

Using Open Data to Characterise Traffic Flows and Operational Concepts across International Aerodromes

Jean Benevides, Rainer Koelle, Quinten Goens, and Enrico Spinielli

Abstract

Common performance indicators can conceal differences in how airport arrival and departure flows are organised and used. Describing this operational context alongside the indicators supports a more informed interpretation of international benchmarking. Using OpenSky Network observations for 1–14 July 2026, this work-in-progress derives linked milestone tables and ordered spatial sequences. The proposed framework connects adapted additional time in arrival and departure airspace and time flown level during climb and descent to the operating arrangement, recurring flows and variation within them, retaining the observational support of each measure. Application to seven airports in the Performance Benchmarking Working Group (PBWG) study demonstrates how the linked evidence supports these classical performance measures, while differences in surveillance coverage and measurement requirements lead to different analytical populations. Recurring patterns identified from ordered spatial sequences provide a basis for examining corridor adherence, repeated traversal and the location of level flight. The examples show that similar corridor adherence can accompany different path use, and that changes in runway configuration can introduce flows poorly represented by an existing reference set. These findings show how operational context qualifies aggregate comparisons, although the short samples and measure-specific selection do not establish representative airport performance or the causes of observed differences. The framework provides a common basis for organising performance and operational-context evidence, supporting comparison with operational experts and further assessment across periods and operating conditions.

Keywords: OpenSky Network; milestones; airport flows; vertical flight efficiency; contextual indicators

1. Introduction

Operational flight efficiency is an established element of the ICAO performance framework. Common performance indicators provide a basis for monitoring developments and comparing air navigation performance across regions [1]. Their interpretation also requires an understanding of the operational context in which the observed performance is achieved. Differences in traffic demand, runway use and the organisation of arrival and departure flows are therefore relevant to international comparisons.

International collaboration on operational performance has expanded through bilateral and multilateral comparison activities. Examples include comparisons between the United States and Europe [2], Brazil and Europe [3], and initial work between China and Europe [4]. These activities develop commonly agreed metrics, comparable datasets and a shared interpretation of similarities and differences between air navigation systems and their operational arrangements. The work undertaken through the Performance Benchmarking Working Group (PBWG) provides a practical setting for further developing this approach [3, 4].

At airport level, additional time in terminal airspace and vertical flight efficiency describe complementary aspects of observed flight operations [5, 6]. Similar indicator values may nevertheless arise from different arrival and departure patterns. An aggregate value does not indicate where level flight occurs, how consistently aircraft follow recurring corridors, or how their paths vary within those corridors. Describing these features alongside the indicators can support a more informed comparison of operational performance. The objective of this paper is to examine how the trajectory observations used for performance measurement can also support a common characterisation of the associated operational context.

OpenSky Network observations provide a common source of trajectory data for this investigation [7]. The approach derives operational milestones and trajectory segments that support both indicator calculation and the description of arrival and departure flows. Adapted arrival and departure sequencing and metering area (ASMA/DSMA) timing measures and vertical flight efficiency (VFE) measures are complemented by descriptors of corridor adherence, repeated path use, vertical structure and local flow composition. Observed runway movement counts describe observed throughput. Shared movement identities allow these descriptions to be related while retaining the observational support available for each analysis.

Previous work has linked arrival performance to entry conditions and individual flows [8]. The present study contributes a linked representation of indicators and ordered path use, retaining the observational support of each component:

- **A common evidence base for performance analysis:** a reusable movement representation connecting adapted terminal timing and vertical flight efficiency measures across selected international airports.
- **Complementary flow-based measures:** ordered spatial descriptions of corridor adherence, repeated traversal, vertical structure and local flow composition.
- **An operational-context characterisation framework:** integration of indicators and flow descriptions with measure-specific support, demonstrated through operational comparisons across the selected airports.

The demonstration covers seven airports selected for the PBWG study for 1-14 July 2026. Weekly ASMA/DSMA and VFE comparisons and configuration-conditioned profiles retain separate samples; first-week diagnostics illustrate the additional information provided by the flow measures. Differences in observational and reference support remain explicit. The results provide an initial basis for examining observed flow organisation alongside performance indicators and for directing further analysis with operational experts. The framework is intended to support more consistent interpretation in international benchmarking and to guide its subsequent extension to additional periods and operational settings.

2. Background

2.1 Performance-based approach and global harmonisation

ICAO's Global Air Navigation Plan (GANP) provides the strategic framework for the evolution of air navigation towards a safe, sustainable, and interoperable system [9]. The performance-based approach links desired outcomes to objectives and targets, with monitoring providing the evidence needed to assess progress and inform operational improvement [1]. In this setting, performance measurement supports a continuing process of reviewing the effects of operational changes and identifying where further action may be required.

The GANP Key Performance Indicator (KPI) catalogue provides a *lingua franca* for international performance benchmarking. It establishes a common vocabulary through which States and regions can express performance, while their monitoring arrangements and operational environments may differ. The accompanying guidance supports the application and adaptation of indicators to specific performance questions [10]. Common indicator names alone do not ensure comparable results: the population of flights, measurement boundaries, reference conditions and aggregation conventions also need to be understood.

The distinction between basic and advanced indicators reflects differences in data and processing requirements. Operational timing and flight-plan information can support many basic measures, while advanced measures of horizontal and vertical flight efficiency require trajectory observations and more detailed processing [11]. This creates a practical requirement for documenting how observations become performance evidence. Harmonised definitions provide the starting point for comparison; their application also requires a clear account of the observations, analytical boundaries and operating conditions represented by the resulting values.

2.2 International benchmarking and operational performance indicators

International benchmarking puts this common framework into practice through agreement on definitions, the identification of comparable data and joint examination of the results. The regional comparisons introduced above illustrate this process: technical collaboration includes reviewing data sources and processing conventions, interpreting differences and identifying questions for further analysis [2, 3, 4]. Worked examples provide a means of examining how indicators behave under different conditions. PBWG offers a setting for sharing these experiences and supporting a harmonised interpretation of the GANP indicators.

Airport throughput describes the number of movements served over a defined interval. Its interpretation depends on the traffic demand, runway arrangements and observation period represented. Peak throughput is therefore distinct from declared capacity, and counts from a partial surveillance sample require an explicit account of their coverage [3].

Additional time in terminal airspace compares an observed transit time with a reference intended to represent unimpeded conditions. Additional ASMA time applies this principle to arrivals; the choice of reference population is central to its interpretation [5]. A corresponding departure perspective considers the outbound transit, with its reference conditions and measurement boundaries stated separately.

Vertical flight efficiency describes level-flight exposure during climb and descent and supports assessment of the objectives associated with continuous climb and descent operations. Time or distance flown level and the proportion of flights meeting a specified continuity criterion answer related but distinct questions [6]. Together, throughput, terminal timing and vertical measures describe complementary aspects of airport operations. Their joint interpretation benefits from information on the traffic served and the horizontal and vertical organisation of the associated flows.

2.3 Open performance data and operational milestones

Open surveillance observations extend the opportunities for performance analysis across airports and regions. The OpenSky Network provides access to aircraft observations collected through a distributed receiver network [7]. Tools such as *traffic* support the processing and analysis of trajectory data [12]. These developments support common analytical methods across locations, while usable evidence depends on reception coverage and observation characteristics. Rahman et al. model reception probability in relation to distance and traffic-related conditions [13]; message reception and movement-population completeness remain distinct questions.

For performance monitoring, access to observations needs to be accompanied by documented processing and an account of data quality. A position report is one observation within a flight; deriving a transit time or a vertical profile requires identifying the relevant movement and the portion of its trajectory represented. Missing observations can affect those quantities differently. Reproducibility consequently depends on retaining the relationship between source observations, derived events and the samples used for each measure.

Operational milestones provide a useful way of expressing that relationship. Examples include an estimated take-off or landing event, passage through a distance boundary and the limits of a climb or descent segment. Schultz, Rosenow and Olive demonstrate the derivation of operational milestones from ADS-B messages for airport performance applications aligned with Airport Collaborative Decision Making [14]. The Open Performance Data Initiative provides reusable flight events and associated measurements through trajectory reduction [15]. The present study adopts a related approach, linking milestones to ordered spatial visits and measure-specific observation support. A common data source supports consistency, but comparability also depends on how movements are identified, which trajectory portions are retained and how missing observations are treated.

2.4 Flow characterisation for operational comparison

Trajectory analysis provides methods for identifying recurring patterns within a traffic sample. Grouping trajectories according to their spatial or temporal characteristics can reveal flows that are less apparent in aggregate movement counts. Basora, Morio and Mailhot describe a trajectory-clustering framework for identifying and analysing traffic flows, including applications to airspace management and traffic-pattern analysis [16]. For terminal flows, Corrado et al. address convergence and divergence through a weighted trajectory distance, illustrating how the comparison method affects the flows resolved [17].

Hardell et al. combine horizontal and vertical efficiency, additional fuel burn and metering effort, including per-flow analysis at three European airports [8]. Their results establish the relevance of entry conditions to performance interpretation. The present study complements this approach by linking adapted timing measures to ordered path use, vertical structure and local branch composition, with explicit support for each measure.

For operational comparison, two complementary questions are relevant. The first concerns **how flows are organised**: the use of separate corridors, the convergence or divergence of branches, and the sharing of route portions. The second concerns **how flights vary within those flows**: their lateral distribution, repeated or extended path use, and vertical progression. An airport can have several well-defined corridors and still exhibit substantial variation within individual flows. Conversely, a broad overall footprint may reflect the use of several concentrated routes.

These perspectives allow performance observations to be examined at a more informative level. A timing distribution can be considered by recurring flow, and level-flight exposure can be located along a route or within range and height bands. Interpretation depends on the flights and trajectory portions contributing to each description; differences in sample composition need to remain visible when the measures are combined.

Observed geometry does not by itself establish a published procedure, a controller instruction or the cause of a performance difference. These interpretations require supporting operational information. The value for benchmarking lies in linking the measured outcome to the observed organisation of traffic, providing a firmer basis for understanding similarities and differences and for identifying where operational investigation is warranted.

3. Concept, Methods, and Data

3.1 Concept and operational-context framework

The proposed framework links operational performance indicators to a structured description of the observed traffic. Comparisons are organised by airport, movement phase and reporting period, with further differentiation by runway and recurring flow where supported by the data. Observed runway movements and runway use describe the traffic sample; timing indicators describe transit relative to an observed reference; flow measures describe horizontal organisation and variation; and vertical profiles describe the distribution of climb, descent and level flight. Coverage and sample support accompany each component, allowing differences in observed performance to be considered alongside differences in operational context.

Operating direction and exact runway configuration describe the observed operating situation. Within that context, the framework addresses two complementary questions: how is traffic organised, and how do trajectories vary within that organisation? Recurring paths and their shared corridors describe organisation. Corridor adherence, repeated path use and horizontal and vertical distributions describe variation within flows. Together with the indicators, these components form an operational profile for comparison, without combining them into a single score or assigning airports to fixed categories.

The architecture separates trajectory preparation from repeated analysis (Figure 1). A common movement identity connects operational-event estimates, range crossings and observed intervals to ordered spatial visits. Analysts can select measurement envelopes and examine where time or level flight occurs using these reusable products. The resulting profile retains the sample and observation support of each measure; a missing component does not invalidate other supported evidence.

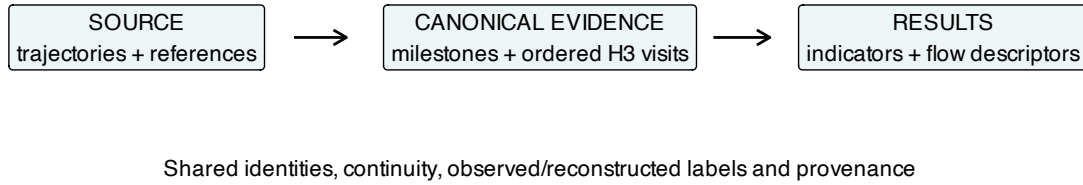


Figure 1. The common representation connects preparation, indicators and context. Quality and eligibility accompany each movement and interval rather than being applied as an airport-wide exclusion.

3.2 Trajectory preparation and reusable evidence

Operational milestones provide the link between surveillance observations and performance analysis [14]. Preparation identifies the movements of interest, establishes the usable trajectory segments and derives evidence for the subsequent analyses:

- **Movement identification.** The classifier identifies arrival and departure candidates and associates runway information where supported. Accepted movements form the analytical payload; unresolved assignments and rejected candidates retain their status for audit. Ambiguous near-repeat movement identities are quarantined rather than merged into an assumed flight.
- **Observation support.** Position updates are checked for temporal order, continuity and plausible displacement. Repeated updates do not create additional observed exposure. Observed intervals are limited to 30 seconds; gaps and source-surface discontinuities remain explicit.
- **Controlled healing.** Trajectory fragment linkage and labelled reconstruction preserve usable trajectory context where supported by the preparation rules. Linking fragments does not establish observation within a gap. Reconstructed geometry remains distinguishable from observed segments and does not supply observed level-flight time; timing eligibility is assessed separately from geometric continuity.
- **Reusable products.** Operational-event estimates and range crossings support observed throughput and transit-time calculations. Observed intervals, level-flight episodes and profile limits support vertical analysis. Spatial discretisation uses H3, the hierarchical geospatial index developed at Uber, to represent positions within a hexagonal tessellation [18]. Ordered cell visits retain traversal order and revisits for flow analysis. Movement, source and trajectory-surface identifiers link these products and their quality decisions.
- **Reference conventions.** Published airport reference points (ARPs) define analytical range and height, while runway geometry supports event estimation. Heights are reported as MSL altitude minus ARP elevation, without pressure or geoid correction, and are not terrain-relative heights. Detailed preparation rules, data lineage, reference provenance and implementation specifications are retained in the companion Technical Note and its methods supplement.

3.3 Performance indicators and flow-based measures

The performance analysis focuses on two GANP indicator families: additional terminal transit time, represented by additional ASMA time (KPI08), and vertical flight efficiency during climb and descent (KPI17 and KPI19) [11]. The study also applies the additional-time algorithm to outbound traffic, yielding an analogous DSMA descriptor, as explored within PBWG. This departure application is distinguished from the catalogue’s additional-ASMA indicator.

Additional ASMA/DSMA time and observed throughput. Arrival transit time extends from the correctly directed fresh range crossing to estimated landing; departure transit time extends from estimated take-off to the outward

crossing. Events and crossings must belong to the supported trajectory surface. For movement i in reference group g , additional time a_i and its airport/phase average are

$$a_i = T_i - Q_{0.20}(T_j : j \in g), \quad \bar{A} = \frac{1}{n} \sum_{i=1}^n a_i.$$

The result is reported in minutes per arrival (ASMA) or departure (DSMA). Groups combine airport, phase, runway end/scope, crossing-bearing quadrant and horizon, with at least 30 movements on three UTC operational days. Timing support covers 40/100/200 NM; the main comparisons use 40 and 100 NM. References are estimated separately for each week. These short-period adaptations use weekly 20th-percentile (P20) references rather than the 5th-percentile (P05) reference from a longer historical sample in the published additional-ASMA method [5]. Sector and inclusion conventions also differ, as documented in the supplement. Movement-level differences below the reference are retained in the average. Additional time is an observed-reference comparison, without attributing it to a particular control action. Rolling 60-minute counts of timing-eligible events, stepped every 15 minutes, describe observed runway throughput; incomplete surveillance coverage may limit the recorded counts, which do not establish total airport throughput or capacity.

VFE during climb and descent. Time flown level and the number of level segments provide familiar measures for assessing continuous climb and descent operations (CCO/CDO) [6]. Let L_i be qualifying level-flight time, N_i the number of qualifying segments and O_i observed time in the measurement envelope. The headline measure is average time flown level, in seconds per arrival or departure; the mean number of level segments describes frequency. The percentage of total observed time flown level provides complementary information:

$$\bar{L} = \frac{\sum L_i}{n}, \quad \bar{N} = \frac{\sum N_i}{n}, \quad V_{\text{exposure}} = 100 \frac{\sum L_i}{\sum O_i}.$$

Candidate level episodes have an absolute altitude rate no greater than 300 ft/min. After clipping to the envelope, continuous components qualify at 20 seconds or more with an altitude excursion no greater than 200 ft. Gaps split episodes; these qualifying components supply N_i . Qualification precedes allocation to range, height or spatial cells, preserving short contributions from longer episodes. Range/height contributions use the whole selected cohort's observed exposure as denominator and sum to V_{exposure} . A mean of individual level-time percentages is retained for the matched first-week diagnostics and distinguished from the pooled percentage in the supplement.

The airport overview uses each indicator's own eligible sample. Linked comparisons additionally require movements eligible for both timing and VFE at 40/100 NM. Conventional lower height limits are 1,800 ft for descent and 3,000 ft for climb, expressed here above ARP. Profiles require at least 60 seconds in the temporal envelope, at least 90% qualifying observed temporal exposure and positive observed distance. VFE spans the range and height gates; timing spans the range crossing and operational event. Changing the horizon changes both exposure and sample support. These bounded, adapted profiles support a CCO/CDO perspective, but a profile without a detected level segment does not establish continuity of the entire climb or descent. Full envelope definitions and differences from the published methodology are retained in the supplement.

Flow organisation and variation. Ordered H3 resolution-6 visits represent paths while preserving repeated traversals and observation gaps. For the first-week diagnostic, recurring patterns are fitted using earlier source days; the last eligible source day is reserved for a descriptive assessment and excluded from reference construction. The baseline retains up to eight patterns per airport, phase and horizon, requiring at least 30 preliminary members on three reference-sample days. Assignment uses exact-cell sequence distance and a separation margin from the next-best template; unmatched, ambiguous and unretained-pattern cases remain outside the assigned sample. Each retained pattern has a representative path (medoid) from the reference sample. For the descriptors, a separate ordered, neighbourhood-aware alignment recognises identical or adjacent cells without removing additional traversals. This provides the basis for the complementary measures in Table 1.

Table 1. Complementary flow measures. Each measure retains its eligible movements and observation support.

Aspect	Measure	Interpretation
Organisation	Recurring patterns and local branch shares	Use of alternative paths and shared corridors
Adherence	Corridor-aligned share of candidate path	Proximity to a recurring spatial pattern
Repeated path use	Substantial return incidence	Supported revisits to an earlier path region
Section path use	Candidate/reference distance ratio	Traversal relative to the representative section
Vertical structure	Height distributions and level occurrence	Location of vertical variation and recurrent level flight

For candidate i , corridor adherence A_i and section path ratio P_{is} are

$$A_i = \frac{{}_k \ell_{ik} \mathbf{1}\{k \text{ aligned within the corridor}\}}{{}_k \ell_{ik}}, \quad P_{is} = \frac{D_{is}}{D_s^{\text{ref}}}.$$

Here ℓ_{ik} is the accumulated chord-distance contribution of visit k on the selected candidate path; corridor alignment accepts the corresponding reference cell or an immediate neighbour. Unaligned traversals remain in the denominator. D_{is} and D_s^{ref} are distances between corresponding ordered section boundaries, requiring same-or-neighbouring-cell matches. The reference is a representative observed path, not a shortest or unimpeded route. A substantial return requires an exact-cell revisit after at least four distinct interior cells, 60 seconds and 4 NM of supported path; it is a spatial signature rather than a verified holding manoeuvre. Dedicated holding detection uses labelled trajectories and a trained model [19]; the present descriptor does not assign that operational label.

Profiles locate lateral spread, height and level occurrence along the reference. Per-movement summaries precede cross-movement distributions, and unsupported bins remain gaps. Level occurrence is the fraction of assessable visitors contributing part of a qualifying level episode, distinct from the time-based VFE percentage. Local branch shares describe supported passages through two arms and a directed shared corridor, excluding ambiguous both-arm visits. Selection rules, alignment costs, geometric gates and support thresholds are specified in the companion supplement. These descriptions retain separate denominators and observation masks, allowing horizontal and vertical evidence to be considered jointly without assuming identical samples.

Runway operating context. Operating direction describes the orientation of runway use, such as westerly 27 or easterly 09 operations at EGLL (London Heathrow). Exact configuration records the arrival/departure runway combination within that direction, including parallel-runway or subset changes. Direction provides a descriptive grouping; exact configurations remain the units for fitting flow families.

Configurations are derived from 15-minute arrival/departure runway sets and a forward rolling hour with at least three matching slots. Support checks require source evidence, qualified events in both phases and agreement with each event's own slot. Major exact configurations account for at least 3% of supported busy slots in either week. Up to eight supported families are fitted per airport, phase, horizon and exact configuration using earlier days from both weeks; 7 and 14 July form separate assessment samples and are excluded from reference construction. The frozen week-1 transfer test remains separate. These are observed event contexts; continuity of the configuration throughout a path is audited separately.

The reference sample provides observations for constructing spatial references; the assessment days provide descriptive checks rather than independent validation of representativeness. The three-day rule screens recurrence: a meaningful configuration may occur too briefly in two weeks to support a family or weekly profile. Thresholds are retained for consistency, with insufficient support reported explicitly. Timing baselines remain weekly; VFE uses the stated exposure envelopes. Conditional medoids provide fixed local spatial references without adding their movements to the other week's denominator.

3.4 Study data and analytical populations

The demonstration uses OpenSky surveillance data [7] for seven airports in the PBWG study: EGLL (London Heathrow), SBGR (São Paulo Guarulhos), CYYZ (Toronto Pearson), VTBS (Bangkok Suvarnabhumi), WSSS (Singapore Changi), RJAA (Tokyo Narita) and KATL (Atlanta). The reporting period is 1-14 July 2026 UTC: week 1 is 1-7 July and week 2 is 8-14 July. This provides geographically varied operational examples; two weeks are not treated as representative of annual airport performance.

The source inventory covers 98 airport-day inputs (7 airports $\times$ 14 days), with a nominal 205 NM extraction radius and five-second snapshots. The canonical releases retain 53,617 movement records in week 1 and 53,894 in week 2, including boundary events. Analytical results require operational anchors within the respective week. The spatial inventories contain 108,180 and 115,880 classifier phase hypotheses, including nonaccepted candidates; these are different counting units. Preparation lineage and temporal quality checks are documented in the companion note.

Movement retention, trajectory eligibility and flow-pattern assignment describe different stages. Retained movements supply prepared trajectories, including labelled reconstruction where supported. Indicator samples and spatial samples are then selected separately. Timing requires a supported range crossing and operational event; additional-time calculation also requires a supported reference group. Joint timing/VFE comparisons further require eligible vertical evidence. Spatial eligibility concerns the represented path within the horizon and does not require a complete ring-to-event transit. Flow descriptors use unambiguous assignments to the retained patterns.

The analyses retain three views: weekly ASMA/DSMA and climb/descent VFE comparisons, first-week diagnostic examples, and two-week configuration-conditioned profiles. Timing support covers 40/100/200 NM; the main indicator overview and joint timing/VFE diagnostics cover 40/100 NM. Spatial and observed-height evidence can extend to 200 NM, where matched level occurrence remains unavailable. Reported retention rates use retained

movement records as the denominator; they do not measure surveillance coverage relative to independently established airport totals. Horizons, reference support and flow assignments can select different movements; their counts are not pooled. Missing observations or unmet measurement requirements yield an unavailable result, rather than zero traffic or zero level flight.

4. Results and discussion

4.1 Movement retention and measurement coverage

Table 2 distinguishes retained movement records from the trajectories supporting transit-time measurement at each horizon. Percentages describe analytical retention relative to those records, rather than surveillance coverage of total airport traffic. EGLL retains high timing support across all three horizons. SBGR shows a greater reduction with increasing range, particularly at 200 NM. These differences concern the support for a complete transit measurement; they do not establish equivalent losses of usable spatial or vertical evidence. Limited low-level surveillance coverage at WSSS prevents arrival transit-time measurement, hence the zero arrival retention rates in Table 2.

Table 2. Retained movement records and timing-measurement retention rates by week. Paired entries are week 1 / week 2; horizon columns give the percentage of retained records supporting transit timing. Retained records include boundary anchors, while timing uses anchors in the reporting week. These shares are not airport traffic coverage rates.

Airport	Phase	Retained	40 NM (%)	100 NM (%)	200 NM (%)
EGLL	ARR	4,692 / 4,672	96.4 / 96.0	95.1 / 95.2	88.4 / 87.3
EGLL	DEP	4,678 / 4,765	97.2 / 92.9	95.4 / 91.1	89.0 / 83.5
SBGR	ARR	2,929 / 2,990	74.8 / 74.1	61.7 / 63.8	29.0 / 33.8
SBGR	DEP	2,680 / 2,595	94.0 / 94.4	69.2 / 71.8	19.0 / 21.4
CYYZ	ARR	3,987 / 4,186	95.3 / 95.3	93.1 / 93.1	79.9 / 77.2
CYYZ	DEP	3,991 / 4,111	89.6 / 88.8	86.8 / 86.2	74.6 / 73.0
VTBS	ARR	3,194 / 3,412	75.7 / 38.7	59.4 / 29.0	11.4 / 0.6
VTBS	DEP	3,221 / 3,445	96.2 / 86.6	59.7 / 28.3	12.7 / 0.5
WSSS	ARR	2,727 / 1,448	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
WSSS	DEP	1,546 / 1,410	23.6 / 22.9	20.0 / 17.7	4.1 / 4.3
RJAA	ARR	2,341 / 2,262	98.1 / 97.9	93.7 / 92.6	52.9 / 58.0
RJAA	DEP	2,347 / 2,276	98.5 / 97.5	95.5 / 93.7	73.8 / 67.6
KATL	ARR	7,624 / 8,135	96.7 / 96.0	93.5 / 89.9	70.5 / 66.7
KATL	DEP	7,660 / 8,187	86.9 / 88.4	83.5 / 80.9	65.8 / 61.4

Joint timing/VFE samples are narrower where reference-group or vertical-profile requirements are not met and remain separate by week. In week 1, WSSS contributes 328 departures to the 40 NM joint comparison. Independently of event timing, its 40 NM VFE envelope retains 25 arrivals and 563 departures. The missing arrival timing therefore does not preclude vertical analysis.

The observations available for transit-time measurement and flow characterisation differ. Table 3 shows how many arrival paths meet the configuration-conditioned analysis requirements and how many can subsequently be assigned to a recurring flow. The two assessment days are reported separately.

Table 3. Configuration-qualified arrival paths and unambiguous flow assignments within 40 NM, by airport. Paired entries are 7 / 14 July. Qualification requires an eligible trajectory and a supported major runway configuration; assignment additionally requires a match to a retained flow family. These configuration-conditioned samples differ from the first-week baseline diagnostic populations. Zero assignments do not imply zero traffic.

Airport	Qualified	Assigned
EGLL	548 / 516	303 / 290
SBGR	225 / 270	105 / 179
CYYZ	109 / 110	0 / 47
VTBS	68 / 96	21 / 23
WSSS	0 / 0	0 / 0
RJAA	118 / 105	53 / 52
KATL	402 / 60	342 / 0

Qualification and assignment are separate analytical steps. Qualified paths can remain unassigned because reference-sample or matching requirements are not met; this is not evidence of missing traffic. WSSS retains spatial observations but lacks the runway-event evidence needed for configuration qualification. The accompanying Technical Note and its supporting material [20] provide detailed measurement requirements, airport/phase/horizon results and complementary descriptor sensitivity analyses.

4.2 Additional ASMA/DSMA time and climb/descent VFE

Figure 2 presents average additional ASMA/DSMA time and average time flown level at 40/100 NM for both weeks. Each indicator retains its own eligible population; the shared milestone identities support joint comparisons where both are available. The departure additional-time application follows the arrival logic and remains an adapted DSMA descriptor trialled within PBWG.

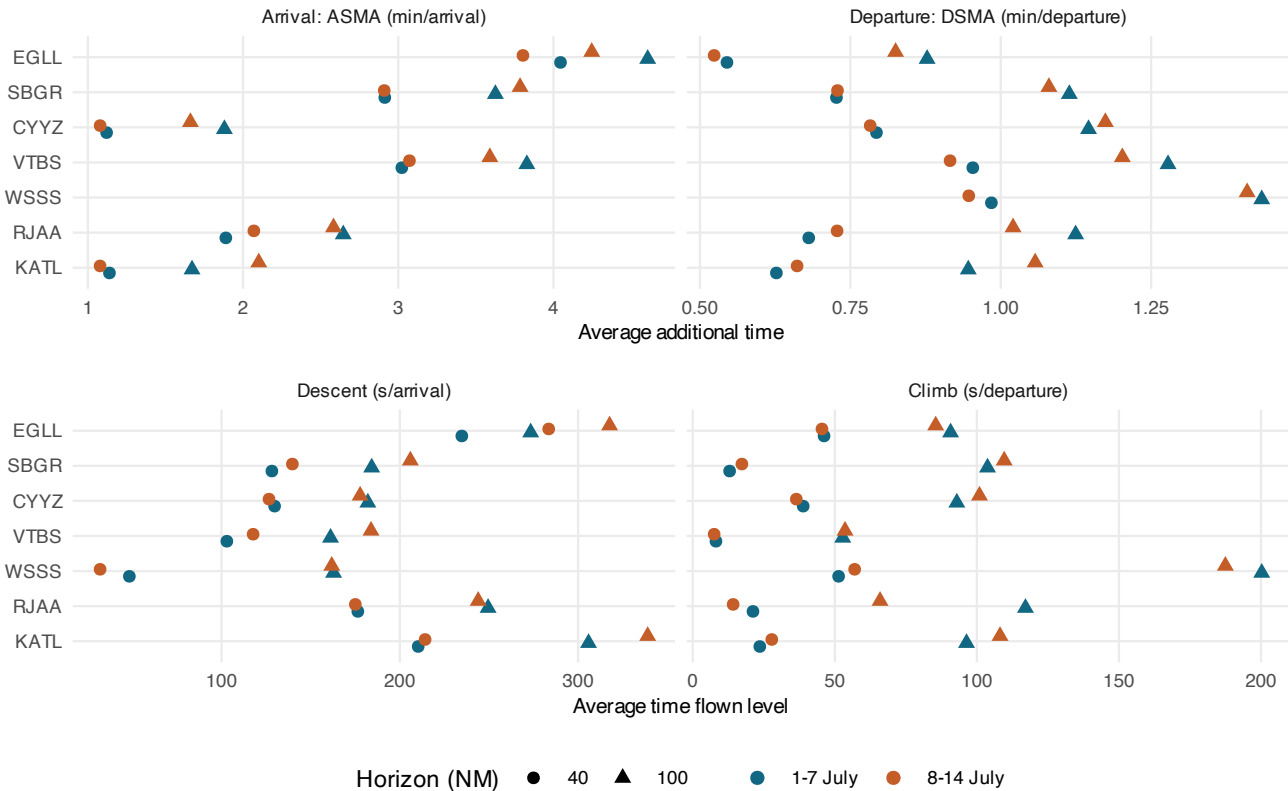


Figure 2. Classical performance measures from milestone evidence, by airport, phase, horizon and week. Top: adapted additional ASMA/DSMA time, in minutes per arrival/departure relative to the respective weekly reference. Bottom: average time flown level during descent/climb, in seconds per arrival/departure within the defined envelope. Colours distinguish weeks and symbols 40/100 NM. Each indicator uses its own eligible sample; exact counts and complementary level-segment/percentage results are in the companion note. Blank arrival timing at WSSS indicates unavailable reference-supported evidence.

Within 40 NM at EGLL, average additional ASMA time is 4.05 and 3.80 min/arrival, while average time flown level during descent is 235 and 283 s/arrival. The average number of qualifying level segments is 2.58 and 2.73. Additional time therefore decreases slightly while the observed level-flight duration and frequency increase. At SBGR, average descent time flown level is 128 and 140 s/arrival. These describe supported weekly samples and separate timing baselines, rather than isolating a causal operational effect.

The phase comparison also remains explicit. EGLL's 40 NM DSMA values are 0.54 and 0.52 min/departure; average time flown level during climb is 46 and 45 s/departure. At 100 NM, average descent time flown level is 273 and 318 s/arrival. Larger envelopes change the trajectory portion observed and the eligible population; these are range-specific comparisons, not within-flight effects. The companion retains matched-sample summaries for examining timing and VFE on identical movements.

The runway-specific decomposition locates the contributions to level-flight exposure (Figure 3). For the selected first-week EGLL cohort, the largest cell is 10-20 NM and 6,000-10,000 ft above ARP, contributing 8.7 percentage points. SBGR's largest contribution is farther out and higher, at 20-40 NM and 10,000-20,000 ft (7.0 points); CYYZ's is at 10-20 NM and 1,800-3,000 ft (5.2 points). The aggregate identifies the extent of observed level flight, while the decomposition identifies where it contributes. These selected runway cohorts differ from the airport-wide matched samples above.

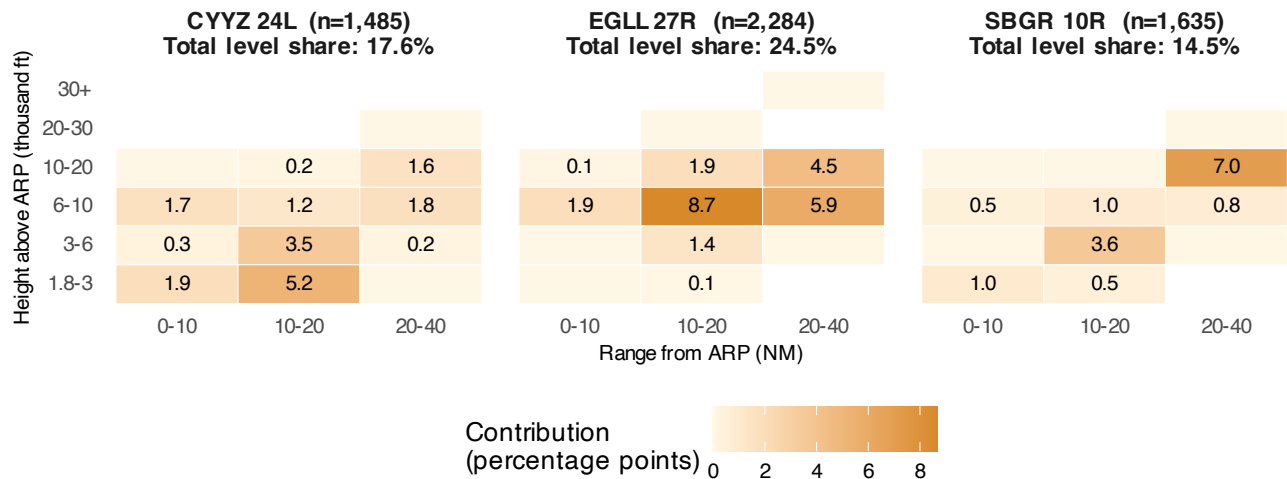


Figure 3. First-week additive range/height explanation of the selected runway cohorts. Panel headings show runway, sample size and total exposure-weighted level share. Each value is a contribution to the cohort’s exposure-weighted level-time percentage; all cells in a panel sum to that percentage. Values below 0.05 percentage points are unlabelled. White cells have no allocated exposure; coloured cells with zero level contribution retain observed exposure.

4.3 Flow organisation and variation within flows

Additional time and time flown level describe complementary aspects of observed performance. Flow characterisation adds information on the paths used, repeated traversal and the location of vertical variation. The following first-week examples illustrate what this spatial context contributes to an operational comparison. Figure 4 locates selected reference-sample patterns and their level-flight regions; the assessment samples provide separate descriptive checks.

Similar corridor adherence, different path use at EGLL and SBGR.

The assessment arrival samples within 40 NM contain 288 EGLL and 164 SBGR assignments. Median one-cell corridor adherence is **95.5% and 95.1%**: both samples follow their respective corridors consistently at this scale. Their use of those corridors differs. Substantial returns occur in **157 of 288 EGLL cases (54.5%)**, compared with **0 of 164 SBGR cases**. All detected assessment EGLL returns have observed trajectory evidence and pass the stricter 90-second/6-NM screen. Reference-sample incidence is 31.5% and 0.2%, respectively. These first-week pattern-assigned samples demonstrate why corridor adherence and repeated path use are separate descriptors; they do not establish airport-wide holding rates.

Spatial patterns qualify a timing comparison at KATL.

The KATL F03/F06 maps locate contrasting approach paths and level-flight regions. In the matched reference samples, median additional times relative to the respective timing references are **0.71 and 0.74 min**, while mean per-movement percentages of time flown level are **40.9% and 20.0%** (441 and 380 movements). Selection deliberately sought this reference-sample contrast; these medians and percentages are matched diagnostic summaries, distinct from the airport-level averages in Section 4.2.

On the assessment day, the vertical difference persists: **40.9% versus 22.7%**, with **206 and 123 movements**. Median additional times are 0.97 and 1.31 min. The timing values do not remain identical; the spatial grouping nevertheless reveals a vertical distinction that the similar reference-sample timing values conceal.

4.4 Operational-context characterisation: a worked synthesis

The framework links four questions: which operating arrangement is observed, which recurring flows serve it, how aircraft use those flows, and what timing and vertical performance accompany them. Shared movement identities connect these descriptions, while each measure retains its observation and sample requirements. Table 4 brings these components together; Figure 5 illustrates comparison along fixed local flow references.

Figure 5 links selected ARR40 reference paths to height distributions and level occurrence. References are selected by membership and day coverage, preferably in both weeks; outcomes do not enter selection. Weeks and reference/assessment samples remain separate. Profile bins require 30 assessable candidates; weekly reference summaries additionally require three source days. Other bins remain gaps. The Technical Note provides the full profiles and supporting counts.

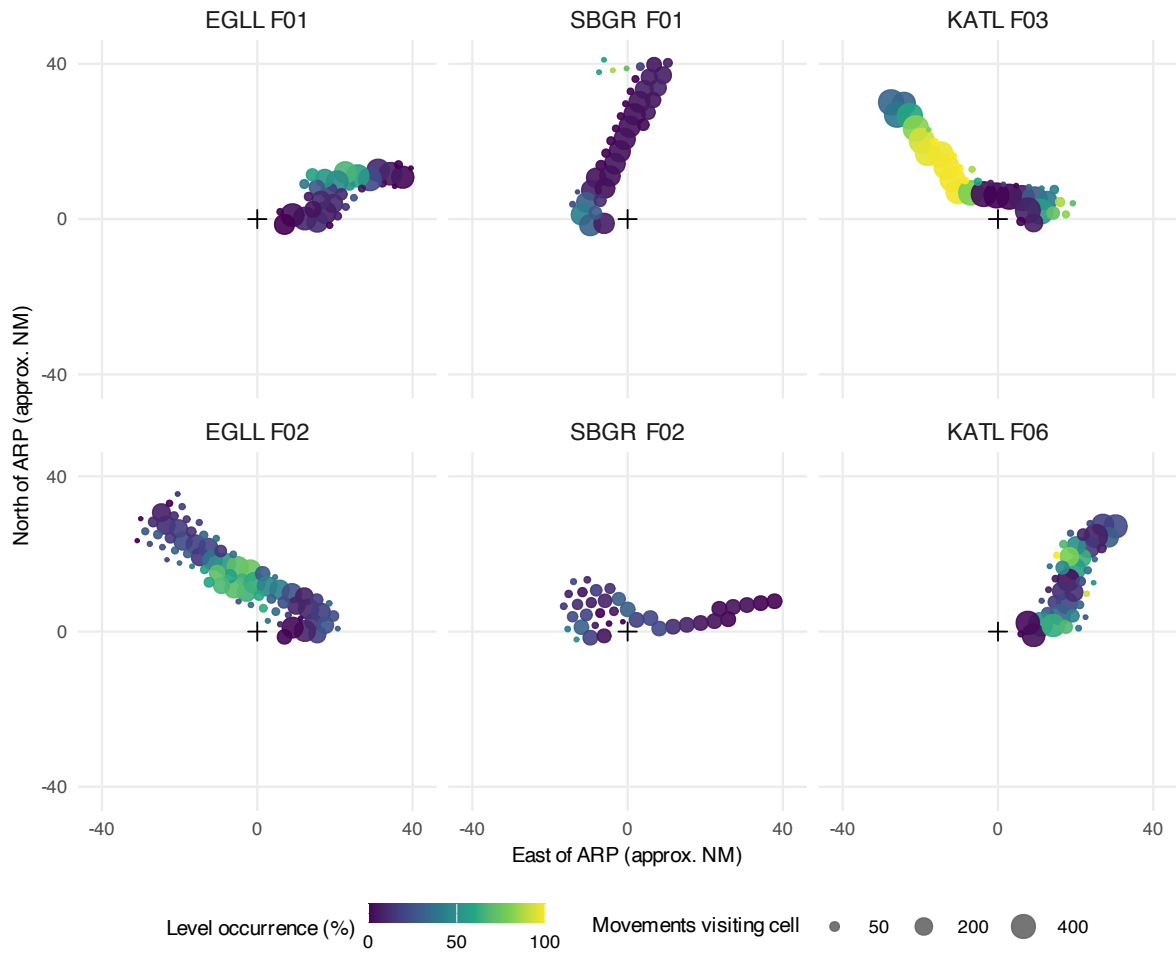


Figure 4. First-week arrival-pattern examples within 40 NM. EGLL/SBGR show two prominent reference-sample patterns; KATL shows the F03/F06 pair selected for the timing/VFE contrast. Colour gives the percentage of assessable movements visiting a cell that contribute qualifying level flight there; point area gives visiting movements (at least five per displayed cell). The ARP is marked +; all panels use the same distance scale. These spatial illustrations do not themselves establish repeated traversal. Pattern identifiers are local to the first-week analysis, distinct from the configuration-conditioned catalogue in Section 4.5.

Table 4. Operational-context synthesis with explicit weekly and conditional scope.

Dimension	Evidence in the demonstration	Interpretation
Observed movements and measurement coverage	Separate weekly retained populations; timing, reference and VFE gates	Establish the contributing movements before comparison.
Timing and vertical outcomes	EGLL ARR40 average additional time decreases slightly while time flown level increases	Read additional time and climb/descent VFE with their own support and measurement envelopes.
Flow organisation	Conditional families represent EGLL easterly operations poorly represented by the first-week bank	Distinguish a new arrangement from missing organisation.
Variation within flows	EGLL/SBGR conditional return contrast; weekly height and level profiles	Examine how a recurring path is used and where variation occurs.
Operational interpretation	Exact configurations, source weeks, conditional assignments and observation masks	Retain scope; seek operational explanations with local expertise.

The selected SBGR and KATL references meet the reference-sample requirements in both weeks. KATL’s earlier reference region (bins 3–5) has reference-sample level occurrence of 98–100% in week 1 and 97–100% in week 2, whereas SBGR’s main reference-sample peak occurs later (bin 8: 46.3% and 52.3%). These are conditional visitor shares, with separate observation envelopes. EGLL’s selected westerly reference has six reference-construction days in week 1 and two in week 2. This limited recurrence coincides with the direction change described below; withholding the weekly summary does not establish instability of the flow. The comparison retains a common geographic reference while allowing weekly support and profiles to differ. Missing curves can reflect configuration, assignment, observation or sample-size limits; they do not establish an absence of organised flow.

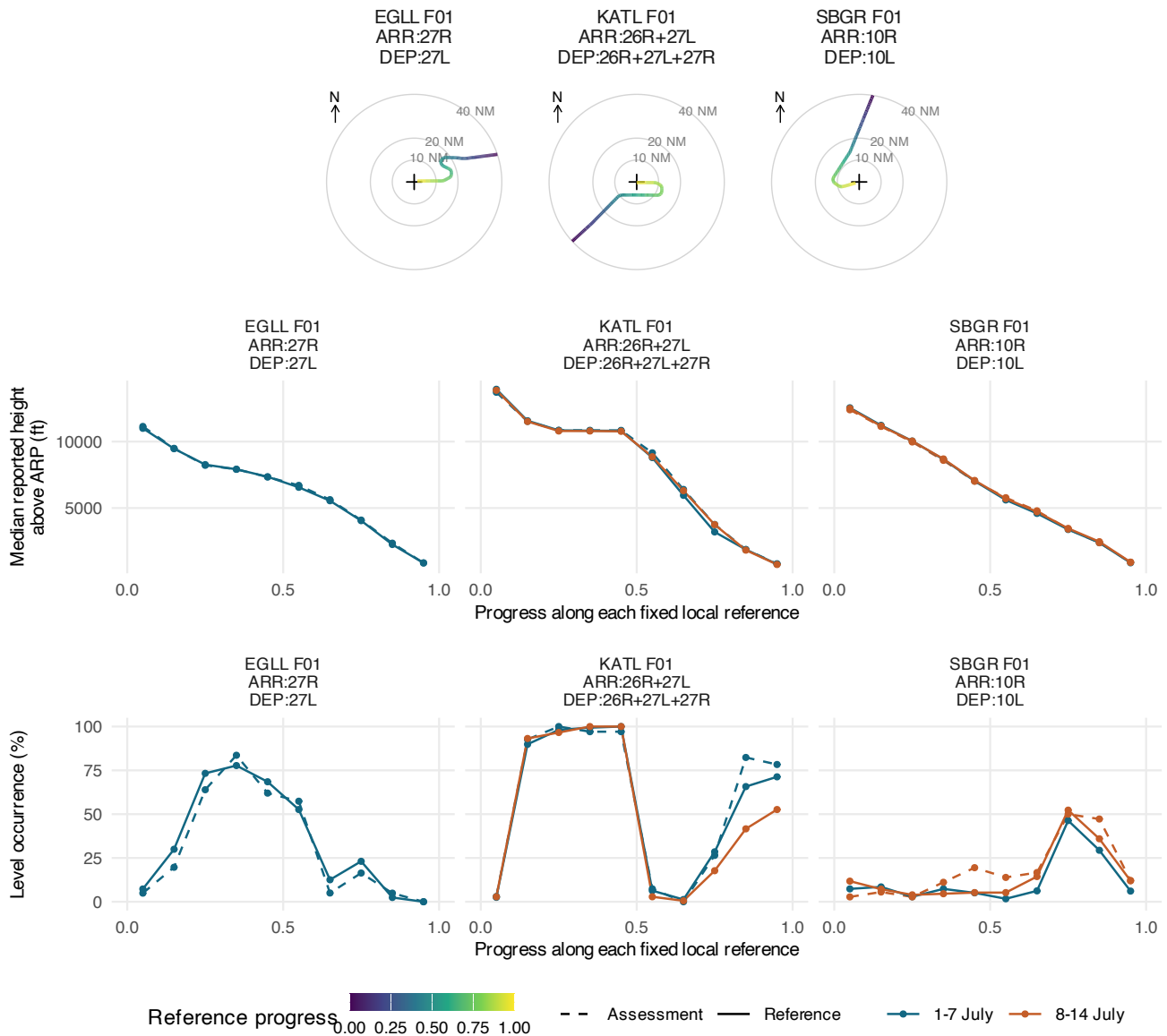


Figure 5. Configuration-conditioned ARR40 profiles at EGLL, SBGR and KATL. Top: retained coordinate geometry of the fixed local reference movement, with ARP (+) and 10/20/40 NM rings; dashed portions denote reconstruction. Middle: median across per-movement observed-height medians above ARP. Bottom: fraction of assessable visitors contributing a qualifying level episode. Colours distinguish weeks; solid/dashed lines distinguish reference/assessment samples. References are selected using reference-sample counts and day coverage; configuration and flow identifiers are local. Missing bins fail the stated sample/day support gates. Height and level profiles have distinct exposure envelopes and denominators.

4.5 Runway configuration changes the flow context

Runway configuration provides an observed operating context for interpreting recurring flow families and their path use. The two-week comparison distinguishes changes in that arrangement from variation within a recurring corridor.

A change in operating direction reveals different flow families at EGLL. Supported busy-period configurations are westerly (27) on 1–9 July and easterly (09) on 10–14 July. Week 1 is entirely westerly within this supported sample; week 2 comprises 29.6% westerly and 70.4% easterly slots. Observation gaps prevent identifying the exact changeover time. Within easterly operations, the exact ARR09L/DEP09R configuration accounts for 62.5% of all supported busy slots in week 2. The configuration-conditioned maps show recurring routes for both directions (Figure 6). On 14 July, **290 of 516 configuration-qualified ARR40 cases** are assigned to the new conditional families, compared with **1 case** under the frozen week-1 bank on that identical cohort. The easterly traffic therefore exhibits organised flows that the reference bank represents poorly. The new catalogue conditions on configuration and adds reference-construction days; the assignment contrast demonstrates additional representation rather than a controlled accuracy improvement.

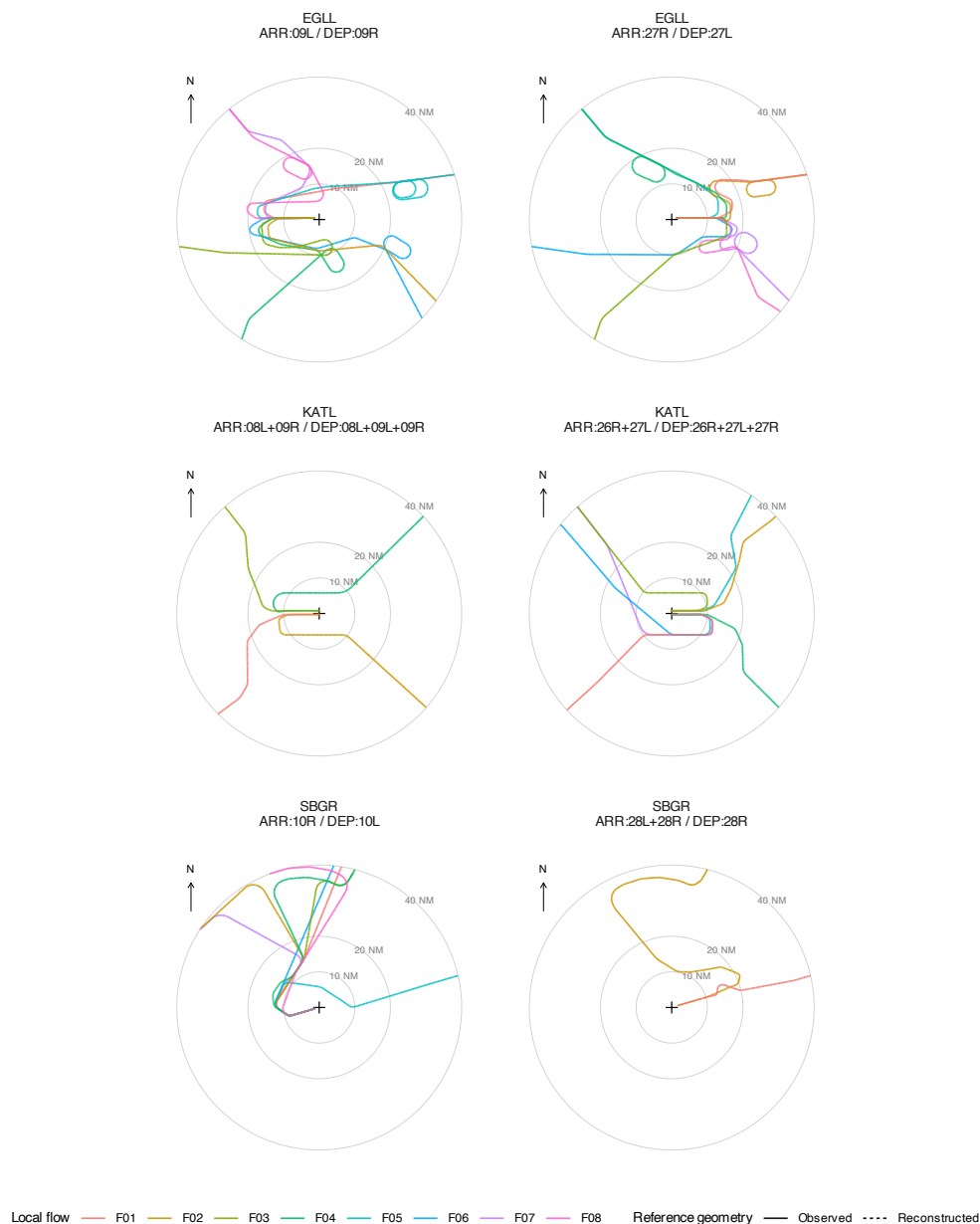


Figure 6. Configuration-conditioned ARR40 reference paths at EGLL, SBGR and KATL. Columns group arrival ends 01–18 and 19–36 for display; each panel uses the most populated supported exact configuration in that direction family. Paths follow retained coordinate intervals of the same medoid movements; gaps remain unconnected, with observed/reconstructed portions shown solid/dashed. Rings mark 10, 20 and 40 NM from the ARP (+); arrows indicate north. Local labels and colours do not identify corresponding routes across panels.

Recurring corridors support different path use at EGLL and SBGR. SBGR’s main ARR10R/DEP10L configuration accounts for 71.4% and 72.3% of supported busy slots in the two weeks. Opposite-direction configurations also have recurring families, so a pooled map combines distinct operating situations. The conditional references describe recurring arrival structures within the dominant configuration. On the assessment day, 14 July, substantial spatial returns remain frequent in EGLL’s main easterly configuration and absent in SBGR’s main configuration (Table 5). This provides a second conditional example of the earlier path-use contrast: configuration identifies the arrangement, while repeated traversal remains a separate characteristic of its use.

Table 5. Configuration-conditioned, pattern-assigned arrivals on the assessment day, 14 July. Returns are substantial spatial excursions under the stated sequence rules, not certified holding manoeuvres or airport-wide holding rates.

Airport and observed configuration	Assigned ARR40 cases	Cases with substantial returns
EGLL: ARR09L / DEP09R	279	98 (35.1%)
SBGR: ARR10R / DEP10L	179	0 (0%)

Direction changes and runway-subset changes add different context at KATL. The principal west-direction configuration, ARR26R+27L/DEP26R+27L+27R, accounts for 66.7% and 69.7% of supported busy slots. The variant adding arrival runway 28 also recurs, at 8.5% and 11.1%. Opposite-direction reference paths show a different collection of approach corridors, whereas a subset variant retains the broader direction. The exact labels let us examine these distinct changes in organisation without treating every small variant as a fundamentally different operating concept. This configuration layer complements the weekly timing/VFE comparison and conditional profiles; the earlier selected KATL timing contrast remains a first-week diagnostic.

Together, these contrasts make operational comparisons more specific: **which arrangement is in use, which recurring flows serve it, and how aircraft use those flows.** The companion Technical Note retains all seven airports, both phases and 40/100/200 NM, including fringe and unsupported configurations, observed-only comparisons and whole-path context checks. The direction/configuration distinction applies across airports, although a two-week sample need not contain a change in direction; stable orientation can still accommodate runway-subset changes. Small exact subsets may lack reference support. WSSS retains spatial evidence while its runway-event support cannot establish this layer. These limits qualify the relevant comparison without implying an absence of organised traffic.

5. Conclusions

The seven-airport PBWG demonstration for 1–14 July 2026 shows how OpenSky observations can link performance measurement to a common characterisation of operational context. It contributes a reusable movement evidence base, complementary ordered-path descriptors and a framework connecting the operating arrangement, flow organisation, path use and associated performance. Shared movement identities retain the observations available for each component.

Adapted additional ASMA/DSMA time and climb/descent VFE describe complementary aspects of operations. At EGLL within 40 NM, average additional ASMA time decreases from 4.05 to 3.80 min/arrival between the two weeks, while average descent time flown level increases from 235 to 283 s/arrival, accompanied by more qualifying level segments. Separate eligible populations and weekly timing references qualify this comparison; path use and the location of level flight provide additional context.

The first-week diagnostics show what the flow descriptors add. Similar corridor adherence at EGLL and SBGR accompanies different repeated path use. At KATL, spatial grouping distinguishes vertical profiles whose reference-sample timing values are similar; the vertical contrast persists in the separate assessment sample. In the configuration-conditioned analysis on 14 July, substantial returns occur in 98 of 279 assigned arrivals in the selected EGLL configuration and none of 179 in the selected SBGR configuration. Ordered spatial evidence preserves these additional traversals; mapped level-flight regions locate vertical variation. Return incidence describes selected flows, rather than airport-wide holding rates.

The worked synthesis makes the operating arrangement explicit before interpreting flow variation. Heathrow's observed 27-to-09 change introduces easterly flows poorly represented by the first-week reference bank. Configuration-conditioned references provide additional representation, although the changed reference-construction sample prevents treating this as a controlled accuracy improvement. KATL illustrates changes in runway subsets within a broader operating direction. Weekly profiles locate level-flight regions along fixed local references. Together, these examples distinguish changes in the operating arrangement from variation within recurring flows and relate both to performance indicators.

Two weeks and the three-day recurrence screen do not establish representativeness. WSSS illustrates that usable vertical and spatial observations can remain available where low-level event timing or runway configuration cannot be established. Measurement retention and flow assignment therefore require separate interpretation; unavailable results do not imply zero traffic or zero level flight. Uneven coverage, sample composition, adapted references and spatial/height approximations constrain comparisons; unknown movement completeness precludes total-throughput or capacity conclusions. Numerical checks establish internal consistency and traceability; operational expertise and further evidence are required to explain differences or attribute causes.

Further work should extend the periods and operating conditions, examine threshold sensitivity and review the contextual dimensions with PBWG participants. Compatible operational milestones could strengthen timing and runway-context evidence where surveillance observations are insufficient. Broader application could inform common reporting conventions and GANP performance-management work, including GANP-PEG. The practical contribution is a common way to organise observed performance and operational context, making the evidence and the questions for operational review explicit.

Acknowledgement

We acknowledge the OpenSky Network and its data contributors. We also thank the operational experts participating in PBWG for sharing insights into the specific operational characteristics of the study aerodromes.

Author contributions

Jean Benevides and Rainer Koelle contributed to the conceptual design, characterisation framework and writing. Quinten Goens and Enrico Spinielli contributed to data preparation, the analytical pipeline and writing.

Funding statement

No specific funding was received for this work. The contribution comprises exploratory analysis and data preparation undertaken within the framework of PBWG. The authors declare no conflict of interest.

Open data statement

The data and code supporting this study are organised as follows:

- **Source observations:** OpenSky Network surveillance data [7] for the airports and period described in Section 3. The [research intake documentation](#) links the extraction provenance and describes authenticated source retrieval. Raw surveillance extracts are not redistributed.
- **Processed datasets:** milestone and ordered H3 evidence for 1–14 July 2026, together with frozen derived results, are being deposited in Zenodo under reserved DOI [10.5281/zenodo.23669884](https://doi.org/10.5281/zenodo.23669884). The two-week timing/VFE and profile extension is maintained separately for a subsequent dataset version. Archive inventories, provenance and checksums identify the inputs and their analytical scope.
- **Analysis and publication resources:** the [research repository](#) contains preparation and analytical workflows; the [publication repository](#) contains manuscript sources, result tables, figure producers and companion material. Project-authored code is licensed under MIT; the manuscript and project-authored data are licensed under CC BY 4.0.

Reproducibility statement

The workflow follows Figure 1: preparation produces linked movement, milestone and ordered spatial evidence; analytical scripts derive indicators and flow descriptors; publication scripts generate figures, tables and the manuscript. The repositories' [reproduction guides](#) document the required inputs, parameters, software and commands. The publication repository supports rebuilding the paper from its result tables; analytical recomputation additionally requires the corresponding processed datasets and research-code versions.

Use of Machine-Learning Models

Machine-learning models assisted with data-preparation code and checking of the analytical pipeline and sensitivity tests. A final manuscript and reference check complemented the work. The authors reviewed the outputs, verified the references, and retain responsibility for the methods, interpretation and final manuscript.

References

- [1] International Civil Aviation Organization. *Global Air Navigation Plan: Performance Management*. URL: <https://ganportal.icao.int/PerformanceManagement> (visited on 10/26/2026).
- [2] Federal Aviation Administration and EUROCONTROL. *Comparison of Air Traffic Management-Related Operational and Economic Performance: U.S.–Europe*. Tech. rep. FAA and EUROCONTROL, Jan. 2024. URL: <https://www.eurocontrol.int/sites/default/files/2024-01/eurocontrol-comparison-atm-related-performance-us-europe.pdf> (visited on 10/26/2026).
- [3] DECEA Performance Section and EUROCONTROL Performance Review Unit. *Operational Comparison of ANS Performance in Brazil and Europe*. Tech. rep. DECEA and EUROCONTROL, Apr. 2024. URL: <https://ansperformance.eu/global/brazil/bra-eur-2023.pdf> (visited on 10/26/2026).
- [4] Civil Aviation Administration of China Operation Supervisory Center, Civil Aviation University of China, Aviation Data Communication Corp., and EUROCONTROL Performance Review Unit. *China/Europe: Comparison of Operational ANS Performance*. Tech. rep. CAAC and EUROCONTROL, 2021. URL: <https://www.caac.gov.cn/XWZX/HYDT/202112/P020211224325749901413.pdf> (visited on 10/26/2026).
- [5] EUROCONTROL. *Additional ASMA Time Performance Indicator Document*. Tech. rep. Edition 01.00. EUROCONTROL, Mar. 16, 2023. URL: https://ansperformance.eu/library/AddASMA_indicator_documentation_mar23.pdf (visited on 10/20/2026).

- [6] EUROCONTROL. *Vertical Flight Efficiency during Climb and Descent*. URL: <https://ansperformance.eu/methodology/cd-vertical-flight-efficiency-pi/> (visited on 10/20/2026). 482
- [7] Matthias Schäfer, Martin Strohmeier, Vincent Lenders, Ivan Martinovic, and Matthias Wilhelm. “Bringing Up OpenSky: A Large-scale ADS-B Sensor Network for Research”. In: *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*. 2014, pp. 83–94. DOI: [10.1109/IPSIN.2014.6846743](https://doi.org/10.1109/IPSIN.2014.6846743). 483
- [8] Henrik Hardell, Anastasia Lemetti, Tatiana Polishchuk, Lucie Smetanová, and Karim Zeghal. “Towards a Comprehensive Characterization of the Arrival Operations in the Terminal Area”. In: *SESAR Innovation Days*. 2021. URL: https://www.sesarju.eu/sites/default/files/documents/sid/2021/papers/SIDs_2021_paper_101.pdf (visited on 10/26/2026). 484
- [9] International Civil Aviation Organization. *Global Air Navigation Plan*. URL: <https://ganportal.icao.int/> (visited on 10/26/2026). 485
- [10] International Civil Aviation Organization. *Adapting GANP KPI*. URL: <https://ganportal.icao.int/PerformanceManagement/DataAnalytics/AdaptingGanpKpi> (visited on 10/26/2026). 486
- [11] International Civil Aviation Organization. *GANP Performance Management: KPI Overview*. URL: <https://ganportal.icao.int/PerformanceManagement/DataAnalytics> (visited on 10/26/2026). 487
- [12] Xavier Olive. “traffic, a toolbox for processing and analysing air traffic data”. In: *Journal of Open Source Software* 4.39 (2019), p. 1518. DOI: [10.21105/joss.01518](https://doi.org/10.21105/joss.01518). 488
- [13] Muhammad Fazlur Rahman, Joost Ellerbroek, and Jacco Hoekstra. “Modelling ADS-B Reception Probability Using OpenSky Data”. In: *Journal of Open Aviation Science* 2.2 (2025). DOI: [10.59490/joas.2024.7895](https://doi.org/10.59490/joas.2024.7895). 489
- [14] Michael Schultz, Judith Rosenow, and Xavier Olive. “Data-driven airport management enabled by operational milestones derived from ADS-B messages”. In: *Journal of Air Transport Management* 99 (2022), p. 102164. DOI: [10.1016/j.jairtraman.2021.102164](https://doi.org/10.1016/j.jairtraman.2021.102164). 490
- [15] Quinten Goens, Rainer Koelle, Enrico Spinielli, Tatjana Bolić, Andrew Cook, and Martin Strohmeier. “The Open Performance Data Initiative: A Foundation Supporting the European Open Science Alliance for ATM Research”. In: *SESAR Innovation Days*. 2024. URL: https://sesar.eu/sites/default/files/documents/sid/2024/papers/SIDs_2024_paper_100%20final.pdf (visited on 10/26/2026). 491
- [16] Luis Basora, Jérôme Morio, and Corentin Mailhot. “A Trajectory Clustering Framework to Analyse Air Traffic Flows”. In: *SESAR Innovation Days*. 2017. URL: https://www.sesarju.eu/sites/default/files/documents/sid/2017/SIDs_2017_paper_45.pdf (visited on 10/26/2026). 492
- [17] Samantha J. Corrado, Tejas G. Puranik, Oliva J. Pinon, and Dimitri N. Mavris. “Trajectory Clustering within the Terminal Airspace Utilizing a Weighted Distance Function”. In: *Proceedings* 59.1 (2020), p. 7. DOI: [10.3390/proceedings2020059007](https://doi.org/10.3390/proceedings2020059007). 493
- [18] H3 contributors. *H3: Hexagonal Hierarchical Geospatial Indexing System*. Version 4.x documentation; tables of cell statistics across resolutions. URL: <https://h3geo.org/docs/core-library/restable/> (visited on 10/26/2026). 494
- [19] Xavier Olive, Luis Basora, Junzi Sun, and Enrico Spinielli. “Training a Machine Learning Model to Detect Holding Patterns in Aircraft Trajectories”. In: *Journal of Open Aviation Science* 2.2 (2025). DOI: [10.59490/joas.2024.7943](https://doi.org/10.59490/joas.2024.7943). 495
- [20] *Technical Note: Organisation and variability along the flow*. Companion working document, with linked methods and descriptor analyses. Oct. 20, 2026. URL: <https://github.com/euctrl-pru/paper-2026-OSN-PBWG-flow-characterisation/blob/main/docs/flow-profiles-v2.qmd> (visited on 10/26/2026). 496