ON THE PERFORMANCE MEASURE ANALYSIS OF THE RESIDUAL SERVICE TIME IN THE M/G/1 QUEUEING MODEL

Supported by tetfund

S. O. Agboola[1]* and G. B. Karokatose[2]

[1]Department of Mathematical and Computing Sciences, KolaDaisi University, Ibadan, Nigeria
[2]Department of Mathematics, Obafemi Awolowo University, Ile – Ife, Nigeria
*Corresponding author: larrysoa7519@yahoo.com

Received: March 14, 2019 **Accepted:** August 10, 2019

Abstract: This present study, the residual service time on the M/G/1 queueing system where the arrival process is Poisson with rate λ and service times of customers are independent and identically distributed and obey an unspecified arbitrary or general distribution function is analysed. Our analysis is based on the fact that, an arrival is more likely to occur during a large service time than a small service interval since the service is a random variable having a general distribution. Using the imbedded Markov chain technique, the stochastic transition probability matrix f_{ij} and average residual service time in [0, t] are obtained by considered the area under the curve $R(t)$ divided by t. Finally, we obtained the probability distribution function of the residual service time conditioned on the server being busy and expected residual service time. The numerical illustration is considered to show its applications in solving real life problem on $M/D/1$queue for which $\lambda = \frac{1}{2}\ and\ \ \mu = 1$. The elements of the transition probability matrix are obtained as $\alpha_1 = 0.303265$; $\alpha_2 = 0.075816$;$\alpha_3 = 0.075816$; $\alpha_4 = 0.075816$; $\alpha_5 = 0.075816$;$\alpha_6 = 0.075816$; $\alpha_7 = 0.075816$. Also, by the use of the recursive procedure and setting$\rho_0 = 0.5$, We obtain the following$P_1 = 0.32436$: $\sum_{i=0}^{1} p_i = 0.824361$, $P_2 = 0.1226$: $\sum_{i=0}^{2} p_i = 0.94696$, $P_3 = 0.037788$: $\sum_{i=0}^{3} p_i = 0.98475$, $P_4 = 0.01091$: $\sum_{i=0}^{4} p_i = 0.995658$, $P_5 = 0.003107$: $\sum_{i=0}^{5} p_i = 0.998764$, $P_6 = 0.000884$: $\sum_{i=0}^{6} p_i = 0.999648$.

Keywords: General distribution, performance measure, Pollaczek-Khintchine equation residual service

Introduction

The M/G/1 queue is a single-server queue of Poisson arrival, the service times of customers are independent and identically distributed and obey an unspecified arbitrary or general distribution function. In particular, the remaining service time may no longer be independent of the service already received. When a customer arrives at an M/G/1 queue and finds at least one customer already present, at that arrival instant a customer is in the process of being served. In this study our concern is with the time that remains until the completion of that service, the so called residual (service) time. In the more general context of an arbitrary stochastic process, the terms residual lifetime and forward recurrence time are also employed. The time that has elapsed from the moment service began until the current time is called the backward recurrence time. If the system is empty; then $\Re = 0$. The mean residual service time is then obtained from the Pollaczek-Khintchinemean value formulae in the context of a first-come, first-served scheduling policy. Stochastic application of queueing theory which involves various probability distributions and its application is discussed in Law and Kelton (2000) and new Convergence Results on Functional Techniques for the Numerical Solution of M/G/1 Type Markov Chains is established. Agboola (2007) studied a single server queue where the inter-arrival time is Markovian time and service time is general. This model generalizes the well known M/G/1 queue. The waiting time process is directly analysed by solving the lindley's equation using transform method.

The Laplace Stieltjes transforms (LST) of the steady state waiting time and queue length distribution are both derived, and used to obtain recursive equations for the calculation of moments. k - server queue is Discussed in Agboola (2010) where the inter arrival time is Markovian and service time is Markovian, General and Erlang distributed. This model generalizes the M/M/K, M/G/K and M/Er/K queues. The departure distribution is directly analysed by using probability generating function to derive the service time, waiting time and sojourn time distribution under single server with general service time. The recursive equation is then used to obtain blocking probability for the Markov inter arrival with K – server under general service time. G/M/1 and G/M/K are discussed in William (2009) where the service process has exponential distribution with mean service time $\frac{1}{\mu}$. i.e.$B(x) = 1 - \exp(-\mu x)$, $x \geq 0$, while the arrival process is general with mean inter arrival time equal to $\frac{1}{\lambda}$. Customers arrive individually and their inter arrival times are independent and identically distributed. To represent this system by a Markovian, it is necessary to keep track of time that passes between arrivals, since the distribution of inter arrival times does not in general possess the memoryless property of the exponential. As was the case for the M/G/Q queue, a two – component state descriptor may be used; the first to indicate the number of customers present and the second to indicate the elapsed time since the previous arrival. In this way, the G/M/1 queue can be solved using the method of supplementary variables. It is also possible to define a Markov chain embedded within the G/M/1 queue. The embedded time instants are precisely the instants of customer arrivals, since the elapsed inter arrival time at these moments is known as zero. This allows us to form a transition probability matrix and to compute the distribution of customers as seen by an arriving customer.

Charan (2012) investigated the single server queueing system where in the arrival of the units follow a Poisson process with varying arrival rates in different states. The server may take a vacation of a fixed duration or may continue to be available in the system for next service. Probability generating function of the units present in the system and various performance indices such as expected number of units in the queue and in the system, average waiting time, etc. are obtained. Michiel (2017) analysed a non-classical discrete time queueing model which is based on complex contour integration to obtain the probability generating functions, the mean values and the tail probabilities of the customer delay and the system steady state with numerical example illustration. Charan (2019) considered a single server queueing system with batch arrival. The supplementary variable approach with probability generating function is applied to analyse the system to find the system performance quantity and numerical illustration is considered to obtain the system state probabilities and queueing reliability indices.

Materials and Methods

The M/G/1 queue is a single-server queue, illustrated graphically in Fig. 1.

Fig. 1: The M/G/1 queue

The arrival process is Poisson with rate λ, while the service times of customers are independent and identically distributed and obey an unspecified arbitrary or general distribution function. In particular, the remaining service time may no longer be independent of the service already received. The mean service rate is denoted by $\mu$and the service time distribution function is denoted as $B(x) = Prob[S \leq x]$, where S is the service time random variable with density function denoted $b(x)$, given by;

$$b(x)dx = Prob[x < s \leq x + dx].$$

The arrival process distribution function is;

$$A(t) = 1 - \exp(-\mu t), \quad t \geq 0.$$

The service or scheduling discipline is first come first serve (FCFS). In $M/M/1$ queue, where both the inter-arrival time and service time distributions are Poisson, all that required to summarised its entire past history is a specification of the number of customers present, $N(t)$, and, in this case, the stochastic process $N(t) \geq 0$ is a Markov process. In $M/G/1$ queue, the stochastic process $N(t) \geq 0$ is not a Markov process since, when $N(t) \geq 1$, a customer is in service and the time already spent by that customer in service must be taken into account as a result of the fact that the service process need not possess the memoryless property of the exponential distribution.

Let $C(x)$ be the conditional probability that the service time finishes before $(x + dx)$

Knowing that its duration is greater than x; therefore,

$$C(x) = Prob[S \leq x + dx | S > x] = \frac{Prob[S \leq x + dx]}{Prob[S > x]} = \frac{b(x)dx}{1 - B(x)}$$

Generally $C(x)$ depend on x, however $B(x)$ is an exponential distribution such that

$B(x) = 1 - \exp(-\mu x)$, $b(x) = \mu\exp(-\mu x)$

and

$$C(x) = \frac{\mu\exp(-\mu x)}{1 - 1 + \exp(-\mu x)} = \mu dx$$

which is independent of x. In this particular case;

if we start to observe the service in progress at an arbitrary time x, the probability thatthe service completes on the interval $(x, x + dx]$ does not depend on x, the duration ofservice already received by the customer. In other words, the probability that a transition will occur depends on its past history, and therefore the process is non-Markovian. This is implying that, if at some time t we want to summarise the complete relevant past history of an $M/G/1$ queue, we must specify both

(a) $N(t)$, the number of customers present at time t, and

(b) $S_0(t)$, the service time already spent by the customer in service at time t.

Since $N(t)$ is not Markovian $[N_0(t), S_0(t)]$ is a Markov process and it provides all the history necessary for describing the future evolution of an M/G/1 queue. The component$S_0(t)$ is called a supplimentary variable and the approach of using this state descriptionto solve the M/G/1 queues is called the method of supplementary variable, which involves working with two components, the first discrete and the second continuous but rather we shall seek an alternative solution based on a single discrete component known as the embedded Markov chain approach.

Results and Discussion

In the embedded Markov chain approach, we look for a Markov chain within (i.e., at certain instants within) the stochastic process $[N_0(t), S_0(t)]$ and solve for the distributionof customers at these times. One convinient set of time instants is the set of departure instants. These are the times at which a customer is observed to terminate service and leave the queueing system. The two dimensional state description $[N_0(t), S_0(t)]$ can be replace with one dimensional description N_k, where N_k denotes the number of customersleft behind by the kth departing customer.Let $A_k$be the random variable that denotes the number of customers that arrive during the service time of the kth customer. The relationship among the number of customersleft behind by the $(k+1)th$ customer in terms of the number left behind by its predecessor and

$$A_{k+1}\text{for}k > 0 \text{ is} N_{k+1} = N_k - 1 + A_{k+1},$$

Since there are N_k present in the system when the $(k+1)th$ customer enters service, andadditional A_{k+1} arrive while this customer is being served, and the number in the systemis reduced by 1 when this customer finally exists.

Similarly if $N_k = 0$, we have;

$$N_{k+1} = A_{k+1}$$

We may combine these into a single equation with the use of the function $\delta(N_k)$ defined as;

$$\delta(N_k) = \begin{cases} 1\,; & N_k > 0 \\ 0\,; & N_k = 0 \end{cases}$$

Since

$$N_{k+1} = N_k - 1 + A_{k+1}, N_k > 0$$
$$N_{k+1} = A_{k+1}, \quad N_k = 0.$$

We have

$$N_{k+1} = N_k - \delta(N_k) + A$$

The stochastic transition probability matrix for the embedded Markov chain,

$$N_k; \ k = 1,2,\cdots$$

The $(ij)th$ element of this matrix is given by $\delta(N_k) = Prob[N_{k+1} = j | N_k = i]$.

The single-step transition probability matrix is given by,

$$F = \begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \cdots \\ \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \cdots \\ 0 & \alpha_0 & \alpha_1 & \alpha_1 & \alpha_3 & \cdots \\ 0 & 0 & \alpha_0 & \alpha_1 & \alpha_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Where α_i is the probability that i arrivals occur during an arbitrary service.

Given that the random variable A_k are independent and identically distributed, and assuming that the kth departing customer leaves at least one customer behind

$$[N_k = i > 0],$$

we have

$$Prob[N_{k+1} = j | N_k = i] = f_{ij}(k) = \alpha_{j-i+1} \ \forall \ j = i-1, i, i+1, i+2, \cdots$$

In the case of the the kth departure leaving behind an empty system $[N_k = i = 0]$, we have

$$Prob[N_{k+1} = j | N_k = 0] = f_{ij}(k) = \alpha_j \ \forall \ j = 0, 1, 2, \cdots$$

Thus the transition probability $f_{ij}(k)$ do not depend on k, which means that the Markov Chain $[N_k; \ k = 1,2,3,\cdots]$ is homogeneous.

The $M/G/1$ residual service time

When a customer arrives at an M/G/1 queue and finds at least one customer already present, at that arrival instant, a customer is in the process of being served to obtain the time that remain until the completion of that service. In the M/M/1 queue, the mean residual service time given that we have an expression for the expected time an arriving customer must wait until its service begins, W_q, and another expression for the expected number of customers waiting in the queue, L_q, i.e.

$$W_q = \frac{\lambda E[S^2]}{2(1-\rho)}$$

and

$$L_q = \frac{\lambda^2 E[S^2]}{2(1-\rho)}$$

The mean residual time i

$$E[\Re] = W_q - \frac{1}{\mu} L_q$$

$$E[\Re] = \frac{\lambda E[S^2]}{2(1-\rho)} - \frac{\lambda^2 E[S^2]}{2\mu(1-\rho)}$$

$$E[\Re] = \frac{\lambda E[S^2]}{2(1-\rho)}(1-\rho) = \frac{\lambda E[S^2]}{2}$$

This expression provides the expected residual service time as seen by an arriving customer and this is also the expected residual service time as seen by a random observer. Since the arrival process is Poisson. The relationship between $E[\Re]$and W_q is given by;

$$E[\Re] = (1-\rho)W_q,$$

Where $(1-\rho)$ is the probability that the server is idle.

Considering the Fig. 2 below;

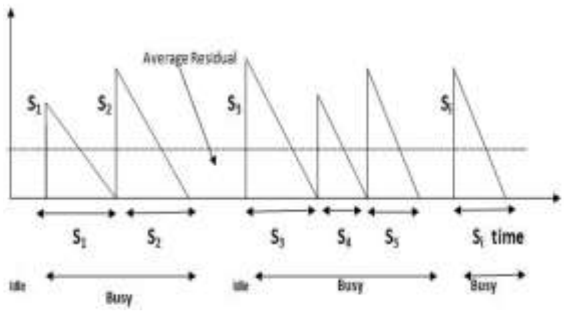

Fig. 2: Residual service time in M/G/1 queue

Immediately prior to the initiation of service for a customer, residual time, $\Re(t)$ equal zero. The moment the server begins to serve a customer, the residual service time mustbe equal to the total service requirement of that customer, i.e., S_i for customer i. Astime passes, the server reduces this service requirement at the rate of one unit per time, hence the slope of -1 from the moment service begins until the service requirement of the customer has been completely satisfied, at which the remaining service time is equal to zero. If at this instant another customer is waiting in the queue, the residual service time jumps an amount equal to the service required by that customer. Otherwise, the server becomes idle and remains so until a customer arrives to the queueing system. Let us assume that at time $t = 0$ the system is empty and let us choose a time t atwhich the system is once again empty. If $\lambda < \mu$, we are guaranteed that such times occur infinitely often. Let$M(t)$ be the number of customers served by time t. The averageresidual service time in $[0, t]$ is the area under the curve $\Re(t)$ divided by t and since thearea of each right angle triangle with base and height equal to $S_t$is$\frac{S_t^2}{2}$, we find;

$$E[\Re] = \frac{1}{t}\int_0^t \Re(t)dt = \frac{1}{t}\sum_{t=1}^{M(t)} \frac{S_t^2}{2}$$

$$= \frac{1}{2} \times \frac{M(t)}{t} \times \sum_{t=1}^{M(t)} \frac{S_t^2}{M(t)}$$

As $t \to \infty$

$$E[\Re] = \frac{1}{t}\int_0^t \Re(t)dt = \frac{1}{2} \times \lim_{t\to\infty} \frac{M(t)}{t} \times \lim_{t\to\infty} \sum_{t=1}^{M(t)} \frac{S_t^2}{M(t)}$$

$$E[\Re] = \frac{1}{2}\lambda E[S^2]$$

Where,

$\lambda = \lim_{t\to\infty} \frac{M(t)}{t}$ and $E[S^2] = \lim_{t\to\infty} \sum_{t=1}^{M(t)} \frac{S_t^2}{M(t)}$

λ is set equal to the mean output rate since at equilibrium, the arrival rate is equal to the departure rate.

Let x be a random variable that denotes the service time of the customer in service when an arrival occurs. The service received by some customers will be long while that received by other customers will be short. Therefore, it is apparent that an arrival is more likely to occur during a large service time than in a small service interval.

Therefore,

$f_x(X) = Prob[x \le X \le x + dx] = \alpha(x)b(x)dx$

Where the role of α is to ensure that this is a proper density function, i.e.

$$\int_0^\infty \alpha(x)b(x)dx = 1.$$

Since

$$E[S] = \int_0^\infty xb(x)dx,$$

It follows that

$$\alpha = \frac{1}{E[S]}$$

And

$$f_x(X)dx = \frac{xb(x)}{E[S]}$$

Since arrivals are Poisson, hence random, an arrival is uniformly distributed over the service interval $(0, x)$. This means the probability that the remaining service time is lessthan or equal to t, $0 \le t \le x$, given that the arrival occurs in a service period of length$x$, is equal to $\frac{t}{x}$. i.e.,

$$Prob[\Re_b \le t | X = x] = \frac{t}{x}$$

Therefore,

$$Prob[t \le \Re_b \le t + dt | X = x] = \frac{dt}{x}, \quad t \le x$$

Removing the condition, by integrating over all possible x, allows us to obtain probability distribution function for the residual service time conditioned on the server being busy.

We have,

$$Prob[t \le \Re_b \le t + dt] = f_{\Re_b}(t)dt$$

$$= \int_t^\infty \frac{dt}{x} dx = \int_t^\infty \frac{b(x)}{E[S]} dxdt$$

$$= \frac{1-B(t)}{E[S]} dt$$

And hence

$$f_{\Re_b}(t) = \frac{1-B(t)}{E[S]}$$

The mean residual service time is found from

$$E[\Re_b] = \int_t^\infty f_{\Re_b}(t)dt = \frac{1}{E[S]}\int_t^\infty t(1-B(t))dt$$

Taking

$u = 1 - B(t); \quad du = -B(t)dt. dv = tdt; v = \frac{t^2}{2}$

Using integration by part method

$$E[\Re_b] = \frac{1}{E[S]}[1 - B(t)]\left|\frac{t^2}{2}\right|_0^\infty + \int_t^\infty \frac{t^2}{2}B(t)dt$$

$$E[\Re_b] = \frac{1}{2E[S]}\int_t^\infty t^2 B(t)dt = \frac{\mu E[S^2]}{2}$$

The higher moment of residual time is given by

$$E[\mathfrak{R}_b{}^{k-1}] = \frac{\mu E[S^k]}{2}, \quad k = 2, 3, \cdots$$

Also, $E[\mathfrak{R}] = \rho E[\mathfrak{R}_b]$, where $\rho = 1 - \rho_0$ is the probability that the server is busy.

Given that the first and second moments of a random variable having an Erlang - r- distribution (r exponentially phases each with parameter $r\mu$) are $E[S] = \frac{1}{\mu}$and$E[S^2] = \frac{r(r+1)}{(r\mu)^2}$ respectively

The expected residual time in an $M/E_r/1$ queue is

$$E[R] = \frac{(1+\frac{1}{r})\big/\mu}{2/\mu} = \frac{(1+\frac{1}{r})}{2\mu}$$

As $r \to \infty$,

$$E[\mathfrak{R}_b] = \frac{1}{2\mu} = \frac{E[S]}{2}$$

which is the expected residual service time in an$M/G/1$queue when the process is deterministic.

Consider an $M/D/1$queue for which $\lambda = \frac{1}{2}$ and $\mu = 1$. The element of the transition probability matrix can be found from

$$\alpha_i = \frac{(0.5)^i}{i\,!} \exp(-0.5)$$

This yields the following results:

$\alpha_0 = 0.606531$; $\alpha_1 = 0.303265$; $\alpha_2 = 0.075816$;$\alpha_3 = 0.075816$; $\alpha_4 = 0.075816$; $\alpha_5 = 0.075816$;$\alpha_6 = 0.075816$; $\alpha_7 = 0.075816$

Using the recursive procedure and beginning with $\rho_0 = 0.5$,

We obtain the following (Agboola, 2011):

$P_1 = 0.32436$:$\sum_{i=0}^{1} p_i = 0.824361$

$P_2 = 0.1226$: $\sum_{i=0}^{2} p_i = 0.94696$

$P_3 = 0.037788$: $\sum_{i=0}^{3} p_i = 0.98475$

$P_4 = 0.01091$: $\sum_{i=0}^{4} p_i = 0.995658$

$P_5 = 0.003107$: $\sum_{i=0}^{5} p_i = 0.998764$

$P_6 = 0.000884$: $\sum_{i=0}^{6} p_i = 0.999648$

Conclusion

In this study, we obtained the probability distribution function of the residual service time conditioned on the server being busy and expected residual service time. We as well demonstrate that, as $r$tends to infinity, the first and second moments of random variable having Erlang - r distribution is the expected residual service time in an $M/G/1$ queuewhen the service process is deterministic, examples and results were given to show itsapplications in solving real life problem on $M/D/1$queue for which $\lambda = \frac{1}{2}$ and $\mu = 1$. The element of the transition probability matrix can be are obtained as $\alpha_1 = 0.303265$; $\alpha_2 = 0.075816$;$\alpha_3 = 0.075816$; $\alpha_4 = 0.075816$; $\alpha_5 = 0.075816$;$\alpha_6 = 0.075816$; $\alpha_7 = 0.075816$ and by using the recursive procedure and beginning with $\rho_0 = 0.5$, We obtain the following $P_1 = 0.32436$: $\sum_{i=0}^{1} p_i = 0.824361$, $P_2 = 0.1226$: $\sum_{i=0}^{2} p_i = 0.94696$, $P_3 = 0.037788$: $\sum_{i=0}^{3} p_i = 0.98475$, $P_4 = 0.01091$: $\sum_{i=0}^{4} p_i = 0.995658$, $P_5 = 0.003107$: $\sum_{i=0}^{5} p_i = 0.998764$, $P_6 = 0.000884$: $\sum_{i=0}^{6} p_i = 0.999648$.

Nomenclature

X: Service time random variable when arrival occurs

x: Duration of service time random variable X

μ: Service rate

$\mathfrak{R}$: Residual service time random variable

$f_{\mathfrak{R}}(x)$: Density function of Residual service time

λ: Arrival rate

R: Response time

$f_{\mathfrak{R}}(x)$: Density function of random variable X

$b(x)$: Density function of service time

$B(x)$: Service time distribution fuction

$b(x)dx$: Frequency of occurrences of service interval having length x

$M(x)$: Number of customers served by time t

W_q: Expected time arriving customer must wait until its service begins

L_q : Expected number of customers waiting in the queue

$E[\mathfrak{R}]$: Mean residual service time

ρ : Workload intensity

$N(t)$: Then number of customers present at time t

Conflict of Interest

Authors declare that there is no conflict of interest in this study.

References

Agboola SO 2017. *Repairman Problem with Multiple Batch Deterministic Repairs*. Unpublished Ph.D. Thesis, Obafemi Awolowo University, Ile-Ife, Nigeria, 331 pp.

Agboola SO 2007. *On the Analysis of M/G/1 Queues.* Unpublished M.Sc. Thesis, University of Ibadan, Ibadan, Nigeria, 125 pp.

Agboola SO 2010. *The Analysis of Markov Inter-arrival Queues Model with K – Server under Various Service Points.* Unpublished M.Sc. Thesis, Obafemi Awolowo University, Ile-Ife, Nigeria, 234 pp.

Bolch G, Greiner S & Trivedi KS 1998. *Queueing Networks and Markov Chains*. Wiley Interscience, New York, USA, pp. 24 – 78.

Charan J *et al* 2012. *Analysis of M/G/1 Queueing Model with State Dependent Arrival and Vacation*. Research Gate.

Charan J *et al* 2019. *Queue with Optional Service and Server Breakdowns.* Performance and Safety Management, Research Gate.

Kleinrock L 1975. *Queueing systems*, Volume I, Wiley – Interscience Publication, New York, USA, pp. 60 – 158.

Law AM & Kelton WD 2000. *Simulation Modeling and Analysis*. Third Edition, McGraw-Hill, New York, USA, pp. 12 – 45.

Lucantoni DM 1993. *The BMAP/G/1 Queue: Models and Techniques for performance evaluation of computer and communications systems*. Springer, New York, USA, pp. 1-46.

Medhi J 1980. *Stochastic Application of Queueing Theory*. Guahandi University, Guwa-hata, Indian, pp. 23 – 120.

Meini B 1997. *New convergence results on functional techniques for the numerical solution of M/G/1 type Markov chains*. *Numerical Mathematical,* 78: 39-58.

Michiel D *et al* 2017. *Analysis of a Discrete – Time Queue with General Service Demands and Phase – Type Service. J. Ind. and Mgt. Optimisation*, 13(4): 1901 – 1926.

Ross SM 1997. *Simulation*. Second edition, Academic Press, New York, USA, pp. 23 – 46.

Saad Y 2003. *Iterative Methods for Sparse Linear Systems.* Second edition, Society for Industrial Application of Mathematics (SIAM). Publications, Philadelphia, USA, pp. 12 – 35.

William S 2009. *Probability, Markov Chains, Queues and Simulation*, Princeton University press, Princeton and Oxford, UK, pp. 1 – 38.