# Queueing Networks with Path-Dependent Arrival Processes

Kerry Fendick
Johns Hopkins University Applied Physics Laboratory, Kerry.Fendick@jhuapl.edu
Ward Whitt
Industrial Engineering and Operations Research, Columbia University, ww2040@columbia.edu

This paper develops a Gaussian model for an open network of queues having a path-dependent net-input process, whose evolution depends on its early history, and satisfies a non-ergodic law of large numbers. We show that the Gaussian model arises as the heavy-traffic limit for a sequence of open queueing networks, each with a multivariate generalization of a Polya arrival process. We show that the net-input and queue-length processes for the Gaussian model satisfy non-ergodic laws of large numbers with tractable distributions.

*Key words*: path-dependent stochastic processes, generalized Polya process, Gaussian Markov process, diffusion approximations, queues, heavy-traffic limit

*History*: November 28, 2022

## 1. Introduction

The standard model for a queueing system has arrivals and service completions occurring as discrete events. That leads to the number in system over time being a stochastic process with a pure-jump net input process. Nevertheless, stochastic models with continuous net input processes have proven to be very useful to understand the behavior and manage the performance of queueing systems. Prominent among these are the Gaussian queues, the focus of this special issue, where the net input process is modelled as a Gaussian process; see Mandjes [16] and the other papers in this special issue.

Continuous Gaussian queueing models typically arise as limits of standard queueing models as the scale increases. The classic case is the heavy-traffic limit for an open network of queues, as discussed in Whitt [19], leading to reflected Brownian motion. Models with continuous Gaussian net input processes also have been applied directly as approximations, as they were for queueing

networks in Harrison [12] using Brownian motion and for large-deviation limits for queues with strongly dependent input processes in Mandjes [16] using fractional Brownian motion and related processes.

We here focus on a Gaussian queueing model that is the heavy-traffic limit of a standard queueing model with an arrival process that is a path-dependent stationary point process. For a stationary point process, path dependence is characterized by satisfying a non-ergodic law of large numbers. The long-run behavior of path-dependent arrival processes depends strongly on their early histories. Path-dependent processes commonly result from self-reinforcing behavior (e.g., epidemics and financial contagion) but they are of more general interest for queueing models whenever there is significant uncertainty about the long-run time average of the net input process; see our previous paper [9] for more discussion. For models with path-dependent arrival processes, useful descriptions focus on the transient distributions.

In contrast to the Gaussian models in [12, 16], for the path-dependent arrival processes considered here, the variance in the number of arrivals in a given interval grows faster than the expected value. For our Gaussian model of the net input process, and indeed for all counting processes considered in this paper, the variances grow as order $t^2$, while the expected value is proportional to $t$. Consequently, the powerful approximation for the steady-state queue length of a Gaussian queue in Section 5.4 of Mandjes [16] then does not apply. Nevertheless, Functional Central Limit Theorems (FCLTs) and Heavy Traffic Limit Theorems (HTLTs) do apply for our model and lead to useful approximations of the transient behavior.

In our previous paper [9], we established a HTLT for a queue with an arrival process that is a path-dependent stationary point process. In this paper, we extend our previous results to queueing networks. Queueing networks have been applied in a variety of contexts including traffic, network, manufacturing, risk, and reliability theory. In an example from reliability theory described in [18], different failed component may require service at different sequences of service stations for diagnosis, repair, assembly, and testing. Here, we study open queueing networks comprised of an

arbitrary number or queues and Markovian routing with an arbitrary routing matrix determining the sequence of queues that must be visited.

Following Cha and Badia [2], we allow dependence between the external arrival processes to the different queues in the network by modeling them as superpositions of independent path-dependent processes. We show that the workload and queue-length processes then converge to heavy-traffic limits that are reflected Gaussian processes. We also show that the limit processes themselves satisfy non-ergodic Laws of Large Numbers (LLN) with tractable distributions. Our prior simulation results in [9] for a single queue with path dependent arrivals in heavy traffic suggest that the LLN limits are useful for approximating the distribution of the queue-length distribution at finite times to first order.

The arrival processes to our queueing network are superpositions of independent Generalized Polya Processes (GPPs). In Cha and Finkelstein [3], GPPs are described as suitable models for failure processes in reliability and risk theory. As discussed in [9, 10], GPPs are a generalization of the classical Polya Process derived in Feller [7] from the Polya urn model, an early example of a path-dependent process. Stationary GPPs were shown in [9] to have non-ergodic LLN limits.

In Cha and Badia [2], superpositions of GPPs were proposed as models of failure processes when failures of different components can occur simultaneously. For the model in [2], the failure process for each component is modeled as the superposition of one GPP representing failures caused by external shocks and a second GPP representing failures from other causes. The first GPP, which appears in the superpositions of all components, results in their mutual dependence through simultaneous failures. We generalize the model in [2] to allow general superpositions of GPPs but focus here on the case in which the GPPs are stationary.

A FCLT was developed for stationary GPPs in [9] with a stationary Gaussian Markov limit process. We extend that result here to show that a multivariate process built from superpositions of stationary GPPs has a stationary multivariate Gaussian Markov limit. The dependence between the superpositions is captured by the covariance structure of that limit. When centered, the Gaussian limit process satisfies the self-reinforcing multivariate linear Stochastic Differential Equation (SDE),

$$\boldsymbol{G}(0) = 0 \quad \text{and} \quad d\boldsymbol{G}(t) = -B\left(A - Bt\right)^{-1}\boldsymbol{G}(t)\,dt + A^{1/2}d\boldsymbol{W}(t), \quad t \geq 0, \tag{1}$$

where $A$ and $-B$ are positive definite matrices and $\boldsymbol{W}$ is standard multivariate Brownian motion.

We then derive an HTLT where such superpositions are the exogenous inputs to a queueing network. There are then two sources of dependence between the queue-length or workload processes of the network's queues: dependence between the exogenous arrival processes, as described above, and dependence because of the routing of the same customers through multiple queues. We show that the limit processes for the queue-length and workload are reflected Gaussian Markov processes expressing all sources of dependence through their parameters. The results are an extension of results for a single-server queue derived in [9], where an exact transient distribution and a non-ergodic LLN were obtained for the queue-length limit process. The transient distribution of the limit process obtained here for a network of queues remains to be determined, but we succeed in extending the result from [9] for the non-ergodic LLN.

We also describe how the non-ergodic LLN limits for the Gaussian net input and associated queue length process are modified by conditioning on an observation of their states at an intermediate time. We show that the later the observation, the smaller the dispersion of the LLN limit for the conditioned net input process, the closer the conditioned net input process becomes to a Brownian motion with drift, and the slower its variance grows with time; see Corollary 3. The results therefore show how how the system becomes more predictable over finite intervals as our knowledge of its history grows.

The remainder of the paper is organized as follows: Section 2 reviews the definition and properties of GPPs and describes their superpositions. Section 3 develops a FCLT for multivariate superpositions of stationary GPPs with a continuous Gaussian process limit. Section 4 then develops two HTLTs for a network of queues with such superpositions as the exogenous arrival processes; one is for the workload process in a fluid model, while the other is for the standard queue length process. Section 5 derives a non-ergodic LLN for the HTLT limit processes and associated conditional processes associated with observation later in time. Finally, Section 6 concludes with a brief discussion of extensions and remaining problems.

## 2. The Stationary Generalized Polya Superposition Process ($\psi$-GPSP)

In this section, we review the definition and properties of univariate GPPs, as developed in [1], [15], [9], and [10]. We then define a new class of multivariate point processes constructed by superposing stationary univariate GPPs.

A univariate GPP with parameter triple $(\kappa(t), \gamma, \beta)$ is defined in [1] as the orderly point process $\{N(t) : t \geq 0\}$ with $N(0) = 0$ and stochastic intensity function

$$\lambda^*(t \,|\mathcal{H}_t) = (\gamma N(t-) + \beta)\,\kappa(t),$$

where $\mathcal{H}_t$ denotes the internal history of $N$ up to time $t$ and $\kappa(t)$ is a positive integrable real-valued function, while $\beta$ and $\gamma$ are positive real numbers. For any time $t \geq 0$, $N(t)$ is a count of the number of arrivals from the GPP up to $t$. The point process $N$ is an element of the space $D \equiv D[0,\ \infty)$ of right-continuous real-valued functions with left-hand limits on $[0,\ \infty)$, endowed with one of the Skorohod topologies and Borel sigma-field, as in [19]. By the definition of an orderly point process, a univariate GPP is regular, which means that the probability of simultaneous arrivals is zero. A Non-Homogenous Poisson Process (NHPP) is the special case of a GPP where $\gamma = 0$. For background on point processes and their intensity functions, see Section 3.3 and 7.2 of [5].

A GPP $N$ is stationary (meaning that it has stationary increments) if

$$\kappa(t) = \frac{1}{(\gamma t + 1)}, \quad t \geq 0. \tag{2}$$

As in [9], we then say that $N$ is the $\psi - GPP$ with parameter pair $(\gamma, \beta)$, where $\psi$ is a mnemonic for "stationary increments". A univariate GPP is a $\psi$-GPP if and only if it has a constant rate, as Remark 2 of [10] discusses. When (2) holds,

$$\mathbb{E}[N(t)] = \beta t \quad \text{and} \quad \mathbb{C}ov[N(s), N(t)] = \beta s + \beta\gamma s t, \quad \text{and} \quad 0 \leq s \leq t, \tag{3}$$

so that $N$ has constant rate $\beta$ and $Var(N(t) = \beta t + \beta\gamma t^2$, which is of order $t^2$ as $t$ increases.

Before turning to multivariate processes, we describe vector and matrix notation that we will use throughout this document.

*Notation.* We use the convention that a vector $x \equiv (x_1, x_2, \ldots, x_m)$ is a column vector with coordinates that are indexed through subscripts or superscripts. Superscripts are used for the coordinate processes of multivariate stochastic processes, which are written in bold font, e.g., $\boldsymbol{X} \equiv (X^1, X^2, \ldots, X^m)$. Subscripts are used for the coordinates otherwise as in $x$ above. The transpose of a vector or matrix $x$ is written as $x^T$. For a vector $x$, we let $diag\,(x)$ denote the square diagonal matrix with $(diag\,(x))_{i,i} = x_i$ and $(diag\,(x))_{i,j} = 0$ for $i \neq j$. For a real matrix $\Sigma$, we let $\Sigma^{1/2}$ denote another real matrix of the same dimension satisfying $\left(\Sigma^{1/2}\right)\left(\Sigma^{1/2}\right)^T = \Sigma$. (When $\Sigma$ is real positive definite, $\Sigma^{1/2}$ always exists.) Let $a \vee b \equiv max\,(a, b)$ for $a$ and $b$ in $\mathbb{R}$ and $(c \vee d) \equiv ((c_1 \vee d_1), (c_2 \vee d_2), \ldots, (c_k \vee d_k))$ for $c$ and $d$ in $\mathbb{R}^k$. Let $D^m$ be the $m$-dimensional product space of functions in $D$, endowed with the product topology. If $x$ and $y$ are in $D^m$ for some $m \geq 1$, let $x \circ y$ denote their coordinate-wise composition in $D^m$, i.e., $(x \circ y)_i = x_i \circ y_i$. We define Hadamard notation for other coordinate-wise operations on two vectors. In particular, for two vectors $x$ and $y$ of the same size, let $x \odot y$ be their coordinate-wise product, i.e., $(x \odot y)_i = x_i y_i$. Also let $x^{\odot v}$ denote coordinate-wise exponentiation of the vector $x$ by the real scaler $v$, i.e., $(x^{\odot v})_i = x_i^v$.

We now define a stationary multivariate point process constructed from superpositions of independent univariate $\psi$-GPPs. We define a multivariate $\psi - GPP$ as any multivariate point process with coordinate processes that are univariate $\psi - GPP$s. Let $\boldsymbol{V} \equiv (V^1, V^2, \ldots, V^m)$ for $m \geq 2$ be the multivariate $\psi - GPP$ in $D^m$ with independent coordinate process $V^i$ that is a univariate $\psi - GPP$ with parameter pair $(\gamma_i, \beta_i)$ for $i = 1, \ldots, m$. Let $\gamma \equiv (\gamma_1, \gamma_2, \ldots, \gamma_m)$ and $\beta \equiv (\beta_1, \beta_2, \ldots, \beta_m)$. We then say that $\boldsymbol{V}$ has parameters $(\gamma, \beta)$.

The paper [2] provides motivation for the superposition process that we will define. There, $m+1$ independent GPPs are mapped into m superpositions with each superposition of the form $N_i = V_i + V_{m+1}$ for $1 \leq i \leq m$. Here, we consider only superpositions of $\psi - GPPs$, but we otherwise generalize the model in [2] in two ways. First, we map $m$ independent $\psi$-GPPs into $k$ superpositions for any $1 \leq k \leq m$ by multiplying a vector of independent $\psi - GPPs$ by a matrix. Second, the elements of the matrix are non-negative integers, so that the coefficients used by the superposition

can be greater than one. The resulting superpositions are therefore counting processes (integer-valued) but can have jumps greater than one.

DEFINITION 1. A process $\boldsymbol{N}$ in $D^k$ for $k \geq 1$ is a stationary Generalized Polya Superposition Process ($\psi$-GPSP) with parameters $(\gamma, \beta, M)$ if

$$\boldsymbol{N} \equiv \left(N^1, N^2, \ldots, N^k\right) = M\boldsymbol{V}$$

where $\boldsymbol{V}$ is a multivariate $\psi$-GPP in $D^m$ for $m \geq k$ with independent coordinate processes and parameters $(\gamma, \beta)$, and $M$ is a matrix of non-negative integers with dimension $k \times m$.

The superpositions are mutually dependent when they contain univariate $\psi$-GPPs in common. They are not, in general, regular. When superpositions contain univariate $\psi$-GPPs in common, an arrival from a common $\psi$-GPP results in simultaneous arrivals for the superpositions in which it appears. When the matrix $M$ contains elements greater than one, the individual superpositions have jumps greater than one, which also correspond to simultaneous arrivals.

The mean and covariance functions of a $\psi$-GPSP depend on time in the same way as a $\psi$-GPP.

PROPOSITION 1 **(Mean and covariance function of a $\psi$-GPSP)**. *For a $\psi$-GPSP $\boldsymbol{N}$ with parameters $(\gamma, \beta, M)$,*

$$\mathbb{E}\left[\boldsymbol{N}(t)\right] = M\beta t \quad \textit{and} \quad \mathbb{C}ov\left[\boldsymbol{N}(s), \boldsymbol{N}(t)\right] = M\ diag\left(\beta\right) M^T s + M\ diag\left(\beta \odot \gamma\right) M^T st$$

*for $0 \leq s \leq t$.*

*Proof.* Using Definition 1, $\boldsymbol{N} = M\boldsymbol{V}$, where

$$\mathbb{E}\left[\boldsymbol{V}(t)\right] = \beta t \quad \text{and} \quad \mathbb{C}ov\left[\boldsymbol{V}(s), \boldsymbol{V}(t)\right] = diag\left(\beta s + \left(\beta \odot \gamma\right) st\right) \quad \text{for} \quad 0 \leq s \leq t.$$

The result for the mean then follows trivially, and the result for the covariance follows because

$$\begin{aligned}\mathbb{C}ov\left[\boldsymbol{N}(s), \boldsymbol{N}(t)\right] &= \mathbb{E}\left[\left(M\boldsymbol{V}(t) - M\beta t\right)\left(M\boldsymbol{V}(s) - \beta Ms\right)^T\right] \\ &= M\mathbb{C}ov\left[\boldsymbol{V}(s), \boldsymbol{V}(t)\right] M^T, \quad 0 \leq s \leq t. \quad \blacksquare\end{aligned}$$

The coordinate processes of a multivariate $\psi$-GPP need not be independent. Here is an example where the coordinate processes are dependent.

PROPOSITION 2 **(A multivariate GPP with dependence)**. *When $\boldsymbol{N}$ in $D^k$ is the $\psi$-GPSP with parameters $(\gamma, \beta, M)$, where $\gamma_1 = \gamma_2 = \ldots = \gamma_m = \hat{\gamma} > 0$, and the matrix $M$ contains only zeros and ones, then the coordinate process $N^j$ is a univariate GPP with parameter pair $\left(\hat{\gamma}, (M\beta)_j\right)$ for $j = 1, \ldots, k$.*

*Proof.* This result follows immediate from Theorem 1 of [2]. ∎

Because the coordinate processes in Proposition 2 are univariate GPPs, they are individually regular. Such processes are called "marginally regular" in [2], where "marginal process" has the same meaning as "coordinate process".

We conclude this section by giving a concrete example from [2].

EXAMPLE 1. (motivating example) Consider a multivariate $\psi - GPP$ with independent marginal one-dimensional $\psi - GPP$'s of dimension $m = 3$. Let $\boldsymbol{V} \equiv (V^1, V^2, V^3)$ be the multivariate $\psi - GPP$ in $D^3$ with independent coordinate processes $V^i$ that are univariate $\psi - GPP$'s with parameter pairs $(\gamma_i, \beta_i)$ for $i = 1, \ldots, 3$. Let $N$ be the associated $\psi$-GPSP with parameters $(\gamma, \beta, M)$, where $M$ is the $2 \times 3$ matrix

$$M = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \tag{4}$$

Then Proposition 1 holds with

$$\mathbb{E}\left[\boldsymbol{N}(t)\right] = \left(\left(\beta_1 + \beta_3\right) t, \quad \left(\beta_2 + \beta_3\right) t\right) \quad \text{and} \tag{5}$$

$$\mathbb{C}ov\left[\boldsymbol{N}(s), \boldsymbol{N}(t)\right] = \begin{pmatrix} \left(\beta_1 + \beta_3\right) s + \left(\beta_1\gamma_1 + \beta_3\gamma_3\right) st & \beta_3 s + \beta_3\gamma_3 st \\ \beta_3 s + \beta_3\gamma_3 st & \left(\beta_2 + \beta_3\right) s + \left(\beta_2\gamma_2 + \beta_3\gamma_3\right) st \end{pmatrix} \tag{6}$$

for $0 \leq s \leq t$.

For further insight, we now consider the symmetric case with parameter pairs $(\gamma_i, \beta_i) = (\hat{\gamma}, \hat{\beta})$ for $i = 1, \ldots, 3$. Hence, there is only the single parameter pair $(\hat{\gamma}, \hat{\beta})$. Then Proposition 1 holds with $\mathbb{E}\left[\boldsymbol{N}_i(t)\right] = 2\hat{\beta}t$ for $i = 1, 2$ and

$$\mathbb{C}ov\left[\boldsymbol{N}(s), \boldsymbol{N}(t)\right] = \begin{pmatrix} 2\hat{\beta}s(1 + \hat{\gamma}t) & \hat{\beta}s(1 + \hat{\gamma}t) \\ \hat{\beta}s(1 + \hat{\gamma}t) & 2\hat{\beta}s(1 + \hat{\gamma}t) \end{pmatrix}, \tag{7}$$

so that

$$Var\left[\boldsymbol{N}_i(t)\right] = 2\hat{\beta}t(1+\hat{\gamma}t) \quad \text{and} \quad Var\left[\boldsymbol{N}_1(t)+\boldsymbol{N}_2(t)\right] = 6\hat{\beta}t(1+\hat{\gamma}t). \tag{8}$$

The dependence makes the variance of the sum 3 times the variance of one term. The index of dispersion (IDC) of $\boldsymbol{N}_i(t)$ is thus $I_i(t) \equiv Var\left[\boldsymbol{N}_i(t)\right]/\mathbb{E}\left[\boldsymbol{N}_i(t)\right] = 1+\hat{\gamma}t$ for each $i$. The index of dispersion of the sum is thus $1.5(1+\hat{\gamma}t)$. Hence, for small $\hat{\gamma}t$, either due to small $\hat{\gamma}$ or small $t$ or both, the processes behave locally like a Poisson process, but for larger $\hat{\gamma}t$, the process is much more highly variable.

It is significant that the structure of the $\psi - GPSP$ exposed by this example also applies to the mean and variance of the limiting Gaussian $\psi - GMP$ net input process, introduced in the next section, because that structure is inherited by the limit; see Remark 2 at the end of §3.2. Recall that a Gaussian process is fully determined by its mean and covariance functions.

## 3. Functional central limit theorem for $\psi$-GPSPs

We now derive a FCLT for a sequence of $\psi$-GPSPs. We will show that these converge in distribution to a multivariate Gaussian Markov Process with stationary increments, called a $\psi$-GMP, studied in [8].

### 3.1. $\psi$-GMPs

We first review the definition and properties of a $\psi$-GMP from [8].

Definition 2. A process $\boldsymbol{G}$ in $D^k \equiv D^k[0,\ \infty)$ for $k \geq 1$ is a $\psi$-GMP with parameters $(A, B)$ if it is a zero-mean Gaussian process with $\boldsymbol{G}(0) = 0$ and

$$Cov\left[\boldsymbol{G}(s), \boldsymbol{G}(t)\right] = E\left[\boldsymbol{G}(t)\,\boldsymbol{G}^t(s)\right] = s\left(A - Bt\right), \quad 0 \leq s \leq t < \infty,$$

where $A$ and $B$ are symmetric matrices of $k \times k$ real scalars, $A$ is positive definite, and $B$ is negative definite.

The definition of a $\psi$-GMP in [8] requires $B$ only to be symmetric, but it may then be necessary to restrict $\boldsymbol{G}$ to $D^k[0,T]$ for some $T < \infty$. By Theorem 3 of [8], $\boldsymbol{G}$ has a representation as a solution to the linear stochastic differential equation (SDE) in (1), where $\boldsymbol{W}$ is standard $k$-variate Brownian

motion (with mean zero and covariance matrix $I$, the identity matrix). It follows that a $\psi$-GMP has almost surely continuous sample paths. When $A$ and $B$ have the assumed properties, $A^{1/2}$ exists, and $(A-Bt)^{-1}$ always exists because $A-Bt$ is positive definite for all $t \geq 0$. If we relax the negative definite assumption for $B$ and assume that $B=0$, then $\boldsymbol{G}$ is a multivariate Brownian motion with zero drift and $\mathbb{C}ov\left[\boldsymbol{G}(s), \boldsymbol{G}(t)\right] = sA$ for $0 \leq s \leq t < \infty$.

The following results describe properties of $\psi$-GMPs:

PROPOSITION 3 **(Linear map of a $\psi$-GMP)**. *If $\hat{\boldsymbol{G}}$ is a $\psi$-GMP in $D^m$ with parameters $\left(A=\hat{A}, B=\hat{B}\right)$ as defined in Definition 2, and $M$ is a real matrix of dimension $k \times m$ with $rank(M)=k$, where $1 \leq k \leq m$, then $\boldsymbol{G}=M\hat{\boldsymbol{G}}$ is a $\psi$-GMP in $D^k$ with parameters*

$$\left(A=M\hat{A}M^T, B=M\hat{B}M^T\right).$$

*Proof.* Because $\hat{\boldsymbol{G}}$ is a zero-mean Gaussian process with $\hat{\boldsymbol{G}}(0)=0$, the process $\boldsymbol{G} \equiv M\hat{\boldsymbol{G}}$ has those same properties. Using the definition of a $\psi$-GMP,

$$\begin{aligned}\mathbb{C}ov\left[\boldsymbol{G}(s), \boldsymbol{G}(t)\right] = \mathbb{E}\left[\boldsymbol{G}(t)\boldsymbol{G}^T(s)\right] &= \mathbb{E}\left[M\hat{\boldsymbol{G}}(t)\hat{\boldsymbol{G}}^T(s)M^T\right] \\ &= s\left(M\hat{A}M^T - M\hat{B}M^T t\right) \quad \text{for} \quad 0 \leq s \leq t.\end{aligned}$$

With the assumed properties of $\hat{A}$, $\hat{B}$, and $M$, the parameter matrix $A=M\hat{A}M^T$ is symmetric positive definite and $B=M\hat{B}M^T$ is symmetric negative definite. Therefore, $\boldsymbol{G}$ satisfies the definition of a $\psi$-GMP with parameters $(A,B)$. ∎

If $\boldsymbol{G}$ is a $\psi$-GMP in $D^k$ and $\omega$ is a vector in $\mathbb{R}^k$, then the process $\boldsymbol{X}(t) \equiv \omega t + \boldsymbol{G}(t)$ for $0 \leq t < \infty$ is called a $\psi$-GMP with drift $\omega$. The next two lemmas describe conditional $\psi$-GMPs with drift as defined by their conditional finite-dimensional distributions. The first of those lemmas states that conditioning a $\psi - GMP$ with drift on its state at the end of an interval results in a multivariate Brownian bridge with a new drift on the interval.

PROPOSITION 4 **(Lemma 2 of [8])**. *If $\boldsymbol{X}(t) \equiv \omega t + \boldsymbol{G}(t)$ in $D^k$ for $t \geq 0$, where $\omega$ is a vector in $\mathbb{R}^k$ and $\boldsymbol{G}$ is a $\psi$-GMP in $D^k$ with parameters $(A,B)$ as defined in Definition 2, and*

$$\boldsymbol{X}_t(s) \equiv \left(\boldsymbol{X}(s) \,|\, \boldsymbol{X}(t)\right) \quad \textit{for} \quad 0 \leq s \leq t,$$

*then* $\boldsymbol{X}_t(s) = st^{-1}\boldsymbol{X}(t) + \boldsymbol{U}_t(s)$ *a.e. for* $0 \le s \le t$*, where* $\boldsymbol{U}_t$ *is a zero-mean Brownian bridge (a Gaussian process) in* $D^k$*, independent of* $\boldsymbol{X}(t)$*, with* $\boldsymbol{U}_t(0) = 0$ *and*

$$\mathbb{C}ov\left[\boldsymbol{U}_t(s_1), \boldsymbol{U}_t(s_2)\right] = \mathbb{E}\left[\boldsymbol{U}_t(s_2)\boldsymbol{U}_t^T(s_1)\right] = s_1\left(A - s_2 t^{-1} A\right)$$

*for* $0 \le s_1 \le s_2 \le t$.

Remark 1. Under the more permissive definition of a $\psi$-GMP from [8], where the parameter matrix $B$ need only be symmetric but need not be negative definite, $\boldsymbol{U}_t$ is a $\psi$-GMP with parameters $(A, t^{-1}A)$. Because the conclusions of Proposition 4 do not depend on $B$, the same process $\boldsymbol{X}_t$ is obtained as $B \to 0$ and $\boldsymbol{X}$ then approaches a multivariate Brownian motion with drift. The process $\boldsymbol{X}_t$ therefore can be described as a Brownian motion with drift conditioned on its end state and therefore a Brownian bridge.

The next proposition is analogous to the restart property for GPPs from [1]; see also Proposition 1 of [9] for a statement of that result. In this case, the $\psi$-GMP is conditioned on its state at the start of an interval.

Proposition 5 **(Lemma 4 of [8])**. *If* $\boldsymbol{X}(t) \equiv \omega t + \boldsymbol{G}(t)$ *in* $D^k$ *for* $t \ge 0$*, where* $\omega$ *is a vector in* $\mathbb{R}^k$ *and* $\boldsymbol{G}$ *is a* $\psi$*-GMP in* $D^k$ *with parameters* $(A, B)$ *as in Definition 2, and*

$$\boldsymbol{X}^s(t) \equiv \left(\boldsymbol{X}(t+s) - \boldsymbol{X}(s) | \boldsymbol{X}(s)\right) \quad \textit{for} \quad 0 < s \le t+s,$$

*then* $\boldsymbol{X}^s(t) = \omega_s t + \boldsymbol{G}^s(t)$ *a.e. for* $t \ge 0$*, where* $\omega_s = \omega - B\left(A - Bs\right)^{-1}\left(\boldsymbol{X}(s) - s\omega\right)$ *and* $\boldsymbol{G}^s$ *is a* $\psi$*-GMP in* $D^k$*, independent of* $\boldsymbol{X}(s)$*, with parameters* $\left(A,\ B_s = B\left(A - Bs\right)^{-1}A\right)$.

### 3.2. Convergence to a $\psi$-GMP

We show convergence of a normalized sequence of $\psi$-GPSPs to a $\psi$-GMP with zero drift. For $k \ge 1$, let $(D^k, WM_1)$ denote the space $D^k \equiv D^k[0,\infty)$ endowed with the Skorohod weak $M_1$ topology, and let $\Rightarrow$ denote convergence in distribution in $(D^k, WM_1)$; see Sections 12.3 and 12.9 of [19] for background. When the limit processes are continuous, as all limits in this paper will be, the $WM_1$ topology reduces to the topology of uniform convergence on compact sets (u.o.c.) in each

of the coordinates. We will use the $WM_1$ topology primarily to avoid measurability issues with the uniform topology discussed in Section 11.5.3 of [19]. The Borel $\sigma-$field generated by the $WM_1$ topology coincides with the Kolmogorov $\sigma-$field generated by the coordinate projections. Throughout, $n$ will always refer to the sequence index used for limit theorems. When $n$ is used as a superscript for a process, it is not to be confused with a coordinate index or an exponent.

The following result for convergence of multivariate $\psi$-GPPs with independent coordinate processes is the immediate consequence of a result for convergence of univariate $\psi$-GPPs from Theorem 4 of [9] and presented in a more convenient form for the application here in Proposition 4 of [10]. Theorem 4 of [9] describes a FCLT for sums of i.i.d. $\psi$-GPPs as the number becomes becomes large. Because a single $\psi - GPP$ with parameter $\beta = nb$ has the same distribution as a sum of $n$ $\psi$-GPPs with parameter $\beta = b$, Proposition 4 of [10] provides a FCLT with the same limit as in Theorem 4 of [9] for a sequence of individual $\psi$-GPPs, but where the parameter $\beta$ is scaled by the sequence index. The scaling therefore differs from the scaling of time used by Donsker's theorem to obtain a Brownian motion limit for ergodic processes. Donsker's theorem is applied in Section 4 to obtain Brownian motion limits for the service and routing process for queueing network models.

LEMMA 1 **(Proposition 4 of [10])**. *If $\boldsymbol{V}^n$ is a multivariate $\psi$-GPP in $D^m$ with independent coordinate processes and parameters $(\gamma, \beta = nb)$, where $n \geq 1$ is a positive integer and $b \geq 0$ is a vector in $\mathbb{R}^m$, and $\boldsymbol{V}_n(t) \equiv n^{-1/2}(\boldsymbol{V}^n(t) - nbt)$, then $\boldsymbol{V}_n \Rightarrow \boldsymbol{V}$ in $(D^m, WM_1)$ as $n \to \infty$, where $\boldsymbol{V}$ is the $\psi$-GMP with parameters $\left(A = diag(b),\ B = -diag(b \odot \gamma)\right)$.*

*Proof.* Using the assumed independence of the coordinate processes $\boldsymbol{V}^{n,i}$ for $i = 1, \ldots, m$, the result follows from Proposition 4 of [10] and Theorem 11.4.4 of [19]. ∎

We can now state and prove a result for convergence of $\psi$-GPSPs to a $\psi$-GMP.

THEOREM 1 **(Convergence to a $\psi - GMP$)**. *If $\boldsymbol{N}^n$ in $D^k$ is the $\psi$-GPSP with parameters $(\gamma, \beta = nb, M)$, where $n \geq 1$ is a positive integer and $rank(M) = k$, and if*

$$\boldsymbol{N}_n(t) \equiv n^{-1/2}(\boldsymbol{N}^n(t) - nMbt), \tag{9}$$

*then $\boldsymbol{N}_n \Rightarrow \boldsymbol{N}$ in $(D^k, WM_1)$ as $n \to \infty$, where $\boldsymbol{N}$ is the $\psi - GMP$ with parameters $(A = M(diag(b))M^T, B = -M(diag(b \odot \gamma))M^T)$.*

*Proof.* Because $(D^k, WM_1)$ is a product topology, mappings on it are continuous if they are continuous in each coordinate. Multiplication by a matrix $M$ can be viewed for each coordinate process in $D$ as a combination of addition and multiplication by constants. In general, addition is not a continuous mapping on $(D, WM_1)$, but Corollary 12.7.1 of [19] states that it is continuous when the limit process is continuous, as it is here. Remark 12.71 on page 411 of [19] implies that addition is measurable on $(D, WM_1)$. Hence, in the case that is relevant here, multiplication by the matrix $M$ is a continuous measurable mapping on $(D^k, WM_1)$. By Definition 1, $\boldsymbol{N}^n = M\boldsymbol{V}^n$, where $\boldsymbol{V}^n$ is defined in Lemma 1. Lemma 1 and Theorem 3.4.3 of [19] then imply that $\boldsymbol{N}_n = M\boldsymbol{V}_n \Rightarrow M\boldsymbol{V}$,where $\boldsymbol{V}$ is also defined in Definition 1. The result that $M\boldsymbol{V} = \boldsymbol{N}$ then follows from Proposition 3. Clearly, $A = M\left(diag\left(b\right)\right)M^T$ is symmetric positive definite, and $B = -M\left(diag\left(b \odot \gamma\right)\right)M^T$ is symmetric negative definite. ∎

Remark 2. A comparison of Proposition 1 and Theorem 1 shows that the prelimit process $\boldsymbol{N}_n$ has the same covariance function as the limit process $\boldsymbol{N}$ for all $n \geq 1$.

## 4. HTLTs for queueing networks with $\psi$-GPSP arrivals

We now state and prove HTLTs for the multivariate workload (buffer content) and queue-length processes for queueing networks with infinite buffers.

For our model of the workload process, work arrives to each queue exogenously (from outside the network) in discrete quanta but departs from each queue as a fluid. Proportions of work departing each queue are routed out of the network or to other queues for service.

For our model of the queue-length process, customers arrive to each queue as a point process, receive service, and then are routed out of the network or to other queues for service.

### 4.1. The reflection map

The HTLTs for both models will involve the multidimensional reflection map, which we now describe. Let $P$ be a substochastic matrix (non-negative with row sums less than or equal to one) of dimension $k \times k$. Let $Q = P^T$ be such that $Q^p \to 0$ as $p \to \infty$. The multidimensional reflection

map $(\phi,\psi) \equiv (\phi,\psi)_Q : D^k \to D^{2k}$ is a mapping of any $x$ in $D^k$ into a unique $(y,z) = (\phi(x), \psi(x))$ in $D^{2k}$ such that

$$z = x + (I - Q)\, y \geq 0,$$

$$y^i \text{ is nondecreasing with } y^i(0) = 0 \quad \text{for} \quad i = 1, 2, \ldots, k, \quad \text{and}$$

$$\int_0^\infty z^i(t)\, dy^i(t) = 0 \quad \text{for} \quad i = 1, 2, \ldots, k.$$

The element $y$ is called the regulator component and $z$ is called the content component of the reflection map. The element $x$ is called the reflection map's net input. For background, including proof of existence and uniqueness of the multidimensional reflection map, see Chapter 14 of [19].

### 4.2. Queueing network fluid model

Our model for the queueing-network workload process is based on the fluid model developed in Sections 14.2 and 14.6 of [19]. Work will arrive to each queue from outside the network in successive quanta. Each queue's server will process and output work continuously at a constant rate whenever the queue is not empty. The quanta sizes for any queue will have a unit mean and finite variance, but the quanta variance and rate at which work is served may vary from queue to queue. A proportion $0 \leq P_{ij} \leq 1$ of the output of work from queue $i$ will be routed to queue $j$ to serve as input, and a proportion $1 - \sum_j P_{ij} > 0$ will leave the network. The matrix $P \equiv (P_{ij})$ will be called the routing matrix. Because we represent multivariate processes as column vectors, it will be convenient to define $Q \equiv P^T$, which we assume has the property that $Q^p \to 0$ as $p \to \infty$ so that work eventually leaves the network. The transposed matrix $Q$ is called the reflection matrix. A queue's workload (or total buffer content content) at each point in time is the work that has arrived to the queue but has not yet been served.

**4.2.1. Net input process for the fluid model.** We define the queueing network's net input process and prove a FCLT for it. The limit will be a multivariate $\psi - GMP$ with drift.

We consider a sequence of models indexed by $n \geq 1$. Each model has a network of $k$ queues and the $k \times k$ reflection matrix $Q$, which is the same for all models. In the $n^{th}$ model, quanta

of work arrive to the queueing network from outside as a $\psi$-GPSP $\boldsymbol{N}^n$ in $D^k$ with parameters $(\gamma, \beta = nb, M)$. The coordinate process $N^{n,i}$ is the exogenous arrival process of work quanta to queue $i$. The sequence of work quanta from successive arrivals to queue $i$ is the sequence $\{V_{j,i} : j \geq 1\}$ of i.i.d random variables with $\mathbb{E}[V_{j,i}] = 1$ and $\mathbb{V}ar[V_{j,i}] = c_{si}^2$. The same sequence will apply for queue $i$ in all models. We assume that the sequences $\{V_{j,i} : j \geq 1\}$ for different queue indices $i$ are mutually independent and independent of the exogenous arrival process. Let $\boldsymbol{S}^n \equiv (S^{n,1}, S^{n,2}, \ldots, S^{n,k})$, where $S^{n,i}(t) \equiv \sum_{j=1}^{\lfloor nt \rfloor} V_{j,i}$, and

$$\boldsymbol{S}_n(t) \equiv n^{-1/2}(\boldsymbol{S}^n(t) - ntI), \ t \geq 0.$$

Then, the classical Donsker's theorem in Section 4.3 of [19] implies that

$$\boldsymbol{S}_n \Rightarrow diag(c_s)\boldsymbol{W} \text{ in } \left(D^k, WM_1\right) \text{ as } n \to \infty, \tag{10}$$

where $c_s = (c_{s1}, c_{s2}, \ldots c_{sk})$, and $\boldsymbol{W}$ is standard $k$-variate Brownian motion. Our assumptions about the distribution of work quanta enter into the results that follow only through that limit. As discussed in [10], the same limit holds with a different interpretation of $c_s$ when the assumption that $\{V_{j,i} : j \geq 1\}$ is an i.i.d. sequence is relaxed in various ways.

The total exogenous input process of work to model $n$ is $\boldsymbol{T}^n \equiv (T^{n,1}, T^{n,2}, \ldots, T^{n,k})$, where

$$T^{n,i}(t) \equiv \sum_{j=1}^{N^{n,i}(t)} V_{j,i} \quad \text{for} \quad i = 1, \ldots, k.$$

The random variable $T^{n,i}(t)$ represents the total service requirements of all exogenous arrivals in $N^{n,i}$ to queue $i$ over the interval $[0, t]$. For the $n^{th}$ model, let $nr_i^n$ be the service rate of work at queue $i$, where $r^n \equiv (r_1^n, r_2^n, \ldots, r_k^n) > 0$ is in $\mathbb{R}^k$. Then the net input process of work for model $n$ is defined as

$$\begin{aligned} \boldsymbol{X}^n(t) &\equiv \boldsymbol{T}^n(t) + Qnr^n t - nr^n t \\ &= \boldsymbol{T}^n(t) - (I - Q)nr^n t, \ t \geq 0. \end{aligned} \tag{11}$$

The random variable $X^{n,i}(t)$, which can be negative, represents what the content of queue $i$ would be at time $t$ if the queues were initially empty and always busy, so that the output from all queues

occurred continuously at their respective rates $nr_j$ for $j = 1, 2, \ldots, k$ without interruption due to idleness.

We now state and prove a FCLT for joint convergence of the net input process and other processes. Let $e \equiv e(t) = t$, and

$$\boldsymbol{N}_n(t) \equiv n^{-1/2}\left(\boldsymbol{N}^n(t) - nMbt\right), \quad \boldsymbol{T}_n(t) \equiv n^{-1/2}\left(\boldsymbol{T}^n(t) - nMbt\right),$$
$$\text{and } \boldsymbol{X}_n(t) \equiv n^{-1/2}\boldsymbol{X}^n(t), \; t \geq 0 \text{ and } n \geq 1. \tag{12}$$

LEMMA 2 **(FCLT for the net input process)**. *If* $(\boldsymbol{N}_n, \boldsymbol{T}_n, \boldsymbol{X}_n)$ *in* $D^{3k}$ *are defined as in* (12)*, where* $\boldsymbol{N}^n$ *from* (9) *is the* $\psi$*-GPSP with parameters* $(\gamma, \beta = nb, M)$*, and*

$$n^{1/2}\left(Mb - (I - Q)\, r^n\right) \to \omega \quad \text{in} \quad \mathbb{R}^k \quad \text{as} \quad n \to \infty, \tag{13}$$

*then* $(\boldsymbol{N}_n, \boldsymbol{T}_n, \boldsymbol{X}_n) \Rightarrow (\boldsymbol{N}, \boldsymbol{T}, \boldsymbol{X})$ *in* $(D^{3k}, WM_1)$ *as* $n \to \infty$*, where* $\boldsymbol{N}$ *is a* $\psi$*-GMP with parameters* $(A = M(diag(b))M^T, B = -M(diag(b \odot \gamma))M^T)$*,* $\boldsymbol{T}$ *is the* $\psi$*-GMP with parameters* $(A = M(diag(b))M^T + diag(c_s^{\odot 2} \odot Mb), B = -M(diag(b \odot \gamma))M^T)$*, and* $\boldsymbol{X} = \boldsymbol{T} + \omega e$*.*

*Proof.* By Theorem 1, $\boldsymbol{N}_n \Rightarrow \boldsymbol{N}$. Coordinate-wise composition is continuous on $(D^k, WM_1)$ because $(D^k, WM_1)$ is a product topology, and composition is continuous in $(D, WM_1)$ in each coordinate under the conditions of Theorem 13.3.1 of [19]. Then

$$\begin{aligned}
\boldsymbol{T}_n &\equiv n^{-1/2}\left(\boldsymbol{T}^n - nMbe\right) \\
&= n^{-1/2}\left(\boldsymbol{S}^n \circ \left(n^{-1}N^n\right) - nMbe\right) \\
&= n^{-1/2}\left(\left(n^{1/2}\boldsymbol{S}_n + Ine\right) \circ \left(n^{-1}\boldsymbol{N}^n\right) - nMbe\right) \\
&= \boldsymbol{S}_n \circ \left(n^{-1/2}\boldsymbol{N}_n + Mbe\right) + \boldsymbol{N}_n \Rightarrow diag(c_s)\,\boldsymbol{W} \circ (Mbe) + \boldsymbol{N},
\end{aligned}$$

where the convergence follows from (10) and Theorem 3.4.3 of [19]. Because

$$\mathbb{C}ov\left(\boldsymbol{W}(s), \boldsymbol{W}(t)\right) = sI \quad \text{for} \quad s \leq t,$$

it follows that $\boldsymbol{T}_n \Rightarrow \boldsymbol{T}$. Using the above result and (13),

$$\begin{aligned}
\boldsymbol{X}_n \equiv n^{-1/2}\boldsymbol{X}^n &= n^{-1/2}\left(\boldsymbol{T}^n - nMbe\right) + n^{1/2}\left(Mb - (I - Q)\, r^n\right)e \\
&\Rightarrow \boldsymbol{T} + \omega e.
\end{aligned}$$

Joint convergence follows from the continuous mapping theorem in Theorem 3.4.3 of [19]. ∎

**4.2.2. HTLT for the fluid model.** The potential output rate from each queue is equal to its constant service rate when the queue has positive content. Some of that potential output will be lost when the queue is empty. The multivariate workload process for a network of queues is obtained from the net input process by adjusting for the cumulative lost potential output process. If, for the $n^{th}$ model, $\boldsymbol{Z}^n$ in $D^k$ is the workload process, $\boldsymbol{L}^n$ in $D^k$ is the cumulative lost potential output process, and $\boldsymbol{X}^n$ is the net input process from (11), then

$$\begin{aligned}\boldsymbol{Z}^n(t) &= \boldsymbol{Z}^n(0) + \boldsymbol{T}^n(t) + Q\left(nr^n t - \boldsymbol{L}^n(t)\right) - \left(nr^n t - \boldsymbol{L}^n(t)\right) \\ &= \boldsymbol{Z}^n(0) + \boldsymbol{X}^n(t) + (I-Q)\,\boldsymbol{L}^n(t) \geq 0 \quad for\ all \quad t \geq 0.\end{aligned}$$

For $\boldsymbol{L}^n$ to have the interpretation as the cumulative lost potential output, its coordinate processes $L^{n,i}$ must each be non-decreasing with $L^{n,i}(0) = 0$ and must increase only at times when $Z^{n,i}$ is equal to zero. We recognize those properties from the description of the reflection map $(\phi, \psi) : D^k \to D^{2k}$ in Section 4.1 and define

$$(\boldsymbol{Z}^n, \boldsymbol{L}^n) \equiv \left(\phi\left(\boldsymbol{Z}^n(0) + \boldsymbol{X}^n\right), \psi\left(\boldsymbol{Z}^n(0) + \boldsymbol{X}^n\right)\right)$$

and

$$(\boldsymbol{Z}_n, \boldsymbol{L}_n) \equiv \left(n^{-1/2}\boldsymbol{Z}^n, n^{-1/2}\boldsymbol{L}^n\right). \tag{14}$$

THEOREM 2 **(HTLT for the workload process)**. *Under the assumptions and definitions of Lemma 2 and (14), if $\boldsymbol{Z}_n(0) \Rightarrow \boldsymbol{Z}(0) \geq 0$ in $\mathbb{R}^k$, then*

$$(\boldsymbol{N}_n,\ \boldsymbol{T}_n, \boldsymbol{X}_n, \boldsymbol{Z}_n, \boldsymbol{L}_n\ ) \Rightarrow (\boldsymbol{N},\ \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{L}) \quad in \quad \left(D^{5k}, WM_1\right) \quad as \quad n \to \infty,$$

*where $(\boldsymbol{Z}, \boldsymbol{L}) \equiv \left(\phi\left(\boldsymbol{Z}(0) + \boldsymbol{X}\right), \psi\left(\boldsymbol{Z}(0) + \boldsymbol{X}\right)\right)$.*

*Proof.* The result follows from Lemma 2, Theorem 14.5.2 of [19] for the continuity of the reflection map, and Theorem 3.4.3 of [19] for the joint convergence. ∎

### 4.3. Standard open queueing network model

The standard Open Queueing Network (OQN) model defined below for the queue-length process is based on the framework developed in Section 14.7 of [19] for queueing networks with service interruptions that occur regardless of whether queues are idle. Here, we consider the special case without such service interruptions. Each queue has a single server with the first-in first-out discipline. We again consider a sequence of models indexed by $n \geq 1$. The queueing network for all models will consist of $k$ queues. For each pair of queues $(i, j)$, let $R^{i,j}(p) \geq 0$ for $p \geq 1$ be the total number of customers immediately routed to queue $j$ among the first $p$ departures from queue $i$, such that $\sum_{j=1}^{n} R^{i,j}(p) \leq p$. We call $\boldsymbol{R}$ the routing process, which will be the same for all models in the sequence of models.

For the $n^{th}$ model, customers arrive to the queueing network from outside as the $\psi$-GPSP $\boldsymbol{N}^n$ in $D^k$ with parameters $(\gamma, \beta = nb, M)$. The coordinate process $N^{n,i}$ is the exogenous arrival process of customers to queue $i$.

To specify the service time process for queue $i$, first let $\left\{\check{S}^i(t) : t \geq 0\right\}$ in $D$ for $i = 1, 2, \ldots, k$ be a renewal counting process associated with a sequence of positive i.i.d random variables having mean $\mu_i^{-1} > 0$ and coefficient of variation (standard deviation divided by the mean) $c_{si}$. We also assume that the processes $\check{S}^i$ for $i = 1, \ldots, k$ are mutually independent and independent of the arrival process. We next introduce an additional scaling depending on $n$ to introduce a drift in the limit, as we will need for the HTLT in Theorem 3 below. In particular, let

$$\widetilde{S}^{n,i}(t) \equiv \check{S}^i(\eta_{n,i}t), \quad S^{n,i}(t) \equiv \widetilde{S}^{n,i}(nt), \quad \text{and} \quad S_n^i(t) \equiv n^{-1/2}\left(S^{n,i}(t) - \mu_i nt\right), \quad t \geq 0. \tag{15}$$

The counting process $S^{n,i}$ is the potential service process for queue $i$, i.e., the process of service completions that would occur at queue $i$ if queue $i$ were always busy. Let $\mu \equiv (\mu_1, \mu_2, \ldots \mu_k)$, $c_s \equiv (c_{s1}, c_{s2}, \ldots c_{sk})$, $\eta_n \equiv (\eta_{n,1}, \eta_{n,2}, \ldots, \eta_{n,k})$, $\boldsymbol{S}^n \equiv (S^{n,1}, S^{n,2}, \ldots, S^{n,k})$, and $\boldsymbol{S}_n \equiv (S_n^1, S_n^2, \ldots, S_n^k)$.

LEMMA 3 **(Convergence of service process to multivariate BM with drift)**. *If*

$$\sqrt{n}\,(\eta_n - 1) \to \eta \quad in \quad \mathcal{R}^k \quad as \quad n \to \infty, \tag{16}$$

*where* 1 *is the vector of* 1*s in* $\mathbb{R}^k$*, then*

$$\boldsymbol{S}_n \Rightarrow diag\left(\mu^{\odot 1/2} \odot c_s\right) \boldsymbol{W}_s + \mu \odot \eta e \quad in \quad \left(D^k, WM_1\right) \quad as \quad n \to \infty.$$

*where* $\boldsymbol{W}_s$ *is standard* $k-$*dimensional Brownian motion.*

*Proof.* For the $n^{th}$ model, let $\check{S}_n^i$ be the standard scaling of $\check{S}^i$ for its FCLT involving translation, time scaling by $n$ and spatial scaling by $\sqrt{n}$, i.e.

$$\check{S}_n^i(t) \equiv n^{-1/2}\left(\check{S}^i(nt) - \mu_i nt\right), \quad t \geq 0.$$

By Donsker's theorem, Theorem 4.3.1 of [19], and the equivalence between FCLT's for partial sums and associated counting processes, from Corollary 7.3.1 of [19],

$$\check{S}_n^i \Rightarrow \mu_i^{1/2} c_{si} W_s^i \quad \text{in} \quad (D, WM_1) \quad \text{as} \quad n \to \infty,$$

where $W_s^i$ is standard Brownian motion independent of $W_s^j$ for $j \neq i$. The result in Lemma (16) then follows coordinate-wise from Theorem 13.3.1 of [19]; see the proof of Corollary 3 of [9] for essentially the same argument. Convergence in $D^k$ then follows from the assumed independence of the coordinate processes using Theorem 11.4.4 of [19]. ∎

In Section 14.7 of [19] the routing process is assumed to satisfy a FCLT with a long term rate (translation term) $P_{i,j}$ corresponding to the long-term proportion of customers routed from queue $i$ to queue $j$. To obtain a limit for the queue-length process that is a multivariate reflected $\psi$-GMP with drift, we must make more specific assumptions about the routing process than those required in Section 14.7 of [19]. In particular, we will consider the common case of Markovian routing, where the probability a customer departing from queue $i$ is routed to queue $j$ is equal to $P_{i,j}$, independently of prior routing decisions and the arrival and service processes, where $P$ is a nonnegative substochastic $k \times k$ matrix with transpose $Q \equiv P^T$ satisfying the assumptions for a reflection matrix of Section 4.1. Then,

$$\begin{aligned} \boldsymbol{R}_n(t) \equiv n^{-1/2}\left(\boldsymbol{R}(\lfloor nt \rfloor) - Pnt\right) \Rightarrow \hat{\boldsymbol{R}}(t) \\ \equiv \left(\left(\Gamma^1\right)^{1/2} \boldsymbol{W}_{r1}(t), \left(\Gamma^2\right)^{1/2} \boldsymbol{W}_{r2}(t), \ ..., \left(\Gamma^k\right)^{1/2} \boldsymbol{W}_{rk}(t)\right)^T \end{aligned} \quad (17)$$

in $\left(D^{k^2}, WM_1\right)$ for $t \geq 0$, where $\boldsymbol{W}_{ri}$ for $i = 1, \ldots, k$ are independent standard Brownian processes in $D^k$, independent of the arrival and service processes, and $\Gamma^i$ for $i = 1 \ldots, k$ are $k \times k$ covariance matrices with

$$\Gamma^i_{j,j} = Q_{j,i}\left(1 - Q_{j,i}\right) \quad \text{and} \quad \Gamma^i_{j,l} = -Q_{j,i}Q_{l,i} \quad \text{for} \quad j \neq l;$$

(Recall that the Markovian routing induces the multinomial distribution and that $Q$ is the transpose of $P$; see (7.6) on page 178 of [11].)

With the above definitions, the arrival process $N^{n,i}(t)$ and potential service process $S^{n,i}(t)$ already grow at a mean rate proportional to $nt$. For consistency with the time scaling used in Theorem 14.7.4 of [19], we define $A^{n,i}(t) \equiv N^{n,i}(t/n)$, $t \geq 0$, and note that $\widetilde{S}^{n,i}(t) = S^{n,i}(t/n)$, $t \geq 0$ by (15). We will thus get our original processes back without change when we scale time by $n$ as in Theorem 14.7.4 of [19] and Theorem 3 below. We can now describe the queue-length process $Z^{n,i}$ for the $i^{th}$ queue in the $n^{th}$ model as

$$Z^{n,i}(nt) = Z^{n,i}(0) + A^{n,i}(nt) + \sum_{j=1}^{k} R_{j,i}\left(\widetilde{S}^{n,j}\left(B^{n,j}(nt)\right)\right) - \widetilde{S}^{n,i}\left(B^{n,i}(nt)\right), \quad t \geq 0, \tag{18}$$

where

$$B^{n,i}(t) \equiv \int_0^t 1_{\left\{Z^{n,i}(s) > 0\right\}} ds. \tag{19}$$

The process $B^{n,i}(t)$ has the interpretation as the cumulative busy time of server $i$ during the interval $[0, t]$. Let $\boldsymbol{Z}^n \equiv (Z^{n,1}, Z^{n,2}, \ldots, Z^{n,k})$ and $\boldsymbol{B}^n \equiv (B^{n,1}, B^{n,2}, \ldots, B^{n,k})$. By Theorem 14.7.1 of [19], there is a unique solution $(\boldsymbol{Z}^n, \boldsymbol{B}^n)$ in $D^{2k}$ with those properties.

A key step in Section 14.7 of [19] is representing the queue length process defined above as the image of the reflection map applied to an appropriate potential net input process. Theorem 14.7.2 of [19] constructs such a potential net input process, which appears in (7.5)-(7.8) on p. 498 of [19]. It's $j^{th}$ coordinate process is

$$\begin{aligned} X^{n,j}(nt) \equiv & \left(\sum_{i=1}^{k} M_{j,i}b_i + \sum_{i=1}^{k} P_{i,j}\mu_i - \mu_j\right) nt \; + \; \left[A^{n,j}(nt) - n\sum_{i=1}^{k} M_{j,i}b_i t\right] \\ & + \sum_{i=i}^{k}\left[R_{i,j}\left(\widetilde{S}^{n,i}\left(B^{n,i}(nt)\right)\right) - P_{i,j}\widetilde{S}^{n,i}\left(B^{n,i}(nt)\right)\right] + \sum_{i=i}^{k} P_{i,j}\left[\widetilde{S}^{n,i}\left(B^{n,i}(nt)\right) - \mu_i B^{n,i}(nt)\right] \\ & - \left[\widetilde{S}^{n,j}\left(B^{n,j}(nt)\right) - \mu_j B^{n,j}(nt)\right] \end{aligned} \tag{20}$$

for $1 \leq j \leq k$. In (20) above as in (7.5)-(7.8) of [19], there are terms that cancel in order to group the terms to correspond to those of the net input limit process obtained below. Theorems 14.7.2 of [19] shows that queue-length process in (18) can be represented as $\boldsymbol{Z}^n = \phi(\boldsymbol{X}^n)$, where $\boldsymbol{X}^n$ is defined by (20) and $\phi$ is the content component of the reflection map from Section 4.1 with $Q = P^T$. The representation in (20) therefore will support a HTLT showing joint convergence of the net input and queue-length processes.

We now state and prove the HTLT for the standard OQN with Markovian routing and the $\psi$-GPSP arrival process. In the statement of the result, $\boldsymbol{X}$ is the limit for the sequence of normalized net input processes, and its reflection $\boldsymbol{Z}$ is the limit of the sequence of normalized queue length processes. The limit for the net input process is remarkably tractable because it is a $\psi$-GMP and thus a Gaussian process.

THEOREM 3 **(HTLT for the net input and queue-length processes)**. *If*

$$\boldsymbol{N}_n(t) \equiv n^{-1/2}(\boldsymbol{N}^n(t) - nMbt), \quad \boldsymbol{X}_n(t) \equiv n^{-1/2}\boldsymbol{X}^n(nt), \quad \boldsymbol{Z}_n(t) \equiv n^{-1/2}\boldsymbol{Z}^n(nt),$$
$$\boldsymbol{S}_n(t) \equiv n^{-1/2}(\boldsymbol{S}^n(t) - \mu nt), \quad \hat{\boldsymbol{X}}_n(t) \equiv n^{-1}\boldsymbol{X}^n(nt), \quad \hat{\boldsymbol{Z}}_n(t) \equiv n^{-1}\boldsymbol{S}^n(nt),$$
$$\hat{\boldsymbol{S}}_n(t) \equiv n^{-1}\boldsymbol{S}^n(t), \quad \textit{and} \quad \hat{\boldsymbol{B}}_n(t) \equiv n^{-1}\boldsymbol{B}^n(nt) \quad \textit{for} \quad t \geq 0,$$

*where* $\boldsymbol{N}^n$ *is the* $\psi$*-GPSP with parameters* $(\gamma, \beta = nb, M)$ *as in Theorem 1 and the definitions in (15) and (18)-(20) apply, and if* $\boldsymbol{Z}_n(0) \Rightarrow \boldsymbol{Z}(0) > 0$ *in* $\mathbb{R}^k$, $\mu \equiv (I - Q)^{-1} Mb$, *and* (16) *and* (17) *apply, then*

$$X_n^j = N_n^j + \sum_{i=1}^k (R_n)_{i,j} \circ \hat{S}_n^i \circ \hat{B}_n^i + \sum_{i=1}^k P_{i,j} S_n^i \circ \hat{B}_n^i - S_n^j \circ \hat{B}_n^j \quad \textit{for} \quad j = 1, 2, \ldots, k, \tag{21}$$

$\boldsymbol{Z}_n = \phi(\boldsymbol{Z}_n(\boldsymbol{0}) + \boldsymbol{X}_n)$ *where* $\phi$ *is the content component of the reflection map* $(\phi, \psi) \equiv (\phi, \psi)_Q : D^k \to D^{2k}$, *and*

$$(\boldsymbol{N}_n, \boldsymbol{S}_n, \boldsymbol{X}_n, \boldsymbol{Z}_n) \Rightarrow (\boldsymbol{N}, \boldsymbol{S}, \boldsymbol{X}, \boldsymbol{Z}) \quad \textit{in} \quad \left(D^{4k}, WM_1\right) \quad \textit{as} \quad n \to \infty,$$

*where*

$\boldsymbol{N}$ *is the* $\psi$*-GMP with parameters* $\left(A = M\left(diag\left(b\right)\right)M^T, \quad B = -M\left(diag\left(b \odot \gamma\right)\right)M^T\right)$,

$\boldsymbol{S} = diag\left(\mu^{\odot 1/2} \odot c_s\right)\boldsymbol{W_s} + \mu \odot \eta e$,

$\boldsymbol{W_s}$ *is standard Brownian motion in* $D^k$,

$\boldsymbol{X} \equiv \boldsymbol{N} + \hat{\boldsymbol{R}}^T \mu^{\odot 1/2} - \left(I - Q\right)\boldsymbol{S}$ *is the* $\psi$*-GMP with drift* $\omega \equiv -\left(I - Q\right)\left(\mu \odot \eta\right)$,

*and parameters* $\left(A = M\left(diag\left(b\right)\right)M^T + \sum_{i=1}^{k} \Gamma^i \mu_i + \left(I - Q\right) diag\left(\mu \odot c_s^{\odot 2}\right)\left(I - Q\right)^T\right.$,

$\left.B = -M\left(diag\left(b \odot \gamma\right)\right)M^T\right)$, *and*

$\boldsymbol{Z} = \phi\left(\boldsymbol{Z}\left(\boldsymbol{0}\right) + \boldsymbol{X}\right)$.

*Proof.* The assumptions of Theorem 14.7.4 of [19] hold with $H = 1/2$. We elaborate on the proof there because the net input process is not clearly identified on lines 6-7 on p. 502. From the convergence established on line 3 from the bottom of p. 501 and (7.5)-(7.7), we get $\hat{X}^i_{n_j} \Rightarrow 0e$ through the subsequence where the busy-time process converges. By Theorems 14.2.5 and 14.7.2, we then get $\hat{B}^i_{n_j} \Rightarrow e$ and apply Theorem 11.4.5 to establish the Functional Weak Law of Large Numbers (FWLLN) $(\hat{\mathbf{X}}_{n_j}, \hat{\mathbf{Z}}_{n_j}, \hat{\mathbf{B}}_{n_j}) \Rightarrow (0, 0, 1e)$ (where 0 and 1 in that context are the vectors of all zeros and ones in $\mathbb{R}^k$). Since that limit holds for all subsequences, we get the full sequence converging as stated there. The essential mathematical tool for proving convergence of (21) to the net input limit then is the preservation in the limit of composition with linear centering as in Section 13.3 of [19], in particular by application of Corollary 13.3.2.

In (7.18) of [19], $\lambda \equiv Mb$, $\mu$ and $R$, and $P$ are independent of $n$, while the displayed limits in (7.20) are established earlier in this paper. Since we have assumed that there are no service interruptions when the content of each queue is positive, $\boldsymbol{D}_n \equiv 0$ in (7.20) there. The condition that the joint limit is in $D_1$ w.p.1 (i.e., that it can have discontinuities in only one coordinate at a time) is satisfied because the limit processes have no discontinuities. The limits in (7.21) there holds because the matrix $P$ and the normalization constants used for the definitions of $\boldsymbol{N}_n$ and $\boldsymbol{S}_n$ do not depend on n. The limit in (7.22) there holds with $c \equiv 0$ since we have assumed that

$\mu \equiv (I-Q)^{-1} Mb$, i.e., $\mu$ is the solution of the traffic rate equation described in (5.7.1) on page 271 of [17]. The conclusions then follow from (7.23) and (7.24) there and the results of Theorem 1 and Lemma 3 here. The drift for the net input process limit is the result of the drift from the potential service process limit, as established by Lemma 3. ∎

EXAMPLE 2. (a symmetric two-queue example.) We start by constructing a symmetric two-queue example with Markovian routing. We first consider the standard case with Brownian motion limits for the arrival, service, and routing processes discussed in Remarks 14.7.1-14.7.4 of [19] and then afterwards extend to the case of a $\psi$-GMP limit for the arrival process. Let the $2 \times 2$ routing matrix $P$ be symmetric with $P_{1,2} = P_{2,1} = \theta$, $0 \leq \theta < 1$, and $P_{1,1} = P_{2,2} = 0$, so that there is no immediate feedback. We adopt all the simplifications in Remarks 14.7.1-14.7.4 of [19], so that the limit process for the net input process will be Brownian Motion (BM) with drift, while the limit for the queue length process will be Reflected Brownian Motion (RBM). In particular, we assume that $N_n^i(t) \equiv n^{-1/2}(N^{n,i}(nt) - \lambda_i nt) \Rightarrow c_{ai} W_a^i(t)$, $i = 1, 2$, for the arrival process, where $W_a^1$ and $W_a^2$ are independent standard Brownian motions. To simplify expressions, we let $\mu = (1,1)$ and $\lambda = (I-Q)\mu = (1-\theta, 1-\theta)$, as will be needed so that traffic intensities approach one for the sequence of systems. As in Lemma 3, we then assume that $S_n^i(t) \equiv n^{-1/2}(S^{n,i}(t) - nt) \Rightarrow c_{si} W_s^i(t) + \eta_i t$ for the service process, where $W_s^1$ and $W_s^2$ are independent standard Brownian motions.. Let $\boldsymbol{R}_n$ be the normalized routing process defined in (17), and let $\hat{\boldsymbol{R}}$ be its limit.

Because the arrival, service, and routing processes are assumed to be independent,

$$(\boldsymbol{N}_n, \boldsymbol{S}_n, \boldsymbol{R}_n) \Rightarrow \left(\boldsymbol{N}, \boldsymbol{S}, \hat{\boldsymbol{R}}\right) \quad \text{in} \quad \left(D^2 \times D^2 \times D^{2\times 2} \equiv D^8, WM_1\right) \quad \text{as} \quad n \to \infty,$$

where $\boldsymbol{N} = diag(c_a)\boldsymbol{W}_a$, $\boldsymbol{S} = diag(c_s)\boldsymbol{W}_s + \eta e$, and $\left(\hat{\boldsymbol{R}}^T\right)^i = (\Gamma^i)^{1/2} \boldsymbol{W}_{ri}(t)$ with $\boldsymbol{W}_a$, $\boldsymbol{W}_s$, and $\boldsymbol{W}_{ri}$ for $1 = 1, ..., k$ being independent standard $k$-dimensional Brownian motions, while $\Gamma_{j,j}^i = P_{i,j}(1 - P_{i,j})$ and $\Gamma_{j,l}^i = -P_{i,j}P_{i,l}$ for $j \neq l$. (For each $i$, the routing produces a $k$-dimensional multinomial distribution.) The model is therefore determined by three parameter vectors $(c_a, c_s, \eta)$ and the probability $\theta$.

It now remains to exhibit the limit process $\boldsymbol{X}$ for the net input process. Applying Theorem 3 with the modification to the assumptions about the arrival process (and recalling that $\boldsymbol{X}$ depends on $-(I-Q)\boldsymbol{S}$), it follows that $\boldsymbol{X}$ is a 2-dimensional BM having a Gaussian distribution with mean vector $\mathbb{E}[\boldsymbol{X}(t)] = -(\eta_1 - \theta\eta_2, \eta_2 - \theta\eta_1)\, t$ and covariance function

$$\mathbb{C}ov\left[\boldsymbol{X}(s), \boldsymbol{X}(t)\right] = s \begin{pmatrix} c_{a1}^2 + c_{s1}^2 + \theta^2 c_{s2}^2 + \theta(1-\theta) & -\theta(c_{s1}^2 + c_{s2}^2) \\ -\theta(c_{s1}^2 + c_{s2}^2) & c_{a2}^2 + c_{s2}^2 + \theta^2 c_{s1}^2 + \theta(1-\theta) \end{pmatrix} \tag{22}$$

for $s \le t$.

Now we change the model. We now assume that the arrival process is a $\psi$-GPSP with parameters $(\gamma, \beta = nb, M)$, which makes its limit process from Theorem 1 a $\psi$-GMP. We assume that the matrix $M$ is the one in (4) of Example 1. The random variable $\boldsymbol{X}(t)$ again has a Gaussian distribution with mean vector $\mathbb{E}[\boldsymbol{X}(t)] = -(\eta_1 - \theta\eta_2, \eta_2 - \theta\eta_1)\, t$. By Theorem 3, the covariance function changes to

$$\begin{aligned} \mathbb{C}ov\left[\boldsymbol{X}(s), \boldsymbol{X}(t)\right] = {} & s \begin{pmatrix} b_1 + b_3 + c_{s1}^2 + \theta^2 c_{s2}^2 + \theta(1-\theta) & b_3 - \theta(c_{s1}^2 + c_{s2}^2) \\ b_3 - \theta(c_{s1}^2 + c_{s2}^2) & b_2 + b_3 + c_{s2}^2 + \theta^2 c_{s1}^2 + \theta(1-\theta) \end{pmatrix} \\ & + st \begin{pmatrix} b_1\gamma_1 + b_3\gamma_3 & b_3\gamma_3 \\ b_3\gamma_3 & b_2\gamma_2 + b_3\gamma_3 \end{pmatrix} \quad \text{for} \quad s \le t. \end{aligned} \tag{23}$$

## 5. Non-ergodic law of large numbers for the limit processes

In Theorem 2 for the fluid model and Theorem 3 for the OQN model, the limit $\boldsymbol{X}$ for the net input process is a $\psi$-GMP with drift, and the limit $\boldsymbol{Z}$ for the workload or queue length process is a multidimensional reflection of $\boldsymbol{X}$. For $(\boldsymbol{X}, \boldsymbol{Z})$ with those properties, we state and prove a non-ergodic LLN, which then applies for the limit processes from both theorems. The LLN depends on the parameters of $\boldsymbol{X}$, which are specified differently by Theorems 2 and 3. The result and its proof generalize the result and proof of Corollary 7 in [9].

For the statement of the result, let $\boldsymbol{N}(m, \Sigma)$ denote a normal random vector in $\mathbb{R}^k$ with mean vector $m$ and covariance matrix $\Sigma$, so that

$$P\left(\boldsymbol{N}(m, \Sigma) \le y\right) = \int_{x \le y} n(x; m, \Sigma)\, dx$$

for $y$ in $\mathbb{R}^k$, where $n(x;m,\Sigma)$ is the multivariate normal density. Let $\boldsymbol{N}^f(m,\Sigma)$ denote the random vector with cumulative distribution function

$$P\left(\boldsymbol{N}^f(m,\Sigma)\leq y\right)=\int_{f(x)\leq y} n(x;m,\Sigma)\,dx,$$

where $f:\mathbb{R}^k\to\mathbb{R}^k$.

THEOREM 4 **(Non-ergodic LLN for the limit processes)**. *If $\boldsymbol{X}$ in $D^k$ is a $\psi$-GMP with parameter matrices $(A,B)$ and drift vector $\omega$, $(\boldsymbol{Z},\boldsymbol{L})\equiv(\phi(\boldsymbol{Z}(0)+\boldsymbol{X}),\psi(\boldsymbol{Z}(0)+\boldsymbol{X}))$, where $(\phi,\psi)\equiv(\phi,\psi)_Q:\ D^k\to D^{2k}$ is the reflection map, and $\boldsymbol{Z}(0)$ is independent of $\boldsymbol{X}$, then*

$$n^{-1}(\boldsymbol{X}(n),\boldsymbol{Z}(n),\boldsymbol{L}(n))\Rightarrow\left(N(\omega,-B),N^{f_Z}(\omega,-B),N^{f_L}(\omega,-B)\right)\quad in\quad\mathbb{R}^{3k}$$

*as $n\to\infty$, where $f_Z(x)\equiv x+(I-Q)^{-1}(-x)\vee 0$ and $f_L(x)\equiv(I-Q)^{-1}(-x)\vee 0$.*

The proof of Theorem 4 will rely on the following FWLLN for the net input process.

LEMMA 4 **(FWLLN for a tied-down $\psi$-GMP with drift)**. *If $\boldsymbol{X}$ in $D^k$ is a $\psi$-GMP with parameter matrices $(A,B)$ and drift vector $\omega$,*

$$\bar{\boldsymbol{X}}_{n,x}(s)\equiv n^{-1}(\boldsymbol{X}(ns)\,|\,\boldsymbol{X}(n)=nx)\quad for\quad 0\leq s\leq 1\quad and\quad n\geq 1,$$

*and*

$$\left(\bar{\boldsymbol{Z}}_{n,x},\bar{\boldsymbol{L}}_{n,x}\right)\equiv\left(\phi\left(n^{-1}\boldsymbol{Z}(0)+\bar{\boldsymbol{X}}_{n,x}\right),\psi\left(n^{-1}\boldsymbol{Z}(0)+\bar{\boldsymbol{X}}_{n,x}\right)\right),$$

*where $(\phi,\psi)\equiv(\phi,\psi)_Q:\ D^k\to D^{2k}$ is the reflection map, and $\boldsymbol{Z}(0)$ is independent of $\boldsymbol{X}$, then*

$$\left(\bar{\boldsymbol{X}}_{n,x}.\bar{\boldsymbol{Z}}_{n,x},\bar{\boldsymbol{L}}_{n,x}\right)\Rightarrow\left(xe,xe+(I-Q)^{-1}(-xe)\vee 0,(I-Q)^{-1}(-xe)\vee 0\right)$$

*in $(D^{3k}[0,1],WM_1)$ as $n\to\infty$.*

*Proof.* Because $\bar{\boldsymbol{X}}_{n,x}$ is continuous, it suffices to prove joint convergence in the space $C^{3k}[0,1]$ of continuous functions on the interval $[0,1]$ with the uniform topology. Because $\bar{\boldsymbol{X}}_{n,x}(s)\equiv n^{-1}\boldsymbol{X}_{n,x}(ns)$, where $\boldsymbol{X}_{n,x}(t)\equiv(\boldsymbol{X}(t)\,|\,\boldsymbol{X}(n)=nx)$ for $0\leq t\leq n$, Proposition 4 implies that $\bar{X}^i_{n,x}(s)=x_is+\bar{U}^i_n(s)$ for $0\leq s\leq 1$, where $\bar{U}^i_n$ is a zero-mean Gaussian Markov process (a Brownian

bridge) in $C[0,1]$ with $\bar{U}_n^i(0)=0$ and $\mathbb{C}ov\left[\bar{U}_n^i(s_1),\bar{U}_n^i(s_2)\right]=n^{-1}A_{ii}s_1(1-s_2)\to 0$ for $0\leq s_1\leq s_2\leq 1$ as $n\to\infty$. It follows that the covariance matrix for $\left(\bar{U}_n^i(s_1),\bar{U}_n^i(s_2),\ldots,\bar{U}_n^i(s_p)\right)$ converges to zero as $n\to\infty$ for all positive integers $p$ and all time points $0\leq s_1\leq s_2\leq\ldots\leq s_p\leq 1$. By Levy's convergence theorem, $U_n^i\to_{f.d.}0$, where $\to_{f.d.}$ denotes convergence of finite dimensional distributions. From the covariance function for $\bar{U}_n^i$, it also follows that $\bar{U}_n^i=_d n^{-1/2}\bar{U}_1^i$, where $=_d$ denotes equality of distributions. If $\nu\left(\bar{U}_n^i,\delta\right)$ is the modulus of continuity defined in (6.2) on page 388 of [19], then $\nu\left(\bar{U}_1^i,\delta\right)\to 0$ as $\delta\to 0$ because $\bar{U}_1^i$ is continuous, and $\nu\left(\bar{U}_n^i,\delta\right)$ for $\delta>0$ is a decreasing function of $n$. It follows from Theorem 11.6.4 of [19] that $\bar{U}_n^i\Rightarrow 0e$ in $C[0,1]$, so that $\bar{\boldsymbol{X}}_{n,x}^i\Rightarrow x_ie$ in $C[0,1]$. We conclude that $\bar{\boldsymbol{X}}_{n,x}\Rightarrow xe$ in $C^k[0,1]$ using Theorem 11.6.4 of [19]. We then easily deduce the other limits using the continuous mapping theorem and properties of the reflection map. ∎

As a corollary of Lemma 4, we obtain a WLLN for a $\psi-GMP$ with drift and its multivariate reflection.

COROLLARY 1 **(WLLN for the conditioned process)**. *Under the assumptions of Theorem 4, if*

$$\boldsymbol{X}_{n,x}(t)\equiv(\boldsymbol{X}(t)\,|\boldsymbol{X}(n)=nx)$$

$$\boldsymbol{Z}_{n,x}(t)\equiv(\boldsymbol{Z}(t)\,|\boldsymbol{X}(n)=nx)\quad \textit{and}\quad \boldsymbol{L}_{n,x}(t)\equiv(\boldsymbol{L}(t)\,|\boldsymbol{X}(n)=nx)$$

*for* $0\leq t\leq n$ *and* $n\geq 1$, *then*

$$\left(\frac{\boldsymbol{X}_{n,x}(n)}{n},\frac{\boldsymbol{Z}_{n,x}(n)}{n},\frac{\boldsymbol{L}_{n,x}(n)}{n}\right)\Rightarrow\left(x,x+(I-Q)^{-1}(-x)\vee 0,(I-Q)^{-1}(-x)\vee 0\right)$$

*in* $R^{3k}$ *as* $n\to\infty$.

*Proof.* By the definitions of $\boldsymbol{X}_{n,x}$ above and $\bar{\boldsymbol{X}}_{n,x}$ from Lemma 4,

$$\boldsymbol{X}_{n,x}=n\bar{\boldsymbol{X}}_{n,x}\circ\left(n^{-1}e\right)\quad\text{on}\quad[0,n].$$

Using the rescaling properties of the reflection map from Theorem 14.2.6 of [19],

$$\begin{aligned}
\left(n^{-1}\boldsymbol{X}_{n,x}, n^{-1}\boldsymbol{Z}_{n,x}, n^{-1}\boldsymbol{L}_{n,x}\right) &= \left(n^{-1}\boldsymbol{X}_{n,x}, n^{-1}\phi\left(\boldsymbol{Z}\left(0\right)+\boldsymbol{X}_{n,x}\right), n^{-1}\psi\left(\boldsymbol{Z}\left(0\right)+\boldsymbol{X}_{n,x}\right)\right) \\
&= \left(\bar{\boldsymbol{X}}_{n,x}\circ\frac{e}{n}, \phi\left(n^{-1}\boldsymbol{Z}\left(0\right)+\bar{\boldsymbol{X}}_{n,x}\circ\frac{e}{n}\right), \psi\left(n^{-1}\boldsymbol{Z}\left(0\right)+\bar{\boldsymbol{X}}_{n,x}\circ\frac{e}{n}\right)\right) \\
&= \left(\bar{\boldsymbol{X}}_{n,x}\circ\frac{e}{n}, \phi\left(n^{-1}\boldsymbol{Z}\left(0\right)+\bar{\boldsymbol{X}}_{n,x}\right)\circ\frac{e}{n}, \psi\left(n^{-1}\boldsymbol{Z}\left(0\right)+\bar{\boldsymbol{X}}_{n,x}\right)\circ\frac{e}{n}\right) \\
&= \left(\bar{\boldsymbol{X}}_{n,x}\circ\frac{e}{n}, \bar{\boldsymbol{Z}}_{n,x}\circ\frac{e}{n}, \bar{\boldsymbol{L}}_{n,x}\circ\frac{e}{n}\right) \quad \text{w.p.1} \quad \text{on} \quad [0,n].
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left(\frac{\boldsymbol{X}_{n,x}\left(n\right)}{n}, \frac{\boldsymbol{Z}_{n,x}\left(n\right)}{n}, \frac{\boldsymbol{L}_{n,x}\left(n\right)}{n}\right) &= \left(\bar{\boldsymbol{X}}_{n,x}\circ\frac{e}{n}\left(n\right), \bar{\boldsymbol{Z}}_{n,x}\circ\frac{e}{n}\left(n\right), \bar{\boldsymbol{L}}_{n,x}\circ\frac{e}{n}\left(n\right)\right) \\
&= \left(\bar{\boldsymbol{X}}_{n,x}\left(1\right), \bar{\boldsymbol{Z}}_{n,x}\left(1\right), \bar{\boldsymbol{L}}_{n,x}\left(1\right)\right) \quad \text{w.p.1}
\end{aligned}$$

in $R^{3k}$ as $n\to\infty$, and the claimed result follows from Lemma 4. ∎

We now apply Corollary 1 to provide an elementary proof for Theorem 4.

*Proof.* (Theorem 4) Using the definition of a $\psi$-GMP $\boldsymbol{X}$ with parameter matrices $(A,B)$ and drift vector $\omega$,

$$n^{-1}\boldsymbol{X}\left(n\right) = \boldsymbol{N}\left(\omega, \frac{n\left(A-Bn\right)}{n^2}\right) \Rightarrow \boldsymbol{N}\left(\omega,-B\right) \quad \text{in} \quad \mathbb{R}^k \quad \text{as} \quad n\to\infty,$$

as follows from Levy's convergence theorem since the covariance matrix for $n^{-1}\boldsymbol{X}\left(n\right)$ converge to $-B$ as $n\to\infty$. Likewise,

$$\begin{aligned}
\mathbb{P}\left(n^{-1}\boldsymbol{Z}\left(n\right)\le z\right) &= \int \mathbb{P}\left(n^{-1}\boldsymbol{Z}\left(n\right)\le z | n^{-1}\boldsymbol{X}\left(n\right)=x\right)\mathbb{P}\left(n^{-1}\boldsymbol{X}\left(n\right)\in dx\right) \\
&= \int \mathbb{P}\left(n^{-1}\boldsymbol{Z}\left(n\right)\le z | n^{-1}\boldsymbol{X}\left(n\right)=x\right) n\left(x;\omega,\frac{n\left(A-Bn\right)}{n^2}\right)dx \\
&\to \int 1_{x+(I-Q)^{-1}(-x)\vee 0\le z} n\left(x;\omega,-B\right)dx \quad \text{as} \quad n\to\infty
\end{aligned}$$

by Corollary 1 and the bounded convergence theorem. The limit for $n^{-1}\boldsymbol{L}\left(n\right)$ is obtained in the same way, and the joint limit follows from the continuous mapping theorem. ∎

Remark 3. Our previous simulation results in Figure 1 of [9] illustrate that the time average of a $\psi-GPP$ rapidly approaches a limiting value for each sample path. We have verified that

the empirical distribution of those values (for an ensemble of independent sample paths) is well approximated by the LLN limit. Figure 2 of [9] suggests that the same is true for the time average of the queue-length limit process for a single queue with $\psi - GPP$ arrivals in heavy traffic. We expect similar results for queueing networks with $\psi$-GPSP arrivals. The LLN limit provides a first-order approximation for the transient queue-length distribution with details lost because of its normalization of the queue-length by time. The LLN limit in Theorem 4 has an atom at zero that describes the probability that the queue-length limit process remains finite as time increases. The rest of the LLN distribution then describes the rate at which the queue-length limit process explodes when it does not remain finite. The probability of an explosion provided by the LLN distribution is fundamental to the understanding of queueing networks with path-dependent arrivals.

Our next result will show how conditioning on an intermediate state induces a changes of parameters for the LLN limit. For real $s, t \geq 0$, let

$$\hat{\boldsymbol{X}}^s(t) \equiv (\boldsymbol{X}(t+s) - \boldsymbol{X}(s) \,|\boldsymbol{X}(s), \boldsymbol{Z}(s)),$$

$$\hat{\boldsymbol{Z}}^s(t) \equiv (\boldsymbol{Z}(t+s) \,|\boldsymbol{X}(s), \boldsymbol{Z}(s)) \quad \text{and} \quad \hat{\boldsymbol{L}}^s(t) \equiv (\boldsymbol{L}(t+s) - \boldsymbol{L}(s) \,|\boldsymbol{X}(s), \boldsymbol{Z}(s)).$$

Corollary 2 **(conditional LLN)**. *If* $(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{L}) \equiv (\boldsymbol{X},\ \psi(\boldsymbol{Z}(0) + \boldsymbol{X}), \phi(\boldsymbol{Z}(0) + \boldsymbol{X}))$ *in* $D^{3k}$ *where* $\boldsymbol{X}$ *is a* $\psi$*-GMP with parameter matrices* $(A, B)$ *and drift vector* $\omega$, $(\psi, \phi) \equiv (\psi, \phi)_Q : D^k \to D^{2k}$ *is the reflection map, and* $\boldsymbol{Z}(0)$ *in* $\mathbb{R}^k$ *is independent of* $\boldsymbol{X}$*, then* $\hat{\boldsymbol{X}}^s$ *is a* $\psi$*-GMP with parameters* $(A,\ B_s)$ *and drift* $\omega_s$*, and*

$$\left( \frac{\hat{\boldsymbol{X}}^s(n)}{n}, \frac{\hat{\boldsymbol{Z}}^s(n)}{n}, \frac{\hat{\boldsymbol{L}}^s(n)}{n}, \right) \Rightarrow \left( \boldsymbol{N}(\omega_s, -B_s), \boldsymbol{N}^{f_Z}(\omega_s, -B_s), \boldsymbol{N}^{f_L}(\omega_s, -B_s) \right)$$

*in* $\mathbb{R}^{3k}$ *as* $n \to \infty$*, where* $B_s = B(A - Bs)^{-1} A$, $\omega_s = \omega - B(A - Bs)^{-1} s(s^{-1}\boldsymbol{X}(s) - \omega)$, *and* $f_Z$ *and* $f_L$ *are defined as in Theorem 4.*

*Proof.* Defining $\boldsymbol{X}^n$ as in Proposition 5,

$$\begin{aligned} \left( \hat{\boldsymbol{X}}^s, \hat{\boldsymbol{Z}}^s, \hat{\boldsymbol{L}}^s \right) &= \left( \hat{\boldsymbol{X}}^s, \phi\left( \boldsymbol{Z}(0) + \hat{\boldsymbol{X}}^s \right), \psi\left( \boldsymbol{Z}(0) + \hat{\boldsymbol{X}}^s \right) \right) \\ &=_d (\boldsymbol{X}^s, \phi(\boldsymbol{Z}(0) + \boldsymbol{X}^s), \psi(\boldsymbol{Z}(0) + \boldsymbol{X}^s)), \end{aligned} \tag{24}$$

where the equality follows from the memoryless property of the reflection map from Theorem 1 of [14], and the equality in distribution follows from the Markov property of $\boldsymbol{X}$. The result then follows by applying Theorem 4 with the modified parameters from Proposition 5. ∎

The next corollary shows that a multivariate $\psi$-GMP with drift, conditioned on its time average at time $s$, behaves increasingly like a multivariate Brownian motion with drift equal to that time average as $s$ becomes large. In the statement of the result, we say that $f(s) = o(g(s))$ for a scaler-valued function $g$ and vector- or matrix-valued function $f$ if $f(s)$ is asymptotically equal to the zero vector or matrix for large $s$ after dividing by $g(s)$.

COROLLARY 3 **(Asymptotic behavior of conditional process)**. *If $\boldsymbol{X}$ is a $\psi$-GMP in $D^k$ with parameter matrices $(A, B)$ and drift vector $\omega$, then*

$$\left(\boldsymbol{X}(t+s) - \boldsymbol{X}(s) | s^{-1}\boldsymbol{X}(s) = x\right) = \omega_{s,x} t + \boldsymbol{G}^{s,x}(t) \text{ a.e.} \quad \text{for} \quad s, t \geq 0, \tag{25}$$

*and*

$$\left(n^{-1}\boldsymbol{X}(n) \mid s^{-1}\boldsymbol{X}(s) = x\right) \Rightarrow \boldsymbol{N}(\omega_{s,x}, -B_{s,x}) \quad \text{as} \quad n \to \infty, \tag{26}$$

*where $\boldsymbol{G}^{s,x}$ is a $\psi$-GMP in $D^k$ with parameters $(A, B_{s,x})$, $\omega_{s,x} = x + s^{-1}AB^{-1}(x - \omega) + o(s^{-1})$ and $-B_{s,x} = s^{-1}A + o(s^{-1})$ as $s \to \infty$.*

*Proof.* If $s$ is greater than the spectral radius of $AB^{-1}$, then

$$\begin{aligned}(A - Bs)^{-1} = -s^{-1}B^{-1}\left(I - s^{-1}AB^{-1}\right)^{-1} &= -s^{-1}B^{-1}\sum_{j=0}^{\infty}\left(s^{-1}AB^{-1}\right)^j \\ &= -s^{-1}B^{-1}\left(I + s^{-1}AB^{-1} + o\left(s^{-1}\right)\right) \text{ as } s \to \infty. \end{aligned} \tag{27}$$

The result in (25) follows from Proposition 5 and (27). Since $\boldsymbol{X}^n =_d \hat{\boldsymbol{X}}^n$ as in (24), the result in (26) follows from Corollary 2 and (27). ∎

When the time $s$ of observation is large, Corollary 3 implies that the subsequent mean growth rate for the conditioned process is approximately equal to the historic average growth rate, and the parameter matrix $B_{s,x}$ that distinguishes the conditioned process from a Brownian motion is approximately equal to zero, both with error of order $1/s$.

## 6. Concluding discussion

We conclude by discussing extensions to the results from this paper and open questions.

### 6.1. Domain of attraction for a $\psi$-GMP

A goal for future work is to determine the class of stochastic arrival processes that lead to the FCLTs and HTLTs with the $\psi$-GMP limit process. The results of this paper build on the result for the convergence of a sequence of one-dimensional $\psi$-GPPs to a $\psi$-GMP from Theorem 4 of [9]. That result in turn exploits Hahn's FCLT for a sum of i.i.d. processes, reviewed in Section 7.2.1 of [19]. We now observe that there are other processes in addition to $\psi$-GPP's that will lead to a $\psi$-GMP limit when we apply Hahn's theorem. To describe such processes, let $\{N^i : i \geq 1\} = \{\{N^i(t) : t \geq 0\} : i \geq 1\}$ be a sequence of i.i.d. stationary point processes in $D$, and let

$$G_n(t) \equiv n^{-1/2} \sum_{i=1}^{n} \left( N^i(t) - \mathbb{E}\left[N^i(t)\right] \right) \quad \text{for} \quad t \geq 0.$$

THEOREM 5 **(Sufficient conditions for convergence to a $\psi - GMP$)**. *If a sequence of i.i.d. stationary point processes $\{N^i : i \geq 1\}$ in $D$ has the properties (i) $\mathbb{E}\left[N^i(t)^j\right] < \infty$ for $t \geq 0$ and $j = 1,2,3,4$, (ii) $\mathbb{V}ar\left[N^i(1)\right] > \mathbb{E}\left[N^i(1)\right]$ for $t \geq 0$, and (iii) the process $\{Y(s) \equiv (N^i(s) | N^i(t)) : 0 \leq s \leq t\}$, defined for any $t > 0$, has the same distribution as the empirical process for $N^i(t)$ i.i.d. uniformly distributed random variables on [0,t], then $G_n \Rightarrow G$ in $(D, WM_1)$ as $n \to \infty$, where $G$ is the $\psi$-GMP with parameters $(A = \mathbb{E}\left[N^i(1)\right], B = -\left(\mathbb{V}ar\left[N^i(1)\right] - \mathbb{E}\left[N^i(1)\right]\right))$.*

*Proof.* Using well-known properties of empirical processes, $\mathbb{E}[Y(s)] = st^{-1}N^i(t)$ and $\mathbb{V}ar[Y(s)] = N^i(t)\, st^{-1}(1 - st^{-1})$, so that $\mathbb{E}\left[Y(s)^2\right] = N^i(t)\, st^{-1}(1 - st^{-1}) + s^2t^{-2}N^i(t)^2$ for $0 \leq s \leq t$. Therefore,

$$\begin{aligned} \mathbb{V}ar\left[N^i(s)\right] &= \mathbb{E}\left[\mathbb{E}\left[Y(s)^2\right]\right] - \mathbb{E}\left[\mathbb{E}\left[Y(s)\right]\right]^2 \\ &= \mathbb{E}\left[N^i(t)\right] st^{-1}\left(1 - st^{-1}\right) + s^2t^{-2}\mathbb{E}\left[N^i(t)^2\right] - s^2t^{-2}\mathbb{E}\left[N^i(t)\right]^2 \end{aligned} \tag{28}$$

for $0 \leq s \leq t$. Because $N^i$ is stationary, $\mathbb{E}\left[N^i(t)\right] = t\mathbb{E}\left[N^i(1)\right]$. Substituting $t = 1$ and $s = u$ into (28) then shows that

$$\mathbb{V}ar\left[N^i(u)\right] = \mathbb{E}\left[N^i(1)\right] u + \left(\mathbb{V}ar\left[N^i(1)\right] - \mathbb{E}\left[N^i(1)\right]\right) u^2 \tag{29}$$

for $0 \leq u \leq 1$. Substituting $s = 1$ and $t = u$ into (28) then shows that (29) also holds for $u > 1$. By assumption (ii), $\mathbb{V}ar\left[N^i(1)\right] > \mathbb{E}\left[N^i(1)\right]$, so that (29) is defined (i.e., is positive) for all $u \geq 0$. Therefore (29) holds for all $u \geq 0$. By the same logic used to derive equation (4) of Theorem 1 in [9], the assumed stationarity and the variance function in (29) imply that

$$\mathbb{C}ov\left(N^i(s), N^i(t)\right) = \mathbb{E}\left[N^i(1)\right] s + \left(\mathbb{V}ar\left[N^i(1)\right] - \mathbb{E}\left[N^i(1)\right]\right) st = \mathbb{C}ov\left(N(s), N(t)\right) \text{ for } 0 \leq s \leq t.$$

The remainder of the proof, which verifies that the conditions of Hahn's theorem are met, is then essentially the same as the proof of Theorem 4 in [9], except that the constant $c$ in (31) there is expressed in terms of the first four moments of $N^i(1)$ instead of their particular expressions for a $\psi$-GPP from (33) there. ∎

Remark 4. If we relax the requirement that $B$ is negative for the $\psi$-GMP limit and allow $B$ to be any real scaler, then we can eliminate assumption (ii). Theorem 5 would then apply when $\mathbb{V}ar\left[N^i(1)\right] = \mathbb{E}\left[N^i(1)\right]$ with no other changes. It would also apply when $\mathbb{V}ar\left[N^i(1)\right] < \mathbb{E}\left[N^i(1)\right]$ with the change that the time domain is limited to the finite interval on which the variance function in (29) remains positive.

We now give an example showing that Theorem 5 applies to more processes than $\psi$-GPP's.

Example 3. A different process for which we can apply Hahn's theorem to get convergence to a $\psi$-GMP is the Poisson Generalized Gamma Process (PGGP) discussed in [4]. An orderly point process $\{N(t) : t \geq 0\}$ is a Poisson Generalized Gamma Process (PGGP) with parameters $(\lambda(t) > 0, \nu \geq 0, k > 0, \alpha > 0, l > 0)$ if it has the stochastic intensity function

$$\lambda^*(t|\mathcal{H}_t) = \frac{1}{\alpha + \Lambda(t)} \frac{\Gamma_\nu\left(k + N(t-) + 1, (\alpha + \Lambda(t))\, l\right)}{\Gamma_\nu\left(k + N(t-), (\alpha + \Lambda(t))\, l\right)}\ \lambda(t),$$

where $\lambda(t)$ is integrable, $\Lambda(t) = \int_0^t \lambda(s)\, ds$, and $\Gamma_\nu(k, \beta)$ is the generalized gamma function defined for $k, \beta > 0$ by

$$\Gamma_\nu(k, \beta) = \int_0^\infty \frac{y^{k-1} exp(-u)}{(y + \beta)^\nu} dy.$$

By Proposition 3.1 of [4], a GPP is a special case of a PGGP where $\nu = 0$. PGGPs are describe in [4] as a more flexible model than GPPs for point processes with dependent increments and

overdispersion relative to a Poisson process. When $\lambda(t) = \lambda t$ for $t \geq 0$, where $\lambda$ is a positive real constant, condition (iii) of Theorem 5 is satisfied by Theorem 3.3 of [4]; and the distribution of the PGGP's increment over an interval depends only on the interval's length and not on its position by Theorem 2.1 (ii) of [4]. The PGGP is then stationary, as follows from the same logic as used by Theorem 1 of [9] to show when a GPP is stationary. A stationary PGGP satisfies the other conditions of Theorem 5 as well: condition (i) is satisfied by Theorem 2.2 (ii) of [4], and condition (ii) is satisfied by Proposition 2.2 of [4] (excluding the special case where the PGGP is a Poisson process with $\mathbb{V}ar[N_i(t)] = \mathbb{E}[N_i(t)]$). Using Theorem 2.2 (iii) of [4], we verify that a stationary PGGP has the mean and variance functions deduced above for a process satisfying the conditions of Theorem 5.

Theorem 5 describes the limit for univariate superpositions of i.i.d. univariate stationary point processes. A superposition of $n$ i.i.d. univariate $\psi$-GPPs has the same distribution as a single $\psi$-GPP with the same combined rate. That property enabled Theorem 4 of [9] to be restated in the form from Proposition 4 of [10], which was applied in Lemma 1 here. That property has not been established for all processes satisfying the conditions of Theorem 5. In particular it has not been established for PGGPs in the general case. An open question is the domain of attraction for a $\psi$-GMP when the limit is obtained by scaling the rate parameter of a process rather than by constructing superpositions.

It is straightforward to generalize Theorem 1 of this paper to apply for the processes in the domain of attraction described by Theorem 5, including PGGPs. We omit a formal statement of that result in the interest of brevity.

### 6.2. Transient distribution

For possible applications, the most obviously useful results in this paper are (i) the Gaussian approximation for the transient distribution of the net input process (because it is a $\psi$-GMP) from Theorem 1 and Lemma 2, and (ii) the non-ergodic LLN for the queueing processes and the net input process in Theorem 4 (which can serve as a good approximation for the conditioned process

by Corollaries 1 and 2). However, it remains to find useful expressions for the exact transient distribution for a reflected multivariate psi-GMP with drift and general reflection matrix. Corollary 6 of [9] provided an exact expression for the transient distribution of a univariate reflected $\psi$-GMP with drift generalizing the well-known transient distribution for reflected univariate Brownian motion with drift (e.g., from Section 6 of [12]). The special case of the transient distribution of reflected multivariate Brown motion with drift is itself still an open question.

A promising approximation method for open queueing networks is robust queueing as in [20, 21]. New approximations for the steady-state distribution of an open queueing network are developed in [21] and older ones by [6, 13] are reviewed, but those are not applicable for path-dependent arrivals. Approximations for the transient distribution of the queue length process are developed in [20], but so far, just as for the transient distribution of $\psi$-GMP in [9], those are limited to a single queue. Extension to networks remains an important open problem.

## References


[1] Cha, J. H. (2014) Characterization of the generalized Polya process and its applications. *Advances in Applied Probability* **46** (4) 1148-1171.

[2] Cha, J. H. and Badia, F. G. (2020) On a multivariate generalized Polya process without regularity property. *Probability in the Engineering and Informational Sciences*, **34** (4) 484-506.

[3] Cha, J. H. and Finkelstein, M. (2018) *Point Processes for Reliability Analysis*, Springer, New York.

[4] Cha, J. H. and Mercier, S. (2021) Poisson generalized gamma process and its properties. *Stochastics* **93** (8) 1123–1140.

[5] Daley, D.J. and Vere-Jones, D. (2003) *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*, Springer, New York.

[6] Dai, J., Nguyen, V., and Reiman, M. I. (1994) Sequential bottleneck decomposition: An approximation method for generalized Jackson networks. *Operations Research* **42** (1) 119-136.

[7] Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. I, 3rd edn. John Wiley, New York.

[8] Fendick, K. W. (2020) Brownian motion minus the independent increments: representation and queueing application. *Probability in the Engineering and Informational Sciences* **36** (1) 144–168.

[9] Fendick, K.W. and Whitt, W. (2021) Queues with path-dependent arrival processes. *Journal of Applied Probability* **58** 484–504.

[10] Fendick, K.W. and Whitt, W. (2022) Heavy traffic limits for queues with non-stationary path-dependent arrival processes. *Queueing Syystems* **101** 113–135.

[11] Glynn, P.W. (1990) Diffusion approximations. *Handbooks in Operations Research and Management Science, Volume 2*, Elsevier, New York, 145–198.

[12] Harrison, J. M. (1985). *Brownian Motion and Stochastic Flow Systems*, Wiley, New York.

[13] Harrison, J.M. and Nguyen, V. (1990) The QNET method for two-moment analysis of open queueing networks. *Queueing Systems* **101** (1) 1–32.

[14] Harrison, J.M. and Reiman, M.I. (1981) Reflected Brownian motion on an orthant. *The Annals of Probability* **9** (2), 302–308.

[15] Konno, T. H (2010) On the exact solution of a generalized Polya process. *Advances in Mathematical Physics* **2010** Article ID 504267.

[16] Mandjes, M. (2007). *Large Deviations for Gaussian Queues: Modelling Communication Networks*, John Wiley, New York.

[17] Ross, S. (1996). *Stochastic Processes, Second Edition*, John Wiley, New York.

[18] Simeu-Abazi, Z., Di Mascolo, M. and Gascard, E. (2012) Performance evaluation of centralized maintenance workshop by using Queueing Networks. *IFAC Proceedings Volumes* bf 45 (31) 175–180.

[19] Whitt, W. (2002). *Stochastic Process Limits*, Springer, New York.

[20] Whitt, W. and You, W. (2019) Time-varying robust queueing. *Operations Research* **67** (6) 1766–1782.

[21] Whitt, W. and You, W. (2022) A robust queueing network analyzer based on Indices of Dispersion. *Naval Research Logistics* **69** (1) 36–56.