

Shadow-Mode Meteorological Concordance Between Remote and On-Site Observation at a Brazilian Regional Aerodrome

Guilherme Vieira Monteiro Amadeu^{1*} 

1. AMD Administração, Participações e Concessões – Diretoria Técnica – Juiz de Fora (MG), Brazil.

*Corresponding author: guilherme.vm.amadeu@gmail.com

ABSTRACT

This study assessed whether a remote aerodrome flight information service could reproduce on-site aeronautical weather observations with operationally acceptable concordance at a Brazilian regional aerodrome. A shadow-mode campaign ran from March to November 2025 at a non-towered aerodrome in southeastern Brazil, compiling 749 paired routine and special weather reports produced independently by on-site and remote operators over 118 operational days. Six meteorological elements were evaluated using a weighted concordance index, a pre-defined divergence framework, Cohen's kappa, intraclass correlation coefficients, and a day-clustered bootstrap. The weighted concordance index was 97.41% (95% bootstrap confidence interval 96.8–97.9%), above the 95% benchmark used in European shadow-mode validation. Agreement was almost perfect for clear-weather coding and moderate for cumulonimbus/towering-cumulus detection, intraclass correlation coefficients were at least 0.975 for continuous variables, and post-commissioning concordance reached 97.57% across 555 messages with no convective-cloud divergences after April 2025. These results indicate high concordance between remote and on-site observation at a daytime regional aerodrome under visual-flight conditions and support further regulatory evaluation of remote aerodrome service in Brazil, while broader generalization still requires multi-site, clustered, and operator-stratified validation.

Keywords: Remote aerodrome flight information service; Aerodrome weather observation; Shadow mode validation; Non-towered aerodromes; Remote tower.

INTRODUCTION

Extending air traffic and flight information services to low-density regional aerodromes remains a persistent operational and economic challenge. European remote tower programs showed that camera-based surveillance and networked operational systems can support remote service provision when local validation demonstrates adequate technical and human performance, as described in the European Union Aviation Safety Agency (EASA) Guidance Material on Remote Aerodrome Air Traffic Services, Issue 2 (2019), and in SESAR validation reports (SESAR Joint Undertaking 2016; 2019). Evidence from tropical and developing operational environments, however, remains limited.

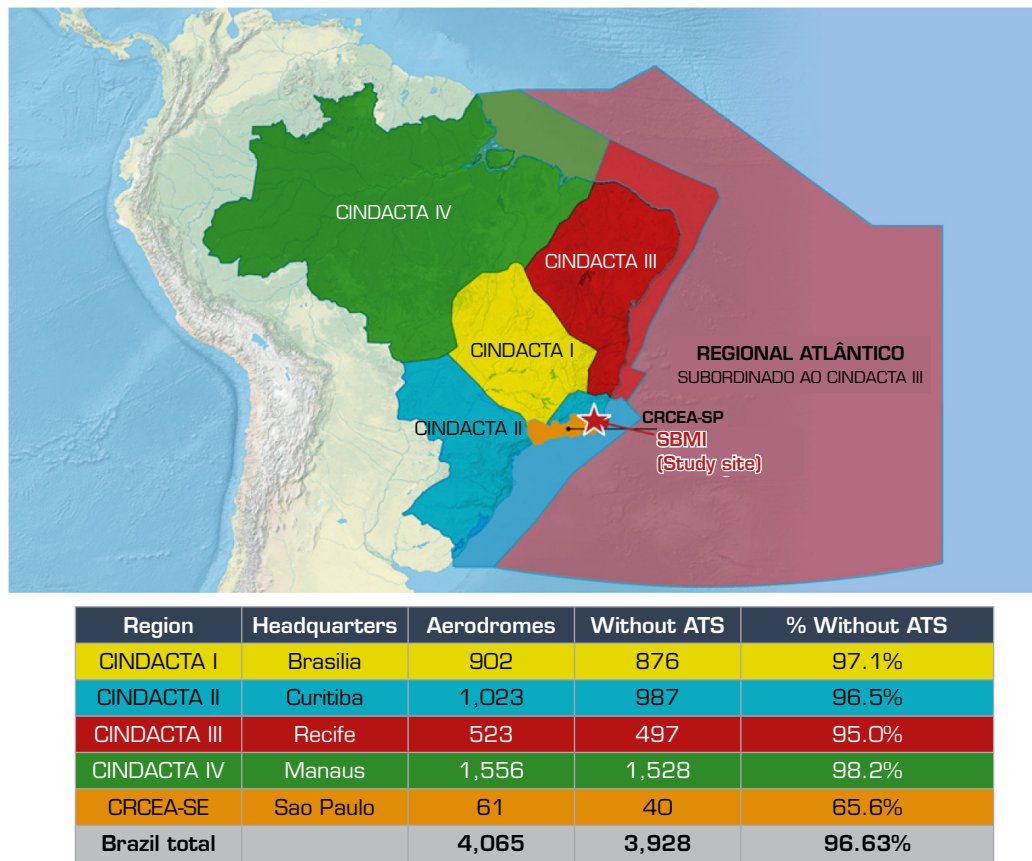
Brazil illustrates that gap (Fig. 1). National performance reporting indicates that only a small fraction of Brazilian aerodromes are served by air traffic services, leaving a large service deficit across public and private infrastructure (DECEA 2024). Remote aerodrome flight information service offers a scalable alternative to the conventional model of fully staffed local facilities by centralizing visual surveillance, communications, and meteorological support while preserving local operational reporting.

Received: Apr. 7, 2026 | **Accepted:** July 31, 2026

Peer Review History: Single Blind Peer Review.

Section editor: Dirceu Herdies 





Source: Adapted from DECEA (2024; 2025).

Figure 1. Geographic and operational context of the study site and the Brazilian aerodrome air traffic services (ATS) coverage gap (all registered aerodromes).

The scientific question is narrower than the policy debate. Before remote aerodrome service can inform broader deployment decisions, it must show that remote and on-site operators produce closely aligned aeronautical weather observations under an explicit validation protocol. To the best of my knowledge, no indexed study has reported a Brazilian shadow-mode validation focused on meteorological concordance for remote aerodrome service.

This study reports a nine-month shadow-mode campaign at the SBMI aerodrome (ICAO location indicator; Maricá, Rio de Janeiro, Brazil), comparing 749 paired observations produced independently by on-site and remote aeronautical station operators (Fig. 1). The objectives here were to quantify concordance across six meteorological elements, examine the commissioning-to-post-commissioning transition, and benchmark the results against international remote tower validation.

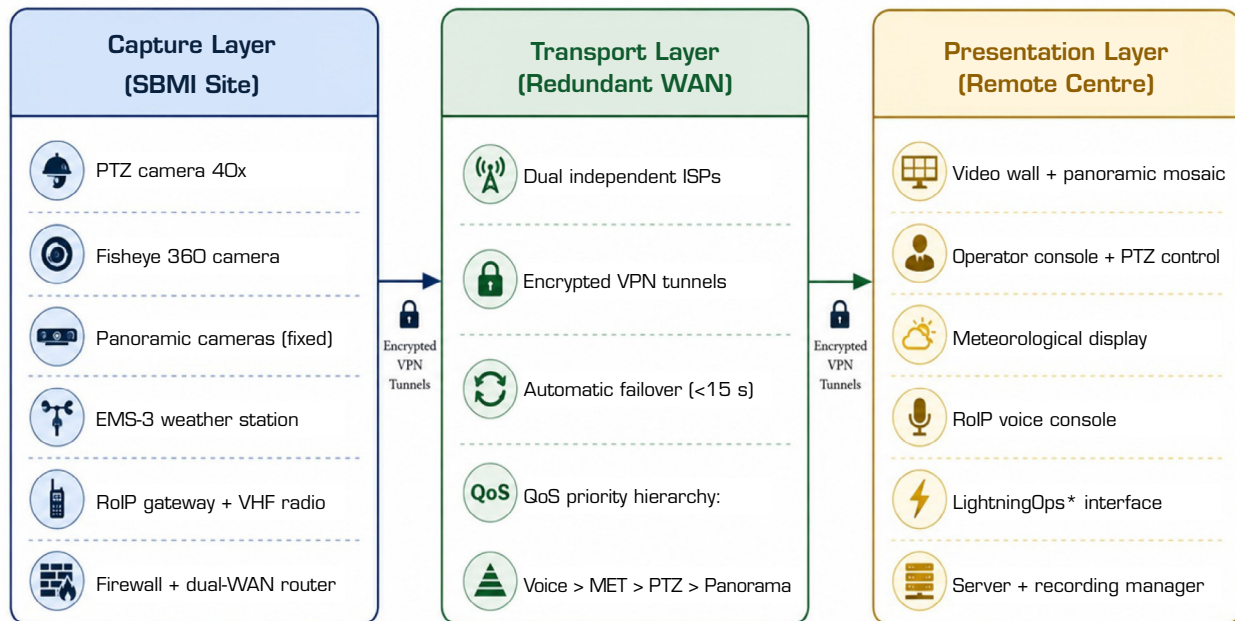
METHOD

Study design and setting

The validation was conducted at SBMI, a regional non-towered aerodrome in Maricá, Rio de Janeiro, Brazil (22°55'05"S, 042°49'44"W; elevation 2 m/7 ft), approximately 60 km east of the city of Rio de Janeiro (Fig. 1). Operations were limited to daytime activity under visual-flight conditions (6 a.m.–6 p.m. local time, UTC-3). The study was performed under the operational framework of ICA 63-10 (DECEA 2023). A memorandum of understanding (MOU No. 033/ICEA/2024) with the Institute of Airspace Control (Instituto de Controle do Espaço Aéreo, ICEA) sets out general lines of cooperation between the parties and did not define or govern the validation protocol reported here.

Remote system and shadow-mode procedure

The remote aerodrome flight information service installation followed a remote-center architecture with three functional layers. The capture layer (Fig. 2, blue) included a pan-tilt-zoom camera, a 360-degree fisheye camera, fixed panoramic cameras, and an EMS-3 automatic weather station. The transport layer (Fig. 2, green) used dual-WAN connectivity with encrypted tunnels and quality-of-service prioritization for voice, telemetry, and video. The presentation layer (Fig. 2, yellow) integrated video, meteorological display, voice communications, and, from November 2025 onward, a network-based lightning-monitoring application.



* Integrated in November 2025. Source: Elaborated by the author.

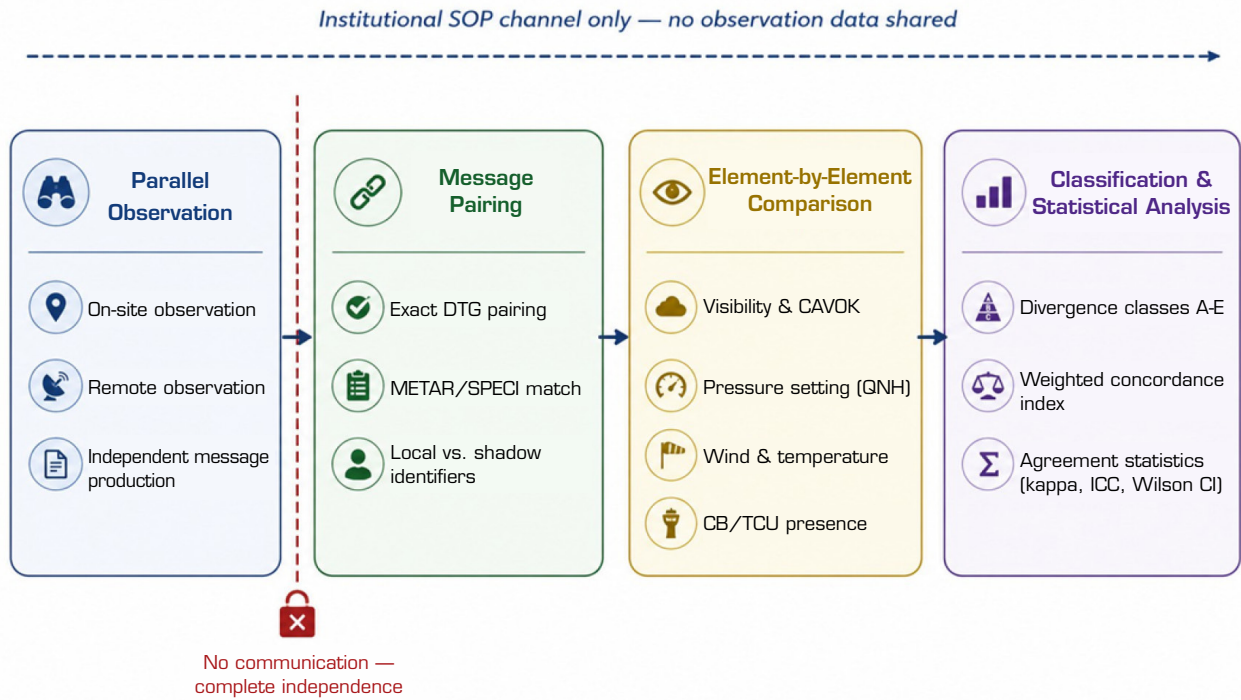
Figure 2. Functional architecture of the remote service installation. Box colors denote the three functional layers: capture (blue), transport (green), and presentation (yellow).

Three remote aeronautical station operators rotated during the campaign. Each held current aeronautical qualification and completed a 40-hour remote-service qualification program before operational participation. During shadow mode, the local operator remained responsible for the official report, while the remote operator produced a simultaneous non-operative shadow report without access to the local observation. Both operators observed the same operational events, used different information channels, and did not exchange observation judgments during routine operation (Fig. 3).

Data preparation and outcome definitions

The final archive contained 749 paired routine (METAR) and special (SPECI) aerodrome weather reports, encoded according to the ICAO METAR/SPECI code format (ICAO 2018), from 118 operational days between 17 March and 28 November 2025. Messages were paired by event and report type within a maximum 10-minute window. Pairs associated with maintenance windows, waivers, communication failures, or corrected/NIL messages were excluded before final tabulation. The paired messages analyzed here are openly available (see the data availability statement).

Monthly sampling is reported in Amadeu (2026). The weighting-scheme sensitivity analysis, the contingency table for the phase comparison, monthly visibility concordance among non-CAVOK (ceiling and visibility OK) pairs, the distribution of divergences by class and element, and descriptive cloud-type and weather-phenomena reporting frequencies are reported in Amadeu (2026). July 2025 contributed 30 paired messages across four operational days, and October 2025 contributed 35 messages across five days, reflecting local operational downtime rather than selective case removal.



Source: Elaborated by the author.

Figure 3. Shadow-mode workflow from paired observation to divergence classification and statistical analysis.

Six meteorological elements were evaluated: visibility, clear-weather coding (CAVOK, ceiling and visibility OK), QNH (altimeter setting reduced to mean sea level), wind direction, temperature, and cumulonimbus/towering-cumulus reporting. Reporting resolution and units were as follows: visibility, in meters, with CAVOK used as the clear-weather code; QNH, in whole hectopascals (hPa); wind direction, in degrees (10° steps); temperature, in whole degrees Celsius (°C); cumulonimbus/towering-cumulus as presence or absence; and wind speed in knots (kt) as a continuous telemetry variable. Visibility concordance treated CAVOK-concordant pairs as visibility-concordant, because CAVOK replaces element-level visibility reporting in the code form. Wind-direction concordance used 678 valid pairs after excluding 71 calm-wind and variable-wind pairs. Wind speed was described as a continuous telemetry variable and assessed through the intraclass correlation coefficient (ICC), not through the weighted concordance table.

The divergence framework classified discrepancies from class A to class E. For visibility, class A denoted identical reporting; and classes B, C, D, and E denoted progressively wider distance bands. For QNH, class A denoted exact agreement of the coded integer hectopascal value, class B denoted a 1-hPa discrepancy, class C denoted 2 hPa, and classes D/E denoted discrepancies of 3 hPa or more or any discrepancy judged capable of crossing an operational threshold. Temperature concordance was defined as a difference of at most 1°C. CAVOK was treated as binary. For cumulonimbus/towering-cumulus reporting, remote non-detection was classified as critical and remote false-positive reporting as significant.

Statistical analysis

The overall concordance index was calculated as a weighted average of the parameter-specific concordance rates, as in Eq. 1:

$$CI_{\text{overall}} = \sum_{i=1}^n w_i CI_i \quad (1)$$

Where: CI_{overall} : the overall weighted concordance index (%); CI_i : the parameter-specific concordance rate (%) of element i ; w_i : the *a priori* weight of element i (dimensionless, with $\sum_{i=1}^n w_i = 1$); $n = 6$: the number of meteorological elements.

Weights were fixed *a priori* as follows: visibility 25%, wind direction 20%, QNH 15%, CAVOK 15%, temperature 15%, and cumulonimbus/towering-cumulus 10%. These weights were fixed before the statistical analysis, reflecting the relative operational relevance of each element for visual-flight-rules service and its reporting base rate, rather than being tuned to the observed data. The lower weight assigned to cumulonimbus/towering-cumulus reflects its low base rate relative to continuously reported elements. Its operational consequence is addressed separately through the divergence framework, which classifies remote non-detection of convective cloud as a critical divergence.

Cohen's kappa was used for CAVOK and cumulonimbus/towering-cumulus reporting (Cohen 1960; Landis and Koch 1977). Wilson's confidence intervals were used for visibility and wind-direction concordance proportions (Wilson 1927). The ICC for QNH, temperature, and wind speed was reported in the two-way, absolute-agreement, average-measures form—ICC(A,k) in the McGraw and Wong notation (Shrout and Fleiss 1979; Koo and Li 2016)—quantifying the agreement between the paired on-site and remote observation streams, with the two observing roles treated as fixed rather than sampled from a broader rater population. The composite concordance index was accompanied by a day-clustered bias-corrected and accelerated bootstrap confidence interval based on 10,000 resamples, with pairs grouped by operational day (118 clusters) to preserve intraday correlation structure. The one-sided Fisher's test direction for the phase comparison was specified before the analysis on the expectation that commissioning-phase divergence would be higher than post-commissioning divergence. A one-sided test was appropriate because the direction of effect was fixed in advance and reflects an expected familiarization trajectory; the two-sided result is also reported to avoid overstating significance.

RESULTS

Table 1 summarizes the main concordance results. The weighted concordance index was 97.41%, above the 95% benchmark. Visibility and CAVOK were the most informative perceptual outcomes, while continuous telemetry variables showed very high agreement ($\text{ICC(A,k)} \geq 0.975$; Table 2).

Table 1. Concordance results by meteorological parameter.

Parameter	Pairs analyzed	Concordant	Divergent	Concordance rate (%)
Visibility	749	727	22	97.06
CAVOK	749	726	23	96.93
QNH	749	700	49	93.46
Wind direction	678	670	8	98.82
Temperature	749	744	5	99.33
CB/TCU	749	743	6	99.20
Overall weighted concordance index	749	-	-	97.41

CAVOK: ceiling and visibility OK; QNH: altimeter setting reduced to mean sea level; CB: cumulonimbus; TCU: towering cumulus.
Source: Elaborated by the author.

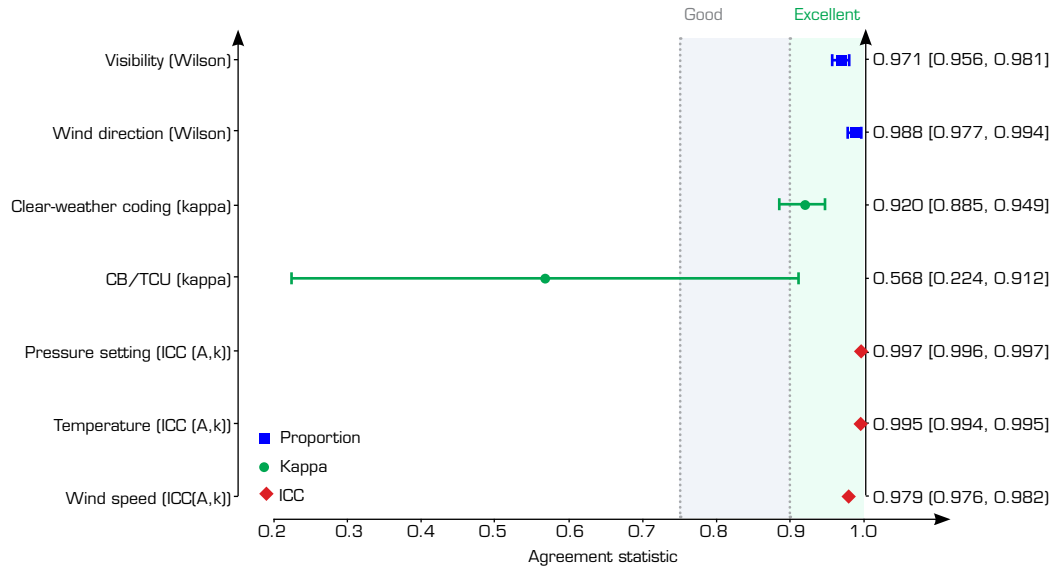
Table 2. Agreement statistics by parameter.

Parameter	Statistic	Value	95%CI	Interpretation
CAVOK	Cohen's kappa	0.920	[0.885–0.949]	Almost perfect
CB/TCU	Cohen's kappa	0.568	[0.224–0.912]	Moderate
QNH	ICC(A,k)	0.997	[0.996–0.997]	Excellent
Temperature	ICC(A,k)	0.995	[0.994–0.995]	Excellent
Wind speed	ICC(A,k)	0.979	[0.976–0.982]	Excellent
Visibility	Wilson interval	97.06%	[95.59–98.05]	-
Wind direction	Wilson interval	98.82%	[97.69–99.40]	-
Overall concordance index	Bias-corrected and accelerated bootstrap interval	97.41%	[96.8–97.9]	Exceeds threshold

CAVOK: ceiling and visibility OK; CB: cumulonimbus; TCU: towering cumulus; QNH: altimeter setting reduced to mean sea level; ICC: intraclass correlation coefficient; 95%CI: 95% confidence interval. Source: Elaborated by the author.



Table 2 reports the complementary agreement statistics, illustrated in Fig. 4. Agreement was almost perfect for CAVOK and moderate for cumulonimbus/towering-cumulus reporting, while ICC(A,k) was at least 0.975 for QNH, temperature, and wind speed. The lower bound of the bias-corrected and accelerated bootstrap interval remained above the 95% benchmark.



CB: cumulonimbus; TCU: towering cumulus; ICC: intraclass correlation coefficient. Source: Elaborated by the author.

Figure 4. Summary display of agreement statistics across meteorological parameters.

The QNH element had the lowest concordance rate in the weighted table, because exact agreement of the coded QNH group was required for class A concordance. The continuous-scale analysis nevertheless remained excellent, indicating that most message-level QNH disagreements were small: out of the 49 QNH divergences, 44 (90%) were 1-hPa class B differences (Amadeu 2026). In operational terms, QNH discrepancies were predominantly minor one-hectopascal differences rather than large pressure disagreements.

Phase comparison showed the expected stabilization after commissioning (Table 3). The commissioning phase comprised 194 messages, whereas the post-commissioning phase comprised 555. Post-commissioning concordance reached 97.57%, against 96.94% during commissioning, and no cumulonimbus/towering-cumulus divergences were observed after April 2025 (Table 3).

Table 3. Commissioning *versus* post-commissioning comparison.

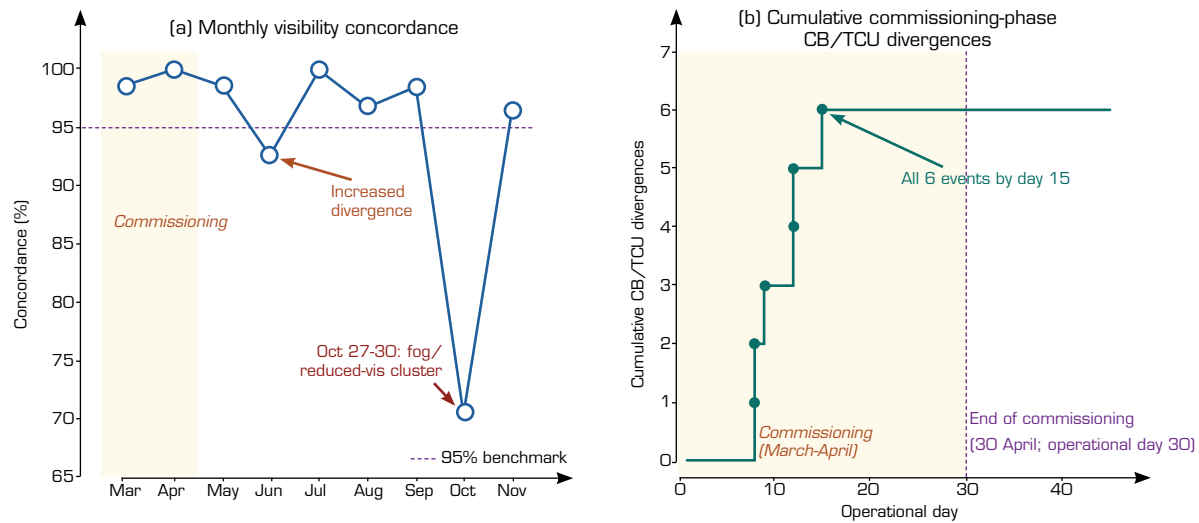
Phase	Period	Messages	Overall concordance [%]	CB/TCU divergences [%]
Commissioning	March–April 2025	194	96.94	6 cases (3.09)
Post-commissioning	May–November 2025	555	97.57	0 cases (0)
Total	March–November 2025	749	97.41	6 cases (0.80)

CB: cumulonimbus; TCU: towering cumulus. Source: Elaborated by the author.

Fisher's exact test for cumulonimbus/towering-cumulus divergences yielded $p = 0.0003$ for the pre-specified one-sided test and $p = 0.0003$ for the two-sided test (Amadeu 2026). All six divergences occurred within the first 18 operational days. The pattern is consistent with early familiarization effects, although the design does not isolate operator learning from concurrent procedural stabilization.

The only notable temporal dip in an otherwise stable visibility series occurred in October 2025 (Fig. 5), when 10 of 34 non-CAVOK pairs showed divergences within a four-day cluster of reduced-visibility and fog events (Amadeu 2026). None of the

divergences during that cluster was associated with an operational consequence. A sensitivity analysis excluding the five October operational days raised the overall weighted concordance index by 0.22 percentage points (to 97.63%), confirming that the main result is not driven by this atypical cluster.



CB: cumulonimbus; TCU: towering cumulus. Source: Elaborated by the author.

Figure 5. Monthly visibility concordance and cumulative commissioning-phase convective-cloud divergences.

Technical performance remained within the applicable limits throughout the validation period (Table 4); the lightning-detection indicator applies from the integration of that application in November 2025. All ten indicators met or exceeded their requirements, most by a wide margin, including glass-to-glass video latency below 400 ms (requirement < 1,000 ms) and communication-link availability of 99.98% (requirement > 99.5%), supporting the technical adequacy of the remote installation underlying the meteorological results.

Table 4. Technical performance indicators.

Indicator	Requirement	Achieved	Reference
Video latency (glass-to-glass)	< 1,000 ms	< 400 ms	EUROCAE ED-240A
Video frame rate	≥ 20 FPS	≥ 30 FPS	EASA GM
VHF one-way voice latency	< 250 ms	< 150 ms	ITU-T G.114
Communication-link availability	> 99.5%	99.98%	EASA GM
Failover time	< 30 s	< 15 s	EASA GM
Lightning detection accuracy	≤ 1 NM	< 1 km	FAA AC 150/5220-16E
Voice jitter	< 50 ms	< 30 ms	ITU-T G.114
Packet loss (voice)	< 3.0%	< 0.1%	VoIP standard
NTP synchronization offset	< 100 ms	< 50 ms	Audit requirement
UPS autonomy / backup continuity	≥ 15 min	25 min	Local acceptance requirement

FPS: frames per second; GM: guidance material; VHF: very high frequency; NM: nautical mile; AC: advisory circular; VoIP: voice over internet protocol; NTP: network time protocol; UPS: uninterruptible power supply. Source: Elaborated by the author.

None of the 113 registered divergences was judged to have produced an operational consequence after review against message logs and local aircraft-movement records.

DISCUSSION

The study supports a restrained conclusion: at a daytime regional aerodrome operating under visual-flight conditions, remote observation achieved high meteorological concordance with conventional on-site observation. The overall index exceeded the 95% benchmark commonly cited in remote tower shadow-mode validation, and the lower confidence bound also remained above that threshold.

The 95% threshold derives from European remote tower validation literature; no equivalent national benchmark is established in Brazil. Because the threshold was adopted before the statistical analysis, and the result exceeded it by a margin of 2.4 percentage points, the conclusion is robust to modest variations in threshold choice. These results are consistent with the broader European remote tower literature, while extending the evidence base to a Brazilian operational setting (Fürstenau 2016; SESAR Joint Undertaking 2016) and with the acceptance logic described in the EASA Guidance Material on Remote Aerodrome Air Traffic Services, Issue 2 (EASA 2019).

The most important interpretive point concerns QNH. Message-level concordance for QNH was lower than for the other elements because exact agreement of the coded integer pressure value was treated as full concordance in the weighted table, whereas one-hectopascal differences were classified as minor divergences. Continuous agreement, by contrast, remained excellent. The combined picture is therefore not one of substantive pressure disagreement, but of a strict message-level rule applied to a variable read at high numerical fidelity.

The six cumulonimbus/towering-cumulus divergences were confined to the commissioning period. That temporal concentration is consistent with early adaptation to camera-based cloud assessment and structured scanning routines. The later integration of network-based lightning monitoring may strengthen future operations, but it cannot explain the elimination of those early divergences, because it was introduced only in November 2025, after the commissioning-phase events had already occurred.

The regulatory relevance of the study is practical rather than dispositive. The findings may inform future regulatory evaluation of remote aerodrome service for regional aerodromes with limited traffic and resource constraints. They do not establish cross-context interchangeability, and they do not by themselves justify generalization to instrument-flight-rules, nighttime, or multi-aerodrome remote-center operations.

Several limitations should remain explicit. This was a single-site study conducted at a subtropical coastal aerodrome under daytime visual-flight conditions. Operator-level heterogeneity was not estimated in the final analysis set; with three rotating operators across 118 days, stratified estimates would lack the within-operator sample size required for stable inference.

No co-located on-site baseline was available to quantify ordinary observer-to-observer variability at SBMI independently of remote observation. Concordance was also the weakest during a brief late-October cluster of reduced-visibility and fog events; since low-visibility conditions are the regime in which the service is most safety-critical, dedicated validation under reduced visibility remains a priority for future work. These constraints do not negate the current finding of high concordance, but they do define the boundary of inference.

CONCLUSION

This shadow-mode validation showed high meteorological concordance between remote and on-site observation at SBMI over 749 paired reports collected during nine months of operation. The weighted concordance index exceeded the 95% benchmark, agreement statistics were consistently strong, and commissioning-phase convective-cloud divergences disappeared after April 2025.

The evidence may inform future regulatory evaluation of remote aerodrome flight information service for comparable low-density regional aerodromes. It should not be generalized beyond the daytime visual-flight context studied here yet. Multi-site studies, clustered uncertainty analysis, and operator-stratified evaluation remain necessary before broader generalization.

CONFLICT OF INTEREST

The remote aerodrome flight information service (R-AFIS) system validated in this study was deployed and operated by a specialized private service provider, the author's employer (AMD Services), contracted by Companhia de Desenvolvimento de Maricá (CODEMAR). The validation was designed and conducted by the service provider. The author also discloses a memorandum of understanding (MOU No. 033/ICEA/2024) with ICEA, a public institution independent of the service provider; the MOU concerns general cooperation and did not govern the validation reported here. The paired observational records are openly archived in a public repository (Zenodo; see the data availability statement), which allows independent verification. The author acknowledges this commercial interest as a potential conflict of interest and mitigated it through concordance thresholds fixed before the statistical analysis, the use of the official on-site reports as the reference, and open availability of the data and analysis code.

DATA AVAILABILITY STATEMENT

The paired METAR/SPECI dataset and the Python analysis code are openly available at Zenodo (<https://doi.org/10.5281/zenodo.19462685>). The supplementary material (Tables S1–S7) is deposited in the same Zenodo record (Amadeu 2026), under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

FUNDING

Not applicable.

DECLARATION OF USE OF ARTIFICIAL INTELLIGENCE TOOLS

During the preparation of this work, the author used generative artificial intelligence tools (Anthropic Claude, GitHub Copilot and OpenAI ChatGPT) for drafting assistance, language editing, formatting of tables and references, review of the analysis code, and review of the proof corrections. After using these tools, the author reviewed and edited the content as necessary and takes full responsibility for the publication's content.

ACKNOWLEDGMENTS

Not applicable.

REFERENCES

Amadeu GVM (2026) Paired METAR/SPECI dataset, analysis code and supplementary material (Tables S1–S7): remote aerodrome flight information service shadow-mode validation at SBMI (2025) [dataset and supplementary material]. Zenodo. <https://doi.org/10.5281/zenodo.19462685>

Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20(1):37–46. <https://doi.org/10.1177/001316446002000104>

Departamento de Controle do Espaço Aéreo (DECEA) (2023) ICA 63-10: Estações Prestadoras de Serviços de Telecomunicações e de Tráfego Aéreo (EPTA). Rio de Janeiro: DECEA. Available at: <https://publicacoes.decea.mil.br/publicacao/ica-63-10>



Departamento de Controle do Espaço Aéreo (DECEA) (2024) Relatório de *performance* ATM do sistema de controle do espaço aéreo brasileiro – SISCEAB 2024. Rio de Janeiro: DECEA. https://pesquisa.icea.decea.mil.br/performance_report/2024/

Departamento de Controle do Espaço Aéreo (DECEA) (2025) Quais as áreas de jurisdição dos órgãos regionais do DECEA? Rio de Janeiro: DECEA. Available at: <https://ajuda.decea.mil.br/base-de-conhecimento/quais-as-areas-de-jurisdicao-dos-orgaos-regionais-do-decea/>

European Union Aviation Safety Agency (EASA) (2019) Guidance Material on Remote Aerodrome Air Traffic Services, Issue 2. Annex I to ED Decision 2019/004/R, 15 February 2019. Cologne: EASA. Available at: <https://www.easa.europa.eu/en/document-library/agency-decisions/ed-decision-2019004r>

Fürstenau N, editor (2016) Virtual and remote control tower: research, design, development and validation. Cham: Springer. <https://doi.org/10.1007/978-3-319-28719-5>

International Civil Aviation Organization (ICAO) (2018) Annex 3 to the Convention on International Civil Aviation: meteorological service for international air navigation. Montreal: ICAO. Available at: <https://store.icao.int/en/annex-3-meteorological-service-for-international-air-navigation-1>

Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15(2):155-63. <https://doi.org/10.1016/j.jcm.2016.02.012>

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159-74. <https://doi.org/10.2307/2529310>

SESAR Joint Undertaking (2016) Remote tower - solution 71: validation report. Brussels: SESAR Joint Undertaking. Available at: <https://www.sesarju.eu/sesar-solutions/atc-and-afis-service-single-low-density-aerodrome-remote-cwp>

SESAR Joint Undertaking (2019) PJ05 Remote Tower for Multiple Airports: final project report. Deliverable D1.2, edition 01.00.00, 27 November 2019. Brussels: SESAR Joint Undertaking. Available at: <https://www.sesarju.eu/node/3564>

Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86(2):420-8. <https://doi.org/10.1037/0033-2909.86.2.420>

Wilson EB (1927) Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 22(158):209-12. <https://doi.org/10.1080/01621459.1927.10502953>