pyModeS: Decoding Mode-S Surveillance Data for Open Air Transportation Research

Junzi Sun, Huy Vû, Joost Ellerbroek, Jacco M. Hoekstra

***Abstract*—The availability of low-cost Automatic Dependent Surveillance-broadcast (ADS-B) receivers has given researchers the ability to make use of large amounts of aircraft state data. This data is being used to support air transportation research in performance study, trajectory prediction, procedure analysis, and airspace design. However, aircraft states contained in ADS-B messages are limited. More performance parameters are downlinked as Mode-S Comm-B replies, upon automatic and periodic interrogation of air traffic control secondary surveillance radar. These replies reveal aircraft airspeed, turn rate, target altitude, and so on. They can be intercepted using the same 1090 MHz receiver that receives ADS-B messages. However, a third-party observer does not know the interrogations, which originated the Comm-B replies. Thus, it is difficult to decode these messages without knowing the type and source aircraft. Furthermore, the parity check also cannot be performed without knowing the interrogations. In this paper, we propose a new heuristic-probabilistic method to decode Comm-B replies, and to check the correctness of the messages. Based on a reference dataset provided by air traffic control of the Netherlands, the method yields a success rate of 97.68% with an error below 0.01%. The performance of the proposed method is further examined with data from eight different regions of the world. The implementation of the inference and decoding process, *pyModeS*, is shared as an open-source library.**

***Index Terms*—aircraft surveillance, air traffic control, Mode-S, ADS-B, Comm-B, Enhanced Mode-S**

I. Introduction

In air transportation research, studies related to aircraft performance are often dependent on the airspeed of the aircraft. This speed information is used in the dynamic model of the aircraft to perform, for example, state estimations [1] and trajectory predictions [2]. In addition to airspeed, the performance model also takes into account other trajectory state information, such as positions, ground speeds, and altitudes.

Many of these states in the dynamic model can be openly observed using modern aircraft surveillance technology, for instance, the *Automatic Dependent Surveillance-Broadcast* (ADS-B). ADS-B provides information on aircraft position, speed, and vertical rate. The speed contained in ADS-B refers to the ground speed rather than the airspeed, unless in (rare) cases when the location cannot be determined from the Global Navigation Satellite Systems. The advantage of ADS-B is that the signals can be openly intercepted and decoded using a simple ground receiver set-up.

When airspeed is not available, there are two ways to adopt the ground speed for performance analysis. The first simple approach is to assume the ground speed as airspeed by ignoring the wind. This may cause errors in performance calculations when a strong wind is present. The second approach is to integrate the wind from numerical forecast models. However, wind data from these models often cannot reflect local wind variations accurately.

On the other hand, air traffic controllers are also interested in the same performance parameters. These parameters are constantly interrogated by surveillance radars under the *Enhanced Mode-S Surveillance* (EHS) technology. Corresponding messages are downlinked using the *Comm-B* protocol. Air traffic controllers make use of this data to better monitor and predict flights, and to make better traffic control decisions. Within these downlinked messages, information such as true airspeed, indicated airspeed, Mach number, and true heading of the aircraft are transmitted.

If available, direct access to air traffic control data would provide the most accurate information. However, due to licenses and data agreement processes, obtaining this data can be challenging for third-party researchers. Even when the access is granted, the information is often extracted from historical data archives, which makes it difficult to perform real-time performance analysis.

Nevertheless, it is possible to obtain the downlinked Comm-B data openly just as ADS-B data with the same ground receiver. However, many difficulties arise when one tries to decode these reply messages. The biggest barriers for decoding are the unknown aircraft source represented by the ICAO transponder address and the interrogation type represented by the *Comm-B Data Selector* (BDS) code. Even though the structure of messages follows open standards [3], [4], without knowing the ICAO code and BDS type, one cannot extract useful information from these messages.

The goal of this paper is to enable open and real-time access to these Mode-S messages. The main research questions of this study are defined as:

1) How to determine the source aircraft of a Comm-B message?
2) How to identify the BDS type and decode a Comm-B message without knowing the original interrogation?
3) How to detect errors in Comm-B messages under incomplete information?

The authors are with Faculty of Aerospace Engineering, Department of Control and Simulation, Delft University of Technology, Delft, 2629HS, the Netherlands.

Digital Object Identifier 10.1109/TITS.2019.2914770

To illustrate these inference efforts, Fig. 1 shows the process related to the methods that are proposed in this paper. In this figure, the downlink surveillance signal (containing ADS-B and Comm-B) from aircraft is received by a software-defined radio (SDR) receiver first. The signal is converted to a raw binary data stream, which is then further decomposed into a sequence of message frames. For ADS-B messages, information can be decoded directly. For Comm-B replies, we first use the inference methods proposed in this paper to determine the BDS code, source aircraft, and errors. Finally, the information contained in these messages is decoded.

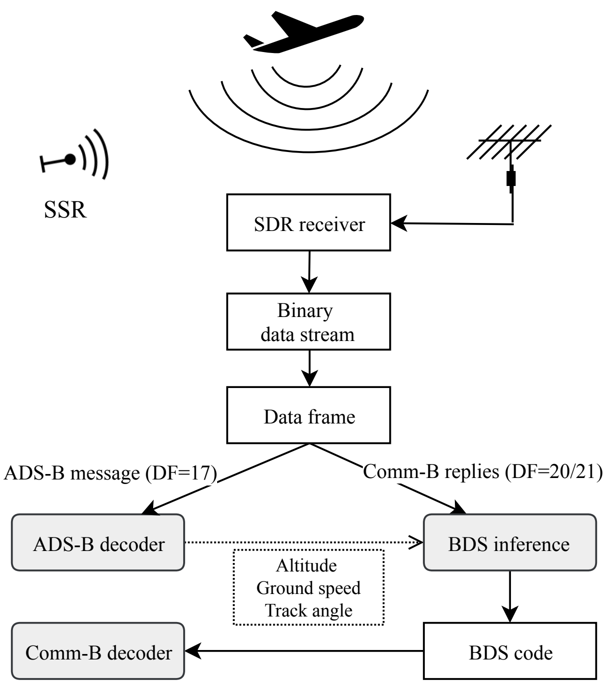

Fig. 1. The Mode-S inference and decoding pipeline

In the remaining sections of this paper, we first explain the background related to ADS-B and EHS. Next, we discuss the identification processes and error detection in detail. Several tests are proposed in this paper. The methods are also validated with a reference dataset provided by Air Traffic Control the Netherlands (LVNL). Finally, we discuss the use cases, implementation, and recommendations, as well as the conclusions at the end of the paper.

In addition to the identification and decoding process described in this paper, an open-source decoder library, *pyModeS*, which is implemented using Python programming language has also been made public [1].

II. Background

A. Fundamentals of Mode-S, ADS-B, and Comm-B

As the demand for air transportation increases, airspace over the world is becoming more crowded. To efficiently make use of the airspace and increase the traffic capacity, air traffic controllers need to rely on accurate flight trajectory predictions. Communications between ground and aircraft are becoming more frequent in order to obtain accurate updates regarding aircraft states beyond the basic position that is provided by the primary radar. Since the beginning of aircraft surveillance, different methods and standards have been developed for downlinking aircraft data. The two most common methods are based on interrogation and broadcast.

Traditionally, when an air traffic controller requires information in addition to the aircraft position, the Mode-S selective interrogation [5] is used. This is performed by the *secondary surveillance radar* (SSR). Numerous aircraft states can be interrogated by the SSR. The most common downlinked messages are Comm-B replies. The content of interrogation is identified by the BDS code, which is a two-digit hexadecimal code (8 bits) that indicates the information desired by the air traffic controller. In total, 255 BDS codes can be defined. The reply data is encoded in a 112-bit Comm-B downlink message. Among all these BDS codes, several BDS codes are grouped and identified as Mode-S enhance surveillance (EHS), which is consist of *selected intention report* (BDS 40), *track and turn report* (BDS 50), and *heading and speed report* (BDS 60).

The simpler ADS-B is an implementation of Mode-S extended squitter [4]. It is a newer technology compared to the interrogation-based Mode-S, which allows the automatic broadcast of the aircraft state information at a constant rate. In many regions, aircraft are being required to be equipped with Mode-S transponders that are compatible with ADS-B. When it is enabled, ADS-B allows aircraft to automatically report the identification, location, speed, and operational status. The update interval of critical states (such as position and speed) is designed to be around 0.5 seconds.

Both ADS-B messages and Comm-B replies are transmitted using the 1090 MHz transponder. Downlinked signals can be intercepted freely using low-cost commercial off-the-shelf ground receivers. Several crowd-sourced initiatives have been constructing global networks of ground receivers, for example, *ADS-B Exchange*, *FlightRadar24*, *FlightAware*, and *OpenSky-network*. The quantity of data gathered by these networks is enormous, which leads to great potential in air transportation researches. For example, in recent research, this data has been used for operational performance studies [6] and trajectory prediction [7]. The ground receiver networks also enable the possibility to determine aircraft location by using multilateration [8].

ADS-B is designed as an independent communication protocol, where the message itself contains all information needed for decoding. On the contrary, Comm-B communication is designed as a dependent protocol. Only the air traffic controller who initiated the interrogation can identify the source aircraft and decode the content of the replies. To this extent, third-party observers have no information on the interrogated aircraft or corresponding BDS code. However, it is has been shown that some information is possible to be extracted in an earlier research [9], which is used to provide meteorological observations.

[1]Available at: https://github.com/junzis/pyModeS

B. Regulation and availability

Several Mode-S capabilities are mandatory for aircraft flying in European airspace since 2009. Two different categories of Mode-S surveillance are defined, which are elementary surveillance (ELS) and enhanced surveillance (EHS) [10]. According to European regulation, all aircraft that fly Instrument Flight Rules (IFR) in general air traffic (GAT) must be ELS compliant. In addition, all fixed-wing aircraft flying IFR in GAT with a maximum take-off mass greater than 5.7 ton or a maximum cruising true airspeed greater than 250 knots must be EHS compliant. ADS-B is a newer surveillance technology and has also been adopted broadly. Regulators in both Europe and the United States have set the agenda for obligatory compliance.

Since ADS-B does not require active interrogations from surveillance radar, the messages are broadcast and available at all times everywhere. These messages can be received by ground receivers or satellites [11].

In Mode-S ELS, only a limited number of parameters are reported, including aircraft identity, altitude, flight status, and related supporting parameters. In Mode-S EHS, more aircraft states are interrogated, such as indicated airspeed, Mach number, vertical rate, magnetic heading, track angle, roll angle, selected altitude, and ground speed.

Depending on the location of the (third-party) ground receiver, the number of received replies varies. The availability and quantity of Comm-B messages also depend on air traffic density and the number of secondary surveillance radars in the area, as well as the rate of interrogation.

C. Data structure

In this paper, we focus on two types of messages, which are ADS-B message and Comm-B message from Mode-S EHS. The structures of ADS-B and Comm-B messages are defined in ICAO Annex 10 [12]. ADS-B and Mode-S data are constructed using the data frame shown in Fig. 2, with a total message length of 112 bits. The number of bits of each segment is indicated with parentheses in this figure. Each message starts with the *downlink format* (DF), followed by a 27-bit header with different components. Then, the crucial 56-bit data is appended with the downlink information encoded. Lastly, 24 bits are dedicated to the parity checksum.

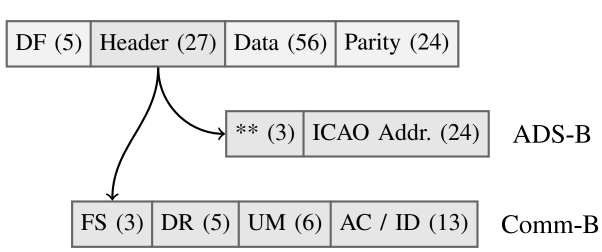

Fig. 2. The structures of ADS-B and Mode-S Comm-B messages

ADS-B messages are identified by a DF number of 17 (`10001` in binary format). In the header of an ADS-B message, the address of the aircraft transponder is indicated. This is a 24-bit address assigned by ICAO and categorized according to geographic region and country. The leading 3 bits are sub-type or category in different types of ADS-B messages. The *Type Code* (TC) is set using the first 5 bits of the 56-bit data segment. It defines the general type of message, for example, airborne position, airborne velocity, surface position, identification, etc.

In a Comm-B reply message, the DF number can be either 20 or 21 (`10100` or `10101` in binary format). In the case of DF=20, the last 15 bits of the header indicates the *Altitude Code* (AC). When DF=21, the last 15 bits represent the *Identification Code* (ID) (a.k.a: the *squawk code*). The leading three segments in the header are *Flight Status* (FS), *Downlink Request* (DR), and *Utility Message* (UM). Unlike ADS-B, there is no indication of ICAO transponder address nor the BDS code in a message, except for a few cases. [2]

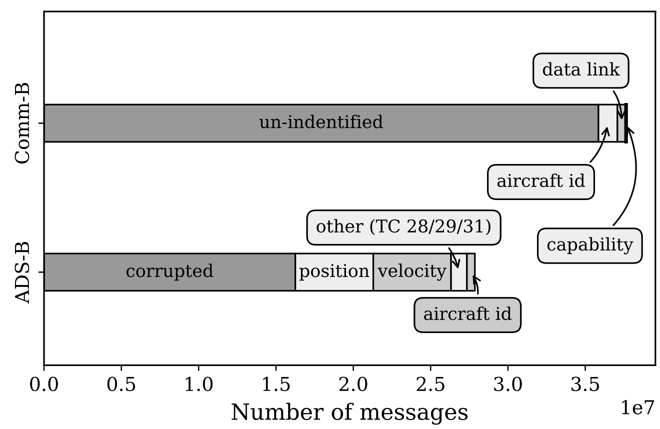

Fig. 3. Number of ADS-B message and Mode-S Comm-B replies received in 24 hours, May 30, 2018

Fig. 3 illustrates the distribution of ADS-B and Comm-B replies, together with their distinct types, for a 24-hour period of data from a ground receiver situated in Delft, the Netherlands. We can see there are more Comm-B messages than ADS-B messages. About 35 out of 38 million EHS messages are not directly identifiable. Unlike ADS-B, none of the Comm-B messages can be checked for corruption due to the incomplete information on aircraft source and BDS code.

It is worth pointing out that more than half of the ADS-B messages (16 out of 28 million messages) are corrupted. The corruption of messages is also analyzed in the later sections of this paper.

III. Basic decoding of ADS-B data

As mentioned earlier, the content of an ADS-B message is directly identified by its Type Code. The primary parameters downlinked by each type are listed in Table I.

Within the same group, Type Code values also have their own indications. For example, in the identification group, a TC of 1 to 4 indicates different aircraft categories defined by

[2]The exception cases are BDS 10, 20, and 30 messages. The BDS code is indicated in these messages. However, together, they only represent a very small percentage of all messages.

TABLE I
BASIC ADS-B MESSAGE TYPES (VERSION 0, 1, 2)

TC	Content	Primary parameters
1 - 4	Aircraft identification	Call sign
5 - 8	Surface position	Latitude Longitude Speed Track angle
9 - 18 20 - 22	Airborne position	Latitude Longitude Altitude
19	Airborne velocity	East-west component North-south component Vertical rate

the Mode-S standard according to [3]. In position groups, TC defines the level of accuracy and uncertainty.

Since ADS-B was first introduced up until now, there have been different versions of implementation: version 0, version 1, and version 2. In version 1 and 2, the accuracy and uncertainty are re-classified using two additional supplement bits. They offer more levels of *Navigation integrity category* (NIC). In addition, from ADS-B version 1 onward, three more types are introduced, as shown in Table II.

TABLE II
ADDITIONAL ADS-B MESSAGE TYPES (FOR VERSION 1, 2)

TC	Content	Primary parameters
28	Emergency/priority status ACAS RA Broadcast	Status value Resolution advisories Threat identity
29	Target state and status	Target altitude Target heading / track Navigation accuracy category Surveillance integrity level
31	Operational status	Airborne capacity class Surface capacity class ADS-B version number NIC supplement Navigation accuracy category Surveillance integrity level

The version number 1 or 2 can be retrieved in TC=31 messages. However, for version 0 transponder, no information regarding the version number is transmitted. In order to identify the exact version of the ADS-B transponder, we have to first check whether the aircraft is broadcasting any message with TC of 28, 29 or 31. If not, then the ADS-B version is assumed to be version 0. When a TC 31 message is received, the ADS-B version (1 or 2) can be found at the 73$^{\text{th}}$ to 75$^{\text{th}}$ bit of the message.

Knowing the Type Code, in combination with the ADS-B version, it is possible to identify the uncertainties in broadcast aircraft states. At the same time, payload information can be decoded according to the *Technical Provisions for Mode Service and Extended Squitter*, which is published by ICAO [3], [4]. Some of the ADS-B states, such as ICAO address and speed, are used for the Comm-B BDS inference in this paper.

IV. INFERRING THE COMM-B DOWNLINK PARAMETERS

Without knowledge of the original interrogation request, most of the Comm-B replies cannot be identified directly due to the lack of BDS information. It is also not possible to know whether a message is corrupted because the source aircraft ID (ICAO address) is unknown. This is because the checksum in Comm-B reply is overlaid with the 24-bit ICAO address. In some cases, it is overlaid again with the BDS code.

The two-digit hexadecimal BDS code can support up to 255 different types of messages. In practice, only a small portion of these types are interrogated. According to the European ELS and EHS mandate, the most commonly interrogated BDS codes are 10, 17, 20, 30, 40, 50, and 60. In Table III, parameters contained in these reply messages are listed.

TABLE III
COMMON MODE-S COMM-B BDS CODE AND CONTENT

Type	BDS	Content	Parameters
ELS	10	Data link capability	Transponder capabilities
	17 20	GICB capability Aircraft identification	Mode-S data capabilities Call sign
	30	ACAS resolution	Resolution advisories Threat identity
EHS	40	Vertical intention	Select altitude Barometric pressure setting
	50	Track and turn	Roll angle Track angle Ground speed Track angle rate True airspeed
	60	Heading and speed	Magnetic heading Indicated airspeed Mach number Vertical velocity

In order to decode these messages, we first need to recover the ICAO address of the source aircraft and examine whether the messages are corrupted without full knowledge of parity. In parallel, we also implement the process of identifying the BDS codes, which is used for further error detection. As a result, the decoding of a message can be accomplished.

A. Source aircraft identification and error detection

Thanks to a process called *Address Parity* [13], we can recover the ICAO addresses for most of the Comm-B messages. In the last 24 bits of the message, a parity checksum is inserted. The checksum is computed using a *cyclic redundancy check* (CRC) algorithm [14], which is a common error-detecting scheme in telecommunications.

The generator code used for Mode-S CRC encoding is specifically designed in the following binary format:

$$G = 1111111111111010000001001 \tag{1}$$

which is used to compute the checksum by the encoder and to validate the message by the decoder. In the polynomial form, the generator is expressed as:

$$G(x) = x^{24} + x^{23} + x^{22} + x^{21} + x^{20} + x^{19} + x^{18} + x^{17} + x^{16} + x^{15} + x^{14} + x^{13} + x^{12} + x^{10} + x^{3} + 1 \quad (2)$$

Similarly, the binary format of the first 88 bits of the message can also be written in the polynomial format with the highest order of 87 (x^{87}). The message polynomial is denoted as $M(x)$. The checksum is computed using polynomial division between the message and generator. By combining the checksum with parity $P(x)$, we can compute the (possible) ICAO address:

$$\begin{aligned} M(x) &= \sum_{i=0}^{87} a_i x^i, \quad a_i \in (0,1) \\ R(x) &= M(x) \ \% \ G(x) \\ A(x) &= R(x) + P(x) \end{aligned} \quad (3)$$

where $R(x)$ is the remainder (checksum) of the division $M(x)$ by $G(x)$. In Fig. 4, we illustrate this ICAO address reversal process with an example message.

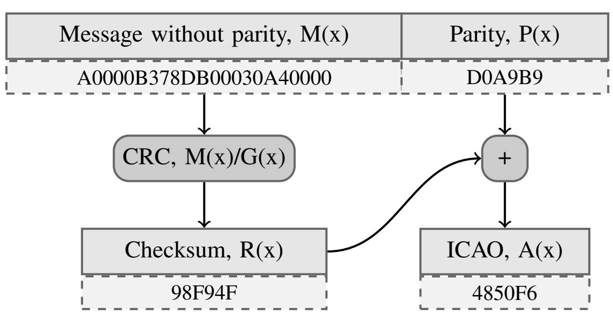

Fig. 4. The ICAO address recovery logic

After the reversal process, the resulting 24 bits would be one of the following possibilities:

- Candidate ICAO address (CA): The correct ICAO address of an aircraft.
- Modified ICAO address (MA): The ICAO address with the first 8 bits overlaid with the BDS code. [3] The parity under this condition is identified as *Data Parity*.
- Impossible ICAO address (IA): An ICAO address that is not assigned. This indicates an error in the message.

If corruption occurs in a reply, the reversal process will generate an incorrect ICAO address. This address can still be either CA, MA, or IA. Since we do not know the actual aircraft that has been interrogated, it is not possible to detect the error just using the parity. Instead, we have to identify new methods to identify corrupted messages.

Our proposed error detection mechanism consists of three checks. The first check is to identify impossible ICAO addresses. The unassigned blocks of addresses are listed in Table IV. If the resulting ICAO belongs to one of these blocks, it is likely that the message is corrupted. However, the possibility of Data Parity cannot be ruled out in this case.

[3]Only a small percentage of current aircraft to date are equipped with this capability.

TABLE IV
UN-ASSIGNED ICAO ADDRESS BLOCKS

Starting bits	Address block	Geographic region
00100	200000 - 27FFFF	Africa-Indian Ocean
00101	280000 - 28FFFF	South America
0101	500000 - 5FFFFF	Europe and North Atlantic
01100	600000 - 67FFFF	Middle East
01101	680000 - 6F0000	Asia
1001	900000 - 9FFFFF	North America and Pacific
111011	B00000 - BFFFFF	Caribbean
1101	D00000 - DFFFFF	Reserved
1111	F00000 - FFFFFF	Reserved

Next, the address is cross-validated with the ICAO addresses included in all ADS-B messages from the same time period. Since ADS-B messages can be properly error checked with the CRC process, we are confident about the obtained addresses. After this step, the correct messages are identified. An unidentified ICAO address indicates either corruption in a message or that the aircraft is not equipped with ADS-B capability.

The third error check is the ICAO-to-Squawk comparison. For Comm-B reply message with DF=21 (identity reply), the squawk code of the transponder can be obtained. By comparing the squawk code and the ICAO address, we can identify error messages, such as the ones associated with low ICAO-to-Squawk combination frequency. Since it is still possible that the message is overlaid with a BDS code, we also use the inferred BDS code to compare the overlaid ICAO address and squawk code.

Summarizing all previous steps, we can construct an error detection model, as illustrated in Fig. 5.

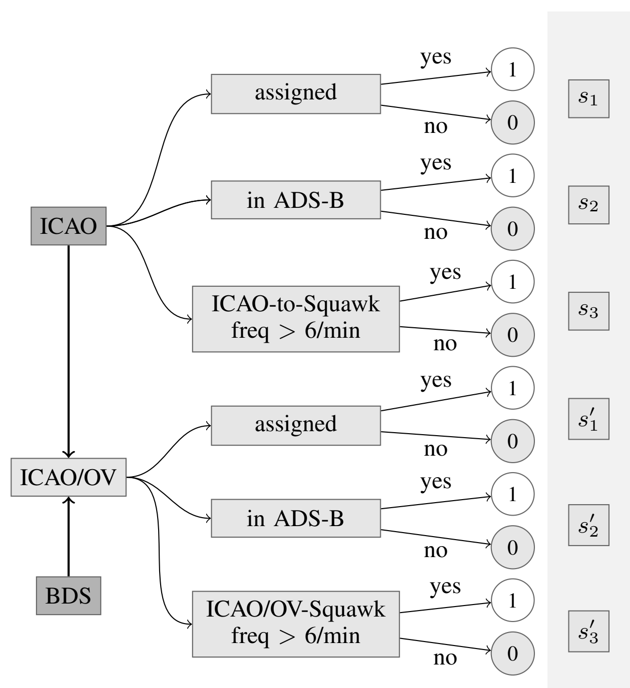

Fig. 5. EHS message error identification process

In this model, we first compute three binary scores based on the following conditions:

1) The ICAO address is assigned.
2) The ICAO address appears in the pool of ADS-B addresses.
3) The ICAO address has the ICAO-to-Squawk frequency for more than a certain (six) times per minute.

The resulting scores are denoted as s_1, s_2, and s_3 respectively. Then the inferred BDS code [4] is overlaid with the ICAO address to examine the possibility of a Modified ICAO Address. If the MA is applied, the resulting ICAO address with over-lapping BDS code (ICAO/OV) satisfies some of these conditions. Similarly, three more scores are computed, denoted as s'_1, s'_2, and s'_3 respectively. The total correctness score of S is computed as follows:

$$S = \Big[(s_1 \vee s'_1) \vee (s_2 \vee s'_2)\Big] \wedge (s_3 \vee s'_3) \tag{4}$$

where $\vee$ is the logic *OR* operation and $\wedge$ is the logic *AND* operation. The error messages are identified with $S = 0$, while correct messages are identified with $S = 1$. It is important to point out that the error model takes into consideration the situations of 1) aircraft not being equipped with an ADS-B transponder and 2) the Modified ICAO Address is used in a parity checksum.

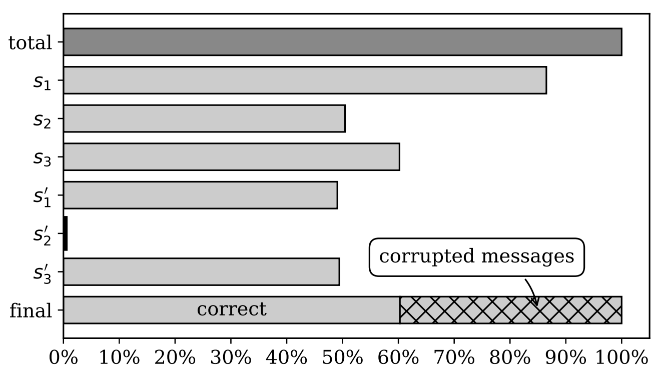

Fig. 6. Error identification statistics, based on Comm-B replies received at TU Delft from 12:00 to 13:00 UTC, September 7, 2017

Fig. 6 shows the resulting decoding statistics based on applying the error detection model to the same one-hour dataset as before. In total, 60% of the messages are identified as correct messages. The remaining 40% are corrupted. We can also see that there are only a few messages that satisfy the condition $s'_2 = 1$, which implies that the BDS overlay is not frequently requested by ATC in our airspace. In addition, only aircraft with overlay capability are able to support this feature. Such information can be found in the BDS 17 (GICB capability) messages of the aircraft.

B. BDS inference - the Heuristic-Probabilistic process

Not all aircraft have the same BDS capabilities enabled. To determine which BDS capabilities are available for an aircraft, the SSR would first initiate a Common usage GCIB capability interrogation (BDS 17). There are 24 common BDS capabilities that are reported in the BDS 17 message. SSR will only interrogate the ones that are enabled for this aircraft.

To infer the BDS code request by ATC, a two-step inference process is designed. It consists of a heuristic step and a probabilistic step. The heuristic logic is inspired and developed upon the method proposed by [9], which gives the first estimation of possible BDS codes. The probabilistic identification is introduced to identify messages with both BDS 50 and 60 codes from the previous step. Together, the Heuristic-Probabilistic (HP) process is able to deal with all common ELS and EHS replies.

1) Heuristic logic: All Comm-B payloads consist of multiple data blocks. Many of these blocks are combined with respective status bits. A status bit must be set to zero when no information is present in the data block. While evaluating the possibility of a specific BDS code, if any of the data blocks violate this rule, this BDS code is discarded.

In the structures of some Comm-B types, some bits can be reserved (or not used). These bits are required to be zeros at all times. If any of these predefined zero bits are set to one, the corresponding BDS code is also discarded.

The heuristic logic also checks parameter values that are decoded for different assumed BDS codes in a parallel fashion. These values need to be within their physical boundaries. For example, in BDS 60, the Mach number cannot be higher than one (for commercial aircraft). If a decoded Mach number is larger than one, BDS 60 is rejected as a BDS code. All evaluations are performed for each BDS code. Deterministic conditions for each parameter in all seven common BDS codes are listed in Table V.

2) Probabilistic identification: Of all available surveillance messages, BDS 50 (track and turn report) and BDS 60 (heading and speed report) messages are most frequently interrogated. However, they have very similar structures, as shown in Table V. Due to this similarity, some messages can be considered as both BDS 50 and 60 after the heuristic logic. In order to differentiate these two codes, the probabilistic step is designed.

The principle is to construct a probability density function (PDF) using current aircraft states that are observed from ADS-B data. The probabilities of BDS 50 and 60 are computed with this function. The BDS code that results in a higher probability is considered to be the correct one.

From ADS-B, we use aircraft ground speed (v_g) and track angle (χ) to construct the x and y components of the velocity. They are treated as the means for the joint probability density function. These two components, denoted as v_{gx} and v_{gy}, as calculated as:

$$\begin{aligned} v_{gx} &= v_g \cdot \sin\chi \\ v_{gy} &= v_g \cdot \cos\chi \end{aligned} \tag{5}$$

Then we construct a bi-variate normal probability density function considering these values as mean values. To simplify the problem, we assume there is no correlation between the x and y speed components, the function of the probability

[4]The BDS code is inferred in parallel using the methods described in the following section. This is possible because the deterministic BDS identification can be performed independently.

TABLE V
STATUS BIT AND PARAMETER BITS

BDS	Bits	Parameter	Logic rule	Value rule
10	1 - 8	BDS	Bits equal to 00010000	
	10 - 14	Reserved	Bits must all be zeros	
17	7	BDS 20 enabled [I]	Bit equals to 1	
	29 - 56	Reserved	Bits must all be zeros	
20	1 - 8	BDS	Bits equal to 00100000	
	9 - 56	Call sign		Only contain 0-9, A-Z, or space
30	1 - 8	BDS	Bits equal to 00110000	
	29 - 30	Threat type	Bits must not equal to 11	
	16 - 22	ACAS III		Less than 48
40	1 : 2-13 [II]	MCP/FCU selected altitude	Status and value bits consistent [III]	
	14 : 15-26	FMS selected altitude	Status and value bits consistent	
	27 : 28-39	Barometric pressure setting	Status and value bits consistent	
	40 - 47	Reserved	Bits must all be zeros	
	52 - 53	Reserved	Bits must all be zeros	
50	1 : 2-11	Roll angle	Status and value bits consistent	Between -60 and 60 (degrees)
	12 : 13-23	True track angle	Status and value bits consistent	
	24 : 25-34	Ground speed	Status and value bits consistent	Between 0 and 600 (knots)
	35 : 36-45	Track angle rate	Status and value bits consistent	
	45 : 46-56	True airspeed	Status and value bits consistent	Between 0 and 500 (knots)
60	1 : 2-12	Magnetic heading	Status and value bits consistent	
	13 : 14-23	Indicated airspeed	Status and value bits consistent	Between 0 and 500 (knots)
	24 : 25-34	Mach number	Status and value bits consistent	Between 0 and 1
	35 : 36-45	Barometric vertical rate	Status and value bits consistent	Between -6000 and 6000 (feet/minute)
	46 : 47-56	Inertial vertical rate	Status and value bits consistent	Between -6000 and 6000 (feet/minute)

[I] BDS 20 is the code that has to be enabled for all transponders to provide the minimum Mode-S capabilities.
[II] This format b1:b2-b3 indicates the status bit at b1 with value bits from b2 to b3.
[III] Consistency indicates that when status bit is zero, all value bits must also be zeros.

(the Gaussian PDF without normalization) can be expressed as follows:

$$p(v_{ax}, v_{ay}) = \exp\Big\{ -\frac{1}{2}\Big[\frac{(v_{ax} - (v_{gx} - v_{wx}))^2}{\sigma_{vx}^2} + \frac{(v_{ay} - (v_{gy} - v_{wy}))^2}{\sigma_{vy}^2}\Big]\Big\} \tag{6}$$

where v_{ax} and v_{ay} are the x and y components of airspeed. v_{wx} and v_{wy} are wind speed components. σ_{vx}^2 and σ_{vy}^2 are the variances. Accurate wind speed is not required for the purpose of BDS identification. For example, it can be obtained using weather forecast data. In case the wind information is not available or at calm wind conditions, these terms may be set to zero. However, this assumption may affect the accuracy of the identification.

For the variances, both are empirically set to be 20 knots in our experiments. This choice is based on the common magnitude of accuracy in speed which is included in ADS-B data [15]. Using Equation 6, the probabilities of BDS 50 and BDS 60 are compared. Finally, corresponding BDS code to the higher $p(v_{ax}, v_{ay})$ value is accepted. Fig. 7 shows an example of the identification based on the possible speeds decoded as BDS 50 and BDS 60.

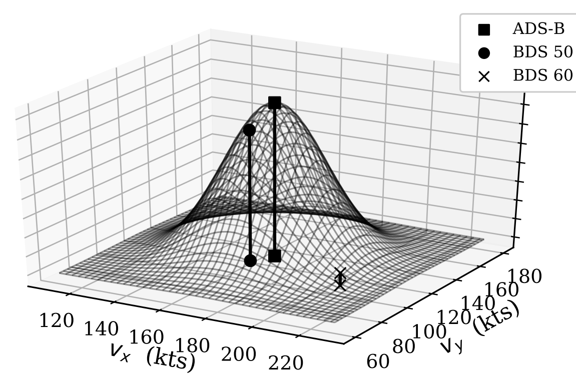

Fig. 7. Example of BDS 50/BDS 60 identification

V. EXPERIMENT AND VALIDATION

In this section, we use three different datasets to verify, validate, and analyze the Mode-S Comm-B replies. The first two datasets are from the same hour, with one collected by our receiver (located at the Delft University of Technology) and the other provided by the Air Traffic Control the Netherlands (LVNL). The last one is a global dataset provided by the ADS-B Exchange receiver network.

A. Experiment 1 - examining the heuristic logic

This experiment focuses on examining the effectiveness of the HP process of the BDS identification presented in this paper. A one-hour dataset collected by our receiver is used as a test set (from 12:00 to 13:00 UTC, September 7, 2017). The result of the heuristic logic is shown in Fig. 8. In total, 1.5 million Comm-B replies are received during this one-hour period. The most common three BDS codes are 40, 50, and 60. In each bar plot of Fig. 8, the hatch pattern marks the corrupted messages.

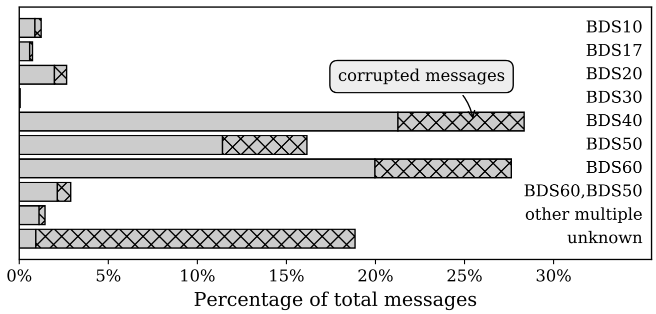

Fig. 8. Heuristic BDS detection, based on Comm-B replies received from 12:00 to 13:00 UTC, September 7, 2017, Delft, the Netherlands

In this figure, we can see that around 20% of the messages are not identified (with unknown status). It is also noticeable that around 90% of these unidentified messages are corrupted.

For all other messages with a BDS code identified, BDS 10, 17, and 20 combined represent around 5% of the total number of messages. BDS 40 and 60 account for about 27% each, and BDS 50 around 16%. Due to the similar message structures, around 4% of the messages identify as both BDS 50 and 60 codes. Less than 0.1% of messages are BDS 30. Since BDS 30 encodes emergency status and ACAS resolutions, it is expected that this category is less frequently interrogated during normal operations.

B. Experiment 2 - identification accuracy

To examine the accuracy of the HP process for BDS identification, we obtain a true reference dataset of the Dutch airspace from the Air Traffic Control the Netherlands (LVNL). This data is collected in the same hour as the previous test dataset. The original data is presented in ASTERIX format, from which we extract the raw messages. As such, raw messages only contain messages that are of BDS 40, 50 and 60 (Enhanced Mode-S only). Raw messages for other BDS types are not included in this ASTERIX dataset.

After applying both heuristic and probabilistic steps, we compute the accuracy at each stage. The results are shown in Table VI. In total 253,059 messages are used. Unidentifiable messages refer to the ones with multiple BDS codes.

TABLE VI
VALIDATION STATISTICS OF MODE-S BDS CODE IDENTIFICATION USING PROPOSED *HP process*

Result	Heuristic logic	Probabilistic id
Correct	235,772 (93.169%)	247,192 (97.682%)
Unidentifiable	17,281 (6.829%)	5847 (2.311%)
Incorrect	6 (0.002%)	20 (0.008%)

Using only heuristic logic, we achieve a correctness rate of 93.2% for a total of 250 thousand messages. 6.8% of the messages have been identified with more than one BDS code. Using additional probabilistic identification, we can further reduce the rate of uncertain messages to 2.3%, and increase the success rate to 97.7%.

In each step, the number of incorrectly decoded messages is almost negligible. We have counted that only 6 (heuristic step) and 20 (probabilistic step) out of 253,059 messages were incorrectly identified. Regarding the significance of the probabilistic step, we found that among all 11,434 identifiable messages (with BDS 50 and BDS 60 combination) only 14 errors occurred. Upon further investigation, we notice that errors happen often when part of the flight information is missing from the message.

C. Experiment 3 - analyze the global data

Using multiple receivers provided by the ADS-B Exchange receiver network, we are able to examine the decoding performance of Mode-S Comm-B data across different regions of the world. A set of data consisting of one hour (local time 10:00 AM to 11:00 AM, on 24 or 25 August of 2018) is collected at eight different locations around the world. Using the method proposed in this paper, we decode the messages and show the composition of Mode-S responses for these different regions. In Fig. 9, the percentages of BDS codes are illustrated.

Mode-S data transmitted to ADS-B Exchange network is filtered by default. Only messages with s_2 state in Fig. 5 are kept. These are Comm-B messages with ICAO addresses that appear in ADS-B data. In Table VII, the exact percentages of BDS codes in Comm-B responses are listed. In all eight regions, the unidentifiable messages are between 1% to 6% of received messages. This result is in line with the test performed using our own receiver after corrupted messages are discarded (as shown in Fig. 8). In most of these regions, common interrogated BDS codes are BDS40, BDS50, and BDS60, with the exception of Dallas in the USA and Tel Aviv in Israel. The diversity of interrogations is one of the most challenging elements in the inference process. However, the proposed methods in this paper are able to cope with interrogation variations and produce a large percentage of identifiable messages.

VI. Discussions

This study aims at efficiently and accurately making use of open aircraft surveillance data. The focus has been on designing a path to decipher Comm-B replies that are part of the Mode-S enhanced surveillance, and cannot be decoded directly. The experiments and validations show a high accuracy of the identification process and error detection strategy that are proposed in this paper. These refined identification methods provide the possibility to conduct aircraft performance studies with better accuracy when using open aircraft surveillance data.

Accurate identification of BDS 50 and 60 enables precise airspeed observations from Comm-B data. One possible extension of this research is its potential contribution to atmospheric modeling. By combining ADS-B ground speed with Mode-S true airspeed, wind and temperature can be computed. With globally around ten thousand aircraft airborne at any given

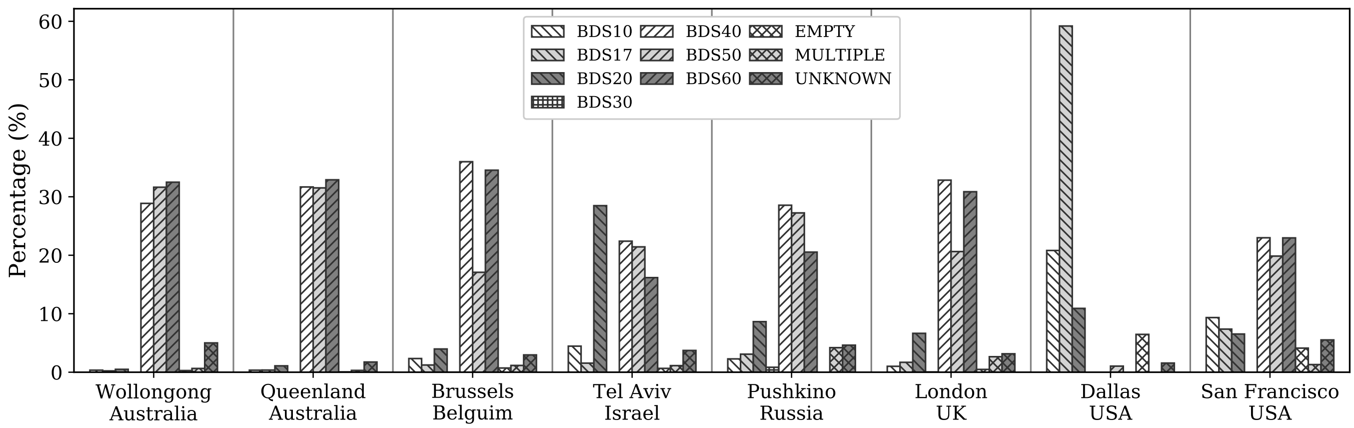

Fig. 9. Statistics of BDS codes in Comm-B replies at different geographic regions

TABLE VII
MODE-S COMM-B RESPONSE SUMMARY IN DIFFERENT REGIONS

Country	City	BDS10	BDS17	BDS20	BDS30	BDS40	BDS50	BDS60	EMPTY	MULTIPLE	UNKNOWN
Australia	Queenland	0.3%	0.4%	1.1%	-	31.7%	31.5%	32.9%	0.1%	0.3%	1.7%
Australia	Wollongong	0.3%	0.2%	0.5%	-	28.9%	31.6%	32.5%	0.3%	0.6%	5.0%
Belguim	Brussels	2.3%	1.2%	4.0%	0.0%	36.0%	17.1%	34.5%	0.7%	1.2%	3.0%
Israel	Tel Aviv	4.5%	1.6%	28.5%	-	22.4%	21.4%	16.2%	0.7%	1.1%	3.7%
Russia	Pushkino	2.3%	3.1%	8.6%	0.8%	28.6%	27.2%	20.5%	-	4.2%	4.6%
UK	London	1.0%	1.7%	6.6%	0.0%	32.8%	20.6%	30.9%	0.5%	2.6%	3.1%
USA	Dallas	20.8%	59.2%	10.9%	-	-	1.0%	-	6.5%	-	1.6%
USA	San Francisco	9.3%	7.4%	6.5%	-	23.0%	19.9%	23.0%	4.1%	1.3%	5.5%

time, a large number of meteorological measurements can be provided. This idea has been proposed in earlier research [16], where data are supplied by air traffic controllers. The methods and tools from this paper allow anyone to gain such ability. For example, in our earlier study [17], wind and air temperature were derived based on this open data, which demonstrates the use of aircraft for weather observations.

Fig. 3 and 6 in earlier sections show the percentage of corrupted messages found in this study. This strongly suggests a frequency congestion problem on the 1090 MHz channel. Of all the ADS-B and Comm-B messages that we received, more than half are corrupted. Based on the percentage of errors, severe message corruption during daily operations is shown. In Fig. 10, the percentage of corrupted ADS-B messages and the number of aircraft flying in a 24-hour period are plotted.

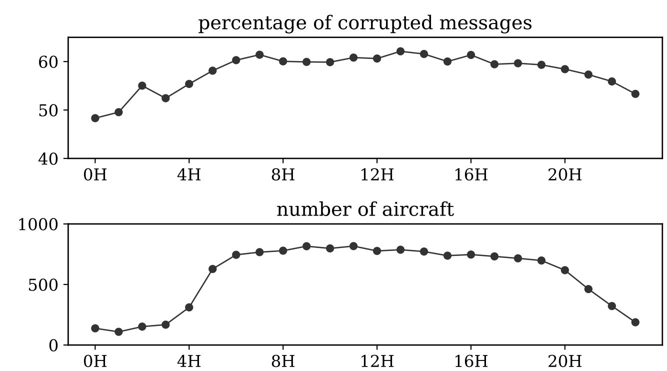

Fig. 10. Percentage of corrupted ADS-B messages and number of aircraft per hour on May 30, 2018, in Delft, the Netherlands

It can be seen that during the night time, when the airspace is less saturated, the percentage of corrupted messages decreases. However, the decrease in corrupted messages ($\sim$10%) is not proportional to the decrease in the number of flights ($\sim$65%). These corrupt messages not only present a challenge for obtaining more accurate data but also indicate a constant frequency congestion in busy airspace.

Other than third-party researchers, air traffic controllers may also benefit from the methods proposed in this paper. With a high level of frequency congestion as shown in Fig. 8, air traffic controllers tend not to over-interrogate. This creates a dilemma where an ATC has to reduce the number and frequency of interrogations, but at the same time, more information is preferable for making better traffic control decisions. Using the proposed BDS identification process, one ATC can intercept replies that originate from interrogations by another ATC center. As one example, based on the one-hour reference dataset, we found that the replies from one specific secondary radar from the air traffic control of the Netherlands account for only around 17% of the total number of interrogation replies. The proposed identification method can therefore significantly increase the amount of information available to each ATC directly, even for aircraft that are outside of their airspace.

Finally, it is worth emphasizing that Mode-S Comm-B messages are passive replies originating from SSR interrogations.

For oceanic and remote areas where secondary surveillance radars are not present, only ADS-B data is available. In this paper, not all BDS codes are addressed. Instead, only the types related to ELS and EHS are investigated. However, this is the situation for many regions around the world as illustrated in Fig. 9 using the global Mode-S data.

VII. Conclusions

In this paper, we propose a set of inference methods that allow third-party observers to identify and decode Mode-S Comm-B replies. The inference is based solely on surveillance replies without any knowledge of Mode-S interrogation. This paper contributes to a missing area of knowledge on handling interrogation-based surveillance data. It gives researchers broader access to accurate aircraft state updates that are transmitted through Enhanced Mode-S. The implementation is based on existing low-cost commercial off-the-shelf ADS-B receivers, with no additional hardware. Using a reference dataset, the proposed process reveals a correctness rate of around 97.7%, with 2.3% unidentifiable and 0.008% error for Enhanced Mode-S messages.

Furthermore, the process proposed in this paper is also likely to be beneficial for air traffic controllers, since it enables the ability to collect more data without the need for increasing interrogation frequency. Finally, we have made the decoding process into an open-source programming library, *pyModeS*, which includes the decoding and inferences discussed in this paper. We hope this will also enable other researchers to make use of this valuable aircraft surveillance data source.

VIII. Acknowledgments

We would like to thank Air Traffic Control the Netherlands (LVNL) for providing us the Enhanced Mode-S validation dataset. We would also like to thank James Stanford from ADS-B Exchange for providing the data for global data analysis.

References

[1] J. Sun, J. Ellerbroek, and J. M. Hoekstra, “Aircraft initial mass estimation using bayesian inference method,” *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 59–73, 2018.

[2] J. Rudnyk, J. Ellerbroek, and J. Hoekstra, “Trajectory prediction sensitivity analysis using monte carlo simulations,” in *2018 Aviation Technology, Integration, and Operations Conference*, p. 3669, 2018.

[3] ICAO, *Technical Provisions for Mode S Services and Extended Squitter*. International Civil Aviation Organization, 2008.

[4] ICAO, *Technical Provisions for Mode S Services and Extended Squitter, 2nd Edition*. International Civil Aviation Organization, 2012.

[5] R. Trim, “Mode s: an introduction and overview (secondary surveillance radar),” *Electronics & Communication Engineering Journal*, vol. 2, no. 2, pp. 53–59, 1990.

[6] R. Koelle, “Open source software and crowd sourced data for operational performance analysis,” in *Proceedings of the 12th ATM Seminar*, 2017.

[7] Z. Wang, M. Liang, and D. Delahaye, “Short-term 4d trajectory prediction using machine learning methods,” in *SID 2017, 7th SESAR Innovation Days*, 2017.

[8] R. Kaune, C. Steffes, S. Rau, W. Konle, and J. Pagel, “Wide area multilateration using ads-b transponder signals,” in *Information Fusion (FUSION), 2012 15th International Conference on*, pp. 727–734, IEEE, 2012.

[9] S. de Haan, M. de Haij, and J. Sondij, “The use of a commercial ads-b receiver to derive upper air wind and temperature observations from mode-s ehs information in the netherlands,” Tech. Rep. TR-336, KNMI, the Netherlands, 2013.

[10] R. D. Grappel, G. S. Harris, M. J. Kozar, and R. T. Wiken, “Elementary surveillance (els) and enhanced surveillance (ehs) validation via mode s secondary radar surveillance,” *Project Report ATC-337, Lincoln Lab., MIT*, 2008.

[11] P. Noschese, S. Porfili, and S. Di Girolamo, “Ads-b via iridium next satellites,” in *Digital Communications-Enhanced Surveillance of Aircraft and Vehicles (TIWDC/ESAV), 2011 Tyrrhenian International Workshop on*, pp. 213–218, IEEE, 2011.

[12] ICAO, *Annex 10 to the Convention on International Civil Aviation, Aeronautical Telecommunications*. International Civil Aviation Organization, 2002.

[13] J. L. Gertz, “Fundamentals of mode s parity coding,” tech. rep., Massachusetts Institute of Technology, Lincoln Laboratory, 1984.

[14] J. S. Sobolewski, “Cyclic redundancy check,” 2003.

[15] RTCA, “Minimum operational performance standards for 1090mhz extended squitter automatic dependent surveillance-broadcast (ads-b) and traffic information services-broadcast (tis-b)[j],” *RTCA DO-260B with Corrigendum*, vol. 1, no. 1, pp. 1365–1372, 2011.

[16] S. de Haan, “High-resolution wind and temperature observations from aircraft tracked by mode-s air traffic control radar,” *Journal of Geophysical Research: Atmospheres*, vol. 116, no. D10, 2011.

[17] J. Sun, H. Vû, J. Ellerbroek, and J. M. Hoekstra, “Weather field reconstruction using aircraft surveillance data and a novel meteo-particle model,” *PloS one*, vol. 13, no. 10, p. e0205029, 2018.

Junzi Sun is a Ph.D. candidate at the Delft University of Technology (TU Delft). He received his M.Sc in Aerospace Engineering from the Polytechnic University of Catalonia (UPC). His doctoral research topic is constructing open aircraft performance models using open data. His main research focuses are aircraft performance modeling, ADS-B and Enhanced Mode-S surveillance, trajectory analysis, machine learning, data mining, and statistical inferences.

Huy Vû received his M.Sc. Aerospace Engineering from the Delft University of Technology in 2018 for his research in improving aircraft performance models by using Mode-S data. In 2016, he received his M.Sc. Econometrics and Operations Research from Maastricht University. Currently, he is working as a data analyst at Air Traffic Control the Netherlands (LVNL).

Joost Ellerbroek received the M.Sc. (2007) and Ph.D. (2013) degrees from the Delft University of Technology, The Netherlands. He is currently an assistant professor with the CNS/ATM chair at the faculty of Aerospace Engineering of TU Delft. His current work includes, among other topics, urban airspace design for drones, topics of airspace complexity and capacity analysis, data mining applications for ATM, analysis of future ATM concepts such as the extended arrival manager, and new approach procedures, separation algorithms for UAS, and several other ASAS-related studies.

Jacco M. Hoekstra is a full professor at the faculty of Aerospace Engineering of the Delft University of Technology. After obtaining his M.Sc. and pilots license in 1990, and a brief excursion in telecommunications, he started working at the National Aerospace Laboratory (NLR). He then obtained his Ph.D. on Air Traffic Management and headed the NLR Air Transport Operations division, before joining the TU Delft faculty. He served two terms as a dean and now heads the CNS/ATM chair and holds a seat at the administrative board of SESAR.