Received: 31 May 2020 | Accepted: 22 February 2022

DOI: 10.1111/rssc.12559

ORIGINAL ARTICLE

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Series C Applied Statistics

Modelling the extremes of seasonal viruses and hospital congestion: The example of flu in a Swiss hospital

Setareh Ranjbar[1] | **Eva Cantoni**[2] | **Valérie Chavez-Demoulin**[1] | **Giampiero Marra**[3] | **Rosalba Radice**[4] | **Katia Jaton**[5]

[1]Faculty of Business and Economics, University of Lausanne, Lausanne, Switzerland

[2]Research Center for Statistics, GSEM, University of Geneva, Geneva 4, Switzerland

[3]Department of Statistical Science, University College London, London, UK

[4]Faculty of Actuarial Science and Insurance, Bayes Business School, City, University of London, London, UK

[5]Institute of Microbiology, Lausanne University Hospital, Lausanne, Switzerland

Correspondence
Eva Cantoni, Research Center for Statistics, GSEM, University of Geneva, 1211 Geneva 4, GE, Switzerland.
Email: Eva.Cantoni@unige.ch

Abstract

Viruses causing flu or milder coronavirus colds are often referred to as 'seasonal viruses' as they tend to subside in warmer months. In other words, meteorological conditions tend to impact the activity of viruses, and this information can be exploited for the operational management of hospitals. In this study, we use 3 years of daily data from one of the biggest hospitals in Switzerland and focus on modelling the extremes of hospital visits from patients showing flu-like symptoms and the number of positive flu cases. We propose employing a discrete generalized Pareto distribution for the number of positive and negative cases. Our modelling framework allows for the parameters of these distributions to be linked to covariate effects, and for outlying observations to be dealt with via a robust estimation approach. Because meteorological conditions may vary over time, we use meteorological and not calendar variations to explain hospital charge extremes, and our empirical findings highlight their significance. We propose a measure of hospital congestion and a related tool to estimate the resulting CaRe (Charge-at-Risk-estimation) under

different meteorological conditions. The relevant numerical computations can be easily carried out using the freely available `GJRM R` package. The empirical effectiveness of the proposed method is assessed through a simulation study.

KEYWORDS

distributional regression, extreme values, flu outbreak, outliers

1 | INTRODUCTION

As demonstrated by the current health crisis with the new virus SARS-CoV-2 and its COVID-19 disease, which is currently wreaking havoc worldwide, congestion is of paramount concern for most large size hospitals. At the time of writing, hospitals across the world are experiencing congestion due to the coronavirus pandemic. There is a necessity to assist in hospitals operational management. Ideally this can be achieved by the understanding of the dynamic nature of the phenomena, such as the propagation of an infectious disease and the dynamic needs of hospital management. Once a patient is tested he/she may or may not stay in the hospital, impacting the hospital resources in terms of bed occupation and staff load. Predicting both the number of visits to the hospital and potential pandemic considerably helps the operational management of hospitals. On one hand, contrarily to the flu, historical data of hospital visits for COVID-19 are rather short and their reliability is still questionable. On the other hand, like for the flu, there are reasons to think that the COVID-19 follows a seasonal pattern with many fewer cases in the warm months. Therefore, we use flu data to address these operational management questions.

In this paper, we concentrate on the 'number of positive' and the 'number of negative' flu cases recorded at the Lausanne University Hospital[1] (CHUV). The operational motivations in studying the number of positive cases is that it indicates the hospital congestion. Alerts regarding potential congestion will help hospitals to increase capacity by different means and avoid public health hazards related to the shortage of resources. On the other hand the operational reasons behind analysing the number of negative tests are twofold. First, there is a high cost related to flu testing (approximately 180 CHF per test at CHUV). Second, there is a risk related to the possible congestion this extra unnecessary load can cause in the laboratory, which would imply that those who are truly positive have to wait longer to get their test results and be treated. This was the case in the beginning of the COVID pandemic, when running the tests was long and therefore potentially infected individuals would go around and infect others, before knowing they were positive.

Empirical evidence suggests that the flu virus is more vulnerable in warm weather, hence making it more common for epidemics to proliferate in fall and winter months in the northern hemisphere. The role of weather in the spread of flu is not yet fully understood and researchers have attempted to address this question. Roussel et al. (2016) studied the impact of six climate variables (related to temperature, humidity and sunshine) on flu spread, whereas Towers et al. (2013) analysed waves of influenza and climate patterns. Davis et al. (2012) examined the

[1]https://www.lausanneuniversityhospital.com/home

hypothesis that cold and/or dry weather enhances human pneumonia and influenza mortality, whereas Firestone et al. (2012) quantified the association between the hazard of flu infection and air temperature, humidity, rainfall and wind velocity. It has been generally found that flu transmission is mostly dependent on humidity and temperature (Lowen & Steel, 2014), with cold and dry weather making flu more active. In this paper, we approach the problem from the point of view of hospitals facing the risk of congestion and hence the need for assessing the efficiency of the flu testing process. Specifically, we aim at understanding and quantifying the impact of weather related variables on the probability of obtaining: a high number of positive flu tested patients (epidemic), a high number of negative flu tested patients (inefficiency).

We use 3 years (2016/2017, 2017/2018 and 2018/2019) of daily data[2] from 1 July 2016 to 21 June 2019 which give us $n = 1086$ observations on the number of visits and positive cases for flu at CHUV, one of the largest hospitals in Switzerland, with a capacity of 1000 somatic beds.

Figure 1 shows the number of positive cases for the 3 years considered in this study. Recording a case of flu is per se an extreme event in the sense that there are generally no flu cases registered on a 'normal' day. This justifies the use of exceedances of positive cases over the threshold of 1, and then to model these 290 exceedances (above or equal to this threshold) using extreme value theory. Although flu positive cases are registered roughly between November and May, the epidemic appears at a different time each year, with patterns that differ across the years. Similar considerations apply to the negative cases shown in Figure 2, where the chosen threshold is 15, which gives us 230 exceedances. The choice of threshold corresponds to a certain level of inefficiency of the flu testing process (each test is expensive and costs 180 CHF). The different patterns observed across years suggest that the calendar day variable is not a good predictor for use within the framework of hospital management of flu (positive or negative) cases. This may be in part due to meteorological variation across years and perhaps also to climate change. As for the latter, we do not have enough data to test for a long-term effect. Regarding the meteorological aspect, we propose to build a model for non-identically distributed discrete extremes where covariate effects can be accounted for. The discrete generalized Pareto distribution (D-GPD) provides a theoretically justified law for discrete extremes (Hitz et al., 2017), whereas, in the same spirit of generalized additive models for location, scale and shape (Rigby & Stasinopoulos, 2005), distributional parameters are made dependent on meteorological effects. The estimation approach needs to account for outlying observations as elaborated further in the next paragraph.

It is important to stress that testing patients for flu in hospital is a process that requires human intervention. As such, the recording process on certain days (e.g. 31 December) will be different as compared to that of other days. This contaminates the underlying distribution of the data by creating outliers which have to be dealt with. Dealing with outliers does not simply amount to truncating the distribution or cutting the largest observations. Rather, it consists in down-weighting the observations that are not compliant with the assumed (here the D-GPD) distribution, which are not necessarily the largest ones. This is also the case in the extreme values setting, as confirmed in the paper by Dell'Aquila and Embrechts (2006) who discusses the seemingly contradiction of using robust methods for extremes. Their Message 1 says that "Robust methods do not downweight 'extreme' observations if they conform to the majority of the data". Other papers have considered robust approaches to model extremes, confirming the needs of such an approach: see, for example, Dupuis and Field (1998), Dupuis and Victoria-Feser (2006),

[2]The dataset has been provided by the CHUV Institute of Microbiology. The access was acquired by one of our co-authors (Dr. Jaton-Ogay) who works at the Institute of Microbiology. The anonymized dataset includes all the visits to the Institute laboratory for the flu test. Then the data were aggregated on a daily basis for our analysis.

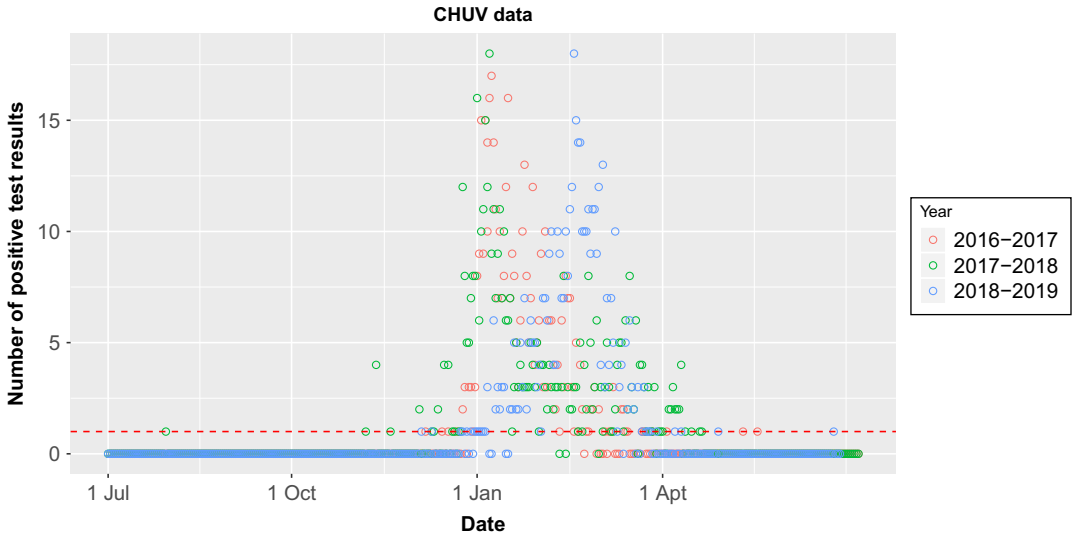

FIGURE 1 Daily number of flu positive cases tested patients from 1 July 2016 to 21 June 2019. The red line shows the threshold defining exceedances

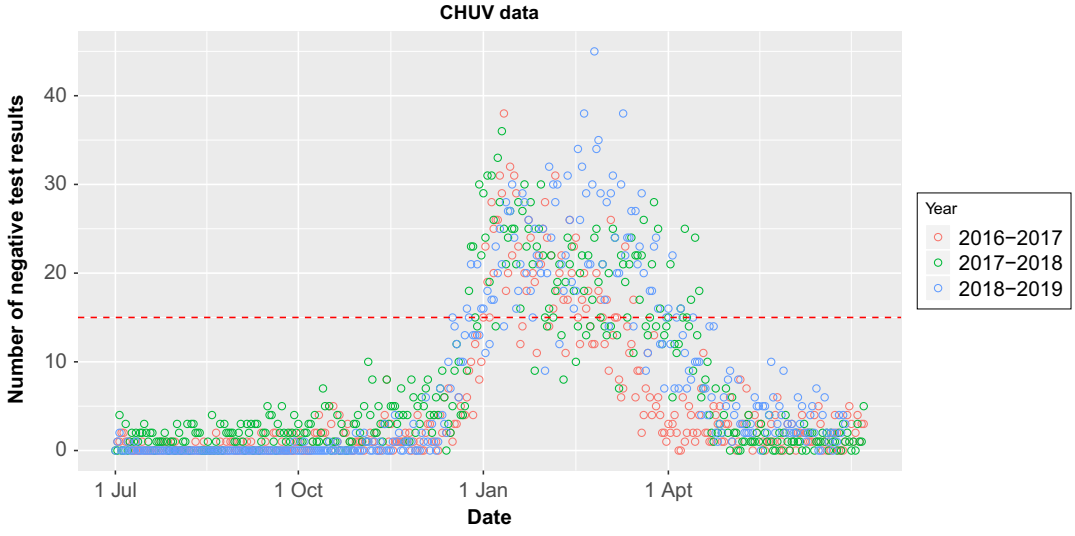

FIGURE 2 Daily number of flu negative tested patients from 1 July 2016 to 21 June 2019. The red line shows the threshold defining exceedances

Dell'Aquila and Embrechts (2006) and La Vecchia et al. (2012). This is also well exemplified in our application (see for instance, Figure 7, and related comment). We adapt the methodology introduced by Aeberhard et al. (2021) to the specific context of this paper. Although robust models for extremes have been developed in the literature, to the best of our knowledge no previous work has attempted to build extreme value models based on D-GPD, where the distributional parameters are allowed to be specified as functions of covariates and the presence of outlying observations is

accounted for by using a theoretically founded robust estimation approach of the type discussed in this paper.

The newly introduced models are available through the `GJRM R` package (Marra & Radice, 2021) which greatly simplifies the implementation of our approach, making it as simple as a canned procedure.

When we apply our proposed approach to the hospital data, we find that our two responses of interest depend on meteorological conditions. Specifically, our results suggest that the risk of congestion (extreme positive cases) increases when temperatures go down, and in periods of no sun and no rain. The risk of testing inefficiently (extreme negative cases) significantly increases in periods of no sun and no rain. We quantify these results in Section 3.

In Section 2, we describe the proposed robust regression methodology for extremes. Section 3 reports the results of the empirical analysis for the flu hospital data while Section 4 presents the findings from a simulation study. We conclude in Section 5.

2 | ROBUST REGRESSION METHODOLOGY FOR DISCRETE PEAKS-OVER-THRESHOLD

Extreme value theory is the field of statistics dedicated to the study of events with low occurrence frequencies and large amplitudes. Such events are rare in relation to the bulk of a population, which makes them hard to model and difficult to predict. Section 2.1 discusses the concepts of peaks-over-threshold and Charge-at-Risk-estimation (CaRe), and describes the approach used to derive the extreme distribution used to analyse the discrete outcomes of this paper. Section 2.2 discusses the robust approach used to estimate D-GPD model whose distributional parameters are allowed to depend on covariate effects.

2.1 | D-GPD and CaRe

One of the classical theories of extremes for a common continuous random variable is based on the probabilistic limit result for exceedances of high thresholds. The so-called peaks-over-threshold (POT) consists of a limiting result for the times and sizes of exceedances over some high level. Letting $(Y_i)_{i\geq 1}$ be a sequence of independent and identically distributed (iid) random variables in $[0, y_F)$, with common continuous distribution F, we concentrate on the sizes of exceedances over some high threshold. In other words, the focus is on the right tail of the distribution F. The result in Balkema and de Haan (1974) allows for the approximation of the conditional distribution of the exceedances above a high threshold u. If there exist normalizing sequences $\{a_n > 0\}$ and $\{b_n\}$ such that for $u \to y_F$,

$$a_n^{-1}(Y-u)|Y \geq u \xrightarrow{d} Z \tag{1}$$

for some Z following a non-degenerate probability distribution on $[0, \infty)$, where $\xrightarrow{d}$ denotes weak convergence, then it is possible to model the limiting distribution of the exceedances $Y-u|Y>u$ with a generalized Pareto distribution (GPD), that is,

$$\Pr(Y > u+y|Y>u) \underset{u\to y_F}{\to} \overline{G}_{(\sigma,\xi)}(y) = \begin{cases} (1+\xi y/\sigma)_+^{-1/\xi}, & \xi \neq 0, \\ \exp(-y/\sigma), & \xi = 0, \end{cases} \tag{2}$$

where $\xi \in IR$ is the shape parameter and $\sigma > 0$ the scale parameter. The case $\xi = 0$ is interpreted as the limit. Parameter ξ provides information about the heaviness of the tail of the underlying distribution F. More formally, condition (1) means that Y belongs to the maximum domain of attraction of an extreme value distribution with shape parameter ξ and we write $Y \in \text{MDA}(\xi)$. Some examples are the Pareto distribution,

$$F(y) = 1 - \left(\frac{\kappa}{\kappa + y}\right)^{\alpha}, \quad \alpha, \kappa > 0, \quad x \geq 0,$$

for which we take $a_n = \kappa n^{1/\alpha}/\alpha$, $b_n = \kappa n^{1/\alpha} - \kappa$, and is in MDA$(1/\alpha)$ (Fréchet case). The exponential distribution, $F(y) = 1 - e^{-\lambda y}$, $\lambda > 0, y \geq 0$, for which we take $a_n = 1/\lambda$, $b_n = (\log n)/\lambda$, and is in MDA(0) (Gumbel case).

Essentially all commonly encountered continuous distributions are in the maximum domain of attraction of an extreme value distribution. If the tail of F decays like a power function then F is in MDA(ξ) with $\xi > 0$. Distributions such as Burr, log-gamma, Cauchy, Pareto and Student-t as well as various mixture models are heavy tailed. The Gumbel class characterized by $\xi = 0$ contains light-tailed distributions such as the Gaussian, log-normal, exponential and gamma whose tails decay roughly exponentially. The so-called Weibull class, defined by $\xi < 0$, contains distributions that are bounded above (e.g. uniform and Beta distributions). In other words, for a wide class of distributions, the distribution of the excesses over a high threshold can be approximated by the GPD. This result suggests that if we choose u high enough then we can assume that result (2) holds for some parameters ξ and σ. In practice, such parameters are estimated by fitting a GPD to the excess amounts over the threshold u, relying on the standard properties of maximum likelihood estimators for $\xi > -0.5$.

Now consider a discrete random variable R. The use of the GPD to approximate the distribution's tail behaviour can be inappropriate. As pointed out by Hitz et al. (2017), many common distributions such as the Poisson, geometric and negative binomial are not in any maximum domain of attraction. Hitz et al. (2017) proposed two methods for modelling the tails of discrete observations from distributions with infinite support. In this paper, we briefly recall one of them. Defining the discrete maximum domain of attraction as D-MDA, we write $R \in$ D-MDA(ξ) with $\xi \geq 0$ if there exists a continuous random variable Y such that $P(R \geq r) = P(Y \geq r)$ for $r = 0, 1, 2, \ldots$ Then, for large integers u, we have $P(R - u = r | R \geq u) = P(Y - u \geq r | Y \geq u) - P(Y - u \geq r + 1 | Y \geq u)$, which, from (2), tends to a discrete generalized Pareto distribution (D-GPD) defined by

$$DG_{(\sigma,\xi)}(r) = \overline{G}_{(\sigma,\xi)}(r) - \overline{G}_{(\sigma,\xi)}(r + 1), \tag{3}$$

for $r = 0, 1, 2, \ldots$ (see Hitz et al. (2017) and references therein). The limiting case where $\xi = 0$ will be referred as D-GP0 here after. The constraint $\xi \geq 0$ comes from the fact that the discrete random variable $R \in$ D-MDA(ξ) for some $\xi \geq 0$ if and only if $R \in$ D-MDA(ξ) and R is long-tailed (Shimura, 2012).

Taking the point of view of the hospital's risk management, an important quantity is the quantile of R given that $R > u$ because it quantifies the information about the charge load the hospital has to be ready for under different scenarios of congestion. In practice, for a fixed threshold u, given that $R > u$, and a horizon of h days, the p-quantity, called Charge-at-Risk-estimation (CaRe), is the value of R that might be exceeded one time in h days. This definition, however, assumes stationarity and in our case it is better to use the alternative definition of the p%-CaRe as the value which on any single day is exceeded with probability $1/h$. In what follows, we derive such an

expression for the D-GPD. Let $p \in (0, 1)$ be the probability for which we seek a quantile We then solve for cases where $\xi \neq 0$

$$p = \sum_{r=0}^{q} \left(1 + \frac{\xi r}{\sigma}\right)^{(-1/\xi)} - \sum_{r=0}^{q} \left(1 + \frac{\xi(1+r)}{\sigma}\right)^{(-1/\xi)}$$
$$= 1 + \sum_{r=1}^{q} \left(1 + \frac{\xi r}{\sigma}\right)^{(-1/\xi)} - \sum_{z=1}^{1+q} \left(1 + \frac{\xi z}{\sigma}\right)^{(-1/\xi)} = 1 - \left(1 + \frac{\xi(1+q)}{\sigma}\right)^{(-1/\xi)},$$

and the following where $\xi = 0$

$$p = \sum_{r=0}^{q} \left\{\exp\left(\frac{-r}{\sigma}\right)\right\} - \sum_{r=0}^{q} \left[\exp\left\{\frac{-(r+1)}{\sigma}\right\}\right] = 1 - \exp\left(\frac{-(1+q)}{\sigma}\right).$$

The p%-CaRe for the discrete D-GPD is

$$\text{CaRe}(p)_{\text{D-GPD}} = \begin{cases} \left\lceil u + \frac{\sigma}{\xi}\left\{(1-p)^{(-\xi)} - 1\right\}\right\rceil - 1, & \xi \neq 0, \\ \lceil u - \sigma \log(1-p)\rceil - 1, & \xi = 0, \end{cases} \tag{4}$$

where $\lceil\ \rceil$ denotes the ceiling function (the smallest integer greater than or equal to) and we have $h = 1/(1-p)$. In a financial (and continuous) context, a related value is the co-called Value-at-Risk imposed by the Basel committee and used, for instance, to measure the risk of loss on a specific portfolio of financial assets.

The next section provides details on the robust estimation approach employed to fit extreme value models based on the D-GPD, where σ and ξ can be specified as functions of covariate effects.

2.2 | Covariate effects and parameter estimation

In the context of POT models for continuous variables through GPD excess size approximations, techniques that allow for flexible forms of dependence on covariates are very attractive in empirical applications (Davison & Smith, 1990). To this end, Chavez-Demoulin and Davison (2005) employed the framework of generalized additive models (Hastie & Tibshirani, 1990; Wood, 2017) to flexibly estimate the shape and scale parameters of an orthogonal reparametrization of the GPD. Yee and Stephenson (2007) proposed, instead, the use of vector generalized additive models. These modelling strategies are philosophically consistent with generalized additive models for location, scale and shape (Rigby & Stasinopoulos, 2005), where, for any continuous or discrete distribution F_θ, with $\boldsymbol{\theta}$ being a d-dimensional parameter vector with virtually any $d > 0$, all distributional parameters are allowed to depend on covariate effects. This type of modelling has received a great deal of interest since its introduction and some researchers also refer to it as distributional or multi-parameter regression. The classical and perhaps most commonly known software implementation of such models is the `gamlss R` package (Rigby & Stasinopoulos, 2005). Another implementation is available via the `gamlss()` function from the `GJRM R` package (Marra & Radice, 2021) which has been extended to incorporate the models developed in this paper.

For $i = 1, \ldots, n$, where n denotes the sample size, let Y_i be independently sampled from F_{θ_i} (with density or probability function f_{θ_i}), where $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{id})$ and $\boldsymbol{x}_i$ is a vector of covariates of dimension p (which can include binary, categorical and continuous variables, for instance). The distributional assumption of Y_i is understood to be conditional on all covariates. This is achieved by assuming for each parameter θ_{ij}, for $j = 1, \ldots, d$, that

$$g_j\left(\theta_{ij}\right) = \beta_{j0} + f_{j1}(\boldsymbol{x}_{ij1}) + \cdots + f_{jk}(\boldsymbol{x}_{ijk}) + \cdots + f_{jK_j}(\boldsymbol{x}_{ijK_j}), \tag{5}$$

where the g_j are one-to-one transformations or link functions (ensuring that the parameters range restrictions are met), $\beta_{j0} \in \mathbb{R}$ are overall intercepts, $\boldsymbol{x}_{ijk}$ denotes the kth sub-vector of covariates pertaining to term j and observation i, and the K_j functions $f_{jk}(\cdot)$ represent generic covariate effects (which can be of any pre-specified parametric form such as linear or quadratic, or can be non-parametric). Each of these functions are approximated by a linear combination of J_{kj} basis functions $b_{kjl}(\boldsymbol{x}_{ikj})$ and regression coefficients $\beta_{kjl} \in \mathbb{R}$, that is, $f_{jk}(\boldsymbol{x}_{ijk}) \approx \sum_{l=1}^{J_{kj}} \beta_{dkj} b_{kjl}(\boldsymbol{x}_{ikj})$. This (regression spline) approach allows for a vast variety of covariate effects. We refer the reader to Wood (2017) for all the options available and that are supported by our implementation. Note that, in our case study, a forward stepwise procedure based on RAIC found that linear specifications were sufficient to model the variation in the response variables of interest.

For D-GDP, we have that $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}) = (\xi_i, \sigma_i)$, hence $d = 2$. The choices of one-to-one transformations have to guarantee that the parameters lie in their admissible definition spaces. For D-GDP, with probability function given by Equation (3), we employ

$$g_1(\xi_i) = \log(\xi_i) \text{ and } g_2(\sigma_i) = \log(\sigma_i).$$

The above choice ensures $\xi_i > 0$, which is more restrictive than the constraint $\xi_i \geq 0$ stated in Hitz et al. (2017) for the D-GPD. The cases where $\xi = 0$ will result in an estimated value for ξ very close to zero, when fitted with the D-GPD. The data analyst will then have the choice of reverting to the fit of the limiting D-GDP0 distribution. We illustrate this way of proceeding in our application.

Let $\boldsymbol{\delta}$ be the vector of the model's parameters to be estimated. This includes the coefficients associated with (5). Model fitting is performed by maximizing the log-likelihood function $\ell(\boldsymbol{\delta}) = \sum_{i=1}^{n} \ell(\boldsymbol{\delta})_i = \sum_{i=1}^{n} \log f(y_i|\boldsymbol{\theta}_i)$. Note that, although not required for our case study, our implementation supports the presence of non-parametric components. In this case, the objective function is augmented by a penalty term usually defined as $1/2\boldsymbol{\delta}'\mathbf{S}\boldsymbol{\delta}$, where $\mathbf{S}$ is a matrix that depends on the choice of basis functions for the non-parametric terms, and on a set of smoothing parameters that controls the trade-off between fit and smoothness.

If outlying observations occur in the data, classical model fitting will suffer from a lack of robustness, which will adversely affect parameter estimates. To deal with this, we adopt Aeberhard et al. (2021) methodology which essentially consists of reducing the likelihood contributions of low log-likelihood values while leaving large log-likelihood evaluations essentially unchanged. This is achieved through a function ρ_c applied to the log-likelihood components, so that the objective function becomes $\tilde{\ell}(\boldsymbol{\delta}) = \sum_{i=1}^{n} \rho_c(\ell(\boldsymbol{\delta})_i) - b_\rho(\boldsymbol{\delta})$, where

$$b_\rho(\boldsymbol{\delta}) = \sum_{i=1}^{n} b_\rho(\boldsymbol{\delta})_i = \sum_{i=1}^{n} \int \rho_c^\star(\log f(y|\boldsymbol{\delta}))\, \mathrm{d}y$$

is a correction factor ensuring Fisher consistency, and $\rho_c^\star$ is directly derived from the specified ρ_c through $\rho_c^\star(z) = \int_{-\infty}^{z} \exp(s)\rho_c'(s)ds$ with $\rho_c'(s) = \partial\rho_c(s)/\partial s$. For brevity, we use integrals in the preceding lines, but those are sums when y is discrete.

The tuning constant $c > 0$ in ρ_c regulates the trade-off between loss of estimation efficiency (should the data exactly come from the assumed model) and the magnitude of the maximum estimation bias (should the data not come from the postulated model). For any given c, ρ_c is assumed to be convex, monotonically increasing and twice continuously differentiable over $\mathbb{R}$, and to have bounded first derivative ρ_c' within $[0, 1]$. The latter can be interpreted as a multiplicative robustness weight, as one would add when weighting the estimating equations in robust M-estimation. An advantage of the approach is that it leads to a natural definition of robust information criteria.

Regarding the choice of ρ_c, Aeberhard et al. (2021) recommend using the log-logistic function first proposed by Eguchi and Kano (2001): $\rho_c(z) = \log \frac{1+\exp(z+c)}{1+\exp(c)}$, for $c > 0$, with corresponding $\rho_c^\star(z) = \exp(z) - \exp(c)\log(1 + \exp(z + c))$ and first derivative $\rho_c'(z) = \exp(z + c)/(1 + \exp(z + c))$. It holds that $\lim_{c\to\infty} \rho_c(z) = z$ so that an increasingly large c value leads to the (non-robust) original $\ell(\boldsymbol{\delta})$. The value of c is tuned via a simulation based procedure that controls how the robustness weights at the score level (represented by ρ_c') behave under data generated from the assumed model. The proportion of down-weighting can be thought of as a proxy for efficiency. The user can decide the level of down-weighting to be achieved with respect to maximum likelihood and subsequently find the value of c that meets this target. The exact error-and-trial simulation procedure is fully described in Steps 1.-4. in section 3.5 of Aeberhard et al. (2021). In our simulation and real data analyses we have targeted 95%, because this is the gold standard in the robust statistics literature. A sensitivity analysis (not shown) has confirmed that results are stable across a reasonable range of efficiencies.

Aeberhard et al. (2021) established the Fisher consistency of $\hat{\boldsymbol{\delta}}$ as well as its asymptotic Gaussian distribution and asymptotic variance–covariance matrix which can be used to construct confidence intervals. The authors also discussed a Bayesian inferential result. This is advantageous because such a result does not rely on asymptotic considerations, and intervals for non-linear functions of the model's parameters (e.g. CaRe) can be reliably and efficiently obtained via posterior simulation. The adopted estimation framework allows for the elegant construction of robust information criteria such as the robust AIC, that is,

$$\text{RAIC}(\lambda) = -2\tilde{\ell}(\boldsymbol{\delta}) + 2\texttt{edf}, \tag{6}$$

where `edf` denotes the effective degrees of freedom which provided by the trace of a matrix that depends on components of the asymptotic variance–covariance matrix of $\hat{\boldsymbol{\delta}}$ (Aeberhard et al., 2021).

In order to estimate the model's coefficients, we have extended the efficient and stable trust region algorithm proposed by Marra and Radice (2020) to our context. One of the many advantages of such an algorithm is that it does not require the orthogonality of the distributional parameters (in this case, ξ and σ). The implementation of the trust region approach required the analytical score and Hessian of the model's log-likelihood which were derived and are reported in Supplementary Section A.

While the construction and estimation of the proposed model rely on the infrastructure and modelling framework of `GJRM`, extending the software to accommodate the developments needed to address the challenges of our case study required a great deal of work. Supplementary Section B provides details on the usage of function `gamlss()` from the `GJRM R` package.

3 | MODELLING FLU EXTREMES

Flu is contagious and it can spread by airborne respiratory droplets, saliva or skin-to-skin contact and by touching a contaminated surface. In Switzerland, a sentinel surveillance system and a mandatory reporting system are used to register flu data. Flu monitoring in hospitalized patients has also been in a testing phase since 2018. From a health care managerial point of view, the number of negative cases among tested patients showing flu-like symptoms is as important as the number of positive cases.

As shown in Figures 1 and 2, the annual calendar variable does not seem to provide important insights into managing and/or preventing congestion due to a flu epidemic. Based on the literature on flu transmission (see, for instance, Roussel et al., 2016), we use meteorological variables for modelling the extremes of positive flu cases and negative cases (visiting the hospital for a flu check). The meteorological variables are represented by $L^3X_t = X_{t-3}$, where L is the usual lag operator and X_t can be each of the variables in Table 1 at day t. For each meteorological factor, we consider the respective L^3X_t value, because, for flu, the incubation time is usually between 24 and 48 h, and sometimes 72 h. When discussing the models used for the analyses, extension `L3` at the end of each covariate's name refers to L^3X_t values.

As for the flu data, the meteorological variables were measured in Lausanne (Switzerland) from 1 July 2016 to 21 June 2019 and are available at https://gate.meteoswiss.ch/idaweb/. Figure 3 shows a correlation plot between the meteorological variables, highlighting, as expected, that humidity and radiation are highly and negatively correlated, whereas radiation and minimum temperature are positively correlated. Recall that the aim is to quantify the effect of meteorological factors on the extremes of positive and negative cases. For these discrete responses, we specify two different models (estimated using the approach described in Section 2.2) based on the D-GPD. The models' robustness tuning constants c were set to 6.1 and 6.7, respectively, to achieve a level of down-weighting of 95% (see the description of the procedure in Section 2.2).

Forward variable stepwise selection based on the RAIC defined in Equation (6) was performed for the two models. We first added the covariates sequentially to model σ while keeping ξ constant and then proceeded similarly with the covariates selection to model ξ. In both processes, the considered covariate was added if it decreased the RAIC. We checked for linear and non-linear covariate effects. As already mentioned in Section 2.2, linear specifications were found to be adequate for the modelling purposes of our dataset. For the two responses, a constant model was selected for the shape parameter ξ with estimated value close to 0 meaning an underlying light tail that in fact many common discrete distributions, including geometric, Poisson and negative binomial distributions have (Hitz et al., 2017). Since for both responses, the estimated value for

TABLE 1 Potential meteorological covariates

Name	Definition	Unit
Mintemp	Daily minimum temperature at 2 m above ground level	°C
Radiation	Daily mean radiation	W/m^2
Humidity	Daily mean relative air humidity at 2 m from the ground	%
Wind	Daily maximum wind (integration 1 s)	m/s
Precipitation	Daily sum of precipitation	mm
Pressure	Daily mean atmospheric pressure with (QNH)	hPa

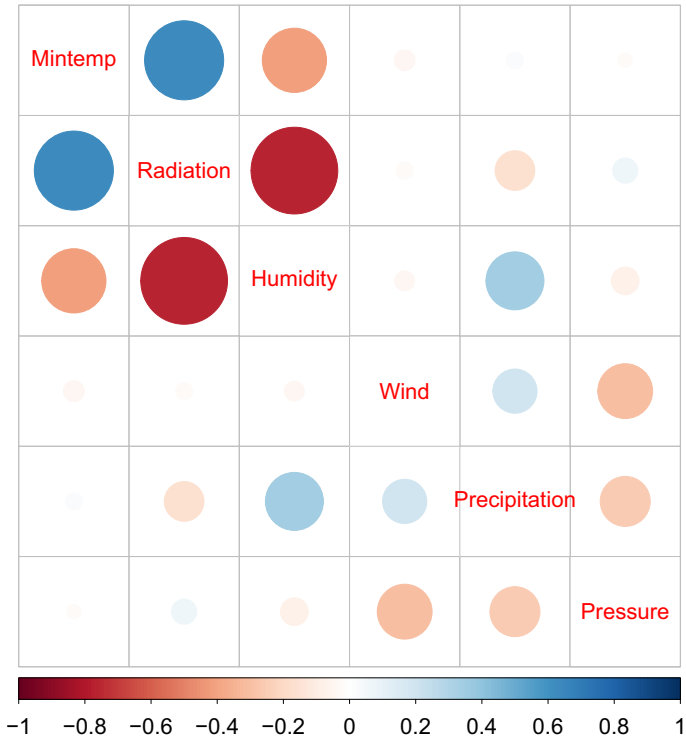

FIGURE 3 Correlation between meteorological covariates

TABLE 2 Estimated coefficients for the final D-GPD0 model fitted to positive cases

$\log(\hat{\sigma})$	Estimate	Std. error	Z-Value	*p*-value	Signif.
Intercept	2.037	0.138	14.79	0.000	***
`MintempL3`	−0.054	0.021	−2.54	0.011	*
`RadiationL3`	−0.005	0.001	−3.62	0.000	***
`PrecipitationL3`	−0.009	0.004	−2.11	0.035	*

Signif. codes: $0 < p\text{-value} < 1e^{-3}$: ***, $p\text{-value} < 0.01$: **, $p\text{-value} < 0.05$: *.

the shape parameter ξ was very close to zero, we discarded the D-GPD model and instead fitted the simplest D-GPD0 one. On the one hand, by assuming a D-GPD0, we ignore the uncertainty about ξ. On the other hand, fitting the D-GPD led to dramatically large confidence intervals for ξ.

Number of positive cases

The final stage of model selection in this case, indicates that $\xi = 0$ and the the final fit is a D-GPD0 model based on the following equations and estimates:

$$\log(\hat{\sigma}) = \hat{\beta}_0 + \hat{\beta}_1 \texttt{MintempL3} + \hat{\beta}_2 \texttt{RadiationL3} + \hat{\beta}_3 \texttt{PrecipitationL3}.$$

The estimated coefficients are reported in Table 2 and their effects graphically shown in Figure S1 in Supplementary Section C.

TABLE 3 Estimated coefficients for the final D-GPD0 model fitted to negative cases

$\log(\hat{\sigma})$	Estimate	Std. error	Z-value	*p*-Value	Signif.
Intercept	2.483	0.151	16.43	0.000	***
`RadiationL3`	−0.003	0.001	−2.15	0.031	*
`PrecipitationL3`	−0.010	0.005	−2.01	0.045	*

Signif. codes: $0 < p\text{-value} < 1e^{-3}$: ***, $p\text{-value} < 0.01$: **, $p\text{-value} < 0.05$: *.

The equation for the scale parameter σ explains both the variable's variance and mean. A broad interpretation of the results is that the warmer and nicer the weather, the lower the number of extreme positive cases, variability and mean. Interestingly, radiation seems to better explain the response than humidity, which is the factor commonly used to explain flu spread (Lowen & Steel, 2014). As highlighted by Figure 3, radiation and precipitation provide complementary proxies of humidity.

Number of negative test results

The final stage of model selection in this case, indicates that $\xi = 0$ and the final fit is a D-GPD0 model based on the following equations and estimates:

$$\log(\hat{\sigma}) = \hat{\beta}_0 + \hat{\beta}_1 \texttt{RadiationL3} + \hat{\beta}_2 \texttt{PrecipitationL3}.$$

The estimated coefficients are reported in Table 3 and their effects shown in Figure S2 in Supplementary Section C.

As compared to the previous model, minimum temperature does not seem to explain the variability and mean of the number of negative extremes. This may be due to the fact that cold weather activates the virus which in turn leads to more positive cases. Both the effects of radiation and precipitation are less important than those found when modelling positive cases. This may be explained by the fact that better meteorological conditions (warmer months and sun) simply decrease the number of test cases. For positive cases, favourable meteorological conditions also decrease the probability of catching the flu during the autumn/winter time. In other words, the evidence suggests that radiation and precipitation, which commonly affect the positive and negative cases, mostly influence the total number of tests that are carried out. During warmer months, summer or when the weather is good during winter time, individuals tend not to go the hospital. The analysis suggests that minimum temperatures influence only the spread of flu.

3.1 | Hospital congestion

From a risk management point of view, a quantity of high interest is the CaRe defined in Equation (4). This value constitutes an important risk measures for the hospital. Intervals are also crucial as they provide information about the estimates' uncertainty.

For the positive cases, the p%-CaRe corresponds to the regime of congestion. Figures 4–6 show some p%-CaRe estimated curves and respective intervals obtained via posterior simulation, based on different values of the meteorological factors. For instance, the bottom panel of Figure 4 shows the 86%-CaRe, corresponding to a time horizon of 7 days. This point estimate decreases from 18, for a minimum temperature of −10°C, to 11, for 0°C. The bottom left panel of Figure 5 shows that the 86%-CaRe decreases from 16, in the case of no sun, to 3, when radiation is at its highest value.

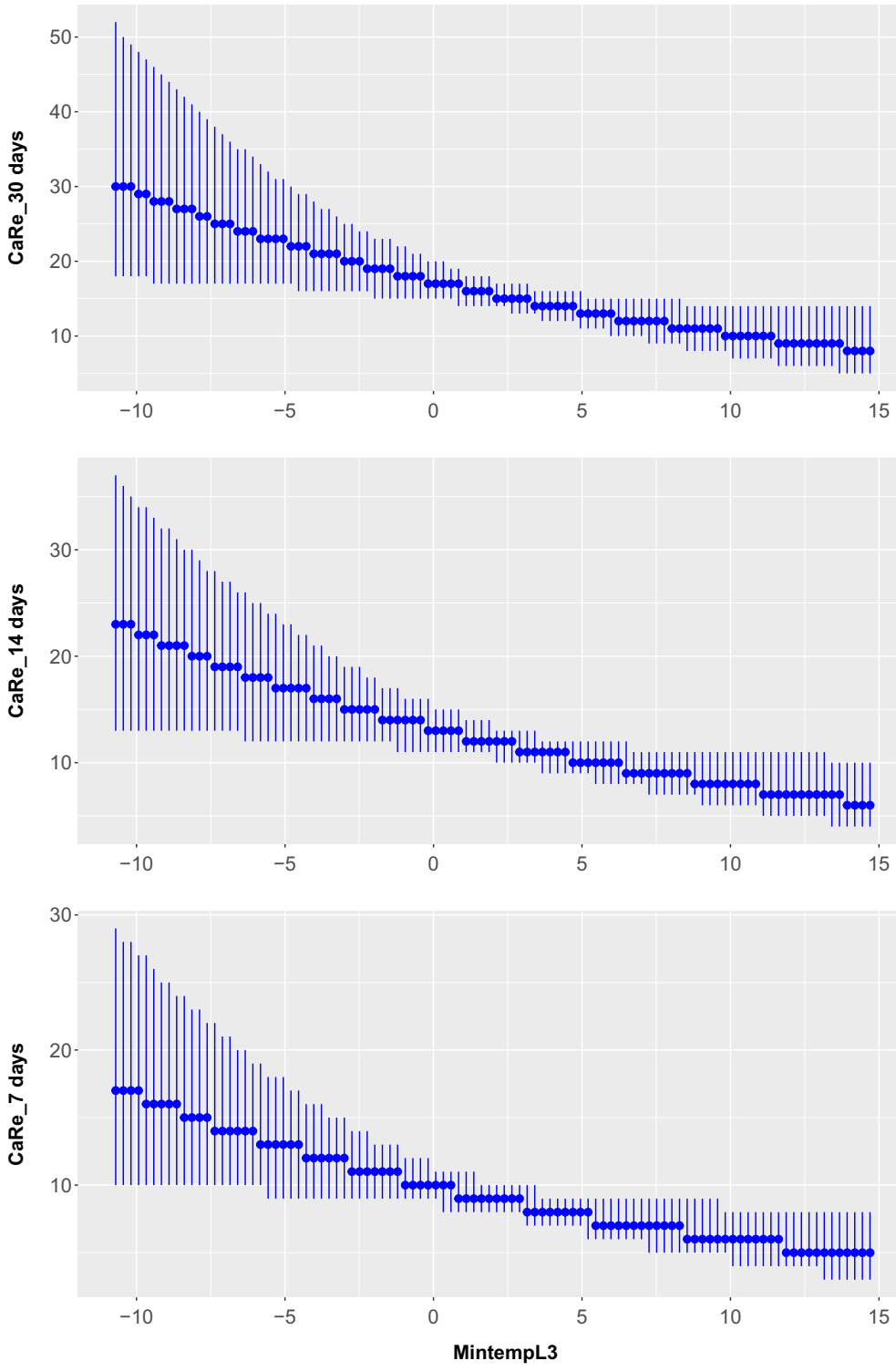

FIGURE 4 Positive cases: estimated CaRe for $h = 30, 14, 7$ days (panels from top to bottom respectively) with respect to lagged minimum temperature. The bars correspond to 95% intervals. The other covariates are fixed to their mean values

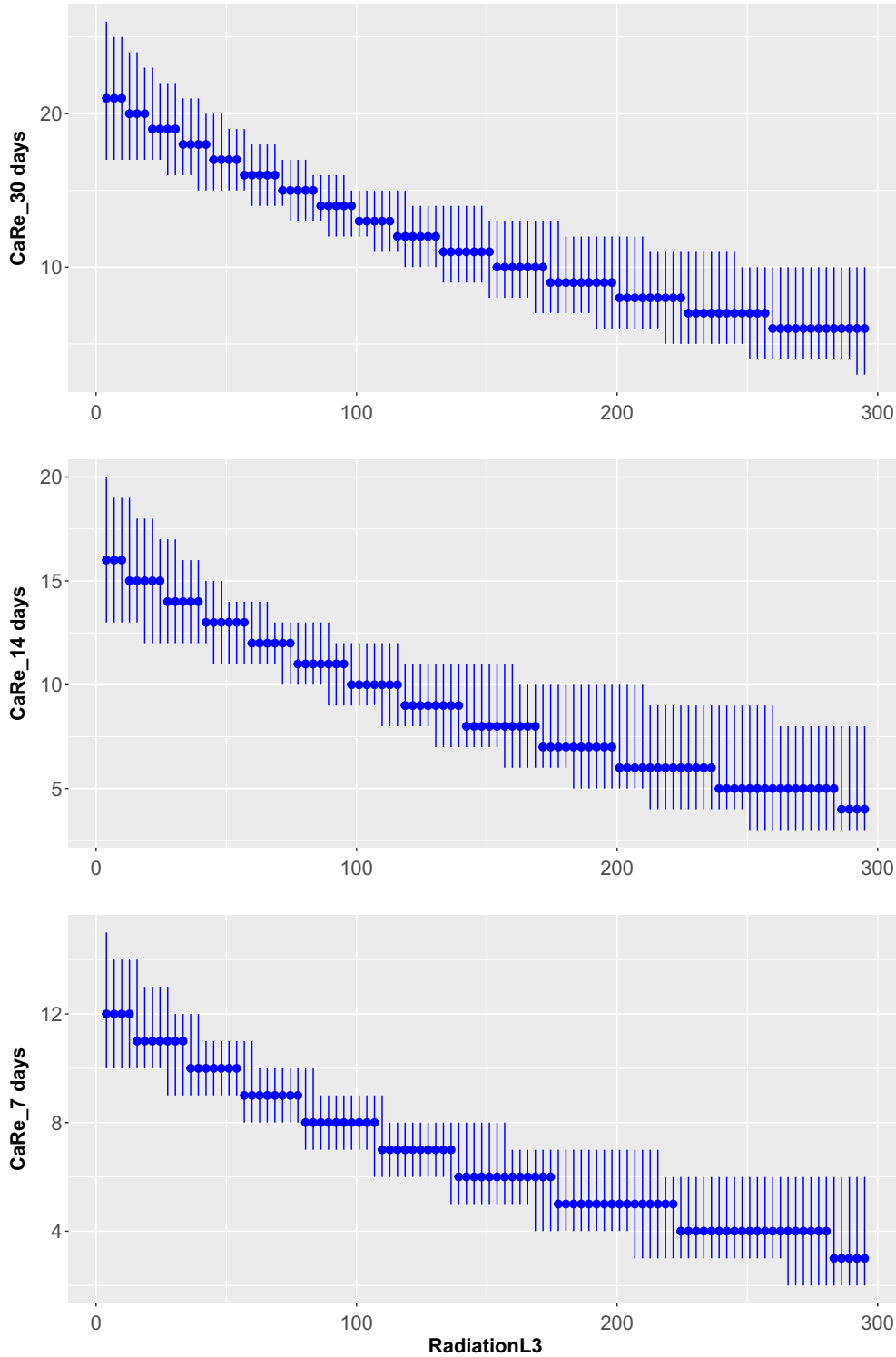

FIGURE 5 Positive cases: estimated CaRe for $h = 30, 14, 7$ days (panels from top to bottom) with respect to lagged radiation. The bars correspond to 95% intervals. The other covariates are fixed to their mean values

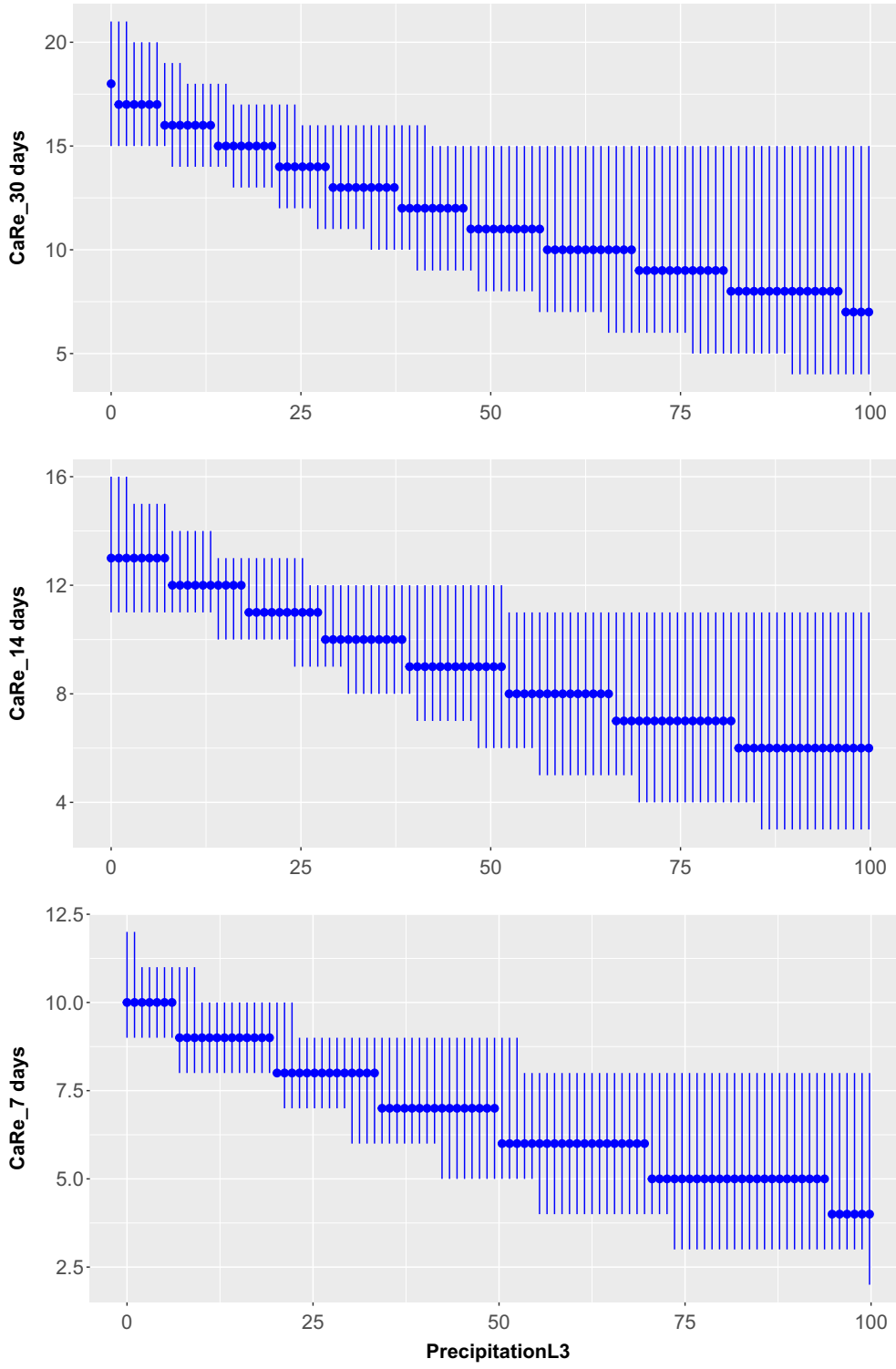

FIGURE 6 Positive cases: estimated CaRe for $h = 30, 14, 7$ days (panels from top to bottom) with respect to lagged precipitation. The bars correspond to 95% intervals. The other covariates are fixed to their mean values

This evidence should, however, be interpreted bearing in mind the widths of the intervals which are large.

Supplementary Figures S3 and S4 show the CaRe estimated values with their simulated intervals as a function of meteorological predictors, for different time horizons (or probability levels) for negative cases. These values convey information on the risk of inefficiency in the flu testing process, revealed by a very high number of negative cases. The risk of inefficiency considerably decreases both in sunny and rainy periods. The top panel of Figure S4 shows the estimated number of negative cases which on any single day is exceeded with probability 1/30 in terms of amount of precipitation. For the 30-day horizon, CaRe goes from 50, in the scenario of no rain, to 30 when it rains.

3.2 | Outlier detection

The hospital flu testing process depends on managerial and/or decision making instances and may lead to outlying records; our robust methodology is capable of detecting these abnormal values. Figures 7 and 8 show the robustness weight $w = \rho_c'$ (see Section 2.2) for each observation. These are obtained as a by product of the parameter estimation process. The size of the circles is proportional to $(1 - w)$. These weights can be used to identify outliers: the lower the weight, the more likely the observation is to be outlying. We identify the points with the smallest weights (largest circles), accompanied by their date of occurrence, from the figures. The observations with the smallest weights are not systematically those with large observed values, and, conversely, the largest observations are not systematically considered outliers.

Dates such as 2 January 2018 (which is right after Christmas) are typically regarded as special days in the flu recording process. Days in February 2019 such as 24 February 2019 and 26 February 2019 in Figure 7 relate to positive cases and correspond to school holidays. At school, children are super-spreaders. This is not the case during holidays when children are with their families. This and the fact that people are less tested during holidays contribute to a significant slow-down of the flu epidemic.

4 | SIMULATION STUDY

To assess the empirical properties of our proposal in finite samples, we designed a simulation study inspired by our data analysis in Section 3. We will look at the quality of the estimated parameters and CaRe obtained using the proposed approach, under the assumed D-GPD model and under contamination (i.e. in the presence of observations that deviate from the assumed model).

We generated 'clean' datasets from the model

$$\log(\xi) = \alpha_0 \quad \text{and} \quad \log(\sigma) = \beta_0 + \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \beta_3 \boldsymbol{x}_3, \tag{7}$$

where the covariates distributions were chosen to mimic the behaviour of `MintempL3`, `RadiationL3` and `PrecipitationL3` in the application. More precisely, we simulated the covariates as follows: $\boldsymbol{x}_1 \sim \mathcal{N}(2.3, 14)$, $\boldsymbol{x}_2 \sim \Gamma(1.55, 0.02)$ and $\boldsymbol{x}_3 \sim \text{lognormal}(0.71, 3.12)$. These were kept fixed throughout the simulation replicates. The parameters were set to $\alpha_0 = -2$ (hence $\xi = 0.135$), and $\boldsymbol{\beta} = (2, -0.05, -0.005, -0.01)^T$. Contaminated datasets were obtained by randomly setting 5%

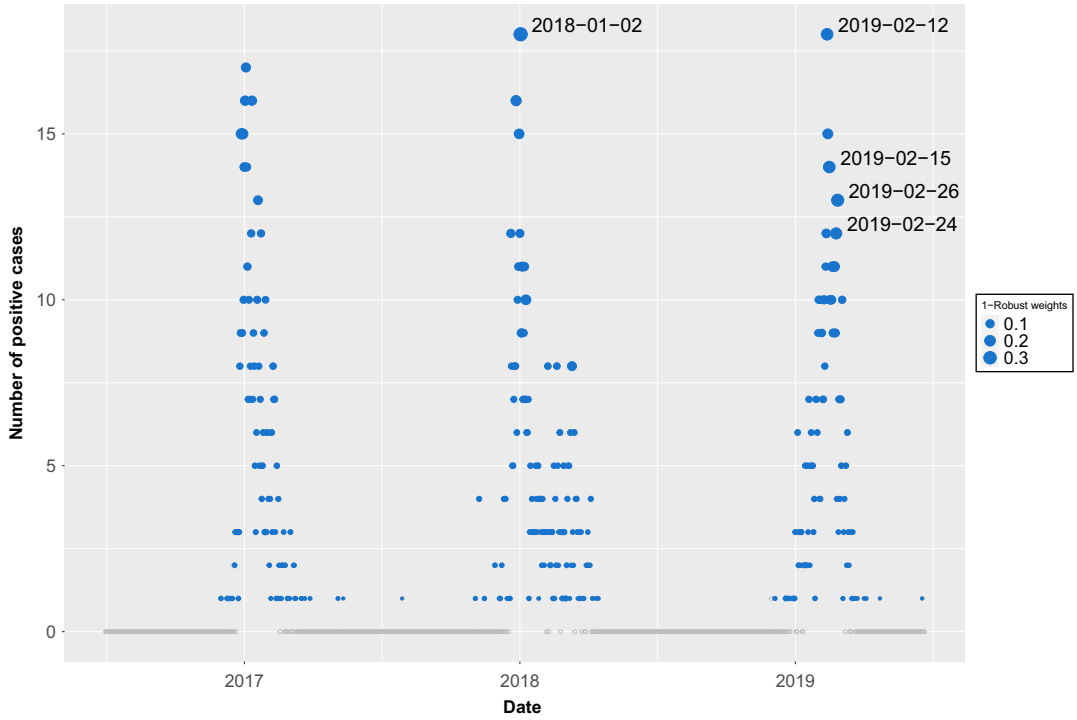

FIGURE 7 Robustness weights from the model fitted to positive cases

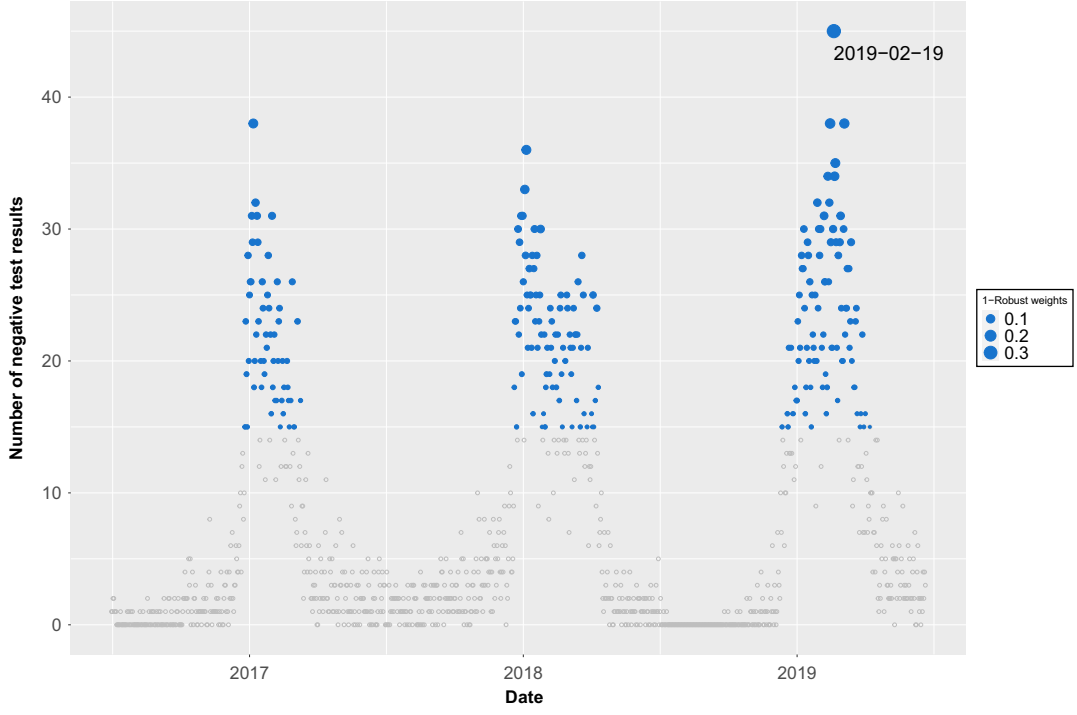

FIGURE 8 Robustness weights from the model fitted to negative cases

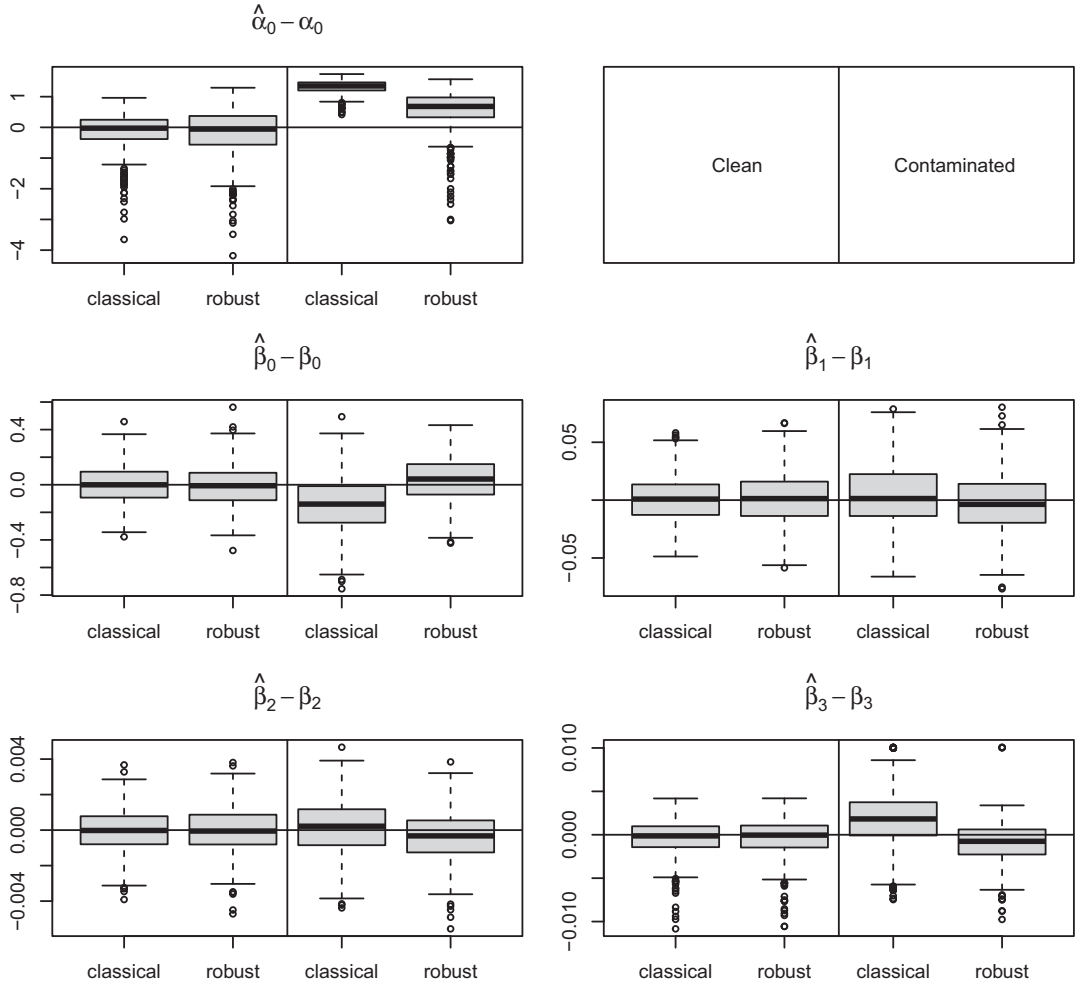

FIGURE 9 Boxplots of the centred parameter estimates for model (7). In each panel, the left boxplots refer to the clean data setting, and the right ones to the contaminated data setting

of the response values to the maximum value observed in the sample. This was inspired by the empirical application.

The sample size was set to $n = 250$ (hence consistent with our application) and the number of replications to 500. We fitted models based on the robust and classical maximum likelihood estimators. For the robust approach, we choose $c = 6$ to achieve a down-weighting of 0.95, as per the tuning procedure described in Section 2.2.

Figure 9 shows the centred estimated parameter estimates for model (7). By looking at the left boxplots of each panel, corresponding to the clean data setting, we see that both boxplots are centred around zero suggesting unbiased estimation of the model parameters. We also see that for the clean data setting the variability of the robust estimator's estimates are slightly larger than the variability of those obtained using its classical maximum likelihood counterpart; this is expected because of the loss of efficiency of the robust estimator with respect to maximum likelihood at the model. The story is different when looking at the right boxplots of each panel, corresponding to the contaminated data. All in all, the robust estimator performs much better than the classical

one. For all the components of $\boldsymbol{\beta}$, the robust estimator shows almost no bias. This is not the case for the classical estimator, which is influenced by the outliers, sometimes heavily, notably for β_0 and β_3. Also note that contamination affects the variability of the classical estimator, which is larger than that of the robust estimator for all the components of β. The estimation of α_0 under contamination seems more difficult. Both estimators show bias, even though that bias is much smaller for the robust estimator. It is worth recalling that a robust estimator guarantees that the bias under contamination does not explode, but it does not guarantee that it will vanish. Had we

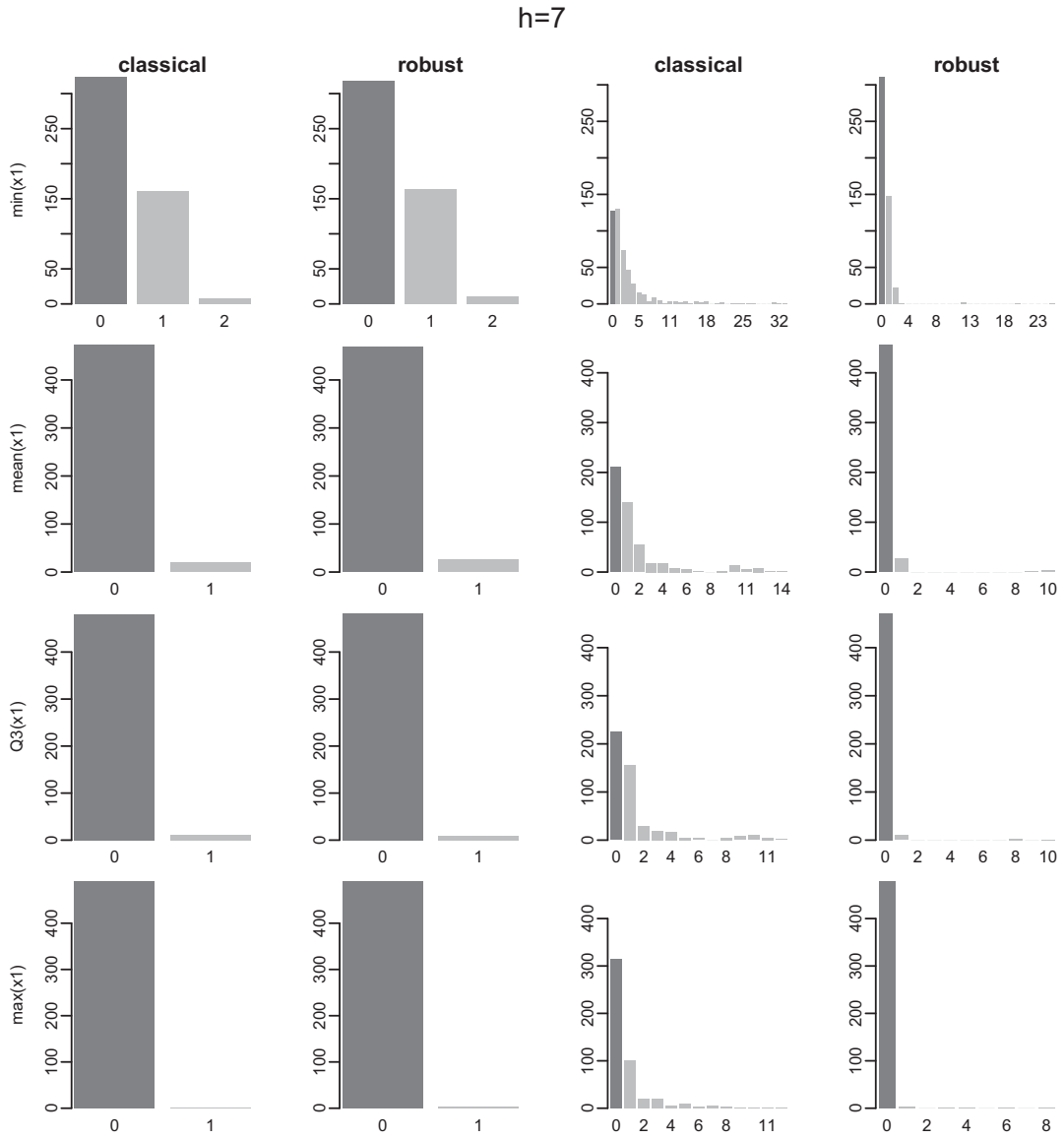

FIGURE 10 CaRe estimates for $h = 7$ for four values of x_1: its minimum value (first row), its average (second row), its third quartile (third row) and its maximum (fourth row). The other covariates are fixed at their mean values. The true population value is identified by a darker bar. The first two columns correspond to the clean data setting, whereas the last two columns correspond to the contaminated data setting

increased the amount of contamination or its strength, we would have expected the bias of the classical estimator to explode, but not that of the robust approach.

Figure 10 reports the estimated CaRe for $h = 7$ as a function of x_1, with x_2 and x_3 fixed at their mean values. CaRe is a discrete positive value (with relatively few different values in our simulation setting) and we depict it using barplots (histograms) for four representative values across the range of x_1: its minimum (first row), its average (second row), its third quartile (third row) and its maximum (fourth row). In each panel, the true population value is identified by the dark bar. The comparison of the first two columns confirms that the robust and classical estimators perform equally well and their estimates either match the population value or are just off by one unit, or very rarely two. Looking at the results under contamination and comparing the last two columns of the figure, we observed that the robust estimator performs much better than the classical one in that it identifies the true population value more often (more than 95% of the times, except for $\min(x_1)$ where this is about 65%). In contrast, the classical CaRe estimator shows large variability and misses the population target quite often. The quality of the CaRe estimates is not uniform across the range of x_1, with lower values of the covariate being more challenging.

We provide in the supplementary material similar figures for all the other combinations of horizons h and covariates; the conclusions do not change. Moreover, as expected, we observe that the performance of both estimators worsens as h increases, and that the CaRe seems to be difficult to estimate at the lower boundary of the range of the covariates, more particularly so for x_3.

To sum up, our simulation study demonstrates that our distributional regression approach is effective in estimating models with D-GPD responses, and that the robust version of the estimator can successfully cope with contaminated data.

However, in reality we never know a priori if the data are contaminated or not. Therefore, our results show that to be always on the safe side it is favourable for the practitioner to use the robust method. For that reason we only used the robust method in the application.

5 | CONCLUSION

Seasonal epidemics may lead to hospital congestion. In this paper, we use extreme value theory to study the occurrence of large numbers of flu cases in a hospital. We developed and implemented in `GJRM` a robust regression-type methodology that allows for non-identically distributed discrete extremes and that deals with outlying data. The response variables of interest (the positive and negative cases) are statistically explained by meteorological variables. Although the models selected for this case study are based on parametric covariate effects, our software implementation allows for very general non-parametric functional forms, which would most likely be required for larger datasets. Even without asymptotic arguments, the case $\xi < 0$ for the D-GPD can be used in practice and is implemented in `GJRM`. Note also that in our software, the robust regression-type methodology for non-identically distributed continuous extremes, relying on the GPD model, is also implemented. An example of its use is given in the Supplementary material.

Taking the point of view of the hospital, which needs to manage admission capacities, we introduced the notion of charge-at-risk whose estimation, based on meteorological factors, can serve as a quantitative tool to alert the hospital and allow time to prepare for a possible congestion. The introduced approach could be applied to several types of seasonal virus data such as those deriving from the new virus SARS-CoV-2.

ACKNOWLEDGEMENTS

The simulations were performed at the University of Geneva using the Baobab cluster. The research was partially funded by the Swiss National Science Foundation SNF (first and third authors). We thank the reviewer and the associate editor for an in-depth reading of our manuscript and the constructive comments.

DATA AVAILABILITY STATEMENT

The data are not publicly available. The dataset has been provided by the Institute of Microbiology of CHUV and is not publicly accessible. The access was acquired by one of our co-authors (Dr. Jaton) who works at the Institute of Microbiology. The anonymised dataset includes all the visits to the Institute laboratory for the u test. The data were then aggregated on a daily basis for our analysis.

AUTHOR CONTRIBUTIONS

SR is the leading author. EC, VCD, GM and RR contributed equally. KJ provided the data. All authors reviewed the final manuscript.

ORCID

Eva Cantoni https://orcid.org/0000-0002-6614-7772

REFERENCES

Aeberhard, W.H., Cantoni, E., Marra, G. & Radice, R. (2021) Robust fitting for generalized additive models for location, scale and shape. *Statistics and Computing*, 31, 1–16.

Balkema, A.A. & de Haan, L. (1974) Residual life time at great age. *The Annals of Probability*, 2, 792–804.

Chavez-Demoulin, V. & Davison, A.C. (2005) Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54, 207–222.

Davis, R.E., Rossier, C.E. & Enfield, K.B. (2012) The impact of weather on influenza and pneumonia mortality in New York city, 1975-2002: a retrospective study. *PLoS One*, 7. Available from: http://doi/org.10.1371/journal.pone.0034091

Davison, A.C. & Smith, R.L. (1990) Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society, Series B*, 52, 393–442.

Dell'Aquila, R. & Embrechts, P. (2006) Extremes and robustness: a contradiction? *Financial Markets and Portfolio Management,* 20, 103–118.

Dupuis, D. & Field, C. (1998) Robust estimation of extremes. *Canadian Journal of Statistics*, 26, 199–215.

Dupuis, D. & Victoria-Feser, M.P. (2006) A robust prediction error criterion for Pareto modeling of upper tails. *Canadian Journal of Statistics*, 34, 639–358.

Eguchi, S. & Kano, Y. (2001) *Robustifing maximum likelihood estimation by psidivergence*. Research Memorandum 802, Institute of Statistical Mathematics (ISM), Tokyo, Japan.

Firestone, S.M., Cogger, N., Ward, M.P., Toribio, J.A.L.M.L., Moloney, B.J. & Dhand, N.K. (2012) The influence of meteorology on the spread of influenza: survival analysis of an equine influenza (a/h3n8) outbreak. *PLoS One*, 7. Availabe from: http://doi/org.10.1371/journal.pone.0035284

Hastie, T.J. & Tibshirani, R.J. (1990) *Generalized additive models*. New York, NY: Chapman & Hall/CRC.

Hitz, A., Davis, R. & Samorodnitsky, G. (2017) Discrete extremes. Available from: https://arxiv.org/pdf/1707.05033.pdf

La Vecchia, D., Ronchetti, E. & Trojani, F. (2012) Higher-order infinitesimal robustness. *Journal of the American Statistical Association*, 107, 1546–1557.

Lowen, A.C. & Steel, J. (2014) Roles of humidity and temperature in shaping influenza seasonality. *Journal of Virology*, 88, 7692–7695.

Marra, G. & Radice, R. (2020) Copula link-based additive models for right-censored event time data. *Journal of the American Statistical Association*, 115, 886–895. http://doi/org.10.1080/01621459.2019.1593178

Marra, G. & Radice, R. (2021) GJRM: Generalised Joint Regression Modelling. Available from: http://CRAN.R-project.org/package=GRJM r package version 0.2-5.1.

Rigby, R.A. & Stasinopoulos, D.M. (2005) Generalized additive models for location, scale and shape (with discussion). *Applied Statistics*, 54, 507–554.

Roussel, M., Pontier, D., Cohen, J.M., Lina, B. & Fouchet, D. (2016) Quantifying the role of weather on seasonal influenza. *BMC Public Health*, 16, 1–14.

Shimura, T. (2012) Discretization of distributions in the maximum domain of attraction. *Extremes*, 15, 299–317.

Towers, S., Chowell, G., Hameed, R., Jastrebski, M., Khan, M. & Meeks, J. (2013) Climate change and influenza: the likelihood of early and severe influenza seasons following warmer than average winters. *PLoS Currents*, http://doi/org.10.1371/currents.flu.3679b56a3a5313dc7c043fb944c6f138

Wood, S.N. (2017) *Generalized additive models: an introduction with R*, 2nd edition. Boca Raton, FL: Chapman and Hall/CRC.

Yee, T.W. & Stephenson, A.G. (2007) Vector generalized linear and additive extreme value models. *Extremes*, 9, 1–19.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Ranjbar, S., Cantoni, E., Chavez-Demoulin, V., Marra, G., Radice, R. & Jaton, K. (2022) Modelling the extremes of seasonal viruses and hospital congestion: The example of flu in a Swiss hospital. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 1–22. Available from: https://doi.org/10.1111/rssc.12559