CEAS Aeronautical Journal
https://doi.org/10.1007/s13272-021-00540-8

ORIGINAL PAPER

A statistical approach for sizing an aircraft electrical generator using extreme value theory

Feriel Boulfani[1] · Xavier Gendre[2] · Anne Ruiz-Gazen[1] · Martina Salvignol[3]

Received: 11 June 2020 / Revised: 23 June 2021 / Accepted: 9 August 2021

Abstract
The sizing of aircraft electrical generators mainly depends on the electrical loads installed in the aircraft. Currently, the generator capacity is estimated by summing the critical loads, but this method tends to overestimate the generator capacity. A new method to challenge this approach is to use the electrical consumption recorded during flights and study the distribution of operational ratios between the actual consumption and the theoretical maximum consumption then size the future aircraft generators by applying a ratio to the theoretical value. This paper focuses on the application of extreme value theory on these operational ratios to estimate the maximal capacity utilization of a generator. A real data example is provided to illustrate the approach and estimate extreme quantiles and the right endpoint of the distribution of the ratios together with their approximate confidence interval in the nominal configuration. In all situations the right endpoint is proven to be finite and does not depend on the use procedures. This approach shows that ELA overestimates the maximal permanent consumption by 20% with error level of 10^{-3} in the nominal configuration.

Keywords Electrical load analysis · Aeronautic electrical system · Generalized Pareto distribution · Quantile estimation · Endpoint estimation · Diagnostics for threshold selection

Abbreviations

AC	Alternating current
APU	Auxiliary power unit
CI	Confidence interval
ELA	Electrical load analysis
EVT	Extreme value theory
i.i.d.	Independent and identically distributed
GEV	Generalized extreme value
GPD	Generalized pareto distribution
KVA	Kilo–Volt–Ampere
min	minutes
PP-plot	Probability–probability plot
QQ-plot	Quantile–quantile plot
RAT	Ram air turbine
sec	Seconds
UCI	Upper confidence interval

✉ Feriel Boulfani
feriel.boulfani@tse-fr.eu

Xavier Gendre
xavier.gendre@isae-supaero.fr

Anne Ruiz-Gazen
anne.ruiz-gazen@tse-fr.eu

[1] Toulouse School of Economics, Université Toulouse Capitole, 1 Esplanade de l'Université, 31080 Toulouse Cedex 6, France

[2] ISAE-SUPAERO, Université de Toulouse, 10 Avenue Édouard Belin, 31055 Toulouse, France

[3] Toulouse, France

1 Introduction

Driven by the demand to reduce emissions, the aviation industry pushes toward the concept of more electrical aircraft and, ultimately, an all-electrical aircraft [11]. Thus, the electrical network will be more in demand. A new network should be designed, and a new electrical-intensive architecture implemented.

The Electrical Load Analysis (ELA) is an airworthiness requirement. For a given aircraft, it describes the electrical network and shows the total theoretical electrical consumption by generators for the different flight phases and different operational modes. In the ELA, the electrical consumption is computed by summing the component loads under the most unfavourable conditions to get the maximal consumption and under normal operating conditions to get the operational consumption.

Journal : Large 13272	Article No : 540	Pages : 12	MS Code : 540	Dispatch : 1-9-2021

The ELA is provided to the airline at the time of aircraft delivery. The airline must use this report to evaluate the effects of equipment changes on the electrical network to avoid electrical overload.

To avoid oversizing the future electrical network, aircraft manufacturer has to assess the current network and re-evaluate the needs based on operational measurements. According to recent internal measurements of electrical networks recorded during the flight by aircraft manufacturers, the theoretical power consumption appears to be overestimated as illustrated in Fig. 1. This figure shows the proportion of electrical consumption with respect to electrical capacities over time for one generator during a given flight. A large difference is observed between the theoretical maximum consumption given by ELA and the real consumption.

Using operational measurements, we want to justify that the maximal observed consumption is smaller than the maximal consumption given by ELA. The main reason is that the electrical loads do not operate all at the same time whereas they are considered simultaneously in the ELA.

A preliminary work has been done by [10] using Monte Carlo algorithms to simulate the electrical load behavior. This approach is based on simulations and differs from ours as our objective is focused on the extreme behavior of the observed electrical consumption. The approach developed hereafter is based on the Extreme Value Theory (EVT). EVT provides statistical tools to estimate extreme quantiles and right endpoints under two hypotheses. First, the observations are considered as independent and identically distributed (i.i.d.) realizations of random variables. Second, the probability distribution belongs to the domain of attraction of some extreme value distribution. Under these hypotheses, we derive extreme quantiles and endpoints together with their confidence interval. Note that extreme quantiles (resp. endpoints) are values such that the probability of getting a larger value is extremely small (resp. equal to zero).

The distribution assumption is not restrictive and can be checked for many well known distributions including the uniform on interval and the normal ones (see [6]). The results are asymptotic in the sense that they are valid for large sample size. The parametric extreme value distribution is obtained by looking at the limit distribution of standardized maxima. This result is comparable to the central limit theorem that considers the asymptotic behavior of the sum of random variables and leads to a normal distribution.

EVT has already been used to estimate very high quantiles for electrical systems (see [13] and [7]). Among recent applications of the EVT in the aeronautical field, the authors of [9] estimate the probability of occurrence of the position, velocity or altitude errors for the navigation systems, while [12] designs the load spectrum for aircraft hydraulic pumps.

The present paper illustrates the application of EVT to aeronautic electrical systems consumption to challenge the ELA assumption approach in the nominal mode only. The approach presented below can also be applied to the degraded and emergency modes. Nevertheless, the few amount of data available in these modes implies a specific statistical pre-treatment and is beyond the scope of the present paper.

We have a sample of 60,000 flight hours from 18 operational aircraft that we split into 8 groups based on conditions of use of the aircraft. One main goal of our study is to use a limited amount of aircraft records to compute probabilities beyond the observed measurements. The EVT answers this challenge by estimating extreme quantiles and right endpoints. Probabilities associated with extreme quantiles are then converted into probabilities by flight hours. Confidence intervals are built to encompass the non observed aircraft.

As each aircraft has its own configuration, the ELA value may vary. Thus we choose to estimate the maximum ratios between the electrical consumption and the theoretical maximum values given by the ELA rather than estimating the maximal electrical consumption. Applying EVT on these ratios will help us to evaluate a maximal ratio irrespective of the electrical aircraft configuration.

First, we apply EVT to each group separately. Then we compare the results between the different groups by using a statistical test. The null assumption is the equality of the endpoints between groups. Using our sample we do not reject the null assumption at usual error level of 5%. From

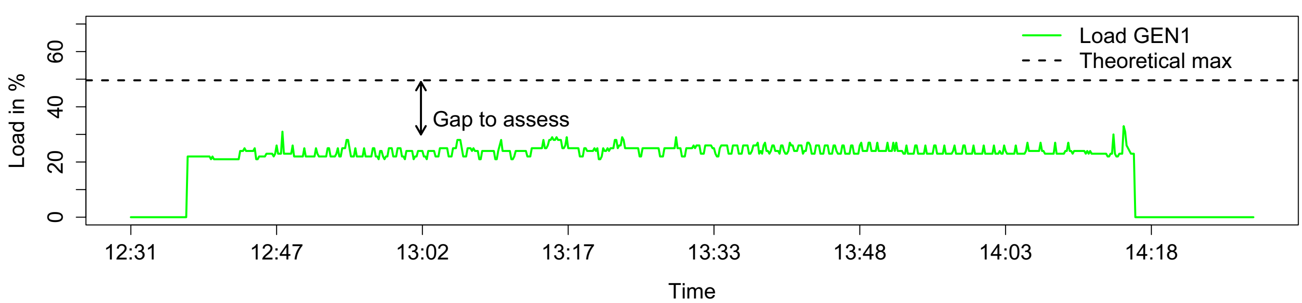

Fig. 1 Example of the consumption in percentage of the capacity of generators as a function of time for a given flight

this result we can suggest a generalized maximal ratio to all operational aircraft and to the future aircraft model. Multiplying the ELA values by the maximal ratio leads to adjusted ELA value that could be used for sizing future generators or adding more loads to operational aircraft.

This paper is organized as follows. Section 2 presents the aircraft electrical network and details the dataset used to assess the electrical network. Section 3 recalls the EVT procedure and the model selection method used to estimate the extreme quantiles and endpoints. Section 4 illustrates this model selection procedure on a given group example. It also shows the results obtained using data from the 8 groups separately and globally after testing the endpoints equality of the ratios between groups. Finally, Sect. 5 concludes the study and proposes possible extensions.

2 Context and data presentation

We are interested in evaluating the extreme electrical consumption with respect to the theoretical ELA value of the generators based on operational measurements.

An aircraft flight is segmented into several phases depending on the altitude and the electrical source used. In our study we consider the flight phase, i.e. where the landing gear is no longer compressed and the altitude is greater than 1.500 feet, and the onground phase and we first analyse these phases separately.

2.1 Aircraft electrical network

Different electrical sources power the electrical network of an aircraft:

– AC (Alternating Current) generators are supplied by the engines. Depending on the aircraft family, the number of AC generators is two or four. Each generator has a capacity of 90 Kilo-Volt-Ampere (KVA) for medium range and 100 KVA for long range.
– APU Generator (Auxiliary Power Unit) is an additional generator that supplies energy. It is used during the onground phase and as a backup in the flight phase to replace one or more AC generators at any time.
– RAT (Ram Air Turbine) is a wind turbine and a power source in case of loss of all electrical sources.
– Batteries have a limited capacity of electric power and are used for temporary actions.

In this paper, we focus the analysis on one of the AC generators.

The generators can support an overload that depends on the load duration. For the AC generator, the percentages

Table 1 Percentage of acceptable overload for an AC generator

	Under 5 sec	Under5 min	> 5 min
AC loads	160% to 183%	120% to 125%	100%

Table 2 Groups description

Group	# of aircraft	# of flight hours	Continent destinations
1	2	10 263	Asia
2	1	1 675	America–Europe
3	4	10 694	Europe
4	2	5 589	Asia
5	5	22 480	Asia
6	2	5 726	America–Europe–Oceania
7	1	1 825	North America
8	1	1 793	Europe–North America
Total	18	60,045	–

stands for the quantity available

of acceptable overload are shown in Table 1. The loads are classified as intermittent or permanent: the loads with a duration less than 5 mins are called intermittent loads; otherwise, they are called permanent loads. In what follows we focus the analysis on the permanent loads only. Moreover, when there is no failure, the electrical network is in the nominal mode and we consider this mode only.

2.2 Data details

We have 8 groups for which we consider 18 operational low-cost aircraft from the same family. Their characteristics are given in Table 2.

For each aircraft, we observe at every second the ratio defined by the electrical consumption divided by the maximal electrical load given by the ELA for the corresponding aircraft and phase. Let Y be a random variable which represents these ratios. The ratios are expressed in percentage but this has no impact on the EVT analysis.

We split the observations into the flight phase and onground phase and independently apply the EVT to each of the two phases.

To remove the intermittent loads, we average Y in a time window of length T by

$$X_k = \frac{1}{T}\sum_{i=1}^{T} Y_{(k-1)T+i}, \quad k \in \{1, \ldots, \tau\} \tag{1}$$

where $\tau = \lfloor n/T \rfloor$ and $\lfloor \cdot \rfloor$ denotes the floor part function. The i.i.d. variables X_k distributed as a variable X are positive and can be greater than 1 if the consumption exceeds the ELA

Journal : Large 13272 | Article No : 540 | Pages : 12 | MS Code : 540 | Dispatch : 1-9-2021

value. On top of that, a special load that generates high peaks for less than 200 milliseconds is removed.

We apply the EVT on these datasets to calculate Q_p the $(1-p)$-quantile associated to a small probability p, i.e. such that $P(X > Q_p) = p$, and the right endpoint x^* of the distribution support. The endpoint is defined by $x^* := \sup\{x : P(X \leq x) < 1\}$ and can be finite or not. If it is finite, this corresponds to the 1-quantile and $P(X > x^*) = 0$.

3 Extreme value theory reminder

EVT is widely used in applied fields such as hydrology, meteorology and insurance (see [1]). The objective is to estimate the probability distributions of the maxima and compute the probabilities associated with rare events.

In this paper, we want to estimate extreme quantiles and endpoint for the observed ratios $x_1, \dots, x_n$ which are considered as realizations of i.i.d. random variables $X_1, \dots, X_n$ with distribution function F. Let Q_p be the $(1-p)$-quantile and x^* the right endpoint of F. Since

$$\mathbb{P}(\max(X_1, \dots, X_n) \leq x) = \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) \\ = F^n(x),$$

$\max(X_1, \dots, X_n)$ converges in probability to x^* as n tends to infinity. To obtain a nondegenerate limit distribution we need to normalize $\max(X_1, \dots, X_n)$. To this end, we assume that there exist deterministic sequences $a_n > 0$ and $b_n \in \mathbb{R}$, such that

$$\frac{\max(X_1, \dots, X_n) - b_n}{a_n}$$

has a nondegenerate limit distribution as $n \to \infty$ given by

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x). \tag{2}$$

G is called extreme value cumulative distribution function and F is in the domain of attraction of G.

The previous assumption is fulfilled under regularity assumption on right endpoint of F. It can be checked for many absolutely continuous distribution functions such as uniform on an interval, normal, log-normal, gamma, beta, etc. (see details in [6], p. 153–157).

EVT is a powerful statistical asymptotic theory that allows us to calculate extreme quantiles and endpoints without parametric assumptions on the distribution F of the data. Thanks to EVT we get a parametrized extreme distribution G. The parameters of G can be estimated using statistical methods such as the maximum likelihood or the moment method as discussed in [5].

The EVT is usually divided into two main approaches. The first approach is the Generalized Extreme Value (GEV) based on the study of the asymptotic distribution of a series of maxima. Under some conditions, this distribution is known to converge to Gumbel, Fréchet, or Weibull distributions. The second approach is the Generalized Pareto distribution (GPD) based on the study of the distribution of excess over a given high threshold.

The two approaches can be used to build an extreme value model for maxima and estimate the parameters. In the GEV approach the selection of the blocks size is a difficult task in practice. From our experience on the flight series data (see Fig. 2), the results strongly depend on the block size and flight length, which makes the fitting difficult. This approach is more adapted to an uninterrupted series of data but is not relevant for flight data. Therefore, we only focus on the GPD approach which better captures all the maxima but recall both approaches in what follows.

3.1 Generalized extreme value approach

The GEV approach consists in dividing the series into non overlapping blocks of identical lengths and taking the maximum of each block. Let $X_1, \dots, X_n, \dots$ be i.i.d. random variables with unknown cumulative distribution. We define a block maximum

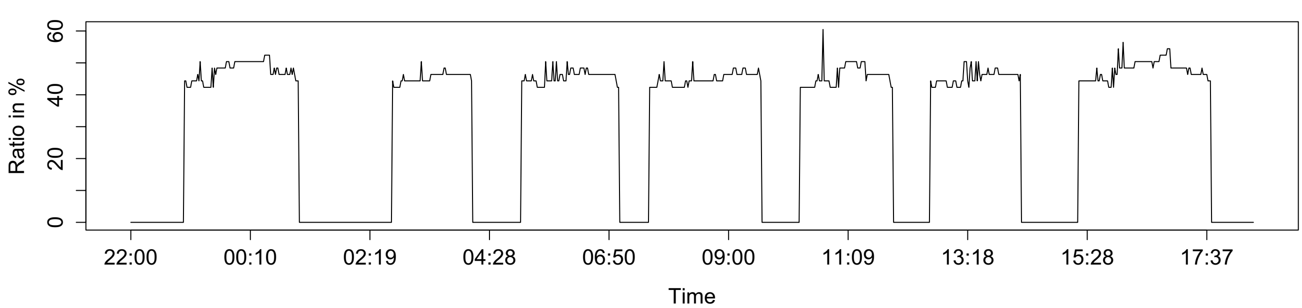

Fig. 2 Record of X for 1 day for a given aircraft; 7 flights were observed during this day

$$M_n = \max\left\{X_1, \dots, X_n\right\}$$

as the observed maximum of the process over n time units. If n is the number of observations in 1 h, then M_n corresponds to the maximum over 1 h.

As stated in [2], the asymptotic cumulative distribution function of block maximum M_n is given by

$$H(x) = \exp\left\{-(1 + \xi\,\frac{x-\mu}{\sigma})^{-1/\xi}\right\}$$

where $1 + \xi\,\dfrac{x-\mu}{\sigma} > 0$. The parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ correspond to location and scale, respectively. The third parameter $\xi \in \mathbb{R}$ is a shape parameter, which corresponds to the thickness of the tail of the distribution:

- $\xi > 0$ corresponds to the heavy-tailed case, and the corresponding distribution converges to Fréchet;
- $\xi = 0$ corresponds to the light-tailed case, and the corresponding distribution converges to Gumbel;
- $\xi < 0$ corresponds to the short-tailed case, and the corresponding distribution converges to Weibull.

The asymptotic distribution of the maximum is always one of these three distributions regardless of the original distribution. The asymptotic distribution of the maximum can be estimated assuming condition (2) but without any parametric assumptions on the distribution of the observations.

3.2 Generalized Pareto distribution approach

The GPD approach consists in selecting a given (sufficiently high) threshold and considering the observations that exceed this threshold. Let $(X_1, \dots, X_n)$ be a sequence of independent random variables with identical distribution as X that satisfies condition (2). The random variables $X_i - u$, for $i \in \{1, \dots, n\}$, are the exceedances over threshold u if this threshold has been exceeded.

For some μ, $\sigma > 0$ and ξ, for u sufficiently large, the cumulative distribution function of $X - u$ conditional on $X > u$ can be approximated by the distribution:

$$H(x) = \begin{cases} 1 - \left(1 + \xi \dfrac{x}{\beta}\right)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\dfrac{x}{\beta}\right) & \text{if } \xi = 0, \end{cases}$$

where $x > 0$, and $\beta = \sigma + \xi(u - \mu) > 0$ is the reparametrized scale.

Note that multiplying the random variable by a positive constant c keeps the parameter ξ unchanged while β is multiplied by c. This means that EVT is equivariant by scale transformation. Estimation of parameters μ, σ and ξ for extreme quantiles and endpoint of the distribution F, with their confidence intervals, are derived from an asymptotic framework where u is replaced by a sequence of upper order statistics depending on n (see [3] for technical details). In order to use these asymptotic results in practice, we have to ensure that the number of observations n is large but also that the ratio between the number n_u of observations larger than u and n is small (see [4] for a detailed application).

The threshold selection involves balancing bias and variance. The threshold u must be sufficiently high to ensure that the asymptotic underlying the GPD approximation is reliable and thus reduce the bias. However, a reduced sample size for high thresholds increases the variance of the parameter estimators.

As discussed in [1], the common graphical diagnostics for threshold selection are the mean residual life, the threshold stability plots and the fitting diagnostic plots. These plots are described below with some guide-lines to use them for threshold selection:

- Mean residual life plot: the empirical mean of the exceedances above threshold u is plotted against u. Above threshold u_0, where the generalized Pareto distribution provides a valid approximation to the excess distribution, the mean residual life plot should be approximately linear in u.
- Threshold stability plots: ξ and β are plotted against a range of thresholds u. For u_0 selected using the mean residual life plot, we look at the stability of the parameter estimates for values of $u > u_0$ and possibly refine the choice of the threshold.
- Fitting diagnostic plots: the Probability-Probability plot and Quantile-Quantile plots, which are named PP-plot and QQ-plot, respectively, are the usual diagnostics tools. If the model fits the data, the points pattern should exhibit a 45-degree straight line for both plots. Once the threshold is selected using the mean residual life and threshold stability plots, the PP and QQ-plots are used to validate our choice.

We propose to estimate the GPD parameters using the maximum likelihood method. The log-likelihood function is given by

$$l(\xi, \beta) = \begin{cases} -n\log(\beta) - \left(\frac{1}{\xi} + 1\right)\displaystyle\sum_{i=1}^{n} \log\left(1 - \xi\,\frac{x_i}{\beta}\right), \\ \text{if } \xi \neq 0, \\ -n\log(\beta) - \dfrac{1}{\beta}\displaystyle\sum_{i=1}^{n} x_i, \\ \text{if } \xi = 0. \end{cases}$$

Journal : Large 13272	Article No : 540	Pages : 12	MS Code : 540	Dispatch : 1-9-2021

In practice, the values $\hat{\xi}$ and $\hat{\beta}$ that maximize $l(\xi, \beta)$ are found by using a gradient descent method (see [5]).

Let X be a random variable that follows a GPD(ξ,β), the quantile Q_p is estimated by

$$\hat{Q}_p = \begin{cases} u + \dfrac{\hat{\beta}}{\hat{\xi}}\left[\left(\dfrac{n_u}{np}\right)^{\hat{\xi}} - 1\right], & \text{if } \hat{\xi} \neq 0, \\ u + \hat{\beta}\log\left(\dfrac{n_u}{np}\right), & \text{if } \hat{\xi} = 0. \end{cases} \tag{3}$$

It is possible to build a $(1-\alpha)$ asymptotic confidence interval (CI) for $\hat{Q}_p$ (see page 150 of [3]). The upper confidence interval (UCI) limit is given by

$$Q_p < \hat{Q}_p + Z_{\alpha/2}\,\hat{\beta}\,q_{\hat{\xi}}\left(\frac{n_u}{np}\right)\sqrt{\frac{\mathrm{Var}(\hat{\xi})}{n_u}}, \tag{4}$$

where $Z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution, an approximation of q_{ξ} for large t (see [3] p. 135) is given by

$$q_{\xi}(t) \approx \begin{cases} t^{\xi}\,\log t/\xi, & \text{if} \quad \xi > 0, \\ (\log t)^2/2, & \text{if} \quad \xi = 0, \\ 1/\xi^2, & \text{if} \quad \xi < 0, \end{cases}$$

and $\mathrm{Var}(\hat{\xi})$ is the variance of $\hat{\xi}$ defined by

$$\begin{cases} (1+\xi)^2, & \text{if} \quad \xi \geq 0, \\ 1 + 4\xi + 5\xi^2 + 2\xi^3 + 2\xi^4, & \text{if} \quad \xi < 0. \end{cases}$$

Let x^* be the right endpoint or the upper limit of the distribution. If the endpoint is known to be finite then $\xi < 0$ and an estimator of x^* can be calculated by letting $p \to 0$ in (3), which leads to

$$\hat{X}^* = u - \frac{\hat{\beta}}{\hat{\xi}}, \text{ for } \hat{\xi} < 0. \tag{5}$$

Replacing q_{ξ} by $1/\xi^2$ in (4), we get $(1-\alpha)$ one sided asymptotic CI

$$x^* < \hat{X}^* + Z_{\alpha}\,\frac{\hat{\beta}}{\hat{\xi}^2}\sqrt{\frac{\mathrm{Var}(\hat{\xi})}{n_u}}, \tag{6}$$

where Z_{α} is the $(1-\alpha)$ quantile of the standard normal distribution. In the next section α is called the error level.

The upper confidence interval values for the quantiles of order p and the endpoint are based on approximations that are valid under certain conditions. These conditions involve that the total number of observations n together with the number of observations that exceed the threshold u are large while the proportion n_u/n is small. Moreover, concerning the UCI of an extreme quantile, the probability p has to be small enough so that np/n_u is small but not too small in order to have a small value for $|log(np)|/\sqrt{n_u}$ (see Remark 4.3.4, p. 135 in [3]). Interested readers could find more details about the CI building in Chapter 4 of [3].

4 Extreme value application on electrical loads

4.1 Illustration of the GPD procedure for one group

In this section, we select one group, apply the GPD approach on the data and compute upper confidence interval values for extreme quantiles and endpoint.

The group under study was observed during more than 10,000 flights between 2016 and 2018. To illustrate the results of the methodology, we select one generator in the permanent mode during the onground phase. For each flight, we apply a mean time window of $T = 150$ seconds as detailed in Eq. (1). We apply the GPD approach using the package *extRemes* [8] in the R software with the maximum likelihood estimation method.

In the first step, we set threshold u using the graphical diagnostics from Sect. 3.2. The mean residual life plot is represented by a solid line in Fig. 3. We look for a linear trend at the extreme right of this curve. For u between 50% and 63%, the data exhibit such a linear trend. This choice is refined using Fig. 4, where we focus on u between 50% and 63%. According to these plots, $\hat{\beta}$ and $\hat{\xi}$ reach stability when $u > 57.5\%$, which indicates that the assumption of GPD is reasonable for $u \in [57.5\%, 60\%]$.

Table 3 gives the maximum likelihood estimates of $\hat{\beta}$ and $\hat{\xi}$ and confirms the stability of the estimates for this range of values. Then, we set $u = 59.5\%$ and check whether the model fits the data by using the fitting diagnostic PP-plot and QQ-plot in Figs. 5 and 6, respectively.

In both Figs. 5 and 6, the point pattern exhibits a 45-degree linear trend. So the GPD assumption appears

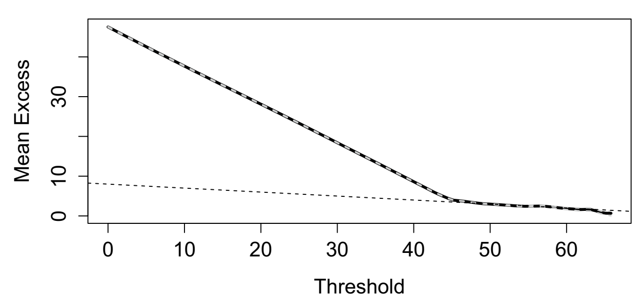

Fig. 3 Mean residual life plot. We plot u against the mean excess for a range of threshold values. A linear trend is observed for $u > 50\%$ represented by the dashed line

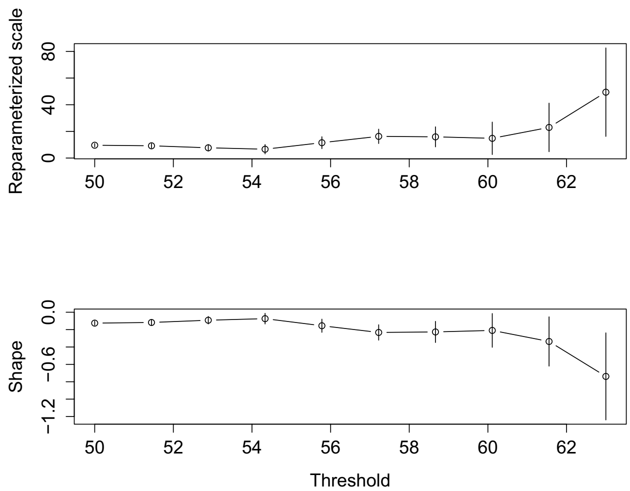

Fig. 4 Threshold stability plots for a threshold between 50% and 63% (top plot for β and bottom plot for ξ). For each value of u the vertical bar represents the confidence interval of the estimators. Stability of estimators is observed for $u \in [57.5\%, 60\%]$

Table 3 Maximum likelihood estimates of β and ξ for different thresholds u

$u(\%)$	$\hat{\beta}$	$\hat{\xi}$
57.5	3	−0.3
58	2.9	−0.2
58.5	2.7	−0.2
59	2.3	−0.2
59.5	2.3	−0.2
60	2.2	−0.2

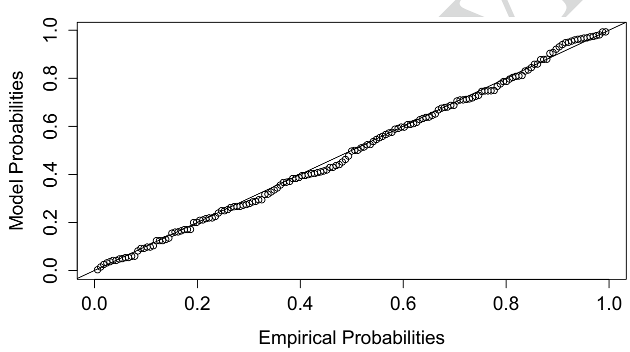

Fig. 5 PP-plot obtained from fitting the GPD using the maximum likelihood method for $u = 59.5\%$. The point pattern falls along the 45-degree line represented by the black line

reasonable for $u = 59.5\%$ and we obtain $n_u = 150$ from $n = 18\,319$.

To align with the safety assumption study, we have to convert our probabilities into probabilities by flight hour. In our case, we recall that the data are preprocessed by taking the mean of the consumption during a time window of

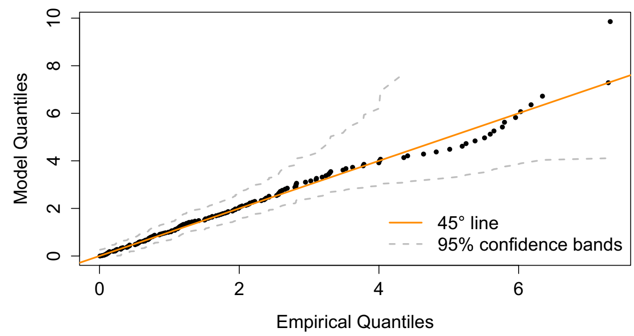

Fig. 6 QQ-plot obtained from fitting the GPD using the maximum likelihood method for $u = 59.5\%$. The point pattern falls along the 45-degree line. The dashed lines represent the 95% confidence bands based on the Kolmogorov-Smirnov statistics

Table 4 Quantile estimation for different values of P_{hour}

P_{hour}	p	Q_p
10^{-3}	10^{-5}	67.1
10^{-5}	10^{-7}	69.4
10^{-7}	10^{-9}	70.3
10^{-9}	10^{-11}	70.7
10^{-12}	10^{-14}	70.9

$T = 150$ seconds (see Sect. 2). Therefore, we have 24 observations per hour.

Let $X_1, \dots, X_{24}$ be the variables observed during a given hour, we want to compute the probability that their maximum is above the quantile Q_p. For a given probability p to exceed Q_p during a period of length T and assuming that $X_1, \dots, X_{24}$ are i.i.d. with the same distribution as X, we can write

$$\begin{aligned}\mathbb{P}\left(\max_i X_i > Q_p\right) &= 1 - \mathbb{P}\left(\max_i X_i \le Q_p\right)\\ &= 1 - \mathbb{P}\left(X \le Q_p\right)^{24} \\ &= 1 - \left[1 - \mathbb{P}\left(X > Q_p\right)\right]^{24}\end{aligned} \tag{7}$$

Let P_{hour} be the probability to exceed Q_p in 1 h. Then Equation (7) becomes $P_{\text{hour}} = 1 - (1-p)^{24}$, and we can compute p for a target probability P_{hour}. Table 4 shows the results obtained using Eqs. (3) and (7) to estimate quantiles associated to the target probabilities.

From Table 4, we select the result corresponding to $P_{\text{hour}} = 10^{-7}$ to respect the aeronautic safety procedure and not increase the probability of losing one generator.

At the probability 10^{-7} by flight hour, the maximum ratio for the selected generator is 70.3%. Using the results from Equation (4) we build UCI at error levels $\alpha = 5 \times 10^{-2}, 10^{-2}, 10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}, 10^{-12}$ and plot these UCI with respect to the error levels. Figure 7 shows a trend from 70.2 to 88.3%.

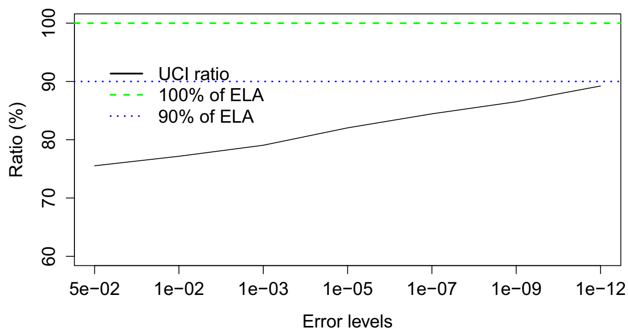

Fig. 7 UCI for the quantile associated to the probability 10^{-7} by flight hour with respect to the error levels. The green dashed line (resp. blue dotted line) represent 100% (resp. 90%) of the ELA value

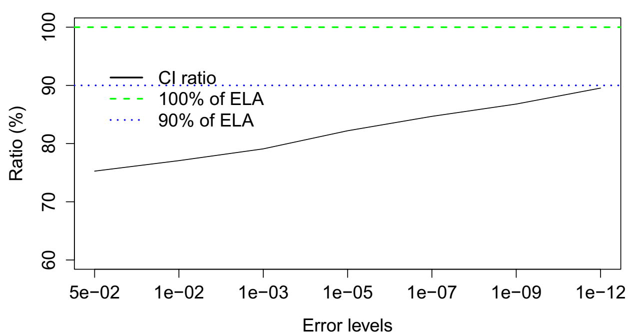

Fig. 8 CI for the endpoint with respect to the error levels. The green dashed line (resp. blue dotted line) represents 100% (resp. 90%) of the ELA value

From Table 3, we see that $\hat{\xi}$ is always negative and so we can assume that the endpoint exists and, from Eq. (5), is estimated at 71%. Using Eq. (6) we can build a CI around the endpoint estimate. Figure 8 gives the endpoint CI with respect to the error levels $\alpha = 5 \times 10^{-2}, 10^{-2}, 10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}, 10^{-12}$. It shows a trend between 75 and 90%.

The results from the quantile at $P_{\text{hour}} = 10^{-7}$ and the endpoint are close. For the group under study, with a reasonable risk error ($\alpha = 10^{-3}$) and to remain in accordance with the aeronautic safety procedures ($P_{\text{hour}} = 10^{-7}$), we can consider a ratio of 80% which means that the ELA is overestimating the electrical network by 20% with an error level of 10^{-3} for this group.

Concerning the assumptions advocated at the end of Sect. 3.2, most are clearly fulfilled in our context, namely that $n = 18\,319$ and $n_u = 150$ are large while $n_u/n = 8 \times 10^{-3}$ and $np/n_u = 5 \times 10^{-5}$ are small. It is not as clear when it comes to the assumption that $|\log(np)|/\sqrt{n_u}$ is small since it equals 0.77. It means that the extrapolation should not be pushed further and results concerning the UCI of extreme quantiles with smaller P_{hour} than 10^{-7} may not be valid anymore.

4.2 Global results

Using the EVT on the sampled groups we want to demonstrate that the ELA is overestimating maximal consumption for all groups. For that, we apply separately the same procedure to the 8 groups for the flight and onground phases to estimate extreme quantile, endpoint and their confidence intervals.

We use the same procedure as described in Sect. 4.1 to set the threshold and fit the GDP. Table 5 shows the parameter estimates for each group by phase. We see that the number of observations for the onground phase is smaller than for the flight phase which is coherent given the length of the two phases. All $\hat{\xi}$ are negative which implies a finite endpoint for all groups in both phases.

To compare the maximal electrical consumption between groups we need to compute the extreme quantiles and endpoint ratios by groups. Let $\hat{Q}_{10^{-7}}$ be the estimated extreme quantile associated to $P_{\text{hour}} = 10^{-7}$ and $\text{UCI}_{10^{-3}}$ its UCI at error level $\alpha = 10^{-3}$. Let $\text{CI}_{10^{-3}}$ be the CI at error level $\alpha = 10^{-3}$ for the estimated endpoint $\hat{X}^*$. The quantiles and endpoints estimates are given in Tables 6 and 7. Concerning the assumptions advocated at the end of Sect. 3.2, we can see that not all of them are fulfilled for all groups. The size n is large and the ratios n_u/n and np/n_u

Table 5 Maximum likelihood estimates $\hat{\beta}$ and $\hat{\xi}$ by group for the flight and onground phases

Group	Flight phase					Onground phase				
	n	n_u	n_u/n	$\hat{\beta}$	$\hat{\xi}$	n	n_u	n_u/n	$\hat{\beta}$	$\hat{\xi}$
1	227,988	316	0.001	2.23	−0.18	18,319	150	0.008	2.47	−0.24
2	35,504	150	0.004	3.46	−0.41	4,701	150	0.032	1.96	−0.24
3	232,296	150	0.001	2.2	−0.33	24,349	13	0.001	2.75	−0.34
4	113,787	200	0.002	1.64	−0.16	20,355	637	0.031	2.19	−0.16
5	455,263	430	0.001	1.91	−0.23	84,267	26	0.000	3.64	−0.44
6	123,430	600	0.005	3.19	−0.24	13,987	500	0.036	2.05	−0.21
7	38,063	150	0.004	2.1	−0.3	5,728	80	0.014	1.76	−0.35
8	40,104	120	0.003	3.15	−0.28	2,935	50	0.017	1.19	−0.22

Journal : Large 13272 | Article No : 540 | Pages : 12 | MS Code : 540 | Dispatch : 1-9-2021

Table 6 Quantiles associated to the probability 10^{-7} by flight hour and its UCI at error level of 10^{-3} by group for the flight and onground phases

Group	Flight phase		Onground phase	
	$\hat{X}_{10^{-7}}$	$\mathrm{UCI}_{10^{-3}}$	$\hat{X}_{10^{-7}}$	$\mathrm{UCI}_{10^{-3}}$
1	67.3	75.2	69.5	75.7
2	70.3	72.2	68.5	73.5
3	69.5	71.8	74.2	83.0
4	65.3	75.4	69.5	77.2
5	69.1	72.2	72.6	76.8
6	70.9	75.2	72.1	76.0
7	67.4	70.4	70.3	72.3
8	71.9	77.7	66.0	73.0

Table 7 Endpoints and its CI at error level 10^{-3} by group for the flight and onground phases

Group	Flight phase		Onground phase	
	$\hat{X}^*$	$\mathrm{CI}_{10^{-3}}$	$\hat{X}^*$	$\mathrm{CI}_{10^{-3}}$
1	68.5	75.9	69.8	75.6
2	70.3	72.1	68.7	73.4
3	69.7	71.8	74.4	82.6
4	66.6	76.1	70.6	77.8
5	69.6	72.5	72.7	76.6
6	71.4	75.5	72.4	76.1
7	67.5	70.3	70.3	72.2
8	72.2	77.5	66.2	72.8

are small in all situations. But the size n_u is quite small and $|\log(np)|/\sqrt{n_u}$ is quite large for the groups 3, 5, 7 and 8 for the onground phase. It means that the results concerning the UCI of the quantiles and the endpoints for these three groups during the onground phase have to be interpreted with caution. It also justifies the interest of gathering the different groups and phases if the results are sufficiently similar.

We observe that $\hat{Q}_{10^{-7}}$ and $\hat{X}^*$ are close. This can be explained by the fact that we are computing quantiles associated to $p = 10^{-9}$ to get the target probability 10^{-7} by flight hours and this probability is so small that we almost reach the endpoint. We see that the CI of the endpoint ratios by groups are aligned in a range of 70–83% which confirms our assumptions that the ELA overestimates the electrical consumption for permanent loads in nominal mode for the observed groups.

The largest endpoint ratio observed is 78% and 83% respectively for flight and onground phases but the ratio varies from one group to another. The final aim of this work is to generalize the observed ratio to all operational aircraft and to size the future aircraft generator. For that, we need to test if the endpoints can be considered the same for the different groups.

To this end we use an asymptotic chi-square test developed in [4]. This test checks the equality of the endpoints for independent random samples. We apply this test to check the equality of the group endpoints. We can consider that the assumption of independence between groups is satisfied as the electrical consumption of one group does not depend on the consumption of another. Let x_j^* be the endpoint of the j^{th} group with $j = 1, \ldots, 8$. We consider the following hypotheses:

$$\begin{cases} H_0 \, : \, x_1^* = \ldots = x_8^* \\ H_1 \, : \, \text{the } x_j^* \text{ are not all equal.} \end{cases}$$

The test statistic is

$$S = d \sum_{j=1}^{8} r_j (\hat{X}_j^* - \tilde{X})^2$$

where $\tilde{X} = \sum_{j=1}^{8} r_j \hat{X}_j^*$, with $r_j = \dfrac{d_j}{d}$, $d = \sum_{j=1}^{8} d_j$, $d_j = \dfrac{n_u^j}{\hat{\beta}_j^2 \, \tau(\hat{\xi}_j^2)}$ and $\tau(\hat{\xi}_j)^2 = 2 + 2\xi_j^{-1} + 5\xi_j^{-2} + 4\xi_j^{-3} + \xi_j^{-4}$, where n_u^j (resp. $\hat{\xi}_j$ and $\hat{\beta}_j$) are the number of observations that exceed threshold u (resp. the shape and the scale estimators) for group j.

Under H_0, [4] demonstrates that the test statistic S follows a chi-square distribution with 7 degrees of freedom. We reject H_0 at level α if $S > q_{\chi^2_7(1-\alpha)}$ where $q_{\chi^2_7(1-\alpha)}$ stand for the $(1-\alpha)$-quantile of the chi-square distribution with 7 degrees of freedom.

The result of this statistical test is given in Table 8. The p values for both phases are greater than 0.05 hence the hypothesis that the endpoints are equal is not rejected with a 5% risk error.

We do not reject that group endpoints are equal (for both phases) which means that the largest possible value of the maximum of electrical network consumption divided by the ELA value does not depend on the groups. This result can also be deduced from Figs. 9 and 10 where the endpoints

Table 8 Chi-square test for groups endpoint equality in flight and onground phases

Group	$\hat{X}^*$ flight	$\hat{X}^*$ onground
1	68.5	69.8
2	70.3	68.7
3	69.7	74.4
4	66.6	70.6
5	69.6	72.7
6	71.4	72.4
7	67.5	70.3
8	72.2	66.2
S	11.7	13.1
p value	0.11	0.07

Journal : Large 13272 | Article No : 540 | Pages : 12 | MS Code : 540 | Dispatch : 1-9-2021

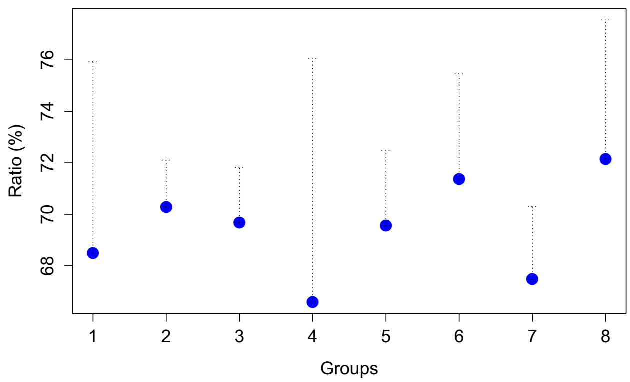

Fig. 9 Endpoints and their CI by group for the flight phase represented by the dashed bars

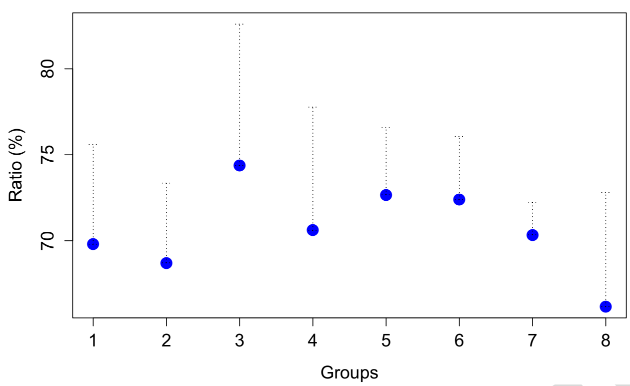

Fig. 10 Endpoints and their CI by group for the onground phase represented by the dashed bars

Table 9 Parameters, quantiles, endpoint estimates and their confidence interval for gathered group by phase

Phase	n	n_u	n_u/n	$\hat{\beta}$	$\hat{\xi}$	$\hat{X}_{10^{-7}}$	$\mathrm{UCI}_{10^{-3}}$	$\hat{X}^*$	$\mathrm{CI}_{10^{-3}}$
Flight	1 266 435	335	0.000	1.84	– 0.24	71.5	74.7	71.6	74.6
Onground	174 640	120	0.001	1.45	– 0.2	73.3	79.7	73.5	79.6

estimates are represented by dots and the corresponding CI by dashed bars. These figures are graphical representations in connection with the chi-square test results and help us to check the equality of endpoints. We confirm graphically the equality of endpoints for both phases since the CI intersect with each other on the two figures.

As the endpoints equality test suggests that there is no effect of the group on the estimated ratio, we gather all groups and estimate a global ratio taking into account all groups. We apply the EVT separately to the flight and the onground phases. The parameters, extreme quantiles, endpoints estimates and their CI are given in Table 9. It shows that we still have a negative $\hat{\xi}$ and thus a finite endpoint. The ratio estimates of extreme quantile and endpoint are close and around 75% for the flight phase and around 80% for the onground phase. Comparing to the ratios found in Table 7 the results are aligned.

To go further in generalizing this ratio and since the endpoints for flight and onground phase are close we check if the endpoints are equal. Table 10 provides the results of the chi-square test of endpoint equality between the flight and the onground phases. The test illustrates that we cannot reject the equality of endpoints at 5% error level and thus the estimated ratio can be considered as independent of the phase.

From this result, we gather also the two phases and apply the EVT on the gathered groups with no distinction between flight and onground phases. Table 11 shows the maximum likelihood estimates of the parameters β and ξ for the gathered groups and phases, we still have $\hat{\xi} < 0$ and thus consider a finite endpoint.

The extreme quantile is estimated at 72.9% and the endpoint at 74.3%. The UCI and CI are given in Figs. 11 and 12 where we vary the error levels $\alpha = 5 \times 10^{-2}, 10^{-2}, 10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}, 10^{-12}$ and plot the UCI of extreme quantile and CI of the endpoint with respect to the error level. As could be expected, we observe an increasing trend for extreme quantile and endpoint ratios. They both vary from 75 to 87%. We see that for the error level $\alpha = 10^{-3}$ we have a ratio of 80% which is in line with the previous results.

In all applications of the EVT, by groups and on gathered data, we get a maximal ratio of 80% for an error level 10^{-3}. From these results we can consider a ratio of 80% for the generator with permanent loads in nominal mode.

Table 10 Chi-square test for phases endpoint equality

Phase	$\hat{X}^*$
Flight	71.6
Onground	73.5
S	0.8
p value	0.381

Table 11 Parameters estimates associated to the gathered groups and phases

n	n_u	n_u/n	$\hat{\beta}$	$\hat{\xi}$
1441 076	500	0.000	1.5	−0.13

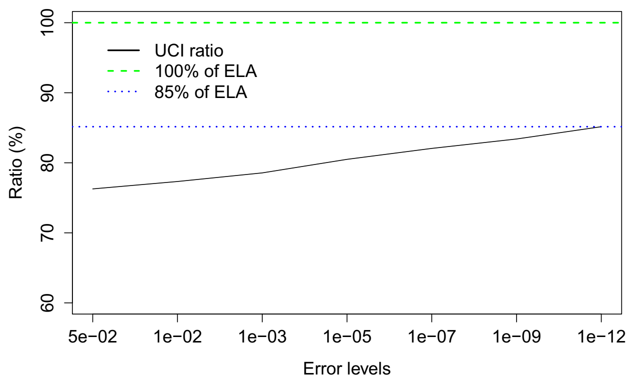

Fig. 11 UCI for the quantile associated to the probability 10^{-7} by flight hour with respect to the error levels for the gathered groups and phases. The dotted line represents the maximum ratio obtained

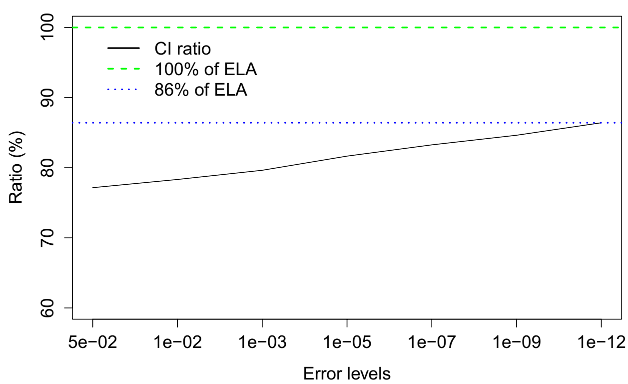

Fig. 12 CI for the endpoint with respect to the error levels for the gathered groups and phases. The dotted line represents the maximum endpoint ratio obtained

5 Conclusion

In this paper, we use the extreme value theory to estimate extreme ratios associated to probability 10^{-7} by flight hour and endpoint ratios, we also build confidence intervals at error level 10^{-3} to check whether the ELA overestimates the maximal consumption. We detail the statistical procedure for permanent loads of a generator in the nominal mode for a specific group. Then, we apply the EVT to 8 groups and demonstrate that the largest ratio is around 83% for the permanent loads in the nominal mode.

To generalize this gap to all operational aircraft and to size the future aircraft generators, we do an asymptotic chi-square test to check that the group endpoints are equal. The endpoints equality is not rejected for both phases which means that there is no group effect on the ratio endpoint. Then we gather all groups to estimate extreme quantiles and endpoint ratios for each of the two phases and we end up with a ratio of 75% for flight phase and 80% for the onground phase. To obtain a global ratio, we check if there is a difference between the flight and onground phases using the endpoint equality test. Again the equality assumption is not rejected and after gathering the two phases, we obtain an endpoint ratio of 80%.

Using a statistical approach, we quantify how much the ELA overestimates the maximal electrical consumption of the generator. For instance, with an error level of 10^{-3} for permanent loads in the nominal mode, our study leads to an excess of 20% for the considered generator.

However, the study only relies on permanent loads in the nominal mode for low-cost aircraft. To complete the electrical network assessment, we need to incorporate also non low-cost aircraft in our analysis and extend the study to the intermittent loads and failure modes. In particular, future work should focus on the degraded mode (loss of generators) to size the generators.

Supplementary Information The online version supplementary material available at https://doi.org/10.1007/s13272-021-00540-8.

Acknowledgements This work has been partly supported by the French Agence Nationale de la Recherche through CIFRE contract 2017/1354 and through the Investments for the Future (Investissements d'Avenir) program, Grant ANR-17-EURE-0010.

References

1. Castillo, E., Hadi, A.S., Balakrishnan, N., Sarabia, J.M.: Extreme Value and Related Models with Applications in Engineering and science. Wiley, Hoboken (2005)
2. Coles, S., Bawa, J., Trenner, L., Dorazio, P.: An Introduction to Statistical Modeling of Extreme Values, vol 208. Springer (2001)
3. De Haan, L., Ferreira, A.: Extreme Value Theory: An Introduction. Springer (2007)
4. Einmahl, J.J., Einmahl, J.H., de Haan, L.: Limits to human life span through extreme value theory. J Am Stat Assoc **114**, 1–10 (2019)
5. El Adlouni, S., Ouarda, T.B.M.J., Zhang, X., Roy, R., Bobée, B.: Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. Water Resour. Res. **43**(3), W03410 (2007)
6. Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling Extremal Events: for Insurance and Finance, vol 33. Springer (2013)
7. Ganger, D., Zhang, J., Vittal, V.: Statistical characterization of wind power ramps via extreme value analysis. IEEE Transa. Power Syst. **29**(6), 3118–3119 (2014)
8. Gilleland, E., Katz, R.W., et al.: extremes 2.0: an extreme value analysis package in *r*. J. Stat. Softw. **72**(8), 1–39 (2016)
9. Larson, J., Gebre-Egziabher, D.: Conservatism assessment of extreme value theory overbounds. IEEE Trans. Aerosp. Electron. Syst. **53**(3), 1295–1307 (2017)
10. Roblot, G.: Méthodologie de pré-dimensionnement de la puissance électrique des générateurs d'un réseau embarqué à partir d'analyses statistiques des consommateurs. PhD thesis, Nantes University (2012)

11. Seresinhe, R., Lawson, C.: Electrical load-sizing methodology to aid conceptual and preliminary design of large commercial aircraft. Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng. **229**(3), 445–466 (2015)
12. Sun, X., Shi, J., Wang, S., Zhang, C.: Design of load spectrum for hydraulic pumps based on extreme value theory. In: 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), IEEE, pp 1321–1326 (2017)
13. Westerlund, P., Naim, W.: Extreme value analysis of power system data. In: ITISE 2019-International Conference on Time Series and Forecasting, 25-27 September 2019 Granada (Spain), vol 1, pp 322–327 (2019)

Journal : **Large 13272**	Article No : **540**	Pages : **12**	MS Code : **540**	Dispatch : **1-9-2021**