



Fig. 2. Current deployment of OpenSky’s sensors. The sensors are operated by volunteers.

particular efforts should concentrate on validating ADS-B data against other technologies in order to improve inaccuracies.

History has shown in many fields that the validation of new technologies is best achieved when many independent research groups work in parallel in order to identify potential weaknesses and propose improvements. In the case of ADS-B, access to large-scale real-world data has only been possible for a few selected industrial and governmental groups so far. While several live radar visualization services based on ADS-B are available on the Internet, they do not provide the raw data that is most valuable for researchers. For that reason, we have developed OpenSky, an open sensor network for research. OpenSky collects and stores all ADS-B traffic as it is being captured by sensor nodes distributed over a large area.

We have started by deploying 11 sensor nodes in Central Europe (see Fig. 2). The sensor network relies on volunteers who deploy sensors at their homes and share the collected data over the Internet. The network is still growing as more volunteers contribute to the system by adding new sensor nodes. OpenSky relies on low-cost commercial equipment, which lowers the barrier of entry for participants. We have been operating the network now for almost two years collecting billions of ADS-B messages for analysis. In its current deployment, the sensing range of OpenSky covers 720,000 km² and is able to capture more than 30 % of the total commercial air traffic in Europe. All collected data is made accessible to the volunteers who contribute with their sensors, and to anyone else on request.

A. Contributions

- We design OpenSky, a low-cost participatory sensor network that collects and provides access to raw ADS-B data. OpenSky records all messages as they are received by sensor nodes, including nanosecond precision time stamps for time-critical evaluations that require tight synchronization.
- To the best of our knowledge, OpenSky is the first open ADS-B sensor network that provides researchers with access to raw ADS-B data for real-time data analysis and arbitrary off-line analysis based on archived data.

- We report on our challenges and lessons learned while deploying, operating, and working with OpenSky over the last two years. During this time, six groups have been using the data for various research projects covering aspects such as security, performance, and applications of ADS-B.
- Based on an extensive data set provided by OpenSky, we evaluate the ADS-B communication channel, characterizing typical reception quality and loss patterns in the real world. We give insights that are relevant for most research and the planned adoption of ADS-B in the future.
- We evaluate the feasibility of performing physical location validation with multilateration based on a low-cost infrastructure such as OpenSky. Even though OpenSky consists of cheap off-the-shelf sensors, we demonstrate that multilateration is able to achieve a localization accuracy with a median error in the horizontal plane of 166 m and a mean of 296 m. This can also be considered as an experimental analysis of OpenSky’s data and sensor quality.

II. APPLICATIONS

During two years of operating OpenSky, we have started working with the data in different ways. To provide an idea about what the data collected by OpenSky can be used for, this section outlines several examples of research applications.

1) *Error and fault diagnosis:* The detection of errors and faults in ADS-B data is important for several reasons. Most importantly, a reliable and fast detection of bad transponder behavior such as odd position reports or message rates can help to improve the overall safety of ADS-B. OpenSky may help to discover misbehaving and erroneous transponders which do not comply with the standard. This way, the chance that safety-related issues are detected prior to wide-scale adoption is increased, since an entire research community may start diagnosing the problem.

2) *Performance evaluation:* Monitoring the air space and the communication channel over longer periods at different locations provides better insights on how the 1090 MHz communication channel performs over time and space. OpenSky may help to assess protocol performance such as the message loss rate or the number of collisions at various locations and times. Thus, we can identify bottlenecks early and apply countermeasures that improve the system capacity.

3) *Data validation:* Since ADS-B depends on the location estimation of the aircraft, it is important to validate that the claimed data in the advertised messages are precise and accurate. Data validation is particularly important for safety reasons, as wrong location estimates may cause situations in which pilots or air traffic controls are confused and route aircraft towards collisions.

4) *Multilateration:* A practical application we show in this paper is multilateration. Multilateration provides additional means for ground-based data validation, as the position may be estimated independently on the satellite localization system of the aircraft. The current deployment of OpenSky allows us to perform wide-area multilateration when the density of the sensors is large enough that the signals from an aircraft are received by multiple sensors.

nication. The uplink is used for interrogations and information services (ground-to-air) while replies and broadcast messages from aircraft are sent via the downlink (air-to-ground). The ICAO specifies two format types for Mode S [5]: short formats with a length of 56 bit and long formats with 112 bit. Long formats include a generic type for broadcasting unspecified data, the so-called Extended Squitter (ES). ES are 112 bit messages providing a 56 bit field that can be filled with arbitrary data. This type is used by ADS-B Out to broadcast messages. The messages are modulated using pulse position modulation (PPM), since PPM is relatively robust against interference and collisions. The modulation and the ES format are depicted in Fig. 4. With respect to the Mode S downlink frequency, the combination of ADS-B and Mode S is also referred to as 1090 ES ADS-B [6]. The protocol and standard hierarchy are depicted at the bottom right corner of Fig. 3.

B. Deployment status

As ADS-B will be mandatory by 2020 in most airspaces, its deployment is in full swing. In fact, the major airlines have already upgraded their fleet and more than 50% of all aircraft support at least rudimentary ADS-B (see Section VI-A for more information). ADS-B is still in the evaluation phase, however, and data provided by the system is not certified and therefore not yet used for air traffic management. In fact, some of the aircraft that broadcast position reports are not even equipped with GNSS sensors and determine their position with less accurate means. This can lead to large errors in some ADS-B position reports and has to be considered when working with this data.

IV. THE OPENSKEY SENSOR NETWORK

OpenSky is a participatory sensor network of ADS-B sensors distributed in Central Europe. The sensors are given to volunteers who deploy them at their homes or organizations.

The goal of OpenSky is to collect and store all ADS-B messages in our reception range for further analysis. There are already other community-based projects using ADS-B data; live radar services freely available on the Internet (e.g., Flight-radar24) offer extensive coverage of world-wide air traffic. However, while these services are able to provide live insights about aggregated flight tracks and abstract information, they do not offer access to the raw historical data that is crucial for research.

While the data collected by OpenSky is not yet publicly accessible due to bandwidth constraints, access is granted to all volunteers and upon request.

A. Architecture

The system architecture of OpenSky is depicted in Fig. 5. The sensors are equipped with an RF interface and an ADS-B decoder that allows the reception of ADS-B messages broadcasted on the 1090 MHz Mode S downlink. On the other end, the sensors have a network interface accessible over the Internet. Once an Internet connection to the sensor is established, it starts forwarding all received ADS-B messages over this link. The main advantage of this design is that the sensors are not responsible for keeping the Internet connection

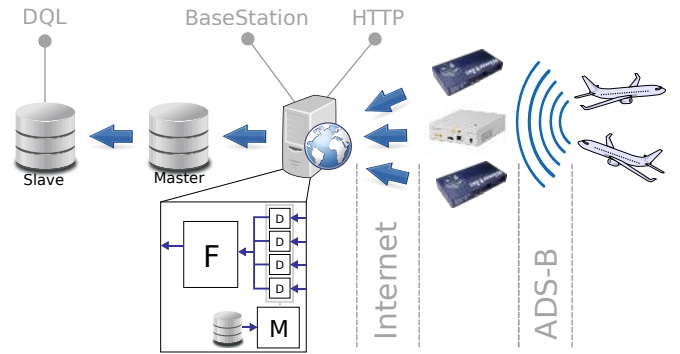


Fig. 5. Overview of OpenSky’s architecture. ADS-B messages are received by sensors deployed by volunteers. The data is collected via the Internet and stored in a central database.

alive and do not have to deal with common problems such as reconnecting after a failure (see Section V).

A central server is responsible for managing the sensors, collecting the received ADS-B messages, and storing all received information in a database. To keep the system flexible and to support different kinds of sensors, the interface to the sensors is implemented by driver modules (D in Fig. 5), which hold the connection to the sensor and transform the incoming message streams from the sensors into a unified data format.

The drivers are managed by a manager module (M), which holds a list of all sensors that are currently deployed, their IP addresses (or hostname) and port, as well as information on the volunteer who deployed the sensor. The manager is responsible for restarting the drivers in case of failures (e.g., broken TCP connection to the sensor). It therefore checks their alive state periodically. It also offers a web (HTTP) interface for adding, removing, activating, or deactivating sensors.

The unified data output of the drivers is processed by a central data fusion unit (F). This unit decodes each message according to the ADS-B standard and generates the respective database queries to store the decoded messages as well as meta data in a database. Since the collected sensor data is intended for research, the database is built after the immutability paradigm: nothing gets updated or changed, new data is simply attached [7]. This, in practice, means that we store every single message, which allows us to reconstruct the state of our data at any given point of time in the past. This also means that changes to the basic message tables are not allowed and abstractions have to be done locally or in separate tables. Besides handling the database capabilities, the data fusion unit also implements an interface to Kinetic Avionics’ freely available BaseStation software. This enables users to visualize the live picture of OpenSky in a radar-style display. An example of the live picture is shown in Fig. 1.

We finally replicated the database and only the slave offers an interface for querying the data. That is for two reasons. First, separating queries by researchers from the insertions by the data fusion unit helps reduce conflicts based on read or write locks. Second, by shifting the workload caused by complex queries to another machine, the threat of the database becoming a bottleneck is mitigated. The sensor network can still collect data without interruption, even when the slave is overloaded. All of OpenSky’s modules are implemented using Python. The driver manager’s database is an SQLite database

while the data provided by OpenSky’s sensors is currently stored on a MySQL database server.

B. Supported sensors

OpenSky currently supports two different types of low-cost sensors: Kinetic Avionics’ SBS-3 station and an ADS-B receiver based on Ettus Research’s software-defined radio USRP. SBS-3 is a commercial all-in-one solution that supports the reception of aircraft data as well as voice communication out of the box. The use of an open protocol to stream the received data over a network and the possibility for remote configuration make it a good fit for OpenSky.

The USRP-based sensor is more complex but provides much more information than the SBS-3. The receiver is implemented in software based on Nick Foster’s Mode S receiver³ for the signal processing framework GNU Radio. We extended this receiver with the capability to additionally measure signal properties such as the received signal strength (RSSI), correlation factor (confidence level) and the signal-to-noise ratio (SNR). This enables researchers to analyze the physical properties of the ADS-B protocol. A glimpse of what is possible with this data is given in Section VI-B.

However, a big disadvantage of USRP-based receivers is the fact that it requires an additional desktop computer running the software receiver and has a much higher energy consumption.⁴ Mainly due to these disadvantages our network consists primarily of SBS-3 stations. The USRP-based receiver is only deployed *ad hoc* for planned experiments.

C. Collected data

As mentioned in the previous section, ADS-B messages are not simply stored as they are delivered by the sensors. The data fusion unit also implements an ADS-B decoder, which can interpret most formats defined by the ADS-B standard. The state vector of each aircraft or vehicle as defined in [2, §3.4.3] is fully supported and stored in our database. It comprises the following information:

- **Identification:** Each ADS-B message contains the transponder’s unique 24bit address which is assigned by the International Civil Aviation Organization (ICAO). Additionally, most aircraft broadcast their call sign, which is an 8-character string and typically assigned by the airline itself.
- **Position:** Aircraft report their position twice per second. The three-dimensional position is stored as decimal latitude, longitude, and altitude. The accuracy of the position depends on the accuracy of the aircraft’s GNSS sensors and the data format. 1090 ES ADS-B currently supports altitude encoding with an accuracy of 25 ft. For a detailed description of the position encoding algorithm CPR and information on its accuracy, please refer to Appendix T of [6]. The accuracy of onboard sensors is provided by the navigation accuracy category (NAC) field of status reports and is typically bound to 92.6 m or 30 m for horizontal positions.
- **Velocity:** The aircraft’s velocity is reported in east-west and north-south velocity and the vertical rate. The encoding

of both horizontal velocities has a granularity of 4 knots while the vertical rate is given in 64 ft/min steps. Again, the accuracy also depends on the underlying sensors used to determine the values. In current practice, however, we observe that the NAC field is set to “unknown” on most transponders.

In addition to the state vector, some aircraft also broadcast status messages that contain information on emergencies, priority, capability, navigation accuracy category, and operational modes.

Besides its content, we also store metadata for each message. This includes a 50 ns rolling timestamp of the reception provided by the SBS-3, a local timestamp that indicates the time when the message was received by the data fusion unit, the receiving sensor’s ID, the ADS-B checksum, and the raw message as a hex string.

We moreover provide an initial abstraction of the data by separating messages from any aircraft into flights. Based on empirical tests, a single flight starts with the first message received when an aircraft enters the sensing range and ends ten minutes after the reception of the last message of this aircraft. Ten minutes have proven to provide a good balance for two typical situations. First, if an aircraft lands within the reception range of OpenSky and departs again shortly after, a threshold longer than the period between landing and departure would falsely join the two flights into one. Secondly, some aircraft leave our reception range at one end and re-enter somewhere else after some time without landing in between. A threshold shorter than this time would incorrectly split such a flight into two flights. However, ten minutes turned out to be a good trade-off for the current structure of OpenSky. If the reception range of OpenSky was separated into multiple parts with large blind gaps in-between, a longer threshold or a more sophisticated definition for flights would be necessary.

D. Data access

The data is accessible via two interfaces. As mentioned above, the data fusion unit implements an interface that enables access to the live picture. The interface implements the protocol specified in [8] and can be accessed for example with Kinetic Avionic’s freely available BaseStation software.

The data stored in the database is accessible via a database dependent query language (QL). The database is currently implemented with Oracle’s MySQL and only the `SELECT` and `SHOW` queries are granted to users to avoid violations of the immutable principle. Furthermore, the QL interface is only accessible on the replicated slave database to avoid read-write conflicts and take load off the crucial master database.

V. CHALLENGES

Besides introducing OpenSky, we would like to provide readers with guidance in building similar networks and help to assess whether participating in OpenSky is reasonable for them or not. Therefore, we summarize several hurdles we encountered during the development of OpenSky, the deployment of the sensors, and the past two years of operation. We distinguish between deployment, system, and research challenges.

³<https://github.com/bistromath/gr-air-modes>

⁴The SBS-3, on the other hand, consumes only about 2 W.

A. Deployment challenges

1) *National regulations*: In some countries regulatory constraints exist that need to be taken into account. In the United Kingdom for example, it is prohibited to “use a wireless telegraphy apparatus with intent to obtain [...] contents, sender or addressee of a message” without being or acting on behalf of the intended receiver.⁵ Such regulations necessitate further measures such as special permissions or artificial delaying of messages. In fact, SBS-3 devices currently induce an artificial delay when the data is streamed to an unauthorized application. However, legal requirements have to be checked independently before deploying sensors.

2) *User-friendly deployment*: It is crucial to a participatory sensor network that the deployment is as user-friendly as possible since the expertise of volunteers varies. Therefore, we created a manual that explains the steps required to connect a new sensor to the network and shifted as much of the complexity as possible to the server side. The SBS-3 station turned out to be sufficiently user-friendly since it does not require any configuration besides the assignment of a static IP address, which can be done by an expert in advance. The only hurdle for end-users is that the sensor must be accessible via the Internet, requiring IP forwarding in most common network setups. However, recent firmware updates enable the SBS-3 to act in client mode, which resolves this problem in the future.

3) *Antenna position*: Finding a good position for the antenna is crucial for the reception range. The reception of ADS-B strongly depends on a good line-of-sight (LOS) connection. Walls attenuate the signal enough to reduce the reception range dramatically. Although many of our volunteers were highly motivated and deployed their boxes outdoors, volunteers are typically not willing to drill holes in their walls to run the antenna cable outdoors. While placing the antennas near windows helps increase the range, there is no perfect solution for this problem. If the reception range is still insufficient, additional sensors should be deployed in this area.

4) *Volunteer’s bandwidth*: Data rates generated by the sensors range from less than 10 kB/s to almost 1 MB/s depending on the altitude of the antenna, LOS conditions, and density of the nearby airspace. The Internet bandwidth available at volunteers homes is usually high enough to handle that traffic. However, especially in rural areas and locations further away from Internet hubs (e.g., in the Alps), Internet connections mostly have low bandwidth. These positions are nevertheless attractive locations for OpenSky. Combined with high message rates at locations with large reception ranges or dense airspaces, slow Internet connections may congest, which causes high delays or even message loss. Even though this is a very rare combination, we have encountered such a case. In these circumstances, rearranging the antenna to a less optimal position (e.g., indoors) helps reduce the data rate at the cost of smaller coverage.

B. System challenges

1) *Internet Effects*: Broken TCP connections turned out to be a frequent problem. Sometimes the TCP/IP implementation of the kernel notices and reports disruptions to the application

layer, sometimes it does not. One case that occasionally leads to ghost sockets is when Internet providers cut the Internet connection. It is in fact a common practice of providers to enforce a reconnect once per night. However, periodically checking the timestamp of the last message that a sensor received helps identify dead links. Since automatically restarting the driver solves the problem, OpenSky’s driver manager checks each connection every three seconds.

2) *Database performance*: The enormous number of messages stretched MySQL to its limits shortly after starting the project. While inserting messages is still fast, the analysis becomes very inefficient. Replicating the database on a dedicated server for analysis and creating appropriate indices mitigated this problem. Based on our experience, indices on time and flights are very helpful for many applications. They even allow use of the database for real-time applications. For example, querying all reported positions over the last 60s takes 0.01s even though the database contains more than 4.1 billion messages. Of course, the performance is expected to deteriorate as more sensors are added and more data is collected.

3) *Logging*: As in most sensor networks, sometimes sensors go offline for a while and go back online after some time. For later analyses of the data, it is important to be able to determine whether a lack of data from a sensor is due to failure of the sensor or due to missing aircraft in range. Therefore, proper logging of all events must be ensured.

C. Research challenges

1) *Node synchronization*: One of the traditional research challenges in distributed systems is the synchronization of participants’ clocks for time-critical applications. There are many practical obstacles in achieving a very tight and accurate synchronization between the sensors used in OpenSky. Low-cost boxes such as SBS-3 do not offer GPS as a feature (although we will consider this in the future) and their strongly drifting internal clocks make a frequent re-synchronization necessary. A centrally coordinated synchronization via a network time server would be considerably affected by the jitter of non-dedicated Internet connections, rendering the achievement of the required accuracy in the low nanoseconds very challenging. Lastly, even when a good synchronization has been achieved, the timestamps are limited by the precision of the internal clocks. See section VI-C for more details on synchronization.

2) *Coverage planning and sensor placement*: There are different aspects to planning the coverage of OpenSky. While at the moment we are constrained by the location of the volunteers we recruited, node density and coverage requirements may become an important point of future research. ADS-B is specified to work at distances of 120 NM and more in the en-route airspace (usually at an altitude of about 30,000 ft). Here, monitoring is easier since the higher the aircraft’s altitude, the larger the area on the ground with a good LOS connection. Indeed, we are able to receive messages by aircraft near the radio horizon, far exceeding the specifications. However, this is only possible under optimal conditions. In more demanding circumstances, more and/or better placed receivers are required, especially if multilateration should serve as a backup system. Examples of such conditions are the flight phases near

⁵See Section 48 of the UK Wireless Telegraphy Act of 2006.

Sensors	11 (SBS-3)
Received messages	>4,000,000,000
Total number of flights	>1,250,000
Unique aircraft	>13,200
Total size of MySQL DB	~800 GB
Covered area	~720,000 km ²
Flights per day	7,000–7,500
Messages per day	>20,000,000
Total network throughput	~1±0.5 MB/s

TABLE I. OPENSky STATISTICS AS OF JANUARY 2014.

airports such as take-off, initial climb, approach, and landing as well as challenging topography such as the Alps. As cost is also an important factor in any ADS-B system, OpenSky can help with the construction of optimal sensor deployments by providing specific coverage, reception quality and loss data for future simulations and coverage planning.

3) *Data validation*: A further issue is the current deployment status of ADS-B. As mentioned above, some aircraft are not yet equipped with GNSS sensors or have defect transponders reporting incorrect positions and wrong (even negative) altitudes. Additionally, the position decoding algorithm used in ADS-B (CPR) has difficulties handling positions close to the radio horizon. We extended the algorithm with filters that detect impossible locations and try to recover the correct position by using both the global and local version of CPR in combination. Nevertheless, some noise remains in the data and must be considered during any analysis. Therefore, means are required to validate the data and to determine the quality of the reported positions. This can be done by comparing ADS-B data with other information sources such as public registers, flight information systems, and multilateration. As we will describe later, multilateration can be performed using OpenSky.

VI. EVALUATION

Based on data collected with OpenSky, this section gives an overview of OpenSky’s performance and provides first insights on the 1090 MHz ADS-B channel as it is perceived by the sensor network. We furthermore implemented wide-area multilateration to demonstrate and analyze the opportunities and limitations of OpenSky. Our results are presented in the following sections.

A. Flight statistics

OpenSky has currently collected more than 4 billion ADS-B messages (see Table I). Summarizing these data, we have seen more than 13,200 different aircraft from over 100 different countries. The majority of aircraft were from Germany (19.03%), the United Kingdom (11.58%) and the United States (11.49%), followed by Switzerland (6.63%), Ireland (6.55%) and France (6.23%). In sum, these six countries made up almost two thirds of all aircraft crossing our sensing range, with up to 7,500 flights doing so every day. Although not all aircraft are using ADS-B at this time (around 50%), this already comprises about 30% of the current flight traffic in Europe, where EUROCONTROL records between 25,000 and 30,000 flights per day [9].

Among those aircraft that are equipped with an ADS-B transponder, the implementation of the standard is still very

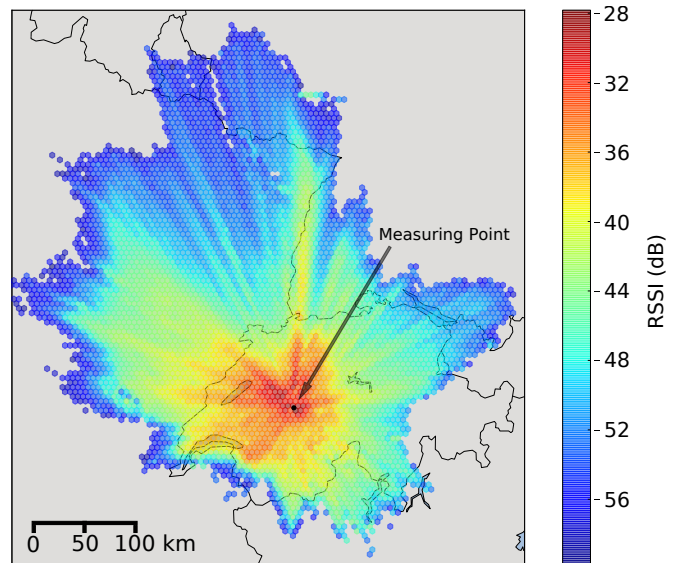


Fig. 6. Heat map showing the RSSI distribution for messages received by a USRP sensor.

patchy. Almost all aircraft (98%) broadcast their call sign, while just under 80% report each velocity and position. Miscellaneous messages are sent out by 27.4% of aircraft currently equipped with transponders, but they occur very rarely, in total making up less than 3% of all received messages. These mainly comprise unknown or non-standard formats and test messages.

Message rates of the regular broadcasts mostly comply with the standard. Position and velocity are each broadcasted twice in a second and the call sign once every five seconds. Overall, 76.9% of all ADS-B capable aircraft broadcast call sign, position and velocity, leading to an average rate of 4.2 messages per second. With further implementation of the standard by the airlines, this will rise to 6.2 messages per second [2].

B. Channel analysis

We deployed a USRP-based receiver for 14 days to record a sample of 53,626,642 messages for further in-depth analysis of the signal characteristics. Utilizing these messages with an RSS indicator (RSSI) and SNR data, we conduct a thorough analysis of the 1090 MHz channel that is used as a data link for commercial ADS-B.

Sensing range and propagation model: A LOS connection is necessary for successful message reception in ADS-B [1]. In fact, we found that aircraft emit sufficient power (up to 500 W [5]) to achieve a SNR high enough for successfully demodulating messages from distances up to the radio horizon. The radio horizon is given by the disruption of the LOS by the earth’s curvature. Assuming a smooth earth and an aircraft’s altitude of 10 km, the radio horizon is at a distance of about 450 km. In accordance with this result, we received position reports over distances up to 440 km with an SNR of 2.7 dB.

However, as further analysis showed, environmental and terrain conditions lead to highly varying sensing ranges depending on the direction. This fact is reflected in the RSSI of the messages depicted in Fig. 6. It clearly shows that

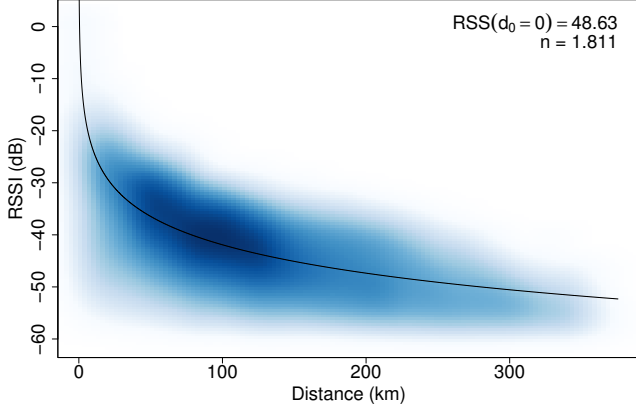


Fig. 7. The RSSI from our 14 days measurement with more than 20 million samples. The RSSI model parameters are derived using linear regression.

the behavior of path loss strongly depends on the direction and the RSSI drops abruptly at the horizon. For instance, the Swiss Alps severely limited the range to the southeast to approximately 130 km while in comparison, the range in north-northwest direction is about 400 km.

For LOS connections, we found that the path loss fits the log-distance path loss model (LDPL) given as follows:

$$PL(d) = PL(d_0) + 10n \log_{10}(d/d_0),$$

where n (usually around 2) is the path loss exponent, $PL(d_0)$ the path loss at the close-in distance d_0 which is determined from measurements and d the distance between sender and receiver [10]. Using this, the following simple model for signal strength can be set up:

$$\begin{aligned} RSS(d) &= P_{tx} - PL(d) - C \\ &= RSS(d_0) - 10n \log_{10}\left(\frac{d}{d_0}\right), \end{aligned}$$

where P_{tx} is the power emitted by the sender and C is a constant covering factors such as antenna gain. Since our model is measurement based, the constants can be combined to the measured signal strength $RSS(d_0)$ at the close-in distance.

Given a set of measured data, the parameters for the above model can be determined using linear regression. Therefore, logarithmic distances are considered in order to establish the linearity between distance and RSSI. The parameters cannot be compared to the parameters of the traditional LDPL model (where $n \approx 2$) since $RSS(d_0)$ summarizes many factors whereas $PL(d_0)$ only describes the path loss at a certain distance. The slope n of the model is interdependent with $RSS(d_0)$ and is also not comparable to the slope of the LDPL.

The result of such an analysis for our 14 days dataset is shown in Fig. 7. It clearly shows the logarithmic dependence of the RSSI on the distance. The explanation for the high variance in the RSSI samples shown here has already been given above: The RSSI depends on the direction, which is neglected by the log-distance path model, resulting in a high variance when considering samples from many directions. Besides this factor, a varying transmission power can also cause variance. According to [5], the transmission power of ADS-B messages is at least 21 dBW (125 W) and not more than 27 dBW (500 W). Since we do not have any information

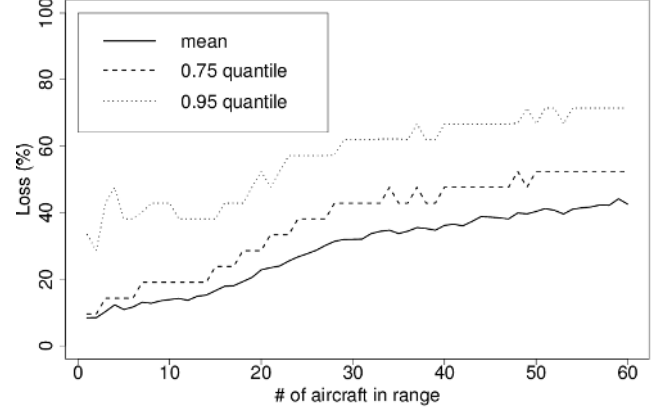


Fig. 8. Packet loss vs. number of senders in transmission range of our ground sensor.

on the actual implementation, we cannot exclude this factor as a source of variance in our RSSI measurements.

Loss rates: To estimate how ADS-B messages get lost, we created message profiles for each aircraft, i.e., we determined how many messages an aircraft sends out on average by considering its supported message types. In five seconds, a typical aircraft sends 21 messages (ten reports for each velocity and position and one identification message) but these numbers vary depending on the transponder's implementation. Loss can be estimated by comparing this average rate with the number of messages received during a given period.

Overall, loss is strongly correlated with the RSSI. Since RSSI is dependent on the distance, loss also increases with increasing distance of the aircraft from the sensor. For signal strengths higher than the demodulation threshold, loss occurs due to interference with SSR and collisions with other ADS-B messages. One way to further analyze the loss related to message collisions and interference is to consider the number of senders (aircraft in reception range) and the occurring loss. The relationship between these two variables is shown in Fig. 8. Initially, loss is about 20% when only a few ADS-B equipped aircraft are in range and increases to more than 50% in the presence of 60 aircraft. As already mentioned, only about 60% of all aircraft in our sensing range are equipped with ADS-B, while 100% are using SSR with much higher message rates, causing frequent message collisions. The activity on the 1090 MHz channel explains why we observed much higher loss rates during peak hours compared to midday or at night.⁶

However, the high loss rate is alarming. As mentioned above, ADS-B was built upon Mode S mostly for economical reasons. The 1090 MHz channel is already used by three legacy secondary radar systems (Mode A/C/S) with much higher message rates. Due to the higher payload and the resulting increased transmission times, message collisions with other ADS-B messages or Mode A/C/S replies are more likely for ADS-B. Furthermore, the loss of one ADS-B message is more severe due to its lower message rate. In other words, ADS-B's reliability (and thus safety) suffers from the high losses caused by the overused communication channel. This

⁶For some further insights on channel behaviour and lost messages, the reader is referred to [11].

# of ADS-B Messages seen	ca. 127,000,000
# of positional ADS-B messages	58,400,372
# of those seen by 4+ sensors	2,749,952
# of those seen by 5+ sensors	427,766
# of those seen by 6+ sensors	21,168

TABLE II. DATA FROM A TWO WEEK TEST PERIOD 15–28 MAY 2013 WITH 7 SENSORS.

calls for reducing the use of legacy systems once ADS-B has been deployed in order to maintain its reliability.

Doughnut effect: There is a noticeable drop in reception quality of messages that are sent in close proximity to a sensor. This effect is caused by clipping as aircraft use very high transmission power of up to 500 W [6]. This effect is reflected in the data where loss rates are higher when aircraft are in very close proximity to the receivers (around 10–20 km) compared to when they are further away. We experienced the lowest loss rates at around 50 km.

Transmitter effects: Aircraft equipped with ADS-B typically use two antennas to alternately transmit messages. This can lead to differences in characteristics such as RSSI values. In fact, we observed that the RSSI of a single flight shows two levels with a gap of up to 5 dB between these levels at close distances. Since there is no indicator on the antennas given in ADS-B messages, filters such as a moving average as threshold or the upper (or lower) 50% percentile of a flight can help mitigate this effect.

Duplicate messages: We regularly receive duplicates of the same message in our test data, i.e., 0.34% of correctly decoded ADS-B messages are identical with one or more previous messages while 0.26% of unique messages have been duplicated. We observe messages being duplicated and received by the same sensor up to 200 times over the course of a few seconds in some cases. While the ADS-B protocol uses a pulse modulation and is hence considered susceptible to multipath effects, particularly in mountainous areas, this is unlikely to be the main cause of such repeated duplicates. We strongly suspect that some transponders are not correctly implementing the ADS-B standard. While this does not affect ADS-B-based air traffic control (ATC), it needs to be taken into account for protocol development and message processing.

Weather effects: We used random samples of 1,000,000 messages from the complete dataset and weather data from a nearby weather station to examine the effects of weather on RSSI, loss and SNR. For example, we compared the RSSI, loss, and SNR distributions of samples from rainy periods with those of dry periods to determine the effect of rain. The random samples were used to eliminate temporary effects.

We found that rain lowers the average RSSI by around 1 dB while high humidity has a slightly larger negative effect. Strong solar activity causes RSSI values to drop at the receiver for a bit more than 2 dB. There was no measurable impact on SNR in our data. Concerning lost packets, we could not conclude any significant effect from our explorative approach, although due to the typically strong correlation with RSSI, we could assume that it follows similar patterns.

C. Wide-area multilateration (WAM)

WAM provides a position estimate not derived by the aircraft itself, it is *independent* (unlike ADS-B), although it does require the aircraft’s *cooperation* by sending out signals. Using the time difference of arrival (TDOA) of a signal between four or more sensors, the position of a wireless sender can be calculated.⁷

We have chosen multilateration as a benchmark for OpenSky and as a tool to evaluate the quality of OpenSky’s data for several reasons:

- **Relevance:** As further described below, multilateration is already used in aviation and its importance grows with the introduction of ADS-B. Commercial dedicated multilateration systems are extremely expensive and complex and therefore not suited for research. For that reason, a low-cost solution and a proper analysis of the achievable accuracy as well as its limits is desirable.
- **Synchronization:** Multilateration requires a tight synchronization of the receiving antennas. Although the sensors in OpenSky are not globally synchronized, it is possible to achieve an *a posteriori* synchronization by using ADS-B’s position reports. The accuracy of this synchronization depends on the accuracy of the positions reported and the clocks of the sensors. Hence, the positional error we observed with multilateration is an appropriate measure for both variables.
- **Completeness:** As demonstrated in this section, we were able to implement WAM by exclusively using data provided by OpenSky. This confirms our design choice to follow the immutable approach, i.e., to store the raw data as it is provided by our sensor network without any level of abstraction.
- **Coverage:** To perform multilateration in a certain area, it needs to be covered by at least four sensors. The implementation and analysis of multilateration provides detailed insights on the structure of OpenSky’s sensing range and challenges the constellation of the sensors positions.

WAM is actively employed in modern ATC systems (e.g., ASDE-X [14]) at some airports in the US and Europe. As multilateration can utilize wireless signals from communication already in place (e.g., ADS-B), no changes to the existing infrastructure are required. On the ground, sufficiently many sensors and processing stations need to be deployed.

Multilateration is currently used primarily for taxiing and in close distances around airports, yet wide-area multilateration is becoming more and more relevant in modern ATC. Compared to PSR, it is relatively easy and cost-effective to install and use on the ground. WAM is considered reasonably accurate, roughly on par with ADS-B and much more precise than PSR, although it is susceptible to multipath effects and duplicate messages. A solution acceptable for use in modern ATC should achieve the same surveillance accuracy category (SAC) as ADS-B for up to 90 NM [15].

In this section, we analyze the application of WAM based on ADS-B messages captured with our sensor network. WAM

⁷For a full explanation of the multilateration process in aviation see, e.g., Savvides et al. [12] or Neven et al. [13].

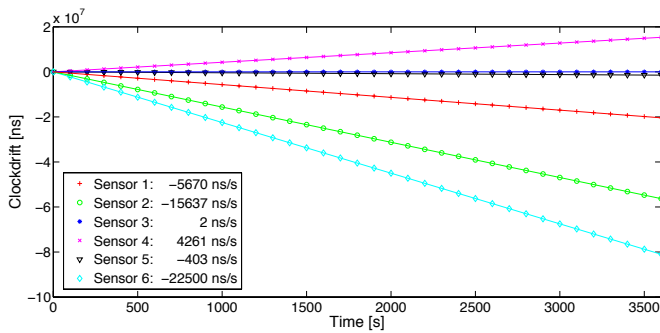


Fig. 9. Drift of the internal clocks of six sensors compared to a reference sensor over 1 hour.

is an interesting and relevant use case for researchers and authorities that are looking to build a sensor network for ADS-B signals. It provides additional location verification without requiring any further cooperation by the aircraft. We examine the potential coverage of WAM in comparison to using pure ADS-B, then analyze the precision our current setup is capable of. We address various challenges in setting up the system, evaluate its accuracy and analyze the reasons for deviations between the two localization systems.

Dilution of precision: Dilution of precision (DOP) is an important metric for the quality and usefulness of the measured data in geomatics engineering.⁸ It specifies the multiplicative effect of the geometry of the senders and receivers on the precision of the measurements. As such it is not a metric for the precision itself but for how noise in the measurements affects the final state estimates and thus the reliability for use in WAM. Depending on the positions of aircraft and sensors, the precision becomes diluted and reduces the quality of the WAM estimates.

DOP values of up to 5 are commonly considered good in satellite navigation where high precision is required; higher values might be tolerable depending on the application. In our case, we calculate DOP for each received ADS-B position message and discard all values over 30, where ADS-B claims cannot be accurately verified with WAM. This value provides us with the lowest errors for our dataset, i.e., the best trade-off between dilution of precision and having enough WAM messages for accurate and frequent synchronization. With external synchronization this value could also be lowered, further improving the results. The error is generally worst around the link from the aircraft to the center of the network [17]. Overall, the deployment of nodes in the field has a direct influence on DOP as both numbers and relative receiver positions play an important role.

Synchronization: For WAM, the sensors must be tightly synchronized. The SBS-3 boxes themselves provide no means of synchronization, but it is possible to synchronize the clocks of the sensors *a posteriori*. The SBS-3 exhibits a 24 bit register that is incremented every 50 ns. This value is extracted and saved with every message. To recover the relative time between two ground stations based on this register, we use a positional ADS-B message. As we know the position of the ground sensors and the advertised location of the aircraft, we can calculate the clock offset of the two registers by accounting

⁸See Zhu [16] for background on DOP calculations.

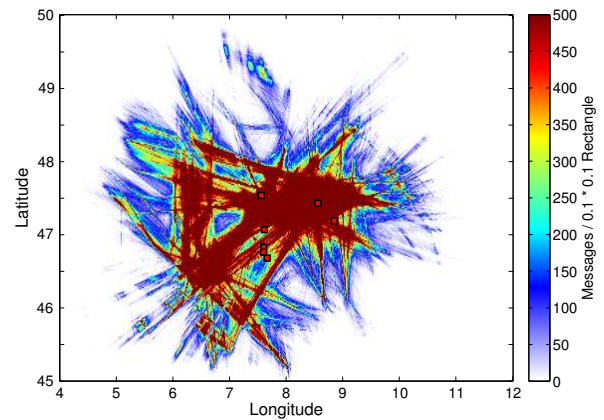


Fig. 10. Coverage of aircraft's positions based on 58,400,372 aggregated positional ADS-B messages. The map focuses on a cluster of 7 closely located sensors and shows the hot spots of received messages in a given rectangle over 5 degrees latitude and 8 degrees longitude.

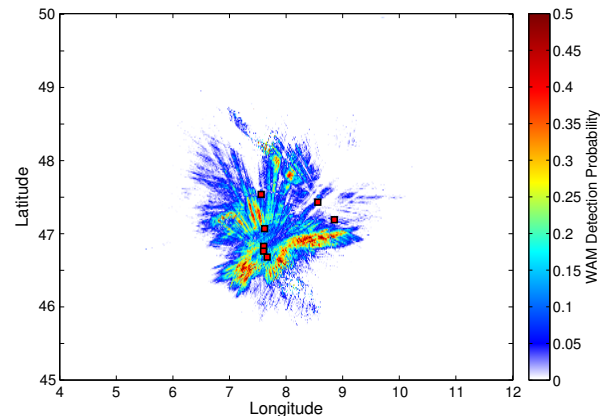


Fig. 11. Probability for ADS-B messages to be received by a sufficient number of sensors (4 or more).

for the signal propagation time between the aircraft and the ground stations. The signal propagation time is approximated by the relative distance divided by the speed of light. This approach allows us to determine the clock offsets of all ground stations and achieve global synchronization. Since the clocks have different drifts (we observed differences of up to $27 \mu\text{s/s}$ in our sensors, see Fig. 9) that are not constant for long and depend on external factors such as temperature, this procedure must be repeated constantly. A robust linear regression over the last received messages of a sensor predicts the clock drift for any new message. More complex synchronization methods can improve accuracy.

Coverage: An interesting question is how the overall ADS-B coverage provided by the sensor network capturing ADS-B messages translates to potential WAM coverage. Since we need multiple sensors for WAM, the coverage area is expected to be in the vicinity of a number of our sensors. This difference in coverage also illustrates the additional cost of a WAM-capable network providing sufficient coverage compared to pure ADS-B handling of the same area. Since one of the main reasons behind the introduction of ADS-B has been cost savings, this is an important factor when considering WAM as an alternative support system for ATC.

Errors	Lat [m]	Lon [m]	Hor [m]	Alt [km]
Mean	206.50	251.96	295.55	50.63
Median	96.40	125.16	165.68	20.23
RMSE	346.51	424.09	451.43	125.63
5 %-Quantile	6.19	7.44	20.65	1.29
95 %-Quantile	833.99	962.64	1,083.72	198.40

TABLE III. SUMMARY OF STATISTICAL DATA ABOUT WAM DEVIATIONS FROM ADS-B POSITIONS.

Fig. 10 illustrates the relative frequency of positional ADS-B messages received by at least one sensor in the testing period, i.e., it represents the network’s ADS-B coverage hot spots. High frequencies in red show metropolitan areas and main air routes. This is illustrated in Fig. 11, which shows the relative frequency with which ADS-B messages reach the WAM quorum of at least 4 sensors in particular areas. Overall, 4.7% of all ADS-B messages we have seen reached that quorum. Naturally, these areas tend to be much closer to the center of our network. In stronger reception areas, more than 30–50% have a detection rate sufficient for WAM. Identifying these areas from a given network is obviously a starting point when deploying a WAM solution. Below 20,000 ft, we received few messages at enough sensors, hindered by the terrain. This is expected as sensor positioning makes OpenSky currently more useful for en-route WAM over airport control.

Data processing: We used a linear algorithm of the well-known multilateration principle to be able to handle large amounts of data with sufficiently high speed.⁹ Linear multilateration requires measurements from at least five sensors, which leaves 427,766 messages to analyze. We used MATLAB to solve the resulting TDOA equations.

A noteworthy processing challenge are duplicate messages. While it is fairly easy to filter a data point that has been received more than once by the same sensor, it is more difficult to tell which of the received duplicates—if any—is the correct one. In the worst case it might even have been destroyed and we only received incorrect signal timings. Since even very small TDOA will create large deviations in position, we can use general plausibility checks (e.g., the calculated WAM position is out of the physical detection range of any of the sensors involved) to ensure that we are not taking duplicated messages into account.

A second problem that is present in the current ADS-B environment is the fact that a low percentage of ADS-B equipped aircraft is not using satellite navigation to calculate their own positions but a method called *dead reckoning*. Dead reckoning advances a previously determined position over time based on speed and course estimates and is significantly less accurate, exhibiting deviations from the real position of up to several kilometers. Unfortunately, it is not possible to tell the position method of an aircraft from the ADS-B data, hence these transponders cause outliers in our WAM processing and deteriorating the accuracy evaluation results.

Accuracy: For all received WAM-capable ADS-B messages with a DOP value of lower than 30, we saw a median error in the horizontal plane of 166 m and a mean of 296 m (see Table III for the full results). The level of accuracy of the

⁹For a short overview of common algorithms, see [18], [19].

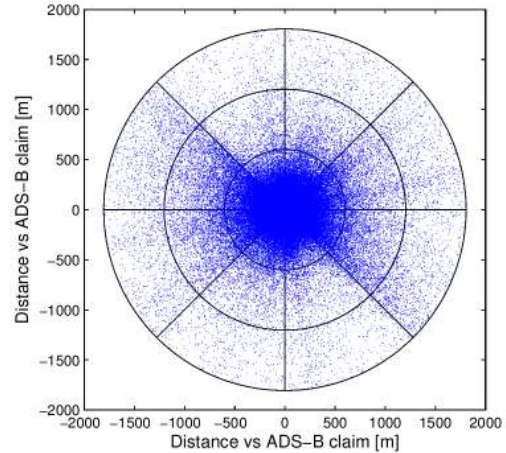


Fig. 12. Horizontal position errors of WAM vs. ADS-B. The errors are largely uniformly distributed with small deviations stemming from imperfect sensor positioning.

results is directly dependent on the message timestamps delivered by the SBS-3 devices [18]. While commercial systems and purpose-built testbeds may be able to perform multilateration with much higher precision¹⁰, we believe our low-cost setup is sufficient to show the challenges and effectiveness of ADS-B-based WAM and evaluate the function of OpenSky.

Error analysis: Fig. 12 shows the overall distribution of the horizontal WAM errors on a radar plot, i.e., the bearing and matching distance of each point against the ADS-B claim. We can see that more than 80% of these errors cluster rather uniformly around the center in a distance of up to 500 m, which is to be expected as this type of noise can usually be modeled as Gaussian. The altitude of the aircraft did not make a significant difference in accuracy. Both errors and confidence intervals are much smaller in the non-core hours (i.e., 10 p.m.–4 a.m.) than during the day. This seems plausible as the channel is much less frequented during these times and thus suffers from fewer artifacts (e.g., multipath or loss).

It is known that with WAM sensors on the ground (i.e., typically in a plane), it is not possible to run 3D localization on airplanes as the altitude is not predictable.¹¹ One of the main reasons for this is high DOP values. Consequently, the altitude estimate of the aircraft was not useful (see Table III).

Lessons learned: There is some potential for improvements on the accuracy of the WAM results which have been out of the scope of OpenSky as a low-cost network using off-the-shelf sensor hardware. They should be taken into account in commercial sensor networks for ADS-B.

- Other ways must be found to improve altitude correction (see, e.g., [17]) or to separately estimate an aircraft’s altitude with different methods. However, this is beyond the scope of the present work.

¹⁰For example, the NAC required by the FAA for ADS-B in surveillance is a 95% bound of 92.6 m [2] and 128 m for WAM, which needs to be delivered by real-world solutions.

¹¹For example, Daskalakis and Martone [20]: “Three-dimensional solutions are not used in this evaluation, and are not considered to be useful for air traffic control purposes, because the accuracy of vertical solutions is quite poor except when an aircraft is nearly above a [sensor].”

- The deployment of sensors plays an important role in improving the accuracy. With sensors being distributed to volunteers, the perfect placement in terms of DOP, reception range and geographic and infrastructure considerations is difficult to achieve.
- The obvious improvements in a commercial system compared to our sensor network include more expensive hardware with much finer clock resolution than 50 ns and synchronization via a cable network or independently through GPS clocks.¹²
- More complex filtering methods such as Kalman filters could help improve the measurements.

VII. FUTURE WORK

In the future, we plan to continue extending both the scale and the scope of OpenSky. Besides distributing more sensors, we are also investigating alternative database structures that will scale well with a growing OpenSky, since the MySQL database is currently the major infrastructural limitation when processing the enormous amount of data.

Furthermore, we are investigating other ADS-B sensors. Although the SBS-3 is a good source for ADS-B data, the different applications of OpenSky might also require different types of data. For instance, a cheaper and more user-friendly alternative to the USRP-based software radio receiver which also measures physical signal properties (RSS, SNR) of ADS-B messages would enable OpenSky to cover these interesting aspects on a much larger scale. Built-in synchronization and other features are appealing, too, although their utility must justify the higher cost of the system.

VIII. CONCLUSION

In this paper we presented OpenSky, a participatory sensor network for monitoring air traffic based on ADS-B. It consists of low-cost sensors and is especially designed for research. Sensors deployed by volunteers receive ADS-B messages and forward them to a central database server. By providing unfiltered and non-abstracted data to the research community, OpenSky opens ADS-B to a broad range of research applications. We furthermore provided insights on OpenSky's data and their limits as well as the ADS-B channel as perceived by the sensor network. By implementing and analyzing wide-area multilateration, we demonstrated its applicability to relevant problems and provided solutions to a variety of issues researchers face when working with data provided by such a sensor network.

ACKNOWLEDGEMENT

We would like to thank our shepherd Prof. John Stankovic and the anonymous reviewers for their feedback and suggestions on how to improve this paper. We are grateful to Prof. Jens B. Schmitt for the valuable discussions and his feedback, and to all the volunteers contributing to OpenSky with a sensor.

¹²GPS has a clock resolution of 1 ns which is more than sufficient for the required service levels.

REFERENCES

- [1] M. Schäfer, V. Lenders, and I. Martinovic, "Experimental analysis of attacks on next generation air traffic communication," in *Applied Cryptography and Network Security*, ser. LNCS, no. 7954. Springer, Jun. 2013.
- [2] RTCA, Inc., "Minimum aviation system performance standards for Automatic Dependent Surveillance – Broadcast (ADS-B)," Report DO-242A (including Change 1), Dec. 2006.
- [3] RTCA, Inc., "Minimum operational performance standards for Universal Access Transceiver (UAT) Automatic Dependent Surveillance – Broadcast," Report DO-282B, Dec. 2011.
- [4] Federal Aviation Administration, "Automatic Dependent Surveillance – Broadcast (ADS-B) Out performance requirements to support Air Traffic Control (ATC) Service; Final Rule," *Federal Register*, vol. 75, no. 103, May 2010, 14 CFR Part 91.
- [5] *Annex 10: Aeronautical Telecommunications*, 4th ed., ICAO, 2007, Volume IV: Surveillance and Collision Avoidance Systems.
- [6] RTCA, Inc., "Minimum operational performance standards for 1090 MHz Extended Squitter Automatic Dependent Surveillance – Broadcast (ADS-B) and Traffic Information Services – Broadcast (TIS-B)," Report DO-260B (with Corrigendum 1), Dec. 2011.
- [7] N. Marz and J. Warren, *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*. Manning Publications Co., Jan. 2012.
- [8] *API Reference Guide*, Kinetic Avionics Ltd., Aug. 2013, revision 1.04.
- [9] Statistics and Forecast Service, "Long-Term Forecast: IFR Flight Movements 2010 – 2030," European Organization for the Safety of Air Navigation (Eurocontrol), EEC Technical Report CND/STATFOR Doc415, Dec. 2010.
- [10] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall PTR, Apr. 1996.
- [11] M. Strohmeier, M. Schäfer, V. Lenders, and I. Martinovic, "Realities and challenges of nextgen air traffic management: The case of ads-b," *Communications Magazine, IEEE*, vol. 52, no. 5, 2014.
- [12] A. Savvides, H. Park, and M. B. Srivastava, "The Bits and Flops of the N-hop Multilateration Primitive for Node Localization Problems," in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*. ACM, 2002, pp. 112–121.
- [13] W. H. L. Neven, T. J. Quilter, R. Weedon, and R. A. Hogendoorn, "Wide Area Multilateration," National Aerospace Laboratory NLR, Report on EATMP TRS 131/04, August 2005, requested by Eurocontrol.
- [14] J. Herrero, J. Besada Portas, F. Rodriguez, and J. R. C. Corredera, "Asde and multilateration mode-s data fusion for location and identification on airport surface," in *The Record of the 1999 IEEE Radar Conference*, 1999, pp. 315–320.
- [15] A. Smith, R. Cassell, T. Breen, R. Hulstrom, and C. Evers, "Methods to provide system-wide ADS-B back-up, validation, and security," in *Proceedings of the 25th IEEE/AIAA Digital Avionics Systems Conference*, Oct. 2006, pp. 1–7.
- [16] J. Zhu, "Calculation of geometric dilution of precision," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 893–895, Jul. 1992.
- [17] M. Garcia, J. Dolan, R. Mueller, and R. Smith, "An enhanced altitude correction technique for improvement of WAM position accuracy," in *Integrated Communications, Navigation and Surveillance Conference*, Apr. 2012, pp. 1–9.
- [18] H. Lee, "Accuracy Limitations of Hyperbolic Multilateration Systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-11, no. 1, pp. 16–29, Jan. 1975.
- [19] P. Xinxin, H. Zhigang, Z. Jun, Z. Yanbo, and L. Wei, "An approach for close-form solution of multilateration equations," in *8th World Congress on Intelligent Control and Automation*, 2010, pp. 5464–5468.
- [20] A. Daskalakis and P. Martone, "A technical assessment of ADS-B and multilateration technology in the Gulf of Mexico," in *Proceedings of the 2003 IEEE Radar Conference*, May 2003.