33RD CONGRESS
OF THE INTERNATIONAL COUNCIL
OF THE AERONAUTICAL SCIENCES
STOCKHOLM, SWEDEN, 4-9 SEPTEMBER, 2022

ICAS 2022 SWEDEN

STATISTICAL DATA SELECTION FOR BETTER FLIGHT CHARACTERISTIC MODELING

Masaru Naruoka[1] & Tetsujiro Ninomiya[1]

[1]Japan Aerospace Exploration Agency

Abstract

To improve the estimation reliability of flight characteristics of aircraft by using flight test data, it is essential to reduce discrepancy between input data and output model. In this paper, the authors propose to use Student's t-distribution in cooperation with Bayesian estimation framework. This method provides the estimated degree of freedom ν of the t-distribution in addition to the model parameter distribution, and we can utilize ν as the metric of the discrepancy. The effectiveness of the method is demonstrated with the modeling of the lift coefficient with actual flight data. The contribution to the lift of the stabilizer deflection, pitch rate, and Reynold's number effect are examined by using the proposed method. Consequently, the pitch rate and Reynold's number effects are successfully extracted from the flight data by investigating the behavior of the estimated ν.

Keywords: Flight characteristics, Bayesian inference, Markov-chain Monte-Carlo, Student's t-distribution

1. Introduction

Estimation of flight characteristics of aircraft with flight data has been actively performed for various applications. The flight characteristics are represented by aerodynamic parameters and stability derivatives. The former such as lift and drag coefficients are required for validation of aircraft design and manufacturing. The latter are essential to build a training simulator.

Typically, the estimation procedure is divided into three steps. First, a model to represent the characteristics is postulated based on mechanical and empirical knowledge. For instance of the lift coefficient modeling, one of the model postulation is that the coefficient is linear to the angle of attack in major cruise conditions. Then, parameters of the model are estimated to fit with the selected data acquired in flight tests. For the example of the lift coefficient, performing flight in which the angle of attack is swept gives a part of the suitable data. Finally, the model with the determined parameters is evaluated its effectiveness with certain criteria such as fitness to other different data from the data used in the previous step. In these steps, the second step has actively been studied, and many methods represented by least square or Kalman filter were developed [1]. The authors also studied a method to enhance the estimation not only for parameter itself but also for its distribution by using a Markov-chain Monte-Carlo (MCMC) sampler, which is one of Bayesian inference methods [2].

Although an estimation method is an important factor to get reasonable and reliable results, the data used for the estimation in the second step is focused in this study. This is because not only an estimation method but also selected data affect the estimation results.When we select inappropriate data which does not include any hint related to the postulated model, the estimation fails even if we use a sophisticated estimation method. Thus, the relationship between the model and data must be considered. This relationship is well-known in the field of general system identification problems. In order to avoid difficult situation, data including various modes of a target system is intentionally utilized as much as possible for those problems. For instance, a M-series random input signal is utilized to excite the target system. However, for the flight characteristic estimation, the arbitrary excitation of aircraft dynamics is restricted due to safety. Therefore, the data selection for the flight characteristic estimation should be a more important topic compared to the general modeling problems.

Actually, estimation failure resulted from a combination of a model and dataset was reproduced with the datasets and models of the authors' previous study [2]. The study was conducted to model the lift coefficient of a fixed wing aircraft. There were two postulated model with and without the contribution of the pitch rate of the body, and there were two dataset with and without pitching motion. The results showed that the combination of the pitch rate considered model and the data including the pitching motion achieved the best estimation which had the steepest distribution of the estimated parameters. On the other hand, the applications of the two datasets to the model without the pitch rate contribution resulted in the comparable moderate distribution. In other words, the combination of the dataset having more information than the range of the postulated model did not work well.
Therefore, in order to gain the robustness to input data, the authors propose utilization of a Student's t-distribution in cooperation with the MCMC method for the flight characteristic estimation. The Student's t-distribution has a heavier tail distribution compared to a normal distribution. How this feature contributes the estimation is discussed in Sec. 2 with brief explanation of the MCMC method. In addition, preliminary effectiveness of the proposed method is demonstrated with simple numerical examples. Moreover, degree of freedom (DOF) of the t-distribution estimated in the proposed method is focused as a metric to measure match between data and model. Then, as a practical problem, lift coefficient modeling of a fixed-wing aircraft by using the proposed method is described in Sec. 3. Followed by introduction of the target aircraft, input data for the modeling is explained. There are four postulated models applied to the data, and the performance of each estimation result is discussed mainly based on the estimated t-distribution DOF. Finally, this study is concluded in Sec. 4.

2. Bayesian inference with Student's t-distribution

In this section, the proposed method used in this study is described in accordance with a framework of the Bayesian inference. First, the MCMC sampling, well summarized in Chapter.11 of Bishop [3], is briefly explained. The MCMC method requires a statistical model to estimate a distribution, and the authors proposes to utilize the Student's t-distribution instead of the typical normal distribution. Thus, following to the MCMC explanation, the Student's t-distribution and its application to the Bayesian inference are elaborated. At the end of this section, the simple numerical examples demonstrate the preliminary effectiveness of the proposed method.

2.1 Markov-chain Monte-Carlo (MCMC) method

The Bayesian inference including the MCMC method provides a posterior, i.e., estimated distribution of parameters $P\left(\underline{\theta}|\underline{y}\right)$ by using the relationship

$$P\left(\underline{\theta}|\underline{y}\right) \sim P\left(\underline{y}|\underline{\theta}\right) P\left(\underline{\theta}\right) , \tag{1}$$

where $\underline{y}$ and $\underline{\theta}$ are the data and parameters, respectively. $P(\cdot)$ represents a probability distribution. The goal $P\left(\underline{\theta}|\underline{y}\right)$ corresponds to a conditional probability distribution of $\underline{\theta}$ in occurrence of $\underline{y}$. $P\left(\underline{\theta}\right)$ is a prior distribution, to which a non-informative prior is typically applied. $P\left(\underline{y}|\underline{\theta}\right)$ is given by a likelihood function to be postulated.
The MCMC method generates $\underline{\theta}$ samples many times like its name includes "Monte-Carlo". This generation is iteratively performed with adjustment of $\underline{\theta}$ based on the calculated likelihood of the right side in Eq. (1). "Markov-chain" indicates this iterative process. In a successful case, the generated $\underline{\theta}$ finally follows to a certain probability distribution, which sufficiently approximates the goal $P\left(\underline{\theta}|\underline{y}\right)$ distribution.
In this study, Stan [4], a sophisticated MCMC sampling software, is utilized as the authors' previous study. Stan has many customization points, and the tuned points for this study are summarized. For each estimation, 2000 iterations are performed to draw the parameter samples from four independent generators (chains). The first 1000 iterations are eliminated from the distribution estimation as "burnin", that is, the startup phase to wait for the parameter distribution being converged. In addition, the thinning parameter is three: every two of three iteration results are excluded for the estimation in order to mitigate autocorrelation. To summarize the numbers, 1336 samples $(= (2000 - 1000) \times 4/3)$ of each estimated parameter are generated par one estimation trial.

The convergence of the distribution of the generated samples must be checked before further analysis. In this study, two methods are mainly used. The first method is visual check of a trace plot which shows history of the samples in the iterations. If the sample is sufficiently random, this check passes. The second method is to use the $\hat{R}$, which is a convergence index calculated for each parameter with variances in a chain and between chains. The convergence is determined when $\hat{R} < 1.1$ according to Chapter 11 of Gelman [5].

2.2 Student's t-distribution

The Student's t distribution, which is utilized to determine fitness of structure of a postulated model to input data in this study, will be explained briefly. The probability density of the distribution is shown in Fig. 1, and its function is described as

$$\text{Student\_t}\left(y|\nu,\mu,\sigma\right) = \frac{\Gamma\left(\left(\nu+1\right)/2\right)}{\Gamma\left(\nu/2\right)\sqrt{\pi\nu}\sigma}\left(1+\frac{1}{\nu}\left(\frac{y-\mu}{\sigma}\right)^2\right)^{\frac{-\nu+1}{2}}, \tag{2}$$

where Γ is the gamma function. ν is the number of degrees of freedom (DOF) of the distribution, and takes a value greater than zero. It regulates the heaviness of the tail distribution, and its effect will be shown later with comparisons of related distribution. μ and σ are the parameters to determine a most probable location and steepness of the distribution, respectively. It is not a rigorous explanation but μ and σ represent mean and standard deviation of the distribution, which are only defined where ν is in a specific range.

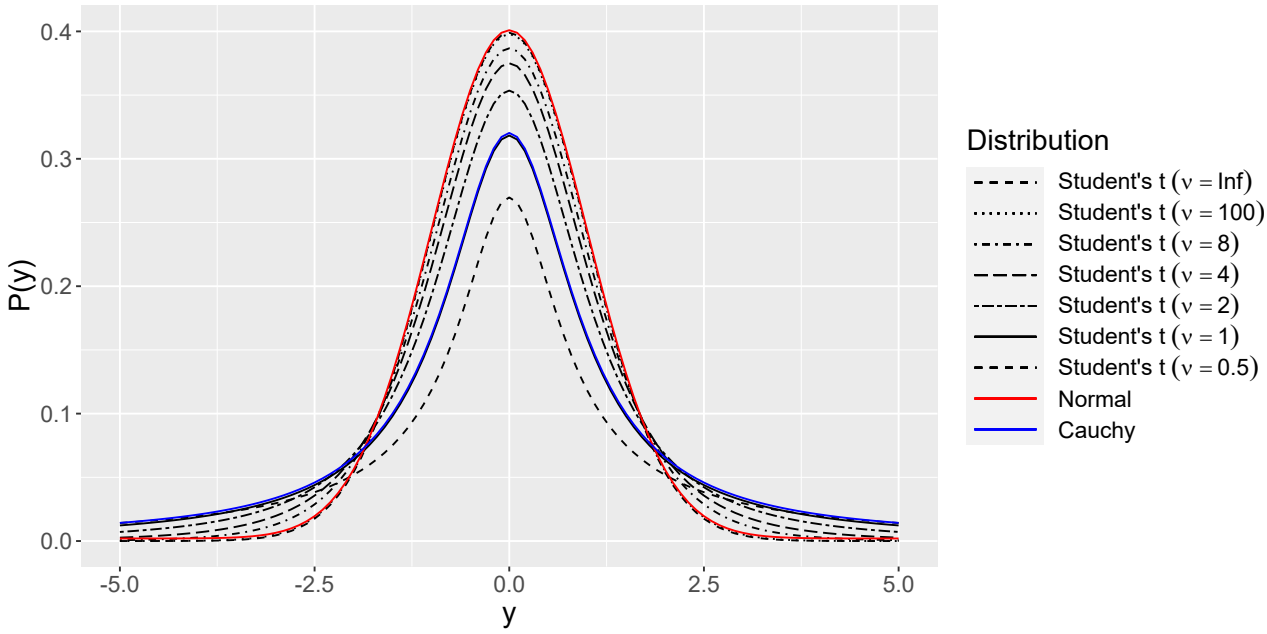


Figure 1 – Probability density of Student's t-, normal and Cauchy distributions

The Student's t-distribution looks like a normal distribution as shown in Fig. 1. Actually, when ν equals to $+\infty$, the distribution is identical to a normal distribution defined as

$$\text{Normal}\left(y|\mu,\sigma\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right). \tag{3}$$

In addition, a $\nu = 1$ t-distribution is a Cauchy distribution of

$$\text{Cauchy}\left(y|\mu,\sigma\right) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\left(y-\mu\right)/\sigma\right)^2}, \tag{4}$$

which has heavier tail than the normal distribution. Table 1 shows comparisons of these three distributions having same $\mu = 0$ and $\sigma = 1$ with various ν for the t-distribution. The values in the table are

the cumulative probabilities in the condition of y is greater than 1, 2, and 3-σs. As the values shows, the Cauchy distribution has the larger probability than the normal distribution. The probabilities of the t-distribution with $\nu = +\infty$ and 1 are the same values as the normal and Cauchy distributions, respectively. Furthermore, the probabilities of the lower ν take the higher values. This functionality of ν to vary the tail distribution heaviness is utilized in this study.

Table 1 – Cumulative probability of normal, Cauchy, and Student's t distributions

Distribution		Cumulative probability		
		$\|y\| \geq 1\sigma$	$\|y\| \geq 2\sigma$	$\|y\| \geq 3\sigma$
Normal		0.3173	0.0455	0.0027
Cauchy		0.5000	0.2952	0.2048
Student's t	$\nu = +\inf$	0.3173	0.0455	0.0027
	$\nu = 100$	0.3197	0.0482	0.0034
	$\nu = 8$	0.3466	0.0805	0.0171
	$\nu = 4$	0.3739	0.1161	0.0399
	$\nu = 2$	0.4226	0.1835	0.0955
	$\nu = 1$	0.5000	0.2952	0.2048
	$\nu = 0.5$	0.6022	0.4455	0.3673

2.3 Application of Student's t-distribution in a Bayesian inference framework

As mentioned, the likelihood function is required to perform estimation with the MCMC method. The likelihood function will be configured by considering the distribution of residual, which is the difference between the observed data and the model outputs. For the flight characteristic estimation, it is typical that the residual is assumed to be approximated by a normal distribution as well as a general parameter estimation problem. This assumption is reasonable, because many traditional estimation methods represented by least square and Kalman filter used the same assumption implicitly or explicitly. The assumption is formulated with a likelihood function as

$$y \sim \text{Normal}\left(f\left(\underline{x}, \underline{\theta}\right), \sigma\right), \tag{5}$$

where f is a model whose arguments are state variables $\underline{x}$, and model parameters $\underline{\theta}$. y is a system output, and σ is residual variance. While $\underline{x}$ and y are observed, $\underline{\theta}$ and σ are estimated.
The proposal of this study is to alter the normal distribution to the Student's t-distribution as

$$y \sim \text{Student\_t}\left(\nu, f\left(\underline{x}, \underline{\theta}\right), \sigma\right), \tag{6}$$

where in addition to $\underline{\theta}$ and σ, DOF ν is also estimated. There are two intended points for this change. Firstly, thanks to the heavier tail of the Student's t-distribution, the estimation will get robustness to outliers in observed data. This is because different from the normal distribution, the t-distribution makes the adjustment of the model parameters insensitive to the outliers, which are located sufficiently far from space where major samples exist. Secondly, the estimated ν will be used as the metric for the better estimation. As far as the true error belongs to the normal distribution, higher ν means more fitness of model outputs to data. In other words, a low ν notifies us that the data may contain disturbing data from the aspect of the postulated model. In that case, one idea is to segment the data and try to achieve higher ν in each segmented estimation. As the consequence, we can improve the model performance comprehensively because the estimated parameters in the segmented estimation have the steeper distribution compared to the unified estimation.

2.4 Numerical examples

The effectiveness to use the t-distribution will be demonstrated by using simple numerical examples before the application to the flight characteristic modeling. The example problem is to estimate a parent distribution with an observed dataset consisting of 1000 samples. The samples are generated by two "true" and "disturbing" distributions, and their portion varies in a different set of the samples.

The true distribution, which is the true goal to be estimated, belongs to a normal distribution, and its mean and standard deviation are configured as zero and one, respectively. On the other hand, the disturbing distribution is a uniform distribution whose samples are configured to populate in a range from -5 to 5. For example, Fig 2 shows a sample set whose 30% are outliers.

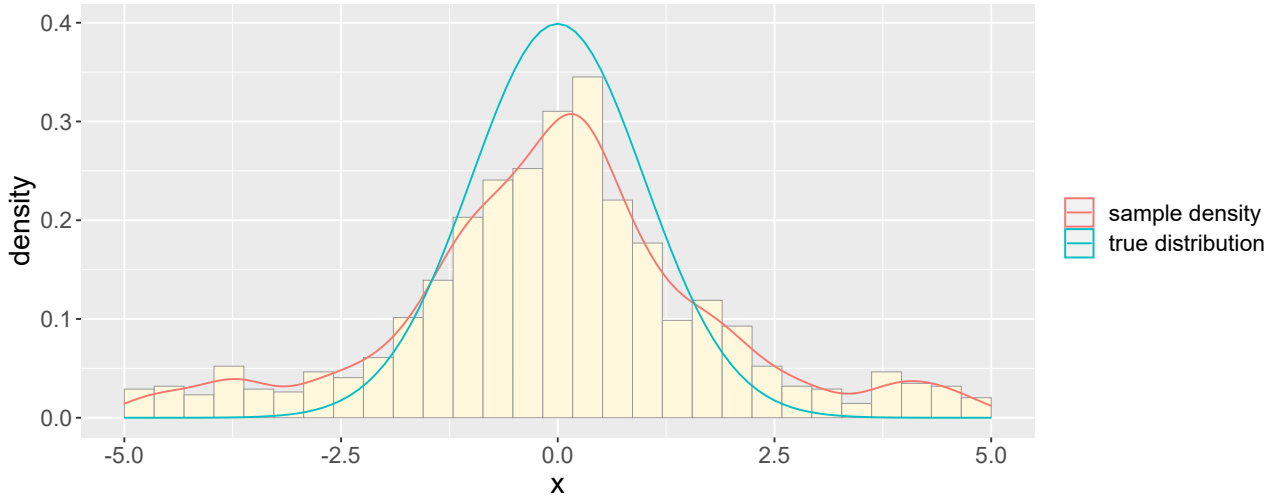


Figure 2 – Sample set occupied with outliers by 30%

The estimation is performed with two postulated models consisting of normal and Student's t-distribution. Table 2 summarizes the results of the estimation. In the table, the mean μ and deviation σ of the estimated distribution are represented by their mean in partition with their standard deviation, while the estimated DOF ν for the Student's t-distribution is represented by the median. Note that the convergence of the estimated parameter distribution has been confirmed with their trace plot and $\hat{R}$ both of which are not included in the paper. In terms of the estimated mean, irrespective of both the postulated models and ratios of outliers, the mean, which is close to the true value, i.e., zero, is well estimated. This is natural because the mean of both the true and disturbing distributions is zero. On the other hand, the accuracy of the deviation of the estimated distribution goes worse as the outlier ratio increases. In addition, this phenomena is severer in the normal distribution compared to the Student's t-distribution. Figure 3a indicating the density of the deviation also highlights the phenomena. The estimation by using the Student's t-distribution is located near from the true value even if the number of the outliers increases. Thus, it concludes that to use the Student's t-distribution as the postulated statistical model relaxes the requirement on the input data. In addition, the estimated ν gives an important queue for the better estimation, because the ν decreases as the ratio of outliers increases. This means we can perform the model adjustment and/or the input data selection so that the estimated ν is increased or at least retained. Moreover, according to Fig. 3b showing the density of the estimated ν, the ν rapidly changes in a range of 0% to 10% of the ratio of outliers. This is the preferable characteristic for fine modeling to include rare cases with low probability of occurrence.

3. Lift coefficient modeling of Cessna 680

In this section, the modeling of the flight characteristic by using the proposed statistical estimation method with the postulation of the Student's t-distribution is described. The modeling target is the lift coefficient C_L of the fixed wing aircraft, and with the utilization of the t-distribution demonstrated in the previous section, the modeling is refined. In the followings, the target aircraft and the flight data used as the inputs are explained firstly. Then, the C_L modeling is elaborated in four steps; base modeling followed by additional consideration of stabilizer deflection, pitch rate and Reynold's number effect.

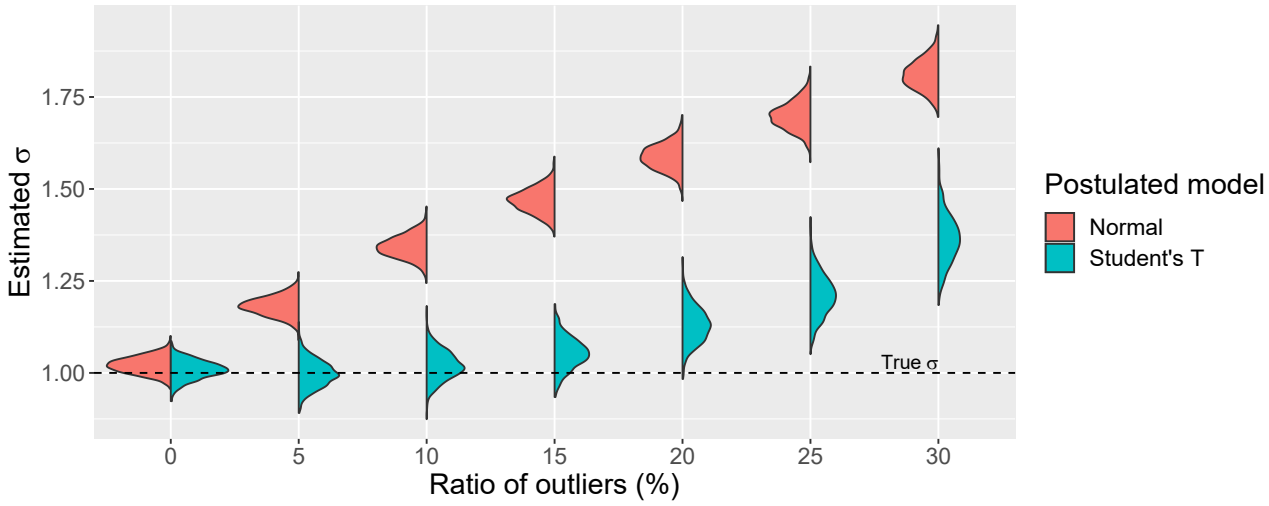
3.1 Target aircraft

The target aircraft is the Cessna 680 Citation Sovereign owned by JAXA shown in Fig. 4. It is twin jet-powered fixed-wing aircraft, and modified for various experimental purpose [6]. In terms of the modification items related to this study, the electrical signals of the aircraft are monitored to record

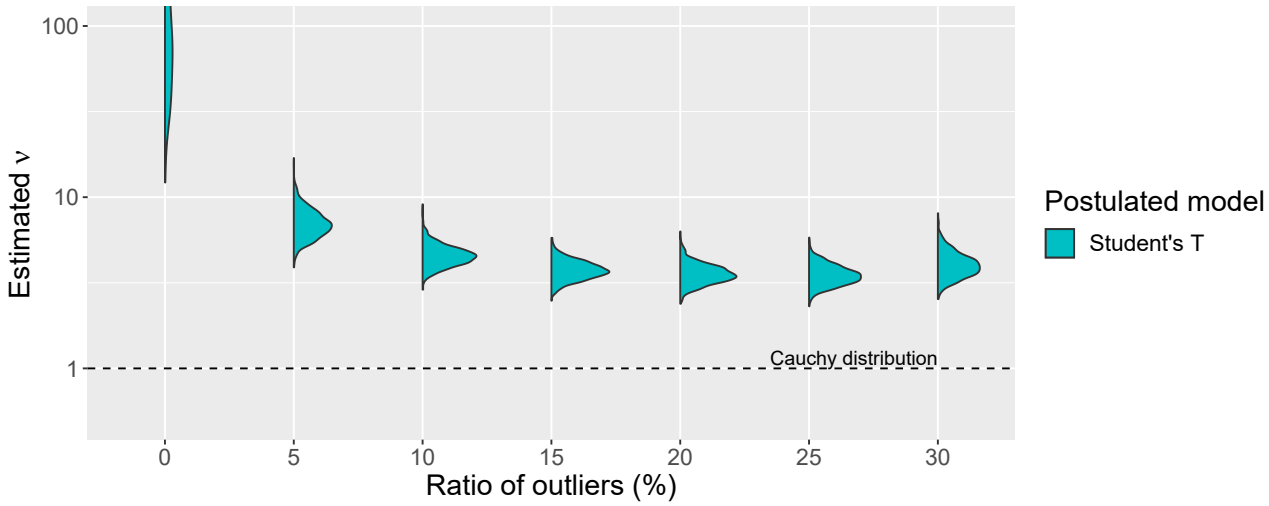
Table 2 – Summary of normal distribution estimation with outliers

Ratio of outliers	Postulated models and results (mean $\pm 1\sigma$ of estimated distribution)				
	Normal		Student's t		
	Mean (μ)	Deviation (σ)	Mean (μ)	Deviation (σ)	DOF (ν) [*1]
0%	-0.011 ± 0.033	1.021 ± 0.023	-0.010 ± 0.031	1.008 ± 0.026	87.119
5%	0.010 ± 0.037	1.182 ± 0.026	-0.012 ± 0.035	0.998 ± 0.037	6.895
10%	-0.011 ± 0.042	1.343 ± 0.031	-0.025 ± 0.037	1.020 ± 0.039	4.506
15%	-0.039 ± 0.046	1.469 ± 0.033	-0.045 ± 0.040	1.053 ± 0.043	3.716
20%	-0.038 ± 0.050	1.587 ± 0.035	-0.036 ± 0.043	1.130 ± 0.050	3.527
25%	-0.012 ± 0.054	1.699 ± 0.038	-0.016 ± 0.045	1.211 ± 0.056	3.477
30%	-0.015 ± 0.058	1.810 ± 0.042	-0.013 ± 0.053	1.365 ± 0.068	3.970

[*1] Instead of mean and standard deviation, median is displayed.



(a) standard deviation σ



(b) DOF ν

Figure 3 – Estimated density

the aircraft state consisting of not only flight conditions but also internal commands. In addition, as the figure shows, the boom is equipped on the nose of the body. It has pitot holes and vanes on the tip to acquire air data information such as airspeed and angle of attack without flow disturbance generated by the body. Furthermore, the aircraft weight in flight is accurately estimated due to precise acquisition of fuel consumption [7]. The thrust is also able to be estimated in a sufficient accuracy with the dedicated method.

Figure 4 – Target aircraft named “Hisho”, which means fly higher in Japanese

3.2 Input data

The input data for the flight characteristic estimation is the lift coefficient C_L and its related variables generated by preprocessing the flight data obtained with the target aircraft. The flight data was obtained in the summers of 2016, 2018 and 2020, and its total flight numbers and length were 22 and 84.7 hours, which were approximate twice as much as 10 and 43.8 hours of the previous study. The data is decimated in order to reduce the computation time of the MCMC estimation to a practical length. Figure 5 indicates the airspeed and altitude histories recorded in the data. The figure shows that the data sufficiently covers the flight envelop of the target aircraft shown in its dotted lines.

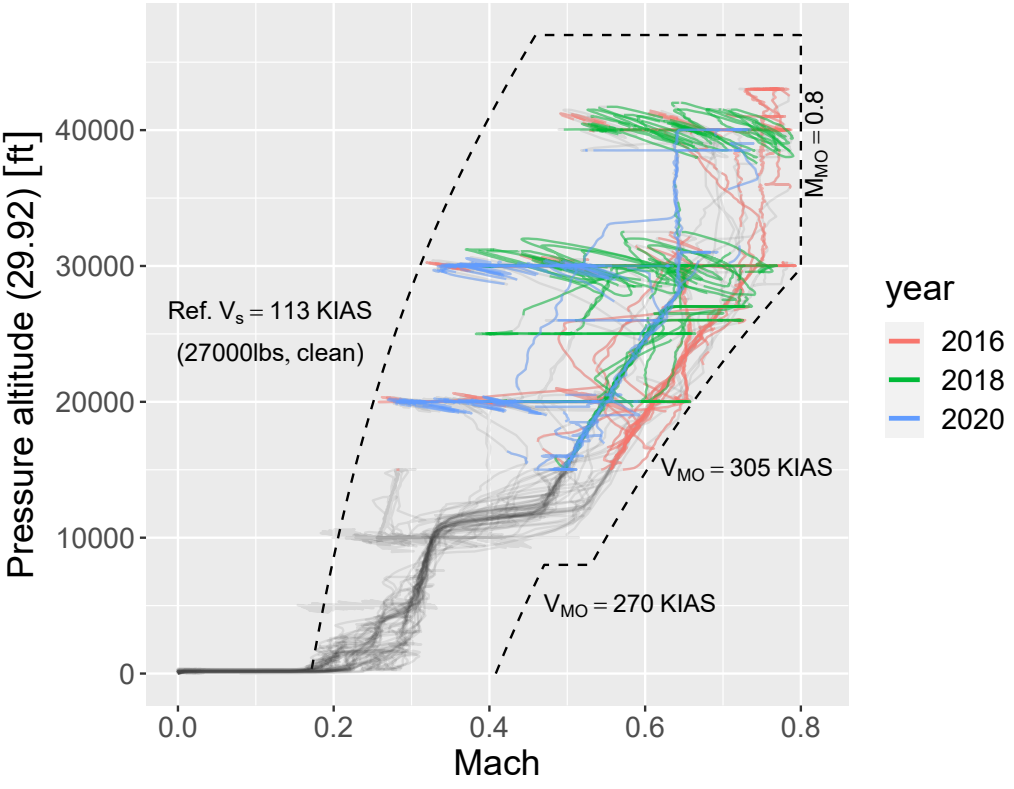


Figure 5 – Flight data

Before the conversion to C_L, the flight data is refined for simplicity. The data on the grey lines in Fig. 5 is removed by the selection, and its length is 31.9 hours corresponding to 12,021 sample points.

The following conditions are firstly used for the data selection; the altitude is greater than 15,000 ft, the gross thrust is greater than 1 kN, the rolling is up to 2 degrees, and spoilers and flaps are not activated. Then, the segmented data whose duration is less than 10 seconds is removed. It is noted that the maximum takeoff thrust and the ceiling altitude of the target aircraft are 25.66 kN and 47,000 ft.
Finally, the filtered flight data is transformed to C_L by using the relation of the equation of motion like

$$C_L = \frac{1}{qS}\left[\mathrm{Rot}\left(\alpha,\beta\right)\left(m\vec{a}-\vec{T}\right)\right]_z, \tag{7}$$

where q and $\vec{a}$ are the dynamic pressure and accelerometer outputs, respectively, both of which are observable. $\mathrm{Rot}\left(\alpha,\beta\right)$ is the matrix to rotate a vector from the body axis to the wind axis through the observed angle of attack α and sideslip angle β. m and $\vec{T}$ are the estimated weight and thrust, and S is the main wing area.
As the related variables of C_L, the angle of attack α, elevator deflection δ_e, stabilizer deflection δ_{stb}, pitch rate q, and Reynold's number Re are selected in this study. Figure 6 is called a pair plot which shows the frequencies and relationship of the C_L and related variables. Note that the absolute values of C_L are intentionally removed. The diagonal elements of the figure are the histograms. The lower triangle elements are the scatter plots of each item pair. The upper triangle elements indicate the scaled Spearman's rank correlation coefficients, which are one of the index values how a pair is linearly correlated. The -100 and 100 of the coefficients represent fully positive and negative correlations.

3.3 Base model with consideration of angle of attack α and elevator deflection δ_e

For the base modeling of C_L, the contribution of the angle of attack α and the elevator deflection δ_e are taken into account. This modeling postulation is the same as the authors' previous study except for the assumed distribution of the normal distribution. At this time, the Student's t-distribution is postulated and the model equation is

$$C_L \sim \mathrm{Student\_t}\left(\nu, C_{L0}+C_{L\alpha}\alpha+C_{L\delta_e}\delta_e, \sigma\right), \tag{8}$$

where C_{L0} and $C_{L\alpha}$ are the estimated zero angle lift coefficient and lift slope. $C_{L\delta_e}$ represents the contribution derived from the elevator deflection. The estimated parameters are summarized in Table 3 with the estimated ν. The results obtained with other postulations mentioned later are also described in the same table. As the representative statistical values of the estimated parameters, the mean, standard deviation, 2.5 and 97.5 quantiles are listed In the table. In addition, all the $\hat{R}$ of the table show that the convergence condition $\hat{R} \leq 1.1$ is satisfied for the all parameters. Note that the estimation convergence is additionally checked with the trace plots not shown here. Another note that instead of the absolute values of the estimated C_{L0}, the differences from the estimated C_{L0} mean are provided in the ΔC_{L0} rows.
In the followings, the means of σ and ν are commonly discussed later, and at this time, they are 1.585×10^{-2} and 4.594, respectively. The preferable σ and ν are smaller one and larger one as discussed in Sec. 2.3. They correspond to the steeper estimated distribution and the smaller number of disturbing samples. For reference, the minimum mean of the σ estimated with the normal distribution postulation was 1.762×10^{-2} in the authors' previous study. The difference is mainly derived from the robustness of the t-distribution to the outliers.

3.4 Stabilizer effect $C_{L\delta_{\mathrm{stab}}}$ consideration

The second model is to consider the effect of the stabilizer deflection as

$$C_L \sim \mathrm{Student\_t}\left(\nu, C_{L0}+C_{L\alpha}\alpha+C_{L\delta_e}\delta_e+C_{L\delta_{\mathrm{stab}}}\delta_{\mathrm{stab}}, \sigma\right), \tag{9}$$

where $C_{L\delta_{\mathrm{stab}}}$ is the added parameter. The stabilizer contribution is anticipated to be hard to model. This is because the stabilizer movement δ_{stab} is the strong linear to the angle of attack α changes shown in Fig. 6, in which their Spearman's rank correlation coefficient is -97. However, the further investigation is conducted. Figure 7 shows the binned scatter plot of δ_{stab} and α. The color of each

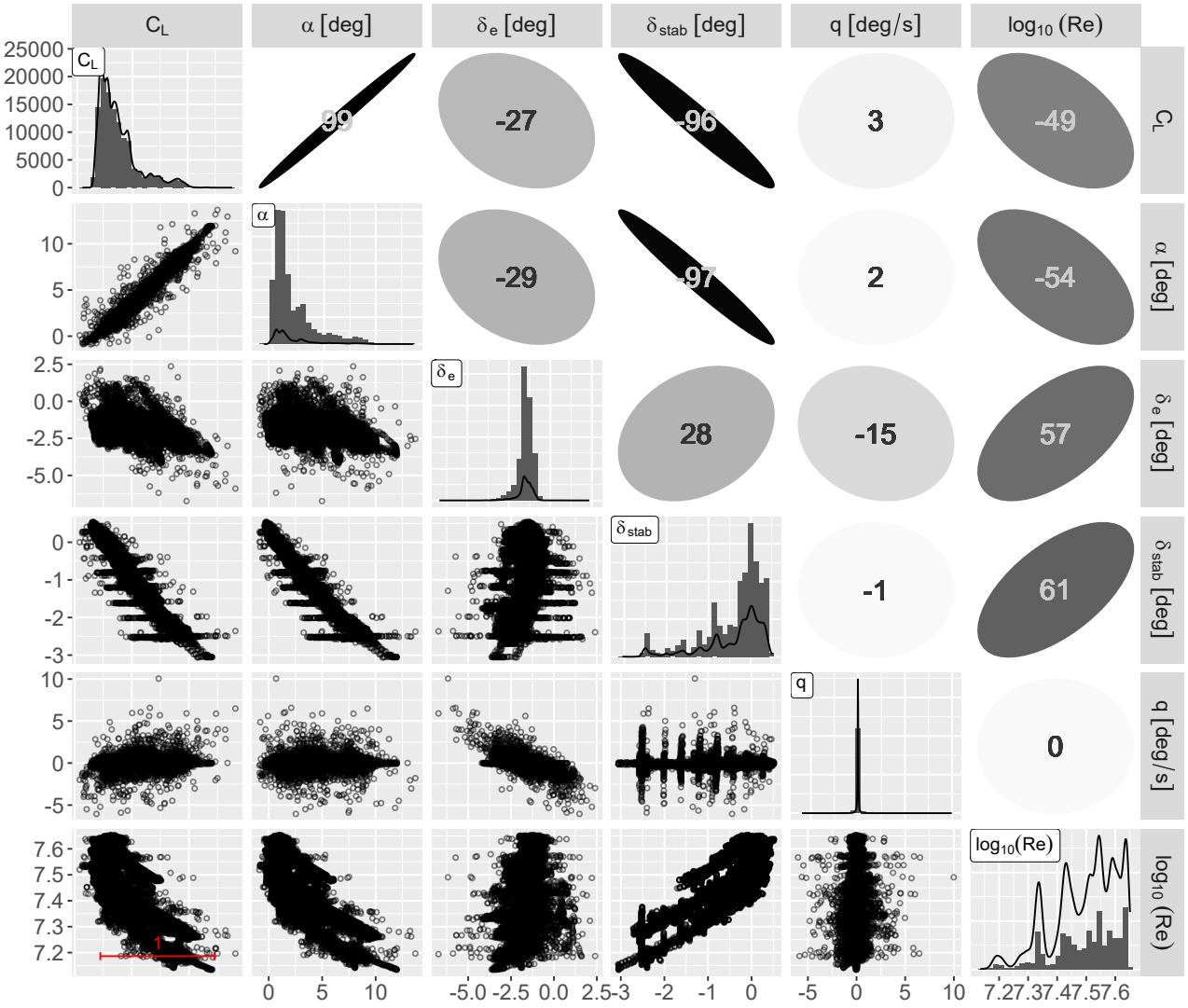


Figure 6 – Pair plot of input data

Table 3 – Estimation results of base model and those of stabilizer and pitch rate effect consideration

Postulated Model	Item	Mean	Standard deviation	2.5%	97.5%	$\hat{R}$
Base model Eq. (8)	ΔC_{L0}	(0.000)	7.489×10^{-4}	(-1.433×10^{-3})	(1.433×10^{-3})	0.999
	$C_{L\alpha}$ *1	8.893×10^{-2}	1.052×10^{-4}	8.873×10^{-2}	8.913×10^{-2}	1.000
	$C_{L\delta_e}$	1.600×10^{-2}	5.083×10^{-4}	1.505×10^{-2}	1.701×10^{-2}	0.998
	σ	1.585×10^{-2}	1.807×10^{-4}	1.551×10^{-2}	1.622×10^{-2}	1.004
	ν	4.594×10^{0}	2.035×10^{-1}	4.221×10^{0}	5.006×10^{0}	1.002
Base + $C_{L\delta_{\text{stab}}}$ Eq. (9)	ΔC_{L0}	(3.042×10^{-6})	7.650×10^{-4}	(-1.445×10^{-3})	(1.489×10^{-3})	0.999
	$C_{L\alpha}$ *1	8.887×10^{-2}	5.294×10^{-4}	8.780×10^{-2}	8.995×10^{-2}	0.999
	$C_{L\delta_e}$ *1	1.597×10^{-2}	5.228×10^{-4}	1.495×10^{-2}	1.702×10^{-2}	0.999
	$C_{L\delta_{\text{stab}}}$ *1	-1.921×10^{-4}	1.477×10^{-3}	-3.165×10^{-3}	2.719×10^{-3}	0.999
	σ	1.584×10^{-2}	1.928×10^{-4}	1.547×10^{-2}	1.620×10^{-2}	1.006
	ν	4.566×10^{0}	2.027×10^{-1}	4.208×10^{0}	4.998×10^{0}	1.008
Base + C_{Lq} Eq. (10)	ΔC_{L0}	(7.187×10^{-3})	7.153×10^{-4}	(5.761×10^{-3})	(8.532×10^{-3})	0.999
	$C_{L\alpha}$ *1	8.905×10^{-2}	1.014×10^{-4}	8.886×10^{-2}	8.925×10^{-2}	1.001
	$C_{L\delta_e}$ *1	2.038×10^{-2}	4.996×10^{-4}	1.936×10^{-2}	2.129×10^{-2}	1.000
	C_{Lq} *2	4.954×10^{-2}	1.083×10^{-3}	4.735×10^{-2}	5.159×10^{-2}	0.999
	σ	1.654×10^{-2}	1.687×10^{-4}	1.619×10^{-2}	1.686×10^{-2}	0.999
	ν	8.568×10^{0}	6.040×10^{-1}	7.484×10^{0}	9.780×10^{0}	0.999

*1 $[1/\text{deg}]$ *2 $[1/(\text{deg/s})]$

dot corresponds to the ratio of the samples acquired in 2020 respect to the whole samples belonging to each bin. The number of the samples of each bin corresponds to the dot size. As the figure shows, some samples acquired in 2020 are located at the position where the linearity is corrupted. They may be helpful for the estimation of the stabilizer contribution.

The estimated results are summarized in the middle of Table 3. The estimated σ is almost identical to the base model results, however, ν is slightly smaller than the base model. In addition, $C_{L\delta_{\text{stab}}}$ is not estimated well due to its relatively large standard deviation 1.477×10^{-3} compared to its mean -1.921×10^{-4}. Therefore, the modeling to include the stabilizer effects is concluded to fail. On the other hand, the t-distribution postulation may be helpful to protect the estimation of the other parameters from this deterioration of the model, because the means of ΔC_{L0}, $C_{L\alpha}$, $C_{L\delta_e}$ and σ are almost unchanged from the base model.

3.5 Pitch rate effect C_{Lq} consideration

The third model is the linear conjunction of the pitch rate contribution to the base model as

$$C_L \sim \text{Student\_t}\left(\nu, C_{L0}+C_{L\alpha}\alpha+C_{L\delta_e}\delta_e+C_{Lq}q, \sigma\right), \tag{10}$$

where C_{Lq} is the new model parameter. The estimation results are described in the bottom of Table 3. Although the mean σ is slightly enlarged to 1.654×10^{-2}, the mean ν is greatly improved to 8.568 compared to the base model. The effect of the inclusion of the pitch rate contribution can be visually recognized with Fig. 8. The figure shows residuals calculated by subtracting the mean model output in condition of each sample. Namely, the residual of the base model is

$$C_L - \left(\overline{C_{L0}}+\overline{C_{L\alpha}}\alpha+\overline{C_{L\delta_e}}\delta_e\right), \tag{11}$$

while that of the third model including C_{Lq} is

$$C_L - \left(\overline{C_{L0}}+\overline{C_{L\alpha}}\alpha+\overline{C_{L\delta_e}}\delta_e+\overline{C_{Lq}}q\right). \tag{12}$$

Every residuals in the third model are located near from the zero residual compared to the residuals in the base model. From another point of view, the regression line $\bar{C_{Lq}}q$ is successfully estimated in

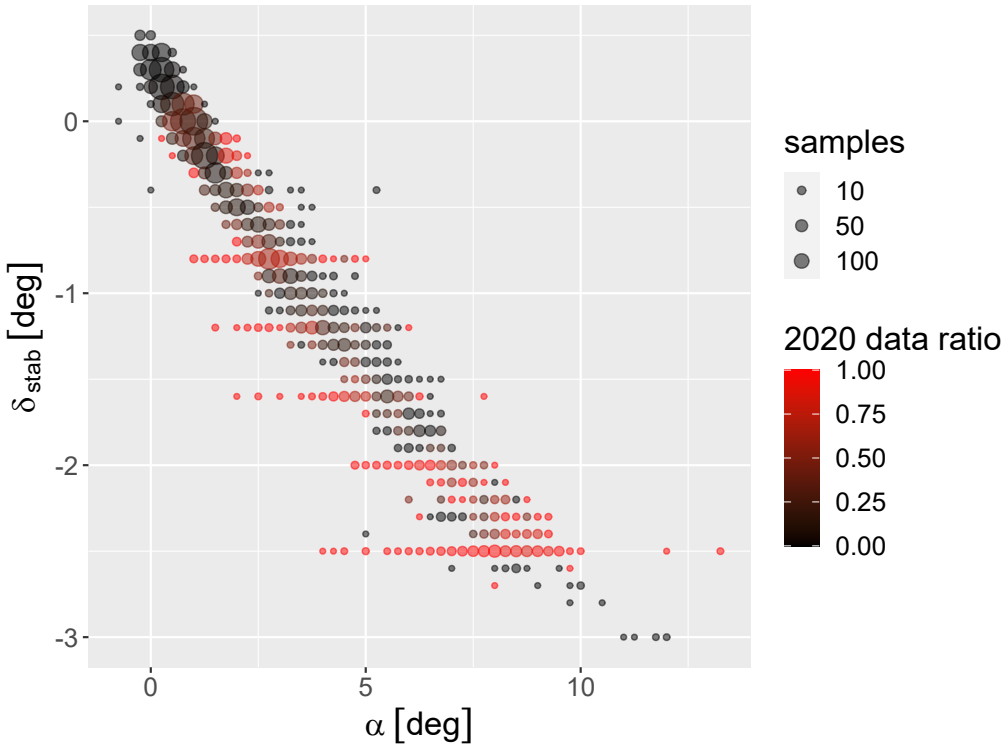


Figure 7 – Binned scatter plot of angle of attack and stabilizer deflection

the third model because the residuals in the base model are located near from the line. Although there are small number of samples which have the information about the q effect shown in the figure, it is concluded that this third model extracts that information and leads to the ν improvement.

3.6 Reynold's effect $C_{L\mathrm{Re}}$ consideration

The forth model cooperates with Reynold's number effect based on the third model like

$$C_L \sim \text{Student\_t}\left(\nu, C_{L0} + C_{L\alpha}\alpha + C_{L\delta_e}\delta_e + C_{Lq}q + C_{L\mathrm{Re}}\left(\log_{10}\mathrm{Re} - \log_{10}\mathrm{Re}_0\right), \sigma\right), \tag{13}$$

where the last term of the right side is that effect. Its linear formulation to logarithm of the Reynold's number is applied according to Wang [8]. Re_0 is the reference Reynold's number defined as the sample mean of the data in which $\alpha \leq 0.1\,[\mathrm{deg}]$ in order to mitigate the change of C_{L0}.

The estimated results are summarized in the top of Table 4. For reference, the results of the third model are duplicated in the bottom of the table. The middle of the table discussed later is the results of the same forth model but the applied data is different from the original data. Compared the third model results, the forth model improves the mean σ from 1.654×10^{-2} to 1.445×10^{-2}, while the mean ν is degraded from 8.568 to 8.259.

According to the discussion in Sec. 2.3, the data may contains the disturbing data in terms of the forth model. It implies that if the data is appropriately segmented, it is possible to improve the model in a specific region. Therefore, the distribution of the Reynold's number in the data is checked as shown in Fig. 9. The figure shows the scatter plot of the Reynold's number respect to the angle of attack α. It indicates that there are many samples varying the Reynold's number in the region of $\alpha \leq 5\,[\mathrm{deg}]$. The data in that low α region are 10791 samples, which is still sufficiently much because of 89.8% of 12021 samples in the original data. The estimation results of the low α data with the forth model are summarized in the middle of Table 4 as mentioned previously. It is natural that the mean σ decreases from 1.445×10^{-2} of the original data to 1.317×10^{-2}, because part of the reason is reduction of the samples. However, the improvement ν from 8.259 to 8.855, which is more than 8.568 of the third model, is especially significant. Therefore, it is proved that although space covered by the model may

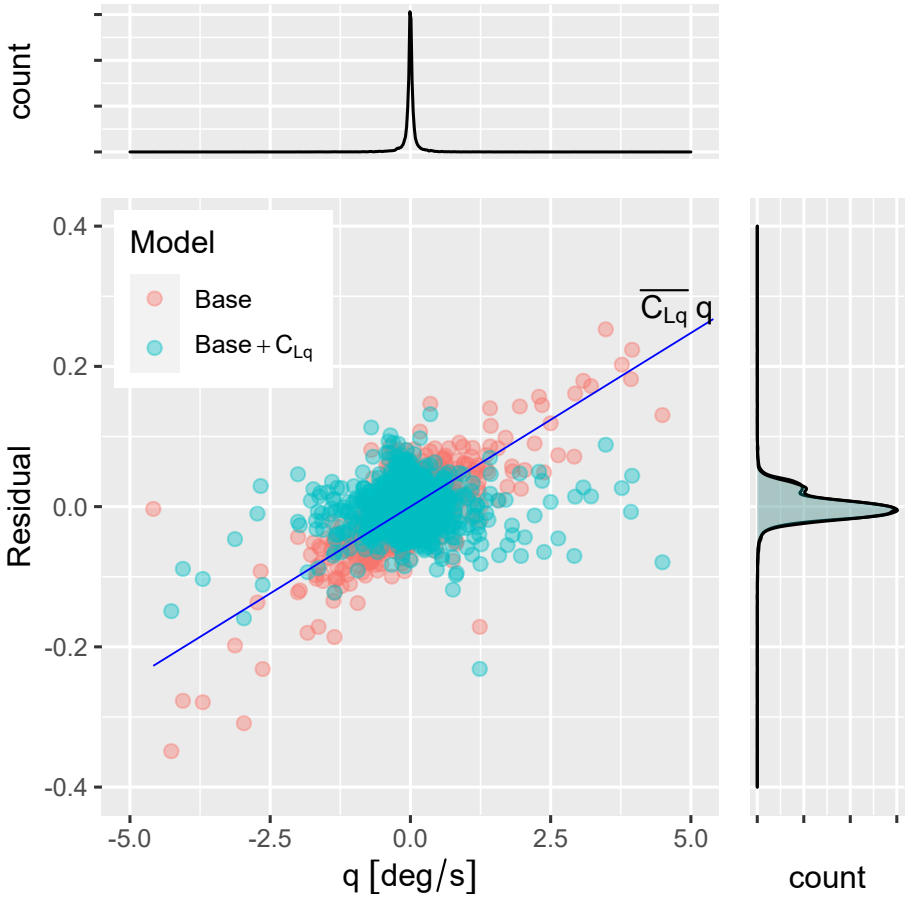


Figure 8 – Residual in models with and without pitch rate contribution

Table 4 – Estimation results with Reynold's number effect consideration

Postulated Model & Data	Item	Mean	Standard deviation	2.5%	97.5%	$\hat{R}$
Base + C_{Lq} + $C_{L\mathrm{Re}}$ Eq. (13) & Original data	ΔC_{L0}	(-2.152×10^{-3})	5.873×10^{-4}	(-3.284×10^{-3})	(-1.004×10^{-3})	0.999
	$C_{L\alpha}$ *1	9.322×10^{-2}	1.159×10^{-4}	9.300×10^{-2}	9.344×10^{-2}	1.000
	$C_{L\delta_e}$ *1	1.279×10^{-2}	4.326×10^{-4}	1.197×10^{-2}	1.361×10^{-2}	0.999
	C_{Lq} *2	3.872×10^{-2}	1.178×10^{-3}	3.635×10^{-2}	4.093×10^{-2}	0.999
	$C_{L\mathrm{Re}}$	1.282×10^{-1}	2.100×10^{-3}	1.242×10^{-1}	1.322×10^{-1}	0.998
	σ	1.445×10^{-2}	1.429×10^{-4}	1.418×10^{-2}	1.473×10^{-2}	1.002
	ν	8.259×10^{0}	4.900×10^{-1}	7.380×10^{0}	9.339×10^{0}	1.000
Base + C_{Lq} + $C_{L\mathrm{Re}}$ Eq. (13) & Data in $\alpha \leq 5$	ΔC_{L0}	(-1.153×10^{-2})	6.492×10^{-4}	(-1.281×10^{-2})	(-1.022×10^{-2})	0.999
	$C_{L\alpha}$ *1	9.772×10^{-2}	1.454×10^{-4}	9.744×10^{-2}	9.801×10^{-2}	0.998
	$C_{L\delta_e}$ *1	9.535×10^{-3}	4.510×10^{-4}	8.690×10^{-3}	1.041×10^{-2}	0.999
	C_{Lq} *2	2.703×10^{-2}	1.099×10^{-3}	2.486×10^{-2}	2.921×10^{-2}	0.999
	$C_{L\mathrm{Re}}$	1.443×10^{-1}	2.050×10^{-3}	1.403×10^{-1}	1.481×10^{-1}	1.003
	σ	1.317×10^{-2}	1.243×10^{-4}	1.293×10^{-2}	1.342×10^{-2}	0.999
	ν	8.855×10^{0}	5.481×10^{-1}	7.859×10^{0}	9.987×10^{0}	1.000
Base + C_{Lq} Eq. (10) & Original data	ΔC_{L0}	(7.187×10^{-3})	7.153×10^{-4}	(5.761×10^{-3})	(8.532×10^{-3})	0.999
	$C_{L\alpha}$ *1	8.905×10^{-2}	1.014×10^{-4}	8.886×10^{-2}	8.925×10^{-2}	1.001
	$C_{L\delta_e}$ *1	2.038×10^{-2}	4.996×10^{-4}	1.936×10^{-2}	2.129×10^{-2}	1.000
	C_{Lq} *2	4.954×10^{-2}	1.083×10^{-3}	4.735×10^{-2}	5.159×10^{-2}	0.999
	σ	1.654×10^{-2}	1.687×10^{-4}	1.619×10^{-2}	1.686×10^{-2}	0.999
	ν	8.568×10^{0}	6.040×10^{-1}	7.484×10^{0}	9.780×10^{0}	0.999

*1 $[1/\mathrm{deg}]$ *2 $[1/(\mathrm{deg/s})]$

be narrower, an appropriate data segmentation is effective for the improvement of the model in terms of the pursuit of the steeper parameter distribution.

4. Conclusion

In this study, the relationship between data and postulated model for the estimation of the flight characteristics of aircraft was discussed. Compared to a general identification problem, the coverage of the available data respect to the postulated model must be carefully checked for the reliable estimation of the flight characteristics. Therefore, using Student's t-distribution with the MCMC sampling method was proposed for the estimation. The two advantages resulted from the method; the robustness to the data, and the estimation improvement by using the ν, that is, the degree of freedom of the t-distribution. The former was the functionality to make the estimated distribution insensitive to infrequent data. The latter ν worked as the metric to improve the model by segmenting the input data in a specific region. These advantages were proved by the simple numerical examples and the lift coefficient modeling with the flight data of the fixed-wing aircraft. Especially, the lift coefficient modeling was successfully performed by taking the pitch rate contribution and Reynold's number effect into account. Moreover, the ν helped to reveal that the data used in this study did not have the meaningful information about the stabilizer contribution to the lift.

Acknowledgment

This work was supported by JSPS KAKENHI Grant JP21381031.

Contact Author Email Address

naruoka.masaru@jaxa.jp

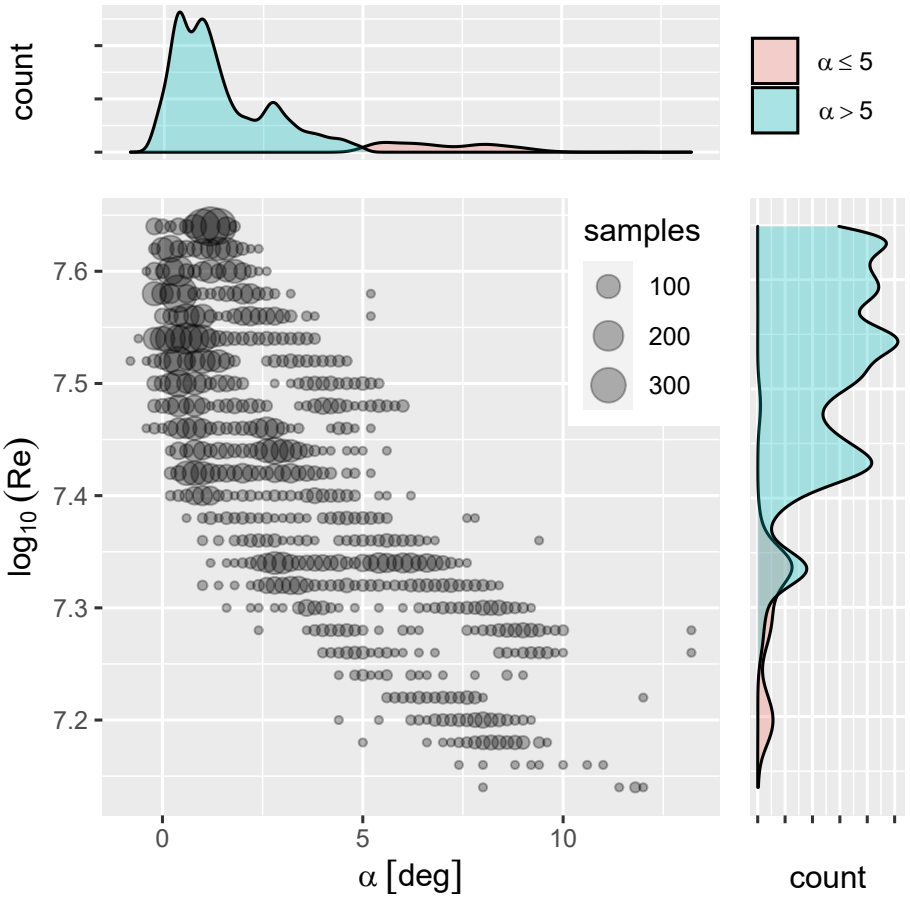


Figure 9 – Binned scatter plot of angle of attack and Reynold's number

References

[1] Ravindra V. Jategaonkar. *Flight Vehicle System Identification: A Time-Domain Methodology, Second Edition*. American Institute of Aeronautics and Astronautics, Inc., January 2015.

[2] M. Naruoka. Bayesian approach to simultaneous static and dynamic flight characteristics modeling. In *ICAS2021, 32th International Congress of the Aeronautical Sciences, Shanghai, China*, Sep 2021.

[3] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[4] Bob Carpenter, Daniel Lee, Marcus A. Brubaker, Allen Riddell, Andrew Gelman, Ben Goodrich, Jiqiang Guo, Matt Hoffman, Michael Betancourt, and Peter Li. Stan: A probabilistic programming language, 2017.

[5] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

[6] Tomita Hiroshi and Naruoka Masaru. Jaxa flying test bed “hisho”. *Aeronautical and Space Sciences Japan*, Vol. 62, No. 6, pp. 195–201, 2014. (in Japanese).

[7] Atsuya Ishikawa, Masaru Naruoka, Tetsujiro Ninomiya, and Shuichi Adachi. Weight estimation of aircraft in flight by sensor fusion with revised forward-backward smoother. *Journal of the Japan Society for Aeronautical and Space Sciences*, Vol. 70, No. 2, pp. 58–66, 2022. (in Japanese).

[8] Yuanjing Wang, Dawei Liu, Xin Xu, and Guoshuai Li. Investigation of Reynolds Number Effects on Aerodynamic Characteristics of a Transport Aircraft. *Aerospace*, Vol. 8, No. 7, p. 177, July 2021.