A Constrained K-Means and Nearest Neighbor Approach for Route Optimization in the Bale Collection Problem

David Zamar [1] Bhushan Gopaluni [1] Shahab Sokhansanj [1,2]

[1] Department of Chemical and Biological Engineering, University of British Columbia

[2] Resource and Engineering Systems Group, Environmental Sciences Division, Oak Ridge National Laboratory, TN

20th World Congress of the International Federation of Automatic Control, 9-14 July 2017

Problem Description and Motivation

- The agricultural industry is now able to collect comprehensive real-time data regarding field operations.
- Novel algorithms and methodologies are needed to make proper use of this data.
- These techniques can be used to improve the coordination of agricultural machines and vehicles.

Problem Description and Motivation

- The agricultural industry is now able to collect comprehensive real-time data regarding field operations.
- Novel algorithms and methodologies are needed to make proper use of this data.
- These techniques can be used to improve the coordination of agricultural machines and vehicles.

Problem Description and Motivation

- The agricultural industry is now able to collect comprehensive real-time data regarding field operations.
- Novel algorithms and methodologies are needed to make proper use of this data.
- These techniques can be used to improve the coordination of agricultural machines and vehicles.

Problem Description and Motivation

- Major field operations require the planned coordination of various farm equipment.
- The bale collection problem (BCP) appears after harvest and baling operations of a crop and consists of defining the sequence in which bales spread over a field are collected.

- Major field operations require the planned coordination of various farm equipment.
- The bale collection problem (BCP) appears after harvest and baling operations of a crop and consists of defining the sequence in which bales spread over a field are collected.

Problem Description and Motivation

- The BCP is concerned with the collaborative operation of loaders and wagons.
- Planned management is required to coordinate the tasks efficiently.
- Collection strategy is typically based on the skills and experience of the operator.
- Inconsistent and subjective nature of decisions based on operator judgment tend to produce suboptimal solutions (Milkman et al., 2009).

- The BCP is concerned with the collaborative operation of loaders and wagons.
- Planned management is required to coordinate the tasks efficiently.
- Collection strategy is typically based on the skills and experience of the operator.
- Inconsistent and subjective nature of decisions based on operator judgment tend to produce suboptimal solutions (Milkman et al., 2009).

Problem Description and Motivation

- The BCP is concerned with the collaborative operation of loaders and wagons.
- Planned management is required to coordinate the tasks efficiently.
- Collection strategy is typically based on the skills and experience of the operator.
- Inconsistent and subjective nature of decisions based on operator judgment tend to produce suboptimal solutions (Milkman et al., 2009).

- The BCP is concerned with the collaborative operation of loaders and wagons.
- Planned management is required to coordinate the tasks efficiently.
- Collection strategy is typically based on the skills and experience of the operator.
- Inconsistent and subjective nature of decisions based on operator judgment tend to produce suboptimal solutions (Milkman et al., 2009).

Problem Description and Motivation

- More efficient bale collection plans are achievable.
- Utilize geo-positioning technology that make it possible to know the exact location of bales and machinery on the field (Amiama et al., 2008).
- We develop optimization models of the BCP, which can be easily integrated into farm management decision support systems.

- More efficient bale collection plans are achievable.
- Utilize geo-positioning technology that make it possible to know the exact location of bales and machinery on the field (Amiama et al., 2008).
- We develop optimization models of the BCP, which can be easily integrated into farm management decision support systems.

- More efficient bale collection plans are achievable.
- Utilize geo-positioning technology that make it possible to know the exact location of bales and machinery on the field (Amiama et al., 2008).
- We develop optimization models of the BCP, which can be easily integrated into farm management decision support systems.

- Solving the BCP involves identifying:
  1. The optimal number and locations of roadside storage sites.
  2. The bale collection routes that minimize the total travel distance.

Part I: Constrained Cluster Analysis of Bales

- The identification of roadside storage sites where bales are to be transported can be expressed as a cluster analysis problem.
- The aim is to partition the bales into k clusters in which each bale belongs to the cluster with the nearest mean, resulting in a partition of the bales on the field.

- The identification of roadside storage sites where bales are to be transported can be expressed as a cluster analysis problem.
- The aim is to partition the bales into k clusters in which each bale belongs to the cluster with the nearest mean, resulting in a partition of the bales on the field.

- If the location of the cluster centers were not constrained to lie on the roadside, then the k-means algorithm (Hartigan, 1975), may be directly applied.
- Even so, the problem is NP-hard.
- The fact that the storage sites are constrained to lie on the roadside makes this part of the BCP an even greater challenge to solve.

- If the location of the cluster centers were not constrained to lie on the roadside, then the k-means algorithm (Hartigan, 1975), may be directly applied.
- Even so, the problem is NP-hard.
- The fact that the storage sites are constrained to lie on the roadside makes this part of the BCP an even greater challenge to solve.

- If the location of the cluster centers were not constrained to lie on the roadside, then the k-means algorithm (Hartigan, 1975), may be directly applied.
- Even so, the problem is NP-hard.
- The fact that the storage sites are constrained to lie on the roadside makes this part of the BCP an even greater challenge to solve.

- Given a set of points $\mathcal{X} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n})$, where each point is a d-dimensional real vector, k-means clustering aims to partition these n points into $k \leq n$ sets $S = \{S_1, S_2, \ldots, S_k\}$ so as to minimize the within-cluster sum of squares.
- In the BCP, each point x_i represents a bale's 2-dimensional field location coordinates.

- Given a set of points $\mathcal{X} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n})$, where each point is a d-dimensional real vector, k-means clustering aims to partition these n points into $k \leq n$ sets $S = \{S_1, S_2, \ldots, S_k\}$ so as to minimize the within-cluster sum of squares.
- In the BCP, each point $\boldsymbol{x_i}$ represents a bale's 2-dimensional field location coordinates.

- We also have m compact sets in $\mathbb{R}^2$ denoted $\{T_1, \ldots, T_m\}$, such that the valid positions for the cluster centers are in $T = T_1 \cup \cdots \cup T_m$.
- Define the auxiliary functions:

$$g_j(x) = \min_{t \in T_j} ||x - t||^2, \;\; j = 1, \ldots, m$$

that calculate the minimum Euclidean distance between a given point, x, and any point in the set T_j.

- We also have m compact sets in $\mathbb{R}^2$ denoted $\{T_1, \ldots, T_m\}$, such that the valid positions for the cluster centers are in $T = T_1 \cup \cdots \cup T_m$.
- Define the auxiliary functions:

$$g_j(\boldsymbol{x}) = \min_{\boldsymbol{t} \in T_j} ||\boldsymbol{x} - \boldsymbol{t}||^2, \;\; j = 1, \ldots, m$$

that calculate the minimum Euclidean distance between a given point, $\boldsymbol{x}$, and any point in the set T_j.

$$\underset{(\boldsymbol{\mu_1},S_1),\ldots,(\boldsymbol{\mu_k},S_k)}{\text{minimize}} \quad \sum_{i=1}^{k}\sum_{\boldsymbol{x}\in S_i} ||\boldsymbol{x}-\boldsymbol{u_i}||^2 \tag{1a}$$

$$\text{subject to} \quad \prod_{j=1}^{m} g_j(\boldsymbol{u_i}) = 0 \quad \text{for} \;\; i = 1,\ldots,k \tag{1b}$$

$$S_1 \cup S_2 \cup \cdots \cup S_k = S \tag{1c}$$

$$S_i \cap S_j = \emptyset \quad \forall i \neq j. \tag{1d}$$

Part I: Constrained Cluster Analysis of Bales

Optimization Algorithm

Step 1 Choose k random points from S to be the initial cluster centers, $\boldsymbol{u} = (\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k})$. The relaxed solution of the k-means algorithm is a good initial starting point.

Step 2 For $i = 1, \ldots, k$, assign points to clusters based on their Euclidean distance to the cluster centers:

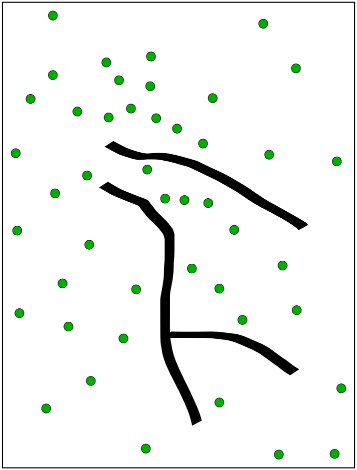
$$S_i = \left\{ \boldsymbol{x} \in \mathcal{X} : ||\boldsymbol{x} - \boldsymbol{u_i}||^2 < ||\boldsymbol{x} - \boldsymbol{u_j}||^2, \;\; j \neq i \right\}.$$

Step 3 For $i = 1, \ldots, k$, update the cluster centers by solving the following minimization problem:

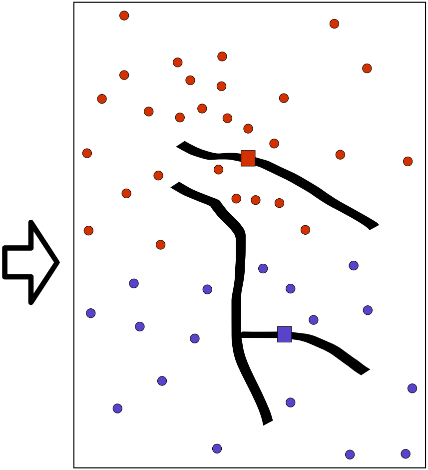
$$\boldsymbol{u_i} = \arg\min_{\boldsymbol{v}} \left\{ f_i(\boldsymbol{v}) = \sum_{\boldsymbol{x} \in S_i} ||\boldsymbol{x} - \boldsymbol{v}||^2 + \gamma \cdot g_i(\boldsymbol{v}) \right\}$$

Step 4 Repeat Steps 2 and 3 until there is no significant change in the clustering criteria, $\sum_{i=1}^{k} f_i(\boldsymbol{u_i})$.

Example of a constrained cluster analysis problem



(a) Road network and bales



(b) Solution

- This part of the BCP may be formulated as a graph theory problem.
- For a given cluster, S_α, $\alpha = 1, \ldots, k$, let $G = (N_\alpha, A_\alpha)$ be an undirected graph, where N_α is the set of nodes (bales) and A_α is the set of edges.
- $N_\alpha = \{0, 1, \ldots, n_\alpha\}$ is an index set for cluster S_α, containing n_α bales and a roadside storage node, denoted by 0.
- $A_\alpha = \{(i, j) \mid i, j \in N_\alpha;\ i < j\}$ represents the set of $(n_\alpha + 1)(n_\alpha + 2)/2$ existing edges connecting the n_α bales and the storage site.

- This part of the BCP may be formulated as a graph theory problem.
- For a given cluster, S_α, $\alpha = 1, \ldots, k$, let $G = (N_\alpha, A_\alpha)$ be an undirected graph, where N_α is the set of nodes (bales) and A_α is the set of edges.
- $N_\alpha = \{0, 1, \ldots, n_\alpha\}$ is an index set for cluster S_α, containing n_α bales and a roadside storage node, denoted by 0.
- $A_\alpha = \{(i, j) \mid i, j \in N_\alpha;\ i < j\}$ represents the set of $(n_\alpha + 1)(n_\alpha + 2)/2$ existing edges connecting the n_α bales and the storage site.

- This part of the BCP may be formulated as a graph theory problem.
- For a given cluster, S_α, $\alpha = 1, \dots, k$, let $G = (N_\alpha, A_\alpha)$ be an undirected graph, where N_α is the set of nodes (bales) and A_α is the set of edges.
- $N_\alpha = \{0, 1, \dots, n_\alpha\}$ is an index set for cluster S_α, containing n_α bales and a roadside storage node, denoted by 0.
- $A_\alpha = \{(i, j) \mid i, j \in N_\alpha;\; i < j\}$ represents the set of $(n_\alpha + 1)(n_\alpha + 2)/2$ existing edges connecting the n_α bales and the storage site.

- This part of the BCP may be formulated as a graph theory problem.
- For a given cluster, S_α, $\alpha = 1, \ldots, k$, let $G = (N_\alpha, A_\alpha)$ be an undirected graph, where N_α is the set of nodes (bales) and A_α is the set of edges.
- $N_\alpha = \{0, 1, \ldots, n_\alpha\}$ is an index set for cluster S_α, containing n_α bales and a roadside storage node, denoted by 0.
- $A_\alpha = \{(i,j) \mid i,j \in N_\alpha;\ i < j\}$ represents the set of $(n_\alpha + 1)(n_\alpha + 2)/2$ existing edges connecting the n_α bales and the storage site.

- The problem is to determine the set of routes that minimize the total travel distance within each cluster identified by the CCA algorithm.
- May be represented as a capacitated vehicle routing problem (CVRP).

- The problem is to determine the set of routes that minimize the total travel distance within each cluster identified by the CCA algorithm.
- May be represented as a capacitated vehicle routing problem (CVRP).

- A weight, q_i, is assigned to each bale i, $1 \leq i \leq n_\alpha$ ($q_0 = 0$).
- Each edge has an associated cost, $c_{ij} > 0$, of sending a vehicle from node i to node j.
- The c_{ij} are assumed to be symmetric and proportional to the Euclidean distance, d_{ij}, between any two nodes, thus $c_{ij} = c_{ji} \propto d_{ij}$, $i, j \in N_\alpha$.
- The collection activities are to be implemented by a fleet of v vehicles, $v \geq 1$, with equal capacity, $\kappa \geq \max\{q_i \mid 1 \leq i \leq n_\alpha\}$.

- A weight, q_i, is assigned to each bale i, $1 \leq i \leq n_\alpha$ ($q_0 = 0$).
- Each edge has an associated cost, $c_{ij} > 0$, of sending a vehicle from node i to node j.
- The c_{ij} are assumed to be symmetric and proportional to the Euclidean distance, d_{ij}, between any two nodes, thus $c_{ij} = c_{ji} \propto d_{ij}$, $i, j \in N_\alpha$.
- The collection activities are to be implemented by a fleet of v vehicles, $v \geq 1$, with equal capacity, $\kappa \geq \max\{q_i \mid 1 \leq i \leq n_\alpha\}$.

- A weight, q_i, is assigned to each bale i, $1 \leq i \leq n_\alpha$ ($q_0 = 0$).
- Each edge has an associated cost, $c_{ij} > 0$, of sending a vehicle from node i to node j.
- The c_{ij} are assumed to be symmetric and proportional to the Euclidean distance, d_{ij}, between any two nodes, thus $c_{ij} = c_{ji} \propto d_{ij}$, $i, j \in N_\alpha$.
- The collection activities are to be implemented by a fleet of v vehicles, $v \geq 1$, with equal capacity, $\kappa \geq \max\{q_i \mid 1 \leq i \leq n_\alpha\}$.

- A weight, q_i, is assigned to each bale i, $1 \leq i \leq n_\alpha$ ($q_0 = 0$).
- Each edge has an associated cost, $c_{ij} > 0$, of sending a vehicle from node i to node j.
- The c_{ij} are assumed to be symmetric and proportional to the Euclidean distance, d_{ij}, between any two nodes, thus $c_{ij} = c_{ji} \propto d_{ij}$, $i, j \in N_\alpha$.
- The collection activities are to be implemented by a fleet of v vehicles, $v \geq 1$, with equal capacity, $\kappa \geq \max\{q_i \mid 1 \leq i \leq n_\alpha\}$.

The following additional constraints are associated with the problem:

1) all routes begin and end at the roadside storage node
2) no two routes visit the same bale
3) all bales are visited exactly once
4) no vehicle can be loaded exceeding its maximum capacity

The following additional constraints are associated with the problem:

1) all routes begin and end at the roadside storage node
2) no two routes visit the same bale
3) all bales are visited exactly once
4) no vehicle can be loaded exceeding its maximum capacity

The following additional constraints are associated with the problem:

1) all routes begin and end at the roadside storage node
2) no two routes visit the same bale
3) all bales are visited exactly once
4) no vehicle can be loaded exceeding its maximum capacity

The following additional constraints are associated with the problem:

1) all routes begin and end at the roadside storage node
2) no two routes visit the same bale
3) all bales are visited exactly once
4) no vehicle can be loaded exceeding its maximum capacity

The following additional constraints are associated with the problem:

1) all routes begin and end at the roadside storage node
2) no two routes visit the same bale
3) all bales are visited exactly once
4) no vehicle can be loaded exceeding its maximum capacity

- Decision vector is $\boldsymbol{x} = (x_{ijr})$, where $i, j \in N_\alpha$, and $r \in R = \{1, 2, \ldots, \tau\}$
- $\tau = \lceil n_\alpha / \kappa \rceil$ is the number of routes needed in order to pick up all of the n_α bales:

$$x_{ijr} = \begin{cases} 1 & \text{if route } r \text{ contains edge } (i,j) \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

- Decision vector is $\boldsymbol{x} = (x_{ijr})$, where $i, j \in N_\alpha$, and $r \in R = \{1, 2, \ldots, \tau\}$
- $\tau = \lceil n_\alpha / \kappa \rceil$ is the number of routes needed in order to pick up all of the n_α bales:

$$x_{ijr} = \begin{cases} 1 & \text{if route } r \text{ contains edge } (i, j) \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

$$\underset{\boldsymbol{u}}{\text{minimize}} \quad \sum_{r\in R}^{t} \sum_{(i,j)\in A_\alpha} c_{ij}x_{ijr} \tag{3a}$$

$$\text{subject to} \quad \sum_{r\in R} \sum_{j\in N_\alpha} x_{ijr} = 1 \qquad \forall i \in N_\alpha \tag{3b}$$

$$\sum_{i\in N_\alpha} \sum_{j\in N_\alpha} x_{ijr} \times q_j \leq \kappa \qquad \forall r \in R \tag{3c}$$

$$\sum_{j\in N_\alpha} x_{0jr} = 1 \qquad \forall r \in R \tag{3d}$$

$$\sum_{i\in N_\alpha} x_{ijr} = \sum_{i\in N_\alpha} x_{jir} \qquad \forall j \in N_\alpha,\ r \in R \tag{3e}$$

Min-min min-max route optimization algorithm (MMROA)

Step 1 Set $k = min(\kappa, |N_\alpha| - 1)$, where κ is the capacity of the wagon . Compute the set of $k-1$ nearest neighbors for each bale $b \in N_\alpha$. Denote the set of $k-1$ nearest neighbors of bale b as Q_b.

Step 2 Define the function

$$M(x, S_\alpha) = \begin{cases} 1 & \text{if } x \in S_\alpha \\ 0 & \text{otherwise} \end{cases}$$

For each bale, $b \in N_\alpha$, compute

$$m_b = \sum_{i \in N_\alpha} M(b, Q_i)$$

Let $b_1^* = \arg\min_b \{m_b : b \in N_\alpha\}$
and $b_2^* = \arg\max_b \{m_b : b \in N_\alpha\}$.
Randomly choose b_1^* or b_2^* with equal probability. Denote the chosen bale as b^*.

Step 3 Among the sets

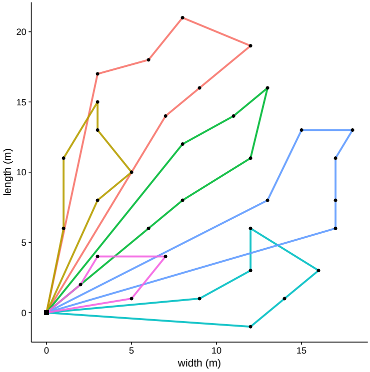
$$\mathcal{Q} = \{Q_b \mid b^* \in Q_b, b \in N_\alpha\},$$

which contain b^* select the one that has the shortest cycle, including the roadside storage node, and denote this set as Q^*.

Step 4 Update the set N_α by:

$$N_\alpha = N_\alpha \setminus Q^*.$$

If $N_\alpha \neq \emptyset$, then go to Step 1.



Obtained tours by the MMROA for a problem previously proposed by Grisso et al. (2007). MMROA yields an additional 6.8% reduction in the total travel distance.

Study Problem

Description

- We consider a study area of size 4800 m $\times$ 4800 m that is composed of 9 equal-sized sections of 256 ha each.
- Bales are spatially distributed across the field according to a Poisson process and assuming a uniform yield.
- Four different wagon capacities are considered (8, 15, 30, and 70).
- Based on an analysis of the within-cluster sum of squares, the bales have been divided into five clusters.

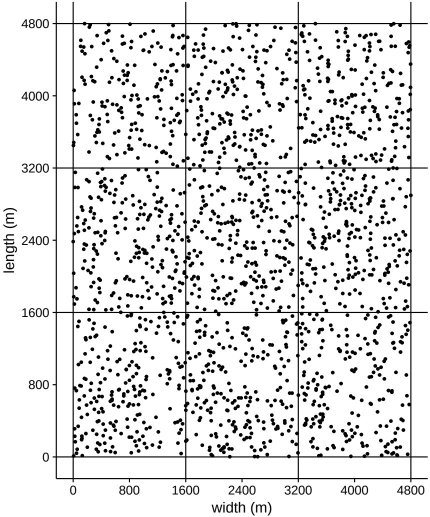
- We consider a study area of size 4800 m $\times$ 4800 m that is composed of 9 equal-sized sections of 256 ha each.
- Bales are spatially distributed across the field according to a Poisson process and assuming a uniform yield.
- Four different wagon capacities are considered (8, 15, 30, and 70).
- Based on an analysis of the within-cluster sum of squares, the bales have been divided into five clusters.

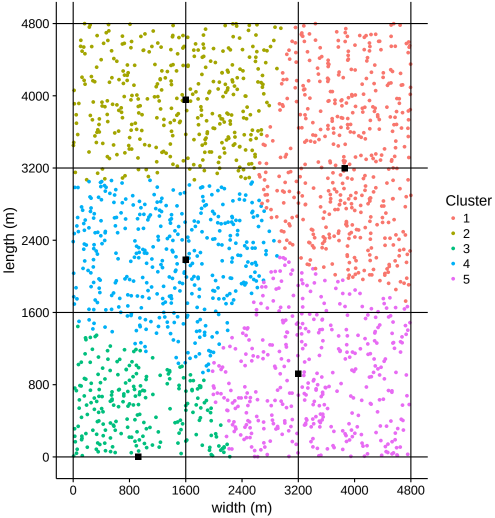
- We consider a study area of size 4800 m $\times$ 4800 m that is composed of 9 equal-sized sections of 256 ha each.
- Bales are spatially distributed across the field according to a Poisson process and assuming a uniform yield.
- Four different wagon capacities are considered (8, 15, 30, and 70).
- Based on an analysis of the within-cluster sum of squares, the bales have been divided into five clusters.

- We consider a study area of size 4800 m $\times$ 4800 m that is composed of 9 equal-sized sections of 256 ha each.
- Bales are spatially distributed across the field according to a Poisson process and assuming a uniform yield.
- Four different wagon capacities are considered (8, 15, 30, and 70).
- Based on an analysis of the within-cluster sum of squares, the bales have been divided into five clusters.

Study Problem

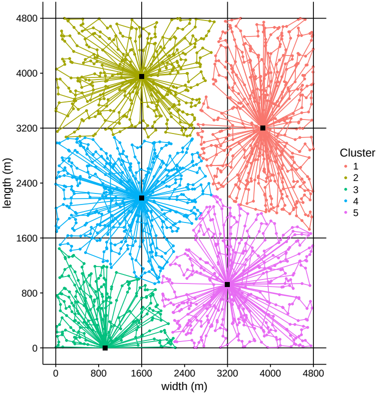
Description



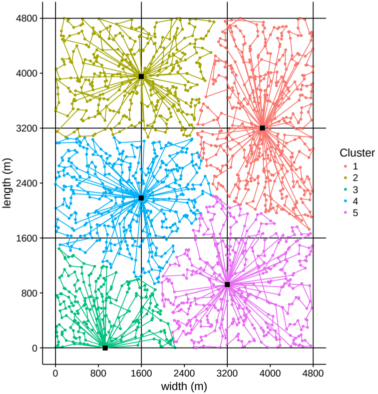

length (m)
4800
4000
3200
2400
1600
800
0
0
800
1600
2400
3200
4000
4800
width (m)
Cluster
1
2
3
4
5

Study Problem

MMROA solutions for wagon capacities of (a) 8 and (b) 15 bales

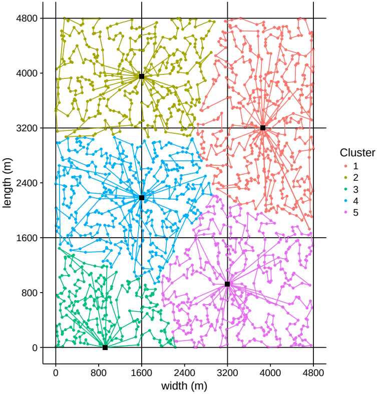


(a) Wagon capacity of 8 bales.

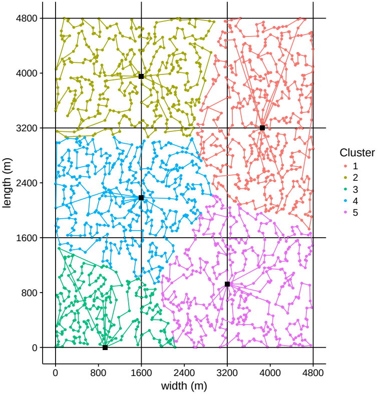


(b) Wagon capacity of 15 bales.

MMROA solutions for wagon capacities of (a) 35 and (b) 70 bales



(c) Wagon capacity of 35 bales.



(d) Wagon capacity of 70 bales.

MMROA Solutions

Wagon Capacity (bales)	Number of Routes	Distance (m)
8	229	525,692
15	124	345,755
35	55	225,206
70	30	184,546

- The potential benefits of our approach to solving the BCP are its scalability and ease of implementation.
- Developed a constrained k-means algorithm and nearest neighbor approach to the BCP, which minimizes travel distance and hence fuel consumption.
- Able to tackle large problems and can be easily incorporated into existing bale collection operations.

- The potential benefits of our approach to solving the BCP are its scalability and ease of implementation.
- Developed a constrained k-means algorithm and nearest neighbor approach to the BCP, which minimizes travel distance and hence fuel consumption.
- Able to tackle large problems and can be easily incorporated into existing bale collection operations.

- The potential benefits of our approach to solving the BCP are its scalability and ease of implementation.
- Developed a constrained k-means algorithm and nearest neighbor approach to the BCP, which minimizes travel distance and hence fuel consumption.
- Able to tackle large problems and can be easily incorporated into existing bale collection operations.

References

C. Amiama, J. Bueno, C. J. Álvarez, and J. M. Pereira. Design and field test of an automatic data acquisition system in a self-propelled forage harvester. *Computers and Electronics in Agriculture*, 61(2):192–200, 2008.

R. D. Grisso, J. S. Cundiff, and D. H. Vaughan. Investigating Machinery Management Parameters with Computer Tools. *ASABE Conf, Paper 071030*, 2007.

J. A. Hartigan. *Clustering algorithms*. John Wiley & Sons, New York, 1975.

K. L. Milkman, D. Chugh, and M. H. Bazerman. How Can Decision Making Be Improved? *Perspectives on Psychological Science*, 4(4):379–383, 2009.