Check for updates

METHOD ARTICLE

Improving Fleet Size Estimation for Factory Logistics Under Demand, Speed, and Efficiency Uncertainty

[version 1; peer review: 2 approved with reservations]

Dionysis Balser-Altanis[1,2], Giorgos Altanis[1]

[1]BILBERRY Private Company, Athens, 11141, Greece
[2]School of Electrical and Computer Engineering, National Technical University of Athens, Zografou, Attica, 15772, Greece

v1 **First published:** 25 Mar 2025, **5**:79
https://doi.org/10.12688/openreseurope.19000.1
Latest published: 25 Mar 2025, **5**:79
https://doi.org/10.12688/openreseurope.19000.1

Abstract

In material handling systems, estimating the required fleet size of automated guided vehicles (AGVs) is a significant task when designing processes to ensure efficient demand fulfillment. Analytic fleet size estimation methods typically assume that key input variables – such as AGV speed, loading / unloading times, material transportation demand between stations, efficiency and availability coefficients – are deterministic averages, overlooking the variability found inherent in real-world operational conditions. In contrast, we investigate the potential benefits of modeling key parameters as random variables with known means and variances, to account for this uncertainty.

The core of the method lies in expanding functions of random variables in terms of deviations from their mean values. Taking expectations, the expansions are expressed in terms of the central moments and mixed moments. The process is facilitated by the introduction of two linear operators, which streamline the development of the method. We thus derive second-order approximations for the expected value and variance of the required number of AGVs, providing more accurate and reliable fleet size estimates. The effectiveness of this approach is demonstrated through Monte Carlo simulations, comparing the results with empirical data from several case studies.

The method has been implemented in the "Agent Optimization" software system, developed as part of the *Grow your manufacturing business (Better Factory)* Horizon 2020 project.

Plain Language Summary

In factories and warehouses, determining the right number of automated guided vehicles (AGVs) is essential to meet demand efficiently. Traditional methods typically assume fixed averages for

Open Peer Review

Approval Status ? ?

	1	2
version 1 25 Mar 2025	? view	? view

1. **Nebojsa Bojovic**, University of Belgrade, Belgrade, Serbia

2. **Cabell Hodge**, National Renewable Energy Laboratory, Lakewood, USA

Any reports and responses or comments on the article can be found at the end of the article.

factors like speed, load times, and demand, ignoring the variability found in real-life operations. This study takes a different approach by treating these factors as random, allowing us to better capture uncertainty.

We use a mathematical method to estimate the average and variability of the AGVs required. Our approach simplifies the process using special tools and provides more reliable results. Testing with computer simulations shows that our estimates align well with real-world data.

The method has been implemented in the "Agent Optimization" software, developed in the Better Factory project.

Keywords
Automated Guided Vehicles (AGV) Fleet Sizing, Analytic models, Factory Logistics, Moment-based Expansion, Material Handling Systems, Expected Value Estimation, Variance Estimation



This article is included in the Horizon 2020 gateway.



This article is included in the Transport gateway.

Corresponding author: Giorgos Altanis (giorgos.altanis@bilberry.gr)

Author roles: Balser-Altanis D: Conceptualization, Formal Analysis, Visualization, Writing – Review & Editing; **Altanis G**: Conceptualization, Formal Analysis, Methodology, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant number 951813(Grow your manufacturing business (Better Factory)).

Copyright: © 2025 Balser-Altanis D and Altanis G. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Balser-Altanis D and Altanis G. **Improving Fleet Size Estimation for Factory Logistics Under Demand, Speed, and Efficiency Uncertainty [version 1; peer review: 2 approved with reservations]** Open Research Europe 2025, **5**:79 https://doi.org/10.12688/openreseurope.19000.1

First published: 25 Mar 2025, **5**:79 https://doi.org/10.12688/openreseurope.19000.1

Introduction

Material handling systems using automated guided vehicles (AGVs) are commonly used in manufacturing plants, warehouses, and other facilities (Le-Anh & de Koster, 2006). An important problem in these systems is determining the minimum *number of AGVs*, or *fleet* size, necessary to satisfy the effective demand of the logistics process. Several approaches have been proposed to tackle this problem: analytic models (Egbelu, 1987; Fitzgerald, 1985; Groover, 2008), discrete-event simulation studies (Gobal & Kasilingam, 1991; Touran & Taher, 1988; Valmiki *et al.*, 2018), network optimization models (Vis *et al.*, 2001), mixed integer optimization models (Erdogan & Tasan, 2023; Rajotia *et al.*, 1998), heuristic and metaheuristic models (Chawla *et al.,* 2018; Kavakeb *et al.,* 2016; Şenaras *et al.,* 2023; Vivaldini *et al.,* 2016), and regression and other statistical approaches (Arifin & Egbelu, 2000; Beamon & Chen, 1998; Johnson & Brandeau, 1996). Recently, fleet sizing problems have been studied in the context of unmanned aerial vehicles (UAVs), as demonstrated by (Hadas & Figliozzi, 2024).

In this paper we extend the calculus – based or analytic fleet sizing models by treating their parameters as random variables with unknown distributions, characterized only by the first moments (mean and variance). Consequently, the fleet size itself becomes a random variable. By leveraging Taylor series expansions (a well-known technique in statistics, underlying the *Delta method* (Casella & Berger, 2002)), we derive approximations for both the mean and variance of the fleet size, offering increased accuracy compared to traditional deterministic methods.

This approach bears resemblance to foundational ideas in probability and statistics. For instance, asymptotic expansions such as the Edgeworth and Gram-Charlier series (Johnson *et al.,* 1994; Wallace, 1958) are widely used to refine approximations for random variable distributions using higher-order moments. These techniques provide corrections to the normal distribution, effectively capturing non-normal features like skewness and kurtosis. Similarly, our method systematically expands functions of random variables in terms of their deviations from their means. Taking expectations, the resulting expansions are expressed in terms of the central moments, enabling second-order approximations tailored specifically to the problem of fleet sizing in material handling systems.

We apply our method to the model presented in (Groover, 2008; Körösi *et al.,* 2022), which initially motivated this research. We demonstrate its validity by directly comparing the analytically derived estimates of mean and variance of the fleet size with results obtained from Monte Carlo simulations.

The method has been implemented in the 2nd version of *Agent Optimization*, a software system developed within the *Grow your manufacturing business (Better Factory)* Horizon 2020 project. Information about the software implementation can be found in Appendix D.

Static deterministic analytic models for the fleet sizing problem

(Egbelu, 1987) describes four analytic models for the (homogeneous) fleet size estimation problem. The parameters of the models are the number of stations (work centers), the number of loaded trips required between the stations (the *flow demand*), the distances between stations, the time length of the period or shift, the AGV speed, the time to load and time to unload an AGV, the average efficiency of vehicle after accounting for breakdowns, and in one particular model the blocking time factor and the idle time factor. All models compute a *total distance* as a sum of the distances between stations weighted by the flow demand between the stations; what mainly distinguishes the models is the way they handle empty-agent trips. Total distance is then divided by the AGV speed and may be adjusted by the efficiency, blocking time, or idle time factors, before adding the loading and unloading times to obtain an *ideal cycle time,* which is finally divided by the *available time* (from which perhaps a battery charge time has been subtracted) to yield the desired fleet size.

The model presented in (Groover, 2008) and later in (Körösi *et al.,* 2022), which is what we concentrate in this research on, is quite similar. The model parameters are collected in Table 1:

Table 1. Model parameters used in (Groover, 2008) and in the present study.

Parameter	Description	Domain
$\mathcal{G}(\mathcal{V},\mathcal{E})$	A *directed graph* where $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ is the set of vertices, corresponding to the layout stations, and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of directed edges corresponding to the direct links between stations, such that $(v_i, v_j) \notin \mathcal{E}$ for $i = j$ (i.e., no reflexive loops or self-loops are allowed).	

Parameter	Description	Domain
D_e	The length of the link (direct path) $e = (v_i, v_j) \in \mathcal{E}$, starting at station i and ending at station j.	$D_e > 0$
F_e	The average flow rate per hour for the link $e = (v_i, v_j) \in \mathcal{E}$, starting at station i and ending at station j.	$F_e \geq 0$
c	The *capacity* of an AGV.	$c > 0$
v	The average speed of an AGV.	$v > 0$
T_L	The average time required to load an AGV.	$T_L > 0$
T_U	The average time required to unload an AGV.	$T_U > 0$
A	The AGV *availability* coefficient. 1 means full availability.	$0 < A \leq 1$
F_t	The *traffic factor*. 1 means no block due to traffic exists.	$0 < F_t \leq 1$
E_w	The *operator efficiency* coefficient. 1 stands for the maximum level of efficiency.	$0 < E_w \leq 1$

Following (Körösi *et al.*, 2022), we define:

The *count of demanded deliveries per hour*:

$$w = \sum_{e \in \mathcal{E}} F_e \tag{1}$$

The average distance per loaded trip:

$$L_d = \frac{z}{cw} = \frac{1}{c} \cdot \frac{\sum_{e \in \mathcal{E}} F_e D_e}{\sum_{e \in \mathcal{E}} F_e} \tag{2}$$

where

$$z = \sum_{e \in \mathcal{E}} F_e D_e \tag{3}$$

The average distance per unloaded trip:

$$L_e = \sum_{\substack{e \in \mathcal{E} \\ F_e = 0}} D_e \tag{4}$$

The *available time per hour per vehicle*:

$$T_{av} = 60 A F_t E_w \tag{5}$$

The *total cycle time per delivery per vehicle*:

$$T_c = T_L + \frac{L_d}{v} + T_U + \frac{L_e}{v} \tag{6}$$

The *ideal cycle time per delivery per vehicle*:

$$T_{ic} = w T_c \tag{7}$$

Then the estimate of the number of vehicles necessary to meet the demand is given by:

$$\boxed{N = \frac{T_{ic}}{T_{av}}} \tag{8}$$

This model provides the mathematical background for the *Agent Optimization module* used in the Horizon 2020 *Grow your manufacturing business (Better Factory)* project (Körösi *et al.*, 2022).

The impact of variability

An important characteristic of the class of models under discussion is that they treat their parameters as *average values*, without however mentioning, let alone exploiting, their statistical essence. However, ignoring the fact that the variables in question are random variables is not without risks; for the sake of argument, assume that the distribution of the computed fleet size is close to a normal distribution; then the average value of the fleet size roughly covers the demand of one half of the days. Moreover, as we will soon demonstrate, exploiting the statistical nature of the problem allows for even better estimate of the average necessary fleet size, and provides an approximate quantification of the uncertainty in this estimation.

In all models mentioned above, f_{ij} or F_{ij} represents the material transportation demand between stations i and j. Observe that all models contain products of the form $f_e\hat{T}$ or $F_e\hat{T}$, where $\hat{T}$ is a time variable (for example, in all models the loading time T_L is multiplied by demand terms), and ratios of the form $\Sigma_{e\in\mathcal{E}}\ F_e/v$ or $\Sigma_{e\in\mathcal{E}}\ f_eD_e/v$. It is therefore important to study such expressions, focusing on the randomness of the variables involved.

Product of two random variables

From the definition of *covariance* of two random variables we immediately get the expression for the mean value of their product:

$$E[XY] = \mu_X\mu_Y + \sigma_{X,Y} \tag{9}$$

This means that the deterministic approximation using average values is exact only when the variables are uncorrelated.

Ratio of two random variables

The situation is quite different in the case of the ratio of two random variables, because the expected value of the inverse of a random variable X does not equal the inverse of its expected value $1/E[X]$.

To illustrate this point, we perform a series of tests (simulations), presented in Appendix C.

Discussion

From the above presentation we can conclude that the operations of multiplication and division have distinctly different behavior. The mean value of the product $Z = XY$ of two random variables differs by the product of their mean values only if these variables are correlated. On the contrary, the mean value of the ratio $Z = X/Y$ of two random independent variables X and Y is mainly influenced by the denominator variance, and it can differ significantly from the simple ratio of the mean values of the random variables X and Y.

Very important is also the range of the support of the denominator Y: if the support is not bounded away from 0, the ratio distribution will have some very large values, which will affect both the skewness (asymmetry) and the kurtosis (fat tails or existence of many outliers) of the ratio distribution.

This means that while the variability of quantities such as the total demand will induce a comparable variability to the number of required vehicles N, leaving the mean value unaffected, non-negligent variability and / or skewness of the AGV speed v or some of the efficiency factors A, F_t, E_w can influence both the variance, and the mean value of N.

Analytical estimations based on low-order perturbations

Next, we derive first- and second-order approximations to the mean $E[f(X,Y)]$, and variance $Var(f(X,Y))$ of a sufficiently differentiable function $f(X,Y)$.

Notation and useful identities. Expansions and the operators $\mathcal{S}$ and $\mathcal{K}$

Partial derivatives

As usual, g_u will denote the partial derivative of a function g with respect to its argument u: $g_u \triangleq \partial g/\partial u$.

Deviations from mean, and their products

Let

$$\epsilon_U = U - \mu_U \tag{10}$$

denote the deviation of a random variable U from its mean.

Some identities related to the expectations of deviations and their products will be used (where we assume that the expectations exist and are finite):

$$E[\epsilon_U] = E[U - \mu_U] = 0 \tag{11}$$

$$E[\epsilon_U^2] = Var(U) = \sigma_U^2 \tag{12}$$

$$E[\epsilon_U \epsilon_V] = Cov(U,V) = \sigma_{U,V} \tag{13}$$

$$E[\epsilon_U^3] = \gamma_U \sigma_U^3 \tag{14}$$

$$E[\epsilon_U^4] = k_U \sigma_U^4 \tag{15}$$

where: $\mu_U = E[U]$, $\sigma_U^2 = Var(U)$, γ_U and k_U are respectively the *mean or expected value*, the *variance*, the *skewness* and the *kurtosis* of the distribution of U, and $\sigma_{U,V} = Cov(U,V)$ is the *covariance* of the joint distribution of U and V.

If, in particular, U and V are *independent*, then for all $j = 1,2, \ldots$:

$$E[\epsilon_U \, \epsilon_V^j] = E[\epsilon_U] E[\epsilon_V^j] = 0 \tag{16}$$

Central moments, mixed central moments, and their order

For all integers $i \geq 2$, the i–th central moment of U is the expected value of the i-th power of ϵ_U:

$$\mu_U^{(i)} = E[\epsilon_U^i] = \int_{-\infty}^{+\infty} \epsilon_U^i \, p_U(u) \, du \tag{17}$$

where $p_U(u)$ is the probability density function (p.d.f.) of the distribution of U.

The exponent i is called the *order* of the moment.

Likewise, for all integers $i, j \geq 1$, the (i, j)–th mixed central moment is the expected value of $\epsilon_U^i \, \epsilon_V^j$:

$$\mu_{U,V}^{(i,j)} = E[\epsilon_U^i \, \epsilon_V^j] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \epsilon_U^i \, \epsilon_V^j \, p_{U,V}(u,v) \, du \, dv \tag{18}$$

where $p_{U,V}(u,v)$ is the *joint probability density function* of the joint distribution of U and V.

The sum $i + j$ is called the *(total) order* of the mixed moment.

Order of approximation, approximation of expectations, and remainder. The operators $\mathcal{S}$ and $\mathcal{K}$.

Suppose we want to approximate a function $f(U,V)$ by a polynomial $\phi(x, y, z)$ of degree i in some variables x, y, z, and denote this by the notation $\simeq_i$. Using the Bachmann–Landau asymptotic notation:

$$f(U,V) \simeq_i \phi(x,y,z) \Leftrightarrow f(U,V) - \phi(x,y,z) = \mathcal{O}(\| [x,y,z]^T \|^{i+1}) \tag{19}$$

The variables, with respect to which this approximation is performed, will generally be clear from the context, but they will occasionally be mentioned in the text for clarity. The integer i is the *order of approximation*.

A common choice of variables is $(x, y) = (\epsilon_U, \epsilon_V)$, in which case $\phi(x, y)$ is the truncated Taylor series of $f(U,V)$ of order i around (μ_U, μ_V), and the expected value of the right-hand side of (19) becomes:

$$\begin{aligned} E[\mathcal{O}(\| [\epsilon_U, \epsilon_V]^T \|^{i+1})] &= \sum_{\substack{m,n \in N, \\ m+n \geq i+1}} \alpha_{m,n} \, E[\epsilon_U^m \, \epsilon_V^n], \; \alpha_{m,n} \in \mathbb{R} \\ &= \mathcal{O}(E[\epsilon_U^m \, \epsilon_V^n]), \; with \; m+n = i+1, \, m,n \in \mathbb{N} \end{aligned} \tag{20}$$

Therefore, **a suitable set of variables for approximating expectations is the set of moments of the distributions p_U and p_V of a specific order, and the mixed moments of their joint distribution $p_{U,V}$ of the same (total) order.**

The term $\mathcal{O}(E[\epsilon_U^m\,\epsilon_V^n])$ in (20) will be called the *remainder* of the expansion. It consists of terms that are of order $i+1$ in the standard deviations and mixed moments:

$$\mathcal{R}_i \stackrel{\text{def}}{=} \mathcal{O}(E[\epsilon_U^m\,\epsilon_V^n]),\ with\, m+n=i,\ m,n\in\mathbb{N} \tag{21}$$

We will denote by $\mathcal{S}_i(f)$ or $\mathcal{S}_i f$ the i-th order expansion of a function f in terms of the standard deviations and mixed moments of its arguments:

$$\boxed{\mathcal{S}_i\big(f(U,V)\big)=\mathcal{S}_i f(U,V)=\left(\sum_{m=0}^{i}\sum_{n=0}^{i-m}\alpha_{m,n}\,\sigma_U^m\,\sigma_V^n\right)+\left(\sum_{m=1}^{i}\sum_{n=1}^{i-m}\beta_{m,n}\,E[\epsilon_U^m\,\epsilon_V^n]\right)} \tag{22}$$

for specific values of the coefficients $\alpha_{m,n}$ and $\beta_{m,n}$, with

$$f(U,V)-\mathcal{S}_i f(U,V)=\mathcal{O}(E[\epsilon_U^m\,\epsilon_V^n])=\mathcal{R}_{i+1},\ \ m,n\in\mathbb{N},\ m+n=i+1 \tag{23}$$

That is, $\mathcal{S}_i f(U,V)$ is the sum of a multivariate polynomial of degree i in the standard deviations σ_U, σ_V, and a linear combination of the mixed moments of total order up to i.

When the more formal notation $\mathcal{S}_i$ is not necessary, we may use the simpler (but not so powerful) notation $\simeq_i$, introduced earlier.

It is easy to verify that $\mathcal{S}_i$ is linear. Consider any two functions f_1, f_2 of the random variables U and V expanded according to (23), which we may rewrite as:

$$f=\mathcal{S}_i f+\mathcal{R}_{i+1} \tag{24}$$

Let:

$$\mathcal{S}_i f_1(U,V)=\left(\sum_{m=0}^{i}\sum_{n=0}^{i-m}\alpha_{m,n}^{(1)}\sigma_U^m\sigma_V^n\right)+\left(\sum_{m=1}^{i}\sum_{n=1}^{i-m}\beta_{m,n}^{(1)}E[\epsilon_U^m\epsilon_V^n]\right)$$
$$\mathcal{S}_i f_2(U,V)=\left(\sum_{m=0}^{i}\sum_{n=0}^{i-m}\alpha_{m,n}^{(2)}\sigma_U^m\sigma_V^n\right)+\left(\sum_{m=1}^{i}\sum_{n=1}^{i-m}\beta_{m,n}^{(2)}E[\epsilon_U^m\epsilon_V^n]\right)$$

Then, with $\kappa,\ \lambda\in\mathbb{R}$:

$$\begin{aligned}\kappa f_1+\lambda f_2 &= \kappa\left(\left(\sum_{m=0}^{i}\sum_{n=0}^{i-m}\alpha_{m,n}^{(1)}\,\sigma_U^m\,\sigma_V^n\right)+\left(\sum_{m=1}^{i}\sum_{n=1}^{i-m}\beta_{m,n}^{(1)}\,E[\epsilon_U^m\epsilon_V^n]\right)+\mathcal{R}_{i+1}\right)\\ &\quad+\lambda\left(\left(\sum_{m=0}^{i}\sum_{n=0}^{i-m}\alpha_{m,n}^{(2)}\,\sigma_U^m\,\sigma_V^n\right)+\left(\sum_{m=1}^{i}\sum_{n=1}^{i-m}\beta_{m,n}^{(2)}\,E[\epsilon_U^m\epsilon_V^n]\right)+\mathcal{R}_{i+1}\right)\\ &=\kappa\left(\sum_{m=0}^{i}\sum_{n=0}^{i-m}\alpha_{m,n}^{(1)}\,\sigma_U^m\,\sigma_V^n\right)+\kappa\left(\sum_{m=1}^{i}\sum_{n=1}^{i-m}\beta_{m,n}^{(1)}\,E[\epsilon_U^m\epsilon_V^n]\right)\\ &\quad+\lambda\left(\sum_{m=0}^{i}\sum_{n=0}^{i-m}\alpha_{m,n}^{(2)}\,\sigma_U^m\,\sigma_V^n\right)+\lambda\left(\sum_{m=1}^{i}\sum_{n=1}^{i-m}\beta_{m,n}^{(2)}\,E[\epsilon_U^m\epsilon_V^n]\right)\\ &\quad+\mathcal{R}_{i+1}\end{aligned}$$

Therefore,

$$\begin{aligned}\mathcal{S}_i\left(\kappa f_1+\lambda f_2\right) &= \kappa\left(\sum_{m=0}^{i}\sum_{n=0}^{i-m}\alpha_{m,n}^{(1)}\,\sigma_U^m\,\sigma_V^n\right)+\kappa\left(\sum_{m=1}^{i}\sum_{n=1}^{i-m}\beta_{m,n}^{(1)}\,E[\epsilon_U^m\epsilon_V^n]\right)\\ &\quad+\lambda\left(\sum_{m=0}^{i}\sum_{n=0}^{i-m}\alpha_{m,n}^{(2)}\,\sigma_U^m\,\sigma_V^n\right)+\lambda\left(\sum_{m=1}^{i}\sum_{n=1}^{i-m}\beta_{m,n}^{(2)}\,E[\epsilon_U^m\epsilon_V^n]\right)\\ &= \kappa\left(\sum_{m=0}^{i}\sum_{n=0}^{i-m}\alpha_{m,n}^{(1)}\,\sigma_U^m\,\sigma_V^n+\sum_{m=1}^{i}\sum_{n=1}^{i-m}\alpha_{m,n}^{(1)}\,E[\epsilon_U^m\epsilon_V^n]\right)\\ &\quad+\lambda\left(\sum_{m=0}^{i}\sum_{n=0}^{i-m}\alpha_{m,n}^{(2)}\,\sigma_U^m\,\sigma_V^n+\sum_{m=1}^{i}\sum_{n=1}^{i-m}\beta_{m,n}^{(2)}\,E[\epsilon_U^m\epsilon_V^n]\right)\\ &= \kappa\mathcal{S}_2 f_1+\lambda\mathcal{S}_2 f_2\end{aligned}$$

We will also make use of the operator $\mathcal{K}_i$:

$$\mathcal{K}_i f=\mathcal{K}_i(f)=\begin{cases}\mathcal{S}_0(f), & \text{if } i=0\\ \mathcal{S}_i(f)-\mathcal{S}_{i-1}(f), & \text{if } i\geq 1\end{cases} \tag{25}$$

Importantly, the operator S_i is *idempotent*:

$$\mathcal{S}_i(\mathcal{S}_i f)=\mathcal{S}_i f \tag{26}$$

For a very simple proof of this (rather obvious) statement: by definition of the $\mathcal{S}_i$ operator and the remainder $\mathcal{R}_{i+1}$, for all $i \geq 0$:

$$f=\mathcal{S}_i f+\mathcal{R}_{i+1} \tag{27}$$

Applying the linear operator $\mathcal{S}_i$ to both sides of (26) yields:

$$\mathcal{S}_i f=\mathcal{S}_i(\mathcal{S}_i f+\mathcal{R}_{i+1})=\mathcal{S}_i(\mathcal{S}_i f)+\underbrace{\mathcal{S}_i(\mathcal{R}_{i+1})}_{0}=\mathcal{S}_i(\mathcal{S}_i f)$$

By extension, the operator $\mathcal{K}_i$ is also linear.

Mean value approximations

The Taylor expansion of $f(X,Y)$ around the point $\theta=(\mu_X,\mu_Y)$ is:

$$f(X,Y)=f(\theta)+f_x(\theta)\epsilon_X+f_y(\theta)\epsilon_Y+\frac{1}{2}(f_{xx}(\theta)\epsilon_X^2+2f_{xy}(\theta)\epsilon_X\epsilon_Y+f_{yy}(\theta)\epsilon_Y^2)+\mathcal{O}(\|[\epsilon_X,\epsilon_Y]^T\|^3) \tag{28}$$

Applying the linear expectation operator yields:

$$\begin{aligned}E[f(X,Y)] &= E[f(\theta)]+f_x(\theta)\underbrace{E[\epsilon_X]}_{0}+f_y(\theta)\underbrace{E[\epsilon_Y]}_{0}+\frac{1}{2}f_{xx}(\theta)\underbrace{E[\epsilon_X^2]}_{Var(X)}+f_{xy}(\theta)\underbrace{E[\epsilon_X\epsilon_Y]}_{Cov(X,Y)}+\frac{1}{2}f_{yy}(\theta)\underbrace{E[\epsilon_Y^2]}_{Var(Y)}\\ &\quad+E[\mathcal{O}(\|[\epsilon_X,\epsilon_Y]\|^T)^3]\\ &= f(\theta)+\frac{1}{2}f_{xx}(\theta)\sigma_X^2+f_{xy}(\theta)\sigma_{X,Y}+\frac{1}{2}f_{yy}(\theta)\sigma_Y^2+\mathcal{R}_3\end{aligned} \tag{29}$$

where the *remainder* $\mathcal{R}_3$ consists of terms that are of third order in the standard deviations and mixed moments:

$$\mathcal{R}_3=\mathcal{O}\left(\sigma_X^3,\sigma_Y^3,\mu_{X,Y}^{(2,1)},\mu_{X,Y}^{(1,2)}\right)$$

From (29), it is straightforward to get the approximations of the mean:

1st order approximation:

$$E[f(X,Y)] \simeq_1 f(\mu_X,\mu_Y) \tag{30}$$

2nd order approximation:

$$E[f(X,Y)] \simeq_2 f(\mu_X,\mu_Y)+\frac{1}{2}f_{xx}(\mu_X,\mu_Y)\sigma_X^2+f_{xy}(\mu_X,\mu_Y)\sigma_{X,Y}+\frac{1}{2}f_{yy}(\mu_X,\mu_Y)\sigma_Y^2 \tag{31}$$

Variance approximations

Since

$$Var(f(X,Y)) = E\left[\left(f(X,Y)\right)^2\right]-E[f(X,Y)]^2 \tag{32}$$

the Taylor expansion of $g(X,Y) = f(X,Y)^2$ is also needed. By the chain rule:

$$g_x = 2ff_x$$
$$g_y = 2ff_y$$

and

$$g_{xx} = 2(f_x^2+ff_{xx})$$
$$g_{xy} = 2f_xf_y+2ff_{xy}$$
$$g_{yy} = 2(f_y^2+ff_{yy})$$

Then,

$$\begin{aligned} f(X,Y)^2 &= g(X,Y)\\ &= g(\theta)+g_x(\theta)\epsilon_X+g_y(\theta)\epsilon_y\\ &\quad+\frac{1}{2}(g_{xx}(\theta)\epsilon_X^2+2g_{xy}(\theta)\epsilon_X\epsilon_Y+g_{yy}(\theta)\epsilon_Y^2)+\mathcal{O}(\|[\epsilon_X,\epsilon_Y]^T\|^3)\\ &= f(\theta)^2+2f(\theta)f_x(\theta)\epsilon_X+2f(\theta)f_y(\theta)\epsilon_Y\\ &\quad+\frac{1}{2}\Big(2\Big(f_x(\theta)^2+f(\theta)f_{xx}(\theta)\Big)\Big)\epsilon_X^2+2\Big(f_x(\theta)f_y(\theta)+f(\theta)f_{xy}(\theta)\Big)\epsilon_X\epsilon_Y\\ &\quad+\frac{1}{2}\Big(2\Big(f_y(\theta)^2+f(\theta)f_{yy}(\theta)\Big)\Big)\epsilon_Y^2\\ &\quad+\mathcal{O}(\|[\epsilon_X,\epsilon_Y]^T\|^3) \end{aligned} \tag{33}$$

Taking expectations:

$$\begin{aligned} E[f(X,Y)^2] = f(\theta)^2&+\Big(f_x(\theta)^2+f(\theta)f_{xx}(\theta)\Big)\sigma_X^2+2\Big(f_x(\theta)f_y(\theta)+f(\theta)f_{xy}(\theta)\Big)\sigma_{X,Y}\\ &+\Big(f_y(\theta)^2+f(\theta)f_{yy}(\theta)\Big)\sigma_Y^2+\mathcal{R}_3 \end{aligned} \tag{34}$$

By (32), to approximate $Var(f(X,Y))$ up to 2nd order, we must find a 2nd order approximation of $E[f(X,Y)]^2$:

$$\begin{aligned} E[f(X,Y)]^2 &= \left(f(\theta)+\frac{1}{2}f_{xx}(\theta)\sigma_X^2+f_{xy}(\theta)\sigma_{X,Y}+\frac{1}{2}f_{yy}(\theta)\sigma_Y^2+\mathcal{R}_3\right)^2\\ &= f(\theta)^2+f(\theta)f_{xx}(\theta)\sigma_X^2+2f(\theta)f_{xy}(\theta)\sigma_{X,Y}+f(\theta)f_{yy}(\theta)\sigma_Y^2+\mathcal{R}_3 \end{aligned} \tag{35}$$

Therefore,

$$\begin{aligned}
Var\big(f(X,Y)\big) &= E\Big[\big(f(X,Y)\big)^2\Big] - E[f(X,Y)]^2 \\
&= f(\theta)^2 + \Big(f_x(\theta)^2 + f(\theta)f_{xx}(\theta)\Big)\sigma_X^2 + 2\Big(f_x(\theta)f_y(\theta) + f(\theta)f_{xy}(\theta)\Big)\sigma_{X,Y} \\
&\quad + \Big(f_y(\theta)^2 + f(\theta)f_{yy}(\theta)\Big)\sigma_Y^2 \\
&\quad - \Big(f(\theta)^2 + f(\theta)f_{xx}(\theta)\sigma_X^2 + 2f(\theta)f_{xy}(\theta)\sigma_{X,Y} + f(\theta)f_{yy}(\theta)\sigma_Y^2 + \mathcal{R}_3\Big) \\
&= \Big(f(\theta)^2 - f(\theta)^2\Big) + \big(f_x(\theta)^2 + f(\theta)f_{xx}(\theta) - f(\theta)f_{xx}(\theta)\Big)\sigma_X^2 \\
&\quad + 2\Big(f_x(\theta)f_y(\theta) + f(\theta)f_{xy}(\theta) - f(\theta)f_{xy}(\theta)\Big)\sigma_{X,Y} \\
&\quad + \Big(f_y(\theta)^2 + f(\theta)f_{yy}(\theta) - f(\theta)f_{yy}(\theta)\Big)\sigma_Y^2 + \mathcal{R}_3 \\
&= f_x(\theta)^2\sigma_X^2 + 2f_x(\theta)f_y(\theta)\sigma_{X,Y} + f_y(\theta)^2\sigma_Y^2 + \mathcal{R}_3
\end{aligned}$$

Hence, we obtain a 2nd order approximation for $Var(f(X,Y))$:

$$Var(f(X,Y)) \simeq_2 f_x(\mu_X,\mu_Y)^2\sigma_X^2 + 2f_x(\mu_X,\mu_Y)f_y(\mu_X,\mu_Y)\sigma_{X,Y} + f_y(\mu_X,\mu_Y)^2\sigma_Y^2 \tag{36}$$

Higher-order approximations

The second-order approximations ignore the effect of higher-order moments (such as *skewness* and *kurtosis*) and mixed moments. Evaluating this effect in the general case requires approximating $E[f(X,Y)^2]$ by a fourth-degree polynomial which involves, apart from the moments of X and Y, the mixed moments of their joint distribution up to (total) order 4. In a practical setting, this may be too much to ask for.

Instead, we will compute higher-order approximations of the mean and the variance of the function $f(X,Y)$ for a few important – especially for the problem we are studying – cases.

Sum of correlated variables: $f(X,Y) = X + Y$

For the sake of completeness, we begin by stating the well – known results:

$$E[X+Y] = E[X] + E[Y] = \mu_X + \mu_Y \tag{37}$$

$$Var(X+Y) = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y} \tag{38}$$

Product of correlated variables: $f(X,Y) = XY$

Equation (12) gives the exact formula for the mean value of a product, and is repeated here for convenience:

$$E[XY] = \mu_X\mu_Y + \sigma_{X,Y}$$

The variance is given by (32):

$$Var(f(X,Y)) = E[f(X,Y)^2] - (E[f(X,Y)])^2$$

The Taylor series expansion of $f(X,Y) = XY$ around (μ_X, μ_Y) is:

$$\begin{aligned}
XY &= \mu_X\mu_Y + f_x\epsilon_X + f_y\epsilon_Y + f_{xy}\epsilon_X\epsilon_Y \\
&= \mu_X\mu_Y + \mu_Y\epsilon_X + \mu_X\epsilon_Y + \epsilon_X\epsilon_Y
\end{aligned} \tag{39}$$

from which the Taylor series of f^2 is obtained:

$$\begin{aligned}
(XY)^2 &= \mu_X^2\mu_Y^2 + \mu_Y^2\epsilon_X^2 + \mu_X^2\epsilon_Y^2 + \epsilon_X^2\epsilon_Y^2 \\
&\quad + 2\mu_X\mu_Y^2\epsilon_X + 2\mu_X^2\mu_Y\epsilon_Y + 4\mu_X\mu_Y\epsilon_X\epsilon_Y + 2\mu_Y\epsilon_X^2\epsilon_Y + 2\mu_X\epsilon_X\epsilon_Y^2
\end{aligned} \tag{40}$$

Therefore, and considering (11 – 16):

$$\begin{aligned} Var(XY) &= E[(XY)^2] - (E[XY])^2 \\ &= \mu_X^2\mu_Y^2 + \mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 + E[\epsilon_X^2\epsilon_Y^2] + 4\mu_X\mu_Y\sigma_{X,Y} + 2\mu_Y E[\epsilon_X^2\epsilon_Y] + 2\mu_X E[\epsilon_X\epsilon_Y^2] \\ &\quad - (\mu_X\mu_Y + \sigma_{X,Y})^2 \\ &= \mu_X^2\mu_Y^2 + \mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 + E[\epsilon_X^2\epsilon_Y^2] + 4\mu_X\mu_Y\sigma_{X,Y} + 2\mu_Y E[\epsilon_X^2\epsilon_Y] + 2\mu_X E[\epsilon_X\epsilon_Y^2] \\ &\quad - \mu_X^2\mu_Y^2 - \sigma_{X,Y}^2 - 2\mu_X\mu_Y\sigma_{X,Y} \\ &= \mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 + 2\mu_X\mu_Y\sigma_{X,Y} + 2\mu_Y E[\epsilon_X^2\epsilon_Y] + 2\mu_X E[\epsilon_X\epsilon_Y^2] + E[\epsilon_X^2\epsilon_Y^2] - \sigma_{X,Y}^2 \end{aligned} \tag{41}$$

Hence, the difference between the exact (4th order) variance of the product and its 2nd order approximation

$$Var(XY) \simeq_2 \mu_Y^2\sigma_X^2 + 2\mu_X\mu_Y\sigma_{X,Y} + \mu_X^2\sigma_Y^2$$

equals $2\mu_Y E[\epsilon_X^2\epsilon_Y] + 2\mu_X E[\epsilon_X\epsilon_Y^2] + E[\epsilon_X^2\epsilon_Y^2] - \sigma_{X,Y}^2$, which, in the case of independence, simplifies to $\sigma_X^2\sigma_Y^2$:

$$\begin{aligned} Var(XY) - \mathcal{S}_2 Var(X,Y) &= 2\mu_Y E[\epsilon_X^2\epsilon_Y] + 2\mu_X E[\epsilon_X\epsilon_Y^2] + E[\epsilon_X^2\epsilon_Y^2] - \sigma_{X,Y}^2 \\ &= \sigma_X^2\sigma_Y^2, \textit{ when } X,Y \textit{ independent} \end{aligned} \tag{42}$$

Inverse of a random variable: $f(X,Y) = 1/Y$

The Taylor series of $1/Y$ around $Y = \mu_Y \neq 0$ is given by:

$$\frac{1}{Y} = \frac{1}{\mu_Y} - \frac{1}{\mu_Y^2}\epsilon_Y + \frac{1}{\mu_Y^3}\epsilon_Y^2 - \frac{1}{\mu_Y^4}\epsilon_Y^3 + \frac{1}{\mu_Y^5}\epsilon_Y^4 + \mathcal{O}(\epsilon_Y^5)$$

from which, and if the distribution Y is bounded away from 0, the mean value approximation of fourth order is:

$$E\left[\frac{1}{Y}\right] \simeq_4 \frac{1}{\mu_Y} + \frac{1}{\mu_Y^3}\sigma_Y^2 - \frac{1}{\mu_Y^4}\gamma_Y\sigma_Y^3 + \frac{1}{\mu_Y^5}k_Y\sigma_Y^4 \tag{43}$$

Hence,

$$\begin{aligned} \mathcal{S}_4\left(\left(E\left[\frac{1}{Y}\right]\right)^2\right) &= \mathcal{S}_4\left(\frac{1}{\mu_Y} + \frac{1}{\mu_Y^3}\sigma_Y^2 - \frac{1}{\mu_Y^4}\gamma_Y\sigma_Y^3 + \frac{1}{\mu_Y^5}k_Y\sigma_Y^4 + \mathcal{R}_5\right)^2 \\ &= \mathcal{S}_4\left(\frac{1}{\mu_Y^2} + \frac{1}{\mu_Y^6}\sigma_Y^4 + 2\frac{1}{\mu_Y^4}\sigma_Y^2 - 2\frac{1}{\mu_Y^5}\gamma_Y\sigma_Y^3 + 2\frac{1}{\mu_Y^6}k_Y\sigma_Y^4 + \mathcal{R}_5\right) \\ &= \frac{1}{\mu_Y^2} + 2\frac{1}{\mu_Y^4}\sigma_Y^2 - 2\frac{1}{\mu_Y^5}\gamma_Y\sigma_Y^3 + \left(\frac{1+2k_Y}{\mu_Y^6}\right)\sigma_Y^4 \end{aligned} \tag{44}$$

For the variance approximation we expand $1/Y^2$:

$$\frac{1}{Y^2} = \frac{1}{\mu_Y^2} - \frac{2}{\mu_Y^3}\epsilon_Y + \frac{3}{\mu_Y^4}\epsilon_Y^2 - \frac{4}{\mu_Y^5}\epsilon_Y^3 + \frac{5}{\mu_Y^6}\epsilon_Y^4 + \mathcal{O}(\epsilon_Y^5)$$

so that

$$E\left[\frac{1}{Y^2}\right] \simeq_4 \frac{1}{\mu_Y^2} + \frac{3}{\mu_Y^4}\sigma_Y^2 - \frac{4}{\mu_Y^5}\gamma_Y\sigma_Y^3 + \frac{5}{\mu_Y^6}k_Y\sigma_Y^4 \tag{45}$$

Then,

$$\begin{aligned}
Var\left(\frac{1}{Y}\right) &= E\left[\frac{1}{Y^2}\right] - \left(E\left[\frac{1}{Y}\right]\right)^2 \\
&= \frac{1}{\mu_Y^2} + \frac{3}{\mu_Y^4}\sigma_Y^2 - \frac{4}{\mu_Y^5}\gamma_Y\sigma_Y^3 + \frac{5}{\mu_Y^6}k_Y\sigma_Y^4 \\
&\quad - \left(\frac{1}{\mu_Y^2} + 2\frac{1}{\mu_Y^4}\sigma_Y^2 - 2\frac{1}{\mu_Y^5}\gamma_Y\sigma_Y^3 + \frac{1+2k_Y}{\mu_Y^6}\sigma_Y^4\right) \\
&\quad + \mathcal{R}_5 \\
&= \left(\frac{1}{\mu_Y^2} - \frac{1}{\mu_Y^2}\right) + \left(\frac{3}{\mu_Y^4} - \frac{2}{\mu_Y^4}\right)\sigma_Y^2 + \left(-\frac{4}{\mu_Y^5} + \frac{2}{\mu_Y^5}\right)\gamma_Y\sigma_Y^3 \\
&\quad + \left(\frac{5}{\mu_Y^6}k_Y - \frac{1}{\mu_Y^6} - \frac{2}{\mu_Y^6}k_Y\right)\sigma_Y^4 + \mathcal{R}_5 \\
&= \frac{1}{\mu_Y^4}\sigma_Y^2 - \frac{2}{\mu_Y^5}\gamma_Y\sigma_Y^3 + \frac{3k_Y - 1}{\mu_Y^6}\sigma_Y^4 + \mathcal{R}_5
\end{aligned}$$

so that

$$Var\left(\frac{1}{Y}\right) \simeq_4 \frac{1}{\mu_Y^4}\sigma_Y^2 - \frac{2}{\mu_Y^5}\gamma_Y\sigma_Y^3 + \frac{3k_Y - 1}{\mu_Y^6}\sigma_Y^4 \tag{46}$$

Ratio of independent random variables: *f(X,Y) = X/Y*

From the previous results and the independence of X and Y it is straightforward to obtain the higher order approximations of the mean value:

$$E[f(X,Y)] = E\left[\frac{X}{Y}\right] = E[X]E\left[\frac{1}{Y}\right] \simeq_4 \mu_X\left(\frac{1}{\mu_Y} + \frac{1}{\mu_Y^3}\sigma_Y^2 - \frac{1}{\mu_Y^4}\gamma_Y\sigma_Y^3 + \frac{1}{\mu_Y^5}k_Y\sigma_Y^4\right) \tag{47}$$

The calculation of variance is performed in a similar manner:

$$\begin{aligned}
Var\left(\frac{X}{Y}\right) &= E\left[\left(\frac{X}{Y}\right)^2\right] - \left(E\left[\frac{X}{Y}\right]\right)^2 \\
&= E[X^2]E\left[\frac{1}{Y^2}\right] - \left(E[X]E\left[\frac{1}{Y}\right]\right)^2 \\
&= (\sigma_X^2 + \mu_X^2)E\left[\frac{1}{Y^2}\right] - \mu_X^2\left(E\left[\frac{1}{Y}\right]\right)^2 \\
&= \mathcal{S}_4\left((\sigma_X^2 + \mu_X^2)E\left[\frac{1}{Y^2}\right] - \mu_X^2\left(E\left[\frac{1}{Y}\right]\right)^2\right) + \mathcal{R}_5 \\
&= \mathcal{S}_4\left(\sigma_X^2 E\left[\frac{1}{Y^2}\right]\right) + \mu_X^2\mathcal{S}_4\left(E\left[\frac{1}{Y^2}\right]\right) - \mu_X^2\mathcal{S}_4\left(\left(E\left[\frac{1}{Y}\right]\right)^2\right) + \mathcal{R}_5
\end{aligned} \tag{48}$$

The expansions of the second and third term have already been computed (44, 45). The first term requires some attention:

$$\begin{aligned}
\mathcal{S}_4\left(\sigma_X^2 E\left[\frac{1}{Y^2}\right]\right) &= \mathcal{S}_4\left(\sigma_X^2\left(\frac{1}{\mu_Y^2} + \frac{3}{\mu_Y^4}\sigma_Y^2 - \frac{4}{\mu_Y^5}\gamma_Y\sigma_Y^3 + \frac{5}{\mu_Y^6}k_Y\sigma_Y^4 + \mathcal{R}_5\right)\right) \\
&= \frac{1}{\mu_Y^2}\sigma_X^2 + \frac{3}{\mu_Y^4}\sigma_X^2\sigma_Y^2
\end{aligned} \tag{49}$$

Therefore:

$$
\begin{aligned}
Var\left(\frac{X}{Y}\right) \simeq_4 & \frac{1}{\mu_Y^2}\sigma_X^2 + \frac{3}{\mu_Y^4}\sigma_X^2\sigma_Y^2 \\
& + \mu_X^2\left(\frac{1}{\mu_Y^2} + \frac{3}{\mu_Y^4}\sigma_Y^2 - \frac{4}{\mu_Y^5}\gamma_Y\sigma_Y^3 + \frac{5}{\mu_Y^6}k_Y\sigma_Y^4\right) \\
& - \mu_X^2\left(\frac{1}{\mu_Y^2} + 2\frac{1}{\mu_Y^4}\sigma_Y^2 - 2\frac{1}{\mu_Y^5}\gamma_Y\sigma_Y^3 + \frac{1+2k_Y}{\mu_Y^6}\sigma_Y^4\right) \\
= & \left(\frac{\mu_X^2}{\mu_Y^2} - \frac{\mu_X^2}{\mu_Y^2}\right) + \left(\frac{1}{\mu_Y^2}\sigma_X^2 + 3\frac{\mu_X^2}{\mu_Y^4}\sigma_Y^2 - 2\frac{\mu_X^2}{\mu_Y^4}\sigma_Y^2\right) + \left(-4\frac{\mu_X^2}{\mu_Y^5}\gamma_Y\sigma_Y^3 + 2\frac{\mu_X^2}{\mu_Y^5}\gamma_Y\sigma_Y^3\right) \\
& + \left(\frac{3}{\mu_Y^4}\sigma_X^2\sigma_Y^2 + 5\frac{\mu_X^2 k_Y}{\mu_Y^6}\sigma_Y^4 - \frac{\mu_X^2(1+2k_Y)}{\mu_Y^6}\sigma_Y^4\right) \\
= & \left(\frac{1}{\mu_Y^2}\sigma_X^2 + \frac{\mu_X^2}{\mu_Y^4}\sigma_Y^2\right) + \left(-2\frac{\mu_X^2}{\mu_Y^5}\gamma_Y\sigma_Y^3\right) + \left(\frac{3}{\mu_Y^4}\sigma_X^2\sigma_Y^2 + \frac{(3k_Y-1)\mu_X^2}{\mu_Y^6}\sigma_Y^4\right)
\end{aligned}
\tag{50}
$$

Summary of the basic approximations

Table 2 summarizes our results.

Table 2. Mean value and variance approximations for *X* + *Y*, *X*·*Y*, 1/*Y*, and *X*/*Y*.

	Z* = *X* + *Y	***Z* = *X*·*Y***	***Z* = 1/*Y***	***Z* = *X*/*Y*** ***X*,*Y* independent**
Mean value				
1st order	$\mu_X + \mu_Y$ **Exact**	$\mu_X\mu_Y$ **Exact when independent**	$\frac{1}{\mu_Y}$	$\frac{\mu_X}{\mu_Y}$
2nd order		1st order + $\sigma_{X,Y}$ **Exact**	1st order + $\frac{1}{\mu_Y^3}\sigma_Y^2$	1st order + $\frac{\mu_X}{\mu_Y^3}\sigma_Y^2$
3rd order			2nd order − $\frac{1}{\mu_Y^4}\gamma_Y\sigma_Y^3$	2nd order − $\frac{\mu_X}{\mu_Y^4}\gamma_Y\sigma_Y^3$
4th order			3rd order + $\frac{1}{\mu_Y^5}k_Y\sigma_Y^4$	3rd order + $\frac{\mu_X}{\mu_Y^5}k_Y\sigma_Y^4$
Variance				
2nd order	$\sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$ **Exact**	$\mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 + 2\mu_X\mu_Y\sigma_{X,Y}$	$\frac{1}{\mu_Y^4}\sigma_Y^2$	$\frac{1}{\mu_Y^2}\sigma_X^2 + \frac{\mu_X^2}{\mu_Y^4}\sigma_Y^2$
3rd order		2nd order + $2\mu_Y E[\epsilon_X^2\epsilon_Y] + 2\mu_X E[\epsilon_X\epsilon_Y^2]$	2nd order − $\frac{2}{\mu_Y^5}\gamma_Y\sigma_Y^3$	2nd order − $2\frac{\mu_X^2}{\mu_Y^5}\gamma_Y\sigma_Y^3$
4th order		3rd order + $E[\epsilon_X^2\epsilon_Y^2] - \sigma_{X,Y}^2$ **Exact**	3rd order + $\frac{3k_Y-1}{\mu_Y^6}\sigma_Y^4$	3rd order + $\frac{3}{\mu_Y^4}\sigma_X^2\sigma_Y^2 + \frac{(3k_Y-1)\mu_X^2}{\mu_Y^6}\sigma_Y^4$
Comments		When independent, the 4th order expansion becomes $\mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 + \sigma_X^2\sigma_Y^2$	Assumes finite skewness and kurtosis of the denominator *Y*	Assumes finite skewness and kurtosis of the denominator *Y*

Estimation of the fleet size statistics

We are now ready to apply our analysis to the fleet size problem. In what follows we work with the model (1 – 7) from (Groover, 2008); the method can also in principle be applied to the models (see Appendix B) from (Egbelu, 1987), although such an attempt is not presented in this work.

We begin by modeling the demand variability: for simplicity, we assume that all individual link demands F_e are *perfectly correlated* and all have the same relative standard deviation δ, so that they all fluctuate by the same percentage around their nominal value. We model this condition by:

$$F_e = \bar{F}_e(1+\Delta) \text{ for all } e \in \mathcal{E} \tag{51}$$

where Δ is a random variable having mean value 0 and standard deviation δ, so that indeed

$$Var(F_e) = \bar{F}_e^2 Var(\Delta) = \bar{F}_e^2 \delta^2 \tag{52}$$

For $e, e' \in \mathcal{E}$, $e \neq e'$:

$$\begin{aligned} Cov(F_e, F_{e'}) &= Cov\left(\bar{F}_e(1+\Delta), \bar{F}_{e'}(1+\Delta)\right) \\ &= \bar{F}_e \bar{F}_{e'} Cov(1+\Delta, 1+\Delta) \\ &= \bar{F}_e \bar{F}_{e'} \left[\underbrace{Cov(1,1+2\Delta)}_{0} + Cov(\Delta,\Delta) \right] \\ &= \bar{F}_e \bar{F}_{e'} Var(\Delta) \end{aligned} \tag{53}$$

so that the correlation between F_e and $F_{e'}$ is indeed

$$\rho_{F_e,F_{e'}} = \frac{Cov(F_e,F_{e'})}{\sigma_{F_e}\sigma_{F_{e'}}} = \frac{\bar{F}_e \bar{F}_{e'} Var(\Delta)}{\sqrt{Var(F_e)}\sqrt{Var(F_{e'})}} = \frac{\bar{F}_e \bar{F}_{e'} \delta^2}{\bar{F}_e \delta \bar{F}_{e'} \delta} = 1 \tag{54}$$

Under this assumption, and treating distances D_e as constants:

$$w = \sum_{e\in\mathcal{E}} F_e = \sum_{e\in\mathcal{E}} \bar{F}_e(1+\Delta) = (1+\Delta)\sum_{e\in\mathcal{E}} \bar{F}_e \tag{55}$$

and

$$z = \sum_{e\in\mathcal{E}} F_e D_e = \sum_{e\in\mathcal{E}} \bar{F}_e(1+\Delta) D_e = (1+\Delta)\sum_{e\in\mathcal{E}} \bar{F}_e D_e \tag{56}$$

Let us compute the mean and variance of these variables:

$$\mu_w = E\left[(1+\Delta)\sum_{e\in\mathcal{E}} \bar{F}_e\right] = \sum_{e\in\mathcal{E}} \bar{F}_e \, E[1+\Delta] = \sum_{e\in\mathcal{E}} \bar{F}_e \tag{57}$$

$$\sigma_w^2 = \left(\sum_{e\in\mathcal{E}} \bar{F}_e\right)^2 Var(1+\Delta) = \left(\sum_{e\in\mathcal{E}} \bar{F}_e\right)^2 \sigma_\Delta^2 = \left(\sum_{e\in\mathcal{E}} \bar{F}_e\right)^2 \delta^2 = \mu_w^2 \delta^2 \tag{58}$$

and

$$\mu_z = E\left[(1+\Delta)\sum_{e\in\mathcal{E}} \bar{F}_e D_e\right] = \sum_{e\in\mathcal{E}} \bar{F}_e D_e \cdot E[1+\Delta] = \sum_{e\in\mathcal{E}} \bar{F}_e D_e \tag{59}$$

$$\sigma_z^2 = \left(\sum_{e\in\mathcal{E}} \bar{F}_e D_e\right)^2 Var(1+\Delta) = \left(\sum_{e\in\mathcal{E}} \bar{F}_e D_e\right)^2 \sigma_\Delta^2 = \left(\sum_{e\in\mathcal{E}} \bar{F}_e D_e\right)^2 \delta^2 = \mu_z^2 \delta^2 \tag{60}$$

Then,

$$L_d = \frac{1}{c}\cdot\frac{z}{w} = \frac{1}{c}\cdot\frac{(1+\Delta)\mu_z}{(1+\Delta)\mu_w} = \frac{1}{c}\cdot\frac{\mu_z}{\mu_w} \tag{61}$$

Hence L_d is constant, as is L_e.

Next, we turn our attention to the ideal cycle time per delivery per vehicle:

$$T_{ic} = w\left(T_L + \frac{L_d}{v} + T_U + \frac{L_e}{v}\right) = w\left((T_L + T_U) + \frac{L_d + L_e}{v}\right) \tag{62}$$

We obtain in succession:

$$\mu_{T_L+T_U} = \mu_{T_L} + \mu_{T_U} \tag{63}$$

$$\sigma^2_{T_L+T_U} = \sigma^2_{T_L} + \sigma^2_{T_U} + 2Cov(T_L, T_U) = \mathcal{S}_2\sigma^2_{T_L+T_U} \tag{64}$$

$$\mu_{(L_d+L_e)/v} = \left(L_d + L_e\right)\mu_{1/v} \simeq_2 \frac{L_d + L_e}{\mu_v} + \frac{L_d + L_e}{\mu_v^3}\sigma_v^2 = \mathcal{S}_2\mu_{(L_d+L_e)/v} \tag{65}$$

$$\sigma^2_{(L_d+L_e)/v} \simeq_2 \left(L_d + L_e\right)^2\frac{\sigma_v^2}{\mu_v^4} = \mathcal{S}_2\sigma^2_{(L_d+L_e)/v} \tag{66}$$

Since $(L_d + L_e)/v$ is independent of $T_L + T_U$, from (6) we obtain:

$$\mu_{T_c} \simeq_2 \mu_{T_L+T_U} + \mathcal{S}_2\mu_{(L_d+L_e)/v} = \mathcal{S}_2\mu_{T_c} \tag{67}$$

$$\sigma^2_{T_c} \simeq_2 \sigma^2_{T_L+T_U} + \mathcal{S}_2\sigma^2_{(L_d+L_e)/v} = \mathcal{S}_2\sigma^2_{T_c} \tag{68}$$

Next, since $T_{ic} = wT_c$, **and** w, T_c are independent, and using the operators $\mathcal{K}_0$, $\mathcal{K}_2$:

$$\mu_{T_{ic}} \simeq_2 \mu_w \cdot \mathcal{S}_2\mu_{T_c} = \mathcal{S}_2\mu_{T_{ic}} \tag{69}$$

$$\begin{aligned}
\sigma^2_{T_{ic}} &= \sigma^2_w(\mathcal{S}_2\mu_{T_c} + \mathcal{R}_3)^2 + \mu^2_w(\mathcal{S}_2\sigma^2_{T_c} + \mathcal{R}_3) \\
&\overset{(58)}{=} \mu^2_w\delta^2(\mathcal{K}_0\mu_{T_c} + \mathcal{K}_2\mu_{T_c} + \mathcal{R}_3)^2 + \mu^2_w(\mathcal{K}_2\sigma^2_{T_c} + \mathcal{R}_3) \\
&= \mu^2_w\delta^2(\mathcal{K}_0\mu_{T_c})^2 + \mu^2_w\mathcal{K}_2\sigma^2_{T_c} + \mathcal{R}_3 \\
&\simeq_2 \mu^2_w\delta^2(\mathcal{K}_0\mu_{T_c})^2 + \mu^2_w\mathcal{K}_2\sigma^2_{T_c}
\end{aligned}$$

Notice that we retained only the 0th order terms from the approximation of μ_{T_C}, since the term $\mu^2_w\delta^2$ is the variance of the quantity w, thus already a 2nd order term, and keeping the full 2nd order approximation $\mathcal{S}_2\mu_{T_C}$ would contain higher order terms.

By (67), (63) and (65),

$$\mathcal{K}_0\mu_{T_c} = \mathcal{K}_0\{\mu_{T_L+T_U} + \mathcal{S}_2\mu_{(L_d+L_e)/v}\} = \mu_{T_L} + \mu_{T_U} + \frac{L_d + L_e}{\mu_v}$$

and so (since $\mathcal{S}_2\sigma_{T_c} = \mathcal{K}_2\sigma_{T_c}$):

$$\mathcal{S}_2\sigma^2_{T_{ic}} = \mu^2_w\delta^2\left(\mu_{T_L} + \mu_{T_U} + \frac{L_d + L_e}{\mu_v}\right)^2 + \mu^2_w\mathcal{S}_2\sigma^2_{T_c} \tag{70}$$

Consider now the available time per hour per vehicle:

$$T_{av} = 60AF_tE_w \tag{5}$$

Since all variables are independent, the approximations of T_{av} are easily obtained:

$$\mu_{T_{av}} = 60\mu_A\mu_{F_t}\mu_{E_w} = \mathcal{S}_2\mu_{T_{av}} = \boldsymbol{\mathcal{S}_0\mu_{T_{av}}} \tag{71}$$

$$\sigma^2_{T_{av}} \simeq_2 60^2(\mu_A^2\mu_{F_t}^2\sigma^2_{E_w} + \mu_A^2\sigma^2_{F_t}\mu^2_{E_w} + \sigma^2_A\mu^2_{F_t}\mu^2_{E_w}) = \mathcal{S}_2\sigma^2_{T_{av}} \tag{72}$$

The last equality in (71) means that the 0th-order approximation of $\mu_{T_{av}}$ is exact.

With $\mathcal{S}_2\mu_{T_{ic}}$, $\mathcal{S}_2\sigma^2_{T_{ic}}$, $\mathcal{S}_2\mu_{T_{av}}$, $\mathcal{S}_2\sigma^2_{T_{av}}$ given by (69) – (72), we can approximate the statistics of the fleet size N (given by (8)), taking into account that the variables defining the numerator are independent from those defining the denominator:

$$\mu_N \simeq \frac{\mathcal{S}_2\mu_{T_{ic}}}{\mu_{T_{av}}} + \frac{\mathcal{S}_2\mu_{T_{ic}}}{\mu^3_{T_{av}}}\mathcal{S}_2\sigma^2_{T_{av}} \tag{73}$$

$$\sigma^2_N \simeq \frac{1}{\mu^2_{T_{av}}}\mathcal{S}_2\sigma^2_{T_{ic}} + \frac{(\mathcal{S}_2\mu_{T_{ic}})^2}{\mu^4_{T_{av}}}\mathcal{S}_2\sigma^2_{T_{av}} \tag{74}$$

The expressions in (73), (74) are approximations based on the 2nd order approximations of the various quantities used. However, *they are not 2nd order approximations themselves*. The actual 2nd order approximations of μ_N and σ_N are (see Appendix A):

$$\mathcal{S}_2\mu_N \simeq \frac{\mathcal{S}_2\mu_{T_{ic}}}{\mu_{T_{av}}} + \frac{\mathcal{K}_0\mu_{T_{ic}}}{\mu^3_{T_{av}}}\mathcal{S}_2\sigma^2_{T_{av}} \tag{75}$$

$$\mathcal{S}_2\sigma^2_N \simeq \frac{1}{\mu^2_{T_{av}}}\mathcal{S}_2\sigma^2_{T_{ic}} + \frac{(\mathcal{K}_0\mu_{T_{ic}})^2}{\mu^4_{T_{av}}}\mathcal{S}_2\sigma^2_{T_{av}} \tag{76}$$

where the operator $\mathcal{K}_0$ returns the 0th order terms of its operand (see definition (26)).

From (69):

$$\begin{aligned}
\mu_{T_{ic}} &= \mu_w\mathcal{S}_2\mu_{T_c} + \mathcal{R}_3 \\
&\overset{(67)}{=} \mu_w\left(\mu_{T_L+T_U} + \mathcal{S}_2\mu_{(L_d+L_e)/v}\right) + \mathcal{R}_3 \\
&\overset{(63,65)}{=} \mu_w\left(\mu_{T_L} + \mu_{T_U} + \frac{L_d+L_e}{\mu_v} + \frac{L_d+L_e}{\mu^3_v}\sigma^2_v\right) + \mathcal{R}_3
\end{aligned}$$

Keeping only the 0th order terms:

$$\mathcal{K}_0\mu_{T_{ic}} = \mu_w\left(\mu_{T_L} + \mu_{T_U} + \frac{L_d+L_e}{\mu_v}\right) \tag{77}$$

Examples

We apply our results in the examples presented in (Körösi *et al.*, 2022) and (Groover, 2008).

Example 1 (Layout from (Groover, 2008))

Assume that the AGVs transport raw parts from station 1 to production stations 2, 3, and 4. Some parts must be transported from station 3 to station 4. All production stations send finished parts to unload station 5.

Table 3 contains the data for examples 1, 2, and 3.

Table 3. Data for examples 1, 2, and 3.

Flow rates (mean values, trips per hour)	$\overline{F}_{12} = 9$, $\overline{F}_{13} = 5$, $\overline{F}_{14} = 6$, $\overline{F}_{25} = 9$, $\overline{F}_{34} = 2$, $\overline{F}_{35} = 3$, $\overline{F}_{45} = 8$
Distances (m)	$D_{12} = 50$, $D_{13} = 120$, $D_{14} = 205$, $D_{25} = 80$, $D_{34} = 85$, $D_{35} = 170$, $D_{45} = 85$
Average distance per unloaded trip (m)	**$L_e = 795$ (example 1) OR $L_e = 100$ (examples 2 and 3)**
Flow rate relative deviation	$\delta = 25\%$
AGV capacity	$c = 1$
AGV speed (m/min)	$\mu_v = 84$, $\sigma_v = 16.8$
Loading time (min)	$\mu_{T_L} = 0.5$, $\sigma_{T_L} = 0.12$
Unloading time (min)	$\mu_{T_U} = 0.5$, $\sigma_{T_U} = 0.12$
Loading – unloading time correlation	$\rho_{T_L, T_U} = 0.3$
AGV availability	$\mu_A = 0.95$ **$\sigma_A = 0$ (examples 1 and 2) OR $\sigma_A = 0.1$ (example 3)**
Traffic factor	$\mu_{F_t} = 0.85$, $\sigma_{F_t} = 0.09$
Operator efficiency	$\mu_{E_W} = 0.95$ **$\sigma_{E_W} = 0$ (examples 1 and 2) OR $\sigma_{E_W} = 0.1$ (example 3)**

Initially, we compute:

$$\mu_z = \sum_{e\in\mathcal{E}} \overline{F}_e D_e = 4360,\ \sigma_z^2 = (\delta\mu_z)^2,\ \mu_w = \sum_{e\in\mathcal{E}} \overline{F}_e = 42,\ \sigma_w^2 = (\delta\mu_w)^2,\ L_d = 103.81.$$

Assume that each loaded trip from station (i) to station (j) is followed by a trip back to station (i) where the AGV travels unloaded. Then $L_e = \Sigma D_e = 795$. In this setup $L_e \gg L_d$, which is very inefficient.

Total cycle time per delivery per vehicle $T_c = T_L + T_U + L_d/v + L_e/v$:

$$\mu_{T_L+T_U} = 1 \qquad \mathcal{S}_2\mu_{(L_d+L_e)/v} = 11.128 \qquad \mathcal{S}_2\mu_{T_c} = 12.128$$
$$\mathcal{S}_2\sigma^2_{T_L+T_U} = 0.19^2 \qquad \mathcal{S}_2\sigma^2_{(L_d+L_e)/v} = 2.14^2 \qquad \mathcal{S}_2\sigma^2_{T_c} = 2.15^2$$

Ideal cycle time per delivery per vehicle $T_{ic} = wT_c$:

$$\mathcal{S}_2\mu_{T_{ic}} = 509.381$$
$$\mathcal{S}_2\sigma^2_{T_{ic}} = 152.44^2$$

The available time per hour per vehicle:

$$\mathcal{S}_2\mu_{T_{av}} = 46.027$$
$$\mathcal{S}_2\sigma^2_{T_{av}} = 4.87^2$$

Finally, the fleet size estimates:

$$\mathcal{S}_2\mu_N = 11.191$$
$$\mathcal{S}_2\sigma^2_N = 3.50^2$$

For comparison, the deterministic model gives

$$\mathcal{S}_0\mu_N = 10.676$$

Example 2: Continuation from example 1, low L_e

Assume now that the process is designed to achieve a low (relative to L_d) L_e, approximately equal to the length of the link from station 5 to station 1. This requires a two-position platform at each production station, so that a vehicle can pick up a load of finished parts right after each delivery of raw parts.

Therefore, assume that $L_e = D_{51} = 100$. Repeating the calculation we find:

Total cycle time per delivery per vehicle T_c:

$$\begin{array}{lll} T_L, T_U \text{ as in example 1} & \mathcal{S}_2\mu_{(L_d+L_e)/v} = 2.523 & \mathcal{S}_2\mu_{T_c} = 3.523 \\ & \mathcal{S}_2\sigma^2_{(L_d+L_e)/v} = 0.49^2 & \mathcal{S}_2\sigma^2_{T_c} = 0.52^2 \end{array}$$

Ideal cycle time per delivery per vehicle $T_{ic} = wT_c$ and resulting fleet size:

$$\begin{array}{ll} \mathcal{S}_2\mu_{T_{ic}} = 147.981 & \mathcal{S}_2\mu_N = 3.250 \\ \mathcal{S}_2\sigma^2_{T_{ic}} = 42.14^2 & \mathcal{S}_2\sigma^2_N = 0.97^2 \\ & \mathcal{S}_0\mu_N = 3.126 \end{array}$$

Example 3: Continuation from example 2, increased denominator variance

Suppose that both AGV availability and operator efficiency have a medium standard deviation: $\sigma_A = \sigma_{E_w} = 0.1$. This affects the available time per hour per vehicle, since its variance increases:

$$S_2\sigma^2_{T_{av}} = 60^2\left(\mu_A^2\mu^2_{F_t}\sigma^2_{E_w} + \mu_A^2\mu^2_{E_w}\sigma^2_{F_t} + \mu^2_{F_t}\mu^2_{E_w}\sigma^2_A\right) = 8.41^2$$

The fleet size estimates are now calculated to be:

Resulting fleet size:

$$\mathcal{S}_2\mu_N = 3.319$$
$$\mathcal{S}_2\sigma^2_N = 1.08^2$$

Observe that the mean value estimate has increased slightly by 2% because of the increased denominator variance alone, accompanied by an 11% increase in the standard deviation estimate. The denominator standard deviation has increased approximately by 70%.

Example 4 (Körösi *et al.*, 2022)

Table 4 contains the data for example 4.

Table 4. Data for example 4.

Flow rates (mean values, trips per hour)	$\overline{F}_{12} = 10$, $\overline{F}_{23} = 15$, $\overline{F}_{31} = 0$
Distances (m)	$D_{12} = 50$, $D_{23} = 60$, $D_{31} = 50$
Average distance per unloaded trip (m)	$L_e = 50$
Flow rate relative deviation	$\delta = 25\%$
AGV capacity	**$c = 1$ OR $c = 2$**
AGV speed (m/min)	$\mu_v = 50$ $\sigma_v = 5$
Loading time (min)	$\mu_{T_L} = 0.75$ $\sigma_{T_L} = 0.08$

Unloading time (min)	$\mu_{T_L} = 0.5$ $\sigma_{T_L} = 0.05$
Loading – unloading time correlation	$\rho_{T_L,T_U} = 0$
AGV availability	$\mu_A = 0.95$ $\sigma_A = 0.09$
Traffic factor	$\mu_{F_t} = 0.9$ $\sigma_{F_t} = 0.09$
Operator efficiency	$\mu_{E_w} = 1.0$ $\sigma_{E_w} = 0$

(i) AGV capacity $c = 1$:

$$\mu_z = \sum_{e\in\varepsilon} \bar{F}_e D_e = 1400,\ \sigma_z^2 = (\delta\mu_z)^2,\ \mu_w = \sum_{e\in\varepsilon} \bar{F}_e = 25,\ \sigma_w^2 = (\delta\mu_w)^2,\ L_d = 56$$

Total cycle time per delivery per vehicle $T_c = T_L + T_U + L_d/v + L_e/v$:

$$\mu_{T_L+T_U} = 1.25 \quad \mathcal{S}_2\mu_{(L_d+L_e)/v} = 2.141 \quad \mathcal{S}_2\mu_{T_c} = 3.391$$
$$\mathcal{S}_2\sigma^2_{T_L+T_U} = 0.09^2 \quad \mathcal{S}_2\sigma^2_{(L_d+L_e)/v} = 0.21^2 \quad \mathcal{S}_2\sigma^2_{T_c} = 0.23^2$$

Ideal cycle time per delivery per vehicle $T_{ic} = wT_c$, available time per hour per vehicle, fleet size:

$$\mathcal{S}_2\mu_{T_{ic}} = 84.78 \quad \mathcal{S}_2\mu_{T_{av}} = 51.3 \quad \mathcal{S}_2\mu_N = 1.684$$
$$\mathcal{S}_2\sigma^2_{T_{ic}} = 21.85^2 \quad \mathcal{S}_2\sigma^2_{T_{av}} = 7.07^2 \quad \mathcal{S}_2\sigma^2_N = 0.48^2$$
$$\mathcal{S}_0\mu_N = 1.642$$

(ii) AGV capacity $c = 2$:

Changing the capacity affects $L_d = \frac{1}{c}\cdot\frac{\mu_z}{\mu_w}$, which in turn affects $T_c = (T_L + T_U) + \frac{L_d + L_e}{v}$, $T_{ic} = wT_c$, and $N = \frac{T_{ic}}{T_{av}}$:

$$L_d = 28 \quad \mathcal{S}_2\mu_{(L_d+L_e)/v} = 1.576 \quad \mathcal{S}_2\mu_{T_c} = 2.826 \quad \mathcal{S}_2\mu_{T_{ic}} = 70.64 \quad \mathcal{S}_2\mu_N = 1.403$$
$$\mathcal{S}_2\sigma^2_{(L_d+L_e)/v} = 0.16^2 \quad \mathcal{S}_2\sigma^2_{T_c} = 0.18^2 \quad \mathcal{S}_2\sigma^2_{T_{ic}} = 18.14^2 \quad \mathcal{S}_2\sigma^2_N = 0.40^2$$
$$\mathcal{S}_0\mu_N = 1.369$$

Example 5 (Körösi *et al.*, 2022)

Table 5 contains the data for example 5.

Table 5. Data for example 5.

Flow rates (mean values, trips per hour)	$\bar{F}_{12} = \bar{F}_{23} = \bar{F}_{67} = 200$ $\bar{F}_{34} = \bar{F}_{36} = \bar{F}_{45} = \bar{F}_{56} = 100$ $\bar{F}_{71} = 0$
Distances (m)	$D_{12} = 65$, $D_{23} = 45$, $D_{34} = 50$, $D_{36} = 140$ $D_{45} = 70$, $D_{56} = 120$, $D_{67} = 65$, $D_{71} = 45$,
Average distance per unloaded trip (m)	$L_e = 45$
Flow rate relative deviation	$\delta = 25\%$
AGV capacity	**$c = 1$ OR $c = 100$**
AGV speed (m/min)	$\mu_v = 50$ $\sigma_v = 5$

Loading time (min)	$\mu_{T_L} = 0.75$ $\sigma_{T_L} = 0.08$
Unloading time (min)	$\mu_{T_L} = 0.5$ $\sigma_{T_L} = 0.05$
Loading – unloading time correlation	$\rho_{T_L T_U} = 0$
AGV availability	$\mu_A = 0.95$ $\sigma_A = 0.09$
Traffic factor	$\mu_{F_t} = 0.9$ $\sigma_{F_t} = 0.09$
Operator efficiency	$\mu_{E_w} = 1.0$ $\sigma_{E_w} = 0$

(i) AGV capacity $c = 1$:

$$\mu_z = \sum_{e\in\varepsilon} \bar{F}_e D_e = 73000,\ \sigma_z^2 = (\delta\mu_z)^2,\ \mu_w = \sum_{e\in\varepsilon} \bar{F}_e = 1000, \sigma_w^2 = (\delta\mu_w)^2, L_d = 73$$

Total cycle time per delivery per vehicle $T_c = T_L + T_U + L_d/v + L_e/v$:

$$\begin{array}{lll} \mu_{T_L+T_U} = 1.25 & \mathcal{S}_2\mu_{(L_d+L_e)/v} = 2.384 & \mathcal{S}_2\mu_{T_c} = 3.634 \\ \mathcal{S}_2\sigma^2_{T_L+T_U} = 0.09^2 & \mathcal{S}_2\sigma^2_{(L_d+L_e)/v} = 0.24^2 & \mathcal{S}_2\sigma^2_{T_c} = 0.25^2 \end{array}$$

Ideal cycle time per delivery per vehicle $T_{ic} = wT_c$, available time per hour per vehicle, fleet size:

$$\begin{array}{lll} \mathcal{S}_2\mu_{T_{ic}} = 3633.6 & \mathcal{S}_2\mu_{T_{av}} = 51.3 & \mathcal{S}_2\mu_N = 72.166 \\ \mathcal{S}_2\sigma^2_{T_{ic}} = 937.60^2 & \mathcal{S}_2\sigma^2_{T_{av}} = 7.07^2 & \mathcal{S}_2\sigma^2_N = 20.69^2 \\ & & \mathcal{S}_0\mu_N = 70.370 \end{array}$$

(ii) AGV capacity $c = 100$:

$$L_d = 0.73 \quad \begin{array}{llll} \mathcal{S}_2\mu_{(L_d+L_e)/v} = 0.924 & \mathcal{S}_2\mu_{T_c} = 2.174 & \mathcal{S}_2\mu_{T_{ic}} = 2173.746 & \mathcal{S}_2\mu_N = 43.174 \\ \mathcal{S}_2\sigma^2_{(L_d+L_e)/v} = 0.09^2 & \mathcal{S}_2\sigma^2_{T_c} = 0.13^2 & \mathcal{S}_2\sigma^2_{T_{ic}} = 556.87^2 & \mathcal{S}_2\sigma^2_N = 12.31^2 \\ & & & \mathcal{S}_0\mu_N = 42.195 \end{array}$$

Comparison with simulation experiments

We complete this study by comparing the estimates presented in the previous section with results of simulation experiments.

Experiment setup:

We model the distributions of the problem data by symmetric triangular distributions with the specified mean value and variance, except for cases where it is not possible to honor the domain of the respective variable. Such are the cases of A, F_t, and E_w coefficients, the domain of which is bounded below 1, and constructing a symmetric triangular distribution with the specified mean and variance while its support is a subrange of [0,1] may be impossible. In such cases, we model the random variable by a beta distribution, which can meet any requirements of this kind, with the important drawback that its support is not bounded away from zero, so that division by such variables may result in fleet size distribution with large and / or unstable higher order moments.

Δ is modeled as a random variable following a symmetric triangular distribution with support $[-\sqrt{6}\cdot\delta, \sqrt{6}\cdot\delta]$. Indeed, $var(\Delta) = (2\sqrt{6})\delta^2/24 = \delta^2$.

Following the *Choleski decomposition* method (Glasserman, 2003), T_L and T_U are modeled as:

$T_L = T'_L$, where T'_L follows a symmetric triangular distribution with the given mean μ_{T_L} and variance $\sigma^2_{T_L}$,

$T_U = \mu_{T_U} + \sigma_{T_U}\left(\rho_{T_L,T_U}\frac{T'_L-\mu_{T_L}}{\sigma_{T_L}} + \sqrt{1-\rho^2_{T_L,T_U}}\cdot\frac{T'_U-\mu_{T_U}}{\sigma_{T_U}}\right)$, where T'_U follows a symmetric triangular distribution with the given mean μ_{T_U} and variance $\sigma^2_{T_U}$. T'_L and T'_U are independent.

It is easy to show that with this definition, T_L and T_U have the prescribed means and variances, as well as the prescribed correlation between them:

Obviously, $E[T_L] = E[T'_L]$ and $\sigma_{T_L} = \sigma_{T'_L}$.

Regarding T_U:

$$\begin{aligned}E[T_U] &= \mu_{T_U} + \sigma_{T_U}\left(\rho_{T_L,T_U}\frac{E[T'_L]-\mu_{T_L}}{\sigma_{T_L}} + \sqrt{1-\rho^2_{T_L,T_U}}\cdot\frac{E[T'_U]-\mu_{T_U}}{\sigma_{T_U}}\right)\\ &= \mu_{T_U} + \sigma_{T_U}\left(\rho_{T_L,T_U}\frac{\mu_{T_L}-\mu_{T_L}}{\sigma_{T_L}} + \sqrt{1-\rho^2_{T_L,T_U}}\cdot\frac{\mu_{T_U}-\mu_{T_U}}{\sigma_{T_U}}\right)\\ &= \mu_{T_U}\end{aligned}$$

and, since T'_L and T'_U are independent:

$$\begin{aligned}Var(T_U) &= Var\left(\frac{\sigma_{T_U}\rho_{T_L,T_U}}{\sigma_{T_L}}\cdot\left(T'_L-\mu_{T_L}\right) + \frac{\sigma_{T_U}}{\sigma_{T_U}}\sqrt{1-\rho^2_{T_L,T_U}}\cdot\left(T'_U-\mu_{T_U}\right)\right)\\ &= Var\left(\frac{\sigma_{T_U}\rho_{T_L,T_U}}{\sigma_{T_L}}T'_L + \sqrt{1-\rho^2_{T_L,T_U}}\cdot T'_U\right)\\ &= Var\left(\frac{\sigma_{T_U}\rho_{T_L,T_U}}{\sigma_{T_L}}T'_L\right) + Var\left(\sqrt{1-\rho^2_{T_L,T_U}}\cdot T'_U\right)\\ &= \frac{\sigma^2_{T_U}\rho^2_{T_L,T_U}}{\sigma^2_{T_L}}\sigma^2_{T_L} + (1-\rho^2_{T_L,T_U})\sigma^2_{T_U}\\ &= \sigma^2_{T_U}\rho^2_{T_L,T_U} + (1-\rho^2_{T_L,T_U})\sigma^2_{T_U}\\ &= \sigma^2_{T_U}\end{aligned}$$

Then, the correlation between T_L and T_U is verified:

$$\begin{aligned}Cov(T_L,T_U) &= Cov\left(T'_L, \mu_{T_U} + \sigma_{T_U}\left(\rho_{T_L,T_U}\frac{T'_L-\mu_{T_L}}{\sigma_{T_L}} + \sqrt{1-\rho^2_{T_L,T_U}}\cdot\frac{T'_U-\mu_{T_U}}{\sigma_{T_U}}\right)\right)\\ &= Cov\left(T'_L, \sigma_{T_U}\left(\rho_{T_L,T_U}\frac{T'_L-\mu_{T_L}}{\sigma_{T_L}} + \sqrt{1-\rho^2_{T_L,T_U}}\cdot\frac{T'_U-\mu_{T_U}}{\sigma_{T_U}}\right)\right)\\ &= Cov\left(T'_L, \sigma_{T_U}\rho_{T_L,T_U}\frac{T'_L-\mu_{T_L}}{\sigma_{T_L}}\right) = Cov\left(T'_L, \sigma_{T_U}\rho_{T_L,T_U}\frac{T'_L}{\sigma_{T_L}}\right)\\ &= \frac{\sigma_{T_U}}{\sigma_{T_L}}\rho_{T_L,T_U}\sigma^2_{T'_L} = \frac{\sigma_{T_U}}{\sigma_{T_L}}\rho_{T_L,T_U}\sigma^2_{T_L}\\ &= \sigma_{T_U}\sigma_{T_L}\rho_{T_L,T_U}\end{aligned}$$

so that

$$\begin{aligned}Corr(T_L,T_U) &= \frac{Cov(T_L,T_U)}{\sigma_{T_L}\sigma_{T_U}}\\ &= \sigma_{T_U}\sigma_{T_L}\rho_{T_L,T_U}\cdot\frac{1}{\sigma_{T_L}\sigma_{T_U}}\\ &= \rho_{T_L,T_U}\end{aligned}$$

Having constructed same-size samples of Δ, T_L, T_U, AGV speed v, AGV availability A, traffic factor F_t, and operator efficiency E_w, the rest of the calculations are straightforward to perform for all sample points.

This way we derive a sample of the distribution of the fleet size N.

The empirical distribution of the ideal cycle time per delivery per vehicle T_{ic} is plotted with a blue solid line, while the one of the available time per hour per vehicle T_{av} is plotted with a green solid line. Notice that these quantities are the numerator and denominator of the Equation (8), which determines the fleet size.

The empirical distribution of the latter is plotted with a red solid line.

Two ranges are also depicted as shaded areas: the one in blue represents the range within ± 1 standard deviation around the empirical mean, derived from the sample data; the one in red represents approximately the ± 1 estimated standard deviation range around the estimated mean, computed as 2nd order approximations. Although the two ranges do not coincide, they overlap significantly.

Finally, the three vertical straight lines (not always distinguishable from one anther) mark: the empirical mean (a dashed blue line), the 2nd order estimation of the mean (a solid green line), and the 0th order estimation of the mean (a dotted black line).

Example 1 (Table 6, Figure 1):

Table 6. Specifications, sample statistics, and theoretical predictions for example 1.

Input statistics vs *spec*ifications:					Derived samples statistics vs estimations:			
Δ	*sample*:	−0.002 ± 0.25	*spec.*:	0 ± 0.25				
v	*sample*:	84.048 ± 16.64	*spec.*:	84 ± 16.8				
T_L	*sample*:	0.500 ± 0.12	*spec.*:	0.5 ± 0.12				
T_U	*sample*:	0.499 ± 0.12	*spec.*:	0.5 ± 0.12				
ρ_{T_L,T_U}	*sample*:	0.30	*spec.*:	0.3				
A	*sample*:	0.950 ± 0.00	*spec.*:	0.95 ± 0				
F_t	*sample*:	0.849 ± 0.09	*spec.*:	0.85 ± 0.09				
E_W	*sample*:	0.950 ± 0.00	*spec.*:	0.95 ± 0				
					T_c	*sample*: 12.159 ± 2.44	2nd *order estim.*:	12.128 ± 2.15
					T_{ic}	*sample*: 509.674 ± 164.86	2nd *order estim.*:	509.381 ± 152.44
					T_{av}	*sample*: 45.988 ± 4.91	2nd *order estim.*:	46.027 ± 4.87
					N	*sample*: 11.223 ± 3.89	2nd *order estim.*:	11.187 ± 3.50
							0th *order estim.*:	10.676

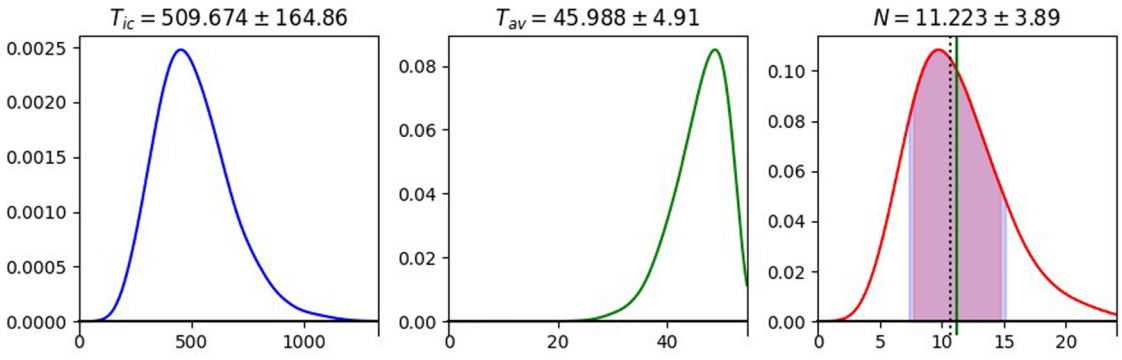


Figure 1. Empirical (sample) distribution of the ideal cycle time per delivery per vehicle (left), the available time (middle), and the fleet size (right) for example 1.

Example 2 (Table 7, Figure 2):

Table 7. Specifications, sample statistics, and theoretical predictions for example 2.

Input statistics vs specifications:					Derived samples statistics vs estimations:			
Δ	*sample* :	0.000 ± 0.25	*spec.* :	0 ± 0.25				
v	*sample* :	83.959 ± 16.78	*spec.* :	84 ± 16.8				
T_L	*sample* :	0.500 ± 0.12	*spec.* :	0.5 ± 0.12	T_c	*sample* : 3.535 ± 0.59	2^{nd} *order estim.* :	3.523 ± 0.52
T_U	*sample* :	0.499 ± 0.12	*spec.* :	0.5 ± 0.12	T_{ic}	*sample* : 148.496 ± 45.32	2^{nd} *order estim.* :	147.981 ± 42.14
ρ_{T_L,T_U}	*sample* :	0.30	*spec.* :	0.3	T_{av}	*sample* : 46.020 ± 4.85	2^{nd} *order estim.* :	46.027 ± 4.87
A	*sample* :	0.950 ± 0.00	*spec.* :	0.95 ± 0	N	*sample* : 3.267 ± 1.08	2^{nd} *order estim.* :	3.250 ± 0.97
F_t	*sample* :	0.850 ± 0.09	*spec.* :	0.85 ± 0.09			0^{th} *order estim.* :	3.126
E_W	*sample* :	0.950 ± 0.00	*spec.* :	0.95 ± 0				

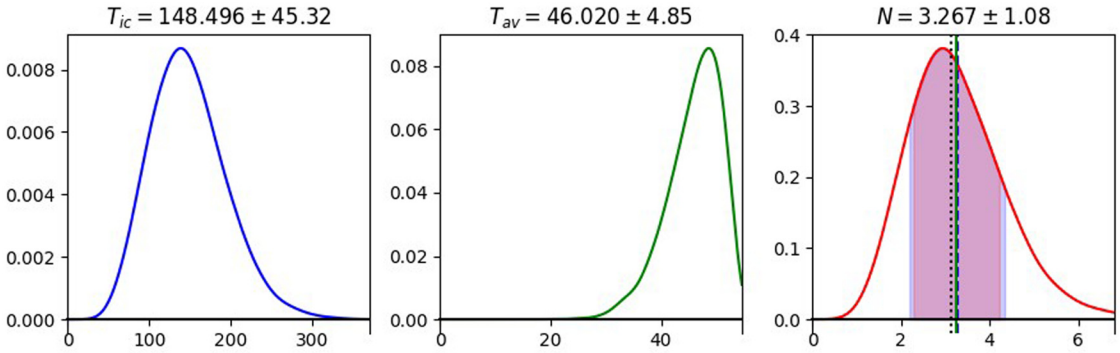


Figure 2. Empirical (sample) distribution of the ideal cycle time per delivery per vehicle (left), the available time (middle), and the fleet size (right) for example 2.

Example 3 (Table 8, Figure 3):

Table 8. Specifications, sample statistics, and theoretical predictions for example 3.

Input statistics vs specifications:					Derived samples statistics vs estimations:			
Δ	*sample* :	0.001 ± 0.25	*spec.* :	0 ± 0.25				
v	*sample* :	84.131 ± 16.93	*spec.* :	84 ± 16.8				
T_L	*sample* :	0.499 ± 0.12	*spec.* :	0.5 ± 0.12	T_c	*sample* : 3.531 ± 0.59	2^{nd} *order estim.* :	3.523 ± 0.52
T_U	*sample* :	0.501 ± 0.12	*spec.* :	0.5 ± 0.12	T_{ic}	*sample* : 148.502 ± 45.25	2^{nd} *order estim.* :	147.981 ± 42.14
ρ_{T_L,T_U}	*sample* :	0.29	*spec.* :	0.3	T_{av}	*sample* : 46.099 ± 8.40	2^{nd} *order estim.* :	46.027 ± 8.41
A	*sample* :	0.949 ± 0.10	*spec.* :	0.95 ± 0.1	N	*sample* : 3.394 ± 1.64	2^{nd} *order estim.* :	3.319 ± 1.08
F_t	*sample* :	0.851 ± 0.09	*spec.* :	0.85 ± 0.09			0^{th} *order estim.* :	3.126
E_W	*sample* :	0.951 ± 0.10	*spec.* :	0.95 ± 0.1				

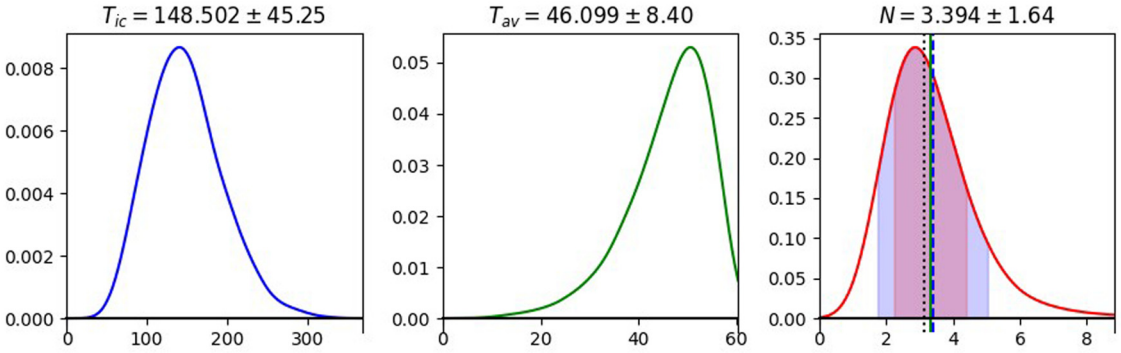


Figure 3. **Empirical (sample) distribution of the ideal cycle time per delivery per vehicle (left), the available time (middle), and the fleet size (right) for example 3.**

Example 4:

i) AGV capacity = 1 (Table 9, Figure 4)

Table 9. **Specifications, sample statistics, and theoretical predictions for example 4i.**

Input statistics vs specifications:					**Derived samples statistics vs estimations:**			
Δ	*sample* :	0.001 ± 0.25	*spec.* :	0 ± 0.25				
v	*sample* :	50.002 ± 4.98	*spec.* :	50 ± 5				
T_L	*sample* :	0.750 ± 0.08	*spec.* :	0.75 ± 0.08	T_c	*sample* : 3.391 ± 0.24	2^{nd} *order estim.* :	3.391 ± 0.23
T_U	*sample* :	0.500 ± 0.05	*spec.* :	0.5 ± 0.05	T_{ic}	*sample* : 84.907 ± 22.14	2^{nd} *order estim.* :	84.780 ± 21.85
ρ_{T_L,T_U}	*sample* :	0.31	*spec.* :	0.3	T_{av}	*sample* : 51.332 ± 7.14	2^{nd} *order estim.* :	51.300 ± 7.07
A	*sample* :	0.951 ± 0.09	*spec.* :	0.95 ± 0.09	N	*sample* : 1.698 ± 0.56	2^{nd} *order estim.* :	1.684 ± 0.48
F_t	*sample* :	0.899 ± 0.09	*spec.* :	0.9 ± 0.09			0^{th} *order estim.* :	1.642
E_W	*sample* :	1.000 ± 0.00	*spec.* :	1 ± 0				

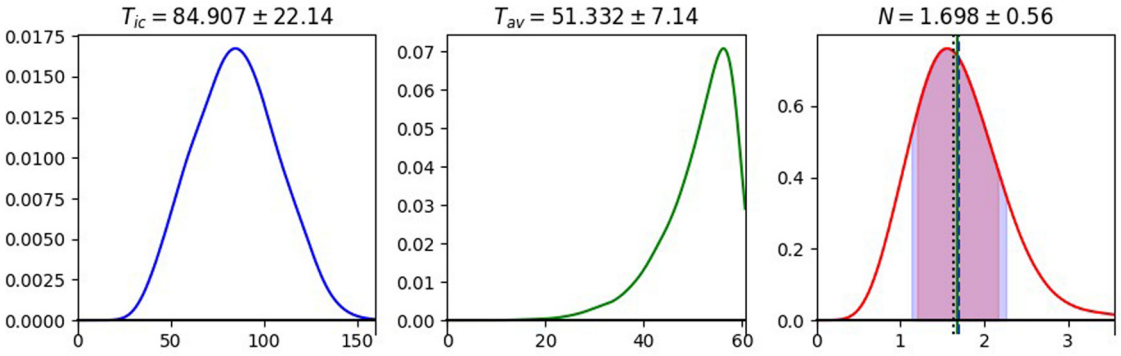


Figure 4. **Empirical (sample) distribution of the ideal cycle time per delivery per vehicle (left), the available time (middle), and the fleet size (right) for example 4i.**

ii) AGV capacity = 2 (Table 10, Figure 5)

Table 10. Specifications, sample statistics, and theoretical predictions for example 4ii.

Input statistics vs specifications:					Derived samples statistics vs estimations:			
Δ	*sample* :	−0.002 ± 0.25	*spec.* :	0 ± 0.25				
v	*sample* :	50.011 ± 5.00	*spec.* :	50 ± 5				
T_L	*sample* :	0.751 ± 0.08	*spec.* :	0.75 ± 0.08				
T_U	*sample* :	0.500 ± 0.05	*spec.* :	0.5 ± 0.05				
ρ_{T_L,T_U}	*sample* :	0.30	*spec.* :	0.3				
A	*sample* :	0.950 ± 0.09	*spec.* :	0.95 ± 0.09				
F_t	*sample* :	0.900 ± 0.09	*spec.* :	0.9 ± 0.09				
E_W	*sample* :	1.000 ± 0.00	*spec.* :	1 ± 0				
					T_c	*sample* : 2.826 ± 0.19	2^{nd} *order estim.* :	2.826 ± 0.18
					T_{ic}	*sample* : 70.507 ± 18.18	2^{nd} *order estim.* :	70.640 ± 18.14
					T_{av}	*sample* : 51.318 ± 7.07	2^{nd} *order estim.* :	51.300 ± 7.07
					N	*sample* : 1.410 ± 0.47	2^{nd} *order estim.* :	1.403 ± 0.40
							0^{th} *order estim.* :	1.369

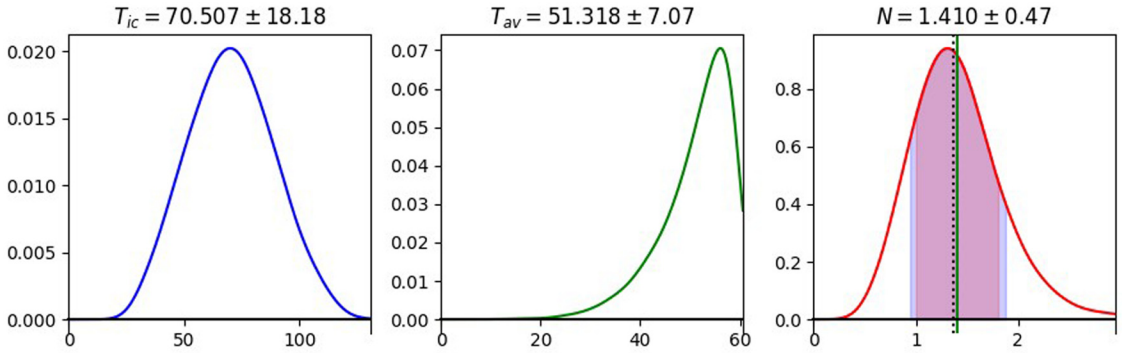


Figure 5. Empirical (sample) distribution of the ideal cycle time per delivery per vehicle (left), the available time (middle), and the fleet size (right) for example 4ii.

Example 5:

i) AGV capacity = 1 (Table 11, Figure 6)

Table 11. Specifications, sample statistics, and theoretical predictions for example 5i.

Input statistics vs specifications:					Derived samples statistics vs estimations:			
Δ	*sample* :	−0.003 ± 0.25	*spec.* :	0 ± 0.25				
v	*sample* :	49.997 ± 5.03	*spec.* :	50 ± 5				
T_L	*sample* :	0.749 ± 0.08	*spec.* :	0.75 ± 0.08				
T_U	*sample* :	0.500 ± 0.05	*spec.* :	0.5 ± 0.05				
ρ_{T_L,T_U}	*sample* :	0.28	*spec.* :	0.3				
A	*sample* :	0.950 ± 0.09	*spec.* :	0.95 ± 0.09				
F_t	*sample* :	0.900 ± 0.09	*spec.* :	0.9 ± 0.09				
E_W	*sample* :	1.000 ± 0.00	*spec.* :	1 ± 0				
					T_c	*sample* : 3.634 ± 0.27	2^{nd} *order estim.* :	3.634 ± 0.25
					T_{ic}	*sample* : 3622.523 ± 949.19	2^{nd} *order estim.* :	3633.600 ± 937.60
					T_{av}	*sample* : 51.266 ± 7.07	2^{nd} *order estim.* :	51.300 ± 7.07
					N	*sample* : 72.483 ± 23.77	2^{nd} *order estim.* :	72.166 ± 20.69
							0^{th} *order estim.* :	70.370

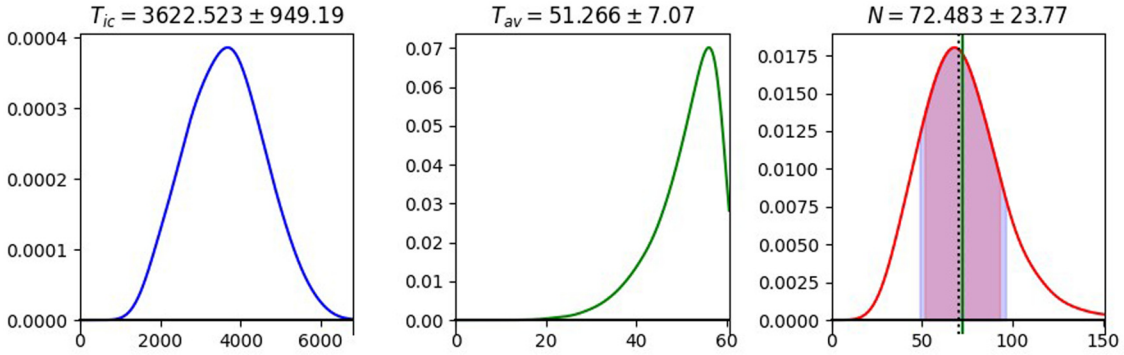


Figure 6. Empirical (sample) distribution of the ideal cycle time per delivery per vehicle (left), the available time (middle), and the fleet size (right) for example 5i.

ii) AGV capacity = 100 (Table 12, Figure 7)

Table 12. Specifications, sample statistics, and theoretical predictions for example 5ii.

Input statistics vs specifications:					**Derived samples statistics vs estimations:**			
Δ	*sample*:	0.003 ± 0.25	*spec.*:	0 ± 0.25				
v	*sample*:	50.027 ± 5.01	*spec.*:	50 ± 5				
T_L	*sample*:	0.751 ± 0.08	*spec.*:	0.75 ± 0.08				
T_U	*sample*:	0.500 ± 0.05	*spec.*:	0.5 ± 0.05				
ρ_{T_L,T_U}	*sample*:	0.30	*spec.*:	0.3				
A	*sample*:	0.951 ± 0.09	*spec.*:	0.95 ± 0.09				
F_t	*sample*:	0.900 ± 0.09	*spec.*:	0.9 ± 0.09				
E_W	*sample*:	1.000 ± 0.00	*spec.*:	1 ± 0				
					T_c *sample*:	2.174 ± 0.14	2^{nd} *order estim.*:	2.174 ± 0.13
					T_{ic} *sample*:	2179.810 ± 565.97	2^{nd} *order estim.*:	2173.746 ± 556.87
					T_{av} *sample*:	51.362 ± 7.03	2^{nd} *order estim.*:	51.300 ± 7.07
					N *sample*:	43.518 ± 14.18	2^{nd} *order estim.*:	43.174 ± 12.31
							0^{th} *order estim.*:	42.195

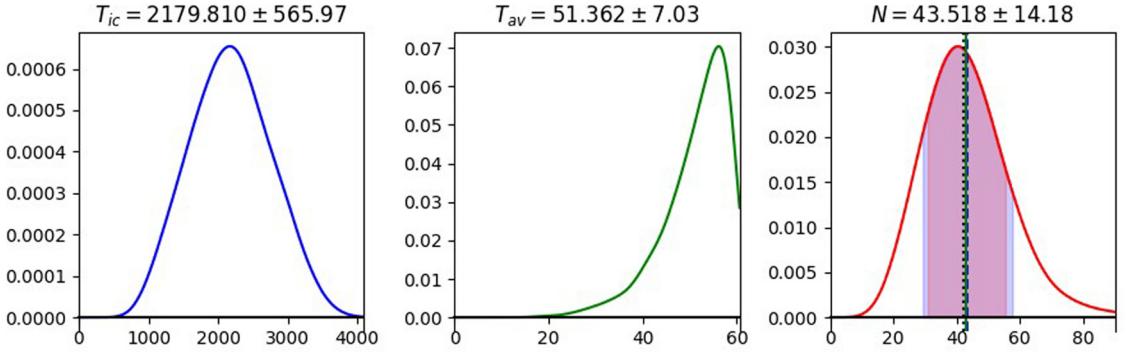


Figure 7. Empirical (sample) distribution of the ideal cycle time per delivery per vehicle (left), the available time (middle), and the fleet size (right) for example 5ii.

Discussion

The simulation experiments verify the validity of the approach. The 2nd order mean value estimation of the fleet size is remarkably accurate. For example, in the last experiment the 0th order (deterministic) estimation is 42.195, the 2nd order estimation is 43.177, while the actual (sample) estimation is 43.518. So, while the deterministic estimation is off by 3%, the 2nd order estimation is less than 1% off. In the third experiment, where the skewness and kurtosis of the N sample are large and perhaps unstable ($\gamma_N = 7.27$, $k_N = 154.83$), the 2nd order mean value estimation is off by only a little more than 2%, a clear improvement over the deterministic estimation which is off by 8%.

The approximation to the fleet size variance is cruder. For example, in the last experiment the variance is estimated to be 12.31^2, while the sample has variance equal to 14.18^2, so that the estimation error to the standard deviation is 13.2%. In the more pathogenic third experiment, $\sigma_{N(sample)} = 1.64$, while the 2nd order variance estimation is 1.10^2 (a 33% error for the standard deviation estimation). However, even with this relatively high error in the estimation, this is valuable information, since it provides the analyst a systematic way to assess, to a first approximation, the process *reliability*. Estimating that the fleet size has a high variance might indicate the need for additional AGVs to ensure a higher level of confidence in meeting the effective demand.

Conclusion

This paper presents an extension to traditional fleet sizing models by modeling key inputs as random variables, improving the model's applicability under uncertainty.

Using the expansion of a function of random variables in terms of the central moments of its arguments, we derive second-order approximations of the expected value and the variance of the required number of automated guided vehicles (AGVs). The method yields more accurate fleet size estimates than the deterministic approach, and also enables better-informed decision making and higher confidence in satisfying demand under uncertain conditions.

Several numerical examples, conducted as Monte Carlo simulations, are used to demonstrate the applicability of the method and illustrate various aspects of the problem. The effects of skewness, kurtosis, and support range of the ratio distribution's denominator are highlighted, and it is pointed out that the mean value is approximated quite satisfactorily, even in cases with high denominator variance, while the variance estimation is accurate only in cases with low denominator variance.

Appendix A – 2nd order approximation of fleet size statistics

Obtaining the 2nd order approximations of μ_N and σ_N requires some attention, because the expressions in (73), (74), which use 2nd order approximations of various quantities, when expanded as polynomials in the various random variable moments contain higher order terms as well. To see this, consider first (73):

$$\mu_N \simeq \frac{\mathcal{S}_2\mu_{T_{ic}}}{\mu_{T_{av}}} + \frac{\mathcal{S}_2\mu_{T_{ic}}}{\mu^3_{T_{av}}}\mathcal{S}_2\sigma^2_{T_{av}}$$

The first term $\mathcal{S}_2\mu_{T_{ic}}/\mu_{T_{av}}$ is indeed of 2nd order, but the second term is of 4th order.

From (25):

$$\mathcal{S}_2 = \mathcal{K}_0 + \mathcal{K}_1 + \mathcal{K}_2$$

Observe that for all quantities studied in this work, $\mathcal{K}_1 \equiv 0$. On the other hand, $\mathcal{K}_0 = \mathcal{S}_0$, so that $\mathcal{K}_2 = \mathcal{S}_2 - \mathcal{S}_0$. Additionally, for the variance of all quantities U it holds that $\mathcal{S}_2\sigma^2_U = \mathcal{K}_2\sigma^2_U$.

The expression for μ_N can be written as

$$\begin{aligned}
\mu_N &= \frac{\mathcal{S}_2\mu_{T_{ic}}}{\mu_{T_{av}}} + \frac{\mathcal{S}_2\mu_{T_{ic}}}{\mu^3_{T_{av}}}\mathcal{S}_2\sigma^2_{T_{av}} + \mathcal{R}_3 \\
&= \frac{\mathcal{S}_2\mu_{T_{ic}}}{\mu_{T_{av}}} + \frac{(\mathcal{K}_0\mu_{T_{ic}} + \mathcal{K}_2\mu_{T_{ic}})\mathcal{K}_2\sigma^2_{T_{av}}}{\mu^3_{T_{av}}} + \mathcal{R}_3 \\
&= \frac{\mathcal{S}_2\mu_{T_{ic}}}{\mu_{T_{av}}} + \frac{\mathcal{K}_0\mu_{T_{ic}}\mathcal{K}_2\sigma^2_{T_{av}}}{\mu^3_{T_{av}}} + \mathcal{R}_3 \\
&\simeq_2 \frac{\mathcal{S}_2\mu_{T_{ic}}}{\mu_{T_{av}}} + \frac{\mathcal{K}_0\mu_{T_{ic}}\mathcal{K}_2\sigma^2_{T_{av}}}{\mu^3_{T_{av}}}
\end{aligned}$$

Consider now (74), which is trickier:

$$\sigma^2_N \simeq \frac{1}{\mu^2_{T_{av}}}\mathcal{S}_2\sigma^2_{T_{ic}} + \frac{(\mathcal{S}_2\mu_{T_{ic}})^2}{\mu^4_{T_{av}}}\mathcal{S}_2\sigma^2_{T_{av}}$$

The first term is already a 2nd order approximation (see also the derivation of (70)). The term $\left((\mathcal{S}_2\mu_{T_{ic}})^2 \big/ \mu_{T_{av}}^4\right)\mathcal{S}_2\sigma_{T_{av}}^2$. is of 6th order, it contains, however, 2nd order terms, which we must keep:

$$\begin{aligned}(\mathcal{S}_2\mu_{T_{ic}})^2\mathcal{S}_2\sigma_{T_{av}}^2 &= \left(\mathcal{K}_0\mu_{T_{ic}} + \mathcal{K}_2\mu_{T_{ic}}\right)^2\mathcal{K}_2\sigma_{T_{av}}^2 \\ &= \left((\mathcal{K}_0\mu_{T_{ic}})^2 + (\mathcal{K}_2\mu_{T_{ic}})^2 + 2\mathcal{K}_0\mu_{T_{ic}}\mathcal{K}_2\mu_{T_{ic}}\right)\mathcal{K}_2\sigma_{T_{av}}^2 \\ &= \left(\mathcal{K}_0\mu_{T_{ic}}\right)^2\mathcal{K}_2\sigma_{T_{av}}^2 + \mathcal{R}_4\end{aligned}$$

Thus, the 2nd order approximation of σ_N^2 is given by

$$\sigma_N^2 \simeq_2 \frac{\mathcal{S}_2\sigma_{T_{ic}}^2}{\mu_{T_{av}}^2} + \frac{\left(\mathcal{K}_0\mu_{T_{ic}}\right)^2\mathcal{K}_2\sigma_{T_{av}}^2}{\mu_{T_{av}}^4}$$

Both expressions differ from (73 – 74) only in that the 2nd order term $\mathcal{S}_2\mu_{T_{ic}}$ has been replaced by its 0th order approximation $\mathcal{K}_0\mu_{T_{ic}}$ (remember that $\mathcal{K}_0\sigma_{T_{av}}^2 = 0$, so that $\mathcal{S}_2\sigma_{T_{av}}^2 = \mathcal{K}_2\sigma_{T_{av}}^2$).

Appendix B – Parameters of the models of (Egbelu, 1987)

For the sake of completeness, the expressions giving the estimate of the required number of vehicles for the four models described in (Egbelu, 1987) are also listed below, somewhat simplified. Besides those used by the previously described model, the variables given in Table B1 are also used:

Table B1. Additional model parameters used in (Egbelu, 1987).

Parameter	Description	Domain
f_{ij}	Expected number of loaded trips required between work center *i* and work center *j* during a shift period. This differs somewhat from F_{ij} of (Groover, 2008), which is flow rate per hour.	
t_b	The *battery charge time* of an AGV.	$t_b \geq 0$
b	The *blocking time factor*	$0 \leq b < 1$
γ	The *idle time factor*. Not to be confused with the skewness of a random variable, introduced in (14).	$0 \leq \gamma \leq 1$
ρ	*Vehicle efficiency*, after accounting for breakdowns	$0 < \rho \leq 1$
$d(\alpha_i,\beta_i)$	The distance between the delivery station α_i and the pickup station β_i within the *i*-th work center.	$d(\alpha_i,\beta_i) \geq 0$

Model 1:

$$N = \left[\frac{\Sigma_{e\in\mathcal{E}} f_e D_e}{v} + \sum_e f_{e\in\mathcal{E}}(t_L + t_U)\right] \Big/ 60(T - t_b) \tag{B.1}$$

Model 2:

$$N = \frac{\Sigma_{e\in\mathcal{E}} f_e}{60T} \cdot \left(\frac{(1+b+\gamma)}{\rho} \cdot \frac{\Sigma_{e\in\mathcal{E}} f_e D_e \big/ \Sigma_{e\in\mathcal{E}} f_e}{v} + t_L + t_U\right) \tag{B.2}$$

Model 3:

$$N = \left[(D_1 + D_2 + D_3)/v + \sum_{e\in\mathcal{E}} f_e(t_L + t_U)\right] \Big/ 60(T - t_b) \tag{B.3}$$

where

$$D_1 = \left(\sum_{e\in\mathcal{E}} f_e D_e \Big/ \sum_{e\in\mathcal{E}} f_e\right) \cdot \sum_{\bar{f}_i > 0} \bar{f}_i \tag{B.4}$$

$$\bar{f}_i = \sum_{(v_j,v_i)\in\mathcal{E}} F_{ji} - \sum_{(v_i,v_j)\in\mathcal{E}} F_{ij} \tag{B.5}$$

$$D_2 = \sum_{i\in\mathcal{V}} \left[\min\left\{ \sum_{j\in\mathcal{V}:e=(v_i,v_j)\in\mathcal{E}} f_e, \sum_{j\in\mathcal{V}:e=(v_j,v_i)\in\mathcal{E}} f_e \right\} d(\alpha_i,\beta_i) \right] \tag{B.6}$$

$$D_3 = \sum_{e\in\mathcal{E}} f_e D_e \tag{B.7}$$

***Model 4*:**

$$N = \left[\sum_{e\in\mathcal{E}} (g_e D_e + f_e D_e)/v + \left(\sum_{e\in\mathcal{E}} f_e \right)(t_L + t_U) \right] \Big/ 60(T - t_b) \tag{B.8}$$

where

$$g(v_i,v_j) = \left(\sum_{\substack{k\in\mathcal{V} \\ (v_k,v_i)\in\mathcal{E}}} f_{(v_k,v_i)} \right) \cdot \frac{\left(\sum_{k\in\mathcal{V}\,(v_j,v_k)\in\mathcal{E}} f_{(v_k,v_i)} \right)}{\sum_{e\in\mathcal{E}} f_e} \tag{B.9}$$

Appendix C – The effect of denominator variability

In the class of problems we study, among all possible arithmetic operations that are involved in the computation of the required fleet size, the division operation has the most significant effect on both the expected value and the variance of the latter, as it cannot be represented exactly by any finite-order expansion $S_i(X/Y)$ of the form (22).

To illustrate this point, we perform a series of tests (simulations). The random variable X is taken to follow a triangular distribution, while the random variable Y is taken to follow either a triangular or a beta distribution, to help illustrate the importance of the support range. We experiment both with the spread of the distributions (the variance), and with their level of asymmetry (skewness).

For each test we define the distributions for X and Y through their parameters: the support endpoints a, b and the mode c for triangular distributions; the shape parameters α and β in the case of the two-parameter beta distributions, the support of which is (0,1).

Then,

- for k = 1,2, ... , M we:
  - draw N samples $(x_i^{(k)}, y_i^{(k)})$, i = 1,2, ... , N from the given distributions
  - for each sample pair $(x_i^{(k)}, y_i^{(k)})$ compute the ratio $z_i^{(k)} = \dfrac{x_i^{(k)}}{y_i^{(k)}}$
  - compute the empirical (sample) mean and variance (the latter using an unbiased estimator):

    $\mu_z^{(k)} = \sum_{i=1}^{N} z_i^{(k)} / N$

    $\left(\sigma_z^{(k)}\right)^2 = \sum_{i=1}^{N} \left(z_i^{(k)} - \mu_Z^{(k)}\right)^2 / \left(N(N-1)\right)$

- compute the average of the empirical means and variances:

  $\mu_z = \sum_{k=1}^{M} \mu_z^{(k)} / M$

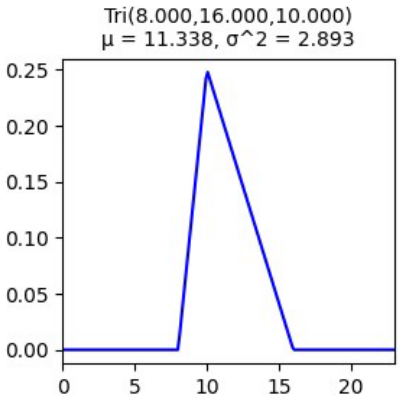
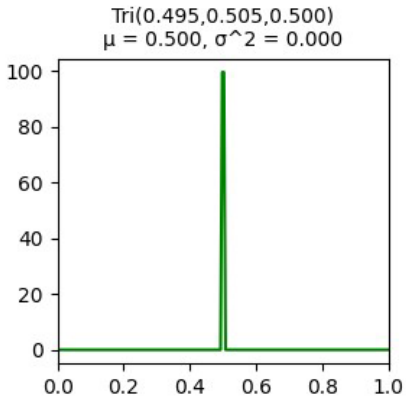
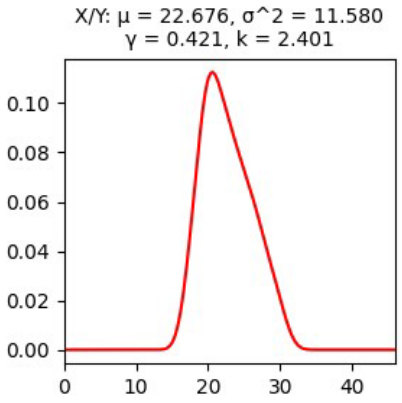
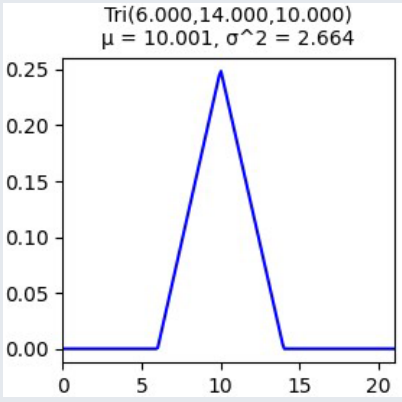
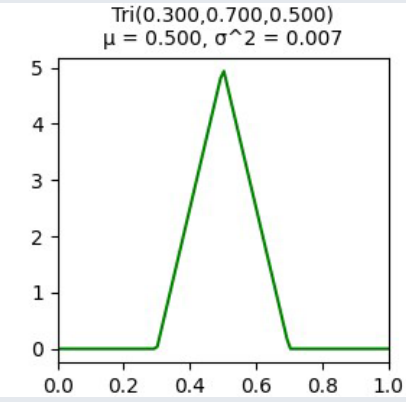
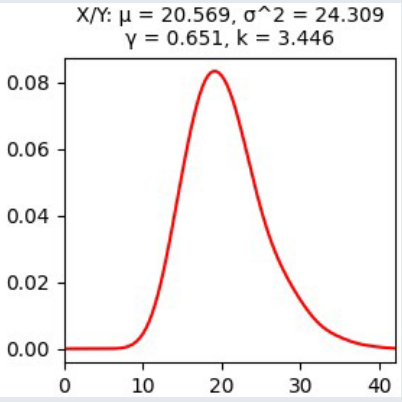
  $\sigma_z^2 = \sum_{k=1}^{M} \left(\sigma_z^{(k)}\right)^2 / M$

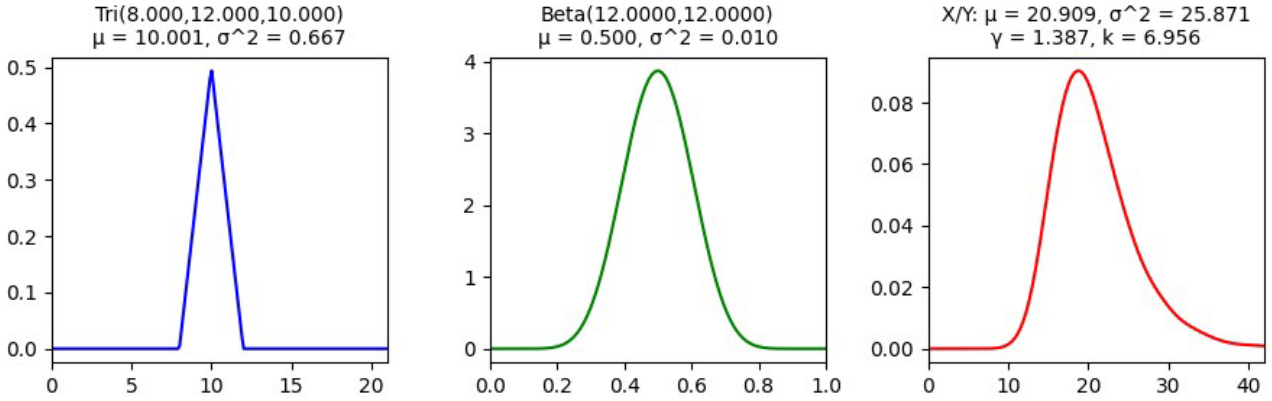
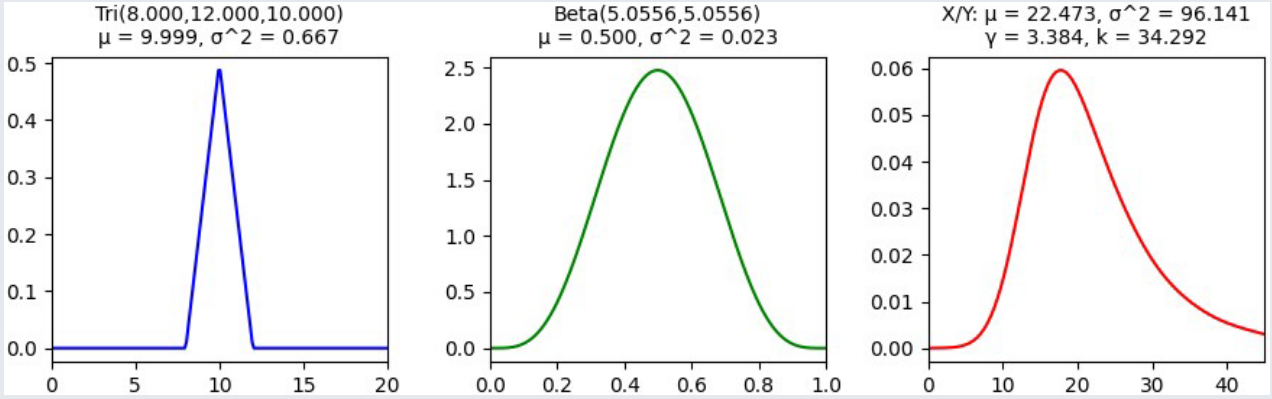
The latter values are the estimations of the mean value and the variance of the distribution of $Z = X/Y$.

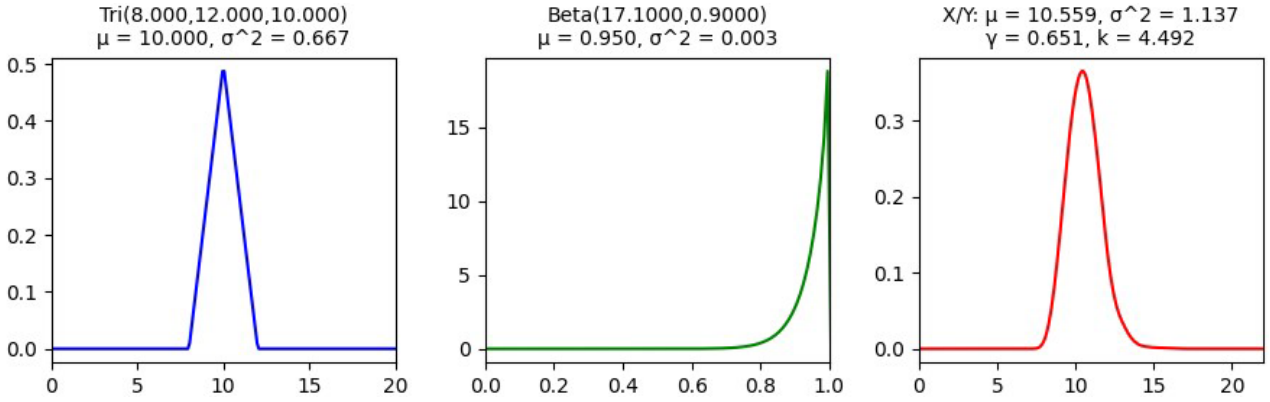
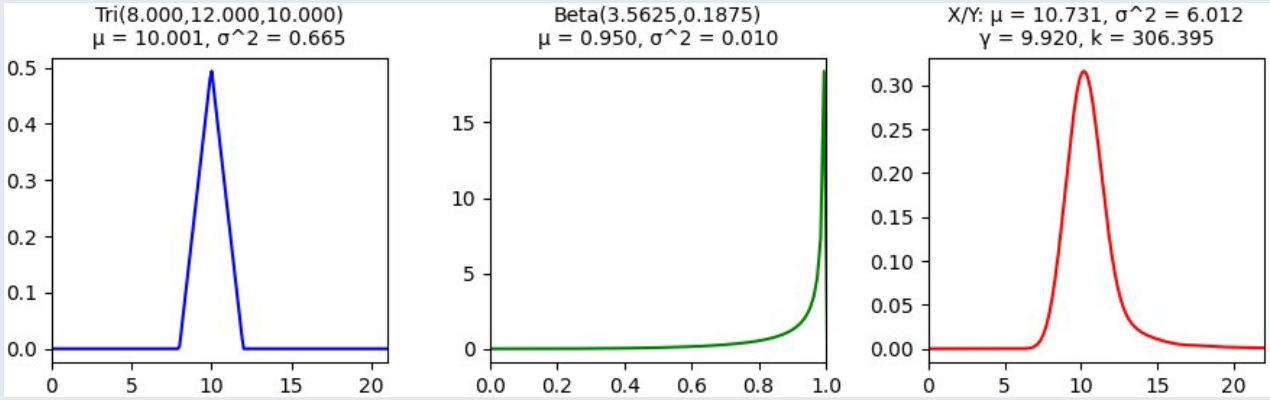
For each test, we also provide the low – order approximations of the mean value and the variance of the ratio, computed according to the rules summarized in Table 2.

The results are summarized in Table C1:

Table C1. Ratio *X*/*Y*: low order approximations of the mean value and variance.

X p.d.f.	*Y* p.d.f.	Ratio distribution (Gaussian Density Estimator)	1st and 2nd order estimations	Comments
			$\mathcal{S}_0\mu_Z = 22.675$ $\mathcal{S}_2\mu_Z = 22.676$ $\mathcal{S}_2\sigma_Z^2 = 11.579$ $\mathcal{S}_3\sigma_Z^2 = 11.579$ $\mathcal{S}_4\sigma_Z^2 = 11.579$	For an almost constant denominator, the 0st order approximations are almost exact, irrespectively of the skewness of the nominator.
			$\mathcal{S}_0\mu_Z = 19.999$ $\mathcal{S}_2\mu_Z = 20.532$ $\mathcal{S}_2\sigma_Z^2 = 21.320$ $\mathcal{S}_3\sigma_Z^2 = 21.889$ $\mathcal{S}_4\sigma_Z^2 = 23.089$	Denominator variance breaks the symmetry. Mean value's approximation error drops from 2.8% to 0.2%. Standard deviation error drops from 6.3% ($\sqrt{\mathcal{S}_2\sigma_Z^2}$) to 2.5% ($\sqrt{\mathcal{S}_4\sigma_Z^2}$).

X p.d.f.	*Y* p.d.f.	Ratio distribution (Gaussian Density Estimator)	1st and 2nd order estimations	Comments
			$\mathcal{S}_0\mu_Z$ = 20.000 $\mathcal{S}_2\mu_Z$ = 20.800 $\mathcal{S}_2\sigma_Z^2$ = 18.675 $\mathcal{S}_3\sigma_Z^2$ = 19.956 $\mathcal{S}_4\sigma_Z^2$ = 23.373	Using a beta distribution results in fatter tails. Mean value's approximation error drops from 4.3% to 0.5%. Std dev. error drops from 15% ($\sqrt{\mathcal{S}_2\sigma_Z^2}$) to 5% ($\sqrt{\mathcal{S}_4\sigma_Z^2}$).
			$\mathcal{S}_0\mu_Z$ = 20.000 $\mathcal{S}_2\mu_Z$ = 21.803 $\mathcal{S}_2\sigma_Z^2$ = 38.717 $\mathcal{S}_3\sigma_Z^2$ = 45.215 $\mathcal{S}_4\sigma_Z^2$ = 60.300	Ratio has very large variance and fat tails. The kurtosis calculation is unstable. Mean value's approximation error drops from 11% to 3%. Error in $\sqrt{\sigma_Z^2}$ drops from 37% ($\sqrt{\mathcal{S}_2\sigma_Z^2}$) to 21% ($\sqrt{\mathcal{S}_4\sigma_Z^2}$).

X p.d.f.	*Y* p.d.f.	Ratio distribution (Gaussian Density Estimator)	1st and 2nd order estimations	Comments
			$\mathcal{S}_0\mu_Z$ = 10.527 $\mathcal{S}_2\mu_Z$ = 10.556 $\mathcal{S}_2\sigma_Z^2$ = 1.047 $\mathcal{S}_3\sigma_Z^2$ = 1.048 $\mathcal{S}_4\sigma_Z^2$ = 1.123	Ratio is mildly skewed with thin tails. Mean value's approximation error drops from 0.30% to 0.03%, std dev error drops from 4% to 0.6%.
			$\mathcal{S}_0\mu_Z$ = 10.526 $\mathcal{S}_2\mu_Z$ = 10.642 $\mathcal{S}_2\sigma_Z^2$ = 1.959 $\mathcal{S}_3\sigma_Z^2$ = 1.986 $\mathcal{S}_4\sigma_Z^2$ = 3.342	Although the ratio has very a fat tail, both mean value estimates work well. Skewness and kurtosis calculations are unstable. Mean value's approximation error drops from 1.9% to 0.8%. Std dev error improves from 43% to 25%.

We may conjecture that with low denominator variance, the low order approximations of the mean value and variance are quite satisfactory, irrespectively of any skewness of the distributions X and Y. For larger values of the denominator variance, the variance estimation (hence, the estimation of the standard deviation) loses accuracy significantly, and the larger the variance and / or the support of the denominator, the larger the effect on the skewness and kurtosis of the ratio distribution and the error on the variance estimate. Still, the accuracy of the 2nd order estimate of $\mu_{X/Y}$ is consistently significantly better than the deterministic estimate, $\mathcal{S}_0\mu_{X/Y} = \mu_X/\mu_Y$.

Appendix D – The Agent Optimization application

The method presented in this paper forms the basis for the second version of the *Agent Optimization* application, developed during our participation in the Horizon 2020 *Grow Your Manufacturing Business (Better Factory)* project. The original version of the application, presented in (Körösi *et al.,* 2022), implemented the (Groover, 2008) model, as described by Equations (1 – 8). It is a command – line interface application, which reads data from a MongoDB database and calculates the *"optimal number of agents"*, and is part of the *Logistics Library*, formerly known as *Open Platform for Innovation in Logistics (OPIL)*.

The new version of the Agent Optimization application, developed by the authors, comes in two variants which can run as Docker containers: a command – line interface (CLI), like the first version, and a Web-based interface (Figure D.1). Both variants use the graph model when editing the quantities F_e and D_e of the links $e = (v_i, v_j)$ (Table 1), provide advanced data validation, allow reading and writing JSON files with the system data directly from / to disk, and encapsulate the underlying database, which becomes an optional component. In a nutshell, the second version facilitates the data manipulation significantly.

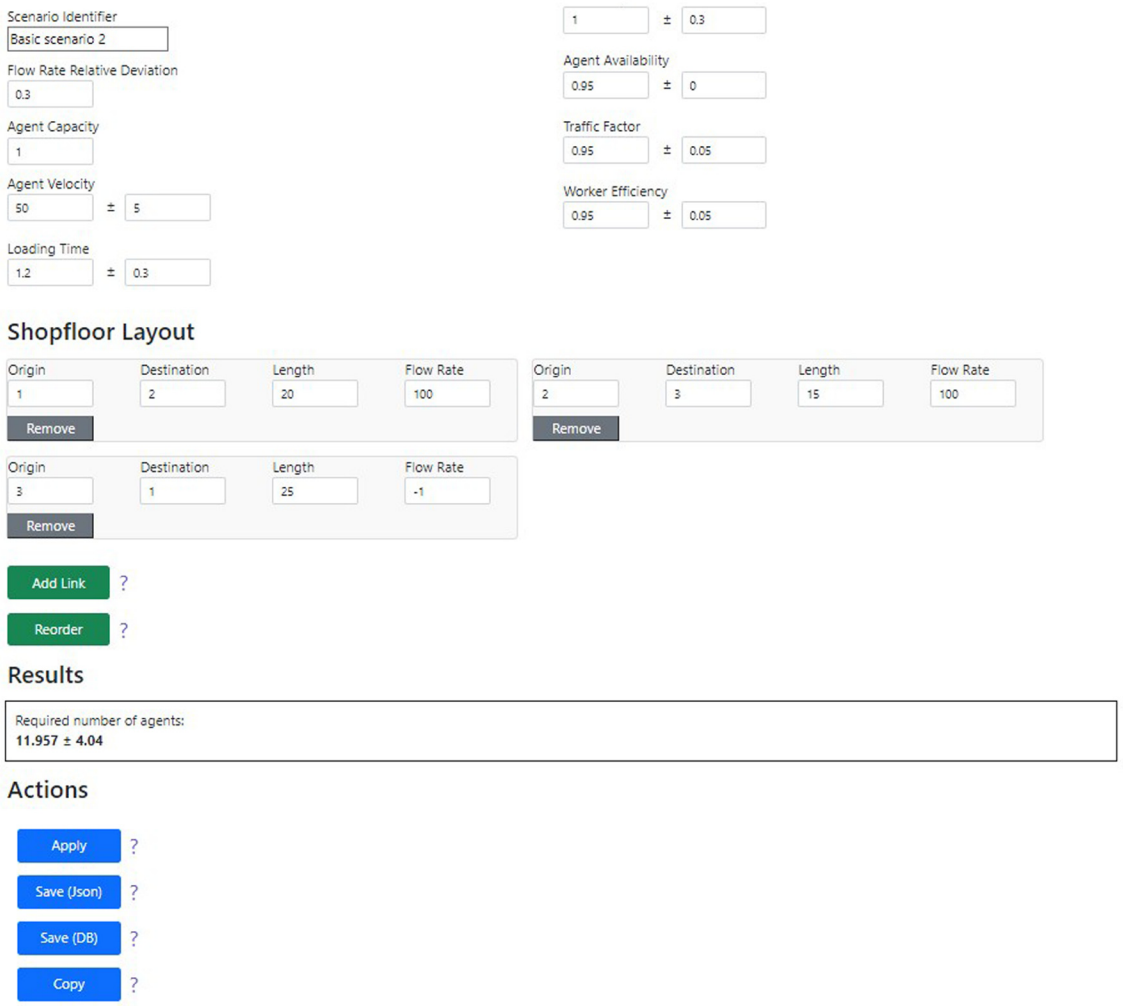


Figure D.1. Partial view of the Web-based user interface.

Figure D.2 depicts a high-level architecture of the system. The *Uncertain Quantity Calculator* subsystem maintains the mathematical model: mean values, standard deviations, finite or infinite ranges, and random variables encompassing a range (their support), a mean value and a standard deviation. The subsystem knows how to perform certain operations on the random variables: addition, subtraction, multiplication, and division. These are implemented according to the rules summarized in Table 1. Both mean values and variances are approximated up to 2^{nd} order, and the standard deviation is approximated by the square root of the (approximate) variance.

The *Agent Optimization Theory* subsystem maintains the model (1 – 8) and provides also any covariances required to compute the additions, multiplications, and divisions necessary, making use of the services of the *Uncertain Quantity Calculator* to perform the actual calculations.

This design allows easy adaptation to other problems that can be solved by a sequence of simple arithmetic operations.

The *Agent Optimization* application (ver. 2) can be downloaded from the *RAMP marketplace* (RAMP, n.d.).

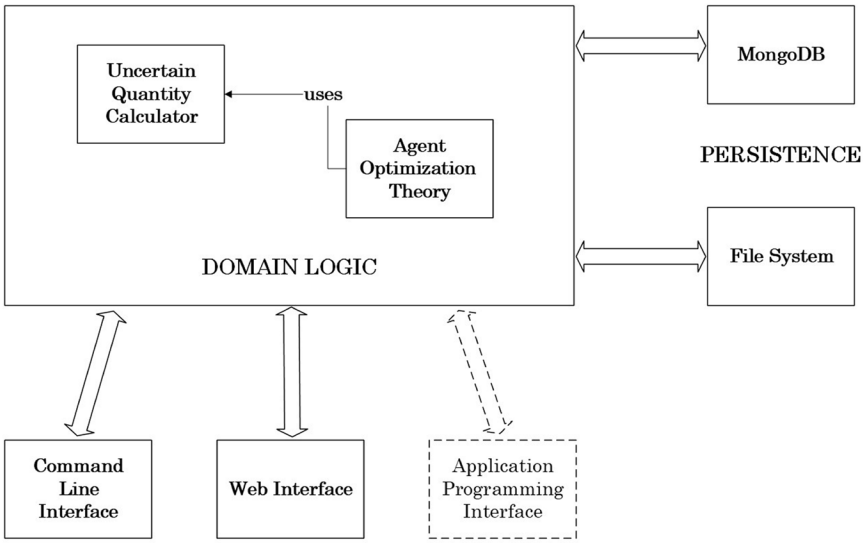


Figure D.2. *Agent Optimization (v.2)* high-level architecture.

Ethics and consent

Ethical approval and consent were not required.

Data availability

No data associated with this article.

References

Arifin R, Egbelu PJ: **Determination of vehicle requirements in Automated Guided Vehicle Systems: a statistical approach.** *Prod Plan Control.* 2000; **11**(3): 258–270.
Publisher Full Text

Beamon B, Chen VCP: **Performability-based fleet sizing in a Material Handling System.** *Int J Adv Manuf Technol.* 1998; **14**: 441–449.
Publisher Full Text

Casella G, Berger RL: **Statistical Inference.** 2nd edn. Duxbury Press, 2002.
Reference Source

Chawla V, Chanda A, Angra S: **Automatic Guided Vehicles fleet size optimization for Flexible Manufacturing System by Grey Wolf Optimization Algorithm.** *Management Science Letters.* 2018; **8**(2): 79–90.
Publisher Full Text

Egbelu P: **The use of non-simulation approaches in estimating vehicle requirements in an Automated Guided Vehicle based transport system.** *Material Flow.* 1987; **4**: 17–32.
Reference Source

Erdogan DE, Tasan SO: **Developing an optimization model to determine fleet size of Automated Guided Vehicle.** In: Mirzazadeh, A., Erdebilli, B., Babaee Tirkolaee, E., Weber, G.W. and Kar, A.K. (eds) *Science, engineering management and information technology: SEMIT 2022.* Communications in Computer and Information Science, Cham: Springer, 2023; **1808**.
Publisher Full Text

Fitzgerald KR: **How to estimate the number of AGVs you need.** *Modern Materials Handling.* 1985; **10**: 79.

Glasserman P: **Monte Carlo methods in financial engineering (Stochastic Modelling and Applied Probability).** Springer, 2003; **53**.
Publisher Full Text

Gobal SL, Kasilingam RG: **A simulation model for estimating vehicle requirements in Automated Guided Vehicle Systems.** *Comput Ind Eng.* 1991; **21**(1–4): 623–627.
Publisher Full Text

Groover MP: **Automation, production systems, and computer-integrated manufacturing.** 3rd edn. Pearson Education, 2008.
Reference Source

Hadas Y, Figliozzi MA: **Modeling optimal drone fleet size considering stochastic demand.** *EURO Journal on Transportation and Logistics.* 2024; **13**: 100127.
Publisher Full Text

Johnson ME, Brandeau ML: **Stochastic modeling for automated Material Handling System design and control.** *Transport Sci.* 1996; **30**(4): 330–350.
Publisher Full Text

Johnson NL, Kotz S, Balakrishnan N: **Continuous Univariate Distributions, Volume 1**. 2nd edn. Wiley-Interscience, 1994.
Reference Source

Kavakeb S, Nguyen TT, Yang Z, *et al.*: **Evolutionary fleet sizing in static and uncertain environments with shuttle transportation tasks - the case studies of container terminals.** *IEEE Comput Intell Mag.* 2016; **11**(1): 55–69.
Publisher Full Text

Körösi L, Duchoň F, Lukáč P: **Open Platform for Innovation in Logistics – Agent Optimization**. 2022.
Publisher Full Text

Le-Anh T, de Koster MBM: **A review of design and control of Automated Guided Vehicle Systems.** *Eur J Oper Res.* 2006; **171**(1): 1–23.
Publisher Full Text

Rajotia S, Shanker K, Batra JL: **Determination of optimal AGV fleet size for an FMS.** *Int J Prod Res.* 1998; **36**(5): 1177–1198.
Publisher Full Text

RAMP: **RAMP**. (n.d.); (Accessed: 24 November 2024).
Reference Source

Şenaras OM, Solmaz E, Öztürk N, *et al.*: **Determination of the fleet size of AGVs with AGV pools using a genetic algorithm and Artificial Intelligence.** *Appl Sci.* 2023; **13**(13): 7994.
Publisher Full Text

Touran A, Taher KAH: **Optimum fleet size determination by queuing and simulation.** *Construction Management and Economics.* 1988; **6**(4): 295–306.
Publisher Full Text

Valmiki P, Reddy AS, Panchakarla G, *et al.*: **A study on simulation methods for AGV fleet size estimation in a Flexible Manufacturing System.** *Mater Today Proc.* 2018; **5**(2, Part 1): 3994–3999.
Publisher Full Text

Vis IFA, de Koster R, Roodbergen KJ, *et al.*: **Determination of the number of Automated Guided Vehicles required at a semi-automated container terminal.** *J Oper Res Soc.* 2001; **52**(4): 409–417.
Publisher Full Text

Vivaldini K, Rocha LF, Martarelli NJ, *et al.*: **Integrated tasks assignment and routing for the estimation of the optimal number of AGVs.** *Int J Adv Manuf Technol.* 2016; **82**: 719–736.
Publisher Full Text

Wallace DL: **Asymptotic approximations to distributions.** *Ann Math Statist.* 1958; **29**(3): 635–654.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 12 June 2025

https://doi.org/10.21956/openreseurope.20560.r54489

© **2025 Hodge C.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Cabell Hodge
National Renewable Energy Laboratory, Lakewood, USA

The model presented herein improves on the estimate of mean vehicle counts over the deterministic model. The math appears sound, and the level of detail provided makes it easy to replicate. However, the stated intention of the article is to improve on calculating the variability of real-world operational conditions, and it's unclear that the methods achieve that goal. The variance referenced in the Examples and Discussion section is off by 13-33% per standard deviation, indicating that it isn't the strongest predictor of the maximum fleet count size. In practical applications, the most important number of warehouse operating vehicles is the maximum fleet count. It would be worth exploring how to calibrate the model around the second or third standard deviation in order to more accurately reflect the higher bounds of the fleet count.

Is the rationale for developing the new method (or application) clearly explained?
Yes

Is the description of the method technically sound?
Yes

Are sufficient details provided to allow replication of the method development and its use by others?
Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Fleet electrification, charger count optimization, and vehicle-grid integration.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 02 June 2025

https://doi.org/10.21956/openreseurope.20560.r53801

© 2025 Bojovic N. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Nebojsa Bojovic
University of Belgrade, Belgrade, Serbia

Review report
Improving Fleet Size Estimation for Factory Logistics Under Demand, Speed, and Efficiency Uncertainty
This paper presents an approach to improving fleet size estimation for Automated Guided Vehicles (AGVs) operating in factory logistics systems. The authors develop a stochastic modelling framework incorporating uncertainty in demand, speed, and operational efficiency. By applying moment-based approximations and operator-based expansions, the proposed model accounts for multiplicative and ratio-based relationships between uncertain parameters.
While the paper offers clear scientific contributions, particularly in its mathematical depth and theoretical modelling, there are notable issues concerning structure, clarity, and certain methodological assumptions that require improvement.
My remarks are as follows:

- Abstract and Introduction: The abstract does not clearly summarize the methodology or the quantitative findings. Additionally, the introduction could be more sharply focused on the problem motivation and research gap, with clearer research questions or hypotheses.
- Literature Review: The paper lacks a distinct section explicitly reviewing the state-of-the-art. The positioning of the work relative to prior deterministic and stochastic fleet sizing approaches is only implied and should be made more explicit.
- Conclusion Section: The conclusion is too brief and lacks reflection on the limitations of the proposed method and avenues for future research. It should include a more critical discussion of model assumptions and potential extensions.
- Assumption of Distributions: The methodology assumes that input variables (demand, speed, efficiency) follow normal or approximately symmetric distributions. In industrial applications, these variables may be skewed or follow empirical distributions. The impact of non-normality is not discussed.
- Scalability and Complexity: While the approach is mathematically elegant, the

computational implications of using operator-based expansions in large-scale systems are not addressed.
- Lack of Real-World Case Study: Although the paper references application in a simulation tool, it does not include an empirical case study or sensitivity analysis based on real-world data, which would strengthen the practical relevance.
- Comment on Section “The Impact Variability”: While the section provides a useful summary of how input variability (e.g., in demand, speed, efficiency) affects output uncertainty, much of the content reflects well-established statistical principles (e.g., propagation of variance through products and ratios). As such, this section may not need to stand on its own.
- Typo on Page 22: “anther”.
- Figure clarity on Page 22: “dashed blue line” – not visible.

Is the rationale for developing the new method (or application) clearly explained?
Yes

Is the description of the method technically sound?
Yes

Are sufficient details provided to allow replication of the method development and its use by others?
Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Operations Research, Transportation, Logistics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.