Safety Science 157 (2023) 105905

Contents lists available at ScienceDirect

Safety Science

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

Modelling the root causes of fatigue and associated risk factors in the Brazilian regular aviation industry

Tulio E. Rodrigues [a,b,*], Frida M. Fischer [c], Otaviano Helene [a], Eduardo Antunes [d], Eduardo Furlan [b], Eduardo Morteo [b], Alfredo Menquini [e], João Lisboa [d], Arnaldo Frank [e], Alexandre Simões [f], Karla Papazian [g], André F. Helene [g]

[a] Experimental Physics Department, Physics Institute, University of São Paulo, P.O. Box 66318, CEP 05315-970, São Paulo, Brazil
[b] Technical Board, Gol Aircrew Association (ASAGOL), São Paulo, Brazil
[c] Department of Environmental Health, School of Public Health, University of São Paulo, São Paulo, Brazil
[d] Flight Safety Board, National Aircrew Union (SNA), São Paulo, Brazil
[e] Technical Board, Brazilian Association of Civil Aviation Pilots (ABRAPAC), São Paulo, Brazil
[f] Safety Board, LATAM Aircrew Association (ATL), São Paulo, Brazil
[g] Department of Physiology, Institute of Biosciences, University of São Paulo, São Paulo, Brazil

ARTICLE INFO

Dataset link: Modelling the root causes of fatigue and associated risk factors in the Brazilian regular aviation industry (Original data)

Keywords:
Fatigue
Aviation
Rostering
Modelling
Fatigue risk assessment
Fatigue risk management

ABSTRACT

This work evaluates the potential root causes of fatigue and its relationships with accident risks using a bio-mathematical model approach and a robust sample (N = 8476) of aircrew rosters from the Brazilian regular aviation, extracted from the *Fadigômetro* database. The fatigue outcomes derive from the software Sleep, Activity, Fatigue, and Task Effectiveness Fatigue Avoidance Scheduling Tool (SAFTE-FAST), which considers the homeostatic process, circadian rhythms and the sleep inertia. The analyses include data from January 2019 until March 2020 and show relevant group effects comparing early 2019 and 2020, with the latter presenting lower fatigue outcomes in most cases. The average minimum SAFTE-FAST effectiveness during critical phases of flight (departures and arrivals) decreases cubically with the number of shifts that elapse totally or partially between mid-night and 6 a.m. within a 30-day period (N_{NS}). As a consequence, the relative fatigue risk increases by 23.3% (95% CI, 20.4–26.2%) when increasing N_{NS} from 1 to 13. The average maximum equivalent wakefulness in critical phases also increases cubically with the number of night shifts and exceeds 24 h for rosters with N_{NS} above 10. The average fatigue hazard area in critical phases of flight varies quadratically with the number of departures and arrivals within 2 and 6 a.m. (N_{Wocl}). These findings demonstrate that both N_{NS} and N_{Wocl} represent potential root causes of fatigue and therefore should be considered key performance indicators and kept as low as reasonably practical when building aircrew rosters, in order to properly manage the fatigue risk. The effectiveness scores obtained at 30-minute time intervals allowed a model estimate for the relative fatigue risk as a function of the time of the day, whose averaged values show reasonable qualitative agreement with previous measurements of pilot errors in the cockpit. Moreover, the 2019 data revealed a risk exposure factor two times (14%) higher than the figures reported by de Mello et al. (2008) (7%), within 0h00 do 05h59. Tailored analyses of the SAFTE-FAST inputs for afternoon naps before night shifts, commuting from home to station and vice-versa, and bedtime before early-start shifts were carried out using the responses of a questionnaire. The average fatigue hazard area in critical phases of flight increases by 43 to 63% switching off the afternoon naps, 14 to 21% increasing the commuting from one to two hours and 35 to 54% switching off the advanced bedtime criterion, with significant group effects in all cases ($p < 0.001$), evidencing the need of a better and more accurate understanding of these parameters when modelling fatigue risk factors. The non-linear relationships between SAFTE-FAST outcomes and key roster's parameters, such as N_{NS} and N_{Wocl}, provided a novel statistical approach for risk assessment and led to few safety recommendations to the aviation sector regarding regulatory reviews and fatigue risk management and mitigation policies.

* Corresponding author at: Experimental Physics Department, Physics Institute, University of São Paulo, P.O. Box 66318, CEP 05315-970, São Paulo, Brazil.
E-mail address: tulio@if.usp.br (T.E. Rodrigues).

https://doi.org/10.1016/j.ssci.2022.105905
Received 4 February 2022; Received in revised form 11 August 2022; Accepted 16 August 2022
Available online 15 September 2022

1. Introduction

Until the COVID-19 outbreak since early 2020, the commercial aviation industry has experienced a solid growth during the last decades and it is particularly characterized as a high standard, strongly regulated and safe transport category (International Civil Aviation Organization, 2021; Janic, 2000). As airplane systems became more reliable, special attention has been given for human errors, which, according to a Boeing summary (Boeing, 2001), have contributed with 66% of commercial jet fleet hull-loss accidents between 1992 and 2000. Among all the relevant aspects related with human factors (Kharoufah et al., 2018), the physiological issue of mental fatigue plays a relevant role in a 24/7 society (Caldwell, 2005). In this regard, countries worldwide are making efforts to establish effective barriers to mitigate the fatigue risk either via prescriptive flight and duty time limitations or via the implementation of fatigue risk management systems (International Civil Aviation Organization, 2016). A recent experiment carried out by the European Aviation Safety Agency (EASA) (European Union Aviation Safety Agency, 2019) investigated the effectiveness of some prescriptive rules and scenarios in Europe using three bio-mathematical models: the Sleep, Activity, Fatigue, and Task Effectiveness Fatigue Avoidance Scheduling Tool (SAFTE-FAST) (Hursh et al., 2004), the Boeing Alertness Model (BAM) (Olbert and Klemets, 2011) and the System for Aircrew Fatigue Evaluation (SAFE) (Belyavin and Spencer, 2004). One of their findings revealed that despite of being necessary, the prescriptive rules are not fully sufficient to mitigate the fatigue risks, especially during disruptive and/or night shifts. Moreover, the Working Time Society has recently stated that prescriptive rules ignore biological aspects and become less effective for working activities outside the normal daytime hours (Honn et al., 2019). Such findings reinforce the importance of a solid safety culture in organizations, as well as the need to strengthen airlines' safety management systems (SMS), bridging the gap between the industry needs and the scientific knowledge in the field.

In Brazil, a new set of prescriptive limits and labour clauses was put in place by August 2017 in law 13.475/17 (Brazil, 2017). After this change, the National Civil Aviation Agency (ANAC) established the criteria and requirements for the implementation by the airlines of a fatigue risk management program/system based on international civil aviation standards and recommendations (International Civil Aviation Organization, 2016) via the Brazilian Civil Aviation Regulation (RBAC 117) (Agência Nacional de Aviação Civil, 2019). Such set of rules from RBAC 117 allows some extensions of the prescriptive limits dictated by the law 13.475/17, as far as the operators comply with additional requirements.

Together with the global efforts of managing fatigue risks, guidelines specially developed for the identification of fatigue as a contributing factor in aviation accidents and serious incidents, like the one recently published by the National Commission of Human Fatigue (Comissão Nacional de Fadiga Humana, 2020) (a Brazilian Commission composed by several stakeholders), also improves the overall knowledge of the likely causes of fatigue, driving future preventive measures and improvements for the aviation system. Moreover, the analysis of speech parameters and its correlations with fatigue and sleepiness under operational circumstances is also a good practical example of research being successfully applied to aviation accident investigations (de Vasconcelos et al., 2019).

Several model-based approaches have been developed in the last decades to provide quantitative estimates of fatigue, sleepiness and performance (Hursh et al., 2004; Olbert and Klemets, 2011; Roma et al., 2012; Ingre et al., 2014; Rangan and Van Dongen, 2013; Raslear et al., 2011; Borbély, 1982; Jewett and Kronauer, 1999; Åkerstedt et al., 2004; Belyavin and Spencer, 2004; Roach et al., 2004; Moore-Ede et al., 2004; Borbély et al., 2016). As described in ICAO's DOC 9966 (International Civil Aviation Organization, 2016), bio-mathematical models are in a continuous improvement process and are useful tools to predict relative fatigue levels and hazards related to scheduling. Most of the models consider the sleep modulation driven by the homoeostatic process, usually called S process, and the circadian rhythm oscillations, usually called C process (Borbély, 1982; Åkerstedt et al., 2004; Roach et al., 2004; Moore-Ede et al., 2004; Borbély et al., 2016). More sophisticated models also take into account the effects of sleep inertia, a phenomenon that can reduce alertness usually after awakening (Hursh et al., 2004; Olbert and Klemets, 2011; Belyavin and Spencer, 2004; Ingre et al., 2014; Jewett and Kronauer, 1999). More recently, Cochrane et al. also emphasized the importance of considering non-linear relationships between fatigue and risk (Cochrane et al., 2021), whereas Flynn-Evans et al. successfully described average changes in performance in a simulated space mission with sleep restrictions (Flynn-Evans et al., 2020). Reviews of the basic features of some of the available bio-mathematical models usually adopted in the aviation industry can be found elsewhere (Mallis et al., 2004; Van Dongen, 2004; Civil Aviation Safety Authority, 2014).

In Brazil, a study carried out in 2012 found strong evidences of a chronic fatigue scenario by the analysis of pilot reports using the SAFTE-FAST model (Licati et al., 2015). Another study, derived from the continued fatigue monitoring effort named *Fadigômetro*, has shown relevant seasonal variations in fatigue indicators comparing high and low season rosters of 2018 (Rodrigues et al., 2020). This first finding demonstrates the potentialities of the *Fadigômetro* project as a reliable tool for the analysis of the impact of regulatory changes, such as the one in effect since March 2020, when the new rules set by RBAC 117 became effective. Unfortunately, the outbreak of COVID-19 in Brazil by the second half of March 2020 coincided with the regulatory change, postponing an unbiased analysis of its impact until the aviation fully recovers (Instituto Brasileiro de Aviação, 2021).

In this work, we present a detailed statistical analysis, based on the three-process SAFTE-FAST model (Hursh et al., 2004), for the investigation of the potential root causes of fatigue associated with night shifts and flight operations within 2 and 6 a.m. and its corresponding relative fatigue risks in a large sample of rosters of the Brazilian regular aviation, extracted from the *Fadigômetro* database. The SAFTE-FAST model was adopted due to its validation against railroad accidents (Hursh et al., 2006, 2011) and good agreements with objective data on psychomotor vigilance tests (PVT) (Roma et al., 2012; Flynn-Evans et al., 2020) and in-flight sleep (Devine et al., 2022) (see the Methods section). As will be presented later, despite of this choice for SAFTE-FAST, some key performance indicators deeply investigated in this paper are of broad application to drive fatigue risk management strategies when building crew schedules, independently of the operator's bio-mathematical model (see the Discussion section). The basic parameters and criteria of the model calculations are the same as described elsewhere (Rodrigues et al., 2020) and are presented in detail in Appendix A.

2. Methods

2.1. The sample

2.1.1. Aircrew rosters

This work included 8476 executed rosters from January 2019 up to mid-March 2020 of aircrew workers from major Brazilian airlines. The period of analysis was chosen with the aim of capturing fatigue outcomes in a recent past scenario before the COVID-19 outbreak in Brazil (mid-March of 2020).

2.1.2. Demographic profile

A questionnaire, including sociodemographic, behaviour and health questions was filled up by 796 aircrew workers from July 19, 2018 until March 02, 2021, without distinction of sex, race, rank, age or years in the job. Both the rosters and the questionnaire were extracted from the *Fadigômetro* database on March 2, 2021. The eligible aircrew workers pertained to airlines that altogether comprised 92.5% of the Brazilian regular aviation market share in 2019 (Agência Nacional de Aviação Civil, 2020).

2.2. Ethical considerations

The present work is derived from the project "Analysis of Fatigue in Brazilian Civil Aviation" approved by the research ethics committee of the Institute of Biosciences, University of São Paulo (Certificate of Presentation for Ethical Appraisal no. 89058318.7.0000.5464). It was ensured confidentiality to all eligible subjects who voluntarily agreed to participate by approving a digital informed consent form. We declare no commercial, labour duality or conflict of interest with any representative institution involved in the experiment, airline or regulator. Confidentiality was ensured for the airlines whose rosters were analysed.

2.3. Rosters: criteria and filtering

As described elsewhere (Rodrigues et al., 2020) rosters were automatically fed into a web-based application, being digitized and analysed by means of an on-line algorithm. Some filters were implemented to extract the comma-separated values (CSV) files, including an internal identification number (Id), event type (*Crewing* for flights and *Working* for non-crewing tasks), departure/arrival times and locations, crew rank, contractual basis and start/end of the duty times. Differently from our previous work (Rodrigues et al., 2020), which included only *Crewing* events, this study also considers all *Working* events in the rosters, such as standbys, training activities, flying as a passenger for airline purposes (dead-heading flights), etc. These non-crewing events do not directly contribute for risk build-up, but may adversely interfere on the sleep opportunities of the subjects, which, in turn, affects their alertness levels and the overall fatigue and/or sleepiness outcomes. The inclusion of all *Working* events in the analyses increased the occurrences of inconsistencies in the extracted CSV files, most frequently associated with erroneous/spurious clock times for departures and/or arrivals, which caused crashes in the SAFTE-FAST (SF) runs. A crash is a condition in which the software is unable to read the file and no output is provided. Furthermore, few rosters also presented warnings in the SF calculations due to the large majority of events being associated with on ground training activities. Warnings are conditions in which SAFTE-FAST is able to provide an output but its internal logic detects that there may be an error. Among all the eligible rosters, 0.64% to 4.5% had crashes, and 0.2% to 6.3% warnings in the SF model, depending on the month. All rosters with crashes and warnings were excluded from the analyses, resulting in a total of 8476 validated rosters.

Considering that only executed rosters were included (past events), home standbys – which are *Working* activities where the crew member stays on-call at the place of his/her choice – were disregarded in the analyses. In that sense these events are indistinguishable from off days, when individuals can sleep in his/her preferable hours. Such procedure avoids bias from the model standard premises that no sleep event will occur during these working periods, which does not seem very likely in a realistic scenario, particularly for home standbys during night times. On the other hand, it is important to take into consideration that this approximation causes an overestimation of the sleep opportunities since some of the participants might have poor sleep quality or actually no sleep event while in home standbys during night, despite of being, for instance, at home. As a consequence, the model predictions for the overall fatigue scores obtained in this analysis should be interpreted as lower bounds, but still a better approximation than considering a total sleep deprivation during these events. Similarly as described elsewhere (Rodrigues et al., 2020), additional filters of minimum crew and narrow body aircraft were also applied in order to focus on the Brazilian domestic flights and few mid-haul international flights within Caribbean, South and North America executed by minimum crew (see Appendix B). Such choice prevents the inclusion of augmented crew flights, which are characterized by inboard sleep opportunities, a feature not included in our SF input criteria for this analysis. Since some of the key performance metrics, such as the fatigue hazard area (*FHA*), strictly depend on the time interval of the analysis, epochs of exactly 30 days were fixed for each month (see Appendix C).

2.4. Modelling the fatigue risk

Following the steps described in our previous work (Rodrigues et al., 2020), some key performance indicators, such as the SF Effectiveness (E_{SF}), are useful to address the probability of mental fatigue and its relationship with the risk of cognitive impairment, mishaps and, ultimately, serious incidents or accidents. The SF model (Hursh et al., 2004) is a three-step algorithm that takes into account the homeostatic process, the circadian rhythms associated with sleep and wakefulness and the sleep inertia. The particular choice for SAFTE-FAST stems from its validation against human factor railroad accidents (Hursh et al., 2006, 2011), as well as consistent agreements with psychomotor vigilance test (PVT) measurements, both under aviation operational environments (Roma et al., 2012) and in a simulated space mission with sleep restrictions (Flynn-Evans et al., 2020). More recently, the SAFTE-FAST model also predicted, with good accuracy, actual sleep data in humanitarian COVID-19 ultra-long flights (Devine et al., 2022) (see the Discussion section). As we have shown recently, the relative probability of railroad accidents caused by human factors, herein denoted as P_{HF}, increases inversely with E_{SF}, such that:

$$P_{HF}(E_{SF}) = b/E_{SF}, \tag{1}$$

with $b = 79.6 \pm 3.0\%$ and E_{SF} given as a percentage from 0 to 100%, where 100% represents an optimum individual performance (Rodrigues et al., 2020).

During the so-called Window of Circadian Low (WOCL), which is considered as default from 2 to 6 a.m. in the SF model, individuals have a higher probability of being fatigued and/or sleepy, with usually lower performance scores related with alertness and attention. However, in the 24/7 aviation industry, flight operations are still needed during these less favourable hours of the day, thus requiring additional protective barriers to mitigate the fatigue risks. Such barriers could involve management strategies when building aircrew rosters or even regulatory changes, as described later on this paper (see the Discussion section). Even with such mitigation strategies, it is unavoidable that individuals will have decreased alertness and increased sleepiness within the WOCL, which may or may not occur simultaneously with more critical phases of flight, such as departures and arrivals. In this regard, it is more useful and effective to determine the E_{SF} scores during these critical phases, which, in the SAFTE-FAST model, comprise the first and last 30 min of each flight sector (Rodrigues et al., 2020). This is a good strategy to map fatigue outputs that have a higher likelihood to generate adverse consequences in flight operations. Consequently, the most degraded fatigue scores, usually called hot spots of fatigue, for a given crewmember within a period of analysis can be given by the minimum E_{SF} score in the critical phases of flight for Crewing events, herein denoted as EM_C.

Additionally with EM_C, the minimum sleep reservoir (R) in the critical phases of flight, herein denoted as RM_C, also represents a key performance variable strictly related with the sleep debt (SD) and wakefulness (t_{awake}). As described elsewhere (Hursh et al., 2004), the sleep reservoir R varies from 0 to 100%, increasing during sleep periods and decreasing during wakefulness, with a score of 75% representing

8 h of sleep debt. So, considering the equations presented in Ref. Hursh et al. (2004), the following linear relationships hold:

$$SD = 32\left(1 - \frac{R}{100}\right) \tag{2}$$

and

$$t_{awake} = \frac{2880}{0.5 \times 60}\left(1 - \frac{R}{100}\right) = 3 \times SD, \tag{3}$$

with SD and t_{awake} in hours. So, given a minimum sleep reservoir in critical phases of flight one can easily obtain the corresponding maximum sleep debt SD^{max} and the maximum equivalent time awake t_{awake}^{max}.

Differently from EM_C and RM_C, which represent plausible metrics to identify fatigue hot spots in crew rosters, the *FHA* brings the concept of a cumulative fatigue score for a given individual within a given period of analysis. This metric was first proposed by Rangan and Van Dongen in 2013 (Rangan and Van Dongen, 2013) and represents the area of the SF effectiveness lineshape along time under a given threshold. Such additive metric represents an overall quantitative fatigue score, which could help to guide preventive actions to mitigate the fatigue risks in hundreds or even thousands of crew rosters. Following the same steps described elsewhere (Rodrigues et al., 2020), we have adopted a SF effectiveness threshold of 77% in order to calculate the *FHA* during critical phases of flight, herein denoted as FHA_C.

The SF parameters and criteria used in our model calculations are described in detail in Appendix A and are strictly the same adopted in our previous work (Rodrigues et al., 2020). However, in this work, we have also investigated – using behavioural information from the questionnaire – three important SF inputs: (1) afternoon naps prior to night shifts, (2) commuting from home to station and vice-versa, and (3) the advanced bedtime feature of the software. For the afternoon naps prior to night shifts, the standard parametrization of the software considers no nap for the individuals with less than 8 h since the last sleep event and a 60-, 90-, 120- and 180-min nap if the individual is within 8 to 10 h, 10 to 12 h, 12 to 14 h or more than 14 h since the last sleep event, respectively. For this input a tailored analysis was carried out switching off the Auto-Nap feature, which means the software will not include afternoon naps regardless of the wakefulness period, for those individuals who declared not being used to take any nap before night shifts [364 out of 796 responders (45.6%)]. For the commuting from home to station and vice-versa, we have also run analyses with two hours (*extended* commuting), in contrast with our *standard* parametrization of one hour. For this input, we have considered the individuals that declared a commuting of two hours [200 out of 796 responders (25.1%)]. For the advanced bedtime feature, the software assumes that individuals go to bed earlier than usual (considering the standard bedtime of 11 p.m. Rodrigues et al., 2020), as a sleep strategy before early-start shifts, typically between 6 and 8 a.m., regardless of having or not any significant sleep deprivation prior to the sleep event. So, in our calculations, we have investigated a scenario without the advanced bedtime feature of the SF model for the individuals who reported not doing any anticipation of the bedtime [262 out of 796 of the responders (31.9%)]. All these customized runs were performed for two high (February and July) and two low (May and June) productivity months of 2019.

2.5. Variables, statistical analyses and fitting procedures

The dependent variables include: EM_C, RM_C, FHA_C, SD^{max}, t_{awake}^{max} and P_{HF}. The independent variables include: time of the day, t_{clock}; duty time, DT; number of night shifts, N_{NS}; number of consecutive night shifts, N_{CNS}; number of *Working* events, N_{work}; number of *Crewing* events (flight sectors), N_{crew}; and number of WOCL events, N_{wocl}. Both N_{NS} and N_{CNS} include *Crewing* and *Working* events where any portion of the duty period occurs between mid-night and 6 a.m. N_{wocl} includes all the departures and arrivals within 2 and 6 a.m. for *Crewing* events only. The independent variables DT, N_{NS}, N_{CNS}, N_{work}, N_{crew} and N_{wocl} are closely related with workload and will be denoted throughout the paper as *productivity metrics*. Among all these important metrics, N_{NS} and N_{wocl} are two strong candidates for potential root causes of fatigue, as will be addressed later on this work. All the analyses were done for exact 30-day epochs to avoid any bias when comparing months of different lengths (see Appendix C for details). Another set of dependent variables derived from our calculations include the relative fatigue risk as a function of N_{NS}, $RFR(N_{NS})$; the SF effectiveness as a function of t_{clock}, $E_{SF}(t_{clock})$; the relative fatigue risk as a function of t_{clock}, $RFR(t_{clock})$; and the flight proportion as a function of t_{clock}, $FP(t_{clock})$.

For the normality hypothesis we have adopted the Shapiro–Wilk test (Shapiro and Wilk, 1965). For the evaluation of group effects we applied the Mann–Whitney test for two independent samples and the Wilcoxon signed-rank test for paired samples, depending on the situation. All the statistical tests were performed with the IBM SPSS software version 25. The calculations of the effect size (dz) were performed by G*Power version 3.1.9.7 (Faul et al., 2007).

All the fitting procedures were carried out using the Least Square Method (LSM) (Helene et al., 2016), where the best fit parameters correspond to the minimum χ^2, defined as:

$$\chi^2 = \sum_{i=1}^{n} \frac{[\tilde{f}(x_i) - y_i]^2}{\sigma y_i^2}, \tag{4}$$

where $\tilde{f}(x_i)$ stands for the fitted function calculated at each x_i value, y_i the corresponding data point, σy_i its respective standard error and n the total number of data points to be fitted. Given that all the data presented throughout this paper depend linearly on the parameters, the optimal properties of the LSM of minimum variance and unbiased fitting are fully satisfied. All the fits are considered successful if the probability of exceeding the χ^2 (p-value) is ≥ 0.05.

The uncertainties of the fitted functions $\sigma_{\tilde{f}}$ where obtained by the propagation of the uncertainties of the fitted parameters taking into account its full covariance matrix, as similarly described in a recent calculation applied for the COVID-19 pandemic spread (Rodrigues and Helene, 2020). The 95% confidence intervals of the fitted functions were assumed as $\sim 2\sigma_{\tilde{f}}$. In some model estimates, standard uncertainty propagation techniques were also applied (Helene et al., 2016).

3. Results

3.1. Sociodemographic parameters of the sample

The sociodemographic parameters from the *Fadigômetro* questionnaire are depicted in Table 1 and include 796 responses, where the number of validated responses for each group differs from the total number of responders, since some of the answers were not provided. Besides the sociodemographic characterization of the sample in terms of gender, rank and age, this work also addressed three questions related with afternoon naps prior to night shifts, commuting from home to station and vice-versa, and bedtime strategies for early-start duties (see Methods).

3.2. SAFTE-FAST outputs and productivity metrics

In this section we present all SAFTE-FAST outputs (software version 4.0.3.207), as well as our *productivity metrics* for 30-day epochs rosters from January 2019 until February of 2020. A 15-day epoch was also used for mid-March of 2019 and 2020. Considering that several filters were adopted to extract the input CSV files (see Methods), the Ids whose rosters were analysed vary from month by month and represent a fraction of the eligible participants of the study. For this reason, the modelling results not necessarily include all the eligible Ids, neither all the Ids that contributed for the questionnaire, but the Ids that fulfilled the filter requirements defined previously.

Table 1
Average ages and respective standard deviations (SD) obtained from the *Fadigômetro* questionnaire. Details in the text.

Groups	Total responders	Age	
		N	Average ± SD (years)
Male	530 (66.6%)	512	40.8 ± 9.5
Female	266 (33.4%)	256	35.6 ± 7.1
Pilots	412 (51.8%)	400	41.8 ± 9.9
Cabin	384 (48.2%)	368	36.0 ± 7.0
Captains	223 (28.0%)	216	47.1 ± 9.2
First Officers	189 (23.7%)	184	35.6 ± 6.5
Total	796	768	39.1 ± 9.1

The results of monthly averages and corresponding standard errors of the SF outputs for EM_C, RM_C, FHA_C are presented in Table 2, together with our estimates of SD^{max} and t^{max}_{awake}, calculated from Eqs. (2) and (3), respectively. The first column of Table 2 shows the total number of validated rosters for each period, which varies between 389 to 680 depending on the month, with a consistent increase since June of 2019. January of 2019 presents the highest fatigue scores with an average EM_C around 71.8%, in contrast with June of 2019, which presented a much higher average of $\sim$ 78.1%. Considering the average minimum sleep reservoir, January of 2019 also presents the lowest score of $\sim$ 75.8%, which is consistent with an average maximum sleep debt of almost 8 h (7.75 h) and an average equivalent maximum time awake of almost 24 h (23.25 h). Considering one standard error, the relative uncertainties of EM_C, RM_C, FHA_C and SD^{max} (or t^{max}_{awake}) vary typically from 0.30 to 0.51%, 0.17 to 0.27%, 5.4 to 9.3% and 0.60 to 0.93%, respectively; thus, evidencing the high precision characteristic of our model estimates.

Table 3 presents the results of our *productivity metrics* extracted by a dedicated filter algorithm for the same group of rosters that generated the SF fatigue outcomes of Table 2. As clearly shown in Table 3, these metrics present large variations from month to month, providing a workload profile for 2019 and early 2020. Once again, Jan-19 presents the highest scores for all metrics, except N_{work}, which is associated with non-crewing and mostly training activities. The typical high season months in South-America of Jan-19 and Jul-19 have the highest scores regarding the average number of night shifts (6.56 ± 0.11 and 6.46 ± 0.10) and the average number of departures and arrivals between 2 and 6 a.m. (4.55 ± 0.15 and 4.51 ± 0.13), respectively. On the other hand, Jan-20, which should also be considered a high season month, presents a lower score for N_{NS} (6.19 ± 0.11), when compared with Aug-19 (6.27 ± 0.10) and Sep-19 (6.23 ± 0.11). These results show that other factors beyond the trivial high/low seasonal variation do play a relevant role in the fatigue outcomes.

Shapiro–Wilk normality tests showed that the SF outputs EM_C, RM_C and FHA_C, as well as the *productivity metrics* N_{NS}, DT, N_{crew} and N_{wocl} are unlikely to be originated from a normal distribution ($p \leq 0.038$ in all cases). For this reason, non-parametric Mann–Whitney tests for independent samples were applied for the investigation of group effects when comparing the early months of 2019 and 2020. The results of the 2020/2019 ratios for the average values of the SF outputs and the *productivity metrics* and the corresponding *p*-values are depicted in Table 4 for 30- (January and February) and 15- (March) day epochs for 2019 and 2020. Except for N_{NS} and N_{wocl} in the comparison between Feb-19&Feb-20 (p = 0.091 and 0.058, respectively) and Mar-19&Mar-20 (p = 0.279 and 0.159, respectively), all the other results show a quantitative decrease in the fatigue scores between early 2019 and 2020. The 2020/2019 ratios for the average values of FHA_C, N_{NS}, DT, N_{crew} and N_{wocl} vary between 0.66 to 0.79, 0.94 to 1.04, 0.89 to 0.94, 0.89 to 0.92 and 0.92 to 0.94, respectively.

3.3. Potential root causes of fatigue

3.3.1. Average minimum SF effectiveness versus N_{NS}

The minimum SF effectiveness during critical phases of flight (EM_C) represents the worst fatigue score (hot spot) for a given subject during a given period of analysis. For this reason, it is likely that crew members with the same amount of night shifts would have similar values of EM_C, since the latter depends on the sleep opportunities of the rosters, which are adversely affected by night shifts. In that sense, crew members with a large number of night shifts are likely to have lower minimum effectiveness scores in critical phases of flight.

The relationship between the average EM_C, herein denoted as $\langle EM_C \rangle$, and N_{NS} is shown in the upper panel of Fig. 1, which includes all 2019 rosters for 30-day epochs with $1 \leq N_{NS} \leq 13$. The average values of $\langle EM_C \rangle$ (data points) are satisfactorily fitted by a third degree polynomial, represented by the solid blue line (χ^2 = 9.00, *d.o.f.* = 9 and p = 0.437). The dashed–dotted red and green lines represent, respectively, the upper and lower limits considering a 95% confidence interval (CI). The fitting was performed using the LSM (Helene et al., 2016) and the confidence intervals were obtained using uncertainty propagation techniques and the covariance matrix of the best-fit parameters, as similarly described elsewhere (Rodrigues and Helene, 2020). It is verified that $\langle EM_C \rangle$ drops more significantly for $1 < N_{NS} \lesssim 5$ and $N_{NS} \gtrsim 10$. The error bars are higher for $N_{NS} \geq 10$, since there are fewer rosters within this range.

3.3.2. Relative fatigue risk versus N_{NS}

Considering that the probability of human factor accidents vary inversely with E_{SF} (see Eq. (1)), one can estimate the relative fatigue risk as a function of N_{NS} by $RFR(x) \cong b/f(x)$, where $f(x)$ stands for the fitted function and $x \equiv N_{NS}$. Under this approximation, one can also calculate the 95% CI's of $RFR(x)$ by the propagation of the uncertainties of $f(x)$ and b. The results of this relative risk estimate and its upper/lower limits are shown by the solid blue and dashed–dotted red/green lines of the centre panel of Fig. 1, respectively. As expected, the relative fatigue risk increases with N_{NS}. Increasing the number of night shifts from 1 to 13 increases the relative risk by 23.3% (95% CI, 20.4–26.2%).

3.3.3. Average maximum equivalent time awake versus N_{NS}

The lower panel of Fig. 1 presents the average maximum equivalent time awake (t^{max}_{awake}) – associated with the average minimum SF sleep reservoir during critical phases of flight – as a function of N_{NS} (data points) and its corresponding standard errors (error bars). Once again, a third degree polynomial fitting (χ^2 = 16.32, *d.o.f.* = 9 and p = 0.060) successfully describes the data. It is also verified that t^{max}_{awake} has a higher slope for $N_{NS} \geq 10$, exceeding more than 24 h of equivalent wakefulness for $N_{NS} \geq 11$.

3.3.4. Fatigue hazard area versus N_{wocl}

The cumulative FHA_C is expected to increase with the number of departures and arrivals within the WOCL period, representing a consistent overall fatigue score for a given subject in a given period of analysis. Fig. 2 shows the average values of FHA_C as a function of N_{wocl} (data points) and their respective standard errors (error bars), for all the 6527 rosters of 2019. At this time, the data are consistently fitted (χ^2 = 16.68, *d.o.f.* = 14 and p = 0.274) by a parabolic function (solid blue line), with the dashed–dotted red/green lines representing the upper/lower limits, respectively, considering a 95% CI. The data point at x = 18.31 ± 0.72 and y = 40.0 ± 3.9 min represents x- and y-average values for all rosters with $16 \leq N_{wocl} \leq 22$. The x-axis error (0.72) corresponds to the standard error of the x-average and was propagated to the y-axis error (3.9 min) in order to perform the fitting procedure. All the fitting results shown in Figs. 1 and 2 are summarized in Table 5.

Table 2
Average values and standard errors for the minimum effectiveness (EM_C), minimum sleep reservoir (RM_C), fatigue hazard area (FHA_C), maximum sleep deficit (SD^{max}) and maximum equivalent time wake (t^{max}_{awake}) during critical phases of flight from January 2019 up to February 2020 in 30-day epochs. Also shown the total number of rosters for each period (N) and the SF results for 15-day epochs (*) for mid-March of 2019 and 2020.

Period	N	$\langle EM_C\rangle$ (%)	$\langle RM_C\rangle$ (%)	$\langle FHA_C\rangle$ (min)	$\langle SD^{max}\rangle$ (h)	$\langle t^{max}_{awake}\rangle$ (h)
Jan-19	419	71.75 ± 0.28	75.78 ± 0.21	8.01 ± 0.48	7.751 ± 0.066	23.25 ± 0.20
Feb-19	435	74.13 ± 0.33	76.90 ± 0.19	5.50 ± 0.36	7.390 ± 0.062	22.17 ± 0.19
Mar-19	404	75.72 ± 0.39	77.95 ± 0.20	4.73 ± 0.38	7.056 ± 0.065	21.17 ± 0.20
Apr-19	389	75.73 ± 0.35	77.92 ± 0.19	4.01 ± 0.32	7.064 ± 0.061	21.19 ± 0.18
May-19	399	76.57 ± 0.34	78.41 ± 0.18	3.30 ± 0.31	6.910 ± 0.057	20.73 ± 0.17
Jun-19	554	78.11 ± 0.35	79.32 ± 0.19	2.85 ± 0.21	6.619 ± 0.059	19.86 ± 0.18
Jul-19	673	74.04 ± 0.26	77.09 ± 0.15	6.02 ± 0.35	7.330 ± 0.049	21.99 ± 0.15
Aug-19	635	74.92 ± 0.25	77.59 ± 0.14	4.37 ± 0.27	7.172 ± 0.045	21.52 ± 0.13
Sep-19	644	74.92 ± 0.28	77.45 ± 0.17	4.79 ± 0.30	7.215 ± 0.053	21.65 ± 0.16
Oct-19	652	75.01 ± 0.28	77.74 ± 0.16	5.04 ± 0.33	7.123 ± 0.051	21.37 ± 0.15
Nov-19	649	76.04 ± 0.29	78.75 ± 0.16	3.92 ± 0.25	6.800 ± 0.051	20.40 ± 0.15
Dec-19	674	74.58 ± 0.25	77.76 ± 0.15	4.79 ± 0.30	7.118 ± 0.049	21.35 ± 0.15
Jan-20	680	73.94 ± 0.22	77.57 ± 0.14	5.29 ± 0.28	7.178 ± 0.043	21.53 ± 0.13
Feb-20	670	75.29 ± 0.26	78.21 ± 0.15	4.32 ± 0.26	6.974 ± 0.047	20.92 ± 0.14
Mar-19 (1/2)*	400	78.45 ± 0.40	79.54 ± 0.19	2.55 ± 0.24	6.547 ± 0.062	19.64 ± 0.19
Mar-20 (1/2)*	599	80.09 ± 0.34	81.06 ± 0.17	1.69 ± 0.15	6.061 ± 0.053	18.18 ± 0.16

Table 3
Average values and standard errors for the number of night shifts (N_{NS}), number of consecutive night shifts (N_{CNS}), duty time (DT), number of *Crewing* events (N_{crew}), number of *Working* events (N_{work}) and number of departures and arrivals within 2 and 6 a.m. (N_{wocl}) from January 2019 until February 2020 in 30-day epochs. Also shown the same *productivity metrics* for 15-day epochs (*) for mid-March of 2019 and 2020.

Period	$\langle N_{NS}\rangle$	$\langle N_{CNS}\rangle$	$\langle DT\rangle(h)$	$\langle N_{crew}\rangle$	$\langle N_{work}\rangle$	$\langle N_{wocl}\rangle$
Jan-19	6.56 ± 0.11	2.072 ± 0.061	116.8 ± 1.2	33.64 ± 0.53	2.73 ± 0.12	4.55 ± 0.15
Feb-19	5.87 ± 0.13	1.795 ± 0.062	110.6 ± 1.5	31.80 ± 0.65	2.98 ± 0.12	4.11 ± 0.15
Mar-19	5.26 ± 0.14	1.554 ± 0.065	105.7 ± 1.6	29.57 ± 0.63	3.68 ± 0.15	3.67 ± 0.16
Apr-19	5.68 ± 0.13	1.617 ± 0.062	110.9 ± 1.5	30.99 ± 0.65	3.92 ± 0.16	3.56 ± 0.15
May-19	5.36 ± 0.13	1.466 ± 0.059	108.1 ± 1.5	30.53 ± 0.65	4.41 ± 0.15	3.25 ± 0.14
Jun-19	4.25 ± 0.13	1.164 ± 0.051	82.5 ± 2.0	23.53 ± 0.68	3.23 ± 0.13	2.64 ± 0.11
Jul-19	6.46 ± 0.10	1.981 ± 0.053	113.1 ± 1.0	32.54 ± 0.52	4.63 ± 0.16	4.51 ± 0.13
Aug-19	6.27 ± 0.10	1.932 ± 0.051	113.9 ± 1.1	32.47 ± 0.51	5.37 ± 0.17	3.95 ± 0.13
Sep-19	6.23 ± 0.11	1.977 ± 0.053	107.1 ± 1.2	30.44 ± 0.52	5.28 ± 0.17	4.08 ± 0.13
Oct-19	6.06 ± 0.11	1.847 ± 0.054	107.3 ± 1.2	30.83 ± 0.53	5.36 ± 0.16	3.99 ± 0.13
Nov-19	5.24 ± 0.11	1.545 ± 0.054	101.8 ± 1.4	27.79 ± 0.54	5.49 ± 0.19	3.21 ± 0.11
Dec-19	5.81 ± 0.10	1.829 ± 0.055	108.2 ± 1.0	29.80 ± 0.46	5.19 ± 0.17	3.61 ± 0.13
Jan-20	6.19 ± 0.11	2.012 ± 0.060	110.0 ± 1.0	30.99 ± 0.49	4.60 ± 0.16	4.28 ± 0.13
Feb-20	5.62 ± 0.10	1.682 ± 0.049	103.7 ± 1.1	28.30 ± 0.48	5.22 ± 0.19	3.79 ± 0.12
Mar-19(1/2)*	2.81 ± 0.09	0.780 ± 0.042	56.19 ± 0.80	15.64 ± 0.33	1.91 ± 0.09	2.03 ± 0.11
Mar-20(1/2)*	2.91 ± 0.07	0.850 ± 0.035	50.23 ± 0.80	14.43 ± 0.32	2.57 ± 0.13	1.88 ± 0.10

Table 4
2020/2019 ratios for $\langle EM_C\rangle$, $\langle RM_C\rangle$, $\langle FHA_C\rangle$, $\langle N_{NS}\rangle$, $\langle DT\rangle$, $\langle N_{crew}\rangle$ and $\langle N_{wocl}\rangle$ for 30-day epochs for early 2019 and 2020 and Mann–Whitney tests for independent samples (p-values). Also shown the comparison of 15-day epochs (*) for March 2019 and 2020.

Variables	Parameters	Groups		
		Jan-19&Jan-20	Feb-19&Feb-20	Mar-19 (1/2)&Mar-20 (1/2)*
EM_C	2020/2019 ratio	1.03	1.02	1.02
	p-value	<0.001	0.001	0.004
RM_C	2020/2019 ratio	1.02	1.02	1.02
	p-value	<0.001	<0.001	<0.001
FHA_C	2020/2019 ratio	0.66	0.79	0.66
	p-value	<0.001	0.004	<0.001
N_{NS}	2020/2019 ratio	0.94	0.96	1.04
	p-value	0.037	0.091	0.279
DT	2020/2019 ratio	0.94	0.94	0.89
	p-value	<0.001	<0.001	<0.001
N_{crew}	2020/2019 ratio	0.92	0.89	0.92
	p-value	0.001	<0.001	0.010
N_{wocl}	2020/2019 ratio	0.94	0.92	0.93
	p-value	0.041	0.058	0.159

3.4. Modelling the monthly-averaged fatigue hazard area

The average FHA_C in 30-day epochs presented in Table 2 represents an overall fatigue metric for any given set of rosters. In this regard, a suitable model to estimate $\langle FHA_C\rangle$ for a given N_{wocl} distribution is highly desirable, given its practical relevance to guide airline policies and management strategies for those involved in crew rostering processes. Consequently, the monthly-averaged fatigue hazard area in critical phases of flight can be written as:

$$\langle FHA^j_C\rangle = \sum_{i=0}^{i_{max}} W^j(x_i) f(x_i), \tag{5}$$

where $W^j(x_i)$ represents the normalized N_{wocl} distribution for a given month j, $f(x_i)$ the fitted parabola of Fig. 2, both calculated at each x_i value, with the sum going from zero up to $x_{i_{max}}$, which represents

Table 5
Best-fit parameters and fitting results for $\langle EM_C \rangle$, $\langle t^{max}_{awake} \rangle$ and $\langle FHA_C \rangle$.

Fitting		$\langle EM_C \rangle$	$\langle t^{max}_{awake} \rangle$	$\langle FHA_C \rangle$
	Model	$f(x) = a + bx + cx^2 + dx^3$ $x \equiv N_{NS}$		$f(x) = a + bx + cx^2$ $x \equiv N_{wocl}$
Results	a	87.38 ± 0.54%	14.95 ± 0.25 h	0.246 ± 0.028 min
	b	−4.55 ± 0.31%	2.26 ± 0.17 h	0.468 ± 0.035 min
	c	0.498 ± 0.054%	−0.250 ± 0.032 h	0.1149 ± 0.0054 min
	d	−0.0204 ± 0.0028%	0.0113 ± 0.0018 h	–
	χ^2	9.00	16.32	16.68
	d.o.f.	9	9	14
	p-value	0.437	0.060	0.274

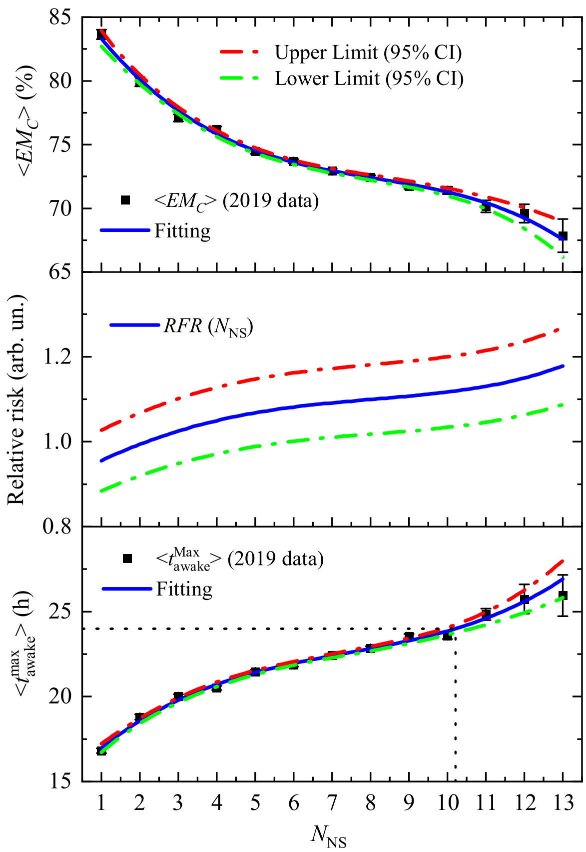

Fig. 1. Upper Panel: Average values and standard errors of EM_C (data points) as a function of N_{NS} and its corresponding third degree polynomial fitting (solid blue line). Centre panel: Relative Fatigue Risk (*RFR*) as a function of N_{NS} (blue line). Lower panel: Average equivalent maximum time awake during critical phases of flight as a function of N_{NS} (data points), its corresponding third-degree polynomial fitting (solid blue line) and a 24 h time awake reference (dotted black line). In all panels the dashed–dotted red and green lines represent the upper and lower limits, respectively, considering a 95% CI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

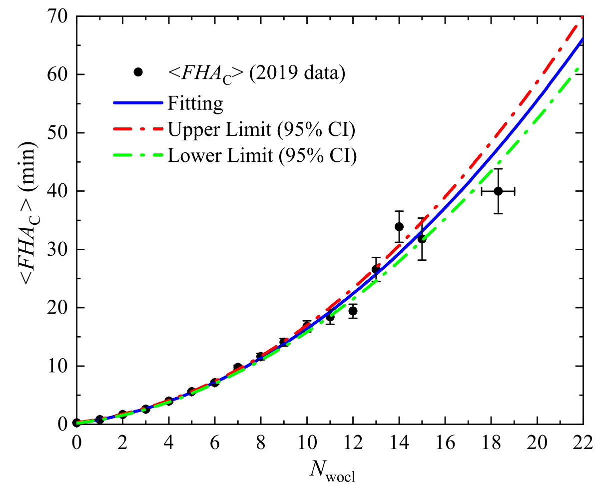

Fig. 2. Average fatigue hazard area in critical phases of flight as a function of N_{wocl} (data points). The solid blue line represents a second-degree polynomial fitting, where the dashed–dotted red and green lines represent, respectively, the upper and lower limits, considering a 95% CI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the maximum N_{wocl} of the distribution. Fig. 3 presents our model estimates for $\langle FHA_C \rangle$ (blue line) with its upper and lower 95% CI given by the dashed–dotted red and green lines, respectively. The insert of Fig. 3 shows, as an example, the normalized N_{wocl} distribution for all the 670 rosters of Feb-2020. The 95% CI of $\langle FHA_C^j \rangle$ was calculated propagating the uncertainties of $f(x_i)$ and $W^j(x_i)$ at each x_i value. The latter is assumed as $\frac{\sqrt{n^{j,i}}}{n^{j,i}} W^j(x_i)$, with $n^{j,i}$ representing the number of events for each x_i value at a given j month. As observed in Fig. 3, the model calculations do reproduce the several structures that appear in $\langle FHA_C \rangle$, with all data falling within the 95% CI, except for Jan-19, which is considerably higher than the model predictions. This latter can be explained considering that January 2019 has a $\langle FHA_C \rangle$ score substantially higher than the average figures of 2019, which were used as a baseline do determine the polynomial best-fit parameters. Both January and February of 2020 are also well reproduced by our model estimates, despite of not being included in the fitting shown in Fig. 2, to avoid any seasonal bias in mixing two months of 2020 with our annual based metric for 2019.

3.5. Circadian variations of fatigue outcomes and associated risk

The SAFTE-FAST software provides effectiveness scores for all the individuals at each 30-min time interval, allowing the verification of circadian oscillations in E_{SF} as a function of the time of the day. The results of this analysis – which includes all the 742 521 30-min crewing events of 2019 – are presented in the upper panel of Fig. 4. The solid black squares represent the average values of E_{SF} for each 30 min bin, considering all effectiveness scores of all individuals during all flights of 2019. The error bars represent the standard errors and range from 0.013 up to 0.056%, showing the high precision of $\langle E_{SF} \rangle$. The dashed–dotted black line is an interpolated curve only to guide the eyes, whereas the dashed–dotted magenta line corresponds to $\langle E_{SF} \rangle = 79\%$, which occurs at 02$h$04 and 06$h$03, considering the Brazilian legal time. Such time interval encompasses the worst fatigue outcomes for all the 30-min crewing assessments, reinforcing our choice to establish the WOCL events within 02h00 and 06h00. As verified, the average effectiveness varies quite significantly as a function of t_{clock} and drops below 90% between 11:45 p.m. and 9:15 a.m. The lower panel of Fig. 4 shows our model estimate for the relative fatigue risk as a function of t_{clock} (solid blue line) calculated as $RFR(t_{clock}) \sim P_{HF}(t_{clock}) \cong b / \langle E_{SF}(t_{clock}) \rangle$. The dashed–dotted red and green lines

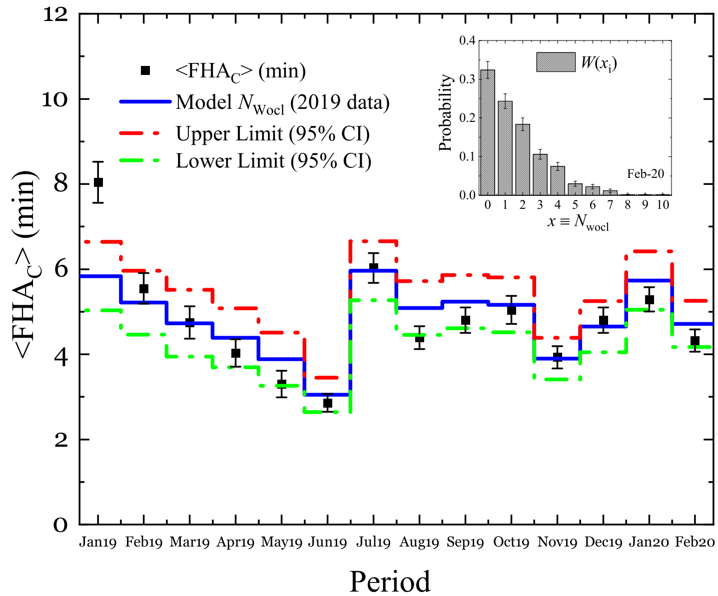

Fig. 3. $\langle FHA_C \rangle$ and its corresponding standard errors for each 30-day epochs (data points) and the model predictions based on the N_{wocl} distributions of each period (solid blue line). The dashed–dotted red and green lines represent the upper and lower limits, respectively, considering a 95% CI. The insert presents the normalized N_{Wocl} distribution for February 2020. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

represent, respectively, the upper and lower limits considering a 95% CI and were obtained propagating the uncertainty of b. The uncertainty of $\langle E_{SF} \rangle$ in the *RFR* was not taken into account, given its negligible values ($\leq 0.056\%$).

In order to compare our model predictions for the relative fatigue risk ratios as a function of the time of the day with previous measurements of pilot errors in the cockpit (de Mello et al., 2008) we firstly averaged the continuous function $RFR(t_{clock})$ within the same time intervals investigated in Ref. de Mello et al. (2008). Secondly, we normalized our results by equalling our lowest *RFR* average (within 18:00 and 23:59) to unit. The results of these procedures are presented by the solid blue histogram in the upper panel of Fig. 5, together with the upper (dashed–dotted red) and lower (dashed–dotted green) limits considering a 95% CI and the normalized risk ratios found by de Mello et al. (2008) (data points). For the latter, we have also assumed an error bar proportional to the ratio of $\frac{\sqrt{N}}{N}$, where N stands for the absolute number of errors within a given time interval. Such approximation holds if the probability of pilot errors follows a Poisson distribution. As verified, our estimates agree qualitatively with the objective measurements performed in Ref. de Mello et al. (2008) (see the Limitations section). The lower panel of Fig. 5 shows the proportion of flights as a function of the time of the day reported by de Mello et al. (2008) (dashed–dotted black histogram) in comparison with our estimates for the proportion of events using all 2019 data (dashed–dotted magenta histogram). As shown in the lower panel of Fig. 5, the risk exposure presents a significant increase (from 7 to 14%) within 0h00 and 5:59, comparing the flight proportion found in 2005 (when the data of Ref. de Mello et al. (2008) were collected) with 2019.

All the results presented in Figs. 1 through 5 are available in electronic format (see Data availability section).

3.6. Tailored analyses of SAFTE-FAST inputs

In this section we demonstrate quantitative variations in some SAFTE-FAST outputs when tailoring some key input metrics related to afternoon naps, commuting, and bedtime constraints.

For the afternoon naps, the SAFTE-FAST console has a standard input parametrization (Auto-Nap) that assumes a nap before a night duty, which depends on the hours of sustained wakefulness since the last sleep event (see Appendix A). This particular input can be switched

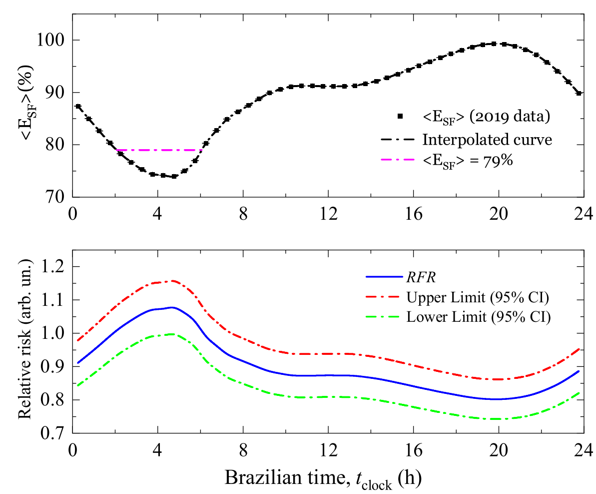

Fig. 4. Upper panel: Average SF effectiveness scores (black squares) and its standard errors (error bars) for all the 30-min assessments during all crewing events of 2019. The dashed–dotted black line is only to guide the eyes and the dashed–dotted magenta line corresponds to $\langle E_{SF} \rangle = 79\%$. Lower panel: Model estimates for the *RFR* as a function of the time of the day (solid blue line) and its upper (dashed–dotted red) and lower (dashed–dotted green) limits considering a 95% CI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

off, which means that no afternoon nap will be added automatically, regardless of the time period since the last sleep event (Auto-Nap OFF).

For the commuting from home to base station and vice-versa, we have adopted a *standard* metric of one hour and also an *extended* commuting profile of 2 h. So, for check-in purposes, we consider one hour of preparation at home, hotel or rest facility and one (*standard*) or two (*extended*) hours of commuting from home to station and vice-versa.

For the bedtime parameter, the SAFTE-FAST algorithm adopts 11 p.m. as default, which is also adopted in our calculations. However, the model also has an advance bedtime function, i.e., a feature that assumes an individual will go to bed earlier than usual should an early start shift is scheduled in the following morning, regardless of the sleep deficit accumulated in previous shifts. This feature can also be switched

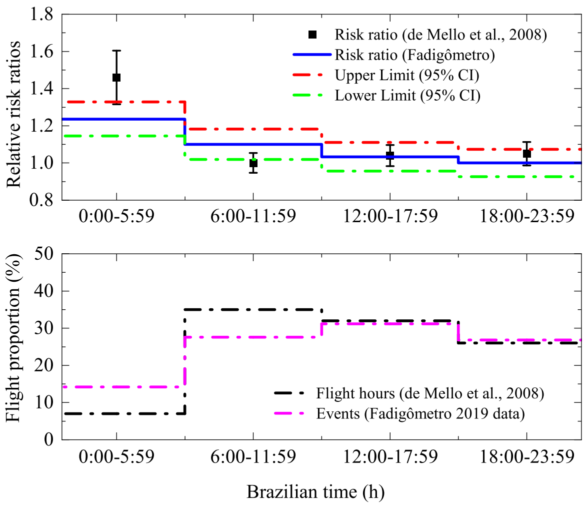

Fig. 5. Upper panel: Relative risk ratios from de Mello et al. 2008 (data points) in comparison with the *Fadigômetro* predictions (solid blue histogram) and its upper (dashed–dotted red) and lower (dashed–dotted green) limits considering a 95% CI. Lower panel: Flight proportion (%) reported by de Mello et al. 2008 (dashed–dotted black histogram) versus the fraction of events from the Fadigômetro data (dashed–dotted magenta histogram). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

off, which means that the standard bedtime of 11 p.m. will be a fixed constraint for all main sleep events during night.

The quantitative effects in few SAFTE-FAST outputs related with the variations of these input criteria are shown in Table 6, considering two months with low (May and June of 2019) and high (February and July of 2019) productivity profiles. The calculations were done using the fractions of the monthly rosters related with the Ids who declared that are not used to get afternoon naps prior to night shifts (within 42.2 to 45.4% of the responders), two hours of commuting from home to station and vice-versa (22.6 to 25.3% of the responders), or not used to advance the bedtime before early-starts (32.1 to 33.3% of the responders). Wilcoxon signed-rank tests for matched pairs show significant group effects ($p < 0.001$) for EM_C, RM_C and FHA_C comparing the groups with and without the afternoon nap, with one or two hours of commuting and with or without the advanced bedtime feature. Pearson's correlations (effect sizes), with and without the Auto-Nap function enabled, are equal or higher than 0.967 (0.763), 0.926 (0.673), and 0.941 (0.545) for EM_C, RM_C and FHA_C, respectively. For the comparison between one and two hours of commuting, Pearson's correlations (effect sizes) are equal or higher than 0.970 (0.470), 0.912 (0.527) and 0.921 (0.304) for EM_C, RM_C and FHA_C, respectively. For the groups with and without the advanced bedtime feature, Pearson's correlations (effect sizes) are equal or higher than 0.970 (0.716), 0.926 (0.849) and 0.890 (0.422) for EM_C, RM_C and FHA_C, respectively. The average fatigue hazard area in critical phases of flight increases by 43 to 63% when switching off the afternoon naps, 14 to 21% when increasing the commuting time from home to station and vice-versa from one to two hours and 35 to 54% when switching off the advanced bedtime criterion.

4. Discussion

The SAFTE-FAST model outputs and most of the *productivity metrics* showed a consistent reduction in the comparison between early 2019 and 2020, as depicted in Table 4. However, such finding does not allow an unambiguous conclusion about possible improvements of fatigue management policies from the operators, given that the *productivity metrics* of *DT* and N_{crew} are also lower ($p \leq 0.010$) for early 2020. So, it is likely that this workload reduction reflected positively in the model outputs of EM_C, RM_C and FHA_C ($p \leq 0.004$ in all cases), but did not change quite substantially the root causes of fatigue given by N_{NS} and N_{wocl} for February ($p = 0.091$ and 0.058) and mid-March ($p = 0.279$ and 0.159), respectively. These results show that the SF outputs are quite sensitive to minor changes in N_{NS} and N_{wocl}. For instance, for January 2020, the average value of N_{wocl} is 6% lower than for 2019, but the average FHA_C is 34% lower. For mid-March 2020, the average N_{wocl} is 7% lower than the 2019 average, but without a significant group effect ($p = 0.159$). On the other hand, the average FHA_C dropped 34% in the same pair of samples ($p < 0.001$).

The fatigue indicators of EM_C and t^{max}_{awake} show non-linear relationships with the *productivity metric* of the number of night shifts N_{NS} (see Table 5), which has an adverse effect on the flight schedules and drives the worst fatigue scores. Exceeding 10 night shifts in a 30-day time interval causes an average maximum equivalent time awake higher than 24 h, or, equivalently, an average maximum sleep deficit of more than 8 h. Indeed, Lamond and Dawson have found (Lamond and Dawson, 1999) that 20–25 h of continuous wakefulness can be associated with decrements in performance scores related with reasoning and attention, and that these degradations could be comparable for individuals with a blood alcohol concentration of 0.10%. The adverse effects of long periods of wakefulness were also pointed as a probable cause for aviation accidents. The National Transportation Safety Board (NTSB) concluded that fatigue was a probable cause for the accident of American International Airways flight 808 in Guantanamo Bay on August 18, 1993 (National Transportation Safety Board, 1993). The NTSB final report determined that "...the probable causes of this accident were the impaired judgement, decision-making, and flying abilities of the Captain and flight crew due to the effects of fatigue...". In fact, the analyses of the sleep/wake periods revealed that the Captain had been awake for 23.5 h within the 28.5 h prior to the accident (National Transportation Safety Board, 1993; Rosekind et al., 1993). Consequently, our finding supports the recommendation that flight schedules should be planned with the lowest achievable number of night shifts, and not exceeding $N_{NS} = 10$ within any 30-day time interval. In fact, the Brazilian regulations (Brazil, 2017; Agência Nacional de Aviação Civil, 2019) allow up to 4 night shifts within a 168-h period, which allows the possibility of an aircrew to have up to 16 night shifts within a four-week period. In this regard, it is highly recommended the revision of current regulations concerning non-augmented passenger flight operations, as summarized in the Safety recommendations section.

The FHA_C also presents a non-linear relationship with the number of departures and arrivals within 2 and 6 a.m., N_{wocl} (see Table 5). This result shows that cumulative fatigue builds up quadratically with the number of WOCL operations, reinforcing its adverse impact on the overall fatigue score for a given individual within a given time interval. Our model estimate for the monthly averaged FHA_C for all the 30-day epochs of 2019 reproduces quite reasonably, except for January 2019, several structures that appear in the data (see Fig. 3), which present huge variations month by month. Consequently, our model approach allows the calculation of $\langle FHA_C \rangle$ for any given N_{wocl} distribution, representing a suitable method for fatigue risk assessment. For the specific case of January 2019, however, our model is not able to reproduce the datum, which is significantly above the average figures found in 2019.

The N_{wocl} distributions (see the insert of Fig. 3 for a typical distribution of February 2020) drive the cumulative fatigue in rosters and should be concentrated within the lowest possible N_{wocl} values. In fact, among all 2019 rosters, only 0.4% present $16 \leq N_{wocl} \leq 22$ and are associated with an average FHA_C of 40.0 ± 3.9 min (see Fig. 2). Consequently, it is highly recommended that scores of $N_{wocl} > 15$ are avoided in rosters, given its significant impact on overall fatigue and its negligible effect on crew productivity. Furthermore, only 3.6% of the rosters had more than 10 WOCL operations within a 30-day time interval, giving room for safety improvements with minimal operational

Table 6
Average values and standard deviations (SD) for EM_C, RM_C and FHA_C with or without the Auto-Nap function, with one or two hours of commuting and with or without the Advanced bedtime (BT) feature for February, May, June and July of 2019. Also shown the total number of responders (N) for each period and input, the corresponding fraction of responders (%), as well as the *p*-values, the Pearson's correlations, the effect sizes (dz) and the ratios of $\langle FHA_C \rangle$ for each matched pair of the analyses.

SF Input			EM_C (%)				RM_C (%)				FHA_C (min)			
			Feb-19	May-19	Jun-19	Jul-19	Feb-19	May-19	Jun-19	Jul-19	Feb-19	May-19	Jun-19	Jul-19
Auto-Nap	ON	Average	73.76	76.54	78.29	74.07	76.92	78.35	79.44	77.19	5.52	3.27	2.67	5.64
		SD	7.33	6.60	8.28	6.49	4.22	3.52	4.48	3.70	6.50	6.68	4.69	6.85
	OFF	Average	72.15	75.22	76.74	72.16	75.80	77.45	78.32	75.83	8.20	4.69	4.35	8.95
		SD	8.07	7.24	9.24	7.35	4.58	3.65	5.02	4.10	8.86	7.91	7.11	9.73
	N		194	181	234	288	194	181	234	288	194	181	234	288
	Fraction (%)		44.6	45.4	42.2	42.8	44.6	45.4	42.2	42.8	44.6	45.4	42.2	42.8
	p-value [a]			<0.001				<0.001				<0.001		
	Pearson's ρ		0.967	0.976	0.986	0.976	0.940	0.931	0.956	0.926	0.960	0.950	0.957	0.941
	Effect size, dz [b]		0.763	0.803	0.886	1.097	0.716	0.673	0.743	0.877	0.840	0.545	0.569	0.823
	Ratio OFF/ON										1.48	1.43	1.63	1.59
Commuting	1 h	Average	73.90	76.48	78.13	74.30	77.11	78.21	79.38	77.52	4.57	4.07	2.66	5.35
		SD	6.20	7.63	8.05	6.90	3.70	3.96	4.15	3.83	4.93	6.22	4.70	6.60
	2 h	Average	73.11	75.62	77.21	73.51	76.17	77.33	78.44	76.58	5.51	4.64	3.04	6.36
		SD	5.95	7.24	7.84	6.72	3.65	3.96	4.32	3.93	5.75	7.16	5.15	7.61
	N		99	101	132	152	99	101	132	152	99	101	132	152
	Fraction (%)		22.8	25.3	23.8	22.6	22.8	25.3	23.8	22.6	22.8	25.3	23.8	22.6
	p-value [a]			<0.001				<0.001				<0.001		
	Pearson's ρ		0.973	0.980	0.970	0.973	0.929	0.941	0.912	0.921	0.921	0.983	0.972	0.959
	Effect size, dz [b]		0.551	0.560	0.470	0.496	0.678	0.647	0.527	0.608	0.414	0.368	0.304	0.446
	Ratio 2 h/1 h										1.21	1.14	1.14	1.19
Advanced BT	ON	Average	73.50	76.62	78.57	74.36	76.43	78.17	79.35	77.20	6.45	3.68	2.81	6.18
		SD	6.90	6.95	7.99	7.22	4.19	3.46	4.39	4.18	9.08	7.61	5.31	9.65
	OFF	Average	72.04	75.24	77.18	73.06	74.88	76.84	78.00	75.85	9.5	4.96	4.31	8.4
		SD	6.76	6.81	8.11	7.09	4.55	3.82	4.83	4.34	12.1	4.96	7.12	12.8
	N		145	128	178	217	145	128	178	217	145	128	178	217
	Fraction (%)		33.3	32.1	32.1	32.2	33.3	32.1	32.1	32.2	33.3	32.1	32.1	32.2
	p-value [a]			<0.001				<0.001				<0.001		
	Pearson's ρ		0.971	0.970	0.971	0.974	0.933	0.926	0.945	0.942	0.894	0.970	0.890	0.930
	Effect size, dz [b]		0.884	0.816	0.716	0.794	0.946	0.921	0.849	0.925	0.537	0.551	0.441	0.422
	Ratio OFF/ON										1.47	1.35	1.54	1.35

[a]Wilcoxon signed-rank tests for matched pairs performed with SPSS version 25.
[b]Effect size dz calculated with G*Power version 3.1.9.7 (Faul et al., 2007).

impact. Both the number of night shifts and the number of WOCL operations, as well as its distributions among aircrew workers, deserve a special attention when building crew rosters. Operators are encouraged to adopt these metrics as key performance indicators to drive management policies and, whenever possible, rostering optimization processes. Under this approach, the core of risk mitigation is not only necessarily related with reducing night shifts and WOCL operations in an absolute manner. Instead, it is related with the dispersion of the distributions. By lowering these dispersions through optimization processes one will find an overall reduced (relative) fatigue risk. The average number of night shifts or WOCL operations per capita do not necessarily need to be reduced, but avoiding extreme situations is certainly a good strategy. This deeper look of different scenarios obtained by additional constraints to the optimization processes is in line with ICAO's statements regarding the usefulness of bio-mathematical models as Fatigue Risk Management tools (International Civil Aviation Organization, 2016). In order to avoid potential cumulative fatigue effects, these optimization processes should consider a dynamical evaluation for any consecutive 30-day time periods. The Brazilian regulations (Brazil, 2017; Agência Nacional de Aviação Civil, 2019) do not limit the number of WOCL operations, indicating the relevance of our findings to support a regulatory review and improve management policies (see the Safety recommendations section).

The contribution and the methods presented herein can also yield benefits for those who have recently implemented or already have well-developed/mature Fatigue Risks Management Systems (FRMS). As there are very few operational initiatives and best practices supported by research to manage fatigue risks in the operational setting, operators could take benefit of the modelling results and recommendations presented in this study by applying them when comparing different scenarios in large-scale optimization processes (International Civil Aviation Organization, 2016). This could also encourage operators to further develop new fatigue metrics as well as to correlate bio-mathematical model outputs, or some of the previously mentioned KPIs, such as N_{NS} and N_{wocl}, with other data derived from Flight Data Analysis Program (FDAP) such as exceedances, hard landings, missed approaches, or with medical data, including absenteeism and number of fatigue calls, to name a few.

The relative fatigue risk as a function of the number of night shifts represents the inverse function of EM_C and N_{NS}, times the parameter b (see the central panel of Fig. 1). Increasing N_{NS} from 1 to 10 increases the relative risk by 16.9% (95% CI, 16.0–17.8%). Additionally, increasing N_{NS} from 10 to 13 increases the risk by 5.5% (95% CI, 3.5–7.5%). Once again, it is quite evident the safety benefit of avoiding more than 10 night shifts within a 30 day's time interval.

The SF effectiveness as a function of the time of the day shows a relevant decrease within 11:45 p.m. and 09:15 a.m., reaching its maximum and minimum scores around 8 p.m. and 4 a.m., respectively. Moreover, the SF effectiveness drops below 79% within the WOCL period (from 2 to 6 a.m.), reinforcing the relevance of effective fatigue risk mitigation policies within these less favourable hours of the day. Among the several fatigue countermeasures deeply discussed in Refs. Caldwell (2005) and Caldwell et al. (2009), controlled rest does represent an effective method to mitigate fatigue. As shown quite extensively in the literature (Caldwell, 2012; Rosekind et al., 1994; Neri et al., 2002; Hartzler, 2014), naps have a consistent positive impact to mitigate the adverse effects of sleep loss and/or circadian disruptions. Unfortunately, the Brazilian regulations (RBAC 117) do not allow the use of controlled rests for minimum crew, which is allowed, for instance, in Australia, Bolivia, Canada, China, Europe, Israel, New Zealand, Turkey, and the United Arab Emirates (Fatigue Countermeasures Working Group, 2018). Another relevant drawback of not having controlled rest allowed by the regulatory framework is the potentially hazardous occurrence of unintentional naps during flight operations. In fact, a recent Brazilian study (Marqueze et al., 2017) found a prevalence of 57.8% of unplanned sleep in a sample of Brazilian pilots, making explicit the need of a regulatory review (see the Safety recommendations section). It is worth-mentioning, however, that the controlled rest should not be adopted with the intent of increasing flight and/or duty time limitations, but exclusively for fatigue mitigation purposes.

The relative fatigue risk as a function of the time of the day is proportional to the inverse of the effectiveness and has a maximum value around 4 a.m. (see the lower panel of Fig. 4). Our averaged values for the *RFR*, normalized to unit within 18h00 and 23:59, agree qualitatively with the objective measurements of pilot errors in the cockpit (de Mello et al., 2008). However, the comparison between these results should be done with caution. Firstly because pilot errors along the time of the day are not exclusively a consequence of fatigue. Secondly because our relative risk estimate is not parametrized as the probability of human errors in a complex aviation environment, but rather by the probability of railroad accidents. In that sense, the pilot errors reported by de Mello et al. (2008) are more closely related with cognitive mishaps, besides several other human factor issues not exclusively related with fatigue (see the Limitations section). Our calculations for the risk exposure, given by the probability of crewing events as a function of the time of the day, are a factor two (14%) higher than the figures reported by de Mello et al. (2008) (7%) within 0h00 and 05h59, showing a relevant change of the Brazilian Commercial aviation flight schedules from 2005 to 2019.

Tailored analyses of the SF inputs related with afternoon naps prior to night shifts, commuting from home to station (and vice-versa) and the advanced bedtime feature of the model were investigated for two low and two high productivity months of 2019 using the responses of the questionnaire. The rosters for the Ids who reported not used to take afternoon naps prior to night shifts, two hours of commuting and not used to advance the bedtime prior to early-starts shifts were run with the standard and the tailored parametrizations, showing relevant group effects ($p < 0.001$) for EM_C, RM_C and FHA_C in Wilcoxon signed-rank tests for matched pairs (see Table 6). The average FHA_C increases by 43 to 63% when discarding the afternoon naps, 14 to 21% when increasing the commuting from one to two hours and 35 to 54% when switching off the advanced bedtime criterion, thus showing the high sensitivity of the SF model to these input parameters.

Differently from the fatigue outcomes herein discussed, which are obviously dependent on the SAFTE-FAST model, the key performance indicators N_{NS} and N_{Wocl} represent independent metrics that can be managed through scheduling policies and strategies. For instance, the BAM model (Olbert and Klemets, 2011) can also be useful to evaluate the potential benefits of reducing the dispersions of N_{Wocl} in its global PA5 metric [see Fig. 1 from Ref. Olbert and Klemets (2011)]. This metric represents the average alertness level for the 5% worst scored flights and its sensitivity (under the analysis of crew rosters, instead of pairings) with the N_{Wocl} distribution could be further investigated. In this regard, the findings and methods herein proposed can be helpful for fatigue risk management, regardless of the operator's choice for any particular bio-mathematical model.

Given that the outbreak of the COVID-19 pandemic in Brazil coincided with the new rules prescribed by the RBAC 117 (March of 2020), the impacts of the new regulatory framework are still unknown, motivating the acquisition of more data to shed light on this issue, as scheduled flights and the commercial aviation industry resume their pre-pandemic levels.

5. Safety recommendations

This section presents a few safety recommendations for minimum crew (non-augmented) passenger flight operations in two main aspects: the regulatory framework and fatigue risk management optimization and best practices.

5.1. Regulatory framework

As for the regulatory aspect, the following parameters and criteria should be revisited:

1. Controlled rest: The use of controlled rest for minimum crew should be thoroughly investigated in further studies, preferably with the collection of objective measures of in-flight sleep (duration and quality) using actigraphy, as well as performance assessments in psychomotor vigilance tests (PVT). After this step, regulatory agencies should consider a review of the regulations, if applicable. However, it is important here to emphasize that the use of controlled rest should not be for the purpose of extending the maximum duty-time limits, but as an additional fatigue countermeasure;
2. Night-shifts: Regulators should consider limiting to a maximum of 10 night-shifts for any consecutive 30-day period for each individual;
3. WOCL operations: Regulators should consider limiting to a maximum of 15 WOCL operations (either departures or arrivals) for any consecutive 30-day time period for each individual.

5.2. Fatigue risk management

As for the fatigue risk management aspect, the following parameters and criteria should be revisited:

1. Operators, managers and scheduling personnel should consider the number of night-shifts (N_{NS}) and the number of WOCL operations (N_{Wocl}) as key performance indicators when building aircrew rosters. These parameters should be kept as low as reasonably achievable during the rostering optimization processes, if applicable, with the lowest possible dispersion between individuals;
2. Operators, managers and scheduling personnel should avoid building rosters with more than 10 WOCL operations within a 30-day time period for the same individual.

6. Limitations of the study

The main limitation of this study is related to the model dependency of all the findings and results. Furthermore, the SF inputs and constraints are set in accordance with subjective assessments from the questionnaire, operational experiences and/or educated guessing. For this reason, objective sleep measurements from actigraphy, for instance, would be highly desirable to provide more accurate estimates for the relevant model inputs and criteria. However, these objective measurements are beyond the scope of this work, which is exclusively dedicated to model analyses.

Other minor limitations are the non-inclusion of home standby duties and unwind periods when analysing rosters. The home standby events were not taken into account in our model calculations given the uncertainties of the expected amount and quality of sleep. This decision was taken to avoid bias, since the SAFTE-FAST model prevents any sleep event during the entire standby activities. This does not seem very realistic during the night time, as most aircrew workers stay at their homes or at an adequate rest facility. The unwind periods, which encompass the elapsed time from the end of commuting (station to hotel, station to home or station to rest facility) up to the start of the rest period, were also not included. These periods may vary considerable from person to person and include personal needs of hygiene, eating, social activities in preparing for sleep. Such limitations make clear that the fatigue outcomes obtained should be interpreted as lower bounds of fatigue, since some of the crew members might have poor or actually no sleep during home standby duties, as well as relevant unwind episodes at home, hotel or rest facility.

Another relevant limitation is related with the extrapolation of the probability of railroad accidents for the aviation scenario. Aviation accidents have a low absolute probability, making it difficult to establish a relationship, for instance, between the SAFTE-FAST effectiveness and human factor accidents with the desirable statistics, as the one obtained in Ref. Rodrigues et al. (2020). Indeed, the investigation of 55 human-factor accidents in aviation (Goode, 2003) demonstrates a relative incidence (accidents proportion per exposure proportion) 5 times higher for duties with 13 h or more, when compared with duties up to 9 h. However, these data do not allow the delineation of statistically relevant relationships between fatigue outcomes and accident risks, since only eleven accidents occurred above 10 h in duty. In this regard, the comparison of our relative risk ratios as a function of the time of the day with objective measurements of pilot errors in the cockpit (de Mello et al., 2008) (see the upper panel of Fig. 5) should be done with caution and under a qualitative approach.

7. Conclusions

This work adopts a modelling approach to provide a comprehensive statistical analysis of the root causes of fatigue in a robust sample of aircrew rosters of the Brazilian regular aviation, derived from the *Fadigômetro* database. The SAFTE-FAST fatigue outputs and some *productivity metrics* delineate an overall fatigue profile for minimum crew, which show a workload decrease comparing early 2019 and 2020. The rosters are fully characterized by non-linear relationships between the SAFTE-FAST variables of minimum effectiveness (and the maximum equivalent time awake) and the number of night shifts, as well as, the fatigue hazard area and the number of departures and arrivals within 2 and 6 a.m. (WOCL period), all considered during the critical phases of flight. The 95% confidence intervals for all the fittings were calculated with the covariance matrix of the fitted parameters and using standard uncertainty propagation techniques. The several structures found for the monthly averaged fatigue hazard areas are consistently interpreted using the distributions of WOCL operations. The relative fatigue risk increases by 23.3% (95% CI, 20.4–26.2%) increasing the number of night shifts from 1 to 13. Moreover, the relative risk ratios as a function of the time of the day agree qualitatively with pilot errors in the cockpit. On the other hand, the risk exposure found in this work (14%) is a factor two higher than the figures reported by de Mello et al. (2008). Tailored analyses of some key SAFTE-FAST inputs were done by switching off afternoon naps prior to night shifts, increasing the commuting from home to station (and vice-versa) from 1 to 2 h and switching off the advanced bedtime criterion of the model, showing significant group effects ($p < 0.001$) for all variables when compared with the standard parametrization. Such finding shows the high sensitivity of the model to these parameters and the need of a deeper investigation to determine more accurately the associated fatigue risk factors. The impact on fatigue caused by the regulatory change with RBAC 117 is still unknown, given the coincidence with the COVID-19 outbreak in Brazil by mid-March 2020. More studies – preferably aggregating objective sleep measures from actigraphy – are very welcome to provide more stringent constraints to the model inputs and criteria. The robustness and accuracy of the methods adopted in this work, including the non-linear relationships between SAFTE-FAST fatigue outcomes and key roster's parameters do support an effective implementation of all safety recommendations, representing a practical contribution for fatigue risk management.

CRediT authorship contribution statement

Tulio E. Rodrigues: Writing – original draft, Validation, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Frida M. Fischer:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Otaviano Helene:** Writing – review & editing, Supervision, Methodology, Formal analysis. **Eduardo Antunes:** Writing – review & editing, Validation, Data curation, Conceptualization. **Eduardo Furlan:** Writing – review & editing, Validation, Data curation. **Eduardo Morteo:** Writing – review & editing, Validation, Data curation, Conceptualization. **Alfredo Menquini:** Writing – review & editing, Validation, Data curation, Conceptualization. **João Lisboa:** Writing – review & editing, Validation, Data curation. **Arnaldo Frank:** Writing – review & editing. **Alexandre Simões:** Writing – review & editing. **Karla Papazian:** Data curation. **André F. Helene:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Data availability

The results presented in Figs. 1 through 5 are available in electronic format at the following link:

Modelling the root causes of fatigue and associated risk factors in the Brazilian regular aviation industry (Original data) (Mendeley Data)

Acknowledgements

We thank Dr. Steven Hursh for the fruitful scientific exchanges and contributions to the manuscript; the staff at Institutes for Behavior Resources (IBR) for the SAFTE-FAST customization for the hazard area calculation; the anonymous reviewers for the relevant comments and suggestions that helped for the improvement of this work; and Mr. Denys Sene, from IASERA, for the development and support to the roster conversion web-based platform. We also thank the National Commission of Human Fatigue (CNFH), the Aeronautical Accidents Investigation and Prevention Center (CENIPA) and Azul Linhas Aéreas Brasileiras for their institutional support in promoting, endorsing and encouraging the study within the aviation community. We finally thank all the crew members who voluntarily agreed to participate in the study.

Conflicts of interest

The authors declare no conflict of interest.

Funding

The development and support of the *Fadigômetro* web-based platform was equally funded by the Brazilian Association of Civil Aviation Pilots (ABRAPAC), Gol Aircrew Association (ASAGOL), LATAM Aircrew Association (ATL) and the National Aircrew Union (SNA). The Article Publication Charge was equaly funded by the Brazilian Association of Civil Aviation Pilots (ABRAPAC), Gol Aircrew Association (ASAGOL), LATAM Aircrew Association (ATL) and Lacourt Fidelity Insurance and Consulting.

Appendix A. SAFTE-FAST parameters and criteria

In this section we describe all the parameters and criteria – adopted to mimic behaviours and operational routines before, during and after the working and crewing activities of the Brazilian civil aviation aircrews (Part 121 Passenger Operations) – for the runs with the SAFTE-FAST software version 4.0.3.207.

For most cases, the start (check-in) and the end (check-out) of the duty periods were captured directly from the rosters. For some rosters format, however, this information were not available and we set 60 min prior to the departure for the check-in and 30 or 45 min after the arrival for the check-out, depending if the flight sector was domestic or international, respectively. These figures follow the usual practice for scheduled flights in Brazil and also comply with current regulations (Brazil, 2017; Agência Nacional de Aviação Civil, 2019).

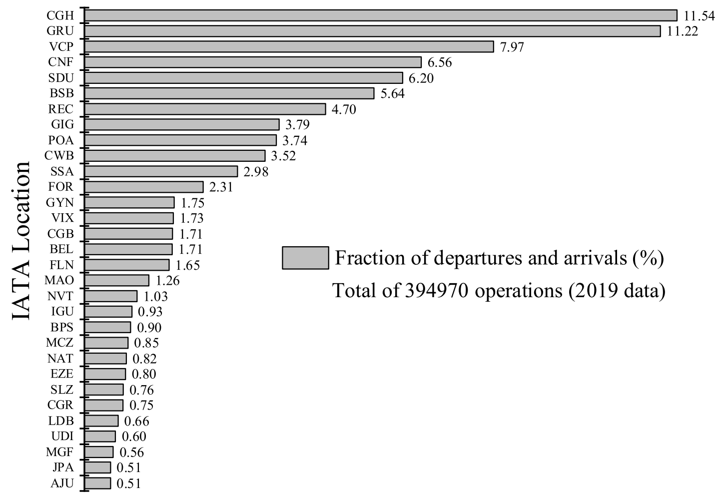

Fig. 6. Percentage of departures and arrivals for the airports with more than 0.5% of the total operations of 2019.

Table 7
Sleep metrics adopted in the SAFTE-FAST runs.

SAFTE-FAST Auto-Sleep Controls and Parameters	Value or Status
AUTO-NAP function	ON (standard) OFF (tailored)
Advanced bedtime function	ON (standard) OFF (tailored)
Normal bedtime	11 p.m.
Minimum Sleep Duration	60 min
Maximum Work Day Sleep	8 h
Maximum Rest Day Sleep	9 h
Max Recovery Nap [a]	210 min
Awake Zone [b]	1 to 8 p.m.
Sleep quality (home, hotel and rest facility)	Excellent
Inflight Sleep	Not included

[a]Recovery Nap is automatically added following work duties that end between Normal bedtime and the start of the Awake Zone if the sleep in the past 16 h is not optimal.
[b]The period of the day that the software prevents sleep events, except for afternoon naps prior to night shifts or advanced bedtime events due to early-start shifts.

Table 8
30 and 15-day epochs adopted for the extraction of flight schedules from the *Fadigômetro* database.

Period	Begin date [a]	End date [a]	Time interval (days)
Jan-19	1/1/19 0:00	1/31/19 0:00	30
Feb-19	1/31/19 0:00	3/2/19 0:00	30
Mar-19	3/2/19 0:00	4/1/19 0:00	30
Apr-19	4/1/19 0:00	5/1/19 0:00	30
May-19	5/1/19 0:00	5/31/19 0:00	30
Jun-19	5/31/19 0:00	6/30/19 0:00	30
Jul-19	7/1/19 0:00	7/31/19 0:00	30
Aug-19	8/1/19 0:00	8/31/19 0:00	30
Sep-19	8/31/19 0:00	9/30/19 0:00	30
Oct-19	10/1/19 0:00	10/31/19 0:00	30
Nov-19	10/31/19 0:00	11/30/19 0:00	30
Dec-19	12/1/19 0:00	12/31/19 0:00	30
Jan-20	1/1/20 0:00	1/31/20 0:00	30
Feb-20	1/31/20 0:00	3/1/20 0:00	30
Mar-19 (1/2)	3/1/19 0:00	3/16/19 0:00	15
Mar-20 (1/2)	3/1/20 0:00	3/16/20 0:00	15

[a]mm/dd/yy hh:mm.

For the input commuting we set a *standard* 60-min time interval from home to station, hotel to station and rest facility to station and vice-versa. We also applied an *extended* commuting of 120 min from home to station and vice-versa for some of the runs, based upon the responses of the questionnaire.

For the preparation time, defined as the average time the crew member usually takes to prepare himself for the flight, we set 60 min either at home, hotel or rest facility.

Owing to the lack of data or reliable information, the unwind time at home, hotel or rest facility was set to zero in our SAFTE-FAST input (see the Limitation section). Regarding the Auto-Sleep controls of the SAFTE-FAST, we have included both the Auto-Nap and the Advanced bedtime functions. The Auto-Nap function adds automatically an afternoon nap prior to night shifts. The amount of nap depends on the continuous wakefulness period until the last sleep event. For 8 to 10 h, 10 to 12 h, 12 to 14 h or more than 14 h since the last sleep event, the software adds 60, 90, 120 or 180 min of nap, respectively. The Advanced bedtime function allows the software to anticipate the beginning of sleep if an early start would significantly shorten the typical sleep quantity of 8 h. Both the Auto-Nap and the Advanced bedtime functions can be switched off, as shown for the tailored analyses presented in Table 6. All the sleep metrics adopted in the SAFTE-FAST runs are summarized in Table 7.

Appendix B. Characteristics of flights of the aircrew rosters

The dominant domestic short-haul flight characteristic of the roster's sample is shown in Fig. 6, which presents the fraction of departures and arrivals for the airports with more than 0.5% of all the 394,970 flight operations of 2019. Buenos Aires (EZE), with a frequency around 0.8%, stands alone as the only foreign destination among the 31 most frequent locations. Congonhas (CGH), Guarulhos (GRU), Campinas (VCP), Confins (CNF), Santos Dumont (SDU) and Brasilia (BSB) altogether comprise 49.1% of the total flight operations (*Crewing* events only). We had also a small fraction of daytime cargo flights operated by Azul Linhas Aéreas that represented, for instance, less than 0.07% (0.06%) of all flight sectors in July (December) of 2019.

Appendix C. Epochs for the analyses

Exact 30 and 15-day epochs were adopted to standardize the extraction of flight schedules, given that fatigue hazard area and duty time have a cumulative character. Moreover, the number of night shifts, consecutive night shifts, crewing/working events and WOCL operations also depend on the time interval. Even the SAFTE-FAST minimum effectiveness and minimum sleep reservoir depend on the time interval, as longer months have a higher probability for lower scores by chance. Table 8 shows the 30 and 15-day epochs adopted in this work.

References

Åkerstedt, T., Folkard, S., Portin, C., 2004. Predictions from the three-process model of alertness. Aviat., Space, Environ. Med. 75 (3), A75–A83.

Agência Nacional de Aviação Civil, 2019. Regulamento Brasileiro de Aviação Civil (RBAC 117): Requisitos para gerenciamento de risco de fadiga humana. Agência Nacional de Aviação Civil, Brasilia, Available at https://www.anac.gov.br/assuntos/legislacao/legislacao-1/rbha-e-rbac/rbac/rbac-117-emd-00 [cited on 11 Aug. 2022].

Agência Nacional de Aviação Civil, 2020. Anuário do Transporte Aéreo 2019. Agência Nacional de Aviação Civil, Brasilia, Available at https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-de-transporte-aereo/anuario-do-transporte-aereo [cited on 11 Aug. 2022].

Belyavin, A.J., Spencer, M.B., 2004. Modeling performance and alertness: the QinetiQ approach. Aviat., Space, Environ. Med. 75 (3), A93–A103.

Boeing, 2001. Statistical Summary of Commercial Jet Airplane Accidents - Worldwide Operations 1959 – 2000. Boeing, Seattle, Available at https://skybrary.aero/bookshelf/books/2726.pdf [cited on 11 Aug. 2022].

Borbély, A.A., 1982. A two process model of sleep regulation. Hum. Neurobiol. 1 (3), 195–204.

Borbély, A.A., Daan, S., Wirz-Justice, A., Deboer, T., 2016. The two-process model of sleep regulation: a reappraisal. J. Sleep Res. 25 (2), 131–143.

Brazil, 2017. Lei n° 13.475, de 28 de agosto de 2017. Available at https://www.anac.gov.br/assuntos/legislacao/legislacao-1/leis/lei-no-13-475-de-28-08-2017 [cited on 11 Aug. 2022].

Caldwell, J.A., 2005. Fatigue in aviation. Travel Med. Infect. Dis. 3 (2), 85–96.

Caldwell, J.A., 2012. Crew schedules, sleep deprivation, and aviation performance. Curr. Dir. Psychol. Sci. 21 (2), 85–89.

Caldwell, J.A., Mallis, M.M., Caldwell, J.L., Paul, M.A., Miller, J.C., Neri, D.F., 2009. Fatigue countermeasures in aviation. Aviat., Space, Environ. Med. 80 (1), 29–59.

Civil Aviation Safety Authority, 2014. Biomathematical Fatigue Models. Civil Aviation Safety Authority.

Cochrane, C., Ba, D., Klerman, E.B., Hilaire, M.A.S., 2021. An ensemble mixed effects model of sleep loss and performance. J. Theoret. Biol. 509, 110497.

Comissão Nacional de Fadiga Humana, 2020. Guidelines Manual for Investigating Human Fatigue in Aeronautical Occurences. Brasilia, Available at https://bit.ly/CNFH-guidelines-manual [cited on 11 Aug. 2022].

de Mello, M.T., Esteves, A.M., Pires, M.L.N., Santos, D.C.d., Bittencourt, L.R.A., Silva, R., Tufik, S., 2008. Relationship between Brazilian airline pilot errors and time of day. Braz. J. Med. Biol. Res. 41, 1129–1131.

de Vasconcelos, C.A., Vieira, M.N., Kecklund, G., Yehia, H.C., 2019. Speech analysis for fatigue and sleepiness detection of a pilot. Aerosp. Med. Hum. Perform. 90 (4), 415–418.

Devine, J.K., Garcia, C.R., Simoes, A.S., Guelere, M.R., de Godoy, B., Silva, D.S., Pacheco, P.C., Choynowski, J., Hursh, S.R., 2022. Predictive biomathematical modeling compared to objective sleep during COVID-19 humanitarian flights. Aerosp. Med. Hum. Perform. 93 (1), 4–12.

European Union Aviation Safety Agency, 2019. Effectiveness of Flight Time Limitation (FTL). European Union Aviation Safety Agency, Available at https://www.easa.europa.eu/document-library/general-publications/effectiveness-flight-time-limitation-ftl-report [cited on 11 Aug. 2022].

Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39 (2), 175–191.

Flynn-Evans, E.E., Kirkley, C., Young, M., Bathurst, N., Gregory, K., Vogelpohl, V., End, A., Hillenius, S., Pecena, Y., Marquez, J.J., 2020. Changes in performance and bio-mathematical model performance predictions during 45 days of sleep restriction in a simulated space mission. Sci. Rep. 10 (1), 1–14.

Goode, J.H., 2003. Are pilots at risk of accidents due to fatigue? J. Saf. Res. 34 (3), 309–313.

Fatigue Countermeasures Working Group, 2018. Controlled Rest on the Flight Deck: A Resourece for Operators. Flight Safety Foundation, p. 7.

Hartzler, B.M., 2014. Fatigue on the flight deck: the consequences of sleep loss and the benefits of napping. Accid. Anal. Prev. 62, 309–318.

Helene, O., Mariano, L., Guimaraes-Filho, Z., 2016. Useful and little-known applications of the Least Square Method and some consequences of covariances. Nucl. Instrum. Methods Phys. Res. A 833, 82–87.

Honn, K.A., van Dongen, H.P., Dawson, D., 2019. Working time society consensus statements: Prescriptive rule sets and risk management-based approaches for the management of fatigue-related risk in working time arrangements. Ind. Health SW–8.

Hursh, S.R., Fanzone, J., Raslear, T.G., et al., 2011. Analysis of the Relationship Between Operator Effectiveness Measures and Economic Impacts of Rail Accidents. United States. Federal Railroad Administration.

Hursh, S.R., Raslear, T.G., Kaye, A.S., Fanzone, Jr., J.F., 2006. Validation and calibration of a fatigue assessment tool for railroad work schedules, summary report.

Hursh, S.R., Redmond, D.P., Johnson, M.L., Thorne, D.R., Belenky, G., Balkin, T.J., Storm, W.F., Miller, J.C., Eddy, D.R., 2004. Fatigue models for applied research in warfighting. Aviat, Space, Environ. Med. 75 (3), A44–A53.

Ingre, M., Van Leeuwen, W., Klemets, T., Ullvetter, C., Hough, S., Kecklund, G., Karlsson, D., Åkerstedt, T., 2014. Validating and extending the three process model of alertness in airline operations. PLoS One 9 (10), e108679.

Instituto Brasileiro de Aviação, 2021. 6° Anuário Brasileiro de Aviação Civil. Instituto Brasileiro de Aviação, Available at https://institutoaviacao.org/anuario/ [cited on 11 Aug. 2022].

International Civil Aviation Organization, 2016. Doc 9966, Manual for the Oversight of Fatigue Management Approaches, second ed International Civil Aviation Organization, Available at https://www.icao.int/safety/fatiguemanagement/FRMS%20Tools/Doc%209966.FRMS.2016%20Edition.en.pdf [cited on 11 Aug. 2022].

International Civil Aviation Organization, 2021. ICAO safety report 2020 edition. Available at https://www.icao.int/safety/pages/safety-report.aspx [cited on 11 Aug. 2022].

Janic, M., 2000. An assessment of risk and safety in civil aviation. J. Air Transp. Manage. 6 (1), 43–50.

Jewett, M.E., Kronauer, R.E., 1999. Interactive mathematical models of subjective alertness and cognitive throughput in humans. J. Biol. Rhythms 14 (6), 588–597.

Kharoufah, H., Murray, J., Baxter, G., Wild, G., 2018. A review of human factors causations in commercial air transport accidents and incidents: From to 2000–2016. Prog. Aerosp. Sci. 99, 1–13.

Lamond, N., Dawson, D., 1999. Quantifying the performance impairment associated with fatigue. J. Sleep Res. 8 (4), 255–262.

Licati, P., Rodrigues, T.E., Wey, D., Fischer, F.M., Menna-Barreto, L., 2015. Correlação dos prognósticos do programa FAST com relatos de fadiga de pilotos da aviação civil brasileira. Revis. ConexãO SIPAER 6 (1), 7–17.

Mallis, M.M., Mejdal, S., Nguyen, T.T., Dinges, D.F., 2004. Summary of the key features of seven biomathematical models of human fatigue and performance. Aviat., Space, Environ. Med. 75 (3), A4–A14.

Marqueze, E.C., Nicola, A.C.B., Diniz, D.H., Fischer, F.M., 2017. Working hours associated with unintentional sleep at work among airline pilots. Revis. Saude Publ. 51.

Moore-Ede, M., Heitmann, A., Guttkuhn, R., Trutschel, U., Aguirre, A., Croke, D., 2004. Circadian alertness simulator for fatigue risk assessment in transportation: application to reduce frequency and severity of truck accidents. Aviat., Space, Environ. Med. 75 (3), A107–A118.

National Transportation Safety Board, 1993. Aircraft Accident Report — Uncontrolled Collision with Terrain, American International Airways flight 808, Douglas DC-8-61, N814CK, U.S. Naval Air Station, Guantanamo Bay, Cuba, August 18, 1993. (NTSB/AAR-94/04). National Transportation Safety Board, Washington.

Neri, D.F., Oyung, R.L., Colletti, L.M., Mallis, M.M., Tam, P.Y., Dinges, D.F., 2002. The flight deck. Aviat., Space, Environ. Med. 73, 654–664.

Olbert, A., Klemets, T., 2011. A Comprehensive Investigation of Flight and Duty Time Limitations and their Ability to Control Crew Fatigue. Jeppesen.

Rangan, S., Van Dongen, H., 2013. Quantifying fatigue risk in model-based fatigue risk management. Aviat., Space, Environ. Med. 84 (2), 155–157.

Raslear, T.G., Hursh, S.R., Van Dongen, H.P., 2011. Predicting cognitive impairment and accident risk. Prog. Brain Res. 190, 155–167.

Roach, G.D., Fletcher, A., Dawson, D., 2004. A model to predict work-related fatigue based on hours of work. Aviat., Space, Environ. Med. 75 (3), A61–A69.

Rodrigues, T.E., Fischer, F.M., Bastos, E.M., Baia, L., Bocces, R., Gonçalves, F.P., Licati, P.R., Menquini, A., Spyer, P., Stefenon, E., et al., 2020. Seasonal variation in fatigue indicators in Brazilian civil aviation crew rosters. Revis. Brasil. Med. Do Trabalho 18 (1), 2.

Rodrigues, T., Helene, O., 2020. Monte Carlo approach to model COVID-19 deaths and infections using Gompertz functions. Phys. Rev. Res. 2 (4), 043381.

Roma, P.G., Hursh, S.R., Mead, A.M., Nesthus, T.E., 2012. 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.

Rosekind, M.R., Graeber, R.C., Dinges, D.F., Connell, L.J., Rountree, M.S., Spinweber, C.L., Gillen, K.A., 1994. Crew factors in flight operations IX: effects of planned cockpit rest on crew performance and alertness in long-haul operations.

Rosekind, M., Gregory, K., Miller, D., et al., 1993. Analysis of crew fatigue factors in AIA Guantanamo Bay aviation accident, Appendix E. To Aircraft Accident Report: Uncontrolled Collision with Terrain, NTSB/AAR-94/04, NTSB, Washington, DC.

Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples) . Biometrika 52 (3–4), 591–611, Available at https://academic.oup.com/biomet/article-pdf/52/3-4/591/962907/52-3-4-591.pdf [cited on 11 Aug. 2022].

Van Dongen, H., 2004. Comparison of mathematical model predictions to experimental data of fatigue and performance. Aviat., Space, Environ. Med. 75 (3), A15–A36.