



A large-scale validation study of aircraft noise modeling for airport arrivals

Thomas C. Rindfleisch,^{1,a)}  Juan J. Alonso,¹  Donald C. Jackson,¹ Brian C. Munguía,¹ and Nicholas W. Bowman²

¹Department of Aeronautics and Astronautics, Stanford University, Stanford, California 94305, USA

²Instrument Flight Software Group, NASA Jet Propulsion Laboratory, Pasadena, California 91011, USA

ABSTRACT:

In the U.S., the Federal Aviation Administration's Aviation Environmental Design Tool (AEDT) is approved to predict the impacts of aircraft noise and emissions. AEDT's critical role in regulatory compliance and evaluating the environmental impacts of aviation requires asking how accurate are its noise predictions. Previous studies suggest that AEDT's predictions lack desired accuracy. This paper reports on a large-scale study, using 200 000 flight trajectories paired with measured sound levels for arrivals to Runways 28L/28R at San Francisco International Airport, over 12 months. For each flight, two AEDT studies were run, one using the approved mode for regulatory filing and the other using an advanced non-regulatory mode with exact aircraft trajectories. AEDT's per aircraft noise predictions were compared with curated measured sound levels at two locations. On average, AEDT underestimated L_Amax by −3.09 dB and SEL by −2.04 dB, combining the results from both AEDT noise-modeling modes. Discrepancies appear to result from limitations in the physical modeling of flight trajectories and noise generation, combined with input data uncertainties (aircraft weight, airspeed, thrust, and lift configuration) and atmospheric conditions.

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/10.0025276>

(Received 29 November 2023; revised 17 February 2024; accepted 23 February 2024; published online 11 March 2024)

[Editor: Anurag Agarwal]

Pages: 1928–1949

I. INTRODUCTION

As in many aviation metroplexes, the deployment of the Federal Aviation Administration's (FAA's) Next Generation Air Transportation System (NextGen) in the San Francisco Bay area in March 2015, resulted in a ground swell of complaints due to changes to air traffic distribution and volume. These complaints have persisted and grown more insistent to the present day. The primary causes of these complaints have been:

- The sudden concentration of previously dispersed air traffic owing to the precision of NextGen navigation technologies such as the Global Positioning System (GPS),
- changes to established aircraft routes and procedures (by distances of up to five miles) that moved noise to communities that were unaccustomed to overflights, and
- the steady increase in overall air traffic volume.

There is an inevitable trade-off between the safety, economic, and environmental factors in the evolution of commercial aviation. These are routinely assessed and balanced using predictive models to estimate the parameters and consequences involved. In the U.S., the FAA Aviation Environmental Design Tool (AEDT) is the prescribed regulatory modeling system for estimating noise and air pollution levels that will result from traffic pattern designs and

changes. For aircraft noise predictions to provide useful information for both community impact assessments and the informed redesign of the airspace in making these trade-offs (Hauptvogel *et al.*, 2021), it is essential that predictions accurately model the level of noise produced by aircraft overflights, both near the airport, where the noise levels can be significant (>DNL 65 dB), and in areas farther away, where the noise levels are typically lower (~DNL 50 dB), and where complaints have increased significantly. There is growing evidence that AEDT has problems accurately predicting noise metrics.

The aims of this study are fourfold:

- (1) To collect a very large, statistically significant set of data, pairing aircraft flight profiles with carefully curated ground sound level measurements over time and identifying intrinsic limitations in the physical measurements.
- (2) To select aircraft study cohorts that control as much as possible many of the flight variables involved while examining AEDT behaviors for a broad, real-world fleet mix.
- (3) To tease out statistically significant measures of AEDT metric prediction accuracy and analysis anomalies, aircraft-type by aircraft-type, to reveal internal computational strengths and weaknesses.
- (4) To provide recommendations for work needed to improve AEDT's noise prediction accuracy.

^{a)}Email: tcr@stanford.edu

The implementation of these goals is based on a data set of over 200 000 flights arriving to runways 28L/28R at San Francisco International Airport (SFO) between July 2021 and June 2022, along with corresponding sound level monitor (SLM) measurements on route segments close to two carefully chosen SLM locations. The data are carefully curated to eliminate compromised sound measurements.

II. RELATED WORK

A number of papers outlining approaches and difficulties for AEDT-versus-ground-measurement validation studies appear in the open literature, but none involve datasets as extensive as the one we have developed (He, 2018). Mavris *et al.* did a literature review in 2020 to understand previous verification and validation studies utilizing the legacy integrated noise model (INM) and newer AEDT systems, as well as the evolution of AEDT's functionalities over time (Mavris and Sparrow, 2021). They note a dated development of an extensive Defense Intelligence Agency database of aircraft noise data in 1997, and specific INM feature additions such as different power-prediction methods, higher altitude operations, thrust levels, and approximations to flap and landing gear configurations.

Two other papers studied examples of metric-estimation results compared to limited real-world flight-profile examples. Gabrielian *et al.* explore the sensitivity of noise-metric predictions to varying assumptions within AEDT (Gabrielian *et al.*, 2021). They use flight performance characteristics from airline flight data records, noise-monitoring data from stations near the SFO airport, and historical weather data to study three each arrival and departure profiles. AEDT noise-metric sensitivity was measured for the different flight procedures and profiles, thrust [including that obtained from a limited and confidential set of actual flight operational quality assurance (FOQA) data, weight, and weather]. The results include statistics from a variable number of sound monitors for each of the six flights, and it is the inter-monitor differences that are used to estimate the statistical stability of AEDT predictions (assuming that the sound-monitor measurements are each stable and accurate as recorded). An important issue is that the sound-profile analysis was not able to distinguish aircraft noise components from background noise (e.g., for monitors close to highway noise). Also, the six examples involve a rather wide range of aircraft types, including the Airbus 320, the Boeing 737, and the Boeing 757. For the departure flight events, some AEDT estimates are found to be 1–2.5 dB below monitor measurements, whereas one in particular showed AEDT estimates 4–6 dB below the monitor values. The conclusion was that “for a majority of noise-monitor readings among the modeled flights, there is a good overall match between AEDT and real-world observations.” That seems hard to reconcile with event errors up to ~6 dB.

Giladi and Menachi take a more rigorous statistical view of the uncertainties in comparing measured and estimated noise levels (Giladi and Menachi, 2020). The statistical

variation in measured noise levels results from compound factors, such as differences in atmospheric conditions, fluctuations in the emitted noise, variations in operational settings, different configurations, diverse pilot practices and procedures, and specific adjustments. Noise measurements confirmed that such variations can be responsible for differences of as much as 12 dB.

Researchers from ASCENT Project 62 also provided a preliminary assessment of the accuracy of AEDT using its most sophisticated trajectory, weather, and operational modeling capabilities as well as the sensitivity of noise predictions to various modeling assumptions (Bendarkar *et al.*, 2022). They found that, depending on the level of modeling, noise underprediction errors for arrival procedures can be as large as 3–4 dB (in the SEL metric) when the best modeling techniques leveraging actual flight data are employed.

A recent conference paper outlines a validation study, using the LAeq metric and compares two European noise prediction models (*sonAIR* and *FLULA2*) and AEDT against measured noise data for Zurich/Geneva airport traffic (Meister *et al.*, 2023). The relevant comparison with AEDT included a total of 6659 aircraft distributed among 24 aircraft types and included both arrivals and departures “close” to the airport. No comparisons of the computed and measured LAmax and SEL metrics are given. The average difference between standard AEDT estimates of LAeq and measurements appears to be ~−1.5 dBA, but this is not broken down by arrival and departure nor by aircraft type, as we have done in our study.

III. STUDY DESIGN AND METHODOLOGY

A. AEDT noise metric predictions

AEDT uses an enhanced aircraft performance modeling system called the Base of Aircraft Data (BADA) (Nuic, 2010), initially developed by EUROCONTROL in the 1990s. In order to predict noise, AEDT uses the BADA performance models, the SAE-AIR-1845 aircraft noise model, AIR1845A (ICAO, 1988; Rhodes and Boeker, 2019; SAE Committee A-21, 2012), and the ICAO Aircraft Noise and Performance (ANP) Database (EASA, 2023; ECAC Secretariat, 2016). An early version of BADA, BADA 3, was *not* intended to optimally model aircraft operations in the terminal area nor during takeoff, climb and descent, which are the current regions of most noise concerns. Instead, on approach segments below altitudes of 10 000 feet, which includes all of the approach segments analyzed in this paper, AEDT computes noise estimates from *standard* altitude and airspeed profiles in conjunction with the latitude and longitude of the aircraft ground track. Despite these acknowledged limits, ANP/BADA 3 modeling is still prescribed by the FAA for routine regulatory use for assessment of noise impacts in airport environments (FAA, 2022).

BADA 4 (announced in 2010) addresses some of the BADA 3 gaps by using both higher fidelity data for existing cruise flight segments and new data and methods for operations in the terminal area. These expansions of the model

and data allow BADA 4 to be used for all flight phases (Poles *et al.*, 2010). BADA 4 capabilities were introduced with the release of AEDT 3b in September 2019. No BADA 4 aircraft models are currently approved for FAA regulatory use (FAA, 2022).

To model an aircraft flight arrival with AEDT, we create a detailed study file specifying all of the appropriate parameters, including calibrated airspeed, derived from Automatic Dependent Surveillance-Broadcast (ADS-B) ground speed using NOAA High-Resolution Rapid Refresh (HRRR) atmospheric data (Alexander *et al.*, 2020).

In this paper we compared AEDT's noise predictions using two different modeling approaches:

- **AEDT-R:** ANP/BADA 3, with only ground-track positions specified; no altitude or airspeed controls are provided. This is the modeling approach required by FAA for regulatory purposes.
- **AEDT-AE:** BADA 4, with both altitude and airspeed controls specified for each ground-track position.

We have attempted to use AEDT to the best of its predictive abilities. We have consulted with FAA and Volpe Center experts to optimize the preparation of flight-profile data to fit AEDT modeling constraints (e.g., smoothing flight profile parameters and limiting profile altitude or speed changes). We have not tried to develop user-defined flight modeling profiles, although such efforts might result in a closer match between AEDT predictions and SLM measurements (Meister *et al.*, 2023; NAS, 2018).

B. Sound measurements

We have developed a sound level analysis system that matches and associates noise peaks with aircraft overflights (via ADS-B trajectory data) to identify the flight that gives rise to each peak detected by the monitor (see the Appendix). In published noise studies, the sound-monitor measurements have generally been considered a “gold standard,” but the real world is more complex:

- Random variations in sound propagation between aircraft and sound monitors can arise due to atmospheric conditions (turbulence, cell structure, stratification, and reflections),
- ambient noise sources of all sorts can distort recorded noise peaks,
- in busy metroplexes it is not uncommon for two aircraft to pass close to an SLM nearly simultaneously, so that one cannot easily distinguish contributions from any single aircraft of interest.

These effects are independent of aircraft noise *per se* and we have tried to exclude “contaminated” data. We eliminate peaks that have contributions from multiple aircraft. We also use a goodness-of-fit (GoF) metric to evaluate each sound peak against an idealized sound model to decide whether to include it in our study data (see the Appendix). Our use of very large data sets allows us to break the data

down into finer subsets to study aircraft-type or ANP-profile-dependent cohorts that more clearly elucidate the strengths and weaknesses of AEDT.

C. Aircraft cohort selection

As shown in Figs. 1 and 2 the number of airports in the Bay Area and the traffic volume produce a complex air traffic pattern. To better explore AEDT performance, we have focused our validation studies on routes that force all aircraft to fly similar flight profiles. We have selected flights on three route segments: final approaches toward each of the two main parallel SFO arrival runways, 28L and 28R, and approach traffic from a highly concentrated, high-volume arrival procedure, SERFR (see Fig. 3).

We have data from about 35 SLMs scattered around the Bay Area, but we focus on two SLMs (SFO NMT-12 and SIDBY) that collect data from flights on our three route segments, as indicated by the circled blue dots in Fig. 3.

The overall fleet mix of aircraft arriving at SFO over the study period (July 2021–June 2022) is shown from two perspectives in Figs. 4 and 5. The first figure shows relative percentages of aircraft types by BADA aircraft identification code and the second shows the percentages of aircraft by ANP performance model type. The majority of aircraft are grouped in approximately 15 models, by count, dominated by regional jets and single-aisle aircraft. A smaller number of large twin-aisle aircraft (also called “heavies”) are also represented.

D. Infrastructure

We have developed a modern hardware/software infrastructure that collects continuous real-time flight and sound data (Jackson *et al.*, 2021). The AEDT study preparation subsystem automatically preprocesses the data and selects an appropriate compact set of GPS, altitude, and calibrated airspeed control points to feed to AEDT via its Microsoft SQL Server database; GPS ground track points only for AEDT-R mode and GPS ground track, altitude, and airspeed points for AEDT-AE mode. These profiles include relevant parameters to estimate noise metrics (L_{Amax} and SEL) at the sound monitor receptors. Individual AEDT-R and AEDT-AE study runs are farmed out in parallel to clusters of virtual machines in a commercial cloud service provider to run thousands of studies for our analyses.

E. Validation process

To validate the noise predictions of AEDT, we compare and analyze the AEDT-estimated noise metrics [using “trajectory-driven flight performance” modeling (FAA, 2021a)] and ground SLM measurements along each of the three chosen route segments. The AEDT-R results incorporate all aircraft in our fleet mix that have ANP/BADA 3 models, whereas AEDT-AE results are limited to the aircraft types in our fleet mix that have BADA 4 models (FAA, 2021b). The noise predictions for flights on final approach to runways 28L and 28R are compared with recordings from



FIG. 1. The SF metroplex includes three international airports (labeled KOAK, KSFO, and KSJC), five regional airports, and two medical centers.



FIG. 2. Typical daylong crowded traffic pattern in the SF metroplex (June 29, 2022). The flight path color codes indicate altitude range.

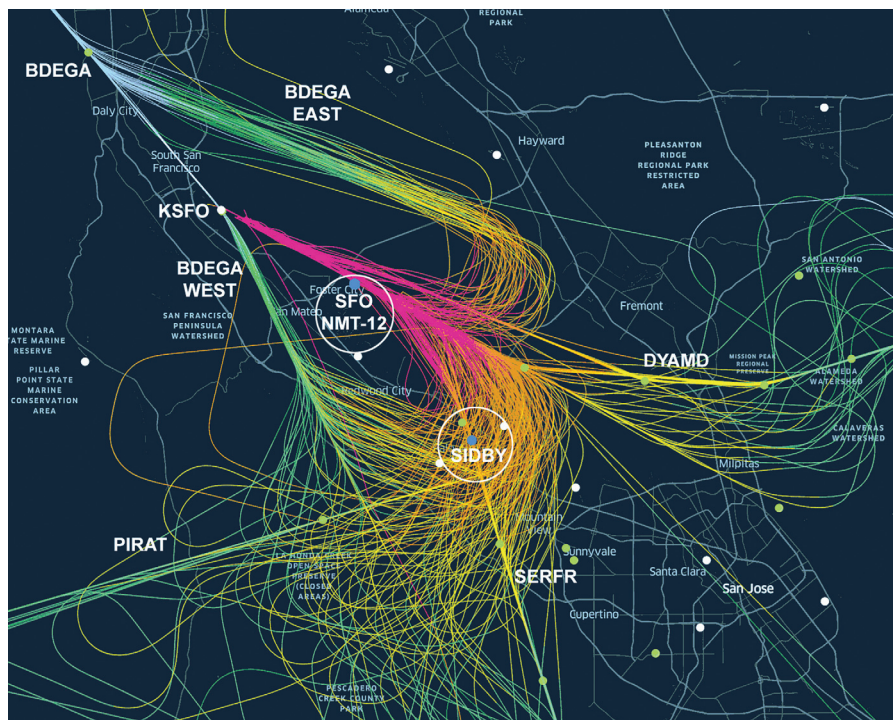


FIG. 3. SFO arrival traffic along the four main routes: BDEGA, PIRAT, SERFR, and DYAMD on June 29, 2022. The color indicates the aircraft altitude along the trajectory.

the SFO monitor NMT-12 in Foster City, ~ 0.4 miles line-of-sight distance from the flight paths. Predictions for approaches along the SERFR-DIRECT route are compared with recordings from one of our project monitors in Palo Alto, ~ 0.9 miles line-of-sight distance from the flight path, at the SIDBY waypoint.

Table I summarizes the total number of flights successfully studied for each modeling approach and how various subsets of flights were discarded for failing quality and relevance criteria.

The rows in the table describe the results of our data screening. The *input AEDT/SLM pairs* represent all flight

profiles for which a successful AEDT study was run with BADA 3 and/or BADA 4 modeling, and which were detected at the SLM associated with the column heading. These counts exclude flights intended to be modeled as BADA 4 but which AEDT downgraded to BADA 3 modeling. The *Number of Pairs skipped as GA* row is the count of those discarded as general aviation flights. The *pairs skipped for low GoF* is the count of pairs for which the SLM peak shape was suspect because it was distorted relative to an analytic model peak using our GoF metric (see the Appendix). The *pairs skipped for multiple PCAs* is the count of pairs for which the SLM peak was suspect because the

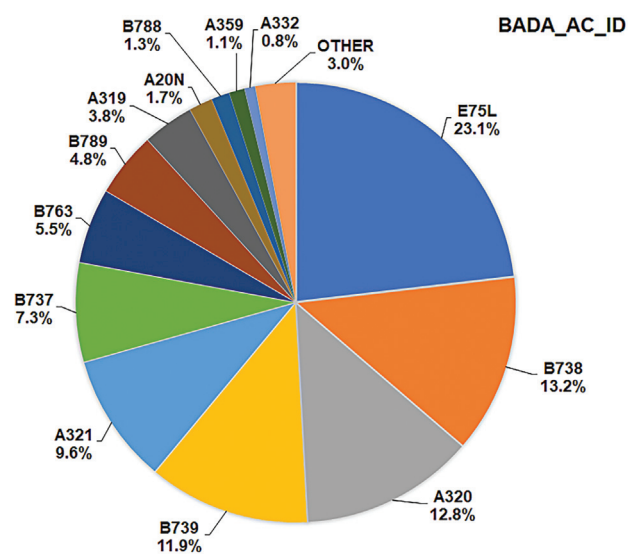


FIG. 4. Relative aircraft mix observed for SFO 28L/R approach by BADA aircraft code.

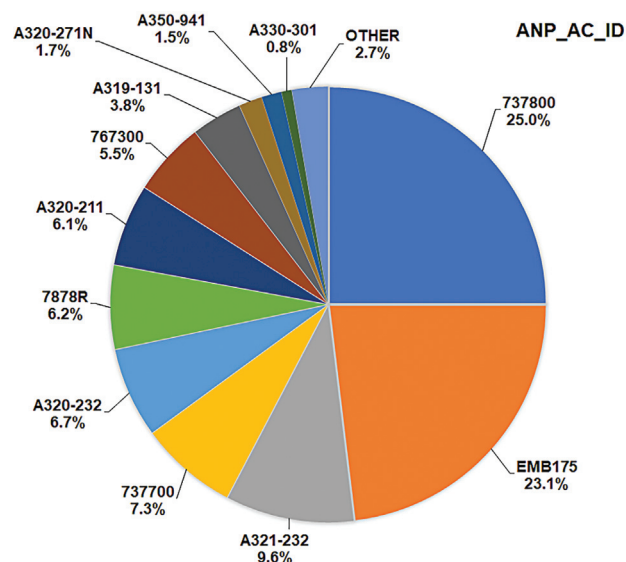


FIG. 5. Relative aircraft mix observed for SFO 28L/R approach by ANP performance model code.

TABLE I. Size of AEDT-R and AEDT-AE cohorts for each route segment.

Data volume statistics:	SFO Runway 28L	SFO Runway 28R	SIDBY SERFR-DIRECT
Number of input AEDT/SLM pairs	226 876	226 876	113 113
Number of pairs skipped as GA	4927	4927	2145
Number of pairs skipped for low GoF	55 861	55 861	70 130
Number of pairs skipped for multiple PCAs	9756	9756	525
Number of pairs skipped for trajectory criteria	58 282	105 988	14 020
Number of post-filter pairs	98 050	50 344	26 293
Number BADA 3 pairs	55 793	30 214	14 112
Number BADA 4 pairs	42 257	20 130	12 181

arrival time of the sound maximum at the *Point of Closest Approach* to the SLM could be attributed to more than one aircraft. The *pairs skipped for trajectory criteria* is the count of pairs that do not conform geometrically to the arrival flight path criteria (i.e., distance, elevation, altitude, speed, heading, etc.)

IV. AEDT-R RESULTS

A. SFO 28L AEDT-R (ANP BADA 3) metric value and difference assessment

In Sec. I we mentioned the FAA “regulatory” mode of AEDT analyses (AEDT-R) that must use ANP BADA 3 modeling with “standard AEDT flight profiles,” i.e., without experimentally observed altitude and speed trajectory control point constraints (FAA, 2021b,c). The use of standard profiles essentially imposes very similar flight paths (using predefined altitude and speed profiles) on all aircraft within a given ANP performance model. This results in essentially constant values (allowing for small changes in aircraft weight) for predicted L_{Amax} and SEL metrics for a given

aircraft type at a given SLM location, independent of the actual flight profile. This procedure ignores the actual physical dynamics of individual flights that BADA 4 modeling seeks to include.

A scatterplot of AEDT-R SEL values versus SLM SEL values from SFO runway 28L traffic is shown in Fig. 6. The purple dots are located at coordinates given by matched pairs of SLM SEL measurements (abscissa) and AEDT-R SEL predictions (ordinate). There are 55 793 such data points in the plot. The diagonal turquoise line is the ideal plot that would result if the SLM measurements and AEDT predictions were in perfect agreement.

As is evident in the figure, for each labeled ANP performance model, there is a horizontal band of nearly constant AEDT-R SEL values, independent of the much broader range of SLM SEL values measured for the actual flight profiles flown. The red dots mark the centroids of the SLM SEL value distributions for each aircraft type and the vertical distance from each dot to the diagonal line indicates how much each AEDT-R SEL prediction would have to change (generally increase) to match the average SLM value.

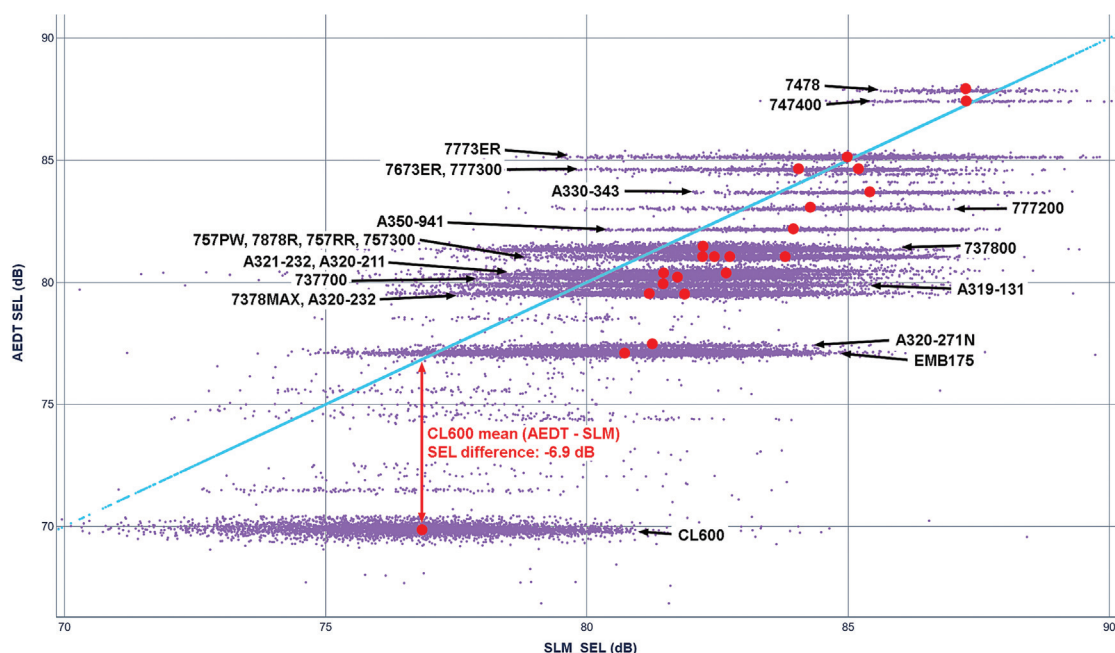


FIG. 6. AEDT-R, SFO NMT-12, Runway 28L. Scatter plot of predicted versus measured SEL values for those ANP models with at least 150 examples in the sub-cohort.

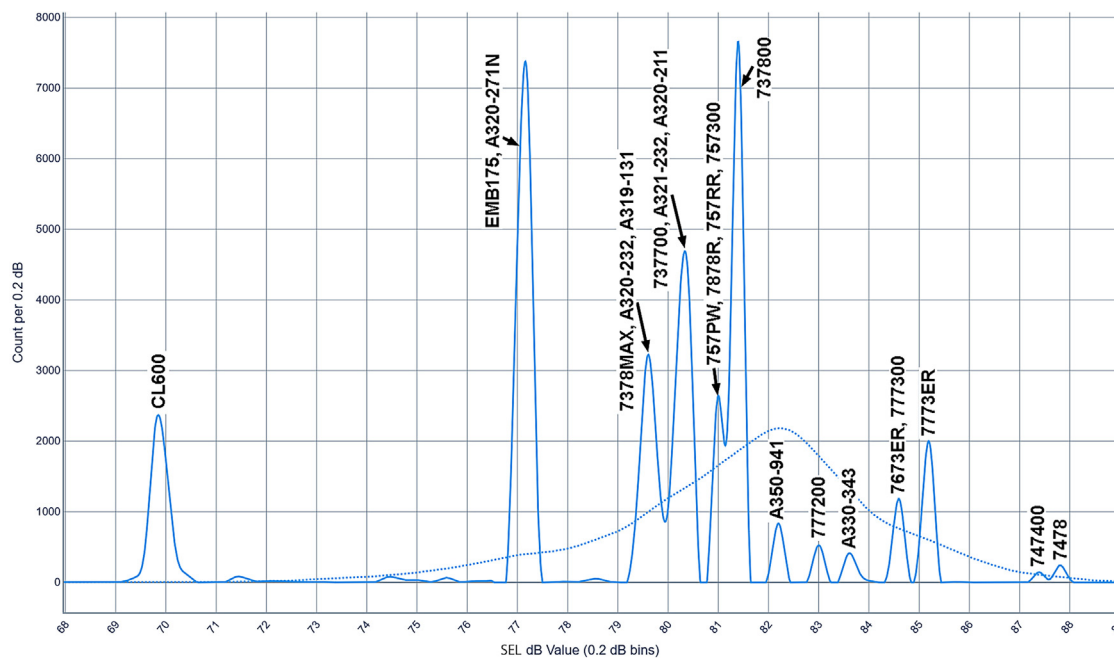


FIG. 7. (Color online) AEDT-R, SFO NMT-12, Runway 28L. Marginal distributions of predicted SEL values showing subpeaks for different ANP models (solid line) and aggregate measured SLM SEL values for all ANP models (dotted line).

To better understand what is going on for individual performance models, Fig. 7 shows histograms of the separate marginal distributions of the scatterplot in Fig. 6.¹ The solid blue curve is the AEDT-R SEL distribution projected along the ordinate. The dotted blue curve is the aggregate SLM SEL distribution for *all* performance models projected along the abscissa. The fact that the AEDT-R SEL histogram peaks are largely separated is a result of the simplistic way in which AEDT-R forces each aircraft performance model type to fly its prescribed model flight profile. The peaks with multiple types assigned means they are not fully resolved with the 0.2 dB histogram bins shown. The much broader aggregate SLM distribution on the other hand shows no resolution of subprofiles by aircraft type in this figure.

Still, there is an SLM SEL subdistribution for each model type that can be derived from the data set by selecting the SLM SEL values from the cohort of AEDT/SLM pairs corresponding to flights with a given aircraft model type. AEDT/SLM paired histogram examples for five major peaks are illustrated in Fig. 8.

Each example model pair is annotated with the ANP performance model AEDT-R used for predicting SEL values (solid lines). The corresponding SLM measurement distribution is shown with a similarly annotated dotted line. The average AEDT-R subpeak standard deviation width is *very* narrow, 0.31 dB, indicating that ANP/BADA 3 uses a very simple fixed flight profile description for each model type. The average SLM subpeak standard deviation width is much broader, 1.60 dB, indicating significant variations in SLM SEL measurements for real-world flight profiles for a given aircraft type. Figure 9 shows histograms of the actual distributions of AEDT – SLM differences for the illustrative ANP models.

As can be seen in Fig. 9, the mean per-model differences range from 0.15 dB to -6.89 dB, with other model differences scattered in between. The average standard deviation for the SLM/model difference distributions is 1.60 dB, a figure dominated by the SLM distribution widths.² The results of the analyses for all of the ANP models seen in the AEDT-R SEL predictions for the SFO 28L arrival fleet mix, are detailed in Fig. 10 (ordered by absolute difference value) and Table II (ordered alphabetically by performance model type).

An analysis of the L_{Amax} metric is done in an exactly analogous way; working with the ensemble of AEDT-R/SLM pairs and separating the pairs according to their ANP performance model. The BADA 3 L_{Amax} difference data by ANP model are also summarized in Fig. 10 and Table II. The average (AEDT – SLM) SEL difference for the entire cohort is -2.37 ± 2.51 dB. The average (AEDT – SLM) L_{Amax} difference is -3.64 ± 2.26 dB. The average model L_{Amax} errors are larger than those for the SEL metric by ~ 1.3 dB, as might be expected for a point-based metric (maximum peak value at the PCA) as opposed to an integrated metric. Still the large relative uncertainty in the average error is about the same, 2.5 dB in SEL versus 2.3 dB in L_{Amax}.

B. Discussion of AEDT-R results

The validation results for the SFO 28R route segment are similar: average (AEDT – SLM) L_{Amax} difference is -3.54 ± 2.60 dB and average SEL difference is -1.76 ± 2.76 dB. The validation results for the SIDBY SERFR-DIRECT route segment are somewhat different: the average (AEDT – SLM) L_{Amax} difference is -1.85 ± 4.08 dB and the average SEL difference is 1.32 ± 3.58 dB. The average differences are

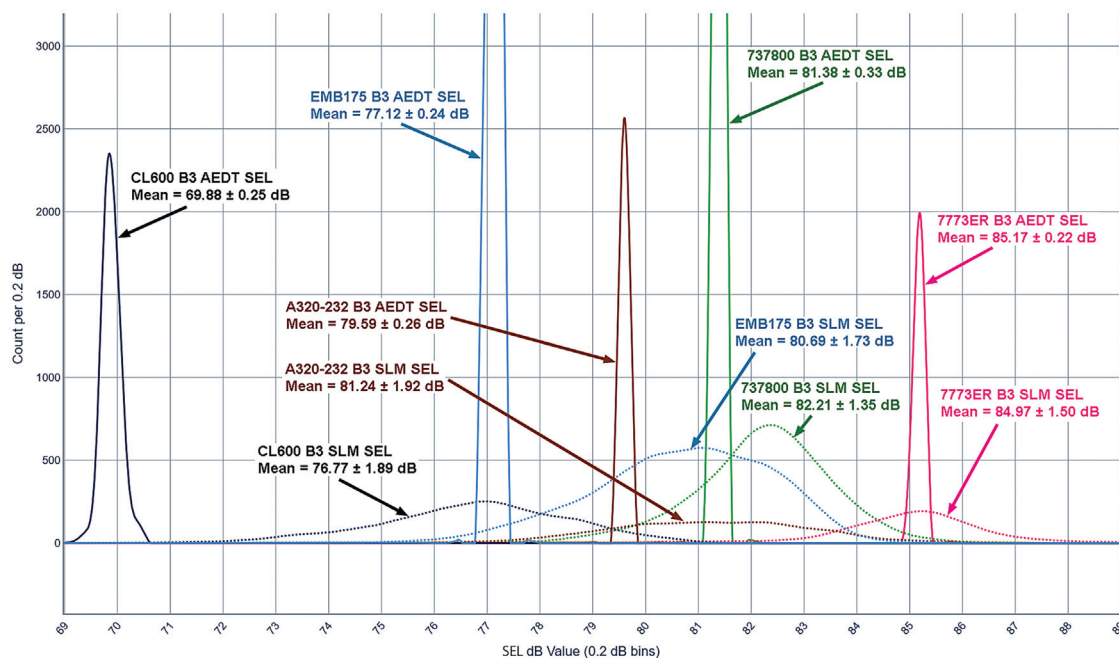


FIG. 8. AEDT-R, SFO NMT-12, Runway 28L. Distributions of predicted SEL (solid lines) and measured SLM (dashed lines) for selected ANP models.

smaller, but the standard deviations are larger. We believe the larger standard deviations are because the SIDBY SERFR-DIRECT route segment is a transition region between the SERFR RNAV procedure and air traffic control guidance. In this region, many pilots decide to deploy auxiliary lift and speed control equipment early, increasing airframe noise, and SLM measurements are more variable because of the longer slant distance. These effects cause more variable ground

noise levels, including higher SLM noise amplitudes for flights with auxiliary equipment deployed. However, this leaves open the question of why the average difference between AEDT-R and SLM noise level is less (rather than greater) in such a region. One hypothesis is that something is amiss in the way AEDT-R estimates the flight altitude and ground speed parameters not supplied by ADS-B measurements in this AEDT-R mode.

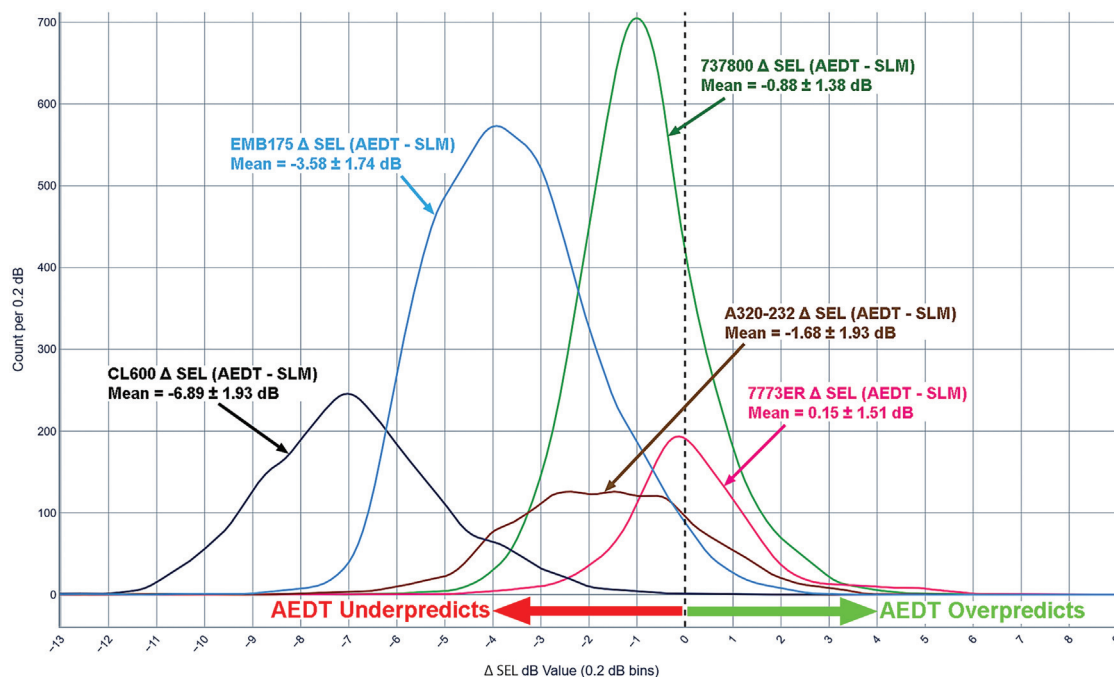


FIG. 9. AEDT-R, SFO NMT-12, Runway 28L. Distributions of SEL differences (AEDT-SLM) for selected ANP models.

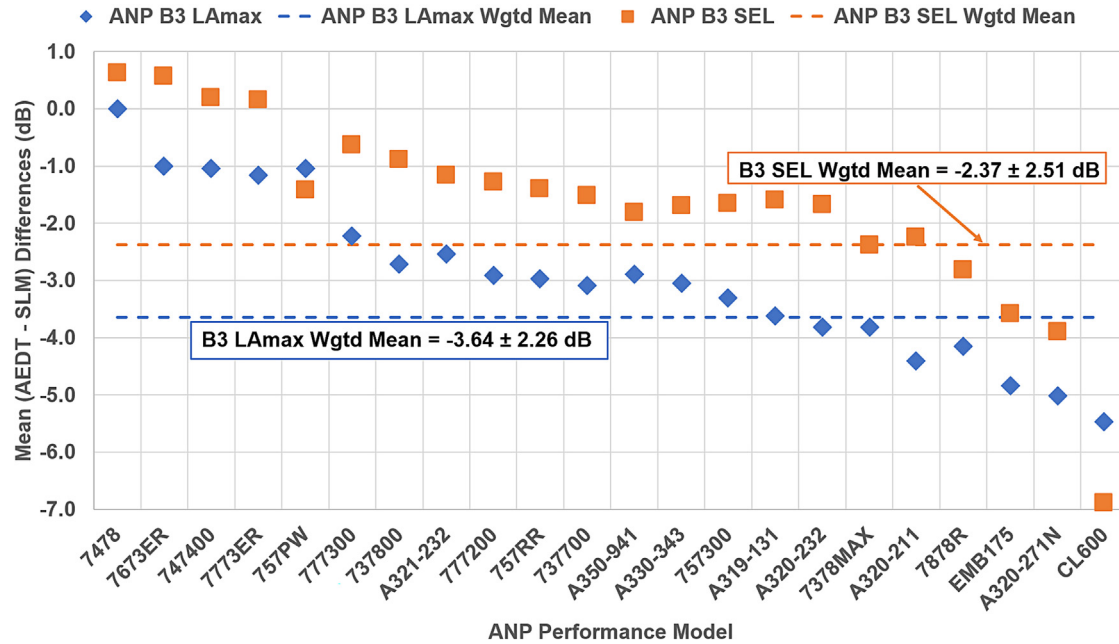


FIG. 10. (Color online) AEDT-R, SFO NMT-12, Runway 28L. Mean LMax and SEL differences (AEDT-SLM), by ANP model.

To explore this hypothesis, we created scatter cross-plots of AEDT estimates of altitude and ground speed against the actual ADS-B measurements for SIDBY SERFR-DIRECT flights. Figure 11 shows the scatter cross-plot for altitudes. The blue line indicates where AEDT and ADS-B altitudes match exactly.

The scatterplot shows the expected horizontal bands where AEDT-R estimates roughly constant altitudes for a

given ANP model. Unexpectedly though, there are three main groups of widely separated bands; one around the nominal RNAV/ATC altitude (4750 feet), another broader set centered at about 3600 feet, and a third centered just above the lower limit (3000 feet) of the class C airspace. We can create a composite histogram of the marginal distributions for AEDT ordinate and ADS-B abscissa altitude values from the scatterplot as shown in Fig. 12.

TABLE II. AEDT-R, SFO NMT-12, Runway 28L. Mean LMax and SEL differences (AEDT-SLM) by ANP model.

ANP code	BADA code	Subcohort size	AEDT-SLM LMax difference	AEDT-SLM SEL difference
ALL	All	55 793	-3.64 ± 2.26	-2.37 ± 2.51
737700	B737	3385	-3.08 ± 1.87	-1.52 ± 1.50
737800	B738, B739	10 224	-2.71 ± 1.65	-0.88 ± 1.38
7378MAX	B38M, B39M, B738	1622	-3.82 ± 1.72	-2.38 ± 1.51
747400	B744	178	-1.04 ± 1.60	0.20 ± 1.36
7478	B748	304	0.00 ± 1.03	0.64 ± 0.89
757300	B753	509	-3.31 ± 1.88	-1.65 ± 1.76
757PW	B752	211	-1.04 ± 2.17	-1.41 ± 1.85
757RR	B752	393	-2.97 ± 1.83	-1.40 ± 1.64
7673ER	B763	1214	-1.01 ± 1.81	0.57 ± 1.58
777200	B772	552	-2.90 ± 2.08	-1.27 ± 1.79
777300	B77L	194	-2.23 ± 1.49	-0.62 ± 1.23
7773ER	B77W	2686	-1.16 ± 1.74	0.15 ± 1.51
7878R	B788, B789, B78X	2487	-4.14 ± 1.81	-2.81 ± 1.63
A319-131	A319	2262	-3.62 ± 2.20	-1.60 ± 1.94
A320-211	A320	2076	-4.41 ± 1.81	-2.24 ± 1.57
A320-232	A320	3260	-3.81 ± 2.22	-1.68 ± 1.93
A320-271N	A20N	662	-5.01 ± 1.69	-3.89 ± 1.48
A321-232	A321	3175	-2.53 ± 1.83	-1.16 ± 1.57
A330-343	A332, A333, A339	577	-3.04 ± 1.97	-1.69 ± 1.55
A350-941	A359, A35K	972	-2.88 ± 2.14	-1.81 ± 1.77
CL600	CL30, CL60	5210	-5.46 ± 1.92	-6.89 ± 1.93
EMB175	E75L	12 635	-4.83 ± 1.89	-3.58 ± 1.74

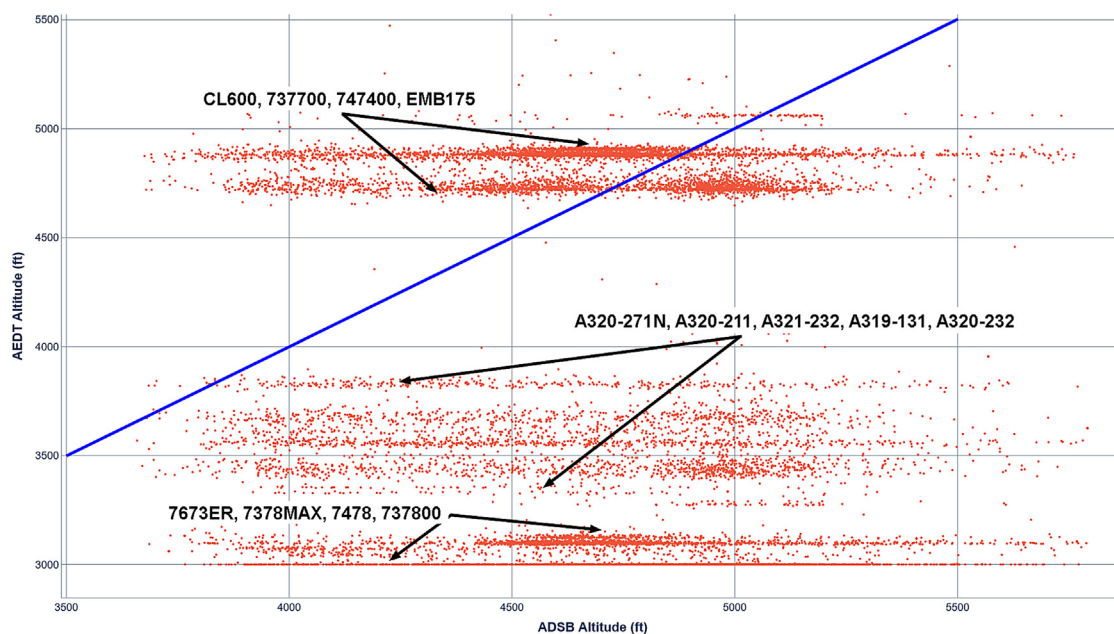


FIG. 11. (Color online) AEDT-R, SIDBY, SERFR-DIRECT. Scatter plot of AEDT altitude versus ADS-B altitude for ANP models with at least 150 examples in the sub-cohort.

The three aircraft clusters—those near the nominal route-segment-specified altitude, those near the bottom of the class C airspace, and those in between—clearly stand out. We also show the ADS-B altitude distribution for all aircraft (light blue dotted trace), which has a mean altitude of 4679 ± 369 feet. The average altitude for the near-nominal types is 192 feet above the ADS-B mean (0.52 times the ADS-B sigma). The average altitude for the cluster

at the class C airspace lower bound is 1570 feet below the ADS-B mean (4.25 sigma). The middle cluster is 973 feet below the ADS-B mean (2.64 sigma). The errors in the latter two anomalous altitude clusters are hard to explain other than that some of the ANP flight profile models do not lend themselves to accurate altitude estimations farther from the airport. The SFO NMT-12 SLM is located about 6 miles upstream from the airport, whereas the SIDBY SLM is

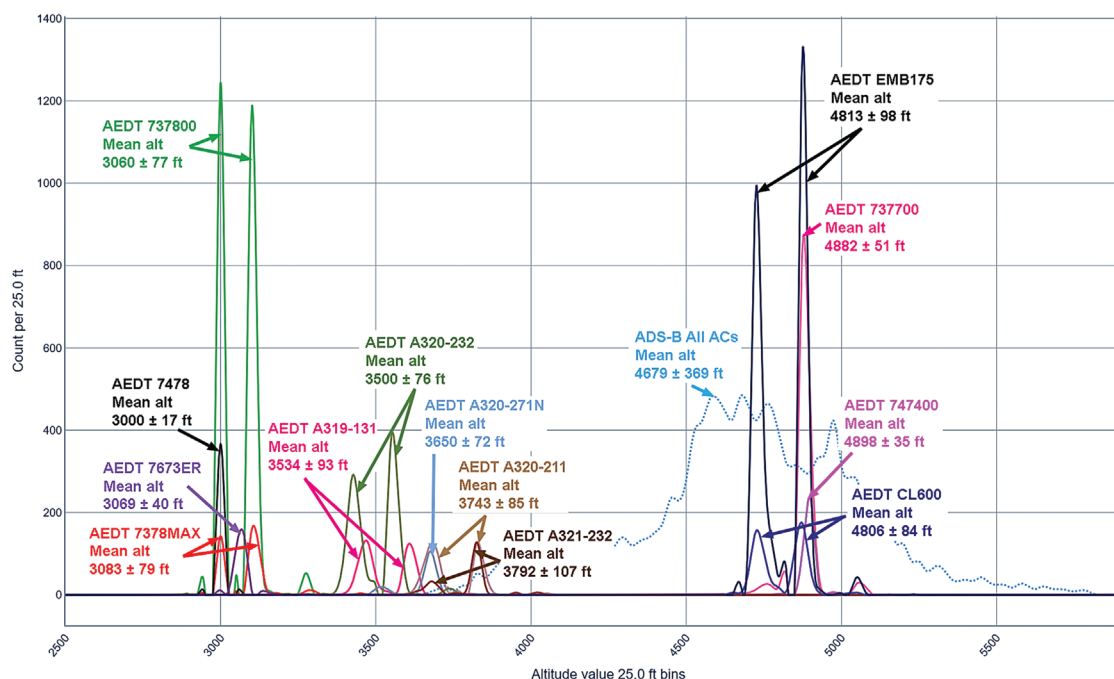


FIG. 12. AEDT-R, SIDBY, SERFR-DIRECT. Distributions of predicted altitudes for selected aircraft (solid lines) and composite measured ADS-B altitudes (dashed line) for all aircraft types.

about 12 miles upstream along the SERFR-DIRECT approach path. Some estimates at SIDBY are close to the ADS-B mean (e.g., for the 747400 model) but others for similar aircraft are much further off (e.g., for the 7478 model). These altitude discrepancies have a strong effect on the AEDT-R estimated L_{Amax} and SEL values because of the $1/r^2$ and atmospheric attenuation effects dictated by physics, i.e., the AEDT-computed noise levels for those lower altitude models will be significantly higher than if they overflew at the nominal higher altitude for the procedure. Therefore, the calculated noise levels for those flights are closer to the measured SLM noise levels and the difference is less. There are indications that these altitude estimation errors are intrinsic to AEDT-R, and probably exist to some extent at all distances from the airport, but are more extreme further upstream. Note that these altitude anomalies do not show up in the AEDT-AE calculations discussed later because explicit ADS-B altitude and airspeed values are included in the control point data passed explicitly to AEDT-AE.

Another interesting feature of the AEDT-R altitude estimates is that for some ANP aircraft types (e.g., 737800, 7378MAX, A319-131, A320-232, EMB175, and others) two altitude peaks are seen in the marginal distributions (solid lines). The elements of the pairs are 100–150 feet apart, which is about the difference in average approach altitudes for SFO runway 28L and 28R. We interpret this as an indication of the separate AEDT-R projections of the ANP flight profiles upstream for flights landing on 28L and 28R, respectively. It is not clear why some aircraft profiles (e.g., 747400, 7478, 7673ER, and 77700) do not make this distinction.

We do not see such large discrepancies in the AEDT-R estimates of ground speeds at the SIDBY SLM. The mean ADS-B ground speed for all aircraft is 229 ± 22 kn. There are two AEDT-R ground speed clusters, one composed of the same aircraft types clustered near the nominal route segment altitude and another composed of the aggregate aircraft types in the two altitude clusters well below the nominal route altitude. The former has an average ground speed about 3 kn above the ADS-B mean and the latter an average ground speed about 26 kn above the ADS-B mean.

These AEDT-R flight parameter estimation errors, dominated by the larger altitude errors, explain why the L_{Amax} and SEL estimates for SIDBY SERFR-DIRECT appear to be closer to the ground SLM measurements. The reason for this result is not because AEDT-R is making a better estimate. In fact, this indicates a serious bug in AEDT-R.

The conclusion drawn from these analyses is that the AEDT-R metric predictions are essentially constant for a given ANP performance model and do not reflect the actual flight or weather conditions. The predictions for a given model would approximate actual SLM measurements only if the mean of the AEDT-R predictions for that model matched the mean of the SLM measurements. At the SFO NMT-12 SLM, this is almost the case for “heavies” such as ANP models for 747400, 7478, 7673ER, and 7773ER types. For other (more frequent) models though, the AEDT-R

predictions fall more randomly short of the measured SLM values. If these means do not match up model by model, AEDT-R will produce unpredictable and unreliable metric estimates that will depend on the fleet mix specifics and distance from the airport. The often-discussed possibility that averaging over very large numbers of flights and aircraft types makes the results of AEDT-R predictions more accurate is mathematically implausible under the observed circumstances. For these reasons, we conclude that AEDT-R has substantial prediction errors on approach trajectories and cannot be considered a reliable methodology for predicting valid aircraft noise impact.

V. AEDT-AE RESULTS

A. SFO 28L AEDT-AE (BADA 4) metric value and difference analysis

We begin the statistical analysis of the pair-by-pair relations between AEDT-AE L_{Amax} estimates and the corresponding SLM measurements by showing a series of histogram plots like we displayed earlier for the AEDT-R runs. Using BADA 4 data from final approaches to SFO runway 28L (42 257 pairs including 14 ANP performance models), Fig. 13 shows histograms of the AEDT-AE L_{Amax} estimates (solid line) and of the SLM L_{Amax} measurements (dotted line). As the histogram labels indicate, the average AEDT-AE L_{Amax} level is 68.59 ± 1.52 dB and that measured by the SLM is 71.32 ± 2.15 dB—on average AEDT underestimates the L_{Amax} value by 2.73 dB. Also, note that the profile of the predicted AEDT values has significant internal structure, indicated by the subpeaks, including one sidelobe at 70.95 dB, and inflection points in the non-Gaussian shape. No discrete subpeaks are present however, as were prominent in the AEDT BADA 3 predictions. This reflects the more sophisticated physics modeled in AEDT BADA 4 and the spread in the trajectories actually flown that are considered in this analysis.

This structure can be elaborated by overplotting histograms for various aircraft subcohorts modeled by particular ANP performance models as shown in Fig. 14. To simplify the plot, only three major aircraft types are shown. One can see that the AEDT-AE calculations produce histogram profiles with quite varied shapes and that the differences in mean L_{Amax} values between the AEDT-AE predictions and the SLM measurements differ significantly (see the green profile for ANP model 737800 in particular).

It is evident from this plot (and by extension for the entire cohort of aircraft types and ANP performance models) that there are important model-based differences between the estimates AEDT-AE makes and what the SLM measures. These are important in that for the most part the AEDT-AE predictions fall short of the SLM measurements—a result analogous to that seen for the AEDT-R BADA 3 analysis above.

We can illustrate the AEDT–SLM differences and their distributions by performance model by computing the metric difference directly for each AEDT/SLM pair and then

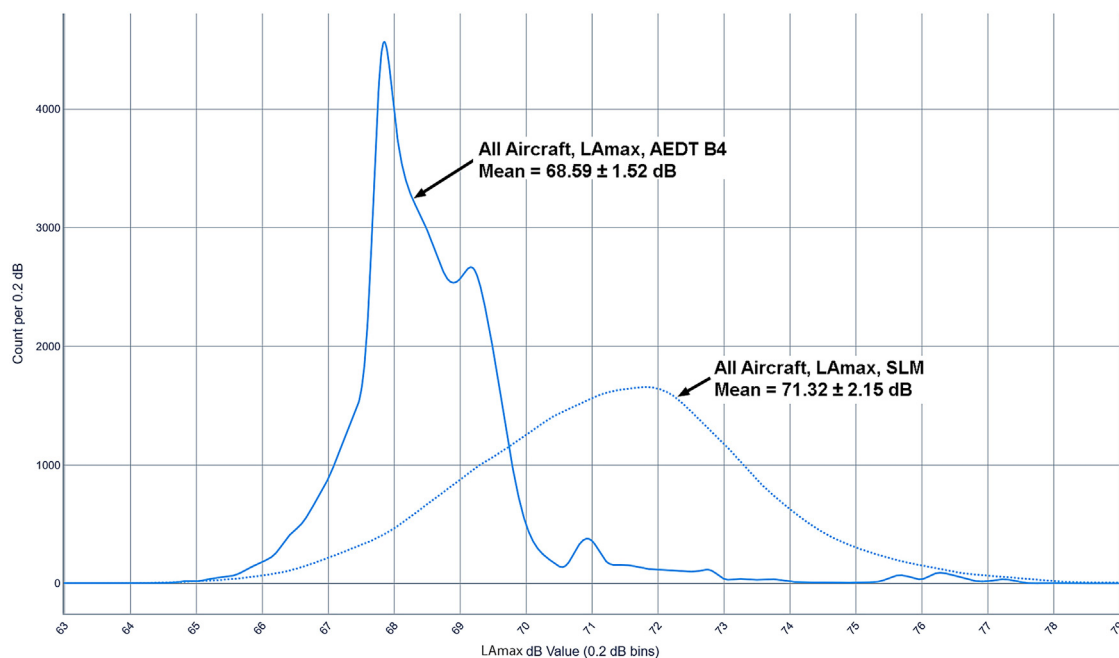


FIG. 13. (Color online) AEDT-AE, SFO NMT-12, Runway 28L. Distributions of predicted LMax (solid line) and measured LMax (dashed line) values for all ANP models.

plotting the histogram of those differences. This comparison is illustrated in the histograms in Fig. 15, both for the overall cohort and for a subset of individual ANP performance models. These differences range from -2.08 dB for the A321-232 performance model to -3.44 dB for the 737800 model, all with standard deviations ~ 2 dB (which result from the relatively broad SLM measurement distributions).

Based on the analyses illustrated in Figs. 13–15, we compute the difference error statistics for all 14 ANP performance models found in the aircraft cohort. Figure 16 shows an ordered plot of the error values and Table III shows the detailed sample flight counts, difference values, and standard deviations. The SEL metric data we have collected from AEDT-AE study runs and SLM measurements are analyzed with exactly the same approach we used for LMax;

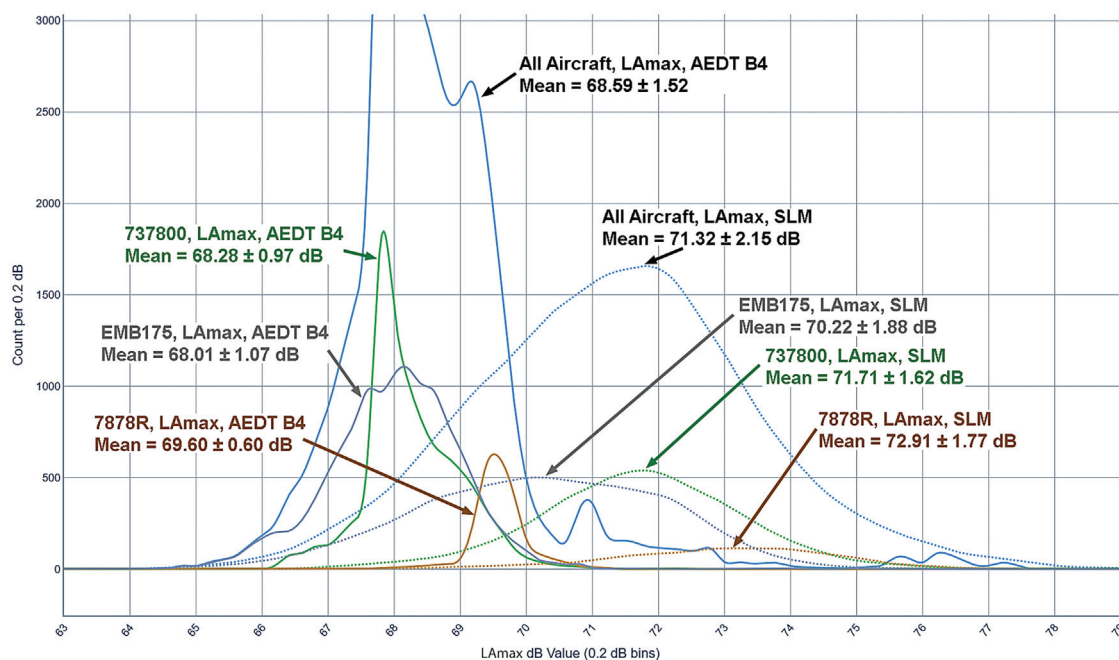


FIG. 14. AEDT-AE, SFO NMT-12, Runway 28L. Distributions of predicted LMax (solid line) and measured LMax (dashed line) values for selected ANP models. Vertical scale has been expanded over that in Fig. 13 to show subcohort data more clearly.

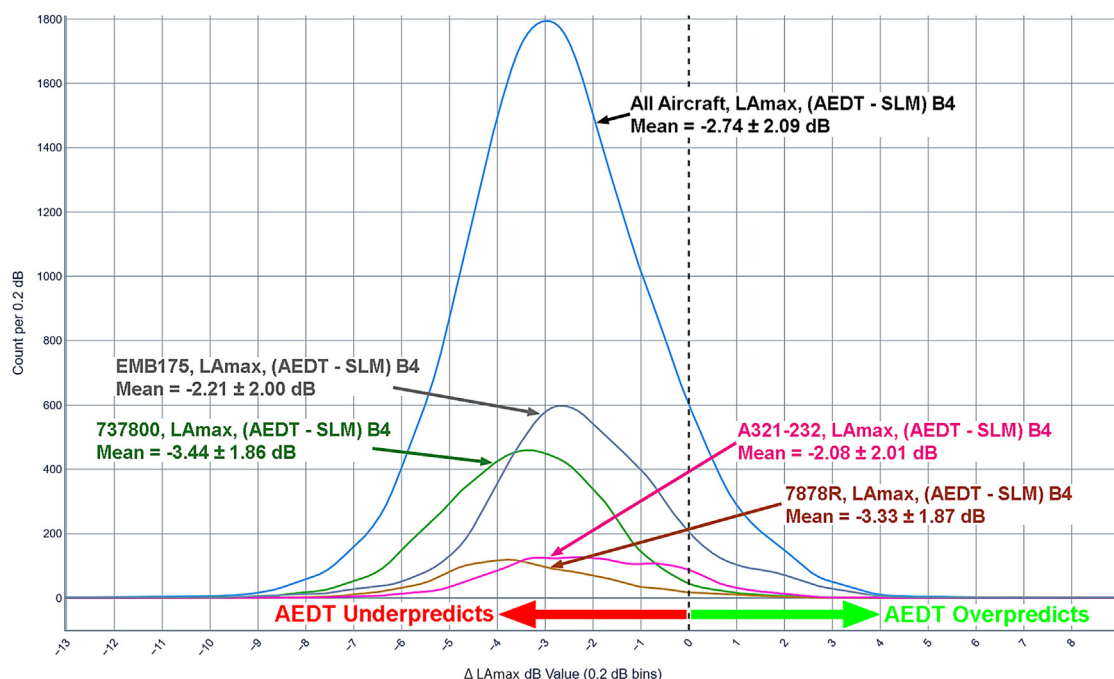


FIG. 15. (Color online) AEDT-AE, SFO NMT-12, Runway 28L. Distributions of LAmix differences (AEDT-SLM) for entire cohort and for selected ANP models.

the results for the SFO 28L approach are also shown in Fig. 16 and Table III.

As is evident in Fig. 16, there are major differences in the accuracy of the AEDT-AE LAmix predictions. For example, the differences for “heavy” aircraft (ANP models 747400 and 7478) are relatively small, indicating that the corresponding models seem to be fairly accurate. For the other aircraft though, the accuracy of the AEDT-AE modeling appears to produce systematically

low estimates, ranging from ~ -1.6 dB to -3.9 dB. The overall LAmix error weighted by frequency counts is -2.74 dB.

The overall difference between AEDT-AE SEL predictions and SLM measurements for the SFO 28L approach is slightly less, -2.48 dB. On the other hand, the same basic observations about their variable nature and the inability of AEDT-AE to model sound metrics accurately for different ANP performance models apply.

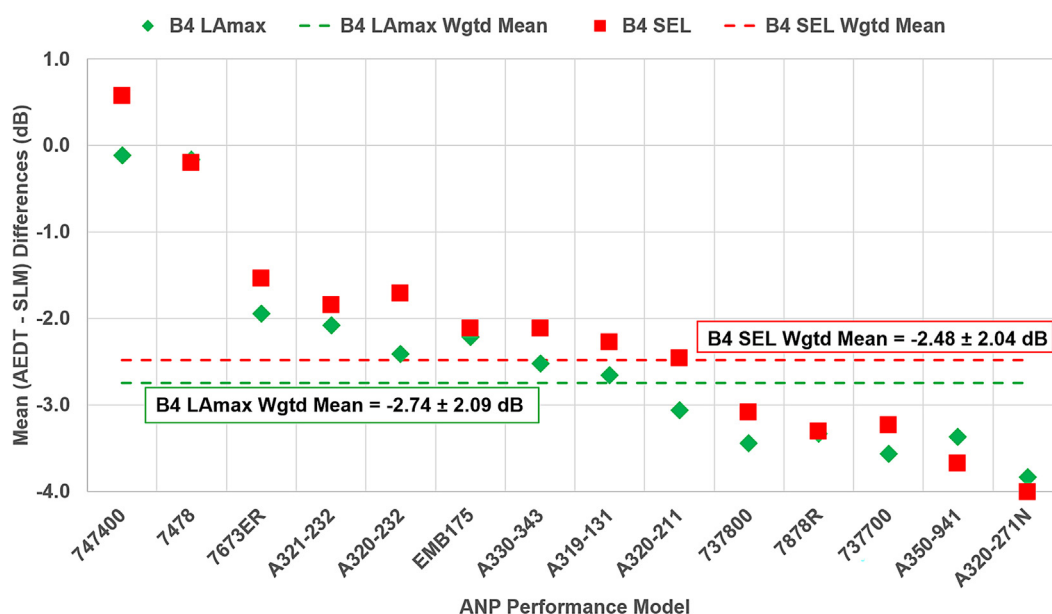


FIG. 16. (Color online) AEDT-AE, SFO NMT-12, Runway 28L. Mean LAmix and SEL differences (AEDT-SLM), by ANP model. The green diamonds for 7478 and 7878 R LAmix are hidden behind the corresponding red SEL squares.

TABLE III. AEDT-AE, SFO NMT-12, Runway 28L. Mean and standard deviations of LAmax and SEL differences (AEDT – SLM), by ANP model.

ANP code	BADA code	Subcohort size	AEDT–SLM LAmax difference	AEDT–SLM SEL difference
ALL	All	42 257	-2.74 ± 2.09	-2.48 ± 2.04
737700	B737	2 685	-3.57 ± 1.78	-3.23 ± 1.66
737800	B738, B739	9 793	-3.44 ± 1.86	-3.09 ± 1.72
747400	B744	177	-0.11 ± 1.65	0.57 ± 1.55
7478	B748	304	-0.16 ± 1.35	-0.20 ± 1.30
7673ER	B763	1 212	-1.94 ± 2.06	-1.54 ± 1.96
7878R	B788, B789, B78X	2 408	-3.33 ± 1.87	-3.31 ± 1.77
A319-131	A319	2 251	-2.65 ± 2.22	-2.27 ± 2.12
A320-211	A320	2 072	-3.06 ± 1.83	-2.46 ± 1.76
A320-232	A320	3 196	-2.41 ± 2.19	-1.71 ± 2.05
A320-271N	A20N	661	-3.84 ± 1.96	-4.00 ± 1.85
A321-232	A321	3 167	-2.08 ± 2.01	-1.84 ± 1.83
A330-343	A332, A333, A339	507	-2.52 ± 1.98	-2.12 ± 1.79
A350-941	A359, A35K	970	-3.37 ± 2.43	-3.68 ± 2.36
EMB175	E75L	12 605	-2.21 ± 2.00	-2.12 ± 2.08

B. SFO 28R AEDT-AE (BADA 4) metric value and difference analysis—Brief summary

We analyzed data for the final approach to SFO runway 28R just as for runway 28L. Figure 17 and Table IV summarize the comparisons between AEDT-AE predictions and SLM measurements of LAmax and SEL values.

C. SIDBY SERFR-DIRECT AEDT-AE (BADA 4) metric value and difference analysis—Brief summary

We analyzed data for the SERFR-DIRECT approach over the SIDBY waypoint just as for the final approach to SFO runway 28L. Figure 18 and Table V summarize the comparisons between AEDT-AE predictions and SLM measurements of LAmax and SEL values.

D. Calibrated air speed effects on modeled metric values

We noted an interesting anomaly between AEDT-AE predicted and SLM measured LAmax and SEL pairs that can be seen in a pair-wise scatterplot against Calibrated Air Speed (CAS). Figure 19 shows data for two single-aisle aircraft types (ANP performance models 737700 and 737800) on final approach to runway 28L, as observed at SFO NMT-12. It is clear that AEDT underestimates both LAmax and SEL as noted in the analyses presented above. However, there is another clearly visible effect.

With increasing CAS and all other factors remaining constant, physics tells us that sound metrics should increase primarily because of the dependence of airframe noise on airspeed. This effect is indeed observed in the SLM

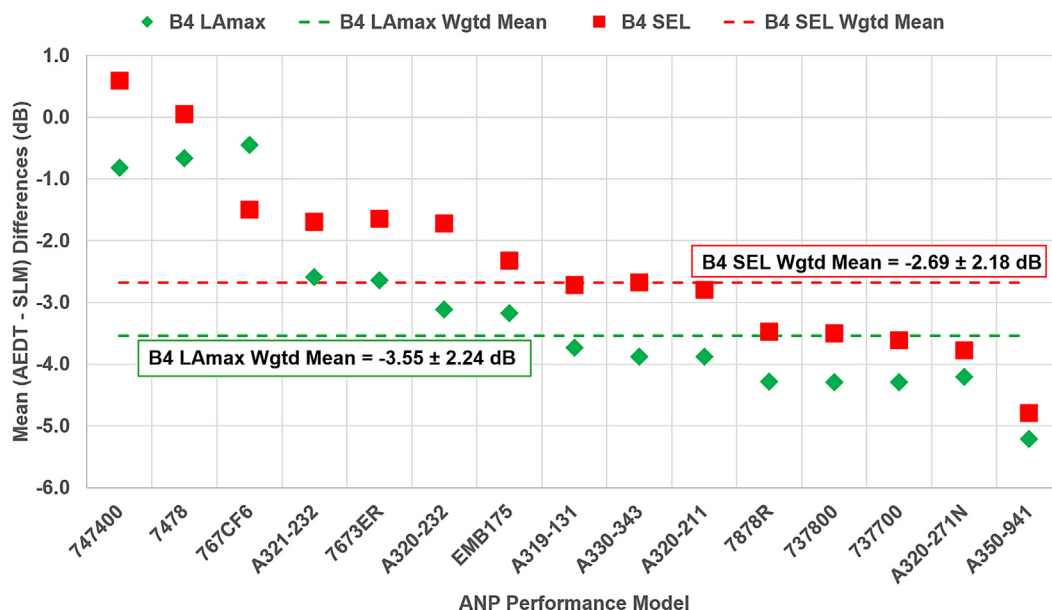


FIG. 17. (Color online) AEDT-AE, SFO NMT-12, Runway 28R. Mean LAmax and SEL differences (AEDT–SLM), by ANP model.

TABLE IV. AEDT-AE, SFO NMT-12, Runway 28R. Mean and standard deviations of LAmax and SEL differences (AEDT – SLM), by ANP model.

ANP code	BADA code	Subcohort size	AEDT–SLM LAmax difference	AEDT–SLM SEL difference
ALL	All	20 130	-3.55 ± 2.24	-2.69 ± 2.18
737700	B737	870	-4.29 ± 1.68	-3.61 ± 1.62
737800	B738, B739	5 051	-4.29 ± 1.79	-3.50 ± 1.73
747400	B744	269	-0.82 ± 1.96	0.59 ± 1.62
7478	B748	296	-0.66 ± 1.58	0.05 ± 1.52
7673ER	B763	1 500	-2.64 ± 2.33	-1.64 ± 2.13
767CF6	B762	190	-0.45 ± 2.24	-1.50 ± 2.12
7878R	B788, B789, B78X	1 883	-4.28 ± 2.10	-3.47 ± 2.04
A319-131	A319	860	-3.73 ± 2.37	-2.72 ± 2.25
A320-211	A320	1 060	-3.88 ± 1.74	-2.79 ± 1.68
A320-232	A320	1 013	-3.11 ± 2.19	-1.72 ± 2.13
A320-271N	A20N	290	-4.20 ± 2.32	-3.77 ± 2.03
A321-232	A321	2 290	-2.59 ± 1.91	-1.69 ± 1.75
A330-343	A332, A333, A339	247	-3.88 ± 1.76	-2.67 ± 1.70
A350-941	A359, A35K	502	-5.21 ± 2.31	-4.79 ± 2.19
EMB175	E75L	3 528	-3.17 ± 2.17	-2.32 ± 2.10

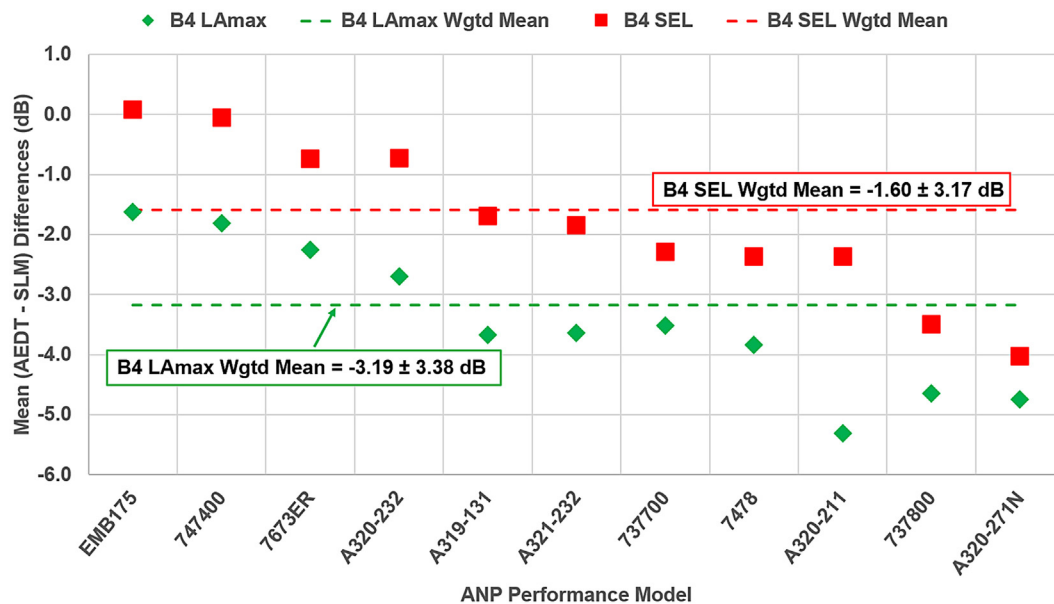


FIG. 18. (Color online) AEDT-AE, SIDBY, SERFER-DIRECT. Mean LAmax and SEL differences (AEDT–SLM), by ANP model.

TABLE V. AEDT-AE, SIDBY, SERFER-DIRECT. Mean and standard deviations of LAmax and SEL differences (AEDT–SLM), by ANP model.

ANP code	BADA code	Subcohort size	AEDT–SLM LAmax difference	AEDT–SLM SEL difference
ALL	All	12 181	-3.19 ± 3.38	-1.60 ± 3.17
737700	B737	1 546	-3.52 ± 3.24	-2.29 ± 2.93
737800	B738, B739	2 895	-4.65 ± 2.99	-3.49 ± 2.72
747400	B744	403	-1.81 ± 2.35	-0.05 ± 2.04
7478	B748	366	-3.84 ± 2.32	-2.36 ± 1.99
7673ER	B763	312	-2.25 ± 3.11	-0.74 ± 2.86
A319-131	A319	576	-3.67 ± 3.66	-1.69 ± 3.17
A320-211	A320	546	-5.31 ± 2.61	-2.36 ± 2.36
A320-232	A320	1 323	-2.70 ± 3.28	-0.73 ± 2.90
A320-271N	A20N	243	-4.74 ± 2.58	-4.02 ± 2.47
A321-232	A321	297	-3.64 ± 3.06	-1.84 ± 2.82
EMB175	E75L	3 608	-1.62 ± 3.28	0.08 ± 2.97

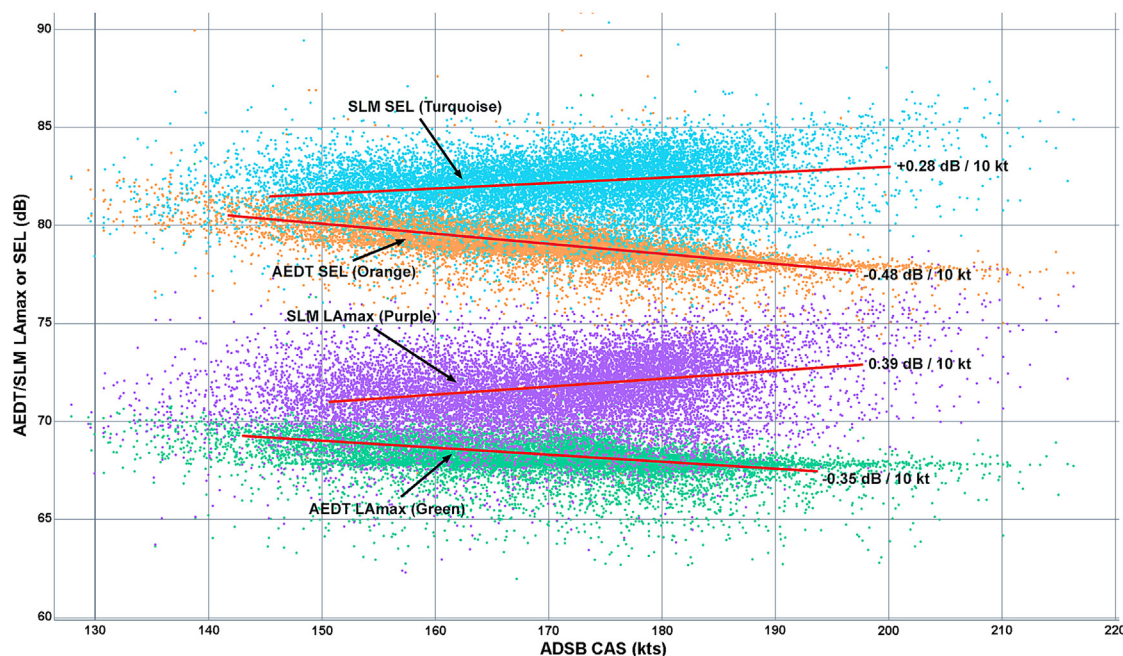


FIG. 19. (Color online) AEDT-AE, SFO NMT-12, Runway 28L. Plots of predicted and measured LMax and SEL against CAS showing decrease in predicted noise metrics at higher speeds versus increase in measured noise metrics.

measurements, but AEDT-AE predicts a *lessening* of noise metrics. The plot shows linear trend lines to illustrate and quantify the rate of metric change, but we note that the effect is more complicated than a linear one. We tentatively attribute this observation to the fact that AEDT-AE modeling is based on engine noise estimates, which may in fact decrease as aircraft airspeed changes due to thrust levels, whereas airframe noise sources, especially from auxiliary lift equipment and landing gear deployment, will increase and eventually overpower engine noise, particularly on approach trajectories (Lopes and Burley, 2011). As discussed in our conclusions, it appears that a more complex and accurate modeling of the physics, particularly of airframe noise, will be needed to make AEDT-AE better conform to ground truth in aircraft arrival situations. Similar results are found for the other route segments (final approach to SFO runway 28R and SERFR-DIRECT over the SIDBY waypoint).

E. Discussion of AEDT-AE results

For all three route segments, our analyses showed similar AEDT/SLM variation by ANP performance model. AEDT-AE consistently predicts LMax and SEL values well below the levels measured by the SLMs.

Given the complex nature of the AEDT internal calculations, it is difficult to divine exactly the reason for the various underpredictions we observe. We believe that the differences are largely attributable to shortcomings in the SAE-AIR-1845/Doc 9911 model as implemented in AEDT, an underestimation of airframe noise contributions, and a lack of exact knowledge of both the aircraft weight and the atmospheric conditions. We believe the SLM measurements are as accurate as physically possible. On the AEDT side, there are many areas of unknown accuracy such as the validity of NPD curves supplied for certification, and the fact that the modeling is relatively simplistic. For example, without FOQA data, we do not know details of an aircraft's weight, the thrust employed along the flight profile, when auxiliary lift equipment is deployed, etc. Most often FOQA data are confidential and tightly held by airlines. Weather and atmospheric conditions are similarly hard to control for and would limit the strength of a validation study based on just a handful of flights.

As before with AEDT-R (ANP BADA 3), one cannot expect or rely on the "law of averages" for error cancellation across the fleet mix so that AEDT might be able to produce a more accurate *average* value. In addition, existing AEDT-AE models seem to contradict the expected physics of

TABLE VI. AEDT-R. Mean and standard deviations of LMax and SEL differences (AEDT–SLM) for all three route segments and all ANP models. The SIDBY row (grey shade) values are suspect because of AEDT-R altitude anomalies discussed earlier.

SLM	Route segment	Cohort size	Mean AEDT–SLM LMax differences	Mean AEDT–SLM SEL differences
SFO NMT-12	SFO 28L	55 793	-3.64 ± 2.26	-2.37 ± 2.51
SFO NMT-12	SFO 28R	30 214	-3.54 ± 2.60	-1.76 ± 2.76
SIDBY	SERFR-DIRECT	14 112	-1.85 ± 4.08	-1.32 ± 3.58

TABLE VII. AEDT-AE. Mean and standard deviations of LAmax and SEL differences (AEDT–SLM) for all three route segments and all ANP models.

SLM	Route segment	Cohort size	Mean AEDT–SLM LAmax differences	Mean AEDT–SLM SEL differences
SFO NMT-12	SFO 28L	42 257	-2.74 ± 2.09	-2.48 ± 2.04
SFO NMT-12	SFO 28R	20 130	-3.55 ± 2.24	-2.69 ± 2.18
SIDBY	SERFR-DIRECT	12 181	-3.19 ± 3.38	-1.60 ± 3.17

approach airframe noise generation by not adequately accounting for increased noise levels with increased calibrated airspeed from auxiliary high-lift equipment and landing gear deployment.

VI. CONCLUSIONS

In this paper we have collected and analyzed a very large dataset of over 200 000 observations of pairs of AEDT noise predictions matched with carefully curated SLM noise measurements. We have focused on SFO arrivals along three high-density route segments and our data covers an entire year of observations from July 1, 2021 through June 30, 2022.

Analyses using the FAA AEDT regulatory mode (AEDT-R) based on ANP/BADA 3 performance modeling with standard profiles and without altitude and speed control data are summarized in Table VI. The resulting statistical data indicate that this type of modeling is overly simplistic and gives far from accurate comparison with ground SLM measurements. It is highly troubling that for flights passing the SIDBY receptor, the estimated altitudes for 9 out of 13 significant aircraft types are significantly below the ADS-B measured altitudes by 2.6 to 4.2 times the standard deviation of the ADS-B altitude distribution. It is noteworthy that, despite these deficiencies, AEDT-R is the only FAA-approved regulatory mode for AEDT use.

Our analyses also point to a systematic underestimation by AEDT-AE in its predictions for LAmax and SEL metrics by significant but highly varied amounts depending on aircraft type and performance model. The results for the three route segments are shown in Table VII.

AEDT-AE modeling must be improved by adjusting the internal representation of the applicable physics, e.g., the NPD curves and modeling of engine and airframe noise at various stages of flight. In at least some instances, AEDT-AE predicts *lower* sound metrics (LAmax and SEL) for aircraft with *higher* calibrated airspeeds. This contradicts the physics involved and suggests AEDT-AE does not adequately account for airframe noise sources when auxiliary high-lift equipment is in use or landing gear is deployed. It would be ideal to have access to FOQA data, but that is often considered confidential. It may be possible to guess factors (weight, auxiliary lift configuration, etc.) affecting airframe noise generation on average in order to use more advanced physical models such as the Aircraft Noise Prediction Program (ANOPP2) to better estimate airframe noise (Geissbuhler *et al.*, 2022; Lopes and Burley, 2011; Mahseredjian *et al.*, 2021). Another approach would be to use machine learning

methods to characterize the modeling error in terms of ADS-B data, SLM data, and other flight parameters so they can be applied to improve the accuracy of airframe noise component prediction (Alonso *et al.*, 2023).

ACKNOWLEDGMENTS

The authors would like to acknowledge other current and former members of our Metroplex Overflight Noise Analysis (MONA) team who have greatly contributed to the development of the infrastructure that has made this work possible, including Sanjaye Narayan, Vikas Munukutla, Chetanya Rastogi, and Priscilla Lui. We would also like to acknowledge the help and guidance of Sean Doyle, former FAA Senior Aviation Noise Specialist and Susumu Shirayama, FAA Project Manager for ASCENT Project 53, who have contributed greatly to elements of AEDT integration into the MONA system and validation of the AEDT noise predictions. Finally, we thank Bert Ganoung, Head of the SFO Aircraft Noise Office, who graciously gave us access to the accumulated SFO noise data with the help of Simon Heath, Director of Strategic Solutions for Airports of Envirosuite Ltd. This research was partially funded by the FAA Aviation Sustainability Center (ASCENT) as Project 053 – *Validation of Low-Exposure Noise Modeling by Open-Source Data Management and Visualization Systems Integrated with AEDT* (FAA Award No. 13-C-AJFE-SU-022).

AUTHOR DECLARATIONS

Conflicts of interest

The authors of this paper have no conflicts of interest to disclose regarding the research reported or the publication of this manuscript.

DATA AVAILABILITY

There are several datasets supporting the research results reported in this manuscript:

- (1) Raw sound-level data recorded at the SFO NMT-12 monitor near runways 28L and 28R. Restrictions apply to the availability of these data, which were used under license from SFO for this study. Contact the SFO's Aircraft Noise Office for access to their sound-level data.
- (2) Raw sound-level data recorded near the SIDBY waypoint by a MONA project sound monitor. These data are available from the corresponding author upon reasonable request.

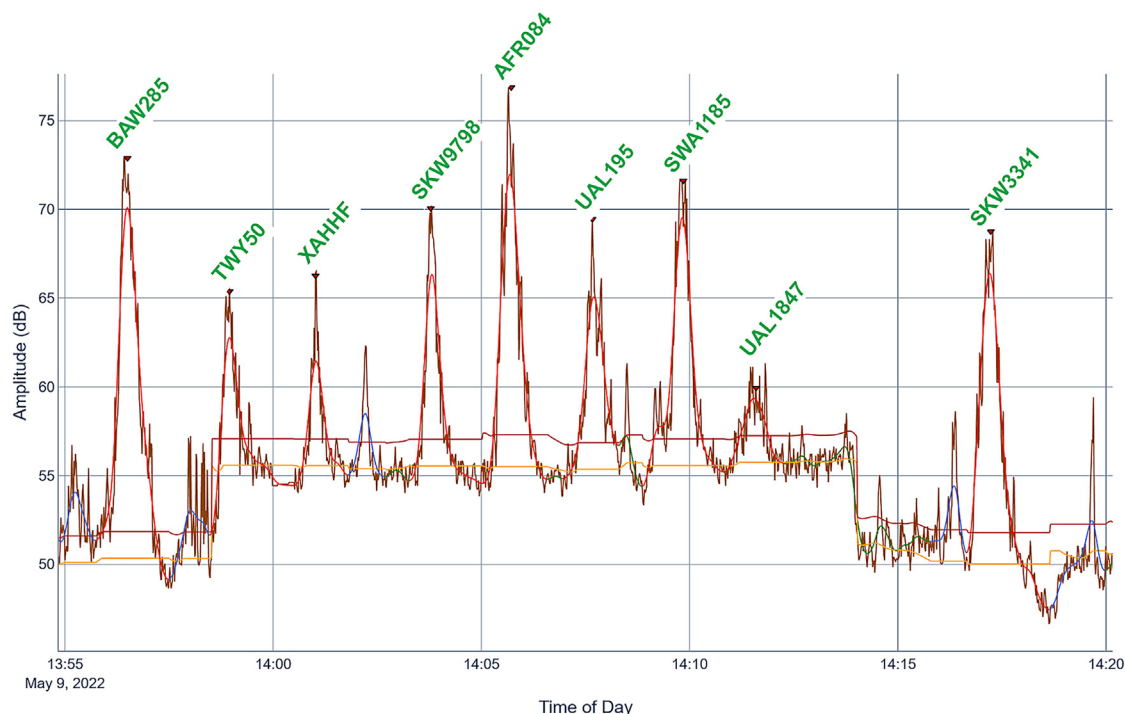


FIG. 20. (Color online) Typical sound profile segment measured by an SLM. Aircraft overflight events are outlined in red and are labeled with the specific flights that caused them. Peaks that were not linked to aircraft have a blue outline. The background is indicated by the gold trace. A peak-detection threshold, derived from the background trace is shown in dark red.

- (3) Annotated raw SFO ADS-B flight trajectory data recorded during the study period of this research (July 2021 through June 2022) by MONA project receivers with annotation information purchased from FlightAware, a subsidiary of Collins Aerospace. ADS-B flight trajectory data for various airports and time periods are available from sources such as OpenSky Network, FlightAware, FlightRadar, or ADSB Exchange. The specific MONA project ADS-B data (SFO, July 2021 through June 2022) are available from the corresponding author upon reasonable request, subject to permission from FlightAware for the integrated annotation data.
- (4) Sound peak data extracted from the raw sound-level data correlated with points of closest approach to relevant SLMs calculated from the annotated ADS-B data using the processing algorithms described in the Appendix. These data, for the study period, are available from the authors upon reasonable request and with the permission from SFO and FlightAware as appropriate.
- (5) The AEDT application is available for license from the FAA.

APPENDIX: SOUND-LEVEL MONITOR DATA PROCESSING

This appendix summarizes the process by which we isolate sound monitor peaks and identify the aircraft that cause them, as well as how we create metrics to curate the sound peaks for quality prior to inclusion in our AEDT metric

prediction analyses. Each isolated peak is compared to an ideal peak model to determine a GoF metric and to identify the possible contributions of simultaneous overflights with similar times and distances of closest approach.

Sound profile filtering and point of closest approach (PCA) determination. Figure 20 shows a 25-min portion of a typical sound profile that was recorded on May 9, 2022, by the SFO NMT-12 SLM located close to the final approach paths to runways 28L/R.

Our processing isolates aircraft peaks from the background and matches each peak with the aircraft that produced it (identified with green labels). We use non-linear time-series filters to estimate the background sound level as a function of time (forward and backward recursive filters) and to locate those peaks above a background-derived threshold in the sound profile that may have been caused by aircraft overflights [cubic spline shape filters (Prilepin and Husheer, 2022)]. These peaks are then precisely time matched with ADS-B closest approach data for all flights observed in the neighborhood of a given SLM to identify the aircraft giving rise to each sound peak.

Our approach to confirming potential aircraft peaks and identifying the corresponding aircraft flight is most closely related to the work of Giladi (2020), but differs in that ours is explicitly time-oriented rather than real-time position or “square-of-interest” oriented. Our approach also does not use extrapolation calculations based only on past aircraft behavior, but rather examines the full flight profile near the sound maximum. This allows us to account explicitly and more accurately for ADS-B errors and atmospheric and

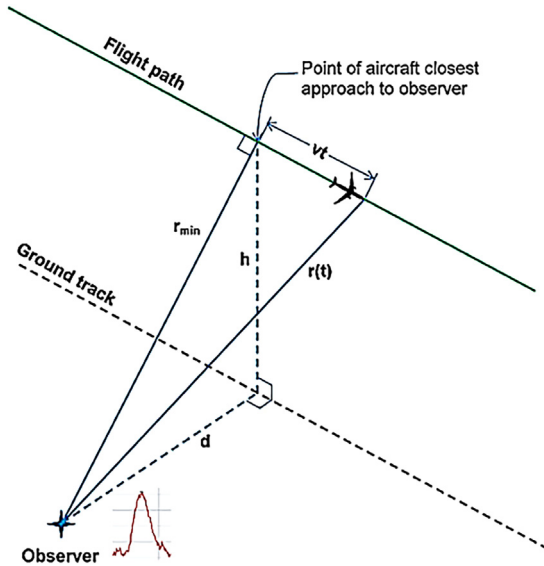


FIG. 21. (Color online) Idealized overflight path relative to a ground receptor or sound-level monitor (labeled *Observer*).

ambient noise distortion of the sound profile. The net profiles of isolated peaks above the background are analyzed to extract desired noise metrics for each identified overflight event (L_{Amax} and SEL).

Figure 21 shows the idealized geometry of an aircraft that passes close to an observer or a receptor (sound monitor) on the ground. The profile of the resulting sound peak is illustrated by the red trace next to the monitor/observer. In this figure, d is the minimum ground distance from the monitor to the ground track of the aircraft, h is the aircraft's altitude, and v is its ground speed. The time, t , marks the progress of the aircraft along its flight path and $r(t)$ is the three-dimensional distance of the aircraft from the monitor as a function of time. The minimum line-of-sight distance from the monitor to the aircraft (PCA) occurs at $t = t_0$ and is labeled r_{min} . For simplicity, we assume that near the point of closest approach the flight path is a straight line with constant altitude and ground speed.

From the ADS-B data for each the aircraft trajectory, we obtain a set of parametric equations for the flight path: latitude, $lat(t)$, longitude, $lon(t)$, altitude, $alt(t)$, and ground speed, $sp(t)$. We interpolate the error-corrected parametric equations, using the cubic-spline formalism to accurately locate the PCA to within ± 0.25 s. At the PCA, h is $alt(t_0)$ and v is $sp(t_0)$. We can calculate d from $lat(t_0)$, $lon(t_0)$, and the known coordinates of the sound monitor. We can calculate the rest of the geometry as shown in Eqs. (A1) and (A2),

$$r(t) = \sqrt{d^2 + h^2 + [v * (t - t_0)]^2}, \quad (\text{A1})$$

$$r_{min} = r(t_0) = \sqrt{d^2 + h^2}. \quad (\text{A2})$$

Once we know r_{min} , we can use the speed of sound, c , to compute the time, t_{slm} , when the aircraft sound emitted at the time of closest approach arrives at the sound monitor, as shown in Eq. (A3),³

$$t_{slm} = t_0 + \frac{r(t_0)}{c}. \quad (\text{A3})$$

In this way, we examine the trajectory of every aircraft that passes overhead close to a given monitor and associate the monitor's potential aircraft peaks with any aircraft that might have caused or contributed to it within a (half-width) window of tolerance. A peak may be the combined result of more than one nearly simultaneous overflight events, such as occur when aircraft approach or depart large airports with multiple parallel runways, or around multiple airports with air traffic layered by altitude.

Data quality assessment, GoF. To facilitate screening sound peaks in terms of their "quality," we evaluated the shapes of confirmed aircraft peaks that we identified in monitor sound profiles. We developed an analytic model for the ideal aircraft overflight peak shape for the trajectory configuration illustrated in Fig. 21. We use this idealized sound model to calculate the GoF metric in order to determine how closely a given sound peak matches the model.

Based on the assumptions of our idealized model, the aircraft-generated sound energy per second (power) recorded by an SLM over time, $E(t_{ac})$, has the form of a Cauchy or Lorentzian distribution, as shown in Eq. (A4). There, E_{max} is the maximum peak amplitude in linear energy space, α is the effective atmospheric attenuation coefficient, $r(t_{ac} - t_0)$ is the line-of-sight distance between the aircraft and the monitor at aircraft time t_{ac} , v is the aircraft ground speed (assumed constant), t_0 is the time at the aircraft when the sound emitted will result in the SLM peak maximum, and r_{min} is the line-of-sight distance of closest approach,

$$E(t_{ac}) = \frac{E_{max} e^{-\alpha * r(t_{ac} - t_0)}}{1 + \left(\frac{v * (t_{ac} - t_0)}{r_{min}} \right)^2}. \quad (\text{A4})$$

The energy distribution in the peak as recorded by the SLM is delayed by the transit time from aircraft to SLM and altered by the Doppler effect. As the aircraft approaches the observer, the sounds have a higher frequency and the leading edge of the peak is steeper. As the aircraft recedes, the sounds have lower frequency and the trailing edge is less steep. The Doppler effect is governed by Eq. (A3).⁴

Equation (A5) shows the relationship between a time interval at the aircraft, dt_{ac} , during which the aircraft emits sound energy that is received at the monitor during its sampling interval, dt_{slm} (nominally one second in our case) as a function of time along its trajectory,

$$dt_{ac} = \frac{dt_{slm}}{1 + \frac{v^2 * (t_{ac} - t_0)^2}{c^2 \sqrt{r_{min}^2 + [v * (t_{ac} - t_0)]^2}}}. \quad (\text{A5})$$

There are only two free parameters in this simple analytical model—the sound peak maximum power, E_{max} , and the time of peak maximum, t_0 . The other model parameters,

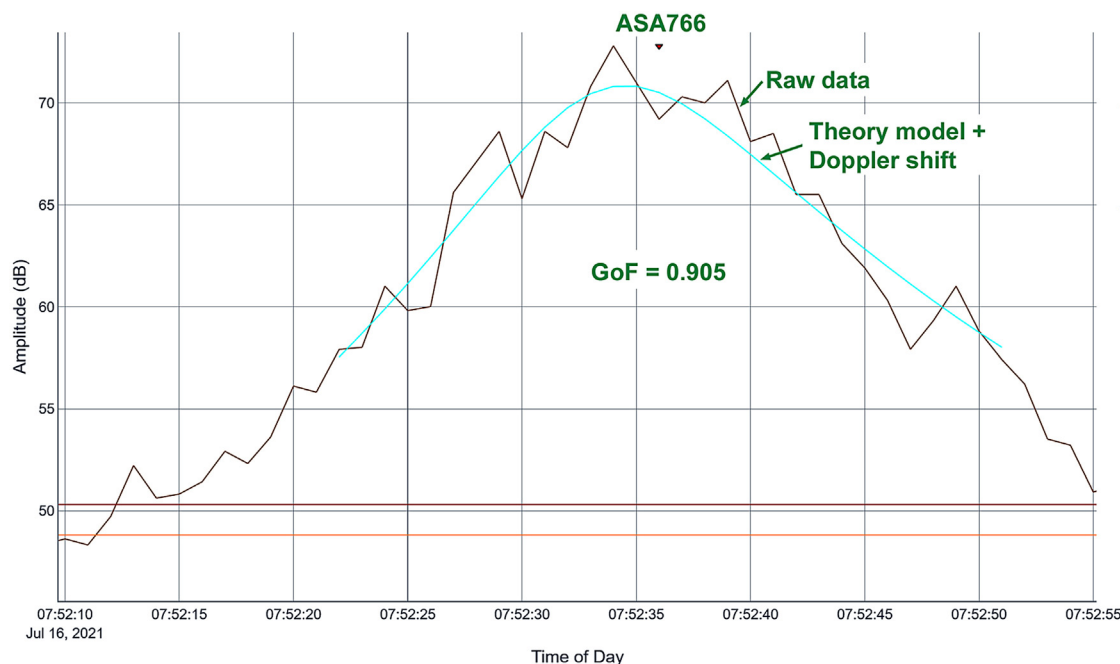


FIG. 22. (Color online) Example of the theoretical aircraft-peak model fit to a clean, raw recorded sound peak.

α , v , and r_{min} , are determined by ADS-B measurements and weather data. Even so, the model fits the real-world data quite well. Figure 22 shows a typical example of the model fit to actual flight data, in this case flight ASA766, a B738 from Seattle overflying monitor SFO NMT-12 on 7/16/2021 at 7:52:36. This peak is quite well-formed in that, while there are evident effects on the peak envelope from atmospheric turbulence and under-sampling, the overall peak shape is not badly distorted. On the other hand, at 7:49:33 as detected by the same SLM, a peak for flight SKW5531, an

E75L, was more strongly affected by a background disturbance as shown in Fig. 23.

This ability of the theoretical model to fit real peak profiles has led us to define a GoF, similar to the definition of the R^2 metric for linear-regression fits, to distinguish well-formed peaks from peaks that are distorted. The metric provides a measure of how well observed peaks are approximated by the model, based on the proportion of total variation of measurements explained by the model.

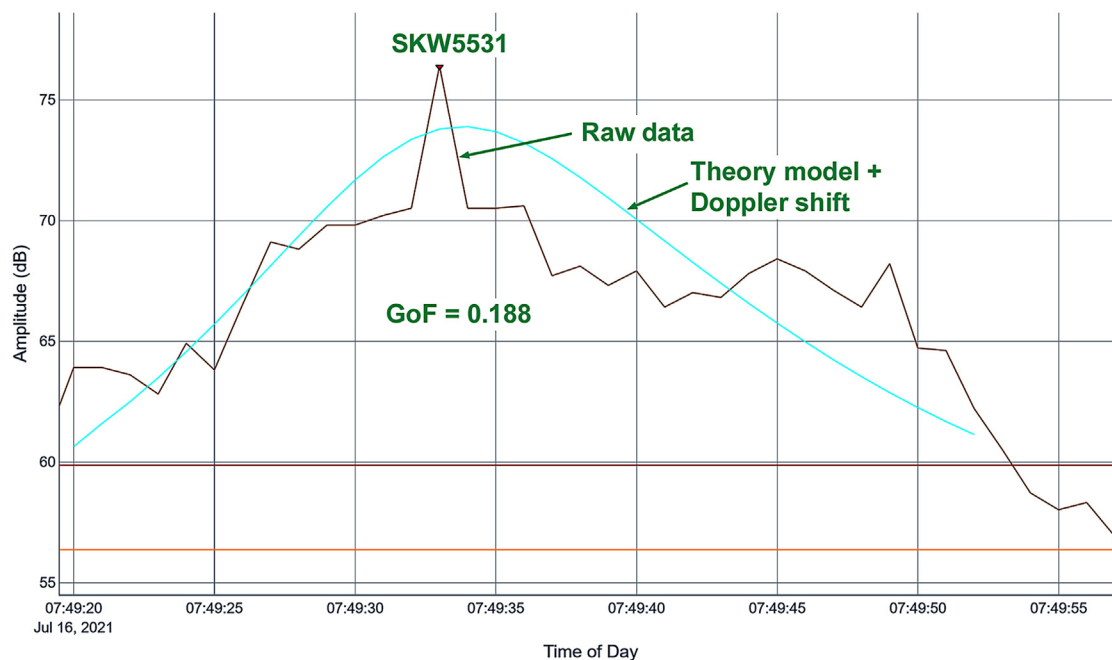


FIG. 23. (Color online) Example of the theoretical aircraft-peak model fit to a distorted, raw recorded sound peak.

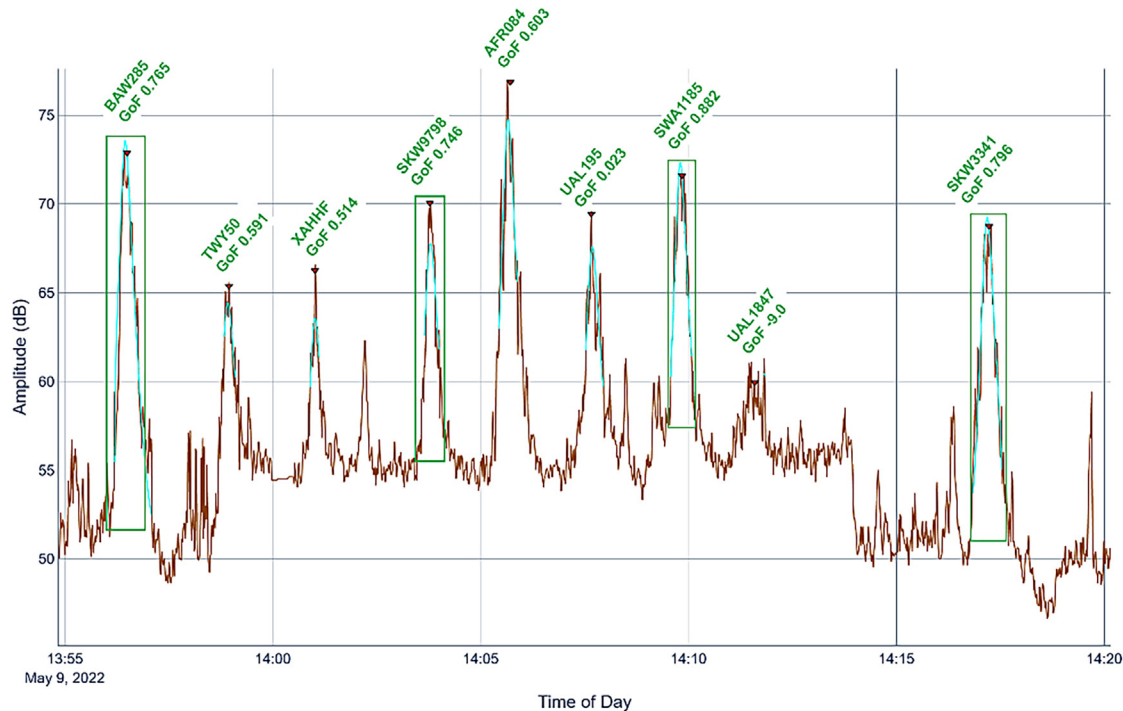


FIG. 24. (Color online) Confirmed aircraft peaks from Fig. 20 overlaid in light green with theoretical model fit to their raw sound data in brown. GoF values are shown above each confirmed peak and the four peaks in dark green boxes have acceptable GoF values (greater than 0.7).

Equation (A6) defines the GoF of a model $f_i = f(x_i)$ to a data set $\{x_i, y_i\}$, where \bar{y} is the mean of y_i . Our GoF metric generally ranges between 0 and 1 but may go negative for particularly poor fits (large numerator),

$$GoF = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}. \quad (A6)$$

Using this definition, we obtain a GoF of 0.905 for the well-formed peak in Fig. 22 and 0.188 for the distorted peak in Fig. 23. We computed GoF values for each confirmed aircraft peak over many days and, by experiment, find that a GoF value of at least 0.7 constitutes an acceptable quality peak. By “acceptable,” we mean that the peak is clean enough to use its maximum aircraft sound level (L_{Amax}) and integrated aircraft sound exposure level (SEL) as a basis for validating noise-prediction software.

Returning to the sound peak profile example we used earlier, Fig. 24 shows the theoretical model fit (in light green) overlaid on the confirmed aircraft peaks from Fig. 20. The four peaks outlined with dark green boxes have GoF values above 0.7. The GoF values for the other five confirmed aircraft peaks fell below this standard primarily because of envelope distortion caused by atmospheric turbulence and background ambient noise.

¹The histogram profiles are produced by counting the frequency distributions of the raw data exactly in the 0.2 dB bins. The profile is then smoothed with a cubic spline routine (Prilepin and Husheer, 2022) to make overplotting and interpretation easier. No important histogram shape information is lost.

²If the AEDT and SLM distributions were Gaussian, the standard deviation of the difference distribution would be given by $\sigma_{diff} = \sqrt{\sigma_{AEDT}^2 + \sigma_{SLM}^2}$.

³In order to compare times as measured by the ADS-B receivers with times measured by the SLMs, we synchronize the two clocks by measuring the average offset for the times of large, well-formed peaks and then adjusting the SLM clock values.

⁴Note that the general form of Eq. (A3) is $t_{slm} = t_{ac} + [r(t_{ac} - t_0)/c]$.

- Alexander, C., Dowell, D., Hu, M., Olson, J., Smirnova, T., Ladwig, T., Weygandt, S., Kenyon, J., James, E., and Lin, H. (2020). “Rapid refresh (RAP) and high resolution rapid refresh (HRRR) model development,” in *100th American Meteorological Society Annual Meeting* (AMS, Boston, MA).
- Alonso, J., Alonso, Y., Shukla, A., Jackson, D., and Rindfleisch, T. (2023). “Improving noise predictions of the aviation environmental design tool (AEDT) using deep neural networks and sound-level monitor data,” in *AIAA SCITECH 2023 Forum*, p. 735.
- Bendarkar, M. V., Bhanpato, J., Puranik, T. G., Kirby, M., and Mavris, D. N. (2022). “Comparative assessment of AEDT noise modeling assumptions using real-world data,” in *AIAA AVIATION 2022 Forum*, p. 3917.
- EASA (2023). “European Union Aviation Safety Agency: Aircraft noise and performance (ANP) data,” <https://www.easa.europa.eu/en/domains/environment/policy-support-and-research/aircraft-noise-and-performance-anp-data> (Last viewed March 2, 2024).
- ECAC Secretariat (2016). “Report on standard method of computing noise contours around civil airports,” Vol. 1, Applications Guide, 4th ed.
- FAA (2021a). *AEDT Version 3d Technical Manual* (Federal Aviation Administration, Washington, DC), Sec. 3.9, p. 145.
- FAA (2021b). *AEDT Version 3d Technical Manual, Section 3.2* (Federal Aviation Administration, Washington, DC), p. 57.
- FAA (2021c). *AEDT Version 3d Technical Manual, Section 3.6* (Federal Aviation Administration, Washington, DC), p. 72.
- FAA (2022). *Aviation Environmental Design Tool Release Notes for AEDT 3e* (Federal Aviation Administration, Washington, DC), p. 37.
- Gabrielian, A. B., Puranik, T. G., Bendarkar, M. V., Kirby, M., Mavris, D., and Monteiro, D. (2021). “Noise model validation using real world operations data,” in *AIAA Aviation 2021 Forum*, No. 2136.

- Geissbuhler, M., Behere, A., Rajaram, D., Kirby, M., and Mavris, D. (2022). "Improving airport-level noise modeling by accounting for aircraft configuration-related noise at arrival," in *AIAA SCITECH 2022 Forum*, p. 1650.
- Giladi, R. (2020). "Real-time identification of aircraft sound events," *Transp. Res. Part D: Transp. Environ.* **87**, 102527.
- Giladi, R., and Menachi, E. (2020). "Validating aircraft noise models," in *MDPI Proceedings*, Vol. 60.
- Hauptvogel, D., Bartels, S., Schreckenberger, D., and Rothmund, T. (2021). "Aircraft noise distribution as a fairness dilemma—A review of aircraft noise through the lens of social justice research," *Int. J. Environ. Res. Public Health* **18**(14), 7399.
- He, H. (2018). "Validation of aircraft noise prediction models," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (Institute of Noise Control Engineering, Reston, VA), Vol. 258, pp. 3359–3367.
- ICAO (1988). *Doc 9911, Recommended Method for Computing Noise Contours around Airports* (International Civil Aviation Organization, Montreal, Quebec, Canada).
- Jackson, D., Rindfleisch, T., and Alonso, J. (2021). "A system for measurement and analysis of aircraft noise impacts," *Eng. Proc.* **13**(1), 6.
- Lopes, L., and Burley, C. (2011). "Design of the next generation aircraft noise prediction program: ANOPP2," in *17th AIAA/CEAS Aeroacoustics Conference (32nd AIAA Aeroacoustics Conference)*, p. 2854.
- Mahseredjian, A., Thomas, J., and Hansman, J. (2021). "Advanced procedure noise model validation using airport noise monitor networks," in *50th International Congress and Exposition on Noise Control Engineering* (Institute of Noise Control Engineering Washington, DC).
- Mavris, D. N., and Sparrow, V. W. (2021). "Project 062 noise model validation for AEDT," ASCENT project report.
- Meister, J., Schalcher, S., Wunderli, J.-M., and Schäffer, B. (2023). "Validation of three inherently different aircraft noise calculation programs," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (Institute of Noise Control Engineering, Washington, DC), Vol. 265, pp. 574–581.
- NAS (2018). *Enhanced AEDT Modeling of Aircraft Arrival and Departure Profiles, Volume 2: Research Report* (The National Academies Press, Washington, DC).
- Nuic, A. (2010). "User manual for the base of aircraft data (BADA) revision 3.10," *Atmosphere* **2010**, 001, available at <http://maartenuijtdehaag.com/bada310-user-manual.pdf/>.
- Poles, D., Nuic, A., and Mouillet, V. (2010). "Advanced aircraft performance modeling for ATM: Analysis of BADA model capabilities," in *29th Digital Avionics Systems Conference* (IEEE, Piscataway, NJ), pp. 1–D.
- Prilepin, E., and Husheer, S. (2022). "CSAPS-cubic spline approximation (smoothing)—CSAPS," https://csaps.readthedocs.io/_/downloads/en/latest/pdf/ (Last viewed March 2, 2024).
- Rhodes, D., and Boeker, E. (2019). "Recommended method for computing noise contours around airports—Recent updates to ICAO doc 9911," in *2019 Environmental Report: Aviation and Environment (Destination Green, the Next Chapter)*, Chap. 2—Aircraft Noise, pp. 62–65.
- SAE Committee A-21 (2012). "Procedure for the computation of airplane noise in the vicinity of airports (stabilized Aug 2012) AIR1845A," technical report.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle		5. Report Date	
		6. Performing Organization Code	
7. Author(s)		8. Performing Organization Report No.	
9. Performing Organization Name and Address		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No.	
12. Sponsoring Agency Name and Address		13. Type of Report and Period Covered	
		14. Sponsoring Agency Code	
15. Supplementary Notes			
16. Abstract			
17. Key Words		18. Distribution Statement	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages	22. Price