Reading Notes Summary

UCL Centre for Advanced Spatial Analysis

MSc Urban Spatial Science, Dissertation

Contents

[Reference Title (Template) 3](#_Toc137664378)

[Sri Lanka Tanks Project 4](#_Toc137664379)

[Anand 2019 – Rehabilitation benefits of tanks in rural India 4](#_Toc137664380)

[Mialhe 2008 – Remote sensing assessment of water reservoirs in India 4](#_Toc137664381)

[Rodrigues 2011 – Estimation of small reservoir storage capacities 5](#_Toc137664382)

[Thenkabail 2009 – Irrigated area maps of India using remote sensing 6](#_Toc137664383)

[Vidanage 2022 – Sri Lanka’s small tank cascade systems 7](#_Toc137664384)

[Population Grids 8](#_Toc137664385)

[Balk 2019 – Population and Urban Classification Grids for India 2011 8](#_Toc137664386)

[Corbane 2017 – Global human settlements mapping 9](#_Toc137664387)

[Spatial Disaggregation Methods 10](#_Toc137664388)

[Comber 2019 – A review of spatial interpolation using areal features 10](#_Toc137664389)

[Eicher 2001 – Dasymetric mapping and areal interpolation 11](#_Toc137664390)

[Holt 2011 – Dasymetric mapping for sociodemographic data redistribution 11](#_Toc137664391)

[Malone 2012 – A general method for downscaling earth resource information 12](#_Toc137664392)

[Monteiro 2018 – Hybrid approach for spatial disaggregation of socio-economic indicators 13](#_Toc137664393)

[Mubareka 2008 – Settlement location and population density estimation 14](#_Toc137664394)

[Qiu 2022 – Disaggregating population data for assessing SDG progress 15](#_Toc137664395)

[Szarka 2022 – Inferring demographic characteristics in population estimation 17](#_Toc137664396)

[Wang 2022 – Global gridded GDP data set 17](#_Toc137664397)

[Wardrop 2018 – Spatially disaggregated population estimates in the absence of census data 18](#_Toc137664398)

[You 2006 – An entropy approach to spatial disaggregation of crop production 19](#_Toc137664399)

[Agriculture dependent populations 20](#_Toc137664400)

[World Bank 2023 – Agriculture and Food 20](#_Toc137664401)

## Reference Title (Template)

|  |  |
| --- | --- |
| Full reference |  |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* |  |
| What is the article about? |  |
| What is the main research question/argument? |  |
| How are the aims explored/tested/presented? |  |
| What are the main findings? |  |
| What gaps does it identify? |  |
| Limitations/critiques |  |
| How does it link to wider learning? |  |
| Additional notes  *(incl. useful quotes)* |  |

# Sri Lanka Tanks Project

## Anand 2019 – Rehabilitation benefits of tanks in rural India

|  |  |
| --- | --- |
| Full reference |  |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Empirical |
| What is the article about? | Assesses the potential of small tanks for strengthening water security in rural areas of India, in a monsoon-driven climate |
| What are the main findings? | * Importance of water security becomes increasingly important with increasingly variable rainfall under climate change * Monsoon rains between June and September contribute to 80 per cent of the total precipitation in many Indian river basins * “Tanks in the Indian context are also inextricably linked to the socio-cultural aspects of rural life and have historically been an indispensable part of the village habitat, sustaining its socio-ecological balance” (p.56) * In the study area (Bhadrachalam) approx. 70% of the population are rural and depend on agriculture-based activities for their livelihoods. |
| How does it link to wider learning? | Key paper for the FL/SA Sri Lanka project; tangentially relevant to main dissertation topic |

## Mialhe 2008 – Remote sensing assessment of water reservoirs in India

|  |  |
| --- | --- |
| Full reference |  |
| Article type | Empirical/methodological |
| What is the article about? | Presents a method to map and quantify water stocks in small irrigation reservoirs (tanks) over large regions, with a case study of India |
| What are the main findings? | * In South India, tanks are traditional water storage reservoirs designed to harvest and store rainfall and surface runoff * Tanks mitigate the impact of seasonal and interannual rainfall variability on water by extending its availability for agriculture through storage * Tank irrigation, and tanks in working condition, are profitable to small farmers and improve livelihoods for those in agricultural regions * Tank systems are widespread in the states of Andhra Pradesh, Tamil Nadu and Karnataka. |
| How does it link to wider learning? |  |

## Rodrigues 2011 – Estimation of small reservoir storage capacities

|  |  |
| --- | --- |
| Full reference |  |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Empirical |
| What is the article about? | Development of a simple method to estimate reservoir storage volumes based on remotely sense reservoir surface area measured with LANDSAT. |
| How are the aims explored/tested/presented? | Method validated on a subset of reservoirs in the Preto River Basin in Brazil, for which a set of ground-based survey measurements had been taken. |
| What are the main findings? | General relationship between measured reservoir volumes and their remotely sensed surface areas found R2 = 83%. |
| What gaps does it identify? |  |
| Limitations/critiques |  |
| How does it link to wider learning? | Reservoir/water security analysis; used in the Sri Lanka paper by Fl/SA. |
| Additional notes  *(incl. useful quotes)* |  |

## Thenkabail 2009 – Irrigated area maps of India using remote sensing

|  |  |
| --- | --- |
| Full reference | Thenkabail, P.S. *et al.* (2009) ‘Irrigated Area Maps and Statistics of India Using Remote Sensing and National Statistics’, *Remote Sensing*, 1(2), pp. 50–67. Available at: <https://doi.org/10.3390/rs1020050>. |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Empirical |
| What is the article about? | Comparison of remote-sensing derived maps of irrigated land with census-derived statistics of irrigated land. |
| How are the aims explored/tested/presented? |  |
| What are the main findings? | * Administrative data on irrigated area are produced separately by 2 departments: Ministry of Agriculture, and Ministry of Water Resources. There are large variations even between the statistics reported by these 2 bodies. * Two recent efforts in mapping global irrigation: GIAM by the International Water Management Institute (IWMI), and the global map of irrigated areas, GMIA, by the UN’s Food and Agriculture Organisation (FAO). GIAM is overwhelmingly remote sensing based, whereas GMIA is overwhelmingly based on national statistics. * In India, traditionally local revenue department officials report irrigation statistics that are in turn compiled at different levels. * Government administrative data is estimated to unreliably account for *minor* irrigation sources, such as ground water, small reservoirs, and tanks. |
| What gaps does it identify? | A need for better harmonisation between different irrigation estimates, and understanding of why the differences in estimates are so large. |
| Limitations/critiques | Uses existing remote-sensing research (the GIAM and GMIA), rather than own methodology; therefore less applicable to developing method for dissertation. |
| How does it link to wider learning? |  |
| Additional notes  *(incl. useful quotes)* |  |

## Vidanage 2022 – Sri Lanka’s small tank cascade systems

|  |  |
| --- | --- |
| Full reference | Vidanage, S.P., Kotagama, H.B. and Dunusinghe, P.M. (2022) ‘Sri Lanka’s Small Tank Cascade Systems: Building Agricultural Resilience in the Dry Zone’, in A.K.E. Haque et al. (eds) *Climate Change and Community Resilience: Insights from South Asia*. Singapore: Springer Nature, pp. 225–235. Available at: <https://doi.org/10.1007/978-981-16-0680-9_15>. |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Empirical |
| What is the article about? | The importance of small tank cascade systems in providing water security and broader ecosystem services to villages in the Dry Zone of Sri Lanka. |
| How are the aims explored/tested/presented? |  |
| What are the main findings? | * Small tank cascade systems harvest rainwater and mitigate floods and drought * These are traditional systems of building farmer resilience to climate variability. * SL government has identified restoration and sustainable management of STCS as a priority adaptation action. Also as a key climate change adaptation mechanism in SL’s National Adaptation Plan. * Benefits of functioning small tank systems include provisioning water for fisheries and livestock, control of soil erosion, flood prevention, water quality control, storage of water for irrigation, reducing vulnerability to drought, and soil health. “For small tank associated agrarian societies, the tank is the most important asset as it provides numerous services in addition to supplying water for irrigation” |
| What gaps does it identify? |  |
| Limitations/critiques |  |
| How does it link to wider learning? | Referenced in FL/SA Sri Lanka paper. Includes useful references to aligned policy goals. |
| Additional notes  *(incl. useful quotes)* |  |

# Population Grids

## Balk 2019 – Population and Urban Classification Grids for India 2011

|  |  |
| --- | --- |
| Full reference | 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>. |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Empirical |
| What is the article about? | Development of a 1km scale population grid for India, to compare with official census urban designations (which underestimate urban population) |
| How are the aims explored/tested/presented? | * 3 sets of input data to generate a gridded population estimate: settlement level demographic data from 2011 census; spatial boundaries from proprietary (private) data product; global human settlement layer (remote sensing derived) * Method followed standard practice of removing waterbodies before creating grids of population distribution and urban areas 🡪 would this be necessary for diss. analysis? |
| What are the main findings? | * 2011 census predicts 31% India’s population live in urban areas; this is far lower than other estimates would suggest * Four urban categories in Indian census: statutory towns, census towns, wards, and outgrowths. Wards and outgrowths are subsets of statutory towns; census towns are defined anew with each census iteration (thus challenges in comparison over time) |
| Limitations/critiques | Boundary data for towns purchased from 3rd-party vendors; limits the reproducibility of this analysis > implications for relevance to dissertation? |
| How does it link to wider learning? | Example of incorporating remote sensing data with census settlement-level data to produce a gridded population estimate; dissertation project could theoretically follow very similar trajectory. |
| Additional notes  *(incl. useful quotes)* | * “Census towns are legally rural settlements but they are designated as urban for the purposes of an upcoming census and grouped with statutory urban areas in the official post-census tabulations” (p.3) * GHSL data available as binary (30m grid cell indicating built or not) or an aggregate at 250m resolution indicating proportion of cell classed as built. |

## Corbane 2017 – Global human settlements mapping

|  |  |
| --- | --- |
| Full reference |  |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Empirical |
| What is the article about? | Methodology behind the development of the Global Human Settlement Layer (GHSL) |
| How are the aims explored/tested/presented? | Presents the processing workflows and results of the two main experiments behind the GHSL |
| What are the main findings? | * Understanding human settlements and built-up areas is crucial to evaluating progress against major international frameworks, notably the UN SDGs * Remote sensing data can be a useful source of information for estimating settlements in a consistent and systematic way. Integration with statistical data contributes to an enhanced overall understanding. * First GHSL released publicly in 2016; the GHSL includes global multi-temporal evolution of built-up surfaces, population densities, and urban-rural delimitation organised in four epochs (1975, 1990, 2000 & 2014) * Earth observation requirements for global settlement mapping: data with worldwide coverage, multi-source, multi-temporal and multi-scale, high-dimensional, highly complex, and unstructured. |
| What gaps does it identify? |  |
| Limitations/critiques | Highly technical paper on calculations and evaluation |
| How does it link to wider learning? | Provides some high-level understanding of the rationale and background behind the GHSL; how does this link to WorldPop, and DynamicWorld? |
| Additional notes  *(incl. useful quotes)* |  |

# Spatial Disaggregation Methods

## Comber 2019 – A review of spatial interpolation using areal features

|  |  |
| --- | --- |
| Full reference |  |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Review (including empirical testing/comparison) |
| What is the article about? | High-level review of different methods for spatial interpolation using areal\* features |
| How are the aims explored/tested/presented? | Tests different methods discussed on a single case study: estimating number of houses in a 500m grid within a US Census Tract in Connecticut. Copies of the R code used to conduct the analysis are included in the supplementary data. |
| What are the main findings? | * Distinguishes point from areal interpolation: point is used for making predictions at locations where values are unknown using other sample points, e.g. temperature surface from point temperature recordings. Areal is used for transferring attribute information from source zones to target zones – aka spatial disaggregation. * Contrasts between interpolation with ancillary information, and interpolation without * Without methods: areal weighting, pycnophylactic, area-to-point * With methods: Dasymetric mapping, street-weighting method, statistical approaches, point-based approaches * Statistical approaches try to establish a functional relationship between the distribution of ancillary data and distribution of the variable of interest. Usually involves a regression analysis at the source zone level. * “Undoubtedly, the robustness of the interpolation depends on the choice of the ancillary data, as well as the interpolation methodological approach” (p.17) * Statistical and Area-to-Point approaches found to vary the most (in terms of output) comparative to others. |
| What gaps does it identify? | * Challenges with dasymetric methods: no one dasymetric method has established as superior (differing performance in different cases); increases in computational cost can limit applicability to large areas; subject to the MAUP |
| How does it link to wider learning? | Provides good literature examples of applications of different interpolation methods, not just related to population. |
| Additional notes  *(incl. useful quotes)* | \*Here *areal* refers to polygons or rasters of space, as opposed to points  - See Table 1 in paper for review of studies applying different interpolation methods |

## Eicher 2001 – Dasymetric mapping and areal interpolation

|  |  |
| --- | --- |
| Full reference |  |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Empirical |
| What is the article about? | Authors tested 5 dasymetric mapping methods for 6 different socioeconomic variables (age structure and house value), and evaluated the error |
| What are the main findings? | * “A dasymetric map depicts quantitative areal data using boundaries that divide the mapped area into zone of relative homogeneity” (p.125) * Dasymetric maps have boundaries that represent sharp changes in the data value being mapped – for example, the likely population density. Choropleth maps have boundaries based on enumeration units or administrative zones. * A necessary condition of dasymetric mapping is the use of an ancillary dataset * The binary method is a special case of dasymetric mapping in raster grids, where cells that have a land use class deemed ‘uninhabitable’ are limited to a value of 0, and then a regular proportional distribution performed on the remaining cells. |
| Limitations/critiques | Published very early in GIS field (2001); therefore discussion of capabilities is in many ways out of date |
| How does it link to wider learning? | Clear definitions of dasymetric mapping, and particularly binary masking/the binary method. |
| Additional notes  *(incl. useful quotes)* |  |

## Holt 2011 – Dasymetric mapping for sociodemographic data redistribution

|  |  |
| --- | --- |
| Full reference |  |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Conceptual/explanatory (book chapter) |
| What is the article about? | Provide an overview of areal interpolation using dasymetric mapping, and illustrate an example in which sociodemographic data (population of individuals in poverty) are derived for different spatial units in Atlanta, Georgia |
| What are the main findings? | * Mapped the population in poverty with LandPro 2001 (land use dataset) as the ancillary data * Applied the zonal statistic tool in ArcGIS to perform an areal interpolation (simple distribution across non-masked pixels) * Little evidence to indicate that multi-class dasymetry performs substantially better than binary |
| Limitations/critiques | Note that LULC data may not recognise institutional populations in universities or prisons, which can be classed as non-residential |
| How does it link to wider learning? | Example of an alternative sociodemographic variable being disaggregated |

## Malone 2012 – A general method for downscaling earth resource information

|  |  |
| --- | --- |
| Full reference |  |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Empirical |
| What is the article about? | Describes an R package ‘*dissever*’ that facilitates downscaling of coarsely resolved EO information using available finely gridded covariate data. |
| How are the aims explored/tested/presented? | * 2-stage algorithm: initialisation and iteration * Model uses a generalised additive model (GAM) to avoid relying on the assumption of a linear relationship * For evaluation, the weighted root mean square error (wRMSE) provides a quantitative measure of the mass balance deviation between the coarse and downscaled gridded information |
| What are the main findings? | * An implicit assumption of disaggregation methods using covariate data is that the covariates are strongly related to the target variable |
| What gaps does it identify? |  |
| Limitations/critiques | This paper discusses the downscaling of earth observation data – already collected in a grid/raster format, but at a coarser resolution. Contrasts with the requirements of administrative data disaggregation, which converts from vector boundaries to a finer raster grid. |
| How does it link to wider learning? |  |
| Additional notes  *(incl. useful quotes)* | * Describes ‘pycnophylactic’ or ‘mass balance’ methods as those that “attempt to harmonise the arithmetic average of the property values at the fine scale with the single property value at the coarse scale” (p.120) * The *dissever* package is structured as a function, which requires two information inputs – (1) a data table containing the target variable information and covariate data source information, and (2) the GAM formula |

## Monteiro 2018 – Hybrid approach for spatial disaggregation of socio-economic indicators

|  |  |
| --- | --- |
| Full reference |  |
| Article type | Empirical 🡪 methodological |
| What is the article about? | Develops and evaluates a novel methodology for the spatial disaggregation of socioeconomic statistics, extending traditional population-based disaggregation methods. The hybrid approach combines dasymetric mapping and pycnophylactic interpolation, along with ancillary data, in a ML model. |
| How are the aims explored/tested/ presented? | A novel disaggregation method is implemented, based on a downscaling procedure that uses regression to combine different ancillary variables. Different regression methods and sampling methods are tested and compared for performance. Analysis performed in R; packages *pycno* and *dissever*.  See page 195 for a step-by-step breakdown of the proposed methodology. |
| What are the main findings? | * “Spatial disaggregation techniques can be used … to transform data from a set of source zones into a set of target zones, with different geometry and with a higher general level of spatial resolution” (p.189) * Disaggregation methods range widely in complexity; from simple areal weighting (equal proportional spread of population across area) being the most basic. * This study develops a method introduced by Malone (2012) which uses 2 phases: (1) initialisation, performing a resampling through simple nearest neighbour where cells in the finer grid take value from the closest coarse grid cell, followed by random sampling and initial model fit, then (2) iteration, adjustments are made iteratively to the predictions to achieve mass balance (where the finer grid values aggregate to the coarse grid values). Iterations proceed until a stopping criteria is met. * Also discuss estimating primary sector workers (i.e. agricultural labourers); include this in LR |
| Additional notes  *(incl. useful quotes)* | * *Pycnophylactic:* mass-preserving property, in that estimates sum to the original quantities in the source zones. Term from Tobler (1970) * *Mass-preserving areal weighting*: homogenous distribution of data throughout the source area. Most basic form of spatial disaggregation. * *Dasymetric disaggregation*: any method that leverages ancillary information (such as a landcover derived mask). * *Mask areal weighting/Binary dasymetry*: uses a mask to define where, within the target zone, source data should be allocated. For example, identifies built-up areas for allocating population. LULC satellite data is often used to create the ‘mask’ layer. This appears to align closest with FL/SA’s method for Sri Lanka project.   + Weaknesses: binary dasymetry assumes all populated areas have homogenous density; does not reflect reality especially across larger target zones * *Poly-categorical dasymetry*: adapts mask weighting so that multiple categories can be assigned weights – percentages are applied to each of the categories for the source area, representing the % of data that is likely to be contained within that category per source area. Main challenge in this sense is devising an appropriate set of weights. * See pages 193-194 for a short bibliography of relevant spatial disaggregation studies |

## Mubareka 2008 – Settlement location and population density estimation

|  |  |
| --- | --- |
| Full reference | Mubareka 2008 – Settlement location and population density estimation using remote sensing |
| Related module | Dissertation (project proposals) |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Empirical |
| What is the article about? | Develops and tests a method for estimating settlement location probability, and population density, at 90m resolution using remote sensing data. Study area is a region in northern Iraq. |
| How are the aims explored/tested/presented? | Trains a model using 50% of a fieldwork collected dataset on settlements in Northern Iraq, combined with remote sensing data of the region. The remaining 50% is used to validate the model predictions. |
| What are the main findings? | * Can build the probability and pop density model using solely data from Shuttle Radar Topographic Mission (SRTM) digital terrain model (for elevation, etc.), and Landsat Enhanced Thematic Mapper (ETM). * Population density layer uses land cover and topographic features for input; Settlement probability layer uses distance from roads and water bodies, and land cover for input. * Model is validated using field collected data set from the Rapid Assessment Program (RAP), from humanitarian agencies. * R2 of approx. 0.3; is this sufficient for practice/application? (may need to look deeper into the interpretation of R2 in this context) |
| Limitations/critiques | Methodology is very complex, and requires extensive data cleaning calculations before incorporation in model; would be challenging to replicate in a different context. |
| Additional notes  *(incl. useful quotes)* | * “Surface modelling refers to allocating population or their attributes … to a fine scale in a regularly spaced grid” (p.2340) * Pop density estimates are more predictable in rural areas, compared with urban environments * Pioneers of remote sensing population estimation:   1. Deichmann & Eklundh 1991   2. Dichmann 1996   3. Sweitzer & Langaas 1995   4. Dobson et al. 2000   5. Tian et al. 2005 |

## Qiu 2022 – Disaggregating population data for assessing SDG progress

|  |  |
| --- | --- |
| Full reference |  |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Review |
| What is the article about? | Summarises the research advances of population disaggregation in terms of methodology, ancillary data, and output products. |
| What are the main findings? | * Population disaggregation into grids can be divided into 2 categories: bottom-up and top-down. Bottom-up is applicable for situations where a country lacks census data, or where a region has high-quality geocoded data. Top-down is the more common and simpler method, and the focus of Sri Lanka project + dissertation * Methods in spatial disaggregation: see next page for classification tree * Common sources of ancillary data for population estimation: land cover, night time lights, infrastructures, and environmental factors. * Some classes of land cover (water, glaciers, protected habitats) are not viable for human habitation; they can be used as masks to reduce errors in the disaggregation process. Q: Should I create a water bodies mask for India? Or is this implicitly already covered by only selecting crop for the binary masking. A: Not necessary; both the crop land boundaries and the world pop estimates already mask out water bodies. * See table 1 (p.10) for a summary of global disaggregated population products (such as GHSL, WorldPop, LandScan, etc.) * Challenges: selecting an optimal scale for the disaggregated population data. Need to rationalise and justify this – align with Sri Lanka project?   1. “The range of grid size should be determined according to research objectives, study area, ancillary data, method, accuracy requirement, etc.” (p.17) |
| What gaps does it identify? |  |
| Limitations/critiques |  |
| How does it link to wider learning? | Good overview of spatial disaggregation methods specific to population. Outlines binary dasymetry, which is the process to be used. |
| Additional notes  *(incl. useful quotes)* | * Spatial disaggregation vs spatial downscaling: both are a process of converting information on a coarse spatial scale to a fine scale. Disaggregation applies to additive variables (such as population), where the values can aggregated in target areas, whereas downscaling refers to non-additive variables (such as temperature), which cannot be aggregated/summed. |

|  |  |
| --- | --- |
| **Term** | **Description** |
| Areal interpolation | Transfer data from one set of objects (source zones) to another (target zones) based on the spatial relationship between the two. |
| Areal weighting | Also known as proportional reallocation; assumes that the population is evenly distributed in each source zone (i.e. the administrative district), and therefore the population can be allocated according to the overlapping area of the target zone (the grid) and the source zone. The simplest method and requires no ancillary data. |
| Pycnophylactic interpolation | Extension of areal weighting; assumes that the population of an area tends to be like the population of nearby areas (spatial autocorrelation), and attempts to produce a smooth/continuous surface, rather than severe jumps in estimated population at the source zone borders. The weighted average of its nearest neighbours is used to iteratively smooth the population values in the grids; in each iteration, the total population of the target zones is adjusted to ensure the same population count as the source zone. |
| Dasymetric Mapping | Subdividing of the source zone into smaller areas that reflect the spatial changes in population density based on ancillary data. The final step of dasymetry it to apply areal interpolation to generate the population distribution. |
| Binary Dasymetric Mapping | Also called mask area weighting. Uses ancillary data to divide a source zone into sub-zones – usually populated and unpopulated areas. The unpopulated area is masked out, and all the population is distributed to the populated area. Can also be used to categorise urban vs rural areas, and distribute population accordingly. |
| Multi-class Dasymetric Mapping | AN extension of binary, where the source zone is divided into several sub-zones. It is necessary to estimate the density of each sub-zone. Often directly uses land cover class as a sub-zone category, and a regression analysis to determine the population density of each class. Finally, the calculated population density of each class is scaled proportionally to ensure volume-preserving properties. |
| Intelligent Dasymetric Mapping | Most complex modelling method; utilises machine learning. Define and optimise the relationship (weighting factors) between the population data and ancillary data; subsequently, the population distribution is dasymetrically allocated. |

## Szarka 2022 – Inferring demographic characteristics in population estimation

|  |  |
| --- | --- |
| Full reference |  |
| Article type | Empirical |
| What is the article about? | Extending traditional population estimation models by predicting not only the number of residents but related demographic characteristics (i.e. age). |
| How are the aims explored/tested/ presented? | Implemented and compared different ML models (RF, SVM, Linear Regression) using point of interest (POI) and real estate data, applied to a case study in administrative areas of Singapore. |
| What are the main findings? | Results reveal that the age pattern of residents can be predicted using real estate information rather than with amenities, which is in contrast to estimating population counts. |
| What gaps does it identify? | Persistent gap in the literature in estimating socio-economic characteristics beyond the population count |
| Limitations/critiques | RQ is promising, but the focus on predicting an age and using real estate data weakens the relevance to dissertation topic |
| How does it link to wider learning? | Good discussion of the gaps in current population estimation focused literature |

## Wang 2022 – Global gridded GDP data set

|  |  |
| --- | --- |
| Full reference |  |
| Article type | Empirical |
| What is the article about? | Produces global gridded estimates of GDP over a regular time series, using night-time lights based spatial disaggregation methods |
| How are the aims explored/tested/ presented? | * Method combines population grid data from LandScan Global Population Database with night-time lights (NTL) * LandScan pop estimates chosen over others (e.g. WorldPop) due to its relatively higher accuracy at the county scale (?) and the details of the remote sensing and dasymetric techniques used in its derivation. * Compared three methods: Population-based disaggregation, NTL based, and hybrid NTL-population based. * The hybrid method (named *LitPop*) combines NTL with pop grids using a function in which GDP is proportional to NTL luminosity and pop count (NTL and pop count images are multiplied together to produce output image) |
| What are the main findings? | * GDP is most commonly used indicator to compare economic development of regions; therefore has huge application in understanding spatial distribution sub-nationally * LitPop approach superior to just population based disaggregation, especially at finer spatial scales. * Pure pop based approach relies on the assumption of a uniform GDP per capita distribution across a given administrative unit, which limits it. |
| What gaps does it identify? | No previous global estimates at this level of resolution and accuracy |
| How does it link to wider learning? | Example of disaggregation of alternative variables (GDP as opposed to population) |

## Wardrop 2018 – Spatially disaggregated population estimates in the absence of census data

|  |  |
| --- | --- |
| Full reference | 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>. |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Review/Editorial |
| What is the article about? | Reviews existing methods of spatial disaggregation for population estimates, and propose new methodology for future work, in the context of poor or absent census data. |
| What are the main findings? | * Population is never uniformly distributed within an areal unit; thus, aggregate population data (particularly at larger administrative areas) do not accurately represent the true spatial distribution of the population * Accurate disaggregated population data is important in development work, especially related to public health outreach and service delivery * Top-down vs Bottom-up disaggregation:   1. **Top-down:** dasymetric methods, such as equal weighting of census counts across grid cells in a given area, or distribution across cells identified as ‘built’ from satellite imagery – method used by the GPW   2. **Bottom-up:** approach proposed by Wardrop; relies on complete counts of population within small, defined areas (microcensus) which are then statistically extrapolated to cover the chosen area. Designed for cases where census data is unavailable or out-of-date. * “Top-down disaggregation [estimates] are only as good as the census data on which they are based” (p.3531) * BU approaches incorporate ancillary data such as built-up areas, land use type, dwelling counts, spectral characteristics, and socioeconomic or physical characteristics, to produce the most accurate estimates |
| What gaps does it identify? | Most work in this area has been in high-income, data rich countries; lacking application in areas with the greatest need. |
| Limitations/critiques | * Acknowledges that BU estimation is challenging to validate, due to being designed for prediction of non-sampled areas. * Requires the collection of a ‘micro-census’ of data; not in the scope of dissertation project |
| How does it link to wider learning? | Explores methodologies; especially related to asymmetric mapping |
| Additional notes  *(incl. useful quotes)* |  |

## You 2006 – An entropy approach to spatial disaggregation of crop production

|  |  |
| --- | --- |
| Full reference |  |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Empirical |
| What is the article about? | Development of an entropy-based spatial disaggregation method to estimate fine-scale crop production from national or regional level statistics |
| How are the aims explored/tested/presented? | Tabular crop production statistics are combined with an array of secondary data (farming system, LULC, biophysical crop suitability assessments, population density) to model spatial disaggregation to the pixel level (approx. 80km2). Case study on Brazil.  The model was then evaluated by comparing aggregated pixel values to the original regional area statistics.  Finally, performance was compared to simpler methods (such as dasymetric spread across the total land area of the region, or the total cropland area of the region (masked binary method). For all crop types, the cross-entropy method performed best. |
| What are the main findings? | * Annual crop production data usually only available on a national or regional scale   **Spatial Allocation Model:**   * Reinterpret classified land cover imagery into crop land and non-crop land (analogous to binary masking), and then utilises the input data in a cross entropy approach * “The intent of the spatial allocation model is not to try to match the real world pixel by pixel, but rather to derive a substantially more information picture of the likely distribution … than the state level statistics alone can reveal” (p.339) |
| What gaps does it identify? | * Identified weakness in the reliability of input data; many inconsistencies between the different input layers. Response: devised a set of rules to progressively relax the constraints until the allocation could be completed successfully. * Highlights the modifiable areal unit problem (MAUP); read further on this; relevance to my project? |
| Limitations/critiques | Does not explicitly cover what a ‘cross-entropy’ approach is in this paper; but references other papers describing the entropy modelling process.  Also, the focus on crop production (as opposed to the human population) lessens the relevance to dissertation. |
| How does it link to wider learning? | Alternative methodologies for spatial disaggregation. Specifically compares performance with the masked binary method, which is used by FL/SA for Sri Lanka. |
| Additional notes  *(incl. useful quotes)* |  |

# Agriculture dependent populations

## World Bank 2023 – Agriculture and Food

|  |  |
| --- | --- |
| Full reference |  |
| Article type  *(e.g., Empirical, Theoretical, Conceptual)* | Web page |
| What is the article about? | High level info page introducing the topic of agriculture and food as it relates to poverty, and the work of the World Bank |
| What are the main findings? | * “Agricultural development is one of the most powerful tools to end extreme poverty, boost shared prosperity, and feed a projected 9.7 billion people by 2050”   **World bank agriculture-related goals:**   * Improving food security for all, including access to safe and nutritious food * Making agriculture and food more sustainable and more resilient to climate change |
| How does it link to wider learning? | Will need to reference World Bank as part of broader scope and background of project. Good to include their strategy and how this research would assist in achieving specific goals/outcomes |
| Additional notes  *(incl. useful quotes)* |  |