A joint chance-constrained programming approach for call center workforce scheduling under uncertain call arrival forecasts

M. Excoffier, C. Gicquel, O. Jouini

A joint chance-constrained programming approach for call center workforce scheduling under uncertain call arrival forecasts

Mathilde Excoffier[1], Céline Gicquel[1], Oualid Jouini[2]

[1] Laboratoire de Recherche en Informatique, Université Paris Sud, Orsay, France
[2] Laboratoire Génie Industriel, Ecole Centrale Paris, Chatenay-Malabry, France

A joint chance-constrained programming approach for call center workforce scheduling under uncertain call arrival forecasts

Abstract

We consider a workforce management problem arising in call centers, namely the shift-scheduling problem. It consists in determining the number of agents to be assigned to a set of predefined shifts so as to optimize the trade-off between manpower cost and customer quality of service. We focus on explicitly taking into account in the shift-scheduling problem the uncertainties in the future call arrival rates forecasts. We model them as independent random variables following a continuous probability distribution. The resulting stochastic optimization problem is handled as a joint chance-constrained program and is reformulated as an equivalent large-size mixed-integer linear program. One key point of the proposed solution approach is that this reformulation is achieved without resorting to a scenario generation procedure to discretize the continuous probability distributions. Our computational results show that the proposed approach can efficiently solve real-size instances of the problem, enabling us to draw some useful managerial insights on the underlying risk-cost trade-off.

Keywords: Personnel planning, Call center shift scheduling, Customer abandonment, Stochastic programming, Probabilistic constraints, Mixed-integer linear programming

1. Introduction

Call centers can be broadly defined as facilities designed to support the delivery of some interactive service via telephone communications ([9]). Applications include among others telemarketing, customer service, help desk support and emergency dispatch. In most cases, the primary function of a call center is to receive phone calls that have been initiated by customers.

Such operations, known as "inbound" call centers, are the topic of the present paper.

Personnel planning is a key issue in call center management. Namely, as reported in [1], call centers are labor-intensive operations in which the cost of the staff members handling the phone calls (known as the agents) typically accounts for 60% to 80% of all the operating expenses. An efficient workforce management is thus crucial to achieve profitability in a call center.

Call center workforce management involves three main levels of decision-making (see e.g. [23]). Long-term planning decisions (6-12 months ahead) include the determination of how many agents to hire and train at what times based on aggregate long-term forecasts of demand for services. Short-term decisions (1-2 weeks ahead) involve the scheduling of an available pool of agents over an horizon typically spanning one week. These decisions are based on detailed short-term forecasts of agent requirements. Finally, real-time adjustment decisions, such as agent schedule updating and call routing, have to be made on a intra-day basis.

The present work is related to short-term workforce management decisions in call centers. These decisions usually involve two main steps. First, a range of possible shift patterns is defined and managers have to determine the number of agents to be assigned to each shift. Second, a rostering procedure combines shifts into rosters and assigns rosters to individual employees.

In the present paper, we focus on the first step of this process and consider the shift scheduling problem. We thus seek to determine the number of agents to be assigned to a set of predefined shifts so as to meet two potentially conflicting objectives, namely minimizing the manpower cost and delivering a high quality of service to the call center customers. To achieve this, we will look for shift schedules where the supply of agent resources is matched as closely as possible with the demand for services. Namely, understaffing would lead to customer dissatisfaction due to a poor quality of service while overstaffing would result in useless over-service to customers and higher than needed operating costs.

One of the major difficulties to be tackled while trying to match supply and demand in a call center is that the level of demand for services, i.e. the workload, is highly variable. This is mainly due to the fact that the call arrival rate (number of calls reaching the call center per unit of time) is subject to strong fluctuations over the course of a day or a week. This difficulty is usually handled in practice by dividing the scheduling horizon into a number of time periods of 15 to 60 minutes. In each period, the call arrival pro-

cess is modeled as a Poisson process with a constant and deterministically known arrival rate and the call center is treated as an independent queuing system in stationary state (see e.g. [11]). This allows to determine the required staffing level for each period which is set to be the minimum number of agents for which the target quality of service (expressed e.g. as the maximum allowed proportion of customers hanging up before being answered) is reached. A deterministic optimization problem has then to be solved to find the minimum-cost shift schedule ensuring that the required number of agents is staffed in each period of the horizon.

However, as mentioned e.g. by [1], at the time when decision on shift schedules is made, call arrival rates are most often not deterministically known. We only have estimations obtained via a forecasting procedure whose outcome is a point forecast and some probabilistic representation of the forecasting errors. The input data of the shift scheduling problem are thus subject to uncertainty: not taking this into account while building the shift schedule might lead to significant discrepancies between the call center performance targeted at the time scheduling decisions are made and the one actually obtained in practice (see [10]).

In the present paper, we propose a stochastic programming based approach to explicitly take into account the uncertainty on the call arrival rates in the shift scheduling problem. Our contributions are threefold. First we model the forecasting error on the call arrival rate in each period as a random variable following a continuous probability distribution. This is in contrast with most previously published approaches which rely on discrete probability distributions through the use of a finite set of scenarios to represent the uncertainty. Second, we propose to model the stochastic shift scheduling problem as a one-stage stochastic program involving a joint chance constraint. Such a model is particularly relevant when the call center is evaluated based on its ability at reaching, on all periods of the scheduling horizon, the target quality of service and when the call center management focuses on the risk of not meeting this objective. Another advantage is that it does not require introducing a penalty cost to be incurred when the target quality of service is not reached, the value of which might be difficult to estimate in practice. Third we present an efficient solution approach based on the reformulation of the stochastic program as an equivalent deterministic large-size mixed-integer linear program. This approach relies on the assumption that the random variables modeling the forecasting errors are independent from one another and comprises two main steps: (1) the reformulation of the joint

3

chance-constrained program into an equivalent deterministic program involving a series of non-linear terms, (2) the building of a numerical representation for these non-linear terms exploiting the fact that their underlying mathematical expressions involve non-increasing piecewise constant functions. Our computational experiments show that the proposed solution approach is capable of solving real-size instances of the problem within computation times compatible with an industrial use.

The remainder of the paper is organized as follows. We provide in Section 2 an overview of the related literature. We then introduce in Section 3 the proposed joint chance-constrained stochastic program. We explain how, under the assumption of independence between the forecasting errors, it can be reformulated as a stochastic program involving a set of individual chance constraints. Section 4 is devoted to the presentation of the proposed solution approach. Its main idea is to exploit the fact that, all other parameters being fixed, the minimum number of agents needed to handle the phone calls is a non-decreasing piecewise constant function of the call arrival rate. A small illustrative example is discussed in Section 5. The results of computational experiments carried out on real-size instances are provided in Section 6.

2. Literature review

Given the size of the call center industry and the complexity associated with its operations, call centers have emerged as a fertile ground for Operations Research. We refer the reader to [1] and [9] for a general introduction to this field and focus in what follows on the recently emerged research stream on stochastic call center shift scheduling. We distinguish three main features to classify the related papers: the call center setting, the representation of the uncertainty and the risk management measures.

In terms of call center architecture, the simplest case consists in a setting where a single pool of homogeneous agents handles a single class of infinitely patient calls. This amounts to using an Erlang C model to represent the call center in each period of the scheduling horizon (see [17] and [18]). However, the importance of modeling customer impatience and abandonment in call centers has been underlined in several papers such as [9] and [22]. Thus, similarly to [10] and [27], we use in the present paper a representation of the call center as an Erlang A model. For both the Erlang C and the Erlang A models, the performance evaluation of the call center can be done by exploiting analytical results available in the queuing theory literature. A

more complicated setting corresponds to skill-based routing call centers. In this case, the performance evaluation of the call center has to be made by relying either on simulation or on approximations under various asymptotic regimes. Stochastic shift scheduling for skill-based routing call centers has been studied by [4], [12], [13] and [32].

Another important feature to be considered is the way the uncertainty is described in the stochastic optimization model. In the present case, the uncertainty comes from the difficulty in exactly forecasting the future call arrival rates. The forecasting procedure provides a point estimate of the call arrival rate for each scheduling period as well as a probabilistic description of the forecasting error, e.g. in terms of a normal distribution with a zero mean and a standard deviation reflecting the forecast quality. However, in most cases, explicitly handling continuous probability distributions in a stochastic optimization problem is very computationally challenging as it requires computing multi-dimensional integrals during the solution procedure. Moreover, in the specific case under study, one is faced with an additional difficulty: the need to translate the probabilistic description of the uncertainty on the call arrival rates into a probabilistic description of the uncertainty on the agent requirements. Even for the simplest call center settings, there is no known analytical expression to directly carry out such a translation. This implies that the agent requirement in a given scheduling period is a random variable, whose probability distribution cannot be described by an analytical mathematical expression. This might explain why, to the best of our knowledge, all previously published approaches for stochastic call center shift scheduling rely on the use of discrete probability distributions to represent the uncertainty on the call arrival rates and translate each corresponding call arrival rate scenario into an agent requirement scenario in a pre-optimization step. Thus, [17] and [18] consider that the information on uncertainty is directly provided in the form of a discrete probability distribution. Scenario generation to discretize a continuous probability distribution is carried out by random sampling in [4], [12], [13], [27], [32] and by exploiting Gaussian quadrature in [10]. As explained in [27], this discretization implies that the obtained solution is a biased estimate of the true solution which would be obtained using a continuous distribution. In the present work, we model the forecasting errors as independent random variables, each following a continuous probability distribution, and keep this representation throughout the solution approach without resorting to a scenario generation procedure to discretize the probability distributions.

Finally, one may also classify the papers based on how the consequences of uncertainty are dealt with in the model. Namely, as the main shift scheduling decisions have to be made prior the realization of the call arrival process, we cannot guarantee that the workforce schedule will succeed in providing the target quality of service for every possible realization of the call arrival rates. Several modeling alternatives can be considered to account for this in the optimization problem. Most existing approaches rely on single-stage stochastic programming approaches where all scheduling decisions have to be made prior to the realization of call arrival rates. [17] and [27] introduce a penalty to be incurred when the target quality of service is not reached. Similarly, the authors of [13] do not use a target quality of service but consider a profit per answered call. All three approaches amount to finding the shift schedule optimizing the expected cost or profit. Constraints ensuring that the expected quality of service over the whole horizon is above a certain limit are used in [10] and [18]. Finally joint chance-constrained programming formulations have been used in [12] and [32]. Two-stage stochastic programming approaches allowing for the use of recourse actions once the arrival rate realization is known have been considered in [4] and [10]. In the present paper, we propose a one-stage stochastic programming approach using joint chance constraints. Our work thus shares some similarities with the one of [12] and [32]. However, these two papers consider a single period staffing problem for a multi-class multi-skill call center and focus on the probability of reaching the target quality of service for every customer class while we investigate a multi-period shift scheduling problem in a single-class single-skill call center and focus on the probability of reaching the target quality of service for every period of the scheduling horizon.

3. Joint chance-constrained programming model

This section is devoted to the detailed presentation of the problem under study in the present paper: the stochastic shift scheduling problem in a single-class single-skill call center. We first provide a description of the deterministic variant of the problem which can be formulated as a mixed-integer linear program. We then consider the stochastic variant of the problem and introduce the proposed joint chance-constrained programming model. We finally show how, under the assumption of statistical independence between the random variables representing the forecasting errors on the future call arrival rates, this model can be reformulated as an equivalent deterministic

6

mathematical program involving some non-linear terms.

3.1. Deterministic formulation

We consider the shift scheduling problem for a single-class single-skill call center. The scheduling horizon consists of T periods, the typical duration of which is between 15 and 60 minutes.

We assume that a set of S predefined overlapping shifts is provided. A shift s corresponds to a possible work schedule for an agent on the scheduling horizon and define in which periods an agent assigned to it will be active. Shifts are created based on work regulations (maximum number of hours per day, lunch and coffee breaks...). We define $A_{ts} = 1$ if an agent assigned to shift s works in period t, 0 otherwise. The cost of an agent assigned to shift s is denoted c_s.

We first focus on the deterministic case where the mean call arrival rate in period t, denoted λ_t, is assumed to be deterministically known. The call arrival process during a period t is thus modeled as a Poisson process with rate λ_t. Service times are assumed to be independent and exponentially distributed with rate μ. Customers are served in the order of their arrivals, i.e. under the First Come-First Served (FCFS) discipline of service. The queue capacity is assumed to be infinite. Finally, customers patience is limited, i.e. a customer placed in the queue might hang up before starting service. Patience times are assumed to follow an exponential distribution of parameter γ. This results in an Erlang A model (M/M/s+M).

Following the SIPP method (see [11]), we assume that each period is sufficiently long to allow the queuing system to reach its stationary regime. We measure the customer quality of service by the stationary fraction of customers that abandon before service (i.e. the fraction of the customers that hang up before being answered by an agent because they have reached their patience threshold). Given a target value p^* for the probability of abandonment, a value for the service rate μ and the mean time before abandonment $1/\gamma$, we can compute $n_t = \phi_{\mu,\gamma,p^*}(\lambda_t)$, the minimum number of agents needed in period t to reach the target quality of service. Note that an analytical expression of the function ϕ_{μ,γ,p^*} is not available. However, for a given value of λ_t, we can obtain the corresponding value n_t by resorting to an algorithm exploiting known results on the performance evaluation of Erlang A systems: see subsection 4.1 and [15] for more detail.

Finally we introduce the integer decision variables x_s defined as the number of agents assigned to shift s.

With this notation, the deterministic shift scheduling problem has the following mixed-integer linear programming (MILP) formulation:

$$
\begin{cases}
Z = min \sum_{s=1}^{S} c_s x_s & (1) \\[2em]
\sum_{s=1}^{S} A_{ts} x_s \geq n_t & \forall t = 1...T & (2) \\[1.5em]
x_s \in \mathbb{Z}^+ & \forall s = 1...S. & (3)
\end{cases}
$$

The objective function (1) corresponds to minimizing the total cost of the shift schedule. Constraints (2) are quality-of-service constraints. They guarantee that, in each period t, the total number of working agents ($\sum_{s=1}^{S} A_{ts} x_s$) is greater than $n_t = \phi_{\mu,\gamma,p^*}(\lambda_t)$, the minimum number of agents required to ensure that the probability of abandonment is below the prescribed limit p^*. (3) are integrality constraints on the decision variables.

3.2. Joint chance-constrained programming formulation

In practice, the future call arrival rates $(\lambda_1, ..., \lambda_t, ..., \lambda_T)$, are not deterministically known as the time when the decisions on the shift schedules have to be made. Their value can only be estimated through a forecasting procedure relying e.g. on statistical methods. We assume in the present paper that this forecasting procedure provides point estimates $(\overline{\lambda_1}, ..., \overline{\lambda_t}, ...\overline{\lambda_T})$ as well as a probabilistic description of the forecasting errors $(\epsilon_1, ..., \epsilon_t, ..., \epsilon_T)$. It amounts to modeling the call arrival rates as a random vector $(\Lambda_1, ..., \Lambda_t, ..., \Lambda_T)$ whose multivariate probability distribution is assumed to be known.

This implies that the minimum number of agents required in each period t to reach the target quality of service cannot be considered as deterministically known. It is in fact a random variable (denoted by N_t in what follows) which is linked to the forecasting error ϵ_t through the expression: $N_t = \phi_{\mu,\gamma,p^*}(\overline{\lambda_t} + \epsilon_t)$. We first note that N_t represents an integer number of agents and is thus a discrete random variable defined over the set of positive integer numbers $\mathbb{N}$. However, we would like to point out that, even if we have on hand a probabilistic description of the random variable ϵ_t, deriving an explicit mathematical description of the probability distribution of N_t is not possible due to the fact that an analytical expression of the function ϕ_{μ,γ,p^*} does not exist.

The deterministic parameter n_t involved in constraints (2) has thus to be replaced by the stochastic parameter N_t. As a consequence, we will not be able anymore to guarantee that these constraints will be respected, i.e. that the target quality of service will be reached, for every possible realization of the random variables $(N_1, ..., N_t, ..., N_T)$. As mentioned in Section 2, a first possible way to manage this situation is to allow the constraints to be violated for some realizations of the random variables and to limit the amount of these violations either by penalizing them in the objective function or by introducing constraints on their expected value. The present work relies on another modeling alternative which consists in limiting the probability that these violations occur. This leads to the introduction of a probabilistic constraint in the formulation which imposes a lower bound π on the probability that the quality-of-service constraints are satisfied by the shift schedule. In this case, the value $1 - \pi$ can be understood as the maximum acceptable risk level, i.e. as the maximum acceptable likelihood that the target quality of service is not met on all periods of the scheduling horizon. While setting the value of $1 - \pi$, call center managers can trade-off between risk aversion and cost minimization.

This leads to the reformulation of the problem (1)-(3) as the following chance-constrained program JCCP.

$$
\begin{cases}
Z = min \sum_{s=1}^{S} c_s x_s & (4) \\
\Pr(\sum_{s=1}^{S} A_{ts}x_s \geq N_t, \forall t = 1...T) \geq \pi & (5) \\
x_s \in \mathbb{Z}^+ & \forall s. \quad (6)
\end{cases}
$$

We note that in the probabilistic constraint (5), the deterministic values n_t have been replaced by their random counterparts N_t. Moreover, we point out that (5) is a joint chance constraint: it imposes a lower bound π on the probability that all quality-of-service constraints are satisfied simultaneously, i.e. on the probability that in all periods $t = 1...T$, the number of working agents is sufficient to meet the target quality of service.

One of the main advantages of the use of chance-constraints in the present case is that it avoids the need to introduce a penalty cost to be incurred when the quality-of-service constraints are violated. Namely, it might be difficult

for the call center management to estimate the value of such a penalty cost. Moreover, as mentioned in [12], the use of a constraint on the expected value of the quality of service over the whole scheduling horizon such as it is done in [10] and [18] is relevant in the situations where the evaluation of the call center performance is based on the average value of the quality of service over the course of a day or a week. In this case, shortfalls in the quality of service in some periods can be compensated by a quality of service higher than the target value in other periods. On the contrary, a chance-constrained programming model such as model JCCP is particularly adequate in the situations where the call center performance is assessed on each period of the scheduling horizon and where high quality-of-service periods cannot be used to make it up for low quality-of-service periods. In this case, the call center management will focus on reaching the target quality of service in every period of the scheduling horizon and on managing the risk of failing to meet this objective.

Except for a few exceptions, mathematical programs involving a joint probabilistic constraint such as (5) are still largely intractable. We refer the reader to [30] for a general introduction to the field of chance-constrained programming. In terms of solution approaches, a variety of tractable approximations have been proposed to handle general joint chance-constrained problems. They mostly rely on conservative convex approximations (see e.g. [28], [25]) or on a discretization of the probability distribution through sampling (see e.g. [6], [19]). Stability of the solution with respect to perturbations of the probability distribution has been investigated among others by [5].

The present work is mostly related to a specific class of joint chance-constraint programs, namely those involving separable probabilistic constraints. This corresponds to the case where the probabilistic constraint features inequalities where only the right hand sides are subject to uncertainty. A key concept in this area is the p-efficiency introduced by [26]. This concept has been widely used to develop methods to solve chance-constraint problems in which the random right hand sides have discrete probability distributions. Those methods require the generation of a subset of the p-efficient points of the probability distribution through an enumeration scheme (see e.g. [8], [3] and [16]). The present work belongs to this line of research. However, we assume that the random variables $N_1, ..., N_T$ are statistically independent from one another. In this case, there is an explicit characterization of the set of p-efficient points (see lemma 2.7 in [8]), which we exploit in our solution approach (see reformulation EDetF below) to avoid generating them through

a computationally intensive enumeration.

3.3. Equivalent individual chance-constrained programming formulation

In what follows, we assume that the random variables representing the forecasting errors on the future call arrival rates are statistically independent from one another. Note that this is consistent with the use of a simple call arrival forecasting method such as the fixed-effects model investigated in [14]. This method allows to take into account both the intra-day and the intra-week seasonality in the call arrivals and makes use of independent and identically normally distributed random variables with mean 0 to represent the forecasting residuals. In the simulation carried out in [14] on data coming from real case studies, the fixed-effects forecasting model is shown to be quite efficient at providing two-week-ahead and one-week-ahead forecasts, which typically corresponds to the lead time with which shift schedules are built in practice. We thus focus on the case where the forecasting errors are modeled as independent random variables and leave the development of a solution approach capable of explicitly handling the statistical dependencies between the forecasting errors open for further work.

We now show how, under the assumption of statistically independent forecasting errors, problem JCCP can be reformulated as a stochastic program involving a set of individual chance constraints and explain how an equivalent deterministic mathematical program involving some non-linear terms can be obtained.

The fact that the forecasting errors $(\epsilon_1, ..., \epsilon_t, ..., \epsilon_T)$ are independent random variables implies that the random events $(E_t : \sum_{s=1}^{S} A_{ts} x_s \geq N_t, t = 1...T)$ are independent from one another. We thus have: $\Pr(\bigcap_{t=1}^{T} E_t) = \prod_{t=1}^{T} Pr(E_t)$.

This leads to the following equivalent formulation for problem JCCP:

$$
\begin{cases}
Z = min \sum_{s=1}^{S} c_s x_s & (7) \\
\prod_{t=1}^{T} \Pr(\sum_{s=1}^{S} A_{ts} x_s \geq N_t) \geq \pi & (8) \\
x_s \in \mathbb{Z}^+ \qquad\qquad \forall s. & (9)
\end{cases}
$$

A first possible way to handle constraint (8) is by decomposing the joint chance constraint into a series of T individual chance constraints and by

11

pre-allocating to each individual chance constraint the same fraction $\frac{1}{T}$ of the risk. This leads to the following formulation which provides a feasible solution of problem JCCP.

$$
\begin{cases}
Z = min \sum_{s=1}^{S} c_s x_s & (10) \\[2ex]
\Pr(\sum_{s=1}^{S} A_{ts} x_s \geq N_t) \geq \pi^{1/T} & \forall t & (11) \\[2ex]
x_s \in \mathbb{Z}^{+} & \forall s. & (12)
\end{cases}
$$

By using $F_{N_t}^{-1}$ the inverse cumulative probability distribution of the random variable N_t, problem (10)-(12) has a deterministic equivalent formulation in the form of the following mixed-integer linear program denoted EDetB.

$$
\begin{cases}
Z = min \sum_{s=1}^{S} c_s x_s & (13) \\[2ex]
\sum_{s=1}^{S} A_{ts} x_s \geq F_{N_t}^{-1}(\pi^{1/T}) & \forall t & (14) \\[2ex]
x_s \in \mathbb{Z}^{+} & \forall s. & (15)
\end{cases}
$$

However, deciding *a priori* to allocate the same fraction $\frac{1}{T}$ of the risk to all the periods of the scheduling horizon is somewhat arbitrary and might lead to a solution of problem JCCP significantly more expensive than the optimal one. Hence, as proposed in [7], instead of preallocating the risk between the periods before solving the optimization problem, one can introduce the portion of the risk allocated to period t as a decision variable (denoted by y_t in what follows) and impose that $\sum_{t=1}^{T} y_t = 1$ to ensure that the maximum allowed risk level is not greater than π. This additional flexibility in the risk management enables us to come up with a minimum-cost feasible solution of JCCP.

This leads to the following individual chance-constraint program whose feasible region is the same as the one of problem JCCP.

$$\begin{cases} Z = min \sum_{s=1}^{S} c_s x_s & (16) \\[3mm] \Pr(\sum_{s=1}^{S} A_{ts} x_s \geq N_t) \geq \pi^{y_t} & \forall t & (17) \\[3mm] \sum_{t=1}^{T} y_t = 1 & (18) \\[3mm] x_s \in \mathbb{Z}^+ & \forall s & (19) \\[2mm] y_t \in]0;1] & \forall t. & (20) \end{cases}$$

Similarly to problem (10)-(12), problem (16)-(20) can be reformulated as the following equivalent deterministic program EDetF.

$$\begin{cases} Z = min \sum_{s=1}^{S} c_s x_s & (21) \\[3mm] \sum_{s=1}^{S} A_{ts} x_s \geq F_{N_t}^{-1}(\pi^{y_t}) & \forall t & (22) \\[3mm] \sum_{t=1}^{T} y_t = 1 & (23) \\[3mm] x_s \in \mathbb{Z}^+ & \forall s & (24) \\[2mm] y_t \in]0;1] & \forall t. & (25) \end{cases}$$

We first note that both problem EDetB and problem EDetF make use of $F_{N_t}^{-1}$, the inverse cumulative probability distribution of the discrete random variable N_t, for which no analytical expression is known. A first step towards solving these two problems thus consists in building a numerical representation of $F_{N_t}^{-1}$ by exploiting the relation $N_t = \phi_{\mu,\gamma,p^*}(\overline{\lambda_t} + \epsilon_t)$. This is the purpose of Subsections 4.1 and 4.2 below. This should be enough to solve problem EDetB as it only involves computing the value of $F_{N_t}^{-1}$ for a given value $\pi^{1/T}$. However, solving problem EDetF requires handling a set of terms $F_{N_T}^{-1}(\pi^{y_t})$ whose value depends on the decision variables y_t and therefore cannot be computed in a preprocessing step. In subsection 4.3, we show that the functions $\Psi_t : [y_t \mapsto F_{N_T}^{-1}(\pi^{y_t})]$ are non-increasing piecewise constant functions.

This enables us to reformulate problem EDetF as a large size mixed-integer linear program (see subsection 4.4).

4. Solution approach

4.1. Minimum number of agents required as a function of the call arrival rate

The first step of our solution approach consists in building a numerical representation of the inverse cumulative probability distribution $F_{N_t}^{-1}$ of the random variable N_t representing the minimum number of agents required in period t to reach the target quality of service. As N_t is linked to the random call arrival rate $\overline{\lambda_t} + \epsilon_t$ through the relationship $N_t = \phi_{\mu,\gamma,p^*}(\overline{\lambda_t} + \epsilon_t)$, we first focus on studying the basic mathematical properties and on obtaining a numerical description of function ϕ_{μ,γ,p^*}.

We defined ϕ_{μ,γ,p^*} in subsection 3.2 as the function of the call arrival rate λ providing the minimum number of servers n needed to reach the target service level p^* when the service and patience threshold rates are μ and γ, respectively.

An analytical expression of function ϕ_{μ,γ,p^*} is not available. However, it is possible to compute $\phi_{\mu,\gamma,p^*}(\lambda)$ for a given value of λ by resorting to a simple algorithm.

This algorithm exploits previously published results on the performance evaluation of Erlang A systems (see e.g. [15] and [21]). In particular, given a call arrival rate λ, a service rate μ, a patience threshold rate γ and a number of agents C in the call center, the probability of abandonment p, i.e. the fraction of customers abandoning before being served, is given by the following formula (see [15]):

$$p(\lambda, \mu, \gamma, C) = \frac{\gamma}{\lambda} \sum_{k=C+1}^{+\infty} (k - C) q_k. \qquad (26)$$

In Equation (26), q_k is defined as the stationary probability that there are k customers in the system and the term $\sum_{k=C+1}^{+\infty} (k - C) q_k$ computes the expected number of customers in the queue.

For $k \geq C + 1$, q_k is given by:

$$q_k = \frac{\lambda^k}{C! \mu^C \prod_{j=1}^{k-C} (C\mu + j\gamma)} \times \frac{1}{\sum_{i=0}^{C} \frac{\lambda^i}{i! \mu^i} + \frac{\lambda^C}{C! \mu^C} \sum_{i=C+1}^{+\infty} \frac{\lambda^{i-C}}{\prod_{j=1}^{i-C} (C\mu + j\gamma)}}. \qquad (27)$$

Note that both Equations (26) and (27) involve infinite summations, consistently with the assumption that the queue capacity is infinite. In order to be able to numerically estimate p, we truncated these infinite summations and considered a finite number of terms in the summation. In our numerical experiments, the truncation point K_{max} was set to a value large enough to ensure that its impact on the accuracy of the numerical evaluation of p was below 0.001.

Moreover, for given values of λ, μ and γ, the abandonment probability p is known to be non-increasing in the number of agents C (see e.g. [2] and [10]).

This allows us to compute the value of $\phi_{\mu,\gamma,p^*}(\lambda) = min\{C \in \mathbb{N} | p(\lambda, \mu, \gamma, C) \leq p^*\}$ through the following simple iterative procedure. For fixed values of λ, μ and γ, we first start with evaluating the abandonment probability for a call center setting involving $C = 1$ agent (note that our call center model is stable for any $\gamma > 0$) and iteratively adds one agent ($C = C + 1$) until the abandonment probability falls below the target value. When the procedure stops, the value of C is equal to the $\phi_{\mu,\gamma,p^*}(\lambda)$.

This provides us with a point estimation of function ϕ_{μ,γ,p^*} for a given value of λ. However, in order to develop our solution approach, we need to study some of the basic mathematical properties of this function. In particular, we will make use of the following proposition.

Proposition 1
Function ϕ_{μ,γ,p^*} is a piecewise-constant, non-decreasing function of λ.
Let $l \in \mathbb{N}$ and denote by $\tilde{\lambda}_l$ the value of the λ such that $p(\tilde{\lambda}_l, \mu, \gamma, l) = p^*$.
We have:

$$\phi_{\mu,\gamma,p^*}(\lambda) = \begin{cases} 0 & \text{for } \lambda = 0 \\ l & \text{for } \lambda \in]\tilde{\lambda}_{l-1}; \tilde{\lambda}_l] \end{cases}$$

Proof. See appendix. $\square$

The current state of the art on Erlang A models does not provide us with an analytical expression to compute the threshold values $\tilde{\lambda}_l$ involved in the definition of function ϕ_{μ,γ,p^*}. We thus use in what follows a numerical description of ϕ_{μ,γ,p^*} over a finite interval $[0; \lambda_{max}]$ which is obtained by computing conservative estimations of the threshold values $\tilde{\lambda}_l$. To achieve this, we go through interval $[0; \lambda_{max}]$ by using small steps $\Delta\lambda$ and record the values $\tilde{\lambda}_l$ where function ϕ_{μ,γ,p^*} changes level.

ALG
   | Set $\tilde{\lambda}_0 = 0$.
   | Set $\lambda = \Delta\lambda$ and $l = 1$.
   | While $\lambda \leq \lambda_{max}$:
   | compute $n = \phi_{\mu,\gamma,p^*}(\lambda)$.
   | if $n = l + 1$:
   | $\tilde{\lambda}_l = \lambda - \Delta\lambda$
   | $l = l + 1$
   | $\lambda = \lambda + \Delta\lambda$

Running algorithm ALG can be time-consuming depending on the values of $\Delta\lambda$ and λ_{max}. However, for given values (μ, γ, p^*), it has to be done only once in a preprocessing step before optimization. The corresponding computation time is thus not included in the numerical results presented in Section 6.

4.2. Inverse cumulative probability distribution of random variables N_t

We now study F_{N_t}, the cumulative probability distribution of the random variable $N_t = \phi_{\mu,\gamma,p^*}(\overline{\lambda_t} + \epsilon_t)$. We first note that, as N_t is a discrete random variable, F_{N_t} is a non-decreasing piecewise-constant function defined over $\mathbb{R}$ by:

$$F_{N_t}(x) = \Pr(N_t \leq x) = \begin{cases} 0 & \text{for } x \leq 0 \\ F_{N_t}(\lceil x \rceil) & \text{for } x > 0 \end{cases} .$$

Function F_{N_t} is thus fully described by giving its values for the set of positive integer values of x. We therefore focus on computing the value of F_{N_t} over the set $\mathbb{N}^*$.

Let $l \in \mathbb{N}^*$. As ϕ_{μ,γ,p^*} is a non-decreasing piecewise constant function of λ (see Proposition 1), there is a bi-univocal correspondence between the two random events $\{N_t \leq l\}$ and $\{\overline{\lambda_t} + \epsilon_t \leq \tilde{\lambda}_l\}$. Besides, we assume that the forecasting error ϵ_t follows a normal distribution $\mathcal{N}(0, \sigma_t)$. We thus have:

$$F_{N_t}(l) = \Pr(N_t \leq l) = \Pr(\overline{\lambda_t} + \epsilon_t \leq \tilde{\lambda}_l) = F_{StdG}\left(\frac{\tilde{\lambda}_l - \overline{\lambda_t}}{\sigma_t}\right) \qquad (28)$$

where F_{StdG} denotes the cumulative probability distribution of the standard normal distribution.

By denoting $\alpha_{t,l} = F_{StdG}\left(\frac{\tilde{\lambda}_l - \overline{\lambda}_t}{\sigma_t}\right)$, we obtain the following description of $F_{N_t}^{-1}$, the inverse cumulative probability distribution of the discrete random variable N_t.

$$F_{N_t}^{-1}(a) = \begin{cases} 0 & \text{if } a = \alpha_{t,0} = 0 \\ l & \text{if } a \in]\alpha_{t,l-1}; \alpha_{t,l}], \ \forall l \in \mathbb{N}^*. \end{cases} \tag{29}$$

The quantity $F_{N_t}^{-1}(a)$ can be understood as the minimum number of agents required in period t to ensure that the risk of not reaching the target quality of service p^* is less than $1 - a$. We note that $F_{N_t}^{-1}$ is defined over $[0; 1[$ and that its limit when a approaches 1 is $+\infty$. In the numerical experiments presented in Section 5, we thus limit ourselves to a description of $F_{N_t}^{-1}$ on the interval $[0; \alpha_{max}]$ with $\alpha_{max} < 1$.

As mentioned at the end of subsection 3.3, by setting $a = \pi^{1/T}$, we can use the expression (29) to compute the right hand side value of the constraints (14) involved in problem EDetB in a pre-optimization step. Solving the resulting mixed-integer linear program then provides us with a feasible solution of problem JCCP.

4.3. Functions $\Psi_t : [y_t \mapsto F_{N_T}^{-1}(\pi^{y_t})]$

However, solving problem EDetF is more complex as it requires handling a set of terms $F_{N_T}^{-1}(\pi^{y_t})$ where y_t is a decision variable of the optimization problem representing the portion of the total allowed risk allocated to period t. This subsection is thus devoted to the study of the functions $\Psi_t : [y_t \mapsto F_{N_T}^{-1}(\pi^{y_t})]$, for $t = 1...T$. Note that $\Psi_t(y_t)$ can be understood as the minimum number of agents required in period t to ensure that the risk of not reaching the target quality of service p^* is below $1 - \pi^{y_t}$.

Proposition 2
Function Ψ_t is a piecewise-constant, non-increasing function defined over interval $]0; 1]$.

Proof. See appendix. □

There thus exist $\beta_{t,0} = 1 > \beta_{t,1} > ...\beta_{t,m-1} > \beta_{t,m} > \beta_{t,m+1}... > 0$ such that $\Psi_t(y_t)$ is constant over each interval $[\beta_{t,m+1}; \beta_{t,m}[$. We denote ν_{tm} the integer value of $\Psi_t(y_t)$ over the interval $[\beta_{t,m+1}; \beta_{t,m}[$.

This leads to the following description of Ψ_t:

$$\Psi_t(y_t) = \begin{cases} \nu_{t,0} & \text{if } y_t \in [\beta_{t,1}; \beta_{t,0}] \\ \nu_{t,m} & \text{if } y_t \in [\beta_{t,m+1}; \beta_{t,m}[, \; \forall m \in \mathbb{N}^*. \end{cases} \tag{30}$$

The numerical value of ν_{tm} can be computed by the following simple mathematical induction on m:

$$\begin{cases} \nu_{t,0} = \Psi_t(1) = F_{N_t}^{-1}(\pi) \\ \nu_{t,m+1} = \nu_{t,m} + 1 & \forall m \in \mathbb{N}^*. \end{cases} \tag{31}$$

As for the threshold value $\beta_{t,m}$, its value can be deduced from:

$$\Psi_t(y_t) = \nu_{tm} \Leftrightarrow \qquad F_{N_t}^{-1}(\pi^{y_t}) = \nu_{tm} \tag{32}$$

$$\Leftrightarrow \qquad \alpha_{t,\nu_{tm}-1} < \pi^{y_t} \leq \alpha_{t,\nu_{tm}} \tag{33}$$

$$\Leftrightarrow \qquad \frac{ln(\alpha_{t,\nu_{tm}})}{ln(\pi)} \leq y_t < \frac{ln(\alpha_{t,\nu_{tm}-1})}{ln(\pi)}, \tag{34}$$

which leads to:

$$\beta_{t,m} = \frac{ln(\alpha_{t,\nu_{tm}-1})}{ln(\pi)}. \tag{35}$$

Note that Ψ_t is not defined for $y_t = 0$ and that its limit when y_t approaches 0 is is $+\infty$. In our numerical description of Ψ_t, we therefore introduce a value y_{min} ($y_{min} > 0$) representing the smallest value for which Ψ_t is numerically defined. We denote $M_t = \Psi_t(y_{min}) - \Psi_t(0)$ the finite number of steps involved in the numerical description of function Ψ_t and we set $\beta_{t,M_t+1} = y_{min}$.

4.4. Reformulation of problem EDetF as a large-size MILP

Proposition 2 implies that the right hand side of constraints (22) is a non-increasing piecewise constant function of y_t. In this subsection, we exploit this result to reformulate problem EDetF as a mixed-integer linear program (MILP) involving a large number of binary variables and constraints.

We introduce binary variables z_{tm} defined for $t = 1..T$ and $m = 1...M_t$ by:

$$z_{tm} = \begin{cases} 1 & \text{if } y_t \leq \beta_{t,m} \\ 0 & \text{otherwise.} \end{cases}$$

and reformulate EDetF as:

$$Z = min \sum_{s=1}^{S} c_s x_s \tag{36}$$

$$\sum_{s=1}^{S} A_{ts} x_s \geq \nu_{t,0} + \sum_{m=1}^{M_t} z_{tm} \qquad \forall t \tag{37}$$

$$y_t \geq \beta_{t,1} + \sum_{m=1}^{M_t} (\beta_{t,m+1} - \beta_{t,m}) z_{tm} \qquad \forall t \tag{38}$$

$$z_{t,m} \leq z_{t,m-1} \qquad \forall t, \forall m \tag{39}$$

$$\sum_{t=1}^{T} y_t = 1 \tag{40}$$

$$x_s \in \mathbb{Z}^+ \qquad \forall s \tag{41}$$

$$y_t \in]0; 1] \qquad \forall t \tag{42}$$

$$z_{tm} \in \{0; 1\} \qquad \forall t, \forall m. \tag{43}$$

Note that the non-linear constraints (22) have been replaced by the large set of linear constraints (37)-(39) which can be understood as follows. In constraints (37), the non-linear term $F_{N_t}^{-1}(\pi^{y_t})$ has been replaced by the linear expression $\nu_{t,0} + \sum_{m=1}^{M_t} z_{tm}$. $\nu_{t,0}$ represents the minimum number of agents required in period t to ensure that the risk of not reaching the target quality of service is below $1 - \pi$. $\sum_{m=1}^{M_t} z_{tm}$ provides the number of additional agents needed to decrease the risk in period t from its maximum acceptable value $1 - \pi$ to a smaller value $1 - \pi^{y_t}$.

Now, to ensure that $\nu_{t,0} + \sum_{m=1}^{M_t} z_{tm} = F_{N_t}^{-1}(\pi^{y_t})$, the value of the binary variables $z_{t,m}$ has to be linked to the value of y_t. This is the purpose of constraints (38)-(39) which impose that y_t stays above a lower bound, the value of which depends on the values of the $z_{t,m}$ variables. Thus, in case no additional agent is staffed in t, i.e. in case all variables $z_{t,m}$ are set to 0, constraints (38) impose that y_t stays above the lower bound $\beta_{t,1}$, which corresponds to the first threshold value appearing in the definition of function Ψ_t (see expression 30). In case $A > 0$ additional agents are staffed in t, constraints (39) ensure that $z_{t,1} = ... = z_{t,A} = 1$ and $z_{t,A+1} = ... = z_{t,M_t} = 0$ so that constraints (38) are equivalent to $y_t > \beta_{t,A+1}$. Note that the values of $\beta_{t,m}$ are strictly decreasing with m, i.e. $\beta_{t,A+1} < \beta_{t,1}$. Thus by staffing A additional agents in period t, we allow a potential decrease of y_t from $\beta_{t,1}$ to

$\beta_{t,A+1}$, which corresponds to allowing a potential decrease of the risk from $1-\pi^{\beta_{t,1}}$ to $1-\pi^{\beta_{t,A}}$. In other works, constraints (38)-(39) make sure that the more additional agents are staffed in period t, the larger the allowed decrease of y_t and consequently of the risk level $1-\pi^{y_t}$ assigned to period t.

5. A small illustrative example

We introduce a small instance of the call center shift scheduling problem in order to illustrate the solution approach and compare between the two formulations EDetB and EDetF discussed in Section 4.

We consider a scheduling horizon of $T = 10$ periods, corresponding to a single day (8:00 to 18:00) divided into 1-hour periods. The expected service time is $\frac{1}{\mu} = 1$ minute and the mean time before abandonment is $\frac{1}{\gamma} = 1.25$ minute. The target quality of service is defined as the maximum allowed abandonment probability: $p^* = 5\%$. Point forecasts $(\overline{\lambda}_1, ..., \overline{\lambda}_t, ...\overline{\lambda}_T)$ for the expected call arrival rates are provided for the scheduling horizon. They display an intra-day seasonality as frequently encountered in call centers (see e.g. [9]) with peak hours in the late morning and early afternoon and off-peak hours in the early morning, midday and late afternoon. The forecasting errors $(\epsilon_1, ..., \epsilon_t, ..., \epsilon_T)$ are modeled as independent random variables, each one following a normal distribution $\mathcal{N}(0, \sigma_t)$. We set $\sigma_t = 0.5\lambda_t$. $S = 5$ predefined overlapping shifts are available: 3 correspond to full-time positions working 7 hours a day, 2 correspond to half-time positions working 4 hours a day. Each full-time position shift involves a one-hour lunch break scheduled between 12:00 and 14:00. Table 1 displays the values of $\overline{\lambda}_t$ and σ_t as well as the shift matrix A. The maximum acceptable risk level $1-\pi$ is set to 10%.

The solution approach proposed to solve problems EDetB and EDetF comprises four main steps. First, we build a numerical representation of function ϕ_{μ,γ,p^*} by using algorithm ALG with $K_{max} = 10000$, $\lambda_{max} = 2000$ and $\Delta\lambda = 0.001$. Second, for each period t, we build a numerical representation of function $F_{N_t}^{-1}$ with $\alpha_{max} = 0.999999$ and use it to compute the right hand side value $F_{N_t}^{-1}(1-\pi^{1/T})$ of the constraints (14) involved in problem EDetB. Third, we set $y_{min} = 0.0001$ and compute for each period t the threshold values β_{tm} used in the description of function Ψ_t, by relying on expression (35): see Figure 1 for a graphical representation of function Ψ_1 for the illustrative example. This provides the input data needed to define problem EDetF. Four, we solve the mixed-integer linear programs corresponding to problems EDetB and EDetF using a mathematical solver.

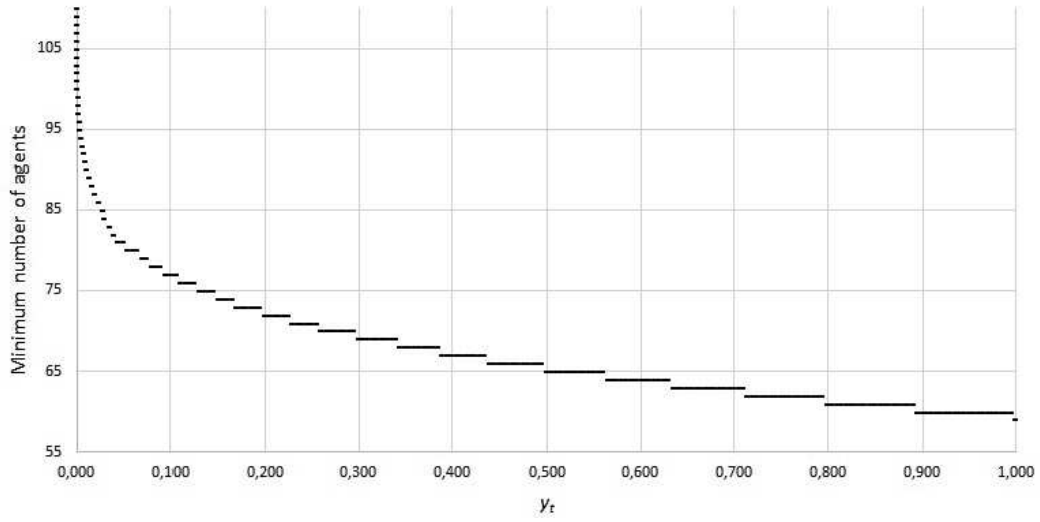| Time slot	t	λ_t	σ_t	$A_{t,1}$	$A_{t,2}$	$A_{t,3}$	$A_{t,4}$	$A_{t,5}$
8:00- 9:00	1	36	18	1	0	0	0	0
9:00-10:00	2	75	37.5	1	1	0	1	0
10:00-11:00	3	80	40	1	1	1	1	0
11:00-12:00	4	39	19.5	1	1	1	1	0
12:00-13:00	5	15	7.5	0	1	1	1	0
13:00-14:00	6	51	26.5	1	0	0	0	0
14:00-15:00	7	73	36.5	1	1	1	0	1
15:00-16:00	8	62	31	1	1	1	0	1
16:00-17:00	9	51	26.5	0	1	1	0	1
17:00-18:00	10	20	10	0	0	1	0	1

Table 1: Illustrative example: input data

Figure 1: Illustrative example: function Ψ_1

Solving problem EDetB to optimality provides the shift schedule $x^*_{EDetB} = (110, 57, 0, 0, 53)$ with $Z^*_{EDetB} = 1381$ whereas solving problem EDetF to optimality provides the shift schedule $x^*_{EDetF} = (100, 50, 0, 0, 49)$ with $Z^*_{EDetF} = 1246$. A detailed description of the corresponding solutions is provided in Table 2.

We note how the use of a flexible sharing out of the risk between the scheduling periods, as is done in problem EDetF, leads to a cost reduction by 10.8% as compared to the use of a predefined allocation of the risk, as is done in problem EDetB. This can be mainly explained as follows. In formulation EDetB, the portion of risk allocated to period t is arbitrarily fixed to $y_t = \frac{1}{T}$ in a pre-optimization step and as a consequence, there is no possibility to adjust the required number of agents needed in period t during the optimization step. The value of the total number of agents assigned to each shift will thus be driven mainly by the number of agents needed in a few "difficult to staff" periods. These periods typically corresponds either to peak hours where the mean call arrival rate is large or to off-peak periods where only a few shifts are working. This leads to a significant overstaffing in many periods as the agents used to staff these few difficult periods will also be working on the other periods defined by their assigned shifts. Thus, we note that in the optimal solution of problem EDetB displayed in Table 2, we have $\sum_{s=1}^{S} A_{ts} x^*_s = F^{-1}_{N_t}(\pi^{y_t})$ in the "difficult to staff"' periods 3, 6 and 9 whereas $\sum_{s=1}^{S} A_{ts} x^*_s >> F^{-1}_{N_t}(\pi^{y_t})$ in the "easy to staff" periods 1, 4, 5, 7, 8, 10. In contrast, by using formulation EDetF, we do not pre-allocate the risk to the scheduling periods and let the optimization problem decide how to share it out so as to minimize the total cost. This means in particular that a level of risk higher than $1 - \pi^{\frac{1}{T}}$ might be accepted in "difficult to staff" periods while a much smaller level of risk (sometimes as low as $1 - \pi^{y_{min}}$) will be imposed in "easy to staff" periods. Thus, note how, in the optimal solution of EDetF displayed in Table 2, y^*_3, y^*_6 and y^*_9 are significantly greater than $\frac{1}{T} = 0.1$ whereas y^*_1, y^*_4, y^*_5, y^*_7, y^*_8 and y^*_{10} are significantly smaller than $\frac{1}{T} = 0.1$.

6. Numerical results

We carried out some computational experiments on real data coming from an anonymous health insurance company in order to evaluate the solution approach presented in Section 4 and to compare it with a scenario-based approach. The results of this computational study are then used to derive

Time slot	t	EDetB			EDetF		
		y_t	$F_{N_t}^{-1}(\pi^{y_t})$	$\sum_{s=1}^{S} A_{ts}x_s^*$	y_t^*	$F_{N_t}^{-1}(\pi^{y_t^*})$	$\sum_{s=1}^{S} A_{ts}x_s^*$
8:00- 9:00	1	0.1	77	110	0.0092	100	100
9:00-10:00	2	0.1	156	167	0.1544	150	150
10:00-11:00	3	0.1	167	167	0.2876	150	150
11:00-12:00	4	0.1	83	167	0.0001	119	150
12:00-13:00	5	0.1	34	57	0.0001	48	50
13:00-14:00	6	0.1	110	110	0.2513	100	100
14:00-15:00	7	0.1	152	220	0.0011	199	199
15:00-16:00	8	0.1	130	220	0.0001	187	199
16:00-17:00	9	0.1	110	110	0.2759	99	99
17:00-18:00	10	0.1	44	53	0.0202	49	49

Table 2: Illustrative example: optimal solutions of problems EDetB and EDetF

some managerial insights on the risk-cost trade-off in stochastic call center shift-scheduling.

6.1. Instances

To carry out our computational experiments, we generated 400 instances based on real data coming from an anonymous health insurance company. More precisely, the various instances tested have the following features.

The call center under study is open on Monday to Friday from 8:30 to 18:30 and on Saturday from 8:30 to 12:00. The scheduling horizon corresponds to a week and is divided into 30-min periods. It thus comprises $T = 107$ periods.

The expected service time is $\frac{1}{\mu} = 1$ minute and the mean time before abandonment is $\frac{1}{\gamma} = 1.25$ minute. The target quality of service is defined as the maximum allowed abandonment probability: $p^* = 5\%$.

Point forecasts $(\overline{\lambda}_1, ..., \overline{\lambda}_t, ...\overline{\lambda}_T)$ for the expected call arrival rates are provided for the scheduling horizon. They display an intra-day seasonality similar to the one featured by the illustrative example presented in Section 5.

We investigated 5 different cases:

1. the reference case provided by our case study where $\overline{\lambda}_t$ is comprised between 4 and 84 calls per minute,
2. a case where $\overline{\lambda}_t$ is divided by 2 in every period t,

3. a case where $\overline{\lambda}_t$ is multiplied by 2 in every period t,
4. a case where in each period t, $\overline{\lambda}_t$ is multiplied by a randomly generated coefficient belonging to $[0.75; 1.25]$,
5. a case where in each period t, $\overline{\lambda}_t$ is multiplied by a randomly generated coefficient belonging to $[0.50; 1.50]$.

The forecasting errors $(\epsilon_1, ..., \epsilon_t, ..., \epsilon_T)$ are modeled as independent random variables, each one following a normal distribution $\mathcal{N}(0, \sigma_t)$. The value of the standard deviation σ_t reflects the quality of the forecasts: the smaller the value of σ_t, the better and the more reliable the forecasts. We considered 4 levels for the quality of the forecasts: $\sigma_t = 0.5\lambda_t$, $\sigma_t = \lambda_t$, $\sigma_t = 1.5\lambda_t$, and $\sigma_t = 2\lambda_t$.

The call center under study uses a set of $S = 19$ predefined overlapping shifts. 5 shifts correspond to part-time positions working 21 hours a week, 4 shifts correspond to part-time positions working 28 hours a week and 10 shifts correspond to full-time positions working 39 hours a week. Each shift involves a one-hour lunch break per working day scheduled between 12:00 and 14:00. We used these data to generate a larger set of $S = 120$ shifts: these shifts correspond to part-time and full-time positions similar to the ones used in our case study, but with more flexibility to place half-days and/or days off within the week. In both cases, the cost of a shift s is set to the corresponding number of worked hours, i.e. $c_s = \sum_{t=1}^{T} A_{ts}$

For each expected call arrival rate profile, each forecast quality level and each set of predefined shifts, we generated 10 instances by varying the acceptable risk level $1 - \pi$ from 50% to 1% (i.e. π varies from 0.5 to 0.99).

6.2. Numerical assessment of the proposed solution approach

We use the solution approach presented in Section 4 (with the values of the parameters K_{max}, λ_{max}, $\Delta\lambda$, α_{max} and y_{min} provided in Section 5) to solve problems EDetB an EDetF.

The corresponding mixed-integer linear programs were solved using the mathematical programming solver CPLEX 12.6. We implemented our algorithms in C++, using the ILOG Concert Technology libraries as an interface with the solver. We used the solver default settings except for the relative MIP gap tolerance which we increased from 0.01% to 0.1%. Our preliminary computational experiments namely showed that this enabled to significantly decrease the computation time while providing near optimal solutions guaranteed to lie within 0.1% of the optimal value. All tests were carried out on

an Intel Core i5 (2.6 GHz) with 4 Go of RAM, running under Windows 7 [1].

The numerical results obtained on the 400 studied instances with formulation EDetB are provided in Table 3 while those obtained with formulation EDetF are provided in Tables 4-7. As the level of the forecasts quality, $\frac{\sigma_t}{\lambda_t}$, appears to have a strong impact on the computation time needed to solve problem EDetF, we chose to group the instances into 4 subsets, each one corresponding to a given value of $\frac{\sigma_t}{\lambda_t}$, and to display the corresponding results in 4 different tables.

We provide, for each set of instances:

- V and C: the average number of binary variables and constraints C involved in the MILP formulation,

- G and $maxG$: the average and maximum values (expressed as percentages) of the integrality gap, i.e. of the relative difference between the value of the linear relaxation of the problem and the value of an optimal integer solution,

- N and $maxN$: the average and maximum number of nodes explored by the Branch & Bound algorithm,

- CT and $maxCT$: the average and maximum computation time in seconds.

Results from Table 3 show that solving problem EDetB does not pose any numerical difficulties: all the considered instances could be solved in less than 30 seconds. This might be explained by the medium size of the corresponding mixed-integer linear programs (involving S integer variables and T constraints) and by the tightness of the integrality gap (0.25% on average).

In contrast, the mixed-integer linear programs corresponding to problem EDetF involve, in addition to the S integer variables and T quality-of-service constraints, a large number of binary variables z_{tm} and related constraints (39). However, results from Tables 4-7 show that, despite their size, these mixed-integer linear programs could be solved within a reasonable computation time for all the considered instances generated from our real-life

---

[1] All data related to our experiments (description of the instances, C++ source files and numerical results) are available upon request from the corresponding author.

π	0.99	0.98	0.97	0.96	0.95	0.90	0.80	0.70	0.60	0.50
V	70	70	70	70	70	70	70	70	70	70
C	107	107	107	107	107	107	107	107	107	107
G	0.20%	0.18%	0.20%	0.23%	0.21%	0.25%	0.26%	0.26%	0.20%	0.20%
$maxG$	0.67%	0.68%	0.54%	0.65%	0.48%	0.59%	0.93%	0.84%	1.08%	1.33%
N	0	0	0	0	0	0	0	0	0	0
$maxN$	0	0	0	0	0	0	0	0	0	0
CT	2.2	2.4	2.5	2.6	2.6	2.9	3.6	4.2	4.7	5.1
$maxCT$	10.2	11.4	12.4	12.2	12.9	15.0	18.5	20.7	24.7	28.0

Table 3: Resolution of problem EDetB - instances with $\frac{\sigma_t}{\lambda_t} \in \{0.5; 1; 1.5; 2\}$

π	0.99	0.98	0.97	0.96	0.95	0.90	0.80	0.70	0.60	0.50
V	6407	6747	6970	7144	7298	7832	8527	9064	9544	10018
C	6622	6962	7185	7359	7513	8047	8742	9279	9759	10233
G	0.06%	0.04%	0.05%	0.06%	0.06%	0.06%	0.06%	0.06%	0.06%	0.07%
$maxG$	0.25%	0.09%	0.14%	0.15%	0.13%	0.17%	0.12%	0.12%	0.12%	0.24%
N	139	0	1	8	1	21	10	641	2	84
$maxN$	1390	4	8	67	14	200	79	6387	9	680
CT	1.1	0.9	1.0	1.2	1.2	1.5	2.6	3.8	1.8	3.9
$maxCT$	3.6	2.3	2.3	2.3	4.7	3.7	10.3	12.0	2.2	10.4

Table 4: Resolution of problem EDetF - instances with $\frac{\sigma_t}{\lambda_t} = 0.5$

π	0.99	0.98	0.97	0.96	0.95	0.90	0.80	0.70	0.60	0.50
V	12710	13392	13840	14190	14471	15543	16945	18013	18979	19918
C	12925	13607	14055	14405	14686	15758	17160	18228	19194	20133
G	0.04%	0.06%	0.05%	0.06%	0.05%	0.06%	0.05%	0.06%	0.05%	0.06%
$maxG$	0.07%	0.12%	0.08%	0.10%	0.09%	0.09%	0.10%	0.10%	0.10%	0.12%
N	0	52	0	24	0	40	88	34	0	37
$maxN$	2	473	0	244	0	400	599	288	0	318
CT	4.7	13.1	4.4	4.6	4.6	14.3	16.9	13.5	7.1	17.7
$maxCT$	24.5	98.7	29.3	23.3	16.0	117.9	104.6	90.5	16.7	124.5

Table 5: Resolution of problem EDetF - instances with $\frac{\sigma_t}{\lambda_t} = 1$

π	0.99	0.98	0.97	0.96	0.95	0.90	0.80	0.70	0.60	0.50
V	19021	20030	20711	21232	21677	23283	25371	26971	28416	29821
C	19236	20245	20926	21447	21892	23498	25586	27186	28631	30036
G	0.03%	0.04%	0.02%	0.05%	0.06%	0.05%	0.04%	0.05%	0.06%	0.04%
$maxG$	0.05%	0.08%	0.04%	0.10%	0.10%	0.09%	0.10%	0.09%	0.08%	0.08%
N	0	63	42	26	0	46	72	83	0	58
$maxN$	4	627	413	259	0	448	367	647	0	578
CT	6.7	31.3	24.5	22.2	7.0	25.8	33.5	34.9	12.6	34.8
$maxCT$	21.0	258.8	196.6	152.5	17.1	188.1	171.9	214.2	29.6	245.6

Table 6: Resolution of problem EDetF - instances with $\frac{\sigma_t}{\lambda_t} = 1.5$

π	0.99	0.98	0.97	0.96	0.95	0.90	0.80	0.70	0.60	0.50
V	25340	26687	27591	28291	28880	30998	33797	35930	37861	39732
C	25555	26902	27806	28506	29095	31213	34012	36145	38076	39947
G	0.05%	0.06%	0.04%	0.03%	0.04%	0.04%	0.05%	0.05%	0.05%	0.05%
$maxG$	0.09%	0.07%	0.09%	0.10%	0.07%	0.09%	0.08%	0.08%	0.09%	0.08%
N	0	52	40	121	78	125	54	147	113	142
$maxN$	0	797	404	420	432	506	535	855	722	951
CT	7.8	13.1	23.9	50.0	41.2	66.2	40.4	95.0	87.7	97.6
$maxCT$	19.0	283.7	162.3	56.6	182.2	247.4	286.7	435.1	402.7	445.0

Table 7: Resolution of problem EDetF - instances with $\frac{\sigma_t}{\lambda_t} = 2$

27

case-study data. Indeed, the average computation time over all considered instances is less than 23s while its worst-case value is 445s (i.e. less than 8 minutes). This is mainly explained by the fact that the integrality gap of the MILP formulation (36)-(43) is nearly equal to 0% (0.05% on average). This means that the linear relaxation of the formulation provides an excellent approximation of the integer optimal solution value so that the Branch & Bound procedure embedded in CPLEX 12.6 has to explore a limited number of nodes of the search tree before finding a solution guaranteed to be within 0.1% of the optimal value.

Moreover, results from Tables 4-7 also show that two features seem to have a strong impact on the computation times, namely the forecast quality and the maximum acceptable risk level. The main reason for this is that they both have a significant influence on the MILP size, in terms of the number of involved binary variables z_{tm} and the number of the related constraints (39). Indeed, the poorer the forecast quality, the larger the standard deviation σ_t of the probability distributions of random variables N_t and the more steps $(M_t = F_{N_t}^{-1}(\pi^{y_{min}}) - F_{N_t}^{-1}(\pi))$ are involved in the description of functions Ψ_t. Similarly, the smaller the probability π, the smaller the value of $F_{N_t}^{-1}(\pi)$ and the more steps M_t are involved in the description of functions Ψ_t.

6.3. Comparison with a scenario-based solution approach

In order to further assess the proposed solution approach, we compare it with a scenario-based approach, namely the sample approximation approach presented in [19]. This approach aims at providing approximate solutions for joint chance-constrained programs. Its main idea consists in replacing the original continuous probability distribution of the random vector by an empirical discrete finite probability distribution obtained by Monte Carlo sampling. This approximation enables to reformulate the stochastic problem as a large-size mixed-integer linear program.

Let Λ^1, ..., Λ^i, ... Λ^I be a Monte Carlo sample of the random vector Λ. Each scenario i describes a possible realization of the call arrival rates over the scheduling horizon, i.e. $\Lambda^i = (\Lambda_1^i, ..., \Lambda_t^i, ..., \Lambda_T^i)$. The corresponding minimum numbers of agents $N^i = (N_1^i, ..., N_t^i, ..., N_T^i)$ can be computed in a pre-optimization step as $N_t^i = \phi_{\mu,\gamma,p^*}(\Lambda_t^i)$ thanks to the simple iterative procedure presented in Subsection 4.1. As the I sampled scenarios are independent and identically distributed observations of the vector Λ, the probability of each scenario i is assumed to be equal to $\frac{1}{I}$.

The sample approximation approach relies on the idea that the value of the probability featured in the joint chance-constraint (5) can be approximately computed by:

$$\Pr\left(\sum_{s=1}^{S} A_{ts}x_s \geq N_t, \forall t\right) \approx \frac{1}{I}\sum_{i=1}^{I}\mathbb{I}\left(\sum_{s=1}^{S} A_{ts}x_s \geq N_t^i, \ \forall t\right), \qquad (44)$$

where $\mathbb{I}(.)$ denotes the indicator function taking the value 1 when . is true and 0 otherwise.

By introducing a set of binary variables defined as $v^i = 1$ if at least one of the quality-of-service constraints is violated in scenario i and $v^i = 0$ otherwise, the approximation (44) enables to reformulate problem JCCP as the following mixed-integer linear program denoted by SA:

$$
\begin{cases}
Z = min \sum_{s=1}^{S} c_s x_s & (45) \\[2ex]
\sum_{s=1}^{S} A_{ts}x_s \geq N_t^i(1 - v^i) \qquad \forall t, \forall i & (46) \\[2ex]
\sum_{i=1}^{I} v^i \leq \lfloor I(1-\pi)\rfloor & (47) \\[2ex]
x_s \in \mathbb{Z}^+ \qquad\qquad\qquad \forall s & (48) \\[1ex]
v^i \in \{0,1\} \qquad\qquad\quad \forall t, \forall i. & (49)
\end{cases}
$$

Constraints (46) make sure that, if $v^i = 0$, the T quality-of-service constraints corresponding to scenario i are respected. Constraint (47) is the joint probability constraint: it limits the number of violated scenarios to $\lfloor I(1-\pi)\rfloor$, thus ensuring that the ratio $\frac{1}{I}\sum_{i=1}^{I}\mathbb{I}\left(\sum_{s=1}^{S} A_{ts}x_s \geq N_t^i \ \forall t\right)$ is above π.

We note that the formulation (45)-(49) involves I binary variables and TI big-M type constraints (46). Hence, in practice, for large sample size I, its resolution by a mathematical programming solver is likely to pose some computational difficulties. The authors of [20] investigated several alternative formulation strengthening techniques for this mixed-integer linear program and showed that the strong extended reformulation they proposed led to significantly reduced computation times. We thus used it in our computational experiments.

29

In order to compare the solution approach presented in Section 4 with the sample approximation approach, we carried out additional computational experiments on the 100 instances which were shown to be the most difficult to solve using formulation EDetF, i.e. on the instances corresponding to the level of forecast quality $\frac{\sigma_t}{\lambda_t} = 2$. We considered 4 possible values for the sample size: $I \in \{100, 500, 1000, 5000\}$. For each sample size, we randomly generated a sample assuming that each random variable Λ_t^i is normally distributed with mean λ_t and standard deviation σ_t. The mixed-integer linear programs corresponding to each solution approach were solved using CPLEX 12.6. We implemented our algorithms in C++, using the ILOG Concert Technology libraries as an interface with the solver. We used the solver default settings except for the relative MIP gap tolerance which we increased from 0.01% to 0.1%. All tests were carried out on an Intel Core i5 (2.6 GHz) with 4 Go of RAM, running under Windows 7.

The numerical results are displayed in Table 8. For each considered solution approach and each set of 10 instances, we provide:

- Ct: the average cost of the optimal integer solution.

- P: the average value of the probability $\Pr\left(\sum_{s=1}^{S} A_{ts} x_s^* \geq N_t, \forall t\right)$ where x_s^* denotes the optimal integer solution of the corresponding mixed-integer linear program. P is computed in a post-optimization step using: $P = \prod_{t=1}^{T} F_{N_t}(\sum_{s=1}^{S} A_{ts} x_s^*)$.

- CT: the average computation time in seconds.

Results from Table 8 show that, for the considered sample sizes, the sample approximation approach fails at providing feasible solutions of problem JCCP. Namely, we note that the average value of P is significantly below π for $I \in \{100, 500, 1000\}$ and that its value gets close to π only for the large sample size $I = 5000$. This means that the shift schedules x^* obtained through this approach are not feasible with respect to the joint chance-constraint (5). This can be explained by the fact that formulation SA relies on an approximate representation of this constraint. There is therefore no definite guarantee to it will provide a feasible solution to problem JCCP. The authors of [19] showed that, under certain conditions, the optimal value of SA converges to the optimal solution of JCCP with probability 1 when I tends to infinity. However, it seems that, for the problem under study here, the sample size required to get such a near-optimal solution of JCCP is too large

	π	0.99	0.98	0.97	0.96	0.95	0.90	0.80	0.70	0.60	0.50
EDetB	Ct	31726	30388	29543	28985	28506	26959	25246	24115	23125	22417
	P	0.998	0.996	0.995	0.993	0.991	0.979	0.955	0.927	0.893	0.848
	CT	2.8	2.3	3.2	3.3	3.3	3.9	4.7	5.5	6.2	6.7
EDetF	Ct	27339	25916	25052	24423	23923	22306	20541	19389	18481	17692
	P	0.990	0.980	0.970	0.960	0.950	0.900	0.801	0.701	0.602	0.502
	CT	7.8	13.1	23.9	50.0	41.2	66.2	40.4	95.0	87.7	97.6
SA100	Ct	23774	23088	22543	22159	21965	20783	19378	18346	17548	16810
	P	0.853	0.845	0.824	0.810	0.804	0.738	0.619	0.502	0.402	0.312
	CT	0.1	0.2	0.2	0.3	0.3	1.1	10.4	135.6	812.5	1334.5
SA500	Ct	26638	25000	24288	23623	23160	21708	20146	_	_	_
	P	0.963	0.946	0.935	0.916	0.908	0.849	0.746	_	_	_
	CT	0.3	0.7	1.1	1.4	2.6	40.3	1212.7	_	_	_
SA1000	Ct	26351	25172	24429	23862	23387	21850	_	_	_	_
	P	0.973	0.956	0.949	0.937	0.924	0.867	_	_	_	_
	CT	0.5	1.4	4.1	9.8	17.2	226.8	_	_	_	_
SA5000	Ct	26605	25501	24730	24157	23707	_	_	_	_	_
	P	0.985	0.974	0.963	0.953	0.943	_	_	_	_	_
	CT	7.1	19.8	47.2	170.0	252.5	_	_	_	_	_

"_"indicates that no guaranteed optimal solution could be found within one hour of computation.

Table 8: Comparison with the sample approximation approach - instances with $\frac{\sigma_t}{\lambda_t} = 2$

to allow a resolution of formulation SA within reasonable computation time, especially for the small values of π.

In contrast, the solution approach presented in Section 4 consistently provides feasible solutions of problem JCCP. Namely, for both formulations EDetB and EDetF, the average value of P is above π for every considered set of instances. However, we note that the shift schedules obtained with formulation EDetB are overly conservative. This can be seen by the fact that, in this case, the average value of P is significantly greater that π and that the cost of the corresponding optimal solutions is on average significantly larger than the cost of the optimal solutions provided by formulation EDetF. This shows the interest of using formulation EDetF to solve problem JCCP as it is capable of providing near-optimal feasible solutions of the problem within reasonable computation times, and this for all considered values of π.

6.4. Discussion and managerial insights

We now seek to derive from the results of our computational study some useful insights for call center managers faced with the problem of scheduling workforce under uncertain call arrival forecasts.

We first compare the two variants of the proposed solution approach: the one based on problem EDetB and the one based on formulation EDetF. Our numerical results show that the cost of the shift schedule obtained by solving problem EDetB is significantly higher that the cost of the shift schedule obtained by solving formulation EDetF. Namely, for the 400 considered instances, the total number of worked hours is reduced on average by 19% thanks to the use of the optimal sharing out of the risk between the scheduling periods carried out in problem EDetF. Furthermore, we note that this reduction is particularly significant (on average 23.5% for the corresponding 200 instances) when there is a small number of available shifts ($S = 19$), i.e. when there is a rather low flexibility in terms of workforce management. This clearly shows the benefit of using a variable level of risk in each scheduling period as it is done in the proposed model EDetF, especially when there are strong constraints on the workforce availability.

A second outcome of our computational study that could be of interest for call center managers is the possibility to quantify to some extent the risk-cost trade-off. We mentioned in subsection 3.2 that by setting the value of the maximum acceptable risk level $1 - \pi$, call center managers should seek to reach a trade-off between risk aversion and cost minimization. Setting small values for $1 - \pi$, i.e. adopting a risk-conservative approach, will lead to
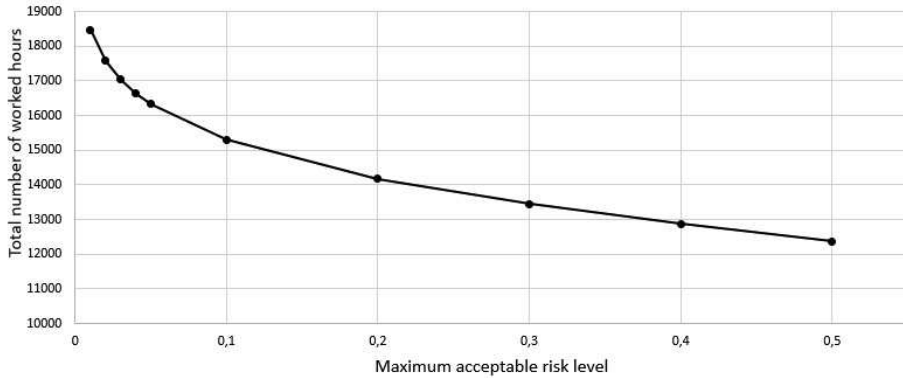
Figure 2: General risk-cost trade-off analysis

expensive shift schedules. On the contrary, increasing the value of $1-\pi$ might lead to significant cost savings. Thus, Figure 2 displays, for each maximum acceptable risk level $1-\pi$, the average value of the total number of worked hours for the corresponding 40 instances. We note in particular that, in the present case, decreasing the acceptable risk level from 50% to 20% leads to a relatively limited increase in the total number of worked hours (+14%). However, decreasing it from 50% to 1% leads to a much more significant cost increase (+49%). Providing call center managers with such a quantified representation of the risk-cost trade-off might help them decide upon the risk level that they are ready to accept.

Finally, our results also show the strong impact of forecast quality on this risk-cost trade-off. Thus, Figure 3 displays, for each forecast quality level (0.5, 1, 1.5 et 2) and each maximum acceptable risk level $1-\pi$, the average value of the total number of worked hours for the corresponding 10 instances. We note how, for a given maximum acceptable risk level $1-\pi$, the cost of the corresponding shift schedule is significantly decreased when the quality of the forecasts is improved. For instance, improving the forecast quality from $\sigma_t = 2\lambda_t$ to $\sigma_t = \lambda_t$ leads on average to a cost reduction of 42%. This information could serve as a useful incentive to invest in more powerful forecasting tools enabling call center managers to improve, as much as possible, the quality of the forecasts on future call arrival rates.
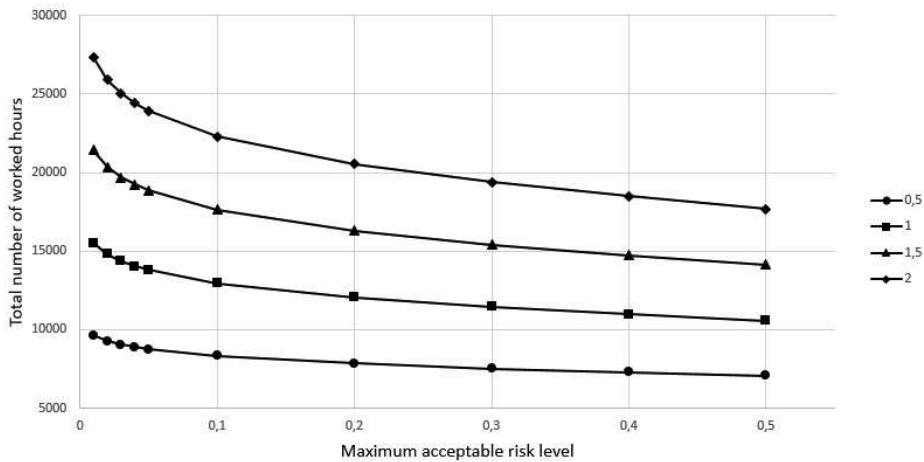
Figure 3: Impact of forecast quality on the risk-cost trade-off

7. Conclusion and research perspectives

We studied the shift scheduling problem for a single-class single-skill call center with impatient customers and focused on explicitly taking into account in the related optimization problem the impact of the uncertainties in the call arrival rates forecasts. We proposed to model this stochastic optimization problem using a joint chance-constrained program and presented a solution approach where the forecasting errors are represented as independent random variables, each one following a continuous probability distribution. Our computational results show that the proposed approach can efficiently solve real-size instances of the problem, enabling us to draw some managerial insights on the underlying risk-cost trade-off.

A first interesting direction for further research could consist in relaxing the assumption of statistical independence between the forecasting errors. Namely, the fact that there exist significant correlations between the call arrival volumes in different periods of the scheduling horizon has been highlighted in several works and forecasting models seeking to exploit this knowledge to improve the forecast quality have been proposed (see e.g. [10], [31] and [14]). Developing a solution approach capable of explicitly handling correlations between the forecasting errors in different periods could thus prove useful. Moreover, one might also question the assumption that over a finite-length scheduling period, the call center can be modeled as a queuing system in stationary state and that its performance can be measured based

34

on the long-term expected value of the service level. In their recent work, [29] namely show that, all other parameters being constant, there is a significant variability in the quality of service of a call center observed over a finite-length time interval and propose to model the quality of service as a random variable following a normal distribution. It might be worth proposing an extension of the present work where this additional source of uncertainty is accounted for.

[1] Aksin, O.Z., Armony, M., and Mehrotra, V. (2007). The modern call-center: a multi-disciplinary perspective on Operations Management research. *Production and Operations Management*, 16:665-688.

[2] Armony, M., Plambeck, M. and Seshadri, S. (2009). Sensitivity of optimal capacity to customer impatience in an unobservable M/M/S queue (Why you shouldn't shout at the DMV). *Manufacturing and Service Operations Management*, 11(1): 19-32.

[3] Beraldi P. and Ruszczynski A. (2002). A branch and bound method for stochastic integer problems under probabilistic constraints. *Optimization Methods and Software*, 17(4), 359-382.

[4] Bodur, M. and Luedtke, J. (2014). Integrated service system staffing and scheduling via stochastic integer programming. Working paper available at $http : //www.optimization - online.org/DB_FILE/2014/01/4200.pdf$.

[5] Branda, M. and Dupacová, J. (2012). Approximation and contamination bounds for probabilistic programs. *Annals of Operations Research*, 193: 3-19.

[6] Calafiore G. and Campi M.C. (2005). Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102: 25-46.

[7] Cheng, J. and Lisser, A. (2012) A second-order cone programming approach for linear programs with joint probabilistic constraints. *Operations Research Letters*, 40:325-328.

[8] Dentcheva, D., Prékopa, A. and Ruszczynski A. (2000). Concavity and efficient points of discrete distributions in probabilistic programming. *Mathematical Programming*, 89: 55-77.

[9] Gans, N., Koole, G., and Mandelbaum, A. (2003). Telephone call centers: tutorial, review, and research prospects. *Manufacturing and Service Operations Management*, 5:79-141.

[10] Gans, N., Shen, H., Zhou, Y.P., Korolev, N., McCord A. and Ristock H. (2015) Parametric stochastic programming models for call-center workforce scheduling. Accepted for publication in *Manufacturing and Service Operations Management*.

[11] Green, L.V, Kolesar, J.V. and Soares, J. (2001). Improving the SIPP approach for staffing service systems that have cyclic demands. *Operations Research*, 49(4):549-564.

[12] Gurvich, I., Luedtke, J. and Tezcan, T. (2010). Staffing call-centers with uncertain demand forecasts: a chance-constrained optimization approach. *Management Science*, 56(7):1093-1115.

[13] Helber, S. and Henken, K. (2010). Profit-oriented shift scheduling of inbound contact centers with skills-based routing, impatient customers, and retrials. *OR Spectrum*, 32:109-134.

[14] Ibrahim, R. and L'Ecuyer, P. (2013). Forecasting call center arrivals: fixed-effects, mixed-effects and bivariate effects. *Manufacturing and Service Operations Management*, 15(1):72-85

[15] Jouini, O. and Dallery, Y. (2007). Monotonicity properties for multi-server queues with reneging and finite waiting lines. *Probability in the Engineering and Informational Sciences*, 21:335-360.

[16] Lejeune, M. and Ruszczynski A. (2007). An efficient trajectory method for probabilistic production-inventory-distribution problems. *Operations Research*, 55(2): 378-394.

[17] Liao, S., Koole G., van Delft, C. and Jouini, O. (2012). Staffing a call center with uncertain non-stationary arrival rate and flexibility. *OR Spectrum*, 34:691-721.

[18] Liao, S., van Delft, C. and Vial, J.P. (2012). Distributionally robust workforce scheduling in call centers with uncertain arrival rates. *Optimization Methods and Software*, 28(3):501-522.

[19] Luedtke J. and Ahmed S. (2008). A sample approximation approach for optimization with probabilistic constraints, *SIAM Journal of Optimization*, 19(2): 674-699.

[20] Luedtke J., Ahmed S. and Nemhauser G.L. (2010). An integer programming approach for linear programs with probabilistic constraints, *Mathematical Programming*, 122: 247-272.

[21] Mandelbaum, A. and Zeltyn, S. (2009). The M/M/n+G queue: summary of performance measures. Technical note, Technion, Israel Institute of Technology.

[22] Mandelbaum, A. and Zeltyn, S. (2009). Staffing many-server queues with impatient customers: Constraint satisfaction in call centers. *Operations Research*, 57(5):11891205.

[23] Mehrotra, V., Ozluk, O., and Saltzman, R. (2010). Intelligent procedures for intra-day updating of call center agent schedules. *Production and Operations Management*, 19:353-367.

[24] Nemirovski A. and Shapiro A. (2005). Scenario approximations of chance constraints, *Probabilistic and Randomized Methods for Design Under Uncertainty*, G. Calafiore and F. Dabbene, eds., Springer, London, 3-48.

[25] Nemirovski A. and Shapiro A. (2006). Convex approximations of chance constrained programs, *SIAM Journal of Optimization*, 17: 969-996.

[26] Prékopa, A. Dual method for the solution of a one-stage stochastic programming problem with random RHS obeying a discrete probability distribution (1990). *Zeitschrift for Operations Research*, 34:441-461.

[27] Robbins, T.R. and Harrison, T.P. (2010). A stochastic programming model for scheduling call centers with global service level agreements. *European Journal of Operational Research*, 207:1608-1619.

[28] Rockafellar R. T. and Uryasev S. P. (2000). Optimization of conditional value-at-risk, *Journal of Risk*, 2: 21-41.

[29] Roubos, A., Koole, G. and Stolletz, R. (2012). Service level variability of inbound call centers. *Manufacturing and Service Operations Management*, 14(3): 402-413.

[30] Shapiro, A., Dentcheva, D. and Ruszczynski A. (2009). *Lectures on stochastic programming: modeling and theory*, MPS/SIAM Series on Optimization 9, Society for Industrial and Applied Mathematics, Philadelphia.

[31] Shen, H. and Huang, J.Z. (2008) Interday forecasting and intraday updating of call center arrivals. *Manufacturing and Service Operations Management*, 10(3): 391-410.

[32] Ye H., Luedtke, J. and Shen, H. (2014). Forecasting and staffing call centers with multiple interdependent uncertain arrival streams. Working paper.

Appendix

Proof of Proposition 1

Proof of Proposition 1 comprises two steps.

1. We show that ϕ_{μ,γ,p^*} is a non-decreasing function of λ.

Let λ_a, λ_b be two values of the call arrival rate such that $\lambda_a \leq \lambda_b$.
For given values of μ, γ and C, p is known to be an increasing function of λ, i.e. with all the other parameters fixed, the abandonment probability is known to be increasing in the arrival rate (see e.g. [12]).
This implies that:

$$p(\lambda_a, \mu, \gamma, l, C) \leq p(\lambda_b, \mu, \gamma, l, C) \quad \forall C \in \mathbb{N}. \tag{50}$$

Now, as with all the other parameters fixed, p is non-increasing in the number of agents C, we have:

$$min\{C \in \mathbb{N}|p(\lambda_a, \mu, \gamma, l, C) \leq p^*\} \leq min\{C \in \mathbb{N}|p(\lambda_b, \mu, \gamma, l, C) \leq p^*\}. \tag{51}$$

Hence:

$$\lambda_a \leq \lambda_b \Rightarrow \phi_{\mu,\gamma,p^*}(\lambda_a) \leq \phi_{\mu,\gamma,p^*}(\lambda_b). \tag{52}$$

2. We show that ϕ_{μ,γ,p^*} is a piecewise-constant function of λ.

Let $l \in \mathbb{N}$ be a number of agents. We define:
- $\tilde{\lambda}_l$, the value of the call arrival rate such that $p(\tilde{\lambda}_l, \mu, \gamma, l) = p^*$,
- $\tilde{\lambda}_{l-1}$, the value of the call arrival rate such that $p(\tilde{\lambda}_{l-1}, \mu, \gamma, l-1) = p^*$.
Note that $\tilde{\lambda}_{l-1}$ and $\tilde{\lambda}_l$ exist and are unique thanks to the fact that the probability of abandonment is continuous and strictly monotone in the arrival rate.
Moreover, with all the other parameters fixed, the abandonment probability p is known to be non-increasing in the number of agents.
This implies:

$$
\begin{aligned}
p(\tilde{\lambda}_l, \mu, \gamma, l-1) &\geq p(\tilde{\lambda}_l, \mu, \gamma, l) \\
p(\tilde{\lambda}_l, \mu, \gamma, l-1) &\geq p^* \\
p(\tilde{\lambda}_l, \mu, \gamma, l-1) &\geq p(\tilde{\lambda}_{l-1}, \mu, \gamma, l-1).
\end{aligned}
\tag{53}
$$

As, with all the other parameters fixed, the abandonment probability is known to be increasing in the arrival rate, we can deduce from inequality (53) that: $\tilde{\lambda}_{l-1} \leq \tilde{\lambda}_l$, i.e. that $[\tilde{\lambda}_{l-1}; \tilde{\lambda}_l[$ is a non-empty interval in $\mathbb{R}$.
We now show that ϕ_{μ,γ,p^*} is constant and equal to l over interval $]\tilde{\lambda}_{l-1}; \tilde{\lambda}_l]$.
We again use the fact that, with all the other parameters fixed, the abandonment probability is known to be increasing in the arrival rate. This gives:
- $\forall \lambda \leq \tilde{\lambda}_l$, $p(\lambda, \mu, \gamma, l) \leq p^*$, i.e. $\forall \lambda \leq \tilde{\lambda}_l$, $\phi_{\mu,\gamma,p^*}(\lambda) \leq l$,
- $\forall \lambda > \tilde{\lambda}_{l-1}$, $p(\lambda, \mu, \gamma, l-1) > p^*$, i.e. $\forall \lambda > \tilde{\lambda}_{l-1}$, $\phi_{\mu,\gamma,p^*}(\lambda) > l-1$.

This shows that: $\phi_{\mu,\gamma,p^*} = l, \forall]\tilde{\lambda}_{l-1}; \tilde{\lambda}_l]$.

Proof of Proposition 2

Function $\Psi_t : [y_t \mapsto F_{N_T}^{-1}(\pi^{y_t})]$ is a composite function of the form $F_{N_T}^{-1} \circ g$ where:

- g is the exponential function defined by $g : [y_t \mapsto \pi^{y_t}]$ with $\pi \in [0, 1[$. g is a strictly decreasing function of y_t.

- $F_{N_T}^{-1}$ is the inverse cumulated probability distribution of the discrete random variable N_t. It is thus is a non-decreasing piecewise constant function defined on the interval $[0; 1[$.

Function Ψ_t is thus a non-increasing piecewise-constant function of variables y_t defined on the interval $]0; 1]$.

39