



OPEN

## Aircrew rostering workload patterns and associated fatigue and sleepiness scores in short and medium haul flights in Brazil

Tulio E. Rodrigues<sup>1,2✉</sup>, Eduardo Furlan<sup>2</sup>, André F. Helene<sup>3</sup>, Otaviano Helene<sup>1</sup>, Eduardo Pessini<sup>4</sup>, Alexandre Simões<sup>5</sup>, Maurício Pontes<sup>6</sup> & Frida M. Fischer<sup>7</sup>

Relationships between workload and fatigue or sleepiness are investigated through the analysis of rosters and responses to questionnaires from Brazilian aircrews, taken from *Fadigômetro* database. The approach includes temporal markers—coinciding with Samn–Perelli (SP) and Karolinska Sleepiness Scale (KSS) responses—where SAFTE-FAST software outcomes are calculated. The latter results follow the increase of fatigue and sleepiness perceptions during the dawn (0h00 to 05h59), but underestimate self-rated scores during the evening (18h00 to 23h59). Importantly, the KSS scores and the inverse of SAFTE-FAST effectiveness fit the relative risk of pilot errors, representing interesting proxies for risk assessment. Linear relationships obtained between workload metrics, computed within 168-h prior to the responses, and self-rated SP and KSS scores provide a consistent method to estimate accumulated fatigue and sleepiness. Considering 7149 rosters of 2023, the duty time ( $DT$ ), the number of flight sectors ( $N_{CREW}$ ) and the sum of flight sectors with sit periods longer than one hour ( $N_{CREW} + N_{SIT}$ ) are associated with 70.1%/60.6% of the highest predicted scores of SP/KSS. Applying the mitigations  $DT \leq 44h$ ,  $N_{CREW} \leq 15$  and  $N_{CREW} + N_{SIT} \leq 19$  for every 168-h interval yields a significant decrease in higher values of SP/KSS with minimal impact on roster coverage.

**Keywords** Aircrew rosters, Fatigue, Sleepiness, Workload, Regulatory aspects, Fatigue risk management

According to the International Civil Aviation Organization (ICAO)<sup>1</sup> fatigue is defined as “a physiological state of reduced mental or physical performance capability resulting from sleep loss, extended wakefulness, circadian phase, and/or workload (mental and/or physical activity) that can impair a person’s alertness and ability to perform safety related operational duties”. Consequently, ICAO recommends that operators apply methods to manage fatigue risks, either via a Prescriptive Approach or via the implementation of a Fatigue Risk Management System (FRMS)<sup>1</sup>. The Prescriptive approach requires that operators comply with prescriptive limits defined by the State, while managing fatigue hazards using the Safety Management System (SMS) processes, and the FRMS approach is a performance-based system that also need to be approved by the State. Both approaches are based on scientific principles and knowledge and operational experience.

Despite been necessary in most cultures, the Prescriptive Approach is not sufficient to mitigate fatigue and/or sleepiness in all scenarios, especially during disruptive and/or night shifts, as recently announced by the European Aviation Safety Agency (EASA)<sup>2</sup>. Such finding endorses the need for improvements in Safety Management Systems in order to identify potential hazards and apply effective mitigation barriers, regardless of being fully compliant with the regulations. More specifically, the “interactions between fatigue and workload in their effects on physical and mental performance” should also be taken into account<sup>1</sup>.

As pointed out recently<sup>3</sup>, biomathematical models have been developed in the last decades to estimate fatigue and sleepiness levels in industry, but still have their limitations to provide quantitative results. The models usually take into account the sleep-awake cycle, governed by the homeostatic process, and the circadian rhythm

<sup>1</sup>Experimental Physics Department, Physics Institute, University of São Paulo, P. O. Box 66318, CEP 05315-970 São Paulo, Brazil. <sup>2</sup>Technical Board, Gol Aircrew Association (ASAGOL), São Paulo, Brazil. <sup>3</sup>Department of Physiology, Institute of Biosciences, University of São Paulo, São Paulo, Brazil. <sup>4</sup>Aeronautics Institute of Technology (ITA), São Paulo, Brazil. <sup>5</sup>Safety Board, LATAM Aircrew Association (ATL), São Paulo, Brazil. <sup>6</sup>Technical Board, Brazilian Association of Civil Aviation Pilots (ABRAPAC), São Paulo, Brazil. <sup>7</sup>Department of Environmental Health, School of Public Health, University of São Paulo, São Paulo, Brazil. ✉email: tulio@if.usp.br

oscillations along the day<sup>4–7</sup>. The effect of sleep inertia—a transient reduction in alertness shortly after waking up—is also considered in some models<sup>8–20</sup>. Despite being interesting and useful tools for estimating fatigue and/or sleepiness, models may not quantitatively address the complex relationships between task-related workload patterns and fatigue or sleepiness outcomes, as recently presented in the Fatigue management resource guide published by the Australian Civil Aviation Safety Authority (CASA)<sup>21</sup>.

Considering Brazilian regulation, a new set of prescriptive rules was implemented by the local regulator Agência Nacional de Aviação Civil (ANAC) in March 2020 (Regulamento Brasileiro de Aviação Civil 117, RBAC 117)<sup>22</sup>, following the requirements of the new Labour Law in force since August of 2017<sup>23</sup>.

In 2017, a new collaboration amongst several stakeholders—named *Fadigômetro*—was initiated, with the aim of estimating fatigue and sleepiness outcomes in the analysis of thousands of Brazilian aircrew rosters, as predicted by the Sleep, Activity, Fatigue, and Task Effectiveness Fatigue Avoidance Scheduling Tool (SAFTE-FAST) software, which is based on the SAFTE model described in Ref.<sup>10</sup>. So far, two studies have been produced under the scope of *Fadigômetro*, either investigating the sensitivity of SAFTE-FAST outputs when comparing rosters of low- versus high-season months of 2018<sup>24</sup>, or providing a comprehensive analysis of the root causes of fatigue related with night-shifts and departures and/or arrivals within 02h00 and 06h00<sup>3</sup>. This latest study analysed 8476 rosters in a pre-COVID-19 period (2019 and early 2020), proposing important safety recommendations for both the regulatory framework and fatigue risk management policies, which were subsequently approved and endorsed—as preventive measures—by the National Committee for the Prevention of Aeronautical Accidents (CNPAA) in November 2022.

Another recent study under RBAC 117 rules included 51 Brazilian airline pilots and demonstrated higher odds for high scores of self-rated fatigue and sleepiness comparing shifts encompassing 0h00 and 06h00 and early-start shifts (start of duty between 06h01 and 07h59) with the other types of shifts, besides other relevant findings<sup>25</sup>.

Brazilian regulator ANAC is currently conducting a review of the RBAC 117 rules, proposing considerable increases in several prescriptive limits (including maximum flight duty periods per day, maximum flight hours within 28 days, among others), which makes it necessary that current fatigue levels and associated risks are accurately estimated and appropriately managed. Such premisses depend on a robust and rigorous test of current regulations, which has not yet being accomplished. In this regard, the present work proposes a novel approach to bridge this gap and evaluate the correlations and relationships between several workload patterns in aircrew rosters along each 168-h loop and the self-rated fatigue and sleepiness scores in terms of the Samn-Perelli<sup>26</sup> and Karolinska Sleepiness Scale<sup>27</sup>, respectively. While this approach does not aim to exhaust the extensive literature on fatigue measures and models (see Ref.<sup>28</sup>), it provides insights into best practices for using bio-mathematical models to guide management decisions, especially when evaluating fatigue and/or sleepiness outcomes in high workload scenarios. As highlighted by the Australian Civil Aviation Safety Authority<sup>21</sup>, biomathematical models differ in scope, have important limitations, and should only be used as supportive tools rather than the sole basis for roster design. Finally, based on our analysis, we propose mitigation strategies in some key roster metrics that are likely to generate small impacts on roster coverage during optimization processes while yielding a remarkable reduction in accumulated fatigue and sleepiness.

## Methods

### The sample and its requirements

The sample includes the responses to a questionnaire and executed rosters from aircrew workers pertaining to major Brazilian airlines, without distinction of sex, race, age, rank or years in the job. The sample was split into four components:

- Sample 1: 585 aircrew responses of the questionnaire since its launch in June 2021.
- Sample 2: 263 executed rosters from June 2021 until December 2023 overlapping the questionnaire responses dates for each aircrew.
- Sample 3: 216 executed rosters satisfying the requirement of Sample 2 with valid SAFTE-FAST outputs.
- Sample 4: 7149 executed rosters of 2023.

The questionnaire and the overlapping rosters were extracted from the *Fadigômetro* database on January 15, 2024 and all the 2023 rosters were extracted on March 21, 2024.

### Ethical and legal aspects

This work was approved by the ethics committee of the Institute of Biosciences, University of São Paulo (Certificate of Presentation for Ethical Appraisal no. 89058318.7.0000.5464). The data acquisition methods and analyses ensure confidentiality to all eligible aircrew volunteers who agreed to participate by approving a digital informed consent form. We declare no conflict of interest with any professional association involved in the experiment, airline or local regulator. The confidentiality of the airlines was also assured. All methods were performed in accordance with the relevant guidelines and regulations.

### Questionnaire and roster data: acquisition, criteria and markers

The *Fadigômetro* questionnaire was designed as a cross-sectional instrument, answered only once by each participant. It comprises 30 items covering a broad range of aspects, including sociodemographic characteristics, behavioral and health-related information, sleep habits and quality in different contexts, commuting routines, and self-perceived fatigue and sleepiness in operationally relevant scenarios. The responses were collected through a web application, where each volunteer accessed a personal and exclusive link. For each submission, the timestamp was automatically recorded. In the present study, the analysis was focused on the self-reported fatigue

(Samn-Perelli scale, SP) and sleepiness (Karolinska Sleepiness Scale, KSS), while the remaining questionnaire information will be explored in future analyses.

Rosters were also acquired through a web application and, more recently, through an application available on IOS and Android mobile platforms. These rosters were then extracted from the *Fadigômetro* platform as comma-separated values (CSV) files, already formatted for direct import into the SAFTE-FAST software. Rosters with long-haul flights performed with augmented crew (three or four pilots) within 168 h prior to the questionnaire response time (hour and minute) were excluded (sample 2), thus ensuring that all rosters were executed with minimal crew (only two pilots).

Markers of 15 min, ending at the exact questionnaire response time for each aircrew, were merged with the respective roster's files in SAFTE-FAST console (version 6.6; see Supplementary Section for details of parameters and criteria). The SAFTE-FAST average values of Effectiveness, Sleep Reservoir, Samn-Perelli Scale and Karolinska Sleepiness Scale were then calculated within the marker periods. For details of the SAFTE-FAST model refer to Ref.<sup>10</sup>. In 47 of the 263 rosters with markers, the SAFTE-FAST software (with all Auto-Sleep functions activated) predicted sleep periods overlapping with questionnaire response times, which prevented valid outputs. These markers were then discarded in sample 3 (N = 216) in order to avoid any arbitrary manipulation of SAFTE-FAST predictive sleep. The SAFTE-FAST parameters and criteria are the same as those adopted in Ref.<sup>3</sup> and are discussed in subsection 1 of the Supplementary Section.

Sample 4 included all rosters of 2023 (N = 7149), except those with long-haul international flights, thus ensuring that flights were carried out with minimum crew. This eliminates flights with augmented crews that allow in-flight sleep events, which are not considered in the SAFTE-FAST configuration in these analyses. Differently from our previous 30-day epochs for rosters<sup>3</sup>, this work adopts a 28-day epoch, which is in line with the RBAC 117 parameter for accumulated block hours<sup>22</sup>. Moreover, this sample also excluded rosters with more than 21 consecutive days without any event, which are likely associated with vacation periods encompassing fractions of two consecutive months. As described in more detail elsewhere<sup>3</sup>, all working and flying activities of the executed rosters were included in the analyses, except home standbys, where the aircrew stays on-call at the place of his/her choice.

### Fatigue and sleepiness outcomes, workload patterns and statistical analyses

The model dependent variables include the SAFTE-FAST outputs of Effectiveness [ $E_{SF}(\%)$ ], Sleep Reservoir [ $R_{SF}(\%)$ ], Samn-Perelli Scale ( $SP_{SF}$ ) and Karolinska Sleepiness Scale ( $KSS_{SF}$ ), all calculated within the 15-minute markers (sample 3). They also include the SAFTE-FAST minimum effectiveness [ $EM_C(\%)$ ], minimum sleep reservoir [ $RM_C(\%)$ ], and fatigue hazard area [ $FHA_C(\text{min})$ ], all within critical phases of flight, considering a standardized 28-days epoch for all rosters of 2023 (sample 4). The self-rated fatigue and sleepiness scores include the Samn-Perelli (SP)<sup>26</sup> and Karolinska Sleepiness Scale (KSS)<sup>27</sup> from the questionnaire (samples 2 and 3). Given that sample 3 had few results for  $SP \geq 6$  and  $KSS \geq 8$ , the SP results were grouped from 1 to 5 and a sixth group combining the scores 6 and 7, whereas the KSS scores grouped from 1 to 7 and an eighth group combining the scores 8 and 9. The independent workload metrics were chosen considering the authors' operational expertise and include:

- Number of night-shifts ( $N_{NS}$ ) – shifts that encompass any fraction between 0h00 and 06h00.
- Number of consecutive night-shifts ( $N_{CNS}$ ).
- Number of early-start shifts ( $N_{ES}$ ) – shifts that start between 06h01 and 07h59.
- Total duty time in hours ( $DT$ ).
- Number of flight sectors ( $N_{CREW}$ ).
- Number of departures and/or arrivals between 02h00 and 06h00 ( $N_{WOCL}$ ).
- Number of long sit events ( $N_{SIT}$ ) – intervals between consecutive flight sectors longer than 1 h.
- Number of short rest periods ( $N_{REST}$ ) – rest periods shorter than 16 h.
- Number of long duties ( $N_{DUTY}$ ) – duty periods longer than 9 h.

All metrics were computed by a dedicated Fortran algorithm (Microsoft Visual Studio 2019, Intel® Fortran Compiler v.2021.1.1) specifically developed to parse, integrate, and consolidate the roster CSV files and marker datasets. From the rosters, the algorithm retrieved non-identifiable internal ID, exclusion criteria for long-haul flights, duty type, flight leg times, and duty period boundaries; from the markers, it incorporated the same internal ID, submission timestamp, self-reported SP and KSS, and SAFTE-FAST outcomes ( $E_{SF}$ ,  $R_{SF}$ ,  $SP_{SF}$  and  $KSS_{SF}$ ). By automating this process, the algorithm ensured full traceability from raw inputs to consolidated workload metrics and fatigue/sleepiness outcomes, calculated over standardized intervals of 168 hours (sample 2) or 28 days (sample 4).

For the normality hypothesis we have applied the Shapiro-Wilk test<sup>29</sup>, and for the null hypothesis the Mann-Whitney test for independent samples. All statistical tests, descriptive statistics, calculations of Pearson's correlations and risk ratios analyses were performed using the IBM SPSS version 25. The linear fittings were done using the Least Squares Method (LSM) in matrix formalism<sup>30</sup> with Mathcad 15. The 95% confidence intervals of the fitted functions were obtained using uncertainty propagation methods and the covariance matrix of the fitted parameters, as described in more detail elsewhere<sup>3</sup>.

## Results

### Questionnaire general statistics

The second questionnaire of *Fadigômetro* was answered by 585 Brazilian aircrew workers (sample 1) from June 14, 2021 to January 15, 2024. The participants were 69.2% male, 26.5% Captains, 22.9% First Officers and 50.6% Cabin Crew. The average ages and respective standard deviations in years were:  $41.3 \pm 11.8$  (male),

$37.0 \pm 8.6$  (female),  $46.9 \pm 12.2$  (Captains),  $36.4 \pm 8.6$  (First Officers),  $38.0 \pm 9.8$  (Cabin Crew) and  $42.0 \pm 11.9$  (Pilots: Captains and First Officers).

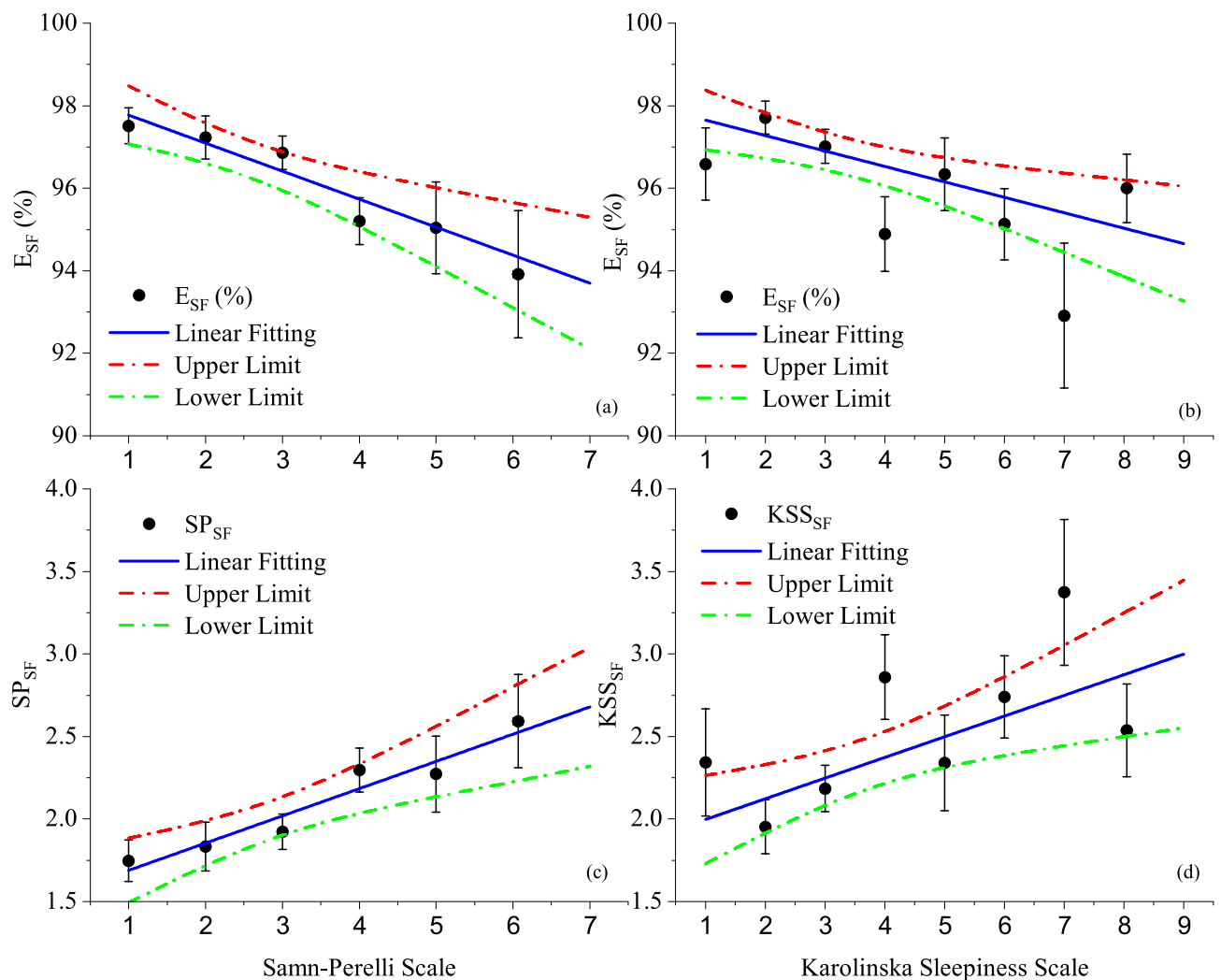
As a baseline for subsequent analyses, the mean Samn-Perelli (SP) and Karolinska Sleepiness Scale (KSS) scores, and their standard errors, were  $3.499 \pm 0.058$  and  $4.439 \pm 0.082$  for sample 1 ( $N = 585$ ),  $3.494 \pm 0.085$  and  $4.39 \pm 0.12$  for sample 2 ( $N = 263$ ), and  $3.458 \pm 0.095$  and  $4.38 \pm 0.14$  for sample 3 ( $N = 216$ ), respectively.

### SAFTE-FAST outputs and self-rated fatigue and sleepiness scores

The SAFTE-FAST Effectiveness ( $E_{SF}$ ) and Sleep Reservoir ( $R_{SF}$ ) represent two important software outputs to estimate fatigue and sleep debt, respectively<sup>5</sup>. They are usually calculated during critical phases of flight (departures and/or arrivals), indicating potentially hot spots of fatigue and/or excessive wakefulness during high demands of cognitive tasks.

The averages of these outputs—calculated for each 15-minute marker of sample 3 ( $N = 216$ )—also demonstrate statistically significant Pearson's correlations with the self-rated fatigue and sleepiness scores from the questionnaire:  $SP \times E_{SF}$  ( $r = -0.242, p < 0.001$ ),  $SP \times R_{SF}$  ( $r = -0.320, p < 0.001$ ),  $KSS \times E_{SF}$  ( $r = -0.184, p = 0.007$ ) and  $KSS \times R_{SF}$  ( $r = -0.168, p = 0.014$ ). These findings show that the higher the SF Effectiveness (or Sleep Reservoir), the lower are the SP and KSS scores reported by the participants.

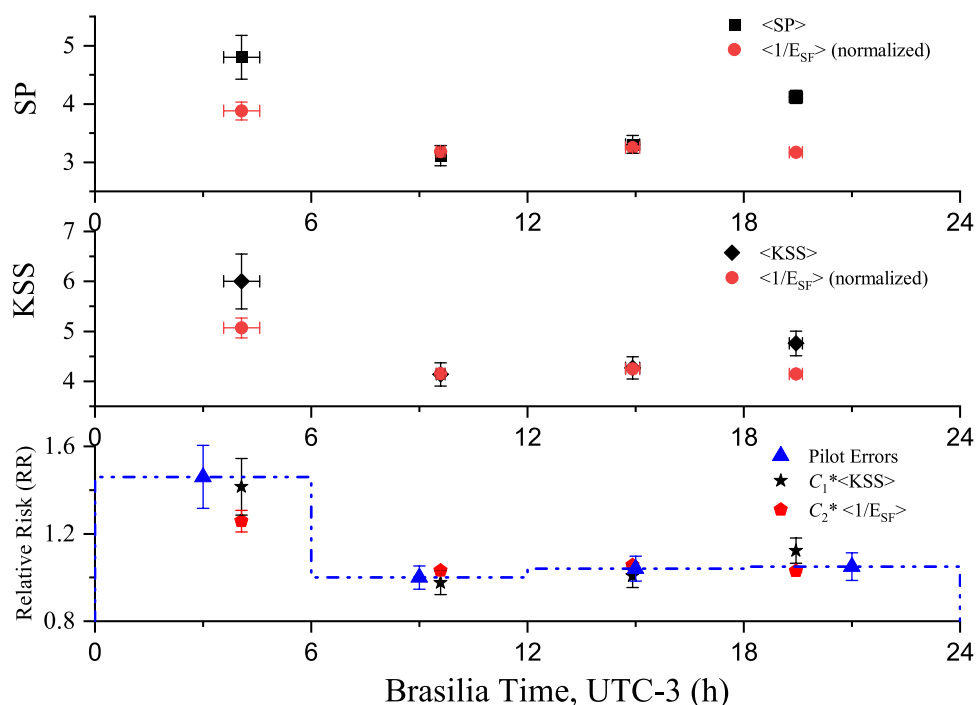
Linear relationships were also obtained using the LSM<sup>30</sup> and grouping the SP and KSS results into 6 and 8 intervals, respectively (sample 3, see Methods). Figure 1 presents the fittings (blue lines) of SP (upper-left) and KSS (upper-right) versus  $E_{SF}$ , as well as,  $SP \times SP_{SF}$  (lower-left) and  $KSS \times KSS_{SF}$  (lower-right). The dashed-



**Fig. 1.** Upper panels: Self-rated SP (left) and KSS (right) scores versus average SAFTE-FAST Effectiveness (black circles) and associated standard errors (error bars). Lower panels: Self-rated SP (left) and KSS (right) scores versus average SAFTE-FAST SP (left) and KSS (right) (black circles) and associated standard errors (error bars). The blue lines represent the fitted functions and the dashed-dotted red and green lines represent, respectively, their upper and lower limits considering a 95% confidence interval (CI) ( $N = 216$ , sample 3).

Function	Fitting parameters				
	Slope	Intercept	N.D.F.	$\chi^2$	Probability(%)
SPx $E_{SF}$	-0.68±0.18%	98.45±0.51%	4	2.57	63.1
KSSx $E_{SF}$	-0.37±0.12%	98.03±0.47%	6	10.06	12.2
SPxSP $_{SF}$	0.165±0.043	1.52±0.14	4	1.94	74.6
KSSxKSS $_{SF}$	0.125±0.041	1.87±0.17	6	10.02	12.4

**Table 1.** Slopes and intercepts of the fitted functions presented in Fig.1. Also shown the number of degrees of freedom (N.D.F.), the  $\chi^2$  and the probability of exceeding the  $\chi^2$ .



**Fig. 2.** Upper panel: Averages of SP scores (black squares) and their standard errors (black error bars) and the averages of  $1/E_{SF}$  (red circles), normalized to the SP averages of groups 2 (morning) and 3 (afternoon), and their standard errors (red error bars), as a function of the average response times for each clock-time group in Brasilia (Brazilian capital) time (UTC-3). Middle panel: The same for the KSS average scores (black diamonds) and the averages of  $1/E_{SF}$  (red circles) normalized to the KSS averages of groups 2 and 3. Lower panel: Relative Risk of pilot errors in the cockpit for the four clock-time groups [dashed blue histograms with each data point (blue triangles with error bars)] and the fitted averages of KSS (black stars with error bars) and  $1/E_{SF}$  (red pentagons with error bars) as a function of the average response times for each clock-time group. All plots refer to sample 3 (N = 216). Details in the text.

dotted red and green lines represent, respectively, the upper and lower limits of the fitted function considering a 95% Confidence Interval (CI). The results of the fittings are summarized in Table 1.

Circadian oscillations in the SAFTE-FAST Effectiveness and self-rated SP and KSS scores were also investigated by splitting sample 3 into four clock-time periods, herein defined as: dawn (from 0h00 to 05h59, group 1), morning (from 06h00 to 11h59, group 2), afternoon (from 12h00 to 17h59, group 3) and evening (from 18h00 to 24h59, group 4), all considering Brasilia (Brazilian capital) time (UTC-3). This choice is particularly useful since night shifts are defined by Brazilian regulations<sup>22,23</sup> as any shift encompassing any fraction between 0h00 and 05h59, with the other groups defined with the same time interval. For each clock-time group, the averages of the response times and associated standard errors were calculated. Since the relative fatigue risk is proportional to the inverse of  $E_{SF}^3$  - which derives from the finding that  $1/E_{SF}$  also scales with the relative likelihood of human-error rail road accidents<sup>24</sup>—it is plausible also to compare this transformed variable with the self-rated SP and KSS scores. These results are presented in Fig.2 both for the SP (upper panel) and KSS (middle panel) scores, where the averages of  $1/E_{SF}$  (with  $E_{SF}$  in decimal units) have been multiplied by a normalization constant that fit the SP (and KSS) averages in groups 2 and 3 simultaneously. Such criterion was adopted after retaining the null hypothesis (Mann–Whitney tests for independent samples) between groups

2 (morning) and 3 (afternoon) both for SP ( $p = 0.310$ ) and KSS ( $p = 0.669$ ) and also considering the flat behaviour of the average values of  $1/E_{SF}$  for these periods of the day.

Following the same steps described elsewhere<sup>3</sup>, the SAFTE-FAST software was also used to calculate the average values of minimum effectiveness  $EM_C$ , minimum sleep reservoir  $RM_C$ , and fatigue hazard area  $FHAC$ , all within critical phases of flight, considering a standardized 28-days epoch for all rosters of 2023 (sample 4). The results of this calculation are presented in subsection 1 of the Supplementary Section.

### Pilot errors versus time of the day

This section investigates, as a function of time of day, the associations between pilot errors<sup>31</sup> and two quantities: the self-rated sleepiness scores from the *Fatigômetro* questionnaire (KSS) and the inverse of the SAFTE-FAST effectiveness ( $1/E_{SF}$ ). The pilot errors were collected in a large Brazilian airline in 2005 via the analysis of 155,327 flight hours using the Flight Data Monitoring (FDM) system. Overall, de Mello et al.<sup>31</sup> found 1065 pilots errors, distributed along the time of the day in the same time periods adopted in the previous section.

The averages of KSS scores and the inverse of SAFTE-FAST effectiveness ( $1/E_{SF}$ ) within the same four clock-time groups (sample 3,  $N = 216$ ) were fitted to the relative risk of pilot errors (normalized to unit within 06h00–11h59) using the Least Square Method. The uncertainties for  $\langle KSS \rangle$  and  $\langle 1/E_{SF} \rangle$  were taken as their respective standard errors, whereas the uncertainty in the relative risk of pilot error, herein denoted as relative risk ( $RR$ ), was assumed as  $\sigma_{RR} = RR \cdot \sqrt{N}/N$ , where  $N$  stands for the total number of errors (this assumption holds if the distribution of pilot errors along time follows a Poisson distribution). The fitted constant for KSS was  $C_1 = 0.240 \pm 0.010$ , with a  $\chi^2$  of 1.05 for 3 degrees of freedom ( $p = 0.790$ ). For the inverse of  $E_{SF}$ , the constant was  $C_2 = 1.001 \pm 0.031$ , with a  $\chi^2$  of 2.30 for 3 degrees of freedom ( $p = 0.513$ ). In both cases, the averages are statistically indistinguishable from the pilot errors, indicating that KSS and  $1/E_{SF}$  can serve as interesting proxies for risk assessment. The bottom panel of Fig. 2 presents these results, showing the fitted KSS (black stars with error bars) and  $1/E_{SF}$  (red pentagons with error bars), along with the relative risk of pilot errors<sup>31</sup>, represented by the dashed blue histogram with each data point (blue triangles with error bars) centered at each clock-time group.

### Workload patterns and self-rated fatigue and sleepiness scores

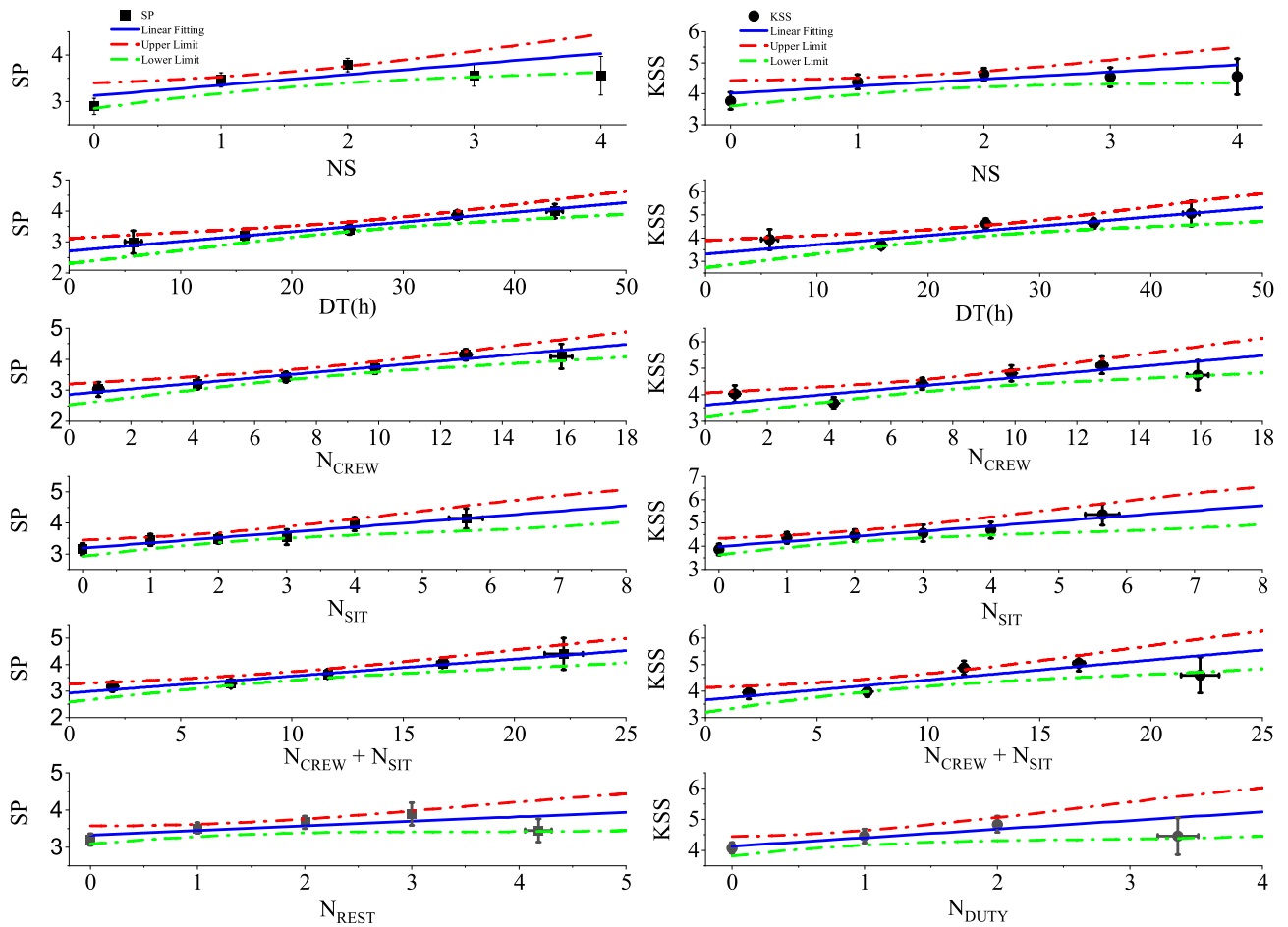
To shed light on the complex interactions between workload patterns and fatigue/sleepiness outcomes, several rosters metrics of sample 2 ( $N = 263$ ) were carefully analysed through a dedicated computer algorithm within the 168 h period prior to the questionnaire response times (see Methods). The Pearson's correlations between all the independent metrics and the fatigue/sleepiness outcomes are presented in Table 2, with statistically significant results highlighted in bold, according to the criterion of  $p < 0.05$ . Among all metrics, the number of night-shifts ( $N_{NS}$ ), total duty time [ $DT(h)$ ], number of flight sectors ( $N_{CREW}$ ) and number of sit times longer than one hour ( $N_{SIT}$ ) have statistically significant correlations with both SP and KSS scores. Additionally, the number of short rests ( $N_{REST}$ ) and long duties ( $N_{DUTY}$ ) are also significantly correlated with SP and KSS, respectively. Although the absolute values of Pearson's  $r$  were modest, statistically significant correlations at the 5% level were obtained for some metrics, which is explained by the relatively large sample size.

Linear relationships were obtained between  $N_{NS}$ ,  $DT(h)$ ,  $N_{CREW}$ , and  $N_{SIT}$ , as well as the sum of flight sectors with long sit events  $N_{CREW} + N_{SIT}$  with both SP and KSS scores, and also  $N_{REST}$  versus SP and  $N_{DUTY}$  versus KSS, as shown in Fig. 3. All the fittings were done by grouping the workload metrics into suitable intervals, where the number of bins ( $N_{Group}$ ) equals the number of degrees of freedom ( $N.D.F.$ ) plus two. The dispersions in the x-axis, whenever applicable, were also taken into account, such that the total variance of each data point (SP or KSS) could be written as:

$$\sigma_y^2 = \langle \partial y / \partial x \rangle^2 \cdot \sigma_x^2 + \sigma_{y0}^2, \quad (1)$$

Workload metric	Fatigue/sleepiness self-rated scores			
	SP		KSS	
	$r$	$p$ -value	$r$	$p$ -value
$N_{NS}$	<b>0.166</b>	<b>0.007</b>	<b>0.125</b>	<b>0.042</b>
$N_{CNS}$	0.056	0.368	0.044	0.478
$N_{ES}$	0.075	0.223	0.036	0.561
$DT(h)$	<b>0.208</b>	<b>0.001</b>	<b>0.176</b>	<b>0.004</b>
$N_{CREW}$	<b>0.260</b>	<b>&lt;0.001</b>	<b>0.211</b>	<b>0.001</b>
$N_{WOCL}$	0.007	0.907	0.035	0.572
$N_{SIT}$	<b>0.180</b>	<b>0.003</b>	<b>0.163</b>	<b>0.008</b>
$N_{REST}$	<b>0.134</b>	<b>0.030</b>	0.077	0.215
$N_{DUTY}$	0.096	0.122	<b>0.122</b>	<b>0.048</b>
$N_{CREW} + N_{SIT}$	<b>0.253</b>	<b>&lt;0.001</b>	<b>0.209</b>	<b>0.001</b>

**Table 2.** Pearson's correlations ( $r$ ) and respective  $p$ -values between the workload metrics and fatigue/sleepiness self-rated scores (SP and KSS) of sample 2 ( $N = 263$ ).



**Fig. 3.** Average SP (left) and KSS (right) scores (data points) and their standard errors (error bars) as a function of workload metrics computed within 168 h prior to the questionnaire response times (sample 2,  $N = 263$ ). The fitted functions are shown by the solid blue lines, where the dashed-dotted red and green lines represent, respectively, their upper and lower limits considering a 95% CI. Details in the text.

where  $\sigma_{y_0}$  is the standard error for SP (or KSS),  $\sigma_x$  the standard error of the workload metric ( $x$ -error) and  $(\partial y/\partial x)$  the slope of the fitted function, which was calculated iteratively until its convergence. The results of all the fittings are shown in Table 3, where all slopes proved to be statistically different from zero (uncertainties smaller than the estimates), thus reinforcing the robustness of the relationships.

The fittings presented in Fig. 3 represent meaningful results to understand the relationships between several workload metrics and the expected SP (or KSS) average values, considering an accumulated period of 168 h before the assessment. On the other hand, these findings do not clarify a crucial question regarding fatigue risk management: What is the relative probability of having a high fatigue (or sleepiness) score considering different workload profiles? So, to address this question we conducted risk ratio calculations comparing groups of rosters of sample 2 with different workload levels, also assuming that scores of  $SP \geq 6$  or  $KSS \geq 7$  do represent high fatigue or sleepiness outcomes<sup>26,27,32</sup>.

According to Shapiro–Wilk Normality tests, almost all metrics - except  $DT$  ( $p = 0.325$ )—are unlikely to be generated by a normal distribution ( $p \leq 0.002$ ). Consequently, the sample was divided into two Groups, namely:  $G_-$ , which included all the workload values below the average (or at or below the medians), and  $G_+$ , which included all values above the averages (or medians) for each distribution. The results of Risk Ratios between  $G_+$  and  $G_-$ , obtained with 95% CI using SPSS version 25 are presented in Table 4, where we found statistically significant figures for  $DT$ , considering both SP [7.111 (95% CI 1.659–30.481)] and KSS [2.844 (95% CI 1.503–5.384)] scores, as well as, for  $N_{CREW}$  [2.515 (95% CI 1.399–4.519)] and  $N_{CREW} + N_{SIT}$  [2.450 (95% CI 1.380–4.349)] considering KSS scores. Moreover, Table 4 also shows threshold values above which there are less than 2.5% of events in each distribution, together with the actual fraction of events above these cutoff parameters. As will be discussed in detail later, these thresholds are promising metrics do address fatigue and sleepiness mitigation strategies in aircrew rosters.

Function	Fitting parameters				
	Intercept	Slope	N.D.F.	$\chi^2$	Probability (%)
$DT \times SP$	2.71±0.22	0.0312±0.0075 $h^{-1}$	3	0.74	86.3
$DT \times KSS$	3.31±0.31	0.040±0.012 $h^{-1}$	3	3.99	26.3
$N_{NS} \times SP$	3.13±0.14	0.227±0.076	3	6.77	8.0
$N_{NS} \times KSS$	4.02±0.21	0.23±0.11	3	2.41	49.2
$N_{CREW} \times SP$	2.86±0.17	0.090±0.019	4	1.16	88.4
$N_{CREW} \times KSS$	3.61±0.24	0.103±0.029	4	5.01	28.6
$N_{SIT} \times SP$	3.18±0.13	0.171±0.052	4	0.97	91.4
$N_{SIT} \times KSS$	3.98±0.18	0.220±0.074	4	1.08	89.7
$(N_{CREW} + N_{SIT}) \times SP$	2.93±0.17	0.064±0.015	3	1.09	80.0
$(N_{CREW} + N_{SIT}) \times KSS$	3.67±0.23	0.075±0.022	3	4.88	18.1
$N_{REST} \times SP$	3.33±0.12	0.122±0.070	3	3.08	38.0
$N_{DUTY} \times KSS$	4.13±0.16	0.28±0.13	2	1.56	45.8

**Table 3.** Intercepts and slopes of the linear functions between workload metrics—computed within 168 h prior to the questionnaire response times—and the self-rated fatigue (SP) and sleepiness (KSS) scores. Also shown the number of degrees of freedom (N.D.F.), the  $\chi^2$  and the probability of exceeding the  $\chi^2$  for each fitting.

Variable	Risk ratio ( $G_+/G_-$ )		Mean or	Threshold	Fraction of
	$SP \geq 6$	$KSS \geq 7$	median	values	excluded events
$DT$	7.111 (1.659–30.481)	2.844 (1.503–5.384)	25.4h	43.9h*	2.3%
$N_{NS}$	0.960 (0.287–3.209)	0.995 (0.495–2.003)	2	4	0%
$N_{CREW}$	1.677 (0.658–4.270)	2.515 (1.399–4.519)	7	15	1.9%
$N_{SIT}$	1.691 (0.668–4.280)	1.249 (0.711–2.196)	2	5	2.3%
$N_{CREW} + N_{SIT}$	1.810 (0.695–5.120)	2.450 (1.380–4.349)	9	19	1.9%
$N_{REST}$	1.521 (0.607–3.813)	0.718 (0.395–1.304)	1	4	0.8%
$N_{DUTY}$	0.900 (0.304–2.666)	1.513 (0.866–2.644)	1	3	1.5%

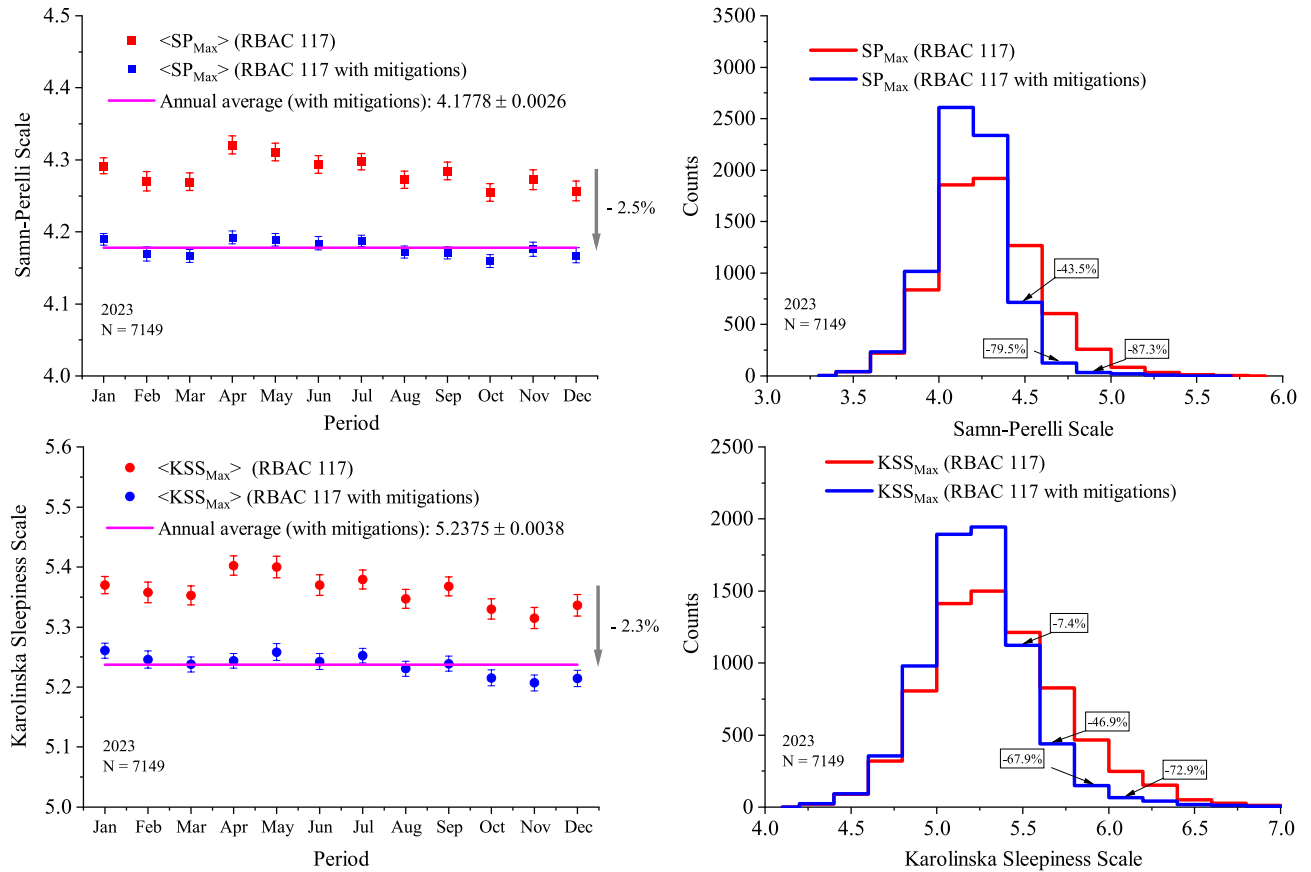
**Table 4.** Risk ratios  $G_+/G_-$  for high fatigue ( $SP \geq 6$ ) and sleepiness ( $KSS \geq 7$ ) scores and their 95% confidence intervals for all workload variables with statistically significant Pearson's correlations with SP and/or KSS (see Table 2). Also shown the mean (or median), the threshold values and the actual fraction of excluded events for each distribution. (\*) For practical purpose the  $DT$  threshold was fixed at 44h.

### Estimating maximum cumulative fatigue and sleepiness scores from roster workload patterns: a novel Fatigue Risk Management approach

In this section, a novel fatigue risk management approach is proposed in order to estimate the maximum cumulative fatigue and sleepiness scores by analysing roster workload patterns. The method also addresses the impacts of mitigation strategies on both operational safety (i.e: reduction in the predictive scores of SP and KSS), as well as on roster coverage (i.e: impacts on roster production). The approach is based on the implementation of mitigations in some key workload metrics, such as  $DT$ ,  $N_{CREW}$  and  $N_{CREW} + N_{SIT}$ , keeping them below or equal the threshold values that exclude less than 2.5% of the higher figures of sample 2 (see Table 4). These workload metrics were chosen given their statistically significant risk ratios for either high fatigue ( $SP \geq 6$ ) or sleepiness ( $KSS \geq 7$ ) scores found between  $G_+$  and  $G_-$ . Considering that these thresholds affect less than 2.5% of each distribution one would expect small impacts on roster coverage during optimization processes.

The analyses are then performed via the implementation of the following steps:

1. First, we pick all non-augmented crew rosters of 2023 (sample 4) using a standardized 28-days epoch for all months.
2. Second, we calculate the expected values of SP and KSS for each 168-h loop and each workload metric using the corresponding fitted functions (see Table 3) plus a random value (negative or positive) normally distributed around zero ( $\mu = 0$ ) with standard deviation equals to  $\sigma_f$ , which is the propagated uncertainty of the fitted functions (95% CI  $\approx 1.96\sigma_f$ ).
3. Third, we collect the maximum value of SP and KSS considering all 168-h loops within 28 days for each roster of 2023 (total of 3,610,145 loops).



**Fig. 4.** Left upper/lower panels: Average maximum SP/KSS scores and their standard errors under the current RBAC 117 rules (red squares/circles) and with the proposed mitigations (blue squares/circles). Also shown the 2023 annual averages with mitigations (magenta lines) both for SP (upper) and KSS (lower) scores. Right upper/lower panels: Maximum SP/KSS distributions under the RBAC rules (red histograms) and with the proposed mitigations (blue histograms).

4. Fourth, we perform the same calculations of the previous steps, but include a mitigation strategy that restricts some key workload metrics at their thresholds shown in Table 4, such that  $DT \leq 44h$ ,  $N_{CREW} \leq 15$  and  $N_{CREW} + N_{SIT} \leq 19$  for each 168-hour loop.

Under this framework, it is possible to estimate the maximum values of SP and KSS (with or without mitigations) by the combination of the analysis of all workload variables in each successive 168-hour loops for all the rosters of 2023 ( $N = 7149$ ) and the fitted functions with their respective standard errors found in the previous section (see Fig. 3 and Table 3 for details).

Considering the current RBAC 117 rules, the predicted average values and standard errors of the maximum SP/KSS scores for each 28-day epochs of 2023 are presented by the red squares/circles in the upper/lower left-hand panels of Fig. 4. The blue squares/circles in the upper/lower left-hand panels represent the average SP/KSS maximum values for the same epochs after applying the mitigations for  $DT$ ,  $N_{CREW}$  and  $N_{CREW} + N_{SIT}$ . The magenta lines show the mitigated annual averages for  $SP_{MAX}$  ( $4.1778 \pm 0.0026$ ) and  $KSS_{MAX}$  ( $5.2375 \pm 0.0038$ ), obtained via the Least Squares Method (SP:  $\chi^2 = 17.4$ ,  $N.D.F. = 11$ ,  $p = 9.8\%$  and KSS:  $\chi^2 = 19.5$ ,  $N.D.F. = 11$ ,  $p = 5.3\%$ ).

The distributions of the maximum SP/KSS scores considering all 7149 rosters of 2023 under the RBAC 117 rules are presented by the red histograms in the upper/lower right-hand panels of Fig. 4, whereas the blue histograms show the SP/KSS maximum scores distributions after applying the mitigations.

Considering all rosters of 2023 under RBAC 117 rules (without mitigation), the workload metrics of  $DT$ ,  $N_{CREW}$  and  $N_{CREW} + N_{SIT}$  were associated with 70.1% (60.6%) of the highest predicted scores of SP (KSS), reinforcing the relevance of these variables in the proposed mitigation. An overall analysis of the workload metrics associated with the predicted highest values of SP and KSS can be found in subsection 2 of the Supplementary Section.

Potential impacts of the proposed mitigations on roster coverage were also investigated, where we found that over the 3,610,145 loops of all rosters of 2023, only 3.9%, 3.1% and 3.8% exceed the thresholds  $DT \leq 44h$ ,  $N_{CREW} \leq 15$  and  $N_{CREW} + N_{SIT} \leq 19$ , respectively. Combining the three workload metrics, only 6.7% of all loops exceed at least one of the proposed limitations. The complete analysis of this impact can be found in subsection 2 of the Supplementary Section.

For the sake of completeness, the workload metrics were also calculated considering a 28-days epochs for all rosters of 2023 (sample 4). The results are presented in subsection 3 of the Supplementary Section.

## Discussion

The self-rated fatigue (SP) and sleepiness (KSS) scores obtained from the questionnaire responses showed statistically significant Pearson's correlations with SAFTE-FAST Effectiveness. Although the correlation coefficients were modest, their statistical significance at the 5% level ( $p < 0.05$ ) reflects the relatively large sample size. Importantly, the linear relationships obtained (upper panels of Fig. 1) demonstrate small variations of  $E_{SF}$  for large variations in SP or KSS. In fact, varying the SP scores from 1 to 7, or KSS from 1 to 9, yields, respectively, small decreases in  $E_{SF}$ , typically from  $\sim 97.8\%$  to  $93.7\%$ , or  $\sim 97.7\%$  to  $94.7\%$ . Furthermore, the slopes found between SP and  $SP_{SF}$  ( $0.165 \pm 0.043$ ) and KSS and  $KSS_{SF}$  ( $0.125 \pm 0.041$ ) (see Table 1) indicate that just small fractions of the variations in the self-rated scores are being reflected in the software outcomes (ideally, one would expect slopes close to unit and intercepts close to zero). Such results suggest the limitations of the model to accurately reproduce perceptions of fatigue and sleepiness, as well as, evidences of software outcomes with high false negative for fatigue/sleepiness considering the Brazilian scenario. As a consequence, these results should be used with caution and not as the sole source of information to rule out the likelihood of potentially high levels of fatigue and/or sleepiness in actual aircrew rosters. A possible explanation for this finding would be that the model does not take workload effects into account when calculating  $E_{SF}$  (or  $SP_{SF}$  and  $KSS_{SF}$ ), probably causing an underestimation of fatigue and/or sleepiness perceptions in successive 168-h loops. Moreover, the inverse of  $E_{SF}$  in sample 3 does not vary considerably between 06h00 and 24h00, in contrast with the circadian oscillations found for SP and at some extent also for KSS (see upper and middle panels of Fig. 2, respectively). In fact, the averages of  $1/E_{SF}$ , normalized to the SP (or KSS) scores of groups 2 (from 06h00 to 11h59) and 3 (from 12h00 to 17h59) present modest peaks within 0h00 and 05h59, but do not reproduce the increase in fatigue (or sleepiness) perceptions between 18h00 and 23h59. Once again, the peak in the self-rated SP—and qualitatively also in KSS—during the evening could be associated with the effects of workload, a feature not included in the calculation of  $E_{SF}$ . Importantly, recent versions of SAFTE-FAST include the Combined Capacity metric, which integrates Effectiveness, Sleep Reservoir, and Workload to identify conditions of high compound fatigue risk when cognitive demands exceed operator capacity. While this functionality is valuable for operational insights and fatigue risk management, it does not modify the fatigue indicators of the model itself—particularly Effectiveness, which remains the main variable used by operators for decision-making.

The circadian oscillations observed in both the self-rated sleepiness scores and the inverse of SAFTE-FAST effectiveness (sample 3,  $N = 216$ ) are quantitatively associated with the relative risk of pilot errors in the cockpit (see the lower panel of Fig. 2), underscoring the ability of both metrics to reproduce objective FDM data. Despite the large uncertainties involved (see Limitations), this finding suggests that errors in the complex cockpit environment aligns with both perceived sleepiness (KSS) and  $1/E_{SF}$ .

The linear relationships found between seven workload metrics—computed within 168 h prior to the questionnaire responses—and both SP and KSS scores of sample 2 (see Fig. 3 and Table 3) demonstrate how fatigue/sleepiness perceptions increase with respect to roster's characteristics. Furthermore, the risk ratios for high fatigue ( $SP \geq 6$ ) or sleepiness ( $KSS \geq 7$ ) scores when comparing groups  $G_+$  and  $G_-$  (see Table 4) indicate statistically significant figures for the total duty time ( $DT$ ) both for SP and KSS, and for the number of flight sectors ( $N_{CREW}$ ) and the aggregated metric ( $N_{CREW} + N_{SIT}$ ) with KSS. Such findings make clear the benefits of mitigation strategies to eliminate extreme situations via the inclusion of threshold values that exclude only few percent (herein fixed at or below 2.5%) of the respective metric distributions (see Table 4). Obviously, these thresholds based on a cut-off of less than 2.5% of each distribution are arbitrary and likely extreme values, but they demonstrate the potential of this new approach to mitigate the risk of high scores of fatigue and sleepiness with the likelihood of minimal impact on roster optimization processes. In this sense, for the purpose of fatigue risk management, operators should analyse the benefits of the mitigations, increasing cut-off values and verifying their impacts on roster coverage (see the Safety Recommendations section).

The analysis of all 2023 rosters (sample 4,  $N = 7149$ ) via the linear relationships found for sample 2 ( $N = 263$ ) propitiated a robust method to estimate maximum values of fatigue (SP) and sleepiness (KSS), as presented in Fig. 4. The novel approach calculates the maximum fatigue and sleepiness scores for each 168-h loop and selects the global maximum within a 28-day epoch for each roster. It also takes into account the uncertainties of the fitted functions (see Fig. 3) using a Monte Carlo sampling technique, thus ensuring that the final results reflect the covariance matrix of the fitted parameters and, ultimately, the variance of the fitted data of sample 2 (see also the Limitations section).

The results of the averaged maximum values of SP (KSS) for each month of 2023 are presented by the red squares (circles) in the left-upper (lower) panels of Fig. 4. The predicted averages of  $SP_{MAX}$  ( $KSS_{MAX}$ ) scores after mitigations are 2.5% (2.3%) lower, when compared with the actual RBAC 117 rules, as shown by the fitted magenta lines in the left-upper (lower) panels of Fig. 4. Indeed, the mitigated annual averages found for  $SP_{MAX}$  ( $4.1778 \pm 0.0026$ ) and  $KSS_{MAX}$  ( $5.2375 \pm 0.0038$ ) do represent very precise figures, with uncertainties of 0.06 and 0.07%, respectively. Despite the modest decrease in the annual averages of  $SP_{MAX}$  ( $KSS_{MAX}$ ) after applying the mitigations, their effectiveness and benefits to reduce higher scores are evident via the inspection of the upper (lower) right panels of Fig. 4. The blue histograms show a significant reduction of the higher values of the distributions both for  $SP_{MAX}$  (upper) and  $KSS_{MAX}$  (lower), when compared with the non-mitigated scenario (upper and lower red histograms, respectively). The combined thresholds for  $DT$ ,  $N_{CREW}$  and  $N_{CREW} + N_{SIT}$  are exceeded in only 6.7% of the total 3,610,145 loops analysed, thus demonstrating the applicability of the mitigations with minimal impact on rosters production.

## Limitations of the study

Despite that this study includes several rosters from the same individual, the self-rated SP and KSS assessments were made just once in time. Consequently, the individual variations along time and with respect to different roster's characteristics could not be captured. Such cross-sectional method limits the statistical robustness of the sample when associating workload metrics with fatigue and sleepiness outcomes (sample 2). In fact, considering the multinomial character of both SP and KSS scores, one would expect sigmoid shapes between these data and the workload metrics, instead of linear relationships, which are not bounded within asymptotic maximum/minimum values. On the other hand, the limitations related with the linear assumptions were at some extent taken into account in the uncertainty ranges of the fitted functions, which depend on the variance matrix of the data. In this regard, it is still feasible and very insightful to extrapolate the results found for sample 2 for the analysis of all 7149 rosters of 2023. Ideally, longitudinal experiments with objective sleep measurements and multiple fatigue and sleepiness assessments for the same individual—as the one proposed recently<sup>25</sup>—are very welcome to capture individual random effects and provide a better understanding of the complex relationships between workload and fatigue or sleepiness. The association of pilot errors in the cockpit with self-rated KSS scores and the inverse of SAFTE-FAST effectiveness (see lower panel of Fig. 2) should be interpreted with caution. Uncertainties in the data (error bars)—particularly during the dawn—limit a more rigorous test (typically  $\sim 10\%$  for pilot errors,  $\sim 9\%$  for KSS and  $\sim 4\%$  for  $1/E_{SF}$ ). Specifically for perceived sleepiness, a dedicated experimental design with multiple responses from the same individual, focused only on shift days, would provide more robust statistics and allow more granular clock-time intervals.

Albeit being a useful approach to estimate cumulative self-rated SP and KSS scores based on roster's workload metrics, the results herein obtained might have been affected by the COVID-19 outbreak, mostly in 2021 and 2022. In addition, the proposed task-related workload metrics do not exhaust the cognitive demands in the complex aviation environment that also depends on weather conditions, air traffic, abnormal airplane configurations, flight disruptions, to name a few.

Other minor limitation of the study is related with the non-inclusion of home standby duties, which, as mentioned previously<sup>3</sup>, are indistinguishable from days off in our analyses.

## Safety recommendations

This section summarizes our main findings and the associated safety recommendations applicable for non-augmented minimum crew passenger flights dedicated to short/medium haul operations in Brazil. We also provide insights for best practices for the use of bio-mathematical models, as well as a novel data-driven fatigue risk management approach.

### Mitigations in the RBAC 117 rules

Considering our predictions for cumulative fatigue and sleepiness, it is recommended that the following limitations are considered in planned and executed rosters:

1. Total Duty Time (*DT*): The total duty time, excluding home standby duties, should be limited to 44 h for each 168-h period in the roster. This limit should not apply for long-haul international flights.
2. Flight Sectors ( $N_{CREW}$ ): The number of flight sectors should be less than 16 for each 168-h period of the roster.
3. Flight Sectors plus Sit Times longer than 1 h ( $N_{CREW} + N_{SIT}$ ): The number of flight sectors plus the number of sit times longer than 1 h should be less than 20 for each 168-h period in the roster.

### Best practices of using bio-mathematical models

According to ICAO DOC 9966<sup>1</sup>, bio-mathematical models do not constitute an FRMS on their own, but are only one tool of many that may be used within an FRMS". Additionally, ICAO brings an implication for States, which "should not rely on bio-mathematical models as the sole means of evaluating the effectiveness of a Service Provider's FRMS". In this regard, bio-mathematical models should not be used as the sole information for "GO" decisions in the following scenarios (given the likelihood of high false negative results for fatigue):

1. Decisions that allow more flexible prescriptive limits.
2. Decisions that disqualify a fatigue report from aircrew.
3. Decisions that discard the likelihood of fatigue in rosters with relevant workload.

### Data-driven fatigue risk management approach

For the purpose of managing workload, fatigue and sleepiness, Operators and Scheduling Personnel should:

1. consider the workload metrics of *DT*,  $N_{CREW}$  and  $N_{CREW} + N_{SIT}$  as Key Performance Indicators (KPI) during the rostering optimization processes. These parameters should be kept as low as reasonably achievable for every 168-h period for each individual.
2. consider limiting the total duty time to 44 hours or less, the total number of flight sectors to less than 16 and the sum of flight sectors and sit periods longer than one hour to less than 20 at each 168-h loop during the rostering optimization processes.
3. evaluate the benefits of fatigue risk mitigations in *DT*,  $N_{CREW}$  and  $N_{CREW} + N_{SIT}$ —computed at every 168-h period—via the analysis of their relationships with the available operational indicators, such as, exceedances in Flight Data Monitoring, hard landings, missed approaches, as well as medical records, among others.

## Conclusions

This work proposes a novel approach to investigate the complex relationships between workload metrics and self-rated fatigue (Samn-Perelli, SP) and sleepiness (Karolinska, KSS) scores. The study also addresses the limitations of bio-mathematical models that may provide high false negative results that do not accurately reflect the perceptions of fatigue and sleepiness, which are most likely associated with roster's characteristics, in addition to the main model ingredients related with the homeostatic sleep process and circadian rhythms.

Conversely, after fitting a single multiplicative constant, the time-of-day averages of both KSS and the inverse of the SAFTE-FAST effectiveness fit the relative risk of pilot errors in the cockpit<sup>31</sup>, indicating that these metrics can serve as interesting proxies for risk assessment.

The linear relationships found between workload metrics and SP (or KSS) scores provided a consistent method to estimate peak values of accumulated fatigue and sleepiness for every 168-h loop. The risk ratios for high fatigue ( $SP \geq 6$ ) or sleepiness ( $KSS \geq 7$ ) comparing groups  $G_+$  and  $G_-$  (see Table 4) are statistically significant for the total duty time ( $DT$ ) both for SP and KSS, and for the number of flight sectors ( $N_{CREW}$ ) and the sum of flight sectors with sit times longer than one hour ( $N_{CREW} + N_{SIT}$ ) with KSS. These workload variables are also associated with 70.1% (60.6%) of the highest predicted scores of SP (KSS) considering all rosters of 2023, representing key metrics for safety recommendations either for the regulatory framework, as well as for fatigue risk management approaches. Applying the mitigations  $DT \leq 44h$ ,  $N_{CREW} \leq 15$  and  $N_{CREW} + N_{SIT} \leq 19$  for every 168-h interval yields a significant decrease in the higher values of SP and KSS with minimal impact (6.7% of all loops of 2023) on roster coverage.

Further longitudinal experiments with objective sleep measurements and multiple fatigue and sleepiness assessments are very welcome to provide more robust results with the possibility of implementing daily analyses and mitigations, in addition to the cumulative (weekly) approach proposed here.

## Data availability

The results presented in Figs. 1, 2, 3 and 4 can be accessed in electronic format upon request to [tulio@if.usp.br](mailto:tulio@if.usp.br).

Received: 7 February 2025; Accepted: 23 September 2025

Published online: 29 October 2025

## References

- International Civil Aviation Organization. *Doc 9966, Manual for the Oversight of Fatigue Management Approaches*, 2 edn. (International Civil Aviation Organization, 2016).
- European Union Aviation Safety Agency. *Effectiveness of Flight Time Limitation (FTL)*. (European Union Aviation Safety Agency, 2019).
- Rodrigues, T. E. et al. Modelling the root causes of fatigue and associated risk factors in the Brazilian regular aviation industry. *Saf. Sci.* **157**, 105905 (2023).
- Borbély, A. A. A two process model of sleep regulation. *Hum. Neurobiol.* **1**, 195–204 (1982).
- Roach, G. D., Fletcher, A. & Dawson, D. A model to predict work-related fatigue based on hours of work. *Aviat. Space Environ. Med.* **75**, A61–A69 (2004).
- Borbély, A. A., Daan, S., Wirz-Justice, A. & Deboer, T. The two-process model of sleep regulation: a reappraisal. *J. Sleep Res.* **25**, 131–143 (2016).
- Ramakrishnan, S. et al. A unified model of performance for predicting the effects of sleep and caffeine. *Sleep* **39**, 1827–1841 (2016).
- Folkard, S. & Åkerstedt, T. A three-process model of the regulation of alertness-sleepiness. *Sleep, Arousal, and Performance* 11–26 (1992).
- Jewett, M. E. & Kronauer, R. E. Interactive mathematical models of subjective alertness and cognitive throughput in humans. *J. Biol. Rhythms* **14**, 588–597 (1999).
- Hursh, S. R. et al. Fatigue models for applied research in warfighting. *Aviat. Space Environ. Med.* **75**, A44–A53 (2004).
- Åkerstedt, T., Folkard, S. & Portin, C. Predictions from the three-process model of alertness. *Aviat. Space Environ. Med.* **75**, A75–A83 (2004).
- Belyavin, A. J. & Spencer, M. B. Modeling performance and alertness: the qinetiq approach. *Aviat. Space Environ. Med.* **75**, A93–A103 (2004).
- Moore-Ede, M. et al. Circadian alertness simulator for fatigue risk assessment in transportation: application to reduce frequency and severity of truck accidents. *Aviat. Space Environ. Med.* **75**, A107–A118 (2004).
- Folkard, S., Robertson, K. A. & Spencer, M. B. A fatigue/risk index to assess work schedules. *Somnol.-Schlaforsch. Schlafmed.* **11**, 177–185 (2007).
- Olbert, Å. & Klemets, T. A comprehensive investigation of flight and duty time limitations and their ability to control crew fatigue. *Jeppesen* (2011).
- Raslear, T. G., Hursh, S. R. & Van Dongen, H. P. Predicting cognitive impairment and accident risk. *Prog. Brain Res.* **190**, 155–167 (2011).
- Roma, P. G., Hursh, S. R., Mead, A. M. & Nesthus, T. E. Flight attendant work/rest patterns, alertness, and performance assessment: Field validation of biomathematical fatigue modeling. *Federal Aviation Administration Oklahoma City Ok Civil Aerospace Medical Inst* (2012).
- Rangan, S. & Van Dongen, H. Quantifying fatigue risk in model-based fatigue risk management. *Aviat. Space Environ. Med.* **84**, 155–157 (2013).
- McCauley, P. et al. Dynamic circadian modulation in a biomathematical model for the effects of sleep and sleep loss on waking neurobehavioral performance. *Sleep* **36**, 1987–1997 (2013).
- Ingre, M. et al. Validating and extending the three process model of alertness in airline operations. *PLoS ONE* **9**, e108679 (2014).
- Civil Aviation Safety Authority. *Fatigue Management Resources Guide—Publications and Applications* (Civil Aviation Safety Authority, 2024).
- Agência Nacional de Aviação Civil. *Regulamento Brasileiro de Aviação Civil (RBAC 117): Requisitos para gerenciamento de risco de fadiga humana* (ANAC, 2019).
- Brazil. *Lei nº 13.475, de 28 de agosto de 2017* (2017).
- Rodrigues, T. E. et al. Seasonal variation in fatigue indicators in Brazilian civil aviation crew rosters. *Rev. Bras. Med. Trabalho* **18**, 2 (2020).

25. Sampaio, I. T. A. *Regulação e trabalho em jornadas irregulares: o caso de pilotos brasileiros. Implicações para o trabalho e para a saúde*. Ph.D. thesis, Universidade de São Paulo (2023).
26. Samn, S. W. & Perelli, L. P. *Estimating Aircrew Fatigue: A Technique with Application to Airlift Operations* (USAF School of Aerospace Medicine, Aerospace Medical Division, 1982).
27. Åkerstedt, T. & Gillberg, M. Subjective and objective sleepiness in the active individual. *Int. J. Neurosci.* **52**, 29–37 (1990).
28. Laverghetta, A. et al. A survey of fatigue measures and models. *J. Defense Model. Simul.* **22**, 147–173 (2025).
29. Shapiro, S. S. & Wilk, M. B. An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611. <https://doi.org/10.1093/biomet/52.3-4.591> (1965).
30. Helene, O., Mariano, L. & Guimaraes-Filho, Z. Useful and little-known applications of the least square method and some consequences of covariances. *Nucl. Instrum. Methods Phys. Res. Sect. A* **833**, 82–87 (2016).
31. Mello, M. T. et al. Relationship between Brazilian airline pilot errors and time of day. *Braz. J. Med. Biol. Res.* **41**, 1129–1131 (2008).
32. Åkerstedt, T., Anund, A., Axelsson, J. & Kecklund, G. Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *J. Sleep Res.* **23**, 242–254 (2014).

## Acknowledgements

We would like to thank Dr. Izabela Sampaio for the fruitful scientific discussions and for the kind review of this manuscript, Mr. Denys Sene, from IASERA, for the development and support to the *Fadigômetro* web-based platform, and Mr. Marcelo Brugnera for the development of the *Fadigômetro* App. We also thank the National Commission of Human Fatigue (CNFH), the Aeronautical Accidents Investigation and Prevention Center (CENIPA) and the National Committee for the Prevention of Aeronautical Accidents (CNPAA) for their support and endorsement of the study within the aviation community and all the anonymous crew members who voluntarily participated in the study. Frida Marina Fischer receives a productivity grant, number 306963/2021-3 from CNPq (a Brazilian Federal Research Agency).

## Author contributions

TR: Conceptualization, Methodology, Validation, Formal analysis, Data Curation, Writing—Original Draft and Project Administration. EF: Validation, Data Curation, Writing—Review & Editing. AH: Conceptualization, Methodology, Writing—Review & Editing, Supervision. OH: Methodology, Formal analysis, Writing—Review & Editing, Supervision. EP: Data Curation, Writing—Review & Editing. AS: Writing—Review & Editing. MP: Writing—Review & Editing. FF: Conceptualization, Methodology, Writing—Review & Editing, Supervision.

## Funding

This project is equally funded by the Brazilian Association of Civil Aviation Pilots (ABRAPAC), Gol Aircrew Association (ASAGOL) and LATAM Aircrew Association (ATL).

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-21705-z>.

**Correspondence** and requests for materials should be addressed to T.E.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025