Exploration of Relationship between Social Status and Mortality Rates in England

Xin Zhang, K. Tocque, J. Boothby, P. Cook, MENGSHI LI
Centre for Public Health
Liverpool John Moores University
Email: x.zhang@2008.ljmu.ac.uk

M. S. Li
Department of Electrical Engineering and Electronics
The University of Liverpool

1 Introduction

It is well established that health is related to Socio-Economic Etatus (SES) or deprivation status (Adler (1994)). Socio-Economic models or multiple deprivation models of health and inequalities are widely used by Public Health practitioners. Since 2000, the Department of Health has developed a pathway, which approach to determine the important influences on health within a population. A lot of researchers think social position or socioeconomic status as the fundamental cause of ill health (Davey Smith and Hart (1998)). There is now a large literature which shows that the socioeconomically disadvantaged experience higher mortality rates for most major causes of death, and this inequality exits at every stage of life-course (Turrell and Mathers (2001)).

In this paper, three different approaches, Ordinary Least Square (OLS) regression, Geographically Weighted Regression (GWR) and Artificial Neural Network (ANN) (Rojas (1996)), are applied for exploring the association between the Socio-Economic status or deprivation and mortality data in England. Researchers, with the spatial data being in use, are increasingly aware of the limitation of global regression techniques, such as Ordinary Least Square regression, which by generating 'global' outputs may mask local variations and ignoring the spatial relationship between variables (Huang and Leung (2002)). GWR, a local regression technique, has been applied in many instances to successfully reveal local relationship as well as spatial association between variables (Fotheringham *et al.* (2002)). However, the linear functional relationship of the OLS and GWR assumptions can not totally reflect reality. ANN, as a powerful tool and robust predictor which is capable of modelling non- linear relationships, is used gradually in health research. This research attempts to compare with the results among three different methods to provide guidelines for future analysis of socioeconomic status and mortality.

2 Data Sources

The models are based on data from the 2007 all-cause mortality cases and Index of Multiple Deprivation (IMD) 2007 for England, at the spatial scale of Lower Super Output Areas.

Domains	Percentage (total 100 %)
Income	22.5%
Employment	22.5%
Health Deprivation and Disability	13.5%
Education, Skills and Training	13.5%
Barriers to Housing and Services	9.3%
Living Environment	9.3%
Crime	9.3%

Table 1: The English indices of deprivation 2007 summary

Super Output Areas (SOAs) are a new national geography created by the Office for National Statistics (ONS) for collecting, aggregating and reporting statistics. SOAs are currently available at two levels - Lower Level and Middle Level - and are defined using groups of Output Areas. Bacically, a Lower Super Output Area (LSOA) contains approximately 1500 people and 750 households.The government of England produced IMD in 2000 then updated in 2007. The overall IMD is conceptualised as a weighted area level aggregation of these specific dimensions of deprivation. Seven domains are included in the IMD2007 listed in Table 1.

2.1 Selection of dependent variables

This paper's models are built to assess the age-specific mortality rates (per-millage) throughout England. The mortality rates are classified into five age groups, *i. e.*, 0-19, 20-64, 65-85, 20-44, and 45-65 at each Lower Supper Output Area (LSOA) respectively .

2.2 Selection of explanatory variables

The overall IMD 2007 and the four sub-domains of IMD 2007, which include income index, employment index, education index, and health index, are chosen as a proxy for socioeconomic status to explain the five age groups mortality rates. Meanwhile, the mean values of IMD 2007 of each LSOA, which is calculated by average the values in surrounding areas (First order adjacent neighbourhood), are indicated as a part of explanatory variables. The ratio between the population in five age groups (0-19, 20-64, 65-85, 20-44, and 45-65) and total population at LSOA is employed to control the dependent variables.

3 Methodologies

The Ordinary Least Squares (OLS) for global regression is adapted initially in this study to assess mortality rates of each age group in Arcgis 9.3. This method is constructed as

Age groups	Moran' I	Standardised Z scores
0-19 years	0.189	1.886
20-64 yesrs	0.385	2.833
65-85 years	0.209	1.989
20-44 years	0.223	1.994
45-65 years	0.397	2.889

Table 2: Moran's I statistics and standardised Z scores of model residuals by using OLS regression

below:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_i X_i + \varepsilon \quad i \in N, \tag{1}$$

where Y indicates mortality rates for each age group at LSOA, X_i indicates the i_{th} explanatory variables, β_i indicates the i_{th} regression coefficients, and ε indicates the residual.

Then the Moran's I statistics is used to examine any exiting spatial autocorrelation in the residuals of five models (Fotheringham EL.,2002).

Table 2 implies that residuals have spatial autocorrelation. Therefore, the GWR is applied to modify the models. In geographically weighted regression, separate β-coefficient values are calculated for each LOSA in the data set. For the j_{th} instance at a LOSA with the coordinates (u_j, v_j), the model is rewritten as:

$$Y_j = \beta_0(u_j, v_j) + \frac{\beta_1(u_j, v_j)}{X_j} + ... + \frac{\beta_i(u_j, v_j)}{X_j} + \varepsilon_j, \quad i \in N. \tag{2}$$

Compared with traditional linear regression algorithms, Artificial Neural Networks (ANNs) has been widely applied due to its advance in non-linear regression (Gemitzi *et al.* (2008)). As a result, a multi-layer feed forward ANN are used in this research. The function of the ANN is expressed as:

$$Y = \psi\left(\sum_{k=1}^{6} w_k \phi\left(\sum_{l=1}^{10} w_{kl} \phi\left(\sum_{m=1}^{7} w_{lm} X_m\right)\right)\right), \tag{3}$$

where ψ indicates the active function for the output neuron, which is set to be linear function, ϕ indicates the active function for the hidden neuron, which are set to be sigmoid functions, and w_{lk} indicates the wights between the k_{th} layer and the l_{th} layer.

4 Results

Results of three approaches exploring the association between the socioeconomic statues and morality rates of five age groups are displayed in the Table 3. Initially, from the comparison of R^2, the performances of three approaches, which predicts the mortality rate of

Variables	0-19 OLS 0-19 GWR	20-64 OLS 20-64 GWR	65-85 OLS 65-85 GWR	20-44 OLS 20-44 GWR	45-64 OLS 45-64 GWR
IMD 07	0.05 -1.11∼0.68	0.15 -2.57∼2.67	5.46 0.05∼8.96	0.11 -3.86∼1.25	0.47 -4.08∼6.72
IMD 07 1st Order	0.01 -0.02∼0.11	-0.13 -0.26∼0.06	-2.06 -6.10∼2.25	-0.05 -0.12∼0.75	-0.24 -0.49∼0.16
Income Index	6.54 -6.3∼17.00	-27.59 -46.15∼28.88	33.55 -40.90∼118.34	-6.01 -15.25∼26.32	-21.03 38.56∼101.66
Employment Index	-9.98 -21.03∼24.56	74.84 -16.11∼146.03	-1012 -1100.01∼690.22	24.01 0.12∼80.29	58.81 -99.63∼165.89
Education Index	-0.01 -0.08∼0.55	0.06 -0.18∼0.19	-1.37 -7.82∼2.11	-0.02 0.10∼0.07	-0.03 -0.29∼0.36
Health Index	0.11 -0.77∼1.86	2.58 -0.07∼9.85	91.79 71.52∼325.67	0.57 -0.98∼3.48	7.04 4.64∼35.87
Population Rate	-5.1 -8.42∼3.53	-33.34 -38.56∼7.90	347.68 8.04∼398.77	-6.91 -10.33∼5.86	-41.23 65.93∼12.86
Intercept	3.03 -0.27∼4.96	33.83 5.79∼29.29	340.58 299.03∼626.85	6.27 0.35∼8.47	38.65 20.41∼45.66
Diagnostics					
OLS R^2 GRW R^2	0.09 0.11	0.55 0.57	0.17 0.22	0.26 0.31	0.55 0.58
ANN R^2	0.17	0.63	0.28	0.35	0.64
OLS AICc	167402	212035	427059	179212	253387
GWR AICc	166013	200930	400480	178833	202908

Table 3: Model results and statistic analysis for OLS, GWR and ANN

Urban-Rural type	Number of LSOA
Urban $>$ 10K - Sparse	70
Town and Fringe - Sparse	152
Village, Hamlet & Isolated Dwellings - Sparse	227
Urban > 10K - Less Sparse	26385
Town and Fringe - Less Sparse	2929
Village, Hamlet & Isolated Dwellings - Less Sparse	2719

Table 4: Description of Urban-Rural types

20 to 64 year old age group and 45-65 year old age group, is above 0.5 and is better than the other groups, *i.e.*, over 50% of morality rate is associated with the explanatory variables.

From Table 2, residuals in each age group obtained by OLS regression exhibits significantly positive spatial autocorrelation, which means any inferences drawn from the global regression are questionable. However , by running GWR using the same variables, the GWR delivers an improvement in predicting mortality rates of five different age groups, which can be confirmed by obtaining an improved R^2 and the decline of AICc values. The model with the lower AICc value provides a better fit to the mortality data.

Compared with previous methods in each age group model, ANN is a non-parametric model to improve R^2 value by 5%. Evidently, ANN is a better predictive tool because of its accurate description of the functional relationship between variables, whether that relationship is linear or not.

In order to observe results intuitively, residuals of models are displayed in ArcGIS by using three methods. The Rural and Urban Area Classification data in Table 4 is joined to the attribute table to analyse the residuals. Figure 1, 2 and 3 highlight the "Urban$>$10K - Less Sparse" and "-3$<$Residual$<$3" for mortality rate of 20-64 age group in each LSOA by applying OLS, GWR and ANN. The probability distribution of errors in OLS methods, GWR methods and ANN methods for 20-64 age group are illustrated in Figure 4. Obviously, the small residuals for three methods cluster around the cites in England, such as London, Birmingham, Manchester, Liverpool, New Castle, *etc*. The ANN has a stable performance compared with other two methods.

5 Conclusion

This research attempts to explore the association between deprivation and mortality rate in England by applying three methods. Primarily, the seven variables are chosen to predict the all-cause mortality rate for all age population through three methods. Nevertheless,so far the results indicate a weak relationship between deprivation and all-cause mortality rate for all age groups of the population. Then the mortality rate is calculated dependent on different age groups. It can be found the deprivation do not strongly affect to mortality

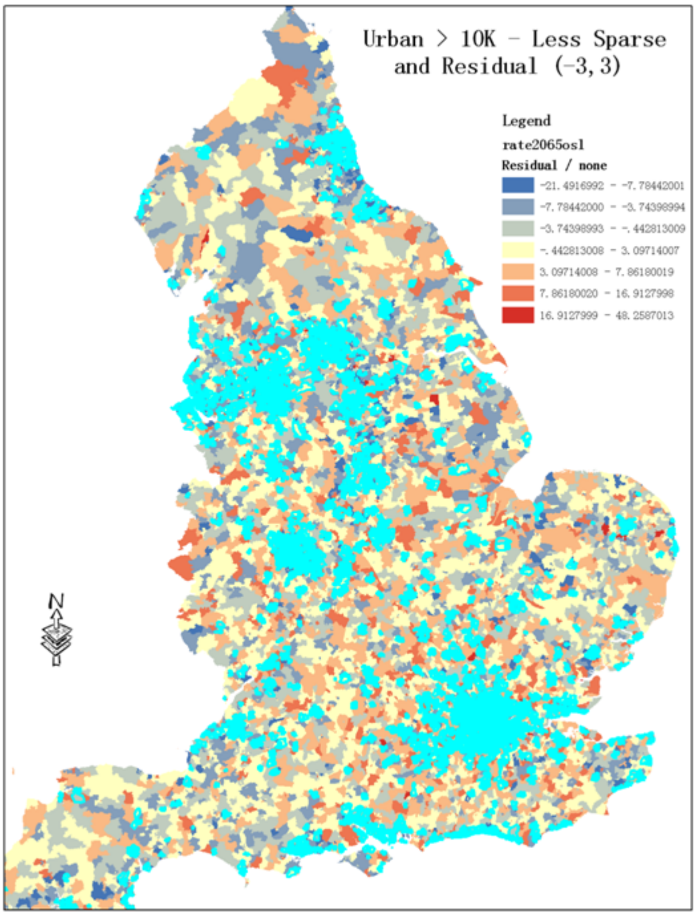

Figure 1: Highlighted urban>10k - less sparse and residual by using OLS

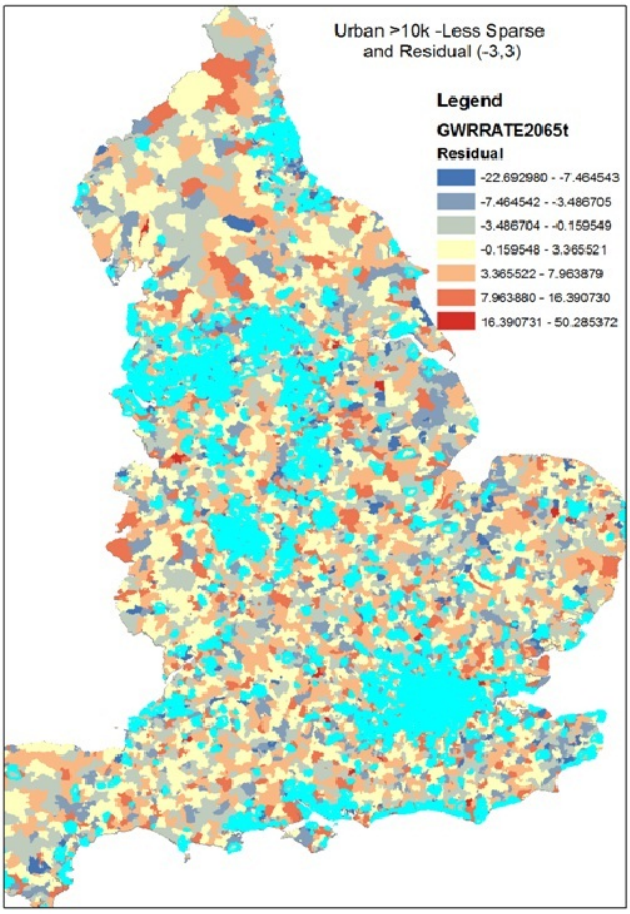

Figure 2: Highlighted urban>10k - less sparse and residual by using GWR

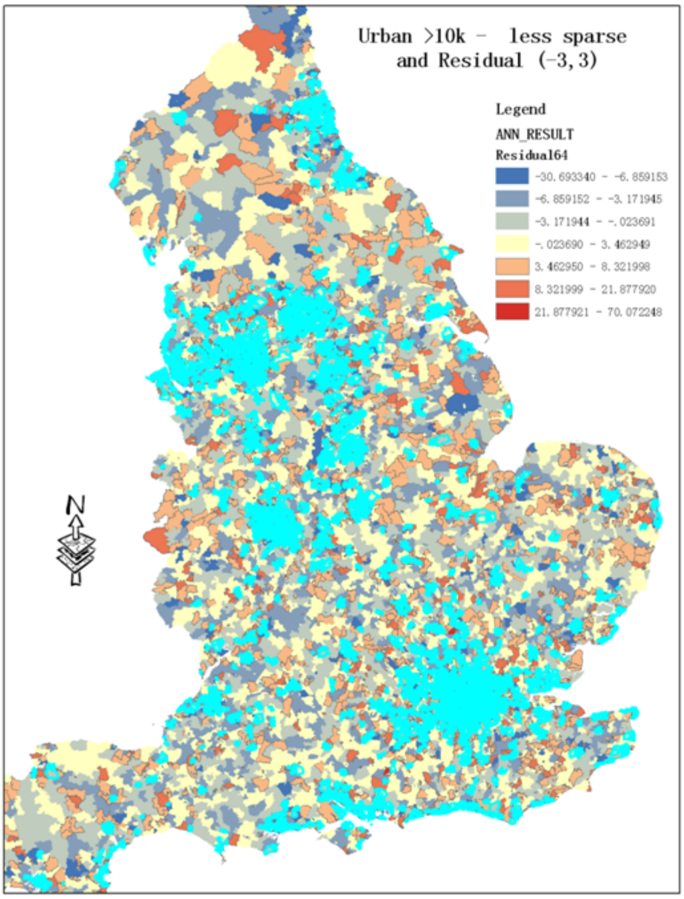

Figure 3: Highlighted urban>10k - less sparse and residual by using ANN

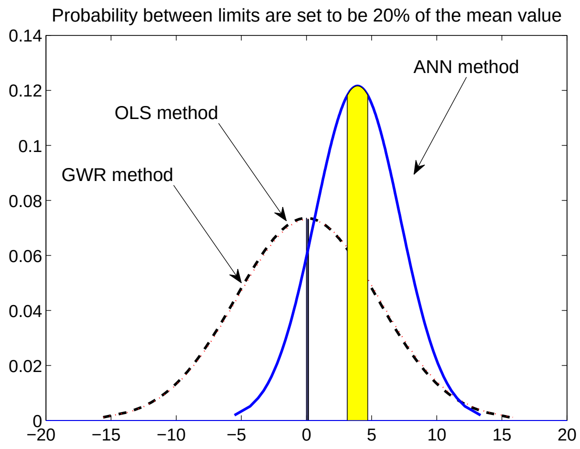

Figure 4: The probability distribution of errors in OLS methods, GWR methods and ANN methods for 20-64 age group

rates for young people and older people. Meanwhile, over 50% of mortality rate for middle aged people in LSOA could be interpreted to associate with socioeconomic status. The results also shows that ANN has the minimal error and a stable performance as compared with the other two methods.

6 Acknowledgements

This paper is associated with a PhD research thesis X. Zhang currently in progress titled deprivation and health inequalities.The research has been funded (2008-2011) by Centre for Public Health, Liverpool John Moores University. The thesis is supervised by Dr. Karen Tocque, Dr. Penny Cook and Dr. John Boothby at the Liverpool John Moores University.

References

Adler, N. (1994). Socioeconomic status and health. *American Psychologist*, **49**(1), 15–23.

Davey Smith, G. and Hart, C. (1998). Individual social class, area-based deprivation, cardiovascular disease risk factors, and mortality: the renfrew and paisley study. *Journal of Epidemiology and Community Health*, **52**, 399–405.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. E. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, Chichester.

Gemitzi, A., Petalas, C., Pisinaras, V., and Tsihrintzis, V. A. (2008). Spatial prediction of nitrate pollution in groundwaters using neural networks and gis: an application to south rhodope aquifer. *Hydrological Processes*, **23**(3), 372–383.

Huang, Y. and Leung, Y. (2002). Analysing regional industrialisation in jiangsu province using geographically weighted regression. *Journal of Geographical Systems*, **4**(2), 233–249.

Rojas, R. (1996). *The backpropagation algorithm of Neural Networks - A Systematic Introduction*. Springer-Verlag, Berlin, New-York.

Turrell, G. and Mathers, C. (2001). Socioeconomic inequalities in all-cause and specific-cause mortality in australia: 1985 - 1987 and 1995 - 1997. *International Journal of Epidemiology*, **30**, 231–239.