# Multiscale Geographically Weighted Discriminant Analysis (Short Paper)

**Alexis Comber**[1]
School of Geography, University of Leeds, UK
a.comber@leeds.ac.uk

**Nick Malleson**
School of Geography, University of Leeds, UK
n.s.malleson@leeds.ac.uk

**Hang Nguyen Thi Thuy**
VNU Vietnam Japan University, Hanoi, Vietnam
nguyen.t.thuyhang@gmail.com

**Thanh Bui Quang**
Faculty of Geography, VNU University of Science, Hanoi, Vietnam
qthanh.bui@gmail.com

**Minh Kieu**
Faculty of Engineering, University of Auckland, New Zealand
minh.kieu@auckland.ac.nz

**Phe Hoang Huu**
R&D Consultants, Hanoi City, Vietnam
hoanghuuphe@gmail.com

**Paul Harris**
Sustainable Agriculture Sciences,Rothamsted Research, UK
paul.harris@rothamsted.ac.uk

## Abstract

This paper describes the novel development and application of a multi-scale geographically weighted discriminant analysis (MSGWDA). This is applied to a case study of survey data of attitudes to a proposed motorbike / scooter ban in Han Noi, Vietnam. It uses discriminant analysis to examine attitudes to the ban in relation to travel purposes, distances, respondent age and so on. The main part of the paper focuses on describing the novel MSGWDA approach, and the results indicate the varying scales of relationship between the different input variables and the categorical responses variable. The paper also reflects on the pervasive logic of the approaches used to fit multiscale geographically weighted bandwidths (for example in regression). These have historically been based on the iterative back-fitting approaches used in GAMs, but risk missing potentially important variable interactions amongst un-evaluated bandwidths because of the sequence of their application. It is argued that although pragmatic in the 1990s, it may be possible to apply more deterministic approaches with increased memory and readily accessible computing power in order to better navigate such highly dimensional search spaces.

**2012 ACM Subject Classification** Computing methodologies → Model development and analysis

**Keywords and phrases** Multiscale; Ha Noi, Vietnam; LDA; Spatially varying coefficient models

**Digital Object Identifier** 10.4230/LIPIcs.CVIT.2016.23

**Funding** This work has received funding from the British Academy under the Urban Infrastructures of Well-Being programme [grant number UWB190190]

[1] Corresponding author


42nd Conference on Very Important Topics (CVIT 2016).
Editors: John Q. Open and Joan R. Access; Article No. 23; pp. 23:1–23:8
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

# 1 Introduction

This paper describes an approach for classification using a novel multi-scale geographically weight discriminant analysis (MS-GWDA). Discriminant analysis (DA) or discriminant function analysis [6] as generalised by Rao [15], is a commonly used technique for predicting membership or class for discrete groups as an alternative to multinomial logistic regression [11]. More recently DA has recently gained much attention in the context of machine learning [10, 12] and real time analyses [16, 4] because it can be used as an information learning technique as well as classification, particularly in the context of pattern recognition.

Conceptually, in DA the independent data used as input to discriminant analysis can be thought of having been drawn from different populations of each class [2]. The discriminant functions are extracted from independent variables and are then used to generate class membership probabilities for each observation. If there are $k$ groups, the aim is to extract $k$ discriminant functions. Each observation is assigned to the class $j$ for which value for the discriminant function for the group is the smallest, thereby determining which population each observation is likely to have come from. Under the assumption that the data are multivariate normal, then if $\sum_j$ is the variance-covariance matrix for the members of class $j$, $q$ is the number of predictor variables in $\mathbf{x}$, $\mu_j$ is the mean vector for the observations in class $j$, and $p_j$ is the prior membership probability of class $j$, the linear assignment (linear discriminant analysis, LDA) multivariate Gaussian decision rule can be written as:

$$k = \arg\max_{j\in(1,\ldots,m)} \left[ p_j \frac{1}{(2\pi|\Sigma_j|)^{q\backslash 2}} exp\left( -\frac{1}{2}(\mathbf{x}-\mu_j)' {\sum}_j^{-1} \mathbf{x}-\mu_j) \right) \right] \quad (1)$$

LDA was extended from the linear to the quadratic case by Marks and Dunn [13]. They and Wahl and Kronmal [17] examined the behaviour of linear and quadratic discriminant function using data with unequal covariances. In cases where the samples are small, the linear DA function has been found to provide more reliable assignment, whereas for large samples, the quadratic DA function may be preferred.

DA was further extended to the spatial case by [2] who proposed a geography weighted DA (GWDA). In a GWDA, the idea is that the decision to allocate an observation to group or population is made whilst taking the geographic location of the observation into account. Whereas a standard DA (LDA and QDA) uses the mean vector and covariance matrix, a GWDA uses geographically weighted means and covariances as described in Brunsdon et al [1] and Fotheringham et al [7]. It uses the same geographical weighted framework as GWR, in which a series of local models are constructed at locations through the study area rather than one global model. In these, observations falling under a moving window or kernel are weighted by their distance to the observation location and used to construct the local model. The key challenge in any geographically weighted model is the determination of the size of the kernel or bandwidth. A standard GWR identifies a single optimal bandwidth or spatial scale over which the process under investigation is assumed to vary. This is usually done through some evaluation of different bandwidths using cross validation or a model parsimony measure such a s AIC.

However, thinking around geographically weighted frameworks and specifically geographically weighted regression (GWR) has matured considerably in recent years. This has driven a number of developments and extensions to the original GWR. Not least of these is multiscale GWR (MSGWR), which seeks to identify variable specific bandwidths rather than using a single best on average bandwidth to construct local models. The idea is that individual response-to-predictor relationships may operate over different spatial scales and the use of a single bandwidth in a standard GWR may under- or over-estimate those. As a result, the recommended approach for any GWR is to first undertake a MSGWR to determine these variable specific scales, in order to guide further model choice [5]. This, as is nicely summed up by Oshan et al [14], eliminates the need for predictor-response relationships to vary at the same spatial scale, can reduce over-fitting and bias in the parameter estimates, and can overcome local collinearity. Thus, MGWR has been suggested as the default GWR approach [5]. Such thinking and logic has potential relevance for all geographically weighted frameworks, including geographically weighted discriminate analysis, hence the method proposed in this paper

## 2 Multiscale Geographically Weighted Discriminant Analysis

A GWDA extends QDA and LDA into geographically weighted methods. Here the decision population probabilities now depend on the spatial location of the observation – ie the the variance-covariance matrix $\sum_j$, the prior membership probabilities of class $j$, $p_j$ or the $\mu_j$ the mean vector for the observations in class $j$, are assumed to vary with spatial location $\mathbf{u}$. Thus, the probabilities used to derive the decision rules are conditional on $\mathbf{u}$, reworking Equation 1 above to estimate the local likelihoods for the population models:

$$f_p(\mathbf{x}|\mathbf{u}) = \frac{1}{(2\pi|\Sigma_j(\mathbf{u})|)^{q/2}} exp\left(-\frac{1}{2}(\mathbf{x}-\mu_j(\mathbf{u}))'\sum\nolimits_j^{-1}(\mathbf{u})(\mathbf{x}-\mu_j(\mathbf{u}))\right) \tag{2}$$

MSGWDA extends GWDA but allows the bandwidths for each input parameter to vary individually. The key objective in all multiscale geographically weighted models is to determine the matrix of parameter specific weights. These in this case will be used to weight each input variable at location $\mathbf{u}$, as defined by the kernel bandwidth. As an example, Figure 1 shows some of the data from the case study described below, with the shading indicating the different bandwidths and potential scales of relationship between the grouping / classification and different variables, at the location being considered.

## 3 Case study: Travel Survey in Ha Noi

Ha Noi like many major cities in emerging economies, suffers serious traffic congestion and air pollution due to rapid urbanization rates, increases in private transport. In Ha Noi, motorbikes are the preferred transportation mode: almost everyone in the city owns a motorbike. In 2015, Ha Noi had 4.9 million motorbikes and 0.5 million cars and 11 million motorbikes are projected by 2025. As public transport does not meet the city's requirements, increases in personal traffic are inevitable, resulting in acute welfare problems, especially air quality. Pollution is chronic, with PM2.5 and ozone concentration regularly exceeding safe.

As a result the government in Vietnam is exploring the possibility of implementing a ban that will stop motorised scooters from entering the city centre of Ha Noi. A survey has been undertaken to capture data on the travel behaviour and preferences of residents in Ha Noi. This includes information on trips made (origins and destinations), their frequency, the transport modes used for different trips, the reasons for those modal choices as well as demographic information about respondents' home locations, and critically respondent attitudes concerning the proposed travel ban and how they would respond to it. This paper explores the survey responses regarding the ban and the factors associated with those responses in order to illustrate a multiscale geographically weighted discriminant analysis.

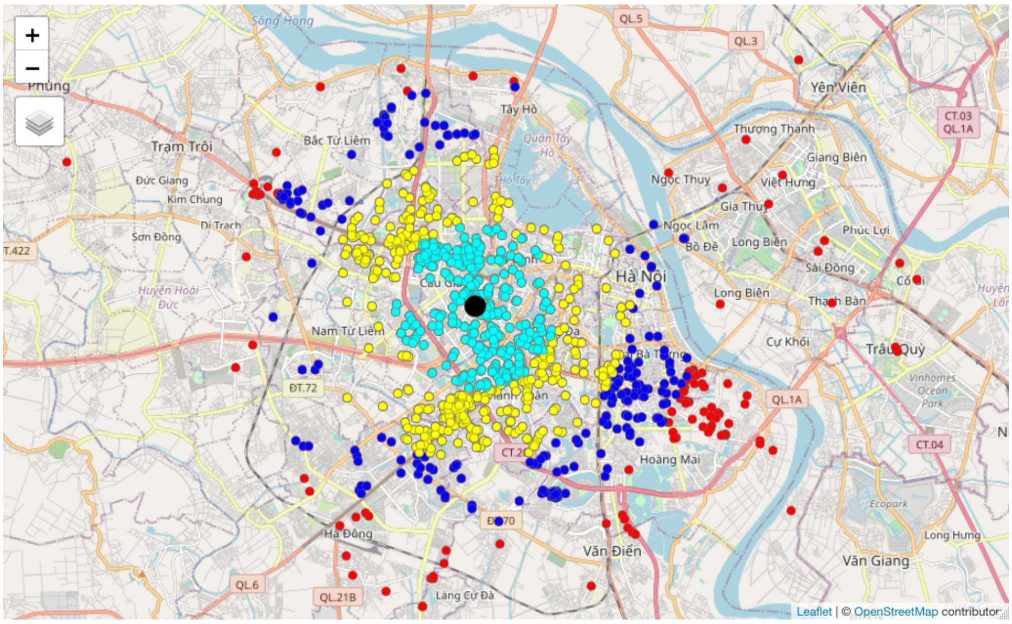


**Figure 1** An illustration of the different adaptive bandwidths, shaded in cyan (30%), yellow (70%), blue (90%) and red (100%), for 4 different variables, for a location marked in black, with an OpenStreetMap backdrop.

Data from some 1191 respondents was obtained and a subset of their responses was used in the analyses as described below. Specifically the aim was to examine create a MSGWDA of attributed to the proposed van on motorised scooters from categorical variables describing:

- respondent age group;
- respondent gender;
- the purpose of the main regular journey they make;
- the network distance of that journey, as derived from a shortest path analysis of OSM route data with snap distances.

The resident locations are shown in Figure 2.

To demonstrate the MSGWDA approach combinations of adaptive bandwidth sizes for each variable were defined as sequences running from 20% to 100% in steps of 10%. For 4 variables, this resulted in $9^4$ bandwidth combinations to evaluate. Each combination of variable specific bandwidths was used to wight inputs into a linear discriminant analysis function (`lda` part of the `MASS` R package). For simplicity a boxcar weighting was used, generating weights of 1 for observations underneath the kernel and 0 for those outside. The were used to create a locally weighted LDA at each observation location in the study area, and each local, under each combination of bandwidths model was used to predict make a local classification prediction. The whole map set of classification predictions were then evaluated and metrics describing classification reliability were extracted (including overall accuracy, Kappa etc). The best performing combinations of bandwidths was then identified.

## 4 Results

Two results are used to illustrate the potential inferential advantages of the MSGWDA: an ordinary global LDA and a novel multiscale GWDA. A geographically weighted LDA was not

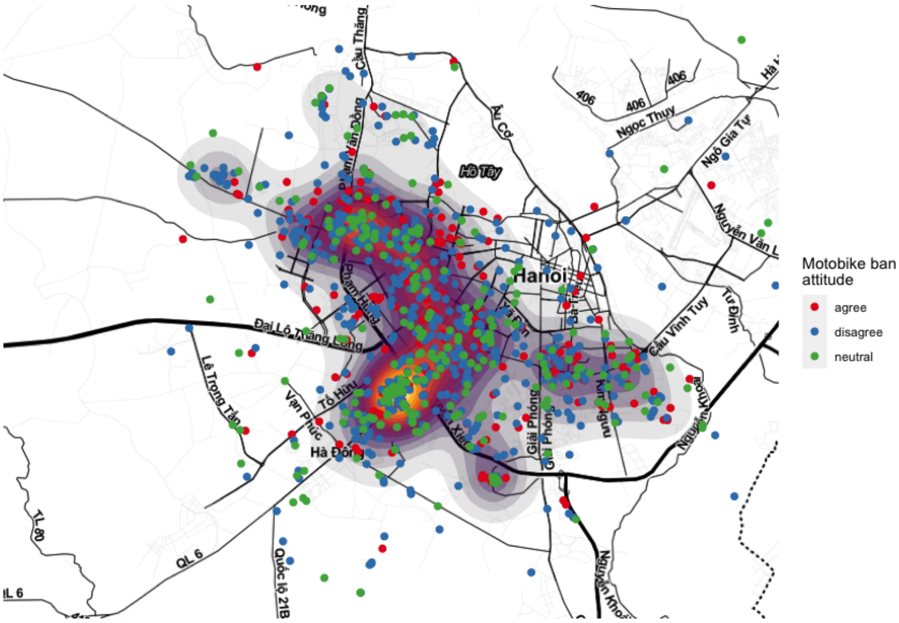


**Figure 2** The motorbike ban attitudes of the survey respondents, with a heatmap showing the density of respondents, and a Stamen toner backdrop.

undertaken because of the inability of the `gwda` function in the `GWmodel` to handle categorical predictor variables.

The standard LDA model is relatively weak, with an overall accuracy of 0.548 and a Kappa statistic of 0.115. The correspondence table is shown in Table 1 and indicates high specificity (ie good a true negatives) and low sensitivity (ie poor at true positives).

| | **Observed** | | |
|---|---|---|---|
| **Predicted** | agree | disagree | neutral |
| agree | 42 | 24 | 24 |
| disagree | 222 | 585 | 238 |
| neutral | 14 | 16 | 26 |

**Table 1** The correspondence matrix of the LDA classification of survey responses regarding a proposed motobike ban in the city centre.

The MSGWDA examined combinations of adaptive bandwidths for each variable. For each of these, a geographically weighted LDA model was created at each of the 1191 respondent home locations. At each location weighted LDA model was used to predict the motorbike ban attitude, such that a vector of 1191 predicted ban attitudes were created from 1191 local models. For each set of predictions, a correspondence matrix of predicted against observed ban attitudes was created an evaluated using overall accuracy and Kappa statistics. The best performing combinations were found to be the following sets of bandwidths when evaluated using Overall accuracy and Kappa statistics:

- Overall accuracy: gender 80%, trip purpose 50%, age 40% and network distance 10%.
- Kappa statistic: gender 40%, trip purpose 20%, age 20% and network distance 10%.

These are illustrated in Figure 3 for the same example location as in Figure 2. Here we can see the different bandwidths indicated by different fit or accuracy measures. The

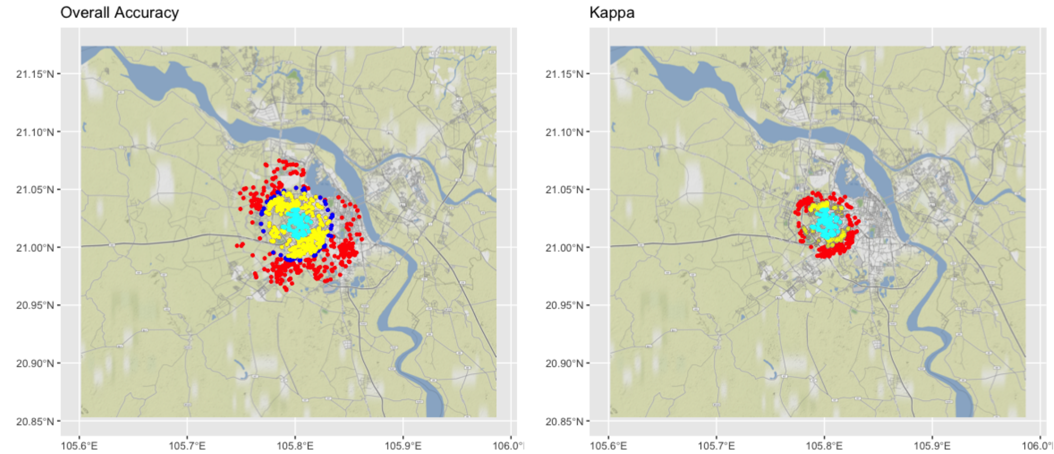


**Figure 3** The best fitting multiscale bandwidths when evaluated using Overall accuracies and Kappa statistics.

correspondences are summarised in Table 2 and result in Overall accuracies and Kappa statistics of 0.579, 0.199 and 0.575, 0.207, respectively.

| | **Overall** | | | **Kappa** | | |
|---|---|---|---|---|---|---|
| **Predicted** | agree | disagree | neutral | agree | disagree | neutral |
| agree | 59 | 27 | 22 | 65 | 37 | 30 |
| disagree | 199 | 573 | 209 | 191 | 556 | 194 |
| neutral | 20 | 25 | 57 | 22 | 32 | 64 |

**Table 2** The correspondence matrices of the MSDWDA classifications of survey responses, when evaluated using Overall accuracy and Kappa statistics.

## 5 Discussion

The MSGWDA approach improves the classification accuracy compared to a standard global LDA. This is to be expected and is a feature of all geographically weighted models. Of more potentially interest and relevance to this specific case study, are the variations in the spatial scale at which categorical data are associated with the outcome: whether evaluated by Kappa or Overall accuracy, the gender variable tends towards the global, with trip purpose, age and distance highly localised in their effect. This understanding of scale will inform future project work in relation to the transport and behaviour simulation models being developed within this project. The MSGWDA results are key to that understanding.

There are a number of areas of further work required to exploit the functionality of this MSGWDA apprach.

1. The bandwidths in this poof of concept were at relatively coarse intervals, evaluating different adaptive bandwidth sizes in steps of 10%. This needs to be refined and potentially, potentially using a refined search heuristic and consideration of bandwidth interactions – see discussion point below.
2. Bandwidths will be extended to the fixed distance case.

**3.** The weighting of categorical variables was done using a box car approach, with the response variables in observations falling under the kernel simply weighted to 1. This needs to be explored in more depth, especially if the data variables are ordered in some way.

Whilst the results of the MSGWDA are interesting and potential refinement on both standard approaches top DA and GWDA in terms of inference and understanding, perhaps the major discussion point to arise from this work has been due to the need to unpick the mechanisms of multiscale geographically weighted models. The key question arising from the back-fitting methods they employ is this:

> How confident can we be that that potentially important variable interactions are not being missed by this *fix the first variable bandwidth, then fix the second, then the next, etc, etc ...* approach, rather than looking at all possible combinations of bandwidths?

The answer to this is uncertain: the multivariate bandwidth search space to determine the optimal set of weights to be passed to the local model at location on $\mathbf{u}$ is potentially huge. In the past, pragmatic short-cuts were needed to be able to move through it. But times and computing power have both changed.

The original MSGWR [18, 8] and subsequent refinements in other packages use a back-fitting algorithm to determine optimal parameter specific bandwidths and thus weights at each location, for each variable. This was based on the approach taken in generalized additive models (GAMs) [9, 3]. Essentially what these do is to determine the optimal set of bandwidths is to determine the bandwidth for each variable sequentially, using smoothing functions that assume the other terms are known. We suspect that this approach was developed by the GAM team as a pragmatic way overcoming the difficulty in searching through a highly dimension solution space comprised of all possible bandwidths for all possible variables. It was then adopted by the initial work into multiscale geographically weighted regression. The reason is the high dimensionality of the solution search space: given explanatory 5 variables, in a dataset with 2000 observations and an adaptive bandwidth approach (ie based on the number of observations to include rather than a fixed distance), this would potentially require $2000^6 = 6.4 \times 10^{19}$ solutions to be evaluated for a regression (including the intercept) and $2000^5 = 3.2 \times 10^{16}$ for a discriminant analysis

In this research, with potentially greater computing power available than at the time the GAM and MSGWR approaches were being developed, a grid of possible combinations of parameter specific bandwidths was used to demonstrate the MSGWDA, as described. Here the aim was to demonstrate how MSGWDA could be undertaken and to explore some of the issues arising from this. Also, this is philosophically preferable: the specification of musicale bandwidths a parameter at a time potentially ignores variable interactions at scales not considered in previously fixed bandwidths.

Future work will definitely explore this in greater detail!

## References


1 Chris Brunsdon, AS Fotheringham, and Martin Charlton. Geographically weighted summary statistics—a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems*, 26(6):501–524, 2002.

2 Chris Brunsdon, Stewart Fotheringham, and Martin Charlton. Geographically weighted discriminant analysis. *Geographical Analysis*, 39(4):376–396, 2007.

3 Andreas Buja, Trevor Hastie, and Robert Tibshirani. Linear smoothers and additive models. *The Annals of Statistics*, pages 453–510, 1989.

4 Dong-Wei Chen, Rui Miao, Wei-Qi Yang, Yong Liang, Hao-Heng Chen, Lan Huang, Chun-Jian Deng, and Na Han. A feature extraction method based on differential entropy and linear discriminant analysis for emotion recognition. *Sensors*, 19(7):1631, 2019.

5 Alexis Comber, Chris Brunsdon, Martin Charlton, Guanpeng Dong, Rich Harris, Binbin Lu, Yihe Lü, Daisuke Murakami, Tomoki Nakaya, Yunqiang Wang, et al. The gwr route map: a guide to the informed application of geographically weighted regression. *arXiv preprint arXiv:2004.06070*, 2020.

6 Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

7 A Stewart Fotheringham, Chris Brunsdon, and Martin Charlton. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, 2003.

8 A Stewart Fotheringham, Wenbai Yang, and Wei Kang. Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers*, 107(6):1247–1265, 2017.

9 Trevor Hastie and Robert Tibshirani. Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398):371–386, 1987.

10 Tyler L Hayes and Christopher Kanan. Lifelong machine learning with deep streaming linear discriminant analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 220–221, 2020.

11 David W Hosmer and Stanley Lemeshow. Confidence interval estimates of an index of quality performance based on logistic regression models. *Statistics in medicine*, 14(19):2161–2172, 1995.

12 Seung-Pyo Jun, Hyoung Sun Yoo, and Jeena Hwang. A hybrid recommendation model for successful r&d collaboration: Mixing machine learning and discriminant analysis. *Technological Forecasting and Social Change*, 170:120871, 2021.

13 Sidney Marks and Olive Jean Dunn. Discriminant functions when covariance matrices are unequal. *Journal of the American Statistical Association*, 69(346):555–559, 1974.

14 Taylor M Oshan, Ziqi Li, Wei Kang, Levi J Wolf, and A Stewart Fotheringham. mgwr: A python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information*, 8(6):269, 2019.

15 C Radhakrishna Rao. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society: Series B (Methodological)*, 10(2):159–193, 1948.

16 Daniel Szostak, Krzysztof Walkowiak, and Adam Włodarczyk. Short-term traffic forecasting in optical network using linear discriminant analysis machine learning classifier. In *2020 22nd International Conference on Transparent Optical Networks (ICTON)*, pages 1–4. IEEE, 2020.

17 Patricia W Wahl and Richard A Kronmal. Discriminant functions when covariances are unequal and sample sizes are moderate. *Biometrics*, pages 479–484, 1977.

18 Wenbai Yang. *An extension of geographically weighted regression with flexible bandwidths*. PhD thesis, University of St Andrews, 2014.