A MILP-based clustering strategy for integrating the operational management of crude oil supply

Leonardo S. Assis[a], Eduardo Camponogara[a,1], Ignacio E. Grossmann[b]

[a]*Department of Automation and Systems Engineering, Federal University of Santa Catarina, Florianópolis, Brazil*
[b]*Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, USA*

Abstract

In this paper, we present an MILP clustering formulation for tackling the operational management of crude oil supply (OMCOS) proposed by de Assis et al. (2019). The OMCOS consists of defining the scheduling of vessels between offshore platforms and a crude oil terminal, combined with the scheduling of operations in a terminal to supply crude oil to distillation columns. The benefits of using the clustering solution as a pre-step before solving the OMCOS are: (a) reduces the number of routes for vessels; (b) simplifies offloading and unloading operations; (c) imposes rules for crude mixtures in clusters of storage tanks that minimize property variations; and (d) produces bounds on crude properties inside storage tanks that are used to linearize bilinear terms in blending constraints. Through the combination of clusters and a MILP-NLP decomposition, near optimal solutions were obtained for a set of representative instances of OMCOS at a reduced computational cost.

Keywords: MILP, Crude Oil Supply, Blending, Bilinear Terms, Clustering

1. Introduction

The main goal of the *Operational Management of Crude Oil Supply* (OMCOS) (de Assis et al., 2019) is to define the schedule of vessel trips and crude oil operations in the terminal in order to deliver to the distillation columns (CDUs) the required feed of crudes within the specification. In this problem (see Figure

*Corresponding Author
Email address: eduardo.camponogara@ufsc.br (Eduardo Camponogara)

1), vessels travel between a crude oil terminal and floating platforms (FPSOs) located in deep water oil fields. After offloading FPSOs, vessels unload crude oil into the storage tanks (STs) located at the terminal. The tanks feed the charging tanks (CTs) in order to achieve a certain crude blend specification, which is then sent to the CDUs.

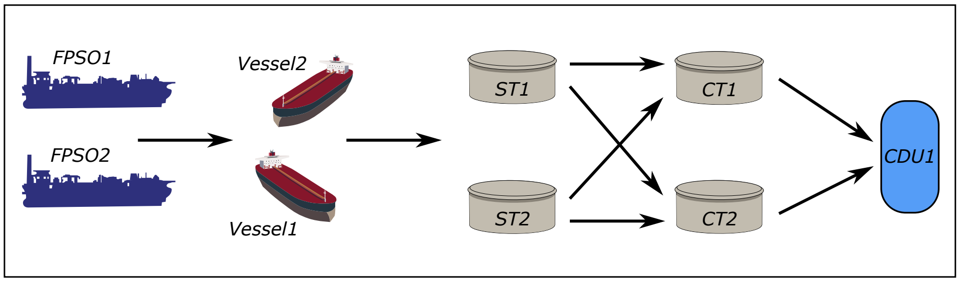


Figure 1: Illustration of the OMCOS. Adapted from de Assis et al. (2019)

Apart from the rules and constraints defined for OMCOS, there are no specific rules on how to perform vessel trips between FPSOs and storage tanks, or how the unloading of crude oil into storage tanks should be performed. This means that vessels are free to travel back and forth all FPSOs and unload their cargo in all storage tanks, leading to a broad range of possible blends in these tanks. As shown in the work of de Assis et al. (2019), this lack of structure creates a highly combinatorial problem, which can be hard to tackle depending on the size of the instance. Additionally, by allowing random mixtures in the storage tanks, crudes with distinct property values may be mixed (e.g., a mixture between crude c_1 with a sulfur content of 0.060 and crude c_2 with 0.010), decreasing the flexibility of operations at the terminal to produce the required blends in the charging tanks.

Ideally, each crude oil arriving at the terminal would have a dedicated storage tank to be unloaded. If this is the case, crudes would only be mixed at the charging tanks in order to produce blends within the specifications required by the CDUs. Nevertheless, when the number of crudes is higher than the number of storage tanks, the pigeonhole principle (Kelly et al., 2017b) suggests that eventually two or more crudes will be mixed in a storage tank. Mixtures may also happen when a vessel needs to unload its cargo in two or more tanks due to the lack of storage capacity in a single one. Finally, a storage tank may be unavailable to receive crudes (i.e., the tank is under maintenance or in operation feeding a charging tank), which forces the vessel to unload in another tank, thereby leading to mixtures.

Since mixtures in the storage tanks seam unavoidable, a mathematical formulation is proposed to define clusters (or groups) of crudes and storage tanks, such that the difference among the property values of the crudes assigned to a cluster of storage tanks is as low as possible. From the solution of the clustering problem, bounds on the crude property can be defined for the storage tanks. These bounds are used to build the relaxation of non-convex blending constraints, resulting in a relaxed MILP formulation, which is then used in a MILP-NLP decomposition scheme. Further, by knowing the crudes that can be grouped or clustered, their origin platforms and the storage tanks they are assigned to, the number of traveling operations of vessels can be limited, which decreases the number of logistic decisions (binary variables) and consequently the complexity of the MILP problem.

The remainder of this work is organized as follows. A review of the literature is presented in Section 2. The problem definition is given in Section 3. The mathematical formulation for clustering crudes is described in Section 4, while the linearization of the blending constraints and the proposed solution strategy is presented in Section 5. Problem instances and computational results are shown in Section 6. Finally, the conclusion and directions for future work are described in Section 7.

2. Literature Review

This section presents a review on the use of clustering models in combination with mathematical optimization strategies for solving Inventory Routing (IRP), Vehicle Routing (VRP), Crude Oil Scheduling (COS) and other problems problems. Regarding IRP and VRP, clusters are typically designed to reduce the complexity of solution strategies for the routing problems. On the other hand, clusters are employed in COS to impose rules on crude mixtures and define groups of tanks to feed CDUs. Nevertheless, few works from the literature consider clusterization for problems that involve COS.

In the VRP literature, Gillett and Miller (1974) were among the first to propose the use of the *cluster-first* and *route-second* strategy. In their work, the solution strategy consists of two sequential steps: (a) define groups of customers according to their polar coordinates and assign vehicles according to capacities; and (b) solve a TSP for each group.

Mathematical programming techniques are also used by Mulvey and Beck (1984) to model the Capacited Clustering Problem (CCP), which has applications in salesforce allocation and VRP. The problem consists in constructing clusters that are as homogeneous as possible (i.e., minimize the sum of distances between each element in a cluster) without violating the capacity of each cluster.

Liu (1999) make use of clusters to tackle the stock location and order-picking problems in a distribution center. In this problem, the goal is to cluster items in the slots of racks and to sequence the picking lists by customers in order to minimize the total travel distance of a picker in the distribution center.

An extension of the (CCP) is the Capacitated Centred Clustering Problem (CCCP) (Negreiros and Palhano, 2006), which does not consider necessarily as center value of a cluster one of the elements' attributes in that cluster. Instead, the center value is defined with respect to all elements of the cluster, which introduces non-linearities into the formulation.

Dondo and Cerdá (2007) tackle a multi-depot VRP with a heterogeneous fleet and time windows. The cluster-based solution strategy is a combination of three sequential steps: (a) identify the set of feasible clusters of customers that are cost-effective; (b) assign clusters to vehicles and sequence them on each tour; and (c) define within a cluster the order of nodes and the schedule of vehicles arrival times at customer locations for each tour.

Ganesh and Narendran (2007) address the VRP with delivery and pick-up nodes. The authors proposed a solution strategy where nodes are first clustered based on proximity, then routes are defined for each cluster of nodes, and finally vehicles are allocated to the routes. Qi et al. (2012) tackle a large-scale VRP with time windows. In this work, clusters of customers are defined based on both their spatial location and temporal information. The authors manage to represent time and space in the same coordinate system, and therefore measure the space-temporal distance between two customers.

Nambirajan et al. (2016) extend the classic IRP formulation by considering replenishment tasks at a central depot and different warehouses in a three echelon supply chain. First, the replenishment schedules from suppliers to a single depot is defined using dynamic programming (DP). Then, the routing of vehicles from the central depot to multiple warehouses is planned using an extension of a three-stage heuristic based on clustering, allocation, and routing (Ramkumar, 2011).

Kelly et al. (2017b) propose an MILP model for defining the assignment of crudes, considering different properties, from external sources to storage tanks in a crude oil terminal. The assignment is done such that the deviation of properties of crudes assigned to a cluster is minimized. Despite considering a large number of crudes and properties, the clustering model does not take into account availability of crudes, flow rate limits, timing, number of storage tanks and storage capacity limits for defining the clusters. Further, Kelly et al. (2017a) discuss how to use their clustering formulation (Kelly et al., 2017b) as part of a pre-scheduling step to reduce the original search space and tackle large scale instances of COS problems.

Cerdá et al. (2018) also make use of a clustering strategy for tackling a COS problem which considers charging tanks, pipelines and CDUs. They proposed a decomposition strategy based on two decision levels. First, an MINLP model is solved to: (a) grouping charging tanks into as many clusters as the number of CDUs, and (b) assign each cluster of charging tanks to feed a unique CDU. Then, an MINLP model for each cluster-CDU pair is solved to defined the scheduling of crude oil operations. The results show for the tested instances a reduced degradation in solution quality and a strong reduction of the computational burden.

When it comes to food grains procurement and their transportation, the use of clusters can play a major role in order to decrease the complexity of resulting optimizations models. Mogale et al. (2019) propose an (MILP) formulation to determine the number and location of procurement centers while minimizing the total supply chain network costs. The first step of the solution strategy consists of using genetic algorithms to group grain suppliers to clusters and then allocate each cluster to a candidate location of procurement center. Then, the (MILP) is solved.

More recently, an extension of the storage assignment and warehouse order-picking problems is proposed by Lee et al. (2020). The solution strategy consists of two steps: clustering and assignment. In the clustering stage, an optimization model to group items balances both travel time and picking delays caused by traffic congestion, and it is solved by evolutionary algorithms. The latter step (assignment) consists of distributing items in each cluster to empty storage locations.

Usually, clustering formulations are tailor-made for problems like IRP, VRP or COS, and cannot be fully applied to an integrated approach as the operational management of crude oil supply (OMCOS).

Therefore, this work proposes an MILP clustering formulation for OMCOS that has the following benefits: (a) decreasing the number of routes available for the vessels; (b) decreasing offloading and unloading operations; and (c) defining rules for crude mixtures in clusters of storage tanks such that the property deviation is minimized. Futher, in order to define the clusters, the proposed MILP formulation takes into account the availability of crudes, flow rate limits, timing, availability of resources (i.e., FPSOs, STs, CTs and CDUs), storage capacity limits and demand satisfaction.

The use of clusters also plays a part in the optimization. After defining clusters, bounds on crude properties inside storage and charging tanks can be inferred. These bounds are used to linearize the bilinear terms in blending constraints, which yields an MILP linearization of the OMCOS MINLP formulation. Through the combination of clusters and an MILP-NLP decomposition strategy, OMCOS is solved for a set of instances, presenting near optimal solutions and reduced computational complexity.

3. Problem Statement

Figure 2 (a) illustrates a problem instance with 7 platforms ($FPSO1$ to $FPSO7$), 7 crudes ($C1$ to $C7$), 5 storage tanks ($ST1$ to $ST5$), 3 charging tanks ($CT1$ to $CT3$) and 2 CDUs ($CDU1$ and $CDU2$). For the sake of simplicity, the operations between resources are not shown (i.e., arrows in Fig. 1).

Figure 2 (b) illustrates the network to be considered for the clustering formulation. Notice that platform-cluster 1 is linked to st-cluster 1, platform-cluster 2 is linked to st-cluster 2, and so forth. Further, all st-clusters are connected to the CT Group (i.e., aggregate of all charging tanks), which is linked to the CDU Group.

As an example, assume that one wants to cluster the set of platforms $FPSO1$ to $FPSO7$, and consequently the set of crudes $C1$ to $C7$, in two groups (i.e., platform-cluster 1 and platform-cluster 2). Similarly, the set of storage tanks $ST1$ to $ST5$ will be clustered in two groups (i.e., st-cluster 1 and st-cluster 2), and therefore, it is possible to define which group of platforms are allowed to feed each group of storage tanks. Each platform-cluster will have a combined production rate, storage and flow rate capacity that correspond to the sum of the rates and capacities of the platforms assigned to the cluster. The same holds for the clusters of storage tanks. Each st-cluster will have a combined capacity and flow rate. As indicated in Figure 2 (b), all charging tanks are grouped into a single CT Group and, in the same manner, the CDUs are integrated in a single CDU Group. The crude oil demand of the CDU Group is the sum of the individual demands of all CDUs. Likewise, the storage and flow rate capacities of the CT Group depend on characteristics of the charging tanks.

The solution of the clustering problem can be seen in Figure 2 (c). The objective of satisfying the demand of the CDU Group, while minimizing the deviation among the crude property values in a cluster drives a solution where platforms $FPSO1$, $FPSO4$, $FPSO6$ and $FPSO7$ are clustered and assigned to the group of storage tanks $ST1$, $ST2$ and $ST3$. As for the remaining, the group of platforms $FPSO2$, $FPSO3$ and $FPSO5$ is assigned to storage tanks $ST4$ and $ST5$.

Note that for the instance presented in Figure 2 (a), if the number of clusters is defined as 1 (one), all storage tanks would be assigned to this cluster, and all platforms would be allowed to feed all storage tanks, leading to the original problem instance. At the other extreme, if the number of clusters is defined as 5 (five), each cluster would have only one storage tank. Thus, the baseline operational management problem of crude oil supply can be reduced to its cluster-based version, meaning that the latter is more general than the former.

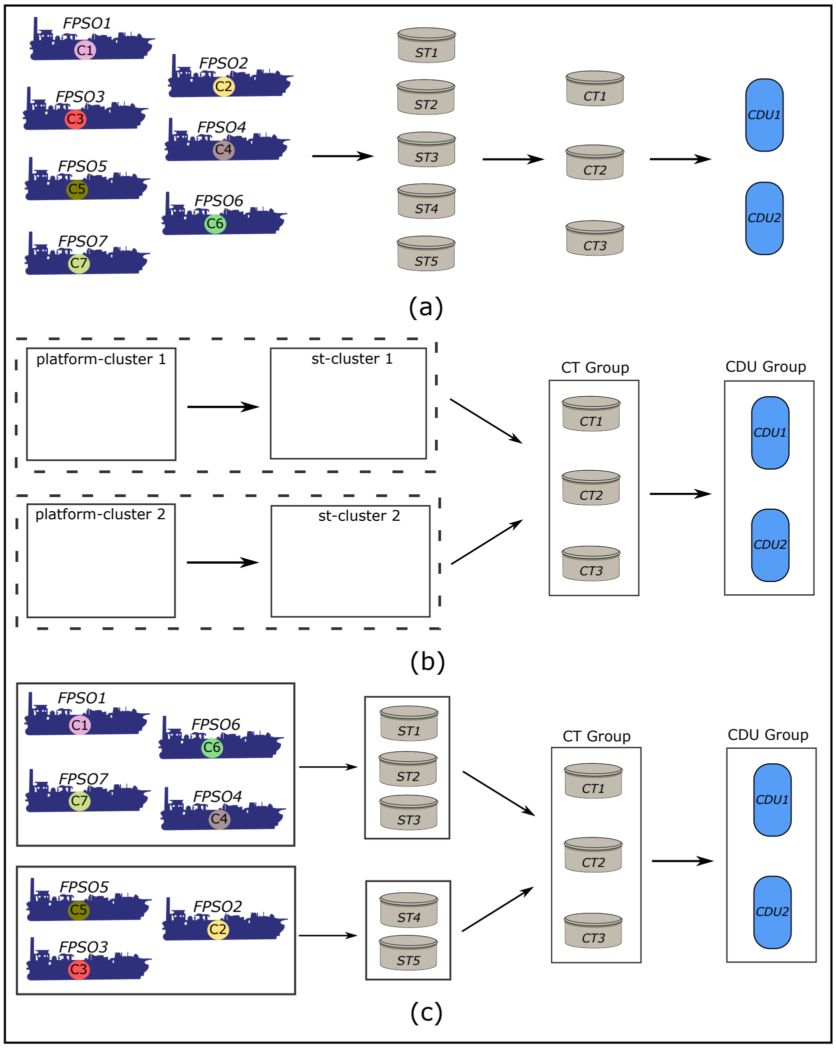


Figure 2: Clustering procedure considering 2 clusters.

Main operational rules can be defined as follows:

(a) a platform must be assigned to a unique platform-cluster. Likewise, a storage tank must be assigned to a unique st-cluster;

(b) a platform-cluster must contain at least one platform. The same holds for the st-cluster. At least one storage tank must belong to a st-cluster;

(c) a st-cluster can perform at most one (i.e., receiving crudes from a platform-cluster or sending crudes to the CT Group) operation during the same time period;

(d) at most one st-cluster can feed the CT Group during the same time period;

(e) a distillation operation (i.e., sending crudes from the CT Group the CDU Group) must be carried out in all time periods.

The optimization problem consists in determining, for the planning horizon, the optimal cluster of platforms, storage tanks, and consequently crude oils, while maximizing the flow of crudes to the CDU Group and minimizing the deviation of crude property values in a st-cluster. To this end, we propose a discrete-time MILP model, whose main decisions consist in selecting the assignments of storage tanks and platforms to clusters, what operations take place at each time, the level of crudes in each cluster and group of resources, and the volumes transferred between clusters and groups of resources.

4. Mathematical Model for Cluster Design

Before presenting the constraints and objective function of the mixed-integer clustering formulation, the required sets, parameters and variables are defined.

1. **Sets**.

    - $\mathcal{T}$. Set of periods. Index i.
    - $\mathcal{RF}$, $\mathcal{RS}$, $\mathcal{RC}$, $\mathcal{RD}$. Set of platforms, storage tanks, charging tanks and CDUs. Index r.
    - $\mathcal{R} = \mathcal{RF} \cup \mathcal{RS} \cup \mathcal{RC} \cup \mathcal{RD}$. Set of all resources. Index r.
    - $\mathcal{RFCS} = \{1, \ldots, NCS\}$. Set of platform-clusters. Cardinality $|\mathcal{RFCS}| = NCS$, where NCS is the number of clusters. Index rr and rrr.
    - $\mathcal{STCS} = \{1, \ldots, NCS\}$. Set of st-clusters. Cardinality $|\mathcal{STCS}| = NCS$, where NCS is the number of clusters. Index rr and rrr.
    - $\mathcal{NRC}$. Single-element set representing all charging tanks (namely, CT Group). Index rr and rrr.

- $\mathcal{NRD}$. Single-element set representing all CDUs (namely, CDU Group). Index rr and rrr.
- $\mathcal{NR} = \mathcal{RFCS} \cup \mathcal{STCS} \cup \mathcal{NRC} \cup \mathcal{NRD}$. Set of all new resources. Index rr and rrr.
- $\mathcal{N} \subset \mathcal{NR} \times \mathcal{NR}$. Links between aggregated resources to represent the network illustrated in Figure 2 (b).
- $\mathcal{C}$. Set of crudes. Index c.
- $\mathcal{K}$. Set of properties. Index k.

2. **Parameters**.

- $RATE_r$. Production rate of platform $r \in \mathcal{RF}$ in 10^3 bbl/day.
- $PR_{k,c}$. Value of property k associated to crude c.
- $CFPSO_{c,r} \in \{0, 1\}$. Indicates if crude c is produced in platform $r \in \mathcal{RF}$.
- $[\underline{CAP}_r,\ \overline{CAP}_r]$. Lower and upper bounds on the capacity of each resource $r \in \mathcal{R}$.
- $[\underline{DEM}_r,\ \overline{DEM}_r]$. Lower and upper bounds on the total volume of crude oil demanded from each charging tank $r \in \mathcal{RC}$ by the CDUs.
- TIL_r. Initial level of crude oil in resource $r \in \mathcal{R} \setminus \mathcal{RD}$.
- $[\underline{\widetilde{CAP}}_{rr},\ \overline{\widetilde{CAP}}_{rr}]$. Lower and upper bounds on the capacity of new resource $rr \in \mathcal{NR}$.
- $[\underline{\widetilde{DEM}}_{rr},\ \overline{\widetilde{DEM}}_{rr}]$. Lower and upper bounds on the total demand of the CDU Group $rr \in \mathcal{NRD}$ over the planning horizon. Notice that $\underline{\widetilde{DEM}}_{rr} = \sum_{r \in \mathcal{RC}} \underline{DEM}_r$ and $\overline{\widetilde{DEM}}_{rr} = \sum_{r \in \mathcal{RC}} \overline{DEM}_r$.
- $[\underline{\widetilde{FR}}_r,\ \overline{\widetilde{FR}}_r]$. Lower and upper bounds on the outlet flow rate of resource $r \in \mathcal{R} \setminus \mathcal{RD}$.
- $[\underline{PR}_k,\ \overline{PR}_k]$. Lower and upper bounds on property k among all crudes c.
- $[\underline{PRST}_{k,r},\ \overline{PRST}_{k,r}]$. Lower and upper bounds of property k in each storage tank $r \in \mathcal{RS}$. These bounds are defined after finalizing the cluster optimization, as calculated by Eqs. (30) and (31) which will be introuced later, and then used to linearize the blending constraints defined in the OMCOS formulation (de Assis et al., 2019).

3. **0-1 Variables**.

   - $assignSTCS_{r,rr} \in \{0,1\}$. Is 1 if storage tank $r \in \mathcal{RS}$ is assigned to st-cluster $rr \in \mathcal{STCS}$.
   - $assignRFCS_{r,rr} \in \{0,1\}$. Is 1 if platform $r \in \mathcal{RF}$ is assigned to platform-cluster $rr \in \mathcal{RFCS}$.
   - $assign_{i,rr,rrr} \in \{0,1\}$. Is 1 if there is flow of crude oil between resources $rr, rrr \in \mathcal{NR}$ in period i.
   - $crudeSTCS_{c,rr} \in \{0,1\}$. Is 1 if crude c is assigned to st-cluster rr.

4. **Continuous Variables**.

   - $\widetilde{lr}_{i,rr} \geq 0$. Total level of crude oil in the cluster resource $rr \in \mathcal{NR}$ in period i.
   - $\widetilde{vt}_{i,rr,rrr} \geq 0$. Flow of crude between cluster resources $(rr, rrr) \in \mathcal{N}$ in period i.
   - $\widetilde{il}_{i,rr} \geq 0$. Initial level of crude oil in cluster resource $rr \in \mathcal{NR} \setminus \mathcal{NRD}$.
   - $tg_{k,rr} \geq 0$. Value associated to property k in st-cluster $rr \in \mathcal{STCS}$ such that the difference between $tg_{k,rr}$ and the property of crudes assigned to rr is minimized.
   - $epr_{i,r,k} \geq 0$. Is the value of property k in storage or charging tank r in period i.

Having introduced the notation, we are in a position to present the constraints that define the formulation for clustering platforms and storage tanks.

4.1. Clustering Rules

Eq. (1) states that at least one platform $r \in \mathcal{RF}$ must be assigned to a platform-cluster $rr \in \mathcal{RFCS}$.

$$\sum_{r \in \mathcal{RF}} assignRFCS_{r,rr} \geq 1, \; rr \in \mathcal{RFCS}. \tag{1}$$

Also, Eq. (2) defines that a platform r is assigned to one platform-cluster rr.

$$\sum_{rr \in \mathcal{RFCS}} assignRFCS_{r,rr} = 1, \; r \in \mathcal{RF}. \tag{2}$$

Similar rules can be defined for the storage tank clusters. At least one storage tank $r \in \mathcal{RS}$ must be assigned to a st-cluster $rr \in \mathcal{STCS}$.

$$\sum_{r \in \mathcal{RS}} assignSTCS_{r,rr} \geq 1, \; rr \in \mathcal{STCS}. \tag{3}$$

Further, a storage tank r must be assigned to one st-cluster rr.

$$\sum_{rr \in \mathcal{STCS}} assignSTCS_{r,rr} = 1, \; r \in \mathcal{RS}. \tag{4}$$

Eqs. (5)-(6) track which crude c is assigned to each st-cluster rrr. If platform r, which produces crude c, is assigned to platform-cluster rr and the connection from the platform-cluster rr to st-cluster rrr is defined in the network $\mathcal{N}$, then crude c will be delivered to st-cluster rrr. Put another way, this constraint states that $crudeSTCS_{c,rrr} = 1$ when: platform r produces crude c, $CFPSO_{c,r} = 1$; the platform is assigned to platform-cluster rr, $assignRFCS_{r,rr} = 1$; and the platform-cluster rr feeds st-cluster rrr, a condition established by the link $(rr, rrr) \in \mathcal{N}$.

$$\begin{aligned} & crudeSTCS_{c,rrr} \geq assignRFCS_{r,rr}, \\ & \qquad c \in \mathcal{C}, \; r \in \mathcal{RF}, \; rrr \in \mathcal{STCS}, \; (rr, rrr) \in \mathcal{N} : CFPSO_{c,r} = 1, \end{aligned} \tag{5}$$

$$\begin{aligned} & crudeSTCS_{c,rrr} \leq \sum_{\substack{r \in \mathcal{RF}: \\ CFPSO_{c,r}=1}} assignRFCS_{r,rr}, \\ & \qquad c \in \mathcal{C}, \; rrr \in \mathcal{STCS}, (rr, rrr) \in \mathcal{N}. \end{aligned} \tag{6}$$

4.2. Inventory Control

Eq. (7) defines the initial volume il_{rr} of crude, in each platform-cluster $rr \in \mathcal{RFCS}$, as the sum of initial volume TIL_r in each platform $r \in \mathcal{RF}$ assigned to rr (i.e., $assignRFCS_{r,rr} = 1$). Similarly, Eq. (8) defines the initial volume in each st-cluster $rr \in \mathcal{STCS}$. The proposed cluster framework considers a unique group $rr \in \mathcal{NRC}$ of charging tanks (see Fig. 2 (b)). Therefore, Eq. (9) defines the initial volume il_{rr} of the CT Group as the the sum of the initial volume TIL_r

in each charging tank $r \in \mathcal{RC}$.

$$il_{rr} = \sum_{r \in \mathcal{RF}} TIL_r \ assignRFCS_{r,rr}, \ rr \in \mathcal{RFCS}, \tag{7}$$

$$il_{rr} = \sum_{r \in \mathcal{RS}} TIL_r \ assignSTCS_{r,rr}, \ rr \in \mathcal{STCS}, \tag{8}$$

$$il_{rr} = \sum_{r \in \mathcal{RC}} TIL_r, \ rr \in \mathcal{NRC}. \tag{9}$$

The inventory $\widetilde{lr}_{i,rr}$ of crude of platform-cluster rr in period i, as described by Eqs. (10) and (11), is defined as the inventory in the previous period i-1, $\widetilde{lr}_{i-1,rr}$ (or il_{rr} for $i = 1$) added to the production rate $RATE_r$ of all platforms r assigned to platform-cluster rr (i.e., $assignRFCS_{r,rr} = 1$) and subtracted the flow $\widetilde{vt}_{i,rr,rrr}$ of crude oil entering the st-cluster rrr.

$$\widetilde{lr}_{i,rr} = il_{rr} + \sum_{r \in \mathcal{RF}} RATE_r \ assignRFCS_{r,rr} - \widetilde{vt}_{i,rr,rrr}, \qquad rr \in \mathcal{RFCS}, \ (rr, rrr) \in \mathcal{N}, \ i = 1, \tag{10}$$

$$\widetilde{lr}_{i,rr} = \widetilde{lr}_{i-1,rr} + \sum_{r \in \mathcal{RF}} RATE_r \ assignRFCS_{r,rr} - \widetilde{vt}_{i,rr,rrr} \qquad rr \in \mathcal{RFCS}, \ (rr, rrr) \in \mathcal{N}, \ i \in \mathcal{T} \setminus \{1\}. \tag{11}$$

Likewise, Eqs. (12) and (13) enforce the inventory control in each st-cluster and in the CT Group for the initial period $i = 1$ and the remaining planning horizon, respectively. For each st-cluster $rr \in \mathcal{STCS}$, the level of crude $\widetilde{lr}_{i,rr}$ in period i is defined as its previous level $\widetilde{lr}_{i-1,rr}$ (or il_{rr} for $i = 1$) plus the inlet flow from platform-cluster $rrr \in \mathcal{RFCS} : (rrr, rr) \in \mathcal{N}$, subtracted the outlet flow to the CT Group $rrr \in \mathcal{NRC} : (rr, rrr) \in \mathcal{N}$.

$$\widetilde{lr}_{i,rr} = il_{rr} + \sum_{(rrr,rr) \in \mathcal{N}} \widetilde{vt}_{i,rrr,rr} - \sum_{(rr,rrr) \in \mathcal{N}} \widetilde{vt}_{i,rr,rrr}, \qquad rr \in (\mathcal{STCS} \cup \mathcal{NRC}), \ i = 1, \tag{12}$$

$$\widetilde{lr}_{i,rr} = \widetilde{lr}_{i-1,rr} + \sum_{(rrr,rr) \in \mathcal{N}} \widetilde{vt}_{i,rrr,rr} - \sum_{(rr,rrr) \in \mathcal{N}} \widetilde{vt}_{i,rr,rrr}, \qquad rr \in (\mathcal{STCS} \cup \mathcal{NRC}), \ i \in \mathcal{T} \setminus \{1\}. \tag{13}$$

In the case of rr being CT Group, the level $\widetilde{lr}_{i,rr}$ of crude in period i is defined as its previous level $\widetilde{lr}_{i-1,rr}$ (or il_{rr} for $i = 1$) plus the inlet flow from st-cluster $rrr \in \mathcal{STCS} : (rrr, rr) \in \mathcal{N}$, minus the outlet flow to the CDU Group $rrr \in \mathcal{NRD} : (rr, rrr) \in \mathcal{N}$.

4.3. Resource Capacity

The capacity bounds on each platform-cluster, st-cluster and the CT Group are enforced by Eqs. (14)-(16).

$$\widetilde{lr}_{i,rr} \leq \sum_{r \in \mathcal{RF}} \overline{CAP}_r \; assignRFCS_{r,rr}, \; i \in \mathcal{T}, \; rr \in \mathcal{RFCS}, \tag{14}$$

$$\widetilde{lr}_{i,rr} \leq \sum_{r \in \mathcal{RS}} \overline{CAP}_r \; assignSTCS_{r,rr}, \; i \in \mathcal{T}, \; rr \in \mathcal{STCS}, \tag{15}$$

$$\widetilde{lr}_{i,rr} \leq \sum_{r \in \mathcal{RC}} \overline{CAP}_r, \; i \in \mathcal{T}, \; rr \in \mathcal{NRC}. \tag{16}$$

4.4. Flow Rate Limits

Eq. (17) defines that when there is flow of crude oil out of platform-cluster rr to st-cluster rrr in period i (i.e., $assign_{i,rr,rrr} = 1$), it is limited by the sum of maximum flow rate $\overline{\widetilde{FR}}_r$ allowed out of each platform r that is assigned to platform-cluster rr (i.e., $assignRFCS_{r,rr} = 1$).

$$\widetilde{vt}_{i,rr,rrr} \leq \left(\sum_{r \in \mathcal{RF}} \overline{\widetilde{FR}}_r \; assignRFCS_{r,rr} \right) assign_{i,rr,rrr}, \\ i \in \mathcal{T}, rr \in \mathcal{RFCS}, \; (rr, rrr) \in \mathcal{N}, \tag{17}$$

Notice that this equation is non-linear as it involves the product of two 0-1 variables, but it can be linearized by the set of Eqs. (18).

$$\begin{cases} \widetilde{vt}_{i,rr,rrr} & \geq 0, \\ \widetilde{vt}_{i,rr,rrr} & \leq \sum_{r \in \mathcal{RF}} \overline{\widetilde{FR}}_r \; assignRFCS_{r,rr}, \\ \widetilde{vt}_{i,rr,rrr} & \leq \left(\sum_{r \in \mathcal{RF}} \overline{\widetilde{FR}}_r \right) assign_{i,rr,rrr}. \end{cases} \quad i \in \mathcal{T}, rr \in \mathcal{RFCS}, (rr, rrr) \in \mathcal{N}. \tag{18}$$

A similar constraint can be defined to limit the flow of crudes out of a st-cluster rr as stated in Eq. (19). Like Eq. (17), the flow out of a st-cluster is limited by the sum of maximum flow rate $\overline{\widetilde{FR}}_r$ allowed out of each storage tank r assigned to the cluster rr (i.e., $assignSTCS_{r,rr} = 1$).

$$\widetilde{vt}_{i,rr,rrr} \leq \left(\sum_{r \in \mathcal{RS}} \overline{\widetilde{FR}}_r \; assignSTCS_{r,rr} \right) assign_{i,rr,rrr}, \\ i \in \mathcal{T}, \; rr \in \mathcal{STCS}, \; (rr, rrr) \in \mathcal{N}, \quad (19)$$

Likewise, this equation can be linearized by the set of Eqs. (20).

$$\begin{cases} \widetilde{vt}_{i,rr,rrr} & \geq 0, \\ \widetilde{vt}_{i,rr,rrr} & \leq \sum\limits_{r \in \mathcal{RS}} \overline{\widetilde{FR}}_r \; assignSTCS_{r,rr}, \\ \widetilde{vt}_{i,rr,rrr} & \leq \left(\sum\limits_{r \in \mathcal{RS}} \overline{\widetilde{FR}}_r \right) assign_{i,rr,rrr}. \end{cases} \quad i \in \mathcal{T}, rr \in \mathcal{STCS}, (rr, rrr) \in \mathcal{N}. \quad (20)$$

Finally, the total flow of crude oil from the CT Group into the CDU Group is limited by Eq. (21). Notice that this constraint is linear since there is only one CT Group which contains all charging tanks.

$$\widetilde{vt}_{i,rrr,rr} \leq \left(\sum_{r \in \mathcal{RC}} \overline{\widetilde{FR}}_r \right) assign_{i,rr,rrr}, \; i \in \mathcal{T}, \; rr \in \mathcal{NRC}, \; (rr, rrr) \in \mathcal{N}. \quad (21)$$

4.5. Demand Satisfaction

Eqs. (22) and (23) define respectively the lower and upper bounds on crude oil demand for the group of CDUs. For instance, the lower demand $\underline{\widetilde{DEM}}_{rrr}$ of the group of CDUs $rrr \in \mathcal{NRD}$, in Eq. (22), is defined as the sum of the minimum supply of crude oil $\underline{DEM}_r$ that each charging tank r needs to provide to the CDUs. A similar definition is also valid for the upper bound in Eq. (23).

Eq. (24) ensures that the total flow $\widetilde{vt}_{i,rr,rrr}$ from the CT Group $rr \in \mathcal{NRC}$ to the CDU Group $rrr \in \mathcal{NRD}$, over the planning horizon, must be within the lower and upper bounds $[\underline{\widetilde{DEM}}_{rrr}, \overline{\widetilde{DEM}}_{rrr}]$ on the overall crude demand requested by the CDUs.

$$\widetilde{\underline{DEM}}_{rrr} = \sum_{r \in \mathcal{RC}} \underline{DEM}_r,\ rrr \in \mathcal{NRD}, \tag{22}$$

$$\overline{\widetilde{DEM}}_{rrr} = \sum_{r \in \mathcal{RC}} \overline{DEM}_r,\ rrr \in \mathcal{NRD}. \tag{23}$$

$$\widetilde{\underline{DEM}}_{rrr} \leq \sum_{i \in \mathcal{T}} \sum_{(rr,rrr) \in \mathcal{N}} \widetilde{vt}_{i,rr,rrr} \leq \overline{\widetilde{DEM}}_{rrr},\ rrr \in \mathcal{NRD}. \tag{24}$$

4.6. Operation Rules

Rules on inlet and outlet operations (de Assis et al., 2019, 2017), can be applied for the cluster network in Figure 2 (b). Eq. (25) ensures that at most one st-cluster rr can feed the CT Group rrr in period i.

$$\sum_{\substack{rr \in \mathcal{STCS}: \\ (rr,rrr) \in \mathcal{N}}} assign_{i,rr,rrr} \leq 1,\ i \in \mathcal{T},\ rrr \in \mathcal{NRC}. \tag{25}$$

Eq. (26) states that, in period i, an inlet operation from platform-cluster rr into a st-cluster rrr can not be performed at the same time as an outlet operation from the same st-cluster rrr.

$$\begin{aligned} & assign_{i,rr,rrr} + assign_{i,rrr,rrrr} \leq 1, \\ & \qquad i \in \mathcal{T}, rrr \in \mathcal{STCS}, (rr, rrr) \in \mathcal{N}, (rrr, rrrr) \in \mathcal{N}. \end{aligned} \tag{26}$$

Further, as stated by de Assis et al. (2019), Eq. (27) defines the continuous distillation condition, which means that in all periods of time $i \in \mathcal{T}$ the CT Group $rr \in \mathcal{NRC}$ must be assigned to supply crude to the CDU Group $rrr \in \mathcal{NRD}$.

$$assign_{i,rr,rrr} = 1,\ i \in \mathcal{T},\ rrr \in \mathcal{NRD},\ (rr, rrr) \in \mathcal{N}. \tag{27}$$

4.7. Objective

The objective (28) of the clustering problem involves three terms:

$$
\begin{aligned}
C: \ \max\ f = & \left(\sum_{i \in \mathcal{T}} \sum_{\substack{(rr,rrr) \in \mathcal{N}: \\ rr \in \mathcal{NRC}}} \widetilde{vt}_{i,rr,rrr} \right) / \overline{\widetilde{DEM}}_{CDUGroup} \\
& - blendWeight \sum_{k \in \mathcal{K}} \sum_{c \in \mathcal{C}} \sum_{rr \in \mathcal{STCS}} \frac{|PR_{k,c} - tg_{k,rr}|}{\overline{PR}_k - \underline{PR}_k} crudeSTCS_{c,rr} \\
- \sum_{\substack{(rr,rrr) \in \mathcal{N}: \\ rr \in \mathcal{RFCS}}} & \sum_{\substack{(rrrr,rrrrr) \in \mathcal{N}: \\ rrrr \in \mathcal{RFCS}}} \left| \frac{\sum_{r \in \mathcal{RS}} assignSTCS_{r,rrr}}{\sum_{r \in \mathcal{RF}} assignRFCS_{r,rr}} - \frac{\sum_{r \in \mathcal{RS}} assignSTCS_{r,rrrrr}}{\sum_{r \in \mathcal{RF}} assignRFCS_{r,rrrr}} \right|
\end{aligned}
\tag{28}
$$

- In the first term, similar to the baseline operational management problem defined in de Assis et al. (2019), the clustering problem aims to maximize the flow of crude supplied to the CDU Group. Since this term has as denominator the maximum demand of the CDU Group, it will assume a maximum value of 1. This is done to bring this term to an order that is comparable with the other terms in the objective.

- In the second term, the objective aims to define a single target value for each property k, and for all crudes to be stored in a given st-cluster rr, namely the value $tg_{k,rr}$. Then, the objective seeks to minimize the deviation of the property k of each crude c, $PR_{k,c}$, that can be delivered to the ST cluster rr from the target value $tg_{k,rr}$. As the property k may vary depending on the type of crude c, this objective seeks to group FPSOs with similar crudes to feed the same st-cluster.

  This term, which is the L_1 distance metric, is typically found in K-medoids MILP formulations for building clusters (Papageorgiou and Trespalacios, 2018; Nemhauser and Wolsey, 1988; Kaufman and Rousseeuw, 1987). The choice of the L_1 distance metric is also endorsed by Kelly et al. (2017b), which propose a clustering MILP model for assigning crudes from external sources to storage tanks in a crude oil terminal.

- Finally, the last term of the objective balances the proportion between the number of storage tanks and platforms in a (platform-cluster, st-cluster) pair with the proportion of these resources in the remaining pairs.

The detailed linearization of Eq. (28), which is described in Appendix Appendix A.

Consequently, the objective can be reformulated as the linear function given by Eq. (29), with parameter $blendWeight$ being a weighting factor for the second term of the objective.

$$C: \ \max \ f = \left(\sum_{i \in \mathcal{T}} \sum_{\substack{(rr,rrr) \in \mathcal{N}: \\ rr \in \mathcal{NRC}}} \widetilde{vt}_{i,rr,rrr} \right) / \overline{\widetilde{\widetilde{DEM}}}_{CDUGroup} \\ - blendWeight \sum_{k \in \mathcal{K}} \sum_{c \in \mathcal{C}} \sum_{rr \in \mathcal{STCS}} \widehat{deviation}_{k,c,rr} \\ - \sum_{\substack{(rr,rrr) \in \mathcal{N}: \\ rr \in \mathcal{RFCS}}} \sum_{\substack{(rrrr,rrrrr) \in \mathcal{N}: \\ rrrr \in \mathcal{RFCS}}} proportionDiff_{rr,rrr}^{rrrr,rrrrr} \quad (29)$$

Having introduced the definitions above, the MILP formulation for the clustering problem consists in minimizing the objective (29) subject to the constraints (1)-(16), Eqs. (18), (20)-(27), (A.1)-(A.2), (A.4), (A.5)-(A.6), (A.9)-(A.10), (A.13)-(A.14), and (A.16)-(A.17).

4.8. Remarks

There are two main consequences of clustering crudes.

- First, by restricting the crudes allowed in each st-cluster rr, lower and upper bounds $[\underline{PRST}_{r,k}, \ \overline{PRST}_{r,k}]$ can be derived for the value of property k for the storage tank r assigned to st-cluster rr.

$$\underline{PRST}_{k,r} = \min\{PR_{k,c} : crudeSTCS_{c,rr} \cdot assignRFCS_{r,rr} = 1, \\ c \in \mathcal{C}, \ rr \in \mathcal{STCS}\}, \ k \in \mathcal{K}, \ r \in \mathcal{RS}. \quad (30)$$

$$\overline{PRST}_{k,r} = \max\{PR_{k,c} : crudeSTCS_{c,rr} \cdot assignRFCS_{r,rr} = 1, \\ c \in \mathcal{C}, \ rr \in \mathcal{STCS}\}, \ k \in \mathcal{K}, \ r \in \mathcal{RS}. \quad (31)$$

- And second, the assignment of a platform-cluster to feed a given st-cluster restricts a vessel to unload in specific storage tanks, and travel only between platforms and storage tanks within these clusters. The full development of these bounds, and their use on the operational management of crude oil supply will be the focus of Section 5. The computational gains arising for this new solution methodology will be analyzed in Section 6.

5. Linear Approximation of Bilinear Terms and Solution Strategy

The solution of the OMCOS MINLP formulation proposed by de Assis et al. (2019) consists of a MILP-NLP decomposition scheme. The MILP is a relaxation of the MINLP, in which the blending constraints are relaxed by McCormick envelopes (McCormick, 1976; Castro, 2015). Further, the logistics decisions (i.e., binary variables) obtained by solving the MILP are fixed in the MINLP, yielding a NLP which is then solved to obtain a primal solution. Despite generating tight relaxations, the use of McCormick envelopes (e.g., univariate or bivariate partitioning) increase the number of binary variables that lead to a significant impact on the solution time of the MILP.

This section presents an alternative way to handle blending constraints, which takes advantage of the structure imposed by solving the clustering formulation.

5.1. Linear Approximation of Bilinear Terms

When a blend of crudes accumulates in a storage or charging tank $r \in \mathcal{RS} \cup \mathcal{RC}$, the total level of crude $lr_{i,r}$ in tank r in period i can also be seen as the sum of volumes of each crude c (i.e., $lr_{i,r} = \sum_{c \in \mathcal{C}} lcr_{i,r,c}$). Variable $epr_{i,r,k}$, which denotes the value of property $k \in \mathcal{K}$ of the blend of crudes in a storage or charging tank r in period i, can be defined by the following non-linear equation:

$$epr_{i,r,k} = \sum_{c \in \mathcal{C}} PR_{k,c} \frac{lcr_{i,r,c}}{lr_{i,r}}, \; i \in \mathcal{T}, \; r \in \mathcal{RS} \cup \mathcal{RC}, \; k \in \mathcal{K}, \tag{32}$$

where parameter $PR_{k,c}$ is the value of property k in crude oil c, and the non-linear term $\frac{lcr_{i,r,c}}{lr_{i,r}}$ is the volume fraction of crude c in a tank r during time i.

As in the baseline model (de Assis et al., 2019), variables analogous to the total and individual level of crudes in a tank r in period i (i.e., $lcr_{i,r,c}$ and $lr_{i,r}$) can be introduced to track the flow of crudes between resources. While $vt_{i,v}$ is the total volume of crude oil transferred in period i by operation v, $vct_{i,v,c}$ is the volume of crude c transferred in period i by operation v (i.e., $vt_{i,v} = \sum_{c \in \mathcal{C}} vct_{i,v,c}$). The blending constraint states that the proportion of crude c inside a storage or charging tank r, defined by $\frac{lcr_{i,r,c}}{lr_{i,r}}$, must hold whenever there is a flow operation v out of resource r (i.e., operation $v \in \mathcal{O}_r$), a condition imposed by the equation below:

$$\frac{vct_{i,v,c}}{vt_{i,v}} = \frac{lcr_{i,r,c}}{lr_{i,r}}, \tag{33}$$

which can be rewritten as:

$$vct_{i,v,c} = vt_{i,v} \frac{lcr_{i,r,c}}{lr_{i,r}},\ i \in \mathcal{T},\ r \in \mathcal{RS} \cup \mathcal{RC},\ v \in \mathcal{O}_r,\ c \in \mathcal{C}. \tag{34}$$

Equation (34) can be incorporated in Eq. (32) by multiplying both sides by $vt_{i,v}$ to obtain Eq. (35) and replacing $\frac{lcr_{i,r,c}}{lr_{i,r}} vt_{i,v}$ by $vct_{i,v,c}$ to obtain Eq. (36). Notice that the right-hand side of Eq. (36) is linear, while the left-hand side is non-linear.

$$epr_{i,r,k} vt_{i,v} = \sum_{c \in \mathcal{C}} PR_{k,c} \frac{lcr_{i,r,c}}{lr_{i,r}} vt_{i,v}, \tag{35}$$

$$epr_{i,r,k} vt_{i,v} = \sum_{c \in \mathcal{C}} PR_{k,c} vct_{i,v,c},\ i \in \mathcal{T},\ r \in \mathcal{RS} \cup \mathcal{RC},\ v \in \mathcal{O}_r,\ k \in \mathcal{K}. \tag{36}$$

This is valid for all periods of time $i \in \mathcal{T}$, storage and charging tanks $r \in \mathcal{RS} \cup \mathcal{RC}$, crude properties $k \in \mathcal{K}$ and transfer operations $v \in \mathcal{O}_r$ leaving resource r. The steps to reach equation (36) are similar to the ones followed by Méndez et al. (2006), who propose a strategy to tackle blending and short-term scheduling in oil-refinery applications. Equation (36) balances, for every period i, the total ($vt_{i,v}$) and individual ($vct_{i,v,c}$) volumes of crude flowing out from storage or charging tanks, with the overall value of property k in tank r ($epr_{i,r,k}$) and the individual property k of each crude c ($PR_{k,c}$).

Every feeding operation $v \in \mathcal{WD}$ from charging tanks to CDUs is bounded by lower and upper bounds $[\underline{DEMC}_{v,k},\ \overline{DEMC}_{v,k}]$ on the value of property k. This means that variable $epr_{i,r,k}$, which is the value of property k in charging tank r in period i, must be between these bounds when there is a transfer of crudes to a CDU. Eq. (37) takes advantage of the lower and upper bounds $[\underline{DEMC}_{v,k},\ \overline{DEMC}_{v,k}]$ to define a linearization for Eq. (36).

$$\begin{aligned} \underline{DEMC}_{v,k} vt_{i,v} \le \sum_{c \in \mathcal{C}} & PR_{k,c} vct_{i,v,c} \\ & \le \overline{DEMC}_{v,k} vt_{i,v},\ i \in \mathcal{T},\ r \in \mathcal{RC},\ v \in \mathcal{O}_r,\ k \in \mathcal{K}. \end{aligned} \tag{37}$$

For storage tanks, vessels may unload different types of crudes during the planning horizon, making it difficult to derive bounds on properties and consequently linearizations such as Eq. (37). Nevertheless, the solution of the clustering problem restricts the crudes present in each st-cluster rr, and consequently in the storage tanks assigned to cluster rr. As a consequence, Eqs. (30) and (31) define lower and upper bounds $[\underline{PRST}_{k,r},\ \overline{PRST}_{k,r}]$ on the crude property k for storage tank r. Similarly as in Eq. (37), one can take advantage of these bounds to linearize Eq. (36) for the case of storage tanks, as indicated by Eq. (38).

$$\underline{PRST}_{k,r} vt_{i,v} \leq \sum_{c\in\mathcal{C}} PR_{k,c} vct_{i,v,c} \leq \overline{PRST}_{k,r} vt_{i,v},$$
$$i \in \mathcal{T},\ r \in \mathcal{RS},\ v \in \mathcal{O}_r,\ k \in \mathcal{K}. \quad (38)$$

5.2. Solution Strategy

In general terms, the combination of the clustering recommendation with the two-step MILP-NLP solution strategy consists of the following steps:

- First, the clustering formulation (29) of a problem instance is solved. Then, the original problem instance is restricted according to the solution of the clustering problem (29). This means: (a) constrain the domain of vessel's trips variable[1] $s_{i,r,v,u}$ to consider trips only among the platforms and storage tanks that belong to the same (platform-cluster, st-cluster) pair, and (b) constrain the domain of logistics decisions variable[2] $z_{i,v}$ for vessels to only consider offloading crudes from platforms and unloading them into storage tanks that belong to the same (platform-cluster, st-cluster) pair.

- Next, an MILP linearization of the OMCOS problem defined by de Assis et al. (2019) is constructed, which consists of all the original variables and constraints, except the blending constraint (33) which is linearized by Eqs. (37) and (38).

- Finally, the solution of the MILP is used as an initial point and its logistics decisions $z_{i,v}$ and $s_{i,r,v,u}$ (binary variables) are fixed in the MINLP, resulting in a non-linear programming (NLP) problem, which is solved to obtain the final solution.

6. Analysis

The goal of this section is to (a) analyze the results given by the clustering formulation (29) proposed in Section 4; (b) understand how the clustering of re-

[1]Binary variable $s_{i,r,v,u}$ takes on value 1 if, after performing an operation v in period i, vessel r performs an operation u in period $i+1$.

[2]Logistic variable $z_{i,v}$ assumes value 1 if operation v is executed in period i.

sources, and consequently crudes, affects the solution of the three instances considered for the *Operational Management of Crude Oil Supply* (de Assis et al., 2019); and (c) propose and solve new instances.

First, the clustering formulation is run to obtain different clustering schemes for the instances 2F-2V-2ST-2CT-1CDU-2C-1P-15D, 4F-4V-6ST-4CT-3CDU-8C-1P-15D and 4F-4V-10ST-6CT-5CDU-8C-1P-15D. For example, instance 2F-2V-2ST-2CT-1CDU-2C-1P-15D means: 2 FPSOS, 2 vessels, 2 storage tanks, 2 charging tanks, 1 distillation column, 2 crude oils, 1 crude property, and 15 days for planning. Next, the clustering schemes are applied to the original instances, resulting into clustered instances, which are solved according to the MILP-NLP solution strategy proposed in Section 5. Finally, larger instances are proposed and solved using the same MILP-NLP strategy.

The mathematical programming models and solution strategy were implemented in AMPL and solved in a computer with two Intel Core Xeon E5-2630 v4 Processor (2.20 GHz), totaling 20 cores of 2 threads, 64 GB of RAM and a Ubuntu environment. The MILP model is solved with CPLEX (IBM, 2013) and the NLP formulation with CONOPT (Drud, 1985).

6.1. Clusterization of Instances

Figure 3 illustrates the clustering scheme obtained for instance 4F-4V-6ST-4CT-3CDU-8C-1P-15D considering 2 clusters. In this case, the (platform-cluster 1, st-cluster 1) pair consists of platform $FPSO1$, which produces crude cA with sulfur content of 0.01, and storage tank $ST1$. This means that the only platform allowed to supply tank $ST1$ is $FPSO1$, which implies that $ST1$ will only store crude cA. Further, vessels allocated to this cluster can only travel between the terminal and platform $FPSO1$, and unload crude cA into $ST1$.

Similarly, platforms $FPSO2$, $FPSO3$ and $FPSO4$ are assigned to supply storage tanks $ST2$ to $ST6$, which define the (platform-cluster 2, st-cluster 2) pair. These platforms produce respectively crudes cB, cC, and cD, with sulfur content of 0.03, 0.045, and 0.06 respectively. Vessels assigned to cover the routes in this cluster pair will only travel between the terminal and platforms $FPSO2$ to $FPSO4$, and unload crudes cB, cC and cD into tanks $ST2$ to $ST6$.

After assigning resources to the (platform-cluster 1, st-cluster 1) and (platform-cluster 2, st-cluster 2) pairs, bounds are derived on the crude property for each st-cluster and consequently for each of its storage tank. For instance, st-cluster 1 and its tank $ST1$ will receive only crude cA, thus the lower and upper bounds on the sulfur property in tank $ST1$ are $[0.01, 0.01]$. The same implication is also valid for st-cluster 2, whose tanks $ST2$-$ST6$ will have $[0.03, 0.06]$ as bounds for the sulfur content. As given by Eq. (38), parameters $\underline{PRST}_{k,r}$ and $\overline{PRST}_{k,r}$ make use of these bounds to linearize the blending constraint (33), and thereby obtain the MILP linearization of the OMCOS MINLP formulation proposed by de Assis et al. (2019).

At the end of the chain, the st-cluster 1 and st-cluster 2 (of storage tanks) feed the CT Group (group of charging tanks), which has the combined storage and transfer capacity of the entire farm of charging tanks. Finally, the CT Group satisfies the crude demand of the CDUs, which are combined in the CDU Group.

Notice that, according to the objective (29), the cluster scheme illustrated in Figure 3 is such that: (a) the flow of crudes to the CDU Group is maximized; (b) the crude property deviation in a st-cluster is minimized; and (c) the difference of resource proportion (i.e., the ratio between the number of storage tanks and platforms in a cluster) among all (platform-cluster, st-cluster) pairs is minimized.

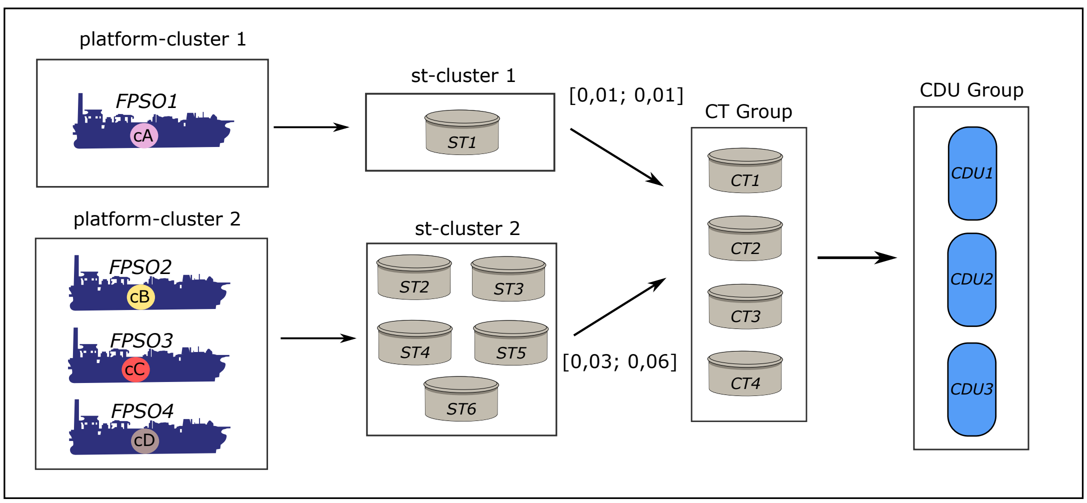


Figure 3: Clustering scheme for instance 4F-4V-6ST-4CT-3CDU-8C-1P-15D considering 2 clusters.

Tables 1, 2 and 3 summarize the clustering schemes for instances 2F-2V-2ST-2CT-1CDU-2C-1P-15D, 4F-4V-6ST-4CT-3CDU-8C-1P-15D and 4F-4V-10ST-6CT-5CDU-8C-1P-15D considering different numbers of clusters. They show the number of clusters, the assignment of platforms and storage tanks to each (platform-cluster, st-cluster) pair and the resulting crudes in each st-cluster. The CPU time is negligible since the solution of the clustering problem takes seconds, and therefore is not a computational burden.

Another important result from solving the clustering problem defined in Section 4 is the value of the target variable $tg_{k,rr}$ and the resulting property deviation in each st-cluster. Notice that for all instances, the target variable $tg_{k,rr}$ assumes the value of the property k of one of the crudes supplied to the st-cluster rr. If an st-cluster receives only one type of crude oil (e.g., instance 2F-2V-2ST-2CT-1CDU-2C-1P-15D in Table 1), the variable $tg_{k,rr}$ will assume the value of property k of this unique crude and the deviation in that st-cluster will be 0. On the other hand, as shown in Table 2, the deviation considering 3 clusters is 0.3 in st-cluster 2, which is selected to receive crudes cB ($PR_{S,cB} = 0.03$) and cC ($PR_{S,cB} = 0.045$). From the objective function, this deviation is computed as $\frac{|PR_{k,c} - tg_{k,rr}|}{\overline{PR}_k - \underline{PR}_k} = \frac{|0.03-0.045|}{(0.06-0.01)} = 0.3$.

Table 1: Cluster schemes for instance 2F-2V-2ST-2CT-1CDU-2C-1P-15D.

Number Clusters	(platform-cluster, st-cluster) pair	Crudes st-cluster	Target $tg_{k,rr}$ st-cluster	Deviat. st-cluster
2	pair1 = ({fpso1}, {st1})	st-cluster 1 = {cA}	0.01	0
	pair2 = ({fpso2}, {st2})	st-cluster 2 = {cB}	0.03	0

Table 2: Cluster schemes for instance 4F-4V-6ST-4CT-3CDU-8C-1P-15D.

Number Clusters	(platform-cluster, st-cluster) pair	Crudes st-cluster	Target $tg_{k,rr}$ st-cluster	Deviat. st-cluster
2	pair1 = ({fpso1}, {st1})	st-cluster 1 = {cA}	0.01	0
	pair2 = ({fpso2, fpso3, fpso4}, {st2, st3, st4, st5, st6})	st-cluster 2 = {cB, cC, cD}	0.045	0.6
3	pair1 = ({fpso1}, {st1, st2})	st-cluster 1 = {cA}	0.01	0
	pair2 = ({fpso2, fpso3}, {st3, st4, st5})	st-cluster 2 = {cB, cC}	0.045	0.3
	pair3 = ({fpso4}, {st6})	st-cluster3 = {cD}	0.06	0
4	pair1 = ({fpso1}, {st1, st2})	st-cluster 1 = {cA}	0.01	0
	pair2 = ({fpso2}, {st3, st4})	st-cluster 2 = {cB}	0.03	0
	pair3 = ({fpso3}, {st5})	st-cluster3 = {cC}	0.045	0
	pair4 = ({fpso4}, {st6})	st-cluster4 = {cD}	0.06	0

Table 3: Cluster schemes for instance 4F-4V-10ST-6CT-5CDU-8C-1P-15D.

Number Clusters	(platform-cluster, st-cluster) pair	Crudes st-cluster	Target $tg_{k,rr}$ st-cluster	Deviat. st-cluster
2	pair1 = ({fpso1}, {st1, st2})	st-cluster 1 = {cA}	0.01	0
	pair2 = ({fpso2, fpso3, fpso4}, {st3, st4, st5, st6, st7, st8, st9, st10})	st-cluster 1 = {cB, cC, cD}	0.045	0.6
3	pair1 = ({fpso1}, {st1, st2})	st-cluster 1 = {cA}	0.01	0
	pair2 = ({fpso2}, {st3, st4, st5})	st-cluster 2 = {cB}	0.03	0
	pair3 = ({fpso3, fpso4}, {st6, st7, st8, st9, st10})	st-cluster3 = {cC, cD}	0.045	0.3
4	pair1 = ({fpso1}, {st1, st2})	st-cluster 1 = {cA}	0.01	0
	pair2 = ({fpso2}, {st3, st4})	st-cluster 2 = {cB}	0.03	0
	pair3 = ({fpso3}, {st5, st6, st7})	st-cluster3 = {cC}	0.045	0
	pair4 = ({fpso4}, {st8, st9, st10})	st-cluster4 = {cD}	0.06	0

6.2. Solution of Clustered Instances

Here, an analysis is conducted based on the statistics and solution of the instances 2F-2V-2ST-2CT-1CDU-2C-1P-15D, 4F-4V-10ST-6CT-5CDU-8C-1P-15D and 4F-4V-10ST-6CT-5CDU-8C-1P-15D that were re-defined according with the clusters defined in Section 6.1. As advocated above, the use of a clustering scheme for the original instance will produce a new instance with a more restricted set of possible operations on the offshore side.

Figure 4 illustrates the comparison between the set of all possible operations and clustered operations on the offshore side of instance 2F-2V-2ST-2CT-1CDU-2C-1P-15D. In the original instance (Figure 4 (a)), the two vessels are allowed to travel between the terminal and the two FPSOs, and unload crudes cA and cB in both storage tanks ($ST1$ and $ST2$). By applying the clustering scheme with two clusters (see Table 1) in the original instance proposed in (de Assis et al., 2019), a restricted set of offshore operations is derived (Figure 4 (b)). In this case, a vessel (e.g., *Vessel1*) can travel only between $FPSO1$ and the terminal, and unload crude oil from $FPSO1$ only into storage tank $ST1$. The same holds for the second vessel regarding $FPSO2$ and $ST2$.

For the example in Figure 4 (b), crudes from $FPSO1$ and $FPSO2$ will never get mixed in the storage tanks $ST1$ and $ST2$. Nevertheless, since the connections between storage and charging tanks remain the same, crudes can be normally mixed in the charging tanks to reach the demanded crude specification. Notice that one of the main consequences of using clusters will be the decrease on the number of variables and constraints.

Table 4 indicates the number of clusters, the MILP MIPGAP used in the MILP-NLP solution strategy, the number of variables and constraints, the best known solution (from de Assis et al. (2019)), the solution of the clustered instance, the GAP from the best known solution, and the CPU time.

Table 4: Statistics and solution for clustered instances.

Instances	Num. Clus.	MILP MIPGAP	MILP Stat. Total Vars.	MILP Stat. Total Cons.	MILP Stat. Binary Vars.	NLP Stat. Non-Linear Cons.	Best Known Solution, CPU [10^3, s]	MILP-NLP Solution Sol. [10^3]	MILP-NLP Solution [1]GAP	MILP-NLP Solution CPU Time [s]
2F-2V-2ST 2CT-1CDU 2C-1P-15D	2	0%	1 380	1 875	531	180	22 800, 8s	22 711	0.4%	0.25s
4F-4V-6ST	2	0%	12 900	8 311	4 351	1 920		inf	-	-
4CT-3CDU	3	0%	8 895	6 815	2 098	1 920	38 563, 446s	38 500	0.16%	17s
8C-1P-15D	4	0%	7 380	6 087	1 346	1 920		37 589	2.5%	11s
4F-4V-10ST	2	0.5%	19 650	11 444	7 272	2 880		55 181	1.75%	300s
6CT-5CDU	3	0.5%	13 335	9 534	3 301	2 880	56 167, 8 843s	54 681	2.64%	181s
8C-1P-15D	4	0.5%	10 920	8 558	1 987	2 880		54 285	3.35%	17s

[1] GAP = $\frac{BestSol. - Sol.}{BestSol.}$ 100.

In addition, Table 5 depicts the best known results for the instances considered by de Assis et al. (2019). Also, the table shows the number of variables and constraints, the MILP MIPGAP used in the MILP-NLP solution strategy and the CPU time. Notice that for instance 4F-4V-6ST-4CT-3CDU-8C-1P-15D, experiments were conducted considering both MILP MILPGAPs of 0% and 1.5%, while for instances 2F-2V-2ST-2CT-1CDU-2C-1P-15D and 4F-4V-10ST-6CT-5CDU-8C-1P-15D it was considered a MILP MIPGAP of 0% and 3% respectively.

Table 5: Statistics and solution for the original instances considered in (de Assis et al., 2019).

Instances	Total Vars.	Total Cons.	Binary Vars.	Non-Linear Cons.	MILP MIPGAP	Best Solution	CPU Time [s]
2F-2V-2ST-2CT-1CDU-2C-1P-15D	2 160	2 483	1 170	180	0%	22 800	8
4F-4V-6ST-4CT-3CDU-8C-1P-15D	17 490	11 593	8 310	1 920	0%	38 460.5	2 018
					1.5%	38 563.6	446
4F-4V-10ST-6CT-5CDU-8C-1P-15D	29 175	15 954	15 000	2 880	3%	56 167.9	8 843

The results from Table 4 suggest the following remarks:

1. **Number of Clusters, Problem Size and CPU Time**. As the number of clusters increase, the instance gets more restricted, and there are fewer routes covered by the vessels assigned to handle the flow of crudes in a (platform-cluster, st-cluster) pair. Further, by clustering the instance, there is a limitation of the offloading and unloading operations.

   The constraints on vessel trips, offloading and unloading operations have an impact on the number of variables and constraints when solving the clustered instances. By comparing results in Tables 5 and 4 it is possible to see in numbers the decrease in problem size. For example, the original instance 4F-4V-10ST-6CT-5CDU-8C-1P-15D goes from $29\,175$ to $10\,920$ variables in the version considering 4 clusters. From this total number of variables, the number of binaries drops from $15\,000$ to $1\,987$, which is a decrease of more than 7.5 times. Likewise, the overall number of constraints decreases from $15\,954$ to $8\,558$. The number of non-linear constraints remain the same since the STs-CTs and CTs-CDUs connections do not change.

   The CPU time follows the same trend as the number of constraints and variables. For a reduced number of clusters, the clustered instance gets closer to the original one, and therefore the solution time is higher. As the number of clusters increases, the number of variables and constraints decreases, and so the solution time. The highest drop happens for instance 4F-4V-10ST-6CT-5CDU-8C-1P-15D, with solution time going from more than $8\,000$ seconds for the original instance to less than 20 seconds when considering 4 clusters.

2. **Number of Clusters and MILP-NLP Solution**.
   Table 4 shows the GAP between the clustered solution and the best known solution found for the original instance.

   The results show that the use of clusters affects the solution, besides a computational gain with the reduction of variables, constraints and CPU time. In addition, the GAP tends to increase with the number of the clusters, reaching maximum values close to 3% in the worst-case scenarios (i.e., instance 4F-4V-10ST-6CT-5CDU-8C-1P-15D with 4 clusters) and lower than 1% in the best one (i.e., 4F-4V-6ST-4CT-3CDU-8C-1P-15D with 3 clusters).

   The effect on the results can be explained by the fact that (a) the linear approximation for the blending constraints, proposed in Section 5, is not as strong as the use of McCormick envelopes to relax the same constraints (McCormick, 1976); (b) the use of clusters restricts the problem and potentially excludes feasible solutions.

6.3. Solution of New Instances

As mentioned in the previous section, one can take advantage of the computational gains of using clusterization for solving larger instances.

Table 6 reports the number of variables and constraints for new instances 4F-4V-10ST-6CT-5CDU-8C-4P-15D (see Figure 5) and 8F-8V-10ST-6CT-5CDU-8C-4P-15D (see Figure 6). Instance 4F-4V-10ST-6CT-5CDU-8C-4P-15D considers 4 crude properties (i.e., S, T, U and V with values described in Figure 5) and all possible connections between storage and charging tanks. On the other hand, instance 8F-8V-10ST-6CT-5CDU-8C-4P-15D, (see Figure 6), takes into account the same 4 crude properties and connections between tanks, but also extends the number of platforms, vessels, the demand of the CDUs and the storage capacity of the FPSOs.

Table 7 presents the statistics and solution of instances 4F-4V-10ST-6CT-5CDU-8C-4P-15D for 2, 3, and 4 cluster schemes, and 8F-8V-10ST-6CT-5CDU-8C-4P-15D for 4 and 6 clusters schemes.

Table 6: Statistics for new instances.

Instances	Total Vars.	Total Cons.	Binary Vars.	Non-Linear Cons.
4F-4V-10ST-6CT-5CDU-8C-4P-15D	33 075	23 544	15 390	8 400
8F-8V-10ST-6CT-5CDU-8C-4P-15D	68 040	39 814	31 590	8 400

Tables 6 and 7 show a significant decrease in the number of variables and constraints when comparing the original instances and their clustered versions. This decrease has a direct effect on the CPU time. As mentioned in de Assis et al. (2019), for instances like 4F-4V-6ST-4CT-3CDU-8C-1P-15D and 4F-4V-10ST-6CT-5CDU-8C-1P-15D, off-the-shelf MINLP solvers were not able to find a feasible solution within a maximum CPU time of 10 hours. The same happens with the new instances 4F-4V-6ST-4CT-3CDU-8C-4P-15D and 8F-8V-10ST-6CT-5CDU-8C-4P-15D.

Table 7: Statistics and solution for new instances with clusterization.

			MILP Stat.			NLP Stat.	MILP-NLP Solution		
Instances	Num. Clus.	MILP MIPGAP	Total Var.	Total Cons.	Binary Vars.	Non-Linear Cons.	MILP Sol. $[10^3]$	NLP Sol. $[10^3]$	CPU Time [s]
4F-4V-10ST-6CT 5CDU-8C-4P-15D	2	1%	25 050	20 793	7 107	8 400	56 564	53 758	884
	3	1%	19 785	19 059	3 946	8 400	56 452	55 661	522
	4	1%	17 370	18 083	2 640	8 400	56 407	53 085	180
8F-8V-10ST-6CT 5CDU-8C-4P-15D	4	3%	28 530	24 261	7 811	8 400	59 589	59 230	8 223
	6	3%	22 440	21 739	4 527	8 400	58 250	56 590	3 271

Depending on the clustering scheme, instance 8F-8V-10ST-6CT-5CDU-8C-4P-15D presents a decrease on the number of binary variables from $31\,590$ to $4\,527$, mainly resulting from the restriction on vessel trips, and offloading and unloading operations when using clusterization. When comparing the statistics of instance 8F-8V-10ST-6CT-5CDU-8C-4P-15D for clustering schemes with 4 and 6 clusters, the number of binary variables decreases to almost one half, which is reflected in the CPU time that drops from 2.28 hours to 0.9 hour. Although not providing the same solution quality (i.e., from $59\,230$ to $56\,590$), there are clear computational gains on using clusters.

7. Conclusion

As highlighted in the introduction, the main goal of OMCOS is to coordinate the activities of vessel trips and crude oil operations in the terminal in order to supply crudes to the CDUs. Nevertheless, without clear rules or constraints, all mixtures of crudes are allowed in storage tanks. This may lead to mixtures of crudes with dissimilar properties, which might be non-desirable.

With the goal of coordinating how crudes can be mixed inside storage tanks, this paper proposed an MILP formulation to define the optimal cluster of crudes and resources, such that the difference among their properties is as low as possible. The use of clusters offers the following benefits: (a) reduces the number of routes for the vessels; (b) simplifies offloading and unloading operations; and (c) imposes rules for crude mixtures in clusters of storage tanks that minimize property variations.

The solution of the clustering formulation produces: (a) more restricted problem instances, and (b) lower and upper bounds on crude properties inside each storage tank. These bounds are used to linearize the blending constraints and derive an MILP linearization of the original MINLP, which is used in the MILP-NLP solution strategy.

Although possibly eliminating feasible solutions, the use of clusters allows to reach solutions with a compatible quality, but with far fewer variables and constraints, and at much less computational cost.

Acknowledgment

This work was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) under the project PrInt CAPES-UFSC.

References

de Assis, L.S., Camponogara, E., Menezes, B.C., Grossmann, I.E., 2019. An MINLP formulation for integrating the operational management of crude oil supply. Computers & Chemical Engineering 123, 110–125.

de Assis, L.S., Camponogara, E., Zimberg, B., Ferreira, E., Grossmann, I.E., 2017. A piecewise McCormick relaxation-based strategy for scheduling operations in a crude oil terminal. Computers & Chemical Engineering 106, 309–321.

Castro, P.M., 2015. Tightening piecewise McCormick relaxations for bilinear problems. Computers & Chemical Engineering 72, 300–311.

Cerdá, J., Pautasso, P.C., Cafaro, D.C., 2018. Optimization approaches for efficient crude blending in large oil refineries. Industrial & Engineering Chemistry Research 57, 8484–8501.

Dondo, R., Cerdá, J., 2007. A cluster-based optimization approach for the multi-depot heterogeneous fleet vehicle routing problem with time windows. European Journal of Operational Research 176, 1478–1507.

Drud, A., 1985. CONOPT: A GRG code for large sparse dynamic nonlinear optimization problems. Mathematical Programming 31, 153–191.

Ganesh, K., Narendran, T., 2007. CLOVES: A cluster-and-search heuristic to solve the vehicle routing problem with delivery and pick-up. European Journal of Operational Research 178, 699–717.

Gillett, B.E., Miller, L.R., 1974. A heuristic algorithm for the vehicle-dispatch problem. Operations Research 22, 340–349.

IBM, 2013. IBM ILOG CPLEX Optimization Studio. URL: `http://www.cplex.com`.

Kaufman, L., Rousseeuw, P., 1987. Statistical Data Analysis Based on the L_1-Norm and Related Methods. North-Holland. chapter Clustering by means of Medoids. pp. 405–416.

Kelly, J.D., Menezes, B.C., Engineer, F., Grossmann, I.E., 2017a. Crude-oil blend scheduling optimization of an industrial-sized refinery: a discrete-time benchmark. Foundations of Computer Aided Process Operations .

Kelly, J.D., Menezes, B.C., Grossmann, I.E., Engineer, F., 2017b. Feedstock storage assignment in process industry quality problems. Foundations of Computer Aided Process Operations: Tucson, United States .

Lee, I.G., Chung, S.H., Yoon, S.W., 2020. Two-stage storage assignment to minimize travel time and congestion for warehouse order picking operations. Computers & Industrial Engineering 139, 106129.

Liu, C.M., 1999. Clustering techniques for stock location and order-picking in a distribution center. Computers & Operations Research 26, 989 – 1002.

McCormick, G.P., 1976. Computability of global solutions to factorable nonconvex programs: Part I — Convex underestimating problems. Mathematical Programming 10, 147–175.

Méndez, C.A., Grossmann, I.E., Harjunkoski, I., Kaboré, P., 2006. A simultaneous optimization approach for off-line blending and scheduling of oil-refinery operations. Computers & Chemical Engineering 30, 614–634.

Mogale, D.G., Ghadge, A., Kumar, S.K., Tiwari, M.K., 2019. Modelling supply chain network for procurement of food grains in india. International Journal of Production Research 0, 1–20.

Mulvey, J.M., Beck, M.P., 1984. Solving capacitated clustering problems. European Journal of Operational Research 18, 339 – 348.

Nambirajan, R., Mendoza, A., Pazhani, S., Narendran, T.T., Ganesh, K., 2016. CARE: Heuristics for two-stage multi-product inventory routing problems with replenishments. Computers & Industrial Engineering 97, 41–57.

Negreiros, M., Palhano, A., 2006. The capacitated centred clustering problem. Computers & Operations Research 33, 1639 – 1663.

Nemhauser, G.L., Wolsey, L.A., 1988. Integer and Combinatorial Optimization. John Wiley & Sons.

Papageorgiou, D.J., Trespalacios, F., 2018. Pseudo basic steps: bound improvement guarantees from Lagrangian decomposition in convex disjunctive programming. EURO Journal on Computational Optimization 6, 55–83.

Qi, M., Lin, W.H., Li, N., Miao, L., 2012. A spatiotemporal partitioning approach for large-scale vehicle routing problems with time windows. Transportation Research Part E: Logistics and Transportation Review 48, 248–257.

Ramkumar, N., 2011. Mathematical models and heuristics for a class of inventory routing problems. Master's thesis. Department of Management Studies, Indian Institute of Technology Madras.

Appendix A. Linearization of the Objective Function

In Eq. (28), the deviations from target values for each property k are expressed in terms of absolutes values, and normalized by the range of maximum and minimum values of the corresponding crude property. By definition, the value $PR_{k,c}$ will assume values within the interval $[\underline{PR}_k,\ \overline{PR}_k]$. Thus, the minimization of the second term of Eq. (28) ensures that, at optimality, the target value $tg_{k,rr}$ will also be within the same bounds. For instance, if $tg_{k,rr} < \underline{PR}_k$ then the values $PR_{k,c} - tg_{k,rr} > PR_{k,c} - \underline{PR}_k$ for all c and therefore the objective would be reduced by setting $tg_{k,rr} = \underline{PR}_k$. Similar reasoning leads us to deduce that $tg_{k,rr} \leq \overline{PR}_k$, and thereby the term $\frac{|PR_{k,c}-tg_{k,rr}|}{\overline{PR}_k-\underline{PR}_k} \in [0,1]$.

In order to linearize Eq. (28), the auxiliary variable $deviation_{k,c,rr} \in \mathbb{R}_+$, $k \in \mathcal{K}$, $c \in \mathcal{C}$, and $rr \in \mathcal{STCS}$, is introduced as an upper bound on the value of term $\frac{|PR_{k,c}-tg_{k,rr}|}{\overline{PR}_k-\underline{PR}_k}$. Eqs. (A.1) and (A.2) ensure the consistency of the upper bound induced by $deviation_{k,c,rr}$.

$$\frac{PR_{k,c} - tg_{k,rr}}{\overline{PR}_k - \underline{PR}_k} \leq deviation_{k,c,rr}, k \in \mathcal{K}, c \in \mathcal{C}, rr \in \mathcal{STCS}, \tag{A.1}$$

$$-\frac{(PR_{k,c} - tg_{k,rr})}{\overline{PR}_k - \underline{PR}_k} \leq deviation_{k,c,rr}, k \in \mathcal{K}, c \in \mathcal{C}, rr \in \mathcal{STCS}. \tag{A.2}$$

This leads the second term to be replaced by Eq. (A.3), which remains nonlinear though.

$$\sum_{k\in\mathcal{K}}\sum_{c\in\mathcal{C}}\sum_{rr\in\mathcal{STCS}} deviation_{k,c,rr}\ crudeSTCS_{c,rr}. \tag{A.3}$$

Because the target value $tg_{k,rr} \in [\underline{PR}_k, \overline{PR}_k]$ and $deviation_{k,c,rr}$ is an upper bound for $|PR_{k,c} - tg_{k,rr}|/(\overline{PR}_k - \underline{PR}_k)$, the minimization of the deviations in Eq. (A.3) ensure that $deviation_{k,c,rr} \in [0,1]$.

Notice that the variables $deviation_{k,c,rr}$ are defined only, and only if, crude c is delivered to the st-cluster rr, a condition flagged by $crudeSTCS_{c,rr} = 1$. To take advantage of this condition, another auxiliary variable $\widehat{deviation}_{k,c,rr} \in \mathbb{R}_+$ is introduced to assume the value $deviation_{k,c,rr}$ when crude c is received by st-cluster rr. This is implemented for all $k \in \mathcal{K}$, $c \in \mathcal{C}$, and $rr \in \mathcal{STCS}$ as follows:

$$\begin{cases} \widehat{deviation}_{k,c,rr} & \geq 0, \\ \widehat{deviation}_{k,c,rr} & \leq deviation_{k,c,rr}, \\ \widehat{deviation}_{k,c,rr} & \leq crudeSTCS_{c,rr}, \\ \widehat{deviation}_{k,c,rr} & \geq deviation_{k,c,rr} - (1 - crudeSTCS_{c,rr}). \end{cases} \tag{A.4}$$

The set of Eqs. (A.4) define that if crude c is not assigned to storage tank-cluster rr ($crudeSTCS_{c,rr} = 0$), variable $\widehat{deviation}_{k,c,rr}$ is set to zero. Likewise, if crude c is assigned to cluster rr, variable $\widehat{deviation}_{k,c,rr}$ is set to $deviation_{k,c,rr}$. Finally, variable $\widehat{deviation}_{k,c,rr}$ replaces the bilinear term $deviation_{k,c,rr}\ crudeSTCS_{c,rr}$ in Eq. (A.3)

The third term of the objective balances the relation between the number of storage tanks and platforms in a (platform-cluster, st-cluster) pair. For instance, consider 2 (platform-cluster, st-cluster) pairs, 4 platforms and 8 storage tanks. If the first platform-cluster has 1 platform and the second 3 platforms, then the first st-cluster would contain 2 storage tanks and the second 6 storage tanks to ensure a balanced assignment. Notice that the third term needs to be linearized since it has non-linear fractions and the modulus operator.

Consider a (platform-cluster, st-cluster) pair, with platform-cluster $rr \in \mathcal{RFCS}$ and st-cluster $rrr \in \mathcal{STCS}$. This means that pair $(rr, rrr) \in \mathcal{N}$. Also, variable $rfCluster_{rr}$ is the number of platforms assigned to platform-cluster rr and $stCluster_{rrr}$ is the number of storage tanks assigned to st-cluster rrr. Notice that at least one platform must be assigned to a platform-cluster and at least one storage tank must be assigned to a st-cluster, which means $rfCluster_{rr} \geq 1$ and $stCluster_{rrr} \geq 1$. Moreover, these variables can only assume integer values, which implies that both $rfCluster_{rr},\ stCluster_{rrr} \in \mathbb{N}^*_+$. These variables are defined by Eqs. (A.5) and (A.6).

$$rfCluster_{rr} = \sum_{r \in \mathcal{RF}} assignRFCS_{r,rr},\ rr \in \mathcal{RFCS}. \tag{A.5}$$

$$stCluster_{rrr} = \sum_{r \in \mathcal{RS}} assignSTCS_{r,rrr},\ rrr \in \mathcal{STCS}. \tag{A.6}$$

Then, as defined by Eq. (A.7), variable $proportion_{rr,rrr}$ assumes the value of the ratio between the number of storage tanks and platforms in (platform-cluster, st-cluster) pair $(rr, rrr) \in \mathcal{N}$. This variable is defined as $proportion_{rr,rrr} \in \mathbb{R}_{+}$.

$$proportion_{rr,rrr} = \frac{\sum\limits_{r \in \mathcal{RS}} assignSTCS_{r,rrr}}{\sum\limits_{r \in \mathcal{RF}} assignRFCS_{r,rr}} = \frac{stCluster_{rrr}}{rfCluster_{rr}}, \quad (rr, rrr) \in \mathcal{N} : rr \in \mathcal{RFCS}. \tag{A.7}$$

By introducing the supporting variable $proportion_{rr,rrr}$, the third term of the objective can be cast as:

$$\sum_{\substack{(rr,rrr) \in \mathcal{N}: \\ rr \in \mathcal{RFCS}}} \sum_{\substack{(rrrr,rrrrr) \in \mathcal{N}: \\ rrrr \in \mathcal{RFCS}}} \left| proportion_{rr,rrr} - proportion_{rrrr,rrrrr} \right|. \tag{A.8}$$

To linearize Eq. (A.8), the auxiliary variable $proportionDiff^{rrrr,rrrrr}_{rr,rrr} \in \mathbb{R}_{+}$, $rr,\ rrrr \in \mathcal{RFCS}$, $rrr,\ rrrrr \in \mathcal{STCS}$, and $(rr, rrr),\ (rrrr, rrrrr) \in \mathcal{N}$, is used to upper bound the value of term $|proportion_{rr,rrr} - proportion_{rrrr,rrrrr}|$. The consistency of the upper bound induced by $proportionDiff^{rrrr,rrrrr}_{rr,rrr}$ is ensured by Eqs. (A.9) and (A.10).

$$proportion_{rr,rrr} - proportion_{rrrr,rrrrr} \leq proportionDiff^{rrrr,rrrrr}_{rr,rrr}, \quad (rr, rrr) \in \mathcal{N},\ (rrrr, rrrrr) \in \mathcal{N} : rr \in \mathcal{RFCS}, rrrr \in \mathcal{RFCS}. \tag{A.9}$$

$$-\,proportion_{rr,rrr} + proportion_{rrrr,rrrrr} \leq proportionDiff^{rrrr,rrrrr}_{rr,rrr}, \quad (rr, rrr) \in \mathcal{N},\ (rrrr, rrrrr) \in \mathcal{N} : rr \in \mathcal{RFCS}, rrrr \in \mathcal{RFCS}. \tag{A.10}$$

Finally, the third term of the objective is replaced by Eq. (A.11).

$$\sum_{\substack{(rr,rrr) \in \mathcal{N}: \\ rr \in \mathcal{RFCS}}} \sum_{\substack{(rrrr,rrrrr) \in \mathcal{N}: \\ rrrr \in \mathcal{RFCS}}} proportionDiff^{rrrr,rrrrr}_{rr,rrr}. \tag{A.11}$$

Notice that Eq. (A.7) is non-linear and can be reformulated as in Eq. (A.12), which has the bilinear term $proportion_{rr,rrr}\ rfCluster_{rr}$.

$$proportion_{rr,rrr}\ rfCluster_{rr} = stCluster_{rrr},\ (rr, rrr) \in \mathcal{N} : rr \in \mathcal{RFCS}. \tag{A.12}$$

In order to linearize the bilinear term $proportion_{rr,rrr}\ rfCluster_{rr}$, consider the following elements:

- From previous definitions, $proportion_{rr,rrr} \in \mathbb{R}_+$. Also, the total number of platforms assigned to platform-cluster rr ($rfCluster_{rr}$) and the total number of storage tanks assigned to st-cluster rrr ($stCluster_{rrr}$) are natural numbers (i.e., $rfCluster_{rr},\ stCluster_{rrr} \in \mathbb{N}$).

- Set $\mathcal{J} = \{1, \ldots, (|\mathcal{RF}| - |\mathcal{RFCS}| + 1)\}$. Cardinality $|\mathcal{J}|$ is equal to the maximum number of platforms that can be assigned to a platform-cluster. For instance, if there are six platforms (i.e., $|\mathcal{RF}| = 6$) and two platform-clusters (i.e., $|\mathcal{RFCS}| = 2$), at least one platform must be assigned to each platform-cluster, and a maximum of five platforms (i.e., $|\mathcal{RF}| - |\mathcal{RFCS}| + 1 = 5$) can be assigned to a platform-cluster.

- Binary variable $z_{j,rr} \in \mathbb{B},\ j \in \mathcal{J}, rr \in \mathcal{RFCS}$ is 1 if integer value j in set $\mathcal{J}$, which represents the number of platforms $rfCluster_{rr}$ assigned to platform-cluster rr, is selected. As stated by Eq. (A.13), for each platform-cluster rr, only one integer value $j \in \mathcal{J}$ can be selected. Then, as defined in Eq. (A.14), the integer variable $rfCluster_{rr}$ can be stated as a sum of binary variables multiplied by the integer value j.

$$\sum_{j \in \mathcal{J}} z_{j,rr} = 1,\ rr \in \mathcal{RFCS}. \tag{A.13}$$

$$rfCluster_{rr} = \sum_{j \in \mathcal{J}} j \cdot z_{j,rr},\ rr \in \mathcal{RFCS}. \tag{A.14}$$

  Variable $rfCluster_{rr}$ can be replaced in Eq. (A.12) leading to Eq. (A.15). Notice that in Eq. (A.15), the bilinear term $z_{j,rr} \cdot proportion_{rr,rrr}$ has a binary and a real variable, which can be easily linearized.

$$(\sum_{j \in \mathcal{J}} j \cdot z_{j,rr}) proportion_{rr,rrr} = stCluster_{rrr},\quad (rr, rrr) \in \mathcal{N} : rr \in \mathcal{RFCS}. \tag{A.15}$$

- Auxiliary variable $\theta_{j,rr,rrr} \in \mathbb{R} \geq 0,\ j \in \mathcal{J},\ (rr, rrr) \in \mathcal{N} : rr \in \mathcal{RFCS}$ can be defined such that the bilinear term $z_{j,rr} \cdot proportion_{rr,rrr} = \theta_{j,rr,rrr}$. Eq. (A.16) corresponds to the linearization of bilinear term $\theta_{j,rr,rrr}$, valid for all $j \in \mathcal{J}$ and $(rr, rrr) \in \mathcal{N} : rr \in \mathcal{RFCS}$.

$$\begin{cases} \theta_{j,rr,rrr} & \geq 0, \\ \theta_{j,rr,rrr} & \leq proportion_{rr,rrr}, \\ \theta_{j,rr,rrr} & \leq z_{j,rr}|\mathcal{RS}|, \\ \theta_{j,rr,rrr} & \geq proportion_{rr,rrr} - |\mathcal{RS}|(1 - z_{j,rr}). \end{cases} \tag{A.16}$$

Variable $z_{j,rr} = 0$ drives $\theta_{j,rr,rrr} = 0$. However, if $z_{j,rr} = 1$, $\theta_{j,rr,rrr}$ gets bounded by the cardinality of the set of storage tanks $|\mathcal{RS}|$ and assumes the value of $proportion_{rr,rrr}$.

After replacing $\theta_{j,rr,rrr}$ in Eq. (A.15), it is then reformulated as Eq. (A.17).

$$\sum_{j \in \mathcal{J}} j \cdot \theta_{j,rr,rrr} = stCluster_{rrr},\ (rr, rrr) \in \mathcal{N} : rr \in \mathcal{RFCS}. \tag{A.17}$$

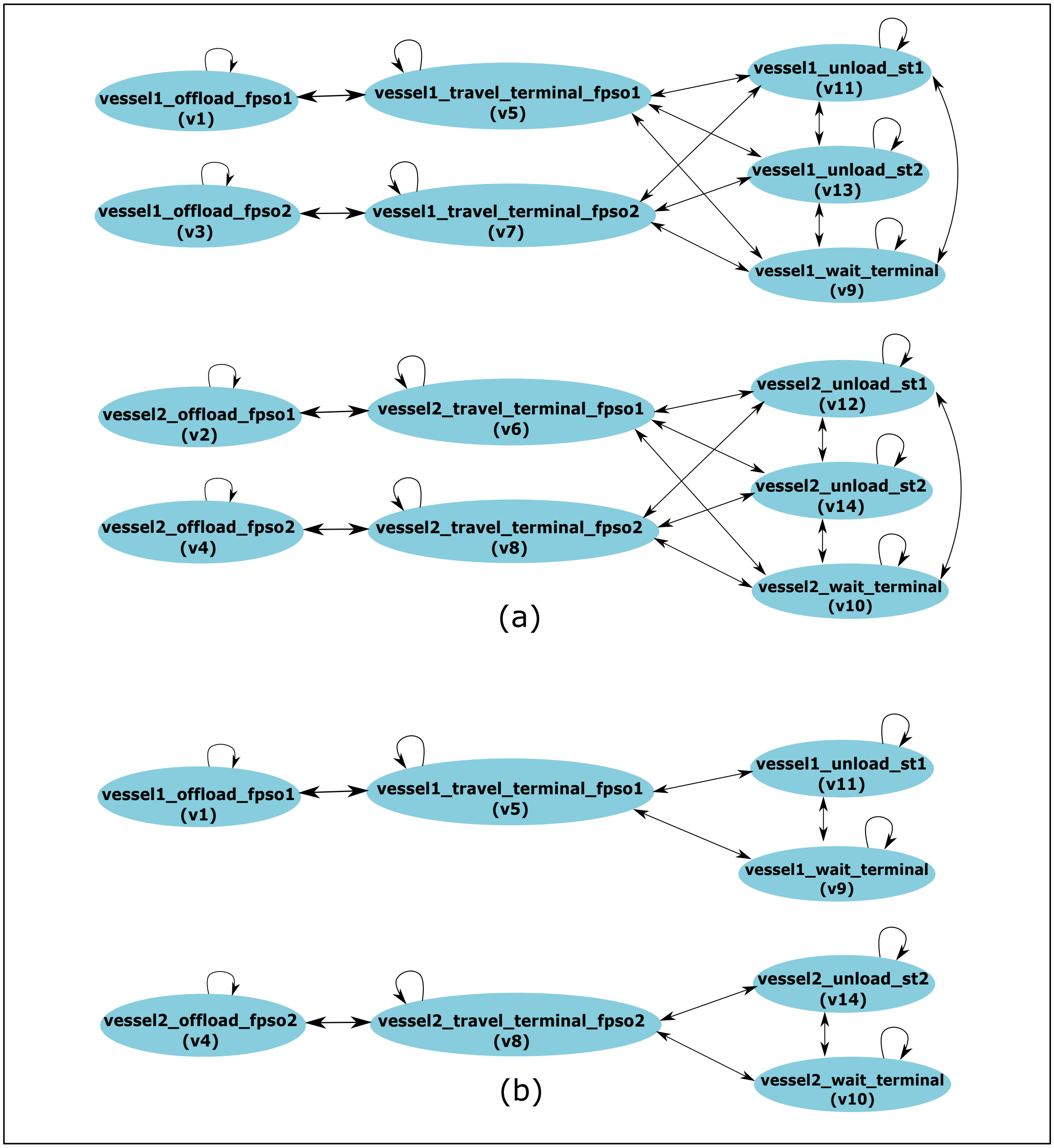
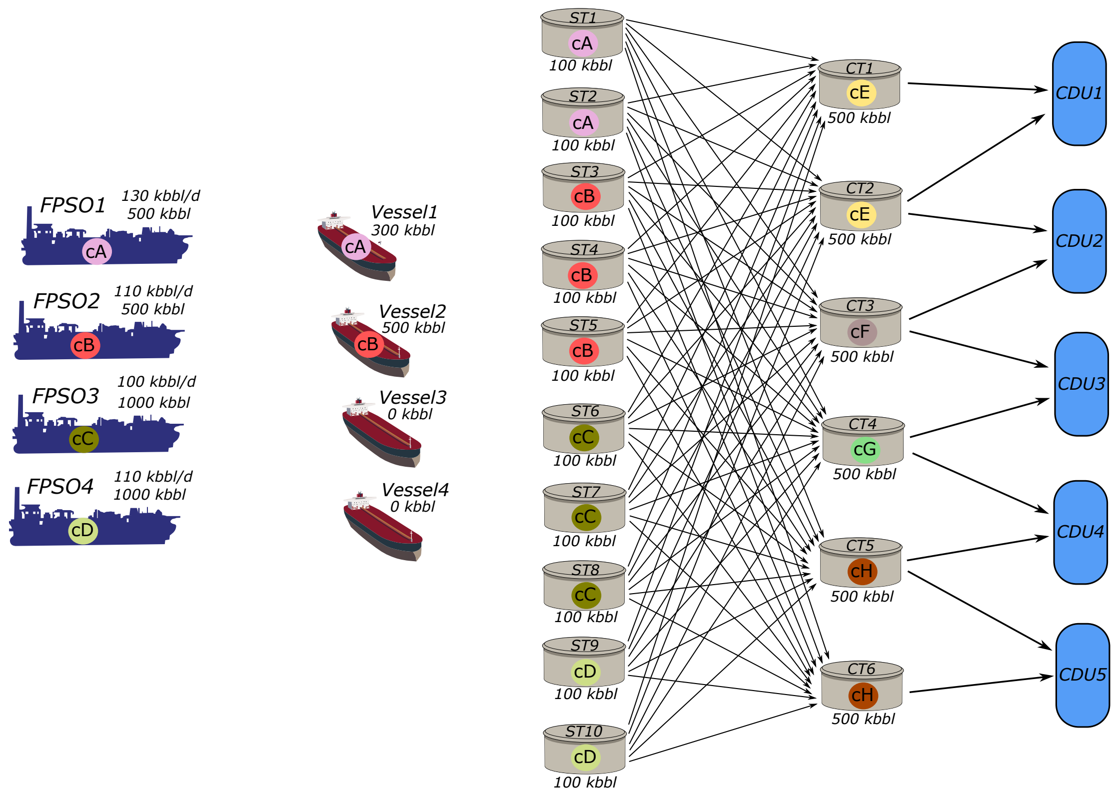


Figure 4: (a) Graph for the flow of offshore operations in the original instance 2F-2V-2ST-2CT-1CDU-2C-1P-15D; (b) Graph for the flow of offshore operations in the clustered instance 2F-2V-2ST-2CT-1CDU-2C-1P-15D.



Crude/Property	S	T	U	V	$/kbbl
cA	0.010	0.0050	0.200	0.150	10,000
cB	0.030	0.0150	0.067	0.050	8,000
cC	0.045	0.0225	0.044	0.033	6,500
cD	0.060	0.0300	0.033	0.025	5,000
cE	0.012	0.0060	0.166	0.125	9,500
cF	0.026	0.0130	0.077	0.057	8,500
cG	0.041	0.0205	0.049	0.036	7,000
cH	0.056	0.0280	0.035	0.026	5,500

Capacity Bounds	kbbl
FPSO	[200, 1500]
VESSEL	[0, 1000]
ST	[0, 1000]
CT	[0, 1000]

Flow Rate Bounds	kbbl/d
FPSO-VESSEL	[0, 500]
VESSEL-ST	[0, 500]
ST-CT	[0, 500]
CT-CDU	[50, 500]

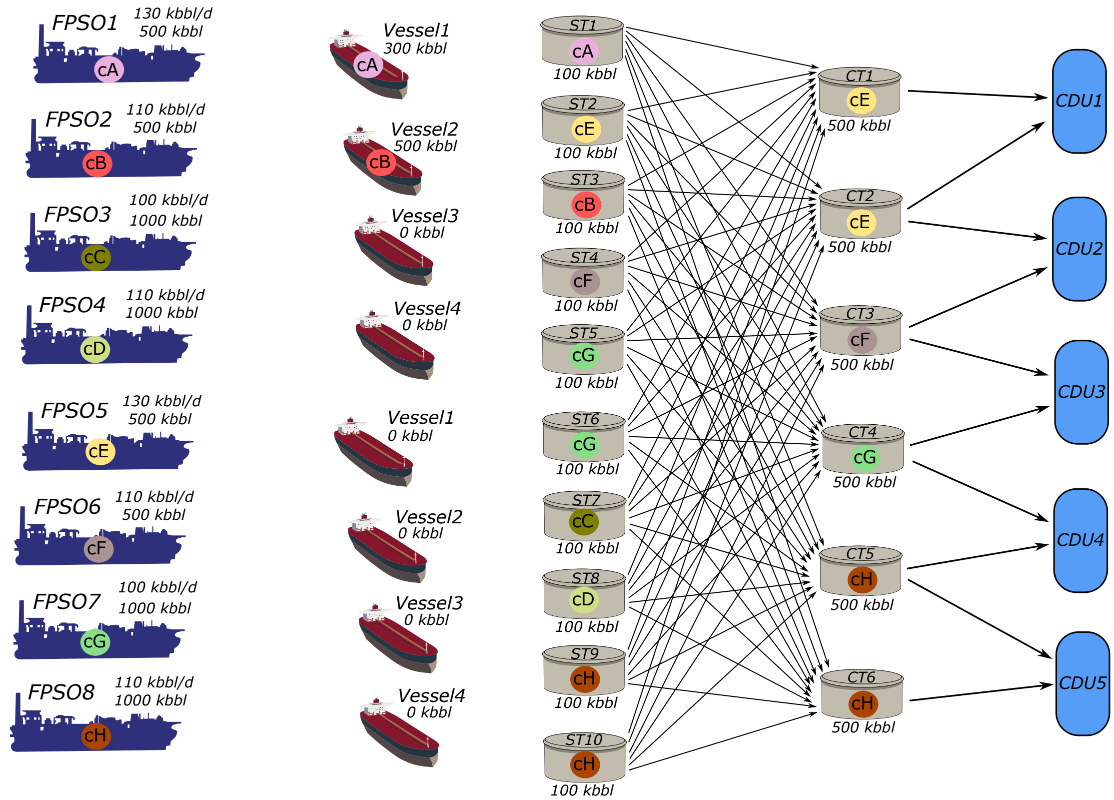
Crude Demand	kbbl
CT1	[800, 1200]
CT2	[800, 1200]
CT3	[800, 1200]
CT4	[800, 1200]
CT5	[800, 1200]
CT6	[800, 1200]

	Initial Operation
Vessel1	unload-ST1
Vessel2	wait-Terminal
Vessel3	wait-Terminal
Vessel4	wait-Terminal

	S	T	U	V
CT1	[0.005, 0.015]	[0.0025, 0.075]	[0.133, 0.400]	[0.100, 0.300]
CT2	[0.005, 0.015]	[0.0025, 0.075]	[0.133, 0.400]	[0.100, 0.300]
CT3	[0.020, 0.030]	[0.0100, 0.0150]	[0.066, 0.100]	[0.050, 0.075]
CT4	[0.035, 0.045]	[0.0175, 0.0225]	[0.044, 0.057]	[0.033, 0.042]
CT5	[0.050, 0.060]	[0.0250, 0.0300]	[0.033, 0.040]	[0.025, 0.030]
CT6	[0.050, 0.060]	[0.0250, 0.0300]	[0.033, 0.040]	[0.025, 0.030]

Planning Horizon = 15 days

Figure 5: Network for instance 4F-4V-10ST-6CT-5CDU-8C-4P-15D.



Crude/Property	S	T	U	V	$/kbbl
cA	0.010	0.0050	0.200	0.150	10,000
cB	0.030	0.0150	0.067	0.050	8,000
cC	0.045	0.0225	0.044	0.033	6,500
cD	0.060	0.0300	0.033	0.025	5,000
cE	0.012	0.0060	0.166	0.125	9,500
cF	0.026	0.0130	0.077	0.057	8,500
cG	0.041	0.0205	0.049	0.036	7,000
cH	0.056	0.0280	0.035	0.026	5,500

Capacity Bounds	kbbl
FPSO	[200, 2500]
VESSEL	[0, 1000]
ST	[0, 1000]
CT	[0, 1000]

Flow Rate Bounds	kbbl/d
FPSO-VESSEL	[0, 500]
VESSEL-ST	[0, 500]
ST-CT	[0, 500]
CT-CDU	[50, 500]

Crude Demand	kbbl
CT1	[500, 1500]
CT2	[500, 1500]
CT3	[500, 1500]
CT4	[500, 1500]
CT5	[500, 1500]
CT6	[500, 1500]

	S	T	U	V
CT1	[0.005, 0.015]	[0.0025, 0.075]	[0.133, 0.400]	[0.100, 0.300]
CT2	[0.005, 0.015]	[0.0025, 0.075]	[0.133, 0.400]	[0.100, 0.300]
CT3	[0.020, 0.030]	[0.0100, 0.0150]	[0.066, 0.100]	[0.050, 0.075]
CT4	[0.035, 0.045]	[0.0175, 0.0225]	[0.044, 0.057]	[0.033, 0.042]
CT5	[0.050, 0.060]	[0.0250, 0.0300]	[0.033, 0.040]	[0.025, 0.030]
CT6	[0.050, 0.060]	[0.0250, 0.0300]	[0.033, 0.040]	[0.025, 0.030]

	Initial Operation
Vessel1	unload-ST1
Vessel2	wait-Terminal
Vessel3	wait-Terminal
Vessel4	wait-Terminal
Vessel5	wait-Terminal
Vessel6	wait-Terminal
Vessel7	wait-Terminal
Vessel8	wait-Terminal

Planning Horizon = 15 days

Figure 6: Network for instance 8F-8V-10ST-6CT-5CDU-8C-4P-15D.