Article:

Comber, A, Dickie, J, Jarvis, C et al. (2 more authors) (2015) Locating bioenergy facilities using a modified GIS-based location-allocation-algorithm: Considering the spatial distribution of resource supply. Applied Energy, 154. pp. 309-316. ISSN 0306-2619

https://doi.org/10.1016/j.apenergy.2015.04.128

Locating bioenergy facilities using a modified GIS-based location-allocation-algorithm: considering the spatial distribution of resource supply

Alexis Comber*, Jennifer Dickie, Claire Jarvis, Martin Phillips, Kevin Tansey

Department of Geography, University of Leicester, Leicester, LE1 7RH, UK
* contact author: email ajc36@le.ac.uk

Abstract

This paper proposes a modification to the classic p-median problem that considers the spatial distribution of supply resources and competition for them by potential facility locations. It is illustrated with a simplified case study to optimally locate community scale anaerobic digesters (ADs) in an area in the East Midlands in the UK. The modification evaluates the spatial distribution of the feedstocks needed by each potential AD unit and only includes new locations if their supply catchments do not overlap with the catchments of the current set of locations being considered. The modified algorithm is implemented using the Teitz and Bart search heuristic. This starts with an initial set of locations that it seeks to improve by swapping poor performing locations with better ones. In this case the modification takes account of the spatial distribution of the feedstocks for a typical AD recipe whilst seeking minimise demand weighted distance. The results demonstrate large improvements over the classic p-median model in location selection that eliminate the overlap of facility catchments. Some wider points relating to the robustness of many analyses reported in the bioenergy literature are discussed, with some observations about commonly found methodological deficiencies, before some areas for further research are suggested.

Keywords: location-allocation, geo-computation, spatial analysis, Geographic Information Systems, GIS

Highlights

- Location-allocation models are frequently used to locate bio-energy facilities
- Current approaches do not account for supply competition between potential sites
- We propose an extension to the p-median model that evaluates supply catchments
- It is demonstrated using a case study to optimally locate AD facilities
- Both resource supply and competition are critical when locating bio-energy facilities

1. Introduction

There is much interest in methodologies for identifying suitable locations to site renewable and bioenergy facilities including biochemical biomass conversion through anaerobic digestion. This is being driven by wider policy recognition of the role renewable energy has in reducing greenhouse gas emissions, increasing energy security, transitioning to a low carbon economy and improving waste management (Adams et al. 2011; Levidow et al. 2014; Perry and Rosillo-Calle 2008; Thorley and Cooper 2008). Geographic Information Systems (GIS) have a large number of tools and functions that can be used to identify the 'best' locations for such facilities. An evolution of methods in the bioenergy literature is evident from multi-criteria evaluations (for example, Ma et al., 2005) to location-allocation models (e.g. Sultana and Kumar, 2012). However, different approaches will result in different answers and in different locations identified as optimal for bioenergy and anaerobic digester (AD) facilities. In this context it is important that researchers using GIS to locate facilities have a clear understanding of the problem they are trying to solve and fully understand the operations and assumptions embedded in the GIS tools that they use. This is difficult because many location-allocation tools within commercial GIS are black boxes and are not fully documented.

Location-allocation analyses seek identify the set of *n* optimal locations by considering the spatial distribution of the potential locations (supply) in relation to spatially distributed demand. The aim of methods such as the p-median model (Hakimi (1964) is to optimise an evaluation function, for example to identify the set of locations that minimise demand weighted distance, travel times or costs under different types of constraints.

One of the underlying assumptions in many commercially available GIS implementations of location-allocation models is that the 'supply' of whatever is being considered is simply present at each location. For example, consider a set of *n* medical facilities located to satisfy a spatially distributed demand. The assumption is that resources needed by each medical facility are simply present: staff, medicines, communications networks etc. For many kinds of facility this assumption is unproblematic: things have to be delivered somewhere and the additional travel costs are marginal. Such assumptions are in part due to the way location-allocation analyses are undertaken: potential facility locations are represented as discrete points, usually as vertices on a road network, although they may be attributed with properties such as supply capacity, resource volumes, etc.

However, the location of bioenergy facilities such as ADs presents a different kind of problem. The supply of resources to any potential AD facility location is a critical consideration when evaluating potential sites for such facilities: it would not make sense to locate ADs in places where the environmental, financial or energy costs of supply severely mitigate their potential energy benefits. Thus in selecting the optimal set of *n* AD locations, the spatial distribution of the feedstocks needed to supply each facility is an important constraint. This is a critical point in renewable energy research as the ultimate evaluation of a potential facility location should be the need for net energy gains (rather than financial cost). This is especially so for community scale AD facilities, where feedstocks typically have to come from the local area for them to be economically viable. Community scale ADs result in multi-level benefits. At a policy level, smaller scale renewable energy facilities promote and develop capacities in sustainable energy as the AD feedstocks are typically supplied locally. In terms of sustainability, the process of setting up and establishing a community scale AD has been found to have an impact on community behaviour (Hoffman and High-Pippert, 2010). It can promote wider engagement with the issues related to renewable energy, with the result that as well as energy production, such community scale activities can transform local energy consumption patterns and other behaviours (Hielscher et al., 2011). AD facilities require a mix of resources or 'feedstocks' from different inputs such as animal carcasses, agricultural residues, domestic food waste and animal manures. At the community scale, the best combination of feedstocks based on the C/N ratios are a mix of agricultural residues and domestic food wastes (Ward et al., 2008).

This paper describes a modification to the p-median algorithm that considers the spatial distribution of different feedstock supplies. In so doing explores some of the critical issues in identifying optimal facility locations through a case study of locating anaerobic digester (AD) facilities. These include the need to consider the spatial distribution of resources, the resource catchments of potentially suitable locations and the need to evaluate sets of potential locations together and not one by one. The paper proceeds with a review of spatial methods that have been used for selecting AD facility locations and the application of location-allocation algorithms in this context (Section 2) before introducing the specific case study used to demonstrate this method extension and providing a detailed description of the proposed p-median modification (Section 3). The results are presented in Section 4 before discussion of the issues arising from this research (Section 5).

2. Background

This review is in two parts. The first describes the methods that have been used in biomass energy facility site selection. These have at their core the desire to identify the 'best' potential locations that, for example, minimise the costs of biomass transportation. The second focuses on the use of and logics of location-allocation methods to determine 'where to put things' that account for the inherent spatial characteristics of the phenomenon under investigation. This review does not describe the many different ways that spatial distribution of biomass has been calculated or the need for such facilities, as these are not the focus of this paper.

2.1 Renewable energy facility location

Studies seeking to identify optimal locations for AD and other renewable energy generation facilities generally use one of 2 GIS-based approaches: *suitability analyses*, also known as Multi-Criteria Evaluations (MCEs), which use different types of buffer and spatial overlay operations (e.g. Ma et al., 2005), and, what have been called *optimality analyses*, which employ search heuristics to satisfy location-allocation problem such as the p-median and seek to match spatially distributed supply and demand (e.g. Shi et al., 2008). In the wider domain of geo-computation and spatial analysis, a number of different suitability approaches are well-established. These range from the simple Boolean overlay, masking out all areas for which any constraint exists, to the classic Weighted Linear Combination (Voogd, 1983; Carver, 1991), Ordered Weighting Averaging (Yager, 1988; Jiang and Eastman, 2000) and the Analytic Hierarchy Process (AHP) (Bania, 1993; Boroushaki and Malczewski, 2008;

Wu, 1998). A description of the rationale, logics, limitations and applicability of the different types of suitability analysis can be found in Jiang and Eastman (2000).

In the bioenergy energy literature suitability analyses have been applied in a number of ways. Ma et al. (2005) used AHP to identify areas suitable for AD facilities and other MCE approaches have been used to examine biomass residues (Recchia et al., 2010), to identify locations suitable for bioenergy production (Baccali et al., 2009) and to identify locations to collect crop residues for bioenergy production (Haddad and Anderson, 2008). The research by Ma et al. (2005) has become the touchstone in the biomass literature on locating AD facilities. Their approach covers the spatial overlay and weighting operations used to determine suitable locations. Critically in the context of this review, the outputs (maps) describe areas that are 'suitable' or their degree of suitability using a suitability index. That is, the results describe where AD facilities *could* be located but do not take demand into account and do not consider the impacts of locating *n* AD facilities within the zones described as suitable.

A second tranche of research has sought to locate optimally locate *n* facilities using location-allocation techniques. For example, Panichelli and Gnansounou (2008) applied location-allocation approach to identify bioenergy (biomass energy) facility locations. They sought to identify the optimal 2 locations for 2 different bioenergy facilities, each with differing capacities, evaluated over biomass farmgate costs (i.e. those related to collection, haulage etc. of biomass). Thompson et al. (2013) identified optimal AD locations amongst farms too small to have an AD facility themselves, using a 'maximum attendance' algorithm to select locations that could be served with manure within 22km network distance. Other work includes Sliz-Szkliniarz and Vogt (2012) used a suitability analysis to identify areas suitable for potential AD facilities with animal manure as the major supply variable and then applied a focal / neighbourhood analysis to determine 'optimal' locations. Shi et al. (2008) identified optimal biomass plant locations using location-allocation techniques. Their method converted land cover to a biomass resource and defined a set of potential facility locations on a road network and then a Voronoi tessellation was used to allocate biomass resources to the nearest facility location. Sultana and Kumar (2012) used a location algorithm, implemented in commercial GIS software, to identify locations for biomass facilities, after selecting candidate locations using AHP. They used a p-median model where the potential supply and demand locations were determined using a standard GIS-based suitability analysis. Their model considered transport costs and determined optimal locations with the constraint that each source was assigned to only one potential facility.

Thus much research has sought to integrate suitability and location-allocation approaches, with a suitability analysis being used first to identify a regions that could support AD facilities before location-allocation is used to identify the set of optimal locations. However, from a geo-computation perspective, there are a number of concerns about logic of the way that location-allocation analyses have been used in relation to the purported study objectives. This will be returned to later in this section but at this point it is instructive to consider location-allocation models.

2.2 Location-allocation

Location-allocation analyses seek to match supply with demand. The aim is frequently to identify best *n* locations and potential sets of locations are evaluated together using an evaluation function of some kind. Typically this is 'demand weighted distance', the default option in location-allocation tools in the ESRI's ArcGIS, perhaps the leading GIS software, but other evaluation functions have been specified depending on the task in hand. Classic location-allocation models include:

- the p-median problem (Hakimi, 1964; ReVelle and Swain, 1970) which minimises the demand weighted distance;
- the maximal covering location problem (Church and Velle, 1974) which seeks to satisfy a distance or travel criterion;
- the location set covering problem (Toregas et al. (1971) which seeks to that minimise the number of facilities to satisfy some demand coverage.

These approaches and their variants – see Owen and Daskin (1998) for a comprehensive review – have been used to consider a number of different types of geographical problems related to facility siting and accessibility. Typical applications include where to site new facilities (e.g. Sasaki et al., 2011), where to enhance capacity in existing facilities (e.g Sasaki et al., 2010), to determine which facilities to close (e.g. Comber et al, 2009) and where to locate new facilities (e.g. Comber et al., 2011). Location-allocation models typically have three inputs: potential supply points, spatially distributed demand often reduced from areas to points and some measure of distance between supply and demand locations such as road distance or travel time. They are used to suggest facility locations, to identify gaps in service provision and to highlight geographic regions with low service coverage.

Location-allocation models are frequently applied within a search heuristic rather than deterministically. This is because very simple location-allocation problems are highly dimensional and it would simply take too long to evaluate every set of potential locations. Classic search heuristics include the Teitz and Bart search heuristic (1968), genetic algorithms (Goldberg, 1989; Holland, 1975) and grouping genetic algorithms (Falkenauer, 1998; Comber et al., 2011). Each of these converges on the optimal set of *n* locations without having to evaluate every possible set of *n*. The reason for avoiding deterministic solutions are illustrated by a simple example: consider a problem where the objective is to identify the best 35 facilities in a study area comprising a 100 by 100 grid (for example of 1km^2) of potential locations. The number of potential sets of 35 in this small area that would have to be evaluated is large:

$$\frac{10000!}{35! \times (10000 - 35)!} = 9.12 \times 10^{99}$$

2.3 Critique

An evolution in GIS and spatial analysis sophistication is evident in research into bioenergy facility location: much of the early work used suitability analyses to identify regions that satisfied a set of criteria. Other research used location-allocation techniques and some studies incorporating both approaches (eg. Voivontas et al., 2001; Sultana and Kumar, 2012). In these, each technique is used to preform a discrete task: typically suitability analyses are used to identify an initial set of potential locations that satisfy a number of constraints, whereas the location-allocation is used to allocate identify a subset of those locations that minimise some evaluation function, such as transport cost. These developments are laudable.

However, some studies that used GIS and location-allocation approaches for facility siting published in the bioenergy literature provide grounds for concern. Specifically there are a number of critical considerations:

1) Pure MCE approaches identify suitable areas. They do not identify discrete locations or consider the resource or demand impact of siting a facility in that area on the suitability of that area for subsequent facilities.

2) Any potential set of *n* suitable locations should evaluate all *n* locations together. It is illogical to identify the best location, then add the second best to the set, then the third best … to the *n*th best. It is self-evident that the single best location may not be in the set of *n* best locations.

3) Evaluation functions are not always clearly specified in research papers where location-allocation approaches have been used and yet they are critical for selecting sets of locations.

4) In some papers no account is taken of the spatial distribution of resources in relation to demand.

5) No account is taken of the possible competition for resources them potential selected locations despite this issue being identified by some authors.

Explicitly addressing these issues is critical if research is to robustly support policy and spatial planning: the map is a very powerful output and precisely because it is increasingly easy to generate sophisticated mapped outputs from the application of very complex (but opaque) GIS functions and tools, there is need to thoroughly consider the operations and inputs of such tools. This paper addresses the considerations listed above and in so doing extends the p-median model.

3. Methods

The objective of the analysis was to determine the ‘best’ locations for *n* AD systems based on the availability of AD feedstocks (supply), the distribution of rural population (demand), a feedstock ‘recipe’ and the input requirements of an AD system that potentially might be suited to community use. The critical consideration in the selection of any potential facility location was to determine how much of the surrounding area would be needed to supply feedstock resources, and then to exclude other nearby potential facility locations that would need the same resource areas if they were selected. The problem description is developed in detail below for a case study around Lincolnshire in the East Midlands region of the UK. The study area and the rural population census areas it contains are shown in Figure 1.

Figure 1. The study area in the East Midlands, in the context of the UK and with rural census areas shaded by population (demand) with a transparency term, all with OpenStreetMap backdrops.

3.1 Problem description and data

In order to illustrate the spatial extension to the p-median model, a community scale AD unit was specified as requiring 2.5 tonnes of dry matter per day (912.5 tonnes per year) and an optimum feedstock of 3 parts (in dry matter) of domestic food waste, 1 part cattle slurry and 1 part wheat straw. These specifications were based on actual AD systems available on the commercial market and expert feedstock recommendations. This equates to feedstocks of 547.5 t yr^{1} of food waste, 182.5 t yr^{1} of slurry and 182.5 t yr^{1} wheat straw. There are some broad assumptions embedded in this specification, for example, the seasonality of the feedstocks has not been considered and they are assumed to be present throughout the year.

The supply data were generated as follows. Spatially distributed data on cattle slurry (t km^{-1}) and wheat (ha km^{-1}) were provided by the ADAS manure and land use database (Proctor et al., 2005; Comber et al., 2008). This 1km resolution dataset uses the Ordnance Survey 1km grid locations. It integrates agricultural land use and animal data from the annual UK June Agricultural Census using the methods described in Comber et al. (2008) and is updated annually. The manure database combines information about local manure management practices such as the timing and location of applications, the proportions of manure applied as slurry and farmyard manure (Smith et al., 2001). Although the database provides monthly breakdowns, the annual total of cattle slurry applied to grass and to arable land from 2010 were used here. The database also describes the amount of land in hectares cultivated as wheat in each 1km^{2} cell. This was converted to tonnes of dry wheat straw per 1km^{2} cell using a residue to product ratio from the Biomass Energy Centre[1] of 3.5t ha^{-1} of wheat. The third feedstock component was derived from a pycnophylactic interpolation (Tobler, 1979) of household number from the 2011 population census in each census small area (Output Area) over the same 1km grid as the agricultural data. The household number was multiplied by a factor of 0.260 (Quested et al., 2013) to estimate the amount of household waste available as AD feedstocks in tonnes per year. The feedstocks all relate to dry material and their spatial distributions are shown in Figure

---

[1] http://www.biomassenergycentre.org.uk/portal/page?_pageid=75,17972&_dad=portal&_schema=PORTAL

2. The spatial distribution of ‘demand’ was also derived from the 2011 population census using the usual residential population (Figure 1).

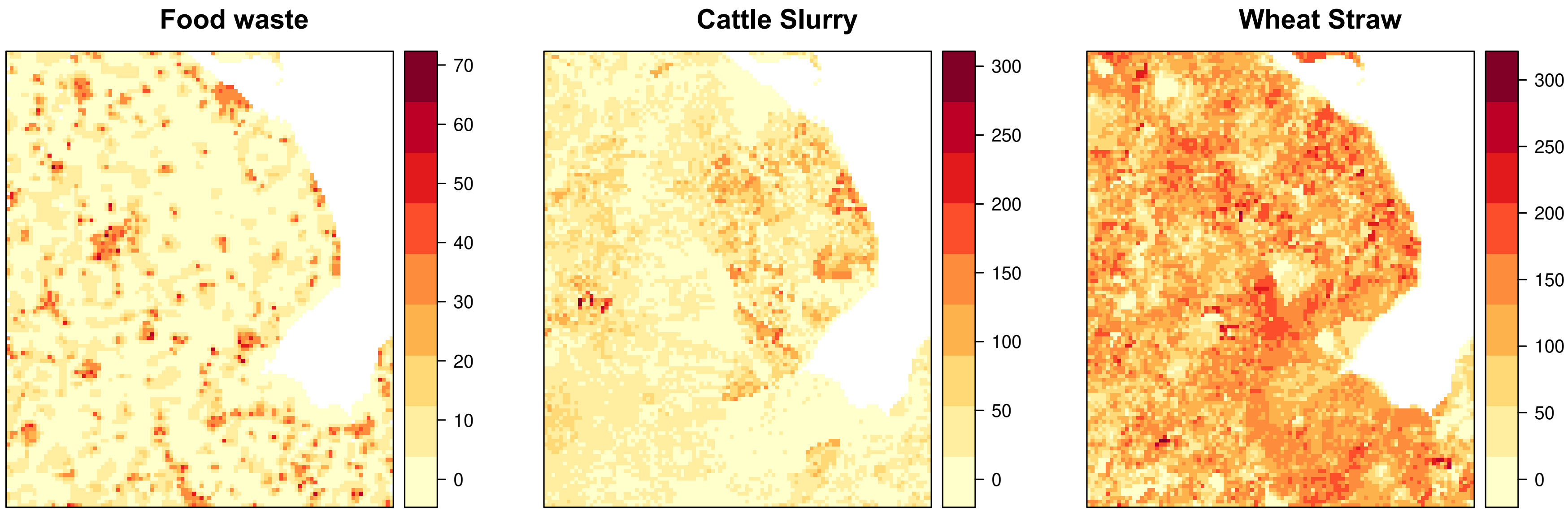


Figure 2. The spatial distribution of feedstocks supplies (t yr^{-1} km^{-2}) in the study area.

In order to test the algorithm a hypothetical objective was formulated. This was to identify the best subset of 35 locations from a set of 524 randomly selected rural village locations (Bibby and Shepherd, 2004). That is, those locations with the greatest potential to sustain a community-level AD system. The choice of 35 locations and the use of 524 locations was to ensure a case study of moderate complexity, that would clearly illustrate the methods and any differences in the results of applying the classic p-median model and the modified model. The locations that were evaluated are shown in Figure 3.

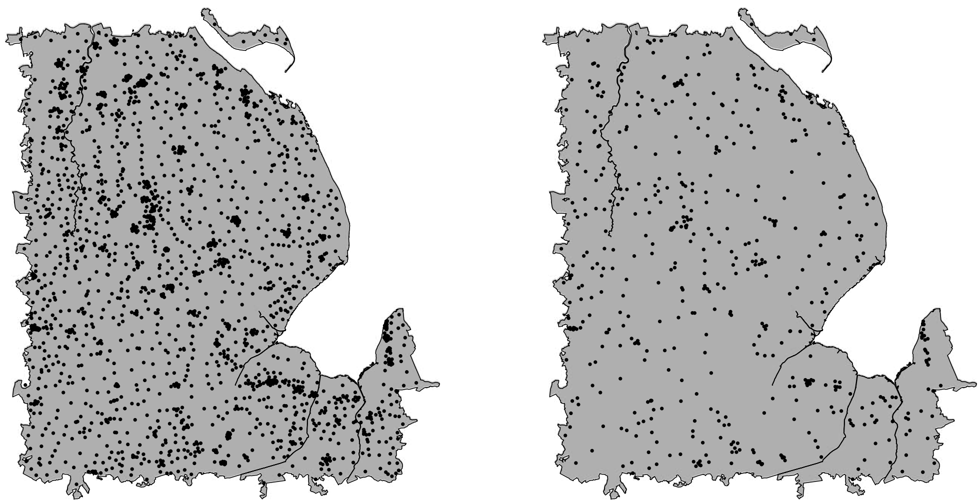


Figure 3. The original 2093 rural demand locations and the random sample of 524 locations used in the analysis.

Next, the feedstock catchments required at each location were determined by calculating the radial distance needed supply the AD with each of the three feedstocks. A circular buffer was iteratively grown at each potential point until the feedstocks needed to supply the AD were satisfied. At each iteration, the catchment was increased by a radius of 1km (the units over which the feedstocks are reported). If the catchment increase resulted in more feedstocks than needed, the distances were rescaled proportionately, and the search distance for each feedstock was attached to each point. It is clear from Figure 2 that food waste is generally the constraining factor, requiring the largest catchment, and many of the largest food waste catchments are found at the edges of study areas (Figure 4).

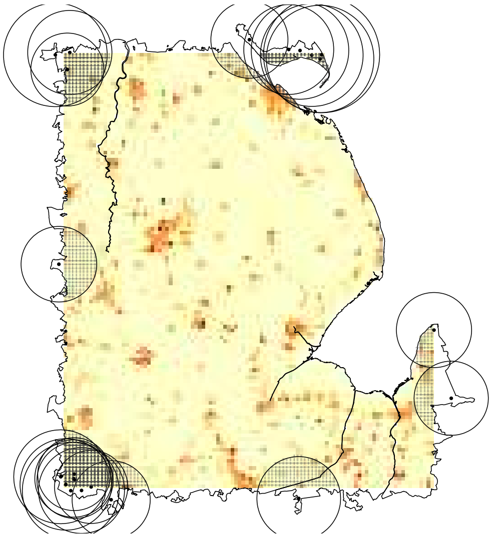


Figure 4. Catchments for food waste slurry whose radius is greater than 10km.

Thus, a set of potential facility locations based on census area centroids were identified. The demand at those locations was described by the total population. In this way, the 524 locations in Figure 3 have both supply and demand constraints: the supply constraints relate to the different feedstock catchments required to satisfy an AD at that location and the demand is driven by the population at that location. The resource catchments needed to support a community scale AD were determined and the Euclidean distance between each supply / demand location was calculated. It is important to note that the analysis is not assessing whether the AD would produce sufficient energy to meet all, or a fixed proportion of the energy needs of each at demand location. Rather the aim of the paper is to demonstrate a method that optimally locates *n* facilities whilst taking account of the spatial distribution of feedstock supplies.

3.2 Modified p-median model

The *p*-median model (Hakimi, 1964; ReVelle and Swain, 1970) seeks to identify sets of supply locations that minimise demand weighted distances. It is formulated as follows:

$$\sum_i^m \sum_j^n a_i d_{ij} x_{ij} \qquad \text{(Equation 1)}$$

where i is the index of demand locations (*1* to m) and j is the index of supply (*1* to n), a_i represents the demand at demand location i, d_{ij} is the distance between i and j and x_{ij} is an allocation decision variable with a value of 1 if demand at location i, is served by a supply j and 0 if otherwise. In this way the p-median model accepts new potential locations if they reduce the overall demand weighted distance between supply and demand locations.

The implementation of the algorithm in the Teitz and Bart (1968) heuristic starts with an initial set of n potential locations and then proceeds to swap these with other locations (an 'exchange operation'), testing for improvement in the evaluation function and accepting the new location in the set if improvement is found. In this case the aim of the evaluation function was to minimise demand weighted distance. Demand weighted distance is calculated from the sum of population weighted distances (distance multiplied by demand), where for each demand location, distance is to the nearest supply location in the set of n locations being evaluated and demand is the population at that demand location.

The problem with the classic p-median approach is that sufficient feedstocks for an AD are not located at each supply location. Rather they will be collected from farms and houses nearby. Also, depending on the spatial distribution of the selected locations, the feedstock resources may be allocated to different locations. Thus in the operation of the algorithm there is a need to exclude potential facility locations that fall within the catchments of the set being considered from the exchange operation. To achieve this, the p-median algorithm was modified as follows. Firstly, for each potential location, all of the supply locations within its catchments were identified for each of the three feedstocks. Secondly, the exchange operation function was modified such that only locations outside of the current catchments were considered as potential locations. In this way the extension constrains the interchange such that the catchments of the candidate locations do not spatially intersect (overlap) with catchments of the current set of locations, V. Formally then Equation 1 is *subject to:*

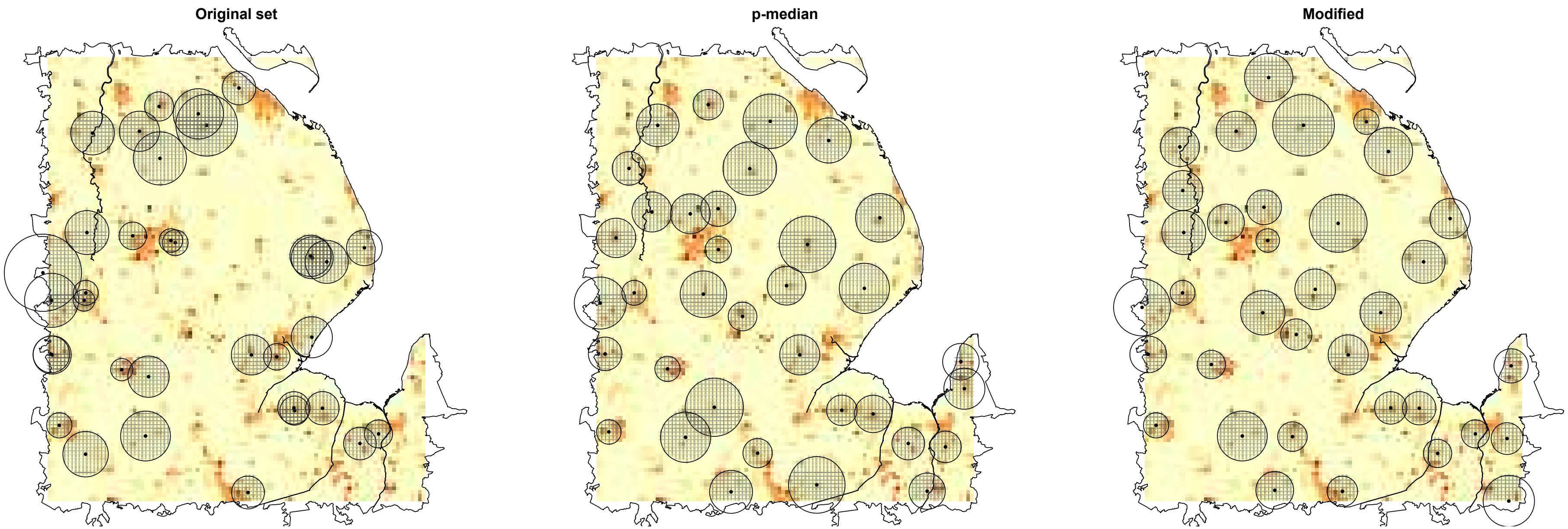
$$x_{ij} \leq y_j; j \cap V = \emptyset \qquad \text{(Equation 2)}$$

4. Results

The AD system and the feedstock ‘recipe’ being considered in this analysis requires 547.5 t yr^{-1} of domestic food waste. Figure 2 shows that this supply variable is the limiting factor in most areas. The p-median modification sought to account for the spatialities of the different feedstocks at each location. The traditional and modified p-median were each run under a Teitz and Bart search with the objective of identifying the best locations for 35 AD facilities from a set of 524 possible locations. At each of the potential locations the catchment needed to supply each feedstock was determined and the total resident population was used to model demand. As a result, the 524 locations have both supply and demand constraints as described in Section 3.1

The classic p-median seeks to optimise supply and demand: it evaluates potential locations by the degree to which they reduce the demand weighted distance between supply and demand. This defines location optimality by the degree to which the most people have to travel the least distance (person kilometres). The modification to the p-median model proposed in this paper also sought to minimise demand weighted distance but also included an additional constraint of only selecting facilities whose catchments do not overlap with the catchments of other selected facility locations. This is important because without it, AD locations identified as optimal would actually be using the same supply resources, which would be impossible. Figure 5 shows the results of applying the modified and original *p*-median models to select 35 (*n*) optimal locations for AD siting from the set of 524 possible locations. Figure 5a shows the starting locations selected by the algorithm and the overlap between some members of this initial subset is evident.

There are significant differences between the results of the two models (Figure 5b and 5c). As expected the classic p-median model minimises the demand weighted distance and takes no account of the spatial overlap of the supply needed for the selected locations (Figure 5b). The modification reduces the catchment overlap to 677 ha (Figure 5c). This residual overlap derives from locations that were randomly selected as part of the initial starting set of locations (‘original set’) and which could not be improved on by swapping under the Teitz and Bart heuristic, a point that is returned to in the discussion. In this way the results show the improvement in the result of the modified algorithm, with no new catchment overlap being introduced as better sites are selected and the cost of the improved spatial footprint of the selected sites and their catchments being minor increases in demand weighted distances.



OL = 130695.4 ha

D_w = 12.1 x 10^9 person km

a)

OL = 11146.0 ha

D_w = 6.8 x 10^9 person km

b)

OL = 677.4 ha

D_w = 7.7 x 10^9 person km

c)

Figure 5. a) the randomly selected original locations, b) locations selected by the standard p-median, and c) those selected by the modified p-median model that considers local feedstock catchments, with the spatial overlap of the results, *OL*, in *hectares* and the total demand weighted distance, D_w, in *person kilometres* arising from each set of locations.

5. Discussion

There are a number of compelling reasons for identifying rural locations that are suitable for the establishment of renewable energy schemes such as community scale AD systems. For example, those living in rural areas typically have higher per capita CO^2 emissions (Commission for Rural Communities 2008, 2010, Fahmy et al. 2011) and higher domestic fuel costs (Smith et al., 2010). Additionally, community scale projects can result in changes in community energy behaviour and attitudes (Hoffman and High-Pippert, 2010; Hielscher et al., 2011). The question of where to locate such facilities will depend on a number of local factors not included in this study: the availability of land, the preparedness of the local community (including farmers) to contribute feedstocks, ease and cost of linking the facility to local properties and the local grid, etc.

A review of the energy, renewables and energy policy literatures suggests that there is sometimes a lack of critical understanding of the operation of different spatial analysis / geo-computational tools. In some instances, the inappropriate use of location-allocation models were found, where the authors appeared to lack a critical understanding of the logics behind the computational operations. For example, Shi et al, 2008 identified optimal biomass plant locations by considering each location and the degree to which it would be served by biomass resources within 100km – that is, each potential site was allocated the resources of each other node within that 100km distance. The locations were then ranked based on the resources available and the best location taken as the first location, followed by the 2nd best, 3rd best etc. and in each case checking to ensure that there is "little or no overlap" in catchments of previously selected locations. This approach plainly runs the risk of selecting sub-optimal sets of *n* locations. Thompson et al. (2013) applied a 'maximum attendance' location-allocation algorithm to select locations that could be served with manure within 22km network distance. However, the resulting locations are driven by demand where is defined on the supply of organic materials with large overlapping service areas for the facilities that were located. No account was taken of the spatial distribution of resources and possible competition for them between selected locations. Whilst

Panichelli and Gnansounou (2008) simply re-allocated any resources that were required by overlapping demands and removed them from further consideration.

In part the origins of problems such as these relate to the availability of very powerful tools and familiar (mapped) outputs available through GIS software. The tools in most commercial GIS software are a 'black box' with the associated computational operations hidden from view and, hence, arguably subject to little critical reflection. This is in contrast to the many open source software and platforms that are available for free (Quantum GIS, various spatial packages in R, etc). Whilst it is encouraging to see the wider use of spatial analysis, geo-computation and the application of GI technologies in many areas of science, it is concerning to sometimes see them used naively: it can be relatively straightforward for those with little geographical or operations-research training to apply very powerful location-allocation tools. The resulting maps are easily generated and provide a very powerful message to those involved in spatial decision-making, planning and policy, but they may have been produced without the operator having a full understanding of the analysis they are undertaking. The clearest evidence of this is when research papers describe the software specific tool they have used and not the generic spatial analysis they have undertaken.

There are a number of areas for future work. In this paper we have simply sought to demonstrate the advantages of a modified p-median model that considers the resources (supply) needed at each potential location. This applied to a simplified case study. It did not consider such things as variations in the actual feedstocks available, variations in transport energy costs for different feedstocks, the main purpose of the facility (heat, electricity etc) or connection to the electricity grid. A number of developments will be included in future. In this analysis, the catchment was defined as a radial distance from the location being considered. Whilst the assumption that straight line distances reflects road distances may hold in many rural areas such as in in this case study, especially where transport of supply materials may be by agricultural vehicle rather than trucks or lorries, this may not be the case in other less rural areas with greater routing complexity. The use of a network analysis to determine distance would provide greater subtly to the analysis in those cases. A Teitz and Bart heuristic was used to search for optimality. This is the most widely applied heuristic and it uses a *1-opt* methodology that swaps members of the facility set for members of the non-facility set. It stops swapping facilities when none of the facility set can be exchanged to improve the evaluation (such as demand weighted distance). However, there is a danger that such algorithms converge on sub-optimal sets of solutions as only one swap is attempted at each iteration. Other work has shown that the selection of the initial set of *n* random locations can severely impact on the optimality of the results, especially in spatially saturated problems. For these reasons, other workers have suggested the use of Genetic Algorithms and Grouping Genetic Algorithms in location-allocation analyses (e.g. Alp et al., 2003; Li and Yeh, 2005; Comber et al., 2010). The key to the success of genetic algorithms is the inclusion of a small random component (mutation) and allowing partially successful solutions to breed. Grouping genetic algorithms also includes some mutation but enables sets of solutions to be identified together. The similarities and differences between GA and GGAs are discussed in Brown and Sumichrast (2003) and both Pitaksringkarn and Taylor (2005) and Xiao et al. (2007) provide a thorough review of algorithmic developments in spatial decision making. A final area of future work will be to modify the way that the model treats the randomly selected initial set of locations. In the results above, the model was unable to swap one of the overlapping initial set as it did not improve the demand weighted distance (i.e. it failed the criterion in the evaluation function). Future work will examine the computational advantages of tweaking the modification such that it seeks to eliminate all overlaps as a first step.

Other research has considered travel times and fuel costs, whereas as this study did not. However, there a number of dangers in using economic costs as a constraint, not least of which is that the costs may change. Ultimately there is a need for renewable energy analyses to consider using net energy costs to determine site locations. Different feedstocks have different transport costs (especially if they are dry or wet). In the future these should logically be reduced to energy costs (kWh) rather than financial costs to provide a more informative and temporally sustainable evidence base upon which to locate energy facilities.

6. Conclusions

In conclusion, the extension to the p-median model proposed in this paper considers the spatial distribution of supply resources and competition for them by potential facility locations. It overcomes the problem of competition for local resources by facilities sited close to each other. As yet, little other research in facility location has included this constraint on site selection. In general, the use of location-allocation models in the renewable energy literature has tended to be focus on the spatial distribution of demand and potential facility locations are evaluated by the degree to which they minimise demand weighted distance. The proposed modification augments the classic the p-median swap operation to consider only those locations outside of the supply resource catchments of the locations being considered.

Acknowledgement

This research was conducted as part of the 'Rural Hybrid Enterprise Energy System' project funded by as part the EPSRC and Department of Science and Technology India 'Bridging the Rural Urban Divide' programme (EPSRC grant ref. EP/J000361/1). All of the statistical analysis and mapping were implemented in R version 3.0.2, the open source statistical software http://cran.r-project.org. The data and code used in this analysis will be provided to interested researchers on request.

References

Adams, P. W., Hammond, G. P., McManus, M. C. and Mezzullo, W. G. (2011). Barriers to and drivers for UK bioenergy development. *Renewable and Sustainable Energy Reviews*, 15: 1217-1227.

Alp, O., Erkut, E. and Drezner Z., (2003). An efficient genetic algorithm for the p-median problem, *Annals of Operations Research, 122,* 21–42

Bania, R. (1993). Fuzziness in geographical information systems: Contributions from the analytic hierarchy process. *International Journal of Geographical Information Science*, 7, 315–329.

Beccali M, Columba P, D'Alberti V and Franzitta V, (2009). Assessment of bioenergy potential in Sicily: A GIS-based support methodology. *Biomass and Bioenergy*, 33:79-87.

Bibby, P. and Shepherd, J. (2004). *Developing a New Classification of Urban and Rural Areas for Policy Purposes - the Methodology.*

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/239084/2001-rural-urban-definition-methodology-technical.pdf [available 11th April 2014].

Boroushaki, S., and Malczewski, J. (2008). Implementing an extension of the analytical hierarchy process using ordered weighted averaging operators with fuzzy quantifiers in ArcGIS. *Computers and Geosciences*, 34, 399–410.

Brown, E.C and Sumichrast, R.T. (2003). Impact of the replacement heuristic in a grouping genetic algorithm. *Computers and Operations Research* 30: 1575–1593

Carver, S. (1991). Integrating multicriteria evaluation with GIS. International Journal of Geographical Information Systems, 5, 321–339.

Church, R., and Velle, C. R. (1974). The maximal covering location problem. *Papers in Regional Science*, *32*(1), 101-118.

Comber, A.J., Brunsdon, C., Hardy, J. and Radburn, R. (2009). Using a GIS–based network analysis and optimisation routines to evaluate service provision: a case study of the UK Post Office *Applied Spatial Analysis and Policy,* 2(1):47-64

Comber, A.J., Proctor, C. and Anthony, S. (2008). The creation of a national agricultural land use dataset: combining pycnophylactic interpolation with dasymetric mapping techniques *Transactions in GIS,* 12(6): 775–791.

Comber, A.J., Sasaki, S., Suzuki, H. and Brunsdon, C., (2011). A modified grouping genetic algorithm to select ambulance site locations. *International Journal of Geographical Information Science*, 25(5): 807–823.

Commission for Rural Communities (2008) *The state of the countryside 2008.* Countryside Agency Publications: Wetherby.

Commission for Rural Communities (2010) *The state of the countryside 2010.* Countryside Agency Publications: Wetherby.

Falkenauer, E. (1998). *Genetic Algorithms and Grouping Problems*. (London: John Wiley and Sons).

Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning* (Reading, MA: Addison-Wesley).

Haddad MA and Anderson PF, (2008). A GIS methodology to identify potential corn stover collection locations. *Biomass and Bioenergy*, 32: 1097-1108

Hakimi, S, (1964), Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph. *Operations Research*, 12:450-459.

Hielscher, S., Seyfang, G., and Smith, A. (2011). *Community innovation for sustainable energy* (No. 2011-03). CSERGE working paper EDM. Paper available from http://www.econstor.eu/bitstream/10419/48797/1/66235690X.pdf

Hoffman, S. M., and High-Pippert, A. (2010). From private lives to collective action: Recruitment and participation incentives for a community energy program. *Energy Policy*, *38*(12), 7567-7574.

Holland, J. H. (1975). *Adaptation in natural and artificial systems* (Michigan: University of Michigan Press).

Jiang, H., and Eastman, R. J. (2000). Application of fuzzy measures in multi-criteria evaluation in GIS. *International Journal of Geographical Information Science*, 14, 173–184.

Li, X. and Yeh, A.G. (2005). Integration of genetic algorithms and GIS for optimal location search. *International Journal of Geographical Information Science, 19,* 581-601.

Ma J, Scott NR, DeGloria SD, Lembo AJ (2005). Sitting analysis of farm-based centralized anaerobic digester systems for distributed generation using GIS. *Biomass Bioenergy,* 28:591–600.

Owen, S. H., and Daskin, M. S. (1998). Strategic facility location: A review. *European Journal of Operational Research*, *111*(3), 423-447.

Panichelli L and Gnansounou E, (2008). GIS-based approach for defining bioenergy facilities location: A case study in Northern Spain based on marginal delivery costs and resources competition between facilities. *Biomass and Bioenergy*, 32: 289-300.

Perry, M. and Rosillo-Calle, F. (2008). Recent trends and future opportunities in UK bioenergy: Maximising biomass penetration in a centralised energy system. *Biomass and Bioenergy*, 32: 688-701.

Pitaksringkarn, L. and Taylor M.A.P, (2005). Grouping Genetic Algorithm In GIS: A Facility Location Modelling, *Journal of the Eastern Asia Society for Transportation Studies*, 6, 2908-2920.

Procter, C., Comber, A., Anthony, S., Lyons, H. and Smith, K., (2005). Spatial data integration: the development of a manure management database for England and Wales. Pp. 565-570 in *Proceedings of the GIS Research UK 13th Annual*

Conference, 6-8 April 2005 (eds Roland Billen, Jane Drummond, David Forrest, Elsa Joao), University of Glasgow

Quested, T., Ingle, R., and Parry, A. (2013). Household food and drink waste in the United Kingdom 2012. *Banbury, UK: Waste and Resources Action Program*. Available from http://www.wrap.org.uk/sites/files/wrap/hhfdw-2012-main.pdf.pdf

Recchia L, Cini E, Corsi S, (2010). Multicriteria analysis to evaluate the energetic reuse of riparian vegetation. Applied Energy 87:310-319

ReVelle, CS and Swain, RW, (1970). Central facilities location. *Geographic Analysis*, 2: 30-42.

Sasaki, S., Comber, A.J., Suzuki, H. and Brunsdon, C., (2010). Using genetic algorithms to optimise current and future health planning - the example of ambulance locations. *International Journal of Health Geographics,* 9: 4. doi:10.1186/1476-072X-9-4

Sasaki, S., Igarashi, K., Fujino, Y., Comber, A.J. and Brunsdon, C., Muleya, C.M. and Suzuki, H. (2011). The impact of community-based outreach immunization services on immunization coverage with GIS network accessibility analysis in peri-urban areas, Zambia. *Journal of Epidemiology and Community Health*, 65: 1171-1178 doi:10.1136/jech.2009.104190

Shi X, Elmore A, Li X, Gorence NJ, Jin H, Zhang X and Wang F (2008) Using spatial information technologies to select sites for biomass power plants: a case study in Guangdong Province, China. *Biomass Bioenergy*, 32: 35–43.

Sliz-Szkliniarz B and Vogt J, (2012). A GIS-based approach for evaluating the potential of biogas production from livestock manure and crops at a regional scale: A case study for the Kujawsko-Pomorskie Voivodeship. *Renewable and Sustainable Energy Reviews* 16: 752–763.

Smith N, Davis A, and Hirsch, D. (2010). *A minimum income standard for rural households: Report to the Commission for Rural Communities and Joseph Rowntree Foundation.* Joseph Rowntree Foundation: York.

Smith, K.A., Brewer, A.J., Crabb, J. and Dauven, A. (2001) A survey of the production and use of animal manures in England and Wales. III. Cattle manures. *Soil Use and Management*, 17 (2), 77-87

Sultana A and Kumar A, (2012). Optimal siting and size of bioenergy facilities using geographic information system, *Applied Energy*, 94: 192-201

Teitz, M. B., and P. Bart (1968), Heuristic methods for estimating generalized vertex median of a weighted graph, *Operations Research*, 16, 955-961.

Thompson E, Wang Q and Li M. (2013). Anaerobic digester systems (ADS) for multiple dairy farms: a GIS analysis for optimal site selection. *Energy Policy.* 61-124

Tobler W R 1979 Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association* 74: 519–36

Toregas, C., Swain, R., ReVelle, C. S. and Bergman, L., 1971. The location of emergency service facilities, *Operations Research* 19, pp.1363-73

Voivontas D, Assimacopoulos D, Koukios EG., 2001. Assessment of biomass potential for power production: a GIS based method. *Biomass and Bioenergy* 20:101–12.

Voogd, H. (1983). *Multi-criteria evaluations for urban and regional planning*. Pion: London.

Ward, A. J., Hobbs, P. J., Holliman, P. J., and Jones, D. L. (2008). Optimisation of the anaerobic digestion of agricultural resources. *Bioresource Technology*, *99*(17), 7928-7940.

Wu, F. (1998). SimLand: A prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules. *International Journal of Geographical Information Science*, 12, 63–82

Xiao, N., Bennett, D.A. and Armstrong M.P. (2007). Interactive evolutionary approaches to multiobjective spatial decision making: A synthetic review. *Computers, Environment and Urban Systems, 31,* 232-252.

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18,183–190.