Human Factors in Robots, Drones and Unmanned Systems, Vol. 93, 2023, 50–57
https://doi.org/10.54941/ahfe1003747

Human Factors Assessment for Drone Operations: Towards a Virtual Drone Co-Pilot

Daniela Doroftei, Geert De Cubber, and Hans De Smet

Royal Military Academy of Belgium, Av. De La Renaissance 30, 1000 Brussels, Belgium

ABSTRACT

Drone operations are prone to incidents caused by human error. Therefore, a careful consideration of the human factors and operator flight parameters that affect the performance of drone pilots is required. Based upon an analysis of these parameters, it is possible to deduce what parameters correspond to behaviour exposed by 'good' pilots and what behaviour corresponds to behaviour exposed by 'bad' pilots. Within this paper, the development of an AI-based expert system that is able to perform such an assessment in real-time is discussed. The expert system is exploited as a virtual co-pilot tool that continuously monitors flight performance and warns pilots when bad flight behaviour is detected, not dissimilar to advanced driver assistance systems in the car industry that warn drivers when not paying attention to the road.

Keywords: Unmanned aerial systems, Human performance evaluation, Simulator systems

INTRODUCTION

As the number of drone operations increases, so does the risk of incidents with these novel, yet sometimes dangerous unmanned systems (Chow et al., 2016, Buric et al. 2017). Research has shown that over 70% of drone incidents are caused by human error (Shively, 2015). Therefore, in order to reduce the risk of incidents, the human factors related to the operation of the drone should be studied. However, this is not a trivial exercise, because on the one hand, a realistic operational environment is required (in order to study the human behaviour in realistic conditions), while on the other hand a standardised environment is required, such that repeatable experiments can be set up in order to ensure statistical relevance. In order to remedy this, within the scope of the ALPHONSE project (Doroftei et al., 2020), a realistic simulation environment was developed that is specifically geared towards the evaluation of human factors for military drone operations.

Within the ALPHONSE simulator, military (and other) drone pilots can perform missions in realistic operational conditions. At the same time, they are subjected to a range of factors that can influence operator performance. These constitute both person-induced factors like pressure to achieve the set goals in time or people talking to the pilot and environment-induced stress factors like changing weather conditions. During the flight operation, the ALPHONSE simulator continuously monitors over 65 flight parameters.

After the flight, an overall performance score is calculated, based upon the achievement of the mission objectives. Throughout the ALPHONSE trials, a wide range of pilots has flown in the simulator, ranging from beginner to expert pilots. Using all the data recorded during these flights, three actions are performed:

- An Artificial Intelligence (AI) - based classifier was trained to automatically recognize in real time 'good' and 'bad' flight behaviour. This allows for the development of a virtual co-pilot that can warn the pilot at any given moment when the pilot is starting to exhibit behaviour that is recognized by the classifier to correspond mostly to the behaviour of inexperienced pilots and not to the behaviour of good pilots.
- An identification and ranking of the human factors and their impact on the flight performance, by linking the induced stress factors to the performance scores
- An update of the training procedures to take into consideration the human factors that impact flight performance, such that newly trained pilots are better aware of these influences.

The objective of this paper is to present the ALPHONSE simulator system for the evaluation of human factors for drone operations and present the results of the experiments with real military flight operators. The focus of the paper will be on the elaboration of the design choices for the development of the AI - based classifier for real-time flight performance evaluation.

PREVIOUS WORK & MAIN CONTRIBUTIONS

The proposed work consists of the development of an advanced pilot assistance system (APAS) for drone operations, based on the assessment of the performance of human operators in a virtual flight simulator environment.

Multiple commercial implementations towards drone flight simulation exist. Examples are the DJI drone simulator, DRL Sim Drone Racing Simulator, the Zephyr Drone Simulator, the Drone Sim Pro Flight Simulator or the RealFlight Drone Simulator. While many of these flight simulators are of excellent quality, the main criticism to most of these implementations is that they lack the capability to log flight & performance parameters during the flight, which renders them unusable for the application at hand. Therefore, we developed the ALPHONSE simulator system, which is introduced in (Doroftei et al., 2019) and validated to state of the art approaches in (Doroftei et al., 2022). As the core developments of this paper heavily depend on this simulator tool, the basic concept of this simulator tool is briefly repeated here in order to enhance the readability of the paper.

The ALPHONSE simulator is a highly realistic simulation environment, where drone operators can perform complex drone operations. In summary, it is based upon the Microsoft AirSim simulation engine (Shah et al., 2017), which is an open-source simulator for drones, built on the Unreal

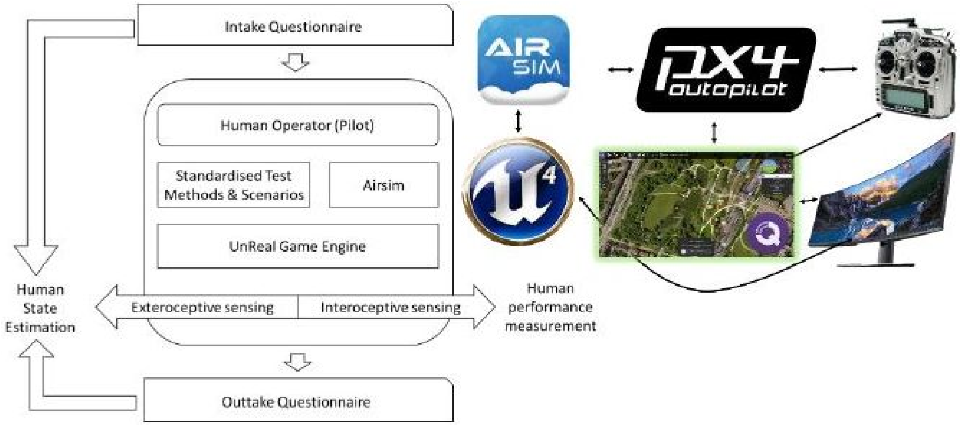


Figure 1: Left: schematic overview of the test procedure where the pilots are subjected to. After taking an intake survey, the pilots have to perform a complex mission in a simulation environment. While doing this, their performance parameters and physiological state are assessed. After completing the mission, they perform an outtake survey. Right: schematic overview of the interplay between the different hardware & software components of the ALPHONSE simulator.

Engine (Epic Games, 2023). This simulation environment is completely open and customizable, which enables us to incorporate the standard test scenarios, to multiple customizable drones and to quantitatively measure the performance of the pilots on-line while executing the mission. Within this simulator, 22 standard operational scenarios are defined within two environments (urban/rural). In these scenarios, the operators need to deal with large-scale dynamic environments, changing environmental conditions and time pressure in order to deliver quality data in a minimal amount of time. These are all factors that can lead to human errors and can have an impact on the performance of the operator.

In the domain of drone simulation, this paper builds upon the earlier developments of the ALPHONSE simulation tool and provides no new contribution, as this paper focuses on the contributions in the domain of APAS.

APAS systems for drone operations do exist. Multiple commercial drone systems are equipped with the capability of returning home automatically or of detecting obstacles and avoiding those. However, these focus on GNSS – based automated guidance and exteroceptive sensing for collision avoidance and do not really investigate the human-drone interaction aspect of the flight operation.

Within the DLR MOSES (More operational flight safety by enhancement of situation awareness) project (Weber et al., 2022), different approaches to measuring and improving situation awareness were investigated, especially in the domain of information gathering depending on different tasks during approach and taxiing. Eye movements of forty pilot students and nine experienced pilots were analysed. Furthermore, the influence of an additional taxi guidance system on eye movements was investigated. While the results are significant, they focus on ground operations, and not on flight operations.

In this paper, we will therefore introduce a novel advanced pilot assistance system for drone operations. The system is based on the analysis of a series of flight patterns by a range of pilots (from novice pilots to heavily trained experts) by an AI system. Based on this analysis, a classifier is constructed that can evaluate in real time the flight patterns exposed by a pilot and – if required – issue a warning signal if the system deems that the flight behaviour is not corresponding to a flight model that is adequate for the mission profile.

In the next section, we will discuss more in detail the concept & design of this so-called AI-based virtual co-pilot.

AI-BASED VIRTUAL CO-PILOT CONCEPT & DESIGN

Multiple drone pilots with varying skill levels have performed flight trials within the ALPHONSE simulator under controlled conditions. Each of the pilots had to perform a multi-target detection mission, while being subjected to auditive distractions, changing weather conditions, etc. At the same time, a series of 66 flight parameters available in the MavLink drone messaging protocol (Koubâa, 2019) is tracked at a rate of 300 samples per second. The problem at hand for the design of the virtual co-pilot is how to determine the relationship between these flight parameters and the skill level of the pilot. In order to do this, a two-step approach was followed.

In a first 'human intelligence' step, the problem space was reduced by incorporating the knowledge from the drone pilots (obtained via intake and outtake questionnaires) from 66 flight parameters to a list of 26 relevant flight parameters that need to be further analysed. These flight parameters include speed & acceleration, gradients for roll, pitch yaw and climb rate, vibrations, control stick inputs, etc.

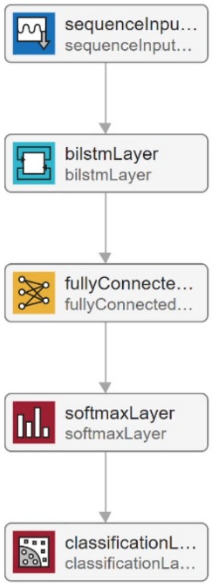
In a second 'artificial intelligence' step, a neural network was trained to model the relationship between the 26 flight parameters and the pilot skill level. However, first the data must be pre-processed. Indeed, as the flight simulator logs the data at a very high sample rate, it is necessary to downscale the data flow. Therefore, we apply an averaging over 100 samples, meaning that each sample covers around 0, 3 seconds. The sequence length is a careful compromise between two contradictory requirements: on the one hand the sequence should be as long as possible in order to incorporate as much data as possible in the decision process. On the other hand, the requirement mustn't be long, as this would induce an unacceptable delay in the skill recognition process. After careful consideration, a sequence length of 50 was chosen, which corresponds to a period of around 15 seconds.

The neural net for the proposed classifier is then composed of 5 base layers:

The network has furthermore 100 hidden layers. The other parameters for the deep learning network architecture are presented in Table 1.

It can be noted from Table 3 that the choice was made to distribute the outputs into three classes. These three classed convey to the following categories:

- **Category 1: Novice pilots.** It should be noted that we include in this category pilots that have experience with e.g. fixed wing drones but not with



- A SequenceInputLayer that manages the sequences of 26 flight parameters.
- A bidirectional long short-term memory (BiLSTM) layer that learns bidirectional long-term dependencies between time steps of time series.
- A fully connected layer that multiplies the input by a weight matrix and then adds a bias vector.
- A softmax layer that applies a softmax function (Bishop, 2006) to the input.
- A classification layer that computes the cross-entropy loss for classification and weighted classification tasks with mutually exclusive classes.

Figure 2: Network architecture.

Table 1. Deep learning network architecture parameters.

ANALYSIS RESULT				
	Name	Type	Activations	Learnables
1	**sequenceInputLayer** Sequence input with 26 dimensions	Sequence Input	26	-
2	**bilstmLayer** BiLSTM with 100 hidden units	BiLSTM	200	InputWeights 800x26 RecurrentWeights 800x100 Bias 800x1
3	**fullyConnectedLayer** 3 fully connected layer	Fully Connected	3	Weights 3x200 Bias 3x1
4	**softmaxLayer** softmax	Softmax	3	-
5	**classificationLayer** crossentropyex with '1' and 2 other classes	Classification Output	3	-

rotary wing drones, as we noted that they expose similar skill levels. This shows that piloting skills seem ill-transferable across drone types.

- **Category 2: Good pilots.** This category includes pilots with experience in flying rotary wing drones.
- **Category 3: Expert pilots.** In this category, we included datasets from highly skilled pilots who practice complex flight operations daily.

We implemented an Adaptive Moment (ADAM) estimation method for stochastic optimization (Kingma and Ba, 2015). ADAM combines two stochastic gradient descent approaches: adaptive gradients, and root mean square propagation. Given this data, the ADAM algorithm is able to converge to a stable solution for the classifier, as shown in Figure 3.

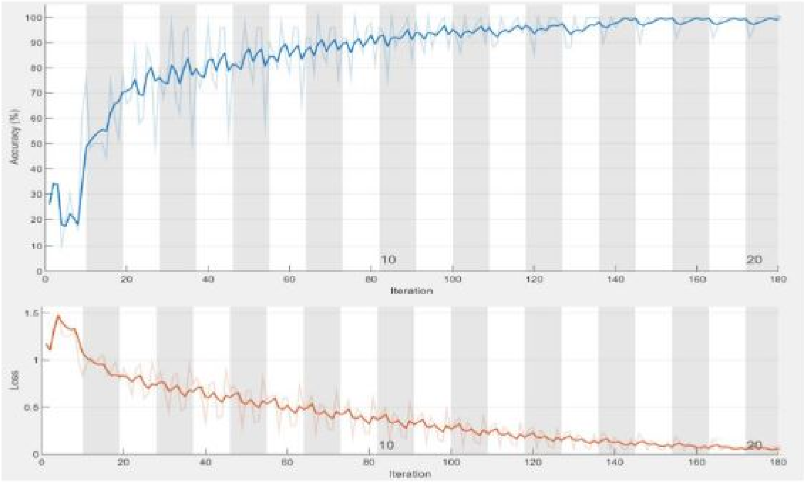


Figure 3: Convergence of the accuracy and loss of the ADAM pilot skill classifier.

HUMAN FACTORS EVALUATION WITH BELGIAN DEFENCE PILOTS

In this section, results are presented of tests with drone pilots from Belgian Defence and civilian Belgian Defence researchers that have flown within the ALPHONSE simulator. These pilots have first acted as data subjects to provide flight data to train the model and have later been used to validate the model. In order to provide this validation, the available flight data from all pilots was split up and 80% was used for the training of the model, while 20% was used for validation.

In order to assess the performance of the pilot skill level classifier, the left side of Figure 4 shows the receiver operating characteristic (ROC) curve for the classifier. Note that this ROC curve is already zoomed in towards the upper left, as the ratio of false positives to true positives is very low for the classification across all classes, as indicated by the ROC curve. This is also visible in the confusion matrix, shown on the right of Figure 4, indicating also the global accuracy of the classifier at 88,5%. However, it must be noted that most ‘confusion’ for the classifier exists between the classes 2 and 3 (respectively good & expert pilots), which is not a big problem for the co-pilot application. Indeed, for the co-pilot application, what really matters is the recognition of novice pilots (corresponding to the first class), for which the classifier attains an accuracy of 92,8%, which is excellent.

The pilot skill level classification is performed in 0,02 seconds, which means that in theory, alerts could be generated at a rate of 50Hz. In practice, it is not necessary to be so fast, as there is in any case already a 15 second latency induced by the need to accumulate a meaningful sequence length of flight parameters. We therefore chose to implement a Kalman Filter to disregard misclassifications.

The output of this Kalman Filter is then an auditive alert at a rate of 1 Hz, issued to the drone operator when (s)he is manifesting flight behaviour that is recognized by the classification system as corresponding to the one of novice

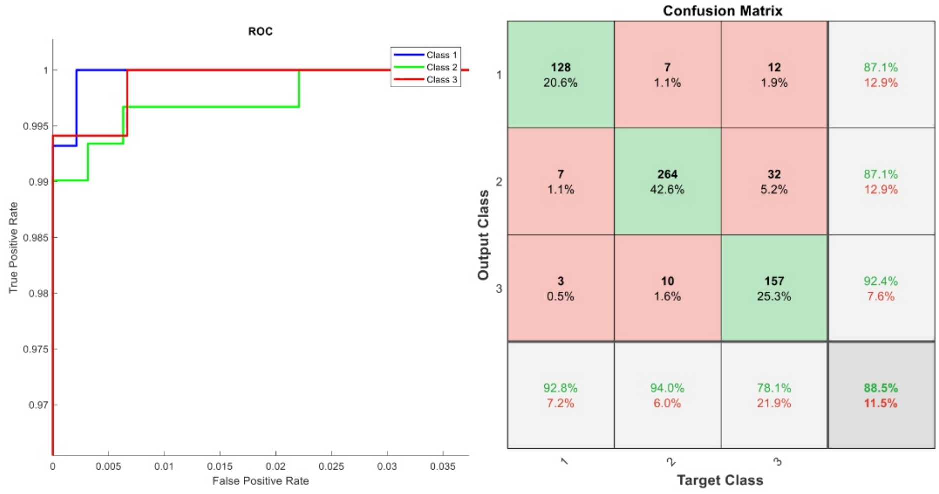


Figure 4: Classifier performance shown by the ROC curve (left) & confusion matrix (right).

pilots. If the reason for the 'bad' flight behaviour is a lack of attention to the piloting task, the drone operator can then take corrective measures. If the pilot is indeed a novice pilot, still learning how to operate a drone with an instructor, then the instructor can guide – based on the alerts by the virtual co-pilot – give advice to the pilot on how to improve the flight performance.

CONCLUSION

In this paper, we have proposed a virtual drone co-pilot system that is able to detect bad flight behaviour and issue warning signals to the drone pilot. The system is based on an AI-based classifier that was trained based upon data obtained in a flight simulator system. However, it is to be stressed that – as the model uses a standard drone control protocol (i.e. MavLink) – it can be used generically for a wide range of real drones for monitoring real flight operations. The validation shows that the virtual co-pilot achieves a very high accuracy and can in 92,8% of the cases correctly identify 'bad' flight profiles in real-time.

The proposed development is highly significant, as it presents a concrete and cost-effective methodology for developing a virtual co-pilot for drone pilots that can render drone operations safer. Indeed, while the initial training of the AI model requires considerable computing resources, the implementation of the classifier can be readily integrated in commodity flight controllers to provide real-time alerts when pilots are manifesting undesired flight behaviours.

ACKNOWLEDGMENT

This research was funded by the Belgian Royal Higher Institute for Defense and the VIAS Institute in the framework of the research study HFM19/05 (ALPHONSE).

REFERENCES

Bishop, C. M. (2006) Pattern Recognition and Machine Learning. Springer, New York, NY, USA.

Buric, Marian and De Cubber, Geert. (2017) Counter Remotely Piloted Aircraft Systems, MTA Review Volume 27. No. 1, Military Technical Academy Publishing House.

Chow, E., Cuadra, A., Whitlock, C. (2016) Hazard Above: Drone Crash Database - Fallen from the skies. The Washington Post.

Doroftei, Daniela and De Smet, Hans (2019). Evaluating Human Factors for Drone Operations using Simulations and Standardized Tests, in 10th International Conference on Applied Human Factors and Ergonomics (AHFE 2019), Washington DC, USA.

Doroftei, Daniela. De Cubber, Geert and De Smet, Hans (2020). Reducing drone incidents by incorporating human factors in the drone and drone pilot accreditation process, in Advances in Human Factors in Robots, Drones and Unmanned Systems, San Diego, USA, pp. 71–77.

Doroftei, Daniela. De Cubber, Geert and De Smet, Hans (2022). Assessing Human Factors for Drone Operations in a Simulation Environment," in Human Factors in Robots, Drones and Unmanned Systems – AHFE International Conference, New York, USA.

Epic Games, Inc. (2023) Accessed 12 February 2023, Website: https://www.unrealengine.com.

Kingma, Diederik, and Ba, Jimmy. (2015) Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations.

Koubâa, A., Allouch, A., Alajlan, M., Javed, Y., Belghith, A. and Khalgui, M. (2019) Micro Air Vehicle Link (MAVlink) in a Nutshell: A Survey. in IEEE Access, vol. 7, pp. 87658–87680.

Shah, S., Dey, D., Lovett, C., Kapoor, A. (2017) AirSim: High-Fidelity Visual and Physical Simu-lation for Autonomous Vehicles. Field and Service Robotics.

Shively, J. (2015) Human Performance Issues in Remotely Piloted Aircraft Systems. In: ICAO Conference on Remotely piloted or piloted: sharing one aerospace system.

Weber, Ute and Attinger, Sabine and Baschek, Burkard and Boike, Julia and Borchardt, Dietrich and Brix, Holger and Brüggemann, Nicolas and Bussmann, Ingeborg and Dietrich, Peter and Fischer, Philipp M. and Greinert, Jens and Hajnsek, Irena and Kamjunke, Norbert and Kerschke, Dorit and Kiendler-Scharr, Astrid and Körtzinger, Arne and Kottmeier, Christoph and Merz, Bruno and Merz, Ralf and Riese, Martin and Schloter, Michael and Schmid, Hans Peter and Schnitzler, Jörg-Peter and Sachs, Thorsten and Schütze, Claudia and Tillmann, Ralf and Vereecken, Harry and Wieser, Andreas and Teutsch, Georg (2022) MOSES: a novel observation system to monitor dynamic events across Earth compartments. Bulletin of the American Meteorological Society, pp. 1–23. American Meteorological Society.