Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Modeling COVID-19 hospital admissions and occupancy in the Netherlands

René Bekker, Michiel uit het Broek, Ger Koole

1. Introduction

The coronavirus has an enormous impact on our health system and today’s society as a whole. On March 11, 2020, the World Health Organization officially characterized COVID-19 as a pandemic. By the end of January 2021, the number of people diagnosed worldwide with COVID-19 crossed the 100 million mark (World Health Organization, 2021), which has put a tremendous strain on scarce hospital capacities. Specifically, the pandemic places a load on clinical bed capacity, and in particular on Intensive Care Units (ICUs), that is sometimes well beyond the currently available bed capacities (IHME COVID team & Murray, 2020a; 2020b). The catastrophic situation in Lombardy, Italy, mid-March 2020 has tragically shown the impact of the lack of health capacities (Rosenbaum, 2020), and the need to manage hospital bed capacities as good as possible. In (Phua et al., 2020), the authors call upon ICU practitioners, hospital administrators, governments, and policy makers to be prepared early for a substantial increase in critical care capacity. Their recommendations relate to, among others, ICU capacity and ICU staffing. More specifically, they recommend to make plans for an increase in capacity as a result of a rapid increase in critically ill COVID-19 patients. A less studied but perhaps even more important issue is the impact on other types of care which are delayed because of COVID-19 patients occupying beds and using other forms of capacity which would otherwise be used for non-COVID care. An early study concerning the impact of only the first wave came to an estimated loss of up to 400 thousand healthy life years in the Netherlands (Gupta Strategists, 2020).

The aim of this paper is to develop a prediction model that helps hospitals in reserving the right number of beds (both ICU and clinical) for COVID-19 patients. It is also used for decisions at the national level for upscaling the number of ICU beds. These decisions are made by the LCPS (Landelijk Coördinatiecentrum Patiënten Spreiding), the national Dutch center responsible for capacity decisions and relocations. Concerning the latter, to balance the pressure on clinical and ICU beds over the Netherlands, patients may be relocated to different regions. Relocations may be necessary due to a lack of capacity, because of an outbreak in a certain region, or to spread the pressure of COVID-19 patients to allow for an equal amount of delayed care in all regions. To make decisions on required capacity and on relocations between regions we were asked to make a model to support these decisions. Regions and local hospitals need a couple of days to modify the number of available COVID-19 beds, and thus require occupancy predictions of a couple of days ahead. This is therefore the main goal of the current work. Furthermore, our model was used for long-term decision making. For different demand scenarios occupation calculations can be made giving insight in required capacity under different policy measures.

The prediction model that we developed consists of three steps. First, we predict arrivals on the basis of historical data. For this, we...
employ a linear programming model that is inspired by smoothing splines that incorporates weekly seasonality and requires little data. The prediction has interpretations in terms of the day-to-day reproduction factor. Then, using information on Lengths-of-Stay (LoS) of previous patients, we determine the LoS distribution for new patients and the residual LoS for patients already present. Note that historical data is censored, because some patients are still present and therefore we do not know their LoS. For uncensoring we used the Kaplan–Meier estimator. Finally we predict hospital occupancy. This third step uses methods stemming from queueing theory, specifically from discrete-time infinite server queues. The whole process is depicted in Fig. 1.

In this paper we use publicly available data to make predictions at the national level. As data of individual patients is more difficult (or sometimes impossible) to obtain, we only require aggregated data, i.e., patient counts. The same model was and is used to make predictions at the regional level at the LCPS. These predictions are communicated twice a week to the regions, or more often if the prediction is very different from the one last communicated. The regions use these predictions to reserve capacity for COVID-19 patients. Next to that, the LCPS uses these predictions to transfer patients between regions to equalize the pressure on the hospitals due to COVID-19 patients. We decided to present results at the national level because of the availability of public data.

Our contribution therefore both has a practical and a theoretical component. From the theoretical point of view we developed a new forecasting model for hospital admissions, and we used queueing theory to translate current occupancy and predicted admissions into a prediction for the short-term occupancy. This model was implemented in R and used successfully by the LCPS. It increased accuracy compared to the simple model used before and gave the regions more trust into the accuracy of the predictions. This likely led to less under- and overcapacity and therefore reduced the amount of delayed care. The model uses little data and can easily be adapted to other situations (such as predicting regular emergency ICU care) and countries. We developed easy-to-use R code including connections to the data warehouse to which all regions upload their most recent data. The model is executed daily with hardly any involvement from our side.

The organization of the paper is as follows. First we discuss related literature. The method for predicting the arrivals is discussed in Section 3. The LoS distribution can be found in Section 4, which is used in the model to predict the bed occupancy in Section 5. The prediction results can be found in Section 6, whereas Section 7 concludes with a brief discussion on delayed care.

2. Related literature

Providing accurate long-term predictions concerning the number of COVID-19 patients requiring hospital capacity is difficult (Ioannidis, Cripps, & Tanner, 2020). This is also supported by the conclusion in Xiang et al. (2021), stating that caution is required when formulating public health strategies based on prediction models; (Xiang et al., 2021) provides a review of COVID-19 epidemic prediction models to study the impact of public health interventions. Although short-term predictions of bed occupancy tend to be more accurate, the amount of literature in this area is still limited.

From Fig. 1 it may be seen that three different elements are involved: (i) admissions, (ii) length-of-stay, and (iii) bed occupancy. By far, most available research has focussed on the area related to the first element, the number of admissions. Typically, the focus is broader and involves the description of the epidemic process rather than only the number of infected (hospitalized) patients. There is a long tradition of epidemiological models (see e.g. Hamer, 1906; Kermack & McKendrick, 1927) aiming to describe the epidemic process such that the impact of public health interventions can be assessed. A classic compartmental model in this area is SEIR (Susceptible-Exposed-Infectious-Recovered). From the systematic review (Shankar et al., 2021) it follows that SEIR is also the most common model used for the COVID-19 pandemic. As an example, an extended version of SEIR has been used in France (Prague et al., 2020) as a simplified representation of the average epidemic process and the impact of a nationwide lockdown. We refer to Guan, Wei, Zhao, & Chen (2020) for an overview of the transmission dynamics of the COVID-19 pandemic. For short-term predictions, the number of patients in each SEIR compartment is relatively stable and provides little additional predictive power. Therefore, in our study we consider the series of the number of hospital admissions directly.

Little work has been done on ICU LoS predictions, the second element. In (Rees et al., 2020; Vekaria, Overton, & Winsiowski, 2020) the LoS distribution of COVID-19 patients has been studied worldwide and in the UK, respectively. For the LoS, we fit the same classical LoS distributions as the authors do in Rees et al. (2020); Vekaria et al. (2020). Moreover, most work is focused on whether or not a patient can be discharged, such as Ma, Si, Wang, & Wang (2020). Our forecasts are, however, not based on individual patients records, thereby also requiring less data. In other settings more research has been done on LoS predictions, such as Armony et al. (2015); Maguire, Taylor, & Stout (1986); Shi, Chou, Dai, Ding, & Sim (2016).

There are some papers that focused directly on predicting the number of occupied beds. For instance, Farcomeni, Maruotti, Divino, Jona-Lasinio, & Lovison (2020) used an ensemble of two forecasting methods for a short-term forecast of occupied COVID-19 beds in Italy. An ensemble of methods is also used in Goic, Bozanic-Leal, Badal, & Basso (2021), which combines autoregressive, machine learning and epidemiological models to provide a short term forecast of ICU utilization. Furthermore, Zhao et al. (2020) applied epidemic models for short-term ICU occupancy forecasts in Switzerland; Massonaud, Roux, & Crépey (2020) uses a similar approach for the situation in France. The authors of
3. Predicting admissions

There are different possibilities for building a model for admissions. An obvious option is a statistical model that uses historical values to make predictions. A disadvantage of such a model is that trend changes caused by external variables cannot be predicted. Also, many classical statistical models require a substantial number of observations in order to produce reliable predictions, whereas data is typically scarce when a new pandemic arises. Furthermore, one would assume that somehow data on positive tests could be used, and that presumably positive tests occur before admissions, such that data on tests can be used to predict later admissions. In Fig. 2 data on admissions and positive tests (at the day of registration) are plotted, the dots are the actual values. Data comes from different publicly available sources, in this case NICE (a Dutch ICU data repository) and RIVM (the Dutch national epidemiological institute), as is conveniently gathered at van Zelst (2020). The red bars indicate policy changes (partial lockdowns), it took 13 and 11 days for the numbers of admissions to go down (the black bars). The decrease in Spring 2021 is due to the vaccinations. Surprisingly, the number of positive tests spikes at the same days as the number of admissions, for all waves except the last. This suggests that this external variable would have little added value. We tested this by looking at the correlation between the admissions and the positive tests with different time lags. Indeed, the highest correlation was obtained for a lag of 0. For this reason, we did not use the number of newly registered positive tests per day to predict trend changes in admissions. Knowing the current reproduction factor would have been very helpful. However, it is determined only retrospectively by the RIVM with a lag of 2 weeks.

Another reason for not using external data are the substantial changes in test policy and behavior. The green bar in Fig. 2 corresponds to one (of many) changes in test behavior; from that day (Dec 1, 2020) on civilians without symptoms could also get a test. We see that this led to a sharp increase in number of positive tests. Hospital admission also increased, but at a lower pace. Even more extreme is the spike in July 2021, due to opening the night clubs too early. Here we do see a delay in admissions, because it took some time before infected young visitors of night clubs infected older vulnerable people. Because of the high vaccination rate the number of admissions remained limited, again an illustration that it is hard to predict admissions by infections.

Variables other than the number of positive tests were tested on their ability to predict admissions, but they were not found to be predictive. For these reasons we focused on predicting admissions without external variables. Note that in the figure we also plotted our model for ICU and clinical admissions, together with a smoothing spline applied to the logs of the positive tests.

A statistical model for predicting daily admissions should have the following properties:

1. it should be smooth but at the same time allow for trend changes;
2. it should have non-negative predictions and exponential growth or decline;
3. it should model the intra-week seasonality present in the data.

For these reason we chose, inspired by smoothing splines, a model with the following features:

1. an additive model on the logs because of the multiplicative effect of time and the intra-week seasonality;
2. that minimizes a weighted sum of errors and second differences to have a smooth trend;
3. that uses absolute values to reduce the impact of outliers and few trend changes, hopefully representing the policy changes.

A similar model is used in van Leeuwen & Koole (2020) to forecast demand in hospitality. As the model is inspired by smoothing splines, it requires little data, which is preferable at the start of a pandemic. In mathematical terms, let \( a_t \) be the realization, either of the admissions at the ICU or the clinics. Our statistical models minimizes the sum of errors and trend changes, thus it is actually a minimization problem. The decision variables are \( s_k \) and \( x_k \), the day factors and the weekly factors, respectively. Let \( w(t) \) be the weekday of day \( t \), thus \( w_{t|1} \) is the day factor of day \( t \). Also define \( \Delta x_k = x_k - x_k-1 \) and \( \Delta^2 x_k = \Delta x_k - \Delta x_{k-1} \), and let \( T \) be the last day with data. Our minimization problem is:

\[
\min_{x} \sum_{t=1}^{T} |x_t + s_{w(t)}| - \log a_t + \lambda \sum_{t=2}^{T} |\Delta^2 x_t|.
\]

Here, \( \lambda \geq 0 \) is a parameter that determines the smoothness of the prediction, the “smoothing parameter”. The first term in (1) gives the difference between the smoothed curve and the data and the second term introduces a penalty for trend changes. For \( \lambda = 0 \) there will be a perfect fit on the data. For higher \( \lambda \), the curve will be smoother and there will be less overfitting.

The fit is given by \( \exp(x_{t} + t(x_{t} - x_{t-1}) + s_{w(t+1)}) \), whereas the \( t \)-day ahead prediction is:

\[
\hat{A}_{T+t} = \exp(x_{t} + t(x_{t} - x_{t-1}) + s_{w(t+1)}).
\]
The solution to (1) can be found using linear programming, as the $|\cdot|$ function can be made linear by a well-known modeling trick involving two additional variables. Specifically, the optimization (1) can be written as

$$\begin{align*}
\min & \quad \sum_{t=1}^{T} (y_t^+ + y_t^-) + \lambda \sum_{t=3}^{T} (z_t^+ + z_t^-) \\
\text{subject to} & \quad y_t^+ - y_t^- = x_t + 2w(t) - \log a_t \quad \forall t \\
& \quad z_t^+ - z_t^- = x_t - 2x_{t-1} + x_{t-2} \quad \forall t \\
& \quad y_t^+, y_t^-, z_t^+, z_t^- \geq 0 \quad \forall t
\end{align*}$$

We used the IpSolve package in R which had negligible running times.

It is interesting to note that $r_1 = e^{r \cdot \log x_1}$ is the fractional de-seasonalized increase or decrease. It can be interpreted as a day-to-day “reproduction factor”. Epidemiologists define the reproduction factor $R_t$ as the amount of people that get infected on average by one infected person at time $t$. As the incubation time is around four days there should be a relation between $R_t$ and $r_t^4$. In Fig. 3, $r_t^{4.5}$ is plotted, both for the ICU and clinical admissions. Note that $r_t$ exactly gets below 1 when the admissions start decreasing (the black bars). We also plotted (in green) the $R_t$ as it is determined by the RIVM, allowing to compare it to our $r_t^{4.5}$. We see a similar shape, and that the biggest correlation is for a lag of around twelve days, which corresponds roughly to the time between infection and hospital admission. Note that this is of little help in predicting admissions, as the final $R_t$ is only known for 2 weeks back.

4. Length of stay

To determine the length of stay (LoS), we use data of NICE again. Specifically, on their website NICE presents data describing the frequencies of number of days that patients spend at the ICU and the clinic. Define $S$ as a random variable denoting the number of hospitalized days taking values in $\{0, 1, \ldots\}$. That is, $S$ may be interpreted as the number of overnight stays at the ICU or the clinic. Some recent studies (Armony et al. 2015; Shi et al. 2016) have described the LoS at two time scales. The LoS in hours depends on many operational factors, whereas the LoS in days is attributed to medical factors. Our focus is on the latter, i.e., the time resolution in days.

Currently, there are still COVID-19 patients present at the ICU and at the clinic, yielding right-censoring of the data. Clearly, the number of patients present is also non-negligible compared to the total number of COVID-19 patients, which in particular holds for the ICU. Therefore, to estimate the LoS distribution, we use the Kaplan–Meier estimator. In particular, we have $\hat{P}(S \geq t) = 1$ and, for $t = 1, 2, \ldots$

$$\hat{P}(S \geq t) = \prod_{s=1}^{t} \left( 1 - \frac{d_s}{n_s} \right).$$
where \( d_s \) is the number of patients that are discharged after \( s \) days, and \( n_s \) is the number of patients that have a LoS of at least \( s \) days (either discharged or still present).

The mean and standard deviation of the LoS can be found in Table 1. We see that the average LoS at the ICU increases with over a day by taking the right-censoring into account. The impact is smaller at the clinic as a smaller fraction of the patients is still present (8.2% at the ICU vs. 3.4% at the clinic).

It is natural to consider the LoS at the time scale of minutes or hours, and model the LoS as a continuous random variable. There is also a considerable body of literature devoted to fitting probability distributions to such a continuous LoS. Specifically, let \( X \) represent a LoS taking values in \( (0, \infty) \). Recall that \( S \) is a random variable denoting the number of hospitalized days taking values in \( [0, 1, \ldots] \). When fitting a distribution to the LoS, we will use a fit to the continuous LoS \( X \), and use a continuity correction to find the distribution of \( S \). In particular, we have, for \( t = 1, 2, \ldots \),

\[
P(S \geq t) = P(S > t - 1) \approx P(X \geq t - 0.5).
\]

In (Armony et al., 2015), a lognormal distribution is found to fit the LoS data well. The authors also pose the challenge to explain why lognormal distributions seem to fit service durations so well. Other common distributions for lengths of stay or survival functions are gamma and Weibull distributions (Marazzi, Paccaud, Ruffieux, & Beguin, 1998); mixtures of exponentials may also be appropriate. We refer to Vekaria et al. (2020) for a study of the LoS of COVID-19 patients in the UK based on a Weibull distribution. In line with the LoS distribution of COVID-19 patients worldwide (Rees et al., 2020), we fit lognormal, gamma, and Weibull distributions. In Figs. A.10 and A.11, these distributions are displayed together with the data adjusted by the Kaplan-Meier estimate. For both the ICU and the clinic, the gamma and Weibull distributions can hardly be distinguished. Interestingly, for the ICU the gamma and Weibull distributions provide visually excellent fits, whereas for the clinic the lognormal distribution provides very good fits.

**Remark 1.** There are different ways to determine parameters of our parametric distribution \( X \). From the perspective of medical specialist and decision makers, the method of moments is especially appealing as the first two moments are relatively easy to interpret. For instance, the impact of changes in the LoS distribution are straightforward to incorporate. For \( X \sim \text{LogNormal}(\mu, \sigma^2) \), we obtain \( \mu = \ln \left( \frac{\bar{x}^2}{\sqrt{s^2 + \bar{x}^2}} \right) \) and \( \sigma^2 = \ln \left( 1 + \frac{s^2}{\bar{x}^2} \right) \), with \( \bar{x} \) and \( s^2 \) denoting the sample mean and the sample variance. For \( X \sim \text{Gamma}(\alpha, \beta) \), we obtain the shape parameter \( \alpha = \bar{x}^2 / s^2 \) and rate parameter \( \beta = \bar{x} / s^2 \). For Weibull distributions, there are no closed-form expressions when using the method of moments.

### 5. Occupancy

To predict the occupancy we use principles from queueing theory to describe the evolution of the number of COVID-19 patients. Essentially, we model the number of patients as a (discretized) infinite-server queueing model with a time-dependent arrival pattern. For the special case of (continuous) time-dependent Poisson arrivals, the \( M/G/\infty \) has well been analyzed with tractable results (Bekker & de Bruijn, 2010; Eick, Massey, & Whitt, 1993; Feldman, Mandelbaum, Massey, & Whitt, 2008; Palomo et al., 2020) uses such an \( M/G/\infty \) model to quantify how flattening the curve affects peak demand for hospital beds. The application of infinite-server models, also in discrete time, is also discussed in Worthington et al. (2020). As our goal is to predict the demand for beds without capacity constraints, the infinite-server assumption is appropriate, albeit we use a discrete-time version. For this, we do not need to make distributional assumptions regarding the arrival process.
When predicting future occupancy, we need to distinguish two groups of patients: (i) patients that are currently present, and (ii) patients that will arrive in the future, see also Fig. 1. For the patients that will arrive in the future, we need a prediction of admissions (as described in Section 3) and the subsequent length of stay (as described in Section 4). For the first group, observe that the patients that are currently present, the total length of stay differs from the one in Section 4 whereas part of the length of stay has elapsed. Since we predict on publicly available data, we cannot use the elapsed length of stay of each individual patient. A reasonable alternative seems to use the stationary residual length of stay (for which \( P(S' \leq t) = \sum_{k=t}^{\infty} P(S > k) / E[S] \)), which follows directly from renewal theory. A disadvantage of the stationary residual length of stay is that the arrival process is obviously not stationary. Therefore, we propose an alternative that takes the past arrival pattern into account.

Next, we derive the residual length of stay \( S' \) of a tagged patient present at time \( T \). Note that the probability that this patient arrived at day \( T - u \) is proportional to \( \alpha_{T-u} P(S \geq u) \), for \( u = 1, \ldots, T \). Hence, the probability that this tagged patient arrived at day \( T - u \) is

\[
\frac{\alpha_{T-u} P(S \geq u)}{\sum_{k=1}^{T} \alpha_{T-k} P(S \geq k)}.
\]

The probability that the residual length of stay of the tagged patient is at least \( s \), when the patient arrived at day \( T - u \), equals \( P(S \geq s + u | S \geq u) = P(S \geq s + u) / P(S \geq u) \). Combining the above, we have

\[
P(S' \geq t) = \sum_{u=1}^{T} \frac{\alpha_{T-u} P(S \geq u)}{\sum_{k=1}^{T} \alpha_{T-k} P(S \geq k)} \frac{P(S \geq t + u)}{P(S \geq u)}.
\]

\[
= \frac{\sum_{k=1}^{T} \alpha_{T-k} P(S \geq t + u)}{\sum_{k=1}^{T} \alpha_{T-k} P(S \geq k)}.
\]

Observe that this is consistent with the stationary residual length of stay by taking a constant and letting \( T \to \infty \).

Now, we turn to predicting the occupancy. As the allocation of COVID-19 patients is based on the occupancy in the morning, we focus on \( N_t \), the number of occupied beds at the beginning of day \( t \). We then have the following relation

\[
N_{T+t} = \sum_{i=1}^{N_T} I[S_i' \geq t] + \sum_{s=0}^{t-1} \sum_{i=1}^{N_{T+s}} I[S_i \geq t - s],
\]

where \( S_i' \) is the residual LoS of the \( i \)th patient present, and \( S_i \) represents the LoS of the \( i \)th patient arriving on that specific day. The first term is due to patients that are currently present at time \( T \), whereas the second term are patients arriving in the future. Of course, that with the relation above it is possible to derive the distribution of \( N_{T+t} \). Focusing on the expectation, it holds that

\[
\hat{N}_{T+t} = E[N_{T+t}] = N_T P(S' \geq t) + \sum_{s=0}^{t-1} E[A_{T+s}] P(S \geq t - s),
\]

providing the \( t \)-day ahead prediction \( \hat{N}_{T+t} \) at day \( T \). Here, \( E[A_{T+s}] \) follows from the predictions in Section 3, \( P(S \geq t - s) \) follows from Section 4, and \( P(S' \geq t) \) from (2). Moreover, using the same relation above and assuming that \( A_{T+s} \) and \( A_{T+t+s} \) are independent for \( s \neq u \), the variance is

\[
\begin{align*}
\text{Var}[N_{T+t}] &= \sum_{u=1}^{T} \alpha_{T-u} \text{Var}(S \geq u) P(S \geq t + u) / P(S \geq u) \\
&= \sum_{s=0}^{t-1} \left( \text{Var}[A_{T+s}] \times \tilde{G}(t - s)^2 + E[A_{T+s}] \times \tilde{G}(t - s)(1 - \tilde{G}(t - s)) \right),
\end{align*}
\]

where \( \tilde{G}(t - s) = P(S \geq t - s) \). Note that the expression above simplifies if the arrivals follow a Poisson process with a known parameter. In that case \( \text{Var}[A_{T+s}] = E[A_{T+s}] \) and \( \text{Var}[N_{T+t}] \) will converge to \( E[N_{T+t}] \), such that \( N_{T+t} \) will behave as a Poisson random variable for \( t \) large enough.

Remark 2. Observe that relation (3) in principle provides the complete occupancy distribution. For infinite-server models, the occupancy can well be approximated by a normal distribution, see Pang & Whitt (2010) for a theoretical foundation. Using the mean \( E[N_{T+t}] \) and variance \( \text{Var}[N_{T+t}] \) given above, an accurate approximation of the full occupancy distribution can be given as well.

6. Results

In this section we present the numerical results that follow from our prediction model. In Section 6.1 we first address...
the choice of the tuning parameter $\lambda$ for the arrival predictions. Section 6.2 visualizes the short-term predictions, whereas its accuracy is addressed in Section 6.3. The implementation at LCPS is described in Section 6.4. As we only have reliable occupancy data of COVID-19 patients from mid October 2020, we will use the time period from November 1, 2020, until February 1, 2021 as an illustration (except for Section 6.4 where we use more recent data). This also involves an interesting period due to the remarkable behavior of infections and hospital admissions during the ‘second wave’. In line with the operations at the LCPS, we use predictions of 3 and 7 days ahead. To assess the accuracy of the predictions, we use the following three evaluation measures: weighted absolute percentage error (WAPE), mean absolute error (MAE), and root mean squared error (RMSE). We note that the WAPE is also referred to as the weighted MAPE. For a period of $n$ days, these measures are defined as

$$\text{WAPE} = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{\sum_{t=1}^{n} y_t}$$

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^{n} |y_t - \hat{y}_t|$$

$$\text{RMSE} = \frac{1}{n} \sqrt{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}$$

where $y_t$ and $\hat{y}_t$ are the actual and predicted values, respectively, at day $t$. Furthermore, we used time-series cross validation, which is based on a rolling prediction origin.

### 6.1. Tuning parameter $\lambda$

In the arrival predictions (1) there is a tunable smoothing parameter $\lambda$. Fig. 4 shows the impact of the smoothing parameter on the WAPE and MAE for both the ICU arrivals (red lines) and occupied ICU beds 3-day ahead predictions. For the ICU beds, we use both the complete prediction model (blue lines), and the occupancy model fed by the actual arrival stream (green lines). The aim of the latter is to obtain insight in the impact of the LoS on the accuracy of the occupancy forecast, as there is no error in the arrival prediction in that case. This also implies that the green lines are not affected by $\lambda$ as this parameter only affects the arrival prediction.

The differences between the green and blue lines should be interpreted as the error in occupancy prediction that is due to the unknown arrival process. Also observe that the arrivals (red) and occupancy (blue) are at a completely different level, as will also become apparent below, explaining the differences in absolute (MAE) and relative (WAPE) errors. Clearly, for $\lambda$ very small the forecast is too responsive, whereas the opposite occurs for large $\lambda$.

We note that the behavior is similar for 7-day ahead predictions and for the clinic. In practice, it is desirable to tune the parameter $\lambda$ based on contextual information, such as measures taken, as this may improve the prediction (Sanders & Ritzman, 2001). For consistency, we use a single smoothing parameter of $\lambda = 10$ in the experiments of Sections 6.2 and 6.3.

### 6.2. Short-term predictions

In this subsection we visualize the predictions for 1, . . . , 7 days ahead for the arrivals and occupancy of both the ICU and the clinic. In Fig. 5 we present the predictions made at December 22, 2020. The arrivals are plotted on the left, with the solid lines the actual values, the blue dotted lines the fit, and the red dotted lines the predictions. The occupancies are plotted on the right, with the solid lines the actual values again, the red dotted lines the predicted values, and the blue dotted lines the predictions when the arrivals are known; the aim of the latter is to obtain insight in the impact of inaccurate predictions for the arrival process.

For the arrivals, we see a very good fit (blue line), with an apparent weekly arrival pattern, in particular for the clinic. The arrival predictions for the clinic are accurate, but for the ICU the model seems to overestimate the number of arrivals. Specifically, the increasing trend does not continue as strongly as suggested by the data up to Dec 22. This also leads to an overestimation of the number of occupied ICU beds (compare the red line with the blue line for the ICU beds). Regarding the occupancy for the clinic, there seems an overestimation of the number of occupied beds for the period from Dec 24 until Dec 28. This is not due to the arrival predictions, as the red and blue lines are rather similar. It seems likely that some patients might be discharged earlier from the clinic in the period around Christmas.

To see how the predictions behave over time, we use a rolling horizon and, for every day, make predictions for 3 and 7 days ahead. In Fig. 6 the 3-day ahead predictions (with corresponding bandwidth) together with their realizations are shown for the arrivals and occupancies for the ICU and the clinic. Overall, the predictions are visually accurate. We see that the predictions tend to deviate from the realizations at moments when the arrival pattern changes, i.e., when the arrivals reach a local peak or valley. When the number of arrivals is at such a local peak or valley, it takes.

Fig. 4. Impact of the smoothing parameter $\lambda$ on accuracy of arrival and occupancy 3-day ahead predictions at the ICU.
Fig. 5. Predictions of arrivals (left) and occupancies (right) for the ICU (top) and clinic (bottom) at December 22, 2020. Left figures show realized admissions (solid), fitted model (blue dashed), and the predictions of number of admissions (red dashed). Right figures show realized occupancies (solid), predicted values (red dashed), and predicted values for realized number of admissions (blue dashed). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 6. 3-day ahead predictions of arrivals (left) and occupancies (right) for the ICU (top) and clinic (bottom). Solid lines are realized values, red dashed lines predicted values, and the grey area the 95% prediction interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
a couple of days for the arrival prediction to detect that the local trend is changing, and this change is not caused by some random realizations. When the predictions are completely based on the time series (without further contextual information), it seems difficult to overcome such an issue. However, the prediction model is able to adapt to such trend changes after a couple of days.

Similar 7-day ahead predictions are shown in Fig. 7. We see similar phenomena for the 7-days ahead predictions as for the 3-days ahead predictions. However, the bandwidth is wider and it naturally takes more time to detect a trend change for the 7-days ahead predictions than for 3-days ahead.

6.3. Accuracy

The accuracy measures of the predictions are presented in Table 2, again for the arrivals and occupancies, and the ICU and the clinic. Clearly, the relative errors (WAPE) are largest for the admissions, which is partly explained by the fact that the number of arrivals is considerably smaller than the number of occupied beds; see also Remark 3 for the impact of scale. Moreover, it reveals that predicting arrivals is complicated for such a volatile process including changes in trend. The 3-day ahead prediction in the required number of ICU beds is remarkably accurate. Given the inherent randomness in the bed census process, see Remark 3, a WAPE of 3% seems to be the best achievable. For the 7-day ahead prediction of ICU occupancy, we see that the error is mainly determined by the error in the arrival process (9% with forecasted arrivals vs. 2% with actual arrivals). Overall, the model performs very well for the most important predictions, i.e., the ICU occupancies. Compared to the ICU, the predictions for the clinic occupancies seem not as good as expected. In particular, even with the actual arrival streams, the WAPE is still 6% and 7% for 3 and 7 days ahead, respectively. These errors can be explained by the discharge behavior at the clinic, where there are only few discharges during the weekend (which are compensated during the week). We like to emphasize that the discharge behavior during the week only has a modest impact on the prediction results in our current practice, as the predictions are only used for at specific days during the week.

Moreover, we consider the impact of the number of days ahead on the accuracy (MAE and WAPE) of the ICU predictions in Fig. 8. The red line concerns the arrivals, whereas the blue line is the prediction of the occupancy; the green line is the occupancy in case the actual arrivals are used (and deviations are due to the LoS). As the scale differs between arrivals and occupancy, the MAE is con-
siderably smaller and the WAPE considerably larger for the arrivals compared to the occupancy. Of course, the predictions become less accurate when the forecast is longer ahead. If the actual number of arrivals are known, we see that the occupancy predictions (green line) remain quite accurate even for 14 days ahead. Hence, prediction of the arrival process is crucial, in particular for predictions that are more than a week ahead.

**Remark 3.** The assessment of the accuracy of predictions is complicated by the inherent randomness in arrivals and LoS. For instance, suppose that our aim is to predict the value of a Poisson random variable with rate \( \mu \); the Poisson distribution typically reflects the randomness in arrivals or occupancy. The most accurate prediction would be \( \hat{y}_t = \mu \). In that case, with \( n \rightarrow \infty \) and using (Crow, 1958), we have \( \text{MAE} = 2\mu \frac{1}{\sqrt{\mu}} + e^{-\mu}/[\sqrt{\mu}] \), \( \text{WAPE} = \mu/\mu \), and \( \text{RMSE} = \sqrt{\mu} \). For example, for \( \mu \) equal to 50, 500, and 2000, the MAE is 5.6, 17.8, and 35.7, respectively, whereas the WAPE is 11.3%, 3.6%, and 1.8%, respectively.

6.4. LCPS implementation

The results in Sections 6.2 and 6.3 are based on a fixed smoothing parameter of \( \lambda = 10 \). In the current LCPS practice there is some manual adjustment, where the smoothing parameter is varied between 1 and 100. More specifically, for each value the 7-day prediction is plotted and used to get more insight in the sensitivity of the model and the likelihood of having a trend break. The decision maker combines the insights from our model with expert knowledge, which is e.g. based on discussions with other organizations such as RIVM, to decide on a final smoothing parameter value for the corresponding week.

Fig. 9 shows the predictions that LCPS made with our model over the last few months. The black line indicates the realized national ICU (left) and clinic (right) number of COVID-19 patients present. The red arrows present the 7-day ahead prediction that is made each Monday. For the ICU prediction, we see that the direction has always been correct except for two weeks. The first week of August, an increase in the number of patients was predicted whereas the occupation started to stabilize. Furthermore, the prediction error is typically well within the daily fluctuation of the number of patients. The accuracy measures for these 7 days ahead prediction are given in Table 3. The accuracy has been found to be satisfactory. Also, observe that the accuracy has clearly im-

![Fig. 8. Accuracy of ICU predictions for 1,…, 14 days ahead.](image)

![Fig. 9. Actual national 7-day ahead occupancy predictions for ICU (left) and clinic (right) used by the LCPS. Black lines are realized occupancies, and red arrows the communicated 7-day ahead predictions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)](image)

|               | WAPE | MAE  | RMSE |
|---------------|------|------|------|
| Beds ICU      | 5.9% | 18.35| 5.75 |
| Beds clinic   | 8.2% | 52.65| 16.15|
proved compared to the situation of a fixed smoothing parameter, as found in Section 6.3.

Note that the prediction made mid-July correctly predicted that a trend break was imminent by stating that the number of patients would increase in the subsequent week. Such insights are particularly important to avoid that hospitals continue to scale down their COVID-19 ICU capacity. The predictions in August and September may all be interpreted as a stable occupancy, such that the same ICU capacity for COVID-19 is kept. The observed minor predictions errors are more than acceptable for this use case. We note that the national ICU capacity for COVID-19 is evidently not adjusted per single ICU bed, but is adjusted in steps of around 100 ICU beds at once. Hence, from a managerial point of view, national prediction errors of about 20 patients are acceptable.

7. Conclusion, future research and discussion

In this paper, we presented a mathematical model to give short-term predictions, in the order of days, of the number of occupied ICU and clinical beds due to COVID-19. The model first predicts the arrivals and then employs a queueing-based method to convert arrivals into occupancy. The predictions for the ICU occupancy are accurate, in particular for 3 days ahead. For the clinical occupations, there is a seasonal component in discharges, with considerably less discharges during the weekend, that affects the performance of the predictions averaged over all days. An interesting topic for further research is to take the seasonal component in discharges into account as well, although this is less relevant for the 7-day ahead prediction.

Another future research direction is to investigate whether enriching the dataset with patient specific characteristics will significantly improve the prediction. For example, the age and day of arrival of each individual patient currently present may provide more information about the remaining LOS. However, this will substantially complicate the data collection as this involves privacy sensitive information. Moreover, with the current IT infrastructure used in the Netherlands, such data is typically only available after a few days, making this data less valuable.

Predictions of a couple of days ahead are crucial to properly manage ICU and clinical bed capacity and to relocate patients across the country. The framework is also suitable for longer-term scenarios, but an appropriate approximation of the behavior of the arrival process is then crucial. A topic of further research could be to use SEIR type of models for predicting the number of admissions over a longer time horizon.

Moreover, COVID-19 admissions consume a considerable part of the resources at the ICUs and clinics in the Netherlands. Additional resources were also used, such as post-anesthesia recovery beds, and anesthesiologists who worked as buddies next to the intensivists. This also reduced other forms of hospital capacity, together leading to reduced capacity for other forms of care leading to waiting lists for multiple forms of care. It is hard to quantify the impact of the delays. For example, van Giessen et al. (2020) reports up to 50,000 “healthy years of life lost” due to the first wave, based on 28% of the specialist medical care. However, some of this loss can be recovered if extra treatments are provided in the future. There is no centralized information on the length of waiting lists and the rate at which lives are lost.

From a mathematical view, it is interesting to study the impact of the second wave on the delayed care. For the moment the daily admissions have not reached the peak level of the first wave, but the rise and decline of the second wave has been much slower, leading to a higher number of patients and days of hospitalization. This inevitably leads to more delayed care, it is highly likely that waiting lists will become at least twice as long. This has a quadratic impact on the years of life lost: if twice as many patients wait on average twice as long before treatment, the total impact is 4 times higher. This amplifies the need for an efficient use of resources and good predictions of required capacity.

Acknowledgments

Part of the work has been carried out during the period that we were affiliated with the LCPS. We would like to thank Marcel de Jong and the LCPS for providing us insight in the management of COVID-19 in the Netherlands and the pleasant and fruitful cooperation.

Appendix A. Length of stay distributions

In Figs. A.10 and A.11, the length of stay distribution is displayed for the ICU and the clinic, respectively. For both cases, the data is plotted after applying the Kaplan–Meier estimator, together with lognormal, gamma, and Weibull fits.

References

Armony, M., Israelit, S., Mandelbaum, A., Marmor, Y. N., Tseytlin, Y., & Yom-Tov, G. B. (2015). On patient flow in hospitals: A data-based queueing-science perspective. Stochastic Systems, 5, 146–194.
Baas, S., Dijkstra, S., Braaksma, A., van Rooij, P., Sijnjers, F. J., Tiemessen, L., et al. (2021). Real-time forecasting of COVID-19 bed occupancy in wards and intensive care units. Health Care Management Science, 24, 402–419.
Bekker, R., & de Bruijn, A. M. (2010). Time-dependent analysis for refused admissions in clinical wards. Annals of Operations Research, 178, 45–65.
Bertsimas, D., Pauphilet, J., Stevens, J., & Tandon, M. (2021). Predicting inpatient flow at a major hospital using interpretable analytics. Manufacturing and Service Operations Management, To appear.
Broyles, J. R., Cochran, J. K., & Montgomery, D. C. (2010). A statistical Markov chain approximation of transient hospital inpatient inventory. European Journal of Operational Research, 207, 1645–1657.
Crow, E. L. (1958). The mean deviation of the Poisson distribution. Biometrika, 45, 556–559.
Davis, S., & Fard, N. (2020). Theoretical bounds and approximation of the probability mass function of future hospital bed demand. Health Care Management Science, 23, 20–33.
Eick, S. G., Massey, W. A., & Whitt, W. (1993). The physics of the M/G/∞ queue. Operations Research, 41, 731–742.
Farcomeni, A., Maruotti, A., Divino, F., Jona-Lasinio, G., & Liveson, G. (2020). An ensemble approach to short-term forecast of COVID-19 intensive care occupancy in Italian regions. Biometrical Journal, 63, 503–513.
Feldman, Z., Mandelbaum, A., Massey, W. A., & Whitt, W. (2008). Staffing of time-varying queues to achieve time-stable performance. Management Science, 54, 324–338.
van Giessen, A. et al. (2020). Impact van de eerste COVID-19 golfe op de reguliere zorg en gezondheid. https://www.rivm.nl/en/biblio/reference/337271.
Goic, M., Bocanic-Leal, M. S., Batal, M., & Basso, L. J. (2021). COVID-19: Short-term forecast of ICU beds in times of crisis. PLoS One, 16, e0245272.
Guan, J., Wei, Y., Zhao, Y., & Chen, F. (2020). Modeling the transmission dynamics of COVID-19 epidemic: A systematic review. Journal of Biomedical Research, 34, 422–430.
Gupta Strategists (2020). COVID goes Cuckoo. https://gupta-strategists.nl/storage/files/200521-COVID-goes-Cuckoo.pdf (downloaded August 24, 2021).
Hamer, W. H. (1906). Epidemic disease in England: The evidence of variability and of persistence of type. Bedford P Press.
IHME COVID team, & Murray, C. J. (2020a). Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. MedRxiv.
IHME COVID team, & Murray, C. J. (2020b). Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European economic area countries. MedRxiv.
Ioannidis, J. P., Cripps, S., & Tanner, M. A. (2020). Forecasting for COVID-19 has failed. International Journal of Forecasting. Available at arXiv: https://doi.org/10.1016/j.ijforecast.2020.08.004.
Joy, M. P., & Jones, S. (2005). Predicting bed demand in a hospital using neural networks and ARIMA models: A hybrid approach. In 13th European symposium on artificial neural networks (pp. 27–29).
Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. Series A, 115, 700–721.
Kortbeek, N., Braaksma, A., Burger, C. A., Bakker, P. J., & Boucherie, R. J. (2015). Flexible nurse staffing based on hourly bed census predictions. International Journal of Production Economics, 161, 167–180.
van Leeuwen, R., & Koole, G. M. (2020). Demand forecasting using smoothed demand curves in hospitality. Journal of Revenue and Pricing Management.
Littig, S. J., & Iken, M. W. (2007). Short term hospital occupancy prediction. Health Care Management Science, 10, 47–66.
Ma, X., Si, Y., Wang, Z., & Wang, Y. (2020). Length of stay prediction for ICU patients using individualized single classification algorithm. Computer Methods and Programs in Biomedicine, 186, 105224.

Maguire, P., Taylor, L., & Stout, R. (1986). Elderly patients in acute medical wards: Factors predicting length of stay in hospital. British Medical Journal (Clinical Research Edition), 292, 1251–1253.

Marazzi, A., Paccaud, F., Ruffieux, C., & Beguin, C. (1998). Fitting the distributions of length of stay by parametric models. Medical Care, 36, 915–927.

Massonnaud, C., Roux, J., & Crépey, P. (2020). COVID-19: Forecasting short term hospital needs in France. MedRxiv.

Nikolopoulos, K., Pumba, S., Schafer, A., Tsionopoulos, C., & Vasilakis, C. (2021). Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. European Journal of Operational Research, 290, 99–115.

Pagel, C., Banks, V., Pope, C., Whitmore, P., Brown, K., Goldman, A., et al. (2017). Development, implementation and evaluation of a tool for forecasting short term demand for beds in an intensive care unit. Operations Research for Health Care, 15, 19–31.

Palomo, S., Pender, J., Massey, W. A., & Hampshire, R. C. (2020). Flattening the curve: Insights from queueing theory. arXiv:2004.09865.

Pang, G., & Whitt, W. (2010). Two-parameter heavy-traffic limits for infinite-server queues. Queueing Systems, 65, 325–364.

Phua, W., Ling, L., Egi, M., Lim, C. M., Divatia, J. V., … Asian Critical Care Clinical Trials Group (2020). Intensive care management of coronavirus disease 2019 (COVID-19): Challenges and recommendations. The Lancet Respiratory Medicine, 8, 506–517.

Prague, M., Wittkop, L., Collin, A., Claire, Q., Dutarte, D., & Moireau, P. et al. (2020). Population modelling of early COVID-19 epidemic dynamics in French regions and estimation of the lockdown impact on infection rate. MedRxiv.

Rees, E. M., Nightingale, E. S., Jafari, Y., Waterlow, N. R., Clifford, S., Pearson, C. A., … CMiMD Working Group (2020). COVID-19 length of hospital stay: A systematic review and data synthesis. BMC Medicine, 18, 1–22.

Rosenbaum, L. (2020). Facing COVID-19 in Italy-ethics, logistics, and therapeutics on the epidemic’s front line. New England Journal of Medicine, 382, 1873–1875.

Sanders, N. R., & Ritzman, L. P. (2001). Judgmental adjustment of statistical forecasts. In Principles of forecasting (pp. 405–416). Boston, MA: Springer.

Shankar, S., Mohakuda, S. S., Kumar, A., Nazneen, P. S., Yadav, A. K., Charterjee, K., et al. (2021). Systematic review of predictive mathematical models of COVID-19 epidemic. Medical Journal Armed Forces India, 77, 5385–5392.

Shi, P., Chou, M. C., Dai, J. G., Ding, D., & Sim, J. (2016). Models and insights for hospital inpatient operations: Time-dependent ED boarding time. Management Science, 62, 1–28.

Vekaria, B., Overton, C., Wisniowski, A., et al. (2020). Hospital length of stay for COVID-19 patients: Data-driven methods for forward planning. BMC Infectious Diseases, 21, 700.

Worthington, D., Utley, M., & Suen, D. (2020). Infinite-server queueing models of demand in healthcare: A review of applications and ideas for further work. Journal of the Operational Research Society, 71, 1145–1160.

World Health Organization (2021). Weekly operational update on COVID-19 - 1 February 2021. https://www.who.int/publications/m/item/weekly-operational-update-on-covid-19-1–february-2021.

Xiang, Y., Jia, Y., Chen, L., Guo, L., Shu, R., & Long, E. (2021). COVID-19 epidemic prediction and the impact of public health interventions: A review of COVID-19 epidemic models. Infectious Disease Modelling, 6, 324–342.

van Zelst, J. M. (2020). COVID-19 repository. https://github.com/mzelst/covid-19 (downloaded Dec 23, 2020).

Zhao, C., Tepekule, B., Criscuolo, N. G., Wendel-Garcia, P. D., Hilty, M. P., Fumeaux, T., et al. (2020). Icu_monitoring.ch: A platform for short-term forecasting of intensive care unit occupancy during the COVID-19 epidemic in Switzerland. Swiss Medical Weekly, 150, w20277.