




# A predictive model for planning emergency events rescue during COVID-19 in Lombardy, Italy

Angela Andreella<sup>1,3</sup>  · Antonietta Mira<sup>1,2</sup> · Spyros Balafas<sup>1,4</sup> · Ernst-Jan C. Wit<sup>2</sup> · Fabrizio Ruggeri<sup>5</sup> · Giovanni Nattino<sup>6</sup> · Giulia Ghilardi<sup>6</sup> · Guido Bertolini<sup>6</sup>

Accepted: 4 September 2023 / Published online: 16 October 2023  
© The Author(s) 2023

## Abstract

Forecasting the volume of emergency events is important for resource utilization in emergency medical services (EMS). This became more evident during the COVID-19 outbreak when emergency event forecasts used by various EMS at that time tended to be inaccurate due to fluctuations in the number, type, and geographical distribution of these events. The motivation for this study was to develop a statistical model capable of predicting the volume of emergency events for Lombardy's regional EMS called AREU at different time horizons. To accomplish this goal, we propose a negative binomial additive autoregressive model with smoothing splines, which can predict over-dispersed counts of emergency events one, two, five, and seven days ahead. In the model development stage, a large set of covariates was considered, and the final model was selected using a cross-validation procedure that takes into account the observations' temporal dependence. Comparisons of the forecasting performance using the mean absolute percentage error showed that the proposed model outperformed the model used by AREU, as well as other widely used forecasting models. Consequently, AREU decided to adopt the new model for its forecasting purposes.

**Keywords** Decision support system · Emergency call data · Emergency departments data · Generalized nonlinear auto-regressive additive model · Predictive models

## 1 Introduction

Emergency medical services (EMS) represent the first point of healthcare access for millions of patients worldwide in need of urgent treatments. Despite the heterogeneous organization of EMS agencies globally, they all share the primary goal of minimizing intervention time. To achieve this goal with limited resources, statistical models for forecasting the number of emergency calls and rescue missions play a crucial role in informing EMS agencies about future activity volumes, which are

subject to significant daily and seasonal fluctuations. These fluctuations are influenced by seasonal weather variations and extreme weather events (Alessandrini et al. 2011; Bassil et al. 2009; Lin et al. 2013; Attia and Edward 1998; Noble et al. 1971), such as heat waves or cyclones (Noji 2000), social and demographic factors (Schuman et al. 1977; Kamenetzky et al. 1982), and epidemiological phenomena (Al Amiry and Maguire 2021; Diaz et al. 2001), such as influenza epidemics. Additionally, the day of the week and the time of the day impact the volume of EMS activity on different timescales (Cantwell et al. 2015; Batal et al. 2001). In this complex framework, predictive models that account for all these factors can assist EMS organizations in optimal staff management and distributing ambulances and rescue equipment.

The importance of reliable predictive models became evident with the emergence of the COVID-19 pandemic. During that period, EMS organizations faced sudden and significant changes in the volume of emergency calls, and existing predictive models failed to capture this phenomenon. Thus, forecasting the number of calls and rescue missions in the near future is essential to optimize resource allocation and address the rapidly evolving needs of the served populations.

In this study, we describe the development of a predictive model to reliably forecast the number of emergency events in the Lombardy region of Italy. This research effort was driven by a request from AREU (*Agenzia Regionale Emergenza Urgenza*), the regional agency responsible for the organization and implementation of EMS, for a reliable predictive model to inform organizational decisions during and after the COVID-19 outbreak. Before the COVID-19 pandemic, AREU routinely used a predictive tool based on an unobserved component model for time series data (Harvey 1990). However, this model was only developed to forecast events in the metropolitan area of Milan, the region's capital and most populous city in the Lombardy region, and it performed poorly when the COVID-19 pandemic began. Therefore, a new family of predictive models covering different areas of Lombardy and accounting for COVID-19-related factors was urgently needed.

The aim was to provide AREU with a predictive model that is (i) computationally efficient and time-effective, allowing for daily updates and new daily predictions, (ii) feasible in terms of accessible and up-to-date covariates, and (iii) flexible to accommodate epidemiological and social changes (Kenett and Gotwalt 2022).

To fulfill this purpose, we propose a generalized additive model (GAM, Hastie and Tibshirani 2017) with a negative binomial family, nonlinear autoregressive terms (Kedem and Fokianos 2002), and exogenous covariates, which can deal over-dispersed count time series (Agresti 2015). The proposed model is flexible, allowing for nonlinear dependence between covariates and outcomes. Although the same family of models (i.e., GAM) has already been used to develop predictive models in a similar framework (e.g., Viglino et al. (2017)), our approach is the first to simultaneously address nonlinearity, auto-regressive effects (Kedem and Fokianos 2002), over-dispersion (Agresti 2015), time (Kedem and Fokianos 2002), weather (Attia and Edward 1998; Noble et al. 1971) and epidemiological (Diaz et al. 2001) factors. For a recent and exhaustive review of the literature on predictive models in the field of EMS, see Al-Azzani et al. (2021); Huang et al. (2021) and the references therein. The developed model was compared, in terms of predictive performance, to

alternative forecasting methods, including the predictive tool used by AREU before the COVID-19 pandemic.

We focused our modeling efforts on predicting the volume of emergency events, defined as the dispatch of the most appropriate transport until the rescue is completed and/or the patient is referred to the most suitable hospital facilities. Planning emergency transportation dispatches is the most crucial organizational effort from a managerial and planning perspective. While most of the literature on forecasting models in EMS has focused on predicting emergency call volumes, trends in emergency call volumes do not always align with changes in actual emergency events, which are the ultimate organizational targets. In Lombardy, approximately half of the emergency calls are not forwarded to the second-level public safety access point because they are deemed non-urgent upon initial assessment (AREU 2022). Various factors, such as media communications and social alerts, influence the proportion of calls that do not require emergency services. For example, immediately after the first COVID-19 case was confirmed in Italy, an unprecedented increase in emergency calls was observed in Lombardy. Many of these calls were requests for information and recommendations about the new disease (Marrazzo et al. 2020; Dattner et al. 2022; Molenberghs 2023).

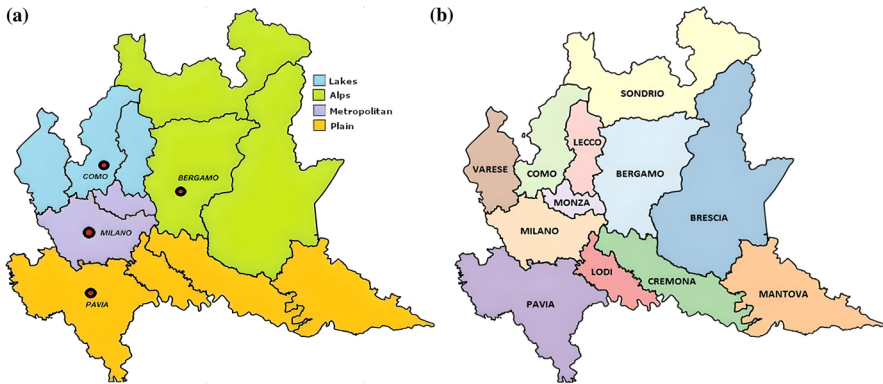
The fact that our study aims to predict the number of future emergency events is a strength. Specifically, having a reliable model to forecast such events allows the central direction unit to activate additional ambulances and increase the number of emergency teams on shift. As reported by AREU, the organization in charge of EMS in Lombardy, implementing such actions requires at least 24 hours. This designated time frame allows sufficient flexibility to adjust the number of ambulances. It ensures effective planning and response to potential emergencies, maintaining an optimal balance of resources and providing timely medical assistance to those in need (Berchi et al. 2010; Gilardi et al. 2021). Therefore, we considered predicting events one day in advance as the shortest forecasting time frame. In addition, since that data is updated hourly, the proposed model allows for considering shorter prediction intervals if deemed necessary. The model presented in this paper has been implemented into user-friendly software. Following the initial testing phase reported here, the software has been successfully used by AREU for several months.

The paper is structured as follows. Subsec. 2.1 describes the Lombardy EMS, which is the target of our research. The data sources utilized in the model development are outlined in Subsec. 2.2. The model development process is described in Subsec. 2.3, while Sect. 3 presents the estimates and predictive performance of the model. In Sect. 4, we discuss the strengths and limitations of our study.

## 2 Methods

### 2.1 EMS organization in Lombardy

Lombardy has a population of 10 million, the most populous region in Italy, accounting for approximately 17% of the Italian people. Before the COVID-19 pandemic, the regional EMS received around 1 million emergency calls annually,



**Fig. 1** Description of the AREU organizational structures: **a** four SOREUs: AREU emergency management; **b** twelve ATTs: AREU management of territorial resources

corresponding to over 900,000 emergency events. AREU, the public agency responsible for coordinating, guiding, managing, and monitoring the out-of-hospital emergency network and emergency number service in Lombardy, oversees these operations. The agency is structured into peripheral units known as Territorial Business Units (AAT, *Articolazioni Aziendali Territoriali*) as well as Regional Emergency Operations Rooms (SOREU, *Sale Operative Regionali dell’Emergenza Urgenza*). The twelve AATs are distributed throughout the region, approximately following the provincial division, while the four SOREUs coordinate rescue operations in super-provincial competence areas (i.e., AAT aggregations). The geographical position and organization of the SOREUs and the distribution of the AATs are depicted in Fig. 1.

## 2.2 Data

A comprehensive set of variables pertaining to emergency events in Lombardy from January 1st, 2015, until May 9th, 2021 were considered potential predictors for model development. We focused on predictors that are continuously available to AREU, enabling decision-makers to effectively utilize the predictive models for short-term forecasting (Kenett and Gotwalt 2022). A total of 134 variables were collected from various data sources: the AREU databases for emergency response data; the Regional Agency for Environmental Protection (ARPA) registry for weather data; the Department of Civil Protection repository for COVID-19 data; the Higher Institute of Health (ISS) estimates for other epidemiological information.

The AREU database contains information about all received calls: the SOREU receiving the call, the exact time of the call (i.e., date and hour/minutes), the call classification (e.g., first aid), the AREU administrative area (i.e., province, zone, AAT), call initiation location (e.g., home, street) with geographic coordinates, the reason for the call, the severity code (i.e., triage), and whether the call triggered an aid response, resulting in an event. The data are aggregated at an hourly level.

Weather data from ARPA data are obtained through the Open Data project (ARPA - Regione Lombardia 2014), which collects temperature, rainfall, and snowfall data from sensors distributed across the Lombardy region. For each AAT, central sensors in the city center are selected, and the weighted average is computed at the SOREU level. The weights equal the proportion of calls (from January 1st, 2015 until May 9th, 2021) received from each province in the SOREU. The weather data are available at an hourly level but were aggregated at the daily level for our analyses, using indicators such as the daily average or the difference between the maximum and minimum temperature of the day.

COVID-19 data from the Department of Civil Protection are obtained from Dipartimento della Protezione Civile (2020). The data cover the Lombardy region, except for the total number of positive cases reported at the province level. The data are available daily. In addition to the basic epidemiological information, we compute the effective reproduction number (Rodpothong and Auewarakul 2012; Dabbaghian and Mago 2014) using the method employed by the ISS, available at Istituto Superiore di Sanità (2020).

Finally, we consider the weekly flu incidence at the national level from 2015 to 2021 as communicated by ISS (Istituto Superiore di Sanità 2020).

### 2.3 Model development

Let  $Y$  be a random vector of dimension  $T \times 1$ , with its components  $Y_t$  representing the counts of events at time  $t \in \mathcal{T}$ ,  $\mathcal{T} := \{1, \dots, T\}$ . In our study, the time index  $t$  is on the hourly scale, and the spatial aggregation unit was the SOREU. Therefore, we targeted the prediction of the number of emergency events for each hour of the day in each of the four SOREUs. Let  $\mathbf{X} = (X_1, \dots, X_K)$  be a set of  $T$ -dimensional covariates associated with the response variable  $Y$ , indexed by  $k \in \mathcal{K}$ ,  $\mathcal{K} := \{1, \dots, K\}$ . Covariates can be year-, month- or day-specific and can be lags of  $Y$ , such as the total number of events in the previous days.

Given the important territorial and urban differences of the SOREUs, characterized by extremely different orographic characteristics and distributions of metropolitan and rural areas, we developed separate models for each area.

We opted for a GAM model to capture nonlinear relationships between  $X$  and  $Y$  with smoothing splines (Wood 2004, 2011, 2017). Since the response was a count variable and preliminary exploratory data analyses showed evidence of over-dispersion (Agresti 2015), we assumed a negative binomial distribution (Casella and Berger 1990) for the response. Specifically, we assumed  $Y_t \sim \mathcal{NB}(r, \pi)$ , where  $t \in \mathcal{T}$ , with  $E(Y_t) = \mu_t = \frac{r\pi}{1-\pi}$  and  $V(Y_t) = \mu_t + \frac{\mu_t^2}{\pi}$ . The expected value  $\mu_t$  of  $Y_t$  is assumed to depend on the covariates as follows:

$$\ln(\mu_t) = \alpha + \sum_{k=1}^K f_k(X_{tk}), \quad \text{where} \quad f_k(X_{tk}) = \sum_{d=1}^{D_k} \beta_{kd} b_{kd}(X_{tk}).$$

Here, the coefficients  $\beta_{kd}$  are the unknown parameters to be estimated, while  $b_{kd}(\cdot)$  are known basis functions, and  $D_k$  is the number of bases for the covariate  $X_{tk}$ .

**Table 1** Description of the five models considered in the cross-validation step

Model	Covariates
Model 1	<p>Hour of the day when the event occurs at SOREU level.★</p> <p>Day of the week when the event occurs at SOREU level.★</p> <p>Interaction between the hour of the day and the day of the week.★</p> <p>Quarter of the year when the event occurs at SOREU level.★</p> <p>Daily average temperature lagged one day at the SOREU level</p> <p>Number of events of the day before delayed 1/2/3 h at SOREU level</p> <p>Number of events aggregated by day and delayed 1/2/7 days at SOREU level</p> <p>Weekly flu incidence lagged one day at the national level</p> <p>COVID-19 reproduction number based on the total amount of positive cases lagged one day at Lombardy level</p>
Model 2	<p>Hour of the day when the event occurs at the SOREU level</p> <p>Day of the week when the event occurs at the SOREU level</p> <p>Interaction between the hour of the day and the day of the week</p> <p>Month of the year when the event occurs at the SOREU level. ★</p> <p>Daily difference between the minimum and maximum temperature ★ lagged one day at the SOREU level</p> <p>Number of events of the day before delayed 1/2/3/4 h at SOREU level</p> <p>Number of events aggregated by day and delayed 1/7/14 days at SOREU level</p> <p>Weekly flu incidence lagged one day at the national level</p> <p>COVID-19 reproduction number based on the total amount of hospitalized patients lagged one day at Lombardy level</p>
Model 3	<p>Hour of the day when the event occurs at the SOREU level. ★</p> <p>Variable indicating if the day when the event occurs at SOREU level is during the weekend or a working day</p> <p>Quarter of the year when the event occurs at the SOREU level. ★</p> <p>Week of the year when the event occurs at the SOREU level. ★</p> <p>Daily difference temperature with respect to 16° C lagged one day at the SOREU level</p> <p>Number of events of the day before delayed 1/2/3 h at SOREU level</p> <p>Number of events aggregated by day and delayed 1/2/7 days at SOREU level</p> <p>Weekly flu incidence lagged one day at the national level</p> <p>Number of hospitalized patients lagged one day at Lombardy level</p>
Model 4	<p>Hour of the day when the event occurs at the SOREU level</p> <p>Variable indicating if the day when the event occurs at SOREU level is during the weekend or a working day</p> <p>Interaction between the hour of the day and working days/weekend dummy variable</p> <p>Month of the year when the event occurs at the SOREU level. ★</p> <p>Daily average temperature lagged one day at the SOREU level. ★</p> <p>Daily average precipitation lagged one day at the SOREU level. ★</p> <p>Number of events of the day before delayed 1/2/3 h at SOREU level</p> <p>Number of events aggregated by day and delayed 1/2/7/14 days at SOREU level</p> <p>Weekly flu incidence lagged one day at the national level</p> <p>Number of intensive care admissions lagged one day at Lombardy level</p>

**Table 1** (continued)

Model	Covariates
Model 5	Hour of the day when the event occurs at the SOREU level. ★ Dummy variable describing the exceptional days, like 1st of January when there are events that cause an increase in calls Dummy variable describing the calendar school Daily average temperature lagged one day at the SOREU level. ★ Daily average precipitation lagged one day at the SOREU level Number of events of the day before delayed 1/2/5 h at SOREU level Number of events aggregated by day and delayed 1/2/7/14 days at SOREU level Weekly flu incidence lagged one day at the national level Number of daily total positive cases at the province level lagged 5 days The sum is computed to have the variables at the level of SOREU

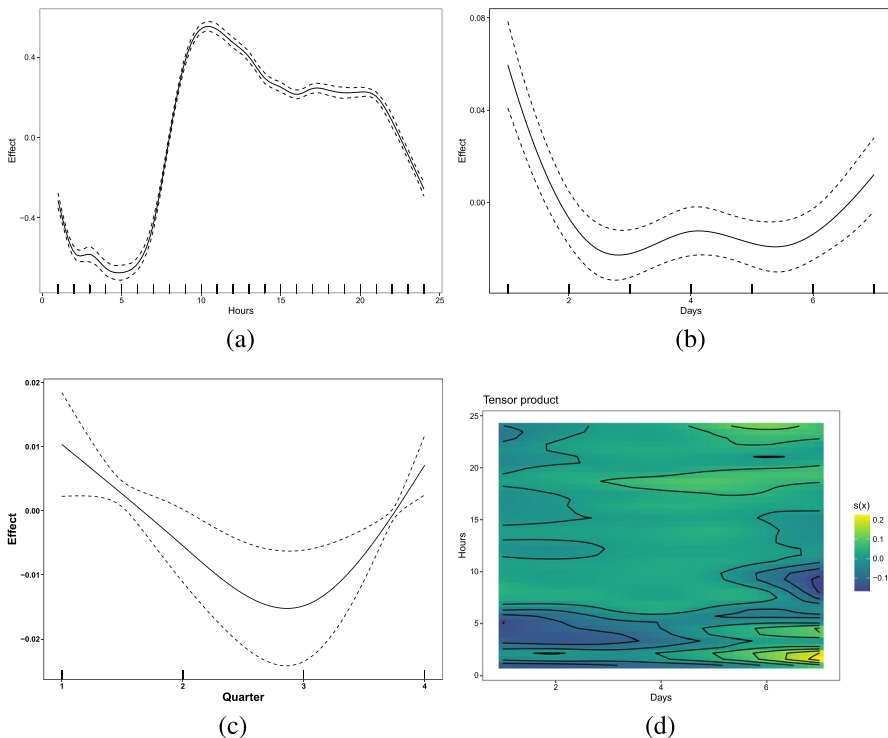
The ★ symbol expresses if the corresponding variable enters by a spline function

**Table 2** MAPE from five different models described in Table 1 calculated across different time periods (30 days) considering respectively one, two, five, and seven days ahead predictions for the *Plain* SOREU

Model	Year	MAPE			
		One day ahead	Two days ahead	Five days ahead	Seven days ahead
Model 1	2016	<b>4.131</b>	<b>4.206</b>	<b>6.047</b>	<b>7.157</b>
Model 2	2016	4.987	5.218	6.886	8.388
Model 3	2016	4.679	5.195	7.063	8.039
Model 4	2016	4.47	4.924	7.088	9.042
Model 5	2016	4.334	4.745	6.293	7.488
Model 1	2018	<b>5.072</b>	4.617	<b>4.67</b>	4.942
Model 2	2018	<b>5.212</b>	<b>4.415</b>	4.772	<b>4.719</b>
Model 3	2018	6.884	6.719	7.891	8.93
Model 4	2018	5.302	4.654	4.959	4.787
Model 5	2018	5.639	4.643	5.284	4.983
Model 1	2020	<b>4.648</b>	<b>5.334</b>	<b>4.573</b>	<b>4.264</b>
Model 2	2020	4.75	5.466	4.951	4.778
Model 3	2020	4.931	6.292	5.888	5.84
Model 4	2020	5.893	6.911	6.489	5.183
Model 5	2020	5.879	6.654	6.671	5.187

We had to perform variable selection to find the optimal model to predict the number of events in each SOREU. The variable selection process was based on the data from the *Plain* SOREU, the second largest SOREU. We did not use the *Metro-politan* SOREU, which is the largest among the four, because AREU already had a predictive model for this area, and the development of a model for the remaining SOREU was their priority.

Given the large number of available explanatory variables (i.e., 134), we relied on assumptions a priori, based on the existing literature, about their effects on the count of events and their possible interactions. We assumed the presence of a time effect at the annual, weekly, and daily levels. In addition, we assumed an effect of weather (Attia and Edward 1998; Noble et al. 1971), epidemiological (Diaz et al. 2001) and auto-regressive factors (Kedem and Fokianos 2002). We constructed several sets of covariates to explain seasonal, weather, epidemiological, and auto-regressive effects with the available variables. For example, we explored whether the seasonal effect was better modeled at the month or quarter levels. In the first case, the variable “month” was entered into the model, while in the second case, we entered the variable “quarter”. Similarly, the meteorological impact was analyzed by considering the average daily temperature and the difference between the daily minimum and maximum temperature. Again, the impact of the COVID-19 spread was analyzed through the number of hospitalized patients with symptoms and the number of positive swabs. These are just a few examples of the combinations analyzed in this variable selection process.

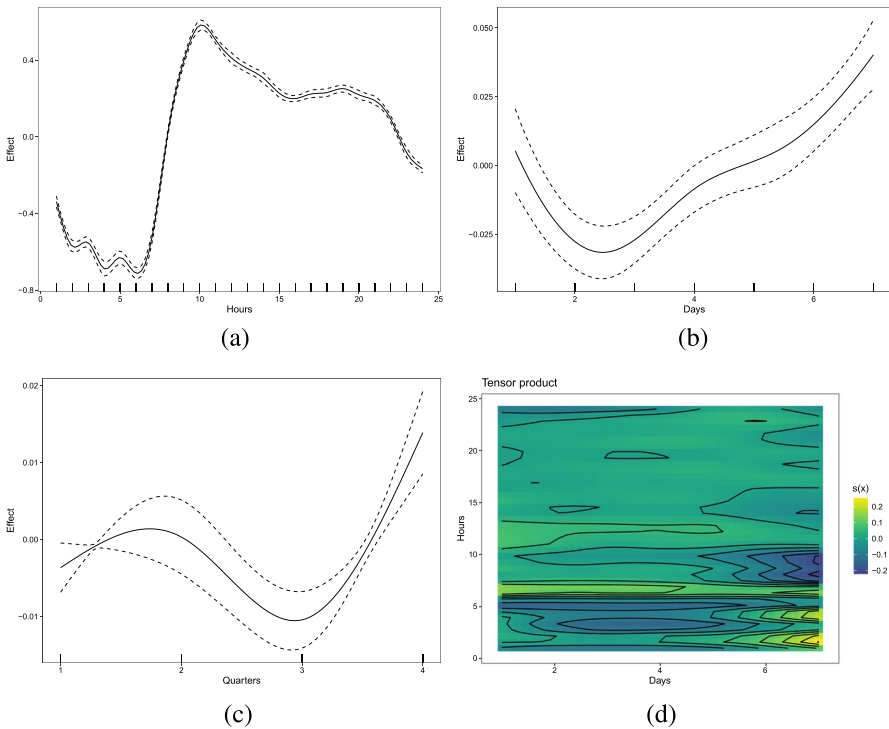


**Fig. 2 Plain model results:** coefficients plots during the COVID-19 pandemic, **a** Effect of the hour of the day; **b** Effect of the day of the week, where 1 stands for Monday; **c** Effect of the quarters of the year, where 1 stands for the first quarter of the year; **d** Effect of the interaction between days and hours, where 1 stands for Monday. The black dotted lines represent two standard error bounds



After that, we analyzed this set of covariates through exploratory analysis to understand the underlying relationships between the dependent variable (i.e., count of events) and covariates and among the covariates themselves. We finally selected five models with different sets of covariates  $\mathbf{X}$  cross-validated across different time periods. Table 1 briefly describes the set of variables inserted in these five final models. Some of them are considered categorical variables in some models, and others as continuous variables inside a spline function. The  $\star$  symbol in Table 1 expresses if the ordinal variable is considered a continuous variable inside a spline function.

The observations in time-series data are temporally dependent, which means that traditional cross-validation procedures do not apply directly. To tackle this issue, a rolling cross-validation approach suitable for temporal data was preferred. Particularly, the candidate models were fitted to data  $Y_1, \dots, Y_t$ , forecasts  $\hat{Y}_{t+1}$  were obtained, and the relative prediction errors  $E_t = \frac{\hat{Y}_{t+1} - Y_{t+1}}{Y_{t+1}}$  were calculated for each model. This was repeated for  $t = m, \dots, m + N - 1$ , where  $m$  corresponds to the minimum size of the dataset used to fit the models and  $N$  is the number of forecasts. The mean absolute percentage error (MAPE) for each model, computed across  $N$  forecasts, was used as a performance metric in the cross-validation step:



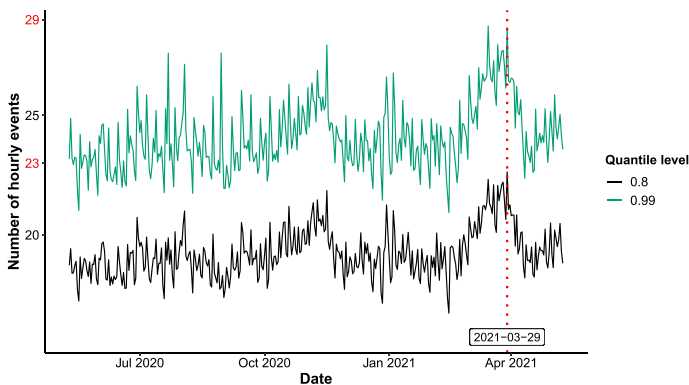
**Fig. 3 Plain model results:** coefficients plots using data before the COVID-19 pandemic, **a** Effect of the hour of the day; **b** Effect of the day of the week, where 1 stands for Monday; **c** Effect of the quarters of the year, where 1 indicates the first quarter of the year; **d** Effect of the interaction between days and hours, where 1 stands for Monday. The black dotted lines represent two standard error bounds

$$MAPE = \frac{\sum_{i=1}^N |E_i|}{N} 100, \quad (1)$$

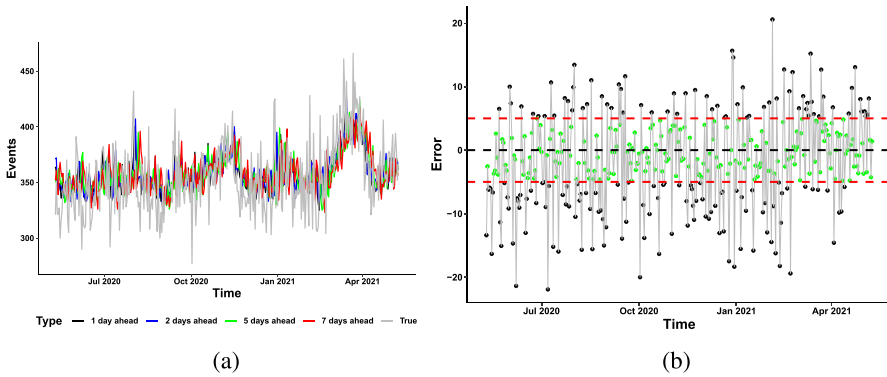
where  $|\cdot|$  is the absolute value, and  $E_i$  is the relative prediction error from forecast  $i$ ,  $i = 1, \dots, N$ . The choice of the relative prediction error for assessing forecasting performance was based on matching the performance measure used by AREU. However, the user-friendly software where the final model is implemented also reports other performance metrics, such as the minimum absolute percentage error, the maximum absolute percentage error, and the median absolute percentage error.

Table 2 shows the MAPE considering the five different models described in Table 1 cross-validated on a rolling basis across different time periods for the *Plain* SOREU. The MAPE is computed across 30 days considering one, two, five, and seven days ahead predictions in 2016, 2018, and 2020. Therefore, for each year, 30 days were randomly selected. Then, in the first step, the training set is composed of the period before the first day of the 30 days randomly selected. The test set comprises the first, second, fifth, and seventh days of the 30 days selected to have predictions one, two, five, and seven days ahead. In the second step, the first day of the 30 days randomly selected is added to the training set, the remaining 29 days are used as a test set, and so on. Model 1 has the lowest cross-validated MAPE overall compared to the other four models.

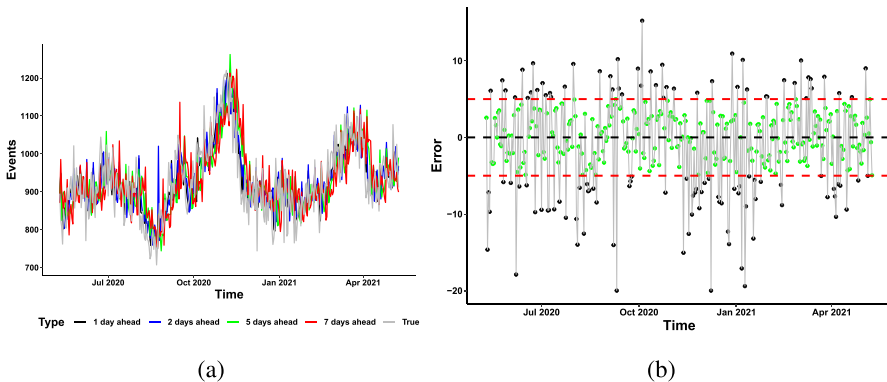
The predictive performance of the developed models was compared to that of other alternative approaches. The first comparator was the model used by AREU for the *Metropolitan* SOREU, which was based on an unobserved component time series model (Harvey 1990). The model included several covariates such as a dichotomous variable for increased calls (e.g., snowfall), a dichotomous variable for decreased calls (e.g., midweek holiday), predicted average daily temperature, flu incidence value (estimated), and the square of the difference between the average daily temperature and a reference temperature (16°C). The model also considered stochastic seasonality at the weekly level. The developed model was also compared



**Fig. 4** *Plain* model results: Quantiles at level 0.8 (black) and 0.99 (light green) considering the predicted events at hourly level one day ahead. The red dotted line shows the maximum quantile at level 0.8 (0.99) corresponding to 23 (29) events expected on March 29, 2021



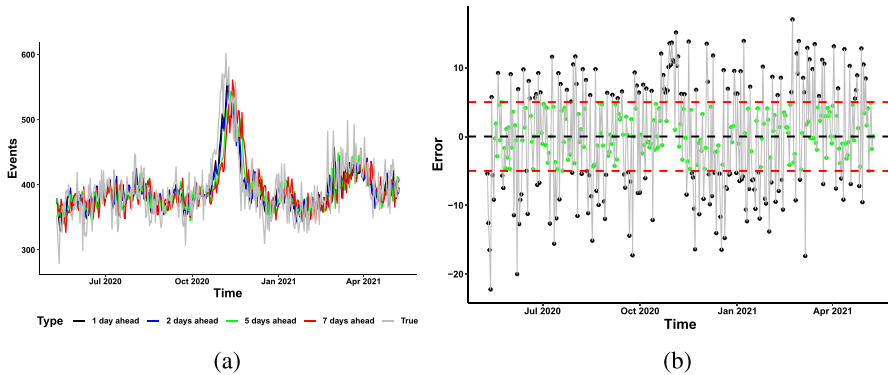
**Fig. 5** *Plain model* results: **a** Predictions one, two, five, and seven days ahead across one year and relative true values; **b** Forecast errors one day ahead, the predictions inside the two red dotted horizontal lines have absolute errors below 5%. The black dotted line represents an absolute error equals 0



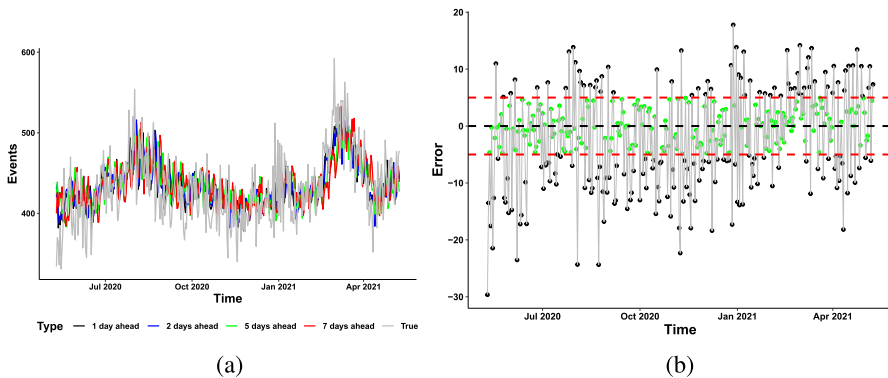
**Fig. 6** *Metropolitan model* results: **a** Predictions one, two, five, and seven days ahead across one year and relative true values; **b** Forecast errors one day ahead, the predictions inside the two red dotted horizontal lines have absolute errors below 5%. The black dotted line represents an absolute error equals 0

to other benchmarking models estimated on the same training data. These included a naive, deterministic model where predictions were simply set to the latest available observation, an ARIMA model based on the time series of previous events (Makridakis and Hibon 1997), and a generalized linear model (GLM) for count time series (Christou and Fokianos 2014).

We use the open-source statistical software R (R Core Team 2015) for preprocessing and model fitting. Among the available R packages for fitting GAMs, we utilized the `mgcv` (Wood 2011) package, which enables the fitting of GAMs on large datasets through the `bam` function. The ARIMA and GLM benchmarking models were fitted using the `auto.arima` R function and the `tsglm` function from `tscount` package (Liboschik et al. 2017), respectively. The full code used in this paper is available at [https://github.com/angeella/Tsunami\\_project](https://github.com/angeella/Tsunami_project).



**Fig. 7 Lakes model results:** **a** Predictions one, two, five, and seven days ahead across one year and relative true values; **b** Forecast errors one day ahead, the predictions inside the two black dotted horizontal lines have absolute errors below 5%. The black dotted line represents an absolute error equals 0



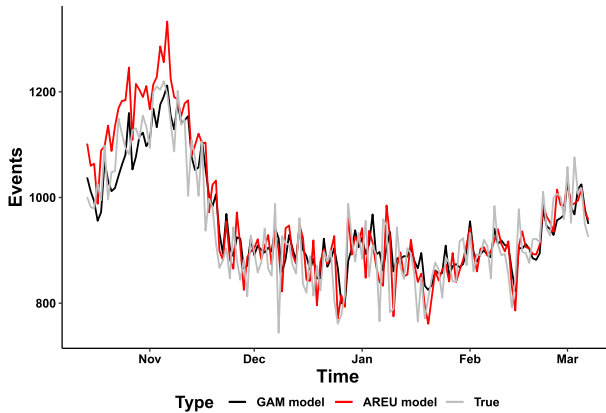
**Fig. 8 Alps model results:** **a** Predictions one, two, five, and seven days ahead across one year and relative true values; **b** Forecast errors one day ahead, the predictions inside the two black dotted horizontal lines have absolute errors below 5%. The black dotted line represents an absolute error equals 0

### 3 Results

#### 3.1 Model estimates

This section will focus on presenting the model developed specifically for the *Plain SOREU*, which was used for variable selection. The covariates included in the optimal model are described in the first model of Table 1.

We employed cubic regression splines (Green and Silverman 2019) to capture the seasonal effects of hours and quarters. The cubic splines were chosen because they match up to the second derivative with their start, which is suitable for modeling cyclical effects. Specifically, we used 24 basis functions for modeling the seasonal effects of hours and 4 basis functions for modeling the seasonal effects of quarters.

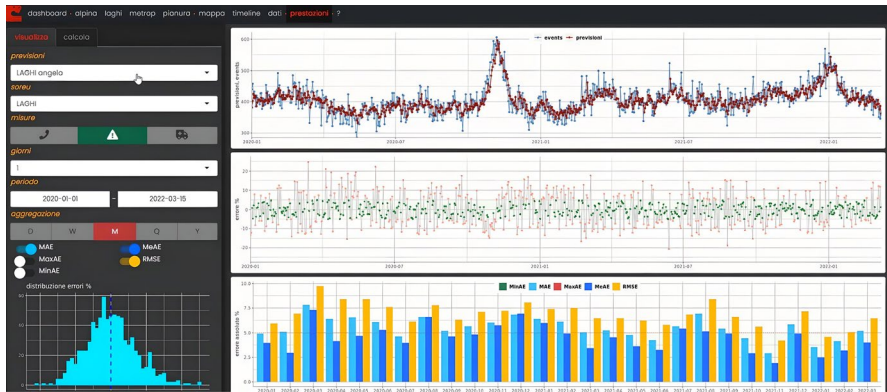


**Fig. 9 Metropolitan model results:** Predictions one day ahead across 145 days and relative true values

For the day variable, we opted for P-splines with 7 basis functions (Eilers and Marx 1996) to capture the day's seasonal effects. Finally, the tensor product smooth (Ramsay et al. 1997) (two-dimensional smooth, where the shape of one dimension varies smoothly over the other dimension) between day and hour covariates is applied to analyze the interaction between these two variables. Linear associations were assumed for the remaining covariates, which were included in the model without any smoothing terms. P-splines performed similarly to cubic regression splines but are numerically more stable and easier to fit. In addition, this permits a fast estimate of the smooth tensor product. The time complexity of the model fitting is essential in this case since AREU needs a computationally fast model to forecast the number of emergency events (Kenett and Gotwalt 2022).

For each covariate, we analyze the estimated association with the outcome. Figure 2a illustrates the estimated seasonal effects of hours within a day. We observe a positive effect during the late morning and afternoon when people are typically awake and active. In contrast, the effect becomes negative during the night and early morning when people are usually asleep. Figure 2b displays the effects of the day of the week. Mondays show the highest incidence of events compared to the other days of the week. Figure 2c presents the seasonal effects of quarters. The number of events is generally lower in the summer when the schools are closed, and people are on vacation. This observation may be associated with rising temperatures during that period. Finally, looking at the interaction between hours and days in Fig. 2d, we can see a positive effect in the morning (9 a.m.–12 p.m.) of weekdays and a strong negative effect in the morning of the weekend. Furthermore, we can see a positive effect during the night on the weekend and a negative effect around 5 a.m. on Monday, Tuesday, and Wednesday.

We estimate the model considering data before the COVID-19 era to see if the temporal dynamics, particularly the seasonal effects, have also changed due to the pandemic. Figure 3 shows the same plots presented in Fig. 2. We can see that the effect of the hour of the day (Fig. 3a) and of the quarter (Fig. 3d) do not change markedly, unlike the effect of the day of the week (Fig. 3b). One possible



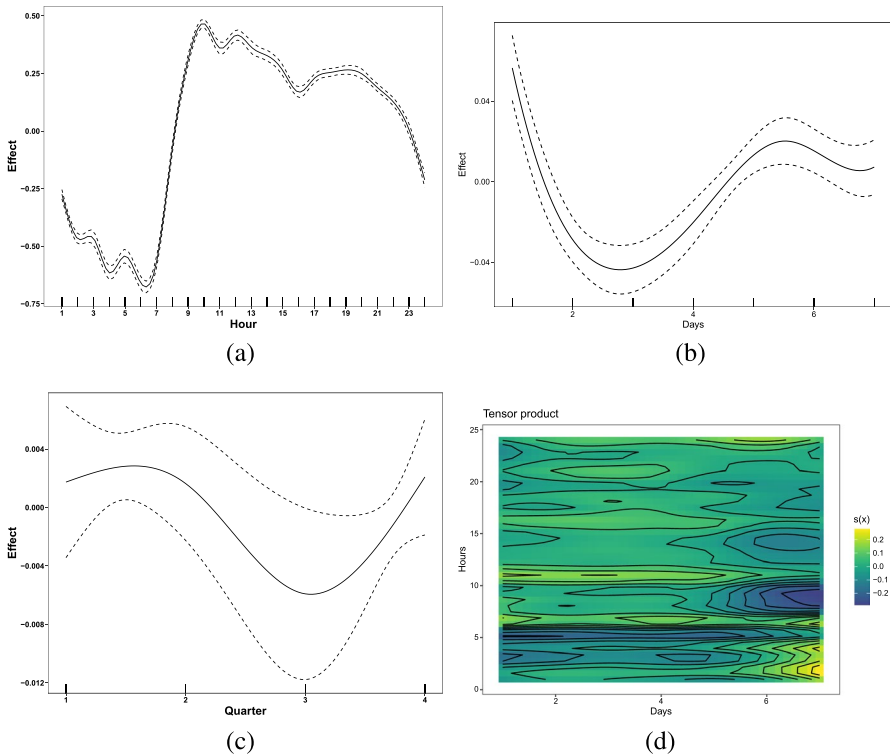
**Fig. 10** Screenshot of one of the many user-friendly interfaces that make the proposed model fully operational on a daily basis to support the decisions by AREU

interpretation could be that before the COVID-19 pandemic, many emergency events were concentrated during the weekend. Therefore, with the arrival of the COVID-19 pandemic, the emergencies referred to AREU have shifted by a few days. People become infected during the week and weekend and seek care on Monday. Before the COVID-19 pandemic, on the other hand, the primary demand, as we have said, was at weekends, when people went out to parties/hiking or when general practitioners were simply not available. Furthermore, Fig. 3d shows a positive effect on early mornings during all weeks and a night effect during the weekend. Again, before the COVID-19 pandemic, people most needed AREU interventions during weeknights and weekends.

Finally, we report the quantiles at levels 0.8 and 0.99 of the predictive distribution considering hourly predictions one day ahead. Figure 4 shows these quantiles for each day from May, 9th, 2020 until May, 9th, 2021. Knowing the quantiles of predicted values can help AREU with resource planning and making informed emergency management decisions as a risk measure. For example, AREU knows that there is 80% probability that on March, 29th 2021, the hourly events will be at most 23 (and 29 at most with probability 99%), represented by the red dotted line in Fig. 4. In other words, 23 (29) represents the cut-off point such that only 20% (1%) of the predicted emergency events exceed this value.

### 3.2 Model performance

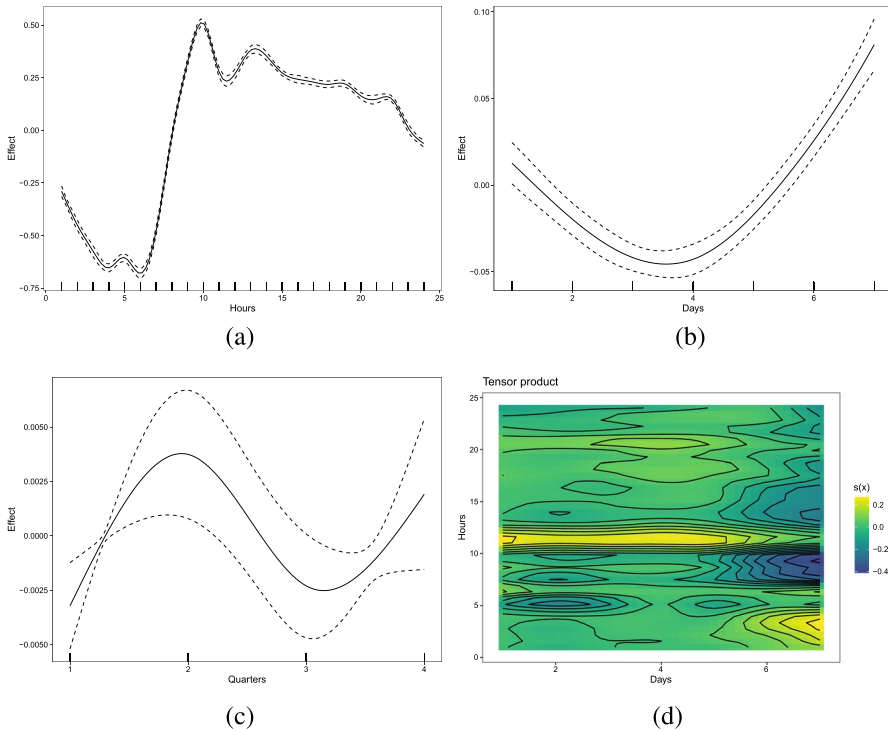
We first focus on the model's performance developed for the *Plain SOREU*. For each date from May, 9th, 2020 until May, 9th, 2021, the model described in the previous section was fitted on one year of retrospective data. The resulting model was used to forecast the number of events one, two, five, and seven days ahead. Figure 5a compares the predictions one, two, five, and seven days ahead with the true number of emergency events, while Fig. 5b shows the forecast errors for the one day ahead



**Fig. 11** *Metropolitan* model results: coefficients plots during the COVID-19 pandemic, **a** Effect of the hour of the day; **b** Effect of the day of the week, where 1 stands for Monday; **c** Effect of the quarters of the year, where 1 stands for the first quarter of the year; **d** Effect of the interaction between days and hours, where 1 stands for Monday. The black dotted lines represent two standard error bounds

prediction. The periods where the error appears to be slightly larger are during the second pandemic wave (November-February 2021) and around the holiday season. In general, the mean absolute percentage error over the whole year was 4.53%. This is a good result, considering that AREU required the mean absolute percentage error to be at most 5% for the predictions to be helpful in efficiently planning future activities. Appendix 2 shows the MAPE values grouped by month and year to have a better view of Fig. 5.

The model selected for the *Plain* SOREU data was then applied to the data of the other three SOREUs. The one day ahead mean absolute percentage error was still acceptable for the *Alps* (6.241%), *Lakes* (5.815%), and *Metropolitan* (4.309%) SOREUs. Figure 6a shows the predictions one, two, five, and seven days ahead for the *Metropolitan* SOREU, while Fig. 6b describes the prediction errors for the one day ahead forecast. Similarly, Fig. 7a and b illustrate the results for the *Lakes* SOREU, and Fig. 8a and b for the *Alps* SOREU. As the *Plain* case, Appendix 2 reports the MAPE values considering all SOREUs grouped by month and year. The



**Fig. 12** *Metropolitan model results*: coefficients plots before the COVID-19 pandemic, **a** Effect of the hour of the day; **b** Effect of the day of the week, where 1 stands for Monday; **c** Effect of the quarters of the year, where 1 stands for the first quarter of the year; **d** Effect of the interaction between days and hours, where 1 stands for Monday. The black dotted lines represent two standard error bounds

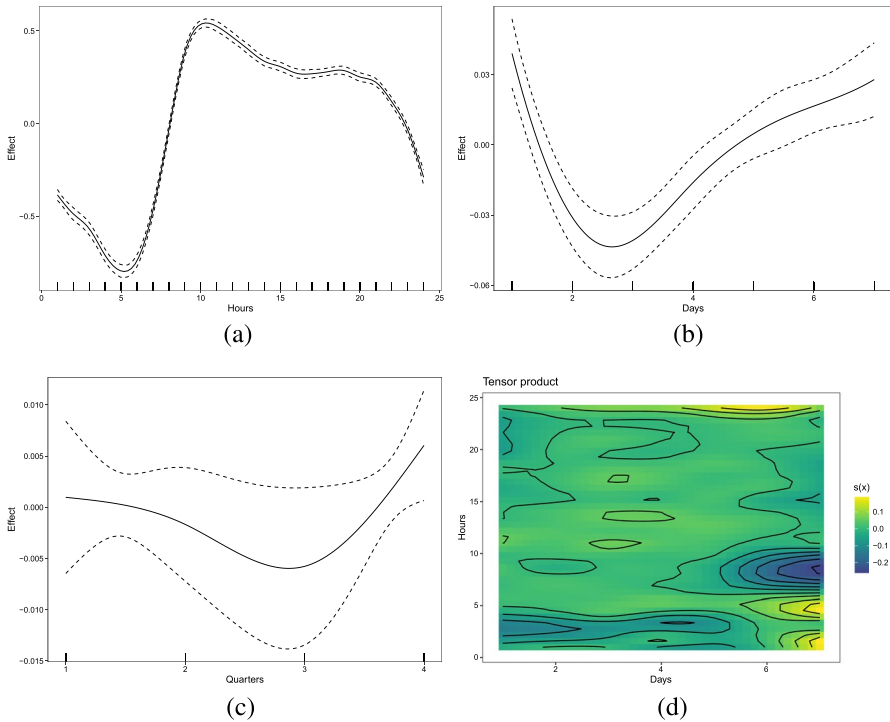
changes in the effects after fitting the model to pre and post-COVID data for the *Metropolitan*, *Lakes* and *Alps* SOREUs are reported in Appendix 1.

Figure 9 compares the one day ahead predictions obtained with the proposed model to the predictions of the model currently used by AREU on the *Metropolitan* SOREU, which was the only SOREU for which the AREU model was calibrated. The predictions are compared across 145 days (i.e., from 2020–10–14 until 2021–03–07). We observed a significant improvement: our model achieved a MAPE of 4.44% whereas the AREU model achieved a MAPE of 5.11%. Therefore, the developed GAM showed a significantly lower error than the unobserved component time series model used by AREU.

Regarding the other benchmarking models, when looking at the one day ahead predictions over one year, the naive deterministic, ARIMA and model and `tsglm` models achieved a MAPE of 7.359%, 11.177% and 10.461% respectively. Thus, the proposed model outperformed all of the considered benchmark models.

To test the flexibility of the proposed model, we trained it using data from the pre-COVID-19 period only, spanning from December 31, 2017, to December 31, 2018. Considering the *Plain* SOREU, we obtained MAPE values (calculated from



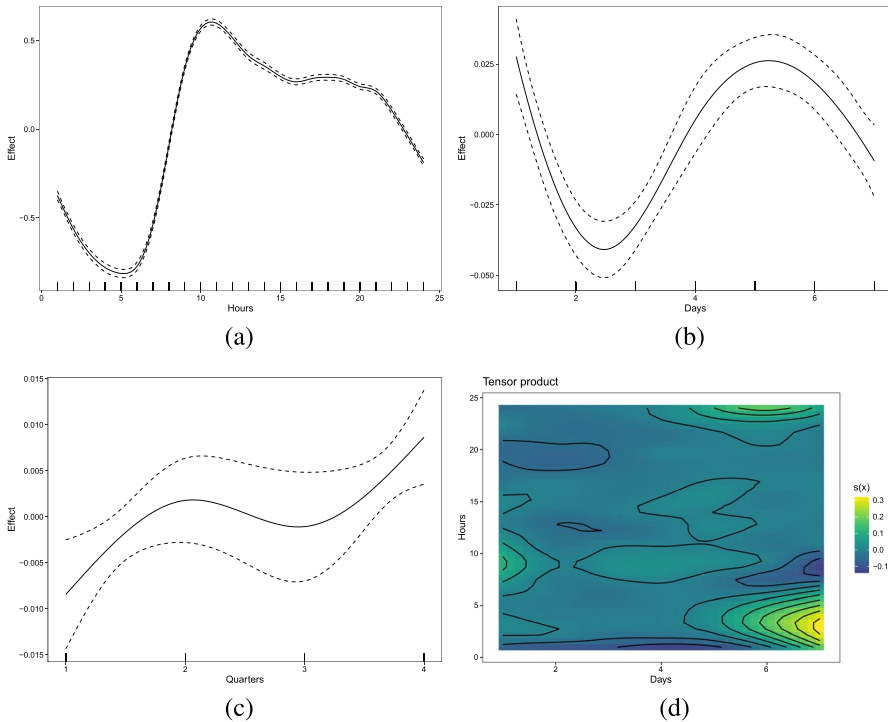


**Fig. 13** *Lakes model results*: coefficients plots during the COVID-19 pandemic, **a** Effect of the hour of the day; **b** Effect of the day of the week, where 1 stands for Monday; **c** Effect of the quarters of the year, where 1 stands for the first quarter of the year; **d** Effect of the interaction between days and hours, where 1 stands for Monday. The black dotted lines represent two standard error bounds

December 31, 2017, to December 31, 2018): 4.093 for one day ahead, 4.341 for two days ahead, 4.391 for five days ahead and 4.465 for seven days ahead. The results for the other SOREUs are provided in Appendix 2. The proposed model demonstrates its ability to capture the dynamic changes in epidemiological and societal circumstances thanks to the incorporation of selected covariates such as the effective reproduction number (Rodpothong and Auewarakul 2012; Dabbaghian and Mago 2014)), as well as the adaptability of the splines in handling variations in the number of emergency events across different periods.

### 4 Discussion

We have described the development of a statistical model developed with state-of-the-art techniques to predict emergency events in the most populous Italian region. Because the forecasting errors were compatible with the organizational needs of AREU, the model is a valuable tool for efficiently planning emergency activities.

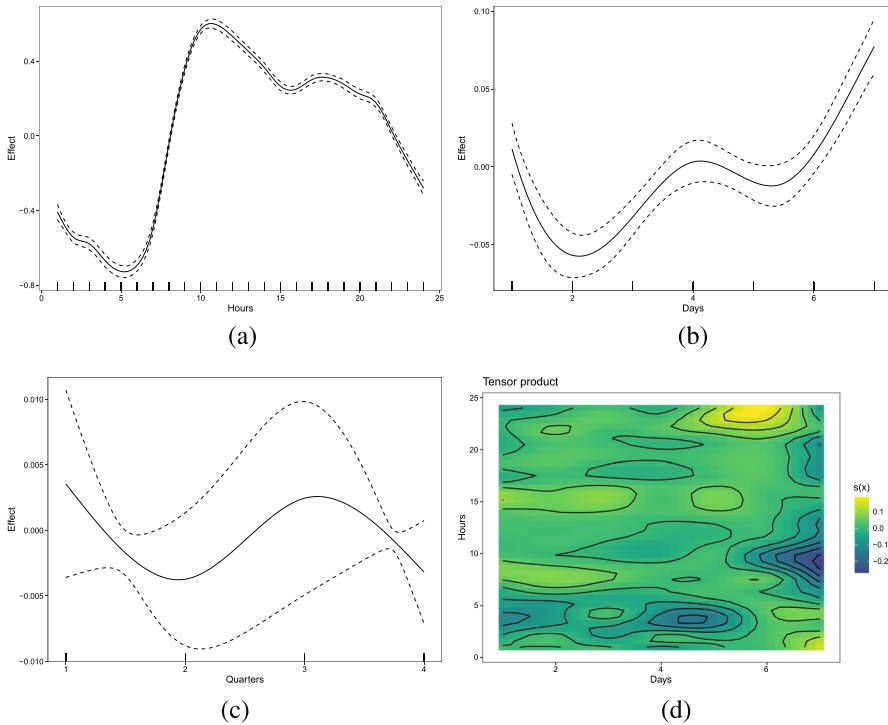


**Fig. 14 Lakes model results:** coefficients plots before the COVID-19 pandemic, **a** Effect of the hour of the day; **b** Effect of the day of the week, where 1 stands for Monday; **c** Effect of the quarters of the year, where 1 stands for the first quarter of the year; **d** Effect of the interaction between days and hours, where 1 stands for Monday. The black dotted lines represent two standard error bounds

A user-friendly dashboard providing the model estimates was implemented to make this model available to decision-makers in AREU, enabling the use of the model in everyday EMS coordination. Such an application is represented in Fig. 10.

Due to the different organization schemes of EMS around the globe, generalizing models predicting emergency events or calls to other contexts is often complex and, in some cases, impossible. However, the general methodology applied in this study and the overall model development process can also be used in other settings. Additionally, we found that our model performed very well on all of the SOREUs in Lombardy, even though the variables were selected based on the *Plain* SOREU data only. This suggests that the form of the model may generalize to slightly different contexts, such as other Italian regions or other European countries. This possibility is already under consideration by AREU decision-makers.

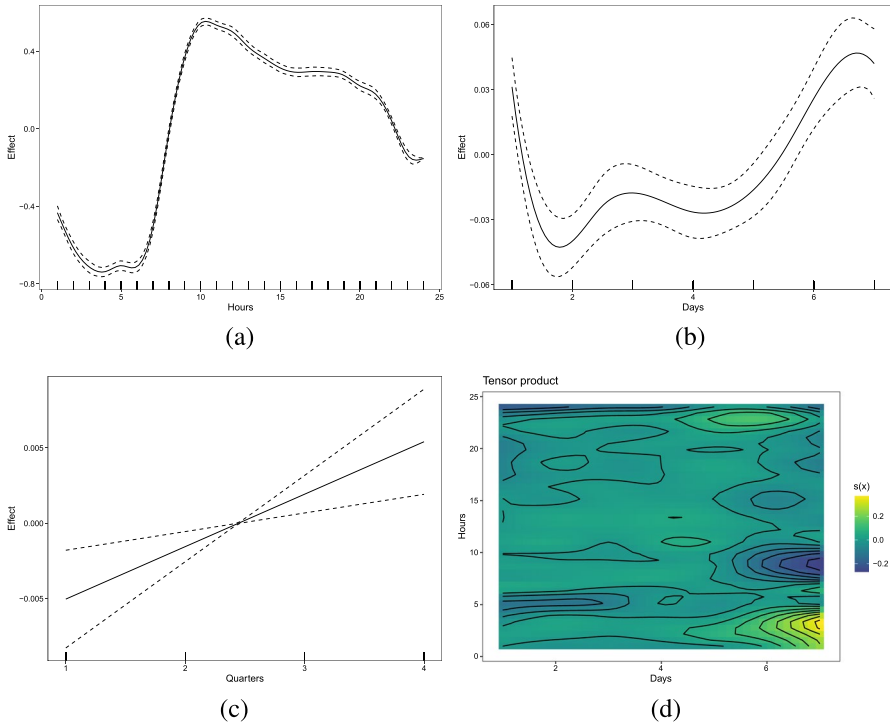
Our study focused on the GAM methodology for model development. Further research should evaluate the application of generalized additive mixed models (GAMMs), i.e., extensions of GAM incorporating random effects. GAMMs are better suited to deal with autocorrelation structures at a price, however, of higher



**Fig. 15** *Alps* model results: coefficients plots during the COVID-19 pandemic, **a** Effect of the hour of the day; **b** Effect of the day of the week, where 1 stands for Monday; **c** Effect of the quarters of the year, where 1 stands for the first quarter of the year; **d** Effect of the interaction between days and hours, where 1 stands for Monday. The black dotted lines represent two standard error bounds

computational costs (Lin and Zhang 1999). In addition, further analyses should explore the possible addition of interaction terms. Bayesian extensions could also be explored. Such models could be used to model trend and seasonal components with appropriate Markov random field priors with different forms and degrees of smoothness (Fahrmeir and Lang 2001). Finally, considering additional epidemiological variables, especially those related to COVID-19 or other emerging health crises, could be valuable and appropriate.

The proposed model was able to capture the dramatic daily and seasonal variations that emerged during the COVID-19 pandemic, attaining good performances both during one of the peaks of the COVID-19 epidemic in Lombardy and in moments when the volume of cases was limited. Nonetheless, we emphasize how predictive models like the one we presented need periodic, rigorous, quality evaluations and, eventually, substantial updates. This is particularly true in the COVID-19 era, when the rapidly increasing proportion of vaccinated individuals and the sudden spread of new viral variants may change the effect of



**Fig. 16 Alps model results:** coefficients plots before the COVID-19 pandemic, **a** Effect of the hour of the day; **b** Effect of the day of the week, where 1 stands for Monday; **c** Effect of the quarters of the year, where 1 stands for the first quarter of the year; **d** Effect of the interaction between days and hours, where 1 stands for Monday. The black dotted lines represent two standard error bounds

**Table 3** MAPE for one, two, five, and seven days ahead predictions of events volume in *Metropolitan, Lakes* and *Alps* SOREUs

SOREU	MAPE			
	One day ahead	Two days ahead	Five days ahead	Seven days ahead
Metropolitan	4.257	4.6	5.804	5.885
Lakes	6.087	6.57	6.482	6.892
Alps	5.338	5.511	6.147	6.336

COVID-19-related predictors in the model. Even beyond COVID-19 factors, structural and organizational modifications in the EMS apparatus may heavily impact the role of the predictors in the model. For these reasons, strict, continuous collaboration between public health stakeholders and data analysts is deemed essential.

**Table 4 Plain model results:** One, two, five, and seven days ahead MAPE for the *Plan* SOREU grouped by month and year

Month	Year	MAPE			
		One day ahead	Two days ahead	Five days ahead	Seven days ahead
5	2021	3.428	3.388	3.903	2.942
4	2021	3.851	4.088	4.409	4.834
3	2021	4.690	4.450	5.265	5.709
2	2021	5.958	6.214	6.882	6.201
1	2021	4.292	5.275	4.258	5.540
12	2020	6.078	7.037	6.095	5.581
11	2020	4.861	5.056	5.626	6.023
10	2020	4.373	4.161	4.263	4.250
9	2020	3.431	3.347	4.026	3.870
8	2020	6.882	7.311	6.836	6.546
7	2020	4.853	5.049	5.640	9.994
6	2020	4.897	5.821	6.475	5.943
5	2020	5.007	7.422	5.859	5.483

**Table 5 Metropolitan model results:** One, two, five, and seven days ahead MAPE for the *Metropolitan* SOREU grouped by month and year

Month	Year	MAPE			
		One day ahead	Two days ahead	Five days ahead	Seven days ahead
5	2021	3.782	5.121	6.613	6.181
4	2021	4.033	4.326	5.915	6.492
3	2021	3.630	4.085	4.344	4.701
2	2021	3.408	4.330	4.186	4.624
1	2021	5.490	5.838	6.346	7.173
12	2020	5.773	5.523	6.015	5.510
11	2020	4.533	5.360	7.753	9.697
10	2020	3.857	4.381	4.972	6.356
9	2020	4.403	9.118	6.980	7.959
8	2020	4.559	6.223	7.297	14.282
7	2020	5.181	6.362	4.173	4.484
6	2020	4.664	5.484	5.230	4.909
5	2020	4.786	4.374	6.453	6.586

## Appendix 1: Model estimates

We show here the plots of the component smooth functions for *Metropolitan* (Figs. 11 and 12), *Lakes* (Figs. 13 and 14) and *Alps* (Figs. 15 and 16) SOREUs using the entire dataset and data pre-COVID pandemic era.

**Table 6 Alps model results:** One, two, five, and seven days ahead MAPE for the Alps SOREU grouped by month and year

Month	Year	MAPE			
		One day ahead	Two days ahead	Five days ahead	Seven days ahead
5	2021	5.167	5.356	4.167	5.019
4	2021	7.385	7.547	7.499	7.919
3	2021	5.941	6.275	6.993	6.701
2	2021	4.677	5.397	5.870	6.423
1	2021	5.801	5.648	5.835	5.881
12	2020	6.254	5.786	5.992	6.157
11	2020	7.585	7.379	7.031	7.300
10	2020	6.154	6.084	5.968	6.095
9	2020	5.184	5.978	6.983	7.325
8	2020	8.265	10.084	13.004	14.502
7	2020	4.686	4.787	5.013	5.061
6	2020	5.754	6.267	5.647	5.788
5	2020	8.134	8.489	10.961	11.182

**Table 7 Lakes model results:** One, two, five, and seven days ahead MAPE for the Lakes SOREU grouped by month and year

Month	Year	MAPE			
		One day ahead	Two days ahead	Five days ahead	Seven days ahead
5	2021	3.333	4.570	6.064	4.471
4	2021	5.808	6.17	6.020	6.311
3	2021	5.447	6.175	5.697	6.091
2	2021	6.044	5.727	5.704	7.038
1	2021	6.368	6.735	6.198	7.259
12	2020	6.769	6.877	7.771	9.065
11	2020	6.189	7.067	11.571	14.247
10	2020	4.798	6.851	6.869	6.855
9	2020	4.883	5.116	6.515	6.405
8	2020	6.313	5.237	6.811	7.773
7	2020	5.391	7.310	5.162	5.470
6	2020	6.511	5.749	6.817	7.493
5	2020	5.401	6.701	7.522	7.553

## Appendix 2: Model performance

Table 3 shows the MAPE computed during the pre-COVID-19 period (i.e., from 2017–12-31 to 2018–12-31) concerning one, two, five, and seven days ahead

predictions of events volume for the SOREUs *Metropolitan, Alps* and *Lakes*.

Tables 4, 5, 6 and 7 show the one, two, five, and seven days ahead MAPE values for the *Plain, Metropolitan, Lakes* and *Alps* SOREUs respectively grouped by month and year. The time period considered is the one represented in Figs. 5, 6, 7 and 8.

**Acknowledgements** This research was supported by the Lombardy region in Italy (DGR N. XI/3017/2020, “TSUNAMI: Prediction of the Impact of the COVID-19 Wave on the Lombardy Emergency and Urgency System”). This funded research projects in the health sector related to the COVID-19 emergency. The authors thank Dr. Andrea Erba, Dr. Andrea Pagliosa, and Dr. Maurizio Migliori for their helpful comments and for creating the user-friendly interfaces that make the proposed model fully operational on a daily basis as a decision support tool. The data supporting this study’s findings are available from the corresponding author upon request.

**Author Contributions** AA: conceptualization, methodology, software, data curation, formal analysis, investigation, and writing of the original draft, review & editing. AM: conceptualization, funding acquisition, project administration, writing - review and editing, and supervision. SB: conceptualization, methodology, data curation, software, investigation, and writing - review and editing. ECW: conceptualization, writing - review and editing, and supervision. FR: conceptualization, writing - review and editing, and supervision. GN: conceptualization, writing - review and editing, and supervision. GG: supervision. GB: conceptualization, funding acquisition, writing - review and editing, and supervision.

**Funding** Open access funding provided by Università degli Studi dell’Insubria within the CRUI-CARE Agreement.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agresti A (2015) Foundations of linear and generalized linear models. John Wiley & Sons, New Jersey
- Al Amiry A, Maguire BJ (2021) Emergency medical services (ems) calls during covid-19: Early lessons learned for systems planning (a narrative review). *Open Access Emerg Med OAEM* 13:407
- Al-Azzani MA, Davari S, England TJ (2021) An empirical investigation of forecasting methods for ambulance calls—a case study. *Health Syst* 10(4):268–285
- Alessandrini E, Sajani SZ, Scotto F, Miglio R, Marchesi S, Lauriola P (2011) Emergency ambulance dispatches and apparent temperature: a time series analysis in Emilia-Romagna, Italy. *Environ Res* 111(8):1192–1200
- AREU (2022) Report of the use of the single european emergency number. <https://www.areu.lombardia.it/web/home/reportnue112>
- ARPA - Regione Lombardia (2014) Open data: Dati sensori meteo. <https://www.dati.lombardia.it/Ambiente/Dati-sensori-meteo/647i-nhxx>
- Attia MW, Edward R (1998) Effect of weather on the number and the nature of visits to a pediatric ed. *Am J Emerg Med* 16(4):374–375


- Bassil KL, Cole DC, Moineddin R, Craig AM, Lou WW, Schwartz B, Rea E (2009) Temporal and spatial variation of heat-related illness using 911 medical dispatch data. *Environ Res* 109(5):600–606
- Batal H, Tench J, McMillan S, Adams J, Mehler PS (2001) Predicting patient visits to an urgent care clinic using calendar variables. *Acad Emerg Med* 8(1):48–53
- Berchi R, Fontana G, Pagliosa A, Bonora R, Sesana G (2010) A five steps methodology for ambulance planning. In: 2010 IEEE workshop on health care management (WHCM), pp 1–5. IEEE
- Cantwell K, Morgans A, Smith K, Livingston M, Spelman T, Dietze P (2015) Time of day and day of week trends in EMS demand. *Prehosp Emerg Care* 19(3):425–431
- Casella G, Berger RL (1990) *Statistical Inference*. Cengage Learning
- Christou V, Fokianos K (2014) Quasi-likelihood inference for negative binomial time series models. *J Time Ser Anal* 35(1):55–78
- Dabbaghian V, Mago VK (2014) *Theories and simulations of complex social systems*. Springer, Berlin
- Dattner I, Gal R, Goldberg Y, Goldstein I, Huppert A, Kenett RS, Manor O, Pfeffermann D, Schechtman E, Di Serio C et al (2022) The role of statisticians in the response to covid-19 in Israel: a holistic point of view. *Israel J Health Policy Res* 11(1):1–9
- Diaz J, Alberdi J, Pajares M, Lopez C et al (2001) A model for forecasting emergency hospital admissions: effect of environmental variables. *J Environ Health* 64(3):9–15
- Dipartimento della Protezione Civile (2020) Dati andamento COVID-19 Italia. <https://github.com/pcm-dpc/COVID-19/blob/master/dati-andamento-covid19-italia.md>
- Eilers PH, Marx BD (1996) Flexible smoothing with b-splines and penalties. *Stat Sci* 11(2):89–121
- Fahrmeir L, Lang S (2001) Bayesian inference for generalized additive mixed models based on Markov random field priors. *J R Stat Soc Ser C (Appl Stat)* 50(2):201–220
- Gilardi A, Borgoni R, Mateu J (2021) A non-separable first-order spatio-temporal intensity for events on linear networks: an application to ambulance interventions. *arXiv preprint arXiv:2106.00457*
- Green PJ, Silverman BW (2019) Nonparametric regression and generalized linear models: a roughness penalty approach. Chapman and Hall/CRC, Boca Raton
- Harvey AC (1990) *Forecasting, structural time series models and the kalman filter*
- Hastie TJ, Tibshirani RJ (2017) *Generalized additive models*. Routledge, Abingdon-on-Thames
- Huang D, Wang S, Liu Z (2021) A systematic review of prediction methods for emergency management. *Int J Disaster Risk Reduct* 62:102412
- Istituto Superiore di Sanità (2020) Sorveglianza integrata COVID-19: i principali dati nazionali. <https://www.epicentro.iss.it/coronavirus/sars-cov-2-sorveglianza-dati>
- Kamenezky RD, Shuman LJ, Wolfe H (1982) Estimating need and demand for prehospital care. *Oper Res* 30(6):1148–1167
- Kedem B, Fokianos K (2002) *Regression models for binary time series. Modeling uncertainty*. Springer, Berlin, pp 185–199
- Kenett RS, Gotwalt C (2022) Functional data analysis and nonlinear regression models: an information quality perspective. *Quality Eng* 35(3):480–492
- Liboschik T, Fokianos K, Fried R (2017) tscount: an R package for analysis of count time series following generalized linear models. *J Stat Softw* 82(5):1–51
- Lin C-H, Hou S-K, Shih FF-Y, Su S (2013) The effect of tropical cyclones (typhoons) on emergency department visits. *J Emerg Med* 45(3):372–379
- Lin X, Zhang D (1999) Inference in generalized additive mixed models by using smoothing splines. *J R Stat Soc Ser B (Stat Methodol)* 61(2):381–400
- Makridakis S, Hibon M (1997) ARMA models and the Box-Jenkins methodology. *J Forecast* 16(3):147–163
- Marrazzo F, Spina S, Pepe PE, D'Ambrosio A, Bernasconi F, Manzoni P, Graci C, Frigerio C, Sacchi M, Stucchi R, Teruzzi M, Baraldi S, Lovisari F, Langer T, Sforza A, Migliari M, Sechi G, Sangalli F, Fumagalli R (2020) Rapid reorganization of the Milan metropolitan public safety answering point operations during the initial phase of the covid-19 outbreak in italy. *J Am Coll Emerg Phys Open* 1(6):1240–1249
- Molenberghs G (2023) The role of biostatistics in the response to covid-19: a Belgian and international perspective. *Israel J Health Policy Res* 12(1):1–3
- Noble JH, LaMontagne ME, Bellotti C, Wechsler H (1971) Variations in visits to hospital emergency care facilities: ritualistic and meteorological factors affecting supply and demand. *Med Care* 9(5):415–427
- Noji EK (2000) The public health consequences of disasters. *Prehosp Disaster Med* 15(4):21–31



- R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Ramsay J, Heckman N, Silverman B (1997) Spline smoothing with model-based penalties. *Behav Res Methods Instrum Comput* 29(1):99–106
- Rodpothong P, Auewarakul P (2012) Viral evolution and transmission effectiveness. *World J Virol* 1(5):131
- Schuman LJ, Wolfe H, Sepulveda J (1977) Estimating demand for emergency transportation. *Med Care* 738–749
- Viglino D, Vesin A, Ruckly S, Morelli X, Slama R, Debaty G, Danel V, Maignan M, Timsit J-F (2017) Daily volume of cases in emergency call centers: construction and validation of a predictive model. *Scand J Trauma Resusc Emerg Med* 25(1):1–7
- Wood SN (2004) Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J Am Stat Assoc* 99(467):673–686
- Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc Ser B Stat Methodol* 73(1):3–36
- Wood SN (2015) Package 'mgcv'. R Package Version 1(29):729
- Wood SN (2017) Generalized additive models: an introduction with R. CRC Press, Boca Raton

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

**Angela Andreella**<sup>1,3</sup>  · **Antonietta Mira**<sup>1,2</sup> · **Spyros Balafas**<sup>1,4</sup> · **Ernst-Jan C. Wit**<sup>2</sup> · **Fabrizio Ruggeri**<sup>5</sup> · **Giovanni Nattino**<sup>6</sup> · **Giulia Ghilardi**<sup>6</sup> · **Guido Bertolini**<sup>6</sup>

✉ Antonietta Mira  
antonietta.mira@usi.ch

Angela Andreella  
angela.andreella@unive.it

Spyros Balafas  
balafas.spyros@hsr.it

Ernst-Jan C. Wit  
ernst.jan.camiel.wit@usi.ch

Fabrizio Ruggeri  
fabrizio@mi.imati.cnr.it

Giovanni Nattino  
giovanni.nattino@marionegri.it

Giulia Ghilardi  
giulia\_irene.ghilardi@marionegri.it

Guido Bertolini  
guido.bertolini@marionegri.it

<sup>1</sup> Università degli studi dell'Insubria, Varese, Italy

<sup>2</sup> Università della Svizzera Italiana, Lugano, Switzerland

<sup>3</sup> Università Ca' Foscari di Venezia, Venice, Italy

<sup>4</sup> Università Vita-Salute San Raffaele, Milan, Italy

<sup>5</sup> Consiglio Nazionale delle Ricerche, Istituto di Matematica Applicata e Tecnologie Informatiche, Milan, Italy

<sup>6</sup> Istituto di Ricerche Farmacologiche Mario Negri, IRCCS, Milan, Italy