# CASA dissertation: Literature Review

Student Number: 22186878

Word count (excluding tables and figures): aim for 1200-1500 words.

**Guidelines from CASA Dissertation Handbook**

*The literature review should evaluate existing research, demonstrate contrasting and/or similar views whilst highlighting research gaps. Synthesise previous work / policy documents and provide a narrative through it whilst trying to show where the research gap is / where your question fits in. Don’t just list what authors have done in the past (e.g., Smith 2009 did x but Jones 2008 did y then Frank 2010 did x). Try to end the literature review with a concluding paragraph that concisely summarises everything within it and states what your work is going to contribute or address.*

*Think of this section as providing a story about what everyone else has done (whilst also showing issues / research gaps) and then what you are going to do.*

## Table of Contents

### Introduction

1. World Bank Water Project: Sri Lanka
2. Justification of extending project to India
3. Research Questions

### Literature Review

1. Agricultural Dependent Populations
   1. How is this defined in the literature?
   2. Why is this term useful? What is the value of understanding this population?
2. Indian context
   1. States/districts/villages formation
   2. National-level agricultural dependence, statistics
   3. Available census data; opportunities and limitations
3. Spatial disaggregation
   1. What does this term mean?
   2. World gridded population
   3. Examples of spatial disaggregation methods; their strengths and weaknesses.
   4. Binary dasymetric mapping: the main method to be used
   5. Considerations: MAUP, etc.
4. Scaling Geocomputation \*\*TBD if this section will be relevant\*\*
   1. Challenges/Limitations of computation at large spatial scales
   2. Methods for overcoming this

## Introduction

Understanding where people live, and the social and economic characteristics of those populations, is core to providing adequate, efficient, and targeted services and investment.

Justification

This study is a novel addition to the field as it extends upon existing methodologies used to estimate total population and applies this to the estimation of the agricultural dependent population. Additionally, the case study of India is designed to assess feasibility and performance at a large spatial scale, comparative to partner research testing proof-of-concept in districts of Sri Lanka (unpublished). Understanding particularly the distribution of agricultural population in a region will provide a more accurate estimate of local demand on water resources.

Stretch out ‘scale’ component further; key aspect of study

Text

## Literature Review

### Preamble

Introduction for the literature review; summarise the findings and topics that will be explored – ADP, India, and Spatial Disaggregation Methodology.

Text

### Agricultural Dependent Populations

Topic: How is ADP addressed in the literature?

Agriculture represents the single largest employer across the globe, as the source of income for 40 per cent of the world’s population (Kondylis *et al.*, 2023). In India, this share is even larger, with XX% estimated to be dependent on agriculture for a living, predominantly in small farms (REF). Agricultural populations in India typically face high rates of poverty and instability, and are identified by the World Bank as a key target for development funding (REF), especially in the context of increasing vulnerability due to the effects of climate change and increased variability of temperature and rainfall (Anand, Kakumanu and Amarasinghe, 2019).

To support effective, context-specific development, it is necessary to understand the spatial distribution of this agricultural population. Specifically in the case of India, tanks …

Topic: LULC for estimating agricultural versus urban/rural land

*Use this paragraph to reflect on comments in from [which?] reading about balance between higher population density and masking out urban areas to ensure that estimates of ADP are not weighted towards the urban – need to fin reading.*

### Indian Context

Paragraph topic: Intro; What is the current situation in India for ADP?

India, the subject of this study, is one of the world’s largest countries by area, the third-largest economy, and is expected to become the most populous country by the end of 2023 (United Nations in India, 2022).

Paragraph topic: Administrative formation.

India is divided into 28 states and 8 union territories, each of which are further subdivided into districts and smaller administrative divisions, variously termed tehsil, taluks, or mandal (Government of India, 2012). Census data for most socioeconomic indicators are published at the district level. In census enumeration, urban areas are divided into four classes – wards, outgrowths, statutory towns, and census towns (Balk *et al.*, 2019).

Paragraph topic: census data – what is available? Opportunities/Limitations?

The Indian census of population is conducted decennially, the last being conducted in 2011 due to COVID delaying a planned census in 2021. There have been considerable changes in population count and distribution in the time since this census, which must be considered when interpreting any analysis.

### Spatial Disaggregation

Paragraph topic: Define spatial disaggregation; Why is it important?

Spatial disaggregation is a broad term which applies to the process of transforming data from a set of source zones into target zones, such as a raster grid, at a finer level of spatial resolution. There is considerable interest in the process across both academic literature and in policy, particularly applied to estimating population at fine spatial scales, as this has important implications for service planning and delivery (Deichmann, 1996), disaster preparation and response (Schneiderbauer and Ehrlich, 2005), monitoring international development goals (Tuholske *et al.*, 2021; United Nations, 2022) and the implementation of public health interventions (Viel and Tran, 2009; Tatem, 2022), among others.

Topic: World gridded population estimates; history, status

On a global scale, the spatial disaggregation of administrative census data has been used to develop world gridded population estimates, providing regularly sized comparable population estimates across regions. Early iterations of this, such as the Gridded Population of the World (GPW) version 1 (Tobler *et al.*, 1997), have inspired a variety of contemporary global grid models, each utilising a specialised methodology and with particular strengths and limitations, bolstered by advances in computational power and the availability of high quality census and earth observation data (Wardrop *et al.*, 2018). Table 1 provides a summary of current global grid models and their key characteristics.

Table : Selected World Population Grid Datasets, adapted from Leyk et al. (2019)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset** | **Source** | **Method** | **Spatial Resolution** | **Ancillary data layers** |
| Gridded Population of the World (GPW) | CIESINa | Areal weighting | 1km | Water bodies |
| Global Human Settlement Layer – Population (GHS-POP) | JRCb and CIESINa | Dasymetric | 250m | Built structures |
| WorldPop | University of Southampton | Statistical/ Dasymetric | 100m | Roads, Land cover, Built structures, Urban areas, Night-time lights, Infrastructure, Climate, Topography, Elevation, Water bodies |
| LandScan Global | ORNLc | Smart interpolation | 30 arcsec | Roads, Land cover, Built structures, Urban areas, Infrastructure, Climate, Topography, Elevation, Water bodies |
| World Population Estimate (WPE) | Esri | Dasymetric redistribution | 150m | Roads, Land cover, Urban areas, Water bodies |

a Centre for International Earth Science Information Network; b Joint Research Centre of the European Commission; c Oak Ridge National Laboratory.

Topic: Spatial disaggregation continued – dasymetric, pycnophylactic

The most straightforward method of spatial disaggregation is areal weighting, where data from the source zone (such as the total population of a district) is evenly distributed across the gridded cells within it. Areal weighting benefits from low computational power and no requirement for ancillary data. However, this approach assumes that populations are evenly distributed across administrative regions, which is rarely the case (Qiu *et al.*, 2022), necessitating the development of more complex models which can incorporate knowledge from additional sources. A popular approach is dasymetric mapping, which divides the area into homogenous zones based on the variable of interest (Eicher and Brewer, 2001). For example, remote sensing data can be used to identify water bodies and other non-inhabitable areas within a zone, and limit these cells to a value of zero, creating a ‘mask’. Population count can then be proportionally distributed across the non-zero cells, to produce a more accurate estimate of the real spatial distribution of population. This method is referred to as binary dasymetric mapping or binary masking (Qiu *et al.*, 2022). Dasymetry is not limited to binary classification; multi-class dasymetric mapping

Topic: Spatial disaggregation methods: Top down vs Bottom up

Alternative to these ‘top-down’ methods of spatial disaggregation, ‘bottom-up’ approaches for small area estimation can be used to produce gridded population estimates. These methods are designed for estimating population distribution in the absence of high-level source data such as a national or regional census, or when such data is out-of-date or known to be inaccurate (Wardrop *et al.*, 2018). However, bottom-up estimation requires the completion of tailored geo-located field surveys, and are generally viewed as complementary to traditional census enumeration in specific data-poor settings (Leyk *et al.*, 2019).

Topic: Spatial disaggregation past population: additional socioeconomic characteristics

Although there has been extensive research and methodological development in the field of population disaggregation, there are fewer studies that extend these methods to estimate additional demographic or socioeconomic characteristics beyond population count or density, despite the methodologies being broadly similar. An early study by Eicher and Brewer (2001) showed the potential for dasymetric mapping to map age structure and housing value in the United States, and more recently novel data sources have been utilised, such as Point of Interest property data in Singapore (Szarka and Biljecki, 2022), to estimate elderly populations at the neighbourhood scale. The WorldPop research unit regularly produce national and regional gridded maps across health and social indicators, such as vaccination coverage (REF), and in the Indian context produced an ‘atlas’ of 19 indicators nationwide at a 5km resolution (Pezzulo *et al.*, 2023). In assessing agricultural populations, only one relevant study was identified which estimated the proportion of primary sector labourers at the parish level in Portugal using a hybrid method of dasymetric mapping and pycnophylactic interpolation (Monteiro, Martins and Pires, 2018) adapted from work by Malone *et al.* (2012).

[How should I argue against using the method in Monteiro 2018, and instead using the simplified population clip proposed by FL/SA?]

Topic: Considerations – MAUP; resolution used; temporal differences between census and ancillary data

Like in most aspects of spatial analysis, the spatial resolution or scale used in a spatial disaggregation can markedly influence the outcomes and interpretation. This challenge is defined as the modifiable areal unit problem (MAUP), where the same analysis performed on data aggregated at different levels or in different zones will produce different results (Wong, 2009).

## References

Anand, S., Kakumanu, K.R. and Amarasinghe, U.A. (2019) ‘Use of Remote Sensing and GIS for Identifying Tanks and Rehabilitation Benefits to the Rural Areas’, *Journal of Rural Development*, 38(1), p. 55. Available at: https://doi.org/10.25175/jrd/2019/v38/i1/121801.

Balk, D. *et al.* (2019) ‘Urbanization in India: Population and Urban Classification Grids for 2011’, *Data*, 4(1), p. 35. Available at: https://doi.org/10.3390/data4010035.

Deichmann, U. (1996) *A Review of Spatial Population Database Design and Modeling*. Santa Barbara, CA: National Centre for Geographic Information and Analysis. Available at: https://escholarship.org/uc/item/6g190671 (Accessed: 28 February 2023).

Eicher, C.L. and Brewer, C.A. (2001) ‘Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation’, *Cartography and Geographic Information Science*, 28(2), pp. 125–138. Available at: https://doi.org/10.1559/152304001782173727.

Government of India (2012) *Census of India 2011: Administrative Atlas*. Delhi, India: Office of the Registrar General and Census Commissioner. Available at: https://censusindia.gov.in/census.website/data/atlas# (Accessed: 1 June 2023).

Kondylis, F. *et al.* (2023) *Agriculture*, *World Bank: Development Impact Evaluation (DIME)*. Available at: https://www.worldbank.org/en/research/dime/brief/agriculture (Accessed: 1 June 2023).

Leyk, S. *et al.* (2019) ‘The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use’, *Earth System Science Data*, 11(3), pp. 1385–1409. Available at: https://doi.org/10.5194/essd-11-1385-2019.

Malone, B.P. *et al.* (2012) ‘A general method for downscaling earth resource information’, *Computers & Geosciences*, 41, pp. 119–125. Available at: https://doi.org/10.1016/j.cageo.2011.08.021.

Monteiro, J., Martins, B. and Pires, J.M. (2018) ‘A hybrid approach for the spatial disaggregation of socio-economic indicators’, *International Journal of Data Science and Analytics*, 5(2), pp. 189–211. Available at: https://doi.org/10.1007/s41060-017-0080-z.

Pezzulo, C. *et al.* (2023) ‘A subnational reproductive, maternal, newborn, child, and adolescent health and development atlas of India’, *Scientific Data*, 10(1), p. 86. Available at: https://doi.org/10.1038/s41597-023-01961-2.

Qiu, Y. *et al.* (2022) ‘Disaggregating population data for assessing progress of SDGs: methods and applications’, *International Journal of Digital Earth*, 15(1), pp. 2–29. Available at: https://doi.org/10.1080/17538947.2021.2013553.

Schneiderbauer, S. and Ehrlich, D. (2005) ‘Population Density Estimations for Disaster Management: Case Study Rural Zimbabwe’, in P. van Oosterom, S. Zlatanova, and E.M. Fendel (eds) *Geo-information for Disaster Management*. Berlin, Heidelberg: Springer, pp. 901–921. Available at: https://doi.org/10.1007/3-540-27468-5\_64.

Szarka, N. and Biljecki, F. (2022) ‘Population estimation beyond counts—Inferring demographic characteristics’, *PLOS ONE*, 17(4), p. e0266484. Available at: https://doi.org/10.1371/journal.pone.0266484.

Tatem, A.J. (2022) ‘Small area population denominators for improved disease surveillance and response’, *Epidemics*, 40, p. 100597. Available at: https://doi.org/10.1016/j.epidem.2022.100597.

Tobler, W. *et al.* (1997) ‘World population in a grid of spherical quadrilaterals’, *International Journal of Population Geography*, 3(3), pp. 203–225. Available at: https://doi.org/10.1002/(SICI)1099-1220(199709)3:3<203::AID-IJPG68>3.0.CO;2-C.

Tuholske, C. *et al.* (2021) ‘Implications for Tracking SDG Indicator Metrics with Gridded Population Data’, *Sustainability*, 13(13), p. 7329. Available at: https://doi.org/10.3390/su13137329.

United Nations (2022) *The Sustainable Development Goals Report 2022*. New York, NY: United Nations. Available at: https://unstats.un.org/sdgs/report/2022/.

United Nations in India (2022) *UN India Annual Report 2021*. New Delhi, India. Available at: https://india.un.org/en/195240-un-india-annual-report-2021 (Accessed: 30 May 2023).

Viel, J.-F. and Tran, A. (2009) ‘Estimating Denominators: Satellite-Based Population Estimates at a Fine Spatial Resolution in a European Urban Area’, *Epidemiology*, 20(2), pp. 214–222.

Wardrop, N.A. *et al.* (2018) ‘Spatially disaggregated population estimates in the absence of national population and housing census data’, *Proceedings of the National Academy of Sciences*, 115(14), pp. 3529–3537. Available at: https://doi.org/10.1073/pnas.1715305115.

Wong, D.W. (2009) ‘Modifiable Areal Unit Problem’, in R. Kitchin and N. Thrift (eds) *International Encyclopedia of Human Geography*. Oxford: Elsevier, pp. 169–174. Available at: https://doi.org/10.1016/B978-008044910-4.00475-2.