



METAMODELLING OF THE WORKLOAD ASSESSMENT IN SIMULATED FLIGHTS USING THE KRIGING METHOD

A. Esposito¹, G. Iacolino¹, C. Orlando¹ & A. Alaimo¹

¹Kore University of Enna, Cittadella Universitaria Enna, 94100, Italy

Abstract

This work focuses on utilizing a supervised surrogate Kriging model, which employs data input from objective metrics derived from heart rate variability sensors to predict workload levels. These signals are collected during simulated flight tasks, with workload levels classified according to the complexity of these tasks. An experimental low-fidelity flight simulator was developed specifically for this study, involving 20 participants from a bachelor's degree program in aerospace engineering. This study represents an application of Kriging for stress assessment in simulated flight environments. The model's performance indicates room for improvement, suggesting the need to expand the participant group for robust validation and broader applicability in future studies.

Keywords: Simulated Flight, Workload, HRV, Kriging

1. Introduction

The human factor (HF) generally refers to incidents resulting from the failure to execute actions prescribed by documentation [1]. Over the years, various idealized models of HF have been developed to enhance the understanding of this phenomenon. As reported in [2], contemporary aviation accidents are attributed to HF by approximately 70-80%. Within the HF domain, excessive workload (WL) in the aviation industry poses a significant contemporary challenge. With increasing air traffic, the need to cover more routes, and a limited number of human resources, pilots frequently find themselves in stressful situations. Over the years, stress assessment methods have been developed and categorized into subjective and objective approaches. Subjective tests represent a subjective method of stress based on questionnaires, typically administered before, during, and after performance. Common methods include the NASA Task Load Index (TLX) [3]. These subjective methods can be complemented by objective metrics capable of measuring the subject's physiological condition. While the subject's psychological condition may influence objective metrics, their value is independent of the subject's will, in contrast to subjective techniques. Various objective metrics are employed in stress evaluation, including eye blink rate (e.g. [4] and [5]), respiration rate (e.g. [6] and [7]) and heart rate variability (HRV) (e.g. [8] and [9]) which is the most widely used one. HRV captures spontaneous fluctuations over time between consecutive heartbeats, measured by the distance between two successive 'RR' peaks on an electrocardiogram. HRV measures in the time domain include the standard deviation of NN interval series (SDNN) [10], as well as nonlinear indexes derived from the Poincaré Plot, such as SD1 and SD2 [11]. According to existing literature, a reduction in these parameters indicates an increase in both cognitive and physical demands [12]. HRV also extends to the frequency domain, employing the Fourier transform to estimate power spectral density associated with frequency bands. These frequency bands provide meaningful information about sympathetic and parasympathetic activities. Furthermore, the HRV-related parameter, Total Power, reflects the total variance of heart rate and requires the analysis of various components at very low, low, and high frequencies of cardiac variability [13].

Thus, in this work, the focus will be on objective metrics obtained using sensors in an in-house Low Fidelity Flight Simulator (LFFS). Additionally, by employing the supervised surrogate Kriging model provided by [14], a pilot study of the workload assessment has been conducted, classifying the workload level between two different levels based on the complexity of the task. Understanding and mitigating the HF and excessive WL plays a crucial role in enhancing safety and performance. In addition, the use of multidimensional approaches based on artificial intelligence (AI) promises a deeper understanding of these phenomena.

2. Kriging Surrogate Model

AI encompasses various subcategories, with Machine Learning (ML) being the prominent one. In particular, Supervised Learning (SL) algorithms are fed with inputs and outputs, also known as *labels*, making SL an accessible and powerful tool for data prediction. Indeed, ML represents a path towards achieving AI [15], which is expected to create an economic value of \$13 trillion worldwide by 2030 [16].

In this work emphasis will be placed on a specific surrogate model known as Kriging. In general, Kriging is a technique for creating prevision models based on datasets, and it was pioneered by the South African engineer Danie Krige for use in geostatistics [17]. Various works related to Kriging can be found in the literature, focusing on topics such as multi-fidelity strategies for engineering design optimization [18], and the use of Kriging in constructing multi-fidelity aerodynamic models [19].

In addition, [20] describes the development and application of an active extremum Kriging-based multi-level linkage method for dynamic multi-component system reliability analysis.

Kriging has also been applied in the medical field. Further literature research reveals several notable works, including [21] which employs Kriging for mapping disease rates, and [22] which introduces a multi-objective optimization method for coronary stents using the aforementioned surrogate model.

Over time, Kriging algorithms have been adapted for both nondeterministic and deterministic simulation models. In the former case, only a limited number of random samples are available from nondeterministic simulations or physical experiments under uncertainty [23]. Conversely, deterministic models are characterized by input and output data [24], each capable of providing varying levels of accuracy. The Kriging method assumes predictions are weighted linear combinations of observed values, with closer input data indicating a higher positive correlation in prediction error [25].

From an analytical point of view, the Kriging is composed of a regression $f(x)$ and a Gaussian process $Z(x)$ with zero mean, variance σ^2 , and correlation matrix.

Therefore,

$$Y(x) = f(x) + Z(x) \quad (1)$$

Where $Y(x)$ is a random function and x is the sample under consideration.

There are different facets of the Kriging method. In particular, when the $f(x)$ is constant and known, it is called Simple Kriging. When the correlation function is constant but not known, it is called Ordinary Kriging. In the most general case, if the function is a multivariate polynomial, then it is called Universal Kriging. Nevertheless, for a comprehensive mathematical treatment of the problem, refer to [26] and [27]. As observed above, the Kriging model has been employed across various fields, from engineering to medicine. This study marks the first instance of utilizing such a model for stress prediction purposes. Hence, the Kriging model has been selected for its predictive capabilities based on dataset analysis. Inputs for the model include HRV values, while the output considered is the WL level, as elaborated further in subsequent sections.

3. Experimental Setup

Evaluating the psychosomatic state of pilots is challenging, especially in real critical flight situations such as takeoff, landing, or emergencies [28]. For this reason, flight simulators can be employed, offering valuable opportunities including training [29] and stress assessment [30]. These simulators can range from basic low-fidelity flight simulators (LFFS) to more advanced and technologically sophisticated FFS [31]. The utilization of FFS, while extremely reliable [30], may pose significant limitations including purchase, maintenance, and operational costs [32]. In this research study, a LFFS was employed with the primary aim of engaging a broader population. As a matter of fact, FFS are

typically restricted to professional pilots. Furthermore, an examination of the existing literature has unveiled several studies centered on LFFSs, for instance [33], [34], [35], [36], which underline that LFFS represent a good tool for verifying WL status. Thus, this research work is based on the LFFS developed at Kore University of Enna. In this study, the LFFS design is based on the three equilibrium equations of motion, together with the navigation equations. These equations are based on four reference systems, namely: Ground Earth Frame, Moving Earth Frame, Body Reference System, and Wind Frame. To formulate the mathematical model, it is assumed that the thrust (T) lies along the X_{BODY} axis and is modeled as an ideal constant thrust system, unaffected by airspeed and only influenced by air density variation with altitude. Fuel consumption depends on a constant specific consumption (C) and thrust intensity. The Earth is considered flat, and the air is assumed to be calm. Under these assumptions, these equations were then executed in Matlab/Simulink along with the International Standard Atmosphere model (to compute air density variation with altitude) and a simplified aerodynamic model.

Thus, the setup is divided into software and hardware components. From a software perspective, three interface subsystems were employed: pilot, instrumentation, and virtual reality subsystems. The pilot subsystem implements the software interface with the pilot's commands, namely the joystick that controls throttle, pitch, and bank. As the mathematical model does not account for rotational dynamics, some transfer functions are introduced between the joystick input signals and the model equations to create the illusion of flight. These transfer functions have been calibrated using characteristic time responses from the FFS of the Kore University of Enna. The instrumentation subsystem connects the simulated flight variables with basic cockpit indicators; namely, the anemometer, the altitude indicator, the attitude indicator, the turn coordinator, the vertical speed indicator and the magnetic compass. In addition to these, the subsystem includes a clock and a 'master caution-like' push button that is used during the HF exercise, as described in the following section. Finally, the VR subsystem connects the variables computed during the flight simulation to the virtual reality environment. In this case, the VR environment is represented by the open-source software FlightGear, which visualizes the flying scenario and serves as the simulator's visual system. On the other side, the hardware part consists of the Joystick (Model T.Flight Stick X), which can generate four input signals corresponding to variations in throttle and pitch, roll, and yaw angles. Additionally, two monitors have been installed. The Polar H10 sensor was also used to assess HRV while the student performed the exercise. During the baseline phase, the student wore the HF sensor and received instructions on basic commands and instrumentation usage during a briefing session. This briefing session was considered a resting phase to establish participants' baseline physiological signals. Afterward, the student was asked to perform two distinct exercises with increasing levels of WL, particularly:

- **TASK 1: Push-Button Exercise (Lower WL level):** starting already at 3000 [ft] with a Mach number of 0.5, the simulation begins in a trimmed state. The subject presses the button as soon as possible, as the 'master-caution-like' warning activates. The exercise lasts for six minutes, and at the end, the student must complete a questionnaire on the subjective perception of WL (more details on the questionnaire are given in the following section).
- **TASK 2: Holding Circuit Exercise (Higher WL level):** the simulation starts in balanced flight conditions at an altitude of 3000 [ft] with a Mach number of 0.5. The student is asked to "press the button" as soon as possible when the 'Master-Caution-like' warning light turns on. Additionally, after one minute of level flight, the subject should execute a standard right turn for one minute, followed by one minute of level flight; then, a second standard right turn is executed for one minute, concluding the exercise in level flight. During the maneuvers, the altitude should be maintained constant and equal to the trim altitude. The exercise lasts for six minutes. Finally, the student completed a questionnaire to evaluate subjective WL.

Figure 1 illustrates the test configuration. The display featuring the simulated VR scenario is outlined in blue, the basic instrument panel in green, and the joystick, along with other simulation controls like the throttle, push button, marker, and the label displaying the current local time, are all delineated in red. Figure 2 represents the briefing scheme designed to clarify the exercise for students involved in the experiment.

At the end of the session, three sets of data will be collected: i) subjective WL assessment; ii) physiological signals, enabling an objective evaluation of WL; iii) flight data, providing insights into piloting performance. Specifically, flight data comprises recorded signals, which enable a step-by-step visualization of piloting performance, and performance indices calculated from simulated FDR data, providing an overall quantification of the entire exercise in terms of an error function relative to the required task. Further details regarding the computation of the performance index and the data recorded are provided in the subsequent sections.

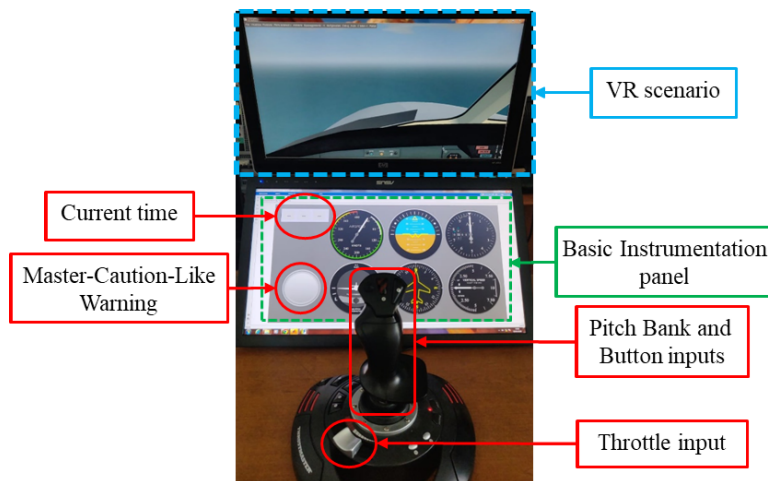


Figure 1 – Low-Fidelity Flight Simulators (LFFS).

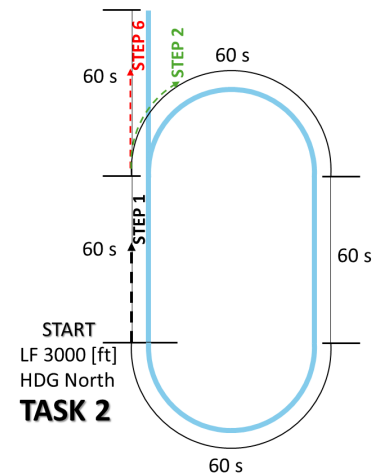


Figure 2 – Task 2 briefing scheme.

3.1 Subjective Test Assessment

As previously mentioned, following the task performed, students were provided with a questionnaire to evaluate their WL after completing the mission. This test was created with the NASA-TLX as a reference point and compiled on a dedicated Google Forms platform. The test begins with the participant's demographic data, including: student ID, age, weight, height, smoking status, years of pilot experience, and whether the participant is familiar with WL assessment or situational awareness questionnaires. Subsequently, participants are asked to compare their usual work activity with the scenario of the exercise performed based on their work experience. Following this, they are prompted to evaluate each individual point on a scale ranging from 'extremely low,' 'low,' 'average,' 'high,' to 'extremely high.' A total of six questions were asked to assess the participants' experience throughout the simulation, with the aim of evaluating different characteristics of WL. Figure 3 shows an overview of the subjective questionnaire.

Workload Task 1					
Based on your work experience, compare the activity you usually carry out with the scenario of exercise performed.					
How would you rate...	Extremely Low	Low	Average	High	Extremely High
1 the overall workload during the last run?	☐	☐	☐	☐	☐
2 the workload level of managing time pressure during the last run?	☐	☐	☐	☐	☐
3 the workload level of managing the number of tasks carried out during the last run?	☐	☐	☐	☐	☐
4 the workload level of managing the overlap of different activities during the last run?	☐	☐	☐	☐	☐
5 your level of frustration during the last run?	☐	☐	☐	☐	☐
6 your level of satisfaction in performing the tasks during the last run?	☐	☐	☐	☐	☐

Figure 3 – Subjective form with workload items.

4. Results

This section presents the results of the workload evaluation. First, the results in terms of subjective evaluation are presented, followed by the performance indices and their results, which are presented and commented on. Finally, the results section describes the HRV indexes, focusing on the supervised surrogate Kriging model.

4.1 Results based on Subjective assessment

The data sample was obtained from a group of 20 students from a bachelor's degree course in aerospace engineering. The population details can be found in Table 1, and the density plots of the sample are presented in Figure 4.

	Mean	Median	Standard Deviation	Min	Max
Age	21.1	21.0	1.36	19	24
Weight	80.6	79	14.7	60	115
Height	179	178	6.63	170	194

Table 1 – Statistical characteristics of the participants.

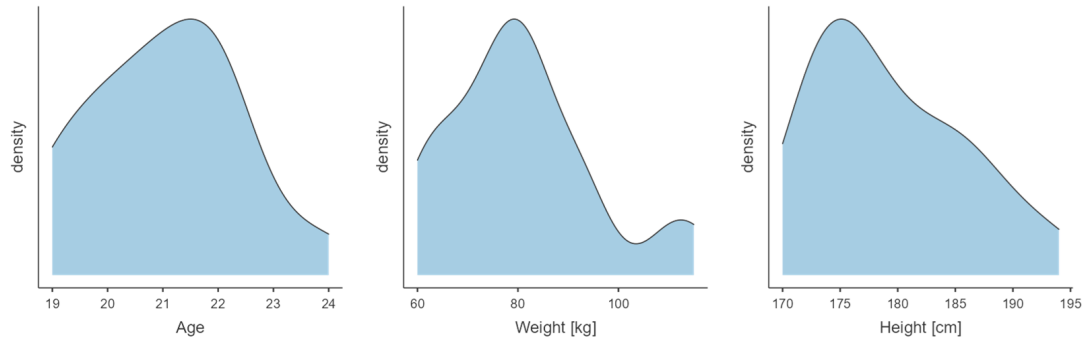


Figure 4 – Density of population characteristics distribution.

As described in section 3.1, the participants were asked to complete a questionnaire on their perceived workload at the end of each task. The results are collected and presented in Figure 4. The Likert scale was converted to a scale from 0 to 5 (extremely low = 0 and extremely high = 5). An exception to this was made for the item relating to performance, which was presented as a reverse item. The obtained value represents, for each participant, the average of the 6 questionnaire items.

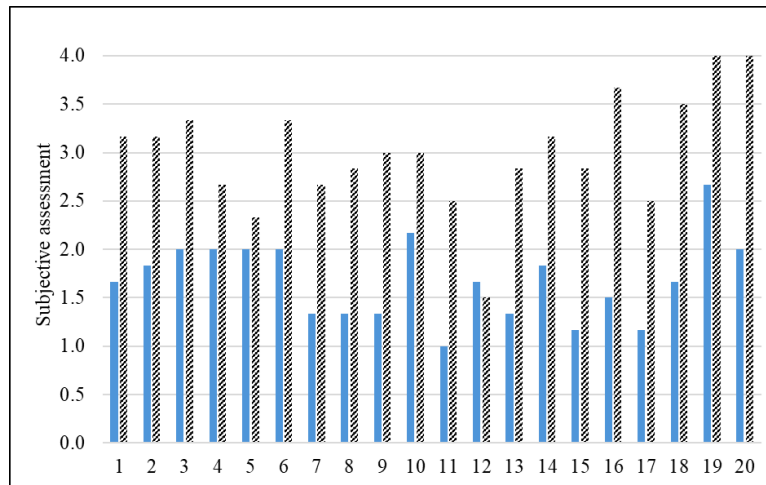


Figure 5 – Subjective perception of workload level for Task 1 (blue) and Task 2 (dashed black).

As illustrated in Figure 4, for the majority of participants, the blue bars (representing task 1) are lower than the black bars (representing task 2). This trend suggests that Task 1 is generally perceived as less complicated than Task 2. An exception is seen with student No. 12, who finds Task 1 slightly more complicated than Task 2, as indicated by the blue bar being marginally higher. Thus, to confirm what was hypothesized in terms of WL levels, the subjective results reveal that participants consider Task 2 more complicated than Task 1.

4.2 Results based on Performance indexes

Two performance indexes are introduced to investigate the pilot's ability to carry out the assigned tasks. The first one refers to task 1 regarding the master caution-like pushbutton exercise and is defined as described in what follows. Figure 6 shows (in black) the time history of the master caution warning that is triggered every 30 seconds and remains turned on for 7.5 seconds, let say it is called $y(t)$; on the other hand, the blue curve, $z(t)$, shows a representative push-button response to turn off the warning light. The performance index for such exercise is defined as

$$I_{pb} = \frac{\int_0^T z(t) dt}{\int_0^T y(t) dt} \quad (2)$$

being $T = 360$ s the simulation time. The index $I_{pb} \in [0, 1]$ and the best is ideally obtained at zero, representing the pilots' full awareness of the warning signal as well as his/her good reaction time.

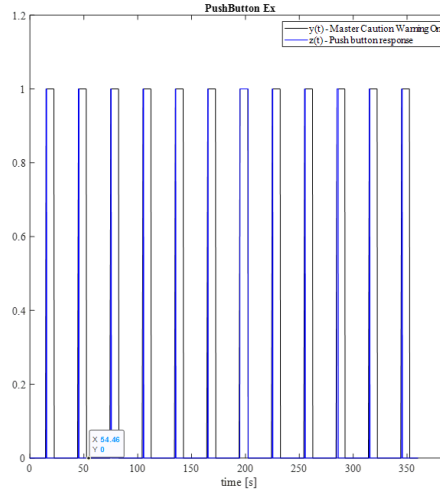


Figure 6 – Example of pushbutton exercise signals.

The second performance index is instead used to have a measure of the standard turn exercise I_{st} . As described before, also during exercise two, the students are asked to turn off the master caution-like warning; thus, the pushbutton index I_{pb} is considered as well. In addition, two more metrics are introduced: one, I_{tr} , refers to the desired turn rate signal to be tracked by the student, while the other, I_{alt} , refers to constant altitude to be maintained during the exercise. More particularly, the turn rate metric I_{tr} is defined as the ratio of the integral absolute error between performed $y_1(t)$ and expected $y_2(t)$ turn rate, see Figure 7 and the integral of the absolute value of the expected turn rate as

$$I_{tr} = \frac{\int_0^T |y_1(t) - y_2(t)| dt}{\int_0^T |y_2(t)| dt} \quad (3)$$

while the constant altitude metric is defined as

$$I_{alt} = \frac{\int_0^T |y_3(t) - y_4(t)| dt}{\int_0^T |y_4(t)| dt} \quad (4)$$

being $y_3(t)$ the actual aircraft altitude and $y_4(t)$ the expected altitude, similar performance indexes were used by Alaimo et al. in [37]. Then, the exercise performance index is obtained by averaging the three metrics as

$$I_{st} = \frac{I_{pb} + I_{tr} + I_{alt}}{3} \quad (5)$$

also in this case, the best performance is obtained when the index I_{st} tends to zero. Representative results of performed exercises are given in Figure 7 where the time histories of turn rate and altitude and the path followed by the aircraft center of mass can be appreciated.

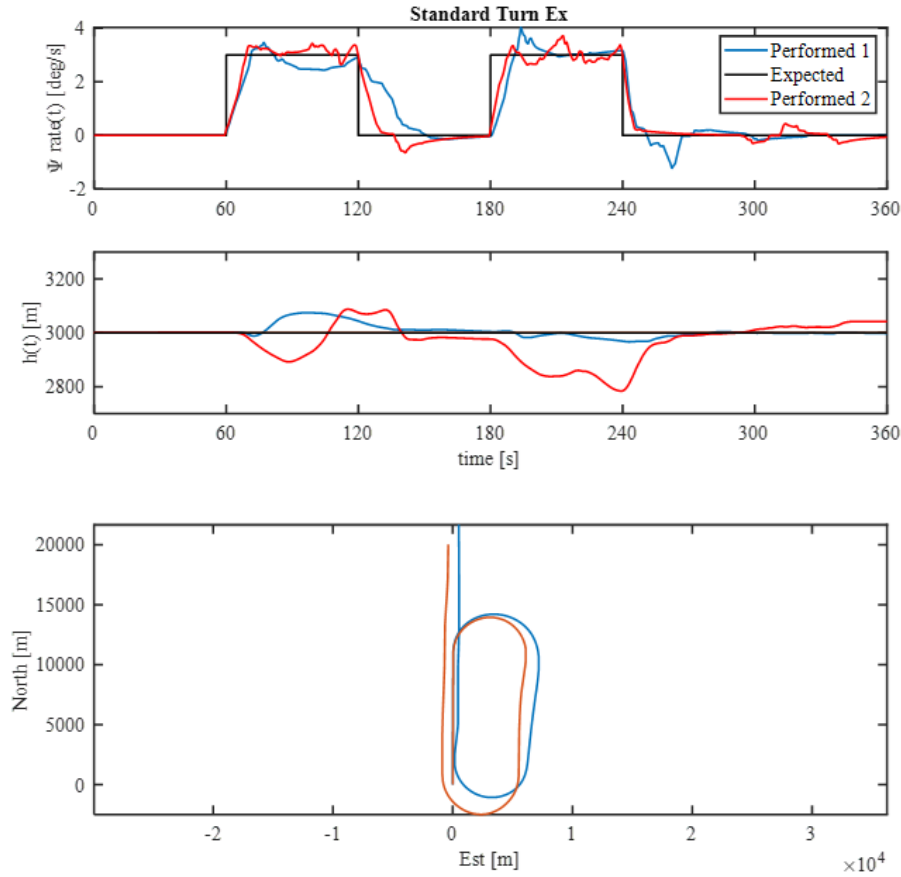


Figure 7 – Example of flight data derived from simulated FDR.

Last, the exercise performance indexes computed for all the students are reported in Figure 8. It can be appreciated that for 17 out of 20 students, the second exercise has been more difficult to perform, confirming that a higher level of difficulty with respect to the first one characterizes it. Thus, it is expected that a higher level of WL is required of students.

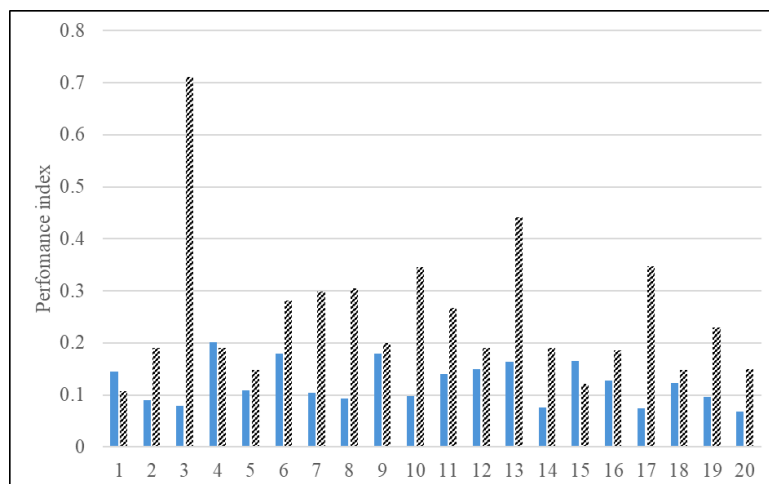


Figure 8 – Performance indexes for Task 1 (blue) and Task 2 (dashed black).

4.3 Results based on HRV

Various indices can be derived from the HRV signal for workload assessment. In this study, three specific indices have been chosen. These selected HRV indexes encompass $SDNN$, LF , and TOT_{pow} . The motivation behind selecting these indices is based on their confirmed validity in workload interpretation, as evidenced by Alaimo et.al in [38]. Similarly, the authors substantiate the utility of employing these HRV indices to effectively approximate the workload tendencies of pilots in Full Flight Simulator sessions across several flight phases.

Table 2 presents the sample's descriptive characteristics for each index for completeness. More in detail, Table 2 reports the mean value and the median for the two Task. The classification of the WL level lower (Task 1) and higher (Task 2) has been performed based on the complexity of the task.

Moreover, Figure 9 shows the boxplot for the two exercises (Task 1 and Task 2) to prove the effectiveness of the three indexes on the sample. It is essential to recognize that the data pertains to a sample comprising all 20 subjects. By boxplots, it is practical to note the uniform tendency of the three indexes that present a decreasing trend correlated to the assumed increase in WL level. The boxplots also present some outliers; more specifically, two students are outliers among the three indexes. Thus, their data are excluded from the subsequent Kriging surrogate model.

	TASK	SDNN	HRV _{TI}	LF	TOT _{pow}	SD2
Mean	1	39.4	10	998	1629	84.5
	2	37.9	9.39	854	1243	73.2
Median	1	40.6	9.52	934	1472	78.7
	2	37.4	9.21	648	1113	72.8

Table 2 – Descriptive sample characteristics in terms of mean value and standard deviation.

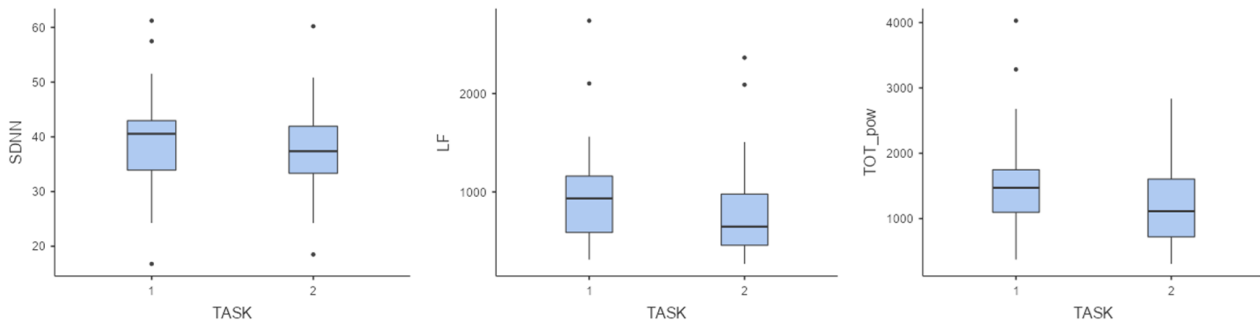


Figure 9 – Boxplot of HRV indexes over the WL level.

4.4 Results by Kriging Surrogate Model based on HRV

The main objective of the present work is to test the Kriging approach in the assessment of WL levels. To achieve this, the 18 remaining participants are divided into two sub-groups: a main group of 17 students used to define the metamodel and a secondary group consisting of one student to predict the value of the tester subject subsequently. In the results that follow, the main group is referred to as the "Predictive Group" (PG), while the secondary one is referred to as the "Tester Subject" (TS).

This approach involved selecting participants through windowing the tester subject in the adjacent row data position. For instance, the first tester subject TS {1} was the attendee 1, and the PG {2 – 18} represented students ID 2 to 18. For the second tester TS {2} and PG {1} $\cup$ {3 – 18} was used to form a PG group of 17 subjects, and this pattern continued sequentially.

Considering the 18 testers (TS), a predictive metamodel was tested for each index. Specifically, Table 3 summarizes the success rate results of metamodels across all 18 testers for the three indexes. All models were evaluated at two different task levels, and Table 3 indicates a value of 1 when the metamodel successfully captures an increase in WL from low to high levels (Task 2 > Task 1). Furthermore, the final row, labeled 'Positive Success Rate,' indicates a green marker for the respective TS when at least two out of three indexes exhibit positive activation. The overall analysis of the results

shows that the metamodels can capture an increase in WL from low to high levels (Task 2 > Task 1) in 56% of cases. Some discrepancy occurs within the indexes with a success rate for *LF* and *TOT_{pow}* equal to 61% and SDNN of 46%. However, at this stage, the detailed sensitivity analysis of the results cannot confirm the reliability of the metamodels due to the small considered sample. Thus, while the developed metamodels can capture the trend associated with increasing WL, the results are also sensitive to the selection of PGs. There could be several reasons for this. The available data are limited and, therefore, need to be expanded. At the same time, discrepancies among individual TS could be attributed to external parameters that may influence both the simulation and the metamodel, especially considering the impact of outliers in a larger sample. For example, the flight experience of a student compared to his colleagues mates was not considered. A student with more simulation experience might exhibit an increasing trend as the flight task difficulty rises, yet his/her heightened awareness during simulations could lead to greater confidence and a lower overall detection of WL compared to less experienced colleagues.

Advancements to the present work will focus on extending the sensitivities analyses to a larger number of students, analyzing input dimensions for the Kriging models to verify the ability to define a mixed metamodel based on more than one index, and additionally evaluating alternative supervised models.

TS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Success rate
SDNN	1	0	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	0	44%
LF	1	1	1	0	1	1	0	1	1	0	1	1	0	0	0	1	1	0	61%
TOT _{pow}	1	1	1	0	1	0	1	1	1	1	0	1	1	0	0	1	0	0	61%
Positive Success Rate	✓ 1	✓ 1	✓ 1	✗ 0	✓ 1	✓ 1	✗ 0	✓ 1	✓ 1	✓ 1	✗ 0	✓ 1	✗ 0	✗ 0	✗ 0	✓ 1	✗ 0	✗ 0	56%

Table 3 – Summary table of different metamodels results based on attendee Tester (TS).

Acknowledgements

The study was partially financially supported by the Italian Ministry of University and Research - M.U.R. under the Sicilian Micro and Nano Technology Research and Innovation Center - SAMOTH-RACE project CUP B73D21015010006.

Contact Author Email Address

Corresponding Author: Giuseppe Iacolino
Email: giuseppe.iacolino@unikorestudent.it

Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

References

- [1] Alexander Rezchikov, Olga Dolinina, Vadim Kushnikov, Vladimir Ivaschenko, Konstantin Kachur, Aleksey Bogomolov, and Leonid Filimonyuk. The problem of a human factor in aviation transport systems. *Indian Journal of Science and Technology*, 9, 2016.
- [2] Dário Antonio Leite Martins de Sant, Adriana Victória Garibaldi de Hilal, et al. The impact of human factors on pilots' safety behavior in offshore aviation companies: A brazilian case. *Safety science*, 140:105272, 2021.
- [3] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.
- [4] Zhen Wang, Lingxiao Zheng, Yanyu Lu, and Shan Fu. Physiological indices of pilots' abilities under varying task demands. *Aerospace medicine and human performance*, 87(4):375–381, 2016.

- [5] Yiyuan Zheng and Yuwen Jie. Study of nasa-tlx and eye blink rates both in flight simulator and flight test. In *Engineering Psychology and Cognitive Ergonomics: 16th International Conference, EPCE 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21*, pages 353–360. Springer, 2019.
- [6] Maria Katsamanis Karavidas, Paul M Lehrer, Shou-En Lu, Evgeny Vaschillo, Bronya Vaschillo, and Andrew Cheng. The effects of workload on respiratory variables in simulated flight: a preliminary study. *Biological psychology*, 84(1):157–160, 2010.
- [7] Mariel Grassmann, Elke Vlemincx, Andreas von Leupoldt, and Omer Van den Bergh. The role of respiratory measures to assess mental load in pilot selection. *Ergonomics*, 59(6):745–753, 2016.
- [8] Lei Wang, Shan Gao, Wei Tan, and Jingyi Zhang. Pilots’ mental workload variation when taking a risk in a flight scenario: a study based on flight simulator experiments. *International journal of occupational safety and ergonomics*, 29(1):366–375, 2023.
- [9] Guido Li Volsi, Ines Paola Monte, Alessandro Aruta, Alfio Gulizzi, Andrea Libra, Stefano Mirulla, Gianluca Panebianco, Giovanni Patti, Ferdinando Quattrocchi, Vincenzo Bellantone, et al. Heart rate variability indices of student pilots undergo modifications during flight training. *Aerospace Medicine and Human Performance*, 94(11):835–842, 2023.
- [10] Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation*, 93(5):1043–1065, 1996.
- [11] Sobhendu Kumar Ghatak and Subhra Aditya. Poincaré parameters and principal component analysis of heart rate variability of subjects with health disorder. *arXiv preprint arXiv:1802.10289*, 2018.
- [12] D. W. Watson. Physiological correlates of heart rate variability (hrv) and the subjective assessment of workload and fatigue in-flight crew: A practical study. 2001.
- [13] Andrzej Bilan, Agnieszka Witczak, Robert Palusiński, Wojciech Myśliński, and Janusz Hanzlik. Circadian rhythm of spectral indices of heart rate variability in healthy subjects. *Journal of electrocardiology*, 38(3):239–243, 2005.
- [14] Ivo Couckuyt, Tom Dhaene, and Piet Demeester. Gode toolbox: a flexible object-oriented kriging implementation. *Journal of Machine Learning Research*, 15:3183–3186, 2014.
- [15] A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 44, 2000.
- [16] Jacques Bughin, Jeongmin Seong, James Manyika, Michael Chui, and Raoul Joshi. Notes from the ai frontier: Modeling the impact of ai on the world economy. *McKinsey Global Institute*, 4, 2018.
- [17] Noel Cressie. The origins of kriging. *Mathematical geology*, 22:239–252, 1990.
- [18] David JJ Toal. Applications of multi-fidelity multi-output kriging to engineering design optimization. *Structural and Multidisciplinary Optimization*, 66(6):125, 2023.
- [19] V Pham, M Tyan, TA Nguyen, and JW Lee. Extended hierarchical kriging method for aerodynamic model generation incorporating multiple low-fidelity datasets. *aerospace* 2024, 11, 6, 2023.
- [20] Hong Zhang, Lu-Kai Song, Guang-Chen Bai, and Xue-Qin Li. Active extremum kriging-based multi-level linkage reliability analysis and its application in aeroengine mechanism systems. *Aerospace Science and Technology*, 131:107968, 2022.
- [21] Che-Chia Hsu, Dai-Rong Tsai, Shih-Yung Su, Jing-Rong Jhuang, Chun-Ju Chiang, Ya-Wen Yang, and Wen-Chung Lee. A stabilized kriging method for mapping disease rates. *Journal of Epidemiology*, 33(4):201–208, 2023.
- [22] Hongxia Li, Junfeng Gu, Minjie Wang, Danyang Zhao, Zheng Li, Aike Qiao, and Bao Zhu. Multi-objective optimization of coronary stent using kriging surrogate model. *BioMedical Engineering OnLine*, 15:275–291, 2016.
- [23] Harok Bae, Daniel L Clark, and Edwin E Forster. Nondeterministic kriging for engineering design exploration. *AIAA Journal*, 57(4):1659–1670, 2019.
- [24] Jack P.C. Kleijnen. Kriging metamodeling in simulation: A review, 2009.
- [25] Wim CM Van Beers and Jack PC Kleijnen. Kriging for interpolation in random simulation. *Journal of the Operational Research Society*, 54:255–262, 2003.
- [26] Alexander Forrester, Andras Sobester, and Andy Keane. *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008.
- [27] Jack PC Kleijnen. Kriging metamodeling in simulation: A review. *European journal of operational research*, 192(3):707–716, 2009.
- [28] Giulia Masi, Gianluca Amprimo, Claudia Ferraris, and Lorenzo Priano. Stress and workload assessment in aviation—a narrative review, 2023.

- [29] Brent Cameron, Hooman Rajaei, Bradley Jung, and RG Langlois. Development and implementation of cost-effective flight simulator technologies. In *International Conference of Control, Dynamic Systems, and Robotics*, number 126, page D01, 2016.
- [30] Giuseppe Iacolino, Antonio Esposito, Calogero Orlando, and Andrea Alaimo. A brief review of pilots' workload assessment using flight simulators: subjective and objective metrics. *Materials Research Proceedings*, 37.
- [31] Matthias Oberhauser and Daniel Dreyer. A virtual reality flight simulator for human factors engineering. *Cognition, Technology & Work*, 19:263–277, 2017.
- [32] Paul L Myers III, Arnold W Starr, and Kadie Mullins. Flight simulator fidelity, training transfer, and the role of instructors in optimizing learning. *International Journal of Aviation, Aeronautics, and Aerospace*, 5(1):6, 2018.
- [33] Kyle J Jaquess, Rodolphe J Gentili, Li-Chuan Lo, Hyuk Oh, Jing Zhang, Jeremy C Rietschel, Matthew W Miller, Ying Ying Tan, and Bradley D Hatfield. Empirical evidence for the relationship between cognitive workload and attentional reserve. *International Journal of Psychophysiology*, 121:46–55, 2017.
- [34] Donald L Lassiter, Daniel G Morrow, Gary E Hinson, Michael Miller, and David Z Hambrick. Expertise and age effects on pilot mental workload in a simulated aviation task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 40, pages 133–137. SAGE Publications Sage CA: Los Angeles, CA, 1996.
- [35] P Archana Hebbar, Kausik Bhattacharya, Gowdham Prabhakar, Abhay A Pashilkar, and Pradipta Biswas. Correlation between physiological and performance-based metrics to estimate pilots' cognitive workload. *Frontiers in psychology*, 12:555446, 2021.
- [36] Liya Tang, Juanning Si, Lei Sun, Gengsheng Mao, and Shengyuan Yu. Assessment of the mental workload of trainee pilots of remotely operated aircraft using functional near-infrared spectroscopy. *BMC neurology*, 22(1):160, 2022.
- [37] Andrea Alaimo, Antonio Esposito, Calogero Orlando, and Andre Simoncini. Aircraft pilots workload analysis: Heart rate variability objective measures and nasa-task load index subjective evaluation. *Aerospace*, 7(9), 2020.
- [38] Andrea Alaimo, Antonio Esposito, Palmira Faraci, Calogero Orlando, and Giusy Danila Valenti. Human heart-related indexes behavior study for aircraft pilots allowable workload level assessment. *IEEE Access*, 10, 2022.