BSCE 20 / WP 18
Helsinki, 21-25 May 1990

A GENERAL STATISTICAL APPROACH TO IDENTIFICATION OF BIRD REMAINS AFTER COLLISION BETWEEN AIRCRAFT AND BIRDS

Presented by
Valery D. Ilyichev, Nikolai A. Nechval and Victor Y. Biryukov

BSCE 20/WP

Helsinki, 21-25 May 1990

A GENERAL STATISTICAL APPROACH TO IDENTIFICATION OF BIRD REMAINS AFTER COLLISION BETWEEN AIRCRAFT AND BIRDS

Valery D. Ilyichev, Nikolai A. Nechval and Victor Y. Biryukov

Severtzov Institute, Civil Aviation Engineers Institute
Moscow, Riga, USSR

SUMMARY

The aim of this paper is to present a general statistical approach to identification of bird remains after collision between aircraft and birds, which is based on bird strike statistics and can be applied in combination with some methods currently used and described by Brom (1988). The proposed approach is, therefore, trying to improve the situation by applying to the bird strike databases more sophisticated analytical techniques generally used to analyse rather variable biological data. The sample sizes of the data set of bird strike reports are small and in some cases, multivariate analysis is not possible. The data are, therefore, statistically weak. However, in order to illustrate a suggested statistical approach to identification of bird remains after collision between aircraft and birds, these weaknesses have been ignored. The approach is based on the results of Nechval (1988, 1989) and is immediately applicable when the alternative distributions have given functional forms but with unspecified parameters. The nonparametric cases can be treated in a similar manner, but no attempt has been made in this paper to offer explicit solutions. The main feature of the approach to identification of bird remains is the population elimination rule. When certain conditions are met, the decision is taken to eliminate specific populations from further consideration, and the identification process is continued with a reduced number of populations. An example is given.

1. INTRODUCTIO

Identification
and birds is t
areas of appli
regulations re
ment of comple
in which a bir
can result in
phic failure t
engines, it is
antity to take
bird ingestion
on. A valuable
in-service inc
tion and accur

The tradition
tion is that
lection. In o
therefore les
adults, a dis
ke statistics
micromorpholo
be distinguis
on macroscopi
of the bird r
hardly any di
der birds, wh
nounced, at l

Bird strike r
lots of aircr
staff who fir
ers who find
on of an airc
craft, the ef
strike occur
staff, on the
craft or the
fect upon th
details abou
on of bird s
self evident
is an essent
and useful d
dination req
relies upon
goodwill and
airport.

The aim of t
ach to ident
aircraft and
and can be a
used and des
therefore, t
bird strike
generally us

1. INTRODUCTION

Identification of bird remains after collision between aircraft and birds is the subject of growing interest and has the many areas of application. For example, civil engine airworthiness regulations relating to bird hazards have required the development of complex and extremely expensive test routines. The way in which a bird of a particular species passes through an engine can result in some cases in no damage and in others to catastrophic failure to the engine. Because of the high costs of testing engines, it is never possible to repeat tests in sufficient quantity to take into account the many variables involved in a bird ingestion, even with the advent of modern computer simulation. A valuable source of information is, therefore, obtained from in-service incidents. This is an area where the complete collection and accurate identification of bird remains is essential.

The traditionally used and most simple way of feather identification is that of comparing unknown feathers with a reference collection. In order to be able to determine whether younger (and therefore less experienced) birds are more accident-prone than adults, a distinction between age classes is needed in bird strike statistics. Since no diagnostic characters are found in the micromorphology of feathers by which juvenile and adult birds can be distinguished, all information on the age of the bird depends on macroscopical criteria, and hence on the size and condition of the bird remains available for examination. For some species hardly any differences in plumage exist between juvenile and older birds, whereas in others these differences are quite pronounced, at least during certain periods of the year.

Bird strike reports arise from one of three sources: 1) from pilots of aircraft which have experienced a strike, 2) from ground staff who find a corpse on the manoeuvring area, 3) from engineers who find evidence of a bird strike during a routine inspection of an aircraft. Pilots will often have details about the aircraft, the effect on the flight, the time of day etc. when a strike occurred, but not the species of bird involved. Ground staff, on the other hand, may know the bird, but not the aircraft or the time of the incident. Engineers often know the effect upon the aircraft, have the bird remains but may have no details about when or where the strike occurred. A high proportion of bird strike reports will, therefore, be incomplete. It is self evident that complete and accurate reporting of bird strikes is an essential prerequisite to the development of a meaningful and useful data base from which to carry out analysis. The coordination required to ensure complete and accurate reporting again relies upon education of the parties involved and also often, the goodwill and cooperation of the air traffic controllers at the airport.

The aim of this paper is to present a general statistical approach to identification of bird remains (after collision between aircraft and birds), which is based on bird strike statistics and can be applied in combination with some methods currently used and described by Brom (1988). The proposed approach is, therefore, trying to improve the situation by applying to the bird strike databases more sophisticated analytical techniques generally used to analyse rather variable biological data.

The sample sizes of the data set of bird strike reports are small and in some cases, multivariate analysis is not possible. The data are, therefore, statistically weak. However, in order to illustrate a suggested statistical approach to identification of bird remains after collision between aircraft and birds, these weaknesses have been ignored.

2. FORMULATION OF THE PROBLEM

Let us suppose we are considering the remains of bird that suffered the collision with aircraft. It is known that this bird belongs to one of m species of birds, but to which of them it belongs is unknown. Our problem is to identify the bird remains with the proper species, on the basis of the values of measurements of p characteristics of these remains available from bird strike.

The problem of identification, that is of assigning the observed remains of bird to the appropriate group, admits a simple solution when the probability distributions of measurements in the alternative groups (populations) of measurements of p characteristics of bird remains are completely specified. The decision rule consists in setting up a correspondence between observed values of measurements of p characteristics of bird remains and one of m distributions of measurements of p characteristics of bird remains, where each distribution corresponds to one of m alternative populations of measurements of p characteristics of bird remains associated with certain species of birds. In practice it is rarely possible to specify completely the distributions of these characteristics, but they may be estimable on the basis of independent samples from each of the alternative distributions.

Let $X(1), \ldots ,X(m)$ be independent samples of observed values of measurements of p characteristics of bird remains from m alternative populations of measurements of p characteristics of bird remains, which may be partially specified, as when the functional forms of the probability densities are given but with unspecified parameters, or completely unspecified. In this paper we consider the identification problem when the alternative distributions have given functional forms but with unspecified parameters. After a p-dimensional vector X (measurement of p characteristics of bird remains) is drawn from a population known a priori to be one of the above set of m populations of measurements of p characteristics of bird remains, the problem is to infer from which population the vector X has been drawn. The decision rule should be in the form of associating X with one of the samples $X(1), \ldots ,X(m)$, and declaring that X has come from the same population as the sample with which it is associated.

In this paper a general statistical approach to identification of bird remains after collision between aircraft and birds has been developed with help of which the identification problem can be solved, utilizing only the sample information. This approach is based on the results of Nechval (1988, 1989) and is immediately applicable when the alternative distributions have given functional forms but with unspecified parameters. The nonparametric cases can be treated in a similar manner, but no attempt has been made in this paper to offer explicit solutions.

3. AN APPROACH TO IDENTIFICATION OF BIRD REMAINS

The main feature of the proposed approach to identification of bird remains is the population elimination rule. When certain conditions are met, the decision is taken to eliminate specific populations from further consideration, and the identification process is continued with a reduced number of populations. The elimination rule is based on some exact tests for discriminant analysis in the presence of unknown parameters and small samples of the data. These tests are based (in the main) on the parameter-free or distribution-free statistics which are obtained by means of suitable transformations on the original observations and possess certain known distributions with known parameters (see Nechval, 1988, 1989).

If all the populations except one are eliminated, we decide that the vector of observations, X, belongs to the remaining population. If however the set of populations not yet eliminated contains more than one element, in this exceptional situation the object (bird remains) is treated as belonging to the population for which the value of the likelihood function of the test statistic is greatest. When there is the possibility that the object does not belong to any of the m populations and all the populations are eliminated from further consideration, we decide that the measurement X does not belong to any of the m populations, i.e., it belongs to the (m+1)th population whose distribution is unspecified.

To illustrate the application of the proposed approach to constructing the procedure of identification of bird remains, when sample sizes of the statistical data are small, an example is given below.

4. AN EXAMPLE

Consider X, a (p x 1) vector of observations on an object (bird remains etc.), which is to be classified as belonging to one of two given populations or to a third population whose distribution is unspecified. Assume that if X is an observation vector from population i then

$$X \sim N_p(b_i, Q), \quad i \in \{1,2\}, \tag{1}$$

where $N_p(b_i,Q)$ denotes p-variate normal distribution with the parameters b_i and Q. Suppose that the means b_1, b_2 and a common variance-covariance matrix Q are unknown and estimated by $\overline{X}_1$, $\overline{X}_2$ and S, respectively. Here S is the pooled estimate of Q from two independent random samples of sizes n_1 and n_2 from populations 1 and 2, respectively, and $\overline{X}_1$ and $\overline{X}_2$ are the corresponding sample means,

$$S = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (X_{ij}-\overline{X}_i)(X_{ij}-\overline{X}_i)'/(n_1+n_2-2), \tag{2}$$

$$\overline{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i, \quad i=1,2. \tag{3}$$

If $X \sim N_p(b_i, Q)$, $i \in \{1,2\}$, and is distributed independently of the two samples then

$$(X-\overline{X}_i) \sim N_p(0, [(n_i+1)/n_i]Q) \tag{4}$$

and is independent of

$$(n_1+n_2-2)S \sim W_p(n_1+n_2-2, Q), \tag{5}$$

where $W_p(r,Q)$ denotes the central Wishart distribution with r degrees of freedom (d.f.). Thus the statistic

$$T_i^2 = [n_i/(n_i+1)](X-\overline{X}_i)'S^{-1}(X-\overline{X}_i) \tag{6}$$

has Hotelling's T^2 distribution and

$$kT_i^2 \sim F(p, n_1+n_2-p-1), \tag{7}$$

where

$$k = (n_1+n_2-p-1)/[p(n_1+n_2-2)] \tag{8}$$

and $F(p, n_1+n_2-p-1)$ denotes the central F distribution with p and n_1+n_2-p-1 d.f.

Let

$$C_i = \left\{X:\ (X-\overline{X}_i)'S^{-1}(X-\overline{X}_i) \leq [(n_i+1)/(n_i k)]F_{a;p,n_1+n_2-p-1}\right\}, \quad \text{for } i \in \{1,2\}, \tag{9}$$

where $F_{a;\dots}$ is the upper (100)a percentage point of the central F distribution with the indicated degrees of freedom. Considering the distribution of kT_i it is noted that

$$\Pr(X \in C_i) = 1-a \quad \text{for } i \in \{1,2\}. \tag{10}$$

Thus C_1 and C_2 are ellipsoids of concentration of size 1-a for the two given populations. For a fixed value of i,

$$\overline{C}_i = \left\{X:\ X \in \text{Complement of } C_i\right\} \tag{11}$$

may be taken as the critical region with exact size a for the sub-hypothesis

$$H_{0i}:\ X \in \text{Population } i, \tag{12}$$

i.e., reject H_{0i} iff

$$(X-\overline{X}_i)'S^{-1}(X-\overline{X}_i) > [(n_i+1)/(n_i k)]F_{a;p,n_1+n_2-p-1}. \tag{13}$$

If one of the two populations is eliminated, we decide that the vector of observations, X, belongs to the remaining population. If however the set of populations not yet eliminated contains the two populations, in this exceptional situation X is treated as belonging to the ith population for which

$$f_i((x-\overline{x}_i)'S^{-1}(x-\overline{x}_i)) > f_j((x-\overline{x}_j)'S^{-1}(x-\overline{x}_j)),$$
$$i,j \in \{1,2\}, \; i \neq j, \tag{14}$$

where $f_i(\cdot)$ is the probability density function of the test statistic $(X-\overline{X}_i)'S^{-1}(X-\overline{X}_i)$.

If the two populations are eliminated from further consideration, we decide that X belongs to the third (unknown) population whose distribution is unspecified. Note that the probability of accepting the hypothesis H_{03}: X does not belong to either population 1 or population 2, when H_{03} is not true, is defined by

$$\Pr(X \in \text{Complement of } C_1 \bigcup C_2) = \Pr(X \in \overline{C_1 \bigcup C_2}) = \Pr(X \in \overline{C}_1 \bigcap \overline{C}_2). \tag{15}$$

From a geometric argument presented in the theorem given below, it is guaranteed that this probability is between a and a/2. This should be sufficient for most practical applications.

<u>Theorem 4.1</u>. The probability of accepting the hypothesis H_{03}: X does not belong to either population 1 or population 2, when H_{03} is not true, satisfies the inequality

$$a/2 \leq \Pr(X \in \overline{C}_1 \bigcap \overline{C}_2) \leq a. \tag{16}$$

<u>Proof</u>. Assume that H_{03} is not true, and without loss of generality, that $X \in$ population 1 (a parallel argument interchanging 1 and 2 holds for the other case). It follows from (11) that

$$\Pr(X \in \overline{C}_1) = a \tag{17}$$

and hence

$$\Pr(X \in \overline{C}_1 \bigcap \overline{C}_2) \leq a. \tag{18}$$

The lower bound on $\Pr(X \in \overline{C}_1 \bigcap \overline{C}_2)$ is more difficult to establish. Let

$$N_a = [(n_1+1)/(n_1 k)]F_{a:p,n_1+n_2-p-1} \tag{19}$$

and $X \in \overline{C}_1$. We have

$$N_a < (X-\overline{X}_1)'S^{-1}(X-\overline{X}_1) = [(X-\overline{X}_2)-(\overline{X}_1-\overline{X}_2)]'S^{-1}[(X-\overline{X}_2)-(\overline{X}_1-\overline{X}_2)]$$
$$= (X-\overline{X}_2)'S^{-1}(X-\overline{X}_2) - (2X-\overline{X}_1-\overline{X}_2)'S^{-1}(\overline{X}_1-\overline{X}_2). \tag{20}$$

Thus, if

$$(2X-\overline{X}_1-\overline{X}_2)'S^{-1}(\overline{X}_1-\overline{X}_2) > 0 \tag{21}$$

then

$$N_a < (X-\overline{X}_2)'S^{-1}(X-\overline{X}_2) \tag{22}$$

and X also belongs to $\bar{C}_2$. (21) implies

$$(X-\bar{X}_1)'S^{-1}(\bar{X}_2-\bar{X}_1) < [(\bar{X}_2-\bar{X}_1)'S^{-1}(\bar{X}_2-\bar{X}_1)]/2 \; . \qquad (23)$$

Note that

$$\bar{C}_1 = \left\{ X:\ (X-\bar{X}_1)'S^{-1}(X-\bar{X}_1) > N_a \right\} = \bar{C}_1^- \bigcup \bar{C}_1^+ \; , \qquad (24)$$

where

$$\bar{C}_1^- = \left\{ X\in\bar{C}_1:\ (X-\bar{X}_1)'S^{-1}(\bar{X}_2-\bar{X}_1) < 0 \right\} \qquad (25)$$

and

$$\bar{C}_1^+ = \left\{ X\in\bar{C}_1:\ (X-\bar{X}_1)'S^{-1}(\bar{X}_2-\bar{X}_1) \geq 0 \right\} \qquad (26)$$

are such that

$$\bar{C}_1^- \bigcap \bar{C}_1^+ = \emptyset \qquad (27)$$

and

$$\Pr(X\in\bar{C}_1^-) = \Pr(X\in\bar{C}_1^+) = a/2 \; . \qquad (28)$$

It follows from (23), (24), (25), and (26) that

$$\bar{C}_1 \bigcap \bar{C}_2 = \bar{C}_1^- \bigcup \bar{C}^+ , \qquad (29)$$

where

$$\bar{C}^+ = \left\{ X\in\bar{C}_1^+:\ (X-\bar{X}_1)'S^{-1}(\bar{X}_2-\bar{X}_1) < [(\bar{X}_2-\bar{X}_1)'S^{-1}(\bar{X}_2-\bar{X}_1)]/2 \right\} \subset \bar{C}_1^+ . \qquad (30)$$

Taking into account (27), (28), (29), and (30), we have

$$\Pr(X\in\bar{C}_1 \bigcap \bar{C}_2) \geq a/2. \qquad (31)$$

This ends the proof.

For the sake of the numerical illustration, we selected the two random samples of size six ($n_1=n_2=6$):

Sample 1

	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
a	200	160	188	190	177	210
b	137	118	134	143	127	140
c	52	47	54	52	49	54
d	144	140	151	141	134	149
e	14	15	14	13	15	13
f	102	99	98	99	105	107

and

Sample 2

	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}
a	186	184	208	199	187	188
b	107	108	125	124	123	114
c	49	43	50	46	47	48
d	120	116	125	119	129	122
e	14	16	14	13	14	12
f	84	75	88	78	75	74

drawn from the populations 1 and 2, respectively. The sample means and pooled sample variance-covariance matrix are given in Table 4.1.

TABLE 4.1. Summary of Computations for Random Samples (1 and 2)

Components		a	b	c	d	e	f
Sample Means	$\overline{X}_1'$	187.5	133.2	51.3	143.2	14.0	101.7
$(n_1=n_2=6)$	$\overline{X}_2'$	192.0	116.8	47.2	121.8	13.8	79.0
Pooled Sample Variance-Covariance Matrix (S)		198.2	99.9	27.2	37.5	-7.8	33.3
		99.9	76.6	14.0	25.2	-5.7	7.0
		27.2	14.0	7.0	10.3	-1.9	5.6
		37.5	25.2	10.3	30.2	-2.5	.23
		-7.8	-5.7	-1.9	-2.5	1.3	.10
		33.3	7.0	5.6	.23	.10	23.1

Utilizing a computer program to evaluate

$$V_1 = (X-\overline{X}_1)'S^{-1}(X-\overline{X}_1) \qquad (32)$$

and

$$V_2 = (X-\overline{X}_2)'S^{-1}(X-\overline{X}_2) \qquad (33)$$

for an arbitrary (6 x 1) vector X, we applied the test recommended above to the following data:

Vector of Observations (X)

	from the population 1			from the population 2			from the population 3		
	No.								
	1	2	3	1	2	3	1	2	3
a	191	173	186	211	201	187	158	146	135

(X) (Continued)

	1	2	3	1	2	3	1	2	3
b	131	127	136	122	114	124	141	119	127
c	53	50	56	49	47	49	58	51	52
d	150	144	148	123	130	129	145	140	140
e	15	16	14	16	14	14	8	11	10
f	104	97	111	95	74	88	107	111	108

The test size selected was a=.05 for each application yielding a critical value of

$$\frac{n_i+1}{n_i k} F_{a;p,n_1+n_2-p-1} = \frac{7}{6(.0833)} F_{.05;6,5} = (14)(4.95) = 69.3. \qquad (34)$$

Table 4.2 contains a summary of results.

TABLE 4.2. Summary of Results of Application of the Test Procedure to the Data of X

Species	X	V_1	V_2	Type I Error	Type II Error
1	1	4.46	146.7	No	
	2	7.43	127.1	No	
	3	14.83	201.3	No	
2	1	99.5	16.9	No	
	2	167.5	12.8	No	
	3	47.5	22.0	No	
3	1	105.2	357.3		No
	2	78.4	344.9		No
	3	113.1	404.7		No

REFERENCES

Brom, T.G. (1988). The analysis of feather remains: evaluation and perspectives. 19th Meeting Bird Strike Committee Europe. Madrid, Working Paper 23.

Nechval, N.A. (1988). Goodness-of-fit testing in the presence of nuisance parameters with applications to feature selection and pattern recognition in digital image processing. SPIE Proceedings Vol. 1001, 9-11 November 1988, Cambridge, Massachusetts USA, Paper No. 1001-124, 8 pages.

Nechval, N.A. (1989). Detection of a change in a sequence of images based on the free statistics. SPIE Proceedings Vol. 1199, Philadelphia, Pennsylvania USA, Paper No. 1199-141, 12 pages.