Citation for published version

Salhi, Said and Moin, Noor Hasnah (2007) Inventory Routing Problems: A Logical Overview. Working paper. Kent Business School, University of Kent, Canterbury

DOI

Link to record in KAR

https://kar.kent.ac.uk/5241/

Document Version

UNSPECIFIED

ISSN 1748-7595 (Online)

Kent Business School

Working Paper Series

Inventory Routing Problems: A Logistical Overview

Noor Hasnah Moin
University of Malaya

Said Salhi
Kent Business School

University of Kent

Working Paper No.138
May 2007

INVENTORY ROUTING PROBLEMS: A LOGISTICAL OVERVIEW

[1]Noor Hasnah Moin and [2]Said Salhi,
[1]Institute of Mathematical Sciences, University of Malaya, Malaysia
[2]Centre for Heuristic Optimisation, Kent Business School,
The University of Kent, U.K.

Abstract: This paper presents an overview of Supply Chain Management while focussing on the area of Inventory Routing. We aim to provide the state of the art in this area while highlighting the usefulness of the models in practice as well as their limitations. We have classified the papers based on the planning horizon employed in the models namely single period, multi period and infinite horizon models which are then complemented by those with stochastic demand patterns. Future research avenues which we believe to be of interest to the OR/MS community are also presented.

Keywords: Logistics, supply chain, transportation, inventory.

Introduction

As more and more companies become aware of their supply chain performance, the coordination and the integration of the various components in the *Supply Chain Management* (SCM) have become critical in gaining competitive advantage. This is usually achieved by identifying and eliminating redundancies (in terms of low capacity utilization) and inefficiencies (in terms of high distribution cost) within the combined supply chain/distribution network. A representative example of this is the merger of Nabisco and Kraft Food Inc., a tasty combination blending Kraft's cheese, dressings, and beverages with Nabisco cookies and crackers. Merging the two giant food chains has been a huge undertaking for the supply chain network in order to ensure a high level efficiency of consolidated distribution (Harp, 2003^a). Many large companies in the United States such as Wal-Mart and Unilever North America, adopt several aspects of SCM and a survey conducted by Harp (2003^b)found that companies with mature supply chain plannings are 38% more profitable than average companies. These companies also reported significant savings in inventory handling cost, increased efficiency and higher delivery performance to requested or committed due dates. The availability of data and information systems that derives from the advances in technology and communication systems has contributed significantly in the creation of successful coordination within the supply chain.

For example, Duffy (2004) illustrates how Gillett after spotting its problem in 2002 re-engineered its supply chain and within 2 years its customer service improved by 10%, their inventory reduced by 25% while costs went also down by 3%. It was also reported by Timme and Williams-Timme (2003) that inventory represents a large proportion of the net operating assets accounting for approximately 37% in manufacturing, 62% in distribution and 56% in retail. This demonstrates the need to integrate inventory and distribution appropriately. Waller `et al.` (1999) indicate that since vendor-managed inventory was popularised by Wal-Mart and Proctor &Gamble in 1980, other companies such as Campbell soup and Johnson and Johnson in the USA, the pasta manufacturer Berille in Europe and others benefited enormously.

The activities within SCM, as shown in Figure 1, are practically interrelated though these are usually treated separately.

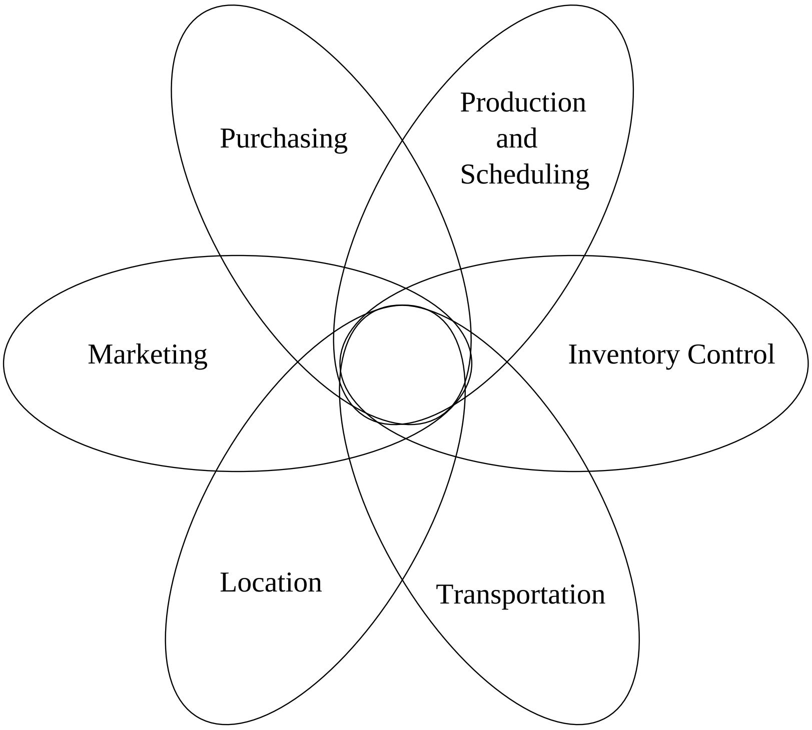

Figure 1: Related activities in Supply Chain Management

These activities have their own objectives which are often conflicting. For instance, the marketing objective of high customer service and maximum sales dollars conflict with manufacturing and distribution goals. On the other hand, a good marketing strategy will increase demand, which will also have indirect consequences on purchasing, production, inventory and distribution. Also, inaccurate customer forecasts for products can affect a variety of process areas, including purchasing, production, inventory and transportation. Clearly, there is a need for these activities to be aligned, coordinated, and synchronised since changes in one part of the SCM are likely to affect the performance of other processes.

This paper focuses on one aspect of the supply chain management, that is the distribution logistics that includes transportation (distribution) management and inventory control. Initial research concentrated on treating these two logistical aspects separately. However, the inventory allocation and vehicle routing decision are interrelated in the following way. In order to determine which customers must be served and the amount to supply each selected customer (the inventory allocation decision), the routing cost information is needed so that the marginal profit (revenue minus delivery cost) for each customer can be accurately computed. On the other hand, the delivery cost for each customer depends on the vehicle routes, which in turn requires information about customer selection and the amount of inventory allocated for each customer. This interrelationship between the inventory allocation and vehicle routing has recently motivated some authors to model these two activities simultaneously. This practical and challenging logistical problem is known as the integrated Inventory Routing Problem (IRP). This combined problem can be seen as an extension of the Vehicle Routing Problem (VRP) where the orders are specified by the customers and the aim of the supplier is to satisfy these customers while minimising its total distribution cost. Whereas in the IRP the orders are determined by the supplier, obviously based on some input from the customers whose aim is to minimise the sum of the inventory cost and the distribution cost. In other words, the IRP is a medium term problem whereas the VRP is a short term one. In the VRP the question is to find (i) the delivery routes, whereas in the IRP the question does not include (i) only but also the quantity to deliver to each customer as well as the time for delivery.

The applications of IRP arise in a variety of industries. For instance Campbell and Savelsbergh (2004) were inspired by the international industrial gaz company, Praxair, where one of their activities is to separate air into gazes such as oxygen, hydrogen, nitrogen and argon which are then transported to their customers. Other applications include petrochemical industry, suppliers of supermarkets and department stores, clothing industry, home products and the automotive industry (see respective references in Campbell and Savelsbergh (2004)). Another application of IRP which is not as commonly referred to as others is in the marine where ships are used instead of trucks, several products are often being shipped in separate compartments, the inventory activities are considered at both the sources and the destinations among other factors. Marine inventory routing differs slightly with most IRP as the transit times are usually much longer (days instead of hours) and the sources and the destinations often involved different countries which incur import taxes that contribute to the inventory holding costs (Ronen, 2002).

Solution techniques for the IRP can be classified into two categories namely the theoretical approach where the derivation of the lower bounds to the problem is sought or a more practical approach where heuristics are employed to obtain the near optimal solutions. Most papers that deal with the theoretical approach employ some strategies that allow the IRP to be decomposed into two underlying problems, namely the inventory and the travelling salesman problem. The inventory problem is solved to determine the replenishment quantities for each customer as well as the replenishment times. Most papers in this category adopt a 2-stage solution approach. They either (i) find the routes first and then solve the IRP formulation which is a simple linear programming based inventory control problem, or (ii) solve the inventory control problem first (sometimes with approximated transportation cost), aggregate (cluster) the customers with the same replenishment time instants and then construct the routes for each cluster. As the modification of routes entails resolving a new inventory allocation problem and vice versa, most algorithms iterate between obtaining a new set of routes and resolving the inventory problem until a suitable stopping criterion is satisfied.

Baita *et al* (1998) presented a review paper for the Dynamic Routing and Inventory Problems. The dynamicity is defined as a situation where repeated decisions have to be taken at different times within a certain horizon, and earlier decisions influence later decisions. This is a common feature in most of the IRP presented in the literature. The problems are usually classified according to their practical characteristics which are defined in terms of decision variables, constraints, objectives and cost factors, and also on the proposed approach to the solution. However, in this paper, the classification very much follows Bramel and Simchi-Levi (1997) whereby the problems are classified as single period models, multi-period models, infinite time horizon and finally those that consider the stochastic demand pattern. The review given in this paper, is by no means exhaustive, but aims to complement that of Baita *et al* (1998). Other relatively older literature reviews can be found in Campbell *et al* (1998), and Dror *et al* (1985).

This paper is organised as follows: The single-period models are presented in section followed by multi-period models in Section . The infinite horizon models and the stochastic IRP models are discussed in sections and respectively. Finally, our conclusion and some potential research avenues are given in Section .

Single-Period Models

Federgruen and Zipkin (1984) were amongst the first to integrate the inventory allocation and routing by treating a single-period, single-item problem with random demands at the retailers. The problem undertaken has a plant with a limited amount of available inventory and for a given day, the problem is to allocate this inventory among the customers. The objective is, therefore, to obtain vehicle routes and replenishment quantities for the retailers so as to minimise the expected one period cost consisting of transportation, inventory holding and shortage costs (the end of the period inventory levels and shortages). The ordering cost is not considered in this problem since the replenishment quantities for each customer is determined at the plant. Using some of the ideas from vehicle routing, they develop a non-linear mixed integer programming model which is solved using a generalised Benders' decomposition approach that has the property that for any assignment of customers to routes, the problem decomposes into a non-linear inventory allocation problem which determines the inventory and shortage costs, and a TSP for each vehicle, which yields the transportation costs. Because of inventory and shortage costs, and the limited amount of inventory available, not every customer will be selected to be visited every day. This is handled in the model by the use of a dummy route that includes all customers not receiving a delivery. The idea is to construct an initial feasible solution and iteratively improve the solution by exchanging customers between routes. Each exchange defines a new customer to the route assignment, which in turn defines a new inventory allocation problem and new TSPs.

Federgruen *et al* (1986) extend the work to the case in which the product considered is perishable. The product in the system is classified into two classes, *old* units which are the ones that will perish in the present period and *fresh* units which are those that are at least one period away from their due dates. The authors adopt the solution approach of Federgruen and Zipkin (1984) with a variation that the inventory allocation subproblems accounts for two product classes. Both papers present a comparison of the integrated to the sequential approach. Federgruen and Zipkin (1984) show that about $6-7\%$ savings can be achieved through a combined approach. The travel costs are found to be substantially less using a combined approach and it was noted that for most instances, the delivery requirements can be met with one vehicle less than those of the sequential approach of Federgruen *et al* (1986).

Chien *et al* (1989) propose an integrated approach to a single period IRP and formulate it as a mixed integer programming problem. The problem consists of a central depot with fixed supply capacities and many customers with deterministic demand. The entire demand needs not be satisfied but there is a penalty cost imposed per unit of unsatisfied demand. The objective is to maximize profit that consists of total revenue less the penalty cost and routing costs. A Langrangean based procedure to generate good upper bounds is also proposed. The procedure provided good quality solutions when tested on a set of randomly generated problems. Although the model was based on a single period approach, it passes some information from one period to the next through the inter-period inventory flow, and hence can be seen to simulate a multiple period planning model which is about to be reviewed next.

Although the single period models may not reflect the long term planning, the models are still of some relevance as they are sometimes used as the basis in the study of multi-period models. Furthermore, in most IRPs, the demand at each retailer is sometimes difficult to predict with certainty and can only be best represented with random variables with known probability. In such situations, long term planning may be prone to adjustments and hence not completely viable.

Multi-period Models

Many short term approaches have the tendency to defer as many deliveries as possible to the next planning period. This flexibility adds complexity to the problem, and hence the proper projection of a long-term objective into a short-term planning period is essential. The objective function needs to capture the costs and benefits of delivering to the customers earlier than necessary. This is often achieved by decomposing a multiperiod problem into a series of single period problems using the single period objective function as a surrogate for long-term cost.

The first effort to study the effect of the short term on the long term planning is due to Dror *et al* (1985) and Dror and Ball (1987). They consider single period models as subproblems and propose a mixed integer programming model where effects of present decisions on later periods are accounted for using penalty and incentive factors. Using the probability that a customer will run out on a specific day in the planning period, the average cost to deliver to the customer, and the anticipated cost of stock out, the optimal replenishment day t^* minimising the expected total cost can be determined for each customer. If t^* falls within the short-term planning period, the customer will definitely be visited, and an expected increase in future cost if the delivery is made on day t instead of on t^*, is computed for each of the days in the planning period. If t^* falls outside the short-term planning period, then a future benefit can be computed for making a delivery on day t of the short term planning period. These two computed values reflect the long term effects of short-term decisions. An integer program, that minimises the sum of these costs plus the transportation costs, is solved to assign customers to vehicles and to particular days. The routing problem is then tackled at the second stage. In their model, the inventory holding costs are not included in the objective function. Here, only customers who will reach their safety stock level during a particular time interval are serviced and the model only considers a fixed number of identical trucks. We note that the delivery amounts are not decision variables as these are dictated by the day of the week on which the delivery is to be made. Dror and Levy (1986) use a similar analysis to yield a weekly schedule, but then they apply a heuristic using node and arc exchanges to reduce costs in the planning period. In both papers the demand that occurs at the retailers/customers is considered to be deterministic.

The same ideas have been extended and improved in Trudeau and Dror (1992) and Dror and Trudeau (1996). In Trudeau and Dror (1992) stochastic demands are considered whilst both deterministic and stochastic demands are employed in Dror and Trudeau (1996). Both papers solved a slightly different model with a specific application to supply and distribution of oil and gas. In these models, a single item has to be delivered from one depot to many retailers/customers, whose demand is different in each period. Trudeau and Dror (1992) develop heuristics, based on linear mixed integer programming sub models to solve their problems. The objective is to develop routing patterns, so as to meet the demands and minimise the long run average transportation costs. In particular, Dror and Trudeau (1996) focus on the maximisation of operational efficiency (average number of units delivered in one hour of operation) and the minimisation of the average number of stock out in one period.

Jaillet *et al* (1997) and Bard *et al* (1998) extend the idea of Dror *et al* (1985) but using a slightly different approach. First, an expected total cost of restocking a customer in a given period is derived, then an optimal frequency to restock a customer is determined by identifying the minimum point on the corresponding cost curve. The assignment of customers to days of the week over the planning horizon is based on the optimal value of the frequency and the incremental cost which is incurred by not servicing a customer on his/her optimal day. If the best day to visit falls within the planning horizon, the customer is selected and the routing of these customers then follows. We note that they used a bi-criteria approach as a performance measure whereby the assignment of customers to days of the week is based solely on the incremental cost whilst the routing of customers is determined using the minimisation of the total distance travelled. The problem they attempted consists of a central depot and customers that need replenishing to prevent stock out due to the high costs of scheduling a special delivery when stocks out occur. They considered the presence of satellite facilities where vehicles can be refilled (other than the depot) and customer deliveries continued until a closing time is reached. They take a rolling horizon approach to the problem by determining a schedule for two weeks. The idea is to model two weeks schedule while implementing only the first week. The problem is solved again in the following week for the next two weeks horizon.

Campbell and Savelsbergh (2004) develop a model, using a vendor managed resupply policies, which considers routing customers together on a day where none of them are at the point of running out of stock, but if combined they can make a full or near-full truckload delivery route. Vendor managed resupply, through the rapid development in communication technology, is the emerging trend in management of logistic systems and it has been claimed to provide a win-win situation for both suppliers and customers. In the vendor managed resupply, the inventory at each customer is monitored by the suppliers and the replenishment of products to different customers is coordinated at the suppliers level. Consequently this eliminates the ordering cost from the objective function. The authors propose a two-phase solution approach implemented in a rolling horizon framework. In phase I an approximation of the problem is constructed based on a k days planning using integer programming to find the customers to serve each day and how much to serve them. The obtained solution is then used as information for phase II where the daily scheduling plan takes place. The distribution plan is constructed for a month to reflect the long-term nature of the planning problem, but implemented only for the first few days. This is repeated as often as necessary using the latest information available. In phase I, a large set of possible clusters are generated and the cost of serving each cluster is estimated. In the second phase the departure times and customer sequence for the different vehicles are carried out using an insertion heuristic for solving the vehicle routing with time windows. In order to balance between the long-term and short-term objectives, the solutions from phase I is treated as suggestion but followed as closely as possible. A small deviation is allowed whereby the deliveries are allowed to be split into small portions, if only it gives a better solution. We note that the inventory cost is not explicitly defined in the objective function since the problem undertaken is for a large industrial gases company. It is argued that in a rolling horizon framework as this, the emphasis should be on the quality and detail of the decisions concerning the first few days of the plan, since it is only on these days where the plan is implemented.

Speranza and Ukovich (1994) study a slightly different problem where several products have to be shipped on a common link (from one origin to one destination) such that the sum of inventory and transportation costs is minimized. The main focus of the paper is to investigate a shipping strategy, that is, a rule stating when shipments must take place and how products must be loaded on vehicles (with finite capacities) in a multi-period scenario. They propose a discrete frequency approach where shipments can only take place at some given discrete times. In addition, the different number of products require a slight modification to the objective function to take into account the quantity of each product requested by a given retailer, the different demand patterns for each product, and in some instances the different holding costs for each product. Most models, as in Speranza and Ukovich (1994), assume the demand rate for each product to be constant. They experimented with different shipment strategies and it was found that allowing products to be split among several shipping frequencies makes trucks travelling at high frequencies to be filled up completely.

Speranza and Ukoivich (1998) compare the results of Speranza and Ukovich (1994) for the discrete frequency approach with the EOQ-based solution for the shipment of several items from one origin to one destination using trucks of a given capacity. It is worth noting that most of the earlier work reported in the literature (see for example Federgruen and Zipkin (1984), Federgruen *et al* (1986), Anily and Federgruen (1990), Anily (1994)) are based on EOQ formulation. One major drawback of these models is that they produced frequencies that can be non-integer (and some even non-rational) and hence it can be impractical to implement in practice (Baita *et al*, 1998). One of the immediate solutions, suggested by Hall (1985), is to round off the EOQ value to the nearest feasible frequency, but Speranza and Ukoivich (1998) show that this strategy leads to larger costs which can be well over 20% above the *ideal* minimum. Despite this drawback such results can be useful as they act as lower bounds for comparison purposes.

Bertazzi *et al* (1997) extend the work of Speranza and Ukovich (1994) to tackle the more general case of several destinations. They introduce the concept of split deliveries where the quantity of a product required at a destination may be split between different shipments, possibly with different frequencies. This allows for a better utilization of the vehicles, thus reducing the transportation cost, but the flexibility obviously adds to the complexity of the problem. For simplicity, most multi-product models assume that each retailer requires only one type of product, and for split deliveries, a new constraint is also introduced to ensure that the total amount of products requested at each retailer is fulfilled. The authors propose a set of decomposition heuristics starting from a link by link (i.e. direct shipping) solution of the problem, and then performing a local search looking for improvement through consolidation. In the second phase, customers visited at the same frequency are considered for aggregation on the same route. The possibility of further reducing costs by choosing different frequency is also investigated. In the final phase, customers with the same frequency are consolidated again, and routes are designed.

Bertazzi *et al* (2002) study a multi-period, multi-product with deterministic demand for order-up-to level inventory policy. The system consists of a common supplier in which a set of products is shipped to several retailers. Each product is made available at the supplier and absorbed by the retailers in a deterministic (the quantity is known at each discrete time instant) and time-varying (the quantity varies from one discrete time instant to the other) way. The order-up-to level policy ensures that, every time a retailer is visited the quantity delivered is such that the maximum level of inventory is reached. Since the inventory costs is imposed at both the supplier and the retailer, the objective function has to be modified accordingly. Two additional constraints have to be included in the formulation to ensure that no shortages occur at the supplier and the inventory level is not lower than the prescribed level. The authors propose a two-step heuristic algorithm to solve the problem. In the first stage, a constructive heuristic inserts customers with the same time instant into the routes based on the minimum cost policy (costs represent the sum of the transportation cost and the inventory costs both at the supplier and the retailer) in which the retailer that incurs the least cost is inserted in the solution.

The second part of the heuristic exchanges customers on two different routes if this leads to an improvement in the total cost. The solution procedure is implemented only for a single product and a single vehicle case. We note that the multi-product case can be handled by replacing each retailer by a set of retailers, one for each product absorbed by the retailer. This leads to setting the transportation costs to zero for retailers in the same set and the transportation costs will only be incurred for nodes from different sets.

Lee *et al* (2003) study the IRP in an automotive part supply chain that comprises several suppliers and an assembly plant. The problem addressed is based on a finite horizon, multi-period, multi-supplier, and a single assembly plant part-supply network where a fleet of capacitated identical vehicles transport parts from the suppliers to meet the demand specified by the assembly plant for each period. This problem represents an in-bound logistic problem of type $many-to-one$ network and is equivalent to the $one-to-many$ under certain conditions. The authors propose a mixed integer programming model to minimise the overall transportation cost and the inventory carrying cost. This mixed integer programming model is decomposed into two subproblems, namely the VRP and the inventory control. A heuristic based on simulated annealing is proposed to generate and evaluate alternative sets of vehicle routes while a linear program determine the optimum inventory levels for a given set of routes. Then, a route perturbation routine is implemented to modify a set of vehicle routes based in some information obtained from the optimal solution to the linear program. The modification of routes entails resolving the linear program to get new inventory levels. This scheme is carried out iteratively until a stopping criteria is reached, namely the specified maximum number of iterations. They also observe an important property that the optimal solution is dominated by the transportation cost regardless of the magnitude of the unit inventory carrying cost. This claim is then proven analytically for a simpler version of the above problem based on an infinite planning horizon and s stationary demand with a single supplier providing either a single part type or multiple part types.

A slight variation of the IRP that involves the use of vehicles for which the purchase or lease agreements may need to be signed months or even years before the start of actual delivery operations has been studied by Webb and Larson (1995). The *Strategic* IRP (SIRP) is motivated by the long lead times between the signing of purchase or lease agreements and the availability of vehicles for delivery operations. The SIRP focuses on estimating the minimum size (cost) vehicle fleet required to supply inventory from a central depot to spatially dispersed customers when only the probability distribution for the per unit time demand at each customer is known. Since the determination is based on information currently known about customers' usage rates, the minimum number of vehicles found must be able to handle a reasonable amount of variation in these usage rates. Most IRP models deal with routing an existing fleet of vehicles to supply customers whose actual demands for replenishment are known or can be estimated. It is easily seen that the SIRP should encompass all possible realisations of IRP.

The work of Webb and Larson (1995) is motivated by Larson's earlier work (1988) undertaken for the New York City Department of Environmental Protection. It was observed that one of the solutions required more visits to one of the customers than necessary. This was attributed to the fact that a single route was used to service a cluster and the replenishment intervals were determined by the customer requiring the most frequent visits. To overcome this, Webb and Larson (1995) propose the use of period and phase of individual customer's replenishment as additional decision variables. Here, the routing solutions are based on customer-specific period and phase of replenishment and it is developed for a simple model of the routing problem. The fleet size estimated is determined by separating the customers into disjoint clusters and creating a route sequence for each cluster. A route sequence is a permanent set of repeating routes. Customers are allowed to be on more than one route in the sequence.

The multi-period models are useful in the sense that they offer a more realistic trade off between the strategic and the operational nature of the IRP models. These approaches generally produce a high quality solution while requiring significantly a larger computing effort. In addition, they allow the effect of the long term cost on the current schedules to be studied. Due to the increase complexity of the problem, most multi-product and multi-period models consider deterministic demand at the retailers and heuristic methods to find solutions for the multi-period models.

Infinite Horizon

The objective function in the infinite horizon models often focuses on minimizing the long-run or the mean average of all the costs involved. In most infinite horizon models, the demand rate is assumed to be constant and deterministic, and the periodic review policies for inventory is often adopted in recent models.

Burns *et al* (1985) are among the first to consider an inventory replenishment problem with vehicle routing costs for an infinite horizon one warehouse multiple retailer system. They adopt an aggregate approach whereby a much detailed information may be neglected without the analytical models losing their capability for devising an implementable solution. Explicit formulas are often obtained in terms of a few measurable parameters. These type of models are often easier to solve and a qualitative insights can easily be achieved as only the most important factors are taken into account. The authors develop an EOQ type formulation and an analytic method is proposed to minimize total replenishment costs. The model considers retailers with constant demand and all holding cost rates are assumed to be identical. The transportation cost is proportional to the Euclidean distance travelled. It treats the transportation cost as part of the constant that measures reorder costs, and generally neglects the dependence of transportation costs on network characteristics and vehicle routing.

Blumenfeld *et al* (1985) use a similar approach to Burns *et al* (1985) in solving a large scale application problem undertaken for the General Motors Corporation. In this problem a large number of different products have to be shipped from 20000 suppliers to 160 plants. The authors analyse the trade off between inventory, transportation and set up costs both in the case of direct shipping and of shipping via intermediate terminals.

One of the commonly used approaches in the infinite horizon model is based on the fixed-partition policies. This strategy solves the problem by performing an a priori partition of the customers into regions, where the demand of each region is roughly equal to a truckload, and then solves the TSP within each region. Note that finding such regions optimally is NP hard. One of the shortcomings of this approach is that the deliveries to different regions are, in most cases, not coordinated. However, assigning to different regions the outlets corresponding to a single retailer and coordinating deliveries may induce further reduction of costs. The lack of coordination of the deliveries does not minimise the rent costs related to space and facilities needed to hold the maximum stock in a retailer warehouse.

It is worth noting that it is naturally easier to coordinate load splitting within the fixed partition approach which often reduces the number of tours and the distance travelled, in addition to maximising the vehicles capacity. However, in most cases, it is found that usually only a few retailers are split among different routes.

Anily and Federgruen (1990) adopted the above ideas to determine an optimal replenishment for the case of a single product, infinite horizon problems with constant but retailer-dependent demand. Here, the demand at each retailer is assumed to be the same, and each retailer consists of a number of outlets located at the same points, thus allowing the original retailer to be assigned to different regions which are then obtained heuristically using the circular partition scheme. The heuristic is shown to be asymptotically optimal when the retailers' locations have independent and identically distributed distance from the depot and the number of outlets grows to infinity. A two-stage heuristic has been proposed. The first stage involves the derivation of a lower bound which is obtained by solving a problem of partitioning the set of demand points into groups. In the second stage, all the groups with the same cardinality obtained by solving the partitioning problem are combined into larger families of demand points. In each of these families, efficient vehicle routes are constructed using regional partitioning heuristics. The case in which the number of regions is a priori fixed is considered in Anily and Federgruen (1991). We note that in both papers the depot (warehouse) acts as break-bulk centre (an outside supplier) where the inventory is only kept at the retailers. Note that in Anily and

Federgruen (1990) a constraint on the number of demand points that can be replenished together is imposed to guarantee an asymptotic convergence of the heuristic developed.

Anily and Federgruen (1993) generalises the above problem considering a central warehouse from which the products are distributed. Such a warehouse pays holding costs and has a limited stock, hence it must be replenished from time to time facing fixed plus proportional costs. Here, the VRP is equivalent to a multi item/multistage inventory control problem with joint set up costs. They considered the same problem as Anily and Fredugruen (1990) but proposed a power-of-two policies, where the replenishment intervals are power-of-two multiples of the base planning period.

Anily (1994) takes a similar approach in solving a slightly different generalisation of the problem where holding costs are retailer dependent. For this problem, a new heuristic, *regional partitioning scheme*, which is asymptotically optimal is introduced. Due to the different holding costs rate, the outlets are first clustered on the base of the ratio between their distance and holding cost and then according to their geographic location. This ensures that the retailers in a single region are in close proximity one to the other and that they have similar holding costs rates. Consequently, it is argued that the effect of the joint replenishment on the holding cost is negligible relative to the savings on routing costs.

Working along similar ideas, Gallego and Simchi-Levi (1990) evaluate the long run effectiveness of direct shipping (separate loads to each customer). They concluded that direct shipping is at least 94% effective over all inventory routing strategies whenever minimal economic lot size is at least 71% of truck capacity. This implies that direct shipping is not an appropriate policy when many customers require significantly less than a truckload, making more complicated routing polices the appropriate choice. Bramel and Simchi-Levi (1997) extended the work to solve a variant of IRP in which customers can hold an unlimited amount of inventory. The problem was first transformed into a capacitated concentrator location problem (CCLP) and having solved that, the problem is transformed back into a solution to IRP. The solution to the CCLP will partition the customers into disjoint sets which translates into a fixed partition in the IRP. These partitions are then used in a similar way to the regions in Anily and Federgruen (1990).

Viswanathan and Mathur (1997) extend the work of Anily and Federgruen (1993) for the multi-period, multi-product problem. They present a new heuristic that generates a stationary nested joint replenishment policy for the problem with deterministic demands. A policy is defined as stationary if the replenishment of each item is made at equally spaced points in time (i.e. with constant replenishment intervals) whilst a nested policy means that if the replenishment interval of a given item is larger than that of another item, the former is a multiple of the latter. They have adopted a power-of-two policies and the objective is to find the optimal replenishment quantities for the products, as well as the vehicle routes to deliver these quantities, so as to minimise the average inventory and transportation costs over a finite time horizon. A fast greedy heuristic, initially developed for the uncapacitated problem, is extended to group the customers into clusters. The replenishment interval for each item is calculated by a modified EOQ formula.

Most of the infinite horizon models are based on the fixed-partition policies, and hence are computationally quite efficient and can solve large instances. Since they can only guarantee asymptotic results and it may sometimes provide trivial or even poor quality solutions to problems of modest size. In addition, coordinated deliveries are very difficult to be scheduled, thus the results obtained are either valid only in the case of independent deliveries, or can be seen as generating an upper bound on the true cost. Although the IRP is an infinite horizon problem, most work in this area is based on the theoretical analysis (with the exception of Blumenfeld *et al* (1985) and Burns *et al* (1985)) and tested on randomly generated problems.

Stochastic IRP

Recently, many researchers have started to investigate the stochastic version of the problem whereby it is assumed that a probability distribution is known for customer usage. The stochastic IRP differs from the deterministic IRP in that the future usage amounts are uncertain. However, obtaining the probability distribution of customer usage in practice is extremely complex. Kleywegt *et al* (1999) formulate the IRP as a Markov decision process and propose approximation methods to find good solutions within reasonable computational time. The computational results are presented for the inventory routing problem with direct deliveries. Most of the stochastic IRP models deal with the transportation and supply of gas and oil (see for example Federgruen and Zipkin (1984), Trudeau and Dror (1992), Jaillet *et al* (1997), Bard *et al* (1998)).

A slightly different application for a multi-item joint replenishment problem, in a stochastic setting, with simultaneous decisions made on inventory and transportation policies is presented in Qu *et al* (1999). The model deals with an inbound material-collection problem, as in Lee *et al* (2003), whereby a central warehouse replenishes its stock by dispatching vehicles to collect the goods from several geographically dispersed suppliers. The demand for each item is assumed to be stochastic, and are independent and identically distributed in the form of a Brownian motion process. The inventory cost consists of a joint fixed replenishment cost to be shared among all the items included in a given replenishment, as well as an item dependent minor cost for including any specific item in that replenishment. Inventory holding cost at the central warehouse is incurred at a constant rate per unit time and total backlogging is assumed with a cost proportional to the total number of unit shorts.

The transportation cost comprises of a fixed cost for each stopover, plus a variable cost proportional to the travel distance. The objective is thus to determine procedures for inventory management and vehicle routing so that the warehouse may satisfy demand at a minimum long-run average cost per unit time.

The mathematical model proposed can be decomposed into inventory (as a master problem) and transportation that acts as a subproblem. The master problem is solved item by item and determines the amount, for each item, to be replenished by which supplier and the period it will be replenished. Then for each period, a TSP is solved. A lower bound to the problem is also constructed to assess the effectiveness of the heuristic. It was observed that a slightly better lower bound could be achieved for a special case when each supplier produces a unique item.

Recently, Ribeiro and Lourenço (2003) incorporate both the stochastic and deterministic demands in their models. Two types of customers are considered. The first is the customer managed inventory (CMI) based on customers with deterministic demands. The second type of customers falls into the vendor managed inventory (VMI) with stochastic demands. The aim of the study is to determine the best routes for all customers and the quantities and the days to be delivered for the VMI customers (since the quantities and days to deliver are known for the CMI customers) over a week (5 days) planning horizon. It is worth noting that the inventory holding cost and the stock out cost are only incurred on the VMI customers. The model considers a single product in a multi-period scenario and the demand for VMI customers is modelled as a continuous random variable. The authors proposed an iterated local heuristic to solve the problem. The model is decomposed into two subproblems, namely inventory and VRP. First, an initial solution is obtained by solving the inventory problem (without considering the transportation cost), then a good feasible solution is obtained by solving the VRP. Subsequently the new quantities and the days to deliver for VMI customers are determined, taking into account the delivery cost calculated from the previous step. This process is carried out iteratively until a satisfactory solution is obtained. The computation study was carried out on randomly generated data and it was found that the reduction on the cost depends on the problem size, proportion of VMI customers and the unit costs chosen for the problems.

Ronen (2002) proposes a combination of heuristic and a mixed integer linear programming model for marine inventory routing. Here, a fleet of ships or barges transport multiple products (in separate compartments) from a set of origins to a set of destinations. The products are produced at the origins according to a production plan and consumed at the destinations according to a demand forecast. The author handles the stochasticity of the demand through determining appropriate levels of safety stocks. The maximal vessel size is both determined by the loading and unloading facilities at the origins and the destinations, respectively. The results presented show that the heuristic solutions tend to ship products as late as possible in order to save money and retain shipping flexibility whereas integer solutions in the linear programming model save shipments later on. As pointed out by Ronen, in real life IRPs are stochastic and any distribution plan covering more than a few days will seldom be executed completely as planned. In other words, in the stochastic scenarios any planning system needs to be flexible enough to cater for the latest changes in the data.

Conclusion and Suggestions

This paper presents a review of the IRP as part of the overall supply chain management. A special emphasis is placed on minimising the total cost which includes the inventory and the routing cost. The papers we reviewed are classified according to the nature of the planning horizon. General comments on the applicability and the usefulness of the models are also provided whenever possible.

It is worth noting that, in general, most activities within IRP can be distinguished depending on the time horizon which can be strategic, tactical or operational (Vidal and Goetschalckx, 1997). The strategic level requires approximate and aggregated data whereas the operational level uses accurate input. On the other hand tactical decisions fall in between those two extremes with respect to time horizon, and the amount and accuracy of data required. The issue of relying on short term decisions while addressing a strategic problem will, in our view, remain a delicate issue and an open question for academics as well as practitioners. Some work on robustness analysis may be one of the ways forward to help addressing this problem. Though the multi-period models are found to be more efficient than their counterpart the single-period ones, some of the intriguing questions which one might ask include how long term information are incorporated into the short term, how the short term is measured and which are the customers assigned in this short term. The infinite horizon models are found to be computationally efficient and capable of solving large instances (up to thousands of customers). Unfortunately these models can only guarantee asymptotic results which can be misinterpreted in practice and hence further studies are needed to determine the applicability of these approaches in real world logistical problems. The stochastic models can be useful in real life as they incorporate stochastic demand but the determination of the probability distribution of customer usage can be, in practice, extremely complex and hence an insight of the problem and extra care is needed when it comes to implementing such models.

In addition to the above comments, we consider the following research avenues to be worthwhile exploring.

Routing-based issues

In most IRP given in the literature, the aim of the problem is to determine for each delivery time instant the set of retailers to visit, the quantity of each product to ship to each retailer, and the route of each vehicle that minimises the total overall costs. The problems considered often assume either the number of vehicles available is unlimited (see for instance Lee *et al* (2003)), or the number of vehicles is determined simultaneously for each time instant. The strategic vehicle fleet size for the whole planning horizon is often not addressed in these problems (with the exception of the paper by Webb and Larson (1995)). The variation in the number of vehicles required from one time instant to the other may result in loss revenues due to the idle vehicles, that may be required in only one particular time instant, or the cost due to additional hired vehicles. This can be partially overcome by considering the fleet size as one of the decision variables. Also in practice, most customers often impose deadlines on goods to be delivered (or picked-up), thus the existing IRP models ought to be extended to reflect such practical restrictions. In nowadays environmentally-based decisions play an important part in our society and hence integrating reverse logistic as part of the overall distribution network is worth the effort. This does not only reduce cost but also the amount of gas emissions and the risk of accidents on our roads.

Inventory-based issues

A majority of the work in IRP have tilted heavily towards the development of deterministic inventory models (with the exception of those used in the application in oil and gas related industries). In practice, the variables that are not known *a priori* may include customers demand patterns, vehicle travel time and the possible variation on the number of customers over time. In light of this, it may be useful to integrate the models of stochastic inventory control in the formulation. Such a result will also provide additional and invaluable information when incorporated in the models of real-time to make them even more efficient.

Effect of Location on IRP

In this review, we concentrate on the inventory-routing while considering the location of the facilities fixed. Research work on location-routing can be found in Salhi and Rand (1989), Min *et al* (1998), and Nagy and Salhi (2007). A recent paper that integrates location and inventory is given by Drezner *et al* (2003). One avenue that is worth further investigation is on the strategic IRP that incorporates the decision on the location of the facilities. The work of Liu and Lee (2003) has been made along this line although the emphasis is more on the location-routing while taking into account the relevant inventory costs into the objective function.

On-Line Logistics

Most of the IRP models developed so far are static and often fail to capture the dynamic nature of the parameters involved. Some of the parameters such as transportation cost and inventory cost change over time. For instance, transportation cost may vary constantly with changes in negotiated contract rates and the extent of empty miles. Similarly, in a given time planning horizon, a company may have to deal with decisions of relocating existing warehouses due to the increasing inventory and related costs (because of high expansion cost, increase in warehousing employee wages and interest rates). As such, IRP parameters are highly time-sensitive and the incorporation of their dynamic nature into IRP models should greatly enhance the realism associated with real-time or on-line logistic operations. A short but interesting editorial comment on issues related to real-time fleet management is given by Gendreau and Potvin (2004).

Benchmarking

One of the commonly used ways in assessing a given method against others is to be able to conduct empirical testing based on benchmark data sets that are easily accessible to the research community. Unfortunately, the vast majority of authors working on the IRP conducted their analysis of their models on randomly generated problems, albeit some data sets are accessible through the internet (see for example Lee *et al* (2003)). This lack of benchmarking stops researchers from making a useful comparison on the efficiency of each solution method proposed, although some took extra effort to encode other methods to be solved on their own instances (Viswanathan and Mathur, 1997). In view of this, a coordinated set of benchmark problems (for several variants of IRP) ought to be constructed in the near future and more interestingly, this would be useful if a proper compilation of the existing problems is maintained at a common site. The above performance measure can obviously be complemented by studying the time complexity of these heuristics under certain conditions. For instance Chan *et al* (1998), Chan and Simchi-Levi (1998) and Federgruen and Simchi-Levi (1995) provide interesting probabilistic results for some IRP algorithms. One way forward is to classify a given IRP into a well defined class within which useful properties and bounds can be obtained. This is a theoretical issue which, in our view as OR/MS open minded academics, needs not to be ignored though its practical use may, in the eyes of some practitioners, be considered less informative and limited in scope.

Acknowledgments- The first author would like to thank British Council for the grant under the Royal Society Malaysian Fellowship (part of the British Chevening Scholarship programme). The authors are also grateful to the referees for their constructive comments.

References

Anily S (1994). The general multi-retailer EOQ problem with vehicle routing costs. *Eur J Opl Res* `79`:451 – 473.

Anily S and Federgruen A (1990). One warehouse multiple retailer inventory systems with vehicle routing costs. *Mngt Sci* `36`:92 – 114.

Anily S and Federgruen A (1991). Structured partitioning problems. *Opns Res* `39`:130 – 149.

Anily S and Federgruen A (1993). Two-echelon distribution systems with vehicle routing and central inventories. In: M. Dror (ed). Special Issue on Stochastic and Dynamic models in Transportation. *Opns Res* `41`:37 – 47.

Baita F, Ukovich W, Pesenti R and Favaretto D (1998). Dynamic routing-and-inventory problem: A review. *Transportation Res -A* `32`:585 – 598.

Bard J, Huang L, Jaillet P and Dror M(1998). A decomposition approach to the inventory routing problem with satellite facilities. *Transportation Sci* `32`:189 – 203.

Bertazzi L, Speranza MG and Ukovich W (1997). Minimization of logistic costs with given frequencies. *Transportation Res - B* `31`:327 – 340.

Bertazzi L, Paletta G and Speranza MG (2002). Deterministic order-up-to level policies in an inventory routing problem. *Transportation Sci* `36`:119 – 132.

Blumenfeld DE, Burns LD, Diltz JD and Daganzo CF (1985). Analyzing trade-offs between transportation, inventory and production costs on freight networks. *Transportation Res - B* `19`:361 – 380.

Bramel J and Simchi-Levi D (1997). *The Logic of Logistics.* Springer: New York.

Burns LD, Hall RW, Blumenfeld DE and Daganzo CF (1985). Distribution strategies that minimize transportation inventory costs. *Opns Res* `33`:469 – 490.

Campbell AM and Savelsberg MWP (2004). A decomposition approach for the Inventory-routing problem. *Transportation Sci* `38`:488 – 502.

Campbell AM, Clarke LW, Kleywegt A and Savelsberg MWP (1998). Inventory routing. In T. Crainic and G. Laporte (eds). *Fleet Management and Logistics.* Kluwer Academic Publisher:Boston, M.A.

Chan LMA, Federgruen A and Simchi-Levi D (1998). Probabilistic analyses and practical algorithms for inventory-routing models. *Opns Res* `46`:96 – 106.

Chan LMA and Simchi-Levi D (1998). Probabilistic analyses and algorithms for three-level distribution systems. *Mngt Sci* `464`:1562 – 1576.

Chien TW, Balakrishnan A and Wong RT (1989). An integrated inventory allocation and vehicle routing problem. *Transportation Sci* 23:67 – 76.

Drezner Z, Scott C and Song J-S (2003). The central warehouse location problem revisited, *IMA J Mngt Math* 14:321 – 336.

Dror M and Ball M (1987). Inventory/routing: Reduction from an annual to a short period problem. *Nav Res Logist Q* 34:891 – 905.

Dror M and Levy L (1986). Vehicle routing improvement algorithms: Comparison of a greedy and a matching implementation for inventory routing. *Comput Opns Res* 13:33 – 45.

Dror M, Ball M and Golden B (1985). Computational comparisons of algorithms for the inventory routing problem. *Ann Opns Res* 4:3 – 23.

Duffy M (2004). How Gillett cleaned up its supply chain. *Supply Chain Mngt Rev* 8:20 – 27.

Federgruen A and Zipkin P (1984). A combined vehicle routing and inventory allocation problem.*Opns Res* 32:1019 – 1037.

Federgruen A and Simchi-Levi D (1995). Analytical analysis of vehicle routing and inventory routing problems. In M. Ball, T. Magnanti, C. Monma and G. Nemhauser (eds). *Handbooks in Operations Research and Management Science, Network Routing.* North Holland: Amsterdam, pp 297-373.

Federgruen A, Prastacos G and Zipkin P (1986). An allocation and distribution model for perishable products. *Opns Res* 34:75 – 82.

Gallego and Simchi-Levi D (1990). On the effectiveness of direct shipping strategy for the one-warehouse multi-retailer R-systems. *Mngt Sci* 36:240 –243.

Gendreau M and Potvin JY (2004). Issues in Real-Time Fleet Management. *Transportation Sci* 38:397 – 398.

Hall RW (1985). Determining vehicle dispatch frequency when shipping frequency differs from suppliers. *Transportation Res - B* 19:421 – 431.

Harp LH (2003^a). The nature of change. *Inbound Logistics* 23:76 – 132.

Harp LH (2003^b). Supply chain best practices: Hitting the Mark. *inboundlogistics.com*, December 2003.

Jaillet P, Huang L, Bard J and Dror M (1997). *A rolling horizon framework for the inventory routing problem.* Working Paper.

Kleywegt AJ, Nori VS and Savelsberg MW (1999). *The stochastic inventory routing problem with direct deliveries.* Technical Report TLI99-01, Georgia Institute of Technology, Atlanta, Ga, USA.

Larson RC (1988). Transporting sludge to the 106-mile site: An inventory/routing model for fleet sizing and logistic system design. *Transportation Sci* 22:186 – 198.

Lee C-G, Bozer YA and White III CC (2003). *A heuristic approach and properties of optimal solutions to the dynamic inventory routing problem.* Working Paper, University of Toronto, Toronto, Ontario, Canada.

Liu SC and Lee SB (2003). A two-phase heuristic method for the multi-depot location routing problem taking inventory control decisions into consideration. *Int J Adv Manuf Tech* 22:941 – 950.

Min H, Jayaraman V and Srivastava R (1998). Combined location-routing problems: A synthesis and future research directions. *Eur J Opl Res* 108:1 – 15.

Nagy G and Salhi S (2007). Location-Routing: Issues, models and methods. *Eur J Opl Res* (accepted).

Qu WW, Bookbinder JH and Iyogun P (1999). An integrated inventory-transportation system with modefied periodic policy for multiple products. *Eur J Opl Res* 115:254 – 269.

Ribeiro R and Lourenço HR (2003). *Inventory-routing model, for a multi-period problem with stochastic and deterministic demand.* Working Paper, Department of Economics and Business and GREL-IET, University Pompeu Fabra, Barcelona, Spain.

Ronen D (2002). Marine inventory routing: shipments planning. *J Opl Res Soc* 53:108 – 114.

Salhi S and Rand GK (1989). The effect of ignoring routes when locating depots. *Eur J Opl Res* 39:150 – 156.

Salhi S and Fraser M (1996). An integrated heuristic approach for the locations of vehicle fleet mix problem. *Stud Loc Anal* 8:3 – 22.

Speranza MG and Ukovich W (1994). Minimizing transportation and inventory costs for several products on a single link. *Opns Res* 42:879 – 894.

Speranza MG and Ukovich W (1998). Analysis and integeration of optimization models for logistic systems. *Int J Prod Econ* 35183 – 190.

Timme SG and Williams - Timme G (2003). The real cost of holding inventory. *Supply Chain Mngt Rev* 7:30 – 37.

Trudeau P and Dror M (1992). Stochastic Inventory routing: route design with stockouts and route failures. *Transportation Sci* 26:171 – 184.

Vidal CJ and Goetschalckx M (1997). Strategic production-distribution models: A critical review with emphasis on global supply chain models. *Eur J Opl Res* 98:1 – 18.

Viswanathan S and Mathur K (1997). Integrating routing and inventory decisions in one warehouse multiretailer, multipproduct distribution systems. *Mngt Sci* 43:294 – 312.

Waller M, Johnson ME and Davis T (1999). Vendor-Management Inventory in the retail supply chain. *J Business Logistics* 20:181 – 203.

Webb IR and Larson RC (1995). Period and phase customer replenishment: A new approach to strategic inventory routing problem. *Eur J Opl Res* 85:132 – 148.

University of Kent

tp://www.kent.ac.uk/kbs/research-information/index.htm