Transportation Research Part A 109 (2018) 65–75

Contents lists available at ScienceDirect

Transportation Research Part A

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

Prediction of U.S. General Aviation fatalities from extreme value approach

Abdoulaye Diamoutene[a,b,c], Bernard Kamsu-Foguem[a,*], Farid Noureddine[a], Diakarya Barro[c]

[a] Université de Toulouse, Laboratoire de Génie de Production (LGP), EA 1905, 47 Avenue d'Azereix, BP 1629, 65016 Tarbes Cedex, France
[b] Institut Polytechnique Rural de Formation et de Recherche Appliquée de Katibougou Mali, BP: 06 Koulikoro, Mali
[c] Université Ouaga 2, UFR-SEG, 12 BP: 417 Ouagadougou 12, Burkina Faso

ARTICLE INFO

Keywords:
General Aviation accident
Extreme value statistics
Graphical diagnosis
Return level
Uncertainty
Nonparametric bootstrap

ABSTRACT

General Aviation is the main component of the United States civil aviation and the most aviation accidents concern this aviation category. Between early 2015 and May 17, 2016, a total of 1546 general aviation accidents in the United States has left 466 fatalities and 384 injured. Hence, in this study, we investigate the risk of U.S. General Aviation accidents by examining historical U.S. General Aviation accidents. Using the Peak Over Threshold approach and Generalized Pareto Distribution, we predict the number of fatalities resulting in extreme GA accidents in the future operations. We use a graphical method and intensive parameters estimates to obtain the optimal range of the threshold. In order to assess the uncertainty in the inference and the accuracy of the results, we use the nonparametric bootstrap approach.

1. Introduction

General Aviation (GA) is the term for all civil aviation operations other than scheduled air service and non-scheduled air transport operations for remuneration or hire, see Crane (1997) and National Transportation Statistics NTS (2016). GA flights range from gliders and powered parachutes to corporate business jet flights, see National Transportation Safety Board NTSB (2011) and Civil Aviation Authority CAA (2013). The majority of the world's air traffic falls into this category and most of the world's airports serve general aviation exclusively, see Crane (1997). The term General Aviation is a catch-all phrase for all aviation activities that do not fall under commercial aviation, major cargo or military operations, and covers a broad spectrum of aviation activity uses.

Between early 2015 and May 17, 2016, a total of 1546 general aviation accidents in the United States has left 466 fatalities and 384 injured, according to data from the NTSB.

In 2011, GA aircrafts were involved in 95% of all aviation accidents and 94% of fatal aviation accidents; accidents involving GA aircrafts accounted for 92% of all U.S. civil aviation fatalities, see NTSB (2011); it is an important component of civil aviation in the United States, see Shetty and Hansman (2012).

GA covers a large range of activities, both commercial and non-commercial, including flying clubs, flight training, agricultural aviation, light aircraft manufacturing and maintenance. GA is particularly popular in North America, with over 6,300 airports available for public use by pilots of GA aircraft in the U.S., see NTSB (2011) and NTS (2016).

In 1990–2014, there is in United States GA a total of 8402 fatal accidents which resulted in about 14984 fatalities according to data from the National Transportation Safety Board (NTSB). Among these 8402 fatal accidents, around 26 accidents can be

* Corresponding author.
E-mail address: bernard.kamsu-foguem@enit.fr (B. Kamsu-Foguem).

https://doi.org/10.1016/j.tra.2018.01.022
Received 9 October 2017; Received in revised form 17 January 2018; Accepted 17 January 2018
0965-8564/

considered as extreme and each of them was responsible for more than 39 fatalities. Moreover, in extreme value approach, an extreme event is an event which has a low probability, but implies serious consequences. In our context, an extreme accident is an accident that has caused over 39 fatalities. Although there is an overall decreasing trend in the annual number of fatalities, see NTS (2016) and Shetty and Hansman (2012), there is no such trend in the number of extreme fatal accidents which cause enormous fatalities. Prediction and risk management in terms of extreme fatality are more important in insurers, reinsurers and aeronautics industry like Boeing and Airbus, see Boeing (2016) and Airbus (2016). Moreover, based on historical data, the return level of fatalities could help management agency of GA to perform aircraft operations, conditions-based maintenance, predictive maintenance and flying conditions.

The safety of civil aviation in the United States is regulated by the Federal Aviation Administration (FAA), for more details see NTSB (2011). The FAA distinguishes between commercial operations and GA operations, see NTSB (2011). There are other organizations like International Civil Aviation Organization (ICAO), Civil Aviation Authority (CAA) of the United Kingdom (UK) and leading manufacturers of commercial jetliners such as Boeing, Airbus give valuable reports on civil aviation accident. These reports mainly focus on descriptive accident statistics such as the rate of fatal accidents, fatalities by year, nature of flight, type of aircraft, transport category, flight hours, for more details, see NTSB (2011) and NTS (2016).

In literature, some works on aviation accidents focus on the cost of GA accidents in U.S., see Sobieralski (2012); Knetch (2015) focuses on the prediction of the accident rate due to pilot total flight hours; Fultz and Ashley (2016), study extreme weather impact in GA accidents in the United States; Farah and Azevedo (1952) focus on safety analysis of passing maneuvers using extreme value theory. Sha et al. (2017) focus on the use of the extreme value theory for forecasting spare parts stocks on the basis of repair data set information, see Romeijnders et al. (2012). Dey and Das (2016) propose an approach to predict extreme aviation fatal injuries. But in literature, it is rare to see research concerning the modeling of the number of fatalities in individual U.S. GA accidents; using extreme value models, notably Peaks Over Threshold (POT) approach. From extreme value statistics, we can model the return level of fatalities due to extreme GA accidents. We distinguish two approaches namely, block maxima approach and Peak Over Threshold (POT) approach, see Coles (2001). The block maxima approach is wasteful of data as only one data point in each block is taken. The second highest value in one block may be larger than the highest of another block and this is generally not accounted for, see Coles (2001) and Castillo et al. (2005). Peak Over Threshold (POT) approach avoids this drawback, by using the Generalized Pareto Distribution (GPD). Indeed, due to advances in extreme value theory, the GPD emerged as a natural family for modeling exceedances over a high threshold. The POT approach has shown its importance and success in a number of statistical analysis problems related to different areas, see Schmidt et al. (2014) and Castillo et al. (2005). However, despite the sound theoretical basis and wide applicability, the fitting of this distribution in practice is not a trivial exercise.

In this paper we focus on U.S. GA fatal accidents from 1990 to 2014 and their resulting fatalities number. A prediction of the possible number of fatalities for an extreme GA accident in the future operation is made using POT approach. We use three estimation methods to estimate the parameters and select the method which standard error is less than in the other two methods. The two main factors which affect the accuracy of the estimations of the return values are the choice of the threshold on one hand and the choice of the parameter estimator on the other hand. An extensive discussion of the various parameter estimation methods has been given by Bermudez and Kotz (2010). Scarrot and Macdonald (2012) reviewed the threshold estimation methods; hence, we use graphical method and intensive estimate of GPD parameters to obtain a reasonable range of the threshold. We also quantify the uncertainty in the inference using nonparametric bootstrap approach, see Carpenter and Bithel (2000) and Yongcheng (2008).

This paper is structured as follows: following this introduction, Section 2 provides background on extreme values statistics, notably both block maxima approach and POT approach. Section 3 exposes GA accident modeling with POT approach. Section 4 presents different uncertainty measurements in the inference of U.S. GA extreme fatalities and finally Section 5 concludes and discusses future challenges.

2. Extreme value theory

Most statistical methods are concerned primarily with what goes on in the center of a statistical distribution, and do not pay particular attention to the tails of a distribution, or in other words, the most extreme values at either high or low end. However, in extreme value analysis, we are not interested in estimating the average we rather want to quantify the behavior of the process at unusually large or small levels, see Coles (2001). Extreme Value Theory (EVT) deals with the extreme deviations from the median of probability distributions and seeks to assess, from a given ordered sample of a given random variable, the probability of events that are more extreme than a certain large value, see Coles (2001) and Castillo et al. (2005).

Broadly speaking, there are two principal kinds of models for extreme values. The first group of model, called Generalized Extreme Value theory (GEV) consists of block- maxima approach; these are models for the largest observations collected from large samples of identically distributed observations. A second group of EVT models contains the Peak-Over-Threshold (POT) models, which model all large observations that exceed a high threshold, for more details see Coles (2001) and Castillo et al. (2005). These POT models are generally considered to be most useful for practical applications, due to their more efficient use of the (often limited) data of extreme values. Within the POT class of models, several styles of analysis exist, see Coles (2001) and Schmidt et al. (2014). Among all these parametric models, the Generalized Pareto Distribution (GPD) has the advantage of being conceptually simple and implemented in software R, see Ribatet (2012).