Optimizing Long-Term Planning of Railway Maintenance

M. Giltaij*, Prof. Dr. R. Goverde**, Dr. N. Bešinović**, Ing. M. Folkerts*, Dr. V. Markine**, Dr. J.A. Annema***

* Arcadis Nederland
** Faculty of Civil Engineering and Geosciences, Delft University of Technology
*** Faculty of Technology, Policy and Management, Delft University of Technology

Abstract

Railway maintenance is currently planned according to a fixed schedule, mostly of one year. This is usually not optimal when considering the cost of maintenance, of possessions and of failure. However, an approach that minimizes the sum of these costs was not yet available. This paper presents a statistical analysis method and an optimization problem that aims to solve this problem. It is applied to fictitious problems and a realistic test case.

I. Introduction

Railways provide affordable and sustainable transportation to many passengers and shippers. However, railways need to be maintained, and this requires train services to be reduced or suspended temporarily. If a reduction of these timetable adjustments can be achieved, rail users profit. When scheduling maintenance, a trade-off has to be made between direct cost, reliability, user preferences, safety and availability. There is a lack of knowledge, coordination and incentive in the parties involved to improve the working methods. Improving and optimizing the way scheduled maintenance is planned could potentially yield benefits to customers and maintenance contractors.

In the Netherlands, the railway network is publicly owned, but maintained by private contractors. Daily maintenance is contracted out on a regional basis. Figure 1 gives an overview of these regions. The maintenance contractors get assigned a certain amount of time to perform maintenance by the infrastructure manager (IM, ProRail in the Netherlands), which is defined in the contract between the two. This contract is based on the tender that is put out by the track manager. Included in this contract is a minimum availability of the infrastructure, expressed in a percentage of weighed time. Weights are allocated to track sections based on importance.

The nature of contracts between the IM and the contractor is shifting from direct specification of work and hourly compensation, to performance targets and fixed sum payment with bonusses and reductions for performance. The task of planning maintenance is also transferred to the contractor, which explains why there is increased interest in efficient maintenance planning.

The offer made by the contractor is a rough estimate of the time needed, and there is a strong incentive to offer a high track availability, as this increases the likelihood of being assigned the contract. The available amount of time is therefore not ideally matched to the maintenance requirement. If contractors have a better indication of the amount of resources that are needed to fulfill the requirement that is proposed in the tender prior to submitting that tender, risk is reduced for all parties involved. One way of doing this is by modelling the maintenance processes, for example in GIS or a mathematical model. The downside of this is that contractors need to invest more in a tender they are unsure to win.

Furthermore, there is a preference by passengers for large maintenance projects, which cannot be performed within a weekend, to be performed during school holidays, when travel demand by commuters is lower. This does mean that maintenance may be performed at a moment in time that is not ideal from an engineering perspective. In addition, because this increases peak demand for maintenance resources, their utilization rates are worse and capital cost higher.

Railways are made up of many components, each of which has to function in order to host the train service. The preferred approach when maintaining these components is to bundle the maintenance operations that affect a service, that is, to reduce the required possessions.

Currently, planning of railway maintenance is done manually, and only for the next year. What's more, maintenance is performed within fixed bounds. There is no tradeoff between cost, reliability and availability. This is not optimal. A better solution may exist than is found by manually creating a solution. An improvement could reduce the number of track failures, the cost of doing maintenance and the number of times the track needs to be closed for performing maintenance. What is desired is a problem formulation that captures all these goals.

Using the latest advancements in data collection, the failure behavior of individual components can be modeled more accurately. When the failure probabilities and costs of components are known, the performance of the system as whole can also be evaluated.

The main research question therefore states: What is the optimal way of planning railway maintenance over a long term? The main research question is subdivided into sub questions. These relate to the characterization, the optimization, the evaluation and the valorization of the problem.

In this paper, a way to analyze the failure characteristics over time and to determine a maintenance schedule that is optimized for the cost of failure, the cost of maintenance and the cost of possessions is introduced. This is more than what has been done in other literature. Ultimately, it could serve to improve the maintenance planning of railways.

Section II discusses the literature that is already available on the topic. This is followed by a brief explanation of the used methodology. Subsequently, in section III the model is presented in two parts; first the statistical analysis and then the optimization model. In section IV the model is evaluated, both on fictitious parameters and on a case study that is based on real data. Finally, a conclusion and recommendations are given.

II. Literature review

Many aspects of railway maintenance have been researched and described in literature. For example, process improvement has been suggested by Nyström (2008) and Vet (2009). These suggestions can provide incremental improvement to the efficiency of railway maintenance.

Many researchers have worked on optimization for railway maintenance. They have usually approached the problem from a deterministic perspective. For example, Budai et al. (2006) consider maximum maintenance intervals, as is currently the practice in maintenance planning, and present an optimal maintenance schedule. Oh et al. (2006) created a solution for a Track Tamping Scheduling Problem (TTSP) in the context of the Korean high-speed railway system.

Optimization of stochastic problems has also taken place. This has been done on individual railway components. The degradation and optimal maintenance strategy were analyzed for longitudinal leveling of the ballast bed (Quiroga and Schnieder, 2010) and (Quiroga and Schnieder, 2012). Consilvio et al. (2016) proposed a more general model, which includes stochastic elements into predictive maintenance scheduling problems.

As can be derived from the literature review, much research has been performed in the fields of railway maintenance, of failure modelling, and of optimization. However, an integrated approach that includes all these elements and applies it to systems consisting of multiple components, is not yet available. The potential for such an approach is that the planning of maintenance can be improved and possibly optimized.

III. Methodology

Several steps are required to get to a well-supported optimization model. First, some definitions and assumptions need to be laid out. This is the foundation for all further modelling. Next, a statistical model that can be used to describe the failure characteristics of railway components is presented. These are then incorporated into a linear optimization model which is able to produce an optimal maintenance schedule.

For a quantitative approach, a way to model degradation of the component is required. Failure of components is a stochastic process. The expected frequency of failure is expressed in the failure rate (Finkelstein, 2008). The failure rate may be constant over time (in which case it is a Poisson point process), or it may be any real and positive function. It can be modelled by defining a failure rate function and estimating its parameters using observed data. For this purpose, a statistical model, the additive Gompertz-Makeham distribution, is presented.

Railway maintenance can be characterized as either preventive maintenance or corrective maintenance. The two are interrelated, because effective preventive maintenance reduces the need for corrective maintenance. Both maintenance types will be included in this research, although corrective maintenance is not modelled.

Preventive maintenance is intended to keep the asset operational. It is planned in advance, and ideally in such a way that normal operations are not affected. The degree to which it can be planned in advance varies. Preferably, it is performed at fixed intervals, but if deterioration of the component is faster than expected, it may have to planned on short notice.

The state of the component is only dependent on the time since last maintenance. Other factors that may influence the failure probability are not considered. After each maintenance operation, the component functions as if it were new.

A type of maintenance is only performed to an element once within one time period. In practice, there are no components that require multiple maintenance operations per week. Therefore, the time step is a week. This means that specific operational details, such as resource assignment and order of maintenance operations, are left out. The model does not have the accuracy needed to incorporate it. As such, the consequences of possessions with individual train services are not relevant.

Corrective maintenance is either a repair of a part that is already broken, or maintenance that has to be performed urgently because the probability of breaking before the next scheduled repair is too high. If a component is broken or exceeds a safety limit that prohibits operations, the asset cannot be used until corrective maintenance is performed, and it is considered to have failed. Even though technically it may still function, the asset is no longer available for use and the cost of repair is incurred. Ideally, corrective maintenance is not needed, but in practice it is not economically efficient to attempt to prevent all failures.

Repairs are considered to be minimal, as defined in Finkelstein (2008). The component that undergoes repair is in the exact state it was in before the failure occurred. This means the failure behavior does not change as a consequence of repairs, and that maintenance schedule does not have to be adapted. In fact, modelling of specific failure occurrences is not needed at all.

An element either functions or fails. There is no state of limited functionality. One instance where limited functionality would be possible in practice is a switch that is locked in a certain position. Trains can still run over the switch, but in one direction only. However, repair is still necessary.

The costs for a possession are dependent on the moment of possession. This is done through using different costs for possessions depending on the week. Furthermore, possessions may have different sets of components and time lengths. This may be introduced by including multiple possessions within one week.

Different individual components that have the same set of characteristics are aggregated into one category. It is not useful to model these separately, as the resulting schedule applicable to the components should be the same.

A maintenance interval is the time between two maintenance operations. The expected number of failures in an interval is a function of only the length of the interval.

III.i. Statistical Model

The failure rate is assumed to follow a Gompertz-Makeham distribution. The failure rate for this function is:

$$\lambda(t) = abe^{bt} + cde^{dt} + f \tag{1}$$

with:

$a, b < 0; c, d > 0$

Integration of this function yields the cumulative failure rate. This is the expected number of failures up to point in time t.

$$\Lambda(t) = \int_0^t \lambda(t)dt = (ae^{bt} + ce^{dt} + ft + C) - (ae^{b\times 0} + ce^{d\times 0} + f \times 0 + C) = ae^{bt} + ce^{dt} + ft - a - c \tag{2}$$

The cost of failure, like failure itself, is stochastic. If the cost of an occurrence of failure is assumed to be constant over time, the total cost of failure over time is directly proportional to the cumulative failure rate.

The failure rate model can be used to optimize the maintenance schedule by expressing the cost due to failures as a function of the maintenance schedule. Ultimately, it is desired to minimize the sum of the cost of maintenance, the cost of failure and the cost of possessions. Linear programming is a widely used tool for solving numerical optimization problems. The underlying assumption is that the relation between each parameter that can be changed and the overall desirability is linear. Even though the problem is not linear in itself, a transformation of the decision variables enables successful problem formulation. In this way, the costs of maintenance, possession and failure can be included in the objective function.

A model is presented that achieves the set goals. The objective function is an expression of the total expected cost as a function of the chosen maintenance strategy. The models are able to solve problems that cannot be solved using models in existing literature. The formulation is very general, and could possibly be applied in many fields other than railways. They have the potential to improve the way maintenance to all kinds of degrading systems is planned.

The model optimizes a situation in which each component in the system has a independently determined maintenance intervals. The costs of maintenance, failure and possessions are taken together and minimized. The result is an efficient maintenance schedule.

The cost of a single maintenance interval to one component can be expressed as a function of the cumulative failure rate. This is defined in equation 3.

$$CMI(t) = CoF \times \Lambda(t) + CoM = CoF \times (ae^{bt} + ce^{dt} + ft - a - c) + CoM \tag{3}$$

III.ii. Optimization Model

The main goal of the optimization is to minimize the total cost over all maintenance intervals for all components, as well as the cost of the possessions. The objective function for the combined model is therefore:

$$\min \sum_{i=1}^{Components} \sum_{j=1}^{Intervals_i} CMI_i(t_{ij}) + \sum_{i=1}^{Possessions} CoP_i \tag{4}$$

with:

This cannot be solved directly in linear programming. A reformulation is applied that enables a solution. The final model is as follows.

$$\min \sum_{i=1}^{Components} \left(\sum_{k=0}^{Values_i} CoF_i \times (a_i e^{b_i k} + c_i e^{d_i k} + f_i k) \times x_{i1k} + \sum_{j=2}^{Intervals_i} \sum_{k=1}^{Values_i} (CoF_i \times (a_i e^{b_i k} + c_i e^{d_i k} + f_i k) + CoM_i) \times x_{ijk} \right) + \sum_{i}^{Weeks} CoP_i y_i \tag{5}$$

$$s.t.$$

$$\sum_{k=0}^{Values_i} kx_{ijk} - \sum_{k=0}^{Weeks-1} (Weeks - k)\, z_{i(j-2)k} + \sum_{k=0}^{Weeks-1} (Weeks - k)\, z_{i(j-1)k} = 0 \ \forall\, i \in Components, j \in \{3 \dots Intervals_i - 2\} \tag{6}$$

$$\sum_{k=0}^{Values_i} kx_{i1k} + \sum_{k=1}^{Values_i} kx_{i2k} + \sum_{k=0}^{Weeks-1} (Weeks - k)\, z_{i1k} = Weeks \ \forall\, i \in Components \tag{7}$$

$$\sum_{i=1}^{Components} x_{i1\,0} - M_1 y_0 < 0.5 \tag{8}$$

$$\sum_{i=1}^{Components}\left(x_{i1k}+x_{i(Intervals_i)(Weeks-k)}+\sum_{j=1}^{Intervals_i-3} z_{ijk}\right) - M_2 y_k < 0.5 \ \forall\ k \in \{1 \dots |Weeks|-1\} \tag{9}$$

$$\sum_{k=0}^{Values_i} x_{i1k} = 1 \ \forall\ i \in Components \tag{10}$$

$$\sum_{k=1}^{Values_i} x_{ijk} < 1.5 \ \forall\ i \in Components, j \in \{2 \dots Intervals_i\} \tag{11}$$

$$\sum_{k=0}^{Weeks-1} z_{ikl} < 1.5 \ \forall\ i \in Components, l \in \{1 \dots Intervals_i - 3\} \tag{12}$$

$$\sum_{k=1}^{Values_i} x_{i(j-1)k} - \sum_{k=1}^{Values_i} x_{ijk} > -0.5 \ \forall\ i \in Components, j \in \{3 \dots Intervals_i\} \tag{13}$$

with:

The objective function is in (5). Unary encoding has been applied to allow the problem to be expressed in a linear programming formulation. Constraints (6) and (7) are auxiliary constraints that are needed to define the possessions. Constraint (8) defines that there needs to be a possession at $t = 0$ if there is any maintenance then. It uses a big M formulation, where $M_1$is the number of components in the system. Constraints (9) defines that if there is any maintenance at $t = k$ there needs to be a possession. It also uses a big M notation, where M_2 is equal to the sum of number of possible intervals minus one for each component. Constraint (10) requires the first interval for each component to have exactly one value. Constraint (11) requires each subsequent interval for each component to have at most one value. If it has no value, this means the interval does not exist. The model is able to produce a solution with fewer maintenance intervals for a component i than the maximum number for that component ($Intervals_i$). Constraint (12) does the same thing as (11), but for the auxiliary decision variables. Constraint (14) is not necessary to arrive at the correct solution and is used to limit the solution space in order to speed up the solution of the model. It does not allow an interval j if there is no interval $j - 1$. Without this constraint, multiple solutions may exist that are the same in practice.

IV.i. Experimental results

A series of performance tests was devised to systematically evaluate several aspects of the model. A limiting factor in the practical application of the model could be the problem size that can be solved with reasonable computing resources. The model should also be able to handle variation in parameter values. Furthermore, changing settings on the computational side could impact the required run time. In total, eight different parameters were varied: optimality gap, presolve setting, number of processor threads, wear out shape parameter, the cost of possession, the planning horizon and the number of components.

The optimality gap is the allowed difference between the best feasible solution that was found and the lower bound for the solution, expressed as a percentage. The larger the optimality gap, the shorter the runtime. The model is responsive to increasing the optimality gap. Should he standard specification of 0 result in an overly long runtime, it will help to decrease the optimality gap.

Using more processor threads does not always yield better performance. In some cases, solving the problem or sub-problems linearly on one thread instead of parallel on multiple threads performs better. An interesting find is that going from an even to an odd number of threads increases runtime. In general, however, the model performs better when increasing the number of processor threads. This means parallel computing can be used to solve bigger problems with this model.

The cost that results from applying a possession is also varied. For a given problem, there is a point of this parameter where the solution time is longest. Increasing or decreasing the cost of possession reduces the difficulty of solving the problem.

Varying cost of possession over time emulates situations in which one week is more expensive than the other. As expected, expensive weeks are avoided. This shows the model can be used to plan maintenance at more beneficial times. The problem becomes faster to solve compared to a constant cost of possession for every week. An overview of the runtimes depending on the cost of possession is in figure 1.

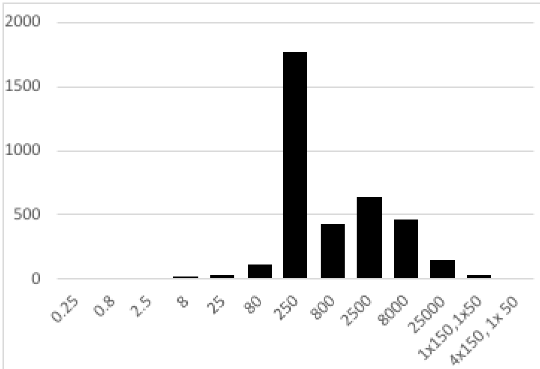

Figure 1, effect of cost of possession on runtime

For the example case, problems of up to 250 time steps could be solved to optimality within one hour. Accepting an optimality gap of 10% increased this figure to 1100. This means maintenance can be planned with enough accuracy for the intended application.

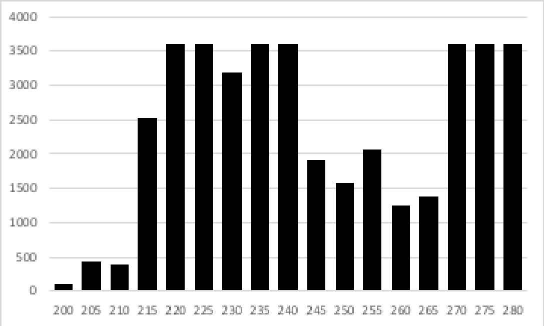

Figure 2,runtime as a function of planning horizon with an optimality gap of 0

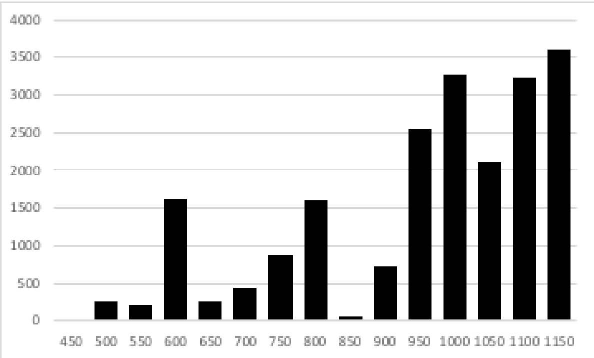

Figure 3, runtime as a function of planning horizon with an optimality gap of 10%

As more components get added to the optimization, the problem size increases. For the example case, the runtime increases from 1.1 seconds for two components and 110.6 seconds for three components to over one hour for five components.

All in all, the model handled changes in these parameters without highly unexpected results.

IV.ii. Test Case

A case study was made of two different components, these being electrical separation joints and switches. Cost values as used in practice were obtained. The statistical model is based on maintenance and failure records for the two components. The vast majority of these components are currently maintained once a year.

The estimation of parameters results in an estimated number of failures. When comparing the estimated and the observed number of failures for switches using a diagram (figure 4), the ability of capturing the falling and rising tendencies in failure probability become visible.

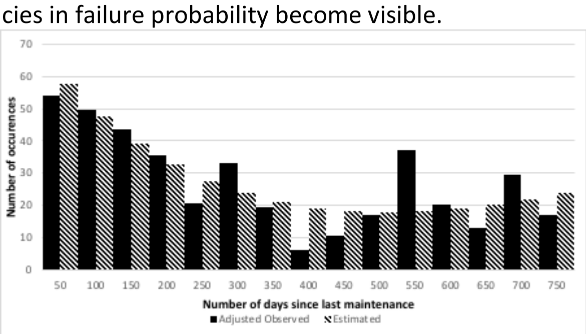

Figure 4, observed and predicted failure counts of switches

The optimization model is able to produce an optimal maintenance schedule for 5 years (261 weeks) in 9.45 seconds. The results show that the maintenance intervals for the components individually, without taking the cost of possession into account, is 126 and 61 weeks for switches and insulated joints respectively. The resulting schedule has one maintenance operation for the switches and three for the insulated joints. The solution is visually expressed in figure 1. The grey horizontal bars are the schedules for component 1 and 2 and for the possessions. On the x axis is the time in weeks. Each broad horizontal bar is a maintenance operation.

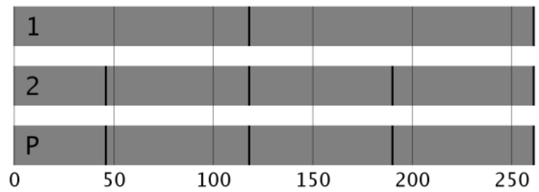

Figure 5, optimized solution to the test case

Table 1, comparison of schedules of the test case.

	Conventional intervals	Optimized intervals	Optimized schedule
Objective value	716.4	703.4	699.9
# possessions	8	6	3
# maintenance operations to switches	5	2	1
# maintenance operations to insulated joints	5	4	3
Compared to conventional intervals	100 %	98.2 %	97.7 %
Compared to optimized intervals	101.8 %	100 %	99.5 %
Compared to optimized schedule	102.4 %	100.5 %	100 %

Comparing the three schedules in table 1, the difference in objective function is small. Even so, small improvements such as these would amount to sizeable profits over the longer term. As the difference in number of maintenance operations and number of possessions is substantial, it seems that the cost of failure is a large factor in this case. Either that is true, or the parameters are based on an overestimation of failures.

Considering that better analysis methods and the use of better data could improve the performance of the model, there is potential for delivering further improvements. Furthermore, differentiated maintenance, as is already applied to switches, can also improve the usefulness of the model. This could be included by splitting components to which it is applied into multiple separate groups, for example based on frequency of use.

VIII. Conclusion and Recommendations

The long-term planning and cost estimation of maintenance can be improved by using the model presented in this paper. As opposed to current methods, a quantification of the expected cost of failure depending on the maintenance strategy is defined.

The model is shown to be an improvement over current working methods. It considers the specific characteristics of the components and, if it is beneficial, combines work to take place in the same week so fewer possessions are needed. Maintenance engineers can make a schedule in which the combined cost of maintenance, failure and possessions is optimized.

Further development that builds on the model as presented could address certain uncertainties that are still present. The main uncertainty is, and probably remains, the predictive quality of the available data. Before major decisions can be based on the model, a more elaborate trial would be advisable.

While the model is only applied to railway components, its possible applications extend beyond this field. As the statistical model can be applied to any system that has maintenance and subsequent degradation, there are many physical systems that could benefit from a maintenance schedule that is optimized by this model.

References

Budai, G., Huisman, D., & Dekker, R. (2006). Scheduling preventive railway maintenance activities. *Journal of the Operational Research Society, 2006*(57), 1035–1044. https://doi.org/10.1057/palgrave.jors.2602085

Consilvio, A., Di Febbraro, A., & Sacco, N. (2016). Stochastic Scheduling Approach for Predictive Risk-Based Railway Maintenance. *2016 IEEE International Conference on Intelligent Rail Transportation (ICIRT)*, Birmingham, 197-203. https://doi.org/10.1109/ICIRT.2016.7588732

Finkelstein, M. (2008). *Failure rate modelling for reliability and risk*. Springer Science & Business Media. https://doi.org/10.1007978-1-84800-986-8

Nyström, B. (2008). *Aspects of improving punctuality: from data to decision in railway maintenance* (Doctoral dissertation, Luleå tekniska universitet). http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A999787

Oh, S. M., Lee, J., Park, B., Lee, H., & Hong, S. (2006). A study on a mathematical model of the track maintenance scheduling problem. *Computers in Railways X*, *88*, 85-96. https://doi.org/10.2495/CR060091

Quiroga, L. M., & Schnieder, E. (2010). A heuristic approach to railway track maintenance scheduling. *WIT Transactions on The Built Environment*, *114*, 687-699. https://doi.org/10.2495/CR100631

Quiroga, L. M., & Schnieder, E. (2012). Monte Carlo simulation of railway track geometry deterioration and restoration. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *226*(3), 274-282. https://doi.org/10.1177/1748006X11418422

Vet, L. (2014). Het verbeteren van predictief liggingsonderhoud bij ASSET Rail: de interactie tussen medewerkers, techniek, organisatie en omgeving op het trans-organisatorisch onderhoudsproces (master thesis, University of Twente). http://essay.utwente.nl/65797/1/VetL_1242032_openbaar.pdf

Appendix 1: notation

$\lambda(t)$	hazard rate at t
a_i	location parameter for break in for component i
b_i	scale parameter for break in for component i
c_i	location parameter for wear out for component i
d_i	scale parameter for wear out for component i
f_i	vertical location parameter for component i
t_{ij}	The length of maintenance interval j to component i
$CMI_i(t_{ij})$	Cost of Maintenance Interval j of component i.
CoF_i	The cost of failure of component i.
$\Lambda(t)$	The cumulative failure rate function of the component.
CoM_i	The cost of performing maintenance once on component i.
$Components$	The set of components in the system under consideration.
$Intervals_i$	The maximum number of maintenance intervals of component i.
$Possessions$	The set of all possessions.
CoP_i	The cost of possession i. Not every possession has the same cost, as some weeks may be preferred over others.
$Values_i$	The possible values of the interval length for component i.
x_{ijk}	The binary decision variable representing interval j of component i being k long.
$Weeks$	The weeks within the planning horizon.
y_i	The decision variable representing a possession at time j.