

# **Techniques for Analysing the Relationship between Population Density and Geographical Features of Interest**

Amanda Johnson and Colin Arrowsmith  
School of Mathematical and Geospatial Sciences

RMIT University  
Email: aldkjohnson@bigpond.com

## **Abstract**

This paper presents a study that aimed to explore a range of techniques for analysing the spatial relationship between population density and geographical features of interest. Three categories of spatial analysis techniques were explored: traditional methods including descriptive statistics and spatial distribution maps, spatial autocorrelation statistics namely Moran's global I and Anselin's local I and regression analysis incorporating ordinary least squares (OLS) regression and error residual testing using the global Moran's I autocorrelation statistic. The correlation between the spatial distribution of Australian cinema screens and a global gridded population density dataset were used as the case study for the analysis.

All three categories of spatial analysis techniques were found to be useful, each with its' own strengths and weaknesses. The analysis was able to visually and statistically identify the spatial distribution of cinema screens and population density, as well as establish the degree to which the two features were correlated. The study concluded that a methodology which utilises the above spatial analysis techniques in conjunction with a global gridded population dataset, provides a sound framework for investigating the correlation between population distribution and a geographical feature of interest.

## **Authors**

Amanda Johnson holds a Master of Applied Science (GIS) from RMIT University and a Bachelor of Economics from Monash University. Amanda's research interests include the application of spatial information systems and spatial statistics in the field of Public Health. She is currently working at the Murdoch Children's Research Institute (MCRI) and has 11 years of experience in the ICT consulting industry as a manager for Accenture.

Colin Arrowsmith is Associate Professor in the School of Mathematical and Geospatial Sciences at RMIT University. He holds a Doctor of Philosophy from RMIT as well as two masters' degrees and a bachelor's degree from the University of Melbourne, and a Graduate Diploma of Education from Hawthorn Institute of Education. Colin has authored more than 40 refereed publications and 6 book chapters in the fields of GIS, tourism analysis and in film studies. Colin's research interests include the application of spatial information systems, including geographic information systems (GIS), to investigating the impact of tourism on nature-based tourist destinations, tourist behaviour, as well as investigating the issue of managing micro-historical data within GIS utilising cinema data.

**Keywords:** Spatial analysis, Moran's I, Anselin's I, Global gridded dataset, Regression analysis, Cinema and screen studies.

## **Introduction**

Spatial analysis is the process by which the locational distribution of a set of features is investigated, in order to understand the underlying processes which led to their generation. It has been applied in multiple fields of study including: epidemiology, (Hay et al., 2004), ecology, (Luck et al., 2010), criminology, (Wu and Grubestic, 2010) and geography, (Getis, 1992). Multiple tools are available for use in spatial analysis and this paper will explore a number of them, with an emphasis on techniques which allow the relationship between population density and the spatial distribution of a geographic feature of interest to be investigated.

Three categories of spatial analysis techniques will be explored, using the correlation between the spatial distribution of Australian cinema screens and a global gridded population density dataset as a case study. Category one is traditional analysis techniques and incorporates descriptive statistics and spatial distribution maps. Category two is spatial autocorrelation statistics and incorporates Moran's Global I, (Moran, 1948) and Anselin's Local I, (Anselin, 1995). The final category is regression analysis and incorporates ordinary least squares (OLS) regression and error residual testing using the global Moran's I autocorrelation statistic.

A literature review of existing spatial analysis studies was unable to identify any studies in the area of cinema and screen studies however a number of Australian spatial analysis studies in other fields were identified. Hu, (2011) used spatial autocorrelation statistics to analysis the spatial distribution of notified dengue fever infections in the state of Queensland and Luck et al., (2010) used ordinary least squares regression to assess the correlation between human population density and bird species richness across Australia.

## **Materials and Methods**

### ***Study area***

Australia is located in the Southern Hemisphere, between the South Pacific Ocean and the Indian Ocean, at latitudes of 10-45° S and longitudes of 113-153° E. The majority of Australia's population is located in the major metropolitan coastal areas, particularly in the east south-east cities of: Brisbane, Sydney, Canberra, Melbourne and Adelaide. Australian Bureau of Statistics data indicates that as at June 2010, 14.3 million people, or 64% of Australia's population lived in the capital cities, (ABS, 2012). In contrast, the central desert regions of the country located more than 1000 km from the coast, have extremely sparse population densities (ABS, 2012).

### ***Data Collection***

Two primary sources of data were used in this study: cinema screen locations and population density.

The cinema screen location dataset was a shapefile containing the longitude and latitude values of Australian cinema screen addresses, along with the number of screens situated at each cinema, as at the start of 2012 stored as point features. In total there were 501 cinemas with 2,014 screens between them, the details of which were collated as part of ARC research project (DP120101940) and obtained from a third party data collector.

Population density data was sourced from the Socioeconomic Data and Applications Centre (SEDAC) at Columbia University. Three Gridded Population of the World version 3 (GPW v3) datasets were downloaded from the SEDAC website, all of which were in a global gridded raster format as at the Year 2000, with a spatial resolution of 2.5 arc-minute grid cells (~5 km<sup>2</sup> at the equator):

- Population Count Grid, v3 (2000); each cell estimates human head count, (Center for International Earth Science Information Network - CIESIN - Columbia University et al., 2005),
- Land and Geographic Unit Area Grids, v3 (2000); each cell measures land in km<sup>2</sup>, (Center for International Earth Science Information Network - CIESIN - Columbia University and Centro Internacional de Agricultura Tropical - CIAT, 2005a) and
- Population Density Grid, v3 (2000); each cell estimates the number of people per km<sup>2</sup>, (Center for International Earth Science Information Network - CIESIN - Columbia University and Centro Internacional de Agricultura Tropical - CIAT, 2005b).

A global gridded population density dataset was used in this study because one of the goals of the ARC project was to develop a methodology for spatial analysis that was globally applicable. Therefore a population density dataset that was consistent across national boundaries was selected. Global gridded density datasets use spatial interpolation to transform native population census data of varying resolutions into a set of standard gridded latitude-longitude cells, (Balk et al., 2006). In turn this provides the analyst with the ability to select the geographic boundaries most appropriate for a given study, rather than being restricted by national administrative boundaries, (Linard et al., 2010).

There are four key global gridded population density datasets available, (Tatem et al., 2012):

- The Global Rural Urban Mapping Project (GRUMP), which has a spatial resolution of ~1km., (Center for International Earth Science Information Network - CIESIN - Columbia University et al., 2011),
- LandScan, with a spatial resolution of ~1km., (Dobson et al., 2000),
- The United Nations Environment Programme (UNEP) Global Population Databases, (UNEP, 2010), with a spatial resolution of ~5km and
- The Gridded Population of the World version 3 (GPW3), with a spatial resolution of ~5km, (Center for International Earth Science Information Network - CIESIN - Columbia University and Centro Internacional de Agricultura Tropical - CIAT, 2005b).

The original aim of the study was to use the smallest available spatial resolution size and therefore the GRUMP dataset was selected because it has a spatial resolution of ~1 km and is freely available, whereas LandScan was only available for a fee. However the software was unable to cope with high volume of data generated and therefore the GPW3 dataset with a coarser spatial resolution of ~5 km was used. The GPW3 dataset was selected above the UNEP dataset because it is based on more recent census data, year 2000 vs. year 1990 respectively.

Studies which have utilised global gridded population datasets for global scale analysis include: Bhatt et al., (2013) who used spatial mapping techniques to analysis of the global distribution of dengue fever and Snow et al., (1999) who used spatial mapping techniques to analysis of the global distribution of malaria.

### ***Data Preparation***

In order to utilise the spatial analysis tools investigated by this study it was necessary to transform and spatially join the four original datasets of varying formats, into one shapefile of consistently formatted data. First the three GPW data files were converted from raster data to integer point data so that it would be possible to spatially join GPW attribute values with cinema screen numbers. A fishnet shapefile was then created with the same spatial resolution as the GWP data files (2.5 arc-minute grid cells) and overlaid on the study area. The fishnet cells were then used as a template for a cell by cell spatial join between the 3 GPW datasets and the cinema screen location dataset. Finally two additional values were calculated for each cell: screen density per km<sup>2</sup> and screen density per person.

The end result was a shapefile of integer points, one for each of gridded cells in the original SEDAC population density dataset. Attached to each point were six cell attribute values: number of cinema screens, land area in km<sup>2</sup>, population count, population density per km<sup>2</sup>, screen density per km<sup>2</sup> and screen density per person.

The software used to prepare the data and subsequently conduct the spatial analysis was ArcGIS® and ArcMap™ by Esri.

### ***Traditional Spatial Analysis Tools***

Traditional spatial analysis tools were utilised in order to gain an understanding of the underlying data structure of the features in the study and to visualise their spatial distribution, using descriptive statistics and spatial mapping techniques respectively. The descriptive statistics utilised were: summary statistics, box plots, histograms and cumulative percentage diagrams. The spatial mapping techniques utilised were: graduated point patterns, spatial density patterns, geographic standard deviation ellipses and geographic mean centres.

Four datasets were analysed using the above tools. The two primary sources of data: cinema screen locations as a graduated point pattern and population density per km<sup>2</sup> and the two datasets created by manipulating the original datasets: screen density per km<sup>2</sup> and screen density per person.

### ***Spatial Autocorrelation Statistics***

Spatial autocorrelation statistics were used to measure the degree of spatial association between the features. They differ from traditional statistics because they simultaneously consider both locational and attribute information and include the concept of space in their mathematical formalisms (Fischer, 2010). Two spatial autocorrelation statistics were used in the study: the global Moran's I statistic and the local Anselin's I statistic. Global statistics measure the degree of spatial association between features across the study region as a whole and local statistics measure the variation in feature spatial patterns within the study area.

Both are inferential statistics with a null hypothesis of complete spatial randomness (CSR). A value close to -1 indicates the presence of strong negative spatial autocorrelation (dispersion) between the features in the study area. A value close to +1 indicates strong positive spatial autocorrelation (clustering) between features and a value near 0 indicates spatial randomness between features, (Anselin, 1995).

Moran's global I was calculated at multiple distance thresholds ranging from 5 km to 100 km in order to identify the distance threshold at which spatial autocorrelation was most pronounced. Five datasets were tested: cinema screen point pattern, cinema screens aggregated per fishnet cell, population density per km<sup>2</sup>, screen density per km<sup>2</sup> and screen density per person. Cinema screen data was categorised and tested in multiple ways to determine what format was the most effective one for identifying spatial autocorrelation; if a dataset contains relatively static values it can be difficult to identify global spatial autocorrelation.

Anselin's local I was calculated using a distance threshold of 20 km, the distance at which the global Moran's I value peaked during the global spatial autocorrelation testing process. Three datasets were tested for local spatial autocorrelation: screens per km<sup>2</sup>, population density per km<sup>2</sup> and screens per person.

### ***Regressions Analysis***

Two regression analysis techniques were used to explore the correlation and relationship between cinema screen density and population density: Ordinary Least Square (OLS) linear regression and spatial autocorrelation testing of the error residual values using the Moran's I statistic. If spatial autocorrelation is present in the error residuals of a model, it is an indication that the explanatory variable is unable to explain the inherent spatial structure of the dependent variable and therefore the model is missing one or more explanatory spatial variables.

Two OLS regression models were tested: one which included all values in the population density per km<sup>2</sup> dataset and a second which excluded the outlier values in the dataset. In both models the explanatory variable was population density per km<sup>2</sup> and the dependent variable was the cinema screen density per km<sup>2</sup>.

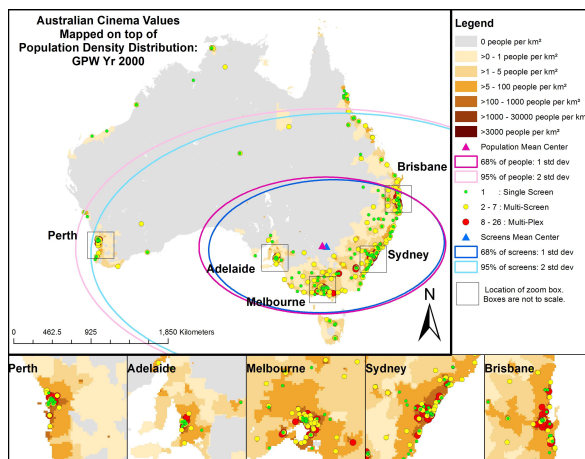
## **Results**

### ***Traditional Spatial Analysis Tools***

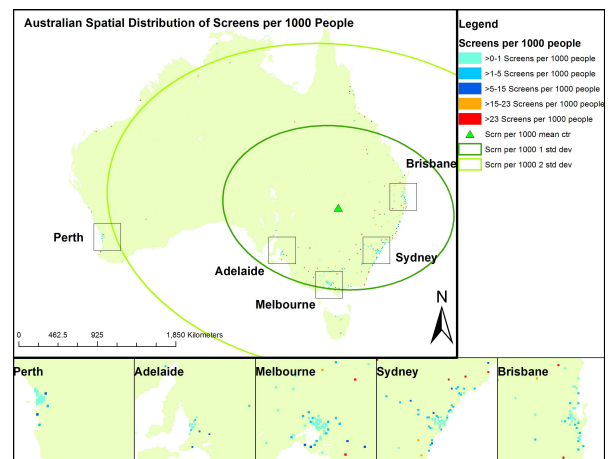
Table 1 shows the summary statistics for the four datasets analysed using traditional methods. The average number of screens per cinema is 4 and the median 3, the maximum number of screens per cinema is 26 and the distribution has a positive skewness value of 1.6. For those fishnet cells which have 1+ cinemas located within them, the average number of screens per km<sup>2</sup> is 0.46 and the median is 0.29, the maximum is 4 screens per km<sup>2</sup> and the distribution has a positive skewness value of 3.29. The average population density is 2.7 people per km<sup>2</sup> and the median is 0, the maximum cell value is 4943 people per km<sup>2</sup> and the distribution has a positive skewness value of 42.45. For those fishnet cells which have 1+ cinemas located within them, the number of screens per 1000 people is 18.3 and the median is 1.1, the maximum is 1500 screens per 1000 people and the distribution has a positive skewness value of 12.1.

<b>Summary Statistic</b>	<b>No. of screens per cinema</b>	<b>No. of screens per km<sup>2</sup></b>	<b>No. of people per km<sup>2</sup></b>	<b>No. of screens per 1000 people</b>
Mean	4.00	0.46	2.70	18.32
Median	3.00	0.29	0.00	1.05
Mode	1.00	0.06	0.00	125.00
Skewness	1.60	3.29	42.45	12.05
Minimum	1.00	0.05	0.00	0.04
Maximum	26.00	4.00	4943.00	1500.00

Figure 1 depicts the spatial distribution of Australian cinema screens as a graduated point pattern, mapped on top of the spatial density distribution of Australia's population as a raster pattern. The one and two standard deviation ellipses and the geographic mean centres of the two datasets are also displayed, screens in pink, population in blue. When these datasets are mapped together there appears to be a strong correlation between population distribution values and cinema points. The dark brown more densely populated areas have multiple cinema points mapped on them, in particular the yellow multiscreen cinemas and red multiplex cinemas. Correspondingly, the lighter brown less populated areas of Australia have fewer cinema points mapped on them and they are often green one screen cinemas. The triangles representing the geographical mean centre of the datasets are located in close proximity in the south-east corner of mainland Australia, as are the standard deviation ellipses.



**Figure 1. Spatial distribution of Australian Cinema mapped on Population Density**



**Figure 2. Spatial distribution of Australian Cinema Screens per 1000 people**

Figure 2 depicts the spatial distribution of Australian cinema screens per 1000 people. In the major metropolitan areas the distribution of screens per 1000 people appears to be the inverse of population density distribution. The central CBD regions have the lowest density of screens per 1000 people, displayed in pale blue. Screen density then increases and changes to darker blue as distance from the CBD increases. The highest screen density per 1000 people cells are displayed in orange and red and are located away from the major CBD areas. In comparison to Figure 1, the geographical mean centre and standard deviation ellipses are located slightly further north but are still near the eastern seaboard.

### ***Spatial Autocorrelation Statistics***

Figure 3 plots Moran's I values against distance and the yellow line represents the expected Moran's I value (approximately 0) for complete spatial randomness. Figure 4 plots the corresponding Z-scores of the Moran's I results against distance and the yellow and orange lines represent the point at which statistically significant clustering is considered to be present in the dataset, at a 95% CI and 99% CI respectively.

The green line maps the Moran's I and Z-score results for the population density per km<sup>2</sup> dataset. For all distances tested, statistically significant global positive spatial autocorrelation (clustering) with >99% CI level was identified. The peak Moran's I value was 0.38 at a distance threshold of 15 km. Of the three cinema screen dataset formats tested for global spatial autocorrelation, the most effective format for identifying spatial autocorrelation was the number of screens per cell dataset, displayed as the purple line in both diagrams. For all distances tested, the trend of the purple cinema screen line mirrored that of the green population density line but at a lower magnitude. Positive spatial autocorrelation with a >99% CI was identified at all distance thresholds and the peak Moran's I value was 0.2 at a distance of 20 km.

Screen density per person is represented by the pink line. The Moran's I values for this dataset were relatively static for all distance thresholds tested and were only marginally higher than the expected Moran's I values. The corresponding Z-scores were well below the 95% CI for positive spatial autocorrelation, indicating that when cinema screens are weighted by population density clustering is removed the dataset and complete spatial randomness is said to exist.

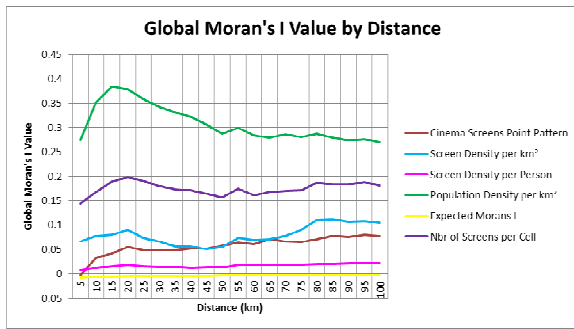


Figure 3. Moran's I Values by Distance.

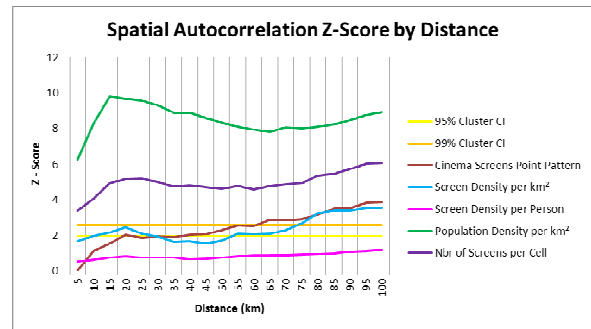


Figure 4. Z-Scores by Distance.

Figure 5 maps the results of the Anselin local I spatial autocorrelation statistic using a distance threshold of 20 km, for each screen venue for screens per km<sup>2</sup> and population per km<sup>2</sup> datasets. Anselin's local I was calculated at a 95% CI for both datasets. Positive spatial autocorrelation was found to be present in both datasets, population density clusters are displayed in red and screen density clusters are displayed in purple. Cells which were found to be statistically non-significant are displayed in grey for the population density dataset and yellow for the screen density dataset. No outliers were found in the population density dataset but some were identified in the screen density dataset and are displayed as in pink and blue.

The spatial distribution of the clustering varies between the two datasets. Screen density clusters are only found in the four major south east metropolitan areas of: Brisbane, Sydney, Melbourne and Perth. Population clusters are found in all of these areas but to a larger extent and are also present in Perth, Darwin, along the eastern seaboard and Tasmania.

Figure 6 maps the results of the Anselin's local I spatial autocorrelation statistics for the screens per person dataset. There are only three locations which have statistically significant results: Kununurra, Karlgoolie and Broome, all of which are considered to be outliers. No positive spatial autocorrelation was found to exist in the dataset, indicating that when screen density is weighted by population density, clustering is removed the dataset and complete spatial randomness is said to exist.

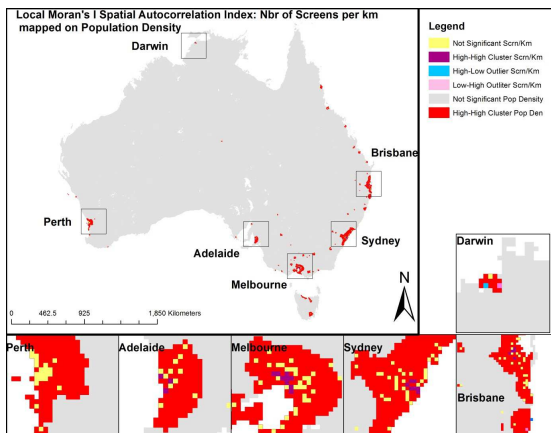


Figure 5. Statistically significant local Anselin's I cells: Screen density is mapped on top of population density.

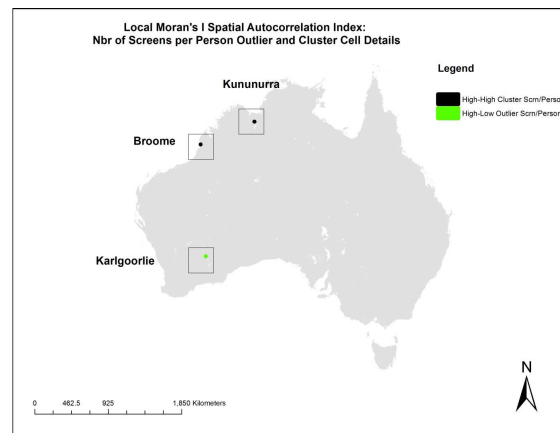


Figure 6. Statistically significant local Anselin's I cells: Screens per person.

### Regression Analysis

Table 2 depicts the diagnostic value results for the two OLS regression models. Model 1 was generated based on all values in the screens per km<sup>2</sup> dataset and has an Akaike's Information Criterion (AIC) value of 485 and an adjusted R<sup>2</sup> value of 19.26%. Model 2 was generated after outlier values were removed from the screens per km<sup>2</sup> dataset and as a result both diagnostic values improved, the AIC value fell to 157 and the adjusted R<sup>2</sup> value increased to 43.41%. The Jarque-Bera (JB) statistic indicates model bias by examining whether the error residuals deviate from a normal distribution. Model 1 has a JB statistic value of 8007 much higher than Model 2's value of 235, however both have a JB p-value of 0.00, indicating that the JB statistic is statistically significant at a >99% CI and therefore both models are biased.

Table 2. OLS Regression Diagnostic Values

Diagnostic Value	Model 1	Model 2
AIC	484.8	157.1
Adj R <sup>2</sup>	0.1926	0.4341
JB	8007.7	235.4
JB-Prob	0.00	0.00

Equations 1 and 2 are the OLS regression equations for Models 1 and 2 respectively. In both models the dependent Y variable is screen density per km<sup>2</sup> and the explanatory X variable is population density per km<sup>2</sup>. For both models when population density increases by one person per km<sup>2</sup>, there is a corresponding increase in screen density of 0.0003 screens per km<sup>2</sup>.

$$Y = 0.2677 + 0.0003X$$

Equation 1: OLS Regression Model 1

$$Y = 0.1730 + 0.0003X$$

Equation 2: OLS Regression Model 2

Figure 7 is a scatterplot of standardised error residual values vs. expected screen density values for Model 2. The cone shape of the dot distribution indicates that the model contains bias due to heteroscedasticity i.e. the relationship between the dependent and explanatory variables is not consistent in data space. Error residual values are smaller for low screen density values and larger for high screen density values and in general the magnitude of under estimation errors (red dots above the line) is greater than the magnitude over estimation errors (blue dots below the line). The colour of the dots matches the legend in Figure 8.

Figure 8 maps the geographic location of the OLS error residuals for Model 2: over estimations are blue, under estimations are red and yellow indicates small errors. The location of the outlier cells excluded from the screen density dataset in pink. Most error residual cells are shaded yellow and located away from the central CBD zones, indicating the model has predicted screen density relatively accurately in these areas. The strongest coloured red and blue cells are located in central CBD regions, indicating that the model predicts screen density values poorly in these areas. The pink outlier cells are primarily located along the coast line.

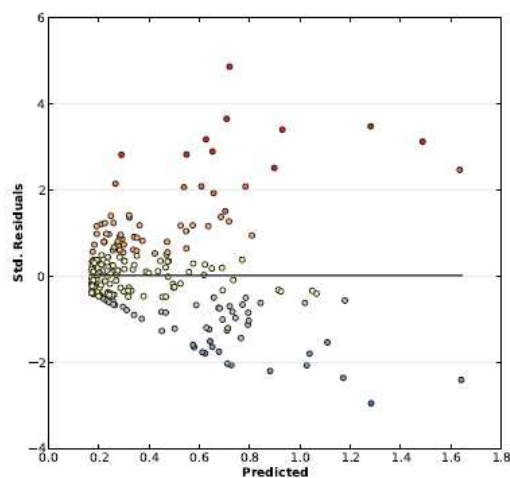


Figure 7. OLS Model 2: Scatterplot of Standard Error Residuals vs Expected Screen Density Values

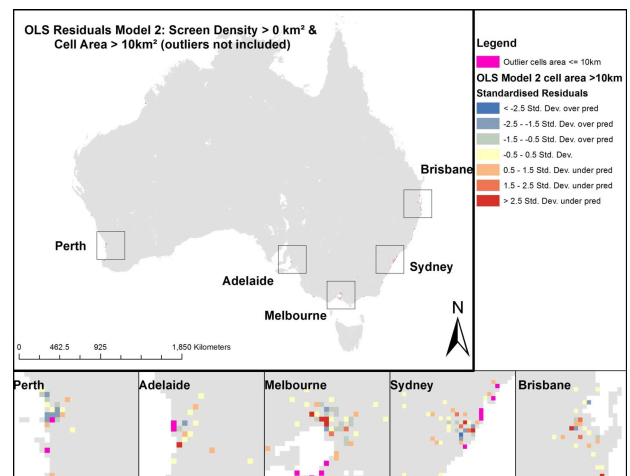


Figure 8. OLS Model 2: Geographical location of Standard Residuals

Table 3 depicts the results for both models when the error residual values were tested for spatial autocorrelation using the Moran's I statistic. Model 1 has a near zero Global Moran's I value of 0.0039 and a P-value of 0.8443, indicating that the distribution of the error residual values is random, i.e. no spatial autocorrelation is present in the error residual values. Model 2 also has a near zero Global Moran's I value of -0.0117 and a p-value of 0.8315, again indicating that the distribution of the error residuals is random. Based on the above results it could be concluded that population distribution adequately explains the inherent spatial structure of screen density distribution in both OLS models and therefore any bias present in the model is not due to a missing explanatory spatial variable.

Table 3. Error Residual Global Moran's I Results

	Model 1	Model 2
Global Moran's Index	0.0039	-0.0117
Expected Index	-0.0031	-0.0035
Variance	0.0013	0.0015
z-score	0.1964	-0.2128
p-value	0.8443	0.8315

## Conclusions

All of the spatial analysis tools explored by this study were found to be useful aids when conducting spatial analysis, each having its own strengths and weaknesses.

The traditional spatial analysis tools explored included descriptive statistics and spatial distribution maps. While descriptive statistics did not prove to be a tool capable of analysing the spatial distribution of a dataset or correlations between the screen density and population density datasets, the insight gained regarding the underlying structure of the input datasets was invaluable input for subsequent analysis processes. In contrast, spatial distribution maps were found to be very useful tools for analysing the spatial distribution pattern of cinema screen locations and population density and the degree to which the two were correlated.

In summary traditional spatial analysis tools provided the following insight. Australia has 501 cinemas with a total of 2,014 screens and they have a spatial distribution pattern which mirrors population density distribution. The vast majority of cinema screens and people are located on or near the Australian mainland coastline in particular the east and south-east mainland coastlines and the major metropolitan cities have the highest densities of screens and population.

Spatial autocorrelation statistics were found to be useful tools for understanding the spatial clustering patterns inherent in the screen density and population density datasets. Moran's I autocorrelation statistic is a global statistic and therefore only provided an average for the whole study area, which to some degree limited its diagnostic value. However it was an invaluable tool for identifying the distance threshold at which spatial clustering was most pronounced, an input parameter required by the local Anselin's I statistic. Anselin's local I spatial autocorrelation statistic was a particularly useful tool not only to answer the question "where are clusters and outlier values located?" but also for deriving some understanding of the correlation between the spatial distribution of cinema venues and population density.

The analysis conducted using the spatial autocorrelation statistics indicated that positive spatial autocorrelation (clustering) was present in the spatial distribution of cinema screens at a 95% CI, in the four largest east coast metropolitan areas: Adelaide, Melbourne, Sydney and Brisbane. The population density dataset was also found to be clustered at these locations, as well as a number of other predominately coastal locations. Additionally when cinema screens were weighted by the number of people per screen no spatial autocorrelation was identified, indicating that a correlation exists between population density and cinema screen density.

Ordinary Least Squares (OLS) linear regression was found to be a very powerful spatial analysis tool for two reasons. Not only was it able to explore the correlation and relationship between screen and population density, it was also able to assess the degree to which outlier values negatively impacted the study.

Regression analysis indicated that population density distribution is a statistically valid explanatory variable for cinema screen distribution, with an adjusted R<sup>2</sup> value of 43.41% when outliers were removed from the dataset. Analysis of the residual error values indicated that bias due to heteroscedasticity was present in the model and therefore one or more key explanatory variables were missing from the model. Specifically the model was less accurate at predicting screen density values in CBD metropolitan areas. The error residuals were also tested for the presence of spatial autocorrelation, none was identified and therefore it was concluded that population density distribution adequately explains the inherent spatial structure of screen density distribution.

In conducting this analysis a number of limiting factors were identified, the two most prevalent being dataset misalignment issues and cinema screen weighting issues. The physical boundaries of the cinema location and SEDAC population datasets did not totally correlate and therefore when they were combined to create the screens per km<sup>2</sup> and screens per person datasets, some cinemas were excluded from the dataset and a number of



abnormally high outlier values were created. Screen weighting issues refer to the fact that all screens in the analysis were given an equal weighting, regardless of how many times per day they were viewed.

Additional limiting factors include a 12 year time misalignment between the Year 2000 SEDAC population datasets and the Year 2012 cinema location dataset and recognised weakness in cartographic mapping techniques. These include unstable abnormally high outlier values due to the small numbers problem and the modifiable areal unit problem which may mask the true correlation between screen distribution and population density.

There are a number of factors that could be incorporated in future studies using the above techniques, which may help mitigate some of the limiting factors discussed. Spatial filtering techniques may help smooth the edge effect issues related to boundary misalignments, heteroscedasticity bias may be improved if cinema screens were weighted by their viewing rate and the utilisation of a more recent population dataset would address the issue of time misalignment.

In conclusion it was found that a methodology which utilises the above spatial analysis tools in conjunction with a global gridded population dataset, provides a sound framework for investigating the correlation between population distribution and a geographical feature of interest. The analysis undertaken was able to visually and statistically identify the spatial distribution of cinema screens and population density, as well as establish the degree to which the two features were correlated.

## Acknowledgements

This study was conducted as part of ARC research project (DP120101940).

## References

- ABS. 2012. *GEOGRAPHIC DISTRIBUTION OF THE POPULATION* [Online]. Canberra, Australia: Australian Government. Available: <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1301.0~2012~Main%20Features~Geographic%20distribution%20of%20the%20population~49> [Accessed 18/08/2014 2014].
- ANSELIN, L. 1995. Local Indicators of Spatial Association - LISA. *Geographical Analysis*, 27, 23.
- BALK, D., DEICHMANN, U., YETMAN, G., POZZI, F., HAY, S. & NELSON, A. 2006. Determining global population distribution: methods, applications and data. *Adv Parasitol*, 62, 119 - 156.
- BHATT, S., GETHING, P. W., BRADY, O. J., MESSINA, J. P., FARLOW, A. W., MOYES, C. L., DRAKE, J. M., BROWNSTEIN, J. S., HOEN, A. G. & SANKOH, O. 2013. The global distribution and burden of dengue. *Nature*.
- CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK - CIESIN - COLUMBIA UNIVERSITY & CENTRO INTERNACIONAL DE AGRICULTURA TROPICAL - CIAT 2005a. Gridded Population of the World, Version 3 (GPWv3): Land and Geographic Unit Area Grids. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).
- CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK - CIESIN - COLUMBIA UNIVERSITY & CENTRO INTERNACIONAL DE AGRICULTURA TROPICAL - CIAT 2005b. Gridded Population of the World, Version 3 (GPWv3): Population Density Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).
- CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK - CIESIN - COLUMBIA UNIVERSITY, INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE - IFPRI, THE WORLD BANK & CENTRO INTERNACIONAL DE AGRICULTURA TROPICAL - CIAT 2011. Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Density Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).
- CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK - CIESIN - COLUMBIA UNIVERSITY, UNITED NATIONS FOOD + AGRICULTURE PROGRAMME - FAO & CENTRO INTERNACIONAL DE AGRICULTURA TROPICAL - CIAT 2005. Gridded Population of the World, Version 3 (GPWv3): Population Count Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).
- DOBSON, J., BRIGHT, E., COLEMAN, P., DURFEE, R. & WORLEY, B. 2000. LandScan: a global population database for estimating populations at risk. *Photogram Eng Rem Sens*, 66, 849 - 857.
- FISCHER, M. G., A 2010. Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. *In: FISCHER, M. G., A (ed.)*. Berlin Heidelberg: Springer.
- GETIS, A. O., J 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24, 18.

- HAY, S. I., GUERRA, C. A., TATEM, A. J., NOOR, A. M. & SNOW, R. W. 2004. The global distribution and population at risk of malaria: past, present, and future. *The Lancet Infectious Diseases*, 4, 327-336.
- HU, W., CLEMENTS, A., WILLIAMS, G., & TONG, S. 2011. Spatial analysis of notified dengue fever infections. *Epidemiology and Infection*, 139, 8.
- LINARD, C., GILBERT, M. & TATEM, A. 2010. Assessing the use of global land cover data for guiding large area population distribution modelling. *GeoJournal*.
- LUCK, G. W., SMALLBONE, L., MCDONALD, S. & DUFFY, D. 2010. What drives the positive correlation between human population density and bird species richness in Australia? *Global Ecology and Biogeography*, 19, 673-683.
- MORAN, P. 1948. The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society.*, 10, 9.
- SNOW, R., CRAIG, M., DEICHMANN, U. & MARSH, K. 1999. Estimating mortality, morbidity and disability due to malaria among Africa's non-pregnant population. *Bull World Health Organ*, 77, 624 - 640.
- TATEM, A., ADAMO, S., BHARTI, N., BURGERT, C., CASTRO, M., DORELIN, A., FINK, G., LINARD, C., JOHN, M., MONTANA, L., MONTGOMERY, M., NELSON, A., NOOR, A., PINDOLIA, D., YETMAN, G. & BALK, D. 2012. Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. *Population Health Metrics*, 10, 8.
- UNEP, U. N. P. D. 2010. World population prospects, 2010. *Book World population prospects, 2010 revision*.
- WU, X. & GRUBESIC, T. H. 2010. Identifying irregularly shaped crime hot-spots using a multiobjective evolutionary algorithm. *Journal of Geographical Systems*, 12, 409+.