Transportation Research Part D 161 (2026) 105620

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Challenges in comparing measured and modelled annual average noise levels around airports

Rebekka C. van der Grift *, Alireza Amiri Simkooei, Mirjam Snellen

Faculty of Aerospace Engineering, Section Operations and Environment, Delft University of Technology, Office 5.06, Kluyverweg 1, Delft, 2629HS, The Netherlands

ARTICLE INFO

Keywords:
Doc 29 aircraft noise model
Noise monitoring terminal
Noise measurements
Validation
L_{den}
Background noise
Noise event detection.

ABSTRACT

Aircraft noise has a significant impact on communities near airports. To regulate noise, annual average levels such as L_{den} are modelled across large areas. This study examines the consistency between L_{den} predicted by the European Civil Aviation Conference Doc 29 model and measurements from noise monitoring terminals around Amsterdam Airport Schiphol from 2017 to 2024. While strong correlations confirm Doc29′s suitability for system-level assessments, systematic biases persist, with nighttime levels L_{night} generally overestimated and L_{den} underestimated. Three potential sources were identified: difference between modelled and flown tracks, background noise influencing measured levels, and undetected flights, particularly for quieter aircraft and in low noise regions. Accounting for these factors improves correlation and reduces deviation between measurements and predictions. These results demonstrate promising methods for enabling noise model validation near airports. Further improvements in detection methods and operational modelling will increase the robustness of noise assessments and better support evidence-based policy development around major airports.

1. Introduction

A growing aviation industry is causing increased annoyance in the communities around large airports (Gjestland, 2019; Janssen et al., 2011). Noise policies are developed to address these concerns and to find a balance between aviation activities and noise pollution. Beyond annoyance, research has demonstrated that aircraft noise exposure can lead to adverse health effects, including sleep disturbance and cardiovascular impacts (Peters et al., 2018). In particular, nighttime aircraft noise has been shown to trigger acute cardiovascular responses and is associated with increased risks of cardiovascular mortality (Saucy et al., 2021). These findings underline the societal relevance of aircraft noise and the importance of accurately quantifying noise exposure.

To address these impacts, noise policies are developed to balance aviation activity with environmental and community constraints. These policies are often based on noise modelling, using best-practice models, such as the Aviation Environmental Design Tool (U. S. Department of Transportation Volpe National Transportation Systems Center, 2023) or the European Civil Aviation Conference (ECAC) Document 29 methodology (Conference, 2016b), short Doc 29. With these models, the noise impact of different aircraft and flight routes can be evaluated. This evaluation is performed through calculations of yearly average Day-Evening-Night noise levels L_{den}, expressed as noise contours. This metric represents the noise experienced on the ground due to all flight movements over a full year. The models are generally accurate within 1–2 dBA (Engineers, 2012) and are widely used for route optimisation, environmental

* Corresponding author.
E-mail address: rebekkavdgrift@gmail.com, r.c.vandergrift@tudelft.nl (R.C. van der Grift).

https://doi.org/10.1016/j.trd.2026.105620
Received 17 February 2026; Received in revised form 4 September 2026; Accepted 4 September 2026
Available online 10 September 2026

impact assessments, and even regulatory decisions such as flight limitations. Given their influence on policy, it is essential that these noise predictions are both accurate and robust. Therefore, validation through direct measurements is a topic of interest.

Dedicated measurement campaigns can be deployed to gain insight into the characteristics of aircraft noise. For example, aircraft noise measurements can be used to investigate the source noise of aircraft as a function of different flight parameters (Snellen et al., 2017; Vieira et al., 2022; Zellmann et al., 2018). At larger measurement distances, the effects of noise propagation and incidence angles also become increasingly important. These effects can also be investigated using such measurements (Filippone and Harwood, 2016; Wunderli et al., 2022). Measurement campaigns are also often used to investigate the effectiveness of specific noise reduction measures (Bertsch et al., 2024).

Studies that use noise measurements to validate best-practice noise models often rely on existing noise measurement systems around airports. A common method is to compare the single-flight noise levels of a specific aircraft type with measurements (Rindfleisch et al., 2024; Giladi and Menachi, 2024). In general, good agreement has been reported, with deviations within 3 dBA. However, such validation efforts are inherently selective. Measurements are usually taken from large aircraft operating close to airports, where noise levels are high, and data quality is more reliable. This introduces a measurement bias, as only certain flight phases and propagation distances are captured (Rhodes, 2018). Consequently, these studies may not fully represent the diverse noise exposure scenarios that support policy decisions. Validating annual average noise levels, such as L_{den}, is therefore essential, yet remains challenging.

Next to this measurement bias, an additional challenge arises from the difficulty of detecting and isolating aircraft noise events in real-world measurement environments. Noise monitoring systems are frequently affected by background noise, which complicates the identification and classification of aircraft-related sound events. Recent studies have therefore explored the integration of surveillance data, such as Automatic Dependent Surveillance-Broadcast (ADS-B), to support the identification and attribution of noise events (Giladi, 2020). These approaches have demonstrated improved detection performance compared to traditional radar-based methods, but they also highlight the complexity of linking acoustic measurements to individual flights.

Moreover, the increasing availability of flight trajectory data has revealed significant discrepancies between prescribed flight procedures and actual operations. Analyses based on ADS-B data show substantial variability in aircraft trajectories and flight profiles, even for standardised procedures such as noise abatement departures (Bhanpato et al., 2021). Such variability can lead to differences in predicted noise levels and spatial exposure patterns, particularly when simplified or representative trajectories are used in modelling frameworks. Incorporating real-world trajectories into noise modelling has been shown to improve the representation of operational variability and improved comparison with noise measurements (Pretto et al., 2024). Recent studies have also highlighted limitations associated with the aircraft performance assumptions embedded in ECAC-based noise modelling approaches. For example, Guo et al. (2025) showed that the use of fixed thrust coefficients can reduce the accuracy of aviation noise estimates and proposed a probabilistic framework to better represent operational variability. Similarly, van der Grift et al. (2025b) demonstrated that uncertainties in thrust estimation can result in differences of several decibels in predicted aircraft noise levels. These findings underline the importance of accurately characterising aircraft performance when comparing modelled and measured noise levels.

These challenges become particularly relevant when considering cumulative noise metrics such as L_{den}. Unlike single-event comparisons, the validation of annual average noise levels requires a comprehensive representation of all contributing flights, including those with low sound exposure levels that may not be detected by measurement systems. Previous work has shown that measurement systems can miss a substantial number of low-level events, especially at larger distances from the airport, leading to potential biases in estimated annual averages (van der Grift et al., 2025a). Consequently, validating cumulative noise metrics remains an open challenge, despite its importance for policy applications (Simons et al., 2022). Recent work has demonstrated the value of long-term noise monitoring datasets for analysing airport noise exposure and quantifying the contribution of different aircraft types to annual noise levels (Simons et al., 2026). Such studies highlight the growing importance of cumulative noise indicators and further emphasize the need for robust validation methodologies for policy-relevant noise metrics.

The objective of this research is to identify and quantify the challenges that arise when comparing measured and modelled aircraft noise levels when the focus is on annual average indicators such as L_{den}. These challenges originate from three main sources: i) the modelling framework, where of10 a mismatch exists between the assumed and actual flown track and used thrust, ii) the measurements, for example, due to background noise, and iii) dataset selection or incomplete detection of flights. Differences between measured and modelled noise levels cannot be attributed to only a single source. This study dives into the three sources by addressing them and estimating how these influence measured annual average noise levels. Finally, it provides a structured interpretation of discrepancies between models and measurements in the context of policy-relevant noise metrics. By addressing these objectives, this research contributes to a clearer understanding of the reliability of annual noise predictions. It also supports the development of validation strategies that can strengthen confidence in noise modelling for both regulatory and community applications.

For the above aims, this research considers the averaged aircraft noise level as predicted by the Doc 29 model with measurements around Amsterdam Airport Schiphol. Section 2 describes the model and measurement data used for this research. In Section 3, an initial comparison between modelled and measured noise levels is presented. Based on these results, several observations are highlighted and investigated in the subsequent sections. In Section 4, the uncertainty introduced by the model input assumptions is discussed by comparing standard procedures to proprietary onboard flight data. Section 5 analyses the effect of background noise on measuring in an uncontrolled environment by quantifying the differences between daytime and nighttime measurements. Lastly, Section 6 addresses the dataset bias by analysing the flights that are not detected by the measurement system and consequently investigates their effect on total annual noise levels. In Section 7, above effects are combined to give a good estimate of the comparison of modelled and measured annual noise levels. Finally, Section 8 summarises the main findings and their implications for future noise modelling and validation efforts.

Table 1
Number of flights performed in each operational year at Schiphol Airport, rounded off.

Year	Night $N_{flights}$	Total $N_{flights}$
2017	32,300	495,000
2018	31,200	497,400
2019	30,100	497,300
2020	16,100	272,600
2021	16,500	238,700
2022	23,900	395,100
2023	24,500	430,800
2024	24,900	473,800

2. Noise model and noise measurements

2.1. Noise metrics

Aircraft noise quantification can be performed using a variety of noise metrics. Most commonly, instantaneous metrics such as the maximum A-weighted noise level $L_{A,max}$, or time-integrated metrics such as the Sound Exposure Level (SEL), are used. These metrics describe the noise intensity of individual aircraft flyovers. To quantify noise exposure for a full year, a Day-Evening-Night average indicator can be used, calculated as

$$L_{\text{den}} = 10 \log \left(\sum_{i=1}^{N} 10^{(SEL_i + W_i)/10} \right) - 10 \log \left(T \right) \quad (1)$$

where SEL_i are the SEL values per aircraft event (measured or modelled) at specific locations, and W_i are the assigned penalties for day (0 dBA), evening (5 dBA) or night (10 dBA) flights. The T is the total reference time (in seconds) over which the period is averaged. This results in a factor of 75 for all seconds within a year. For the night-time metric L_{night}, an eight-hour period (23:00 to 07:00) is used, resulting in a normalisation factor of 70.2.

2.2. Noise model

The model used in this research is the European standard Doc 29 aircraft noise model (Conference, 2016b). The reference (Conference, 2016a) describes the calculation methodology in detail. The Doc 29 methodology utilises the Aircraft Noise and Performance (ANP) database (European Union Aviation Safety Agency, 2026) and works in a manner similar to the Aviation Environmental Design Tool (AEDT) (U. S. Department of Transportation Volpe National Transportation Systems Center, 2023). Both models are denoted as best-practice noise models. The ANP database contains so-called Noise-Power-Distance (NPD) tables for different aircraft types. These tables provide the expected noise level (e.g. SEL) on the ground, as a function of the engine thrust setting and distance from the aircraft.

Often, the most challenging part of best-practice noise modelling is the availability of accurate operational input parameters. For this research, the methodology developed by the Royal Netherlands Aerospace Centre (NLR) is used. References (Heblij and Derei, 2019; Heblij et al., 2019) describe how airport-specific operational procedures are developed for Amsterdam Airport Schiphol. This methodology is a trade-off between accuracy and computational demands.

The model output consists of L_{night} and L_{den} predictions on a spatial grid around the airport, roughly 80×80 km. These noise contours include all flights of commercial traffic and general aviation. Helicopter noise is outside the scope of this research. Commercial traffic refers to flights operated by airlines that are available for individual bookings of passengers, cargo, and/or mail transport. These aircraft movements can be subdivided into scheduled flights, which follow fixed routes according to published timetables, and non-scheduled flights, which include charter services with a non-regular operational character. General aviation traffic comprises all other traffic that does not fall under commercial traffic and is independent of aircraft size. This category includes, for example, small business aviation, technical flights following maintenance, and public service operations.

Aviation policy in the Netherlands is based on operational years, which run from November 1st to October 31st.[1] This research utilises the data from 2017 to 2024.[2] In Table 1, the number of flights operated at the airport of each operational year is given with its corresponding number of nighttime flights. A clear effect of the COVID-19 pandemic is visible in the years 2020 and 2021.

2.3. Noise measurements

Schiphol Airport is equipped with the Noise Monitoring System NOMOS, consisting of 41 Noise Monitoring Terminals (NMTs), placed at different distances from the runways as illustrated in Fig. 1. These NMTs are often installed in and around residential areas

[1] Detailed noise contour assessments are available in the usage forecast evaluation reports: https://www.schiphol.nl/nl/jij-en-schiphol/gebruiksprognose-alles-wat-je-maar-wilt-weten/.

[2] For example, the operational year 2017 spans from 1st November 2016 to 31st October 2017.

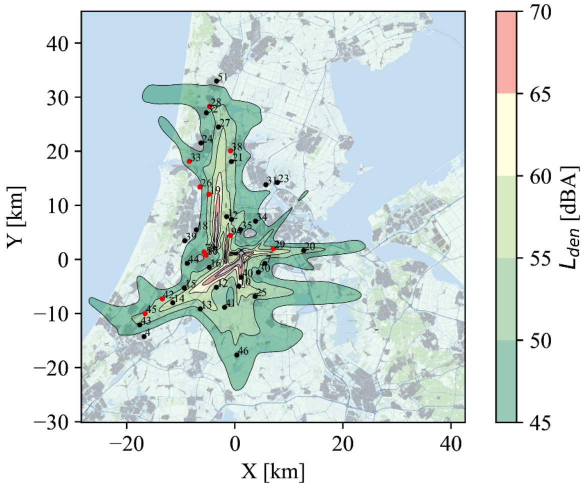


Fig. 1. Locations of the 41 NOMOS NMTs around Schiphol Airport with respect to the modelled L_{den} of 2023. Red NMTs are not used for validation purposes.

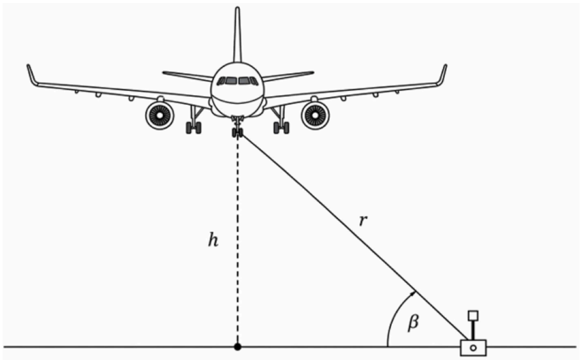


Fig. 2. Visualisation of aircraft-observer geometry.

as their primary goal is to inform the public about aircraft noise exposure. A Dutch national research campaign has investigated the suitability of these NMTs for aircraft noise model validation purposes (Sahai et al., 2021). Each NMT was given a quality score based on criteria such as location, background noise, and the potential for reflection and obstruction. These criteria match the ones outlined in ISO20906 (Standardization, 2009). Only the NMTs considered suitable for noise model validation are used for the research presented in this contribution. The NMTs deemed not suitable are depicted in red in Fig. 1. For example, NMT 33 (marked as red) is mounted on a pole approximately one meter above a roof surface and located near a large air-conditioning unit. This causes a high measured background noise and reflections.

The NMTs are placed on high poles 6 to 10 m tall to minimise ground reflections. A recent study measured differences up to 2 dBA depending on the elevation angle of the aircraft when compared to measurements without ground effect (Jayanthi et al., 2026). The Doc 29 model predicts noise levels for a height of 1.2 m above the ground, but states that *"Comparison of measurements at 1.2 m and 10 m and theoretical calculation of ground effects show that variations of the A-weighted sound exposure level are relatively insensitive to receiver height."* (Conference, 2016a). Jayanthi et al. (2026) confirmed this result by showing minimal differences for measurements at 1.75 m and 4.5 m.

In addition to the noise metrics $L_{A,max}$ and SEL for each noise event, NOMOS gathers operational characteristics such as radar tracks, aircraft type, operation type, tail number, and departure/arrival time. An important parameter used in this study is the Point of Closest Approach (PCA). For each flight-NMT combination, the system calculates the slant distance r, elevation angle β, and altitude h at the moment the aircraft passes closest to the NMT, as visualised in Fig. 2. Each flight thus results in 41 PCAs, one for each NMT. For NMTs where increased noise levels are detected, the corresponding PCA is coupled to the specific measured noise event. At Schiphol, there is a capacity of up to 500,000 flights per year, which results in a large dataset. The availability of PCA data for all NMTs enables the identification of flights with no corresponding noise measurements.

2.3.1. Noise measuring thresholds

NMTs register aircraft noise by identifying increased noise levels, while accompanied software simultaneously uses radar data to check for aircraft nearby. To prevent non-aircraft events such as road traffic, dogs, or alarms from being falsely detected as aircraft

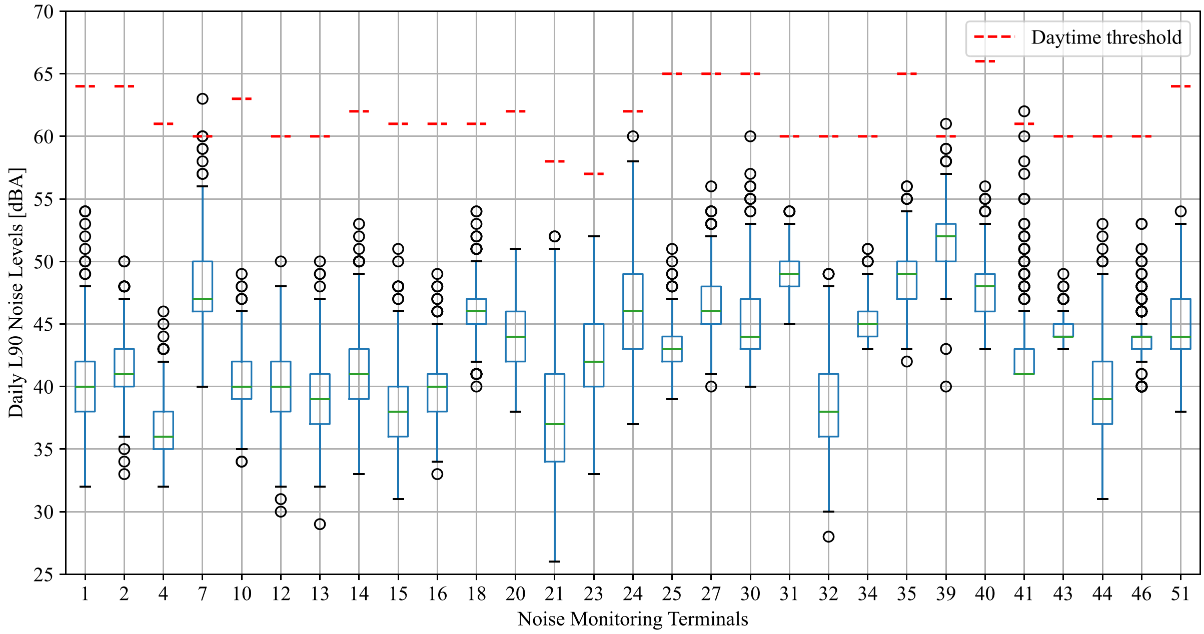


Fig. 3. Boxplot of daily L90 background noise levels per NMT during the year 2023 and their corresponding daytime thresholds.

noise, NMTs are equipped with thresholds. These thresholds are often defined relative to the local background noise level and defined as the noise value in dBA that the peak of the noise of the event $L_{A,max}$ must surpass. A 10 dBA difference with $L_{A,max}$ is the standard to adequately remove background noise levels (Standardization, 2009). Around Schiphol, background noise levels are often around 45 dBA. This results in thresholds ranging from approximately 58 to 65 dBA, depending on location and time of the day. For some NMTs, nighttime thresholds are set 2–5 dBA lower due to less background noise. Average background noise levels can be estimated using an L90 metric, the noise level exceeded during 90% of the measurement period. A boxplot of daily L90 values and their daytime thresholds for each NMT is presented in Fig. 3. The circles outside the boxplot represent outlier days. The figure illustrates the variability in background noise conditions at each NMT throughout the year 2023. This highlights the importance of selecting a high-enough threshold to exclude background noise from real aircraft noise events.

Although thresholds decrease the chance of contamination of aircraft noise events by other factors, they also prevent the detection of aircraft noise events that lie below the threshold. This mainly affects quieter aircraft like the A320-NEO, or flights at larger distances from NMTs. This effect is one of the challenges addressed in Section 6. For some NMTs, like NMTs 10 and 25, the applied threshold is more than 10 dBA higher than the highest recorded L90 values. This indicates that the threshold at these locations is likely set too high and should be readjusted in future monitoring efforts.

2.3.2. Pre-processing and outlier removal

Several pre-processing steps are required to ensure data quality before analysis. Because measurements have been taken in uncontrolled outdoor environments, disturbances or incorrect coupling to aircraft can occur. Aircraft at low elevation angles relative to an NMT are particularly problematic, as signals may be obstructed or reflected. In addition, unrelated noise sources such as road traffic may be incorrectly linked to an aircraft that is still on the runway at an elevation angle of 0°.

ISO 20,906 recommends using only measurements at elevation angles above 60° to limit the effect of ground reflection and lateral attenuation. Applying this criterion, however, would discard a large portion of flights and hence would not provide a realistic image of the noise exposure at certain locations. Fig. 4 shows the measured SEL for the B737-800 versus elevation angle and slant distance for arrivals and departures separately. It highlights increased variability at low elevation angles and indicates background noise influences beyond 10,000 ft. Based on data analysis, a more practical compromise was set in this study. Only events with elevation angles below 10° were excluded. This filter effectively removes the most unreliable measurements, where disturbances are most prevalent, while preserving a sufficiently large dataset for statistical analysis. Although the likelihood of disturbed measurements increases at the lower elevation angles (10 to 60°), these data are retained to ensure representativeness of real operational conditions.

It should be noted that lateral attenuation effects become more pronounced at low elevation angles, as described in Doc 29, where attenuation increases significantly as the elevation angle decreases. While this may influence the spread observed in the measurements by a given distance, the same effects are also inherently included in the Doc 29 modelling results used in this study. Therefore, both measured and modelled datasets are affected in a consistent manner, allowing for a fair comparison. Nevertheless, increased uncertainty in both measurement and modelling is expected in this angular range, which should be considered when interpreting the results.

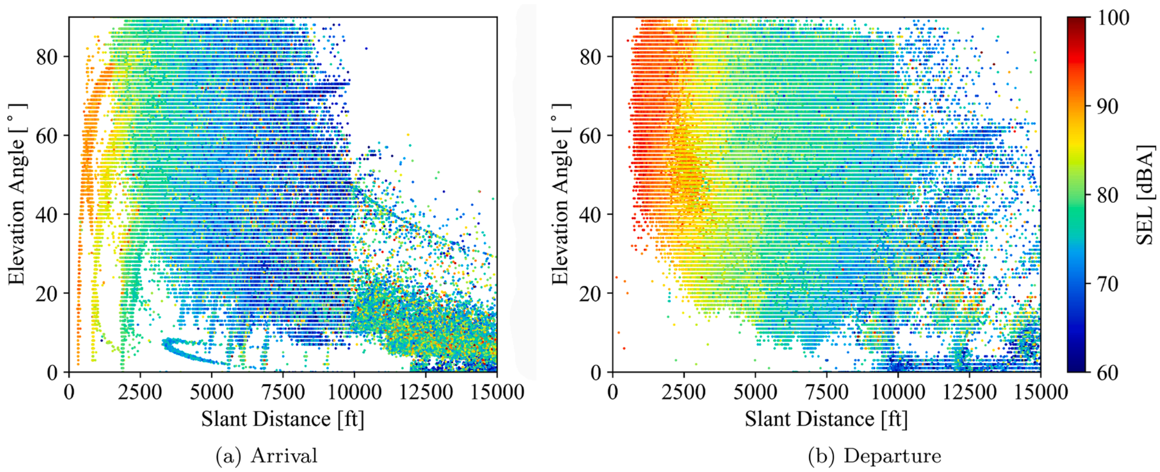


Fig. 4. Complete eight operational years of measured B737-800 SEL levels across NMTs versus elevation angle and slant distance before outlier removal.

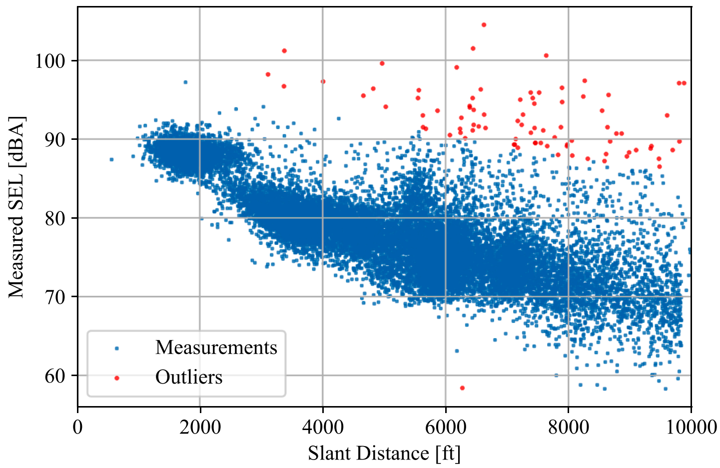


Fig. 5. Complete eight years of arrival measurements of the A21N with respect to slant distance, where outliers are identified by the w-test.

Fig. 4 also shows a clear cut-off in the measurements at 10,000 ft. This is due to a software constraint that limits the correlation of noise events with aircraft beyond this slant distance for all but a small number of NMTs.

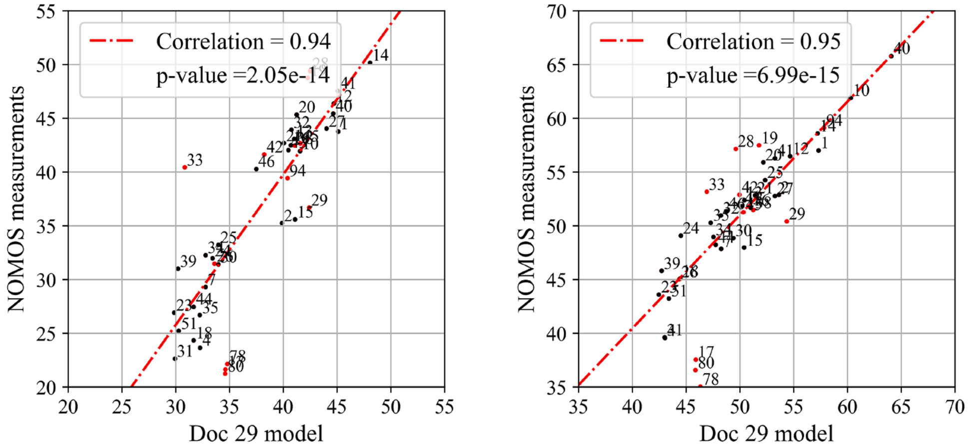
ISO 20,906 also recommends using only measurements recorded under wind speeds below 10 m/s and in the absence of precipitation (Standardization, 2009). For the Netherlands, this would result in a 24% reduction in the dataset. In this study, however, measurements under such less favourable weather conditions are retained, even though this would likely lead to somewhat elevated noise levels during these periods.

Further analysis revealed that some NMTs exhibited unexpected large noise levels. A small number of measurements exhibited physically unrealistic noise levels of approximately 120 dBA, most likely due software-related errors. These observations were removed during an initial quality-control step and are not included in the subsequent statistical analysis. To identify less obvious outliers, a w-test (Teunissen, 2000; Baarda, 1968) was applied to the residuals of a least-squares (LS) regression model. The data were first grouped by aircraft type and operation (arrival or departure), after which an LS model relating the measured SEL to slant distance was fitted for each group. For every observation, the normalized residual was computed and tested using the w-test. The null hypothesis assumes that the observation is free from gross errors, whereas the alternative hypothesis assumes the presence of an outlier. Observations were flagged as outliers when the test statistic exceeded the critical value corresponding to the selected significance level. A significance level of $\alpha = 0.005$ was adopted. This value provides a balance between sensitivity to actual outliers and protection against falsely rejecting valid measurements. Fig. 5 illustrates the procedure for arrival measurements of the A21N aircraft, where observations identified as outliers by the w-test are shown in red.

Lastly, the NOMOS system occasionally experiences interruptions in data availability. In twelve instances, gaps exceeding one week within a given year were identified. For a fair comparison of yearly average results, these specific NMTs and years are excluded from further analysis.

Table 2
Baseline results of comparing modelled L_{night} and L_{den} to measurements around Schiphol Airport. (The number of flights is rounded off).

	L_{night}				L_{den}			
Year	$N_{flights}$	μ [dBA]	σ [dBA]	ρ	$N_{flights}$	μ [dBA]	σ [dBA]	ρ
2017	32,300	1.31	2.92	0.96	495,000	-0.68	1.53	0.96
2018	31,200	0.23	2.20	0.94	497,400	-0.99	1.76	0.94
2019	30,100	0.78	2.37	0.97	497,300	-0.63	1.86	0.95
2020	16,100	0.82	4.88	0.88	272,600	-1.39	2.42	0.92
2021	16,500	1.43	4.08	0.91	238,700	0.10	2.11	0.94
2022	23,900	1.90	3.67	0.93	395,100	-0.14	2.28	0.92
2023	24,500	1.68	3.54	0.94	430,800	-0.51	1.91	0.95
2024	24,900	0.23	3.25	0.92	473,800	-1.07	2.24	0.94
Average		1.05	3.36	0.93		-0.66	2.02	0.94



(a) L_{night} (b) L_{den}

Fig. 6. Measured and modelled L_{night} and L_{den} of all NMTs for the year 2023, including the correlation coefficient and p-value.

3. Baseline modelled-measured results

For each year, the L_{night} and L_{den} noise metrics were calculated from the available measurements and compared with their respective modelled values for all NMTs. The correlation coefficient (ρ) was calculated between the modelled noise levels L_{model} and the measured noise levels L_{meas}. Furthermore, the difference between model and measurement is defined as:

$$\Delta L = L_{\text{model}} - L_{\text{meas}} \tag{2}$$

where a negative ΔL indicates that the model underestimates noise relative to the measurements. For all NMTs, the median difference (μ) and standard deviation (σ) of ΔL were computed. Results are summarised in Table 2.

A strong correlation between modelled and measured noise levels is observed across all years, for both L_{night} and L_{den}. The correlation coefficients are slightly lower for the years 2020 and 2021. These two years coincide with a substantially reduced number of flights due to the COVID-19 pandemic. A clear difference emerges between ΔL_{night} and ΔL_{den} when analysing the μ. While the model generally overestimates nighttime noise levels, it tends to underestimate noise levels over the full day.

Another notable difference includes a larger standard deviation (σ) for ΔL_{night} compared to ΔL_{den}. Fig. 6 illustrates these variations for 2023, showing measured and modelled noise levels per NMT. The correlation coefficient and its corresponding p-value are also given. A p-value smaller than 0.05 indicates statistical significance of the estimated correlation. As mentioned in Section 2, NMTs deemed unsuitable for validation are indicated in red. These NMTs were excluded from the calculation of μ, σ, and ρ, but are shown for completeness.

When analysing the performance of individual NMTs, big differences were observed. Some NMTs, such as NMT 15, showed consistently lower measured noise than predicted by the model. For L_{night} predictions, and to a lesser extent for L_{den}, the differences between measured and modelled noise levels increase in the so-called low-noise regions, defined, for example, as $L_{\text{den}} < 50$. Measurements further from the airport have lower signal-to-noise ratios, making it more difficult to accurately isolate aircraft noise from background noise. As a result, noise measurements may be overestimated due to the presence of background noise, or aircraft noise events may not be detected at all. In the latter case, the noise contribution of some of the flights is entirely omitted from the measurements, leading to the underestimation of L_{den} or L_{night}.

In the next three sections, the main identified causes that challenge this comparison of measured and modelled L_{den} are addressed. Starting with the potential mismatch between the actual track and thrust and the assumed model input in Section 4. After, in Section 5, the influence of ambient noise on the noise measurements is discussed. Finally, in Section 6, the effect of missing flights, due the fact that they have not been detected by the NOMOS system, on the comparison is quantified.

4. Model input discrepancies

The first challenge in comparing measured and modelled aircraft noise levels lies in the discrepancies between the assumed operational procedures in the noise model and the actual flight operations. This section provides a preliminary analysis of these operational differences and their potential impact on noise modelling.

When comparing measured and modelled aircraft noise, it is essential to assess how the modelled flight inputs differ from the actual flown tracks. In best-practice noise modelling, so-called *procedures* are used. These procedures consist of consecutive segments with changing altitude, speed, flap settings, and thrust. The procedures are defined as the distance from the start-of-roll (SOR) for departures and the distance to the landing threshold for arrivals. The ANP database provides default departure and arrival procedures for most aircraft types (European Union Aviation Safety Agency, 2026). As each airport often has its own operational preferences, Schiphol has developed a procedure database specific to the airport in the form of Fixed Point Profiles (FPP) (Heblij and Derei, 2019). To evaluate the discrepancy between these procedures and actual operations, an analysis is performed. The results presented in this section are based on van Pijlen (2026), who examined four aircraft types (A330-300, B737-800, B777-300 and A321neo) using proprietary airline data, consisting of 100 departure and 100 arrival flights per type. The dataset includes flight parameters such as calculated net corrected thrust F_n/δ.

4.1. Departures

Departure procedures typically include acceleration on the runway, an initial climb, a second acceleration, and a subsequent climb. Above an altitude of 10,000 ft, altitude, speed, and thrust profiles are extrapolated for simplicity. FPPs vary based on the aircraft take-off weight, which is estimated from trip length and fuel load. In addition to weight and aircraft type, the specific procedure can differ. The most common departure procedure is the Noise Abatement Departure Procedure 2 (NADP2), which includes an acceleration step at 800 ft. Variants with acceleration at 1000 or 1500 ft are also used. For modelling purposes, the departure procedure for a flight is assigned based on airline questionnaires. This introduces uncertainty because weather conditions or operational restrictions can cause deviations from the selected FPP.

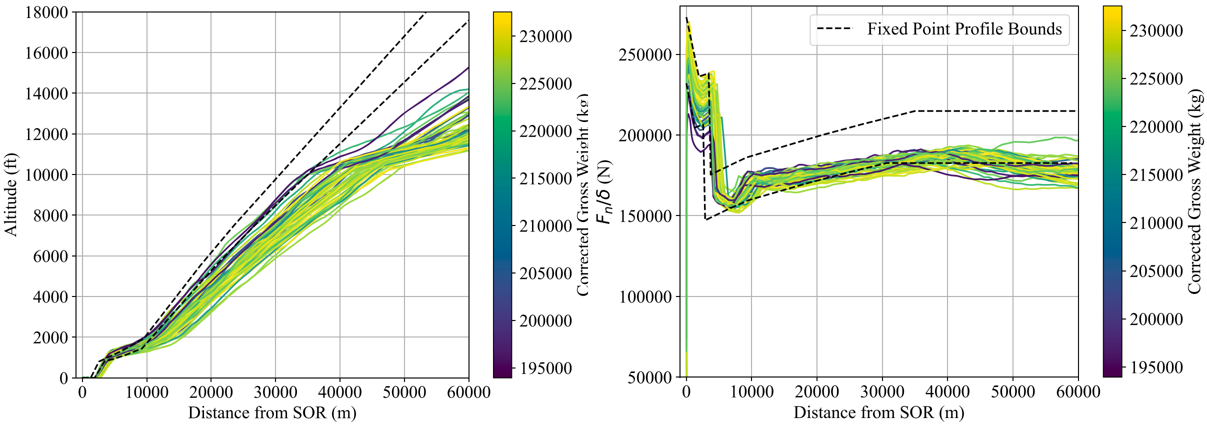
Fig. 7 illustrates example A330-300 departure profiles using airline flight data. The graphs show the lower and upper bounds of the FPP as a function of the aircraft weight. The upper bound of the altitude profile in Fig. 7(a) represents the lowest aircraft take-off weights, while in the FPP thrust profile (Fig. 7(b)), it represents the highest aircraft take-off weights. It is evident that the flown altitudes are significantly lower than those prescribed by the FPP. Additionally, the corresponding thrust settings are also reduced compared to the model assumptions. To assess the effect of these discrepancies on the noise, the summed SEL (not normalised to L_{den}) was calculated for a spatial grid around Schiphol, using both the airline data and the FPPs, respectively. Fig. 8 shows that noise levels are estimated to be lower when actual aircraft performance is used instead of the standard NADP2 procedure. This effect can be significant, depending on NMT location, up to -3 dBA. For this aircraft type, reduced thrust has a larger impact on noise levels than the lower flown altitudes. However, few definitive conclusions can be drawn because performance differences vary among aircraft types. In addition, proprietary flight data listing the thrust is not available for all aircraft types, which restricts a more comprehensive analysis.

4.2. Arrivals

Arrival procedures at Schiphol are classified as Continuous Descent Approach (CDA) or stepped approaches. In stepped approaches, aircraft maintain level flight segments at altitudes such as 2000 or 3000 ft. Such procedures are often used at congested airports to facilitate traffic sequencing. Schiphol's arrival FPP includes both reduced- and full-flap variants for the CDA and stepped approaches. The stepped approach has variants starting the level segment at 5 to 40 km before the runway. For modelling noise at Schiphol, arrival FPPs are matched to flown vertical flight profiles (from radar data) using a weighted least-squares error metric, reducing mismatches between modelled and flown trajectories. However, deviations in the start location and extent of the stepped approach remain. For example, in Fig. 9(a), stepped approaches starting at 50 km or even 85 km before the runway are visible. Thrust prediction during arrivals is challenging because small deviations in altitude or speed require short thrust increases, as shown in Fig. 9(b).

A similar analysis as for departures was performed for arrivals, producing the SEL grid illustrated in Fig. 10. In contrast to departures, using actual flight data results in higher modelled noise levels. This is mainly due to lower flown altitudes and intermittent thrust increases during the approach operations. Because thrust patterns during arrivals are highly irregular, predicting their effect on total L_{night} and L_{den} remains difficult.

Finally, atmospheric conditions introduce additional variation in the noise model input. Doc 29 provides a methodology to account for impedance and atmospheric absorption, but temperature variations also affect aircraft performance. For example, colder conditions or nighttime operations generally require less thrust for similar altitude and speed changes, while hot conditions typically lead to increased thrust requirements. Although average meteorological conditions in the Netherlands are close to ISA, these variations can still significantly influence aircraft performance. For example, in van der Grift et al. (2025b) it is found that the thrust of an aircraft



(a) Altitude profiles.

(b) Thrust profiles.

Fig. 7. Example departure profiles and take-off weights of the A330 compared to the FPPs. Modified from van Pijlen (2026).

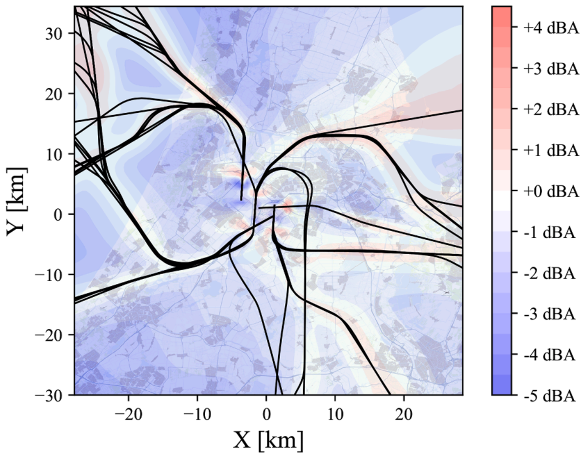
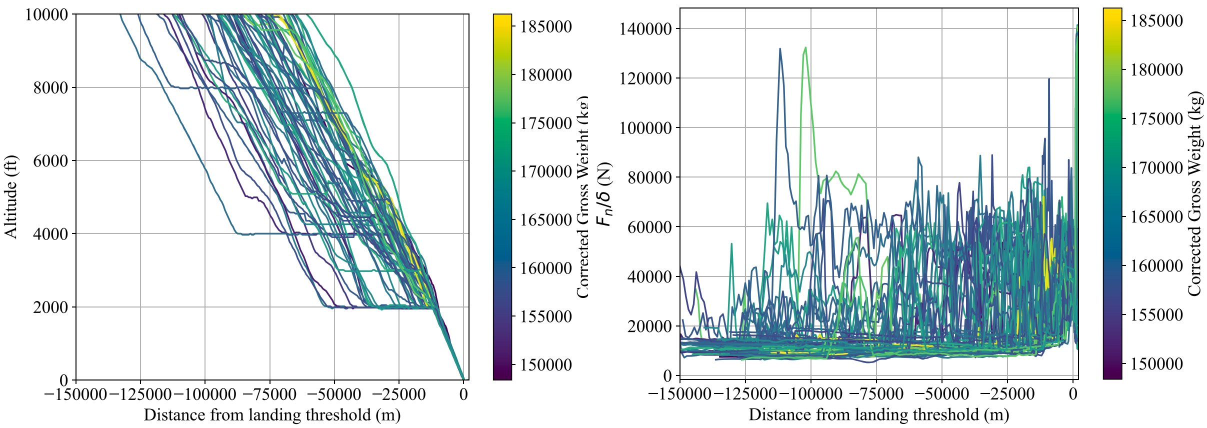


Fig. 8. Effect on summed SEL of actual A330 departure performance compared to standard FPP. Modified from van Pijlen (2026).



(a) Altitude profiles.

(b) Thrust profiles.

Fig. 9. Example arrival profiles and landing weights of the A330 relative to runway threshold. Modified from van Pijlen (2026).

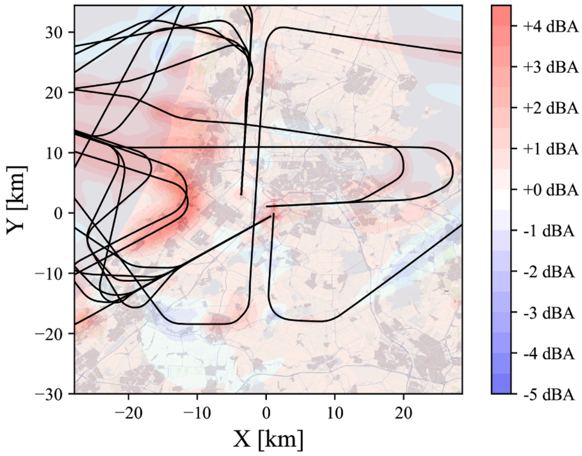


Fig. 10. Effect on summed SEL of actual A330 arrival performance compared to standard FPP. Modified from van Pijlen (2026).

can be reduced by roughly 5% for a 10 C° colder day. This would consequently correspond to a 1 dBA noise decrease for a B738. Quantifying this effect for the annual averages is outside the scope of this research.

5. Offset in measured noise levels due to ambient noise

The second challenge in comparing modelled and measured aircraft noise levels arises from the presence of ambient noise, affecting the measurements of the continuous noise measurement system. In contrast to measurement campaigns with more controlled conditions, NMTs are placed in an uncontrolled environment with background noise. This background noise can cause higher measured noise levels in SEL and thus higher measured L_{den} noise values. Background noise levels differ per NMT and time of day. Analysis showed lower L90 background noise levels during night hours (23:00-7:00) than during day hours (7:00-19:00). This systematic difference between day and night can help analyse the effect of background noise on the measured SEL of the aircraft noise events. In Section 5.1, the differences between measured noise during day and night are quantified. Section 5.2 tries to extract the effect of background noise on the measured SEL and calculate the effect on the total measured L_{den}.

5.1. Quantification of differences in measured day and night noise levels

To quantify the day-night differences, the average measured SEL per NMT were calculated separately for day and night. The resulting differences between SEL_{Day} and SEL_{night} are illustrated in Fig. 11 for the year 2024. Most NMTs show lower measured nighttime noise levels. The average difference is approximately 0.9 dBA. However, this result alone does not confirm a causal relationship between time of day and measured noise levels. Operational differences between daytime and nighttime traffic may also contribute. For example, 72% of nighttime arrivals use CDA, compared to only 38% during the day. CDA procedures maintain higher aircraft altitudes for a longer period. This increases the average PCA distance at night, lowering SEL_{night} compared to SEL_{Day}. Further analysis of average PCA during arrivals revealed substantial differences at NMTs located farther from the arrival paths, such as NMT 39. This can explain the large difference in measured SEL in Fig. 11 for this NMT, as an average altitude difference of approximately 1,500 ft was observed between day and night operations. This difference can correspond to noise level differences of up to 3 dBA. Differences in PCA were negligible for NMTs directly under fixed arrival routes (e.g., NMTs 1, 10, and 14). Aircraft over these latter NMTs typically follow the Instrument Landing System (ILS), maintaining similar altitude profiles during day and night.

Additional operational differences between day and night are observed in fleet composition. Smaller aircraft types, such as the E190, fly less frequently during the night. Conversely, a higher proportion of large (cargo) aircraft, such as the B747-400, fly during the night. This change in fleet mix would be expected to increase the measured nighttime noise levels, which does not seem to be the case. To find the effect of the different ambient noise levels between day and night on the measured SEL, the uncertainties within operational differences need to be removed.

To isolate the effect of background noise, SEL values were analysed across fixed PCA distance bins. Measurements were grouped into bins of 1000 ft. For each aircraft type and operation, median SEL values were calculated separately for day and night. The Mann-Whitney U test was applied to assess statistical significance between these non-normally distributed, unpaired datasets (Mann and Whitney, 1947). Only bins with more than 20 measurements and statistically significant differences were considered. To avoid location bias, comparisons were restricted to measurements from the same NMTs with sufficient nighttime data. The results, averaged across NMTs per distance bin, are shown in Fig. 12 for three different aircraft type groups. For most aircraft, like the B737NG series depicted in Fig. 12(a), the measured differences increase with the PCA distance. This effect is more pronounced for arrivals than departures, which is expected given the lower source noise levels of arriving aircraft. The day-night SEL differences ($\Delta\mathrm{SEL}_{DN}$) were

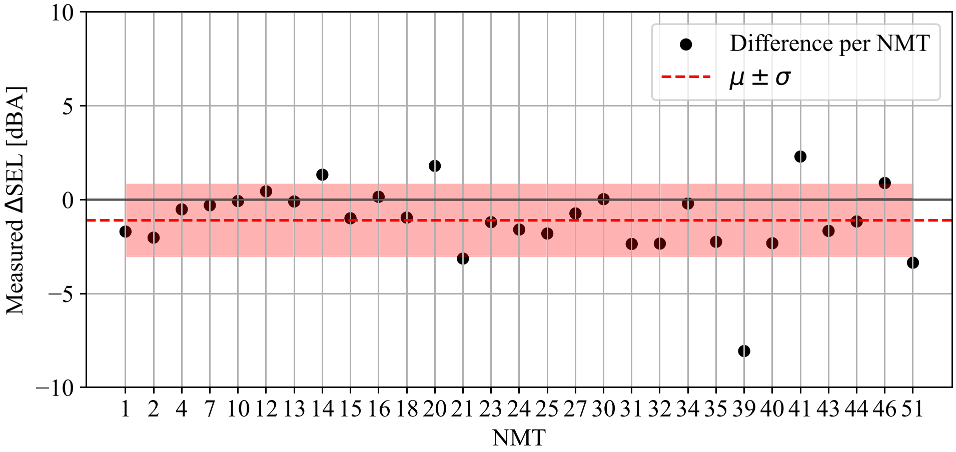


Fig. 11. Mean difference between measured SEL (night - day) per NMT for the operational year 2023.

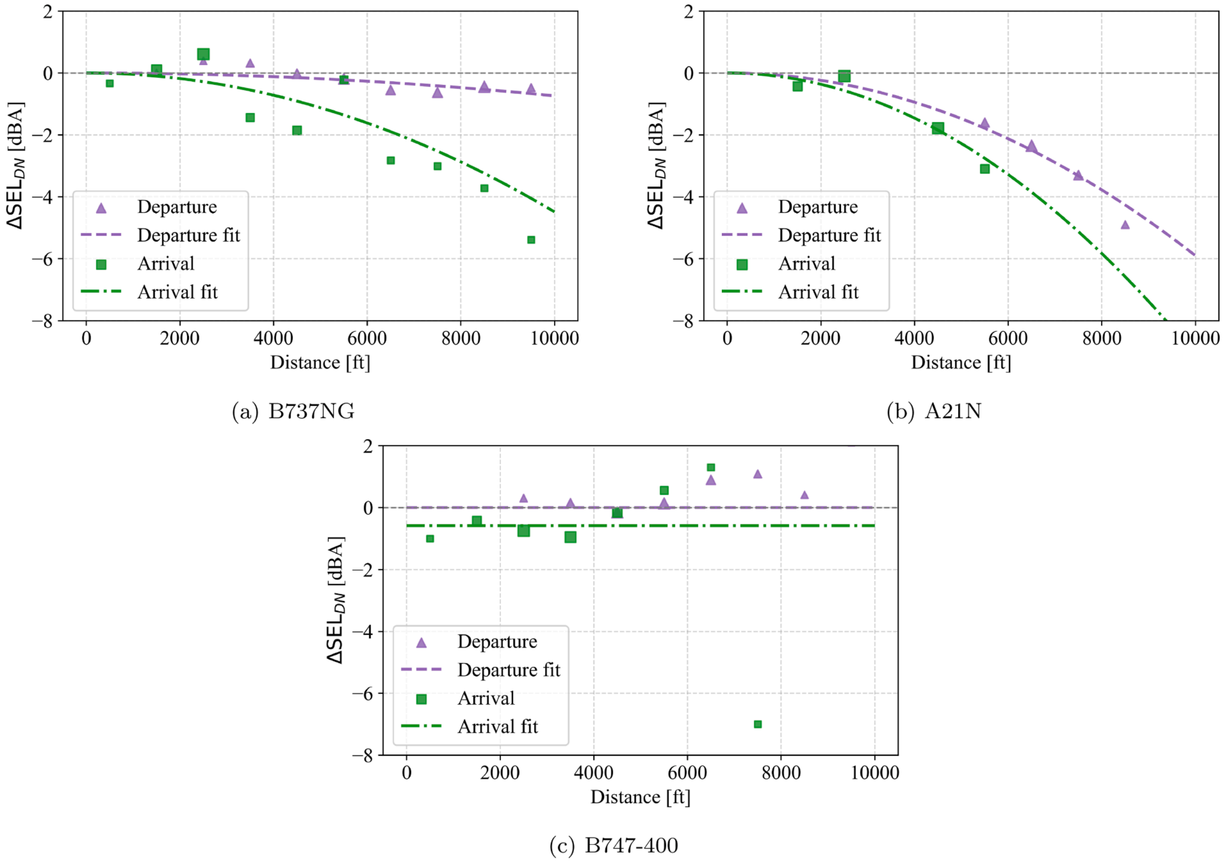


Fig. 12. Weighted average differences in measured SEL between day and night for three aircraft groups for arrival and departure operations with a fitted curve. This includes measurements for all eight operational years. The area size of the points indicates the relative number of measurements in the respective bins.

fitted using second-order polynomial curves (see Fig. 12). The fitted curve through the bins is constrained to start at 0 dBA at the 0 ft bin.

Note that most aircraft types lack sufficient data at either close or far distances. For example, few close-range departure measurements in the 0–1000 ft bin exist due to rapid climb rates and NMT placement away from runways. Similarly, quieter aircraft, such as the A21N or regional jets, often lack long-distance data. This limits the ability to detect statistically significant day-night differences. For the available measurements of the A21N in Fig. 12(b), larger differences in measured SEL between the daytime and nighttime

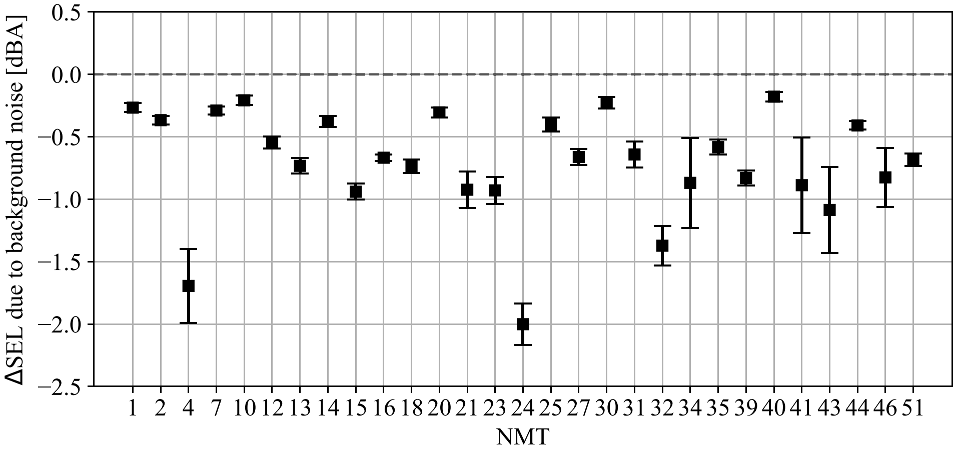


Fig. 13. Estimated effect of background noise on the measured SEL per NMT (arrivals and departures combined), averaged over 8 operational years and the standard deviation between the years.

Table 3
Comparison between modelled and measured L_{den} with estimated effect of ambient noise. Results without ambient noise incorporated are given within brackets.

Year	μ [dBA]	σ [dBA]	ρ
2017	-0.31 (-0.68)	1.55 (1.53)	0.96
2018	-0.57 (-0.99)	1.65 (1.76)	0.94
2019	-0.21 (-0.63)	1.85 (1.86)	0.95
2020	-0.91 (-1.39)	2.40 (2.42)	0.93
2021	0.48 (0.10)	2.07 (2.11)	0.94
2022	0.32 (-0.14)	2.30 (2.28)	0.93
2023	-0.01 (-0.51)	1.99 (1.91)	0.95
2024	-0.53 (-1.07)	2.22 (2.24)	0.94
Average	-0.22 (-0.66)	2.00 (2.02)	0.94

measurements are found due to the lower source noise level of this aircraft type. In contrast, larger aircraft like the B747, shown in Fig. 12(c), show smaller or insignificant differences. This is likely due to their higher source noise levels, which reduce the relative impact of background noise.

From the above results,

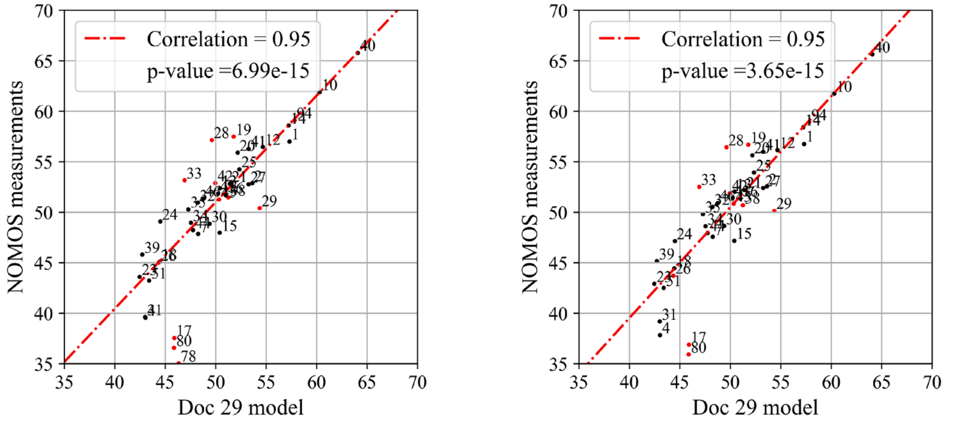
- almost always higher levels per event during daytime compared to nighttime
- consistently smaller differences for aircraft with high source noise levels
- consistently larger differences for aircraft with low noise levels, such as the A21N
- increasing effect with distance

demonstrated that the ambient noise has a significant impact on the measurements during daytime (increased noise levels).

5.2. Effect of background noise

The obtained polynomial curves for ΔSEL_{DN} were used to estimate the effect of background noise on individual daytime measurements by subtracting the ΔSEL_{DN} from the measured levels. This results in the values in Fig. 13 where the effect is averaged over the eight years of measurements. For most NMTs, the effect of background noise is modest, limited to 1 dBA and similar for all evaluated years. For the NMTs further away from the flight path, such as NMTs 4 and 24, the estimated effect is almost twice as large.

The last step is to implement these newfound background noise corrections on all daytime measurements of the total measured L_{den} and compare this to the modelled values at each NMT. The results presented in Section 3 are updated, and the new figures of the year 2023 are visible in Fig. 14(b). The effect on L_{night} is not calculated, as this methodology assumes that background noise at night is already limited. In comparison to Fig. 14(a), the overall impact of the background noise correction on L_{den} is modest, despite larger corrections at NMTs farther from the airport (i.e. > 1 dBA). The average correction across all NMTs is approximately 0.4 dBA. A summary of statistics across years is presented in Table 3, where the values in brackets are the results without the background noise correction. The median difference (μ) between modelled and measured L_{den} decreases after applying the correction, though not entirely eliminated. Standard deviation σ also improves slightly, while the correlation coefficient ρ remains the same for all years.



(a) L_{den} (b) L_{den} with background noise corrected measurements.

Fig. 14. Measured and modelled L_{den} of all NMTs for the year 2023 with and without correction for background noise.

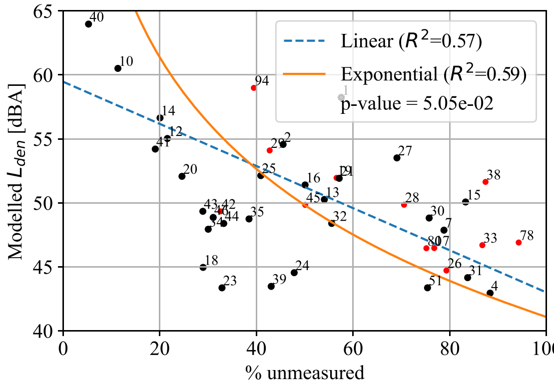


Fig. 15. Percentage of undetected flights per NMT in comparison to the modelled L_{den} averaged over all eight operational years.

In summary, background noise has a significant influence on aircraft noise measurements in uncontrolled environments. Lower nighttime background levels result in lower measured SEL values. These effects are larger for noise events with relatively low SEL values, such as for the quieter aircraft types. The resulting background noise corrections on SEL affect the measured L_{den} only to a small extent, resulting in a modest total correction.

6. Analysis of undetected flights

The third challenge in comparing modelled and measured aircraft noise levels is the fact that not all flights are acoustically recorded by the monitoring system. The modelled noise contours are based on the summation of all flights departing from or arriving at Schiphol. The measured L_{night} and L_{den}, however, only consist of the noise levels of flights that surpass the detection threshold of the respective NMT. Flights that fly past NMTs without being detected are referred to as undetected flights.

6.1. Percentage of undetected flights

To determine the factors contributing to the non-detection of flights, the percentage of undetected flights is calculated with respect to the total number of flights with a PCA within 10,000 ft. The relationship between the percentage of undetected flights and the modelled noise levels L_{den} is shown in Fig. 15. A weak but consistent negative relationship is observed, which can be described by an exponential decay function. To reduce the influence of data scatter, the analysis was performed with a binned representation of the results. In low-noise regions (i.e. lower L_{den} values), flights are more frequently undetected. This is likely because their noise levels remain below the ambient background noise and thus do not exceed the NMT threshold. Despite this trend, the moderate coefficient of determination ($R^2 = 0.59$) and the marginal statistical significance ($p = 5.05 \times 10^{-2}$) indicate that detection performance is also influenced by other factors such as background noise, measurement conditions, and flight variability. A similar trend is observed for L_{night}.

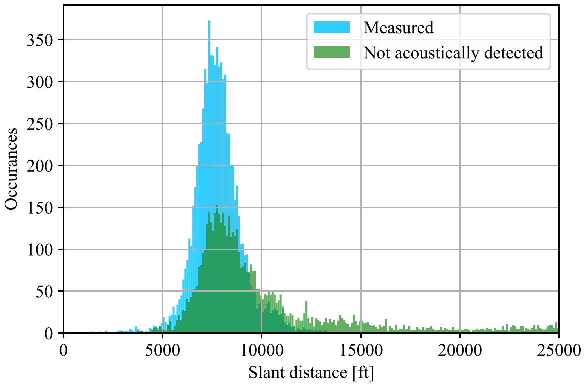


Fig. 16. Slant distances of measured and unmeasured flights of all aircraft types over NMT 18 during the year 2023.

Table 4
Number and percentage of undetected flyovers within 10,000 ft per aircraft type for the operational year 2024.

Aircraft type	$N_{undetected}$	$\%_{undetected}$
B738	183,653	41.4
E190	160,500	45.7
A320	99,749	48.5
E75L	88,979	51.6
E295	78,415	59.1
A20N	68,072	60.9
A319	53,681	50.8
B737	27,622	43.3
B772	27,215	37.2
A21N	25,026	51.6
A333	21,034	33.4

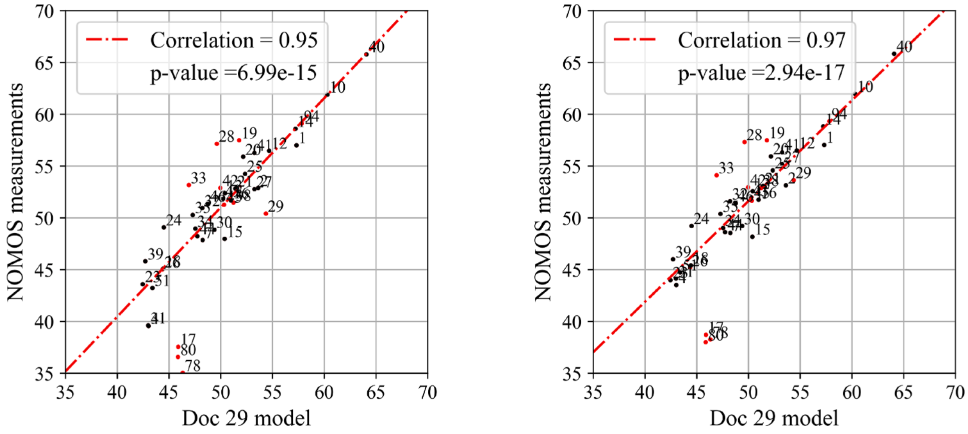
Distance effect: At Schiphol Airport, arrival and departure routes are fixed. As a result, aircraft typically overfly the NMTs at similar slant distances. For NMT 18, located in a low-noise area, flyovers typically occur at a PCA of around 7,000 ft. While flights below this distance are often detected, detection rates drop substantially near 10,000 ft (±3 km). Measurements associated with a PCA larger than 10,000 ft are deemed unreliable, as discussed in Fig. 4 in Section 2.3. Notably, several flyovers at smaller distances remain acoustically undetected. This is illustrated in Fig. 16, showing that distance alone does not fully determine (un)detection.
Aircraft type: Detection, most likely, also depends on aircraft type. Table 4 lists the number and percentage of undetected flyovers for eleven aircraft types. While the B738 accounts for most undetected flights in absolute terms, the relative picture differs. The A20N family shows the highest proportion of undetected flyovers. More than 60% of A20N flyovers within 10,000 ft are not detected. The E295 follows closely with 59% of unrecorded flights. Both aircraft types are relatively newer and quieter.[3] This suggests that their noise emissions often fall below the NMT thresholds. Note that the numbers in Table 4 refer to flyovers rather than flights. A single flight can be measured by an NMT close to the flight path, while it can remain undetected at more distant NMTs.
Operation type: Differences are also observed between the two operations: arrival and departure. When considering flights within a 10,000 ft PCA for all NMTs, 58% of arrival events are not measured on average. In contrast, only 45% of the departures are not recorded. This difference is likely due to the higher noise level of departure operations, which have a greater chance of exceeding the NMT detection thresholds. Analysis also shows a correlation between the difference in measured and modelled noise ΔL_{den} and the percentage of undetected flights at the individual NMTs. This correlation is 0.43 for arrivals and 0.54 for departures, with p-values of 4×10^{-7} and 7×10^{-14}, respectively. These findings highlight the limitations of acoustic monitoring systems in low-noise regions.

6.2. *Assessing the effects of undetected flights on L_{night} and L_{den}*

The effect of undetected flights on the measured L_{night} and L_{den} can be investigated by estimating their noise contributions. In this contribution, this is fully based on measurements and performed using a weighted least-squares method, as described by van der Grift et al. (2025a). The missed 'measurements' are predicted using operational data and the existing measurements.

This method groups flights based on key characteristics influencing noise emission, including aircraft type, operation type (arrival or departure), noise environment, and slant distance at the PCA. For each group, a logarithmic relationship between SEL and distance

[3] Quiet aircraft are defined as those certified under Chapter 14 of EASA noise standards.



(a) Without undetected flights (b) With undetected flights

Fig. 17. Measured and modelled L_{den} with and without the addition of undetected flights of all NMTs for the year 2023.

Table 5
Effect of undetected flights on the measured L_{den} and L_{night} and updated comparison between modelled and measured results. Results without undetected flights incorporated are given within brackets.

Year	ΔL_{night}			ΔL_{den}		
	μ [dBA]	σ [dBA]	ρ	μ [dBA]	σ [dBA]	ρ
2017	0.07 (1.31)	2.07 (2.92)	0.96	-1.18 (-0.68)	1.35 (1.53)	0.97
2018	-0.78 (0.23)	1.48 (2.20)	0.96	-1.48 (-0.99)	1.59 (1.76)	0.95
2019	-0.01 (0.78)	1.65 (2.37)	0.98	-1.17 (-0.63)	1.52 (1.86)	0.96
2020	-0.42 (0.82)	2.68 (4.88)	0.95	-1.86 (-1.39)	1.99 (2.42)	0.94
2021	-0.22 (1.43)	2.97 (4.08)	0.93	-0.72 (0.10)	1.60 (2.11)	0.95
2022	0.36 (1.90)	2.51 (3.67)	0.95	-0.87 (-0.14)	1.73 (2.28)	0.94
2023	0.34 (1.68)	2.62 (3.54)	0.96	-1.11 (-0.51)	1.41 (1.91)	0.97
2024	-1.10 (0.23)	1.79 (3.25)	0.97	-1.65 (-1.07)	1.71 (2.24)	0.98
Average	-0.22 (1.05)	2.22 (3.36)	0.96	-1.25 (-0.66)	1.61 (2.02)	0.96

is fitted using measured data. To ensure reliable extrapolation at larger distances, this fit is regularised using Noise–Power–Distance (NPD) relations, which are incorporated as additional weighted observations in the least-squares formulation. As a result, measured data dominate the fit at short distances, while NPD-based observations constrain the behaviour at larger distances where measurements are sparse or unavailable.

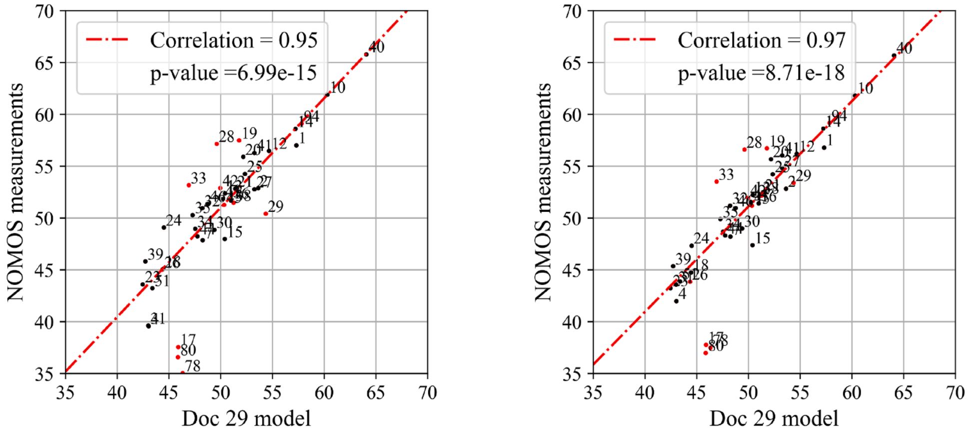
The statistical analysis in van der Grift et al. (2025a) showed that noise levels of undetected flights could be estimated accurately for flights within a 10,000 ft PCA. Noise measurements beyond this distance are considered unreliable and are therefore ignored.

The effect on L_{den} is visualised in Fig. 17 for all NMTs in the year 2023. A clear increase in the measured noise levels is evident for the NMTs in the low-noise regions, such as NMTs 4 and 31. For NMTs with already high measured and modelled L_{den} values, the inclusion of undetected flights has only a limited effect. The correlation between model and measurement also improves after accounting for undetected flights. The results for all operational years are summarised in Table 5.

The results are consistent over the different operational years. Slightly larger effects of undetected flights are observed in 2021 and 2022. For all years, the correlation coefficients improve after accounting for undetected flights. The standard deviation between measured and modelled noise levels is also reduced. This indicates a better agreement between the measured and modelled noise levels for both L_{night} and L_{den}. The updated yearly averages can therefore provide a more realistic representation of the noise exposure in the low-noise regions.

The effect of undetected flights is more pronounced for L_{night} than for L_{den}. On average, this effect is approximately doubled on $\Delta\mu$: 1.27 dBA for L_{night} and 0.59 dBA for L_{den}. This difference is primarily due to a higher percentage of undetected flights at night, particularly for NMTs located in low-noise areas. One possible explanation is that ambient sound during the day amplifies aircraft noise events, allowing them to exceed the NMT detection threshold more frequently. At night, lower ambient noise results in fewer noise events exceeding the threshold. For most NMTs, the threshold is the same during the day and night. For a small number of NMTs (e.g. NMTs 25 and 41), where nighttime thresholds are set lower, similar percentages of undetected flights are observed during both day and night, supporting this hypothesis.

Including undetected flights reduces the model overestimation for L_{night}, while increasing the underestimation of L_{den}. The undetected flights are most prevalent in low-noise regions, where a relatively large proportion of events is initially missing from the measurement dataset. Adding these contributions increases the measured L_{den} values more strongly at these locations than at higher-



(a) Without corrections ($\sigma = 1.91$) (b) With corrections ($\sigma = 1.40$)

Fig. 18. Measured and modelled L_{den} with and without the addition of undetected flights and the ambient noise correction of all NMTs for the year 2023.

Table 6
Estimated combined effect of background noise and undetected flights on L_{den} and updated comparison between modelled and measured results. Results without effects incorporated are given within brackets.

Year	μ [dBA]	σ [dBA]	ρ
2017	-0.79 (-0.68)	1.35 (1.53)	0.97 (0.96)
2018	-1.06 (-0.99)	1.46 (1.76)	0.96 (0.94)
2019	-0.72 (-0.63)	1.47 (1.86)	0.96 (0.95)
2020	-1.37 (-1.39)	1.91 (2.42)	0.95 (0.92)
2021	-0.32 (0.10)	1.55 (2.11)	0.96 (0.94)
2022	-0.39 (-0.14)	1.70 (2.28)	0.95 (0.92)
2023	-0.61 (-0.51)	1.40 (1.91)	0.97 (0.95)
2024	-1.10 (-1.07)	1.66 (2.24)	0.96 (0.94)
Average	-0.80 (-0.66)	1.56 (2.02)	0.96 (0.94)

noise sites, thereby altering the spatial distribution of measured noise levels. As a result, the comparison with the modelled results shifts, leading to a larger apparent underestimation by the model. At the same time, the correlation between modelled and measured levels improves significantly, indicating a more consistent representation of noise exposure across all NMTs. This behaviour is consistent with the previously observed relationship between low L_{den} values and a high percentage of undetected flights.

These findings highlight the importance of accounting for detection limitations when comparing measured and modelled noise levels, particularly for nighttime operations.

7. Combined results

The three main causes challenging the comparison of annual measured and modelled aircraft noise levels are identified in the previous sections. The model input discrepancies are difficult to quantify for the full year due to limited data availability. However, the effects of background noise and undetected flights are quantified and are combined in this section to give the most accurate result. For consistency, these results are plotted in Fig. 18 for the year 2023. The results summarised for all years are presented in Table 6.

Note that combining the corrections is not a simple addition of the two correction factors. As the background noise correction also influences the estimated SEL of the undetected flights, the combined effect is calculated by first incorporating the undetected flights to the measured L_{den} values and then applying the background-noise correction. Nevertheless, the effect of undetected flights is generally stronger than that of background noise. An effect of $+0.14$ dBA averaged over all NMTs over all eight years is found. These combined effects improve the correlation between measured and modelled noise levels, with an average correlation coefficient of 0.96. The average difference μ is slightly larger than the baseline levels depicted in Table 2. At the same time, the standard deviation σ is lowest compared to all previous results. This indicates that jointly accounting for undetected flights and background noise leads to a more consistent comparison between measured and modelled annual noise levels.

While the combined corrections improve consistency, a residual underestimation of L_{den} remains. This suggests that additional factors related to the modelling approach may contribute to the observed discrepancies, such as limitations in the representation of noise propagation, particularly at larger distances from the flight path.

8. Conclusions

This research investigated the challenges associated with comparing the annually averaged aircraft noise levels predicted by the Doc 29 model with measurements around Amsterdam Airport Schiphol. The analysis was performed through noise measurements from the Noise Monitoring System (NOMOS) from the operational years 2017 to 2024. The analysis revealed a strong overall correlation between modelled and measured results, confirming the robustness of Doc 29 as a best-practice tool. However, systematic discrepancies were identified: the model generally overestimates L_{night} while it underestimates L_{den}. To further investigate these discrepancies, this paper investigated three main challenges influencing the comparison between modelled and measured noise levels.

The first challenge stems from differences between the model input parameters and the actual flown flight procedures. Operational analysis demonstrated mismatches between modelled procedures and flown tracks. Variability in flown speed, altitudes, thrust settings, and atmospheric conditions introduces uncertainties that are not fully captured by current fixed-point profiles. Although deviations of up to 3 dBA were found for individual flights, their impact on annual noise metrics cannot be quantified with the available data. In addition, because such deviations can be both positive and negative, their average effect could partially cancel out over a large number of operations. Therefore, it is impossible to determine whether accounting for this source of uncertainty would improve or worsen the agreement between measured and modelled annual noise levels.

The second challenge concerns the measurement offset introduced by background noise. This effect was quantified by examining the differences between day and night noise levels. While lower nighttime background noise contributes to reduced measured SEL values, operational factors such as increased CDA usage and fleet mix differences play a substantial role. Applying background noise corrections led to modest improvements in agreement for L_{den} ($\pm$0.4 dBA), but did not fully eliminate the discrepancies.

The third addressed challenge in this comparison stems from undetected flights. Particularly in low-noise regions and for newer, quieter aircraft types such as the A21N and E295, the NMTs frequently failed to measure the flyovers. On average, 58% of arrivals and 45% of departures within 10,000 ft PCA were not acoustically detected. Correcting for these missing events improved the correlation between modelled and measured levels and reduced variability, with average adjustments of +1.2 dBA for L_{night} and +0.6 dBA for L_{den}. These findings highlight the importance of explicitly accounting for detection thresholds when validating model predictions against measurements. A recommendation is to apply a dynamic threshold methodology that accounts for real-time background noise conditions, thereby allowing a measurement system to detect lower-level noise events.

The findings have important implications for the interpretation and validation of aircraft noise assessments. The strong correlation between measured and modelled noise levels confirms that the Doc 29 methodology provides a suitable representation of annual noise exposure and can therefore support policy applications. At the same time, the results demonstrate that discrepancies between measurements and model predictions cannot be attributed solely to modelling assumptions. Background noise, incomplete flight detection, and operational variability can all contribute substantially to the observed differences. Consequently, comparisons between modelled noise contours and monitoring data should explicitly account for uncertainties related to both modelling and measurement processes.

Some limitations should be acknowledged. This study was performed using data from a single airport and relies partly on airport-specific datasets and correction methodologies. While the identified mechanisms are expected to be relevant for other airports as well, their quantitative contribution may differ depending on local operational characteristics and environmental conditions. In addition, the analysis focused on the annual average noise metrics L_{den} and L_{night}. The implications of the sources of uncertainty, considered in this contribution, for event-based noise metrics require further investigation.

In summary, the validation of annual average noise metrics requires careful treatment of operational variability, background noise effects, and undetected flights. While Doc 29 provides reliable predictions at the system level, measurement-based validation must address biases inherent to acoustic monitoring in uncontrolled environments. The remaining discrepancies may also be partly related to limitations in the modelling approach, such as the use of NPD relations, which are known to become less accurate at larger slant distances where measurements are sparse. Future work should focus on refining the operational model input with real flight data and improving the detection of low-level noise events. Such efforts will strengthen confidence in aircraft noise modelling and support more effective policy development for communities around major airports.

CRediT authorship contribution statement

Rebekka C. van der Grift: Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Alireza Amiri Simkooei:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization; **Mirjam Snellen:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Acknowledgments

This work was mainly supported by the Knowledge and Development Centre (KDC) Mainport Schiphol. The authors are also grateful to the EC for supporting the present work, partly performed within the NEEDED project, funded by the European Union's Horizon Europe research and innovation programme under grant agreement no. 101,095,754 (NEEDED). This publication solely reflects the authors' view and neither the European Union nor the funding Agencies can be held responsible for the information it contains.

Data availability

The data that has been used is confidential.

References

Baarda, W., 1968. A testing procedure for use in geodetic networks. Technical Report. Netherlands Geodetic Commission, Publ. on Geodesy, New Series, Vol. 2(5), Delft.

Bertsch, L., Pott-Pollenske, M., Wienke, F., Kurz, J., Delfs, J., 2024. Retrofit measures for aircraft noise reduction: simulation benchmark and impact assessment. J. Aircr. 62, 191–206. https://doi.org/10.2514/1.C037819

Bhanpato, J., Puranik, T.G., Mavris, D.N., 2021. Data-driven analysis of departure procedures for aviation noise mitigation. Eng. Proc. 13 (1), 2. Presented at the 9th OpenSky Symposium, Brussels, Belgium, 18–19 November 2021. https://doi.org/10.3390/engproc2021013002

Conference, E. C.A., 2016a. Doc 29 Report on Standard Method of Computing Noise Contours around Civil Airports, 4th Edition Report, Volume 2: Technical Guide. Technical Report. ECAC.CEAC Doc29. Neuilly-sur-Seine Cedex, France.

Conference, E. C.A., 2016b. Doc 29 Report on Standard Method of Computing Noise Contours around Civil Airports, 4th Edition, Volume 1: Applications Guide. Technical Report. ECAC.CEAC Doc29. Neuilly-sur-Seine Cedex, France.

Engineers, S. o.A., 2012. Procedure for the Calculation of Airplane Noise in the Vicinity of Airports. SAE-AIR-1845A. SAE International. Warrendale, PA, USA. https://doi.org/10.4271/AIR1845A

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. Accessed: 2026-07-31, 2026.

Filippone, A., Harwood, A., 2016. Flyover noise measurements and predictions of commercial airplanes. J. Aircr. 53. https://doi.org/10.2514/1.C033370

Giladi, R., 2020. Real-time identification of aircraft sound events. Transp. Res. D: Transp. Environ. 87, 102527. https://doi.org/10.1016/j.trd.2020.102527

Giladi, R., Menachi, E., 2024. Validating aircraft noise models: aviation environmental design tool at heathrow. J. Air Transp. Manag. 116, 102557. https://doi.org/10.1016/J.JAIRTRAMAN.2024.102557

Gjestland, T., 2019. Environmental Report 2019: Aviation and Environment. Technical Report. International Civil Aviation Organization. https://www.icao.int/environmental-protection/Documents/EnvironmentalReports/2019/ENVReport2019_pg89-92.pdf.

van der Grift, R.C., Jayanthi, A.S., Snellen, M., Amiri Simkooei, A., 2025a. Analysing the effect of undetected flights in the comparison of measured and modelled aircraft noise. In: Forum Acusticum Euronoise 2025: Proceedings of the 11Th Convention of the European Acoustics Association Forum Acusticum / EuroNoise 2025. European Acoustics Association, EAA, France, pp. 5607–5614. https://doi.org/10.61782/fa.2025.0129

van der Grift, R.C., Snellen, M., Amiri-Simkooei, A., 2025b. The effect of using n1% as input for aircraft noise modelling. Transp. Res. D: Transp. Environ. 142, 104710. https://doi.org/10.1016/J.TRD.2025.104710

Guo, W., Zeng, W., Zhou, Y., 2025. A copula-ECAC model for estimating aviation noise around airports. Transp. Res. D: Transp. Environ. 142. https://doi.org/10.1016/j.trd.2025.104666

Heblij, S.J., Derei, J., 2019. Methodenrapport Doc29. Technical Report. NLR-Nederlands Lucht-en Ruimtevaartcentrum.

Heblij, S.J., Derei, J., Hogenhuis, R.H., 2019. Toepassing ECAC Doc29 voor het bepalen van de geluidbelasting van het vliegverkeer van Schiphol. Technical Report. Royal Dutch Aerospace Centre (NLR). Amsterdam, Netherlands.

Janssen, S.A., Vos, H., Kempen, E. E.M.M.V., Breugelmans, O. R.P., Henk, Miedema, M.E., Miedema, H. M.E., 2011. Trends in aircraft noise annoyance: the role of study and sample characteristics. J. Acoust. Soc. Am. 129, 1953–1962. https://doi.org/10.1121/1.3533739

Jayanthi, A.S., Snellen, M., Amiri-Simkooei, A., 2026. Investigating and quantifying the ground effect using spectrograms of realistic aircraft flyover measurements. CEAS Aeronaut. J.. https://doi.org/10.1007/s13272-026-00947-1

Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. 18 (1), 50–60.

Peters, J.L., Zevitas, C.D., Redline, S., Hastings, A., Sizov, N., Hart, J.E., Levy, J.I., Roof, C.J., Wellenius, G.A., 2018. Aviation noise and cardiovascular health in the united states: a review of the evidence and recommendations for research direction. Curr. Epidemiol. Rep. 2018 5:2 5, 140–152. https://doi.org/10.1007/S40471-018-0151-2

van Pijlen, E. S.D., 2026. Thrust and Weight Estimation for Doc. 29 Noise Models: Using ACMS Data to More Accurately Predict Noise Levels at Amsterdam Airport Schiphol. Master's thesis. Delft University of Technology. Delft, The Netherlands. https://resolver.tudelft.nl/uuid:85eaf1b8-0022-4d38-9cd5-6dc3421cd6c4.

Pretto, M., Dorbolò, L., Giannattasio, P., Zanon, A., 2024. Aircraft operation reconstruction and airport noise prediction from high-resolution flight tracking data. Transp. Res. D: Transp. Environ. 135, 104397. https://doi.org/10.1016/J.TRD.2024.104397

Rhodes, D., 2018. Guidance on comparing calculated aircraft noise levels with measurements. Technical Report. UK Civil Aviation Authority.

Rindfleisch, T.C., Alonso, J.J., Jackson, D.C., Munguía, B.C., Bowman, N.W., Ia, B. C.M., 2024. A large-scale validation study of aircraft noise modeling for airport arrivals. J. Acoust. Soc. Am. 155, 1928–1949. https://doi.org/10.1121/10.0025276

Sahai, A.R., Hogenhuis, R.N., Heblij, S.N., Smetsers, R.R., Verver, G.K., Assink, J.K., 2021. Programmatische Aanpak Meten Vliegtuiggeluid: Nationale Meetstrategie. Technical Report. National Institute for Public Health and the Environment Netherlands (RIVM). https://doi.org/10.21945/RIVM-2021-0001

Saucy, A., Schäffer, B., Tangermann, L., Vienneau, D., Wunderli, J.M., Röösli, M., 2021. Does night-time aircraft noise trigger mortality? a case-crossover study on 24886 cardiovascular deaths. Eur. Heart J. 42, 835–843. https://doi.org/10.1093/EURHEARTJ/EHAA957

Simons, D.G., Amiri-Simkooei, A., Melkert, J.A., Snellen, M., 2026. Contribution of aircraft types to noise levels across the NOMOS network of amsterdam airport schiphol. J. Air Transp. Manag. 131, 102910. https://doi.org/10.1016/J.JAIRTRAMAN.2025.102910

Simons, D.G., Besnea, I., Mohammadloo, T.H., Melkert, J.A., Snellen, M., 2022. Comparative assessment of measured and modelled aircraft noise around amsterdam airport schiphol. Transp. Res. D: Transp. Environ. 105, 103216. https://doi.org/10.1016/J.TRD.2022.103216

Snellen, M., Merino-Martínez, R., Simons, D.G., 2017. Assessment of noise variability of landing aircraft using phased microphone array. J. Aircr. 54, 2173–2183. https://doi.org/10.2514/1.C033950

Standardization, I. O.f., 2009. Acoustics - Unattended monitoring of aircraft sound in the vicinity of airports. Technical Report ISO 20906:2009. International Organization for Standardization. Reviewed and confirmed in 2020.

Teunissen, P.J.G., 2000. Testing theory: an introduction. Delft University Press, Series on Mathematical Geodesy and Positioning. Website: http://www.vssd.nl.

U. S. Department of Transportation Volpe National Transportation Systems Center, 2023. Aviation Environmental Design Tool (AEDT) Version 3f. Technical Manual. U.S. Department of Transportation. Washington, DC, USA.

Vieira, A., von den Hoff, B., Snellen, M., Simons, D.G., 2022. Comparison of semi-empirical noise models with flyover measurements of operating aircraft. J. Aircr., 1–14. https://doi.org/10.2514/1.C036387

Wunderli, J.M., Meister, J., Jäger, D., Schalcher, S., Zellmann, C., Schäffer, B., 2022. Aircraft noise in situations with grazing sound incidence—comparing different modeling approaches. J. Acoust. Soc. Am. 151, 3140–3151. https://doi.org/10.1121/10.0010419

Zellmann, C., Schäffer, B., Wunderli, J.M., 2018. Aircraft noise emission model accounting for aircraft flight parameters. J. Aircr. 55, 682–695. https://doi.org/10.2514/1.C034275