Ô! [•• Á æ ā * Á [| æ ā • Á Á ā | ā Á ^! ç æ ^ •
Á

Á Á Á á! Á Ô [| ^ } Á æ ā Á æ & Á æ à! ^ & @

Cross training policies in field services

P. Colen¹, M. Lambrecht²

¹ Faculty of Business and Economics, Research Center for Operations Management,
K.U.Leuven – Belgium (pieter.colen@econ.kuleuven.be)

² Faculty of Business and Economics, Research Center for Operations Management,
K.U.Leuven – Belgium (marc.lambrecht@econ.kuleuven.be)

Abstract : *As the trend towards more after sales services progresses through the industrial equipment industry, OEMs are now offering comprehensive service contracts in which they take full responsibility for the functioning of their customers' machines. The growing responsibilities allow for more leeway in determining how to maintain the machinery. At the same time the OEM is directly effected by the success of its maintenance operations. This is certainly the case for capacity decisions in its field operations, which is the focus of this paper. One of the most important capacity decisions in field service operations is determining the training level of technicians. A field service organization can employ fully cross trained technicians which are trained to execute any service job. However, training and retaining these employees can be very expensive. Therefore, it can be cost effective to use dedicated technicians who received a focused training enabling them to perform a limited range of service jobs. We conducted a simulation study of the field service operations of an OEM in the compressed air industry to be able to evaluate the possibility to deploy technicians dedicated to preventive maintenance instead of fully cross trained technicians. We analyze the cross training decision for different scenarios in which a considerable part of the installed based is covered by service contracts. Due to these service contracts, maintenance demand as experienced by the field service organization is influenced by the cross training decisions. Our contribution consists of identifying the optimal cross training policy and the factors that influence the benefits of the different workforce configurations while taking into account the effect of the capacity decisions on the maintenance demand. We succeed to do this by characterizing the failure behavior of the individual machines. We find evidence that full cross training might be especially beneficial in a field service setting even when the effect on demand is taking into account.*

Keywords : *Simulation, Cross Training, Field Service, Maintenance, Service Contracts*

1 Introduction

Original equipment manufacturers (OEMs) around the world are recognizing the importance of after sales services (Cohen et al., 2006; Pawar et al., 2009; Sacconi et al., 2007). Excellence in service does not only attract and retain equipment customers but has become an important profit source for companies such as Rolls-Royce, GE, Xerox and many other OEMs. These manufacturers pursue a product service strategy in which they actively promote services to enhance the value proposition of the product offering. As OEMs take over more and more maintenance activities of their customers, the maintenance organizations are becoming larger, more professional and more complex. This complexity is caused by the diversity of the machine park (installed base) that OEMs have to maintain. Typically the installed base will be scattered around the globe, contain a wide range of different machines that are supported by different technologies such as combustion or electronic engines, oil free or oil injected rotors. All these different technologies require specialized skills. In such an environment it is a daunting task to ensure that there is always a properly trained technician nearby to provide service to the customer when required. The easiest way to deal with this problem is to fully cross train all technician, i.e. enabling them to perform any type of service job on any type of machine. With such fully cross trained technicians, skills are no issue when allocating technicians to customers. However, this option may be prohibitive expensive, therefore many service organizations will have technicians that are trained for specific tasks or technologies. In this paper we report on our research concerning the cross training trade-off in a field service organization.

Typically maintenance organizations have to deal with two types of tasks. On the one hand, they will need to fix machines that have failed (repairs or corrective maintenance). On the other hand, they will try to avoid failures by performing preventive

maintenance. Contrary to preventive maintenance, repairs cannot be planned in advance and have to be dealt with in an "as soon as possible" manner. To highlight this difference we will indicate repairs with the term "emergencies" while preventive maintenance jobs are referred to as "non emergencies". Moreover, the professionalization of the aftermarket service industry has led to a surge in the use of multi-period service contracts. In some of these contracts service providers take full responsibility for the functioning of the equipment. They will perform both preventive and corrective maintenance in order to maintain the machine, sometimes they are even advertising up-time guarantees (performance based contracting) (Kim et al., 2007). With such contracts in place, demand for service can be influenced by the OEM by setting proper maintenance policies and corresponding capacity levels. By taking responsibility for the entire maintenance of the machine, the OEM will be able to perform more preventive maintenance which translates itself in less repairs. By doing so, the OEM can steer its service organization towards more planned and less emergency activities. Therefore, comprehensive contracts offer the possibility to lower the uncertainty in the service demand which creates opportunities to optimize the service operations (Colen and Lambrecht, 2010; Oliva and Kallenberg, 2003).

Advanced service providers have to make decisions concerning their workforce and the service product mix. How many technicians of the different skill levels need to be hired/trained? Which service products to offer? These questions are related: by performing more preventive maintenance the amount of breakdowns will go down, allowing a reduction of the workforce that can handle repairs but increasing the need for technicians to perform preventive maintenance. To the best of our knowledge, we are the first to study the impact of selling comprehensive long term service contracts on the capacity requirements in a field service system. We performed a simulation study to investigate the interdependency between the cross training decisions and the service product mix. More specifically, we investigate the consequences of decisions with respect to the maintenance policy, the contract coverage and the machine reliability on the optimal cross training policy. We define a cross training policy as the decision about the number of technicians to train for a specific skill set. A skill set is determined by the machine technologies the technician can handle, e.g. mechanic or electric technology and by the type of service he or she can perform, e.g. installation or repair. In this study, we limit the possible skill sets to two types of service (emergency and non emergency service) and only one technology.

In section two we will present the related literature. In section three the assumptions and performance measures of the simulation model will be explained. Section four clarifies the different scenarios that we tested while the simulation results are presented in section five. In section six, we apply reliability theory in order to study analytically the impact of the maintenance frequency and machine reliability. We conclude in section seven.

2 Literature review

There are two broad fields of literature related to our research. First, our work adds to the literature on capacity planning in service operations with a focus on cross training strategies. Second, we will apply models that are firmly established in the maintenance literature.

In the literature dealing with cross training decisions, there appears to be consensus on the conclusion that cross training of workers can be very beneficial in a manufacturing setting. Several researchers have shown that often even small amounts of cross training can be sufficient to unlock most of the advantages of full cross training (Brusco and Johns, 1998; Cachon and Terwiesch, 2006). Capacity management in services differs from the traditional manufacturing setting by the fact that capacity and demand have to be matched instantaneously. In most service settings unused capacity is lost capacity, i.e. when an employee has no customer to serve (s)he cannot build up inventory to satisfy future demand.

There has been a substantial amount of research done on cross training policies, both for manufacturing and service environments (Brusco, 2008; Hopp and Van Oyen, 2004). Nevertheless, the topic remains sparsely studied in a field service setting. Field service is characterized by the fact that servers have to travel to the customer. Stochastic traveling times and the sequence dependency of these travel times render capacity management in a field service setting especially challenging. Given this complexity it is hardly surprising that most studies apply a simulation model to study cross training decisions (Agnihothri and Mishra, 2004; Agnihothri et al., 2003; Watson et al., 1998; Papadopoulos, 1996). Agnihothri and Mishra (2004) define a cross training index and show that in some field service settings fully cross training technicians can be optimal. The main reason for this divergent conclusion is that in a field service organization the risk of allocating the job to an unqualified field service engineer (FSE) can be significant. If the risk and cost implications for sending an unqualified FSE are substantial, full cross training can be cost effective. Analytical work in the area of field service capacity planning applies queuing theory to determine steady-state performance measures. Tang et al. (2008) derive a travel distance approximation taking into account the state dependent characteristics of these travel distances. The travel distance approximation is dependent on the staffing level which gives the possibility to link staffing levels with service performance measures such as the percentage of

customer serviced within the promised time window. The objective of our research is comparable to that of Chakravarthy and Agnihothri (2005). They consider two customer types and allow employees to be cross trained or dedicated. A dedicated employee will handle only one type of customers while the cross trained employees can serve both types. For this setting they analytically derive insights with respect to the conditions for which either flexible, dedicated or a combination of the employee types is suitable. Our research looks at similar decisions but differs on some key points. First of all, we include traveling in our analysis. Moreover, in our model, demand is determined endogenously based on machine characteristics which makes it possible to study the relationship between maintenance and capacity decisions. All of the above mentioned papers assume a demand that is exogenously determined and as a consequence the demand experienced is independent of the capacity decisions. However, demand and capacity are intertwined for maintenance services: the maintenance policy pursued by the service provider clearly has an impact on the workload. This interrelationship is most obvious when there is insufficient capacity to perform preventive maintenance. The resulting long waiting times for preventive maintenance will cause a burst of emergencies as machines will fail due to lack of timely care.

Next to the literature on capacity management, our work is related to a second research field, viz. the maintenance policy literature. In a maintenance policy the service provider stipulates how the machinery will be maintained. For an overview of the work in maintenance optimization for deteriorating systems we refer to Wang (2002). The maintenance policy has an immediate impact on the capacity decisions. Each preventive maintenance visit foreseen in the maintenance policy creates work for the technicians and the more preventive maintenance is being done the less repairs will be mandated. To design optimal maintenance policies the failure behavior of the machines needs to be modeled. The intensity function characterizes the evolution of the failure rate. Moreover, one has to detail and optimize the effect of preventive maintenance on the failure behavior of the machines, see for example Bartholomew-Biggs et al. (2009), Wu et al. (2010) or Yun et al. (2008). Although we will not explicitly consider the design of maintenance policies, we will model the failure behavior (reliability) of machines including the impact of maintenance on this failure behavior. By explicitly modeling the reliability of the machines covered by a maintenance contract we can evaluate different cross training strategies taking into account the interrelationship between service capacity and demand.

We will apply a competing risk model, similar as discussed by Doyen and Gaudoin (2006). The logic of the competing risk model builds on the fact that a failure can be caused by one of several possible failure modes which are called "risks". However, in reality only one of these potential failures will actually occur. To determine which failure mode will occur, each of the potential failure modes can be represented by a stochastic process. The failure mode that occurs first will become reality while the others are disregarded. For example, a machine failure can be caused by either a mechanic or a electric failure. Both failure modes can be represented by a stochastic process. How to determine the reason and timing of a failure? First, both stochastic processes will generate a time till the next failure: let us assume the mechanic failure process generates a time to failure of 1000 hours while the electric failure process indicates a time to failure of 700 hours. The competing risk model will select the failure type with the shortest time to failure: in our example an electric failure. Consequently, the machine will fail after 700 hours due to a electric failure. After the failure has occurred the logic is repeated with two new inter failure times generated by the stochastic processes. By defining a preventive maintenance visit as one of the risks, the effect of maintenance can be incorporated into the model: when the time till the next preventive maintenance is smaller than the time till the next failure, the failure is prevented from occurring.

Hence, our work is situated on the crossroad between capacity and maintenance planning. Our main focus is on the evaluation of cross training decisions when demand changes in response to these capacity decisions. The response in demand will depend on the the number of machines being maintained (workload), the maintenance policy, the reliability characteristics of the machines and the contract coverage.

3 Field service model

In this section we will present the simulation model that is built in order to evaluate different cross training policies. We captured the essence of a field service organization with a discrete time simulation model created in ARENA 11.0. To safeguard the applicability of our results we used real-life inputs concerning the installed base, the execution times and service capacity. These data were received from an existing service region of an international OEM in the compressed air and generator industry. In the following paragraphs the assumptions and performance measures of the model will be discussed.

3.1 Model assumptions and notation

To facilitate the further explanation of the model, we introduce the following notation

t_{id} travel home time of technician i on day d

a_{id} hours that technician i is still available to work on jobs on day d

t_j travel time towards job j

c_j time at the customer site for job j

$\bar{c}_l$ average time at customer site for jobs of type l ($l \in \{n, e\}$)

with e indicating emergency jobs and

n non emergency jobs

$\bar{r}_l$ average response time for jobs of type l ($l \in \{n, e\}$)

Response time is defined as the time between job arrival and arrival of the technician at the customer site.

b_d time remaining until the end of working day d

o_j agreed upon point in time to perform the non emergency job j (= ordered time)

s_j actual start time of execution of job j (= time of service)

S_l number of technicians skilled for job types l ($l \in \{e, n\}$)

D_{sc} yearly number of jobs originating from machines under contract

D_r yearly number of jobs on machines which are not covered by a contract

U_j indicates whether service demand is an emergency or not

$$U_j = \begin{cases} 1 & \text{in case of emergency, and} \\ 0 & \text{in case of non emergency} \end{cases}$$

X_j time between failures of a specific machine

Z_j time between scheduled maintenance on a machine

Y_j time between consecutive maintenance visits to the same machine ($\min\{X_j, Z_j\}$)

t_{pm} time since the last non emergency job on the machine

M total number of machines covered by a service contract

M_{up} number of machines under contract that are up

M_{down} number of machines under contract that are down

λ failure rate of a machine, i.e. the number of failures per time unit

λ_e rate at which emergency service is executed

λ_n rate at which non emergency service is executed

λ_{up} rate at which machines are restored to an operating state

λ_{down} rate at which machines are shut down, either because of failure or for maintenance

We assume that there are two types of field service engineers (FSE). The first type of technicians which will be indicated by "E FSE" are fully trained and as a consequence can handle both emergency and non emergency jobs. Next to E FSE, the OEM can decide to deploy N FSE which are less trained and as a consequence can only handle non emergency jobs which are less complex and more repetitive. Consequently, a non emergency job can either be executed by an E FSE or a N FSE but an emergency can only be resolved by an E FSE. We assume that there is no efficiency difference between the different types of technicians: the duration to finish a job is independent of the technician executing the job. We assume a cost difference of 33% between E and N FSE, i.e. the OEM can hire three N FSEs at the same expense as would be incurred for two E FSEs. During our simulations we will keep the budget constant and alter the configuration of the FSE within this budget.

In line with a product service strategy the OEM under consideration offers comprehensive service contracts in which the OEM takes care of all maintenance activities, i.e. both emergency and non emergency service. Although a considerable part of the

installed base is covered by such contracts, the organization also has to deal with incoming calls for service on machines that are not covered by a service contract. The demand from the first category of machines (D_{sc}) is more predictable and can be influenced by the maintenance policy of the OEM: the OEM can decide on the optimal preventive maintenance interval. The demand from the machines that are not covered by a contract, service-on-request (D_r), is less predictable and the OEM has no (less) influence on the maintenance policy. In the service region under investigation, approximately 2000 jobs are finished each year of which around 50% are jobs on machines under contract ($D_{sc} \approx 1000$ interventions). Both D_{sc} and D_r can be a request for emergency or non emergency service. The key difference is that D_{sc} is influenced by the maintenance policy and the system state, while D_r is independent of the OEM's maintenance policy and the system state. As a consequence, we let jobs of D_r arrive according to an independent Poisson process and once executed the jobs are discarded. Machines with a service contract are explicitly modeled as entities in a closed-loop queuing system comparable to the approach in Papadopoulos (1996). Non emergency jobs on a machine under contract arrive with a fixed time interval dependent on the contract terms and the yearly operating hours of the machine. If for example the maintenance policy would prescribe that preventive maintenance should be executed after 3000 operating hours, a machine with 2000 yearly operating hours would generate a demand for non emergency service each one and a half year. Contrary to non emergency jobs, emergency jobs on a machine under contract arrive based on a failure process which characterizes the reliability of the machine (see further).

The FSEs all work eight hours a day. Thanks to the use of internet enabled technology, the FSEs can start their working day by going directly to a customer and their working day ends when they arrive back home. As a consequence, the number of hours that a FSE i is available to perform jobs during the remainder of day d (a_{id}) depends on the time he needs to drive back home.

$$a_{id} = b_d - t_{id} \quad (1)$$

To capture the fact that a technician needs to be able to drive back home before the end of the working day, a travel time observation (t_{id}) is subtracted from the remaining working hours (b_d) to obtain the hours that the FSE is available to work on jobs (a_{id}). Traveling is a significant part of the working time of a FSE and the travel time depends on the effectiveness of the planning and the amount of work in the queue. We model these travel times by taking each travel time as an observation from the empirical distribution of the real-life travel times taking over a period of one year. We use the empirical distribution because the data is unsuitable to fit to any parametric distribution (Kolmogorov-Smirnov test P-value $< 0,1$). The travel home times are taking each day and for each FSE separately which introduces variability in the available service capacity. Although we do not make the travel times dependent on the sequence of maintenance jobs, we believe that by using real-life data the travel times in our model should be representative for a situation where planners try, with a given level of success, to minimize total travel time.

It became clear from our data set that there is a difference between emergency and non emergency jobs with respect to the time that the technician spends at the customer site, i.e. the time that the FSE is at the customer site to perform the maintenance. Both durations can be characterized by a parametric distribution, respectively the exponential and beta distribution (Kolmogorov-Smirnov test P-value $> 0,15$). We assume that the time at the site is only used to work on the machine (this is important when calculating the machine availability).

Many theoretical papers neglect the discrete nature of working days and jobs. As a result the field service systems are unrealistically represented. It is common to assume that jobs are allocated based on a FCFS rule. Although this rule is the preferred way of working for our service organization, the discrete character of the working days combined with the traveling requirements make it impractical to use the FCFS rule in a strict sense. Instead, when a FSE needs a job, the planner will first look at the emergency jobs in queue and pick the first job (FCFS) for which the following constraint is valid

$$a_{id} - t_j > 0 \quad (2)$$

Constraint (2) states that a FSE is sent to a job j only if (s)he has enough time to travel to the customer site and start working on the machine. This allocation rule satisfies for emergency jobs. For non emergency jobs, however, the planning is more subtle. The reason is that the customer has to pay for the travel time of the FSE, so the OEM wants to avoid that the customer has to pay for long traveling times, certainly when the FSE does not have enough time to complete the job in one visit. Due to these concerns, a FSE will be sent to a non emergency job only if one of the following conditions hold

$$a_{id} - t_j \geq c_j \quad (3)$$

or

$$a_{id} - t_j > Threshold * (t_j + t_{id}) \quad (4)$$

A FSE is allocated to a job j if either he can finish the job during the same day (3) or if the remaining time when the FSE arrives at the customer site is longer than a threshold percentage of the total travel time (4). In our simulation we took a threshold value of 50%, i.e. the total travel time can run up to a maximum of twice the time at the customer site.

When a job is not finished at the end of the working day, the job will get priority over jobs of the same type (either emergencies or non emergencies) that are not started yet. An additional requirement for the started but unfinished jobs is that the job has to be allocated to the same FSE who started the job.

In a previous paper we identified the need to develop more scheduled maintenance operations in order to be successful in a product service strategy (Colen and Lambrecht, 2010). One key advantage of non emergency jobs is the leeway in timely execution of these jobs. Preventive maintenance jobs are accepted to be on time if they are executed during an interval of 10% around the optimal preventive maintenance timing. The management of the OEM confirmed the applicability of this general rule in their maintenance operations. The flexibility in execution timing allows for non emergency jobs to be executed when there is idle capacity or to be postponed when work is piling up. In effect allowing for a more efficient use of capacity. Based on this reasoning and in order to capture the advantages of performing more preventive maintenance we allow that non emergency jobs can be started as soon as 10% of the maintenance interval before the timing foreseen in the maintenance contract (o_j). However, when the job is postponed for longer than 10% of the maintenance interval, the job is considered to have become an emergency. We assume that when a failure occurs during the 10% interval around o_j both the repair and the preventive maintenance are performed during the same visit (opportunistic maintenance).

Between the different types of jobs we use a non-preemptive allocation priority for emergency jobs. Before any non emergency job is started there will be a check whether there are still emergency jobs that can be allocated. So no E FSE will start a non emergency job if he can work on an emergency order. We assume that the service organization handles D_{sc} and D_r equally: there is no priority rule depending on whether the machine is covered by a service contract or not.

3.2 Reliability of machines

Next to the planning logic that makes it possible to work in discrete time, we need to model the reliability of the machines which are covered by a service contract. To model the maintenance demand originating from these machines, we take the sum of the maintenance interventions on the machines m that are covered by a contract. By dividing this sum by the number of years for which the demand data is gathered, we obtain the yearly number of interventions originating from machines under contract (D_{sc})

$$D_{sc} = \frac{\sum_{m \in R} D_m}{k} \quad (5)$$

with R the set of machines under contract

D_m the demand from machine m during k years

We apply a competing risk framework to model the demand originating from a machine m . Demand for maintenance on a machine is a sequence of emergency and non emergency service interventions in which both types of service can be considered to be a stochastic process. The resulting service demand can then be described by a stochastic process characterized by the time between maintenance Y_j and the indicator U_j for the type of maintenance job.

$$\text{stochastic process} \{Y_j, U_j\} \quad (6)$$

with Y_j equal to the time between two maintenance interventions j and $j+1$ on machine m

U_j indicates the type of maintenance intervention

$$\text{and } U_j = \begin{cases} 1 & \text{in case of emergency, and} \\ 0 & \text{in case of non emergency} \end{cases}$$

$$Y_j = \min\{X_j, Z_j\}$$

with X_j equal to the time till the next failure after maintenance intervention j

Z_j is the time till the next non emergency service after maintenance intervention j

The demand D_m equals the number of maintenance interventions generated by the process (6).

In order to model X_j we have to quantify the impact of maintenance on the failure rate (λ). In the standard competing risk approach it is assumed that each maintenance is perfect, i.e. after receiving maintenance the machine is as good as new (AGAN). In reality this is not the case, therefore we assume that after emergency maintenance the failure rate of the machine is the same as just before the failure which is a more plausible assumption. For non emergency maintenance we keep the assumption of AGAN maintenance. As a consequence the failure intensity is determined by the time that has past since the last preventive maintenance job on the machine (t_{pm}). We assume that the failure intensity function can be modeled by the widely used Power Law (PL) process (Doyen and Gaudoin, 2004).

$$\begin{aligned}\lambda(t_{pm}) &= \frac{\beta}{\alpha} \left(\frac{t_{pm}}{\alpha} \right)^{\beta-1} \\ &= \text{PL}(\alpha, \beta) \\ &\quad \text{with } \alpha = \text{scale parameter} \\ &\quad \beta = \text{shape parameter}\end{aligned}\tag{7}$$

When β is larger than 1, the failure rate increases with t_{pm} . For a Power Law intensity function the time to failure can be represented as a truncated Weibull distribution.

$$(X_j + t_{pm}) \sim \text{Weibull}(x|x > t_{pm})\tag{8}$$

By taking an observation from (8) and deducting t_{pm} time units, we obtain the time to failure after maintenance intervention j (X_j) for the machine which had its last preventive maintenance t_{pm} time units ago. By doing so we can simulate the time to failure taking into account the increasing failure rate. The time to failure obtained in this way is in line with the Power Law process.

By incorporating the failure behavior of the machines under contract into the simulation model and making this failure behavior dependent on the maintenance policy (role of t_{pm} in (8)), we have linked the demand for service (emergency and non emergency) with the maintenance policy followed by the service provider. Moreover, the demand is dependent on the state of the service system. Indeed, when there is too much work to be done for the FSEs, preventive maintenance is stalled which will lead to more machines failures due to the increasing failure rate as a function of the time since the last preventive maintenance (t_{pm}).

3.3 Performance measures

Most researchers use response time as the main measure to evaluate the performance of the service organization. In the context of field service, response time is defined as the time between order arrival and the arrival of the technician at the customer site. However, customers value response time differently in case of emergency and non emergency jobs. As we consider a situation where both emergency and non emergency service coexist and interact with each other, we opted for a penalty minimization approach of the following form

$$P = p_e \bar{r}_e + p_n \bar{r}_n + p_l L\tag{9}$$

The penalty coefficients p_e , p_n and p_l allow for a weighting between the different performance measures: average response time for emergencies ($\bar{r}_e$), average response time for non emergencies ($\bar{r}_n$) and the fraction of non emergencies that have become emergencies (L). The first term ($p_e \bar{r}_e$) indicates the valuation (penalty) of the response time for emergencies which has typically a larger penalty per time unit than the penalization for the response time for non emergency service ($p_n \bar{r}_n$). L is the fraction of non emergencies that have become emergencies due to postponement. Postponing non emergency jobs can free up technicians to execute the more pressing emergency jobs. However, there is a risk that the postponed non emergency job will transform into an emergency. Non emergencies can become emergencies due to two reasons. First of all, the machine can fail during the period between the timing that the job should have been executed (o_j) and the actual execution of the job. Moreover, as mentioned above each non emergency job for which waiting time exceeds the 10% threshold will transform into an emergency. Hence, L is the fraction of non emergency jobs that have become emergencies either by excessive waiting time or because of a failure that could have been avoided by timely maintenance. In (9) the last term does not only incorporate effectiveness of the organization but also conformance. As is already pointed out by Apte et al. (2007) for non emergency

service conformance with the agreed upon timing (o_j) is a principal concern. We comply with this remark by incorporating the fraction of non emergencies that became emergencies (L) in our penalty function. The results reported in this paper were obtained with the following penalty coefficients: 1, 20, 3 for respectively p_n , p_l and p_e . So average emergency response time is three times more penalized than average non emergency response time. While, a five percent increase in the number of non emergencies that become emergencies is equally evaluated to an one hour longer average response time for non emergencies or equivalently a 20 minutes increase in average emergency response time.

The average response time for non emergency service is calculated as follows

$$\bar{r}_n = \frac{\sum_{j \in N} \max(s_j - o_j, 0)}{N_n} \quad (10)$$

As is clear from (10), the response time for non emergency service is calculated as the total time difference between the arrival of the FSE at the site and the agreed-upon timing for the service, divided by the total number of non emergency jobs. Note that if the service is delivered before the agreed upon time ($s_j < o_j$), the numerator is not increased but the denominator is. By doing so we assure that starting jobs early is rewarded (the OEM adheres to his commitments), nevertheless there is no difference between a just in time execution and an execution ahead of time. For emergency service the response time is calculated in a straightforward way because for emergencies the rule of "the sooner the better" remains.

Due to the increased focus on services, many OEMs now offer comprehensive service contracts. These contracts often emphasize availability through up-time guarantees or a payment scheme based on the usage or availability of the machine. As a consequence availability is becoming a key performance indicator for all customer oriented service organizations. We will first show how the average availability can be derived and afterwards we highlight the differences between the penalty and availability performance measures. Availability can be defined as

$$A = \frac{\text{MTBM}}{\text{MTTM} + \text{MTBM}} \quad (11)$$

Figure 1: Representation of the maintenance system of a machine park

The mean time between maintenance (MTBM) is the time between maintenance interventions which can be either emergency or non emergency interventions. The mean time to maintain (MTTM) is the time that the machine is shut down for maintenance. By applying Little's Law on both lead time measures, MTBM and MTTM, in the system as shown in figure 1 we obtain

$$\text{MTBM} = \frac{M_{up}}{\lambda_{up}} \quad (12)$$

$$\text{with } \lambda_{up} = \lambda_e + \lambda_n$$

$$\text{MTTM} = \frac{M_{down}}{\lambda_{down}}$$

If the system is in steady state

$$\lambda_{up} = \lambda_{down} = \lambda_e + \lambda_n \quad (13)$$

$$\begin{aligned}
 A &= \frac{M_{up}}{M_{down} + M_{up}} \\
 &= \frac{M - M_{down}}{M} \\
 &= \frac{M - ((\bar{c}_e + \bar{r}_e) * \lambda_e + \lambda_n \bar{c}_n)}{M}
 \end{aligned}$$

A denotes the average availability of the machines under contract. Based on the above derivation it is clear that with respect to the availability of machines the response time is only important in case of an emergency. While for non emergencies only the time that the FSE really works on the machine is important, not the travel nor the waiting time. Moreover, contrary to the measure of response time, the availability also takes into account the frequency of maintenance (λ_{down}). Therefore, we state that the availability measure gives a more overall and objective evaluation of the service performance than looking solely at the response time. Nevertheless, a company can have good reasons to stress the importance of the response time or more generally the penalty function (9). There can be a major difference both economically and emotionally between an uncontrolled failure and a foreseen shut-down for maintenance. Moreover, the contractual agreement can be such that response time is more relevant for the profitability of the contract than availability. We conclude that availability is an objective overall performance measure but that the penalty function is more suitable to capture the service quality as experienced by the customers. As both measures have their merits, we opted to use both of them and highlight possible differences in capacity decisions when using one of the two measures.

4 The simulation experiment

We want to evaluate the possibility to introduce technicians dedicated to non emergencies (N FSE) in a field service setting and determine which factors (workload, maintenance policy, machine reliability and/or contract coverage) play a role in the attractiveness of this option. The contract coverage indicates the proportion of maintained machines that are covered by a service contract. We determined the optimal cross training policy for several scenarios, which differ according to the above mentioned characteristics. A scenario is defined as a specific combination of a number of machines to maintain, a preventive maintenance interval, a parameter selection for the Power Law intensity function and a certain level of contract coverage. By varying the configuration of FSE within such a scenario we are able to evaluate the cross training policy based on both the penalty function and the availability of the machines under a service contract.

Before discussing the results of our simulation, we will first elaborate on the different scenarios we use. First, each scenario is characterized by a preventive maintenance frequency. This frequency equals 2000 or 3000 operating hours, i.e. preventive maintenance should be performed each time a machine under service contract has run for 2000 or 3000 hours. In order to have a maintenance regime apt for the different levels of utilization of the machines to maintain, the service organization registers the yearly operating hours of each machine to determine the timing of the maintenance based on the actual usage. Also in our simulation, we use the data about the operating hours of the different machines (installed base data) to determine the timing of maintenance for each machine individually. Secondly, the failure behavior of the machines is characterized by one of four potential parameter sets for the Power Law (PL) intensity function (see figure 2). We consider the Power Law intensity function with a scale parameter α of 3175 and a shape parameter β of 5, i.e. PL(3175,5), as the reference intensity function. PL(3157,10) characterizes a failure behavior that is more sensitive (steep) to the operating hours. The intensity functions of PL(2500,3) and PL(3000,2) are less steep but reliability deterioration begins sooner. Highly reliable machines have a failure behavior in line with PL(4000,10) (although deterioration is fast after 3400 hours).

Besides information about the maintenance policy and the failure behavior, a scenario also contains information about the workload measured as the total number of machines to maintain. Chakravarthy and Agnihothri (2005) already stipulated that the workload (utilization rate of technicians) is a key determinant for the suitability of different cross training policies. We evaluate two workload levels. The first workload we apply, is created by 538 machines under contract and a yearly D_r of 802 orders originating from 430 machines without a service contract. This level of service demand will be referred to as "Standard" workload. The second demand level creates a higher workload and constitutes 700 machines under contract and a D_r of 1043 orders per year or equivalently 560 machines. This workload level will be indicated by "High". Although the workload increases, the contract coverage of the installed base being maintained remains unaltered at a level of 56%. Lastly, we are interested in the effect of increased contract selling, i.e. the contract coverage increases. For this goal we will analyze scenarios in which all of the 430 machines that generate D_r ("Standard" workload) are put under contract.

Figure 2: Power Law intensity functions with different parameters

This adds up to thirty scenarios for which we evaluate the different possible workforce configurations in order to select the optimal cross training policy based on either the availability or the penalty function. All simulation results are average values based on 25 simulation runs. In order to assure that the response time measures have a 95% confidence interval smaller than 10% of the average value of the response time measures, we simulated periods of 15 years with a two year warm-up period by doing so we are confident that the system is in steady state.

5 Simulation results

In this section we discuss the results of our simulations and determine which factors play a role in determining the optimal policy in a field service setting with comprehensive service contracts. In section 5.1 we first explain the dynamics at play when substituting E FSE by N FSE. Afterwards, in section 5.2, we investigate the factors influencing the optimal cross training policy including the workload, the maintenance policy, machine reliability and the contract coverage. In section 6, we apply reliability theory to characterize those types of machines that are more suitable to be maintained by a certain workforce configuration. We conclude in section 7.

5.1 Direct and indirect effect of increasing the fraction of N FSE

A workforce configuration (cross training policy) can be characterized by the ratio

$$R_n = \frac{S_n}{S_e + S_n} \quad (14)$$

with S_n the number of N FSE

S_e the number of E FSE

When the number of technicians dedicated to non emergencies (S_n) increases the ratio goes up. We opted to maximize the service performance while keeping the budget for technicians constant. As a consequence, we start with a situation of 10 fully cross trained technicians (budget fully used) and progressively switch fully cross trained technicians for dedicated ones. Thanks to the fact that the cost of a N FSE amount only to 2/3 of the cost of an E FSE, two E FSEs can be replaced by three

N FSE without exceeding the budget. So after switching one E FSE for one N FSE ($R_n = 0, 1$) we can replace two E FSE by three N FSE ($R_n = 0, 27$), ... (we only consider those cross training policies that use the full budget). Among these workforce configurations we then select the one with the highest service performance, measured by either the average availability or the penalty function.

Increasing the number of FSEs that are dedicated to non emergencies (N FSE) has some advantages compared to the situation with only fully trained FSEs (E FSE). Having more N FSEs on the payroll can improve the timely execution of preventive maintenance jobs. This is a consequence of the fact that E FSEs give priority to emergencies. Therefore, preventive maintenance jobs are the first to be postponed, if there is a capacity shortage. This postponement will endanger the timely execution of preventive maintenance. Without timely execution of preventive maintenance, the total number of emergencies increases due to machine breakdowns. This, in turn, will increase the need to postpone jobs, once again endangering timely preventive maintenance and further increasing the number of emergencies ... The use of N FSEs can avoid this "emergency trap" of an escalating number of emergencies due to postponement of preventive maintenance. Replacing E FSEs by N FSEs can have a positive direct and indirect effect. Obviously, the introduction of technicians dedicated to non emergencies will reduce the response time for non emergencies (direct effect), but the lower capacity for emergencies will increase the response time for emergencies. On the other hand, the improvement in timely preventive maintenance will reduce the total number of emergencies which will reduce the response time for the remaining emergencies (indirect effect). So although the response time for non emergencies is less important than the response time for emergencies, the potential indirect effect of timely preventive maintenance on the number of emergencies may tip the balance in favor of employing more dedicated technicians (N FSEs).

Figure 3 shows the average response times under the scenario of a maintenance policy with preventive maintenance after each 3000 operating hours, a intensity function of PL(3175,5) and a high workload. Figure 4 shows the resulting availability and penalty for the same scenario.

Figure 3: Response times evolution when increasing the fraction of N FSE – high workload

The first N FSE that are added to the workforce, strongly reduce the response time of both non emergencies and emergencies. Having more capacity to handle non emergencies, it is hardly surprising that the non emergency response time drops (direct effect). Although the capacity for emergencies is lowered, the emergency response time also decreases. Clearly, the indirect effect is an important element in this scenario. As the N FSEs focus on non emergency service a lot of emergencies will be avoided (indirect effect), offsetting the increase in emergency response time due to the reduction of E FSEs. But as the proportion of N FSE increases, the positive effect on the number of emergencies decreases and at a ratio of 0,27 the emergency response time starts rising again. The response time for non emergencies continues to decrease when N FSE are added but this positive effect is canceled out by the increase in emergency response time. The increasing response time for emergencies deteriorates the availability of the machines and increases the penalty function. From figure 4 it can be seen that a FSE configuration with a R_n of 0,27 is optimal with respect to both the machine availability and the penalty function. With this configuration an average availability of 99,29% is obtained and the penalty function reaches a value of 58,88.

Compared to the scenario with a high workload (figure 4), the scenario with a standard workload, PL(3175,5) and maintenance

Figure 4: Service performance when increasing the fraction of N FSE – high workload

after 3000 hours shows a different evolution in response times (figure 5). Although the non emergency response time goes down, the response time for emergencies immediately starts to increase when R_n increases. At R_n equal to 0,27 the decrease in emergency response time is still strong enough to compensate for the longer emergency response times but at a fraction of 50% N FSE both the availability and service quality deteriorate (figure (6)). So, although it remains optimal to deploy a certain fraction of N FSE, the driving force for deploying N FSE in this standard workload scenario is the direct effect instead of the indirect effect.

Figure 5: Response times evolution when increasing the fraction of N FSE – standard workload

In order to understand why N FSE are more attractive in certain scenarios than in other ones, we will estimate the direct (*DE*) and indirect effect (*IE*) of deploying N FSE. When we consider the direct effect, the attractiveness of N FSE depends on the existence of sufficient preventive maintenance work for the N FSE. If there is a high proportion of preventive work compared to emergency work, adding N FSE has a positive direct effect on the service performance by lowering the response time for non emergencies without effecting the emergency response time. Given this rationale, we suggest to estimate the strength of the direct effect by the yearly number of executed non emergencies divided by the total number of executed jobs during a year.

Figure 6: Service performance when increasing the fraction of N FSE – standard workload

$$DE = \frac{N_n}{N_e + N_n} \quad (15)$$

with N_n the yearly number of executed non emergency jobs

N_e the yearly number of executed emergency jobs

Although we derive (15) from our simulation results, we expect N_n and N_e to be readily available in most maintenance organizations.

The indirect effect of adding N FSE exists of reducing the total number of emergencies by performing timely preventive maintenance. When a large number of non emergency interventions become emergencies due to postponement of the intervention, performing preventive maintenance in a more timely fashion will be very beneficial. Given this insight, the percentage of non emergency demand that becomes emergencies due to postponement, is a suitable measure to gauge the indirect effect of adding N FSE to the workforce as these FSE will enable a timely execution on non emergency jobs.

$$IE = \frac{D_{n_e}}{D_n} \quad (16)$$

with D_{n_e} the yearly number of demanded non emergency jobs that have become emergencies

D_n the yearly number of demanded non emergency jobs

If demanded non emergency jobs are not executed in a timely fashion, they become emergencies. Nevertheless these jobs remain accounted for as non emergency demand, i.e.

$$D_{n_e} \subseteq D_n \quad (17)$$

When D_{n_e} equals D_n , all non emergency demand is postponed until the jobs become emergencies, so not one non emergency job is started. Consequently, in this case the IE will be 100% while the DE will be zero. By using the measurements (15) and (16) we can determine which effect, direct or indirect, plays an important role in the potential benefit of deploying N FSE. Figure (7) gives an overview of the estimates of the direct and indirect effect for the different scenarios with a workforce consisting of 10 E FSE.

Maintenance policy	Workload	Effect type	PL(4000,10)	PL(3175,10)	PL(3175,5)	PL(2500,3)	PL(3000,2)
2000	Standard	DE	65,25%	64,98%	60,74%	45,50%	47,51%
		IE	17,56%	17,44%	18,77%	25,84%	25,16%
	High	DE	17,16%	17,83%	0,08%	0%	0%
		IE	78,62%	77,69%	99,84%	99,40%	99,26%
3000	Standard	DE	60,04%	45,18%	42,64%	0,28%	0,3731%
		IE	16,98%	19,86%	22,51%	0,3544%	27,93%
	High	DE	53,30%	34,57%	23,33%	0%	0,0001%
		IE	26,39%	35,27%	57,97%	91,89%	94,06%

Figure 7: Measures of direct effect and indirect effect with 10 E FSE

If we would take for example the scenario characterized by a PM interval of 2000, a standard workload and machines with reliability PL(3175,5), we obtain values of 60,74% and 18,77% for respectively the *DE* and *IE* (see figure 7). From these results, we can derive that a relative large part (60,74%) of the work being done are non emergencies while still a significant amount of non emergency demand (18,77%) is not timely executed. Consequently, there is a considerable potential to deploy N FSE both with respect the the *DE* and *IE*. The rationale is similar for the other scenarios.

5.2 Factors influencing the cross training policy

In this subsection we will use the optimal cross training policies as derived from our simulation results to identify those factors that influence the cross training policies. Figure 8 gives an overview of the optimal cross training policies, characterized by R_n , for the different scenarios based on a maximization of the availability of the machines. Contrary to figure (8), figure (9) shows the optimal R_n values when maximizing service quality as measured by the penalty function.

Maintenance policy	Workload	PL(4000,10)	PL(3175,10)	PL(3175,5)	PL(2500,3)	PL(3000,2)
2000	Standard	0,5	0,5	0,27	0	0
	High	0,5	0,27	0,27	0,27	0,5
3000	Standard	0,27	0	0	0	0
	High	0	0,27	0,27	0,27	0,27

Figure 8: Optimal R_n with respect to availability

Maintenance policy	Workload	PL(4000,10)	PL(3175,10)	PL(3175,5)	PL(2500,3)	PL(3000,2)
2000	Standard	0,27	0,27	0,27	0,27	0,27
	High	0,27	0,27	0,27	0,27	0,27
3000	Standard	0,27	0,27	0	0	0
	High	0,27	0,27	0,27	0,27	0,27

Figure 9: Optimal R_n with respect to service quality

When looking at figures (8) and (9) it is striking that full cross training is optimal in a considerable amount of scenarios. Furthermore, for those scenarios in which deploying N FSE is advisable, the optimal fractions of N FSE are relatively low. This is surprising as the general consensus in capacity management seems to be that only small amount of cross training are sufficient to capture most of the benefits of flexibility, e.g. Cachon and Terwiesch (2006). The relative low optimal R_n values can be explained by the fact that a considerable number of emergencies are unavoidable and that E FSE have the advantage that allocating fully cross trained FSE to jobs is easier than allocating dedicated FSE. Easier job allocation is especially beneficial

in field services which are characterized by traveling times. From comparison between the optimal cross training policies based on the availability and those based on the service quality, it is clear that the service quality (penalty function) enables the use of E FSE in more scenarios than the availability measure. However, the optimal fraction of N FSE can reach higher values when maximizing availability compared to maximizing service quality. From figure 8, it is clear that all the factors under consideration, i.e. the workload, the maintenance policy and the reliability play a role when determining the optimal cross training policy.

In the following subsections we will use the simulation results as represented in figures (7), (8) and (9) in order to discuss the influence of the different factors on the optimal cross training policy. First, we will examine the role of the workload, maintenance policy and the reliability of the machines. Afterwards in section 5.2.2, we evaluate the consequences of more intensive contract selling by increasing the contract coverage of the machine park.

5.2.1 Impact of workload, maintenance policy and reliability

When we focus on the difference in optimal R_n between scenarios with standard and high workload (see figure (8)), we can conclude that in most cases a high workload favor the use of dedicated technicians more strongly than a low workload. Based on figure (7) we observe that if the workload is "Standard" the direct effect prevails. If the workload is "high", however, the indirect effect becomes more important. In most situations this increased IE tips the balance in favor of deploying more N FSE.

This is conform with the observation we made in section 5.1. With a high workload E FSEs are overwhelmed by work so that they will be pre-occupied by emergencies and neglect non emergency jobs. With the spotlight on emergencies, non emergency jobs will get postponed which further increases the workload as more and more machines fail (high IE). This strong IE that occurs when the workload is high explains the increase in the optimal fraction of N FSE for scenarios with a high workload (figures (8) and 9). However, if the workload is "Standard" the indirect effect is small and adding N FSE to the workforce is only optimal when the direct effect is sufficiently large, e.g. scenarios with PL(4000,10).

We have established that an increased workload shifts the emphasis from the direct effect towards the indirect effect. However, it are the maintenance frequency and the machine reliability characteristics that determine to which extent this shift will occur and how large the effects are. By comparing the scenarios with more reliable machines (PL(4000,10) and PL(3175,10)) with those for which the machines are less reliable (PL(2500,3) and PL(3000,2)), we observe that servicing a machine park that is more reliable collides with a high DE and a low IE (figure (7)). From the optimal ratios R_n (figures (8) and 9), we can state that maintaining more reliable machines allows for the use of more N FSE.

The use of N FSE seems more appropriate with a small maintenance interval compared to a larger interval, this observation reveals a trade-off between the possibility to economize on the number of maintenance visits or reduce the cost of the technician executing the maintenance, viz. a N FSE instead of an E FSE. We will elaborate on the influence of the maintenance frequency and the reliability characteristics in section 6 by performing an analytical quantification.

5.2.2 Impact of increased contract coverage

Companies that adhere to a product service strategy, are stressing the importance of building long lasting relationships with their customers. This has led to a surge in the use of comprehensive service contracts. But do these contracts offer the opportunity to deploy more dedicated FSE? And are they beneficial for the customer?

Instead of having 538 machines under contract and 430 machines without a contract, we assume now that all these machines are covered by a service contract. Figure (10) shows the optimal cross training policies based on the average availability of the machines under contract. It is clear that an increased contract coverage enhances the ability to work with N FSE. This increase in R_n is solely attributed to the direct effect, as a major shift occurs from emergency interventions towards more non emergency interventions. Indeed, this observation reveals the potential to streamline service operations by increasing the contract coverage. The percentage of non emergency tasks that become emergencies due to postponement slumps in all scenarios, once again showing a benefit of increased contract coverage. However, it also means that the indirect effect of deploying E FSE decreases, lowering the incentive to work with N FSE. From our simulation results we can conclude that in general the positive influence of an increased contract coverage on the direct effect of N FSE prevails, making it possible to use more N FSE. So from the perspective of the service provider, increasing the contract coverage is a worthwhile strategy as it shift the workload towards more non emergency (scheduled) work, offering the opportunity to streamline operations.

From the customer's point of view the situation is less obvious. If we evaluate the service quality as measured by the penalty

Maintenance policy	Contract coverage	PL(4000,10)	PL(3175,10)	PL(3175,5)	PL(2500,3)	PL(3000,2)
2000	Normal	0,50	0,50	0,27	0	0
	Full	0,50	0,50	0,27	0,27	0,27
3000	Normal	0,27	0	0	0	0
	Full	0,50	0	0	0,27	0

Figure 10: The optimal R_n with respect to availability for varying contract coverage

function, increasing the contract coverage improves service quality except for a maintenance policy with preventive maintenance after 3000 operating hours and machines with failure behavior characterized by PL(2500,3). If we calculate the availability of machines not under contract based on the interventions executed by the OEM and compare this availability with the availability of the machines when they are put under contract, the average availability improves. Consequently, from an operational point of view customers indeed have an incentive to buy a service contract, viz. higher machine availability. However, if the machine reliability does not pertain a minimum reliability level, which is the case for PL(3000,2) and PL(2500,3), the machine availability of the machines already under contract actually drops when putting more machines under contract. This is caused by the fact that these machines have a high failure rate even before the scheduled execution of preventive maintenance. Due to the high number of "unavoidable" failures, the number of interventions to be executed actually increases, causing increased response times for both emergency and non emergency service and reducing average availability of the machines under contract.

From the perspective of the customer, we conclude that comprehensive service contracts can add value by reducing the number of emergencies which will also increase the availability. However, in order to safeguard this value proposition the service provider should make sure that his maintenance operations are apt to the task. This means the maintenance policy should not only reduce the number of emergencies on the machines put under contract but do so sufficiently in order to lower the total number of interventions required to maintain the machine park. Alternatively, the service provider has to increase his field capacity to avoid that the increasing sales of service contracts erodes the value proposition of the contracts. Therefore, it is important for companies that pursue a product service strategy, to proactively optimize their maintenance policy and capacity decisions to cope with a higher contract coverage.

6 Analytical derivation of the impact of maintenance frequency and machine reliability on the workforce configuration

In this section we will perform an analytical analysis on the role of the maintenance frequency and the machine reliability characteristics on the optimal workforce configuration. For this purpose, we will use reliability concepts applicable to failure processes of repairable systems. This section is meant to theoretically justify and further explain the simulation results from the simulation with respect to the impact of the maintenance frequency and the machine reliability.

On the one hand, the benefit from deploying N FSE depends on the amount of non emergency work compared to the total amount of work. On the other hand, the benefit also depends on the number of non emergency jobs becoming emergencies. In this analytical section we make the assumption that response times and maintenance execution times are zero, as a consequence the demand for service generated by a machine is only depending on the maintenance frequency and the reliability characteristics of the machines. The assumptions of "as bad as old" repairs and "as good as new" preventive maintenance remain valid.

For a Power Law intensity function (equation (7)), the expected number of failures of a machine before a certain point t in time can be expressed as

$$M(t) = \int_0^t \lambda(t) dt = \left(\frac{t}{\alpha}\right)^\beta \quad (18)$$

By taking t equal to the maintenance interval (T) expressed in operating hours, we obtain the expected number of failures of a machine between two preventive maintenance (non emergency) jobs. If we define a maintenance cycle as the period from the

start of a preventive maintenance job on a machine until the start of the next preventive maintenance job on the same machine, we can write the fraction of non emergency interventions compared to the total interventions on a machine as

$$\frac{1}{M(T) + 1} \quad (19)$$

The numerator is one, as a maintenance cycle contains one non emergency job, while the total number of jobs (denominator) equals one non emergency and $M(T)$ emergencies. It is clear that a high fraction (19) or otherwise stated a high proportion of non emergency work compared to emergency work creates a potential to maintain the machines by deploying N FSE. Accordingly, (19) is related to the DE as defined in (15). If we take for example reliable machines with intensity function $PL(4000,10)$, we can calculate the expected number of failures during one maintenance cycle of e.g. 3000 operating hours by applying (18).

$$M(3000) = \left(\frac{3000}{4000} \right)^{10} = 0,0563 \quad (20)$$

Using (19) we obtain the fraction of jobs that are non emergencies

$$\frac{1}{M(3000) + 1} = 0,9467 \quad (21)$$

As could be expected for reliable machines ($PL(4000,10)$) the fraction of jobs that are non emergencies is very high. However, if we calculate (19) for less reliable machines, e.g. machines with reliability characterized by $PL(2500,3)$, the fraction of jobs that are non emergencies drops to 0,3666. These analytical results support the simulation results (figure 7) as they explain why the DE for more reliable machines is higher than for less reliable machines. Equation (20) can be graphically represented as the area determined by the intensity function, the horizontal axis and the maintenance frequency (see figure 11). The larger this area the more emergencies can be expected during one maintenance cycle.

Figure 11: The expected number of failures during one maintenance cycle

Next to a high amount of non emergency work compared to emergency work, deploying N FSE can also become optimal when a lot of non emergencies become emergencies due to postponement. The number of machines that fail when preventive maintenance is postponed, also depends on the maintenance frequency and the reliability characteristics of the machines. The probability that a machine with reliability characteristics determined by a Power Low process (7) will fail during a time period $[t_1, t_2]$ can be calculated as

$$\begin{aligned} F(t_2) - F(t_1) &= \int_{t_1}^{t_2} f(t) dt \\ &= \int_{t_1}^{t_2} \frac{\beta}{\alpha^\beta} t^{\beta-1} e^{-\left(\frac{t}{\alpha}\right)^\beta} dt \end{aligned} \quad (22)$$

$$= e^{-\left(\frac{t_1}{\alpha}\right)^\beta} - e^{-\left(\frac{t_2}{\alpha}\right)^\beta}$$

$F(t)$ represents the probability of a machine failure before t and $f(t)$ is the failure density function which determines the probability of a failure at time point t . By taking a time interval, expressed in operating hours, of 10% around the maintenance frequency (T), (22) expresses the probability that a machine failure can be avoided by executing the preventive maintenance as soon as possible.

$$F(T + 0,1 * T) - F(T - 0,1 * T) = e^{-\left(\frac{T-0,1*T}{\alpha}\right)^\beta} - e^{-\left(\frac{T+0,1*T}{\alpha}\right)^\beta} \quad (23)$$

Graphically (23) can be represented as the area bounded by the failure density function ($f(t)$) and the two points (t_1 and t_2) in time. Figure (12) graphically shows the area calculated by (23) for machines characterized by PL(4000,10) and PL(2500,3). The probability of failure between 2700 and 3300 operating hours equals 11,65% and 18,35% for respectively machines characterized by PL(4000,10) and PL(2500,3). Machines with reliability PL(2500,3) have a higher probability of failure around the maintenance interval of 3000 operating hours. This corresponds with the larger IE found in our simulation results (figure (7)) for the machines with PL(2500,3) compared to those with reliability PL(4000,10).

Figure 12: The expected number of failures around the maintenance interval

The higher the probability of a failure around the preventive maintenance interval is, the more important it is to execute the preventive maintenance in a timely manner. Therefore, the higher the probability (23) is, the more attractive it becomes to deploy N FSE. Equations (19) and (23)), determine how the maintenance frequency and reliability characteristics of the machines contribute to the attractiveness of deploying N FSE. We have demonstrated this for the specific cases of PL(4000,10) and PL(2500,3) with preventive maintenance after 3000 operating hours. More general, we can say that deploying N FSE is most attractive when machines seldom fail at the beginning of the maintenance cycle but have a steep degradation of reliability around the preventive maintenance interval: high values for both (19) and (23). On the contrary, if the failure rate is already substantial at the beginning of the maintenance cycle and steadily augments as the operating hours increase, the attractiveness of deploying N FSE is low (e.g. PL(2500,3)): low values for both (19) and (23).

7 Conclusion

In advanced maintenance service operations a considerable part of the installed base is covered by comprehensive service contracts. In these service contracts the maintenance provider takes full responsibility for all maintenance tasks, rendering the demand dependent on the capacity decisions. We conducted a simulation study incorporating this capacity-demand interrelationship and mimicking the operations of an existing service region of an OEM. Based on the simulation results, we evaluated the possibility to deploy technicians dedicated to preventive maintenance (N FSE) instead of fully cross trained technicians (E FSE).

Employing technicians dedicated to preventive maintenance can be beneficial, as it will lower the response time for these interventions (direct effect). However, the reduced capacity to handle emergencies (failures) will increase the emergency response time. Having dedicated technicians will benefit the timely execution of preventive maintenance, reducing the number of emergencies (indirect effect) and possibly offsetting the negative impact of reducing the number of fully cross trained technicians. In a wide range of scenarios service performance can be improved by employing dedicated technicians. However, the optimal fractions of dedicated technicians remain relatively low and for several scenarios full cross training is optimal.

Concerning the attractiveness of dedicated technicians, we found that with a high workload the optimal fraction of dedicated technicians increases due to a strong indirect effect. Furthermore, highly reliable machines and/or low maintenance frequencies enable the use of dedicated technicians to maintain the machine park. With respect to the maintenance frequency we found that although lowering the maintenance frequency can render cost savings in the form of a lower number of interventions, a low maintenance frequency may prohibit the use of dedicated technicians as the direct effect of N FSE will be lower. So the cost savings due to a less maintenance frequencies is countered by the higher labor cost of an E FSE.

An increasing number of manufacturers are focusing on after sales services by offering comprehensive maintenance contracts to boost their revenues and streamline the service operations. Indeed, our results show that a higher contract coverage offers opportunities to deploy more dedicated technicians while improving the availability of the machines put under contract. Nevertheless caution is warranted, to avoid that the value proposition of the service contracts is eroded the OEM should make sure that his maintenance policy succeeds in decreasing the number of emergencies sufficiently. Otherwise, the total number of interventions required will increase, lowering the availability of the machines already under contract. Therefore, we state that companies actively promoting the sales of comprehensive service contracts, should proactively optimize their capacity, maintenance policy and machine reliability.

Lastly, we used reliability theory to study the impact of the maintenance frequency and machine reliability analytically. We derived that highly reliable machines with fast reliability degradation around the preventive maintenance interval are most suitable to be maintained by dedicated technicians. Machines that fail regularly but for which the failure rate is less sensitive to time, should be maintained by fully cross trained technicians.

References

- Agnihothri, S. R. and A. K. Mishra (2004). Cross-training decisions in field services with three job types and server-job mismatch. *Decision Sciences* 35(2), 239–257.
- Agnihothri, S. R., A. K. Mishra, and D. E. Simmons (2003). Workforce cross-training decisions in field service systems with two job types. *Journal of the Operational Research Society* 54(4), 410–418.
- Apte, A., U. M. Apte, and N. Venugopal (2007). Focusing on customer time in field service: A normative approach. *Production and Operations Management* 16(2), 189–202.
- Bartholomew-Biggs, M., M. J. Zuo, and X. H. Li (2009). Modelling and optimizing sequential imperfect preventive maintenance. *Reliability Engineering & System Safety* 94(1), 53–62.
- Brusco, M. J. (2008). An exact algorithm for a workforce allocation problem with application to an analysis of cross-training policies. *Iie Transactions* 40(5), 495–508.
- Brusco, M. J. and T. R. Johns (1998). Staffing a multiskilled workforce with varying levels of productivity: An analysis of cross-training policies. *Decision Sciences* 29(2), 499–515.
- Cachon, G. and C. Terwiesch (2006). *Matching supply and demand*. McGraw Hill.
- Chakravarthy, S. R. and S. R. Agnihothri (2005). Optimal workforce mix in service systems with two types of customers. *Production and Operations Management* 14(2), 218–231.
- Cohen, M. A., N. Agrawal, and V. Agrawal (2006). Winning in the aftermarket. *Harvard Business Review* 84(5), 129–138.
- Colen, P. and M. Lambrecht (2010). Product service systems: exploring service operations strategies. In *Third International Conference on Information Systems, Logistics and Supply Chain*. Cassablanca, Morocco.
- Doyen, L. and O. Gaudoin (2004). Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering & System Safety* 84(1), 45–56.

- Doyen, L. and O. Gaudoin (2006). Imperfect maintenance in a generalized competing risks framework. *Journal of Applied Probability* 43(3), 825–839.
- Hopp, W. J. and M. P. Van Oyen (2004). Agile workforce evaluation: a framework for cross-training and coordination. *Iie Transactions* 36(10), 919–940.
- Kim, S. H., M. A. Cohen, and S. Netessine (2007). Performance contracting in after-sales service supply chains. *Management Science* 53(12), 1843–1858.
- Oliva, R. and R. Kallenberg (2003). Managing the transition from products to services. *International Journal of Service Industry Management* 14(2), 160–172.
- Papadopoulos, H. (1996). A field service support system using a queueing network model and the priority mva algorithm. *International Journal Management Science* 24, 195–203.
- Pawar, K. S., A. Beltagui, and J. Riedel (2009). The pso triangle: designing product, service and organisation to create value. *International Journal of Operations & Production Management* 29(5), 468–493.
- Saccani, N., P. Johansson, and M. Perona (2007). Configuring the after-sales service supply chain: A multiple case study. *International Journal of Production Economics* 110(1-2), 52–69.
- Tang, Q. H., G. R. Wilson, and E. Perevalov (2008). An approximation manpower planning model for after-sales field service support. *Computers & Operations Research* 35(11), 3479–3488.
- Wang, H. Z. (2002). A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research* 139(3), 469–489.
- Watson, E. F., P. P. Chawda, B. McCarthy, M. J. Drevna, and R. P. Sadowski (1998). A simulation metamodel for response-time planning. *Decision Sciences* 29(1), 217–241.
- Wu, J., M. Xie, and T. Ng (2010). On a general periodic preventive maintenance policy incorporating warranty contracts and system ageing losses. *International Journal of Production Economics* NN, NN.
- Yun, W. Y., D. N. P. Murthy, and N. Jack (2008). Warranty servicing with imperfect repair. *International Journal of Production Economics* 111(1), 159–169.