Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Increasing predictability in the response of an AI-assisted stall recovery system in complex stall conditions by expanding the knowledge-base of AI

To cite this article: C Koopman and D Zammit-Mangion 2024 *J. Phys.: Conf. Ser.* **2716** 012054

View the article online for updates and enhancements.

You may also like

- Flexible needle puncture path planning for liver tumors based on deep reinforcement learning
Wenrui Hu, Huiyan Jiang and Meng Wang
- Automated gadget discovery in the quantum domain
Lea M Trenkwalder, Andrea López-Incera, Hendrik Poulsen Nautrup et al.
- CT2X-IRA: CT to x-ray image registration agent using domain-cross multi-scale-stride deep reinforcement learning
Haixiao Geng, Deqiang Xiao, Shuo Yang et al.

This content was downloaded from IP address 86.153.25.19 on 20/03/2024 at 16:20

Increasing predictability in the response of an AI-assisted stall recovery system in complex stall conditions by expanding the knowledge-base of AI

C Koopman[1] and D Zammit-Mangion[1,2]

[1] Institute of Aerospace Technologies, University of Malta, Msida MSD 2080, Malta
[2] Centre for Aeronautics, Cranfield University, Bedford MK43 0JR, UK

E-mail: cynthia.koopman@um.edu.mt

Abstract. The environment in the cockpit of commercial aircraft is becoming increasingly complex due to the introduction of automation systems. This complexity is especially evident when malfunctions take place, making it difficult for pilots to comprehend the interconnectedness of the systems and potentially leading to loss of control. This paper investigates a novel method for creating an Artificial Intelligence-based stall recovery assistant using Reinforcement Learning by training the agent to generate a stall and subsequently recover from it. This enables training in a large training space with a simple reward function, where the agent has the ability to develop a deep understanding of the environment. Tests show that the agent is able to recover from stall at a variety of altitudes while experiencing unreliable airspeed information originating from a blocked Pitot tube system and with a better response than all baseline agents. The results indicate that restricting AI is not always necessary and, further, that too many restrictions can lead to a system that learns only shallow features, causing it to be unreliable in unforeseen circumstances.

1. Introduction

The operational environment in the cockpit of commercial aircraft has become increasingly complex over the last decades after the introduction of digital systems in the cockpit in the late 1970s and early 1980s. This complexity arises primarily from an increase in automation systems, and this becomes particularly noticeable when such automation systems fail. The risk of disconnect between the pilot and automation even in normal operation is well known, with such disconnects following an automation-related failure that had led to loss of control where the crew was unable to resolve the problem with the available information. Indeed, several aircraft accidents have occurred due to conflicting information that is either generated by automation systems or caused by inadvertent activation of automation systems [1, 2, 3, 4]. Solving problems in such an increasingly complex environment is, therefore, critical to safety. However, this may become too demanding a task for the human pilot in the case of failures such as data errors or when the human pilot is additionally faced with factors such as fatigue and stress.

To mitigate this risk, an 'artificial expert' can be introduced to help the pilot better understand the situation and, consequently, bridge the gap between human and automation systems. Artificial Intelligence (AI) has the ability to create powerful functions that reduce the probability of loss of control and mitigate the risks of consequences when such events occur.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

One way of doing the latter is by using a system based on Reinforcement Learning to assist in upsets such as stall recovery procedures.

AI can be a suitable candidate for this issue since it has the capacity to comprehend the environment and the aircraft's response to it, as well as comprehend the consequences of any malfunctions. To do so, the Reinforcement Learning system is required to have the correct features that allow such deep understanding. In Reinforcement Learning, these features are created in the neural network by interacting with the environment and by a reward that informs the agent if the action was desirable or not. As the knowledge of the Reinforcement Learning agent becomes deeper, the unpredictability in unforeseen situations reduces. Creating such deep connections can become challenging, as it is directly influenced by the environment space in which it is operating and the reward it receives. If the exploration space is too small and the reward too demanding, the good performance of the Reinforcement Learning agent can be deceiving. Instead of learning meaningful connections, the Reinforcement Learning agent learns the reward function. If this happens, the Reinforcement Learning agent will have good performance in the space it was learnt in, but this performance becomes unpredictable outside the training space. As a result, the system created will need very strict limitations on the operational domain, as described by the EASA guidelines [5], to ensure safe behaviour. Currently, most research that uses Reinforcement Learning to improve safety in upset conditions results in a very narrow operational domain.

Previous work that uses Reinforcement Learning for aircraft control in upset recovery can be distinguished by approaches that aim to achieve robustness by training without off-nominal conditions or by including these during training. Research that only tests their Reinforcement Learning system on disturbances, such as wind, effectively expects that their training method is capable of letting the Reinforcement Learning agent learn such a deep understanding that it can recognise that the aircraft system does not behave as expected and hence must adapt its strategy [6, 7, 8, 9]. However, it remains unclear how the Reinforcement Learning agent is supposed to achieve this understanding from the examples on which it was trained. When an Reinforcement Learning agent is trained only on perfect examples, there is no incentive for the agent to intrinsically create, for example, a filter for noise. Therefore, it should not be surprising that it cannot achieve the same performance. The only thing that the Reinforcement Learning agent can do under such disturbances is to depend heavily on the reward, which is why these models are typically trained with a reward function that defines many variables in a dense manner. This method is restrictive and does not take full advantage of the possibilities of AI.

Previous research has acknowledged the importance of including disturbances during training [10, 11]. Cao et al. [11] trained their stall recovery system not only for different weather conditions, but also for different weights. Changing the weight of the aircraft can create very informative examples for the Reinforcement Learning agent to train on, for example, it can learn connections between the response of the controls and angle of attack. These studies modify the training examples in order to enhance the Reinforcement Learning agent's understanding and consequently enhance its response to such conditions.

Our aim is to extend the approach from previous research to increase the knowledge of the Reinforcement Learning agent to enhance control in stall recovery. This paper addresses this challenge to create an AI assistant by focussing on the increase of knowledge of the neural connections in the Reinforcement Learning system to create an improved ability to respond predictably in unforeseen circumstances, specifically in the context of stall with data (sensor) failures.

To expand the understanding of the Reinforcement Learning agent and create more meaningful connections, the aim of this paper is to investigate a technique that does not limit the Reinforcement Learning agent from exploring different parts of the environment. Therefore, the Reinforcement Learning agent is trained in a very large training space, and effectively training

from a sparse representation of this large training space. To increase the focus on learning stall recovery within this large space, and thus making the examples more useful to learn from, the Reinforcement Learning agent is required to train on a related task: learning how to force a stall. This novel method of training for stall recovery is realised by training a single neural network to generate a stall and subsequently recover from it.

The remainder of this paper is structured as follows. Section 2 presents the implementation of the Reinforcement Learning algorithm and training details. Section 3 analyses the ability to recover from a stall by comparing different baseline agents. Finally, the paper is concluded in the last section.

2. Reinforcement Learning architecture and training details

For Reinforcement Learning, the Proximal Policy Optimization (PPO) algorithm [12] was used. This algorithm is able to optimize with continuous control actions and has stable performance with perturbations to the state space compared to other reinforcement algorithms [13]. PPO is an on-policy algorithm, these require more data than off-policy algorithms, which is not an issue when using a fast simulation environment. On-policy algorithms are also guaranteed to converge.

The PPO algorithm is based on an actor-critic architecture. The actor predicts the actions to be taken based on the state information as input. During training, the predicted actions are sampled from a normal distribution to increase exploration. Therefore, the output of the actor network are a mean and standard deviation of the population to be sampled from. For testing and deployment, the means that are predicted are directly used as actions to make the policy deterministic. The critic network is only used during training to optimise the actor network.

For optimisation of the Reinforcement Learning task, the PPO variant with a clipped objective is used, following the work by Schulman et al. [12] for continuous control tasks. This uses a probability ratio between the old and new policy $p_t(\theta) = \frac{\pi_{\theta(a_t|s_t)}}{\pi_{\theta_k}(a_t|s_t)}$, to stabilise learning by limiting the deviation from the old policy. The objective function for PPO follows Eq. 1:

$$\mathcal{L}^{CLIP}_{\theta_k}(\theta) = min\Big(p_t(\theta)A^{\pi_{\theta_k}}(s_t, a_t), clip(p_t(\theta), 1-\epsilon, 1+\epsilon)A^{\pi_{\theta_k}}(s_t, a_t)\Big), \tag{1}$$

where ϵ is a hyperparameter that defines the amount by which the new policy is allowed to deviate from the old and $A^{\pi_{\theta_k}}(s_t, a_t)$ is the estimated advantage function based on the current value function $V_\phi(s_t)$. The critic network is optimised by the mean-squared error loss between the predicted value and the calculated advantages.

The actor and the critic are implemented as neural networks. Both networks have the same architecture of a Long-Short Term Memory (LSTM) layer followed by two fully connected layers. Each layer contains 256 neurons.

The training of the Reinforcement Learning model was performed in the six-degrees-of-freedom simulation software Jsbsim [14], with a A320 commercial aircraft model. The input to the Reinforcement Learning agent is a state space consisting of 24 variables including flight path angles, velocities, altitude, current position of control surfaces, Pitot tube sensor information, and target information for the altitude and airspeed. The action space of the agent was restricted to the control of the elevator and the thrust setting. Turns were not considered due to simulation limitations. The initial states were randomly sampled from a wide range of altitudes and airspeeds. The altitude was sampled between 15000 and 35000 ft. The airspeed was sampled between 420 ft/s and 770 ft/s. To address the ability to recover from stall whilst experiencing data errors, a Pitot tube system was manipulated during training. This Pitot tube system allowed for continuous blockages to modulate severity. Blocking the Pitot tube had the effect of making airspeed information unreliable.

To train the Reinforcement Learning agent to create a stall and subsequently recover from it, the reward was changed accordingly, and this change was then notified to the agent by a flag in the state space. First, an episode was sampled in a trimmed initial state, where the agent was rewarded towards a stall. If this episode resulted in a stall condition or went outside the safe envelope, the episode was terminated. In the next episode, the agent started in the stall state that was previously created and was rewarded towards recovery. The reward to create a stall included reward for low lift, high angles of attack, low vertical acceleration, and low vertical speed. The reward function to recover from the stall consisted of airspeed and altitude errors. When recovering from the stall, no termination conditions were included.

3. Stall recovery results and comparison to Baseline agents

For comparison, a baseline agent was trained that was not allowed to leave the safe flight envelope. Thus, this agent was trained to recover only from a stall and all episodes during training started in a stall. These stalls were generated by reducing thrust and pitching up until a stall occurred at low vertical speed. The reward function for this baseline was the same as our approach, however, several termination conditions were included during training. These are termination conditions that are typically included: high angles of attack, secondary stall, and low airspeed. The resulting agent was tested on stall recovery examples at a variety of altitudes. During testing, early termination conditions were removed. The results of this agent show that it does not generalise well, examples of its strategy to stall recovery are shown in Fig. 1 and Fig. 2. These examples include random blockages of the Pitot tube and therefore have unreliable airspeed information. The agent can recover in the example in Fig. 1, but is unable to recover in the example in Fig. 2 as it creates a secondary stall. This inability to generalise indicates that the features that were learnt during training are superficial.

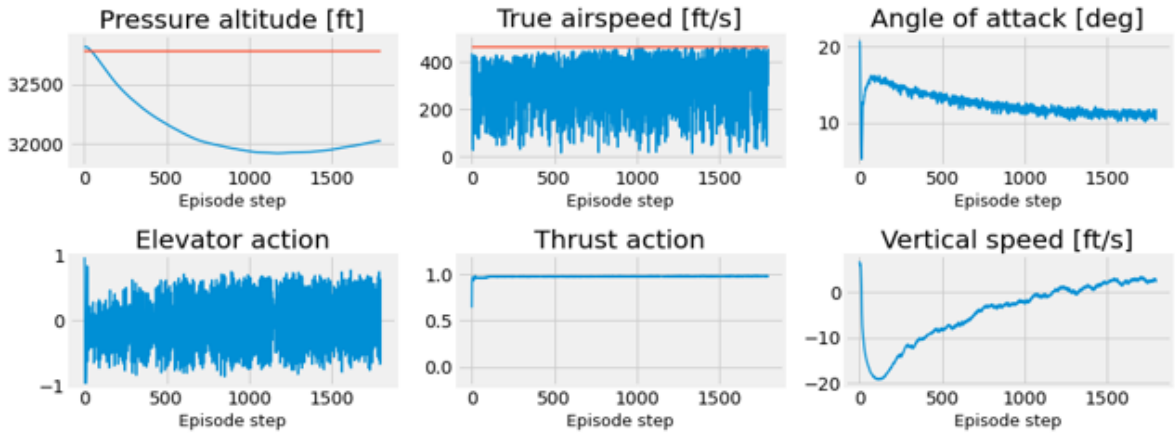

Figure 1. Stall recovery at 33000 ft of the baseline agent trained with a restrictive approach. Random Pitot blockages were included. The altitude and airspeed targets are shown in red.

To highlight these effects, another agent was trained with even stricter termination conditions; a time limit on secondary stall and a maximum condition on loss of altitude. An example of its stall recovery is shown in Fig. 3. Here the agent successfully reduces the angle of attack, however, it introduces strange behaviour in order to remain within the allowed altitude loss requirement. This indicates that the agent has learnt the termination conditions over more meaningful features that would lead the agent to reduce the altitude loss.

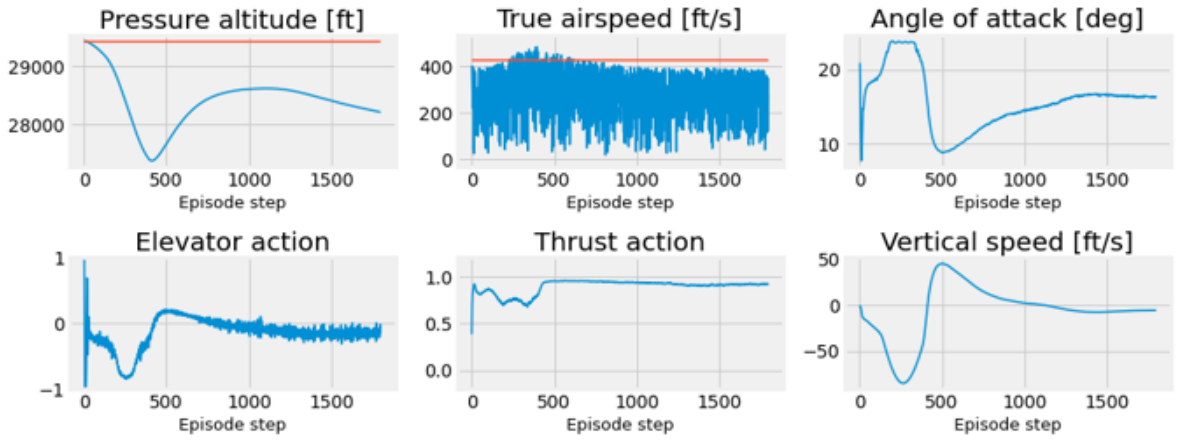

Figure 2. Stall recovery at 29000 ft of the baseline agent trained with a restrictive approach. Random Pitot blockages were included. The altitude and airspeed targets are shown in red.

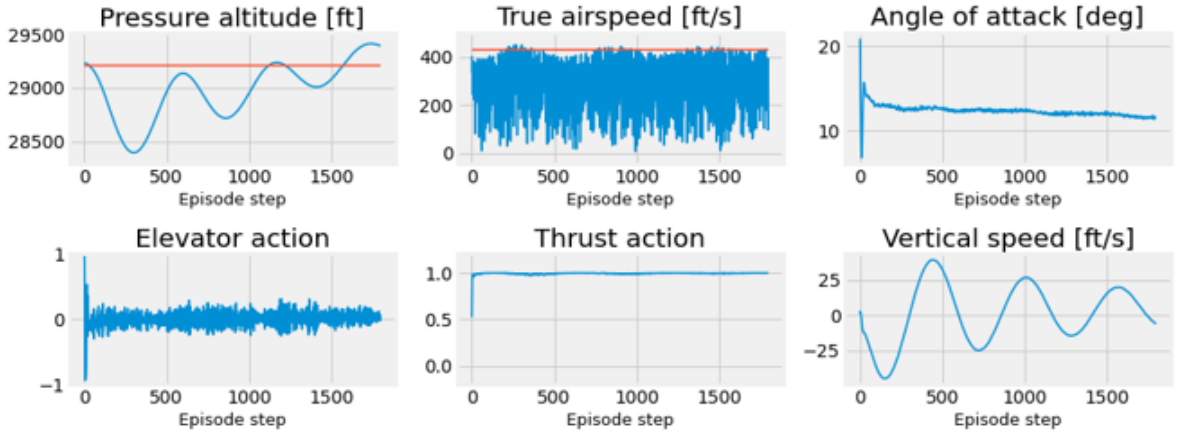

Figure 3. Stall recovery of the baseline agent with more strict termination conditions. Random Pitot blockages were included. The altitude and airspeed targets are shown in red.

To isolate the behaviour of our training method (learning to stall), an agent was trained that did not have any termination conditions. Removing all termination conditions would allow the Reinforcement Learning agent to have a larger exploration space and therefore be able to learn more meaningful features. This was tested by keeping the same training as the previous baseline agent, starting in stall, but removing all termination conditions. Figure 4 shows an example of stall recovery with this agent where the Pitot tube was unblocked. This shows that the agent can recover, but has an unnecessarily large loss of altitude and high airspeed.

The final agent was trained with our approach, where the agent is trained to create a stall and subsequently recover from it without having termination conditions in the reward function for recovery. Figures 5 and 6 show two examples of stall recovery achieved by this method. It was found that this agent is able to recover from stall at a variety of altitudes and with a better response than those shown by the baseline agents. Figure 5 shows that the agent is able to recover from a stall by reducing the angle of attack below critical and an altitude loss of less than 1000 ft. The vertical speed is also within an acceptable range. This agent can also recover similarly when experiencing a blocked Pitot tube. However, it was observed that the agent changes its response with severe Pitot blockages at higher altitudes. Figure 6 illustrates

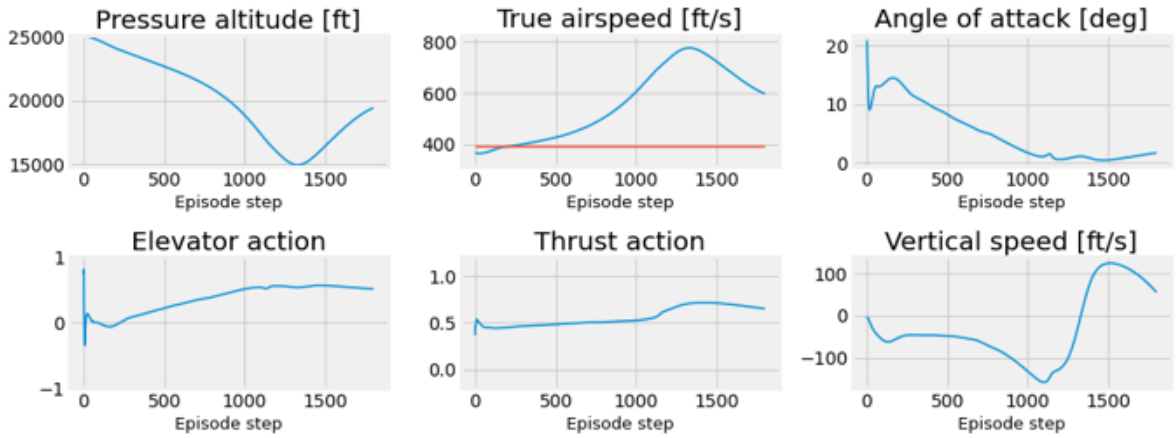

Figure 4. Stall recovery of the baseline agent without termination conditions. Random Pitot blockages were not included. The altitude and airspeed targets are shown in red.

this. The agent is able to recover, but is not able to return to the initial altitude. This is most likely due to the loss of a significant amount of information due to the blocked Pitot tube and the high precision required to manage speed at high altitudes.

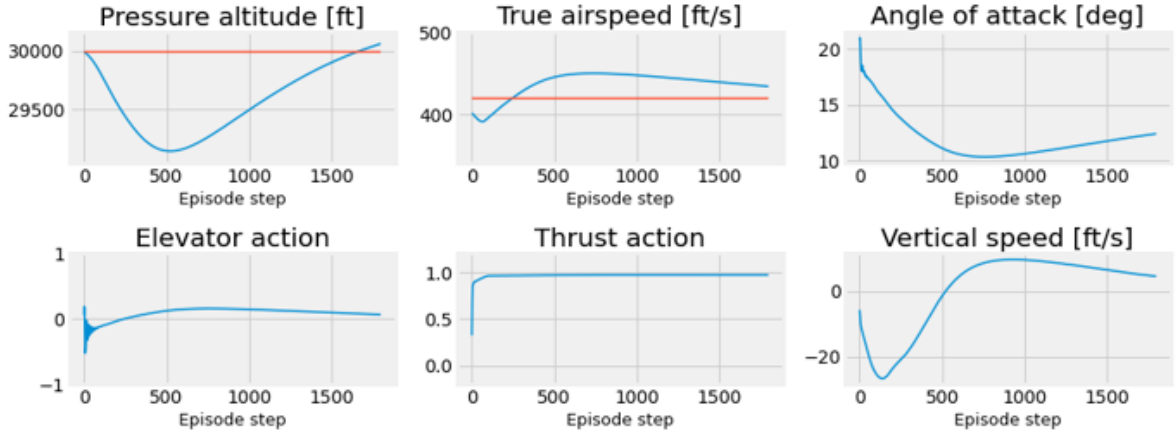

Figure 5. Stall recovery of the Reinforcement Learning agent trained to force a stall. No random Pitot blockages were included. The altitude and airspeed targets are shown in red.

The results showed that restricting AI is not always necessary and, furthermore, that too many restrictions can lead to a system that learns only shallow features causing it to be unreliable in unforeseen circumstances. More importantly, increasing the knowledge through learning related tasks improves its response in situations it has not seen before, thus reducing the unpredictability of the AI system.

Robustness and reliability of the proposed approach are dependent on the agent learning specifically how to get into a stall. Therefore, it becomes dangerous when it does not learn such behavior. However, the difference between the two tasks (forcing a stall and recovering from a stall) is significant enough to not expect them to bleed into each other. Actually, making the distinction between the two tasks known to the agent acts as an enforcement factor to reduce the probability of creating a stall when recovering. The correct response of our approach despite Pitot failures shows that robustness can be addressed by learning for it. Thus, having the ability

Figure 6. Stall recovery at a high altitude of the Reinforcement Learning agent trained to force a stall. Random Pitot blockages were included. The altitude and airspeed targets are shown in red.

to train whilst experiencing failures, shown in this paper by Pitot tube blockages, provides a method to create robustness of the system.

4. Conclusion

This paper investigated a method for automatic stall recovery that was aimed at taking more advantage of the abilities of AI than previous approaches and therefore improving towards an AI assistant in the cockpit to reduce loss of control accidents. This was explored by having a large and sparse training space in combination with making the Reinforcement Learning agent learn a related task; forcing a stall. The results were compared to several baseline agents that have stricter learning environments. Learning to create a stall resulted in successful stall recovery without the need to reduce training space or a very dense reward function. This agent was able to generalise not only to a variety of altitudes but also while experiencing the simulated Pitot tube blockages. Future work will include a deeper analysis from, for example, an aircraft performance perspective. This effective yet controversial method of allowing the AI system to know how to create a potentially dangerous stall also opens up a discussion on whether it is safer if an AI system never learns dangerous situations or if an AI system should precisely understand these dangerous situations to enhance its understanding of the environment and create an intrinsic ability to ensure that it avoids them.

References

[1] Anonymous 1996 (Junta Investigadora de Accidentes Aereos) URL `https://www.faa.gov/sites/faa.gov/files/Birgenair301_Accident_Report.pdf`
[2] Anonymous 2012 (Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile) URL `https://bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf`
[3] Anonymous 2019 (Komite Nasional Keselamatan Transportasi Republic of Indonesia) URL `http://knkt.dephub.go.id/knkt/ntsc_aviation/baru/2018%20-%20035%20-%20PK-LQP%20Final%20Report.pdf`
[4] Anonymous 2020 (The Federal Democratic Republic of Ethiopia Ministry of Transport Aircraft Accident Investigation Bureau) URL `https://bea.aero/fileadmin/user_upload/ET_302__B737-8MAX_ACCIDENT_FINAL_REPORT.pdf`
[5] European Union Aviation Safety Agency 2020 Easa artificial intelligence concept paper - proposed issue 2 URL `https://www.easa.europa.eu/document-library/easa-ai-concept-paper-proposed-issue-2`
[6] Zhen Y and Hao M 2020 *2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS)* (IEEE) pp 912–917
[7] Dally K and Van Kampen E J 2022 *AIAA SCITECH 2022 Forum* p 2078

[8] Tomar D S, Gauci J, Dingli A, Muscat A and Zammit-Mangion D 2021 *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)* (IEEE) pp 1–7
[9] Lang X, Cen F, Li Q and Lu B 2022 *Aerospace Systems* 1–10
[10] Jiang H, Xiong H, Zeng W and Ou Y 2023 *IEEE Transactions on Aerospace and Electronic Systems*
[11] Cao H, Zeng W, Jiang H, Hu H, Li C, Lu W and Xiong H 2021 *2021 China Automation Congress (CAC)* (IEEE) pp 2080–2085
[12] Schulman J, Wolski F, Dhariwal P, Radford A and Klimov O 2017 *arXiv preprint arXiv:1707.06347*
[13] Andrychowicz M, Raichuk A, Stańczyk P, Orsini M, Girgin S, Marinier R, Hussenot L, Geist M, Pietquin O, Michalski M *et al.* 2020 *International Conference on Learning Representations* URL `https://openreview.net/forum?id=nIAxjsniDzg`
[14] Berndt J 2004 *AIAA Modeling and Simulation Technologies Conference and Exhibit* p 4923

https://dspace.org/

Increasing predictability in the response of an AI-assisted stall recovery system in complex stall conditions by expanding the knowledge-base of AI

Koopman, Cynthia; Zammit-Mangion, David

2024-03-13
IOP Publishing
https://doi.org/10.1088/1742-6596/2716/1/012054
Koopman C, Zammit-Mangion D. (2024) Increasing predictability in the response of an AI-assisted stall recovery system in complex stall conditions by expanding the knowledge-base of AI. In: Journal of Physics: Conference Series, Volume 2716, March 2024, Article Number 012054

Downloaded from a DSpace repository. https://dspace.org/