An integrated model of land-use trade-offs and expanding agricultural processing centres

A. Nazari [a], **S. Penazzi**[b], **A.T. Ernst**[a], **S. Dunstall**[a], **B. Bryan**[c], **J. Connor**[c] **and M. Nolan**[c]

[a]*CSIRO, Melbourne, Australia*
[b]*Department of Industrial Engineering, Bologna University, Italy*
[c]*CSIRO, Adelaide, Australia*
Email: *asef.nazari@csiro.au*

Abstract: Climate change and demand for greener energy alternatives are putting increased pressure on the use of agricultural land for not just for food and fibre production but also biofuels, carbon sequestration, biodiversity and other non-traditional uses. A key question is how this competition might impact on not only future land use but also on the composition of the supply chains that process the products of the land. In this paper we address a major part of this question by considering the location of processing centres alongside land use change in an integrated optimisation model. CSIRO has previously developed a model of land-use trade-offs that considers the possible evolution of agricultural land areas in Australia over the next 40 years. This can be modelled as a large scale multi-stage linear programming problem. Here we consider in addition the construction of some processing centres for bio-fuel, bio-energy, livestock facilities and so forth, which introduces a new combinatorial aspect to the model. The decisions of land use and the location of processing centres are interlinked, because transport costs based on distances are often instrumental in determining the economic viability of some of the land uses and conversely economies of scale are necessary to justify investment in processing plants.

In this paper we introduce a model containing both problems of a land allocation and a facility location simultaneously which results in a large scale mixed integer linear programming (MILP) problem and therefore is computationally difficult to solve. We suggest an algorithm to solve the problem which utilises some decomposition techniques including aggregation and disaggregation, column generation and a concept of clustering. Furthermore, some numerical results are provide to empirically show the computational feasibility of the suggested solution methodology.

Keywords: Mixed integer linear programming, land use management, facility location

1 INTRODUCTION

The objective of CSIRO's Australian Land Use Trade-Offs (LUTO) project is to model how global and local economic and environmental drivers influence future use of Australian land currently in agricultural production. Predicting land use change in Australian agriculture in the context of increasing energy prices stimulating bio-fuels and bio-energy land uses, and a carbon policy with possibilities for a carbon price increasing over time, requires solving a large scale linear programming problem. The model covers the domain of Southern and Eastern Australian land currently in intensive agricultural use, on a one square kilometre grid cell resolution spanning $813,000$ square kilometres in total. Starting with the agricultural land present in 2013, the current LUTO model finds the optimal land use for each grid cell and for each year forward in time until 2050, where optimality is defined as maximizing profit and social welfare (the sum of profit and consumers' surplus). For each year the corresponding LP has $7,313,847$ continuous variables and $814,811$ constraints. The focus is on the change from current agricultural production to alternative land uses such as carbon plantings, environmental plantings, bio-energy or bio-fuels. Food prices are computed endogenously following a maximum welfare approach, whereas costs and revenues for non-agricultural commodities depend on scenarios and exogenous modeling. The platform is built to model a range of scenarios involving alternate assumptions about global climate change, world carbon emissions trajectories, emissions limits and prices of carbon credits, price trajectories for energy, world food demand, supply and price trajectories, and agricultural productivity growth Bryan et al. (2014).

In Nazari et al. (2015) a combination of an aggregation-disaggregation technique with the concept of column generation was developed to solve the large scale LP problem originating from the LUTO project. Based on the properties of the problem, such as similar economical and geographical properties of nearby land parcels, the combination of clustering ideas with column generation to decompose the large problem into smaller sub-problems yields a computationally efficient algorithm for the large scale problem. By drawing inspiration from these facts, instead of considering cells of size one square kilometer, the whole of Australia is divided into big chunks of land (lower resolution) called clusters. The economical properties of each cluster were calculated based upon cells in each cluster. This type of aggregation technique in optimisation is a handy tool to create a set of smaller problems out of a large problem. However, the smaller problem size comes at the cost of lower accuracy, which was remedied by using an iterative disaggregation approach. The smaller problem is gradually made a better approximation of the original problem by breaking down bigger clusters into smaller clusters, and adding new clusters as new columns to the small problem.

There is an increasing level of attention being given to the use of optimisation techniques (as used in LUTO) in agricultural and other land use management planning contexts. As an example, in Weintraub and Cholaky (1991) a hierarchical approach is presented for large-scale forest planning. The algorithm is based on solving an aggregate problem, which is of moderate size. Another example, in Seppelt et al. (2013) authors argue the usage of optimisation techniques in combination with scenario analysis can provide efficient land use management options for sustainable land use from global to sub-global scales. In terms of water resource management, in Gaddis et al. (2014) a spatial optimisation technique implemented among four diffuse source pathways in a mixed-use watershed to maximize total reduction of phosphorus loading to streams while minimizing associated costs. An interesting utilization of a multiobjective optimisation technique is reported for identifying optimum land management adaptations to climate change Klein et al. (2013).

The future factors influencing Australian land-use are significant but uncertain in magnitude and potential effect. Changes in land-use will reflect responses to challenges such as rising energy prices, global warming pertaining to the excessive use of fossil fuel, policies to reduce atmospheric greenhouse gasses including possibilities for incentives to encourage land based carbon sequestration, rising demand for food, and so forth. Results could include increasing shifts from agricultural to forest-based carbon sequestration and bio-fuel production, if rising energy and/or carbon prices make these alternative land uses more profitable than agriculture. In other words, climate change and the demand for greener energy alternatives are putting increased pressure on the use of agricultural land for not just food production but also biofuels, carbon sequestration, biodiversity and other non-traditional uses. Additionally, land use outcomes will depend on technological progress in improving agricultural productivity and how climate influences the productivity of different land uses. A key question is how this is going to impact not only the land use but also the agricultural supply chains that process the outputs of the land use.

As an extension of the LUTO project, the purpose of this paper is to model decisions on the optimal allocation of agricultural and energy related activities coupled with finding the optimal locations for processing centre fa-

cilities for bio-fuel, bio-energy, livestock facilities and so forth. Embedding a facility location problem inside the existing large scale LP problem introduces a new combinatorial aspect to the existing LUTO model. The decisions of land use and the location of processing centres are interlinked because transport costs based on distances are often instrumental in determining the economic viability of some of the land uses and conversely economies of scale are necessary to justify investment in processing plants. We provide a mathematical formulation for this problem of considering land use together with the need to cluster land use around processing centres in order to make harvesting, transport and processing activities economically viable. When instantiated with data pertaining to the entire productive land mass of Australia, this model results in a very large mixed integer (linear) programming (MIP) problem which is computationally difficult to solve. We assess this computational difficulty in terms of the computation time required to find an optimal solution and the number of variables and constraints resulting from use of a whole-of-Australia data set.

The rest of this paper is organized as follows. Section 2 introduces a high level description of the facility layout problem. In Section 3 we give the formal mathematical model of the problem we want to solve. In Section 4 we present some computational difficulties and our approaches to obtain preliminary numerical results. Section 5 will provide concluding remarks.

2 FACILITY LOCATION PROBLEMS

The Facility Location Problem (FLP) is a very well established aspect of the strategic design of supply chains and networks. The FLP is concerned with the optimal placement of facilities to minimize costs while considering some constraints regarding the proximity of facilities to different entities based on the context of the application. Formally, in a FLP a set $N = \{1, \ldots, n\}$ of potential facility locations and a set of clients $I = \{1, \ldots, m\}$ are given. A facility placed at j costs c_j for $j \in N$. Each client has a demand for a certain commodity or good, and the total cost of satisfying the demand of client i from a facility at j is h_{ij}. The optimisation problem is to choose a subset of the locations at which to place facilities and then to assign the clients to these facilities so as to minimize total cost, as described in Nemhauser and Wolsey (1988). In addition to binary variables x_j for $j \in N$, continuous variables y_{ij} are introduced for $j \in N, i \in I$ to represent the fraction of demand of client i that is satisfied from a facility located at j. A MIP representation of the uncapacitated FLP is as follows:

$$\min_{\mathbf{x} \in \mathbb{B}^n, \mathbf{y} \in \mathbb{R}_+^{mn}} \left\{ \sum_{j \in N} c_j x_j + \sum_{i \in I} \sum_{j \in N} h_{ij} y_{ij} \quad \text{s.t.} \quad \sum_{j \in N} y_{ij} = 1, \ \forall i \in I \text{ AND } y_{ij} - x_i \leq 0, \ \forall i \in I, j \in N \right\}$$

The FLP is a generalization of a clustering problem where the aim is to find a set of κ centres so that the average distance of all points from the nearest centre is minimised. The FLP is a generalization because h_{ij} need not be euclidean distances (nor, for example, must the values satisfy a triangle inequality), and because κ is determined by cost optimisation rather than being given as a fixed parameter.

There are very many publications related to examining FLP from combinatorial optimisation perspective. Melo et al. (2009) provides a literature review of the application of FLP in the context of supply chain management. Even more relevant to the research expressed here, Zhang et al. (2011) demonstrates that the financial success of producing biofuel is identifying the optimal location for the facility. The location decision is especially important for woody biomass feedstock owing to the distributed nature of biomass and the significant costs associated with transportation. As an example, Jouzdani et al. (2014) discusses a dynamic dairy facility location and supply chain planning through minimizing the costs of facility location, traffic congestion and transportation of rawprocessed milk and dairy products under demand uncertainty. More instances of the application of FLP within environmetal and supply chain decision making can be found in Treitl and Jammernegg (2014), Zangeneh et al. (2014) and Etemadniaa et al. (2015).

In our application we are interested in a variant of the FLP with the following differences: (a) the maximum distance of a point from its centre is limited, because long transport is not economically viable particularly for biofuel/bioenergy; and (b) a minimum number of points must be served by a facility, to give it the required economies of scale.

3 A MODEL FOR LAND ALLOCATION AND FACILITY LOCATION

To make decisions on building new processing facilities, we will focus on energy-related activities. Namely, we consider four types of processing centres in the categories of bio-energy and bio-fuel each including tree-based and wheat-based products. This list could be easily extended to include any type of processing centre.

Let σ be the maximum distance that a commodity can be transported before it becomes unviable (in general this may depend on the commodity). If we allow a facility location at every cell and define y_{ij} variables for all cells j in a σ radius of i, our model would require an excessive number of variables (many hundreds of millions). Hence we simplify the problem by partitioning the area of Australia into regions R_c for $c \in C$ with the distance between any two cells in the region being at most 2σ. A decision will then be made regarding whether or not a processing facility is going to be built in a region (refer Figure 1). Hence we only need to consider $|C|$ possible facility locations, by four types of facility.

For now, let us consider that there are no construction costs for processing centres (e.g., these will be fully subsidized by the government). The main question is where new processing centres will be built. One criterion to have a particular processing facility in a region (cluster) is to have enough production of a particular type in that region. We define the parameter M to be the amount that should be produced in a region that makes the construction and operation of a processing facility financially viable. As this research is an extension of the LUTO project, we use the same terminology and modelling framework. For a full explanation of the LUTO model without facilities see Nazari et al. (2015). The basic model has variables x_{rj} representing area r being allocated to production $j \in F \cup N$ where F is a set of possible commodities that can be grown in the area, and N is the set of non-food activities. Furthermore variables y_{sj} are used to define a piecewise linear curve for price as a function of supply. Parameters D_{sj}, P_{sj} define this curve while Q_{rj} give the quantity of j that can be produced in cell r. The x_{rj} are allowed to be fractional enabling multiple land uses in a cell, though the solutions to the model generally have few areas with such shared land use.

Here we focus on new items for the model. In this context, a processing centre type is defined as $f \in \mathcal{F} = \{BeWP, BeS, BfWP, BfFS\}$ and a new set of binary variables $z_{cf} \in \{0,1\}$ are defined so that each variable is one if we decide to build a processing facility of type $f \in \mathcal{F}$ in the region $c \in C$. Problem LUTO-EX below, which is expressed in terms of a single time period, represents the complete MIP model for the extension of LUTO to consider facility location. Two sets of new constraints are added to the model. Production threshold constraints prevent the construction of a facility centre if there is not enough production in a cluster. In addition, facility force constraints takes into account that if there is not a particular facility inside a cluster, the land inside the cluster should not be allocated for that particular activity.

Problem LUTO-EX

$$\max \sum_{j \in F} \sum_{s \in S} P_{sj} y_{sj} - \sum_{j \in F} \sum_{r \in R} C_{r,\mathrm{ag}_j} Q_{r,\mathrm{ag}_j} x_{r,\mathrm{ag}} + \sum_{j \in N} \sum_{r \in R} \delta_{rj} Q_{rj} x_{rj}$$

s.t.

$$\sum_{r \in R} Q_{rj} x_{rj} \leq \mathrm{Lim}_j \qquad j \in \{\text{carbon, biofuel, bioenergy}\} \subset N \quad \text{Expansion}$$

$$\sum_{r \in R} Q_{r,\mathrm{epB}} x_{r,\mathrm{epB}} \leq \mathrm{Lim}_{\mathrm{epB}} \qquad\qquad\qquad\qquad\qquad \text{Biodiversity}$$

$$\sum_{j \in J} x_{rj} \leq 1 \qquad\qquad \forall r \in R \qquad\qquad\qquad \text{Land-use}$$

$$\sum_{s \in S} y_{sj} \leq \sum_{r \in R} Q_{r,\mathrm{ag}_j} x_{r,\mathrm{ag}} \quad \forall j \in F \qquad\qquad \text{Supply-Production}$$

$$y_{sj} \leq D_{sj} \qquad\qquad \forall s \in S,\, j \in F \qquad \text{Supply-Demand}$$

$$\sum_{r \in R_c} Q_{rf} x_{rf} \geq M_f z_{cf} \qquad \forall c \in C,\, f \in \mathcal{F} \qquad \text{Production-Threshold}$$

$$x_{rf} \leq \sum_{c \in C} z_{cf} \qquad\qquad \forall f \in \mathcal{F},\, r \in R \qquad \text{Facility-Force}$$

$$z_{cf} \text{ binary} \qquad\qquad \forall c \in C,\, f \in \mathcal{F}$$

$$x_{rj}, y_{si} \geq 0 \qquad\qquad \forall r \in R,\, j \in J,\, s \in S,\, i \in F$$

With resolution set to 1, which is the highest resolution with land parcels of size 1 km^2, there are $|R| = 812,383$ regional active cells. Also, there exist $|S| = 100$ price and demand segmentations, $|J| = 9$ non-food activities and $|F| = 24$ food commodities. The MIP model has $|R| * |J| + |S| * |F| \approx 7,313,847$ continuous variables and $4C$ binary variables, where C is the number of clusters (regions). In addition it has
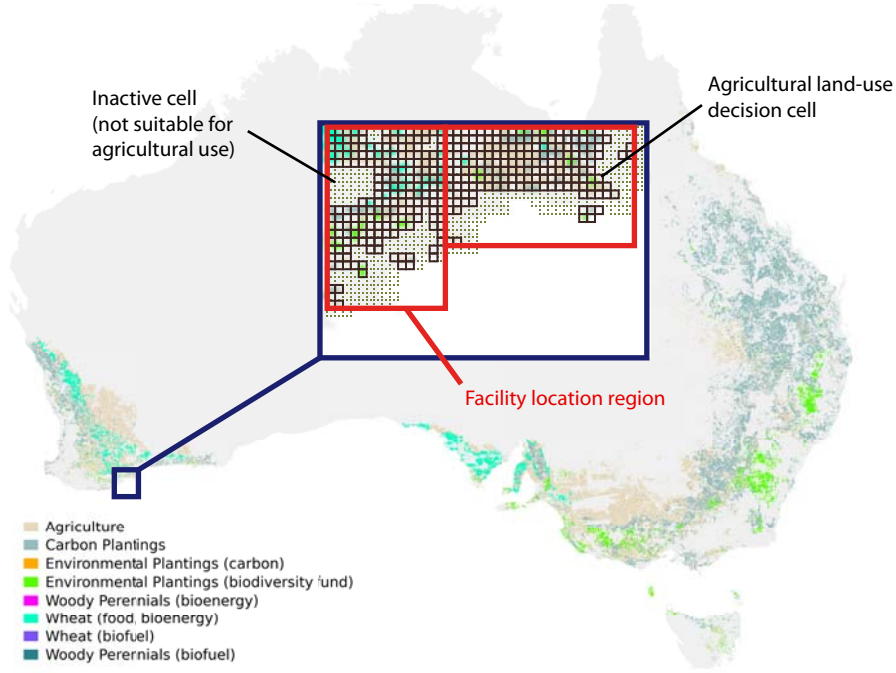
Figure 1. Agricultural areas under consideration for land-use decisions (coloured areas, relating to a LUTO solution for 2050) with an inset showing facility location regions overlaid on the land-use cells. The number of land-use decision cells is much less than the total number of one square kilometre cells for the entire continent, because many cells are considered "inactive" due to not being used for agriculture in 2013.

$|N| + |R| + |F| + |S| * |F| + 4(C + R + 1) \approx 4,064,352 + 4C$ constraints. More detailed information on the LUTO model and its parameters is provided in Nazari et al. (2015).

4 NUMERICAL RESULTS

Considering the way that existing cost, profit and other productivity data of each activity in Australia is managed in a rectangular shape consisting of cells with area of one square kilometre, the division of Australia into C regions (where C is in the order of 10^3) creates large MIP problems. When using a high-end PC for computation we then experienced computer memory issues, this meaning that we could not utilise the IBM CPLEX software package to fully solve the mixed integer linear programs (CPLEX is a well-known and advanced solver for MIP problems). We used two different techniques to reduce the scale of the problem so as to be able to demonstrate some preliminary numerical results.

1. We decided to use two different spatial clustering schemes. In addition to the spatial partitioning that divides Australia into C regions for facility location purposes, we explored the option of aggregating the underlying $1\ km^2$ cells (land parcels) for land-use decisions into $3\ km^2$ (nine cell) or $9\ km^2$ (81 cell) clusters for which a decision of land-use would be made within these regions. The economic and agricultural productivity data for each of these clusters was obtained by the summation or averaging of the underlying data for the cells of the cluster. Using this scheme, we could explore the effect of decreasing the size of R by factors of 1, 9 or 81.

2. We recognize that facility force constraints add a huge number of rows to the model and increase the complexity to the problem. Considering that the regions are designed so that they are not overlapping, instead of having $|\mathcal{F}| * |R|$ constraints, we introduced an alternative set of constraints of size $|\mathcal{F}| * |C|$. By implementing this reduction technique we could make the problem solvable using CPLEX.

$$\sum_{r \in R_c} x_{rf} \leq |R_c| z_{cf} \quad \forall f \in \mathcal{F},\ c \in C$$

For the current purpose of exploring the computational characteristics of the model, regions were defined on a regular square grid with 9, 40 or 81 cells per side. The underlying $1\ km^2$ cells span continental Australia and are surrounded and interspersed with non-agricultural areas for which land-use decisions are not made, thus many facility-location regions have less than the maximum number of "active" cells. In an alternative study where the aim is to directly investigate land-use futures for Australia, regions would be defined on a more complex basis considering topography, transport corridors and other such relevant considerations.

Table 1 represents the dimensions of the problem considering different cluster sizes. In this table, res is the size of smaller clusters (in terms of the number of original $1km^2$ cells per cluster) and Res is the maximum number of original cells in each facility-location region, so that $C \approx R/Res$. After choosing the resolutions for clusters and regions, the table shows number of cells (smaller clusters) and clusters (regions) in that particular model. Also, the number of variables, equations (constraints) and non-zero coefficients are reported for each case. Scripts written in Python 2.7 were used for accessing and manipulating data, and for forming the MIPs ready for solution, and the MIPs were solved using the IBM CPLEX 12.5 solver on a computer running a dual core 64-bit Intel(R) Xeon(R) processor at 2.79 GHz with 64 GB RAM. There is a threshold parameter

Table 1. Problem dimensions considering different clustering

Resolution	num of cells	num of regions	num of vars	num of eqns	num of nonzeros
$res = 1$, $Res = 1600$	812,383	1,405	8,131,850	826,048	35,760,892
$res = 9$, $Res = 81$	138,105	20155	1,444,831	340,015	2,124,949,710
$res = 81$, $Res = 6561$	20,155	415	204,943	43,727	5,412,647

for each type of activities which indicates whether a region would have a processing centre. The threshold in essence is the number of cells (small clusters) allocated to a particular activity. By fixing this parameter to 1, we could find the number of regions in Australia that has at least one cell allocated to that activity. Table 2 is representing this information. The total number of all regions is shown in the second column, and the total number of regions with designated activity is represented in the third column. The last column shows related objective functions. Table 3 summarizes some numerical output of the optimisation process. For these

Table 2. Number of regions with at least one cell allocated to a particular activity

Resolution	num of regions	num centres	obj func value
$res = 1$, $Res = 1600$	1405	510	799,174,282
$res = 9$, $Res = 81$	20155	3642	347,267,145
$res = 81$, $Res = 6561$	415	142	176,433,831

results we fixed the facility threshold to 300, that is, a decision of facility construction will be made if in a region more than 300 original $1\ km^2$ cells are assigned to a particular activity. The second column contains the amount of time in seconds spent in finding an optimal solution using CPLEX, and in the third column, we record the objective function values. In the last column the number of facilities is reported to process related products in the designated regions. All the solutins are obtained in the root node. An interesting reality is understood from the amount of objective functions in the third column. As mentioned before, we suppose that the facility construction cost is fully subsidized by the government, therefore, it does not have any impact in the objective function of the model. However, different value for the objective functions show that different clustering schemes and partitioning Australia into regions with different topologies would yield different objective function values. The optimal value of the LP $res = 1$ without facility location problem is $799, 175, 395$.

5 CONCLUDING REMARKS

In this paper we introduce a model containing both problems of a land allocation and a facility location simultaneously, which results in a large scale mixed integer linear programming problem. The original model is too complicated to be able to solve even on a state-of-the-art computer. Some reduction and aggregation techniques are used in this paper to obtain preliminary numerical results. However, the original problem needs more sophisticated treatments to be solved efficiently. The structure of this problem suggests using specific decomposition techniques including column generation and benders cuts in ongoing future work.

Table 3. Optimisation results

Problem	time (sec)	obj func value	num of facilities
$res = 1,\quad Res = 1600$	3027	798,027,812	88
$res = 9,\quad Res = 81$	229	347,168,574	2606
$res = 81,\ Res = 6561$	10	176,433,704	138

REFERENCES

Bryan, B., M. Nolan, T. Harwood, J. Connor, J. Navarro-Garcia, D. King, D. Summers, D. Newth, Y. Cai, N. Grigg, I. Harman, N. Crossman, M. Grundy, J. Fin-Niganj, S. Ferrier, K. Williams, K. Wilson, E. Law, and S. Hatfield-Dodds (2014). Supply of carbon sequestration and biodiversity services from australias agricultural land under global change. *Global Environmental Change 28*, 166–181.

Etemadniaa, H., S. J. Goetza, P. Canningc, and M. S. Tavallalid (2015). Optimal wholesale facilities location within the fruit and vegetables supply chain with bimodal transportation options: An lp-mip heuristic approach. *European Journal of Operational Research 244(2)*, 648–661.

Gaddis, E., A. Voinov, R. Seppelt, and D. Rizzo (2014). Spatial optimization of best management practices to attain water quality targets. *Water Resources Management 28(6)*, 1485–1499.

Jouzdani, J., S. J. Sadjadi, and F. M. (2014). Dynamic dairy facility location and supply chain planning under traffic congestion and demand uncertainty: A case study of tehran. *Applied Mathematical Modelling 37*(18-19), 8467–8483.

Klein, T., A. Holzkmper, P. Calanca, R. Seppelt, and J. Fuhrer (2013). Adapting agricultural land management to climate change: a regional multi-objective optimization approach. *Landscape Ecology 28(10)*, 2029–2047.

Melo, M., S. Nickel, and F. S. da Gama (2009). Facility location and supply chain management a review. *European Journal of Operational Research 196*(2), 401 – 412.

Nazari, A., A. Ernst, S. Dunstall, B. Bryan, J. Connor, M. Nolan, and F. Stock (2015). Combined aggregation and column generation for land-use trade-off optimisation. *R. Denzer et al. (Eds.): ISESS 2015, IFIP AICT 448 29*, 455–466.

Nemhauser, G. and L. A. Wolsey (1988). *Integer and Combinatorial Optimization*. Wiley and Sons.

Seppelt, R., S. Lautenbach, and M. Volk (2013). Identifying trade-offs between ecosystem services, land use, and biodiversity: a plea for combining scenario analysis and optimization on different spatial scales. *Forest Science 5(5)*, 458–463.

Treitl, S. and W. Jammernegg (2014). Facility location decisions with environmental considerations: a case study from the petrochemical industry. *Journal of Business Economics 84*(5), 639–664.

Weintraub, A. and A. Cholaky (1991). A hierarchical approach to forest planning. *Forest Science 37(2)*, 439–460.

Zangeneh, M., P. Nielsen, A. Akram, and A. Keyhani (2014). Agricultural service center location problem: Concept and a mcdm solution approach. In B. Grabot, B. Vallespir, S. Gomes, A. Bouras, and D. Kiritsis (Eds.), *Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World*, Volume 439 of *IFIP Advances in Information and Communication Technology*, pp. 611–617. Springer Berlin Heidelberg.

Zhang, F., D. M. Johnson, and J. W. Sutherland (2011). A gis-based method for identifying the optimal location for a facility to convert forest biomass to biofuel. *Biomass and Bioenergy 35*(9), 3951 – 3961.