Wait Time Impact Model at Pre-Board Screening Checkpoints for Canadian Airports (with Enhancements)

Technical Leads: Dr. Yiqiang Zhao, Dr. Patrick Boily
Project Consultants: Wenzhe Ye, Dr. Katrina Rogers-Stewart, Shintaro Hagiwara

Centre for Quantitative Analysis and Decision Support

Original Report: August 30, 2013
Revised: December 15, 2013
Enhancements: December 4, 2015

Contents

List of Figures

List of Tables

1 Introduction

Numerous factors influence the wait time at Canadian airports' pre-board screening (PBS) checkpoints: the schedule intensity of departing flights, the volume of passengers on these flights, the number of servers and processing rates at a given checkpoint, etc.

The Canadian Air Transport Security Agency (CATSA) is tasked to ensure that the PBS experience at Canadian airports is made as efficient as possible by minimizing the waiting time at checkpoints.

1.1 Objectives

Queuing Theory (QT) can be used to develop a Wait Time Impact Model (WTIM), which will satisfy the following objectives:

1. provide estimates of the passenger arrival rates, the processing rates and the number of servers at checkpoints from the field data available for all checkpoints;
2. for a given arrival rate, processing rate and number of servers, calculate the Quality of Service (QoS) level under an appropriate queueing model assumption and determine what service level can be achieved at the checkpoint (i.e. the percentage p of passengers which will wait less than x minutes);
3. provide the average number of servers required to achieve a prescribed QoS level, given an arrival profile in the queueing model, and
4. allow for the analysis of various scenarios (such as passenger growth, for instance) via the tweaking of a small number of parameters and whenever the available data is updated.

1.2 Outline

The model establishes a relationship between the arrival rates, the service rates, the number of servers and the service levels. Basic concepts, process descriptions, and limitations are provided in Section 2.

The WTIM is best described *via* the flow-chart of Figure 1 (the various concepts will be defined as they arise in the corresponding section):

1. computation of the arrival rates λ from the raw data (Section 3.2);
2. computation of the distribution of the number of servers c from the checkpoint utilization reports (Section 3.3);
3. computation of the waiting time distribution W_q from the waiting time report (Section 3.4);
4. computation of the QoS levels (p, x) from the waiting time report (Section 3.4);
5. computation of the estimated QoS levels $(\hat{p}_M, x)$ under the $M/M/1$ assumption (Section 3.5);

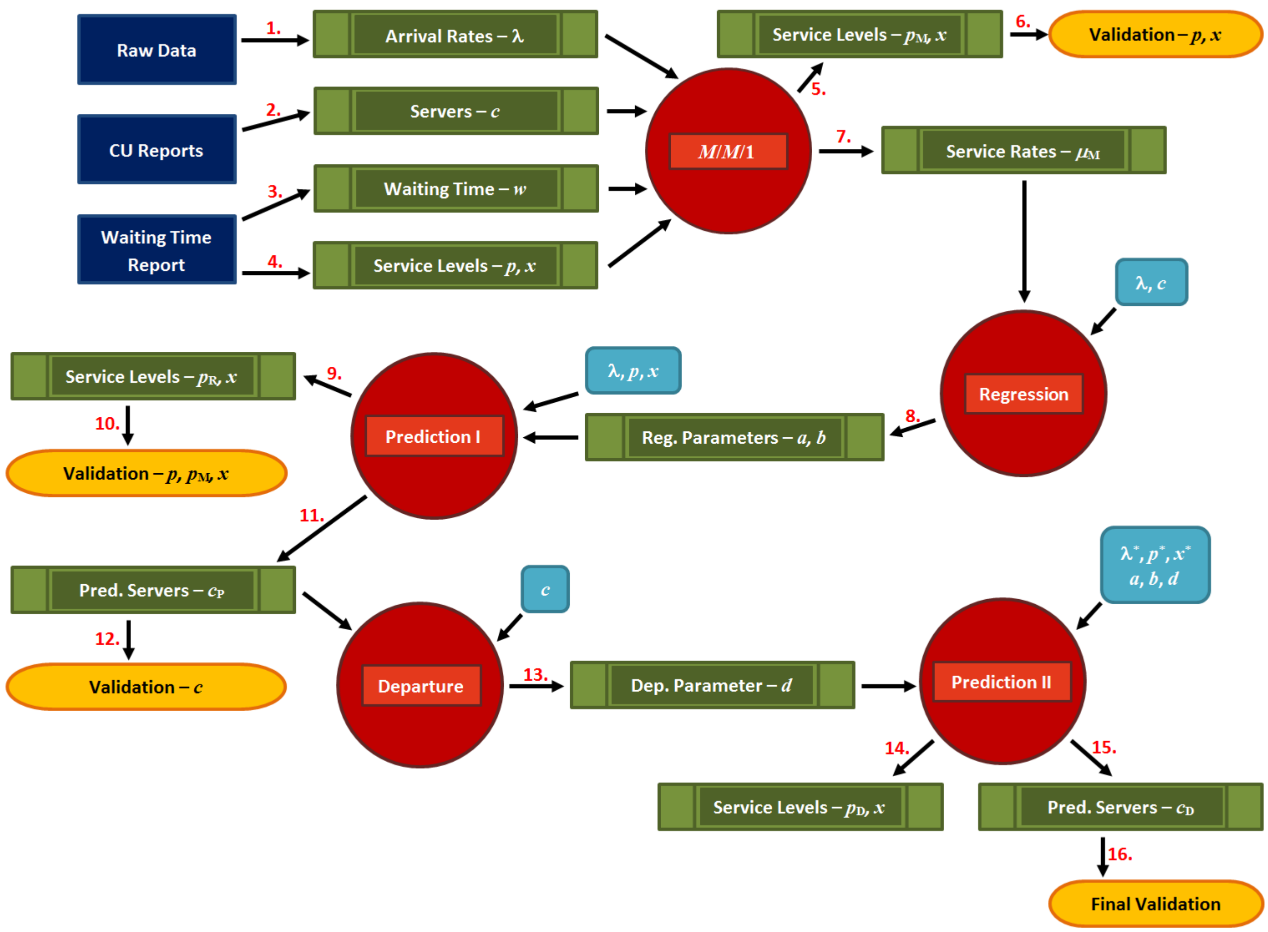


Figure 1: WTIM flow. The dark blue parallelograms are CATSA-provided data inputs; the green boxes indicate computed and derived values; the red circles are conceptual nodes; the light blue boxes represent carry-over values, and the orange cells are validation steps.

6. validation of the $M/M/1$ assumption based on a comparison of $(\hat{p}_M, x)$ and (p, x) (Section 3.6);

7. computation of the estimated service rates $\hat{\mu}_M$ under the $M/M/1$ assumption (Section 3.5);

8. computation of the seasonal checkpoint regression parameters a and b under the combined $M/M/1$ and Regression assumptions (Section 4.1);

9. computation of the estimated QoS levels $(\hat{p}_R, x)$ under the combined $M/M/1$ and Regression assumption (Section 4.2);

10. validation of the combined $M/M/1$ and Regression assumptions based on a comparison of $(\hat{p}_R, x)$, $(\hat{p}_M, x)$ and (p, x) (Section 4.3);

11. prediction of the number of servers c_R under the combined $M/M/1$ and Regression assumptions (Section 5);

12. validation of the combined $M/M/1$ and Regression assumptions based on a comparison of c_R and c (Section 5.3);

13. computation of the checkpoint departure parameters d under the combined $M/M/1$, Regression and Departure assumptions (Section 5.3);

14. computation of the estimated QoS levels $(\hat{p}_D, x)$ for various projected arrival growth rates λ^* under the combined $M/M/1$, Regression and Departure assumptions (Section 6.2);

15. prediction of the number of servers c_D for various projected arrival growth rates λ^* under the combined $M/M/1$, Regression and Departure assumptions (Section 6);

16. final validation of the combined $M/M/1$, Regression and Departure assumptions based on a comparison of $(\hat{p}_D, x)$ and c_D with empirical data (Section 6.3).

In order to illustrate the WTIM process, its details are worked out on a step-by-step basis for the Domestic/International Checkpoint at the Edmonton International Airport (YEG), based on 2012 data. The results are shown at the end of each sections, under the heading **YEG (DI) — 2012 (continued)**.

A summary of results for all checkpoints is also provided, as well as recommendations and suggested next steps.

2 Preliminaries

2.1 Definitions

In this section, the various mathematical concepts to which the report will refer are described.

- An $M/M/c$ **queueing model** describes a system where arrivals form a single queue and are governed by a Poisson process (the first M), units arriving are processed by c servers and service times are exponentially distributed (the second M).

- A **Poisson process** is a stochastic process where the time between any two consecutive event has an exponential distribution with parameter λ.

- The **arrival rate** is the rate at which passengers arrive for PBS (i.e. passengers per minute), the **service rate** is the processing rate at a screening line (i.e. maximal potential throughput), the **number of servers** is the number of screening lines and the **service level** is the percentage of people waiting less than a given number of minutes at a checkpoint.

2.2 Description of PBS Process

At each checkpoint, the PBS process is structurally similar: passengers arriving at the beginning of the main queue may have their boarding passes scanned at the S_1 position, but they are always scanned at the S_2 position (see figure 2).

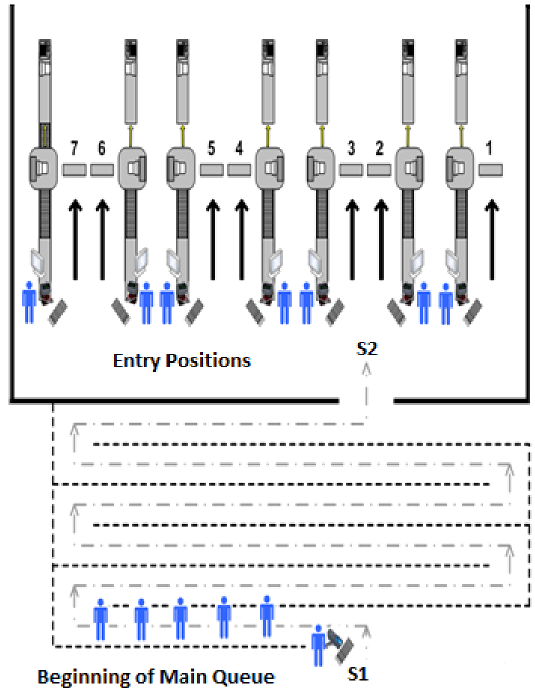


Figure 2: Schematics of pre-board screening (PBS). Passengers enter the main queue, where their boarding pass may be screened at S_1. Once they reach the end of the main queue, their boarding pass is screened at S_2 and they are sent to one of the active lines for processing.

2.3 Available Data Sources

For each checkpoint, CATSA provides three datasets.

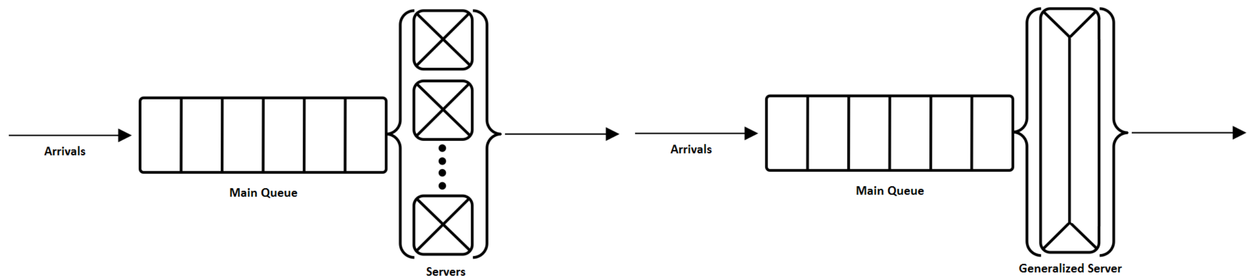
Raw Data: for each passenger scanned once they reach the end of the main queue, this dataset records the date, the scan time upon entering the main queue (S_1), the scan time upon exiting the main queue (S_2), and the wait time between S_1 and S_2. As a passenger may not have been scanned upon entering the main queue, the fields for S_1 and the wait are sometimes empty. The Raw Data contains other variables as well, but they are not used by the WTIM at this stage.

Checkpoint Utilization Report: for each day of the year and each 15—minute block, this dataset records the maximum number of open lines. The CU Report contains other variablesas well, but they are not used by the WTIM at this stage.

Waiting Time Report: consists of the subset of Raw Data for which S_1 and S_2 are both available. Observations for which the wait time exhibits outlying behaviour have also been removed.

3 $M/M/1$ Queueing Model

One of the difficulties for the situation under consideration is that the number of servers varies with time, according to different factors: there are times when all servers are busy, others when a number of open servers are idle, and the number of open servers changes according to some



(a) $M/M/c$ **(b)** $M/M/1$

Figure 3: Conceptual visualization of an $M/M/c$ queueing system as an $M/M/1$ system: the c servers can be considered as 1 generalized server.

vacation policy which it is difficult to model. This is problematic when using an $M/M/c$ model as service rate estimates depend, amongst other things, on the number of open servers.

It is possible to circumvent this issue altogether, without invoking Vacation Models, by noticing that an $M/M/c$ queueing system may be viewed as an $M/M/1$ queueing system, where the servers are hidden behind a generalized server (see Figure 3). Under that interpretation, the service rates can be estimated independently of the number of servers. Furthermore, not only do $M/M/c$ results still hold for $M/M/1$ (simply by setting $c = 1$ in the appropriate theorems), but the quantities to be computed tend to be simpler in the generalized case.

While this conceptual simplification has removed some of the difficulties associated to server vacation, there remains, another problem: the theory of $M/M/1$ systems, alone, is not sufficient to recover (and later predict) the actual (and hidden) number of servers for the checkpoint. This situation can be addressed by finding another way to link the arrival rates, the estimated service rates and the number of servers (see Section 4.1).

3.1 Clustering

In order to better predict the average behaviour of a system and its possible outcomes, a wide range of typical patterns must be considered. When analyzing the behaviour of queues, it may become necessary to group the data into meaningful **clusters** exhibiting similar properties (for example, properties that can be characterized by the same Poisson process).

This approach allows for proper estimation of queuing model parameters (arrival rates, processing rates, etc.), which in turn yields the most reliable results. The selection of the appropriate cluster size relies on finding a balancing point between two extremes.

- In order to properly define the stochastic process, a minimum amount of data with similar properties is required. If clustering is not performed (i.e., if the clusters are too large), the data may present different characteristics which cannot be represented by a single Poisson process.
- On the other hand, if the clusters are too small, they may not contain enough data to capture

the underlying properties. More importantly, clusters that cover too short a period are unlikely to exhibit the statistical behaviour of the process.

A preliminary analysis of the model's accuracy was assessed based on the following clustering criteria:

- Checkpoint
- Weekly patterns (day of week versus weekday/weekend)
- Seasonal patterns (season versus month)
- Daily patterns (2-hour period versus 4-hour period)

The cluster combination that produced the most encouraging queueing results when compared against actual reports was: checkpoint, weekday/weekend, season, 4 hour-period.

Clustering also plays a role in the Regression stage of the model (Section 4), but the optimal regression cluster combination need not be the same as the **queueing cluster combination**.

3.2 Computing the Average Arrival Rate

Since not all boarding passes are scanned at S_1, the Wait Time report (S_1 data) cannot be used to derive the cluster arrival rates.

The S_1—S_2 line-up (main queue) is a birth-death process (i.e. a reversible one-dimensional Markov chain). In particular, the forward chain $S_1 - S_2$ and the reversed chain $S_2 - S_1$ are stochastically identical and so the arrival epochs of the reversed chain are the departure epochs of the forward chain. We can then use Burke's Theorem for $M/M/c$ queues at steady states.

Theorem 1 (BURKE'S THEOREM, [1]) *Consider an $M/M/c$ queue in the steady state with arrivals modeled by a homogeneous Poisson process with rate parameter λ. Then the departure process is also a homogeneous Poisson process with rate parameter λ.*

This does not rule out the possibility that, at a particular time, the arrivals at S_1 could be greater than the departures at S_2, due to the inherent randomness of Poisson processes. But all S_1 arrivals will eventually leave at S_2 and thus the fluctuations at S_2 follow the same statistical property governing arrivals to the queue. Therefore, the arrival rates can be estimated by using data readings at S_2 within a given cluster.

It remains only to show that arrivals follow a homogeneous Poisson process in each cluster (this is a common hypothesis). To do so, one must show, assuming is the number of arrivals in the cluster by time t is denoted by $N(t)$, that (see [4, 3])

1. $N(t)$ is a counting process;
2. $N(t)$ has independent and stationary increments;

Cluster				# of Hours	Count	Avg Arrival Rate
Jan 01 to Mar 31 - YEG (DI) - 2012	Week day	0:00	4:00	260	844	0.055
		4:00	8:00	260	129,069	8.274
		8:00	12:00	260	97,949	6.279
		12:00	16:00	260	84,548	5.420
		16:00	20:00	260	78,964	5.062
		20:00	0:00	260	33,061	2.119
	Week-end	0:00	4:00	104	1,076	0.172
		4:00	8:00	104	39,674	6.358
		8:00	12:00	104	31,200	5.000
		12:00	16:00	104	26,136	4.188
		16:00	20:00	104	28,129	4.508
		20:00	0:00	104	10,013	1.605

(a) First quarter

Cluster				# of Hours	Count	Avg Arrival Rate
Apr 01 to Jun 30 - YEG (DI) - 2012	Week day	0:00	4:00	260	1,068	0.070
		4:00	8:00	260	128,655	8.247
		8:00	12:00	260	106,704	6.840
		12:00	16:00	260	87,208	5.590
		16:00	20:00	260	82,198	5.269
		20:00	0:00	260	34,330	2.201
	Week-end	0:00	4:00	104	626	0.100
		4:00	8:00	104	35,923	5.757
		8:00	12:00	104	35,683	5.718
		12:00	16:00	104	25,564	4.097
		16:00	20:00	104	24,489	3.925
		20:00	0:00	104	11,735	1.881

(b) Second quarter

Cluster				# of Hours	Count	Avg Arrival Rate
Jul 01 to Sep 30 - YEG (DI) - 2012	Week day	0:00	4:00	260	4,256	0.281
		4:00	8:00	260	128,186	8.345
		8:00	12:00	260	113,577	7.394
		12:00	16:00	260	87,439	5.605
		16:00	20:00	260	82,053	5.260
		20:00	0:00	260	44,213	2.834
	Week-end	0:00	4:00	108	1,781	0.285
		4:00	8:00	108	40,218	6.206
		8:00	12:00	108	41,898	6.466
		12:00	16:00	108	30,237	4.666
		16:00	20:00	108	26,675	4.117
		20:00	0:00	108	15,665	2.417

(c) Third quarter

Cluster				# of Hours	Count	Avg Arrival Rate
Oct 01 to Dec 31 - YEG (DI) - 2012	Week day	0:00	4:00	260	1,114	0.074
		4:00	8:00	260	132,094	8.468
		8:00	12:00	260	102,019	6.540
		12:00	16:00	260	87,806	5.629
		16:00	20:00	260	83,881	5.377
		20:00	0:00	260	35,769	2.293
	Week-end	0:00	4:00	104	771	0.134
		4:00	8:00	104	38,196	6.121
		8:00	12:00	104	38,538	6.176
		12:00	16:00	104	26,683	4.276
		16:00	20:00	104	25,399	4.070
		20:00	0:00	104	11,879	1.904

(d) Fourth quarter

Table 1: Total number of hours, count of arrivals and average arrival rates, per cluster, per quarter.

3. The number of arrivals in any time interval of length t is Poisson-distributed with mean λt, i.e. for all $s, t \geq 0$,

$$P(N(t+s)-N(s)=n)=e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n=0,1,\ldots$$

The first assumption is obviously satisfied. The second assumption is satisfied with the introduction of clusters. The third assumption holds if the **inter-arrival times** (the times between consecutive events) are independent and identically distributed (i.i.d.) exponential random variables with the same rate λ: analysis of S_2 in the raw data with EasyFit suggests that this is indeed the case.

YEG (DI) − 2012

The total count of arrivals for each cluster are shown in Table 1. Notice that the arrival rate λ is simply calculated by dividing the count in each cluster by the number of minutes (the number of hours × 60 minutes) in each cluster, independently of whether the checkpoint was always open or not during period spanned by the cluster. A low arrival rate may thus indicate either that checkpoint traffic was low or intermittent for the cluster, or that it was closed for some or all of the period that it spans.

3.3 Computing the Average Number of Servers

The number of active servers (open lines) at each checkpoint can be adjusted at any moment during each time period, in order to accommodate fluctuations in arrivals. The CU reports do not quite record the number of active servers (open lines) or the average number of active servers during each 15-minute block; rather, they record the *maximum* number of simultaneously active servers for each block. It is reasonable to hope that the discrepancy between the actual numbers and the reported number is fairly small, due to the short duration of the blocks.

At any rate, data is not available for smaller time scales. As long as the distinction between the theoretical c and the reported c is kept in mind when interpreting the results, this issue is unlikely to cause serious problems.

YEG (DI) — 2012 (continued)

A cluster-by-cluster distribution of the number of active servers is easy to compute (see Table 2). Clusters for which the average arrival rate is low (as seen in Table 1) tend to have distributions with low number of servers, whereas those with high traffic rarely have a small number of active servers.

3.4 Computing the Average Wait Time and the Empirical QoS

As has been discussed previously, not all wait time data is available since a number of passengers did not get their boarding passes scanned at S_1. However, if the subset of those passengers for which there is an S_1 scan is fairly representative of the larger and more comprehensive raw data, it is reasonable to expect that the parameters and quantiles of the wait time distribution can be estimated from the subset provided by the wait time report.

Of course, since the full wait time data is inaccessible, it is impossible to verify whether this assumption of representativeness is met in reality. But it appears that there are three main reasons why a raw data observation is not included in the wait time report:

1. the passenger was scanned at S_1, but the calculated wait time $w = S_2 - S_1$ is classified as an outlier because it is uncharacteristically large compared to neighbouring passenger scans (the passenger might have left the main queue for any number of reasons);

2. the passenger was not scanned at S_1 because too many ppassengers were entering the main queue at roughly the same time and the S_1 scanner was overwhelmed, or

3. the main queue was empty when the passenger arrived and so the passenger was processed immediately, leading to $w = 0$.

In the first two instances, the absence of wait time data in the subset does not introduce a bias in the estimates. However, that's not necessarily the case for the third instance, as, if a large number of such observations were removed to create the wait time report, the estimates are likely to be biased. This is likely to affect the predicted QoS levels in the small wait time regime. Finally, it should be noted that it is possible that a passenger enters the main queue in one cluster and leaves the main queue in another cluster, especially if the passenger entered near the end of a cluster. In order to remain compatible with the computation of the cluster arrival rates, the cluster in which the wait time $w = S_2 - S_1$ is recorded is the cluster in which S_2 falls.

YEG (DI) — 2012 (continued)

The average wait time and quantiles are shown in Table 3. Note that there are clusters for which no wait time data was collected, and that the number of wait time observations is smaller than the corresponding arrivals in each cluster (see Table 1 for a comparison).

3.5 Estimating the Service Rates and the Performance Levels

Consider an $M/M/c$ queue. The probability that a passenger has to wait upon entering the queue is given by

$$C(c, c\rho) = \frac{\left(\frac{(c\rho)^c}{c!}\right)\left(\frac{1}{1-\rho}\right)}{\sum_{k=0}^{c-1} \frac{(c\rho)^k}{k!} + \left(\frac{(c\rho)^c}{c!}\right)\left(\frac{1}{1-\rho}\right)},$$

where λ is the arrival rate, μ is the **service rate**, c is the number of servers, and $\rho = \lambda/(c\mu)$ is the **traffic intensity** of the system, while the wait time distribution for the queue is the conditional exponential distribution satisfying

$$P(\text{Wait time } \leq x) = P(W_q \leq x) = 1 - C(c, \rho)e^{-(c\mu-\lambda)x}.$$

for $x > 0$ [3]. From this, it is possible to conclude that the average wait time $\overline{W}_q$ of a passenger in such a queue is given by

$$\overline{W}_q = \int_0^\infty xP'(W_q \leq x)\,dx = \frac{C(c, \rho)}{c\mu - \lambda}.$$

For the generalized $M/M/1$ queue, this translates to

$$C(1, \rho) = \rho, \quad p(x) = P(W_q \leq x) = 1 - \rho e^{-(\mu-\lambda)x} \quad \text{and} \quad \overline{W}_q = \frac{\rho}{\mu - \lambda}. \tag{1}$$

In particular, if the processing rate μ is unknown but the arrival rate λ is known and the average wait time $\overline{W}_q$ can be computed by other means, then it is possible to recover μ from the last equality in (1):

$$\hat{\mu}_M = \frac{\overline{W}_q\lambda + \sqrt{\overline{W}_q^2\lambda^2 + 4\overline{W}_q\lambda}}{2\overline{W}_q}, \tag{2}$$

assuming $\overline{W}_q > 0$ (the negative solution being discarded). Note that, in theory, the estimated traffic intensity $\hat{\rho}_M = \lambda/\hat{\mu}_M < 1$, which means that $\hat{\mu}_M - \lambda > 0$ and that the QoS levels can be estimated by

$$\hat{p}_M(x) = 1 - \hat{\rho}_M e^{-(\hat{\mu}_M-\lambda)x} \in (0, 1) \quad \text{for all } x > 0. \tag{3}$$

In practice, however, the queueing system is not an exact (generalized) $M/M/1$ queue, and even if it were, the exact cluster arrival rates and average waiting times can at best estimated from the

Cluster				Avg # of Servers	Distribution of # of Active Servers								
					0	1	2	3	4	5	6	7	8
Jan 01 to Mar 31 - YEG (DI) - 2012	Week day	0:00	4:00	0.14	86.7%	13.0%	0.3%	-	-	-	-	-	-
		4:00	8:00	5.38	-	7.6%	11.3%	4.4%	5.6%	11.3%	21.3%	20.4%	18.1%
		8:00	12:00	4.63	-	-	0.9%	10.3%	35.2%	33.6%	19.0%	1.0%	0.1%
		12:00	16:00	4.19	-	-	3.1%	21.9%	37.6%	27.8%	9.2%	0.4%	-
		16:00	20:00	3.78	-	-	17.7%	30.7%	22.3%	17.6%	9.2%	2.5%	-
		20:00	0:00	0.58	49.4%	42.9%	7.5%	0.2%	-	-	-	-	-
	Week-end	0:00	4:00	0.21	82.5%	13.9%	3.6%	-	-	-	-	-	-
		4:00	8:00	4.56	-	1.9%	9.4%	10.1%	20.7%	32.2%	18.8%	7.0%	-
		8:00	12:00	3.92	-	-	1.0%	31.3%	46.6%	18.0%	2.4%	0.7%	-
		12:00	16:00	3.41	-	-	6.3%	51.9%	36.8%	4.6%	0.5%	0.0%	-
		16:00	20:00	3.60	-	0.7%	17.5%	38.2%	18.8%	15.6%	8.2%	0.5%	0.5%
		20:00	0:00	1.47	0.2%	56.3%	39.7%	3.8%	-	-	-	-	-

(a) First quarter

Cluster				Avg # of Servers	Distribution of # of Active Servers								
					0	1	2	3	4	5	6	7	8
Apr 01 to Jun 30 - YEG (DI) - 2012	Week day	0:00	4:00	0.15	86.3%	12.1%	1.6%	-	-	-	-	-	-
		4:00	8:00	4.85	6.3%	3.8%	11.5%	6.0%	6.0%	15.0%	23.8%	22.4%	5.2%
		8:00	12:00	4.69	-	-	1.0%	10.8%	32.7%	31.4%	22.5%	1.6%	-
		12:00	16:00	4.12	-	-	3.2%	24.4%	38.8%	24.7%	8.8%	0.1%	-
		16:00	20:00	3.82	0.2%	-	3.2%	33.8%	43.5%	16.0%	3.3%	0.2%	-
		20:00	0:00	1.69	1.3%	37.1%	53.4%	8.0%	0.2%	-	-	-	-
	Week-end	0:00	4:00	0.17	84.6%	13.7%	1.7%	-	-	-	-	-	-
		4:00	8:00	4.03	6.3%	1.2%	6.3%	13.7%	26.7%	34.6%	10.3%	1.0%	-
		8:00	12:00	4.08	-	-	1.4%	21.6%	50.5%	20.0%	6.5%	-	-
		12:00	16:00	3.58	-	-	7.0%	40.1%	41.6%	10.6%	0.7%	-	-
		16:00	20:00	3.45	-	-	14.9%	41.1%	30.3%	11.8%	1.9%	-	-
		20:00	0:00	1.84	-	27.4%	61.5%	10.6%	0.5%	-	-	-	-

(b) Second quarter

Cluster				Avg # of Servers	Distribution of # of Active Servers								
					0	1	2	3	4	5	6	7	8
Jul 01 to Sep 30 - YEG (DI) - 2012	Week day	0:00	4:00	0.26	77.5%	19.0%	3.5%	-	-	-	-	-	-
		4:00	8:00	4.34	12.0%	5.2%	8.0%	6.3%	11.6%	15.2%	21.3%	19.4%	0.9%
		8:00	12:00	4.29	-	-	4.0%	17.7%	35.4%	31.7%	10.5%	0.7%	-
		12:00	16:00	3.53	-	0.4%	8.7%	39.6%	40.5%	10.4%	0.5%	-	-
		16:00	20:00	3.32	0.3%	0.3%	9.7%	52.0%	32.5%	4.8%	0.4%	-	-
		20:00	0:00	1.84	1.8%	23.8%	63.5%	10.5%	0.4%	-	-	-	-
	Week-end	0:00	4:00	0.27	75.2%	22.2%	2.5%	-	-	-	-	0.0%	-
		4:00	8:00	3.59	11.6%	0.9%	6.9%	21.5%	26.6%	22.0%	9.7%	0.7%	-
		8:00	12:00	3.75	-	-	9.3%	30.1%	40.3%	17.4%	2.5%	0.5%	-
		12:00	16:00	3.11	-	-	18.5%	54.4%	24.3%	2.8%	-	-	-
		16:00	20:00	3.08	-	-	21.5%	52.1%	24.1%	1.9%	0.5%	-	-
		20:00	0:00	2.13	-	14.6%	60.2%	23.1%	2.1%	-	-	-	-

(c) Third quarter

Cluster				Avg # of Servers	Distribution of # of Active Servers								
					0	1	2	3	4	5	6	7	8
Oct 01 to Dec 31 - YEG (DI) - 2012	Week day	0:00	4:00	0.18	83.8%	15.2%	0.9%	0.2%	-	-	-	-	-
		4:00	8:00	5.17	0.5%	5.8%	12.5%	4.5%	8.6%	15.8%	18.6%	23.0%	10.9%
		8:00	12:00	4.78	-	-	1.1%	10.4%	27.4%	36.5%	20.4%	4.0%	0.2%
		12:00	16:00	4.20	-	-	2.5%	19.1%	42.1%	29.1%	6.7%	0.2%	0.2%
		16:00	20:00	3.74	-	-	17.4%	28.6%	25.1%	21.6%	6.5%	0.8%	-
		20:00	0:00	1.93	0.1%	27.9%	52.9%	17.7%	1.4%	-	-	-	-
	Week-end	0:00	4:00	0.20	82.7%	14.9%	2.2%	0.2%	-	-	-	-	-
		4:00	8:00	4.37	0.2%	6.3%	8.4%	16.1%	21.6%	17.8%	18.5%	9.1%	1.9%
		8:00	12:00	4.44	0.2%	1.7%	1.7%	12.3%	40.6%	25.5%	14.7%	2.9%	0.5%
		12:00	16:00	3.34	-	3.1%	9.4%	44.7%	36.5%	5.8%	0.5%	-	-
		16:00	20:00	3.24	-	4.3%	24.8%	34.9%	21.2%	8.2%	6.7%	-	-
		20:00	0:00	2.03	1.2%	20.7%	53.4%	23.6%	1.2%	-	-	-	-

(d) Fourth quarter

Table 2: Distributions of the number of active servers, per cluster, per quarter. The total number of hours in each clusters is shown in Table 1.

available data. Furthermore, the estimated cluster average waiting times may be biased due to the nature of the missing observations in the wait time report.

Should that bias become too large, it is possible that the estimated traffic intensity $\hat{\rho}_M$ takes on a value greater than 1 for some clusters, which makes it impossible to use (3) to estimate those **unstable** clusters' QoS levels under the $M/M/1$ assumption. This situation can be addressed so that a service rate and QoS level estimates can be produced nonetheless (see Section 4.2).

YEG (DI) − 2012 (continued)

The estimated service rates, traffic intensities and quantiles are shown in Table 4 (compare with Table 3). Note that estimates for those clusters in which no wait time data was collected cannot be provided.

3.6 Validating the $M/M/1$ Assumption

A number of hypotheses have been made concerning the nature of the clusters, the arrival rates, the average wait time and the service rate in order to progress towards an accurate model. These assumptions are not always easy (and in some instances, are actually impossible) to test. The easiest way to validate the (generalized) $M/M/1$ assumption remains to compare its wait time predictions with those of the actual wait time distributions.

In essence, for each cluster, the QoS levels $p = p_n(x)$ and the estimated QoS levels $p = \hat{p}_{M,n}(x)$ represent two families of indexed curves: performance p as a function of the waiting threshold x for the cluster $\mathscr{C}_n$. The $M/M/1$ assumption is validated if those two families are "close" to one another. For each cluster C_n (with non-zero waiting data), consider

1. the largest relative difference ratio

$$\tau_n^M = \max_x \left\{ \frac{|p_n(x) - \hat{p}_{M,n}(x)|}{|p_n(x)|} \right\}$$

   between the QoS level curve and the estimated QoS level curve, and

2. the relative area ratio

$$\alpha_n^M = \left| \frac{\int_0^\infty (p_n(x) - \hat{p}_{M,n}(x))\, dx}{\int_0^\infty p_n(x)\, dx} \right|$$

   of the (signed) area between the curves to the area under the QoS level curve.

If $\hat{p}_{M,n}(x) \approx p(x)$, then both τ_n^M and α_n^M should be small. By construction, τ_n^M is more likely to capture the short wait time bias discussed previously. In order to avoid difficulties linked to outlying clusters (sensitivity of the mean) and to clusters with a small number of arrivals (disproportional influence), it is preferable to not only consider the average and variance of the distributions for τ_n^M and α_n^M, but rather to examine the weighted quantiles of those distributions, with weights ω_n given by the number of arrivals in the cluster $\mathscr{C}_n$.

Cluster				Count	Avg Wait	Performance					
						5m	10m	15m	20m	25m	30m
Jan 01 to Mar 31 - YEG (DI) - 2012	Week day	0:00	4:00	-	-	-	-	-	-	-	-
		4:00	8:00	50,132	6.564	57.1%	79.0%	88.7%	93.5%	96.6%	98.8%
		8:00	12:00	43,033	4.466	68.9%	89.2%	96.5%	99.5%	99.8%	99.9%
		12:00	16:00	32,380	5.374	64.1%	81.8%	92.6%	97.6%	99.3%	99.9%
		16:00	20:00	29,279	5.373	68.0%	81.8%	90.9%	95.8%	97.8%	99.1%
		20:00	0:00	4,511	2.975	86.3%	96.7%	99.9%	100%	100%	100%
	Week-end	0:00	4:00	204	3.992	70.6%	99.5%	100%	100%	100%	100%
		4:00	8:00	14,450	4.520	68.4%	86.4%	96.8%	99.3%	100%	100%
		8:00	12:00	12,638	3.317	82.3%	95.0%	96.8%	98.2%	99.7%	100%
		12:00	16:00	11,938	3.043	83.0%	95.6%	98.5%	99.9%	100%	100%
		16:00	20:00	8,625	5.247	60.5%	80.6%	94.3%	98.9%	100%	100%
		20:00	0:00	1,529	2.382	88.9%	100%	100%	100%	100%	100%

(a) First quarter

Cluster				Count	Avg Wait	Performance					
						5m	10m	15m	20m	25m	30m
Apr 01 to Jun 30 - YEG (DI) - 2012	Week day	0:00	4:00	-	-	-	-	-	-	-	-
		4:00	8:00	57,120	6.718	45.1%	76.2%	91.8%	99.1%	100%	100%
		8:00	12:00	50,006	4.588	63.4%	89.7%	98.8%	99.9%	100%	100%
		12:00	16:00	38,563	4.123	71.5%	89.8%	97.6%	99.8%	100%	100%
		16:00	20:00	27,285	4.524	65.9%	86.9%	97.1%	99.5%	100%	100%
		20:00	0:00	2,370	1.862	98.6%	100%	100%	100%	100%	100%
	Week-end	0:00	4:00	93	2.471	97.8%	100%	100%	100%	100%	100%
		4:00	8:00	18,920	3.666	73.4%	93.8%	99.7%	99.9%	100%	100%
		8:00	12:00	17,151	3.855	73.5%	91.3%	98.7%	99.9%	100%	100%
		12:00	16:00	13,450	2.034	92.7%	98.9%	100%	100%	100%	100%
		16:00	20:00	9,487	1.843	93.6%	99.8%	100%	100%	100%	100%
		20:00	0:00	2,180	1.622	97.8%	100%	100%	100%	100%	100%

(b) Second quarter

Cluster				Count	Avg Wait	Performance					
						5m	10m	15m	20m	25m	30m
Jul 01 to Sep 30 - YEG (DI) - 2012	Week day	0:00	4:00	75	26.115	0.0%	0.0%	8.0%	33.3%	49.3%	58.7%
		4:00	8:00	59,787	7.426	36.8%	70.5%	93.4%	99.2%	99.8%	100%
		8:00	12:00	54,360	6.289	48.3%	78.2%	95.7%	99.5%	99.8%	100%
		12:00	16:00	41,429	4.371	71.2%	88.7%	96.1%	99.1%	99.9%	100%
		16:00	20:00	39,972	3.695	74.9%	91.9%	98.6%	99.8%	100%	100%
		20:00	0:00	11,949	2.410	94.3%	98.9%	99.6%	99.7%	99.8%	99.9%
	Week-end	0:00	4:00	-	-	-	-	-	-	-	-
		4:00	8:00	23,778	4.966	59.5%	87.4%	98.2%	100%	100%	100%
		8:00	12:00	23,377	5.177	56.9%	87.8%	98.4%	99.9%	100%	100%
		12:00	16:00	17,713	3.532	78.1%	93.2%	98.9%	100%	100%	100%
		16:00	20:00	14,773	3.062	80.3%	93.5%	99.4%	100%	100%	100%
		20:00	0:00	5,019	1.656	98.3%	100%	100%	100%	100%	100%

(c) Third quarter

Cluster				Count	Avg Wait	Performance					
						5m	10m	15m	20m	25m	30m
Oct 01 to Dec 31 - YEG (DI) - 2012	Week day	0:00	4:00	244	2.435	88.5%	100%	100%	100%	100%	100%
		4:00	8:00	67,961	6.332	45.8%	80.0%	95.4%	99.4%	100%	100%
		8:00	12:00	62,780	3.950	71.6%	94.8%	99.8%	100%	100%	100%
		12:00	16:00	54,348	5.048	61.7%	87.8%	96.6%	99.1%	99.7%	99.8%
		16:00	20:00	51,770	5.804	55.3%	81.6%	94.4%	98.2%	99.6%	100%
		20:00	0:00	25,190	3.751	78.1%	92.3%	96.7%	98.4%	99.8%	100%
	Week-end	0:00	4:00	147	4.956	59.2%	92.5%	100%	100%	100%	100%
		4:00	8:00	23,490	4.065	68.1%	95.5%	99.6%	100%	100%	100%
		8:00	12:00	23,882	4.482	65.5%	91.5%	98.8%	100%	100%	100%
		12:00	16:00	18,626	3.786	73.9%	94.1%	99.0%	99.8%	100%	100%
		16:00	20:00	17,494	4.907	64.3%	82.7%	96.7%	99.2%	100%	100%
		20:00	0:00	8,276	2.457	89.5%	98.6%	99.3%	99.6%	99.7%	100%

(d) Fourth quarter

Table 3: Average waiting time and service level performances, per cluster, per quarter. Notice that the counts are different than those shown in Table 1.

YEG (DI) − 2012 (continued)

The weighted quantiles are shown in Table 5. The table can be read as follows: for instance, at the checkpoint level,

$$P(\alpha^M \leq 0.0166) = 0.90 \quad \text{and} \quad P(\tau^M \leq 0.2132) = 0.95,$$

whereas $P(\alpha^M \leq 0.0157) = 0.75$ and $P(\tau^M \leq 0.0384) = 0.25$ during the third quarter. The extreme maximum values for both α^M and τ^M are easily identified as outliers. All in all, both measures seem to indicate that the $M/M/1$ assumption is reasonable at the checkpoint level, especially when taking into consideration that the large quantiles for τ^M are due to a poorer performance in the third quarter.

4 Regression Model

Given a Poisson arrival rate λ, an average waiting time $\overline{W}_q$ for an exponential distribution and stable clustering periods, the QoS level curves $\{\hat{p}_n(x)\}$ can be recovered from (1), (2) and (3), under the (generalized) $M/M/1$ assumption. But what about the number of servers c?

4.1 Linking the Service Rate, the Arrival Rate and the Number of Servers

In theory, the service rate is constant in a (generalized) $M/M/1$ queue: each server has a fixed capacity, and it operates, constantly, at that capacity. In practice, however, this is not the case: the (total) service rate for a given arrival rate is likely to increase if the number of open lines in the generalized server increases, and *vice-versa* (assuming of course that there is a non-zero average wait time upon entering the main queue).

It's also likely, given the non-mechanical nature of the servers' operators, that other factors (such as a sudden increase in the arrival rates leading to most or all lines of the generalized server becoming open) could affect the service rate.

The **Regression assumption** is that, on a quarterly level, the cluster service rate $\mu = \mu(c, \lambda)$ is a function of the number of active servers c (hidden behind the generalized server) and the arrival rate λ, and that this functional relationship is the same for all the regression clusters making up each of the quarters. It is economical to re-use the (generalized) $M/M/1$ clusters for the regression (although it is not necessary to do so).

YEG (DI) − 2012 (continued)

The arrival rates per line λ/c and estimated service rates per line μ_M/c are shown in Table 6. Only those clusters for which the average number of servers $c > 1$ are used in the regression (see Section 5.2 for details).

Cluster				Est Serv Rate	Est ρ	Estimated Performance (M/M/1)					
						5m	10m	15m	20m	25m	30m
Jan 01 to Mar 31 - YEG (DI) - 2012	Week day	0:00	4:00	-	-	-	-	-	-	-	-
		4:00	8:00	8.423	0.982	53.5%	78.0%	89.6%	95.1%	97.7%	98.9%
		8:00	12:00	6.495	0.967	67.2%	88.9%	96.2%	98.7%	99.6%	99.9%
		12:00	16:00	5.600	0.968	60.7%	84.0%	93.5%	97.4%	98.9%	99.6%
		16:00	20:00	5.242	0.966	60.7%	84.0%	93.5%	97.3%	98.9%	99.6%
		20:00	0:00	2.414	0.878	79.9%	95.4%	98.9%	99.8%	99.9%	100.0%
	Week-end	0:00	4:00	0.311	0.554	72.3%	86.2%	93.1%	96.5%	98.3%	99.1%
		4:00	8:00	6.572	0.967	66.8%	88.6%	96.1%	98.7%	99.5%	99.8%
		8:00	12:00	5.285	0.946	77.3%	94.5%	98.7%	99.7%	99.9%	100.0%
		12:00	16:00	4.495	0.932	79.8%	95.6%	99.1%	99.8%	100.0%	100.0%
		16:00	20:00	4.691	0.961	61.5%	84.6%	93.8%	97.5%	99.0%	99.6%
		20:00	0:00	1.950	0.823	85.4%	97.4%	99.5%	99.9%	100.0%	100.0%

(a) First quarter

Cluster				Est Serv Rate	Est ρ	Estimated Performance (M/M/1)					
						5m	10m	15m	20m	25m	30m
Apr 01 to Jun 30 - YEG (DI) - 2012	Week day	0:00	4:00	-	-	-	-	-	-	-	-
		4:00	8:00	8.393	0.983	52.7%	77.2%	89.0%	94.7%	97.5%	98.8%
		8:00	12:00	7.051	0.970	66.3%	88.3%	95.9%	98.6%	99.5%	99.8%
		12:00	16:00	5.823	0.960	70.0%	90.6%	97.1%	99.1%	99.7%	99.9%
		16:00	20:00	5.482	0.961	66.8%	88.5%	96.0%	98.6%	99.5%	99.8%
		20:00	0:00	2.647	0.831	91.1%	99.0%	99.9%	100.0%	100.0%	100.0%
	Week-end	0:00	4:00	0.258	0.389	82.3%	91.9%	96.3%	98.3%	99.2%	99.7%
		4:00	8:00	6.018	0.957	74.1%	93.0%	98.1%	99.5%	99.9%	100.0%
		8:00	12:00	5.967	0.958	72.4%	92.0%	97.7%	99.3%	99.8%	99.9%
		12:00	16:00	4.540	0.902	90.2%	98.9%	99.9%	100.0%	100.0%	100.0%
		16:00	20:00	4.408	0.890	92.1%	99.3%	99.9%	100.0%	100.0%	100.0%
		20:00	0:00	2.370	0.794	93.1%	99.4%	99.9%	100.0%	100.0%	100.0%

(b) Second quarter

Cluster				Est Serv Rate	Est ρ	Estimated Performance (M/M/1)					
						5m	10m	15m	20m	25m	30m
Jul 01 to Sep 30 - YEG (DI) - 2012	Week day	0:00	4:00	0.316	0.892	24.8%	36.6%	46.6%	55.0%	62.0%	68.0%
		4:00	8:00	8.478	0.984	49.3%	73.8%	86.5%	93.1%	96.4%	98.2%
		8:00	12:00	7.550	0.979	55.0%	79.4%	90.5%	95.7%	98.0%	99.1%
		12:00	16:00	5.825	0.962	68.0%	89.4%	96.5%	98.8%	99.6%	99.9%
		16:00	20:00	5.518	0.953	73.8%	92.8%	98.0%	99.5%	99.8%	100.0%
		20:00	0:00	3.202	0.885	85.9%	97.8%	99.6%	99.9%	100.0%	100.0%
	Week-end	0:00	4:00	-	-	-	-	-	-	-	-
		4:00	8:00	6.402	0.970	63.5%	86.2%	94.8%	98.0%	99.3%	99.7%
		8:00	12:00	6.653	0.972	62.0%	85.1%	94.2%	97.7%	99.1%	99.7%
		12:00	16:00	4.934	0.946	75.2%	93.5%	98.3%	99.6%	99.9%	100.0%
		16:00	20:00	4.421	0.931	79.6%	95.5%	99.0%	99.8%	100.0%	100.0%
		20:00	0:00	2.918	0.829	93.2%	99.4%	100.0%	100.0%	100.0%	100.0%

(c) Third quarter

Cluster				Est Serv Rate	Est ρ	Estimated Performance (M/M/1)					
						5m	10m	15m	20m	25m	30m
Oct 01 to Dec 31 - YEG (DI) - 2012	Week day	0:00	4:00	0.215	0.343	83.0%	91.6%	95.9%	98.0%	99.0%	99.5%
		4:00	8:00	8.623	0.982	54.8%	79.2%	90.4%	95.6%	98.0%	99.1%
		8:00	12:00	6.784	0.964	71.5%	91.6%	97.5%	99.3%	99.8%	99.9%
		12:00	16:00	5.820	0.967	62.9%	85.8%	94.5%	97.9%	99.2%	99.7%
		16:00	20:00	5.544	0.970	57.9%	81.8%	92.1%	96.6%	98.5%	99.4%
		20:00	0:00	2.534	0.905	72.9%	91.9%	97.6%	99.3%	99.8%	99.9%
	Week-end	0:00	4:00	0.244	0.548	68.5%	81.9%	89.6%	94.0%	96.5%	98.0%
		4:00	8:00	6.358	0.963	70.5%	91.0%	97.2%	99.2%	99.7%	99.9%
		8:00	12:00	6.392	0.966	67.1%	88.8%	96.2%	98.7%	99.6%	99.8%
		12:00	16:00	4.526	0.945	72.9%	92.2%	97.8%	99.4%	99.8%	99.9%
		16:00	20:00	4.265	0.954	63.9%	86.4%	94.8%	98.0%	99.3%	99.7%
		20:00	0:00	2.248	0.847	84.9%	97.3%	99.5%	99.9%	100.0%	100.0%

(d) Fourth quarter

Table 4: Estimated service rates and service level performances under the $M/M/1$ assumption, per cluster, per quarter. Compare with Table 3.

	Quarter	Metric	min	1st	5th	10th	25th	med	75th	90th	95th	99th	max
YEG (DI) - 2012	Jan Mar	Area Ratio	0.08%	0.08%	0.08%	0.10%	0.13%	0.18%	0.47%	0.84%	1.03%	1.15%	5.15%
		Max Abs Diff Ratio	2.42%	2.42%	2.42%	2.42%	2.68%	5.26%	6.29%	8.40%	9.64%	10.62%	13.40%
	Apr Jun	Area Ratio	0.09%	0.09%	0.09%	0.09%	0.27%	0.54%	1.00%	1.14%	1.19%	1.23%	4.43%
		Max Abs Diff Ratio	1.50%	1.50%	1.50%	1.63%	1.78%	2.57%	6.51%	16.39%	16.58%	16.74%	16.78%
	Jul Sep	Area Ratio	0.16%	0.16%	0.16%	0.16%	0.24%	1.06%	1.57%	1.73%	1.76%	1.79%	105.47%
		Max Abs Diff Ratio	1.54%	1.54%	1.54%	1.54%	3.84%	7.95%	13.04%	24.94%	29.68%	33.46%	482.09%
	Oct Dec	Area Ratio	0.09%	0.09%	0.11%	0.27%	0.84%	1.22%	1.38%	1.57%	1.63%	1.68%	5.70%
		Max Abs Diff Ratio	2.04%	2.04%	2.05%	2.13%	2.83%	4.48%	6.17%	17.86%	18.71%	19.39%	19.56%
	All	Area Ratio	0.08%	0.08%	0.09%	0.13%	0.24%	0.88%	1.28%	1.66%	1.71%	1.78%	105.47%
		Max Abs Diff Ratio	1.50%	1.50%	1.54%	1.80%	2.42%	4.69%	9.00%	17.44%	21.32%	31.64%	482.09%

Table 5: Quantiles of the Area Ratio (α^M) and Maximal Difference Ratio (τ^M), per cluster, and for the entire checkpoint.

4.2 Estimating the Service Rates and the Performance Levels

The form taken by the functional relationship determines the estimated service rates $\hat{\mu}_R$ and the QoS level curve $\hat{p}_{R,x}$ for each cluster. Fortunately, the simplest case yields fairly accurate results: set

$$\mu = \mu(c,\lambda) = ac + b\lambda, \quad \text{for some } a, b,$$

where c is as computed in Section 3.3, which can be re-written as

$$\frac{\mu}{c} = a + b\frac{\lambda}{c}, \quad \text{for some } a, b. \tag{4}$$

Evidently, then, in order to determine the optimal constants $\hat{a}$ and $\hat{b}$, one needs to regress the service rate per line against the arrival rate per line. The best available estimate for $\frac{\mu}{c}$ remains $\frac{\mu_M}{c}$, and since the regression clusters are identical to the $M/M/1$ clusters, the estimated service rates μ_M do not need to be re-calculated.

Once $\hat{a}$ and $\hat{b}$ are known, the estimated service rates $\hat{\mu}_R$ are easily computed as

$$\hat{\mu}_R = \hat{a}c + \hat{b}\lambda \tag{5}$$

for each quarterly cluster.

The estimated QoS level curves $\hat{p}_x$ sit is sufficient to substitute (4) into (3) to obtain the QoS level approximations

$$\hat{p}_{R,x} = 1 - \frac{\lambda}{\hat{a}c + \hat{b}\lambda} e^{-(\hat{a}c+\hat{b}\lambda-\lambda)x}. \tag{6}$$

If $\hat{\rho}_R = \lambda/(\hat{a}c + \hat{b}\lambda) > 1$, the cluster is unstable (see comments at the end of Section 3.5) and $\hat{p}_R(x)$ cannot be produced for that cluster.

The unspoken assumptions are that the quarterly regressions produce good fits, and that there is a quarterly characteristic to the service rate.

Cluster				Avg # of Servers	Arrival Rate	Arr Rate / Server	Est Serv Rate	Serv Rate / Server
Jan 01 to Mar 31 - YEG (DI) - 2012	Week day	0:00	4:00	0.14	0.055	0.000	0.405	0.000
		4:00	8:00	5.38	8.274	8.423	1.539	1.567
		8:00	12:00	4.63	6.279	6.495	1.356	1.403
		12:00	16:00	4.19	5.420	5.600	1.292	1.335
		16:00	20:00	3.78	5.062	5.242	1.341	1.388
		20:00	0:00	1.58	2.119	2.414	1.337	1.524
	Week-end	0:00	4:00	0.21	0.172	0.311	0.815	1.471
		4:00	8:00	4.56	6.358	6.572	1.394	1.441
		8:00	12:00	3.92	5.000	5.285	1.276	1.349
		12:00	16:00	3.41	4.188	4.495	1.228	1.318
		16:00	20:00	3.60	4.508	4.691	1.253	1.304
		20:00	0:00	1.47	1.605	1.950	1.091	1.326

(a) First quarter

Cluster				Avg # of Servers	Arrival Rate	Arr Rate / Server	Est Serv Rate	Serv Rate / Server
Apr 01 to Jun 30 - YEG (DI) - 2012	Week day	0:00	4:00	0.15	0.070	0.000	0.452	0.000
		4:00	8:00	4.85	8.247	8.393	1.702	1.732
		8:00	12:00	4.69	6.840	7.051	1.459	1.505
		12:00	16:00	4.12	5.590	5.823	1.357	1.414
		16:00	20:00	3.82	5.269	5.482	1.379	1.434
		20:00	0:00	1.69	2.201	2.647	1.306	1.570
	Week-end	0:00	4:00	0.17	0.100	0.258	0.588	1.510
		4:00	8:00	4.03	5.757	6.018	1.427	1.492
		8:00	12:00	4.08	5.718	5.967	1.400	1.461
		12:00	16:00	3.58	4.097	4.540	1.145	1.269
		16:00	20:00	3.45	3.925	4.408	1.138	1.279
		20:00	0:00	1.84	1.881	2.370	1.021	1.287

(b) Second quarter

Cluster				Avg # of Servers	Arrival Rate	Arr Rate / Server	Est Serv Rate	Serv Rate / Server
Jul 01 to Sep 30 - YEG (DI) - 2012	Week day	0:00	4:00	0.26	0.281	0.316	1.084	1.216
		4:00	8:00	4.34	8.345	8.478	1.924	1.955
		8:00	12:00	4.29	7.394	7.550	1.724	1.760
		12:00	16:00	3.53	5.605	5.825	1.587	1.649
		16:00	20:00	3.32	5.260	5.518	1.584	1.661
		20:00	0:00	1.84	2.834	3.202	1.542	1.742
	Week-end	0:00	4:00	0.27	0.285	0.000	1.045	0.000
		4:00	8:00	3.59	6.206	6.402	1.729	1.783
		8:00	12:00	3.75	6.466	6.653	1.723	1.773
		12:00	16:00	3.11	4.666	4.934	1.499	1.585
		16:00	20:00	3.08	4.117	4.421	1.338	1.437
		20:00	0:00	2.13	2.417	2.918	1.136	1.372

(c) Third quarter

Cluster				Avg # of Servers	Arrival Rate	Arr Rate / Server	Est Serv Rate	Serv Rate / Server
Oct 01 to Dec 31 - YEG (DI) - 2012	Week day	0:00	4:00	0.18	0.074	0.215	0.421	1.226
		4:00	8:00	5.17	8.468	8.623	1.639	1.669
		8:00	12:00	4.78	6.540	6.784	1.369	1.420
		12:00	16:00	4.20	5.629	5.820	1.341	1.386
		16:00	20:00	3.74	5.377	5.544	1.439	1.484
		20:00	0:00	1.93	2.293	2.534	1.191	1.316
	Week-end	0:00	4:00	0.20	0.134	0.244	0.671	1.225
		4:00	8:00	4.37	6.121	6.358	1.400	1.454
		8:00	12:00	4.44	6.176	6.392	1.392	1.440
		12:00	16:00	3.34	4.276	4.526	1.281	1.355
		16:00	20:00	3.24	4.070	4.265	1.255	1.315
		20:00	0:00	2.03	1.904	2.248	0.938	1.108

(d) Fourth quarter

Table 6: Arrival rates (per server) and estimated service rates (per server), per cluster, per quarter. Then entries in red are not used in the regression.

YEG (DI) — 2012 (continued)

The regression graphs and parameters are shown in Figure 4. The assumption that the service rate is affected by both the number of open lines and the arrival rate is clearly met in practice. Note the goodness-of-fit improvement over the year.

The estimated service rates, traffic intensities and quantiles are shown in Table 6 (compare with Tables 3 and 4). Note that estimates for those clusters in which no wait time data was collected cannot be provided.

4.3 Validating the Combined Assumptions

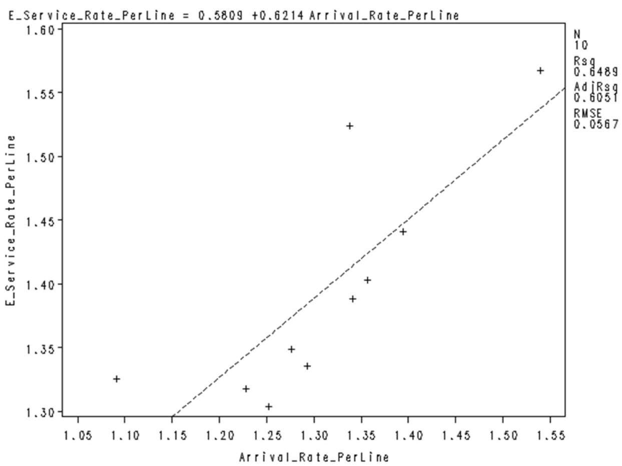
As before a number of hypotheses have been made about the appropriateness of the Regression assumption. The easiest way to validate the combined (generalized) $M/M/1$ and Regression assumptions is still to compare the wait time predictions with those of the actual wait time distributions.

The Regression Area Ratios α^R and Regression Maximal Difference Ratios τ^R are defined in a similar manner as α^M and τ^M.

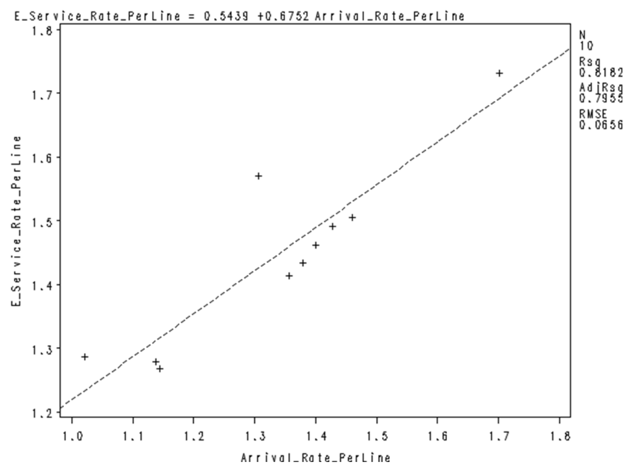
YEG (DI) — 2012 (continued)

The weighted quantiles are shown in Table 8 (compare with Table 5). The extreme maximum values for both α^M and τ^M tend to be lower, but there is a bit more spread among the cluster quantiles, which is not surprising as the regression assumption has introduced some uncertainty.

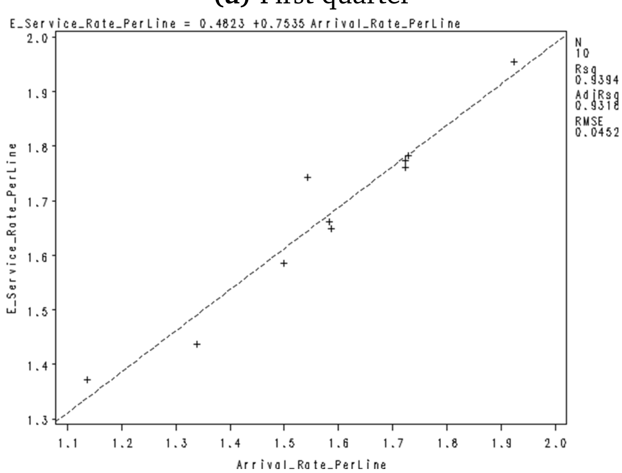
The combined $M/M/1$ and Regression assumptions are not as accurate as the $M/M/1$ queueing system on its own (some of the quantiles seem a bit high), but since there is no way to extract the number of servers c without introducing an external relationship $\mu = \mu(c, \lambda)$, the numbers are still satisfactory. Further regression possibilities are explored in Section 8.



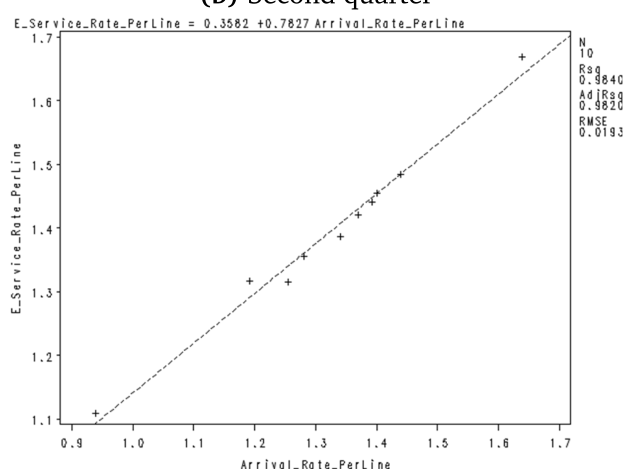
(a) First quarter



(b) Second quarter



(c) Third quarter



(d) Fourth quarter

	Quarter	a	b
YEG (DI)	Jan 01 - Mar 31	0.581	0.621
	Apr 01 - Jun 30	0.544	0.675
	Jul 01 - Sep 30	0.482	0.753
	Oct 01 - Dec 31	0.358	0.783

(e) Regression parameters

Figure 4: Regression of the cluster estimated service rates (per server) against the cluster arrival rates (per server), per quarter. The regression parameters are also gathered.

Cluster				Class	Reg Serv Rate	Reg ρ	Estimated Performance (M/M/1 + Reg)					
							5m	10m	15m	20m	25m	30m
Jan 01 to Mar 31 - YEG (DI) - 2012	Week day	0:00	4:00	-	-	-	-	-	-	-	-	-
		4:00	8:00	1.5	8.264	1.001	-	-	-	-	-	-
		8:00	12:00	1.4	6.591	0.953	80.0%	95.8%	99.1%	99.8%	100%	100%
		12:00	16:00	1.3	5.804	0.934	86.3%	98.0%	99.7%	100%	100%	100%
		16:00	20:00	1.3	5.338	0.948	76.2%	94.0%	98.5%	99.6%	99.9%	100%
		20:00	0:00	1.3	2.237	0.947	47.5%	70.9%	83.9%	91.1%	95.1%	97.3%
	Week-end	0:00	4:00	-	-	-	-	-	-	-	-	-
		4:00	8:00	1.4	6.600	0.963	71.3%	91.4%	97.4%	99.2%	99.8%	99.9%
		8:00	12:00	1.3	5.383	0.929	86.3%	98.0%	99.7%	100%	100%	100%
		12:00	16:00	1.2	4.584	0.914	87.4%	98.3%	99.8%	100%	100%	100%
		16:00	20:00	1.3	4.892	0.922	86.5%	98.0%	99.7%	100%	100%	100%
		20:00	0:00	1.1	1.852	0.867	74.8%	92.7%	97.9%	99.4%	99.8%	99.9%

(a) First quarter

Cluster				Class	Reg Serv Rate	Reg ρ	Estimated Performance (M/M/1 + Reg)					
							5m	10m	15m	20m	25m	30m
Apr 01 to Jun 30 - YEG (DI) - 2012	Week day	0:00	4:00	-	-	-	-	-	-	-	-	-
		4:00	8:00	1.7	8.204	1.005	-	-	-	-	-	-
		8:00	12:00	1.5	7.167	0.954	81.4%	96.4%	99.3%	99.9%	100%	100%
		12:00	16:00	1.4	6.015	0.929	88.9%	98.7%	99.8%	100%	100%	100%
		16:00	20:00	1.4	5.636	0.935	85.1%	97.6%	99.6%	99.9%	100%	100%
		20:00	0:00	1.3	2.403	0.916	66.6%	87.8%	95.6%	98.4%	99.4%	99.8%
	Week-end	0:00	4:00	-	-	-	-	-	-	-	-	-
		4:00	8:00	1.4	6.081	0.947	81.3%	96.3%	99.3%	99.9%	100%	100%
		8:00	12:00	1.4	6.082	0.940	84.8%	97.5%	99.6%	99.9%	100%	100%
		12:00	16:00	1.1	4.713	0.869	96.0%	99.8%	100%	100%	100%	100%
		16:00	20:00	1.1	4.525	0.867	95.7%	99.8%	100%	100%	100%	100%
		20:00	0:00	1.0	2.271	0.828	88.3%	98.3%	99.8%	100%	100%	100%

(b) Second quarter

Cluster				Class	Reg Serv Rate	Reg ρ	Estimated Performance (M/M/1 + Reg)					
							5m	10m	15m	20m	25m	30m
Jul 01 to Sep 30 - YEG (DI) - 2012	Week day	0:00	4:00	-	-	-	-	-	-	-	-	-
		4:00	8:00	1.9	8.380	0.996	16.0%	29.2%	40.3%	49.7%	57.6%	64.3%
		8:00	12:00	1.7	7.640	0.968	71.7%	91.7%	97.6%	99.3%	99.8%	99.9%
		12:00	16:00	1.6	5.927	0.946	81.1%	96.2%	99.2%	99.8%	100%	100%
		16:00	20:00	1.6	5.565	0.945	79.4%	95.5%	99.0%	99.8%	100%	100%
		20:00	0:00	1.5	3.022	0.938	63.3%	85.6%	94.4%	97.8%	99.1%	99.7%
	Week-end	0:00	4:00	-	-	-	-	-	-	-	-	-
		4:00	8:00	1.7	6.408	0.969	64.6%	87.1%	95.3%	98.3%	99.4%	99.8%
		8:00	12:00	1.7	6.682	0.968	67.1%	88.8%	96.2%	98.7%	99.6%	99.9%
		12:00	16:00	1.5	5.017	0.930	83.9%	97.2%	99.5%	99.9%	100%	100%
		16:00	20:00	1.3	4.585	0.898	91.4%	99.2%	99.9%	100%	100%	100%
		20:00	0:00	1.1	2.847	0.849	90.1%	98.8%	99.9%	100%	100%	100%

(c) Third quarter

Cluster				Class	Reg Serv Rate	Reg ρ	Estimated Performance (M/M/1 + Reg)					
							5m	10m	15m	20m	25m	30m
Oct 01 to Dec 31 - YEG (DI) - 2012	Week day	0:00	4:00	-	-	-	-	-	-	-	-	-
		4:00	8:00	1.6	8.478	0.999	5.2%	10.0%	14.6%	18.9%	23.1%	27.0%
		8:00	12:00	1.4	6.830	0.958	77.6%	94.7%	98.8%	99.7%	99.9%	100%
		12:00	16:00	1.3	5.909	0.952	76.6%	94.3%	98.6%	99.7%	99.9%	100%
		16:00	20:00	1.4	5.547	0.969	58.6%	82.3%	92.4%	96.8%	98.6%	99.4%
		20:00	0:00	1.2	2.484	0.923	64.5%	86.4%	94.8%	98.0%	99.2%	99.7%
	Week-end	0:00	4:00	-	-	-	-	-	-	-	-	-
		4:00	8:00	1.4	6.357	0.963	70.5%	90.9%	97.2%	99.1%	99.7%	99.9%
		8:00	12:00	1.4	6.424	0.961	72.1%	91.9%	97.7%	99.3%	99.8%	99.9%
		12:00	16:00	1.3	4.543	0.941	75.2%	93.5%	98.3%	99.5%	99.9%	100%
		16:00	20:00	1.3	4.348	0.936	76.6%	94.1%	98.5%	99.6%	99.9%	100%
		20:00	0:00	0.9	2.217	0.859	82.1%	96.2%	99.2%	99.8%	100%	100%

(d) Fourth quarter

Table 7: Estimated service rates and service level performances under the regression assumption, per cluster, per quarter. Compare with Tables 3 and 4.

	Quarter	Metric	min	1st	5th	10th	25th	med	75th	90th	95th	99th	max
YEG (DI) - 2012	Jan Mar	Area Ratio	1.27%	1.27%	1.27%	1.27%	1.27%	2.70%	7.02%	8.22%	9.39%	14.35%	15.59%
		Max Abs Diff Ratio	4.89%	4.89%	4.89%	4.89%	4.89%	11.92%	17.32%	36.25%	43.19%	44.54%	44.88%
	Apr Jun	Area Ratio	0.21%	0.21%	0.21%	0.21%	0.21%	2.85%	4.00%	4.79%	5.33%	6.66%	6.99%
		Max Abs Diff Ratio	2.24%	2.24%	2.24%	2.24%	2.24%	17.85%	27.33%	28.94%	29.72%	31.90%	32.45%
	Jul Sep	Area Ratio	0.46%	0.46%	0.46%	0.46%	1.23%	3.26%	6.60%	29.42%	39.55%	47.66%	49.68%
		Max Abs Diff Ratio	6.04%	6.04%	6.04%	6.04%	8.49%	15.88%	44.99%	53.72%	56.13%	58.06%	58.55%
	Oct Dec	Area Ratio	0.23%	0.23%	0.23%	0.23%	0.34%	1.35%	3.93%	47.36%	64.44%	78.10%	81.52%
		Max Abs Diff Ratio	1.82%	1.82%	1.82%	1.82%	5.82%	9.60%	23.42%	60.17%	74.40%	85.80%	88.64%
	All	Area Ratio	0.21%	0.21%	0.21%	0.21%	0.53%	3.31%	5.83%	22.80%	53.42%	75.90%	81.52%
		Max Abs Diff Ratio	1.82%	1.82%	1.82%	1.82%	5.92%	15.76%	30.63%	50.61%	62.08%	83.33%	88.64%

Table 8: Quantiles of the Area Ratio (α^R) and Maximal Difference Ratio (τ^R), per cluster, and for the entire checkpoint.

5 Predicting the Number of Servers Under the Combined Assumptions

Given regression parameters a and b, an average arrival rates $\lambda > 0$ and QoS levels $p_x \in (0,1)$ for and $x > 0$, the average number of active servers $c > 0$ can be obtained by solving for c in

$$1-p=\frac{\lambda}{ac+b\lambda}e^{(\lambda-ac-b\lambda)x}. \tag{7}$$

While (7) cannot be solved for positive c using elementary functions, numerical solvers (such as MATLAB's non-linear solver `fsolve`) can be used to find an approximate solution.

However, in instances where such a solver is either unavailable or inconvenient to use (as is the case with SAS), an approximate solution can still be calculated using the Lambert W function [2]. Equation (7) can be re-written as

$$(ac+b\lambda)e^{(ac+b\lambda)x}=\frac{\lambda}{1-p}e^{\lambda x},$$

which, upon multiplication by x, becomes

$$(ac+b\lambda)xe^{(ac+b\lambda)x}=\frac{x\lambda}{1-p}e^{\lambda x}.$$

Setting $y=(ac+b\lambda)x$ and $z=\frac{x\lambda}{1-p}e^{\lambda x}>0$ in this last equation yields $ye^y=z$. The solution for positive z is $y=W_0(z)$, where W_0 represents the main branch of the Lambert W function. Then

$$y=(ac+b\lambda)x=W_0(z)=W_0\left(\frac{x\lambda}{1-p}e^{\lambda x}\right),$$

which can be re-arranged to yield

$$c=-\frac{b\lambda}{a}+\frac{1}{ax}W_0\left(\frac{x\lambda}{1-p}e^{\lambda x}\right)=\frac{1}{ax}\left[W_0\left(\frac{x\lambda}{1-p}e^{\lambda x}\right)-b\lambda x\right]. \tag{8}$$

But $W_0(z)$ cannot be evaluated by elementary means except at special z—values, and so one has to rely on efficient numerical algorithms to recover c [2].

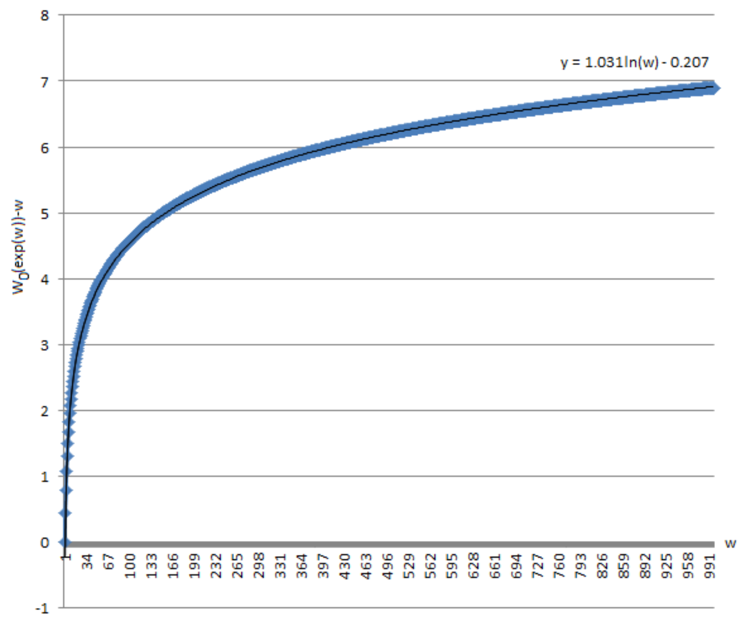


Figure 5: Graph of $W_0(e^w) - w$ (in blue), together with logarithmic trend line (in black).

5.1 Estimating Lambert's W Function

However, in order to predict the number of servers for various service level performances, the formula (8) needs to be implemented $\approx 100,000$ times in SAS for each checkpoint. As SAS doesn't lend itself particularly well to repeated algorithmic computations, it becomes imperative to find a quick and relatively accurate alternative approach. Re-write $z = e^w$. A graph of $W_0(z) - \ln z = W_0(e^w) - w$ for integer values $w = 1, \ldots, 1000$ (computed separately with MATLAB's built-in W function) appears to show logarithmic growth, as can be seen in Figure 5. This suggests that $W_0(e^w) - w$ could be approached by

$$W_0(e^w) - w \approx q_1 \ln w + q_2. \tag{9}$$

Indeed, $q_1 = 1.031$ and $q_2 = 0.207$ provide an excellent fit for $w \in (1, 1000)$. Consequently, the Lambert W function is approximated by

$$W_0(z) \approx \ln z + 1.031 \ln(\ln z) + 0.207, \quad e \leq z \leq e^{1000}, \tag{10}$$

and (8) becomes

$$c_R \approx \frac{1}{ax}\left[\ln\left(\frac{x\lambda}{1-p}e^{\lambda x}\right) + 1.031 \ln\left(\ln\left(\frac{x\lambda}{1-p}e^{\lambda x}\right)\right) + 0.207 - b\lambda x\right], \tag{11}$$

as long as $e \leq \frac{x\lambda}{1-p}e^{\lambda x} \leq e^{1000}$.

YEG (DI) — 2012 (continued)

The first order of business is to determine the parameters to use in (11). The arrival rates λ are found in Table 1, while the regression parameters are those of Figure 4e. The cluster QoS levels are a little bit less obvious to select. Indeed, a given cluster's worth of observations is characterized by an entire QoS level curve $(p(x), x)$, whereas (11) only calls for one pair (p, x).

Cluster				Arr Rate	Service Level		Avg # Servers
					perf	min	
Jan 01 to Mar 31 - YEG (DI) - 2012	Week day	0:00	4:00	0.055	-	-	0.14
		4:00	8:00	8.274	89%	15	5.38
		8:00	12:00	6.279	96%	15	4.63
		12:00	16:00	5.420	93%	15	4.19
		16:00	20:00	5.062	91%	15	3.78
		20:00	0:00	2.119	100%	15	1.58
	Week-end	0:00	4:00	0.172	100%	10	0.21
		4:00	8:00	6.358	97%	15	4.56
		8:00	12:00	5.000	97%	15	3.92
		12:00	16:00	4.188	98%	15	3.41
		16:00	20:00	4.508	94%	15	3.60
		20:00	0:00	1.605	89%	5	1.47

(a) First quarter

Cluster				Arr Rate	Service Level		Avg # Servers
					perf	min	
Apr 01 to Jun 30 - YEG (DI) - 2012	Week day	0:00	4:00	0.070	-	-	0.15
		4:00	8:00	8.247	92%	15	4.85
		8:00	12:00	6.840	99%	15	4.69
		12:00	16:00	5.590	98%	15	4.12
		16:00	20:00	5.269	97%	15	3.82
		20:00	0:00	2.201	99%	5	1.69
	Week-end	0:00	4:00	0.100	98%	5	0.17
		4:00	8:00	5.757	100%	15	4.03
		8:00	12:00	5.718	99%	15	4.08
		12:00	16:00	4.097	100%	15	3.58
		16:00	20:00	3.925	100%	10	3.45
		20:00	0:00	1.881	98%	5	1.84

(b) Second quarter

Cluster				Arr Rate	Service Level		Avg # Servers
					perf	min	
Jul 01 to Sep 30 - YEG (DI) - 2012	Week day	0:00	4:00	0.281	8%	15	0.26
		4:00	8:00	8.345	93%	15	4.34
		8:00	12:00	7.394	96%	15	4.29
		12:00	16:00	5.605	96%	15	3.53
		16:00	20:00	5.260	99%	15	3.32
		20:00	0:00	2.834	100%	15	1.84
	Week-end	0:00	4:00	0.285	-	-	0.27
		4:00	8:00	6.206	98%	15	3.59
		8:00	12:00	6.466	98%	15	3.75
		12:00	16:00	4.666	99%	15	3.11
		16:00	20:00	4.117	99%	15	3.08
		20:00	0:00	2.417	98%	5	2.13

(c) Third quarter

Cluster				Arr Rate	Service Level		Avg # Servers
					perf	min	
Oct 01 to Dec 31 - YEG (DI) - 2012	Week day	0:00	4:00	0.074	89%	5	0.18
		4:00	8:00	8.468	95%	15	5.17
		8:00	12:00	6.540	100%	15	4.78
		12:00	16:00	5.629	97%	15	4.20
		16:00	20:00	5.377	94%	15	3.74
		20:00	0:00	2.293	97%	15	1.93
	Week-end	0:00	4:00	0.134	93%	10	0.20
		4:00	8:00	6.121	100%	15	4.37
		8:00	12:00	6.176	99%	15	4.44
		12:00	16:00	4.276	99%	15	3.34
		16:00	20:00	4.070	97%	15	3.24
		20:00	0:00	1.904	99%	15	2.03

(d) Fourth quarter

Table 9: Actual average arrival rates, service level performances and average number of servers, per cluster, per quarter.

- When the p_{15} service level is available (i.e. when $p_{15} \neq 1$), set $(p,x) = (p_{15}, 15)$.
- If $p_{15} = 1$ and $p_{10} \neq 1$, set $(p,x) = (p_{10}, 10)$.
- If $p_{15} = p_{10} = 1$ and $p_5 \neq 1$, set $(p,x) = (p_5, 5)$.
- Otherwise, discard the cluster.

The parameters are shown in Table 9, and the prediction results c_R for the average number of active servers per cluster are shown in Table 10. The accuracy of the predictions offsets some of the (slight) uncertainty appearing in Table 8. Evidently, then, the (generalized) $M/M/1$ model is better-suited than the combined model to determine the QoS level curves (p_x, x), but the combined model nevertheless provides good estimates for c (at least, for this checkpoint).

5.2 Classifying the Clusters

As can be seen in Table 10, the predicted average number of active servers is not available for all clusters, due to some technical characteristics of the cluster. In Tables 7a and 7b, some clusters do not have an associated estimated QoS level curve, again due to some technical characteristics of the cluster. To simplify the interpretation of the results, clusters are classified according to one of two schemes.

Cluster				Actual # Servers	Pred # Servers
Jan 01 to Mar 31 - YEG (DI) - 2012	Week day	0:00	4:00	0.136	-
		4:00	8:00	5.375	5.643
		8:00	12:00	4.629	4.474
		12:00	16:00	4.193	3.829
		16:00	20:00	3.775	3.572
		20:00	0:00	1.585	2.198
	Week-end	0:00	4:00	0.212	-
		4:00	8:00	4.560	4.535
		8:00	12:00	3.918	3.650
		12:00	16:00	3.411	3.205
		16:00	20:00	3.599	3.264
		20:00	0:00	1.471	1.701

(a) First quarter

Cluster				Actual # Servers	Pred # Servers
Apr 01 to Jun 30 - YEG (DI) - 2012	Week day	0:00	4:00	0.154	-
		4:00	8:00	4.845	5.232
		8:00	12:00	4.687	4.623
		12:00	16:00	4.119	3.792
		16:00	20:00	3.822	3.577
		20:00	0:00	1.686	2.766
	Week-end	0:00	4:00	0.171	-
		4:00	8:00	4.034	4.160
		8:00	12:00	4.084	3.943
		12:00	16:00	3.579	3.402
		16:00	20:00	3.447	3.426
		20:00	0:00	1.841	2.395

(b) Second quarter

Cluster				Actual # Servers	Pred # Servers
Jul 01 to Sep 30 - YEG (DI) - 2012	Week day	0:00	4:00	0.260	-
		4:00	8:00	4.337	4.641
		8:00	12:00	4.289	4.214
		12:00	16:00	3.533	3.309
		16:00	20:00	3.321	3.273
		20:00	0:00	1.838	2.182
	Week-end	0:00	4:00	0.273	-
		4:00	8:00	3.590	3.722
		8:00	12:00	3.752	3.876
		12:00	16:00	3.113	3.008
		16:00	20:00	3.076	2.794
		20:00	0:00	2.127	2.804

(c) Third quarter

Cluster				Actual # Servers	Pred # Servers
Oct 01 to Dec 31 - YEG (DI) - 2012	Week day	0:00	4:00	0.175	-
		4:00	8:00	5.165	5.708
		8:00	12:00	4.777	5.082
		12:00	16:00	4.198	4.038
		16:00	20:00	3.737	3.796
		20:00	0:00	1.925	2.009
	Week-end	0:00	4:00	0.200	-
		4:00	8:00	4.373	4.720
		8:00	12:00	4.438	4.558
		12:00	16:00	3.339	3.441
		16:00	20:00	3.243	3.097
		20:00	0:00	2.029	2.058

(d) Fourth quarter

Table 10: Predicted and actual number of servers under the $M/M/1$ and regression assumptions, per cluster, per quarter.

In the **original** classification, clusters are flagged with

1. if $\lambda = 0$, the flag is 0;
2. if $\lambda > 0$, then
   (a) if $\overline{W}_q = 0$ or $c < 1$, the flag is 0.5;
   (b) otherwise, the flag is 1.

In the **modified** classification,

1. if $\lambda = 0$, the flag is 0;
2. if $\lambda > 0$, then
   (a) if $\overline{W}_q = 0$, the flag is 0.5;
   (b) if $\overline{W}_q > 0$, then
      i. if $\rho_R \geq 1$, the flag is 1.5;
      ii. if $\rho_R < 1$, then
         A. if $c < 1$, the flag is 2;
         B. else, the flag is 1.

Cluster				Original Flag	Modified Flag
Jan 01 to Mar 31 - YEG (DI) - 2012	Week day	0:00	4:00	0.5	0.5
		4:00	8:00	1	1.5
		8:00	12:00	1	1
		12:00	16:00	1	1
		16:00	20:00	1	1
		20:00	0:00	1	1
	Week-end	0:00	4:00	0.5	2
		4:00	8:00	1	1
		8:00	12:00	1	1
		12:00	16:00	1	1
		16:00	20:00	1	1
		20:00	0:00	1	1

(a) First quarter

Cluster				Original Flag	Modified Flag
Apr 01 to Jun 30 - YEG (DI) - 2012	Week day	0:00	4:00	0.5	0.5
		4:00	8:00	1	1.5
		8:00	12:00	1	1
		12:00	16:00	1	1
		16:00	20:00	1	1
		20:00	0:00	1	1
	Week-end	0:00	4:00	0.5	2
		4:00	8:00	1	1
		8:00	12:00	1	1
		12:00	16:00	1	1
		16:00	20:00	1	1
		20:00	0:00	1	1

(b) Second quarter

Cluster				Original Flag	Modified Flag
Jul 01 to Sep 30 - YEG (DI) - 2012	Week day	0:00	4:00	0.5	2
		4:00	8:00	1	1
		8:00	12:00	1	1
		12:00	16:00	1	1
		16:00	20:00	1	1
		20:00	0:00	1	1
	Week-end	0:00	4:00	0.5	0.5
		4:00	8:00	1	1
		8:00	12:00	1	1
		12:00	16:00	1	1
		16:00	20:00	1	1
		20:00	0:00	1	1

(c) Third quarter

Cluster				Original Flag	Modified Flag
Oct 01 to Dec 31 - YEG (DI) - 2012	Week day	0:00	4:00	0.5	2
		4:00	8:00	1	1.5
		8:00	12:00	1	1
		12:00	16:00	1	1
		16:00	20:00	1	1
		20:00	0:00	1	1
	Week-end	0:00	4:00	0.5	2
		4:00	8:00	1	1
		8:00	12:00	1	1
		12:00	16:00	1	1
		16:00	20:00	1	1
		20:00	0:00	1	1

(d) Fourth quarter

Table 11: Original and modified cluster classifications, per cluster, per quarter.

In both schemes, only the clusters for which the flag 1 or higher are used in the regression, which translates into a higher number of regression clusters for the modified scheme. Moreover, c_R is set to 0 when the flag is 0, and to 1 when the flag is 0.5. In any other cases, it is computed according to (11). For the rare quarterly instances where the regression parameters a and b are undefined because too few clusters were included in the regression model, c_R is simply set to c.

The original scheme was implemented at CATSA's behest to be compatible with their schedule optimizer; the modified scheme, which will be implemented in any future iteration of the model, retains this compatibility while allowing for a finer cluster classification.

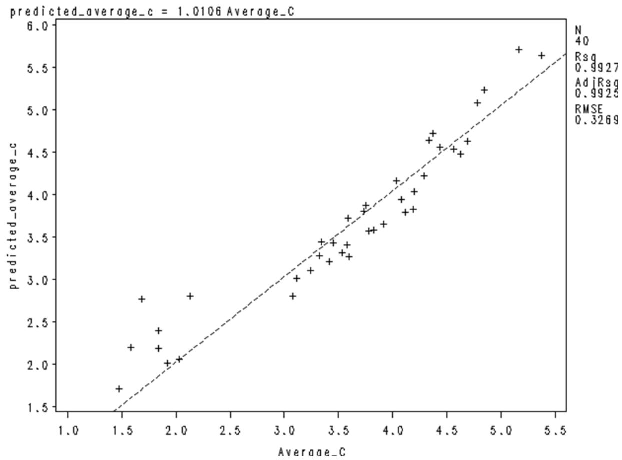
YEG (DI) — 2012 (continued)

The two classification schemes are shown in Table 11. Note that there are no clusters flagged as 0 for this checkpoint.

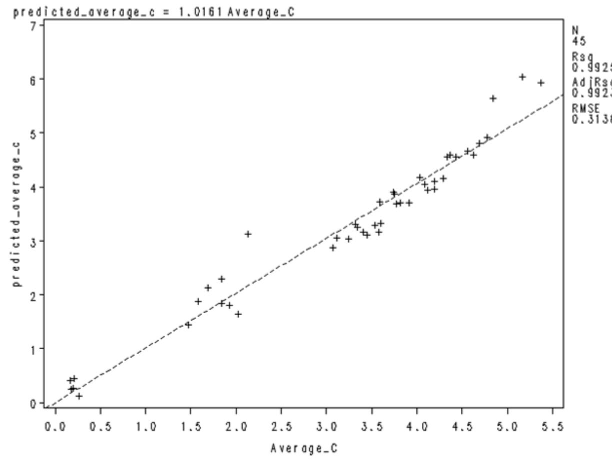
5.3 Validating the Combined Model, Checkpoint Departure

The relative accuracy of (11) suggests another method to validate the combined model. For any given checkpoint, the plot of c_R against c strongly suggests that the variables are linked according to

$$c_R = d \cdot c, \quad \text{for some } d.$$



(a) Original classification



(b) Modified classification

	d_{orig}	d_{mod}
YEG (DI)	1.0106	1.0161

(c) Departure parameter

Figure 6: Regression of the predicted average number of servers against the actual average number of servers, in both the original and the modified cluster classification. The departure parameters are also shown.

Linear regression once again determines the optimal $\hat{d}$ for each checkpoint. The **departure parameter** $\hat{d}$, then, serves as a measure of the predictive model's departure from reality.

If $\hat{d} \approx 1$ (i.e. $c_R \approx c$), then the assumptions that go into the combined model are justified *a postiori*, in the context of predicting the average number of active servers. The modified predictions c_D for a checkpoint where $\hat{d}$ is large or close to 0 may still be accurate, but an analysis should be undertaken to understand if any anomalous activity may be in play.

YEG (DI) — 2012 (continued)

The departure parameters and regressions of c_R against c are shown in Figure 6, for the two classifications. The clusters which were excluded from the regression in the original classification due to their small c values (because the checkpoint was closed for parts these clusters' time periods) appear in the bottom-left corner in the modified classification, and fit the linear pattern quite tightly, which is reflected in the near equal departure parameters for the two classification schemes.

6 Predicting the Number of Servers Under the Departure Assumption

But the departure parameter d plays another role: it can be used to refine the estimates of the predicted number of active servers c_R.

	Quarter	a	b	d
YEG (DI)	Jan 01 - Mar 31	0.581	0.621	1.0106
	Apr 01 - Jun 30	0.544	0.675	1.0106
	Jul 01 - Sep 30	0.482	0.753	1.0106
	Oct 01 - Dec 31	0.358	0.783	1.0106

(a) Regression and departure parameters

	2013	2014	2015	2016	2017	2018	2019
YEG (DI)	1.04919	1.04266	1.04177	1.03714	1.03157	1.03157	1.03157

(b) Forecasted arrival rate growth

Table 12: Regression and departure parameters, per quarter; arrival rate growth parameters, by year.

Given quarterly regression parameters a and b, and the checkpoint departure parameter d, setting

$$c_D = d \cdot c \tag{12}$$

in (5) yields

$$\mu_D = \frac{a}{d} c_D + b\lambda, \tag{13}$$

so that

$$p_{D,x} = 1 - \frac{\lambda}{\frac{a}{d}c + b\lambda} e^{-(\frac{a}{d}c + b\lambda - \lambda)x}, \tag{14}$$

and

$$c_D \approx \frac{d}{ax}\left[\ln\left(\frac{x\lambda}{1-p}e^{\lambda x}\right) + 1.031 \ln\left(\ln\left(\frac{x\lambda}{1-p}e^{\lambda x}\right)\right) + 0.207 - b\lambda x\right]. \tag{15}$$

YEG (DI) — 2012 (continued)

The quarterly regression parameters and the checkpoint parameter is shown in Table 12a.

6.1 Estimating the Service Rates and the Performance Levels

The service rates μ_D and the QoS levels $(p_{D,x}, x)$ can be computed directly from (13) and (14), exactly as in Sections 3.4 and 4.2.

YEG (DI) — 2012 (continued)

Since the focus of the Departure assumption is to provide refined predictions for the average number of active servers, the best estimates for the service rates and QoS levels are those given by the $M/M/1$ model (see Table 4) or the Regression model (see Table 7).

6.2 Forecasting the Number of Active Servers

The only information that is required in order to forecast the average number of active servers c_D for a cluster using (12) is:

- the regression parameters a and b;
- the checkpoint departure parameter d;
- an arrival rate λ and a QoS level (p, x).

For a given cluster, a, b and d are fixed and obtained *via* historical data. For a given QoS level, c_D is thus a function of λ, and this is the parameter for which a forecast is needed in order to provide a prediction.

YEG (DI) — 2012 (continued)

The arrival rate growth forecast for 2013—2019 is shown in Table 12b. The forecasted arrival rate value is obtained simply by multiplying the original cluster arrival rate by the appropriate growth factor. Various scenarios are shown in Table 13, using the modified cluster classification. Note that the predicted average number of active servers is automatically 1 for all clusters classified as 0.5.

6.3 Validating the Forecast

The validation in this case is a bit different: it makes little sense to compare the predicted value c_D with the actual c as the prediction depends not only on the arrival rate forecast (which could be different from the actual arrival rate), but also on the attained QoS levels (for which an independent forecast is unavailable). The best validation alternative is to wait for the data to be collected, determine the actual cluster arrival rate and QoS level and to use (12) to determine a new prediction c_D, which will then be compared with the actual c.

7 Results for All Checkpoints

As there are 26 checkpoints in total, results for each of them following the pattern established in this report for **YEG (DI) — 2012** have not been prepared in order to keep the length of this report down to a manageable size. However, the underlying data has been provided to CATSA in various tables.

8 Supplemental Comments and Recommendations

Perhaps the foremost conclusion is that the $M/M/1$ model on its own provides the best QoS levels estimates, while the best estimates of the average number of active servers are provided by the Departure model.

This discrepancy may be partly explained by the fact that, in any modeling endeavour, some loss of information is inevitable due to the necessity of making simplification assumptions. Here is a list of possible issues which could reduce the WTIM's accuracy.

1. The underlying arrival processes are roughly Poisson, and the wait time distributions are roughly conditionally exponential for each cluster; depending on the distance between the theoretical process and the empirical data, the $M/M/1$ assumption may be inappropriate.

2. The wait time distribution may be seriously biased as not every boarding pass has been scanned at S_1, and there are no easy way to verify how representative the subset of those for which wait time data is available actually is.

Cluster				Modified	Arrival Rate		Service Level				Pred # of Serv	
				Flag	λ_1	λ_2	p_1	p_2	m_1	m_2	1	2
Jan 01 to Mar 31 - YEG (DI) - 2012	Week day	0:00	4:00	0.5	0.058	0.060	85%	95%	15	20	1	1
		4:00	8:00	1.5	8.681	9.051	85%	95%	15	20	6.31	6.59
		8:00	12:00	1	6.588	6.869	85%	95%	15	20	4.81	5.03
		12:00	16:00	1	5.686	5.929	85%	95%	15	20	4.17	4.36
		16:00	20:00	1	5.311	5.537	85%	95%	15	20	3.90	4.08
		20:00	0:00	1	2.224	2.318	85%	95%	15	20	1.69	1.78
	Week-end	0:00	4:00	2	0.181	0.189	85%	95%	15	20	0.20	0.23
		4:00	8:00	1	6.671	6.955	85%	95%	15	20	4.87	5.09
		8:00	12:00	1	5.246	5.470	85%	95%	15	20	3.85	4.03
		12:00	16:00	1	4.394	4.582	85%	95%	15	20	3.24	3.40
		16:00	20:00	1	4.730	4.931	85%	95%	15	20	3.48	3.65
		20:00	0:00	1	1.684	1.755	85%	95%	15	20	1.30	1.37

(a) First quarter

Cluster				Modified	Arrival Rate		Service Level				Pred # of Serv	
				Flag	λ_1	λ_2	p_1	p_2	m_1	m_2	1	2
Apr 01 to Jun 30 - YEG (DI) - 2012	Week day	0:00	4:00	0.5	0.073	0.076	85%	95%	15	20	1	1
		4:00	8:00	1.5	8.653	9.022	85%	95%	15	20	5.96	6.23
		8:00	12:00	1	7.176	7.483	85%	95%	15	20	4.96	5.19
		12:00	16:00	1	5.865	6.115	85%	95%	15	20	4.08	4.27
		16:00	20:00	1	5.528	5.764	85%	95%	15	20	3.85	4.03
		20:00	0:00	1	2.309	2.407	85%	95%	15	20	1.67	1.76
	Week-end	0:00	4:00	2	0.105	0.110	85%	95%	15	20	0.14	0.17
		4:00	8:00	1	6.040	6.298	85%	95%	15	20	4.20	4.39
		8:00	12:00	1	6.000	6.256	85%	95%	15	20	4.17	4.36
		12:00	16:00	1	4.298	4.482	85%	95%	15	20	3.02	3.16
		16:00	20:00	1	4.118	4.293	85%	95%	15	20	2.89	3.03
		20:00	0:00	1	1.973	2.057	85%	95%	15	20	1.44	1.52

(b) Second quarter

Cluster				Modified	Arrival Rate		Service Level				Pred # of Serv	
				Flag	λ_1	λ_2	p_1	p_2	m_1	m_2	1	2
Jul 01 to Sep 30 - YEG (DI) - 2012	Week day	0:00	4:00	2	0.295	0.308	85%	95%	15	20	0.40	0.49
		4:00	8:00	1	8.756	9.129	85%	95%	15	20	4.69	4.94
		8:00	12:00	1	7.758	8.089	85%	95%	15	20	4.19	4.42
		12:00	16:00	1	5.881	6.132	85%	95%	15	20	3.26	3.44
		16:00	20:00	1	5.519	5.754	85%	95%	15	20	3.08	3.25
		20:00	0:00	1	2.974	3.100	85%	95%	15	20	1.80	1.93
	Week-end	0:00	4:00	0.5	0.299	0.312	85%	95%	15	20	1	1
		4:00	8:00	1	6.512	6.790	85%	95%	15	20	3.57	3.77
		8:00	12:00	1	6.784	7.073	85%	95%	15	20	3.71	3.91
		12:00	16:00	1	4.896	5.105	85%	95%	15	20	2.77	2.93
		16:00	20:00	1	4.319	4.503	85%	95%	15	20	2.48	2.63
		20:00	0:00	1	2.536	2.645	85%	95%	15	20	1.59	1.70

(c) Third quarter

Cluster				Modified	Arrival Rate		Service Level				Pred # of Serv	
				Flag	λ_1	λ_2	p_1	p_2	m_1	m_2	1	2
Oct 01 to Dec 31 - YEG (DI) - 2012	Week day	0:00	4:00	2	0.077	0.081	85%	95%	15	20	0.13	0.16
		4:00	8:00	1.5	8.884	9.263	85%	95%	15	20	6.34	6.63
		8:00	12:00	1	6.861	7.154	85%	95%	15	20	4.93	5.16
		12:00	16:00	1	5.905	6.157	85%	95%	15	20	4.26	4.46
		16:00	20:00	1	5.641	5.882	85%	95%	15	20	4.07	4.27
		20:00	0:00	1	2.406	2.508	85%	95%	15	20	1.81	1.91
	Week-end	0:00	4:00	2	0.140	0.146	85%	95%	15	20	0.19	0.22
		4:00	8:00	1	6.422	6.696	85%	95%	15	20	4.62	4.84
		8:00	12:00	1	6.480	6.756	85%	95%	15	20	4.66	4.88
		12:00	16:00	1	4.486	4.678	85%	95%	15	20	3.27	3.43
		16:00	20:00	1	4.271	4.453	85%	95%	15	20	3.12	3.27
		20:00	0:00	1	1.997	2.083	85%	95%	15	20	1.52	1.61

(d) Fourth quarter

Table 13: Predicted average number of servers under the modified cluster classification for two scenarios (arrival rates and service level performances), per cluster, per quarter.

3. The server vacation policy is unknown, and may not be uniformly adhered to (if one even exists).

4. The actual number of active servers is only crudely approximated by the maximum number of active lines within a 15 minute block.

5. The service rate seems to depend on factors other than the number of active servers and the arrival rate, leading to wildly different outputs for similar inputs and contributing to the lessened accuracy of the regression model when estimating QoS levels.

6. Different checkpoints might require different optimal clustering strategies.

It might be possible to minimize some of that information loss simply by selecting a slightly more sophisticated regression functional form in (4). Preliminary analysis suggests that the choice

$$\mu = \mu(c, \lambda) = ac + fc^2 + b\lambda$$

may provide better QoS results. Further analysis is needed in that regard, as it is clear that other factors need to be included in order to get the best possible fit and to minimize the number of clusters which become unstable as a result.

Finally, it is conceivable that while adding more historical data to the model, going too far back into the past may bias the results if policy changes have lead to characteristically distinct underlying data over the years. It seems clear that at least one year's worth of data is needed, but, as the datasets only contained trustworthy data for the year 2012, it is still too early to get a definitive answer on this topic.

A Service Level Curves

For a given checkpoint and cluster, assuming only that the $M/M/1$ model holds, it is straightforward to compute the processing rate μ corresponding to a set arrival rate λ and QoS level (p, x), according to the machinery developped in Section 3: indeed, from

$$p = 1 - \frac{\lambda}{\mu} e^{-(\mu-\lambda)x},$$

one obtains

$$\hat{\mu} = \frac{1}{x} W_0\left(\frac{x\lambda}{1-p} e^{\lambda x}\right), \tag{16}$$

where $W_0(x)$ is the Lambert W function (see Section 5). The cumulative distribution function (cdf) $p(x)$ can then be computed easily by noting that

$$p(x) = 1 - \frac{\lambda}{\hat{\mu}} e^{-(\hat{\mu}-\lambda)x} \quad \text{for } x = 0, 1, 2, \ldots.$$

Note that, theoretically, $p(0) = 1 - \frac{\lambda}{\hat{\mu}}$ represents the number of customers who are served with no wait time in the queue. As discussed previously, the quality of that estimate is directly linked to the quality of the waiting time data. The probability density function (pdf) can be estimated by computing the difference of successive cdf values: $f(x) = p(x+1) - p(x)$, for $x = 1, 2, \ldots$ (with the special exception that $f(1)$ represents the probability of either waiting 1 minute in the queue or no time at all).

Note that the number of servers (hence the three regression parameters a, b, d) do not enter the picture.

The cdf's for various clusters and or checkpoints can be combined by computing a weighted average, where the weights correspond to the number of arrivals: a cluster or checkpoint with a large number of arrivals contributes more to the overall number. An Excel template which computes both the exact value of the processing rates $\hat{\mu}$, and which combines the cdf's along checkpoint, airport and national lines, both quarterly and annually, has been provided.

B Improvements to the Original Regression Model

A number of modifications to the original regression model have been suggested to better reflect the nature of the data under consideration and to help provide stable estimates.

B.1 Clusters Re-classification

The clusters have been re-classified to allow clusters which are only open during parts of the time period (and for which the average number of servers may then be small) to be included in the regression (see Section 5.2).

B.2 Weighted Regression

Another improvement has been to use weighted regression: we still fit a regression model of the form

$$\frac{\mu}{c} = a + b\frac{\lambda}{c},$$

but we weigh the points according to the number of arrivals. The effect of this modification is to increase the importance of clusters with a larger number of arrivals in the regression, and consequently in the combined model. Graphically, this might lead to slopes that seem to "ignore" a certain number of points in Figure 4; numerically, this would be justified as the contribution of these clusters is minimal in the overall picture.

B.3 Outliers Removal

Another issue to consider is that we are seeking the "typical" cluster behaviour for a given quarter and checkpoint. As such, there might be clusters which are, for whatever reason, atypical or anomalous in that they have a much stronger influence on the regression results than would be expected from a typical point: a cluster with a very small arrival rate per server (even if it represents a small number of arrivals) but a very large processing rate (for whatever reason) would be an example of such a point, tending to drag the slope of the regression line away from the "true" slope of the regression model.

Standard methods to identify and remove these unduly influential observations (using the Mahalanobis distance) have been implemented in the SAS code.

B.4 Two-Year Period for Historical Data

In the best-case scenario, the number of clusters that appear in a regression is limited to twelve (there are six 4-hour periods during the day, and two types of day: WeekDay and WeekEnd), which is fairly low (especially considering that this is the best-case scenario). When the number of observations is that small, the appearance of only one new observation in later years (perhaps due to a checkpoint expanding its operating hours) can greatly modify the regression model, leading to wildly different predictions from year to year.

A possible solution is to increase the number of clusters appearing in the regressions, by evaluating data over a two-year period instead of a single year. As patterns may change beyond recognition over even a small time interval, going back further than two years in time is not recommended. A SAS implementation of this procedure, allowing for different weights to be given to either of the years, has been provided.

YEG (TB) — 2012, 2013

The modifications to the original regression model are illustrated using two years' worth (2012, 2013) of Edmonton's Transborder checkpoint data, for each of the quarters. Notice the increased number of data points, together with the effect of the weights in Figure 7.

C Enhancements to the WTIM

After having had the chance to run the WTIM on real data, CATSA requested 4 enhancements in 2015:

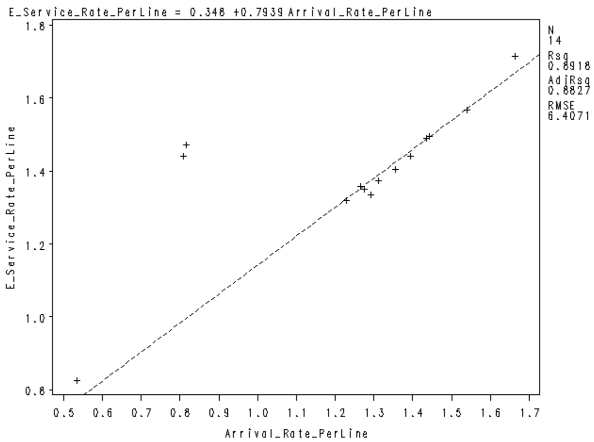
1. Improved accuracy: Enhance CATSAs WTIM 1-year and 2-year versions to increase overall accuracy of the forecast on an airport level, and on quarterly results.

2. Quarterly Rotation: Adjust WTIM to enable producing quarterly output and update on a rotational basis based on the most recent 4 quarters (as opposed to the most recent calendar year only).

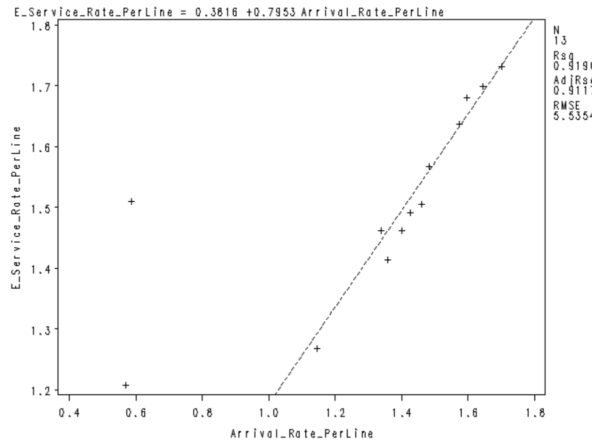
3. Weighting methodology: Determine best possible weighting values for the 2-year WTIM.

4. Code refinement: Overall iterative refinement of current WTIM code in terms of:

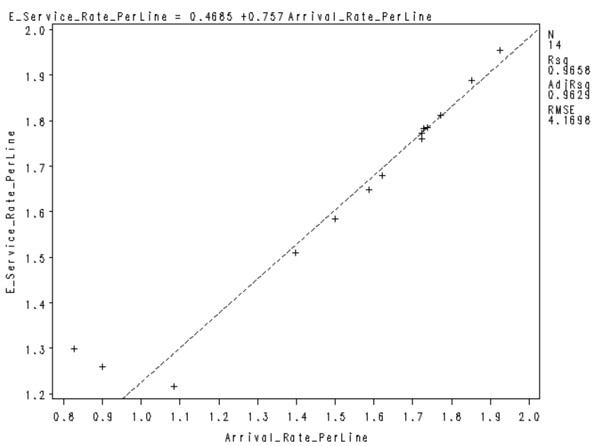
    - eliminating redundancies;
    - streamlining input of new data;
    - enhancing ability to obtain specific output;



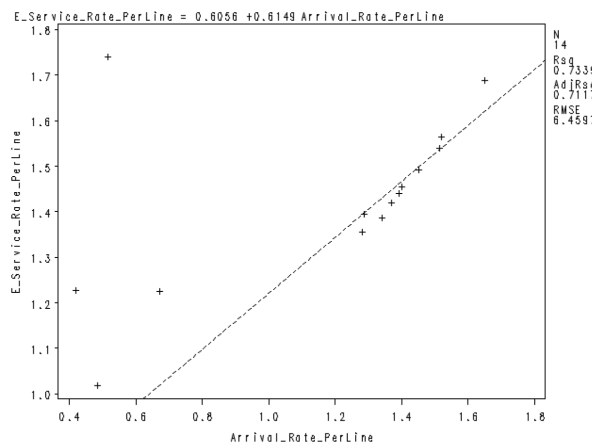
(a) First quarter



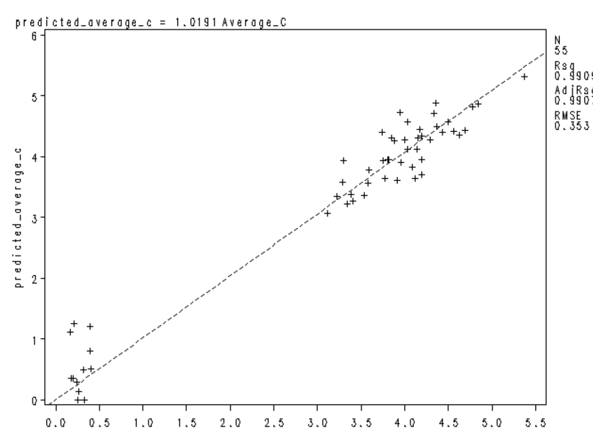
(b) Second quarter



(c) Third quarter



(d) Fourth quarter



(e) Departure regression

Figure 7: Regression of the cluster estimated service rates (per server) against the cluster arrival rates (per server), per quarter. The departure regression (with 55 observations) is also shown.

- flagging of compatibility or formatting issues with new data to enhance usability;
- eliminating the need for manual adjustments and changes throughout the code when including more recent data, such as adjustment for years, fixed parameters, generation of curves, etc.

These four objectives can be divided into two groups:

- objective 2 (quarterly rotation) and objective 4 (code refinement) only require modifications to the code, whereas
- objective 1 (improved accuracy) and objective 3 (weighting methodology) will require additional data exploration and the potential for more sophisticated modeling approaches.

The SAS code was modified and streamlined to meet the objectives of the first group. The process also discovered a number of more efficient paths, allowing for the WTIM run-time to be decreased by one order of magnitude, a significant reduction.

C.1 Improved Accuracy

Given that the original version of the WTIM works on a "one-size-fits-all" basis (in the sense that the same clusters and the same regression framework is used for all the checkpoints, regardless of their size and activity levels), there was some hope that a more flexible approach, taking into consideration not only the average arrivals for a cluster and the maximum number of open servers during each 15 minute period, but also the size and detailed short-term and long-term traffic trends at each checkpoint would provide more accurate estimates.

To that effect, we considered various functional forms $\mu(c,\lambda)$ in (4), separately for each checkpoint, season, and peak-time period. Whereas we were originally fitting

$$\frac{\mu}{c} = a + b\frac{\lambda}{c},$$

for parameters a, b, the enhanced model considers 9 potential functional forms, 7 of which end up being linear in c:

$$\frac{\mu}{c} = a_1 + \beta_1\frac{\lambda}{c} + \nu_1\frac{\sqrt{\lambda}}{c} \tag{17}$$

$$\frac{\mu}{c} = \alpha_2\frac{\lambda}{c^2} + \beta_2\frac{\lambda}{c} + \nu_2\frac{\sqrt{\lambda}}{c} \tag{18}$$

$$\frac{\mu}{c} = a_3 + \beta_3\frac{\lambda}{c} + \nu_3\frac{\sqrt{\lambda}}{c} + \gamma_3\lambda + \eta_3\frac{\lambda^2}{c} \tag{19}$$

$$\frac{\mu}{c} = a_4 + \beta_4\frac{\lambda}{c} + \eta_4\frac{\lambda^2}{c} \tag{20}$$

$$\frac{\mu}{c} = a_5 + \beta_5\frac{\lambda}{c} \tag{21}$$

$$\frac{\mu}{c} = a_6 + \gamma_6\lambda \tag{22}$$

$$\frac{\mu}{c} = \beta_9\frac{\lambda}{c} + \nu_9\frac{\sqrt{\lambda}}{c} + \gamma_9\lambda + \eta_9\frac{\lambda^2}{c}, \tag{23}$$

the last 2 being quadratic in c:

$$\frac{\mu}{c} = a_7 + \alpha_7\frac{\lambda}{c^2} + \beta_7\frac{\lambda}{c} + \nu_7\frac{\sqrt{\lambda}}{c} + \gamma_7\lambda + \eta_7\frac{\lambda^2}{c} \quad (24)$$

$$\frac{\mu}{c} = \alpha_8\frac{\lambda}{c^2} + \beta_8\frac{\lambda}{c} + \nu_8\frac{\sqrt{\lambda}}{c} + \gamma_8\lambda + \eta_8\frac{\lambda^2}{c}. \quad (25)$$

As before, the goal is to fit the functions to the data and ultimately solve

$$1 - p = \frac{\lambda}{\mu(c,\lambda,\sqrt{\lambda},\lambda/c,\lambda c,\lambda^2)} e^{(\lambda-\mu(c,\lambda,\sqrt{\lambda},\lambda/c,\lambda c,\lambda^2))x} \quad (26)$$

for c, given a desired QoS level (p,x). In the original version of the WTIM, such solutions would be estimated using approximations of the Lambert function W_0; in the enhanced version, c is estimated directly, using `proc iml` functionality that only became available with the latest release of SAS and Enterprise Guide.

These 9 functional forms were selected based on our prior experience with the data — other functions could conceivable be used, subject to one requirement: the existence and uniqueness of a solution $c > 0$ over a reasonable interval.

As it happens, the data does not stongly support these models, except for models (20) and (21), although this could change with new data becoming available.

For each checkpoint and each season, the departure parameter d is computed for all admissible models. The best model is then selected according to some criterion. Some of the potential criteria that have been considered include:

1. minimizing $\left|d_{\text{mod}}c_{\text{mod}} - c_{\text{avg}}\right|$
2. minimizing $|d_{\text{mod}} - 1|$
3. minimizing the regression mean square error (RMSE)

Initial tests suggest that the first criterion provides better estimates in the long run, although that is also highly data-dependent and could change when more data becomes available.

C.2 Weighting Methodology

Lastly, there comes the matter of calculating the lagged weights, which depends on whether we are using one year's worth of data (1 season) or two years' worth of data (5 seasons).

In the first case, no weight calculations are necessary. In the second case, we start by regressiing the processing rates μ_t against the lagged processing rates μ_{t-L}, $L = 1,2,3,4$:

$$\mu_t = \alpha + \sum_{L=1}^{4}\beta_L\mu_{t-L}.$$

The weights are then given as relative strength of the Sum of Squares for each lagged processing rate, as can be seen in the SAS file 'Data Summary, Regressions, and Predictions - 2 Years', lines 535-640.

References

[1] Burke, P.J. [1956], "The Output of a Queuing System", *Operations Research* vol **4** (6): 699704.

[2] Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J. and Knuth, D.K. [1996], "On the Lambert W Function", *Advances in Computational Mathematics*, vol **5**: 329—359.

[3] Newell, G.F. [1971], *Applications of Queuing Theory*, Chapman and Hall.

[4] Ross, S.M. [2010], *Introduction to Probability Models*, 10th ed., Academic Press.

[5] Walrand, J. [1983], "A probabilistic look at networks of quasi-reversible queues", *IEEE Transactions on Information Theory*, vol **29** (6): 825831.