



TRIPLE PROBABILITY DENSITY DISTRIBUTION MODEL IN THE TASK OF AVIATION RISK ASSESSMENT

Ivan OSTROUMOV ^{1,*}, Karen MARAIS ², Nataliia KUZMENKO ³, Nicoletta FALA ⁴

^{1, 3}*Air Navigation Systems Department, National Aviation University, Kyiv, Ukraine*

²*School of Aeronautics and Astronautics, Purdue University, West Lafayette, USA*

⁴*Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, OK, USA*

Received 10 January 2020; accepted 02 April 2020

Abstract. The probability of an airplane deviation from pre-planned trajectory is a core of aviation safety analysis. We propose to use a mixture of three probability density distribution functions in the task of aviation risk assessment. Proposed model takes into account the effect of navigation system error, flight technical error, and occurrence of rare events. Univariate Generalized Error Distribution is used as a basic component of distribution functions, that configures the error distribution model from the normal error distribution to double exponential distribution function. Statistical fitting of training sample by proposed Triple Univariate Generalized Error Distribution (TUGED) is supported by Maximum Likelihood Method. Optimal set of parameters is estimated by sequential approximation method with defined level of accuracy. The developed density model has been used in risk assessment of airplane lateral deviation from runway centreline during take-off and landing phases of flight. The efficiency of the developed model is approved by Chi-square, Akaike's, and Bayes information criteria. The results of TUGED fitting indicate better performance in comparison with double probability density distribution model. The risk of airplane veering off the runway is considered as the probability of a rare event occurrence and is estimated as an area under the TUGED.

Keywords: risk, airplane, aviation safety, probability density function, Triple Univariate Generalized Error Distribution, deviation, Maximum Likelihood Method, statistics.

Introduction

Operation of the air transport system is performed in accordance with a certain safety level. The safety of air transportation depends on numerous factors whose action is taken into account by assessing the risks of their operation in the form of the frequency of their occurrence over a certain period of time. A special place in aviation safety is given to assessing the risk of mid-air collision (Mitici & Blom, 2018). Results of the collision risk assessment are used for the separation minima estimation in air traffic management. A risk of mid-air collision can be evaluated as a product of the probabilities of airplanes deviations from the preplanned trajectory of flight (Fujita, 2013). The risk of collision with the terrain at low altitudes can be considered as the probability of airplane proximity to a dangerous point of relief. Consideration of collision risks and events that can lead to unsafe flight (Fala & Marais, 2016) is an important step of airspace design and usage.

Performance of safety assessment depends on the numerous factors that are taken into account in the research model. Simple models of risk assessment based on the statistical processing of airplane trajectory data are commonly used. Also, analytical functions of probability distributions are used in the tasks of assessing the risks of rare events. In particular, the assessment of the risk of airplane veer off the boundaries of the runways can be considered as an assessment of an extremely rare event. The risk associated with rare events is practically impossible to assess by the frequency of occurrence due to the limited statistical data sample. Therefore, estimation of the probability of rare events is performed on the basis of statistical data processing taking assumptions concerning the function of random variables distribution. In this case, parameters estimation of the probability distribution function is performed. Probability of random value is evaluated as an area below probability density function (PDF) limited by the confidence band and reference frame. In the common

*Corresponding author. E-mail: ostroumovv@ukr.net

tasks of risk analysis, a set of confidence bands is used to estimate probabilities of parameter deviation occurrence.

A mixture of two probability density functions is commonly used in the task of risk assessment of airplane deviation from a cleared flight level or pre-planned trajectory in both horizontal and vertical planes. For example, the mixed distribution model (*MD*) consisting of core distribution (*CD*) and tail distribution (*TD*) is considered in numerous studies (Mori, 2011; Nagaoka, 2008):

$$MD(x) = \gamma CD(x) + (1 - \gamma) TD(x), \quad (1)$$

where x is an airplane deviation from pre-planned trajectory, γ is the probability of a gross navigational error.

The *CD* represents errors derived from standard navigation system deviations. These errors depend on the on-board navigation system precision level and human factor. The *TD* corresponds to non-nominal performance and represents result of rare factors degradation influence. *TD* corresponds to rare events simulation.

Different PDFs can be used as *CD* and *TD* in a model (1). Normal Density Function (NDF), Exponential Density Function (EDF), Double Exponential Density Function (DEDF) (ICAO, 1988; Nagaoka, 2008), Laplacian, Freshet, Weibull, Gumbel (ICAO, 2008b), Generalized Pareto Distribution (GPD) are often used in risk assessment within air navigation system (ICAO, 2008b).

The main problem of model (1) usage is a correct selection of α coefficient, which characterizes the *CD* and *TD* mixture level. Incorrect or inaccurate selection of α is the reason for a certain overestimation or underestimation of the risk evaluation by the PDF on the tails.

CD utilizes errors of on-board positioning system and flight technical error which is result of human factor influence. Performance of on-board positioning system has been increasing rapidly during last three decades. An accuracy improving is a result of development of advanced signal processing algorithms, new antenna design, introduction of numerous augmentation systems. Thus, an error of on-board positioning system becomes much smaller than flight technical error in manual mode. Different degrees of these errors make usage of one PDF in *CD* inaccurate. Thus, model (1) is not adequate in the case of imbalanced errors.

The main purpose of current research is to design a new model of risk assessment connected with airplane deviations which take into account imbalanced error contribution of positioning system and flight technical errors. Separation of PDFs in the risk model leads to improvement of risk assessment performance and makes the model adequate to current equipment level.

1. Triple probability density distribution model

In the case of estimating the airplane deviation from the planned trajectory by statistic, the *CD* characterizes the distribution of Total System Error (*TSE*), that is, the total error of maintaining a given trajectory. The *TSE* is a sum

of Navigation System Error (*NSE*), Flight Technical Error (*FTE*), and Path Definition Error (*PDE*) (ICAO, 2006, 2008a):

$$TSE^2 = NSE^2 + FTE^2 + PDE^2. \quad (2)$$

In a general case, the *NSE* is determined by an error of airplane positioning and depends on the type of navigation system. *FTE* characterizes the ability of a pilot or an automatic piloting system to follow a predetermined flight trajectory. In case of manual control, the *FTE* includes errors of indication by flight instruments and data interpretation by pilot. *PDE* utilizes the difference between path defined in Flight Management System and path expected to be flown over the ground. Due to use of on-board computing systems with digital route mapping support, the impact of *PDE* is too small (compared to *NSE* and *FTE* values) that its value can be neglected (ICAO, 2008a).

According to error model (2), the deviations from the pre-planned trajectory should be represented by decomposition of *CD* on the main components corresponding to the *NSE* and *FTE* in tasks of risk assessment. Thus, a proposed model includes a mixture of three PDFs (Figure 1):

$$\rho(x) = \alpha \rho_{NSE}(x) + \beta \rho_{FTE}(x) + (1 - \alpha - \beta) \rho_T(x), \quad (3)$$

where $\rho_{NSE}(x)$ is the PDF utilizing the errors of navigation system; $\rho_{FTE}(x)$ is the PDF characterizing the *FTE*; $\rho_T(x)$ is the PDF characterizing the appearance of rare events; α and β are weight coefficients of $\rho_{NSE}(x)$ and $\rho_{FTE}(x)$ in the mixture.

Global Navigation Satellite System (GNSS) can be considered as a primary on-board positioning system of airplane due to a high level of accuracy, availability, and continuity of measured data. Influence of some factors can degrade performance of GNSS and may lead to positioning lock. Therefore, inertial navigation system or positioning by navigational aids data can be used as a stand-by positioning systems (Ostroumov & Kuzmenko, 2018). Wide-area multilateration system, primary and secondary surveillance radars are used for airplanes coordinates measurement for air traffic control purposes. Also, LIDAR data (Ryu & Young, 2016) or localization by image processing (Naidu & Durgarao, 2012) are used in some risk

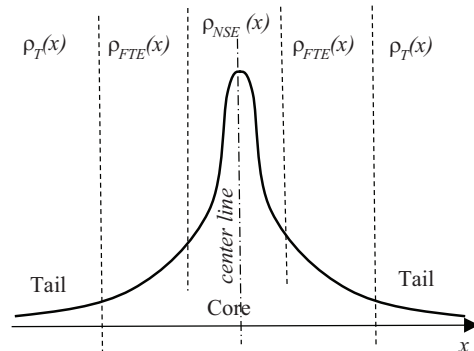


Figure 1. Components of triple probability density distribution model

analysis, but these approaches require specific sensors on-board or in the ground infrastructure.

Typical positioning system errors are described by a normal error distribution model with different parameters for lateral, longitudinal and vertical directions. GNSS error distribution model depends on time and user location, while radar error distribution model depends mostly on slant distance to airplane. Risk assessment of lateral deviations can be performed by a statistical analysis of trajectory data measured by on-board sensor. Position measurements by Global Positioning System (GPS) with Satellite-based (SBAS) or ground-based augmentation systems (GBAS) allow getting precise trajectory data in the terminal area at a level that satisfies the conditions for CAT 3 landing system (less than 1m). Performance of positioning by navigational aids does not reach level of GNSS in normal operational mode (Kuzmenko & Ostroumov, 2018; Ostroumov et al., 2019). However, GNSS error depends on a number of factors such as: geometry of location in space, state of ionosphere, interference of radio waves propagation, interference and unintended jamming of navigation signals (Kutsenko et al., 2018). Some of these factors are difficult to take into account because their actions are related to extremely rare events. For example, the state of the ionosphere depends on the amount of solar radiation and the state of the Earth’s magnetic field. Also, simulation of the probability of unintentional jamming of GNSS signals can be considered in terms of fuzzy sets.

The NDF should be considered as *FTE* model in (2) (Cramer & Rodriguez, 2013). A comparative analysis of *NSE* and *FTE* shows that *NSE* values are smaller than *FTE* due to wide range of admissible deviations from flight path in lateral plane and high accuracy of trajectory data measured by GNSS.

The simulation of rare events $\rho_T(x)$ is usually performed by an EDF, that “moves up” the tails according to input statistics.

We propose to use Univariate Generalized Error Distribution (UGED) as $\rho_{NSE}(x)$, $\rho_{FTE}(x)$ and $\rho_T(x)$ density functions to make model (3) universal due to high flexibility property. The domain of the UGED is $x \in [-\infty, \infty]$ and the function is defined by three parameters: $\mu \in [-\infty, \infty]$ is the mean value of distribution; $a \in [0, \infty]$ defines the dispersion; and, $b \in [0, \infty]$ controls the PDF shape. UGED can be represented in different forms (Giller, 2013; Ayebo & Kozubowski, 2003; Czyżycki, 2013). We use the follows:

$$\rho_{UGED}(x, a, b, \mu) = \frac{1}{2ab\Gamma(b)} \exp\left(-\left|\frac{x-\mu}{a}\right|^{\frac{1}{b}}\right);$$

$$\Gamma(b) = \int_0^{\infty} e^{-t} t^{b-1} dt, \quad (4)$$

where a is a scale factor; b is a shape coefficient; μ is a mean value; $\Gamma(b)$ is a Euler-gamma function.

Flexibility of (4) makes UGED similar to NDF if $b = 0.5$, $a = \sqrt{2}$ and DEDF if $b = 1$, $a = \lambda - 1$ (considering $\Gamma(1) = 1$ and $\Gamma(0,5) = 0.5\pi$):

$$\rho_{UGED}(x, \lambda^{-1}, 1, \mu) = \rho_{DE}(x, \lambda, \mu) = \frac{\lambda}{2} \exp(-\lambda|x-\mu|); \quad (5)$$

$$\rho_{UGED}(x, \sigma\sqrt{2}, 0.5, \mu) = \rho_N(x, \sigma, \mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad (6)$$

where λ is a rate parameter; σ is a standard deviation; μ is a mean value.

In general form Triple UGED (TUGED) model can be represented by (3) with PDF in the form (4):

$$\rho_{TUGED}(x, \alpha, \beta, A, B, M) = \frac{\alpha}{2a_1b_1\Gamma(b_1)} \exp\left(-\left|\frac{x-\mu_1}{a_1}\right|^{\frac{1}{b_1}}\right) +$$

$$\frac{\beta}{2a_2b_2\Gamma(b_2)} \exp\left(-\left|\frac{x-\mu_2}{a_2}\right|^{\frac{1}{b_2}}\right) + \frac{1-\alpha-\beta}{2a_3b_3\Gamma(b_3)} \exp\left(-\left|\frac{x-\mu_3}{a_3}\right|^{\frac{1}{b_3}}\right); \quad (7)$$

$$\alpha + \beta < 1,$$

where $A = [a_1, a_2, a_3]$ is a matrix of scale factors; $B = [b_1, b_2, b_3]$ is a matrix of shape coefficients; $M = [\mu_1, \mu_2, \mu_3]$ is a matrix of mean values.

In certain tasks, the PDF displacement in a horizontal plane can be neglected to make deviations in positive and negative directions of equal probabilities in order to improve computation performance. Then:

$$\mu = \mu_1 = \mu_2 = \mu_3.$$

Parameters in matrixes A and B , obtained after TUGED fitting to a series of data, evaluate each component of PDFs mixture in wide range of curves from NDF to DEDF. This property of TUGED makes its flexible to any input sensor data.

2. Model parameters estimation

The TUGED model in form (7) can be used in statistical analysis of lateral deviations from the pre-planned airplane trajectory. The TUGED is defined by a set of parameters estimated by statistics. Grouping of statistics by different properties (for example, under certain conditions of flight or requirements of area navigation) creates a series of airplane deviation models and gives a list of different probabilities assessment.

Different methods of mathematical statistics can be used to estimate the parameters of TUGED by experimental set of data. Maximum Likelihood Method (MLM) (Mori, 2011; Klein, 1980) or Least Squares Method (Markovsky & Huffel, 2007) can be applied. The choice of evaluation method depends on two basic factors: the statistical properties of a valid estimator and the computational

complexity of the method. Analytical solution is impossible in many cases due to high complexity of the calculation. Moreover, the computation complexity grows with increasing the number of estimated parameters.

The least-squares method is more oriented on a linear model and is undesirable for nonlinear functions. The MLM is based on likelihood function estimation that provides probabilistic relation for parameters evaluation and confidence band data simultaneously. Also, MLM compares the ability of different models to fit the statistics. According to (ICAO, 1988), the main disadvantage of the MLM makes method less sensitive to the data structure within the tail in comparison with structure in the center of distribution. In order to improve MLM performance, tail data uses greater weight than in the central part. The mentioned above advantages make MLM reasonable to use in the task of TUGED fitting to input data.

We assume that X is a discrete random variable represented by n measured values x_1, x_2, \dots, x_n . A variable X is distributed randomly under TUGED (7) with unknown matrix $\theta = [\alpha, \beta, A, B, M]$.

According to MLM, the likelihood function is the following:

$$L(x_i, \theta) = \prod_{i=1}^n f(x_i, \theta), \quad (8)$$

where $f(x_i, \theta)$ is probability that x variable takes a value of x_i ; $i = [1, n]$; n is a number of measurements; θ is a matrix of TUGED parameters.

The likelihood function takes the maximum value at the point of optimal fitting with values of θ . The maximum value of function can be reached at the point of first order derivative from the likelihood function equal to zero:

$$\frac{\partial}{\partial \theta} L(x_i, \theta) = \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i, \theta) = 0. \quad (9)$$

Since the functions $L(x_i, \theta)$ and $\ln L(x_i, \theta)$ take a maximum at one value of θ , we localize the maximum function of $\ln L(x_i, \theta)$ to replace product operator by a sum in order to improve computation performance:

$$\frac{\partial}{\partial \theta} \ln L(x_i, \theta) = \frac{\partial}{\partial \theta} \sum_{i=1}^n \ln f(x_i, \theta) = 0. \quad (10)$$

Finally, we have the system of non-linear equations:

$$\begin{cases} \sum_{i=1}^n \left(\frac{\partial f(x_i, \theta)}{\partial \theta} \cdot \frac{1}{f(x_i, \theta)} \right) = 0 \\ \theta = [\alpha, \beta, A, B, M] \end{cases}. \quad (11)$$

The first order derivatives of (7) by parameters $\alpha, \beta, a_1, a_2, a_3, b_1, b_2, b_3$ are:

$$\begin{aligned} \frac{\partial p_{UGED}(x, \alpha, \beta, A, B, M)}{\partial a_1} &= \frac{-\alpha}{2a_1 b_1 \Gamma(b_1)} \exp \left(- \left| \frac{x - \mu_1}{a_1} \right|^{\frac{1}{b_1}} \right) \\ &\left(\frac{1}{a_1} + \frac{1}{b_1} \left| \frac{\mu_1 - x}{a_1^2} \right| \left| \frac{x - \mu_1}{a_1} \right|^{\frac{1}{b_1} - 1} \right); \end{aligned} \quad (12)$$

$$\begin{aligned} \frac{\partial p_{UGED}(x, \alpha, \beta, A, B, M)}{\partial a_2} &= \frac{-\beta}{2a_2 b_2 \Gamma(b_2)} \exp \left(- \left| \frac{x - \mu_2}{a_2} \right|^{\frac{1}{b_2}} \right) \\ &\left(\frac{1}{a_2} + \frac{1}{b_2} \left| \frac{\mu_2 - x}{a_2^2} \right| \left| \frac{x - \mu_2}{a_2} \right|^{\frac{1}{b_2} - 1} \right); \end{aligned} \quad (13)$$

$$\begin{aligned} \frac{\partial p_{UGED}(x, \alpha, \beta, A, B, M)}{\partial a_3} &= \frac{-(1 - \alpha - \beta)}{2a_3 b_3 \Gamma(b_3)} \exp \left(- \left| \frac{x - \mu_3}{a_3} \right|^{\frac{1}{b_3}} \right) \\ &\left(\frac{1}{a_3} + \frac{1}{b_3} \left| \frac{\mu_3 - x}{a_3^2} \right| \left| \frac{x - \mu_3}{a_3} \right|^{\frac{1}{b_3} - 1} \right); \end{aligned} \quad (14)$$

$$\begin{aligned} \frac{\partial p_{UGED}(x, \alpha, \beta, A, B, M)}{\partial b_1} &= \frac{-\alpha}{2a_1 b_1 \Gamma(b_1)} \exp \left(- \left| \frac{x - \mu_1}{a_1} \right|^{\frac{1}{b_1}} \right) \\ &\left(-\frac{1}{b_1} - \frac{1}{\Gamma(b_1)} \frac{\partial \Gamma(b_1)}{\partial b_1} + \frac{1}{b_1} \left| \frac{x - \mu_1}{a_1} \right|^{\frac{1}{b_1}} \ln \left| \frac{x - \mu_1}{a_1} \right| \right); \end{aligned} \quad (15)$$

$$\begin{aligned} \frac{\partial p_{UGED}(x, \alpha, \beta, A, B, M)}{\partial b_2} &= \frac{-\beta}{2a_2 b_2 \Gamma(b_2)} \exp \left(- \left| \frac{x - \mu_2}{a_2} \right|^{\frac{1}{b_2}} \right) \\ &\left(-\frac{1}{b_2} - \frac{1}{\Gamma(b_2)} \frac{\partial \Gamma(b_2)}{\partial b_2} + \frac{1}{b_2} \left| \frac{x - \mu_2}{a_2} \right|^{\frac{1}{b_2}} \ln \left| \frac{x - \mu_2}{a_2} \right| \right); \end{aligned} \quad (16)$$

$$\begin{aligned} \frac{\partial p_{UGED}(x, \alpha, \beta, A, B, M)}{\partial b_3} &= \frac{-(1 - \alpha - \beta)}{2a_3 b_3 \Gamma(b_3)} \exp \left(- \left| \frac{x - \mu_3}{a_3} \right|^{\frac{1}{b_3}} \right) \\ &\left(-\frac{1}{b_3} - \frac{1}{\Gamma(b_3)} \frac{\partial \Gamma(b_3)}{\partial b_3} + \frac{1}{b_3} \left| \frac{x - \mu_3}{a_3} \right|^{\frac{1}{b_3}} \ln \left| \frac{x - \mu_3}{a_3} \right| \right); \end{aligned} \quad (17)$$

$$\begin{aligned} \frac{\partial p_{UGED}(x, \alpha, \beta, A, B, M)}{\partial \alpha} &= \frac{-1}{2a_1 b_1 \Gamma(b_1)} \exp \left(- \left| \frac{x - \mu_1}{a_1} \right|^{\frac{1}{b_1}} \right) - \\ &\frac{1}{2a_3 b_3 \Gamma(b_3)} \exp \left(- \left| \frac{x - \mu_3}{a_3} \right|^{\frac{1}{b_3}} \right); \end{aligned} \quad (18)$$

$$\frac{\partial \rho_{UGED}(x, \alpha, \beta, A, B, M)}{\partial \beta} = \frac{1}{2a_2 b_2 \Gamma(b_2)} \exp\left(-\left|\frac{x - \mu_2}{a_2}\right|^{\frac{1}{b_2}}\right) + \frac{1}{2a_3 b_3 \Gamma(b_3)} \exp\left(-\left|\frac{x - \mu_3}{a_3}\right|^{\frac{1}{b_3}}\right). \quad (19)$$

Substituting expressions (12 – 19) in (11), we have a nonlinear system of eleven equations, that can be represented in the matrix form:

$$F(\theta) = 0, \quad (20)$$

where θ is a matrix of arguments, and F is a vector of functions.

The Newton's method of iterated local linearization is applied to find a solution of (20). The solution of nonlinear system of equations at i^{th} iteration is:

$$\Theta = \theta_i + \varepsilon_i, \quad (21)$$

where θ_i is a solution at i^{th} iteration; $\varepsilon_i = [\varepsilon_{\alpha}^i, \varepsilon_{\beta}^i, \varepsilon_{a_1}^i, \varepsilon_{a_2}^i, \varepsilon_{a_3}^i, \varepsilon_{b_1}^i, \varepsilon_{b_2}^i, \varepsilon_{b_3}^i, \varepsilon_{\mu_1}^i, \varepsilon_{\mu_2}^i, \varepsilon_{\mu_3}^i]$ is an error matrix.

Substituting (20) into (21), we have:

$$F(\theta) = F(\theta_i + \varepsilon_i) = 0. \quad (22)$$

Function $F(\theta)$ in (22) can be represented in the Taylor series expansion form:

$$F(\theta_i + \varepsilon_i) = F(\theta_i) + W(\theta_i)\varepsilon_i = 0, \quad (23)$$

where $W(\theta_i)$ is a matrix of partial derivatives by estimated parameters:

$$W(\theta_i) = \frac{\partial F(\theta_i)}{\partial \theta}. \quad (24)$$

From (23), we have

$$\varepsilon_i = -F(\theta_i)W^{-1}(\theta_i). \quad (25)$$

Substituting (25) into (21), we have

$$\theta_{i+1} = \theta_i - F(\theta_i)W^{-1}(\theta_i). \quad (26)$$

Matrix θ became closer to maximum of likelihood function in (26) at each iteration. Computation of (26) is iterative until a certain acceptable level of error will be achieved.

Since $F(\theta)$ has a multiple critical points, it is important to localize a search intervals for each of the parameter. Maximal and minimal values of search intervals for TUGED parameters are represented in Table 1. Since positioning errors make the greatest contribution into the measurement results, it is advisable to start search of α from the middle of the general set ($\alpha = 0.5$). We choose starting value for $\beta = 0.1$ much less than α , which characterizes the effect of *FTE*. The initial value for a and b coefficients are chosen from the assumption about the normal error distribution. Also, we use $\mu = 0$, which corresponds to the airplane motion at the preplanned path.

Table 1. The search intervals and starting point for evaluated parameters

Parameter	Minimal value	Maximal Value	Starting point
α	0.1	0.8	0.5
β	0.01	0.5	0.1
a_1, a_2, a_3	1	∞	$\sigma\sqrt{2}$
b_1, b_2, b_3	0.5	1	0.5
μ_1, μ_2, μ_3	$-\infty$	$+\infty$	0

3. Risk of airplane lateral deviation

Risk in aviation safety can be estimated by different approaches. One of them is grounded on evaluation of expected number of incidents or accidents per airplane flying hour (ICAO, 1998). Another approach considers risk as a probability of undesirable event. In this case, risk of airplane lateral deviation from cleared trajectory can be represented as a probability of unplanned airplane deviation out the defined boundaries in horizontal plane. An area navigation requirements specify admissible level of performance for airplane positioning system (ICAO, 2008a), that can be used as a boundary level in risk assessment (Ostroumov & Kuzmenko, 2019).

In case of known PDF of airplane deviation, a risk can be estimated as an area under PDF limited to boundaries of area navigation requirements. A probability area is represented in Figure 2.

Risk of airplane deviation can be estimated by the next formula:

$$R_{LR} = P\{x_L < x < x_R\} = \int_{x_L}^{x_R} \rho_{TUGED}(x, \alpha, \beta, A, B, M) dx, \quad (27)$$

where x_L and x_R are left and right boundaries; $P\{x_L < x < x_R\}$ is a probability of random value x located out of interval $[x_L, x_R]$, R_{LR} is a risk of airplane deviation out of interval $[x_L, x_R]$.

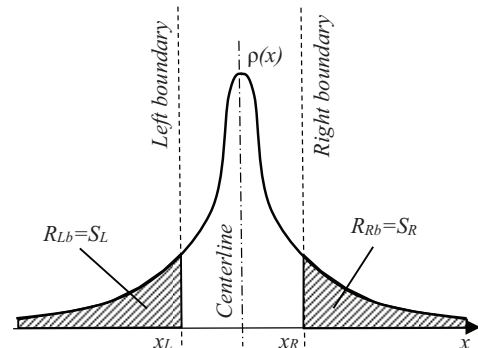


Figure 2. Risk of airplane deviation out of predefined boundaries

4. Risk of airplane lateral deviation during runway operation

Veer off the runway is one of the dangerous events in aviation that can be analyzed in terms of rare events. Let's make a risk level assessment of airplane deviation out the runway by TUGED model (7). As input data, we use statistics of airplane path on the runways of Purdue University Airport (LAF) during take-off and landing phases of flight. The majority of statistical data contains flight data from non-professional pilots with high values of flight technical error, which is a good example of imbalanced levels of FTE and NSE. Airplane navigation is supported by G1000 integrated flight instrument system with GPS receiver supporting Wide Area Augmentation System (WAAS). Statics includes results of airplane coordinates measurements represented as latitude and longitude data with 1 Hz rate. Table 2 shows the basic description of statistical data for 10/28 (6600×150 ft) and 5/23 (4225×100 ft) runways.

A training sample includes 447 unique flights for 5/23 runway with (8229 data length) and only 36 flights for 10/28 runway (only 717 data length). Mean values indicate about left side deviation behavior during runway operations.

Figure 3 shows a histogram of airplane deviations from the runway geometric center using statistical data of both 10/28 and 5/23 runways.

In statistical analysis it makes sense to consider a set of models that support some assumptions about the event. We use an assumption about the airplane motion at the center of the runway that makes $\mu_1 = \mu_2 = \mu_3 = 0$ in a model (7):

$$TUGED_1 = \rho(\alpha, \beta, a_1, a_2, a_3, b_1, b_2, b_3, 0, 0, 0).$$

We consider the case of a mixture of two NDFs and one EDF in TD:

$$TUGED_2 = \rho(\alpha, \beta, a_1, a_2, a_3, 0.5, 0.5, 1, \mu_1, \mu_2, \mu_3).$$

We use a composition of $TUGED_1$ and $TUGED_2$ properties:

$$TUGED_3 = \rho(\alpha, \beta, a_1, a_2, a_3, 0.5, 0.5, 1, 0, 0, 0).$$

Simplified models $TUGED_1$, $TUGED_2$, and $TUGED_3$ help to reduce time of data processing at the fitting stage due to less number of estimated parameters in comparison with $TUGED$.

The results of parameters estimation for the whole training sample (10/28 and 5/23) in the various TUGED options including comparison with DUGED are shown in Table 3. Mean values of each component of PDF are about zero, that indicate identical statistics for left and right sides deviations from the centerline of runway. Performance of statistical data fitting is shown in Table 4.

The validity of the estimated models can be compared by a set of coefficients:

1. Akaike's Information Criteria (AIC). The AIC coefficient indicates the quality of the fitting model by the training sample (Akaike, 1973):

$$AIC = -2 \sum_{i=1}^N \ln(\rho(x_i)) + 2k, \quad (28)$$

where N is a size of training sample; $\rho(x)$ is a fitting model; k is a number of estimated parameters.

2. Bayes Information Criteria (BIC) takes into account the volume of training sample in the form of a fine in distribution parameters evaluation (Schwarz, 1978):

$$BIC = - \sum_{i=1}^N \ln(\rho(x_i)) + 0.5k \ln(n), \quad (29)$$

where x is an observed data; n is a number of data points in x ; k is a number of parameters estimated by the model.

Table 2. Basic parameters of training sample

Runway	Runway width, ft	Amount of sample	Amount of unique flights	Standard deviation, ft	Mean value, ft	Maximal deviation in the left side of runway, ft	Maximal deviation in the right side of runway, ft	Mean data spread (max-min), ft
5/23	100	8229	447	3.15994	-0.31643	-46.602	35.04986	46.60201
10/28	150	717	36	7.038563	-0.86115	-51.8364	19.19464	51.83587
10/28 and 5/23	-	8946	483	3.793449	-0.38248	-51.8364	35.04986	86.88628

Table 3. Results of TUGED parameters estimation

	α	β	μ_1	μ_2	μ_3	a_1	a_2	a_3	b_1	b_2	b_3
TUGED	0.1	0,19	1.2×10^{-3}	1.9×10^{-3}	2.0×10^{-3}	32	6	1	0.65	0.67	0.51
TUGED ₁	0.37	0.01	(0)	(0)	(0)	13	17	1	0.68	0.67	0.61
TUGED ₂	0.1	0.25	$-8.11 \cdot 10^{-5}$	1.3×10^{-3}	2.7×10^{-3}	31	12	1	(0.5)	(0.5)	(1)
TUGED ₃	0.14	0.04	(0)	(0)	(0)	71	10	1	(0.5)	(0.5)	(1)
DUGED	0.74	-	4.9×10^{-3}	-1.1×10^{-4}	-	1	29	-	0.63	0.52	-

Note: values in (*) are predefined by model approximation.

3. Criterion χ^2 (Pearson Criterion). χ^2 coefficient takes into account the divergence of the empirical and theoretical absolute frequencies over the histogram intervals:

$$\chi^2 = \sum_{i=1}^r \frac{(m_i - np_i)^2}{np_i}, \quad (30)$$

where r is a number of histogram intervals; m_i is an absolute frequency at i^{th} interval; p_i is the theoretical probability of getting into the i^{th} interval; n is the total number of experimental data.

4. The sum of the probability difference between the frequency of occurrence in the interval and the probability of occurrence evaluated by fitting function:

$$SPD = \sum_{i=1}^N |v_i - np_i|, \quad (31)$$

where v is a frequency; p is a probability; n is a number of intervals; N is a size of training sample.

The quality of fitting a training sample to the PDF is better when the value of the coefficient is smaller.

Results of risk assessment and performance of approximation by MLM are represented in Table 4 for different models of PDF. Risk of airplane lateral deviation is estimated by (27) for boundaries in half of runway width in 100ft and 150ft according to input data.

Results of using TUGED model gives a better fitting performance in comparison with other models according to AIC, BIC, χ^2 , and SPD criteria. However, TUGED requires more computation time due to fitting complexity.

Results of data fitting by TUGED in comparison with DUGED for different training samples are represented in Figures 4–6.

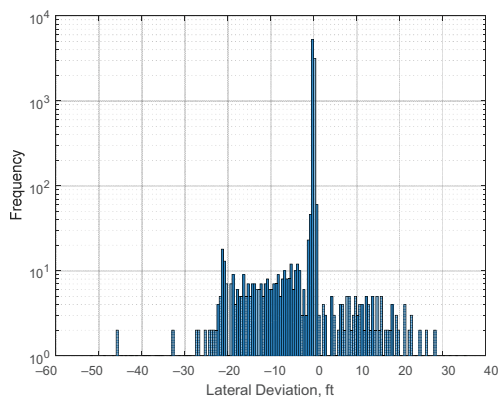


Figure 3. Lateral deviation statistics

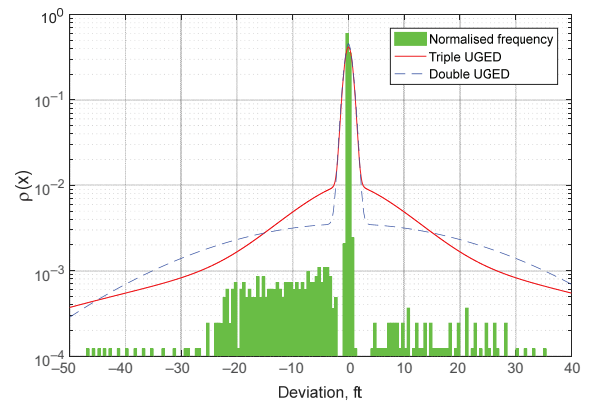


Figure 4. TUGED for 5/23 runway statistics

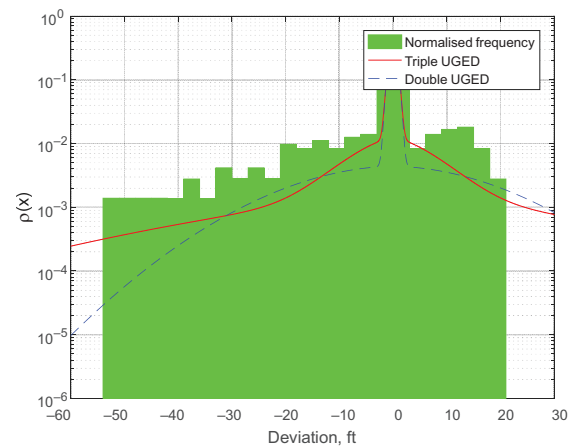


Figure 5. TUGED for 10/28 runway statistics

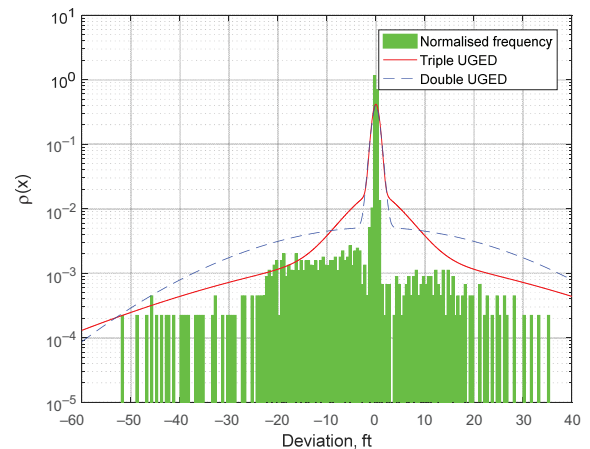


Figure 6. TUGED for the whole statistic sample

Table 4. Performance of approximation and results of risk assessment

	AIC	BIC	χ^2	SPD	Computation time, s	Number of parameters	R_{100}	R_{150}
TUGED	20412	10245	4714	5631	4.13	11	7.2×10^{-3}	1.1×10^{-3}
TUGED ₁	23264	11660	7764	7770	1.08	8	1.0×10^{-4}	5.4×10^{-7}
TUGED ₂	25615	12836	11165	9306	0.38	8	2.3×10^{-3}	6×10^{-5}
TUGED ₃	22266	11151	7063	7123	0.84	5	4×10^{-2}	1.8×10^{-2}
DUGED	20509	10279	5494	6266	0.82	7	4.6×10^{-3}	1.1×10^{-4}

Normalized frequency of airplane deviation approves the assumption of triple component behavior of PDF. Part of core within small deviations from the centerline of runway (± 5 ft) can be a result of airplane positioning error. Deviations within interval ± 30 ft utilize FTE. Fitting with TUGED makes possible to recognize FTE and rare events. Estimation of rare events with TUGED, such as airplane veering off the runway, is more appropriate in comparison with DUGED due to more adequate PDF used in part of rare events.

Conclusions

In the paper, we propose to use a triple probability density distribution function as a model of airplane deviations from preplanned trajectory in the tasks of aviation risk assessment. An airplane deviation is a result of navigation equipment error, flight technical error, and rare events. Affect of these factors is included in the triple probability density distribution function to develop a model adequate to acting errors. Usage of Univariate Generalized Error Distribution allows getting a sufficient level of flexibility from Normal to Double exponential distribution function evolution. Introduction of additional probability density distribution component to airplane deviation model is caused by different degrees of NSE and FTE.

As an example, we use TUGED in the statistical analysis of airplane lateral deviation during runway operations. Training statistical data of airplane deviations contains unbalanced errors caused by a significant difference between NSE and FTE in manual mode. Results of TUGED validation in the statistical analysis of airplane deviations from the center of the runway give the risks of airplane rolling out the runway of LAF airport in levels of 3.7×10^{-5} (runway 5/23) and 1.2×10^{-6} (runway 10/28). Efficiency of TUGED model usage for input data was proved by AIC, BIC, and Pearson criterions

TUGED is a flexible and universal model that can be applied in case of different errors affecting a transport system. Thus, TUGED will give the same result with DUGED in case of the same level of FTE and NSE. TUGED model is recommended to use in different tasks of risk assessment at different flight phases with manual flight control mode.

Also, obtained results can be used for the further development of air traffic safety models and improvement of airplane operation within the runway.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory* (pp. 267–281). Akademiai Kiado.
- Ayabo, A., & Kozubowski, T. (2003). An asymmetric generalization of Gaussian and Laplace laws. *Journal of Probability and Statistical Science*, 1(2) 187–210.
- Cramer, M., & Rodriguez, L. (2013). Analysis of aircraft lateral path tracking accuracy and its implications for separation standards. *Integrated Communications, Navigation and Surveillance Conference, ICNS* (pp. 1–11). Herndon, VA, United States. <https://doi.org/10.1109/ICNSurv.2013.6548526>
- Czyzycki, R. (2013). Using GED (Generalized Error Distribution) for modeling distribution of the rates of return. *International Masaryk Conference for PhD Students and Young Researchers* (pp. 1530–1535). Hradec Králové, The Czech Republic.
- Fala, N., & Marais, K. (2016). Detecting safety events during approach in general aviation operations. *16th AIAA Aviation Technology, Integration, and Operations Conference* (pp. 1–15). Washington, United States. <https://doi.org/10.2514/6.2016-3914>
- Fujita, M. (2013). Iterative bayesian estimation of navigation performance modeled by a mixture of Gaussian and laplace distributions for the application of collision risk modeling. *Transactions of the Japan Society for Aeronautical and Space Sciences*, 56(5), 253–260. <https://doi.org/10.2322/tjsass.56.253>
- Giller, G. (2013). A Generalized Error Distribution. SSRN (August 16, 2005). <https://doi.org/10.2139/ssrn.2265027>
- ICAO. (1988). *Review of the General Concept of Separation Panel (RGSP)* (Doc 9536). ICAO.
- ICAO. (1998). *Manual on airspace planning methodology for the determination of separation minima* (Doc 9689, AN/953). ICAO.
- ICAO. (2006). *Procedures for Air Navigation Services. Aircraft Operation. Construction of Visual and Instrumental Flight Procedure*, Vol. 2, (Doc 8168, OPS/611). ICAO.
- ICAO. (2008a). *Performance-based Navigation (PBN)* (Doc 96131) Manual. ICAO.
- ICAO. (2008b). *Unified framework for collision risk modelling in support of the manual on airspace planning methodology with further applications* (Doc 9689, 181 ed.). International Civil Aviation Organization.
- Klein, V. (1980). Maximum likelihood method for estimating airplane stability and control parameters from flight data in frequency domain. *NASA Technical Paper 1637* (61 p.).
- Kutsenko, O., Ilnytska, S., & Konin, V. (2018). Investigation of the residual tropospheric error influence on the coordinate determination accuracy in a satellite landing system. *Aviation*, 22(4), 156–165. <https://doi.org/10.3846/aviation.2018.7082>
- Kuzmenko, N., & Ostroumov, I. (2018). Performance analysis of positioning system by navigational aids in three dimensional space. *2018 IEEE 1st International Conference on System Analysis and Intelligent Computing, SAIC 2018 – Proceedings* (pp. 101–104). IEEE. <https://doi.org/10.1109/SAIC.2018.8516790>
- Markovsky, I., & Van Huffel, S. (2007). Overview of total least-squares methods. *Signal Processing*, 87(10), 2283–2302. <https://doi.org/10.1016/j.sigpro.2007.04.004>
- Mitici, M., & Blom, H. (2019). Mathematical models for air traffic conflict and collision probability estimation. *IEEE Transactions on Intelligent Transportation Systems*, 20(3), 1052–1068. <https://doi.org/10.1109/TITS.2018.2839344>
- Mori, R. (2011). Identifying the ratio of aircraft applying SLOP by statistical modeling of lateral deviation. *Transactions of the Japan Society for Aeronautical and Space Sciences*, 54(183), 30–36. <https://doi.org/10.2322/tjsass.54.30>
- Nagaoka, S. (2008). A model for estimating the lateral overlap probability of aircraft with RNP alerting capability in parallel RNAV routes. *ICAS Secretariat – 26th Congress of International Council of the Aeronautical Sciences 2008, ICAS 2008, 1* (pp. 3590–3597). Anchorage, AK, United States.

- Naidu, V., & Durgarao, S. (2012). Estimation of aircraft height and lateral deviation with respect to runway using images from un-calibrated camera. *National Conference on Signals and Image Processing*, 1, 81–185.
- Ostroumov, I., & Kuzmenko, N. (2018). Compatibility analysis of multi signal processing in apnt with current navigation infrastructure. *Telecommunications and Radio Engineering* (English translation of *Elektrosvyaz and Radiotekhnika*), 77(3). <https://doi.org/10.1615/TelecomRadEng.v77.i3.30>
- Ostroumov, I., Kharchenko V., & Kuzmenko, N. (2019). An airspace analysis according to area navigation requirements. *Aviation*, 23(3), 36–42. <https://doi.org/10.3846/aviation.2019.10302>
- Ostroumov, I. V., & Kuzmenko, N. S. (2019). Risk analysis of positioning by navigational aids. In *Signal Processing Symposium: SPSympo-2019, International Conference of IEEE* (pp. 92–95). Krakow, Poland. <https://doi.org/10.1109/SPS.2019.8882003>
- Ryu, E., & Young, S. (2016). General aviation runway design evaluation based on aircraft deviations from runway centerline. *7th International Conference on Research in Air Transportation* (pp. 1–4). Drexel University.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>

Notations

Variables and functions

- A – matrix of a scale factors;
 a – scale factor;
 B – matrix of shape coefficients;
 b – shape coefficient;
 F – vector of functions;
 $f(x_i, \theta)$ – probability that x variable takes a value of x_i ; $i = [1, n]$;
 k – number of estimated parameters;
 $L(x_i, \theta)$ – likelihood function;
 M – matrix of mean values;
 m_i – absolute frequency at i^{th} interval;
 n – number of measurements;
 N – size of training sample;
 $P\{x_R < x < x_L\}$ – probability of random value x located out of interval $[x_L, x_R]$;
 p_i – theoretical probability of getting into the i^{th} interval;
 r – number of histogram intervals;
 R_{LR} – risk of airplane deviation out of interval $[x_L, x_R]$;
 $W(q_i)$ – matrix of partial derivatives by estimated parameters;
 x – airplane deviation from a pre-planned trajectory;
 X – discrete random variable;

- x_L – left boundary of deviation;
 x_R – right boundaries of deviation;
 $\Gamma(b)$ – Euler-gamma function;
 α – weight coefficients of $r_{NSE}(x)$ in TUGED;
 β – weight coefficients of $\rho_{FTE}(x)$ in TUGED;
 e_i – error matrix;
 γ – probability of a gross navigational error;
 λ – rate parameter;
 μ – mean value;
 θ – matrix of TUGED parameters;
 $\rho(x)$ – probability density distribution function;
 $\rho_{FTE}(x)$ – PDF characterizing the FTE;
 $r_{NSE}(x)$ – PDF utilizing the errors of navigation system;
 $r_T(x)$ – PDF characterizing the appearance of rare events;
 $r_{UGED}(x)$ – Univariate Generalized Error Distribution function;
 σ – standard deviation;
 υ – frequency.

Abbreviations

- AIC – Akaike’s Information Criteria;
 APR – Approach;
 BIC – Bayes Information Criteria;
 CD – Core Distribution;
 DEDF – Double Exponential Density Function;
 DUGED – Double Univariate Generalized Error Distribution;
 EDF – Exponential Density Function;
 FTE – Flight Technical Error;
 GBAS – Ground-based Augmentation System
 GNSS – Global Navigation Satellite System;
 GPS – Global Positioning System;
 MD – Mixed Distribution;
 MLM – Maximum Likelihood Method;
 NDF – Normal Density Function;
 NSE – Navigation System Error;
 PDE – Path Definition Error;
 PDF – Probability Density Function;
 SBAS – Satellite-based Augmentation System;
 SID – Standard Instrument Departure;
 SPD – Sum of the Probability Difference;
 STAR – Standard Terminal Arrival;
 TD – Tail Distribution;
 TSE – Total System Error;
 TUGED – Triple Univariate Generalized Error Distribution;
 UGED – Univariate Generalized Error Distribution;
 WAAS – Wide Area Augmentation System.