Omega 133 (2025) 103262

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

Bed census prediction combining expert opinion and patient statistics ☆

Hayo Bos [a,b,*], Stef Baas [b,c], Richard J. Boucherie [b,c], Erwin W. Hans [b,d], Gréanne Leeftink [b,d]

[a] Diakonessenhuis Utrecht, P.O. Box 80250, 3508 TG Utrecht, The Netherlands
[b] Center for Healthcare Operations Improvement and Research, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
[c] Stochastic Operations Research, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
[d] Industrial Engineering and Business Information Systems, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

ARTICLE INFO

Keywords:
Operations Research in Health Services
Bed census distribution
Expected Discharge Date
Poisson binomial
Bayesian methods

ABSTRACT

Predictions of bed census are crucial for hospital capacity management choices, encompassing ward sizing, staffing, patient bed assignments, and surgical scheduling. Presently, these predictions heavily rely on doctors' estimated Expected Discharge Date (EDD). This paper introduces two probabilistic models that integrate EDD with Length of Stay (LoS) distributions derived from data. By employing the Poisson binomial distribution and probabilistic convolution, we generate full census distributions. Applying our approach to real hospital data demonstrates its ability to provide precise predictions, leading to valuable managerial insights.

1. Introduction

Hospital ward bed census is an important indicator for hospital capacity decision-making. Bed census data can, for instance, be used to dimension care units [1], or to evaluate ward capacity constraints when developing surgery schedules [2]. For day-to-day ward operations, ward managers could use operational bed census predictions as a workload proxy to determine the required staffing levels and prevent over- and under staffing [3]. A mismatch between bed supply and demand has adverse effects. Not having sufficient bed capacity could cause reduced quality of care and, after a certain tipping point, even increased patient mortality [4]. Having a surplus of bed capacity is inefficient as expensive resources are not optimally used [5]. Predicting the operational, day-to-day bed census characteristics and behaviour is crucial to match supply and demand, which ensures smooth hospital operations [6]. To position our contribution, we analyse existing work on bed census models along the lines of the strategic, tactical and operational decision levels [7].

Strategic bed census models are concerned with describing the overall stationary behaviour of the bed census. These type of bed census models are used for long-term decisions spanning a year or more. The models are often assumed to be stationary and typically do not consider any exogenous variables, and are based on, e.g., queueing [8] or approximations [9]. Strategic census models have different uses. Bravo et al. [10] use queueing models to devise system states in which there is an increased congestion risk for Intensive Care Units (ICUs). Strategic census models are also used to dimension wards [1] or admission lounges [11]. For tactical decisions, more granular models are required, because strategic models do not provide information on different types of seasonality.

Tactical level bed census models provide insight into weekly or monthly census patterns to facilitate tactical decisions, comprising, for instance, cyclic staff and surgical schedules. An example is the development of the Master Surgery Schedule (MSS), which is the (often times cyclic) assignment of specialties to Operating Rooms (ORs). Vanberkel et al. [12] are the first to provide a statistical convolution-based model, which explicitly models and evaluates the distribution of the daily resulting bed census, given an MSS. As the MSS is cyclic, so is their resulting bed census distribution. Kortbeek et al. [13] extends the model of Vanberkel et al. [12] by incorporating bed census of emergency patients. Tactical models are useful for tactical decisions such as nurse staffing [3,14], but could provide bad predictions on the operational level in case of unusual peaks or drops in bed census.

Operational level bed census models provide day-to-day census predictions using the available data on today's bed census. These predictions are necessary, as unexpected peaks in patient numbers could cause, e.g., surgery cancellations [15], the dilemma of the last bed [16] or overcrowding of the Emergency Department (ED) [4]. Having predictions allows managers to take precautionary measures preventing these adverse affects, such as creating more capacity by the temporary deployment of float or agency nurses, or tempering the

☆ Area: Data-Driven Analytics. This manuscript was processed by Associate Editor M. Zhao.

* Corresponding author at: Center for Healthcare Operations Improvement and Research, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
E-mail address: h.bos-1@utwente.nl (H. Bos).

https://doi.org/10.1016/j.omega.2024.103262
Received 23 April 2024; Accepted 11 December 2024
Available online 18 December 2024

inflow by asking neighbouring hospitals to admit more patients. If a decrease in patient numbers is anticipated in the near future, managers can permit staff to take leave, and thereby temporarily reduce capacity.

Operational bed census models have been previously proposed in the literature. Heins et al. [17] propose a two-step approach, consisting of a regression model for arrivals and a simulation model to obtain the resulting bed census. Tello et al. [18] combine K-means clustering with support vector machines to provide short-term bed census predictions. A key difference between the operational level and the other levels is that information about the patients currently residing at the ward is available, which could be valuable for predictions. While the models presented in the former two papers provide point estimates, Davis and Fard [19] provide a full description of the bed census distribution based on the attained LoS per patient. This is referred to as probabilistic forecasting [20], and has gained attention as it facilitates risk quantification of capacity decisions. Examples of such decisions include decisions about patient referrals [21]. Baas et al. [22],Dijkstra et al. [23] present a method to forecast COVID-19 bed census at the ward and ICU based on an estimated queueing network, a prediction of the arrival process in the coming days, and a (residual) LoS distribution estimated using survival analysis. Bekker et al. [24] provide a bed census model, which uses linear programming to predict admissions, and fitted residual LoS and queueing theory to convert this into predictions. While providing the full bed census distribution, the aforementioned models only use the attained LoS to predict the residual LoS, and typically assume identically distributed LoSs for all patient types. However, it is not hard to imagine that other patient characteristics could provide information as about the residual LoS, such as age, specialty, disease, gender, and the recovery progress [25,26].

The drawback of incorporating this type of information is that it is privacy-sensitive, as also noted by Tello et al. [18]. Besides, there might be a delay in data availability, as some patient characteristics can be missing or non-available at the time a forecast is to be made. How to deal with missing or delayed data in such models is not straightforward and poses an additional challenge in the modelling part, and its justification requires additional communication with the hospital. Models that take many patient characteristics into account typically need a lot of data to be trained and require the decision maker to select the best model among many [27]. To overcome these issues, we thus are in need for a proxy for this information, circumventing both the privacy and the modelling challenges.

The EDD could be such a proxy. The EDD is the doctor's best guess of the patient's discharge date [28–30]. It is common practice that doctors provide an estimate for the EDD at the start of the patient's admission and update it throughout the patient's stay in the hospital [28]. The EDD is particularly accurate close to the actual discharge date [29]. Doctors make this estimate based on their expert knowledge and patient characteristics [28], thus providing the required proxy for privacy sensitive information. An additional benefit of using the EDD is that both the residual LoS and the EDD have time as their unit of measurement, making it easy to directly relate these quantities. This inquires models that relate the residual LoS and the EDD to predict the bed census.

Summarising, we underline the importance of bed census prediction models. We identify a gap at the operational level, as existing models do not use all available information, such as the EDD. In this work, we therefore propose a probabilistic method that provides the full bed census distribution. The model is easy to interpret, can be explained to practitioners, and allows for risk quantification. We incorporate the EDD to dynamically adjust bed census forecasts. We propose two models for the residual LoS. The *mixture model* linearly combines the conditional LoS distribution and EDD, where the weight is optimised using Maximum Likelihood Estimation (MLE). As an alternative, we propose a *weighted model*, where the dependence of the prediction on the EDD is automatically weighted by historical deviations between the EDD and the observed LoSs. Building on existing literature, we prove that the resulting census predictions follow a Poisson binomial distribution. Finally, we provide a framework for selecting the appropriate model using simulated data and an easy-to-interpret SVM classifier, ensuring optimal application based on specific hospital data characteristics. We show that the resulting bed census distribution accurately predict bed census on real world hospital data.

This paper is organised as follows. Section 2 introduces our model. Section 3 presents numerical results based on simulated data and a case study. Section 4 discusses managerial insights, and Section 5 provides the overall conclusion where we reflect upon our work and identify directions for further research.

2. Bed census prediction

We start in Section 2.1 with the basis for our census prediction model. In Section 2.2, we provide the theorems to obtain the complete bed census distribution. We then provide two models for the residual LoS. Section 2.3 presents the first model, which uses a mixture distribution, and presents a Newton–Raphson-based algorithm to obtain the MLE estimate for the mixture parameter. Section 2.4 presents the weighted model for the residual LoS.

2.1. Census prediction model

Consider a clinical ward with patients being admitted and discharged over time, for which we know the number of currently present patients. Our goal is to determine the distribution of the number of patients present for the coming days. Let $t = \dots, -1, 0, 1, \dots$ denote discrete time. Without loss of generality, let $t = 0$ be the current time. Let T denote the prediction horizon. We assume ample ward capacity, to allow for a predicted census higher than the ward capacity. Our goal is to obtain the distribution of the number of patients present at time t, $t = 1, \dots, T$. Let L denote the random LoS of a patient. We assume that for all patients, L is independent and identically distributed with finite support $\ell = 1, 2, \dots, \overline{L}$, where $\overline{L}$ denotes the maximum LoS.

Let random variable $N(t)$ denote the total number of patients present at time t. Let $X_0(t)$ denote the number of patients from $N(0)$ that are still residing at the ward at time t, with $X_0(0) = N(0)$. Let $Y_s(t)$ denote the number of patients who enter at time s, who are still present at time t. Then $N(t)$ can be obtained by summing over these quantities, i.e.,

$$N(t) = X_0(t) + \sum_{s=0}^{t} Y_s(t). \tag{1}$$

The distribution of $N(t)$ can be determined by convolving the distributions of $X_0(t)$ and $Y_0(t), \dots, Y_t(t)$. To obtain $X_0(t)$, the residual LoS is necessary. We are considering a patient admitted to the ward at time $t = 0$. Let η represent the attained LoS. τ denotes the doctor's estimate of the expected discharge date (EDD) at time t, which is the doctor's estimation of the residual LoS. A model for the residual length of stay should be dependent on η, ensuring that instances close to τ are more probable if the doctor's estimate is accurate. We will proceed by obtaining the distribution of $X_0(t)$ and $Y_s(t), s = 0, \dots, t$, and subsequently $N(t)$.

2.2. The distribution of $X_0(t)$ and $Y_s(t)$

To derive the distribution of $X_0(t)$, we first introduce the random residual LoS R, with $\mathbb{P}(R = r)$. Without further specifying the distribution of R, we obtain the following result which provides us with the distribution of $X_0(t)$.

Theorem 1. *Let $N(0)$ be the number of patients present at time 0. Assume that each patient's LoS is independent and identically distributed. Let R_i denote the residual LoS of patient i. Then the number of patients still present at time t, $X_0(t)$, has a Poisson binomial distribution with success probabilities $\sum_{s=t}^{\infty} \mathbb{P}(R_i = s)$, $i = 1, \dots, N(0)$.*

Proof. Consider a fixed patient who is at the ward for η time units, with an EDD of τ. The probability that this patient stays for exactly another t days is given by $\mathbb{P}(R(\eta,\tau)=r)$. Hence, if we define a success event as a patient staying for *at least* t days, then this is a Bernoulli trial with success probability $\sum_{s=t}^{\infty}\mathbb{P}(R=s)$.

Let $S\subseteq\{1,\dots,N(0)\}$ denote a size-n subset of patients, and let S_n denote the set of all subsets of n patients selected from $\{1,\dots,N(0)\}$. Let $p_i(s)=\mathbb{P}(R_i=s)$. For the patients in S, we obtain the joint probability of all patients staying at least another t days by

$$\prod_{i\in S}\sum_{s=t}^{\infty}p_i(s)\prod_{j\in S^c}(1-\sum_{s=t}^{\infty}p_j(s)). \tag{2}$$

To obtain the probability that there are n out of $N(0)$ patients remaining at time t, we have to calculate (2) for all size-n subsets of $\{1,\dots,N(0)\}$, and take the sum, i.e.,

$$\mathbb{P}(X_0(t)=n)=\sum_{S\in S_n}\left(\prod_{i\in S}\sum_{s=t}^{\infty}p_i(s)\right)\left(\prod_{j\in S^c}(1-\sum_{s=t}^{\infty}p_j(s))\right), \tag{3}$$

which is the probability density function of a Poisson binomial distribution, with success probabilities $\sum_{s=t}^{\infty}p_i(s),\ i=1,\dots,N(0)$. □

The next step is to obtain the distribution of $Y_s(t)$. The following result is readily obtained from ([13], Sections A.1 and B.1).

Theorem 2. *Let $t=0$ denote the current time. Assume that $Y_0(0)=0$. Let $A(t)$ denote the number of patient arrivals at time t, and let L denote the random LoS of patients. For notational convenience, let $\tilde{p}_t=\mathbb{P}(L=t|L\geq t-1)$. Let $\mathbb{1}(\cdot)$ denote the indicator function. Then we have:*

$$\mathbb{P}(Y_s(t)=n)=\begin{cases}\mathbb{1}(n=0), & \text{if } s>t,\\ \mathbb{P}(A(t)=n), & \text{if } s=t,\\ \sum_{k=n}^{\infty}\binom{k}{n}\tilde{p}_t^{k-n}\left(1-\tilde{p}_t\right)^n\mathbb{P}(Y_s(t-1)=k), & \text{if } s<t.\end{cases} \tag{4}$$

By exploiting this theorem, we can account for the distinct characteristics of planned and emergency patients as follows. For emergency patients, we typically assume Poisson arrivals, implying $\mathbb{P}(A(t)=n)=\frac{e^{-\lambda(t)}\lambda(t)^n}{n!}$. Planned arrivals could be modelled as a deterministic arrival process. For example, if there are c planned patients on day t, we have $\mathbb{P}(A(t)=c)=1$ and $\mathbb{P}(A(t)=n)=0,\ n\neq c$. The next corollary presents the main result of this section.

Corollary 1. *Let $Y_s^A(t)$ and $Y_s^P(t)$ denote the number of acute and planned patients respectively, that arrived at time s which are still present at time t. Combining Theorems 1 and 2, we obtain that*

$$N(t)=X_0(t)+\sum_{s=0}^{t}\left(Y_s^A(t)+Y_s^P(t)\right), \tag{5}$$

of which the distribution can be obtained by probabilistic convolution.

We will proceed by presenting two such residual LoS models, the mixture model $\mathbb{P}_\alpha^{\mathrm{MM}}$ and the weighted model $\mathbb{P}_\beta^{\mathrm{WM}}$. The result presented in Corollary 1 requires a single specification of the residual LoS distribution. To combine $\mathbb{P}_\alpha^{\mathrm{MM}}$ and $\mathbb{P}_\beta^{\mathrm{WM}}$, we use a machine learning based classifier to choose between $\mathbb{P}_\alpha^{\mathrm{MM}}$ and $\mathbb{P}_\beta^{\mathrm{WM}}$ per patient type. We illustrate this approach using a support vector machine classifier and simulated data in Section 3.

2.3. Mixture model

We now assume that the residual LoS depends on the attained LoS η and the EDD τ, i.e., let $R(\eta,\tau)$ denote the residual LoS. In the mixture model, we assume that the EDD is perfectly estimated with probability α, or that τ provides no estimation, i.e., $R(\eta,\tau)$ is equal to LoS $L-\eta\mid L\geq\eta$, with probability $(1-\alpha)$. In this model, α represents our tuning parameter, where high α means high EDD estimation accuracy. We obtain

$$R(\eta,\tau)=\begin{cases}\tau, & \text{w.p. } \alpha,\\ L\mid L\geq\eta, & \text{w.p. } (1-\alpha).\end{cases} \tag{6}$$

The probability $\mathbb{P}_\alpha^{\mathrm{MM}}(R(\eta,\tau)=r)$ is calculated as

$$\begin{aligned}\mathbb{P}_\alpha^{\mathrm{MM}}(R(\eta,\tau)=r)&=\mathbb{P}_\alpha^{\mathrm{MM}}(L=r+\eta|L\geq\eta,\tau)\\&=\alpha\mathbb{1}(r=\tau)+(1-\alpha)\mathbb{P}(L=r+\eta|L\geq\eta)\\&=\alpha\mathbb{1}(r=\tau)+(1-\alpha)\frac{\mathbb{P}(L=r+\eta)}{1-\mathbb{P}(L\leq\eta-1)},\end{aligned} \tag{7}$$

where the third equality follows the definition of conditional probability. If α is high, $R(\eta,\tau)$ will have more probability mass on τ. If α is low, the probability distribution will hardly be influenced by the EDD. Note that $\alpha=1$ corresponds to the case of only using the EDD, which is the common practice in Dutch hospitals up to date. The case of $\alpha=0$ corresponds to only using the LoS distributions, which agrees with the model proposed in [19]. We estimate α using MLE.

Suppose we have a size-N sample of patients, indexed $i=1,\dots,N$, for which we know the realised residual LoS t_i, their attained LoS η_i, and their EDD τ_i. Let $X=(t_i,\eta_i,\tau_i)_{i=1}^N$. The log-likelihood $\ell(\alpha|X)$ of patients is given by

$$\begin{aligned}\ell(\alpha|X)&=\sum_{i=1}^{N}\log\mathbb{P}(R(\eta_i,\tau_i)=r_i)\\&\overset{(3)}{=}\sum_{i=1}^{N}\log\left(\alpha\mathbb{1}(t_i=\tau_i)+(1-\alpha)\frac{\mathbb{P}(L=t_i+\eta_i)}{1-\mathbb{P}(L\leq\eta_i-1)}\right).\end{aligned} \tag{8}$$

The MLE estimate of α is then obtained by

$$\hat{\alpha}=\arg\max_{\alpha\in[0,1]}\ell(\alpha|X). \tag{9}$$

The following theorem shows that $\ell(\alpha|X)$ is concave, which allows evaluation of the arg max in (9).

Theorem 3. *$\ell(\alpha|X)$ is concave on $\alpha\in[0,1]$.*

Proof. Observe that

$$g_i(\alpha|X)=\log\left(\alpha\mathbb{1}(t_i=\tau_i)+(1-\alpha)\frac{\mathbb{P}(L=t_i+\eta_i)}{1-\mathbb{P}(L\leq\eta_i-1)}\right)$$

is the logarithm of a non-negative linear function $c_1\alpha+c_2$ with constants c_1,c_2, and $\alpha\in[0,1]$, and hence concave on $\alpha\in[0,1]$. The result now follows as $\ell(\alpha|X)=\sum_{i=1}^{N}g_i(\alpha|X)$. □

Directly optimising $\ell(\alpha|X)$ using its first order conditions is hard, as $\frac{\partial}{\partial\alpha}\ell(\alpha|X)$ sums N fractional terms with α in the denominator. Therefore, we maximise $\ell(\alpha|X)$ over $\alpha\in[0,1]$ using the Newton–Raphson algorithm ([31], Chapter 5), presented here in Algorithm 1. As we are maximising $\ell(\alpha|X)$ for $\alpha\in[0,1]$ for which $\ell(\alpha|X)$ is concave, the algorithm is guaranteed to converge to the maximiser for $\alpha\in[0,1]$, up to a tolerance of $\epsilon>0$.

Algorithm 1 Newton–Raphson algorithm to find the maximiser $\ell(\alpha|X)$.

input An initial value for $\hat{\alpha}$, N samples $x_1,\dots,x_N$, convergence threshold ϵ.
output An MLE estimate $\hat{\alpha}$.
initialise $\hat{\alpha}\leftarrow 1,\hat{\alpha}_{\mathrm{old}}\leftarrow 0$
while $\hat{\alpha}-\hat{\alpha}_{\mathrm{old}}>\epsilon$ **do**
  $\hat{\alpha}_{\mathrm{old}}\leftarrow\hat{\alpha}$
  $\hat{\alpha}\leftarrow\hat{\alpha}_{\mathrm{old}}-\frac{\frac{\partial}{\partial\alpha}\ell(\alpha|X)}{\frac{\partial^2}{\partial^2\alpha}\ell(\alpha|X)}$
  if $\hat{\alpha}>1$ **then** $\hat{\alpha}\leftarrow 1$
  if $\hat{\alpha}<0$ **then** $\hat{\alpha}\leftarrow 0$
end

2.4. Weighted model

A drawback of the mixture model (6) is that the EDD t only increases the probability mass of outcome $t = \tau$. In reality, also outcomes close to τ should be more likely. Reasoning along those lines, we propose an alternative probability model $\mathbb{P}^{\text{WM}}_{\beta}$ for $R(\eta, \tau)$ that is a weighted version of the conditional LoS $L|L \geq \eta$, where the weights are obtained by the pdf of a Gaussian distribution with mean τ and variance β, i.e.,

$$\begin{aligned}\mathbb{P}^{\text{WM}}_{\beta}(R(\eta,\tau)=r) &= \mathbb{P}^{\text{WM}}_{\beta}(L=t|L\geq\eta,\tau)\\ &= C^{-1}\mathbb{P}(L=t|L\geq\eta)\frac{1}{\sqrt{2\pi\beta}}\exp\left\{-\frac{1}{2}\frac{(t-\tau)^2}{\beta}\right\},\end{aligned} \tag{10}$$

with C being the normalisation constant defined as

$$C=\sum_{t=0}^{\overline{L}-\eta}\mathbb{P}(L=t|L\geq\eta)\frac{1}{\sqrt{2\pi\beta}}\exp\left\{-\frac{1}{2}\frac{(t-\tau)^2}{\beta}\right\}. \tag{11}$$

The variance β of the Gaussian in (11) should be large if, based on realisations of the residual LoS, the estimates of the EDD are inaccurate and vice versa if the estimates are accurate. We proceed with deriving β by using Bayesian estimation. Let $p(\beta)$ denote the pdf of the prior distribution, and $p(\beta|t;\eta,\tau)$ the posterior pdf. As we do not have any prior information about β, we choose the Jeffreys prior [32], i.e., $p(\beta) \propto \frac{1}{\beta}$, where $\propto$ denotes the proportionality operator. By Bayes' rule, we have

$$\begin{aligned}p(\beta|t,\eta,\tau) &= \left(\prod_{i=1}^{N}\mathbb{P}^{\text{WM}}_{\beta}(R(\eta_i,\tau_i)=r)\right)p(\beta)\\ &\propto\left(\prod_{i=1}^{N}C^{-1}\mathbb{P}(L=t_i|L\geq\eta_i)\frac{1}{\sqrt{2\pi\beta}}\exp\left\{-\frac{1}{2}\frac{(t_i-\tau_i)^2}{\beta}\right\}\right)\frac{1}{\beta}\\ &\propto\left(\frac{1}{\beta}\right)^{\frac{1}{2}N+1}\exp\left\{\frac{-\frac{1}{2}\sum_{i=1}^{N}(t_i-\tau_i)^2}{\beta}\right\},\end{aligned} \tag{12}$$

from which we obtain that the posterior distribution of β is an inverse Gamma distribution with shape parameter $\alpha = \frac{1}{2}N$ and scale parameter $\delta = \frac{1}{2}\sum_{i=1}^{N}(t_i-\tau_i)^2$. The Maximum A Posteriori Estimator (MAPE) of β, denoted by $\hat{\beta}$ is given by the mode of the posterior distribution, i.e.,

$$\hat{\beta}=\frac{\delta}{\alpha+1}=\frac{\frac{1}{2}\sum_{i=1}^{N}(t_i-\tau_i)^2}{\frac{1}{2}N+1}. \tag{13}$$

If β is small, the probability mass of $R(\eta,\tau)$ will be high around τ, which corresponds to the case of having accurate EDD estimates. If β is large, the probability mass will be similar to $\mathbb{P}(L|L\geq\eta)$, i.e., the case in which the EDD is inaccurate.

3. Numerical results

This section presents numerical results to validate our model, which can be split into two parts. The results in Section 3.1 are based on generated data, and serve to investigate under what conditions either the mixture model or weighted model performs better to predict $X_0(t)$ per patient type. Section 3.2 presents results based on real-life data, by predicting next week's census for a medium-sized hospital in The Netherlands. All results are obtained using Python 3.9 on a notebook computer with a 10-core M1 CPU and 16 GB of RAM.

3.1. Comparison of the mixture and weighted model using simulated data

In Section 2 we presented two models for $R(\eta,\tau)$ in order to predict $X_0(t)$, the mixture model in Section 2.3 and the weighted model in Section 2.4. To provide insights into the performance of both models, we generated simulated data sets to compare both models. In Section 3.1.1, we provide the parameters for our simulation in Section 3.1.2.

3.1.1. Parameter settings

We assume one ward with $N(0) = 100$ patients present at time $t = 0$. Our prediction horizon is $T = 7$, i.e., we are concerned with predicting $X_0(1),\ldots,X_0(7)$. We assume that each patient's LoS L is i.i.d., and follows a truncated geometric distribution [33] with parameter $p = 0.1$ truncated at the 98th percentile, implying that $\mathbb{E}(L) \approx 10$. We sample the attained η_i from a discrete uniform distribution on $[0, L_i]$. From those two quantities, we obtain the residual LoS $R_i = L_i - \eta_i, i = 1,\ldots,100$.

Sampling the EDD τ_i is more involved. We hypothesise that the mixture model would do well if the EDD is estimated correctly with a high probability. Contrarily, if the EDD is wrong but still close to the real discharge date, it is likely that the weighted model performs better. Our generated data thus needs to incorporate both effects and be spread in terms of precision and variance of the errors of the EDD. To this end, we introduce two parameters $\rho \in \{0, 0.2, \ldots, 1\}$ and $\gamma \in \{0.1, 0.25, 0.5, \ldots, 2.5\}$ to control accuracy and variance. If $R_i = 0$, we let $\tau_i = 0$, because doctors usually know with certainty who will be discharged today. If $R_i \neq 0$, we let $\tau_i = R_i$ with probability $\rho - \frac{1}{100}\sum_{i=1}^{100}\mathbb{1}(R_i = 0)$; we sample uniformly from $\{\max(\lfloor(1-\gamma R_i)\rfloor, 0), \ldots, \lceil(1+\gamma)R_i\rceil\} \setminus \{R_i\}$ otherwise. Following this procedure, the fraction of times in which the doctor is able to accurately predict the EDD is controlled by ρ, and the variance of the EDD by γ. We also ensure that the EDD gets more accurate for smaller values of R_i, by letting the lower and upper bound of the uniform distribution depend on R_i. The procedure is summarised in Algorithm 2.

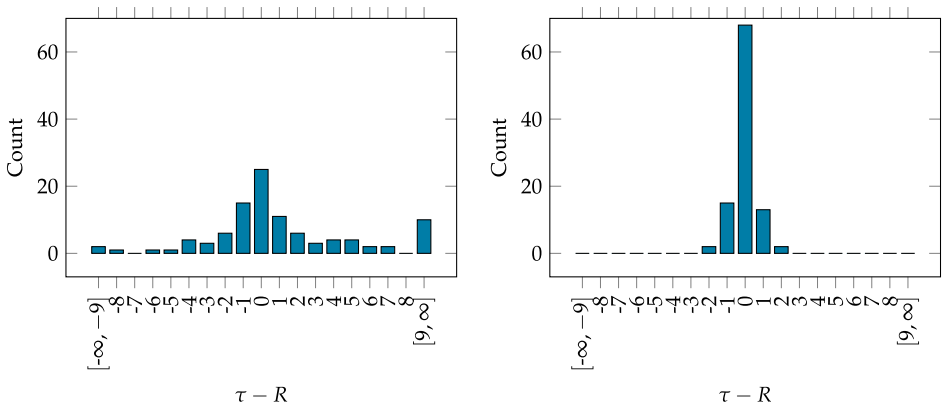
Algorithm 2 Sampling the EDD of a patient.

1: **input** $\rho, \gamma, R_1, \ldots, R_{N(0)}$, a patient i, a random $q \in [0, 1]$.
2: **initialise** $p_0 = \frac{1}{N(0)}\sum_{i=1}^{N(0)}\mathbb{1}(R_i = 0)$.
3: **output** The EDD of patient i, τ_i.
4: **if** $R_i = 0$ **then** $\tau_i = 0$
5: **if** $q \leq \rho - p_0$ **then** $\tau_i = R_i$
6: **else** Sample τ_i from $\mathcal{U}\left\{\max(\lfloor(1-\gamma R_i)\rfloor, 0), \ldots, \lceil(1+\gamma)R_i\rceil\right\} \setminus \{R_i\}$.
7: **end**

Fig. 1 depicts the distribution of the errors of the EDDs, $\tau - R$, generated by Algorithm 2, both for 100 patients for two sets of values for ρ and γ. In Fig. 1(a), for $\rho = 0.25$ and $\gamma = 2$, a fraction of 25% of the EDDs was estimated correctly, and 50% of the incorrect EDDs deviated at most 2 time units. In Fig. 1(b), for $\rho = 0.8$ and $\gamma = 0.25$, 68% was estimated correctly and all of the incorrect EDDs deviated at most 2 time units. This shows that using Algorithm 2, we can generate data corresponding to highly accurate and less accurate EDD estimations.

3.1.2. Comparing both models

To evaluate the two models of $R(\eta,\tau)$, we first draw $L_1,\ldots,L_{300}$ and $\eta_1,\ldots,\eta_{300}$. For each pair of γ and ρ we sample $\tau_1,\ldots,\tau_{300}$. We then obtain $\hat{\alpha}$ following Algorithm 1, and $\hat{\beta}$ using (13). We subsequently evaluate (7) and (10). To obtain predictions for $X_0(t)$, we need to solve Eq. (3), i.e., we need to compute a Poisson binomial distribution. It is notoriously hard to compute this distribution, as the number of subsets S_n over which we calculate the outer sum equals $\binom{N(0)}{n}$. To overcome this computational difficulty, we use an efficient Python implementation [34] of the Discrete Fourier Transform based method as presented in [35]. Using their implementation, the Poisson binomial distribution for $N(0)$ up to 5000 can be obtained in less than a second using the mentioned hardware. This procedure is summarised in Algorithm 3.



(a) Errors of the EDD for $\rho = 0.25$ and $\gamma = 2$. **(b)** Errors of the EDD for $\rho = 0.8$ and $\gamma = 0.25$.

Fig. 1. Histograms of the errors of the EDDs generated by Algorithm 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Algorithm 3 Algorithm to obtain the distribution of $X_0(t)$ for various γ and ρ

Output the distribution of $X_0(t)$
sample $L_1, \dots, L_{300}$, $\eta_1, \dots, \eta_{300}$
for $\gamma \in \{0, 0.25, 0.5, \dots, 2.5\}$ **do**
  for $\rho \in \{0, 0.1, \dots, 1\}$ **do**
    for $i \in \{1, \dots, 5\}$ **do**
      Sample $\tau_1, \dots, \tau_{100}$ using Algorithm 2.
      Obtain α following Algorithm 1, and β using (13) .
      Obtain $\mathbb{P}_{\alpha}^{\text{MM}}(R(\eta, \tau) = r)$ and $\mathbb{P}_{\beta}^{\text{WM}}(R(\eta, \tau) = t)$ using (7) and (10).
      Obtain two distributions of $X_0(t)$ using [34].
end

To compare the two models, we calculate the Mean Squared Error (MSE) and Mean Absolute Error (MAE) between the median of the obtained distributions of $X_0(t)$, and compare this with simulated realisations. The MAE represents by how many beds the predictions were off on average. The MSE squares the errors and therefore puts more weight on larger error values, making it more sensitive to large prediction errors. A high MSE means that some predictions were off by a lot. The realisations of $X_0(1), \dots, X_0(7)$ are obtained by counting the number of patients with $R_i \geq t$ for each $t = 1, \dots, 7$. In a real-world case, the underlying distribution of the residual LoS is unknown, let alone the parameters ρ and γ. The conditions under which either of the two models performs better therefore have to be expressed in terms of α and β, as those quantities can be obtained before deciding on which method to use.

Following Algorithm 3, we have $13 \times 11 \times 5 = 605$ data points for which we can compare the MSE for both models. We proceed with classification, wherein each data point is categorised based on the percentage difference in MSE. If this difference exceeds 15%, the data point is labelled as either "Mixture" or "Bayes", determined by which of the two has a lower MSE. Otherwise, it is labelled as "Indifferent". We then utilise a SVM classifier to find the decision boundaries that shows us for what values of α and β the weighted or mixture model is better, and for which values the SVM classifier is indifferent.

Fig. 2 shows the decision boundaries. The plotting area is segmented into blue, orange, and grey sections, representing the values (α, β) where either the weighted or mixture model excels over the other, or where the SVM classifier is indifferent. The scattered points are newly generated unseen test data. The weighted model tends to work well for low values of α and β. These values indicate that the EDDs are often wrong (low α), but that the errors are small (low β). This matches our intuition for the weighted model, as (10) reshapes the residual LoS distribution as contrary to the mixture model, it assigns more probability mass to both the EDD and realisations close to the EDD. The mixture model outperforms the weighted model if $\alpha > 0.5$, and for lower values of α if β is large. This occurs when the EDD is often accurate; otherwise, if incorrect, the error is significant. Both models perform equally well in the grey area, where α is small and β is large, which happens when the EDD does not tell much about the realisation of the residual LoS. In this case, both models converge to the conditional LoS distribution.

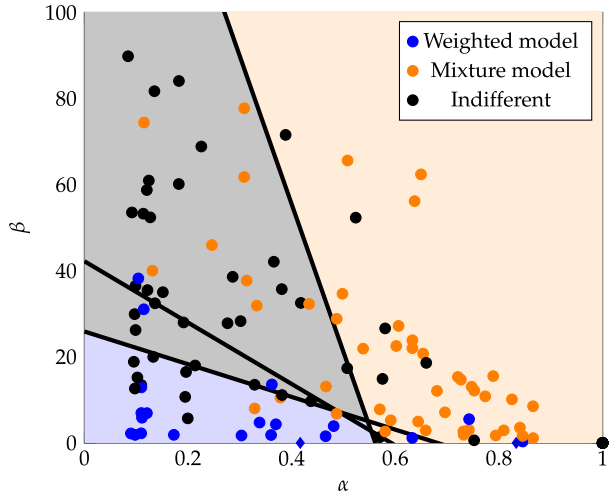


Fig. 2. Maximum margin hyperplanes for SVM classifier for each combination of the three classes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Results for the case study

In this section, we present the results of our model applied to a case study. The data is obtained from a medium-sized Dutch hospital. The hospital has a main location, and a subsidiary location for its lower complexity care. The hospital and its subsidiary location have approximately 500 beds.

3.2.1. Input data

Two data sets used to obtain the results are retrieved from the hospital's Enterprise Resource Planning (ERP) system. The first data set consists of patient-level information at that time, about the date and time of hospitalisation, the EDD, ward, current treatment specialty,

whether the hospitalisation was planned, and the discharge date. To create this data set, for eight days in 2023 and 2024 we created a snapshot of the present patients on a selection of the hospital's wards. As EDDs may get updated as time passes, the EDD in the data set is that day's EDD as recorded in the ERP system. The discharge dates were added after the patients were discharged. This resulted in a data set of 1025 patients.

The second data set consists of 360 325 hospital admissions between 2019 and May 2023, with for each admission information about its admission and discharge date, final specialty, location and whether or not the hospitalisation was planned. This data set is used to obtain the LoS distributions and to calculate, by averaging, the parameter of the truncated Poisson distribution regarding the number of emergency arrivals per weekday. We choose the 98th percentile as the truncation point for the Poisson distribution. For both data sets, we removed any entries with missing data, and dropped duplicate rows. We choose not to remove any outliers from the second data set, as outliers would likely have a limited effect on the empirical LoS distributions. To display the structure of both data sets, we included two tables with dummy data in Tables A.3 and A.4.

To assess the performance of our prediction of $N(t)$, we create a train and test set. The test set consists of the three most recent snapshots, including 541 patients. The training data comprises the remaining five snapshots, covering 484 patients, to optimally account for potential recent developments in EDD estimation behaviour. Note that for the test sets we also included patients that are planned for admission. The training snapshots are only used to estimate α and β, and hence do not include planned admissions. For the test data set, we added the planned and emergency patients for the upcoming seven days.

To account for differences in the accuracy of the EDD for different patient groups and the difference in LoS, we define a patient type as a combination of a specialty, whether the admission was a planned (P) or emergency (E) admission. The location, being either Main (M) or Subsidiary (S), is important because severely ill patients (ASA Score 1–3, Daabiss [36]) are only treated at the Main location. The specialties considered are neurology, geriatrics, gastroenterology, urology, internal medicine, general surgery, pulmonology, gynaecology, orthopaedic surgery, cardiology and otolaryngology. Due to data sensitivity, we have anonymised the specialties using (in no particular order) the numbers 1–12. Type "Specialty 1 P M" thus means planned patients from specialty 1 residing at the main location. This grouping resulted in 32 patient types.

3.2.2. Prediction of bed census

We use the train snapshot data to obtain $\hat{\alpha}$ using Algorithm 1, and $\hat{\beta}$ using (13). Using the trained classifier from Section 3.1.2 we determine which model to use per type, summarised in Table 1. The table displays the patient type, the number of patients of that type in our data set, the estimated values for α and β, and the chosen model based on the trained classifier. The classifier tells us which residual LoS model to use for each patient type in the test set. On average, $\alpha = 0.67$ and $\beta = 36.83$. As on average α is quite high and β rather low, we conclude that doctors are relatively good at predicting the discharge date. Interestingly, when distinguishing between emergency (E) and planned patients (P), we see that α on average is 0.40 and 0.88, and β is 63.35 and 16.93, respectively, indicating that for planned patients it apparently is much easier to estimate the discharge date accurately. This is in agreement with intuition, as emergency patients typically arrive to the hospital with multiple conditions and have to undergo various treatments, making it harder to estimate a patient's discharge date.

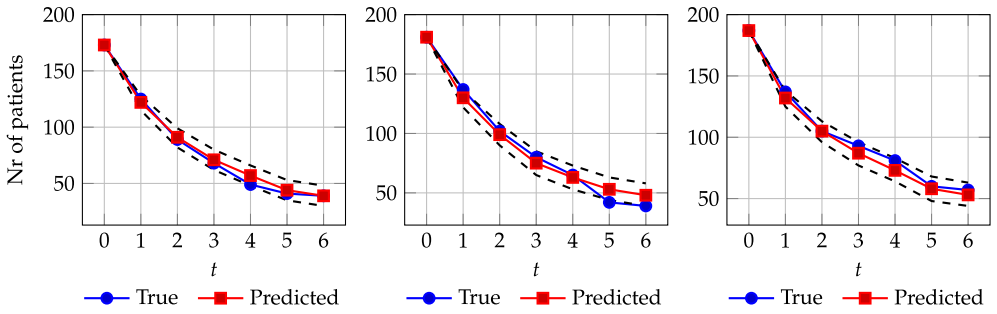
We apply Theorem 1 to find the predicted distribution of $X_0(t)$. We predict $X_0(t)$ for the three snapshots in our test data, starting on October 23, 2023, November 11, 2023 and January 8, 2024. Fig. 3 displays the median of the predicted distribution of $X_0(t)$ (red), the 95%-confidence interval (black, dotted) and the true value of $X_0(t)$ (blue). Based on these results, we conclude that our model accurately predicts $X_0(t)$. For all three snapshots in the test data, the true value lies in or on the boundary of the confidence interval.

Table 1
Count per type, obtained estimates of α, β and selected model for each type in training data.

Type	Count	$\hat{\alpha}$	$\hat{\beta}$	Model
Specialty 1 E M	1	0.60	115.48	Mixture
Specialty 1 E S	6	1.00	0.00	Mixture
Specialty 1 P M	3	0.68	107.23	Mixture
Specialty 1 P S	15	1.00	0.00	Mixture
Specialty 2 E M	2	0.19	27.86	Mixture
Specialty 2 P M	7	0.85	0.09	Mixture
Specialty 2 P S	1	1.00	0.00	Mixture
Specialty 3 E M	124	0.37	31.28	Mixture
Specialty 3 P M	9	0.46	0.91	Weighted
Specialty 3 P S	1	1.00	0.00	Mixture
Specialty 4 E M	43	0.26	116.08	Mixture
Specialty 4 E S	26	0.00	216.00	Mixture
Specialty 4 P M	4	1.00	41.14	Mixture
Specialty 4 P S	5	1.00	0.00	Mixture
Specialty 5 E M	21	0.39	16.24	Mixture
Specialty 5 E S	2	0.00	0.00	Weighted
Specialty 5 P M	3	0.18	46.09	Mixture
Specialty 5 P S	32	1.00	0.00	Mixture
Specialty 6 E M	7	0.00	240.53	Mixture
Specialty 6 E S	4	0.39	118.67	Mixture
Specialty 6 P S	6	1.00	50.00	Mixture
Specialty 7 E M	4	0.34	25.03	Mixture
Specialty 7 P M	41	1.00	0.00	Mixture
Specialty 8 E M	3	0.24	14.29	Mixture
Specialty 8 P M	9	1.00	0.00	Mixture
Specialty 9 E M	1	0.29	28.79	Mixture
Specialty 9 P M	2	0.34	93.22	Mixture
Specialty 9 P S	1	1.00	0.00	Mixture
Specialty 10 E M	16	1.00	0.00	Mixture
Specialty 10 P M	11	1.00	0.00	Mixture
Specialty 10 P S	17	1.00	0.00	Mixture
Specialty 11 E M	3	1.00	0.00	Mixture
Specialty 11 P M	15	1.00	0.00	Mixture
Specialty 11 P S	25	1.00	0.00	Mixture
Specialty 12 P M	19	1.00	0.00	Mixture

We exploited our mixture model to benchmark our model against the current practice where only the EDD is used ($\alpha = 0$) and the situation where only the conditional LoS is used ($\alpha = 1$), as is suggested by Davis and Fard [19]. Table 2 presents results of comparison per snapshot. "True" denotes the true value of $X_0(t)$, "EDD" and "Naive" the results of the EDD and Naive model respectively, and "This paper" with the results obtained by our residual LoS models combined with the trained SVM classifier. We observe that our model outperforms the other two models for all three snapshots in terms of MSE, and in two out of three snapshots in terms of MAE where the in the third case the MAE of our model and the Naive model are almost equal. This indicates that it is indeed helpful to include the EDD when making predictions for $X_0(t)$.

By leveraging Corollary 1, we can easily obtain the predicted distribution of $N(t)$. Fig. 4 displays the historical bed census up to and including $t = 13$ (Monday), and the predicted median (blue) with 95%-confidence interval (black, dashed) together with the realisation (red) for $t = 13, \dots, 19$ (Monday to Sunday) for the snapshot of November 13. The figure shows that the realised bed census falls within the predicted confidence bounds. Given the high volatility of the bed census, we conclude that our model accurately predicts next week's bed census for



(a) Snapshot October 23, 2023. **(b)** Snapshot November 13, 2023. **(c)** Snapshot January 8, 2024.

Fig. 3. Predictions vs. realisations of $X_0(t)$ for three data snapshots. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

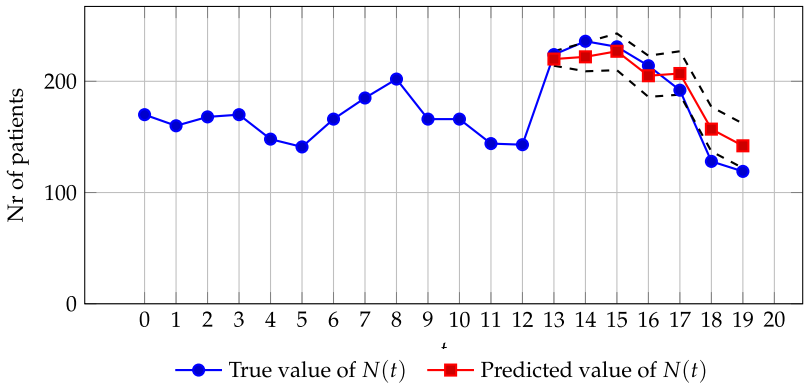


Fig. 4. Predictions vs. true value of $N(t)$ for the snapshot of November 13, 2023. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
The true values together with the predictions made using the EDD ($\alpha = 0$) and naive ($\alpha = 1$) approach, and the combined model presented in this paper for $t = 0, \ldots, 6$ for three data snapshots.

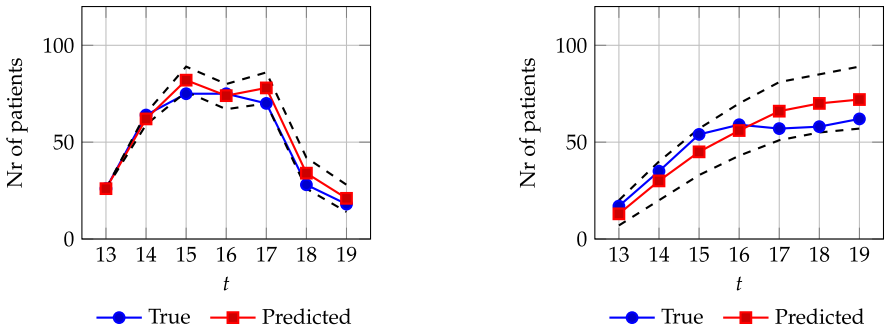
t	October 23, 2023				November 13, 2023				January 8, 2024			
	True	EDD	Naive	This paper	True	EDD	Naive	This paper	True	EDD	Naive	This paper
0	*173*	173	173	173	*181*	181	181	181	*187*	187	187	187
1	*125*	109	133	122	*137*	118	143	130	*137*	119	144	132
2	*89*	87	91	91	*102*	96	104	99	*105*	99	112	105
3	*68*	72	67	71	*80*	80	67	75	*93*	83	88	87
4	*49*	61	52	57	*65*	68	54	63	*81*	71	72	73
5	*41*	51	34	44	*42*	58	39	53	*60*	61	48	58
6	*39*	44	32	39	*39*	51	38	48	*57*	54	48	53
MSE		77.9	25.1	**13.6**		115.1	48.6	**41.3**		81.4	61.3	**20.7**
MAE		7	4	**2.7**		8	5.1	**5.3**		6.9	7	**3.6**

time periods $t = 13, \ldots, 16$. The error increases for periods $t = 17, 18, 19$. The predictions for $N(t)$ for the other two snapshots can be found in Figs. B.7 and B.8 in the Appendix.

In Fig. 5, we decompose the predicted median bed census into elective and emergency patients. This decomposition tells us that the census caused by planned patients arriving on or after $t = 13$ in Fig. 5(a) can be predicted accurately, but that we tend to overshoot at the end of our prediction horizon. The observed emergency arrivals depicted in Fig. 5(b) once more align with the anticipated range, albeit with a greater variance from the median in comparison to scheduled patients. Additionally, it is worth highlighting the considerably broader confidence interval, attributable to the stochastic nature of Poisson emergency arrivals. Exceeding both the projected and emergency predictions towards the end of the prediction interval leads to a greater deviation at that point in Fig. 4. The predictions for the planned and emergency patients for the other two snapshots can be found in Figs. B.9 and B.10 in the Appendix.

4. Managerial insights

At the start of our project, the partnering hospital was only utilising tactical-level predictions to determine bed capacity. Because tactical planning addresses a longer planning horizon, much of the operational-level information like e.g., current ward census and the number of



(a) Predictions vs. true value of planned patients.

(b) Predictions vs. true value of emergency patients.

Fig. 5. Breakdown of prediction into planned and emergency patients for the snapshot of November 13, 2023. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

patients scheduled for admission over the coming days. However, adjustments using operational-level data allow hospital managers to better adapt capacity to what is likely to happen in the near future. Using a participatory design approach with the stakeholders, we implemented our model in the hospital's Business Intelligence (BI) environment. This resulted in the following considerations:

- What predictions to show?
- What is the desired risk level?
- How to act upon predictions?

As our model provides the full distribution, there is a wide range of possible predictions to show. Together with the management, we decided to have the model produce three outputs, which are the median bed census together with the upper and lower fence of a prediction interval.

The hospital management must establish an acceptable level of risk. However, this task is challenging, as interpreting probabilities and making decisions based on probabilistic forecasts is inherently difficult [37]. Deciding on the risk threshold requires careful consideration of hospital-specific factors and operational realities. The choice of risk level is subjective, and based on the risk aversion of the hospital management. We conveyed a risk level corresponding to a prediction interval width of 85%, implying that we show the 7.5th and 92.5th percentiles of the census distribution.

To translate predictions into actions, decision makers have to decide on how to use the model. In our partnering hospital, the hospital management has decided to use the upper bound of the prediction range as the target for bed capacity planning. This ensures a more conservative approach, accommodating potential surges in patient admissions not yet accounted for by the model. We distinguish between two situations. If the predictions show that the bed census will likely be lower than the current bed capacity, the management can decide to reduce bed capacity by giving nurses time-off, asking them complete trainings such as e-learnings or to temporarily reassign nurses to other departments. Alternatively, the management could aim to increase patient inflow by allowing for more surgeries resulting in more admissions. If predictions show that the bed census will overshoot bed capacity, action should be taken to reduce inflow, shorten admissions or create more capacity. The former two can be achieved by, e.g., cancelling non-urgent surgeries or diverting patients to other hospitals, or by proactively discharging patients. Managers can create more capacity by, e.g., hiring flexible staff or temporarily asking nurses from other department for help.

We are currently piloting our model for a cluster of wards at our partnering hospital. The cluster manager monitors the predictions and adjusts capacity accordingly. Fig. 6 shows the prediction for the pilot cluster for the coming 5 days. It depicts the tactical bed plan which is based on cyclic predictions which do not change throughout the year, the operational prediction together with the prediction interval, and the capacity decision. We conclude that the manager decided to deviate from the tactical bed plan saving almost 10 beds per day by letting nurses complete their compulsory e-learning modules, freeing up their capacity at a later stage. Given a nurse-to-patient ratio of 1:6, the prediction model reduced the required nurses by at least one per day. This example thus illustrates the impact of our model.

5. Concluding remarks

In this paper, we have presented a probabilistic bed census prediction model. For this model, we have developed two residual LoS models, both combining the conditional LoS distribution and the EDD. In the first model, we propose a mixture distribution, where we use the Newton–Raphson method to find the MLE estimate of the mixture parameter. In the second model, we weigh the importance of the EDD using a Gaussian distribution, and estimated the variance using Bayesian techniques. We are the first to consider the EDD to analytically model the residual LoS. Regardless of the residual LoS model, we have shown that the bed census for the patients present at time t who were already at the ward at time 0 follows a Poisson binomial distribution. Using results from the literature, we obtained the full bed census distribution, including planned and emergency patients. We showed the conditions under which either of the two residual LoS models performs better, and we have demonstrated the use of our model using real-world hospital data.

Our model outperforms the current hospital practice, which is to solely use the EDD for census predictions. Our model also outperforms using just the conditional LoS, which is an approach suggested in recent literature. The model produces accurate predictions for next week's bed census, which implies that it could be used as support for capacity-related decisions. The decomposition of the census prediction into patients that are still present, scheduled patients and emergency patients could provide necessary decision support in the course of action for the next week.

As the model does not assume any specific LoS or arrival characteristics, can be used in many different situations. From a computational point of view, evaluating a Poisson binomial distribution is demanding, but we have indicated at the end of Section 3.1.2 that 5000 patients takes less than a second. This means that our model is applicable for larger hospitals as well. Very small hospitals with a severely limited bed capacity might pose additional challenges. In those cases, it might be favourable that the model produces a bit more conservative LoS

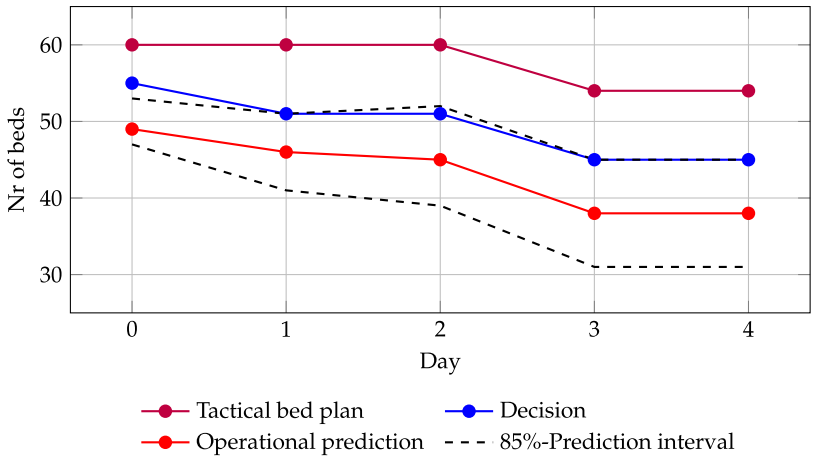


Fig. 6. Tactical and operational predictions and decisions for the pilot cluster, October 24–28, 2024. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

predictions. The decision maker might opt to add a penalty when obtaining the estimates for α and β. Our model can be adapted to include other arrival streams unaccounted for in this paper, such as arrivals from General Practitioners or other hospitals, by fitting an arrival distribution and exploiting Theorem 2.

We did not compare our bed census prediction with Machine Learning (ML) based census predictions models. In comparison to the methods proposed in this paper, ML models are generally more complex and often perceived as black-box models [38]. These methods might offer relief as they are generally considered to be flexible and able to deal with many variables, potentially resulting in better predictions. Known ML based bed census models provide point estimates [39], whereas probabilistic forecasts are increasingly preferred over point forecasts [20]. This frequently makes comparison between the two very hard.

We had to make several limiting assumptions in developing our model. For example, we assumed i.i.d. LoSs throughout this paper, with each patient type having its own LoS distribution. Although this assumption is much less stringent compared to assuming that the LoS is independent across *all* patients, it might still be a bit strong in some cases. In our partnering hospital for example, doctors are encouraged to discharge patients on Friday, as otherwise the patient would have to spend the weekend in the hospital. This might artificially shorten a patient's LoS, creating a dependence on the day of the week.

As another example, (over) crowded wards might also incur doctors to discharge patients sooner than normal. Although we acknowledge that these situations might occur and indeed contradict our assumption, we think that the practical impact is limited as the predictions of bed occupancy are primarily used for weekdays to adjust the planning of scheduled admissions, and not so much during weekends. One might argue that in case of an unexpected event such as a peak in emergency arrivals this is unrealistic. As proposed in Section 2.2, we modelled the arrivals of scheduled admissions as a deterministic process. However, we choose to reflect the information present at that time in our model without adjusting for unexpected events. Manually adjusting for these kind of events would in our opinion introduce bias due to subjectivity.

As opportunities for further research, it might prove useful to consider continuous time to more accurately model intra-day bed census dynamics. A first step may be to consider hourly bed census, that may readily be obtained from our results. Another limitation is the computational complexity of computing the Poisson binomial distribution. There might be better, even faster algorithms that could make our method more scalable. As we provide the full bed census distribution, another avenue for future work is to use our model in staffing optimisation models, which could use our model to obtain decision rules on short-term flexible staff hiring. The model could also be used in other areas of healthcare management, for instance to model workload of the patient base of a therapist whose patients need an unknown number of consultations, and about which the therapist could make an estimate. The proposed model is currently being implemented at our Dutch partnering hospital. Further research should gather evidence on the impact of these models in healthcare practice and the way these more accurate predictions impacted healthcare delivery.

CRediT authorship contribution statement

Hayo Bos: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Stef Baas:** Writing – review & editing, Methodology, Formal analysis. **Richard J. Boucherie:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Formal analysis, Conceptualization. **Erwin W. Hans:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Gréanne Leeftink:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research is funded by and done at Diakonessenhuis Utrecht, the Netherlands, in collaboration with the Center for Healthcare Operations Improvement and Research (CHOIR).

Appendix A. Input data

See Tables A.3 and A.4.

Table A.3
Sample of dummy data of data set 1.

Admission nr	Admission date	Admission time	Discharge date	Discharge time	Ward	Specialty	Emergency
1	13/9/2018	10:00	15/9/2018	10:53	W_1	S_1	Yes
2	04/12/2018	15:10	08/12/2018	11:46	W_1	S_2	No
3	06/12/2018	14:10	12/12/2018	9:36	W_2	S_2	Yes

Table A.4
Sample of dummy data of data set 2.

Admission nr	Admission date	Admission time	Expected discharge date	Ward	Specialty	Emergency	Realised discharge date
4	13/06/2023	10:00	16/06/2023	W_1	S_1	Yes	17/06/2023
5	13/06/2023	9:42	15/06/2023	W_1	S_2	No	19/06/2023
6	13/06/2023	10:37	17/06/2023	W_2	S_2	No	17/06/2023

Appendix B. Predictions

B.1. Predictions of $N(t)$ for remaining snapshots

See Figs. B.7 and B.8.

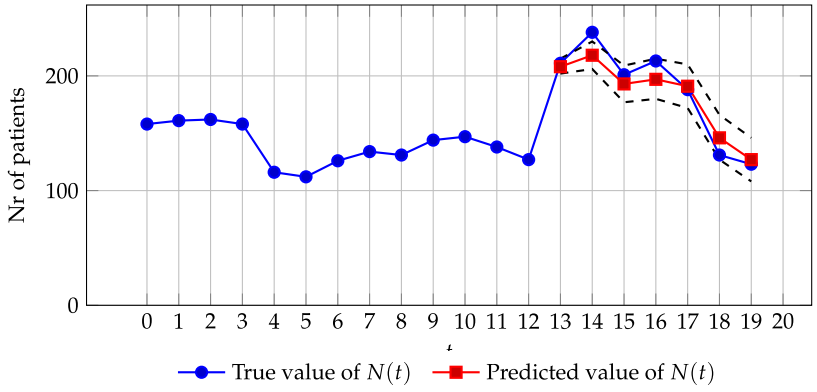


Fig. B.7. Predictions vs. true value of $N(t)$ for the snapshot of January 8, 2024. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

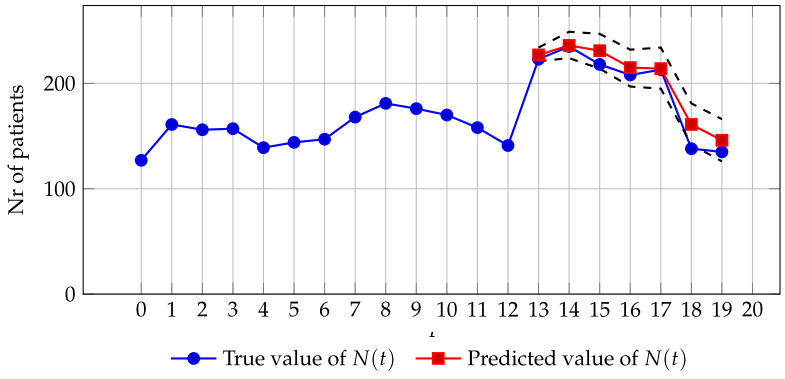
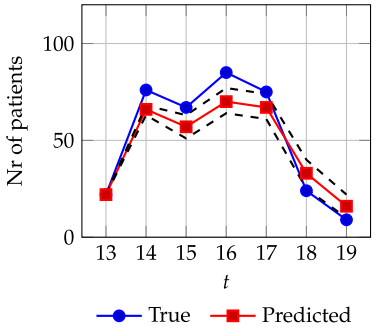


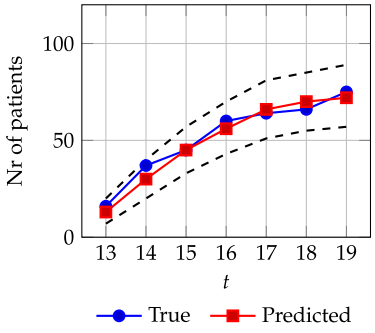
Fig. B.8. Predictions vs. true value of $N(t)$ for the snapshot of October 23, 2023. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

B.2. Predictions for planned and emergency patients

See Figs. B.9 and B.10.

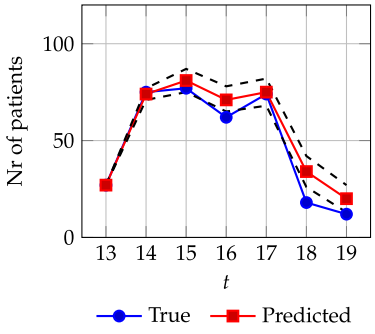


(a) Predictions vs. true value of planned patients.

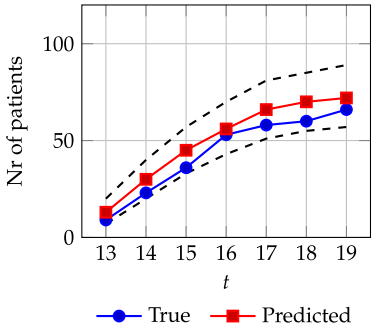


(b) Predictions vs. true value of emergency patients.

Fig. B.9. Breakdown of prediction into planned and emergency patients for the snapshot of October 23, 2023. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



(a) Predictions vs. true value of planned patients.



(b) Predictions vs. true value of emergency patients.

Fig. B.10. Breakdown of prediction into planned and emergency patients for the snapshot of January 8, 2024. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Data availability

The data that has been used is confidential.

References

[1] De Bruin AM, Bekker R, Van Zanten L, Koole GM. Dimensioning hospital wards using the Erlang loss model. Ann Oper Res 2010;178(1):23–43.
[2] Fuegener A, Hans EW, Kolisch R, Kortbeek N, Vanberkel PT. Master surgery scheduling with consideration of multiple downstream units. Eur J Oper Res 2014;239(1):227–36.
[3] Kortbeek N, Braaksma A, Burger C, Bakker P, Boucherie R. Flexible nurse staffing based on hourly bed census predictions. Int J Prod Econ 2015;161:167–80.
[4] Kuntz L, Mennicken R, Scholtes S. Stress on the ward: Evidence of safety tipping points in hospitals. Manage Sci 2015;61(4):754–71.
[5] Giokas DI. Greek hospitals: how well their resources are used. Omega 2001;29(1):73–83.
[6] Harper PR, Shahani AK. Modelling for the planning and management of bed capacities in hospitals. J Oper Res Soc 2002;53(1):11–8.
[7] Hans EW, van Houdenhoven M, Hulshof PJH. A Framework for Healthcare Planning and Control. In: Hall R, editor. Handbook of healthcare system scheduling. Vol. 168, Boston, MA: Springer US; 2012, p. 303–20, Series Title: International series in operations research & management science.
[8] Cochran JK, Roche K. A queuing-based decision support methodology to estimate hospital inpatient bed demand. J Oper Res Soc 2008;59(11):1471–82.
[9] Barado J, Guergué JM, Esparza L, Azcárate C, Mallor F, Ochoa S. A mathematical model for simulating daily bed occupancy in an intensive care unit*:. Crit Care Med 2012;40(4):1098–104.
[10] Bravo F, Rudin C, Shaposhnik Y, Yuan Y. Interpretable Prediction Rules for Congestion Risk in Intensive Care Units. Stoch Syst 2023. stsy.2022.0018.
[11] Veneklaas W, Leeftink A, Van Boekel P, Hans E. On the design, implementation, and feasibility of hospital admission services: The admission lounge case. Omega 2021;100:102308.
[12] Vanberkel PT, Boucherie RJ, Hans EW, Hurink JL, van Lent WAM, van Harten WH. An exact approach for relating recovering surgical patient workload to the master surgical schedule. J Oper Res Soc 2011;62(10):1851–60.
[13] Kortbeek N, Braaksma A, Smeenk FH, Bakker PJ, Boucherie RJ. Integral resource capacity planning for inpatient care services based on bed census predictions by hour. J Oper Res Soc 2015;66(7):1061–76.
[14] Wright PD, Mahar S. Centralized nurse scheduling to simultaneously improve schedule cost and nurse satisfaction. Omega 2013;41(6):1042–52.
[15] Sahraoui A, Elarref M. Bed crisis and elective surgery late cancellations: an approach using the theory of constraints. Qatar Med J 2014;2014(1):1.
[16] Azcarate C, Esparza L, Mallor F. The problem of the last bed: Contextualization and a new simulation framework for analyzing physician decisions. Omega 2020;96:102120.
[17] Heins J, Schoenfelder J, Heider S, Heller AR, Brunner JO. A Scalable Forecasting Framework to Predict COVID-19 Hospital Bed Occupancy. INFORMS J Appl Anal 2022;52(6):508–23.
[18] Tello M, Reich ES, Puckey J, Maff R, Garcia-Arce A, Bhattacharya BS, Feijoo F. Machine learning based forecast for the prediction of inpatient bed demand. BMC Med Inform Decis Mak 2022;22(1):55.

[19] Davis S, Fard N. Theoretical bounds and approximation of the probability mass function of future hospital bed demand. Health Care Manage Sci 2020;23(1):20–33.
[20] Gneiting T, Katzfuss M. Probabilistic Forecasting. Annu Rev Stat Appl 2014;1(1):125–51.
[21] Li N, Pan J, Xie X. Operational decision making for a referral coordination alliance-When should patients be referred and where should they be referred to? Omega 2020;96:102077.
[22] Baas S, Dijkstra S, Braaksma A, Van Rooij P, Snijders FJ, Tiemessen L, Boucherie RJ. Real-time forecasting of COVID-19 bed occupancy in wards and Intensive Care Units. Health Care Manage Sci 2021;24(2):402–19.
[23] Dijkstra S, Baas S, Braaksma A, Boucherie RJ. Dynamic fair balancing of COVID-19 patients over hospitals based on forecasts of bed occupancy. Omega 2023;116:102801.
[24] Bekker R, Uit Het Broek M, Koole G. Modeling COVID-19 hospital admissions and occupancy in the Netherlands. European J Oper Res 2023;304(1):207–18.
[25] Cuadrado D, Valls A, Riaño D. Predicting intensive care unit patients' discharge date with a hybrid machine learning model that combines length of stay and days to discharge. Mathematics 2023;11(23):4773.
[26] Alam MM. An efficient random forest algorithm-based telemonitoring framework to predict mortality and length of stay of patients in ICU. Multimedia Tools Appl 2024;83(17):50581–600.
[27] Boulesteix A-L, Schmid M. Machine learning versus statistical modeling. Biom J 2014;56(4):588–93.
[28] Porter BJ. Estimated discharge dates: Putting theory into practice. In: Leez L, editor. Timely discharge from hospital. Cumbria: M&K Update Ltd; 2011, p. 355–67.
[29] Henry OP, Li G, Freundlich RE, Sandberg WS, Wanderer JP. Understanding the Accuracy of Clinician Provided Estimated Discharge Dates. J Med Syst 2022;46(1):2.
[30] Piniella NR, Fuller TE, Smith L, Salmasian H, Yoon CS, Lipsitz SR, Schnipper JL, Dalal AK. Early Expected Discharge Date Accuracy During Hospitalization: A Multivariable Analysis. J Med Syst 2023;47(1):63.
[31] Beck A. Introduction to nonlinear optimization: Theory, algorithms, and applications with Python and MATLAB. MOS-SIAM series on optimization, 2nd ed.. Society for Industrial and Applied Mathematics; 2023.
[32] Jeffreys H. An invariant form for the prior probability in estimation problems. Proc R Soc Lond Ser A 1946;186(1007):453–61, arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1946.0056.
[33] Santos D, Marques I. Designing master surgery schedules with downstream unit integration via stochastic programming. European J Oper Res 2021.
[34] Straka M. Poisson binomial distribution for Python. 2016, https://github.com/tsakim/poibin.
[35] Hong Y. On computing the distribution function for the Poisson binomial distribution. Comput Statist Data Anal 2013;59:41–51.
[36] Daabiss M. American Society of Anaesthesiologists physical status classification. Indian J Anaesth 2011;55(2):111.
[37] Hájek A. Interpretations of Probability. In: Zalta EN, Nodelman U, editors. The stanford encyclopedia of philosophy. Winter 2023 ed.. Metaphysics Research Lab, Stanford University; 2023.
[38] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1(5):206–15.
[39] Hartwig M, Schiff S, Wolfrum S, Möller R. Aggregating predicted individual hospital length of stay to predict bed occupancy for hospitals. In: BIOSTEC (2). 2024, p. 175–84.