Measuring and understanding the differences between urban and rural areas, a new approach for planners.

Malcom J Beynon, Cardiff Business School, Colum Drive, Cardiff, CF10 3EU, UK.
Tel 00 44 2920 875747; email beynonmj@cf.ac.uk

Andrew Crawley, Cardiff Business School, Colum Drive, Cardiff, CF10 3EU, UK.
Tel 00 44 2920 875079; email crawleyaj@cf.ac.uk

Max Munday, Cardiff Business School, Colum Drive, Cardiff, CF10 3EU, UK.
Tel 00 44 2920 875089; email mundaymc@cf.ac.uk

Abstract

Understanding the factors that makes a location more rural or urban is an important task for planners and policymakers. Traditional individual characteristics of rurality sometimes hide the more complex social, as well as physical dynamics of a locality. In this context, the paper builds on early work such as Cloke (1977), which applied factor analysis to construct a single index of rurality. This approach is developed with a combined metric encompassing multiple measures. These are, capable individually of defining rurality but together they deliver greater insight on more complex patterns and help redefine the simple notion of rurality. The paper then utilises a novel graphical method, the constellation graph, providing a diagnostic and visual framework to aid planners when assessing the spatial dimensions of a locality.

Measuring and understanding the differences between urban and rural areas, a new approach for planners.

1. Introduction

The Oxford English Dictionary defines rural as relating to, or characteristic of, the countryside rather than the town (i.e. following the Latin *ruralis*). The definition of rurality has long been in dispute and comprises an elusive concept (Halfacree, 1993). For example, Weisheit *et al.* (1999) state:

> "*Like concepts such as "truth," "beauty," or "justice," everyone knows the term rural, but no one can define the term very precisely*." (p. 213)

However, defining rurality, and contrasting the rural with the urban is of practical importance. Isserman (2005) shows that rural research and rural policy are often based on ill-defined distinctions between the rural and the urban, and criticises the common use of the metro/non-metro distinction as a proxy for—or even worse—as synonymous with an rural/urban distinction. Gallant and Robinson (2011), Agarwal *et al.* (2009), and Argent (2008), all underline the need to improve the definition of rurality for policy targeting and developmental purposes. Waldorf (1996) criticises the arbitrary nature of traditional rural definitions showing that approaches based on a single scale (index), which is delivering a point estimate value, offers little insight into the true nature of rurality. The work of Waldorf (1996) and Hasse and Totzer (2012), emphasises the continuous and multi-dimensional nature of the rural concept, whereby, different sets of criterion can change the designation of an area from rural to urban (or vice versa). Indeed, Champion and Watkins (1991) argue that a single scale does not always take into account social and economic differences between areas.

In light of the evolving definition of rurality, planners have begun to study the interface between town and country with greater intensity, referred to in the literature as the 'rural-urban fringe' the objective is the development of improved spatial policy (see for example Gallent *et al.* 2006). Given the challenges of identifying something that is dynamic, new methods are required that embrace the complex nature of rurality. This present study aims to do just that, multiple indices are developed, and then a framework designed so that it is possible to understand how different index dimensions affect the overall designation of a region being rural or urban. With a few exceptions (see for example, Harrington and O'Donoghue, 1998), there have been limited attempts to move beyond single index approaches to understanding rurality.

This paper expands the single index approach, as in Cloke (1977), to understanding rurality by combining multiple measures capable individually of defining rurality, but together offering a powerful diagnostic tool for assessing the spatial dimensions associated with rural locations. This study adheres to the approach underpinning Cloke (1977), applying factor analysis to construct indices of rurality. Rather than using a fixed one-factor solution (model), n-factors are identified, based on whether identified index components have eigenvalues greater than one (Hair *et al.*, 2010). This particular use of factor analysis permits a greater amount of information (variance) to be retained from the considered variables in the identified factors. The paper extends the described intermediate multiple index approach, by considering the information using the constellation graph method of data represention (following Wakimoto and Taguri, 1978). This representation provides a unique graphical depiction of the contribution of individual variables in the construction of an index value (or factors from a one-factor or n-factors solution), as well as the ability to view the relative levels of rurality associated with defined areas. The constellation graph method also provides a means of assessing the relative consistency of the variables' information present in the constructed intermediate and single indices. For illustrative purposes we use the 22 local authority areas of Wales as a case in this paper.

The next section examines literature on both rurality and the measurement of rurality. The third section focuses on the method developed by Cloke (1977) and the technical specifications required for the evaluation and representation of intermediate multiple indices using constellation graphs. Comparisons are made between the index values found using the intermediate indices (constellation graph) approach and those from following the approach in Cloke (1977). The results are appraised both in terms of method and implication of using the overall technique for this form of spatial analysis.

2. Exploring rurality indices

One premise of the study of rurality is the assumption that rural areas retain distinctive features, making them desirable for distinct socio-economic analysis (Champion and Watkins, 1991). During the 1970's, there was particularly strong interest in approaches to measuring and classifying rurality, with elements of the research seeking to construct rurality indices based on statistical indicators (see for example in the US, Smith *et al*., 1973).

Table 1 about here

An important step forward was taken by Cloke (1977). This index combined 16 variables (see Table 1). The variables used focused on population density and demography, access to amenities and remoteness from urban centres/commuting distances. In the Cloke index, household amenities were included based on the presumption that those living in rural areas will have less amenities than those in urban areas. The original 'Cloke' index was updated in 1981 by Cloke and Edwards (1986). The later paper demonstrated that replication of the index was possible but given boundary changes and other differences in data showed significant changes in rurality across the UK. The authors proceeded to rerun the analysis constructing a second index incorporating new variables that were thought to give a more contemporary view of rurality. With new data the principle component analysis was rerun reworking the loading scores. This demonstrates that when using the form of analytics it becomes necessary to continually update the method to take account of influence of new data.
Supsiquently the index was updated by Harrington and O'Donoghue (1998) and Cloke and Johnston (2005). These later revisions again included additional variables describing mobility, and numbers of second and holiday homes.

The benefit of the individual indices developed above was the potential to compare one region/county with another in terms of levels of rurality. Another index of rurality was produced by Clevland (1995), but the focus was on mobility, income and employment structure sharing some of the characteristics of the Smith *et al* index (see above and also Table 1). Clevland also included education as a variable. This index has been taken up by other researchers and developed for various needs, most notably by Edmondson and Fontanez (1995), who included a "connected-ness" component. This latter index included measures of economic health, changes in poverty over time and the number of newspapers per county, to proxy for local communication networks. These indices have paved the way for ever more complex approaches to rurality analysis, which over the last 10 years has included work such as Muilu and Rusaneu (2004) that utilise GIS based approaches.

The described rurality indices have some similar themes including population density, variables examining remoteness, economic structure and activity, and variables describing income differentials. Moreover, the construction of each of these indices follows a very similar pattern, with the primary means of development being through the factor analysis method. The approach means that all variables are used initially to determine what degree each contributes to explaining rurality. This approach does not utilise any statistical means to explain more of the variation that is occurring in the data. A related problem is the absence of a correction mechanism. For example, how does one treat with an area that might have low population density but other attributes that firmly place the area in an urban bracket? Notwithstanding these issues, this form of single index, as defined in Cloke (1977), has been highly cited and updated numerous times by a large body of researchers (see for example, Bannister, 1980; Best, 1981; Pacione, 1984; Harrington and O'Donoghue, 1991).

A challenge, deriving from the above review, is to develop the work of Cloke (1977) to construct an index that can overcome the potential of arriving at a single value from the numerous variables that make up a rurality index (Hoggart, 1988), as well as introducing a novel graphical depiction of the contribution of potential intermediate indices (individual factors) to rurality. The indices reviewed above combine a collection of different variables to best describe rurality; indeed these collectively might be considered almost a proxy for rurality. What might be considered is not just the specific variables to capture rurality but a clearer definition of the dimensions of rurality. Dimension and variable may have subtle differences in meaning but this is an important distinction in terms of trying to define something. The dimensions of rurality are the forces that generate it, or that cause it to exist.

Coombes and Raybould (2001) acknowledge the complex patterns that exist in contemporary human geography and note the lack of one single variable that can 'capture' urban/rural settlements. However, the authors do note that three key dimensions are present within modern human settlements, which can be captured and used as proxies for identifying rural areas in index measurements. These are settlement size, concentration or population density, and accessibility or degree of openness. These are similar to the factors in indices constructed previously, such as Clevland (1995), but Coombes and Raybould (2001) make a clear conceptual distinction between these dimensions. This means, for the first time, an area can have characteristics akin to both urban and rural localities, unimpeded by the aggregation of variables into a single metric (index). This is an important step forward, and in part, an acknowledgement of the changing nature of rurality. Hugo *et al*. (2001) agree that the nature of rurality has changed over the last 20 years, due in no little part to improved transport and advances in communications technology.

Further work by. Hugo et al (2003) notes how demographic analysis has historically had a simple typology of what is rural what is urban to differentiate settlements, but as time has passed the meaning of these classifications has become too narrow. The work concludes that there is a need to introduce an " intermediate, or transitional category of space, recognizing a more graduated set of situations between the most urban and the most rural locations"(p.278). This has led a number of authors to purpose alternative ways of considering rurality Pateman (2011) uses 7 classifications of rurality. Whereas work by (Scott, Gilbert and Gelan, 2007) suggest there are as many as 30 stratifications of rurality in the UK. This plethora of contributions in it self presents a problem, by acknowledging the complex nature of rurality there is now a need to understand in greater depth what characteristics are forming these stratifications.

3. Revisiting Cloke (1977)

The intended approach described follows the work in Cloke (1977)[1]. For illustrative purposes we use the 22 local authorities of Wales as a case.[2] The choice of this spatial scale was driven partly as a means to replicate the original work but also by data availability. Since the initial study of Cloke (1977) much more refined spatial analysis has been possible both in terms of smaller and larger geographical areas. Research using Super Output Areas (Smaller) has allowed greater detail to be captured in spatial analysis (See for example the work of Curl et al

[1] The original approach of Cloke (1977) was chosen rather than later editions to preserve the initial spirit of the work with greater concentration on preserve multidimensional information rather than individual variables.

[2] Local authorities in Wales are a single tier of local government, they administer all local functions such as waste collection. Authorities in Wales are only unitary in nature (they are described by the Local Government (Wales) Act 1994).

2015). Equally Labour Market Areas (Larger) make use of more functional demographics particularly important in economic studies (See for example Boschma et al 2014). This present work acknowledges the local authority as an arbitrary geography however the authors consider it to be an illustrative example of the technique. If there exists greater data refinement researchers may modify the spatial scale. The variables considered in this illustrative investigation (see Table 2) are a sample of those used in Cloke (1977) in his analysis of England and Wales Rural Districts.

Cloke (1977) used Census data from the 1960's and 70's and the method involved some cross-table analysis combining variables from different sources. This was supplemented with local area data from the Central Statistics Office (CSO). Replicating this exactly for 2010 (our illustrative year) is difficult given the last available Census records for the UK were for 2001 at the time of writing. It was also found that the UK government no longer collects data on some of the household amenities characteristics as originally defined. To this end, and in this case, an attempt is made to reconstruct the data using ONS and Welsh Government data. The authors do not feel that having an exact match of variables is essential for the present exercise. The variables chosen for 2010 represents a strong match for the characteristics Cloke was trying to capture in his original study. Indeed in replication of the index in (Cloke and Edwards, 1986) the research was also forced to change the variables to match the data available at the time. The goal of this paper is not to critique the choice of variables in constructing a rurality index, but to merely show how multiple indicators maybe utilised in a more effective manner. For two variables, those describing the working population not in agriculture (Non-Working Agriculture %) and working populations not commuting (Non Commute %), these are reverse coded versions of variables originally considered (see descriptions in Table 1 and 2). The reason they were reverse coded is based on the factor analysis next undertaken (original variables had negative loadings so were reverse coded – see Hair *et al.*, 2010). For simplicity, the reverse coded variables are used as the variables in the analysis here.

Table 2 and 3 about here

The information contained in the variables reported in Table 2 is analysed using factor analysis (Hair *et al.*, 2010). Factor analysis examines the patterns of complex multi-dimensional data to determine whether the information can be condensed or summarized in a smaller set of factors (or components). Here, two factor analyses are performed. The first of these is based on identifying factors which have associated eigenvalues greater than one (inferring a factor extracts at least as much variance as the equivalent of one of the original variables - Kaiser, 1960), termed an *n*-factor model (the *n* dependent on how many factors are identified). The second factor analysis is similar to that performed in Cloke (1977), this is termed a one-factor model. The results of these two factor analyses follow, with their details intended for integration into the constellation graph approach (see later).

In Table 3, the factor analyses results are shown for the 10 variables described in Table 2. Based on the principle of identifying components (factors) with eigenvalues greater than one (see discussion earlier), three factors are identified, so termed the three-factor model. Collectively, nearly 80.81% of the variance[3] in the underlying data is contained in the variables (there is a noticeable difference between the eigenvalues of the third (1.50) and fourth (0.78) identified components, the divide between those components used as a factor and those not –

[3] The '% of variance term' relates to what percentage of the variance in the considered 10 variables is explained by the respective number of factors (see Hair *et al.*, 2010).

see Table 3). Once rotated using Varimax with Kaiser Normalization, the '% of Variance' contribution of the three identified components are 34.49%, 25.58% and 20.75% (for technical elucidation see Hair *et al*., 2010). For the one-factor model, the results are also included in Table 3. That is, the first factor becomes the only factor for the one-factor model considered here, with an identified '% of Variance' of 40.40% (see Initial Eigenvalues column – not required to consider rotation on a one-factor model). Based on the '% of Variance' values between the two models, there is twice the amount of information from the 10 variables contained in the three-factor model (80.81% of variance) than in the one-factor model (40.40% of variance).

Following on from the identification of factors using factor analysis, the resulting loadings of the 10 variables, for the three-factor and one-factor models, are presented in Table 4. These loadings estimate the level of contribution of a variable to a factor.

Table 4 about here

The role of the loadings, as presented in Table 4, is to construct factor scores, values representing the factors for each local authority (enabling a form of data reduction). It is a matter for debate on how the loadings should be used to enable factor scores to be evaluated. That is, there are a number of approaches to constructing factor scores for the local authorities (in this case). Hair *et al*. (2010) elucidate this problem, highlighting a number of ways to achieve these factor scores, namely (in brief terms), identifying a single variable (value) to represent each factor, aggregating the values of variables most associated with each factor (averaged or weighted by loadings values), and 'loadings' weighted aggregation of values of all variables associated with each factor. There are advantages and disadvantages to the use of each of these approaches (Hair *et al*., 2010), and all of them have been employed in factor analysis (see for example, Duenckmann, 2010; Barbieria and Mahoney, 2009; Jauhiainen, 2009).

For the three-factor model (initially), each variable is loaded onto the factor it was most associated with (based on largest loading value), and weighted by the loading value (identified in bold face in Table 4).[4] For the one-factor model, the 'loadings' weighted aggregation of values of all variables associated with the one factor was employed (loading values shown in bold in last column in Table 4).

As referred to earlier in the text, the three factors identified in the three-factor model form the intermediate multiple indices of rurality (these would not have been considered in the one factor model approach), each of which offers a dimension on the notion of rurality.[5] In the three-factor model, based on the partition of the variables across the identified three factors, terms used to describe the three factors are next expressed (refer to Table 4):

- **Population and Housing Dynamics (Factor 1):** This factor is a combination of traditional population metrics, such as population density and work activity ratios i.e. Male and Females of Working Age as a % of the total population. This is coupled with the variable of dwelling stock.

[4] The authors acknowledge this approach does not use all the loadings based information able to be used – hence it could be argued a level of information loss has taken place. The authors stress this approach adopted is without loss of generality to the use of other factor score evaluation approaches that exist.

[5] The use of the term intermediate here is since they themselves can be aggregated to form a single factor (undertaken later).

- **Migratory Dynamics (Factor 2):** This factor is a combination of changes in population movement made up of In Migration %, Out Migration % and Balance Migration %.
- **Social Dynamics (Factor 3):** This factor is a combination of two variables both reflecting socio economic trends, Population Change 2000-10 and Non Commuting Population.

The next part of the paper works with the three-factor model to further elucidate the collective information in these intermediate rurality indices. The constellation graph method is adopted here as an analytical tool. Constellation graphs were introduced in Wakimoto and Taguri (1978) and are a means of obtaining a 2D representation of multi-dimensional data (see Mitzuta, 1994, Sekiya *et al.*, 1991, for examples of its application). Here, they are employed to position local authorities in a domain encompassing the limits of rurality, namely rural and urban. In the constellation graph method, multi-dimensional data are represented as connected (elementary) vectors, one for each local authority, in a semicircle with a radius of unity.

For the i^{th} local authority, each of the original variable values describing it, $v_{i,k}$ $k = 1, \ldots, K$, in terms of those forming an individual factor, is transformed by a real valued function $f_k(\cdot)$ given by:

$$f_k(v_{i,k}) = \frac{v_{i,k} - \underline{v_k}}{\overline{v_k} - \underline{v_k}},$$

where $\overline{v_k}$ and $\underline{v_k}$ are the identified maximum and minimum variable values with the k^{th} variable. A subsequent single complex number z_i (vector) is constructed to represent the local authority in the constellation graph domain, given as follows (for $i = 1, \ldots, N$):

$$z_i = \sum_{k=1}^{K} w_k \exp\left(\sqrt{-1} f_k(v_{i,k})\pi\right),$$

and w_k is the weight of the importance/contribution of the k^{th} variable.

In the context of this paper, the constellation graph method is employed on the three-factor and one-factor models (using details in Tables 3 and 4). Each set of loadings for the three-factor and one-factor models (see Table 4) are normalised (so summing to one – i.e. for the three-factor model, Factor 1, (0.926 + 0.602 + 0.859 + 0.877 + 0.751 = 4.015) when normalised, 0.231, 0.150, 0.214, 0.218 and 0.187; and so on for Factors 2 and 3). The normalised values are the weights of contribution (w_k) of the variables to each factor, in the factors' representations as complex numbers (z_i) in a constellation graph.

Using these weights, the respective constellation graph results can be formulated, finding the respective z_i values for each local authority for the three factors and one factor in the three-factor and one-factor models, respectively, see Figure 1.

Figure 1 about here

In each constellation graph shown in Figure 1, a series of joined lines to a single point represents one local authority (the information associated with one local authority). The joined lines show the sequential contribution of the individual variable values to the final constellation coordinate (z_i) for a region. That is, in Figure 1a, for the 'Population and Housing Dynamics' factor in the three-factor model, each joined line moving out from the centre of the base line (shown for the local authorities Cardiff labelled 5 and Flintshire labelled 10), represents the weighted contribution of the Population Density (V1), Non-Working Agriculture (V6), Male Working Age (V7), Female Working Age (V8) and Dwelling Stock (V9) variables, as shown by the labelling on a number of the joined lines. The lengths of the constituent joined lines match the values of the weights (w_k) of contribution (normalised loadings values) of the included variables (since the constellation graph has radius unity). With the relationship of increasing variables value associated with more urban local authorities, the domain to the constellation graph infers increasing association of rural to urban from left to right. Beyond the positions (z_i) of the local authorities in a constellation graph, measures describing aspects of the level of rurality are next described.

A measure/index of rurality is when the point in the constellation graph is mapped down to the base line of the constellation graph, since the origin (middle of base line) is considered (0, 0), and the radius of the constellation graph is unity, then its value actually goes from −1 (bottom left) to 1 (bottom right), to move it to a standard 0 to 1 index domain, the rurality measure ($RlUn_i$) is given by (where $z_i = (x_i, y_i)$):

$$RlUn_i = \frac{x_i + 1}{2},$$

and has constant domain [0, 1], where values near 0 and 1 denote more rural and urban respectively. The term constant here means that irrespective of the number of variables used in the construction of factors, the rural-urban domain of $RlUn_i$ index values will always go between 0 and 1, since the constellation coordinates (z_i) will always be inside the constellation graph domain. For the z_i points in the constellation graphs in Figure 1, the lines mapping them down on the base line between 0 and 1, denote the rurality index based on that factor (see Table 5 for the actual values – and use of labelling 1 to 22 to denote the local authorities).

One additional feature of this approach to constructing a rurality index, or for the moment intermediate rurality indices, is the notion of consistency in the information from the constituent variables used in the individual factors' constructions. In each constellation graph in Figure 1, some of the local authorities' sets of joined lines are more consistently straighter than others. That is, in technical terms, since each variable value $v_{i,k}$, transformed by $f_k(v_{i,k})$ is over the domain 0 to 1, for a single local authority if the constituent joined lines are all in the same direction it follows the original values are the same proportion of the way through their respective domains.

In the limiting case of local authority Cardiff (labelled 5 in constellation graphs in Figure 1), its joined lines go (start) from the centre of the constellation graph (0,0), nearly parallel to the base line to the right hand corner (near coordinate (1, 0)), with the constituent joined lines all but one going horizontally to the right since all but one of the five variable values making up this Population and Housing Dynamics factor for Cardiff were the largest across all the local authorities (so $f_k(v_{5,k}) = 1$ with the exception of the case for Non-Working Agriculture (V6)). In contrast, for local authority Flintshire (labelled 10) the more 'meandering' nature of its constituent joined lines, compared to those for Cardiff, and subsequent final point away from the circular boundary of the constellation graph, means its variable values, once normalised were different proportions along the 0 to 1 domain. Hence they have different levels of contributory information for the region being more rural or more urban etc. An associated consistency (Cns_i) value to measure this is given by:

$$Cns_i = \sqrt{x_i^2 + y_i^2}\ .$$

This measures the consistency of the variables information contributing to a region's final rural/urban position. In other words it measures how close to the boundary and away from circle centre the final constellation coordinate of the region actually is.

The results, in terms of index and consistency values, for the three and one factors in the three-factor and one-factor models are presented in Table 5.

Table 5 about here

For the one-factor model, the given index values are the final rurality index values for the 22 local authorities, based on the constellation graph approach. For the three-factor model, the three intermediate index values, found from the factor analysis, convey index values to different dimensions on rurality, namely Population and Housing Dynamics, Migratory Dynamics and Social Dynamics.

The next analysis, offers an approach to aggregate these three intermediate multiple rurality indices, even though as mentioned earlier it may be pertinent to use all three indices separately in any further analysis. This aggregation process needs to include the levels of information content that the individual factors have associated with them. This information content is contained in the different levels of '% of variance' associated with each factor. For the factors in the three-factor model, from Table 3, the '% of variance' for each factor is, 35.807, 25.912 and 20.207, which can be normalised so they sum to one, giving, factor information weights of 0.437, 0.316 and 0.247.

Bringing together the details from the three factors identified, in the three-factor model, namely Population and Housing Dynamics, Migratory Dynamics and Social Dynamics, new constellation coordinates can be found using the factor information weights previously found, see Figure 2 (the aggregation can be done on the final z_i values from the different factors for each unitary authority).

Figure 2 about here

In Figure 2, the 22 points across the base line of the constellation graph, associated with the 22 labelled constellation coordinates, represent the aggregated rurality indices for the 22 local authorities based on the constellation graph approach and n-factor model. The numerical rurality index values associated with the constellation graph in Figure 2 are presented in Table 6. A further rurality index value can be found, termed the alternative aggregated three-factor model in Table 6, in this case these index values were found by using all the loadings values shown in Table 4 for the three factors, Population and Housing Dynamics, Migratory Dynamics and Social Dynamics. This is one of the other ways of using the loadings to construct factor scores. Also shown in Table 6 is a series of rurality index values following directly the Cloke (1977) one-factor model approach, finding the rural index value using regression (the loadings in the one-factor model are the coefficients in a regression equation with standardised variable values for the other values in the equation). To directly compare the introduced index values and the Cloke based index values, in the final column in Table 6, the normalised versions of the Cloke one-factor regression values are shown (normalised over the same domain as the aggregated three-factor model results).

Table 6 about here

Table 6 reveals that as expected local authority areas more urban on the three factor models include Cardiff, Swansea, Bridgend and Newport. More rural areas in the analysis in Table 6 are flagged up as Anglesey and Powys. Perhaps more interesting are some of the seeming anomalies. For example, Flintshire in North East Wales is classified as more rural because its migratory dynamics are more akin to what might be expected in a more rural area. We return to some of these seeming anomalies later in Figure 3.

One further piece of interesting information from Table 6 is the difference in the consistency (Cns_i) values associated with the local authorities from the two aggregated three-factor solutions given, when only largest loadings values were used (Aggregated Three-Factor column) and when all loadings values were used (Alternative Aggregated Three-Factor column). While these results do not add further to the index results directly, if more inspection of specific local authorities is warranted then these results, when compared across local authorities, may highlight the possibly inconsistent findings from specific variables, later combined to form the index values.

Following from the approaches to finding certain rurality index values, collective information for single local authorities is next considered, with respective constellation graphs for individual local authorities able to be constructed that include all necessary information from a particular model, here the three-factor model. Figure 3 presents a sample of these constellation graphs and associated information, for the local authorities Cardiff (3a), Powys (3b), Anglesey (3c), and Ceredigion (3d).

Figure 3 about here

In Figure 3, each constellation graph shows the rurality information for a single local authority. In each constellation graph, the three rurality indices found from the three factors are shown, namely, Population and Housing Dynamics, Migratory Dynamics and Social Dynamics, from these the aggregated constellation position is shown with concomitant index. For comparative purposes, the one factor value is given as well as the index value found from following Cloke's (1977) approach. For the constellation graph based results, the constellation graph offers a constant domain, for the Cloke 'one factor regression' index, it was found the values ranged from −4.799 up to 14.445, to enable a level of visual comparison of results between the Cloke based index and aggregated three factor values, the domain of the Cloke index is mapped onto the same domain as the three factor index, hence the horizontal line below the constellation graph is positioned as it is (the least and largest values of the two indices are in line with each other).

In Figure 3 the value of the constellation graph approach can be appreciated. Cardiff and Anglesey are shown to 'fit' well into the urban and rural category respectively on the respective factors. However, both Ceredigion and Powys are usually 'understood' as more rural areas in terms of policy. However, the analysis reveals that Ceredigion 'occupies' a more urban position due particularly to the factor migratory dynamics, while Powys while being more firmly in the rural categorisation on the aggregated three factor analysis, has features in terms of social dynamics which sway more towards the urban. This then is a useful visualisation of the information available and points to some of the problems of strict urban/rural categorisations.

Conclusions

This paper has sought to contribute methods for both identifying and classifying rural and urban areas. Identifying and classifying the rural is still very relevant for policy interventions and provides key context for these same interventions. The interest of planners as well as Government agencies in Europe (Gallent et al, 2006), Australasia (Bunker and Huston, 2003) and North America (Audirac, 1999) in constructing development policies at the rural-urban fringe requires the identification of this changing locality. This paper provides a replicable methodology capable of supporting this identification. We believe the analysis here represents an innovative means of both visualising and analysing the urban/rural classification problem particularly at a time when the rural/urban divide (particularly in the UK) becomes more complex to understand because of improvements in ICT and physical infrastructures. The method adopted in this paper provides for an innovative visual depiction of the contribution of individual variables in the construction of an rurality index value (or factors from a one-factor or n-factors solution). The method provides additional value in the ability to visualise the relative levels of rurality associated with defined areas. The constellation graph method then allows us to examine the relative consistency of the variables' information present in constructed intermediate and single rurality indices.

References

Agarwal S, Rahman S, Errington E, 2009, "Measuring the Determinants of relative economic performance of rural areas" *Journal of Rural Studies* **25** 309-321.

Argent N, 2008, "Perceived density, social interaction and morale in New South Wales rural communities" *Journal of Rural Studies* **24** 245-261

Audirac I, 1999, "Unsettled views about the fringe: rural-urban or urban-rural frontiers", in *Contested Countryside: The Rural Urban Fringe in North America* Eds O J Furuseth, M B Lapping (Ashgate, Aldershot, Hants) pp 7 -32

Bannister D, 1980 *Transport mobility and deprivation in inter-urban areas* (Farnborough, Hants, Teakfield)

Barbieri C, Mahoney E, 2009, "Why is diversification an attractive farm adjustment strategy? Insights from Texas farmers and ranchers" *Journal of Rural Studies* **25** 58–66

Best R, 1981 *Land use and living space* (Methuen, London)

Boschma R, Eriksson H, Lindgren U, 2014, " Labour market externalities and regional growth in Sweden: the importance of labour mobility between skill-related industries" *Regional Studies*, **48**(10), 1669-1690.

Bunker R, Houston P, 2003, "Prospects for the rural- urban fringe in Australia: observations from a brief history of landscapes around Sydney and Adelaide" *Australian Geographical Studies* **41** 303 - 323

Champion A, Watkins C, 1991 *People in the countryside: studies of social change in rural Britain* (Paul Chapman, London)

Clevland C, 1995 "Measure of Rurality" *The University of Tennessee Institute of Agriculture, Agricultural Experiment Station, Bulletin* **689** 5-8

Cloke P, 1977, "An index of rurality for England and Wales" *Regional Studies* **11** 31-46

Cloke P, Johnston R, 2005 "Deconstructing human geography's binaries", in S*paces of Geographical Thought: Deconstructing Human Geography's Binaries* Eds Cloke P, Johnston R, (Sage, London) pp 1–41.

Coombes M, Raybould S, 2001, "Public policy and population distribution: developing appropriate indicators of settlement patterns" *Environment and Planning C: Government and Policy* **19** 23–248

Curl A, Nelson D, Anable, J 2015 "Same question, different answer: A comparison of GIS-based journey time accessibility with self-reported measures from the National Travel Survey in England" *Computers, Environment and Urban Systems*, **49**, 86-97.

Duenckmann F, 2010, "The village in mind: applying Q-methodology to re-constructing constructions of rurality" *Journal of Rural Studies* **26** 284–295

Edmondson B, Fontanez W, 1995, "Where are the boonies?" *American Demographics* **17** 60

Gallant N, Robinson S, 2011, "Local perspectives on rural housing affordability and implications for the localism agenda in England" *Journal of Rural Studies* **27** 297–307

Gallant N, Bianconi M, Andersson J, 2006, "Planning on the edge: England's rural -urban fringe and the spatial-planning agenda" *Environment and Planning B: Planning and Design* **33**, 457 - 476

Hair Jr J, Black W, Babin B, Anderson R, 2010 *Multivariate Data Analysis* (Prentice Hall, London)

Halfacree K, 1993, "Locality and social representation: space, discourse and alternative definitions of the rural" *Journal of Rural Studies* **9** 23–37

Harrington V, O'Donoghue D, 1998, "Rurality in England and Wales 1991: A replication and extension of the 1981 Rurality Index" *Sociologia Ruralis* **38**, 178–203

Hasse D, Totzer T, 2012, "Urban - rural linkages - analysing, modelling, and understanding drivers, pressures, and impacts of land use changes along the rural-to-urban gradient" *Environment and Planning B: Planning and Design* **39,** 194 -197

Hoggart K, 1988, "Not a definition of rural" *Area* **20** 35–40

Hugo G, Smailes P, Macgregor C, Fenton M, Brunckhorst D, 2001, *Defining Social Catchments in Non-Metropolitan Australia* (Department of Agriculture, Fisheries and Forestry, Australia)

Hugo G, Champion A, Lattes A, 2003, "Towards a new conceptualization of settlements for demography", *Population and Development Review* **29** 277-97

Isserman A, 2005, "In the national interest: Defining rural correctly for research and policy" *International Regional Science Review* **28** 465-499

Jauhiainen J, 2009, "Will the retiring baby boomers return to rural periphery?" *Journal of Rural Studies* **25** 25–34

Kaiser H. F, 1960, "The application of electronic computers to factor analysis" *Educational and Psychological Measurement* **20** 141-151

Mitzuta M, 1994, "A review of graphical methods in Japan – from histogram to dynamic display" *New Approaches in Classification and Data Analysis*, Springer-Verlag, pp. 36-44.

Muilu T, Rusaneu J, 2004, "Changes in population and industries in the rural areas of Finland: From analysis of administrative regions to GIS based approach". In *Changing Functions of Rural Areas in the Baltic Sea Region* Ed Banski J (Institute of Agricultural and Food Economics and Institute of Geography and Spatial Organization, Warsaw) pp 109–117

Pacione M, 1984 *Rural geography* (Harper and Row, London)

Pateman T, 2011, "Rural and urban areas: comparing lives using rural/urban classifications", *Regional Trends*, **43**

Scott A, Gilbert A, Gelan A, 2007, 'The Urban-Rural Divide: Myth or Reality', The Macaulay Land Use Research Institute.

Sekiya T, Watanabe A, Saito M, 1991, "The use of modified constellation graph method for computer-aided classification of congenital heart diseases" *IEEE Transactions on Biomedical Engineering* **38** 814-820

Wakimoto K, Taguri M, 1978, "Constellation graphical method for representing multidimensional data" *Annals of the Institute of Statistical Mathematics* **30** 97-104.

Waldorf B, 2006 "A Continuous Multi-dimensional Measure of Rurality: Moving Beyond Threshold Measures" Annual Meetings of the Association of Agricultural Economics, Long Beach, CA, July 2006. http://agecon.lib.umn.edu/cgi-bin/pdf_view.pl?paperid=21522&ftype=.pdf

Weisheit R, Falcone D, Wells L, 1999 "Crime and policing in rural and smalltown America" (Waveland Press, Prospect Heights, IL)

List of Tables and Figures

Table 1. Comparison of Variables used in Selected Rurality Indices

Smith *et al*. (1973) Rurality Index Variables	**Cloke (1977) Rurality Index Variables**	**Clevland (1995) Rurality Index Variables**
Population density	Population Density	Metropolitan access via interstate highway
Percent of Persons Living in rural areas	% Population age over 65	Percent in retail employment
Total Population	% Population men age 15-45	Percent in professional employment
Percent employment in agriculture	% Population women age 15-45	Percent in agricultural employment
Percent of persons living on farm	Occupancy rate (household/dwelling)	Median household income
Average annual percent change in population	Occupancy rate (persons per room)	Percent of families in poverty
Percent employment in medical profession	Household amenities	Percent in governmental employment
Percentage employment in entertainment	Occupational structure (% agricultural)	Percent population change
Percentage employment in service work	Commuting out pattern	Percent over 65 years of age
	Population change (in past 10 years)	Population density (per sq. mile)
	In-migration (% population resident <5 years)	Hi/Low education ratio
	In/out migration balance	
	Out Migration (% population moved out in the last year)	
	Distance from nearest urban centre of 50,000 population	
	Distance from nearest urban centre of 100,000 population	
	Distance from nearest urban centre of 200,000 population	

Adapted from Smith *et al*. (1973), Cloke (1977), and Clevland (1995)

Table 2. Description of 10 variables (taken from reference to Cloke, 1977*)

Variable	Description and Source
V1: Population Density	Population/ Area (Office for National Statistics)
V2: Population Change 2000-2010 %	% Change in Population 2000-2010 (Stats Wales)
V3: In Migration %	% Total Population (Stats Wales)
V4: Out Migration %	% Total Population (Stats Wales)
V5: Net Migration %	% Total Population (Stats Wales)
V6: Working Population Excluding Agricultural Employment %	% of Total working population NOT employed in Agriculture (100 – Workinagri %) (Stats Wales)
V7: Male %Working Age	As a % of Total Male Population 18-65 (Stats Wales)
V8: Female %Working Age	As a % of Total Female Population 18-65 (Stats Wales)
V9: Dwelling Stock	The total number of dwellings in local authority area (Stats Wales)
V10: % of Population that does not Commute	% Total working Population who DO NOT commute out of the local authority area to work (100 – Commute %) (Stats Wales)

*Raw data extracted from sources, authors calculations

Table 3. Factor analysis results for 10 variables (described in Table 2)

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.040	40.397	40.397	3.449	34.490	34.490
2	2.542	25.423	65.820	2.557	25.575	60.064
3	1.499	14.993	80.813	2.075	20.749	80.813
4	0.784	7.840	88.653			
5	0.526	5.259	93.913			
6	0.270	2.697	96.610			
7	0.194	1.942	98.552			
8	0.102	1.019	99.571			
9	0.043	0.429	100.000			
10	0.000	0.000	100.000			

Extraction Method: Principal Component Analysis.

Table 4. Variable loadings values for three-factor and one-factor models

Variable	Components (three-factor model)				Component (one-factor model)
	Factor 1	**Factor 2**	**Factor 3**		**Factor 1**
V1: Population Density	**0.926**	0.140	0.015		**0.811**
V2: Population Change 2000-2010 %	0.279	0.175	**0.812**		**0.551**
V3: In Migration %	0.138	**0.972**	0.164		**0.682**
V4: Out Migration %	0.169	**0.950**	–0.020		**0.641**
V5: Balance Migration %	0.032	**0.727**	0.542		**0.576**
V6: Non-Working Agriculture %	**0.602**	–0.094	–0.460		**0.291**
V7: Male %Working Age	**0.859**	0.167	0.092		**0.795**
V8: Female %Working Age	**0.877**	0.279	–0.116		**0.809**
V9: Dwelling Stock	**0.751**	–0.087	0.507		**0.693**
V10: Non-Commute %	–0.179	0.098	**0.777**		**0.138**

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 5 iterations (for three-factor model).

Table 5. *Urbn*$_i$ (and *Cns*$_i$) values for 22 local authorities for the factors in three-factor and one-factor models

	Three-factor				One-factor
Local authority	**Population and Housing Dynamics**	**Migratory Dynamics**	**Social Dynamics**		**All**
1. Anglesey	0.172 (0.754)	0.230 (0.997)	0.321 (0.995)		0.174 (0.868)
2. Blaenau Gwent	0.392 (0.674)	0.248 (0.832)	0.018 (0.982)		0.273 (0.748)
3. Bridgend	0.385 (0.756)	0.533 (0.905)	0.617 (0.962)		0.447 (0.833)
4. Caerphilly	0.471 (0.810)	0.338 (0.998)	0.235 (0.815)		0.396 (0.906)
5. Cardiff	1.000 (1.000)	0.844 (0.999)	0.932 (0.928)		0.948 (0.928)
6. Carmarthenshire	0.204 (0.870)	0.503 (0.948)	0.721 (0.999)		0.360 (0.810)
7. Ceredigion	0.227 (0.728)	1.000 (1.000)	0.733 (0.719)		0.525 (0.395)
8. Conwy	0.133 (0.788)	0.291 (0.995)	0.495 (0.981)		0.194 (0.813)
9. Denbighshire	0.123 (0.918)	0.555 (0.999)	0.651 (0.993)		0.311 (0.836)
10. Flintshire	0.315 (0.903)	0.008 (0.994)	0.178 (0.940)		0.211 (0.864)
11. Gwynedd	0.165 (0.874)	0.374 (0.998)	0.687 (0.685)		0.280 (0.860)
12. Merthyr Tydfil	0.388 (0.600)	0.294 (0.944)	0.171 (0.971)		0.300 (0.787)
13. Monmouthshire	0.124 (0.922)	0.576 (0.851)	0.391 (0.790)		0.312 (0.748)
14. Neath	0.282 (0.919)	0.466 (0.996)	0.292 (0.979)		0.347 (0.932)
15. Newport	0.414 (0.805)	0.551 (0.995)	0.440 (0.976)		0.442 (0.910)
16. Pembrokeshire	0.198 (0.730)	0.154 (0.996)	0.820 (0.805)		0.213 (0.791)
17. Powys	0.063 (0.949)	0.086 (0.996)	0.817 (0.871)		0.148 (0.853)
18. Rhondda Cynon Taff	0.559 (0.755)	0.220 (0.977)	0.227 (0.998)		0.397 (0.830)
19. Swansea	0.595 (0.785)	0.558 (0.995)	0.747 (0.926)		0.577 (0.885)
20. Torfaen	0.310 (0.863)	0.225 (0.930)	0.107 (1.000)		0.236 (0.909)
21. Vale of Glamorgan	0.324 (0.724)	0.628 (0.983)	0.373 (0.520)		0.421 (0.781)
22. Wrexham	0.373 (0.828)	0.098 (0.861)	0.674 (0.996)		0.318 (0.750)

Table 6. Rurality index values from different models

	Aggregated Three-Factor	Alternative Aggregated Three-Factor	Cloke (1977) Approach	Cloke (normalised to fit Aggregated three-factor)
1. Anglesey	0.229 (0.856)	0.220 (0.769)	–4.159	0.207
2. Blaenau Gwent	0.250 (0.727)	0.262 (0.673)	–2.112	0.287
3. Bridgend	0.491 (0.833)	0.464 (0.733)	2.052	0.450
4. Caerphilly	0.368 (0.845)	0.382 (0.789)	1.202	0.417
5. Cardiff	0.933 (0.923)	0.871 (0.784)	14.445	0.933
6. Carmarthenshire	0.431 (0.814)	0.426 (0.686)	–0.166	0.363
7. Ceredigion	0.602 (0.402)	0.580 (0.356)	4.143	0.531
8. Conwy	0.276 (0.797)	0.257 (0.705)	–4.460	0.196
9. Denbighshire	0.395 (0.808)	0.376 (0.674)	–1.218	0.322
10. Flintshire	0.182 (0.854)	0.234 (0.715)	–3.378	0.238
11. Gwynedd	0.365 (0.743)	0.354 (0.717)	–1.466	0.312
12. Merthyr Tydfil	0.303 (0.795)	0.296 (0.721)	–1.140	0.325
13. Monmouthshire	0.336 (0.762)	0.369 (0.613)	–1.100	0.327
14. Neath	0.343 (0.941)	0.374 (0.797)	0.216	0.378
15. Newport	0.464 (0.900)	0.443 (0.804)	2.118	0.452
16. Pembrokeshire	0.344 (0.613)	0.274 (0.654)	–3.445	0.235
17. Powys	0.264 (0.677)	0.245 (0.669)	–4.799	0.182
18. Rhondda Cynon Taff	0.366 (0.830)	0.375 (0.742)	1.126	0.414
19. Swansea	0.622 (0.874)	0.572 (0.769)	4.341	0.539
20. Torfaen	0.231 (0.904)	0.246 (0.791)	–2.107	0.287
21. Vale of Glamorgan	0.433 (0.700)	0.426 (0.674)	1.424	0.425
22. Wrexham	0.363 (0.728)	0.348 (0.634)	–1.517	0.310

Figure 1. Constellation graphs showing rurality indices, based on, Population and Housing Dynamics, Migratory Dynamics and Social Dynamics, with three-factor model and one-factor model

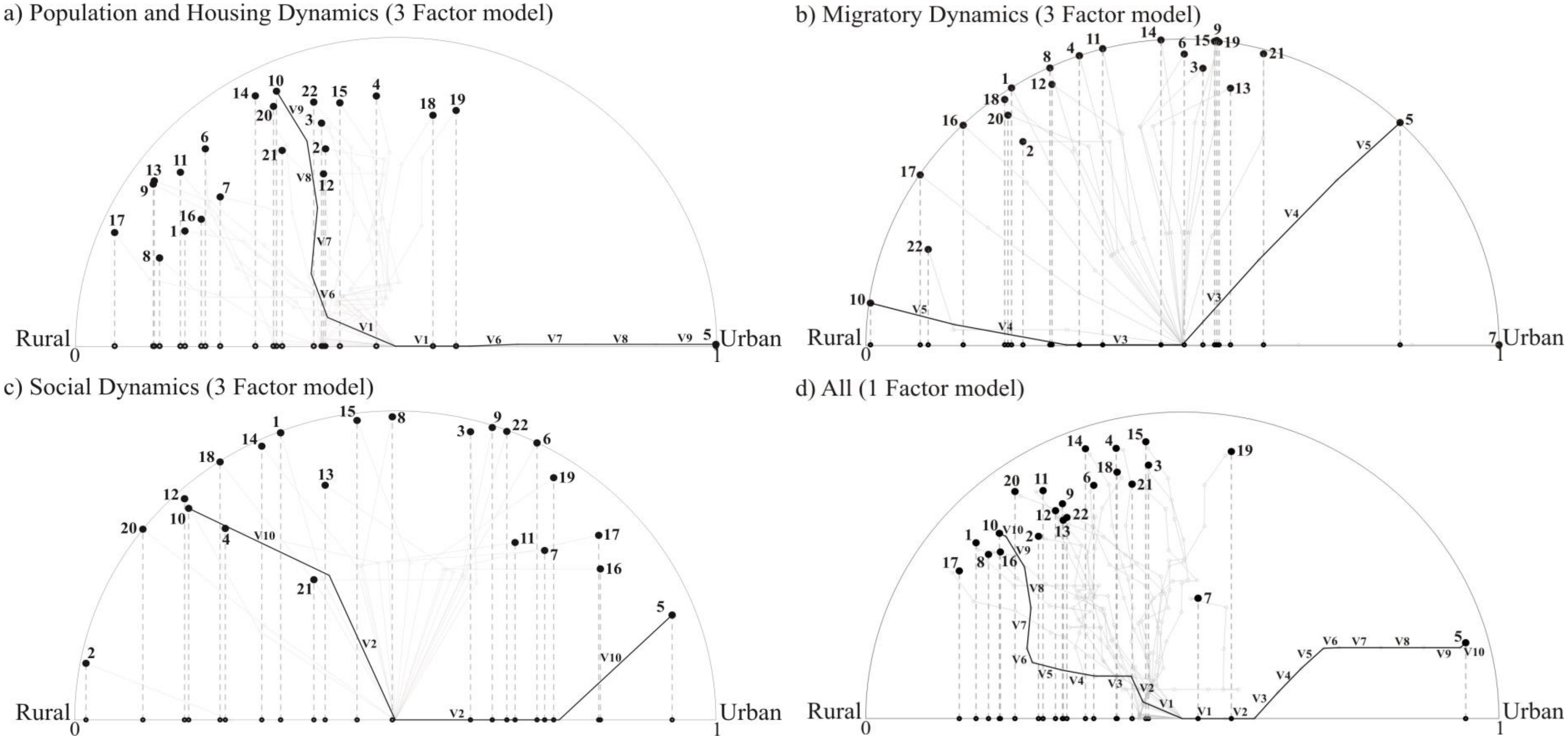

Figure 2. Constellation graph for aggregated three factor rurality index in three-factor model

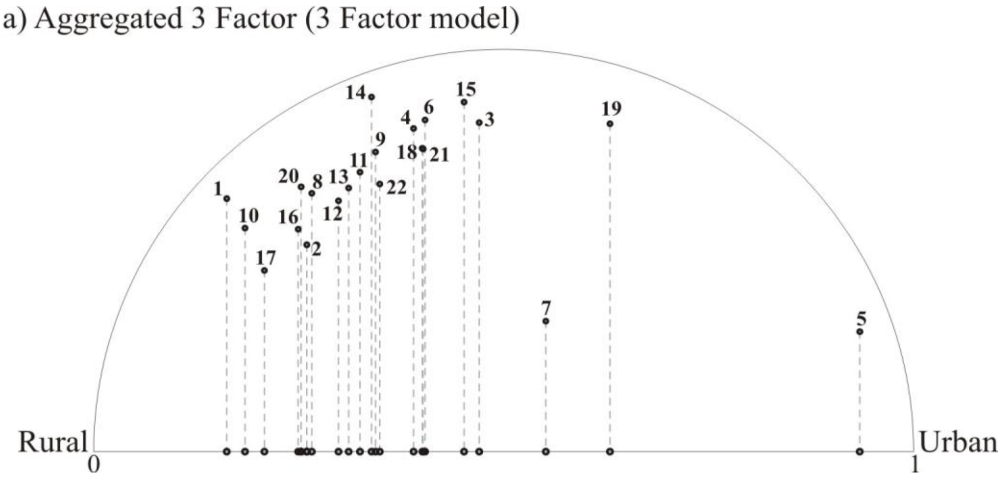

Figure 3. Constellation graphs showing rurality indices, based on, Population and Housing Dynamics, Migratory Dynamics and Social Dynamics, with aggregated three factor index, one factor index and one factor Cloke regression index values also shown