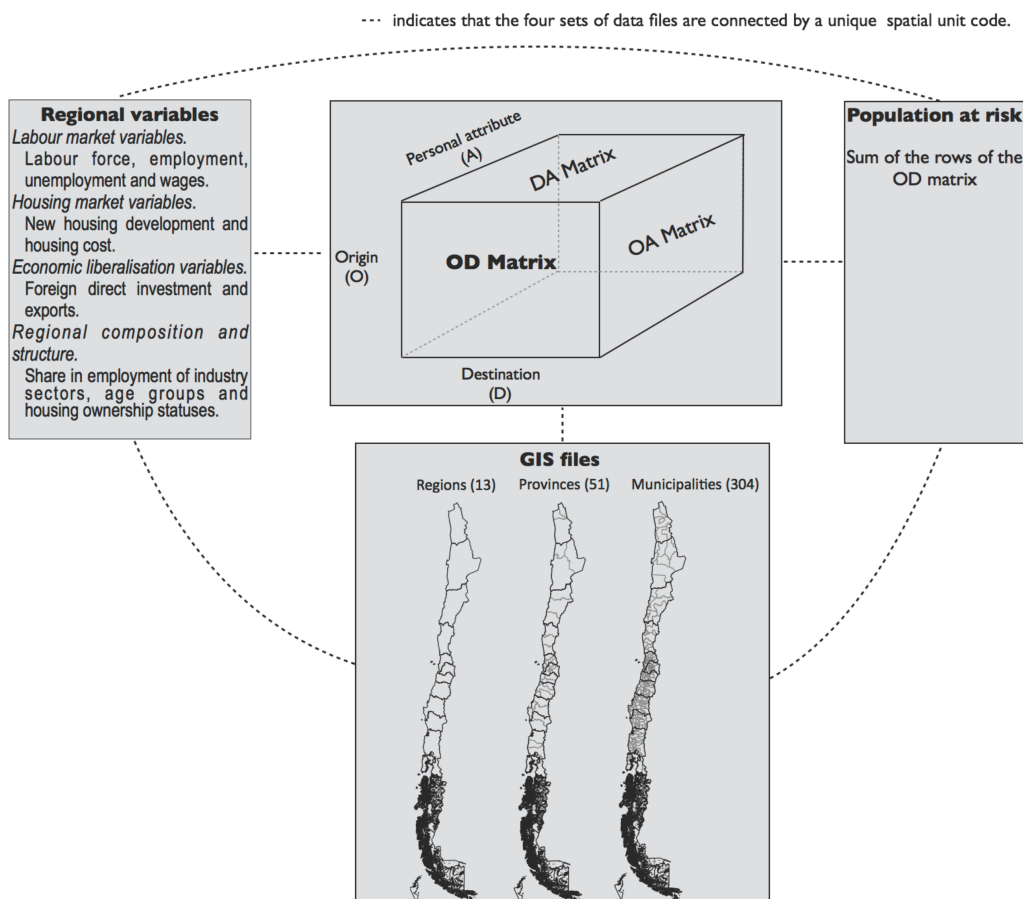


*Notes:* TAR: Tarapacá, ANT: Antofagasta, ATA: Atacama, COQ: Coquimbo, VAL: Valparaíso, MR: Metropolitan Region, OHI: O'higgins, MAU: Del Maule, BIO: Bío-bío, ARA: Araucanía, LAG: Los Lagos, AIS: Aisen, MAG: Magallanes. A full list of province and municipality names and codes is available: <https://s3.amazonaws.com/geoda/data/CHIM.zip>

Figure 1: Temporally consistent geography for Chile

origin-destination migration matrices, (2) populations at risk, (3) digital boundaries, and (4) regional contextual variables. Origin-destination migration matrices and populations at risk were derived from census micro-data based on the temporally harmonized geography described above. Digital boundaries to match this geography were created by using 1992 and 2002 census municipal and census district digital boundaries obtained from the INE and the United Nations' Latin American and Caribbean Demographic Centre (UN-CELADE). The data set on regional attributes was built by drawing on information from the three censuses and external sources – including the Foreign Investment Committee (FIC), the Ministry of Housing and Urban Development (MINVU), the Chilean Chamber of the Construction Industry (CCHC), [Dresdner, Sanhueza \(2009\)](#) and [Prado et al. \(2002\)](#).

versus collected census data ([Bianchini et al. 2013](#)). A recommendation has been made to not use these data ([Bravo et al. 2013](#)).



Note: For a list of personal characteristics available in the CHIM database, see <https://s3.amazonaws.com/geoda/data/CHIM.zip>.

Figure 2: Composite of aggregate data files in the CHIM database

## 2 Description of the resource

This section provides a description of the aggregate data sets from the CHIM database which are made available through this publication. Open access to census micro-data is not possible for confidentiality issues but customized micro-data requests can be made on <http://www.franciscorowe.com>. Four sets of data files are made available: (1) origin-destination migration matrices, (2) populations at risk, (3) digital boundaries, and (4) regional contextual variables. The content of each of these files are now described:

**Origin-destination migration matrices** This set of data files contains the output of 3,211 origin-destination migration matrices derived by cross-tabulating information on individuals' usual place of residence at the census date and five years earlier. The data cover three five-year intervals: 1977-1982, 1987-1992 and 1997-2002, and are stored in long format with each row corresponding to an origin-to-destination pair. They include the diagonal elements of the matrices, showing the number of people staying in the sample geographical area. The data are disaggregated by five-year age groups, 2 gender categories, 5 educational attainment levels, 5 employment classes, 3 labor force statuses, 11 industry sectors, and 10 occupational groups. The data are available at the TM scale and can be aggregated to provincial and regional geographical scales by using the respective identifiers in the metadata file.

**Populations at risk** This set of data files contains place-specific populations which correspond to the row sum of origin-destination migration matrices and indicate the population at the start of each census interval.

**Digital boundary** This is a folder which consists of a set of Geographic Information System (GIS) files, providing digital boundaries for three spatial scales: TMs, Provinces and Regions.

**Regional contextual variables** This data file contains, as indicated in Figure 1, a set of labor market, housing market, economic liberalization, and regional composition and structure variables.

These data have already been deployed to examine the determinants of long-distance commuting (Rowe 2014, Rowe, Bell forthcoming) and to build functional labor market areas for the analysis of labor migration (Casado-Díaz et al. 2017). However, they offer an opportunity for many other existing applications and future work will update the CHIM database to include data from the 2017 census.

### 3 Potential applications

Internal migration is widely recognized as an integral component of national development. As countries modernize and become globally connected, the intensity, forms, and patterns of population movement evolve, reshaping patterns of human settlement. Chile provides an ideal exemplar. Over the last four decades, Chile transitioned from a closed, centrally-planned economic system with a socialist government to a globally integrated, market-driven economy under military and democratic regimes. This has transformed Chile's space economy. Prior work has systematically traced the evolution of mobility in Chile and its connections to the socio-economic and political context, revealing a declining trend in the intensity and redistributing capacity of internal migration (Rowe 2013).

The described database offers a unique opportunity to understand the forces driving these declining trends in the context of a global decline in internal migration intensity. Identification of the causes underpinning the migration decline would provide empirical-based evidence to guide policy makers to ensure appropriate economic and social policy responses to the continuing decline in human mobility. This finding also has the potential to inform sub-national population and housing forecasting, and enable the formulation of policies that ensure that regions with economic opportunities can attract the required workers to enhance productivity and build resilient local economies.

The described database also offers an opportunity to examine long-term associations between internal migration and industry agglomeration patterns, and understand the effects of regional economic performance and policy changes on population movements across the Chilean urban hierarchy. Empirical evidence suggests that as urban systems mature, they develop a hierarchical network of primary, intermediate, and small cities (Kontuly, Geyer 2005). Cities appear to have progressed through cyclical concentration and deconcentration phases, starting with urbanization, followed by polarization reversal, counter-urbanization and re-urbanization. Differences in political and socio-economic conditions are noted to affect the timing of these processes, and each of these processes is characterized by particular patterns of migration, reflecting the spatial diffusion of urban development and economic growth. Urbanization is characterized by net in-migration into major urban areas, echoing the spatial concentration of development and economic growth. Polarization reversal is associated with deconcentration of firms and capital to peripheral areas of smaller population size due to diseconomies of scales in major urban centers. Counter-urbanization comprises net out-migration from urban centers, reflecting the inner city economic decline and mobility to rural locations; and re-urbanization involves gentrification and re-population of city centers through net in-migration. Organizing the TM into a functional urban framework, the CHIM database can be deployed to explore how these processes have played out in Chile and how they have been shaped by contextual changes.

## References

- Bianchini R, Feeney G, Singh R (2013) Report of the international commission on the 2012 population and housing census of Chile. Instituto nacional de estadísticas, Santiago, Chile
- Bravo D, Larrañaga O, Millán I, Ruiz M, Zamorano F (2013) Informe final comisión externa revisora del censo 2012. Instituto nacional de estadísticas, Santiago, Chile
- Casado-Díaz JM, Martínez-Bernabéu L, Rowe F (2017) An evolutionary approach to the delimitation of labour market areas: an empirical application for Chile. *Spatial Economic Analysis*: 1–25. [CrossRef](#).
- Dresdner J, Sanhueza C (2009) Estimación de series de salarios regionales. *Estudios de Economía* 36[2]: 243–261
- Kontuly T, Geyer H (2005) Lessons learned from testing the differential urbanisation model. *Tijdschrift voor economische en sociale geografie* 94: 124–128. [CrossRef](#).
- Prado M, Gatica R, Letelier J (2002) Las exportaciones regionales de Chile en la perspectiva de la globalización: Período 1990-2000. Technical report MIDEPLAN, Santiago, Chile
- Rowe F (2013) Spatial labour mobility in a transition economy: Migration and commuting in Chile. PhD thesis, school of geography, planning and environmental management, The University of Queensland, Brisbane, Australia. [CrossRef](#).
- Rowe F (2014) Micro and macro drivers of long-distance commuting in Chile: The role of spatial distribution of economic activities and population. paper presented at the 53rd congress of the Western Regional Science Association, San Diego, CA
- Rowe F, Bell M (2013) Creating an integrated database for the analysis of spatial mobility in Chile. Working paper 02/2013, Queensland centre for population research, school of geography, planning and environmental management, The University of Queensland, Brisbane, Australia
- Rowe F, Bell M (forthcoming) The drivers of long-distance commuting in Chile: The role of spatial distribution of economic activities and population. In: Poot J, Roskrige M (eds), *Regional science perspectives on population change and impacts in Asia and the Pacific*. Springer
- UNESCO – The United Nations Educational Scientific and Cultural Organization (1997) International standard classification of education. ISCED 1997, 22nd May, 2012. [http://www.unesco.org/education/information/nfsunesco/doc/isced\\_1997.htm](http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm)