Volume 28 (2), pp. 117–135
http://www.orssa.org.za

ORiON
ISSN 0529-191-X

Multi-objective optimisation with stochastic discrete-event simulation in retail banking: A case study

E Scholtz* J Bekker† D du Toit*

Received: 15 March 2012; Revised: 7 November 2012; Accepted: 8 November 2012

Abstract

The cash management of an autoteller machine (ATM) is a multi-objective optimisation problem which aims to maximise the service level provided to customers at minimum cost. This paper focus on improved cash management in a section of the South African retail banking industry, for which a decision support system (DSS) was developed. This DSS integrates four Operations Research (OR) methods: the vehicle routing problem (VRP), the continuous review policy for inventory management, the knapsack problem and stochastic, discrete-event simulation. The DSS was applied to an ATM network in the Eastern Cape, South Africa, to investigate 90 different scenarios. Results show that the application of a formal vehicle routing method consistently yields higher service levels at lower cost when compared to two other routing approaches, in conjunction with selected ATM reorder levels and a knapsack-based notes dispensing algorithm. It is concluded that the use of vehicle routing methods is especially beneficial when the bank has substantial control over transportation cost.

Key words: Retail banking, vehicle routing, knapsack, computer simulation, inventory.

1 Introduction

Retail banks provide one or more services to a large proportion of South Africans and thus affect the lives of many citizens. Despite the drive of retail banks towards a cashless society, cash is still the preferred transaction medium of many — not only in South Africa, but also in the rest of the world. ATMs form a key part in cash provision. It is thus important, from the perspective of retail banks and clients alike, to ensure that cash levels in ATMs are sufficient for as long a period as possible.

Providing a 100% service level for one ATM is possible, but as no bank has only one ATM, providing a high level of service to an entire network of ATMs is challenging, for

*Department of Industrial Engineering, University of Stellenbosch, Private bag X1, Matieland, 7602, Republic of South Africa

†Corresponding author: Department of Industrial Engineering, University of Stellenbosch, Private bag X1, Matieland, 7602, Republic of South Africa, email: jb2@sun.ac.za

various reasons. An ATM network consists of a number of ATMs, each with its own stochastic, seasonal customer demand profile, and a count house (CH) from which ATMs are replenished. Cash must be distributed from the CH to the individual ATMs, which involves distribution cost, delivery delay cost and the cost of carrying inventory. Cash transit requires vehicles, labour and rigorous security, and involves high risk. To a bank, cash in an ATM is not earning interest and is therefore adding to existing inventory costs.

ATM demand varies with time and is subject to trends that follow weekly, monthly and annual cycles. In addition to these cycles, stochastic, time-based behaviour of users is driven by events such as paydays and holidays [13]. Arguing that the first step in effective cash management is the accurate prediction of daily cash withdrawals, Adendorff [1], Seedig [12] and Simutis [13] suggest various prediction models for this purpose. Adendorf [1] suggests the combination of forecasting and a decision support model for improving the replenishment process of a South African retail bank. Simutis [13] investigated two different methods for forecasting the daily cash demand of ATMs: a method based on flexible artificial neural networks (ANN) and a method using the support vector regression (SPR) algorithm. Seedig [12] applied a prediction algorithm, referred to as a Fuzzy c-Neural Network Model (FCNNM), to data representing daily cash withdrawal amounts. The FCNNM was compared to a Fuzzy c-Autoregression Model algorithm and found to be superior.

Another approach to cash management found in the open literature is that of inventory management. Armenise [2], Miranda [9] and Castro [4] follow this approach. Armenise [2] used a genetic algorithm to determine optimal cash replenishment strategies for 30 ATMs by varying reorder points and reorder quantities. The objective was a solution in which inventory levels are kept to a minimum without the cash residual dropping to zero. Miranda [9] developed a decision support system (DSS) combining periodic and continuous review policies for inventory management. The DSS is simulation-based and can be used to determine the optimal values of reorder points and reorder quantities. The stochastic nature of the problem is accounted for by using various statistical distributions. Castro [4] formulated a stochastic program of which the objective is to minimise the total inventory cost involved by determining the amount of cash that should be placed in an ATM during a certain period.

Also taking the inventory management approach, but extending it to include vehicle routing, Wagner [15] developed a conceptual framework for finding the optimal cash deployment strategy for a network of ATMs. For the purpose of his study, 'optimal' is equivalent to 'minimum cost'. Conceptually, Wagner assumes the existence of a perfect forecasting model, effectively eliminating the actual stochastic nature of demand. Focus is placed on the development of primarily deterministic mathematical and simulation models.

In this paper, we propose a DSS for cash management and distribution in an ATM network that can be used by a retail bank. For the purpose of experimentation, we obtained data for a small network of 18 ATMs situated in the rural Eastern Cape, South Africa. The ATMs are serviced from a central point (the CH). Distances between the CH and ATMs, and between ATMs, are long. Several cash-in-transit (CIT) vehicles service an ATM network at costs negotiated between the banking group and a CIT company. The cost structure negotiated and the resulting cost to the bank are dependent on the total distance covered by the CIT vehicles. Costs associated with covering the distances between these ATMs may therefore be significant.

It was found that forecasting was effective for managing cash orders from the CH to its upstream supplier, but not from ATMs to the CH. This could be ascribed to the rate at which ATM inventories were depleted in the area we investigated. Focusing on the cash supply from the CH to ATMs, it was decided that an inventory management approach would be more appropriate. Due to the high costs associated with cash deliveries from the CH to ATMs, transportation cost had to be taken into account in developing a DSS that would improve cash management.

The South African currency is the rand (ISO 4217 currency code 'ZAR') and is abbreviated 'R'. Five denominations are currently in use: R10, R20, R50, R100 and R200. Analyses by other researchers consider daily withdrawal amounts or total notes withdrawn over a period. This effectively ignores the discrete nature of cash dispensing: the inventory level in an ATM does not simply drop by R250 when a customer demands R250, but the numbers of units of for example R200 and R50 notes (if available) become less. We argued that, to meet customer demand, the availability of a specific combination of notes for every withdrawal event must be taken into account. To do so, an algorithm was determined according to which the notes are dispensed to make up the amount of cash demanded by a customer.

Finally, it was reasoned that the single-objective optimisation approach found in literature was not sufficient. Cash management is a multi-objective optimisation (MOO) problem in which service levels have to be maximised at minimum cost. The proposed DSS has four elements, namely a vehicle route planner, a continuous review inventory manager and a cash-dispensing algorithm, all of which are based on a discrete-event computer simulation platform. These are discussed in the next section, followed by an explanation of the experimental setup in Section 3. The results of the experiments are discussed in Section 4, followed by conclusions in Section 5.

2 Decision support system

A DSS was developed to improve cash management in an ATM network while considering both optimisation objectives, namely service level and total cost. This DSS is an integration of:

- discrete-event, stochastic simulation, which was used to compare different strategies while emulating the ATM network dynamics over a three-month business period
- an adaptation of the knapsack problem according to which notes are dispensed
- the vehicle routing problem
- the continuous review policy for inventory management.

The DSS is discussed in the following subsections, based on these four elements.

2.1 Discrete-event, stochastic simulation in the DSS

The ATM network in question is a complex system. The machine dispenses denominations according to a specified algorithm for seasonally variable demand unique to each ATM. The inventory (cash) in an ATM is reduced over time, until a critical inventory level is reached. Given that some ATMs require replenishment and others not, an optimal (or near-optimal) vehicle routing plan needs to be determined. Additionally, this is an MOO problem and not only one, but two optimisation objectives need to be considered (minimum cost and maximum service level). The problem thus yields itself very well to simulation.

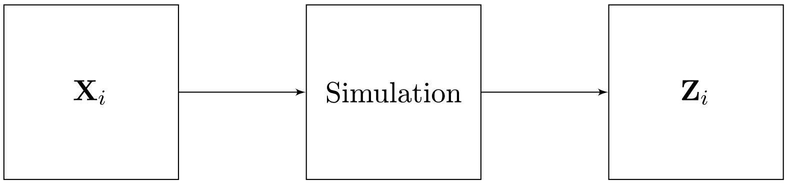


Figure 1: *The schematic shows simulation as part of an MOO decision support system. The model takes an input vector* $\mathbf{X}_i$ *and simulates scenario* i *to obtain the output vector* $\mathbf{Z}_i$*. The elements of* $\mathbf{Z}_i$ *follow certain distributions if the model contains one or more stochastic elements.*

In essence, simulation enables analysts to evaluate the effects of changes in certain variables on the complex system studied. The analyst decides on a finite set of scenarios ($\mathbf{X}_i$) for the simulation study. These scenarios are combinations of variable values chosen from the decision space and serve as an input to the simulation model. For each scenario, the simulation model yields one or more related response distributions. A specific reorder level and vehicle routing strategy would, for example, produce a certain response. These responses make up $\mathbf{Z}_i$, a set of objective function values serving as the objective space for the MOO problem. From the objective space, a Pareto front containing good solutions may be obtained. Figure 1 shows this process: scenarios serve as input to the simulation model which yields the objective space. In this study, the values of several decision variables will be mapped onto a two-dimensional objective space with simulation.

The simulation model of the DSS is made up of two main and several secondary processes. The first primary process is the arrival and servicing of customers. Customers arrive at any of the 18 ATMs according to a unique, seasonal arrival rate. An arbitrary snapshot of the customer arrival rate for a specific ATM in the network is shown in Figure 2. After arrival, customers demand a cash amount. This quantity is drawn from a demand distribution which is also unique for every ATM. Part of such a distribution is shown in Figure 3. Each customer is served using the note-dispensing algorithm. Inventory levels in an ATM decrease with each customer's arrival and successful service. Once inventory levels in an ATM fall below the reorder point s, a reorder flag is raised for that ATM, registering an order. The reorder point may easily be varied so as to evaluate the consequences of changing it.

The second primary process is vehicle routing. The number of reorder flags is monitored continuously. At the beginning of each day all reorder flags are counted and if the number of flags is greater than the routing point value, routes are planned and vehicles are

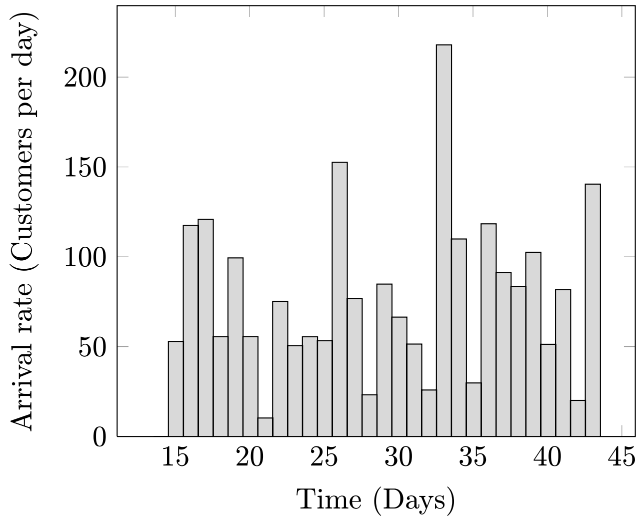


Figure 2: *A snapshot of the customer arrival rate for ATM i shows the time-varying nature of the process.*

dispatched. The routing point is defined as the minimum number of reorder flags that have to be up before a replenishment route will be determined. Also, if more than one route must be planned (two or more vehicles are considered, or one vehicle must service a number of routes over two or more days), then a minimum number of ATMs must be present on a route. For example, if ten ATMs require replenishment, and the first route is created with five ATMs included, and the second with four ATMs, that would leave one ATM. This ATM will be carried over to the next planning event and is thus deliberately neglected.

Determining routes is only the first step in the replenishment process and vehicle availability must also be taken into account: only vehicles that are at the CH may be loaded and sent on a route. In addition, the time required to complete a route must be considered: a vehicle may only be sent out if there is enough time left in the business day to complete a route. If the time left until the end of the business day is less than the expected trip and delivery time, the route may only be initiated and completed the following day. Once these factors have been accounted for, available vehicles are loaded and sent out. Vehicles travel from the CH to all ATMs on their assigned routes, replenish the ATMs and then return to the CH. Secondary processes merely enable primary processes.

The processes were simulated by replicating a time period of three business months.

2.2 Algorithm for dispensing notes

South African ATMs typically do not carry R10 notes. The ATMs in the network that was studied all contain five cash canisters of which three are filled with equal numbers of R20, R50 and R100 notes. The remaining two canisters contain R200 notes. If, for example, a customer requests R500, the ATM can dispense twenty-five R20 notes, or ten R50 notes, or five R100 notes or a multitude of other denomination combinations. As long as all required denominations are in stock, the only constraint is that the sum of the

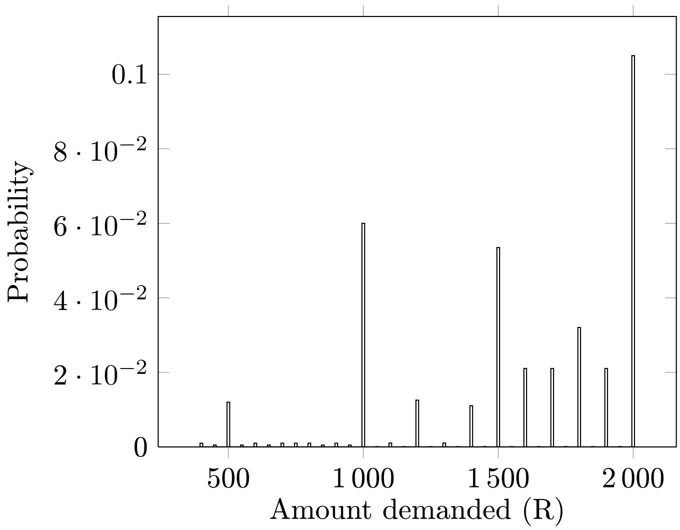


Figure 3: *An example of the demand distribution for an ATM.*

monetary values of the notes dispensed must equal the amount A_c requested.

Two main algorithms for dispensing notes are used in industry: least-note-picking and even-note-picking. Given that an ATM cannot dispense fractions of notes, the field of integer programming has been investigated for possible note dispensing algorithms. The bounded knapsack problem (BKP) [6] lends itself to be used as an algorithm for denomination dispensing. The combination of notes dispensed will occasionally be constrained by the number of notes u_j of a certain denomination j that is available. The quantity available of denomination j is b_j. Every denomination has a profit p_j equal to the monetary (or face) value of the note. There are, however, significant differences between the BKP and the suggested algorithm, of which the problem formulation is shown in (1)–(4). Let u_1, u_2, u_3 and u_4 respectively be the number of R20, R50, R100 and R200 notes that are issued during a cash withdrawal. The objective then becomes to

$$\begin{aligned} \text{minimise} \quad & u_1 + u_2 + u_3 + u_4 && (1)\\ \text{subject to} \quad & 20u_1 + 50u_2 + 100u_3 + 200u_4 = A_c, && (2)\\ & 0 \leq u_j \leq b_j, \quad j = 1, 2, 3, 4, && (3)\\ & u_j \text{ integer}, \ j = 1, 2, 3, 4. && (4) \end{aligned}$$

According to the logic that dispensing as few notes as possible would lead to the fewest replenishment runs and subsequently to the lowest total cash management cost, the number of notes to be dispensed is minimised, instead of maximising profit to be achieved from a selection of items (as in the BKP).

An ATM may never dispense less (or more) than the amount A_c requested by customer c. The constraint in (2) ensures that exactly the requested amount is issued by means of the four types of notes available.

The constraints on the four types of notes are specified by (3) and (4).

2.3 Continuous review policy for inventory management

Dealing with uncertain demand, stochastic models include the news vendor problem and the (s, S) policy [16]. Ravindran [11] mentions stochastic multi-echelon inventory models, and Bellman [3] discusses a stochastic dynamic programming approach referred to as 'the optimal inventory equation'.

The inventory situation in the network studied may be described as a stochastic, multi-echelon system. Inventory levels in the CH may be described as multi-period, while inventory levels in ATMs are managed as either single-period or multi-period. Notes may be added to the cash remaining in an ATM during replenishment (this is called a 'cash-add' system and would be multi-period management) or the remaining notes are removed and replaced with newly packed canisters (a single-period inventory management technique referred to as 'cash-swap').

The scope of this paper does not cover the entire multi-echelon inventory system. It focuses only on inventory management of the cash in ATMs, assuming infinite availability of notes at the CH, because the CH maintains a conservative buffer.

The inventory management model developed uses the cash-swap approach described above. This effectively means that order size is equal to a fixed S which will restore the inventory levels to 100% every time an ATM is replenished. Figure 4 shows a schematic of the (s, S) policy which was implemented in the simulation model.

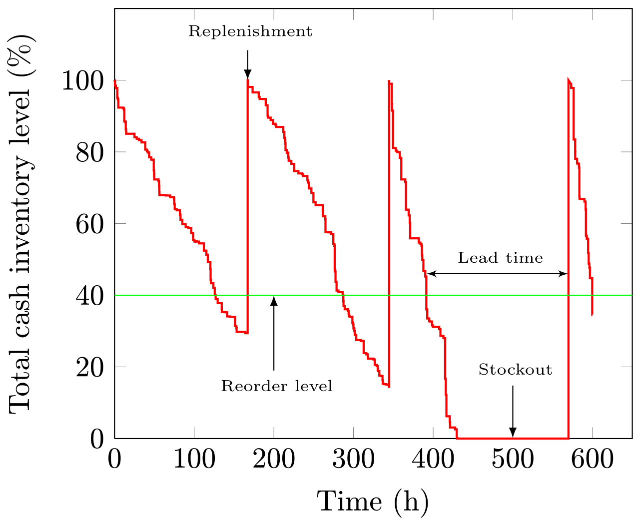


Figure 4: *The schematic of the (s, S) inventory policy shows that the reorder point s can be varied, while the reorder quantity S is fixed because ATMs are filled completely during replenishment. The lines exhibit steps due to the discrete nature of the process.*

Order size S is based on operational data provided by a retail bank: canisters containing cash can at most contain 2 500 notes each, but dispensing problems occur when canisters are filled to the brim. To prevent such problems, newly packed canisters are filled with 2 000 notes each. Canisters filled with R200 notes tend to run empty first. For this reason, two of the five canisters available in an ATM are packed with R200 notes. The maximum numerical value of S is R1 140 000, but a bank may decide to lower the maximum value.

This is to reduce loss when an ATM is damaged because of theft.

In this study, S was not varied; only changes to s were investigated. The reorder point s is either the total monetary value of cash left in the ATM, such as R350 000, or the total number of notes left in the ATM. The reorder point can also be the number of notes left of a certain denomination: an order could, for instance, be triggered once there are fewer than five hundred R50 notes available. The variations in s that were experimented with are discussed in Section 3.4.

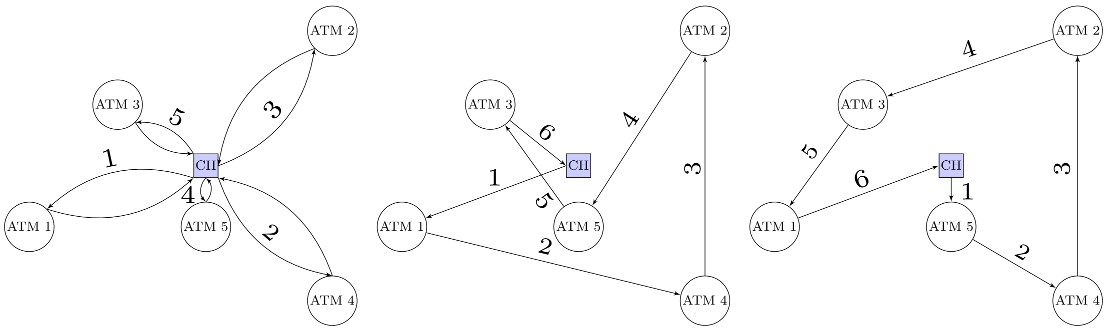
2.4 Vehicle routing in the ATM network

The CIT company usually has two alternatives for cash transportation: scheduled CIT vehicles or dedicated CIT vehicles. A scheduled CIT vehicle is controlled entirely by the CIT company providing the vehicle. A dedicated CIT vehicle is manned by a private CIT company but routed by the bank. Dedicated vehicles provide banks with greater control over cash deliveries (the bank determines the date, time, place and route of a delivery), but at greater cost.

Three routing methods were included in the DSS. The current routing method used in the ATM network is referred to as 'direct replenishment' and dispatches vehicles on a one-ATM-at-a-time basis. A vehicle is sent to the ATM that has requested replenishment and returns to the CH on completion of the task. The next ATM requesting service is then attended to, and so on. A second routing method is intuitive and is called 'first-come-first-served' routing (FCFS). ATMs are replenished by a series of visits from one or more CIT vehicles, in the order in which they request service. The third routing method uses a formal application of the vehicle routing problem (VRP), which we shall refer to as the 'vehicle routing method' (VRM). This method routes vehicles to replenish ATMs according to a shortest route, in the time available.

The three methods are conceptually shown in Figure 5 — assuming the five ATMs indicated require replenishment, and their requesting sequence was 1–4–2–5–3. The schematic given by Figure 5(a) shows the current method of 'One-ATM-at-a-time'. The schematic given by Figure 5(b) shows the 'first-come-first-served' routing method. The numbers on the arcs show the replenishment sequence followed due to the requesting sequence. The schematic given by Figure 5(c) shows a typical shortest route within a time window proposed by a VRP algorithm. This VRP is solved using a branch-and-bound based heuristic. The algorithm used to determine the lower bound is a version of the greedy algorithm, while a backtrack algorithm was expanded from an algorithm provided by Navernouk & Chu [10]. A route combination algorithm was adapted from an algorithm for lexicographic combinations suggested by Knuth [8]. Due to the fact that the network under consideration is located in a rural area, large distances need to be covered between ATMs. Physical distances used in this study are shown in Table 1. CIT vehicles need to be back at the CH by the end of the working day. The VRP, and its implementation as the VRM, is thus constrained by the daily working hours available.

For the purposes of the problem, it is assumed that the constraint on vehicle capacity is insignificant compared to the time constraint. In other words, a CIT truck would be able to carry all the cash required to service a route (however, this increases the safety risk



(a) One-ATM-at-a-time method. (b) First-come-first-served method. (c) Vehicle routing method.

Figure 5: *Three routing methods were used as part of the study. The method in (a) shows that an ATM is serviced when it requests service, and the vehicle returns to the CH on completion. ATM 1 first required service, then ATM 4, ATM 2, ATM 5, and ATM 3. In the case of the method in (b), the ATMs are served in the sequence they request service, but the return trips to the CH are eliminated, except for the last trip labelled '6'. The method in (c) shows application of a vehicle routing method: the requesting ATMs are served in the sequence determined by the route plan.*

of the personnel and the vehicle). The time required to complete a route is a function of distance. Note that the distances between stations are symmetrical. The suggested VRP formulation is thus an adaption of the symmetric distance-constrained vehicle routing problem put forth by Toth [14].

The DSS was implemented in the simulation package Arena [7] and its Visual Basic for Applications (VBA) environment. The experimental setup is discussed in the next section.

3 Experimental setup

The proposed DSS was evaluated by experimentation in the context of specific business parameter settings. The controlled variables that were adjusted during experimentation are discussed next.

3.1 Elementary sensitivity analysis of travelling distance cost

An existing cost structure (the status quo) determines the cost of cash distribution in the ATM network. A number of vehicles are hired per month, each with f free kilometres per month. We realised that the cost structure can be adapted so that the bank receives a better service at lower cost. An adjusted cost structure was thus created: it was decided to halve the cost of vehicles per month, while quadrupling the cost of kilometres falling outside of the free kilometres that make up part of the CIT service. The number of free kilometres was reduced to compensate for the lower capacity cost.

ATM	*CH*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*
CH	0	96	12	10	8	173	82	80	116	85	96	118	50	15	144	12	80	116	96
1		0	96	96	96	227	85	174	209	156	106	81	144	96	216	96	174	209	106
2			0	14	15	173	82	80	116	85	96	118	50	9	144	1	80	116	96
3				0	8	173	82	80	116	85	96	118	50	7	144	14	80	116	96
4					0	173	82	80	116	85	96	118	50	11	144	15	80	116	96
5						0	148	131	125	88	268	290	166	173	53	173	131	125	268
6							0	132	163	88	167	162	123	15	144	82	132	163	167
7								0	35	65	170	194	45	80	80	80	1	35	170
8									0	85	201	223	75	116	75	116	35	1	201
9										0	180	204	86	85	59	85	65	85	180
10											0	37	125	96	240	96	170	201	1
11												0	147	118	262	118	194	223	37
12													0	50	117	50	45	75	125
13														0	144	9	80	116	96
14															0	144	80	75	240
15																0	80	116	96
16																	0	35	170
17																		0	201
18																			0

Table 1: *The table shows typical distances between ATM locations (km). Although there are alternative routes between some destinations, these distances were used during the study.*

3.2 Varying the number of vehicles

Currently, two dedicated CIT trucks are used in the ATM network. We varied the number of vehicles and developed scenarios in which one, two or three vehicles are available.

3.3 Varying the routing point value

Every day, various ATMs request refilling. The number of requesting ATMs depends on the values of other control variables, for example the reorder point values. Dispatching one or more vehicles may be delayed until a set number of ATMs have requested service. This variable is the 'routing point value' discussed in Section 2.1 and refers to the minimum number of ATMs that need to have registered an order before a route will be determined. It is mainly dependent on the routing method. Direct replenishment routes contain only one ATM and routes are subsequently created at a routing point equal to one. On the other hand, the FCFS method and VRM have to contain more than one ATM (the routing point has to be greater or equal to two) and the actual value is determined by the decision maker.

3.4 Varying the reorder point value

The 'reorder point' s refers to the inventory level at which an ATM triggers an order. As discussed in Section 2.3, many variations are possible for defining the reorder point because either the cash value or the number of notes in an ATM may be seen as the primary measurement of inventory level. For the purposes of this study, the cash value of notes left in an ATM serves as measurement of the inventory level in an ATM. The reorder point for all ATMs in the model are the same. Reorder points of R150 000, R350 000 and R550 000 were used to ensure a wide spread of values.

From the above, the number of experiments is as follows: for the VRM, there are two routing point values, three reorder points and three vehicles, giving $2 \times 3 \times 3 = 18$ experiments. The same applies to the FCFS routing method. The direct replenishment routing method implies one routing point value, resulting in $1 \times 3 \times 3 = 9$ experiments. The total is thus 45 experiments for one cost structure, which results in 90 experiments for two cost structures. All experiments (or scenarios) were simulated for a period of three real-world business months, using the appropriate data provided for analysis.

3.5 Output parameters

The following output parameters were studied:

1. Service level
2. Total cost
3. The total distance covered by CIT vehicles
4. Capacity cost
5. Travelling distance cost
6. Rebanking cost
7. Opportunity cost

These parameters will be clarified in the following discussion. Two main model output parameters were measured for decision making, namely the service level and the total cost of a scenario. The *service level* is defined as

$$S_L = \frac{N_S - N_U}{N_S} 100\%,$$

where N_S is the number of simulated customers requesting service, and N_U is the number of simulated customers whom the ATMs could not supply with the desired amounts. Note that if a simulated customer requested say R500, and there was only R400 available, it was considered to be an unsatisfactory transaction. In real life a customer would generally accept the amount available, but we avoided simulation of detailed customer behaviour.

Our *cost* formulation is based on the work of Daganzo [5] as reorganised by Wagner [15]. The total cost is the sum of the four of costs listed above, namely the capacity cost (C_C), the travelling distance cost (C_D), the rebanking cost (C_R) and the opportunity cost (C_O).

These four cost types are now discussed in more detail. Note that all cost types must be calculated over a standard time period, for example a month.

1. The **capacity cost** C_C is a function of the number of vehicles employed (K) and the fixed cost C_V associated with each vehicle per standard cost calculation period. The capacity cost is given by $C_C = C_V K$.
2. The **travelling distance cost** C_D is the sum of the distance cost C_{D_k} per vehicle with $k = 1, \ldots, K$. The distance cost C_{D_k} is a function of the distance D_k travelled by each vehicle during the standard cost calculation period and is dependent on f, the number of free kilometres available, and c_{km}, the unit cost per kilometre. Thus

the distance cost is calculated using

$$C_{D_k}(D_k) = \begin{cases} 0 & \text{if } D_k \leq f, \\ (D_k - f)c_{km} & \text{if } D_k > f, \end{cases} \quad \text{for} \quad k = 1, \ldots, K \text{ and}$$

$$C_D = \sum_{k=1}^{K} C_{D_k}(D_k).$$

3. The **rebanking cost** C_R depends on the cash value left in an ATM at the time of its replenishment, I_r, and the rebanking cost per R100, c_R. It is calculated by means of

$$C_R = \frac{\sum_{r=1}^{n_r} I_r \times c_R}{100}$$

where n_r is the total number of replenishment events that occurred during the standard cost calculation period.

4. The **opportunity cost** C_O is calculated by multiplying the effective daily interest rate r_j with the cash I_{ij} left in ATM i at the end of business day j. Opportunity cost is calculated on a daily basis for each business day in the standard cost calculation period, until the last day in this period, denoted by d. These costs are then summed for each ATM i, using

$$C_O = \sum_{i=1}^{18} \sum_{j=1}^{d} r_j I_{ij}.$$

The experimental results are discussed in the following section.

4 Results

The 90 experiments may be categorised according to any of the control variables. Forty-five experiments were run using each of the two cost structures (the current one and the proposed alternative). Thirty experiments used a reorder point equal to R150 000, another 30 were run with the reorder point set to R350 000, and the remaining 30 used a reorder point of R550 000. Similarly, 36 experiments used the VRM, and 36 experiments used first-come-first-served routing. For the remaining 18 experiments, vehicles were routed according to the direct replenishment approach. For each control variable, however, experiments are categorised to illustrate the effect of the control variable. Table 2 and Table 3 show detailed results for arbitrarily selected experiments from the set of 90 experiments, followed by discussions of summarised results. In Table 2, the VRM is followed, and in Table 3, direct replenishment is performed.

The summarised results are discussed next, by control variable category.

4.1 The effect of the cost structure

The two cost structure experiments each contained 45 scenarios. The difference between results for the status quo and the adjusted cost structures is shown in Figure 6. The status

EXPERIMENT 11					
Details					
Routing method	VRM		Number of vehicles used	2	
Routing point	3		Reorder point for ATMs	R350 000	
Minimum number of ATMs on route	2		Number of replications	80	
Results					
Output	*Average*	*95% CI half-width*	*Output*	*Average*	*95% CI half-width*
OPERATIONAL			COST (R)		
Service level	0.925	0.00237	Total cost	668 190	1 700.90
Number of unsatisfied customers	8 673.3	272.78	Capacity cost	472 260	0
Total distance travelled (km)	21 794	139.51	Cost of distance travelled	33.55	26.37
Number of routes determined	86.587	0.638	Opportunity cost	142 440	566.23
Number of routes completed	84.8	0.615	Rebanking cost	53 454	1 174.50
Average vehicle utilisation	36.3%	0.27%			

Table 2: *The table shows detailed results for Experiment 11. The experimental settings are shown under "Details". The point estimators of the various output parameters are shown with 95% confidence intervals.*

EXPERIMENT 41					
Details					
Routing method	Direct replenishment		Number of vehicles used	2	
Routing point	1		Reorder point for ATMs	R350 000	
Minimum number of ATMs on route	1		Number of replications	80	
Results					
Output	*Average*	*95% CI half-width*	*Output*	*Average*	*95% CI half-width*
OPERATIONAL			COST (R)		
Service level	0.950	0.0021	Total cost	713 440	2 294.60
Number of unsatisfied customers	5 765.1	242.15	Vehicle cost	472 260	0
Total distance travelled	32 313	141.36	Cost of distance travelled	16 945	440.79
Number of routes determined	218.13	1.1716	Opportunity cost	152 560	579.03
Number of routes completed	212.93	1.16	Rebanking cost	71 672	1 439.20
Average vehicle utilisation	53.7%	0.25%			

Table 3: *Detailed results for Experiment 41. The experiment settings are given under 'Details.' It is in contrast to the results of Experiment 11 in Table 2.*

quo cost structure resulted in three bundles of results. This is due to the fact that the capacity cost for this scenario is so large that adding an additional vehicle results in a giant step in cost. These steps made it difficult to analyse the effect of routing methods on the system.

The adjusted cost structure was created in such a manner that it would place greater emphasis on distance cost than the status quo cost structure. To shift the emphasis, capacity cost in the new structure was reduced and the cost associated with the distance covered was increased by a factor of four.

No adjustments were made to rebanking and opportunity costs. All differences observed between the two cost structures are a result of reduced vehicle and increased distance costs. There is also no difference in the way operational performance measures (such as service level, distance travelled, the number of routes determined, vehicle utilisation, *etc.*) were estimated between the two cost scenarios.

There is more cost variation in the results obtained for the adjusted cost structure than for the status quo. Despite the higher levels of variation, the adjusted cost structure yields lower cost results than the status quo.

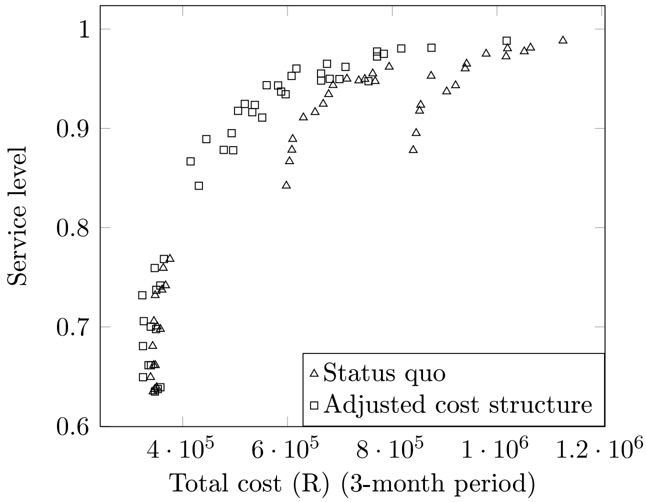


Figure 6: *The overall difference between the status quo and the adjusted cost structures is shown here. The status quo cost structure shows clusters of solutions resulting from the distinct increase in cost when an additional vehicle is added. The results of the adjusted cost structure are more evenly distributed due to the greater emphasis on distance travelled.*

4.2 Varying the number of vehicles

The effect of varying the number of vehicles is discussed in this section.

For one vehicle, the service level cannot be improved by much, while the total cost also does not vary much. Using one vehicle chokes the business process, hence the low service level and little variation in cost. The service level improves significantly when two vehicles are available instead of only one. Achieving the higher service level requires that a greater overall distance be covered. Two vehicles are able to cover a distance that one vehicle is incapable of doing. The increase in service level from two vehicles to three is not as significant and subsequently there is not a significant difference in distance covered. In general, cost related to distance increases from one vehicle to two, but drops from two vehicles to three. The cost due to the distance travelled drops because of the fee structure: for every vehicle used, there are free kilometres available in a month. Adding a vehicle thus increases the number of free kilometres, which leads to a reduction in distance cost.

As Figure 7 illustrates, using three vehicles generally achieves the highest service levels, but at very high cost. Having two vehicles available results in a slight decrease in service level (compared to using three vehicles) at significantly reduced cost. Routing only one vehicle yields poor service levels at low cost.

4.3 Varying the routing method

Vehicle utilisation is highest for the direct replenishment method and lowest for the VRM. Likewise, the number of routes completed using direct routing is the highest, whereas the VRM requires the least number of routes. Looking at Figure 8, it may generally be observed that vehicle routing results in high service levels at low costs. The best service level is achieved using direct routing. However, this very high service level (98.8%) is

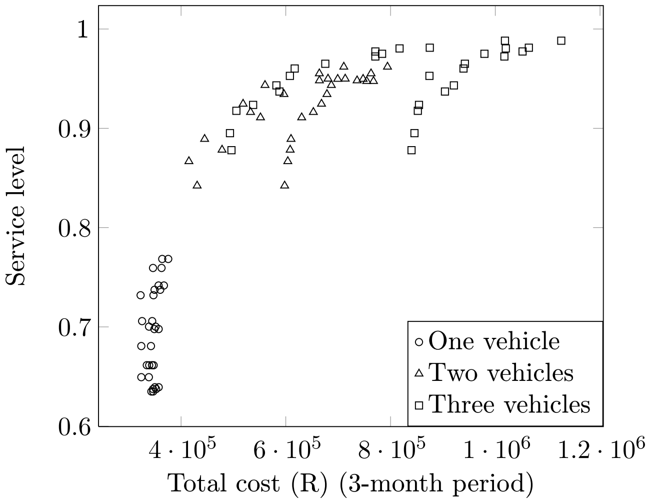


Figure 7: *The figure shows the overall effect of the number of vehicles used in the cash distribution. Adding a vehicle increases the number of free kilometres offered by the service contract, hence a slower increase in cost. One vehicle cannot handle the workload and simply chokes the service level with little variation in cost. A third vehicle adds more to the cost than it improves the service level.*

achieved at a very high cost (R1 126 000). The worst service level also results from direct routing. First-come-first-served routing yields intermediate results over all experiments.

4.4 Varying the routing point value

The effect of the routing point value was small. Values of one, two, or three ATMs affected the service level, but their effect on cost was minimal. To obtain as high a service level as possible, a higher routing point should be used if very few vehicles are used, while a lower routing point value yields better results if more vehicles are employed.

4.5 Varying the reorder point value

The reorder point value has a significant effect on the service level and total cost. It is directly related to both, and the results are shown in Figure 9.

As the reorder point rises, the number of routes determined, and covered, escalates. Due to the fact that more routes need to be traversed, vehicle demand increases as the reorder point is raised. As the number of routes traversed increases, the total distance covered increases. Increased distance leads to higher distance cost.

4.6 Summary of global observations

The categories of results discussed above provide a good indication of the nature of the ATM network that was modelled and how it was affected by certain variables.

An interesting *total cost* $\rightarrow$ *service level* $\rightarrow$ *total cost* cycle was observed: adjusting the routing variables (increasing the number of vehicles or lowering the routing point) increases

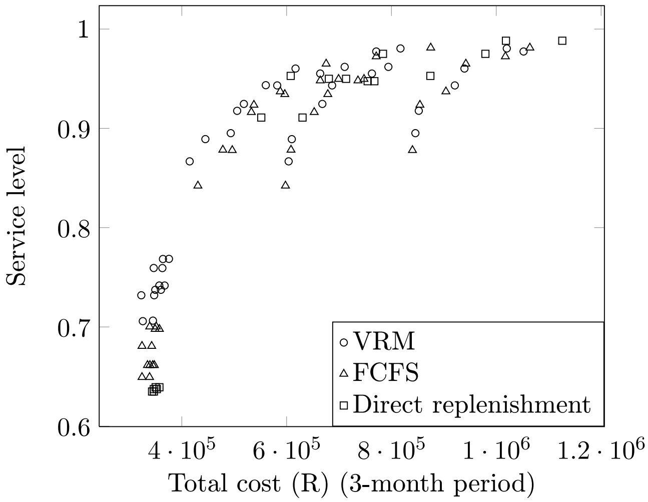


Figure 8: *The figure shows the overall effect of the routing method. The direct replenishment method yields very high service level values but at high cost, while the worst service level values also result from this method. The VRM yields good service levels at reasonable cost. The FCFS routing generally gave mediocre results.*

total cost due to higher transportation costs, but leads to a service level improvement. The improved service level, in turn, results in an escalation of opportunity and rebanking costs.

Other than this observation, main points that deserve to be highlighted are:

- Total cost and service level are both directly related to the number of vehicles available. For one available vehicle, the results exhibit characteristics that differ significantly from scenarios where more than one vehicle is available.
- The VRM consistently provides high service levels at low cost (when compared to cost resulting from the other two routing methods). If more than one vehicle is available, direct replenishment yields very high service levels, but at high total cost.
- Although the routing point has an effect on results, it does not bear a strong relation to service level or total cost.
- The reorder point is strongly related to service level and total cost.

The observation that the system reacts differently when vehicle availability is limited, or not, is important. When only one vehicle is used, routing efficiency is of the utmost importance. When more vehicles are used, high speed of delivery (a measure of the time between order placement and order fulfilment) yields good results.

5 Conclusions

The aim of this paper was to develop and evaluate an integrated multi-objective optimisation decision (MOO) support system that may be used for cash management and distribution in retail banking. This decision support system integrates the continuous review policy for inventory management, the vehicle routing problem (VRP), an adaption of

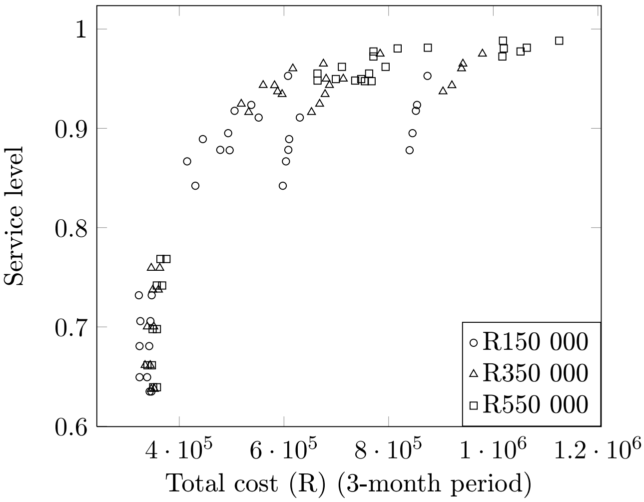


Figure 9: *The overall effect of the reorder point value shows that a low reorder point limits the service level, whereas a high reorder point value increases the service level, but at high cost. When only one vehicle is used for cash distribution, a high reorder point value does not improve the service level because the vehicle cannot cope with the workload.*

the knapsack problem for dispensing notes and stochastic discrete-event simulation, which was used to compare different strategies.

Analysis of experimental results indicates that the application of the VRP to an ATM network results in a low cost, high service level strategy. It is also clear that the continuous review reorder point s is directly related to cost and service level alike. The scenarios including the adjusted cost structure (as opposed to the status quo) provides the majority of the Pareto optimal solutions because of the lowest cost totals. Renegotiating the current cost structure to be the same as the one experimented with, is not necessarily feasible. The concept illustrated is important, though: renegotiating the current cost structure so that the variable component of the transportation cost is larger and the fixed component smaller, would provide the bank with more control over transportation cost. This could result in significant cost savings, depending on how this additional control is managed.

The nature of the MOO problem is such that there is no single optimal cash management strategy: improving the service level leads to escalated total cost. Decision makers therefore need to decide on the service level required and pay the price, or decide on a total cost level and accept the associated service level. The scatter plot in Figure 10 shows the results for the 90 scenarios and the Pareto optimal set of solutions from which decision makers need to choose a scenario best suited to their requirements.

Note that all members of the Pareto optimal set result from the adjusted cost scenario. This reinforces the recommendation made above that greater control over transportation cost could lead to a significant reduction in total cost. Three possible decisions following the results are shown in Table 4. On the one extreme, lowest cost is considered important, while the service level suffers. On the other extreme, a high service level is achieved, but at significant cost. The authors suggest a compromise as shown, which gives a high service level at a cost between the cost values of the two extremes cases.

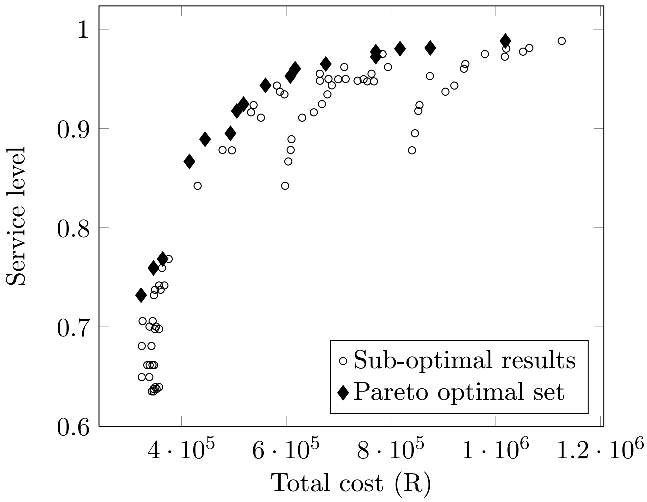


Figure 10: *The set of Pareto optimal solutions for the MOO problem. All solutions in this set come from the subset of adjusted cost scenarios, which shows that the current cost agreement can be improved.*

		Status quo	Adjusted cost
		Experiment 19	*Experiment 49*
	Service level	**0.6493**	**0.7319**
	Total cost	**R338 540**	**R322 730**
Minimum cost	Number of vehicles	1	1
	Routing method	FCFS	VRM
	Reorder point	R150 000	R150 000
	Routing point	2	3
		Experiment 45	*Experiment 90*
	Service level	**0.9882**	**0.9882**
	Total cost	**R1 126 000**	**R1 018 500**
Maximum service level	Number of vehicles	3	3
	Routing method	Direct replenishment	Direct replenishment
	Reorder point	R550 000	R550 000
	Routing point	1	1
		Experiment 9	*Experiment 54*
	Service level	**0.9619**	**0.9619**
	Total cost	**R794 190**	**R710 760**
Suggested compromise	Number of vehicles	2	2
	Routing method	VRM	VRM
	Reorder point	R550 000	R550 000
	Routing point	2	2

Table 4: *When doing multi-objective optimisation, the decision maker is confronted with a set of good solutions, and choosing one for implementation is often done using business considerations. Here we show scenarios resulting in lowest cost, highest service level, and a good compromise. A lowest cost solution will result in a maximum service level of only 73%, while a highest service level choice will increase cost by more than three times. A suggested compromise shows a lower cost with a service level in excess of 96%.*

Further work may include expanding the network in terms of the number of ATMs, and identifying a reorder point for each ATM. ATM networks in urban areas may also be analysed. ATMs in urban areas are usually more densely distributed and closer to the CH in terms of distance, but traffic may cause long delivery times. The effect of a bank client going to another close-by ATM when an ATM is empty is also more prevalent in the urban network.

The results and associated decision support provide evidence that several OR methods may be integrated into a useful decision support system for cash management at ATMs.

References

[1] Adendorff SA, 1999, *A decision support model for the cash replenishment process in South African retail banking*, PhD Dissertation, University of Pretoria, Pretoria.

[2] Armenise R, Cosimo B, Sangianantoni E & Troiano L, 2010, *A generative solution for ATM cash management*, Proceedings of the 2010 International Conference of Soft Computing and Pattern Recognition, pp. 349–356.

[3] Bellman RE, 1957, *Dynamic programming*, Dover Publications, New York (NY).

[4] Castro J, 2009, *A stochastic programming approach to cash management in banking*, European Journal of Operational Research, **192(3)**, pp. 963–974.

[5] Daganzo CF, 2005, *Logistics systems analysis*, Springer, New York (NY).

[6] Kellerer H, Pferschy U & Pisinger D, 2004, *Knapsack problems*, Springer, New York (NY).

[7] Kelton D, Sadowski R & Sturrock D, 2007, *Simulation with Arena*, Better World Books, Mishawaka (IN).

[8] Knuth DE, 2005, *The art of computer programming*, 4th Edition, Addison-Wesley, Boston (MA).

[9] Miranda AK & Muñoz DF, 2005, *DSS to manage ATM cash under periodic review with emergency orders*, Proceedings of the 37th conference on Winter Simulation Conference, pp. 2595–2599.

[10] Naverniouk I & Chu F, 2008, *Backtracking: Branch and bound*, [Online], [Cited November 10th 2012], Available from: `http://www.cs.cornell.edu/ wdtseng/icpc/notes/bt2.pdf`

[11] Ravindran AR, 2008, *Operations research and management science handbook*, CRC Press, Boca Raton (FL).

[12] Seedig HG, Grothman R & Runkler TA, 2009, *Forecasting of clustered time series with recurrent neural networks and a fuzzy clustering scheme*, Proceedings of the International Joint Conference on Neural Networks, pp. 2846–2853.

[13] Simutis R, Dilijonas D & Bastina L, 2008, *Cash demand forecasting for ATM using neural networks and support vector regression algorithms*, Paper presented at the 20th EURO Mini Conference: Continuous optimization and knowledge-based technologies, Neringa.

[14] Toth P & Vigo D, 2002, *The vehicle routing problem*, SIAM, Philadelphia (PA).

[15] Wagner M, 2007, *The optimal cash deployment strategy — modeling a network of automated teller machine*, Master's Thesis, Hanken Swedish School of Economics and Business Administration, Helsinki.

[16] Winston WL, 2004, *Operations research: Applications and algorithms*, 4th Edition, Thomson, Independence (KY).