Construction Operation Simulation

Lecture #10

Output analysis

Amin Alvanchi, PhD

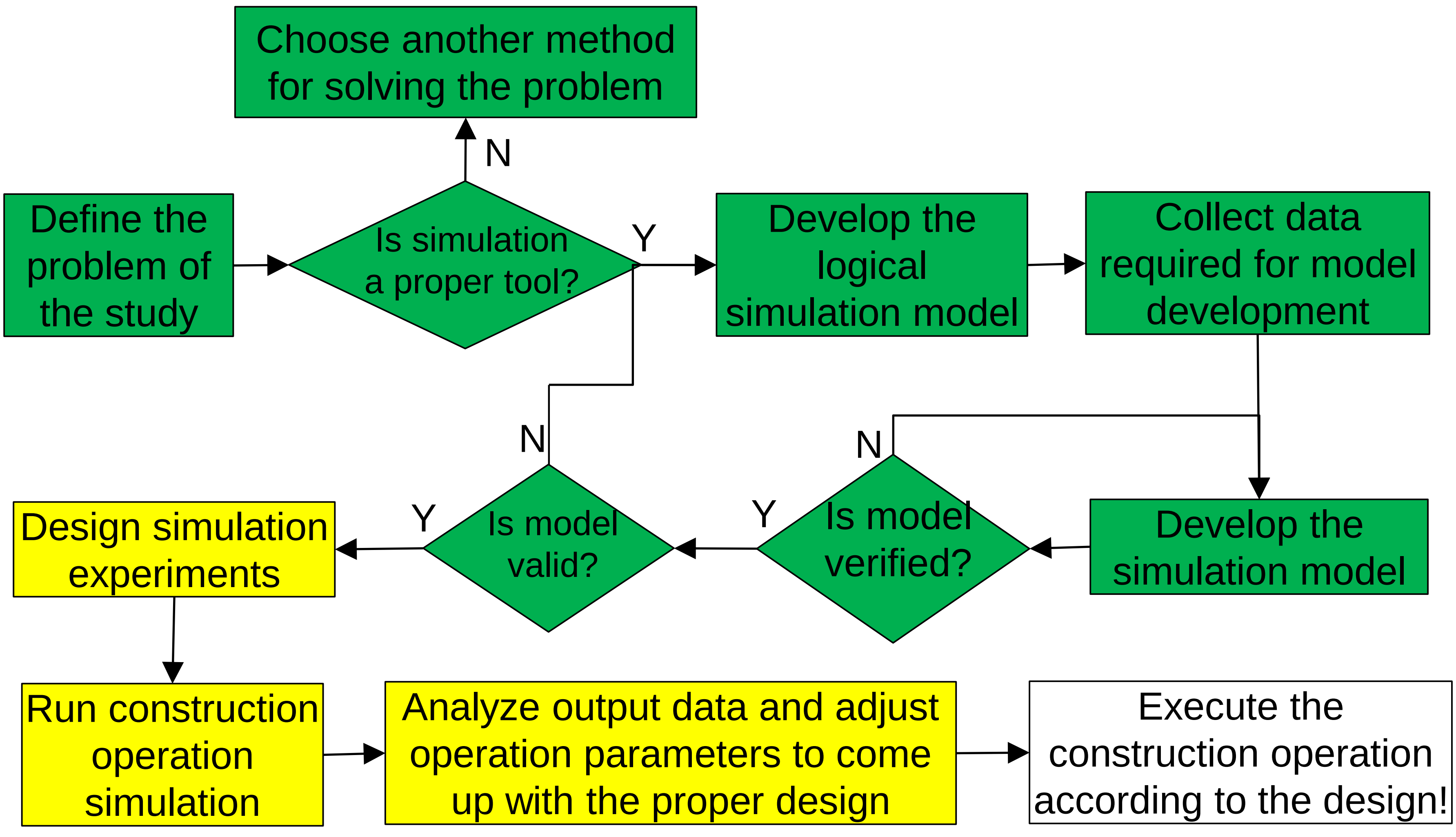
Construction Engineering and Management

LinkedIn Instagram WebPage

Outline

- Introduction
- Two types of simulation for output analysis
- Output parameters
- Confidence interval Vs Prediction interval
- Sample size estimation
- Initialization concerns

Introduction



Two types of simulation for output analysis

- Terminating
  - We simulate the model for a limited time or working hours after which simulation is terminated
  - Example, asphalt plant works from 7 am to 7 pm and we simulate it for 720 minutes to represent daily working hours
- Transient
  - We simulate the model for a long period of time where model reaches a steady condition
  - Example, steel fabrication shop with daily work of two shifts; every shift exactly continues previous shift’s uncompleted jobs, we may want to simulate 20 continuous shifts to be able to asses fabrication shop’s condition after passing initial condition of the work!

Output parameters and point estimation

- What are typical output parameters we need out of our simulation models?

Output parameters and point estimation

- Average queue length, maximum queue length, average waiting time, daily production rate and cycle time are some typical output parameters!
- Many of which are standard outputs reported in simulation programs and some are not and we need to calculate them through directly calculating them.
- How can we point estimate “average queue length”?

Output parameters and point estimation

- Average queue length, maximum queue length, average waiting time, daily production rate and cycle time are some typical output parameters!
- Many of which are standard outputs reported in simulation programs and some are not and we need to calculate them through directly calculating them.

- How can we point estimate "average queue length"? $$\hat{\phi} = \frac{1}{T_E}\int_0^{T_E} Y(t)\,dt$$

  Where Y(t) is length of queue at each point of time and T_E is the simulation duration.

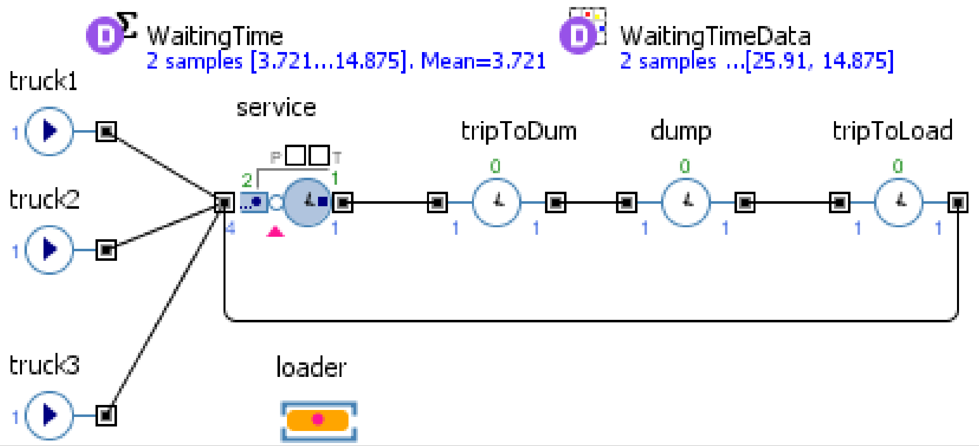
- How can we point estimate "average waiting time"? $$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

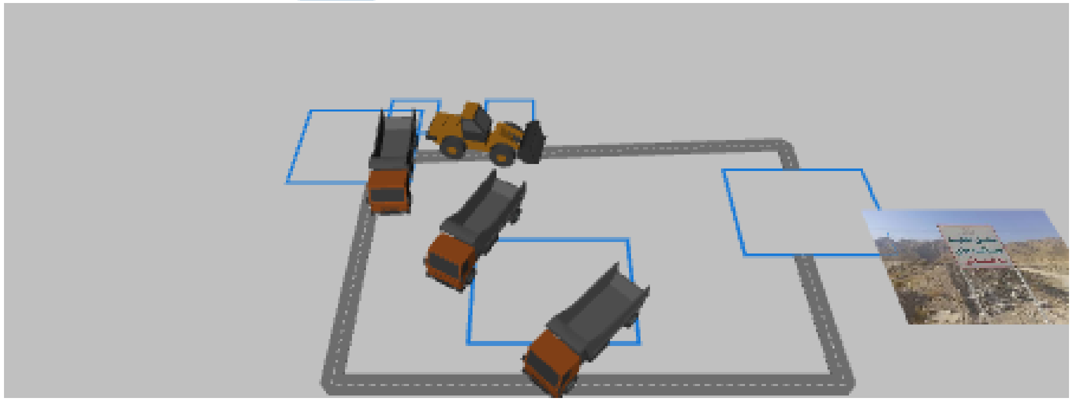
  Where Y_i is i^{th} entity's waiting time and n is number of entities served.

- How can we estimate "cycle time"?

Output parameters and point estimation

Example 1: Calculating entity's waiting time, earth moving example





Confidence interval Vs Prediction interval

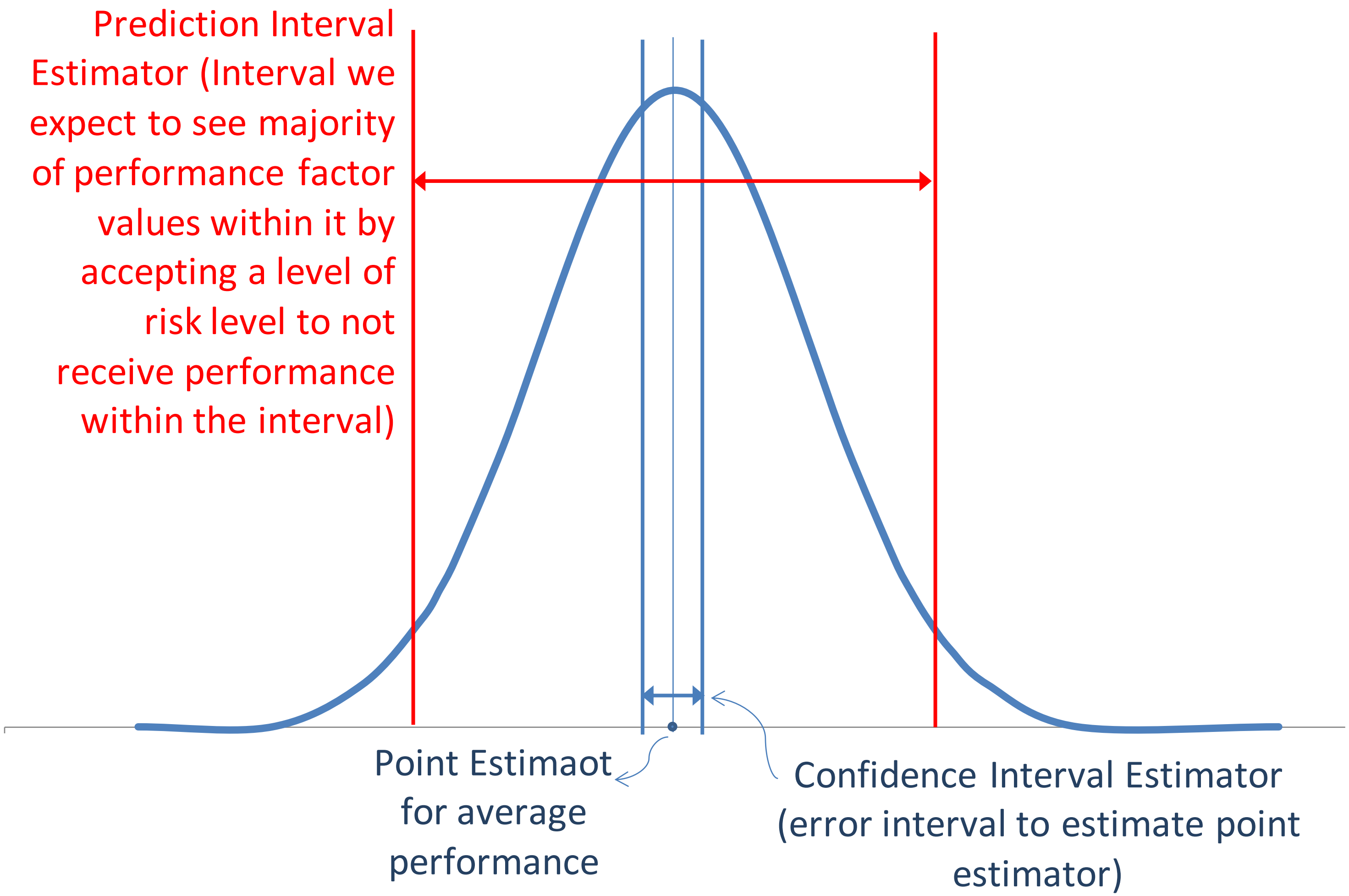
- **Confidence interval** corresponds to the *confidence level* (= 1 - error level) when point estimating an output result!

  Confidence level (= 1- error level) is the chance that the actual value of the point estimator of the output parameter (θ) belongs to the confidence interval.

- **Prediction interval** corresponds to the prediction level (= 1 - risk level) when receiving output results!

  Prediction level (= 1- risk level) represents the chance that observed output parameter results happen within the prediction interval in different model runs!

Confidence interval Vs Prediction interval



Confidence interval Vs Prediction interval

- For an output parameter of Y with R runs of our simulation model we have:

$$\bar{Y}_{..} = \sum_{i=1}^{R} \bar{Y}_{i.} / R$$

Where $\bar{Y}_{i.}$ is estimated average of parameter in each run and $\bar{Y}_{..}$ is the total estimated average. Let suppose:

$$S^2 = \frac{1}{R-1} \sum_{i=1}^{R} (Y_{i.} - \bar{Y}_{..})^2$$

The confidence interval of $\bar{Y}_{..}$ estimator is:

$$\bar{Y}_{..} \pm t_{\alpha/2,R-1} \frac{S}{\sqrt{R}}$$

As it is expected our confidence interval will be decreased (the error value of ϵ or $t_{\alpha/2,R-1} S / \sqrt{R}$ is reduced and converged to zero) by increasing number of observations (R).

Confidence interval Vs Prediction interval

Note: With large enough number of simulation runs (R) we will have our confidence interval converged to zero and value of our point estimator equal to exact value of the parameter: $t_{\alpha/2,R-1}S/\sqrt{R}$ =0 => $\bar{Y}_{..} = \theta$

$$S^2 = \frac{1}{R-1}\sum_{i=1}^{R}(Y_{i.} - \bar{Y}_{..})^2 = \sigma^2$$

According to the central limit theorem, with a large number of R (or our simulation runs) we can suppose $\bar{Y}_{..}$ has a normal distribution. In such case our prediction interval is:

$$\bar{Y}_{..} \pm t_{\alpha/2,R-1} S\sqrt{1+\frac{1}{R}}$$

The prediction interval can be then calculated as: $\theta \pm z_{\alpha/2}\sigma$

Confidence interval Vs Prediction interval

- **Example 2:** Cycle time of our truck in 8 different simulation runs are presented in below:

R	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13
1	22.03	29.70	38.80	30.28	25.13	21.61	28.42	33.53	33.98	34.31	28.04	28.08	24.83
2	28.48	28.37	27.12	29.50	31.26	34.72	33.16	31.17	30.65	32.23	33.71	34.13	29.80
3	29.15	29.96	32.67	35.52	30.17	29.50	28.31	31.97	32.46	31.54	33.88		
4	32.55	28.25	23.63	29.66	29.57	37.79	29.10	32.48	27.47	28.72	31.84	31.76	35.17
5	32.62	30.38	31.38	28.98	30.84	31.01	27.38	31.03	27.24	34.13	31.58	36.91	
6	31.88	29.45	35.71	22.53	31.66	38.73	30.17	28.60	30.48	37.01	33.57	34.17	28.21
7	32.17	31.69	36.32	36.37	32.78	26.80	24.76	27.31	30.17	29.69	30.99	30.83	
8	35.35	35.77	33.72	27.53	25.91	27.01	33.66	33.10	33.03	32.15	34.05	29.07	27.86

Confidence interval Vs Prediction interval

- **Example 2(cont'd):** Unbiased estimator for the average cycle time can be calculated as:

R	Ýi.	Ý..	S
1	29.13	30.91	0.79
2	31.10		
3	31.37		
4	30.61		
5	31.12		
6	31.71		
7	30.82		
8	31.40		

- The confidence interval of $\bar{Y}_{..}$ by accepting 5% error is:

$\bar{Y}_{..}$ ± t(2.5%, 7) * S / √R = 30.91 ± 2.36 * 0.79 / 2.82 = 30.91 ± 0.664 minute

- Prediction interval of observed results ($\bar{Y}_{i.}$) by accepting 5% risk is:

$\bar{Y}_{..}$ ± t(2.5%, 7) * S * √(1+1/R) = 30.91 ± 2.36 * 0.79 * 1.14 = 30.91 ± 2.14 minute

Confidence interval Vs Prediction interval

- **Empirical prediction interval:** When you have K number of output results:
  - Sort output results in an ascending manner
  - Determine risk level you can accept (e.g., $\alpha=5\%$)
  - For two-tail interval, our prediction interval forms by removing first and last ($\alpha/2 * K$) results
  - For one-tail interval, our prediction result forms by removing last ($\alpha * K$) results

Example: With 500 number of results observed in our 500 simulation runs, by accepting two-tail risk level of 10%, when we have all results sorted, our prediction interval forms by removing first and last 25 results (interval includes sorted result number 26 to 475). For one tail risk, our prediction result includes results 1 to 450.

Sample size estimation

- As a part of our output analysis, before start running our simulation model, we need to determine how many simulation runs fulfill our confidence level requirement!

Suppose we require a confidence interval for point estimator of $\bar{Y}_{..}$ with absolute error value of ϵ compared to actual parameter value (θ) and confidence level of 1-α:

$$P(|\bar{Y}_{..} - \theta| < \epsilon) \geq 1 - \alpha$$

This gives: $t_{\alpha/2,R-1} S_0 / \sqrt{R} \leq \epsilon$

$$\Rightarrow R \geq \left(\frac{t_{\alpha/2,R-1} S_0}{\epsilon} \right)^2$$

(Where **R** is number of runs to be estimated, $\mathbf{S_0}$ is an initial estimation of σ calculated based on initial number of runs of $\mathbf{R_0}$.)

Sample size estimation

In general we have: $t_{\alpha/2,R-1} \geq Z_{\alpha/2}$

in addition with large number of R, $t_{\alpha/2,R-1}$ converges to $Z_{\alpha/2}$; so we have:

$$R \geq \left(\frac{z_{\alpha/2} S_0}{\epsilon}\right)^2$$

Example:

t(0.025,7) = 2.36462 t(0.025,1000) = 1.96234

Z(0.025) = 1.95996

Example: In our last example of truck cycle time, what is minimum acceptable number of simulation runs, if maximum error size of 0.1 minute for the average of truck cycle time is required:

$R \geq (t_{(0.025,7)} * S_0 / 0.1)$^2 = (2.36 * 0.79 / 0.1)^2 = 347.6

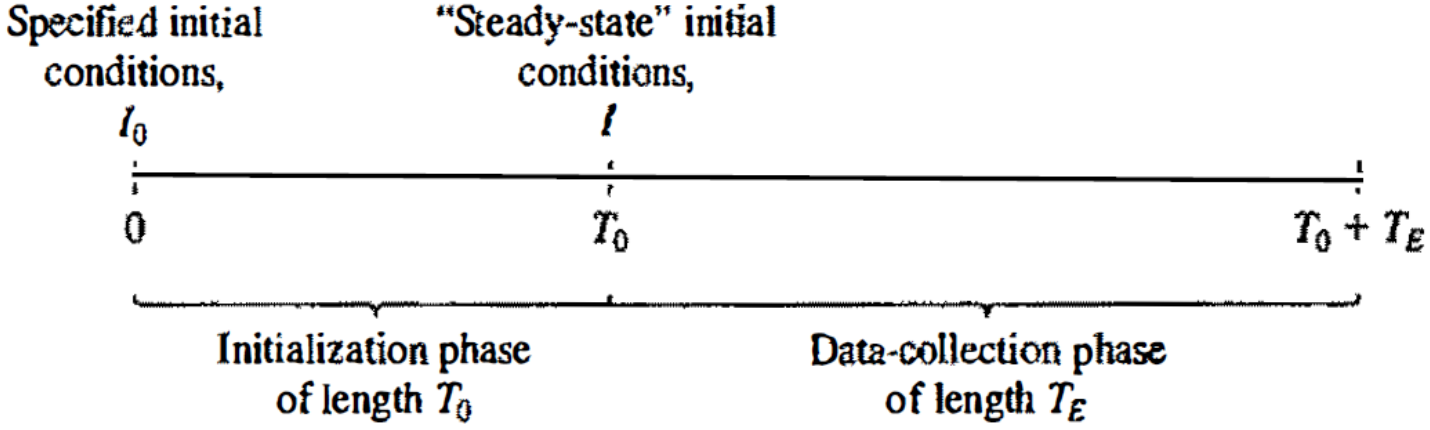
=> minimum number of 348 runs required

Initialization concerns

- In many simulation models initial system condition affects model outputs during initialization phase. Failure to properly considering effects of initial conditions of the system on the model outputs can cause bias calculations of point estimators!
- **Example:** in our pipeline construction example, first batches of pipe arrived to the construction site are welded with no delay since welding crew is idle. Pipe queue and waiting time may grow as welding crew get busy serving previously arrived pipes. If simulation model does not continue long enough, pipe waiting time estimators will be calculated bias.
- For transient models, as they are run for a long duration, for point-estimating output parameters it is important to separate point estimation in initialization state and steady state.
- For terminating models it is important to set initial values of the model close to the actual initial state of the system!

Initialization concerns

- Initialization phase in transient models:
  - One way for eliminating mal-effects of initial conditions in transient models is to continue simulation for long period of time
  - Another way for removing mal-effects of initialization phase is to divide simulation runs into two parts; initialization phase and steady phase. Data collection and point-estimator calculation will be done during the steady phase.



Initialization concerns

- Initialization phase in terminating models:
  - Initial condition is a part of point estimator calculations, since in this type of models, system opens and closes (is terminated) repeatedly after specific period of time (e.g., 10 hours on daily basis). In many cases effects of initial condition last for a significant portion of system time in each period and contributes in actual value of the output parameters.
  - Considering no one in line and idle resources for the initial condition of the system might not be the case for many systems!
  - Study of historical data on possible initial conditions of the system is the best way for correctly capturing the initial conditions and their effects.
  - Asking system experts is another way of capturing initial conditions.
  - Another way for capturing initial conditions, usually in absence of other credible ways, is start running the model with an arbitrary initial condition and then continue running it to reach the steady phase, different model status during the steady phase can be used as initial conditions!

Output analysis example

Example 3: Calculating entity's cycle time, earth moving example

Remember our earthmoving example with 3 x 10-tonne trucks, 1 loader, no limitation in number of dumping sites, working hours from 7 am to 7 pm and following activity durations estimated:

Loading: Triang(8,10,13)(minutes)

Trip to dumping site: U(3, 7) (minutes)

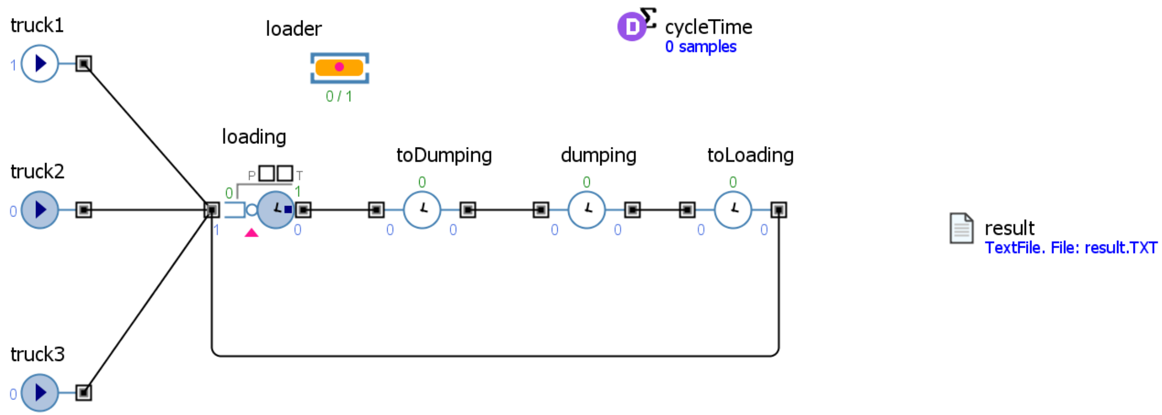
Dumping 2 minutes

Trip from dumping: Uniform (3, 6) (minutes)

Morning arrival of each truck is from 6:50 to 7:20

Output analysis example

Example 3: Calculating entity's cycle time, earth moving example



- We want to answer following questions:

  - How many runs do we need for calculating cycle time with less than 0.01 minute error while accepting confidence level of 95%?

  - What is prediction interval with 90% of chance?

Output analysis example

- I set an initial simulation run number of 10 to be able to calculate initial S!
- Results are exported to a file. In this way I can easily read them from Excel and do the calculations! We need to run the model more than 6161 times!

	Yi.	Y..	S
1	31.12	31.37	0.347
2	30.90		
3	31.19		
4	32.07		
5	31.56		
6	31.15		
7	31.37		
8	31.78		
9	31.29		
10	31.24		

Pilot calculations done in Excel

t(2.5%,9) = T.INV.2T(5%,9) = 2.26

R >= (2.26 * 0.347 /0.01)^2 = 6161

Output analysis example

- Montecarlo simulation experiment is used to run 6200 replications of the model. The resulst are presented in below:

Exported cycle time actual result from AnyLogic

	Yi.	Y..	S
1	31.23	31.36	0.435
2	31.30		
3	31.06		
4	31.16		
5	30.81		
6	31.57		
7	31.77		
8	31.72		
9	31.88		
10	31.24		
11	31.47		
12	32.01		
13	31.05		
14	31.94		
15	31.44		
16	31.15		
17	[illegible]		

t (.025, 6199) =	1.96
Confidence interval of 5% =	0.01
Prediction interval of 5% =	0.85

After class practice

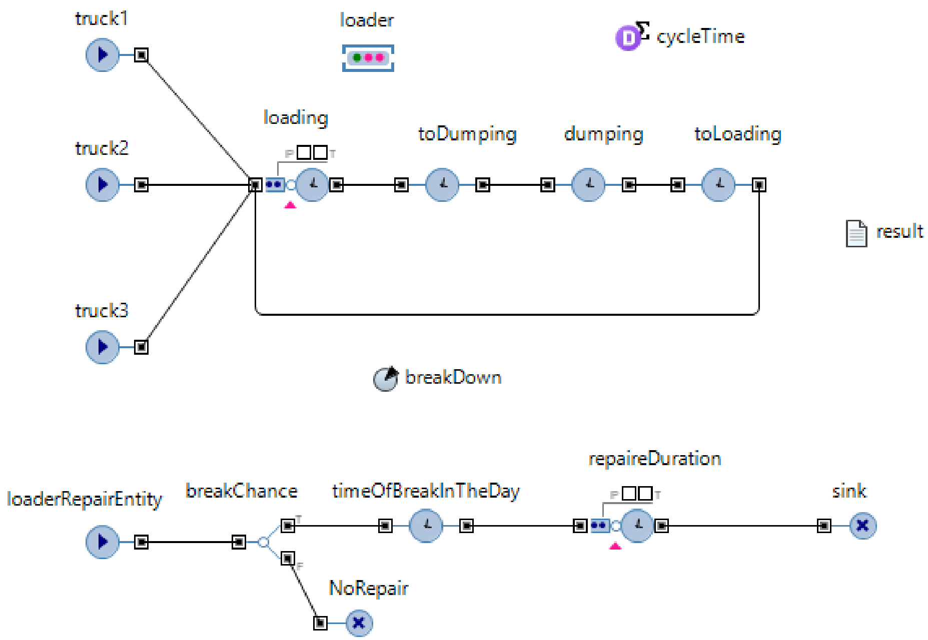
Remember our asphalt plant example. In this case we have:

- Working hours from 7 am to 7 pm
- Time between trucks arrival is distributed exponentially with mean of 12 min
- There 3 different trucks of 8, 10 and 12 tonnes respectively with the chance of 10%, 60% and 30% arriving to the plant
- Loading time has a triangular distribution with the average minimum of 8, maximum of 15 and most likely of 11 minutes
- How many runs do we need for calculating daily production tonnage with less than 1 tonne error while accepting confidence level of 99%?
- What is prediction interval with 95% of chance?

Breakdown model

Example 4: In our last earthmoving example, suppose that:

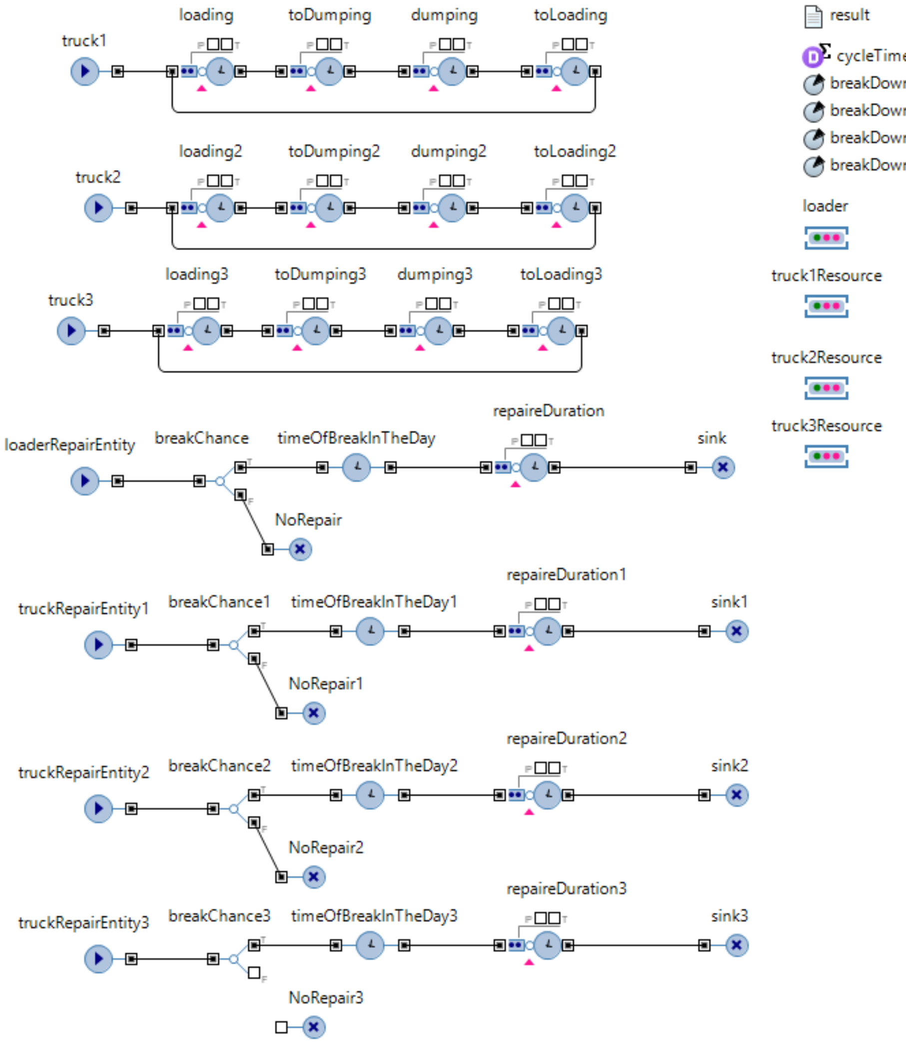
- 1- There is a chance of 1% that loader breaks during a working day! When loader is broken, repairing it takes 1 to 2 hours!



Breakdown model

Example 4:

- 2- There is a chance of 2% for every truck to be broken during a working day! When a truck is broken, repairing it takes 1 to 2 hours!



Reference

- Banks, J., Carson, J.S., Nelson, B.L. And Nicol D.M. (2004) “Discrete event simulation” Prentice Hall, ISBN: 0131446797., chapter 11.

Thank you!