Productivity Analytics in

a Tree Cutting & Earthmoving Project

using a Stochastic Based Decision-Support Tool

- Symphony.net

CE5805QA Construction Productivity Analytics

PREPARED FOR:

Prof. Chua Kim Huat, David

PREPARED BY:

PAN XINXIN

Contents

1. Abstract and Executive summary

In this study, we model a tree cutting and earthwork construction project in Shenzhen, China. The project mainly cleaned 2,000 trees and completed 220,000LCM of earth excavation. We observed the project and described as 1) Timber cutting, 2) tree truck cycle, 3) Soil loading cycle, 4) Soil dump process. To prepare the data to simulate the processes and queues of operations in each resource cycle by the Simphony.NET, recorded the real running time, arrival rate, number of services and other information, and did statistical analysis.

We defined 3 kinds of statistical distributions to estimate the observed data, including normal distribution, beta distribution and gamma distribution. Based on the results of the goodness of fit test, we choose the appropriate distribution type for each process. After running the model simulation and doing trials and tests, we performed a sensitivity analysis of the results and determined that the optimization objective was the number of trucks, a variable that plays a key role in controlling the process duration. Based on our analysis and discussions, we recommend the following management strategies for the project:

1) Considering site capacity and queuing, increase the number of trucks while adjusting the number of excavators to align with it for improved operational efficiency;
2) The tree cutting cycle serves as a prerequisite for the initiation of the 2^{nd} soil cycle, and initially it can allocate a greater number of lumberjacks. Subsequently, as the soil excavation commences, a reduction in the number of lumberjacks may be warranted, as one of the methods for budget control.

2. Background of Project

2.1. Background

The earthwork construction project was located in the central area of Baoan, Shenzhen, China (in Figure 1), and it required earthwork grading for a commercial development. Before the earthwork, there were trees covered in the site with 68,900 cubic meters. Therefore, the project involved mainly 2 tasks as following:

1) Cut and cleaned 2000 trees in the site;
2) Earth excavation around 200,000 m^3.

We do some assumptions and transformation for the following simulation, thus the total soil needed to be excavated is 220000 LCM.

Figure 1. Location of the project

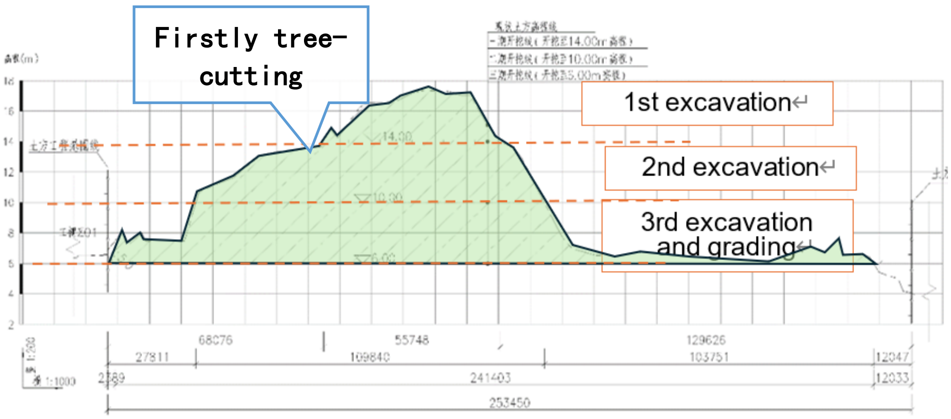

Figure 2. Grading Excavation Section

The original site was covered by coastal clay and filled with construction waste and silt, with a density of about 1.88g/cm^3. The average designed excavation height was 4m, and the maximum excavation height was up to 11.5m. The Figure 2 show a simplified excavation profile.

2.2. Information of Main Individual Process

We simplify the total project as 2 cycles:

2.2.1. Timber Cycle

In the timber cycle, it involves timber cutting and tree truck cycle.

Figure 3.Timber cutting Figure 4. Load and haul trees

The first thing is each crew of lumberjacks enter the site and cutting trees. Then the loader lifts timbers into trucks. After the timber concentration until sufficient trees, the truck hauls to the offloading site and wait for the forklift to offloads trees. After offloading, the idle truck returns the project site.

2.2.2. Soil Excavation Cycle

In the second cycle, the soil excavation cycle, includes soil loading cycle and soil dump.

Figure 5. Soil excavation Figure 6. Truck haul

Figure 7. Truck dump Figure 8. Soil spread

The loading cycle starts from the soil excavation firstly, and then truck loading soil hauls to dump and soil spread, where there is a spotter in the dump process. After all excavated soil from this project is spread by dozer, the project is completed successfully.

3. Resource cycles and integrated simulation network

In this productivity analytics, we use the Simphony.NET constructs to model.
In the project, the total soil needed to be excavated is 220000 LCM, and the total number of abandoned trees are 2000.

3.1. Resource Link

For the resource entering in the system, as the initial resource was the tree timber, we need to define a transformation relationship for resource in the 2nd soil excavation cycle. We need to cut trees firstly, then start excavation in the corresponding area. Consider the capacity volume of truck, that loading soil is 10 LCM, meaning that it needs 220000 LCM /10 LCM = 22000 trucks for soil; the other loading timber is 20 trees, indicating that it needs 2000/20 = 100 trucks for trees.
Relationship: 2000 trees ~ 100 trucks (timber) ~ 220000 LCM soil ~ 22000 trucks (soil)
➔ Density: 1 truck (timber) ~ 220 trucks (soil)
Therefore, set a GEN node after cutting process to present the corresponding density for trucks loading soil.

3.2. Integrated simulation network

Based on the above resource transformation, the link between these 2 cycles is to use density to transfer the initial trees resource in the 1st cycle into the soil in the 2nd cycle. And the flow diagram is the following, where the color of grid corresponds to the kind of distribution and the necessary parameters. The chosen distribution is shown in the following analysis (5 Model Inputs):

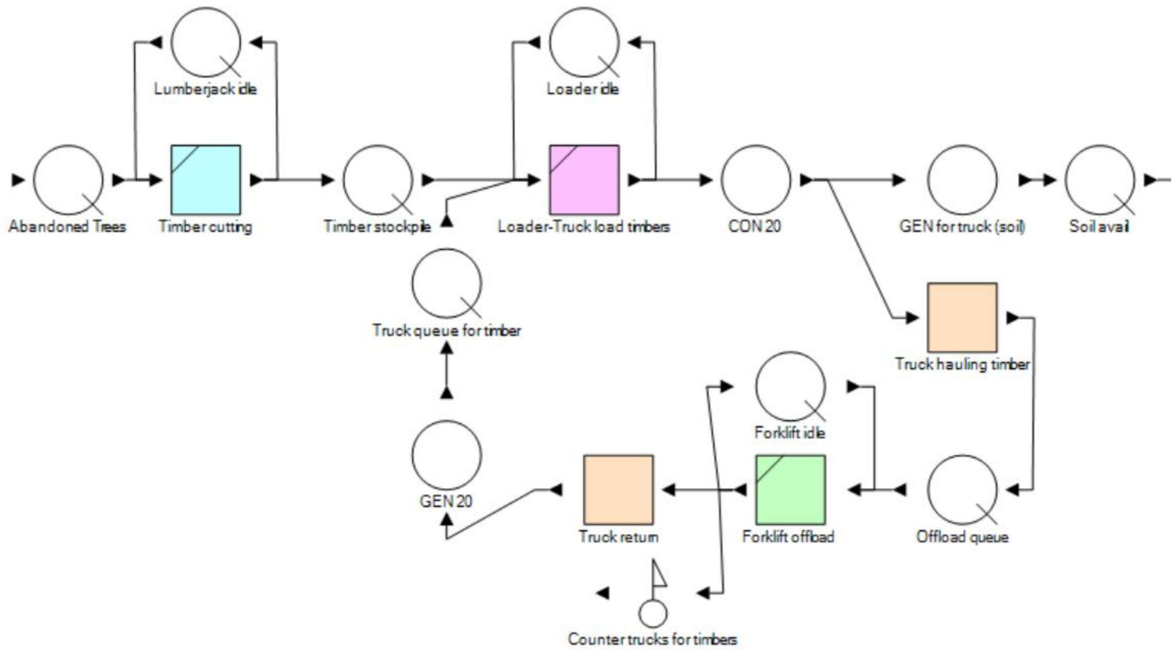

Figure 9. 1st part: Timber Cycle

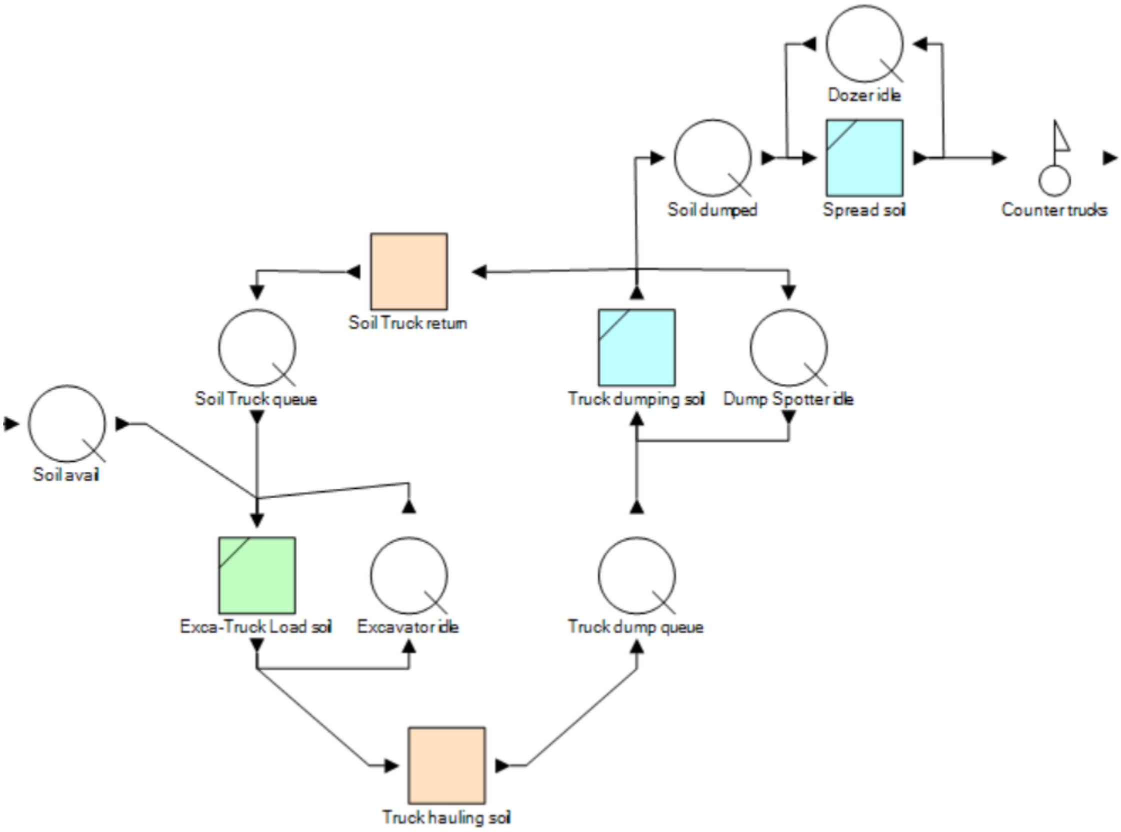

Figure 10. 2nd part: Soil Excavation Cycle

4. Parameters and Distribution Test for Input data

4.1. Examined Distribution

The main statistical inputs in Simphony.NET are the distribution of the processing time of some elements in the overall project. The basic CYCLONE modeling elements that

need to add the processing time contain Normal work task elements and COMBI work task elements, and the other important element is queue node, set as a constant value.
We observed the main processes: 1) Timber cutting, 2) tree truck cycle, 3) Soil loading cycle, 4) Soil dump, and recorded their cycle time.
Based on the preliminary histograms, we defined the 3 kinds of statistical distribution to estimate our observation data, referring to normal distribution, beta distribution and gamma distribution.
The methods for examined distributions are mainly following:

4.1.1. Histograms

In order to construct the histogram, firstly, define the number of bins by Sturges' rule: [1+3.3log$_{10}$(n)], according to the number of observation data (No. = 25~30). Then we can get the number of bins.

$$Number\ of\ Cells = [1 + 3.3log_{10}(n)]$$

$$Width\ of\ Cell = \frac{\max\{x_i\} - \min\{x_i\}}{Number\ of\ Cells}$$

We recorded 30 sets of values for the time taken by the excavator to fill a truck, ranging [6,10] (mins). The number of bins was calculated to be 6 with a width of 2/3. The code for the calculation process is shown in *Appendix I*. The example histogram follows.

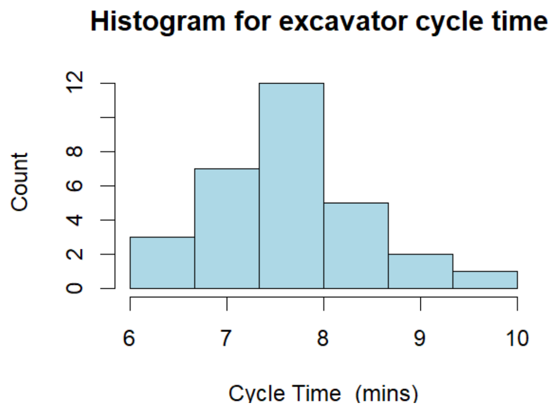

Figure 11. Histogram for Excavator Cycle Time

4.1.2. Fitted with Different Distributions

Only for beta distribution, it is necessary to normalize the given data to range onto [0,1] by the min-max method. For the other two distributions, the main process is to calculate the necessary statistical parameters for randomly generating data within the corresponding distribution.
The calculation process is implemented by statistical functions using R language, like:

1) rnorm() for randomly generating value fitting normal distribution;
2) after obtaining the shape and rate(=1/scale) parameters by fitdist(data, "gamma"),

use rgamma() to randomly generate value fitting gamma distribution;

3) to fit beta distribution, normalize by min-max method and use ebeta() to get alpha and beta parameters; then use rbeta() to randomly generate normalization value fitting beta distribution.

The above calculation is taken by R language, with code shown in 10.1 Appendix I: Derivation for probability distribution.

4.2. Goodness of Fit Tests

To check how good our choice of distribution is, we need to do goodness of fit tests. We use two ways to achieve the test: Q-Q plot and Pearson's Chi-Square test.

4.2.1. Q-Q Plot

If the points on the Q-Q plot closely follow a straight reference line (y=x), it suggests that the data approximates the estimated distribution. The closer the points are to the line, the better the fit. On the contrary, if there is a significant deviation with the line, that means a bad fitting.

4.2.2. Pearson's Chi-Square Test

Use Pearson's Chi-Square test to determine which distribution is the most suitable one. For the three distributions, the null hypothesis H0 and alternative hypothesis H1 are as follows. The significance level alpha is defined as $\alpha = 0.05$.

Table 1. Hypothesis

	Null Hypothesis H0	Alternative Hypothesis H1
Normal Distribution	Distribution is standard normal	Distribution is not standard normal
Gamma Distribution	Distribution is Gamma	Distribution is not Gamma
Beta Distribution	Distribution is Beta	Distribution is not Beta

In R, use chisq.test() function to calculate the p-value for the observation and estimated frequencies. The code is shown in 10.2 Appendix II: Goodness of fit tests.

4.3. Statistical Results

We show the details for 3 kinds of distributions for processes as examples, including 1) Excav-Truck loading soil; 2) Truck hauling soil; 3) Timber cutting. The calculation and analysis are taken by R language, with code shown in 10.1 Appendix I: Derivation for probability distribution. The observation data is following:

Table 2. 3 Types of Observation Data about the Cycle Time (unit: mins)

No.	Excav_Truck Loading Soil	No.	Excav_Truck Loading Soil	No.	Truck Hauling Soil	No.	Truck Hauling Soil	No	Timber Cutting	No.	Timber Cutting
1	7.5	16	7.2	1	15.20	16	16.80	1	3.5	16	3.8

No.	Excav_Truck Loading Soil	No.	Excav_Truck Loading Soil	No.	Truck Hauling Soil	No.	Truck Hauling Soil	No	Timber Cutting	No.	Timber Cutting
2	8	17	6.8	2	16.00	17	16.50	2	3.2	17	3.8
3	8.4	18	10	3	17.00	18	15.80	3	3.3	18	4.3
4	7.5	19	7.2	4	17.20	19	15.60	4	3.6	19	3.7
5	9	20	7.5	5	15.80	20	16.40	5	3.5	20	4.8
6	8	21	6.5	6	15.50	21	15.80	6	3	21	3.6
7	7.4	22	8.5	7	15.00	22	15.20	7	4	22	4.3
8	8	23	6.2	8	16.20	23	15.00	8	4.4	23	4.1
9	6	24	7.4	9	14.80	24	15.50	9	3.5	24	3.7
10	8.5	25	8.2	10	15.50	25	15.80	10	3.8	25	4
11	9.2	26	6.8	11	15.80			11	3.5		
12	8.4	27	7.6	12	18.00			12	3.8		
13	7.2	28	8	13	16.20			13	3.5		
14	7.5	29	7.5	14	14.50			14	3.2		
15	6.8	30	7.2	15	16.50			15	3.6		

4.3.1. Q-Q Plot Results

The following figures show the Q-Q plots, and the analysis of the points fitting and derivation is recorded in the Table 3. And the order of 3 figures is by column for one process.

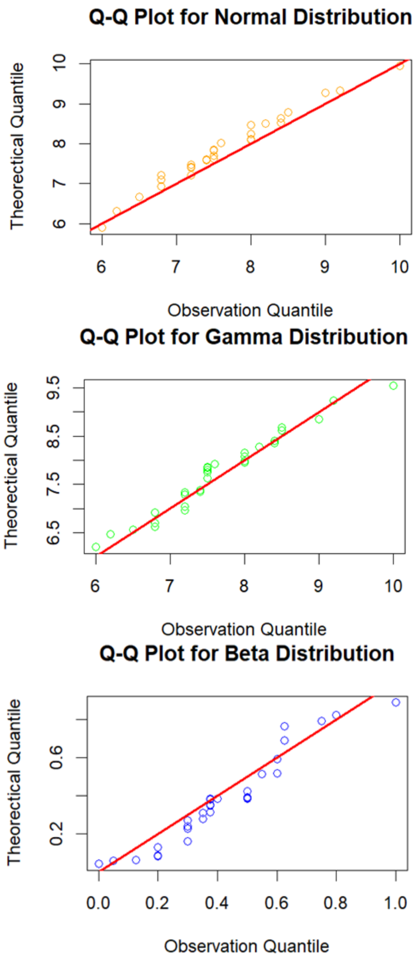

Figure 12. Q-Q Plot for Excav-Truck Loading Soil

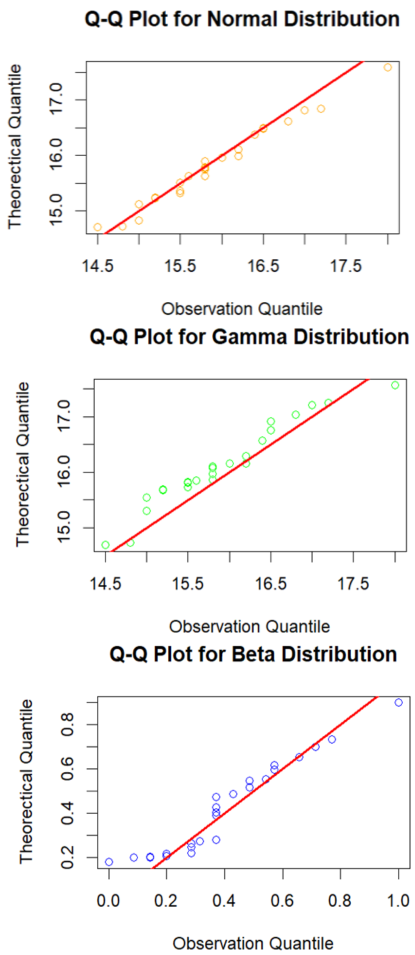

Figure 13. Q-Q Plot for Truck Hauling Soil

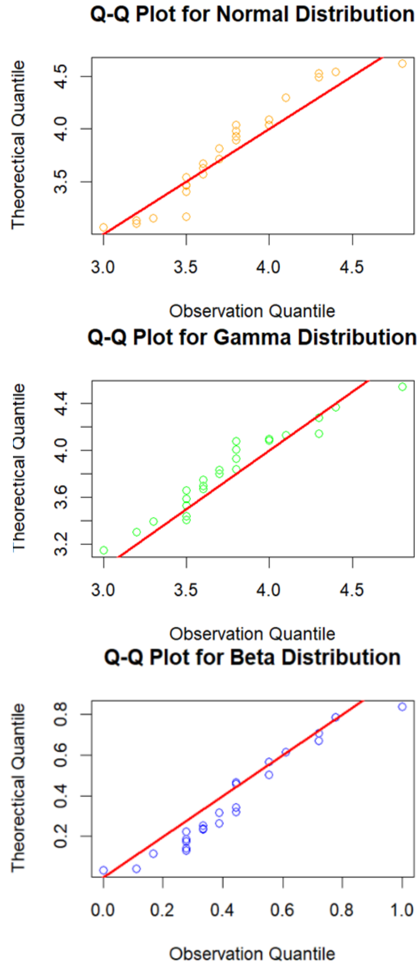

Figure 14. Q-Q Plot for Timber Cutting

4.3.2. Statistical Summary

The results are shown in the table below. For the Q-Q plot, visual inspection is used to check which distribution fits the observed data better. That is, how far the observed data deviates from the straight line.

The closer the data points are to the straight line, the better the fit is represented. For the p-value, as p-value is greater than the significance level α = 0.05, fails to reject the null hypothesis. If less than 0.05, conclude that the two variables are in fact dependent. The distribution with a higher p-value is superior.

Table 3. Summary for Tests

Type	**Q-Q Plot by Distribution**			**P-value by Distribution**			**Chosen Distribution**
	Beta	Gamma	Normal	Beta	Gamma	Normal	
Load Soil	deviate significantly	close to	close to	0.553	0.861	0.299	Gamma

Type	Q-Q Plot by Distribution			P-value by Distribution			Chosen Distribution
	Beta	Gamma	Normal	Beta	Gamma	Normal	
Haul Soil	close to	deviate significantly	close to	0.936	0.021	0.994	Normal
Timber cutting	moderate deviation	deviate significantly	moderate deviation	0.772	0.034	0.437	Beta

After defining the estimated distribution for the specific processes, visualize by plotting the histograms of observation data and the corresponding estimated distributions as following.

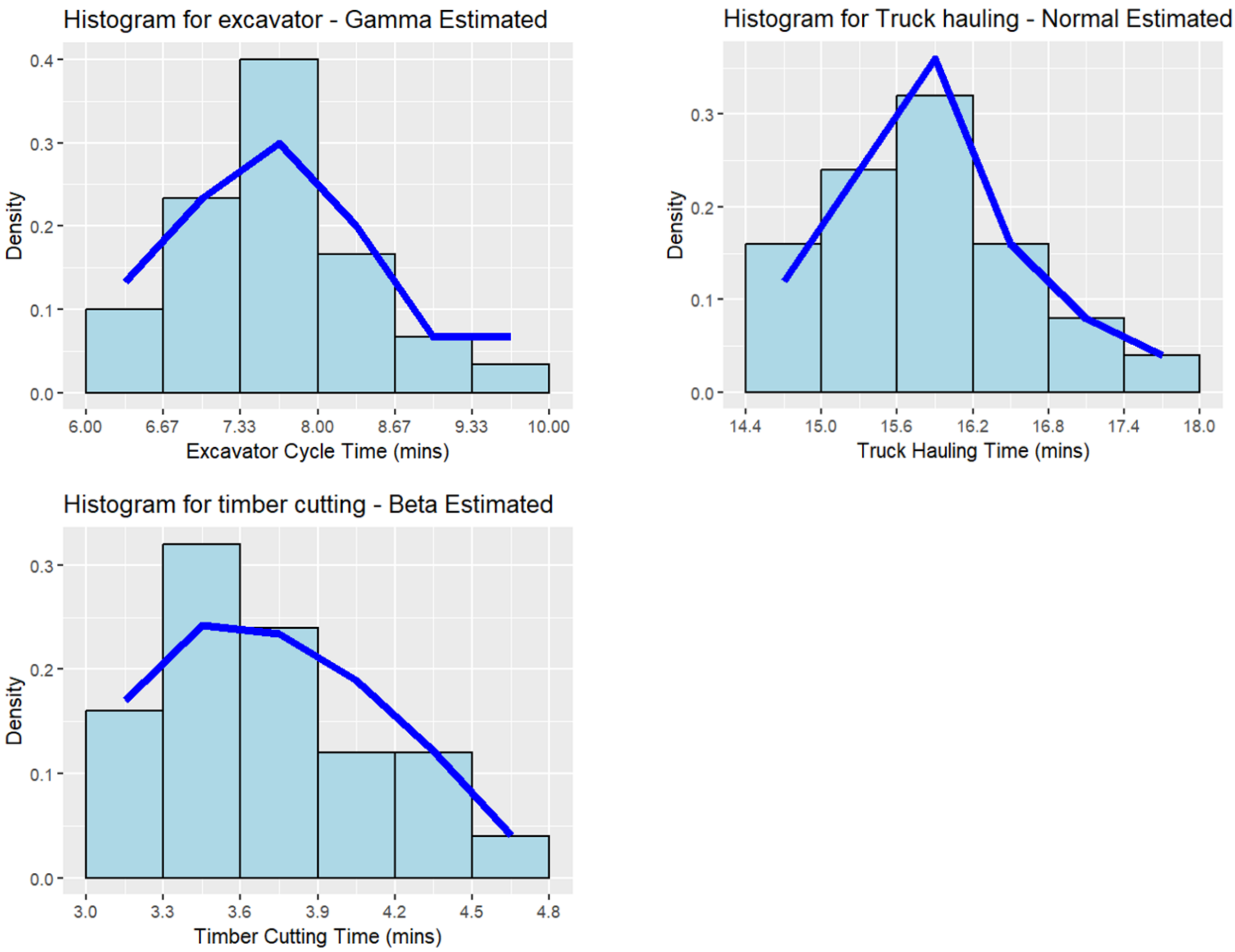

Figure 15. Line-Histograms for Example Processes

5. Model Inputs

After defining the distribution and parameters, the input PDFs as the simulation set-ups for processing time and the initial length for the queue components are in Table 4.

Table 4. Distribution Parameters and Description for Each CYCLONE Modeling Element

NO.	Cycle	Part	Node	Description	Density Type	PDF - Parameters				
						constant	shape	scale	mean	StDev
						BETA:	Alpha	Beta	Low	High
1	**Timber cutting**	Abandoned Trees	Q	total trees	initial-L	2000				
2		Timber cutting	combi	a crew for 5 trees once	Beta		1.458	2.089	3.000	4.800
3		Timber stockpile	Q	in system N=6	constant	0				
4		Lumberjack idle	Q		constant	5				
5	**Tree truck cycle**	Loader-Truck load trees	combi	in system N=40/20+1=3	constant	1				
6		Loader idle	Q		constant	40				
7		truck hauling timber	Norm	same with soil	Normal				15.904	0.808
8		offload q	Q		constant	0				
9		Truck-Forklift offload	combi	for 1 truck	Gamma		830.714	0.008		
10		Forklift idle	Q	Forklift N=2	constant	1				
11		truck return	Norm		Normal				9.744	0.581
12		CON 20/GEN 20	NODE	20 idle virtual truck	Divide/ Multiply	20				
13		Truck q for timber	Q	N=40/20+1=3	constant	40				
14	**Soil loading cycle**	GEN for truck (soil)	NODE	density for tree-soil-truck	multiply	**220**				
15		Soil available	Q		constant	0				
16		Exc-Truck Load soil	combi		Gamma		79.910	0.096		
17		FEL idle	Q	N=4	constant	3				
18		Truck hauling soil	Norm		Normal				15.904	0.808
19		Truck dump queue	Q		constant	0				
20		Soil Truck dump	combi	N=5	Beta		1.499	1.468	3.100	5.020
21		Dump Spotter	Q		constant	4				
22		Soil Truck return	Norm	truck N=15	Normal				9.744	0.581
23		Soil Truck queue	Q		constant	14				
24	**Soil dump**	soil dumped	Q		constant	0				
25		spread soil	combi	dozer N=6	Beta		2.490	2.084	7.000	8.800
26		dozer idle	Q		constant	5				
27		soil spread	Q		constant	0				

Note:

1) The color of grid is related to the kind of distribution and the necessary parameters. And the corresponding color is set in Simphnoy.NET (seen in Figure 9 and Figure 10);
2) There are two counter tools in model for the calibration.

6. Model Running and Results Analysis

6.1. Evaluation and Calibration for Initial Model

To verify the total project is achieved successfully, check the two counts at the end of two main cycles. It can be seen in Table 5 that the final counts for two kinds of trucks are 100 and 22000, respectively for trees and soil excavation, that correspond to the requirement value. Thus, we can conclude that all trees are cut, and soil is excavated.

Table 5. Counters for Scenario No.1

Element Name	Final Count	Production Rate	Average Interarrival	First Arrival	Last Arrival
Counter trucks	22,000.000	0.365	2.736	52.223	60,246.209
Counter trucks for timbers	100.000	0.047	20.991	41.484	2,119.556

Besides, the lasting time for excavation is significantly more than the timber cutting cycle, that means when trees cutting succeeded in a short time, the excavation is still processing. If the total process still counted the lasting time in the system, that would cause the obvious deviation of average waiting time in 1st cycle. Thus, if we want to analyse and adjust the scenario setting in timber cutting cycle, it is necessary to divide the total process into 2 parts, and then the correct waiting time for timber cycle can be visualized.

Also, the initial and current(end) length of abandoned trees can be seen as additional indicators to evaluate project completion (seen in Table 6).

6.2. Initial Scenario Results

In the initial scenario, the total time is 60,248.00 mins, with 8hrs-per-day working efficiency, it needed around 125 working days. And the main waiting time results for excavation in the initial scenario (No.1) are following:

Table 6. Main Waiting Results for Scenario No.1

Element Name	Average Length	Standard Deviation	Maximum Length	Current Length	Average Wait Time
Abandoned Trees	24.816	180.161	2,000.000	0.000	747.572
Dozer Idle	2.086	0.741	5.000	5.000	5.710
Dump Spotter Idle	2.514	0.815	4.000	4.000	6.881
Excavator Idle	0.198	0.432	3.000	3.000	0.537
Soil Avail.	10,608.271	6,133.816	21,238.000	0.000	29,051.398
Soil Dumped	0.000	0.013	1.000	0.000	0.000
Soil Truck Queue	0.347	0.648	14.000	14.000	0.937
Truck Dump Queue	0.000	0.000	1.000	0.000	0.000

To judge whether the input length is too much or too little, compare the average length of each queue element to the initial value.

As the excavation process domains the lasting time, we mainly analyse and adjust the inputs in 2nd cycle.

For the timber cutting cycle, the average length of dozer idle was 2.1, amounting half of the initial value, indicating around half of dozer were in queue instead of working, and the inputs were much, or cannot match the length of other equipment. Thus, if remain other unchanged, the number of dozer idle can reduce to remain the cycle operating at full working capacity. Same as Dump Spotter idle.

On the contrary, the excavator idle and soil truck queue approximated to zero, referring to their almost full working capacity, and they are match in the quantitative relationship. Thus, we can increase both excavator and trucks in soil excavating cycle to reduce the total time, and check the waiting length to match their quantitative relationship.

6.3. Sensitivity Analysis and Regression Analysis

We assume the initial length for the main elements influence the total time and ignore the adjustment for the working efficiency for equipment.

Also, we assume that both the on-site and off-site parking space can accommodate up to 24 trucks for soil, meaning the system can hold a maximum of 25 trucks for soil in total.

Table 7. Different Scenario Results

Process. Cycle	Soil Loading Cycle			Soil Dump	Last Arrival(mins) - Domain
	FEL Idle	Dump Spotter	Soil Truck Queue	Dozer Idle	
Description	N - Excav	Affect dumping	N-Truck for soil	for soil compacting	
1	4	2	18	4	47,440.44
2	8	4	24	6	34,425.47
3	5	3	24	6	35,369.06
4	5	3	24	5	35,372.16
5	5	3	24	6	35,308.40
6	5	3	24	7	35,349.30
7	6	3	24	7	34,826.43
8	7	3	24	7	34,827.23
9	6	3	24	8	34,822.45
10	6	4	24	7	34,539.56
11	7	5	24	7	34,383.18
12	8	6	24	7	34,357.39
13	8	6	24	5	35,154.29

Process. Cycle	Soil Loading Cycle			Soil Dump	Last Arrival(mins) - Domain
	FEL Idle	Dump Spotter	Soil Truck Queue	Dozer Idle	
Description	N - Excav	Affect dumping	N-Truck for soil	for soil compacting	
14	7	5	24	5	35,162.95
15	7	6	24	5	35,174.11

We took the trials and tests, and recorded results from the 15 scenarios in Table 7.

In the construction theory, the total time is limited by the number of trucks. And we verify that by calculating and analysing the correlation matrix using Pearson's r method in Figure 16.

The correlation results shows that the truck for soil has the smallest negative correlation, as close to -1. The negative correlation suggests an inverse relationship between the lasting time and trucks, indicating that as trucks increases, the lasting time tends to decrease significantly. That matches the conclusion of the theory.

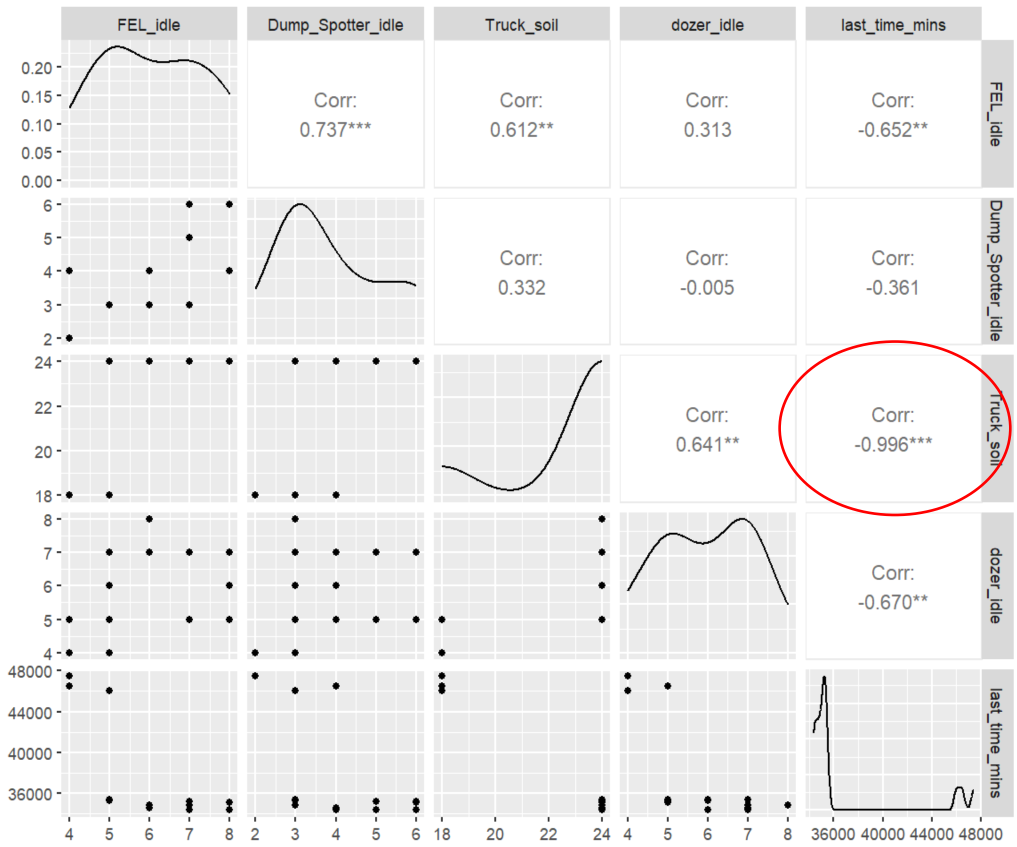

Figure 16. Pearson's Correlation Coefficients and Plots for Excavation Cycle

Besides, considering the correlation of FEL and dozer for soil spreading are similar, the transportation capacity for soil and the speed of soil disposal both impact the overall completion time of the project.

6.4. Optimization

Consider striking a balance between reserving some surplus capacity for equipment and maximizing their utilization to the fullest, to obtain the minimum working time, we take the scenario No.10 as the optimal choice, and the project can be completed within 74 working days with 8hrs-per-day working efficiency.

More details about the statistical report from the software are in 10.3 Appendix III: Statistic Report for optimal Scenario.

Table 8. The Optimal Scenario Inputs

Process.	Soil Loading Cycle			Soil Dump	Last Arrival(mins) -
Cycle	FEL Idle	Dump Spotter	Soil Truck Queue	Dozer Idle	Domain
10	6	4	24	7	34,539.56

Table 9. The Optimal Scenario Results

Element Name	**Average Length**	**Standard Deviation**	**Maximum Length**	**Current Length**	**Average Wait Time**
Abandoned Trees	21.662	168.371	2,000.000	0.000	373.988
Dozer Idle	1.916	1.146	7.000	7.000	3.006
Dump Spotter Idle	1.406	1.032	4.000	4.000	2.204
Excavator Idle	1.112	1.046	6.000	6.000	1.736
Soil Avail.	10,719.667	6,203.714	21,479.000	0.000	16,824.517
Soil Dumped	0.012	0.116	3.000	0.000	0.020
Soil Truck Queue	0.134	0.765	24.000	24.000	0.191

7. Concluding decision and rationale

The project lasting mainly depends on the hauling and return time of truck in the soil cycle, specifically it is limited to the number of trucks. However, we cannot expand the trucks without the consideration of cost, site capacity and queuing space.

Thus, if want to reduce the lasting time, increase the number of trucks within the limited site space, while adjusting the number of excavators to align with it for improved operational efficiency.

8. Development in the future

In this study, we collected data during the non-flood period.

To improve the applicability of the model, we suggest considering the limitation in truck haul and return time as the flooding. In the rainy season, use a parameter to relate flood depth (precipitation intensity) to vehicle speed, showing a reduction in the truck hauling and return time.

If the precipitation intensity is less than a certain standard, the overall process is not affected. On the contrary, if the precipitation is above a certain level, the parameter is introduced into the truck cycle, showing an increase in the project lasting time.

Thus, even that the data is collected in non-flooding period, the model still can apply in the flooding season and the whole year by considering an additional parameter related to the weather changes, instead of modifying the distribution description of the model set-ups.

9. Lessons learned from the study

1) Integer the real process into a simulation model

If there are several resources in the project, it needs to consider the transformation between resources using density relationship or other quantitative functions.

When the corresponding relationship of resource and element is not one-by-one, it is necessary to use the *CON* and *GEN* function to relate. Otherwise, the number relationship between trees and trucks will be unjustifiable.

2) Statistical analysis and distribution test

Assisting with the code or other statistical software, we can analyze the statistical properties of data with an easier method. And it is necessary to do a normalization for the original data ranging into [0,1], as the limitation of the code functions.

Also, the correlation analysis is helpful to quantify the relationship, but it needs more the trials and tests work to collect data.

3) Productivity analytics in the whole process

From this study, we have a better understanding of a whole process required for a project. There are multiple steps involved in the process, each of which can be a key point in limiting the efficiency of the work. By using different parameters and simulating the improvement scenarios, we will be able to better identify the problem areas in practice in the future.

In the optimisation, we found that the process was largely controlled by the number of trucks, and tested the optimisation using a methodology of controlling a single variable. However, in the optimisation tests we did not increase the number of trucks unlimitedly, as there were also constraints of site, money, and queuing.

10. Appendix

10.1. Appendix I: Derivation for probability distribution

1. Define the number of bins and plot corresponding histograms

Take the calculating code for excavator-truck loading cycle time as an example.

```
# histogram of columns ##
# NO. of bins (columns) is defined by the Sturges' rule
no_bins <- 1 + 3.3 * log10(length(cycle_time))
no_bins <- round(no_bins) ; no_bins
bin_break<-seq(6,10, by=2/3)
```

2. Beta distribution

```
# normalized by min-max method
# ranging [0,1], cannot plot with normal without inverse transition
min_max <- function(x){
  y<-(x - min(x, na.rm=TRUE))/(max(x,na.rm=TRUE) - min(x, na.rm=TRUE))
  return (y)
}

nor_cyclets <- min_max(cycle_time)

# Estimate the shape and scale parameters of a beta distribution: ebeta()
library(EnvStats)
ebeta(nor_cyclets) # estimated   shape1 = 1.75; shape2 = 2.45; Estimation Method: mle

## Calculate mean and variance of the data.
mu <- mean(nor_cyclets)
var <- var(nor_cyclets)
# calculations will only work if the variance is less than the mean*(1-mean)
## Define function to estimate parameters of a beta distribution.
estBetaParams <- function(mu, var) {
  alpha <- ((1 - mu) / var - 1 / mu) * mu ^ 2
  beta <- alpha * (1 / mu - 1)
  return(params = list(alpha = alpha, beta = beta))
}

## Apply function to your data.
estBetaParams(mu, var)    # alpha [1] 1.675;   $beta [1] 2.345
shape1_ alpha <-1.675
shape2_beta<-2.345

# randomly generate value within beta distribution.
```

```
theoretical_beta <- rbeta(length(nor_cyclets), shape1_beta, shape2_beta)
```

3. Gamma distribution

```
fit_gamma <- fitdist(cycle_time, "gamma")
fit_gamma
# estimated shape = 79.9095; estimated rate =10.423;
shape1_ga<- fit_gamma$estimate[1]
rate_ga<- fit_gamma$estimate[2]
ga_time_estimated <- rgamma(30, shape = fit_gamma$estimate[1], rate = fit_gamma$estimate[2])
```

4. Normal distribution

```
nor_mean = mean(cycle_time)
nor_stdev = sd(cycle_time)
nor_count_estimated<-rnorm(30, mean = nor_mean , sd = nor_stdev)
```

10.2. Appendix II: Goodness of fit tests

1. Visual Inspection – Q-Q plot

```
# Normal ##
qqplot(cycle_time, nor_count_estimated, col="orange",
        xlab="Observation Quantile", ylab="Theorectical Quantile",
        main = "Q-Q Plot for Normal Distribution")
# Add a reference line y=x
abline(0, 1, col = "red", lty = 1, lwd=2)

# Gamma ##
qplot(cycle_time, ga_time_estimated , col="green",
          xlab="Observation Quantile", ylab="Theorectical Quantile",
          main = "Q-Q Plot for Gamma Distribution")
# Add a reference line y=x
abline(0, 1, col = "red", lty = 1, lwd=2)

# Beta #####
theoretical_beta <- rbeta(length(nor_cyclets), shape1_beta, shape2_beta)
qqplot(nor_cyclets, theoretical_beta, col="blue",
        xlab="Observation Quantile", ylab="Theorectical Quantile", main = "Q-Q Plot
for Beta Distribution")
# Add a reference line y=x
abline(0, 1, col = "red", lty = 1, lwd=2)
```

2. Pearson's Chi-Square test

```
# Normal ##
obs_frequency<-hist_data$density*length(cycle_time) *(2/3)
```

```
obs_frequency
sum(obs_frequency)
# Expected frequencies for a normal distribution - p
hist_expe_nor<-hist(nor_time_estimated, breaks = bin_break, col = "lightpink",
                    xlab = "Estimated cycle time (mins)", ylab = " Count",
                    main = "Histogram for Normal - excavator ")
expected_freq_norm <- hist_expe_nor$density * (2/3)
sum(expected_freq_norm)
# Perform a chi-square goodness-of-fit test
chi_square_test <- chisq.test(obs_frequency, p = expected_freq_norm
*length(cycle_time) , rescale.p = T)
chi_square_test # p-vlaue =0.2993
```

 # Gamma##

```
hist_expe_gam<-hist(ga_time_estimated, breaks = bin_break, col = "lightpink",
                               xlab = "Estimated cycle time (mins)", ylab = " Count",
                               main = "Histogram for Gamma - excavator ", ylim=c(0,12))
expected_freq_gam <- hist_expe_gam$density * (2/3)
# Perform a chi-square goodness-of-fit test
chi_square_test <- chisq.test(obs_frequency, p = expected_freq_gam
*length(cycle_time), rescale.p = T )
chi_square_test # p-vlaue = 0.8606
```

 # BETA##

```
# Set the number of bins and calculate bin edges
num_bins <- 6
bin_edges <- seq(0, 1, length.out = num_bins + 1)
# Calculate frequency
freq_cycle <- table(cut(nor_cyclets, breaks = bin_edges, include.lowest = TRUE))
freq_cycle
freq_estimated <- diff(pbeta(bin_edges, shape1 = shape1_beta, shape2 = shape2_beta))
# test
chi_square_test <- chisq.test(freq_cycle, p = freq_estimated)
chi_square_test # p-value =0.55
```

 # aggregate estimated data into histogram #####

```
ggplot() + geom_histogram(data=excavator, aes(x = cycle_time, y = ..density..*(2/3)),
                          bins=6, fill = "lightblue", color = "black",
                          binwidth =2/3, boundary=6) +
  ggtitle("Histogram for excavator - Gamma Estimated") +
  geom_line(data = df_gamma, aes(x = x, y = y),
                 col = "blue", linewidth=2)    +
  labs(x = "Excavator Cycle Time (mins)", y = "Density")+
```

```
scale_x_continuous(breaks = seq(6, 10, by = 2/3),
                                labels = scales::number_format(digits = 2))
```

10.3. Appendix III: Statistic Report for optimal Scenario

Statistics Report

Date: 2024-04-13

Project: Model

Scenario: Optimal Scenario

Run: 1 of 1

Non-Intrinsic Statistic

Element Name	Mean Value	Standard Deviation	Observation Count	Minimum Value	Maximum Value
initial plan (Termination Time)	34,536.502	0.000	1.000	34,536.502	34,536.502
Forklift offload (InterArrivalTime)	8.205	4.290	99.000	0.248	16.945
Loader-Truck load timbers (InterArrivalTime)	0.408	1.997	1,999.000	0.000	16.473
Spread soil (InterArrivalTime)	1.568	1.194	21,999.000	0.000	8.769
Timber cutting (InterArrivalTime)	0.372	0.348	1,999.000	0.000	3.140
Truck dumping soil (InterArrivalTime)	1.568	1.187	21,999.000	0.000	9.284
Truck hauling timber (InterArrivalTime)	8.194	4.293	99.000	0.202	16.473

Intrinsic Statistic

Element Name	Mean Value	Standard Deviation	Minimum Value	Maximum Value	Current Value
Abandoned Trees (PercentNonempty)	0.022	0.145	0.000	1.000	0.000
Dozer idle (PercentNonempty)	0.902	0.298	0.000	1.000	1.000
Dump Spotter idle (PercentNonempty)	0.785	0.411	0.000	1.000	1.000
Excavator idle (PercentNonempty)	0.661	0.473	0.000	1.000	1.000
Forklift idle (PercentNonempty)	0.997	0.052	0.000	1.000	1.000
Loader idle (PercentNonempty)	1.000	0.000	1.000	1.000	1.000
Lumberjack idle (PercentNonempty)	0.978	0.145	0.000	1.000	1.000
Offload queue (PercentNonempty)	0.000	0.007	0.000	1.000	0.000
Soil avail (PercentNonempty)	0.999	0.037	0.000	1.000	0.000
Soil dumped (PercentNonempty)	0.011	0.103	0.000	1.000	0.000

Soil Truck queue (PercentNonempty)	0.094	0.292	0.000	1.000	1.000
Timber stockpile (PercentNonempty)	0.023	0.149	0.000	1.000	0.000
Truck dump queue (PercentNonempty)	0.043	0.203	0.000	1.000	0.000
Truck queue for timber (PercentNonempty)	0.977	0.151	0.000	1.000	1.000

Counters

Element Name	Final Count	Production Rate	Average Interarrival	First Arrival	Last Arrival
Counter trucks	22,000.000	0.637	1.568	43.010	34,534.688
Counter trucks for timbers	100.000	0.118	8.202	32.012	843.980

Waiting Files

Element Name	Average Length	Standard Deviation	Maximum Length	Current Length	Average Wait Time
Abandoned Trees	21.486	167.661	2,000.000	0.000	371.029
Dozer idle	1.918	1.144	7.000	7.000	3.009
Dump Spotter idle	1.409	1.027	4.000	4.000	2.209
Excavator idle	1.109	1.041	6.000	6.000	1.732
Forklift idle	1.982	0.153	2.000	2.000	10.535
Loader idle	59.942	1.074	60.000	60.000	23.489
Lumberjack idle	9.784	1.454	10.000	10.000	0.000
Offload queue	0.000	0.007	1.000	0.000	0.018
Soil avail	10,722.849	6,204.944	21,478.000	0.000	16,833.168
Soil dumped	0.011	0.109	3.000	0.000	0.018
Soil Truck queue	0.132	0.752	24.000	24.000	0.187
Timber stockpile	2.358	17.460	210.000	0.000	40.720
Truck dump queue	0.046	0.225	3.000	0.000	0.073
Truck queue for timber	78.080	12.173	80.000	80.000	0.673

*** FURTHER STATISTICS CAN BE OBTAINED FROM INDIVIDUAL ELEMENTS ***