CEAS-GNC-2022-005

CEAS EuroGNC 2022
"Conference on Guidance, Navigation and Control"
3-5 May 2022 @ Technische Universität Berlin, Germany

Towards Stall Detection and Learning-Based Stall Prevention for Glider Aircraft

Franziska Hein Research Associate, University of Stuttgart, Institute of Flight Mechanics and Control, 70569, Stuttgart, Germany. franziska.hein@ifr.uni-stuttgart.de

Stefan Notter Research Associate, University of Stuttgart, Institute of Flight Mechanics and Control, 70569, Stuttgart, Germany. stefan.notter@ifr.uni-stuttgart.de

Jan Axthelm Research Associate, University of Stuttgart, Institute of Flight Mechanics and Control, 70569, Stuttgart, Germany. jan.axthelm@ifr.uni-stuttgart.de

Walter Fichter Professor, University of Stuttgart, Institute of Flight Mechanics and Control, 70569, Stuttgart, Germany. walter.fichter@ifr.uni-stuttgart.de

ABSTRACT

Glider pilots often struggle with the problem of stalling during the winch start, turning flight or final approach, which in most cases leads to serious or even fatal accidents. In general aviation, stall warning systems do exist, but they are not applicable and practical in all aircraft. On the one hand, the estimation of the angle of attack and the required measurement variables are not easy to obtain, and on the other hand, these systems often warn very early and are therefore considered rather annoying by some pilots. In the following work, two different types of angle of attack estimation approaches are presented, one of which does not require the measurement of aerodynamic parameters and thus can be installed easily in existing aircraft. Both methods are validated with test flight data. The estimation of the angle of attack is the most important part of stall detection, whereas stall prevention is realized by a controller algorithm. The difficulty is to intervene on time, but not too early, to prevent the stall. Therefore, the idea of a reinforcement learning approach is proposed, which learns to detect an imminent stall in time and intervene with minimal control effort.

Keywords: Stall Detection, Envelop Protection, Stall Prediction, Stall Prevention, Anti Stall System, Reinforcement Learning

1 Introduction

Loss of control due to stall situations has been a major cause of accidents, especially fatal accidents, since the beginning of general aviation. While stall warning and assistance systems already exist in general aviation, in gliders the pilot alone is responsible for detecting and preventing the stall. This frequently leads to serious accidents, as can be read in the *EASA Annual Safety Report for Gliders* [1]. Figure 1 shows the occurrences of different risk scenarios over the last five years. It can be seen, that stall leading to loss of control has both the highest risk potential and the most occurrences.

In contrast to a fly-by-wire control system from military or commercial aviation, the aim here is to create a system that assists the pilot but does not affect the main control system. The current state-of-the-art technology in general aviation is a stall warning system which typically indicates to the pilot

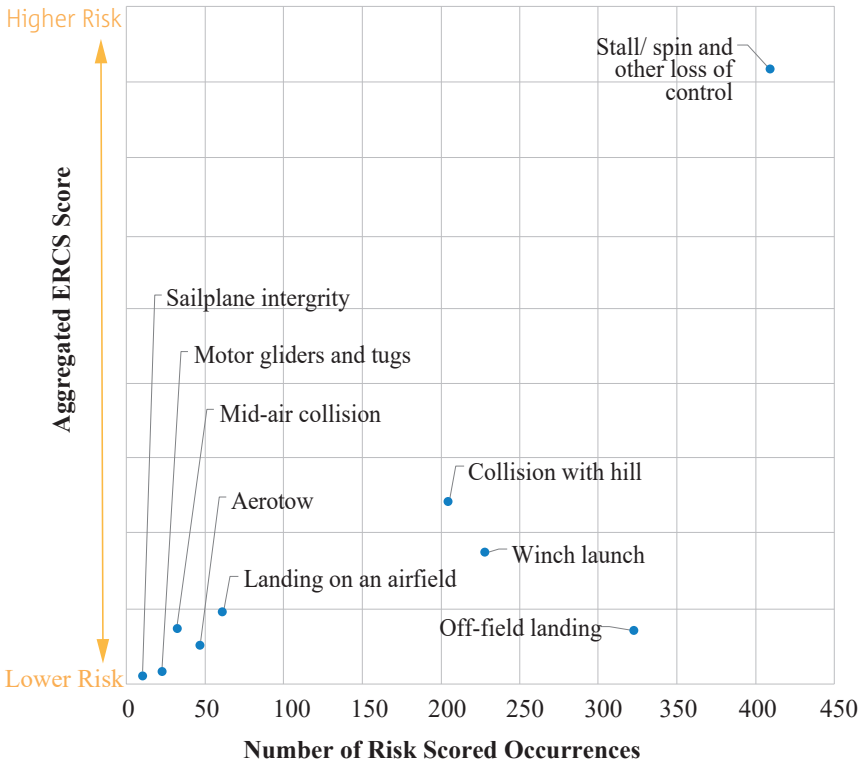

Fig. 1 Sailplane risk groups by aggregated ERCS score and number of risk-scored occurrences involving sailplanes [1].

that the airspeed is close to stalling speed. This warning can be either audible, visual, or haptic. In either case, the pilot must respond to the warning himself and recover the aircraft safely. In literature, different approaches exist to find a method that is more accurate in predicting an approaching stall. An artificial neural network (ANN) featuring a *Long Short-Term Memory* (LSTM) was proposed for predicting an impending stall before the classical warning system even reacts to give the pilot more time to take action [2].

Passive stall prevention methods can be realized by aerodynamic devices (e.g. stall fences, stall shields [3]) which help to reduce the risk of stalling. Active stall prevention mechanisms directly yield the deflection of a control surface (e.g. elevator, additional spoiler). The command must be triggered by the pilot or an automated assistance system. A control algorithm that overrides the pilot input for the elevator in order to keep the angle of attack under a certain limit is proposed in [4]. However, sharing the elevator between the pilot and the anti-stall system is often not desired as it means a major intervention in the normal flight behavior and the pilot loses his control authority.

Within the *Anti-Stall Assistance System* (ASASys) project, a system for stall detection and active stall prevention for glider aircraft is developed. A new feature of the proposed *Anti-Stall Assistance System* is that in addition to warning the pilot (which is still possible), the system directly assists the pilot in restoring a safe flight condition without restraining the elevator control of the pilot. Therefore, an additional, automatically actuated spoiler is integrated in the undersurface of the horizontal stabilizer. The combined problem of finding the perfect time to intervene while not disturbing normal flight situations is an interesting task to apply a reinforcement learning (RL) method within the scope of aerospace control problems. Compared to other methods, it is possible to "learn" the relationship between several state variables and thus react appropriately to different stall approximations. For the detection of the stall, an embedded wall shear sensor (EWSS) is installed on each wing surface to estimate the angle of attack (AoA). The estimate is used among others to detect imminent stall situations on time. Additionally,

a second method to estimate the angle of attack by only using measured accelerations is developed. Both methods are verified with real flight data. The *Anti-Stall Assistance System* is comparable to assistance systems in the automotive sector, such as the *Anti-lock Braking System* (ABS) or the *Electronic Stability Program* (ESP). On the one hand, this increases acceptance in the market of non-commercial users and pilots. On the other hand, a system is created that is easier to certify and can be retrofitted to existing aircraft.

The paper presented below is structured as follows. The experimental glider and the simulation of the modified aircraft are described in Sec. 2. In Sec. 3 the two methods to detect stall situations are presented and the experimental results and theoretical assumptions are compared to real flight data. In Sec. 4, we outline the control problem and present a novel strategy for preventing stall situations through a reinforcement learning approach.

2 Real Flight System and System Model

In this section, the setup of the aircraft with all its additional components is described. Afterwards, the structure of the simulation and how the new components were integrated are explained.

2.1 Experimental Aircraft

The experimental aircraft is an *Arcus E* which was modified as follows. There are three major changes to the glider, shown in Fig. 2, which are necessary for the study: the EWSS to detect stall, the electrical system for the transmission of the signal, and the additional flaps on the elevator to prevent stall. On the left hand side of the figure, additional equipment is shown, which was installed for validation purposes, e.g. the air data probe and the pitot statics. The electrical system was installed by running cables through the aircraft to provide a connection from the cockpit to the actuator system on the elevator. The elevator is adjustable and therefore the flaps move with the elevator as well as separately from it.

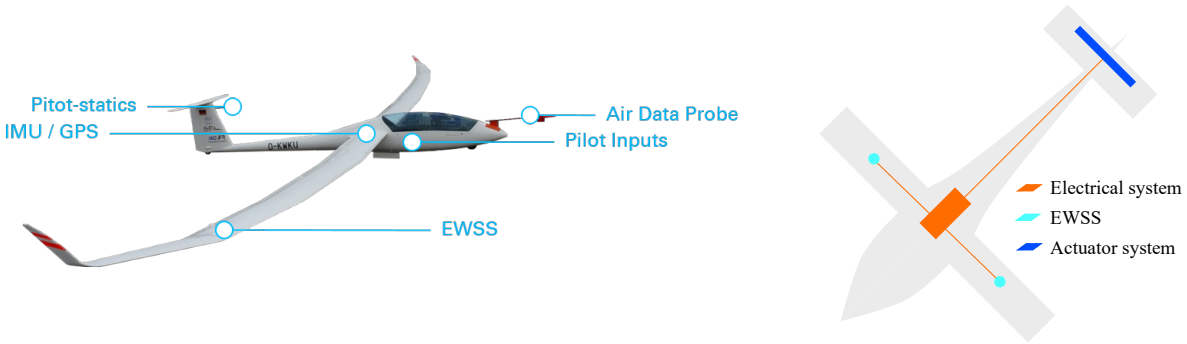

Fig. 2 Flight test glider, equipped with additional sensors and flaps.

The mechanism and arrangement of the additional flaps, so-called spoilers, can be seen in Fig. 3. The spoilers are integrated underneath the elevator and can only be deflected in one direction to produce a pitching moment. On each side of the elevator three flaps are installed. The embedded wall shear sensors were built and analyzed by the Institute of Aerodynamics and Gas Dynamics (IAG) of the University Stuttgart [5] and are placed on the upper side of the wings as shown in Fig. 4. Their basic working principle will be explained in the next chapter.

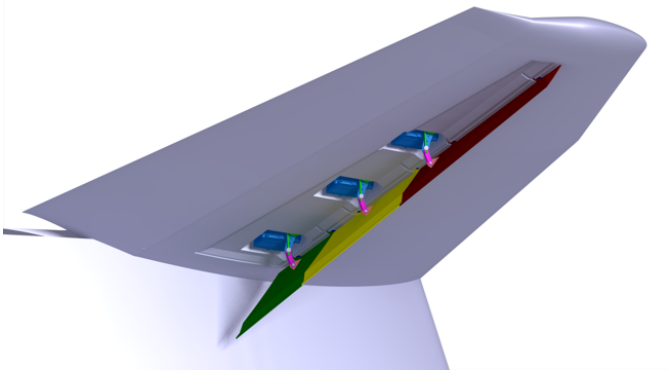

Fig. 3 Additional flaps on the elevator to prevent stall.

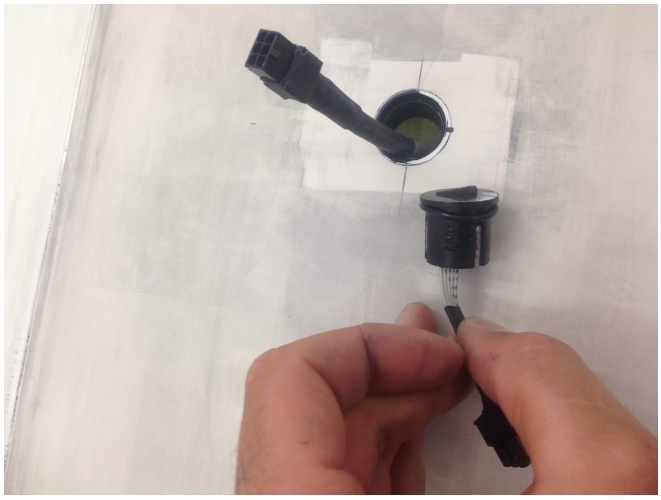

Fig. 4 Installation of the EWSS (picture from the IAG).

2.2 Simulation Environment

The simulation environment is implemented in *MATLAB/Simulink*. We use a 6 DOF model for the implementation of the nonlinear flight dynamics, represented by the following equations:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}), \tag{1}$$

with:

$$\mathbf{x} = [\mathbf{p}, \mathbf{v_f}, \Phi, \omega], \ \mathbf{u} = [\xi, \ \eta, \ \zeta, \ \beta, \ \gamma] \tag{2}$$

The aerodynamic parameters of the *Arcus E* were determined using *XFLR5*. Additional moments of the spoiler are calculated using wind tunnel data. Therefore, the lift coefficient of the whole elevator with and without the additional spoiler flaps was measured and integrated in the simulation using lookup tables. The additional lift coefficient of the spoiler depends on the angle of attack at the elevator, the current position of the elevator and the current position of the spoiler. The additional moment induced by the spoiler is then calculated using the lever arm of the spoiler. For the estimation of the angle of attack, the common method using the measurement of a multi-hole probe is used for validation purpose. Additionally, the new methods which will be explained in the following section are also integrated in the simulation environment.

3 Stall Detection

In this section, two methods for the detection of a stall situation are introduced. First, the EWSS data from the wind tunnel tests are evaluated to link the measured pressure drop to the angle of attack. The second method to estimate the angle of attack uses the acceleration measured by the inertial measurement unit (IMU). Afterwards, both methods are verified in flight tests using air data from a multi-hole probe and the IMU.

3.1 Embedded Wall Shear Sensor

The EWSS measures the difference between the total air pressure and the static air pressure in the boundary layer at the installation location and can thus detect if the airflow is laminar/turbulent-attached or laminar/turbulent-detached. To ensure that the measurement is independent of the airspeed, the pressure difference is normalized with the dynamic pressure Q_{Inf}. It was found that the normalized pressure difference $\Delta p_{\mathrm{EWSS}}/Q_{\infty}$ has the value of around 0.2 if the sensor is in the separated flow for high angles of attack. This value can be used to find the maximum AoA given by the EWSSs.

Fig. 5 shows the normalized pressure difference over the angle of attack, measured in the wind tunnel with a turbulator placed at 5 % / 100 % of the profile depth. It was found that between 9° and 12.5° the signal drops linearly with increasing AoA and is mostly independent from the flap position β and the transition point between laminar and turbulent flow. This linear dependency can be used to calculate the local AoA on the wing.

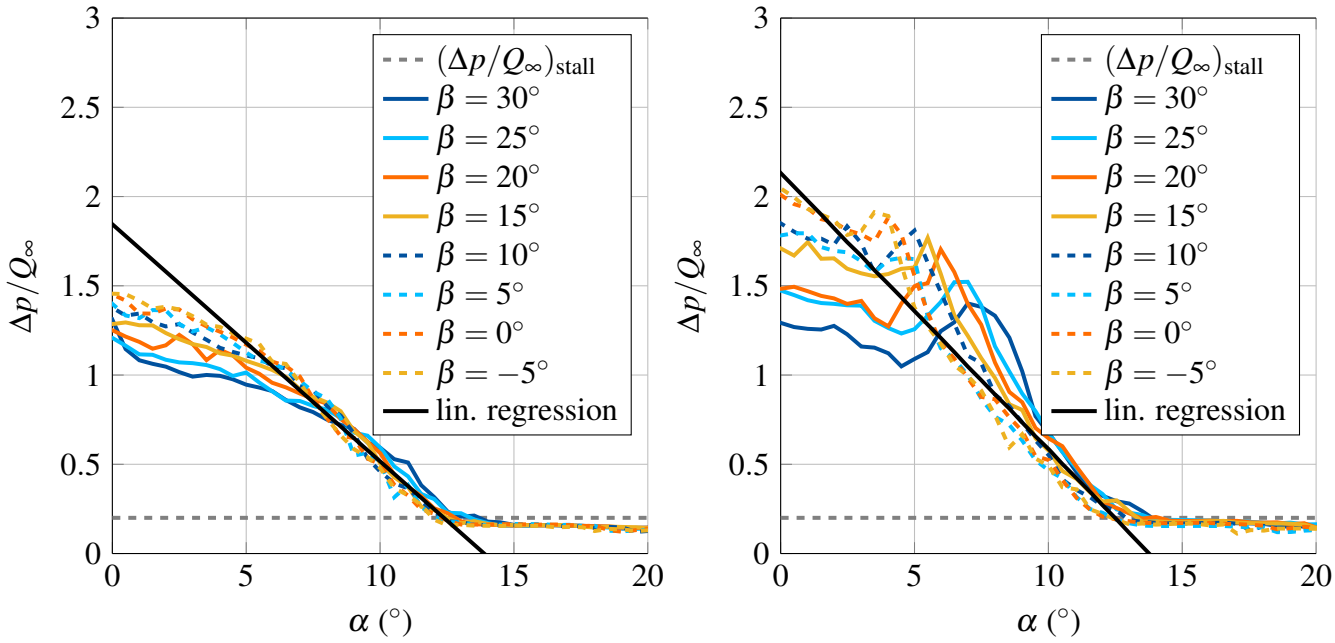

Fig. 5 Wind tunnel measurement of an embedded wall shear sensor at the IAG. Plotted is the difference pressure measured by the EWSS and normed by the dynamic pressure in dependence of the AoA for different flap positions and two different turbulator positions. The left plot shows the data for a turbulator placed at 5 % **of the chord length and on the right side for a turbolator placed at** 100 %**.**

Another important dependency can be seen in Fig. 6. The lines show the AoA limit for different detecting methods. The dark blue line corresponds to the EWSS signal being zero whereas the light blue line shows the maximum AoA coming from the maximum lift coefficient. For a flap angle of around 20° the AoA for the maximum lift coefficient is lower then the one at which the EWSS signal is zero. This has to be taken into account when using the EWSS signal to detect the maximum angle of attack for high flap angles. Otherwise, the stall is detected too late when the maximum lift coefficient already starts dropping.

This behavior will be verified with flight test data in section 3.3. Before that, the theory of another method to determine the angle of attack is introduced.

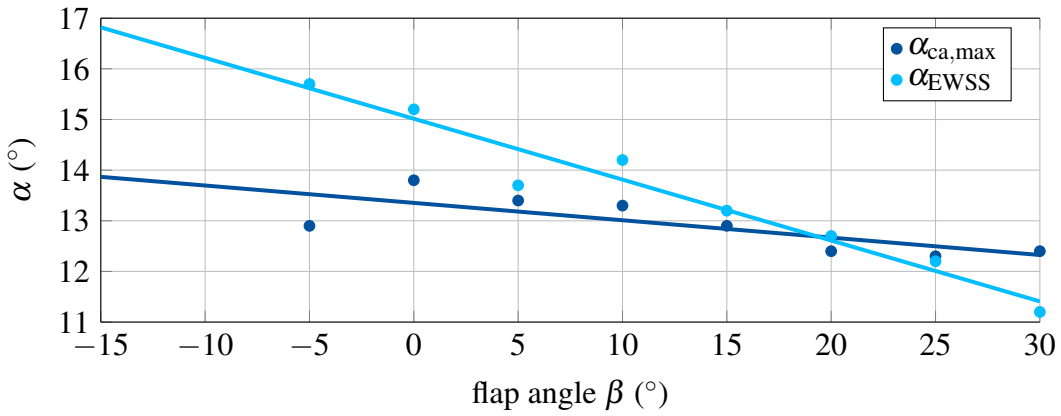

Fig. 6 Relationship between the angle of attack of the maximum lift coefficient and the angle of attack at which the EWSS signal becomes zero.

3.2 Estimation using Inertial Acceleration

Besides using the EWSS, the AoA of gliders can also be estimated without aerodynamic measurements. This is of particular interest with respect to retrofitability into existing gliders and is based on the dependence between the AoA and the acceleration ${}_f a_x$ and ${}_f a_z$ measured by the inertial measurement unit.

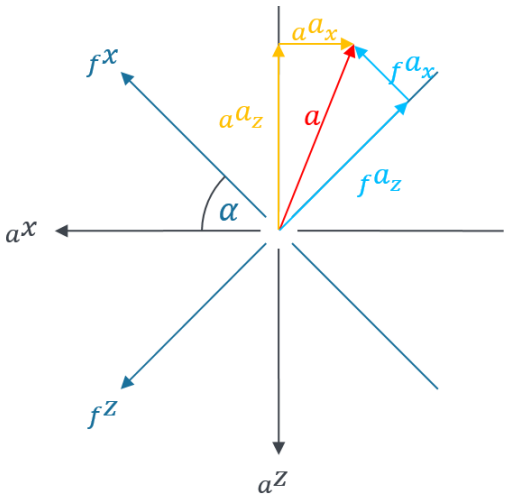

Fig. 7 Basic idea of the angle of attack estimation from inertial measurements in gliding flight.

The basic idea is shown in Fig. 7. While gliding, accelerations measured by the IMU are solely caused by aerodynamic forces. The following preliminary assumptions are made:

- Zero sideslip angle.
- Only valid for gliders, i.e. only gravity and aerodynamic forces are acting on the aircraft.

With this, the following relation is valid:

$$\alpha = \arctan\left(-\frac{{}_f a_x}{{}_f a_z}\right) + \arctan\left(-\frac{{}_a a_x}{{}_a a_z}\right), \tag{3}$$

with ${}_aa_x = C_D \frac{\bar{q}S}{m}$ and ${}_aa_z = C_L \frac{\bar{q}S}{m}$. Omitting $\frac{\bar{q}S}{m}$ leads to:

$$-\frac{{}_fa_x}{{}_fa_z} = \tan(\alpha - \arctan(\frac{C_D(\alpha)}{C_L(\alpha)})) =: f(\alpha) \tag{4}$$

The dependencies of C_D/C_L on α are known and so is $f(\alpha)$ then. Assuming f is invertible, it follows:

$$\alpha = f^{-1}(-\frac{{}_fa_x}{{}_fa_z}) \tag{5}$$

3.3 Verification with Flight Data

In order to compare the measurements of the EWSS sensors in the wind tunnel with the real flight, the angle of attack estimated in flight must be determined at the location of the sensors. For this purpose, the local airflow velocity in the body-fixed system $v_{A,f}$ is calculated and then corrected by the cross product of the rotation rate ω and the lever arm of the EWSS position r_{EWSS}:

$$v_{A,f,\mathrm{EWSS}} = v_{A,f} - r_{\mathrm{EWSS}} \times \omega \tag{6}$$

In addition, for zero side slip angle, the dihedreal Γ must be taken into account. For this purpose, $v_{A,f,\mathrm{EWSS}}$ is transformed into an "airfoil-fixed" system with the index w, which corresponds to a rotation T_{wf} about the body-fixed axis by Γ:

$$v_{A,w,\mathrm{EWSS}} = T_{wf} \cdot v_{A,f,\mathrm{EWSS}} \tag{7}$$

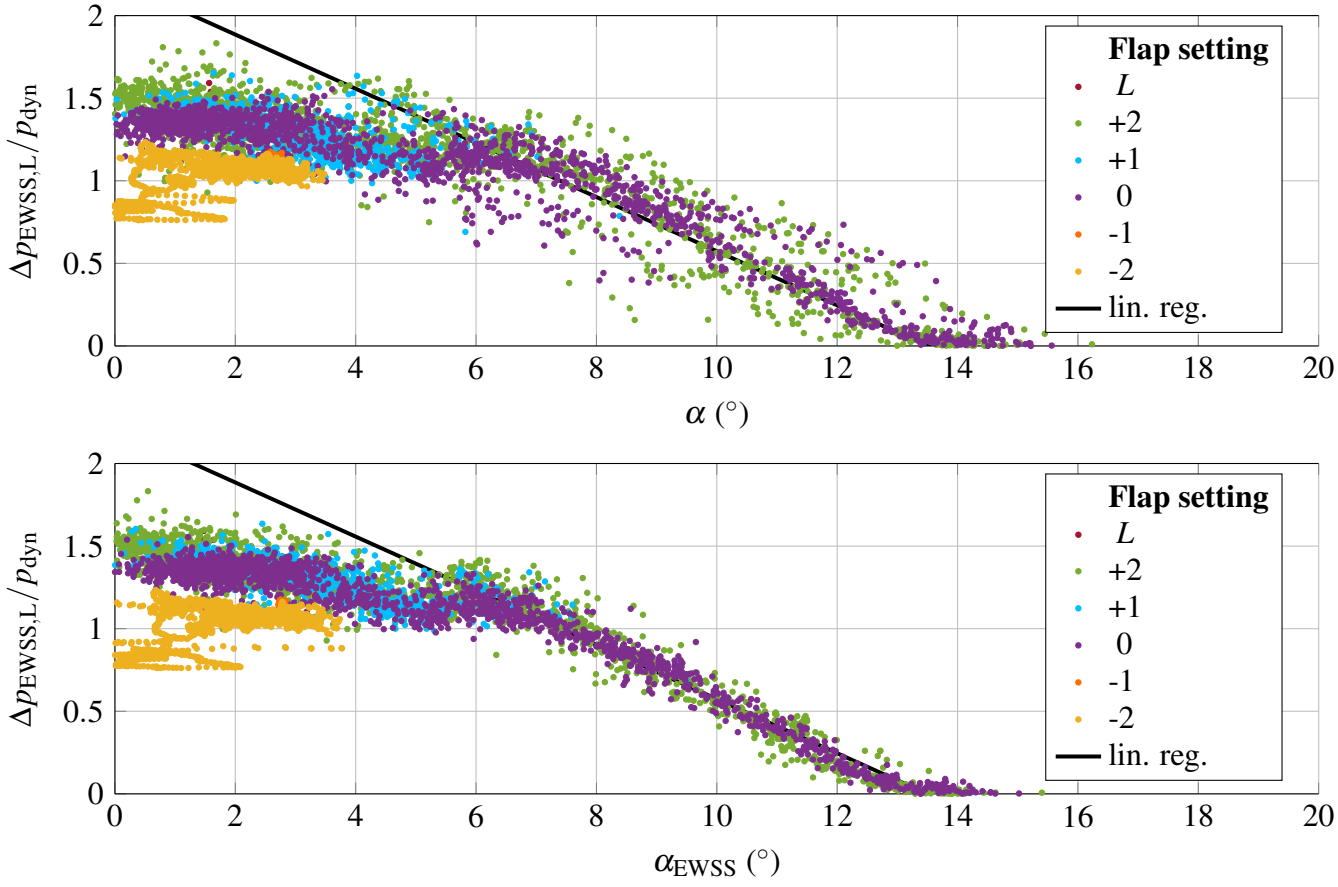

Fig. 8 Top: Uncompensated data from the left EWSS. Bottom: Compensated Data from the left EWSS.

Fig. 8 shows the measurement from the left EWSS, which was measured during a test flight, compared to the local angle of attack at the left wing at the position of the sensor. The black line shows the fitted data from the wind tunnel tests.

As can be seen, the measurements in the flight test coincide with the data from the wind tunnel measurements in the range between $8° \leq \alpha \leq 13°$. Thus, the EWSS are suitable for the detection of high angles of attack. This means that this method is not only suitable for detecting a stall but can also be used to identify an approaching stall.

To verify the second method for angle of attack determination, flight measurement data from quasi-stationary flights with increasing angle of attack for several flap positions were considered. The result is shown as a dependency of $-\frac{a_X}{a_z}$ on α in Fig. 9. A linear dependency between approximately 3° and 11° is obtained for the test aircraft at all flap positions. Therefore, with a simple correction for the differences due to the flap, the angle of attack can be determined very easily in the relevant range. However, this method is only valid for gliders.

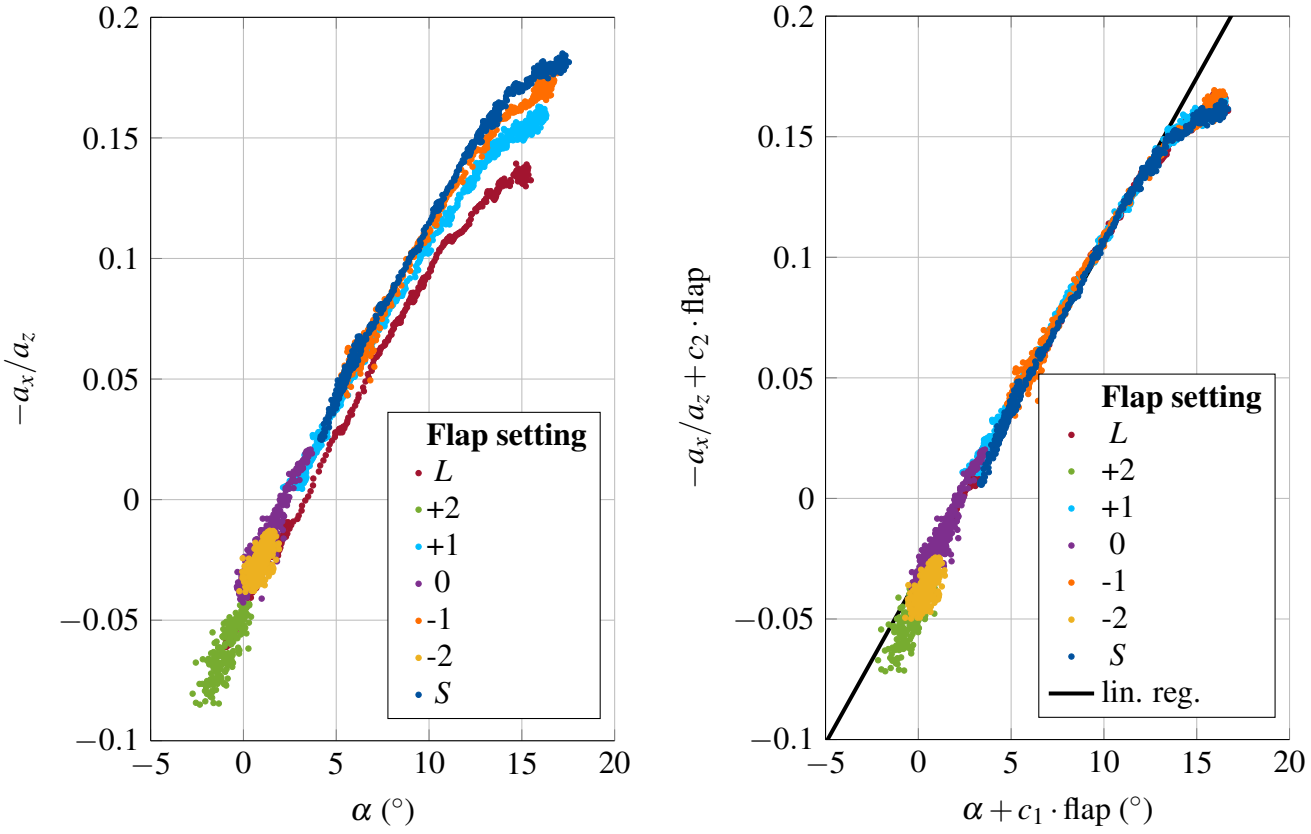

Fig. 9 Dependency of $-\frac{a_X}{a_z}$ on the AoA for different flap positions measured with the *Arcus E*. On the left hand side, the measurement was corrected with a constant multiplied by the flap position.

4 Anti-Stall Prevention Algorithms

In this section the anti-stall control system is derived. First of all, a short problem description will be given to justify the chosen approach and give an idea of the obstacles encountered in the development of this controller. Afterwards, the control algorithm using reinforcement learning is derived and the simulation results are shown.

4.1 Problem Description

The stall prevention system has to keep the angle of attack under a critical value by deflecting the additional spoilers. However, it must be taken into account that the spoiler can only apply a certain maximum pitching moment limited by the full spoiler deflection. The additional pitching moment not only depends on the deflection of the spoiler but also on aircraft state, the flap position and the elevator position. Fig. 10 shows the additional lift coefficient of the stabilizer over the spoiler angle γ for different elevator positions. For positive elevator angles (downward deflection, i.e., stick pushed) the additional lift coefficient even drops if the spoiler is deflected between $0° - 20°$. A major challenge is caused by the behavior for both large spoiler angles ($\gamma > 45°$) and a high angle of attack, where the additional lift coefficient drops. The characteristics shown make it obvious that the effect of the spoiler is nonlinear. Additionally, the effect of the spoiler is limited to positive $\Delta C_{L,Elev}$. Thus, a linear controller approach is likely to show inefficient performance.

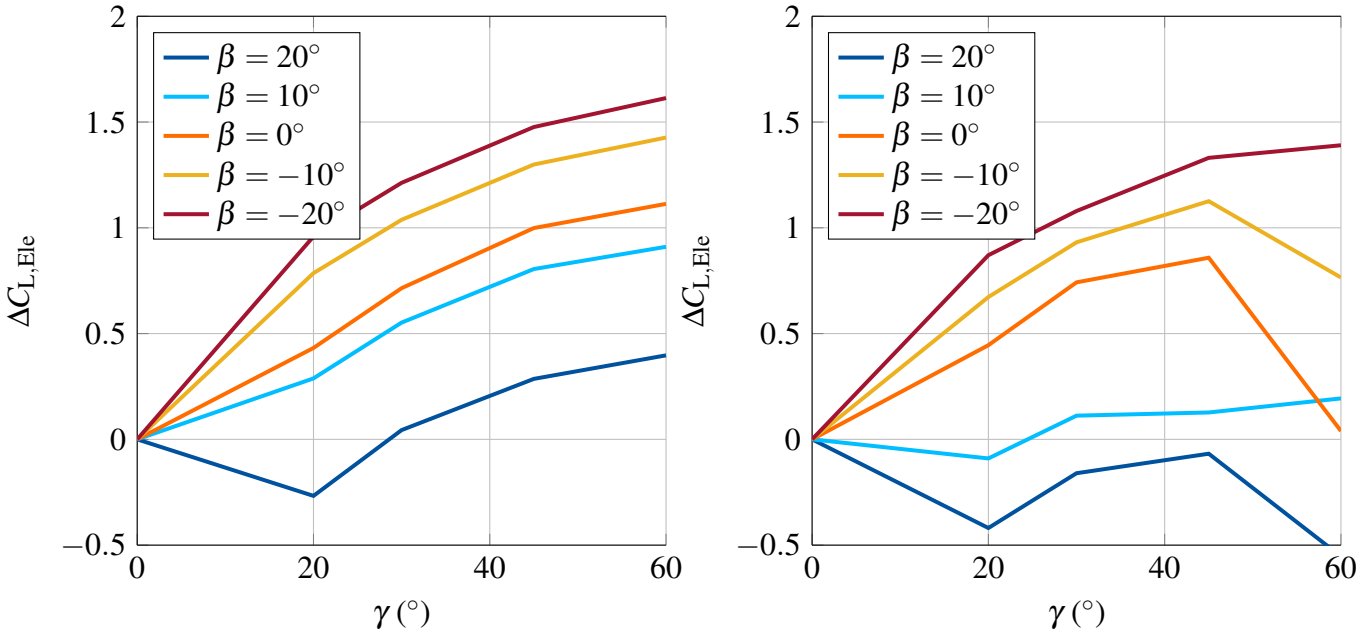

Fig. 10 Additional lift coefficient caused by the spoiler, plotted against the spoiler angle for different elevator positions. Left for $\alpha = 4°$ and on the right for $\alpha = 10°$.

Nevertheless, different control approaches such as a proportional–derivative (PD) controller or direct nonlinear inversion were tested but without achieving the desired anti-stall behavior for a range of different flight conditions. It was found that a proportional controller, which aims to keep the angle of attack under a certain limit, only works in certain flight conditions. A slow drifting to higher AoA coming from straight flight, was preventable. Nevertheless, a critical flight situation for this controller type can easily be created as follows: A positive elevator command yields the glider starting to dive and picking up speed. After pulling the nose up, again, the angle of attack quickly rises beyond $\alpha_{\text{stall}} = 12°$ and the glider stalls. The zero-spoiler-command baseline situation is depicted in Fig. 11. The same situation subject to a PD-controller command is shown in Fig. 12. It can be seen that the angle of attack still rises quickly even though the spoiler command γ is at its maximum. Thus, the stall of the glider cannot be prevented. Lowering the threshold of α_{stall} would help in this case but lead to a premature deflection in other scenarios.

A possible solution to this problem is a control algorithm, which can predict an imminent stall situation by using more than one variable. Deep reinforcement learning yields such a nonlinear, predictive control law. With the prediction shifted to the training phase, evaluating a policy resulting from reinforcement learning is computationally light when compared to a full-fledged nonlinear model predictive control law [6].

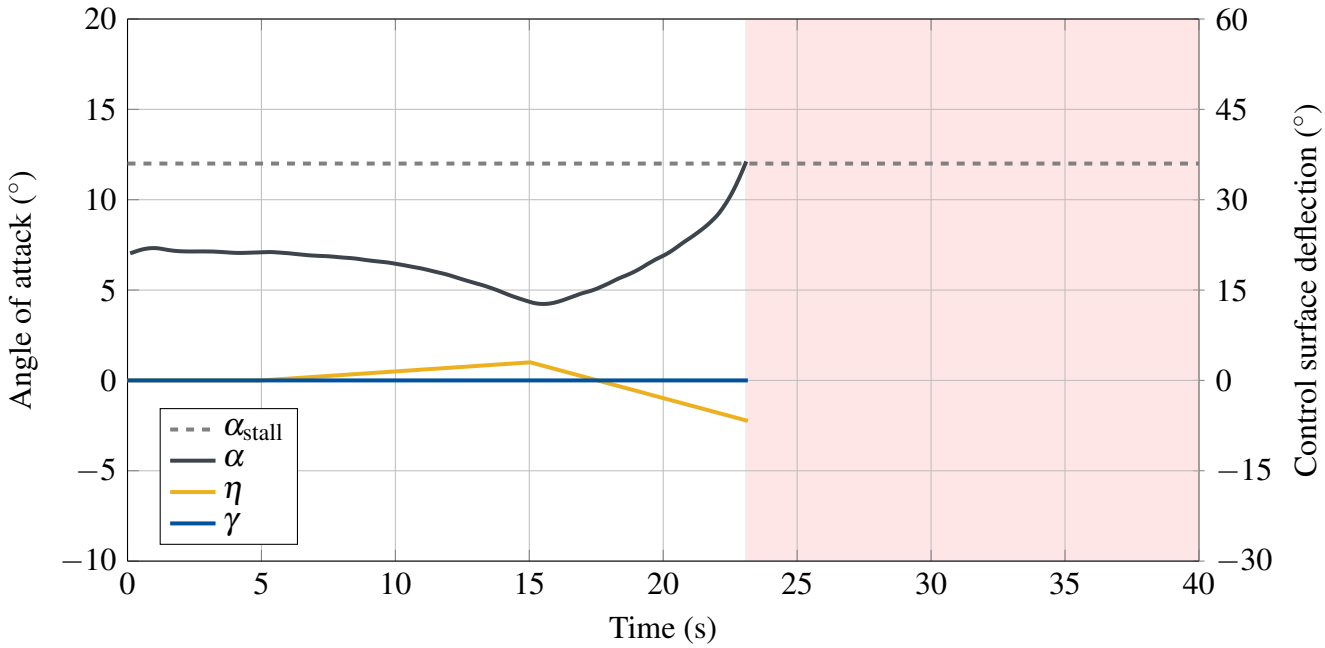

Fig. 11 Baseline: Evolution of the angle of attack for a reference elevator input without additional spoiler deflection.

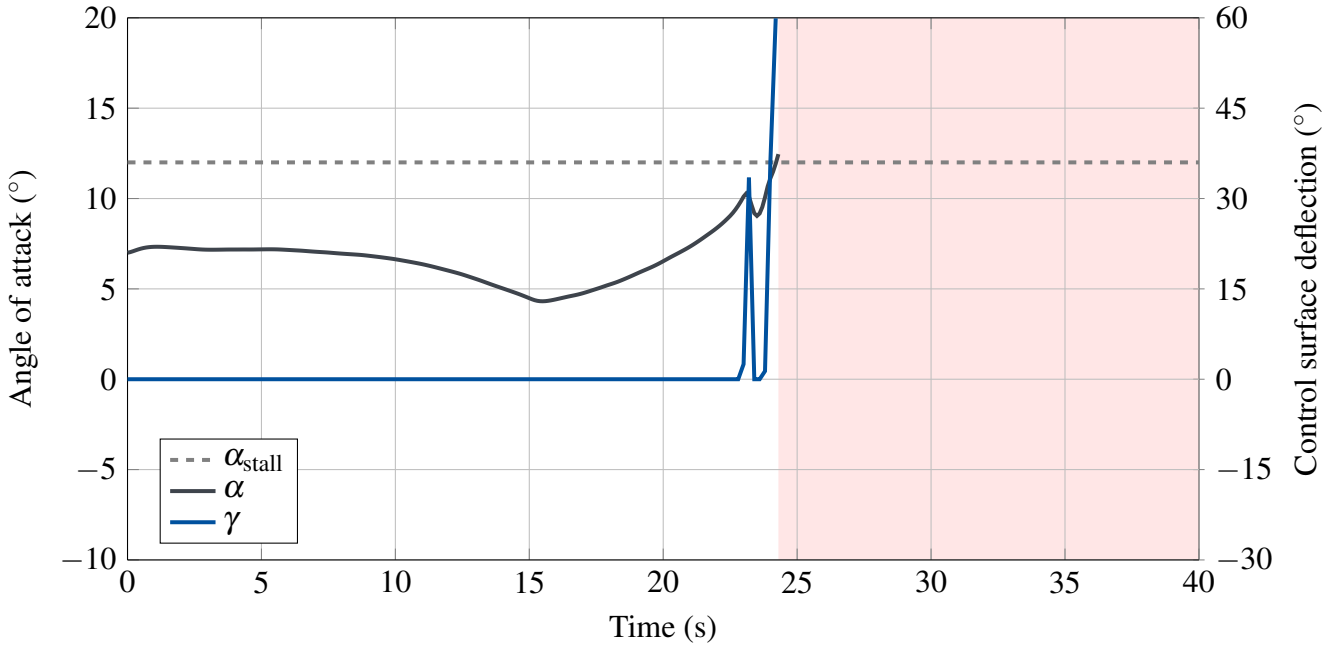

Fig. 12 PD-controller: Evolution of the angle of attack and spoiler deflection commanded by a PD-control approach for the same reference elevator input as in Fig. 11. The spoiler deflection does not prevent the glider from stalling.

4.2 Reinforcement Learning Control Approach

In this section, a first approach for a reinforcement learning algorithm is presented using the state variables of the longitudinal motion (α, θ, V_A) in order to compute a control command for the spoiler deflection. Fig. 13 shows the working principle of the RL algorithm.

RL is a machine learning method in which an agent learns a desired behavior solely by interacting with the environment. To do so, it receives feedback from its own actions, called reward, in order to assess its behavior. The state-action pair is mapped on a certain reward, which is maximized by the

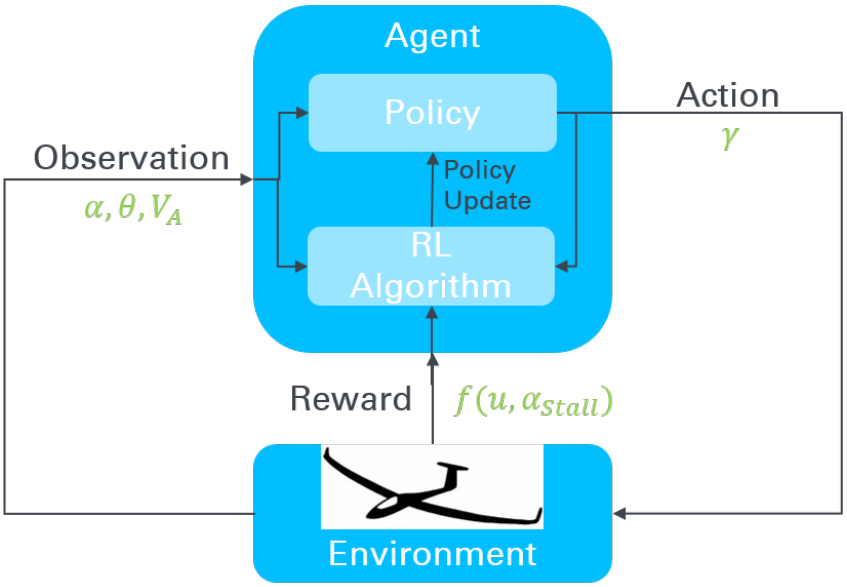

Fig. 13 Structure of the reinforcement learning algorithm

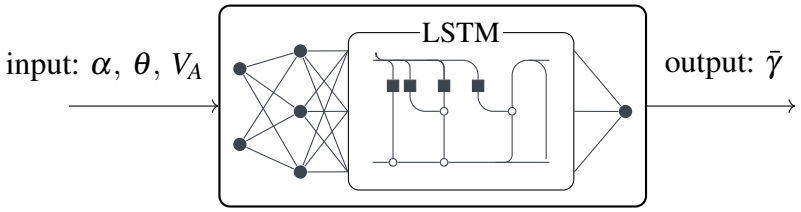

Fig. 14 Recurrent policy architecture for stall prevention (adapted from [8])

algorithm. Compared to supervised learning, RL works with the trial and error principal and thus does not need data generated in advance.

The RL framework applied in this paper is based on the setup from [7, 8][1]. Therein, *Proximal Policy Optimization* (PPO) optimizes a policy for an autonomous soaring control problem. The objective within the paper at hand is to validate the applicability of an RL approach to the problem of stall prevention in a first, preliminary attempt.

PPO is a method which was introduced in 2017 [9]. It is a new approach in the family of policy gradient methods and has the efficiency and robustness of *Trust Region Policy Optimization* while being easier to implement. The policy network chosen features a *Long-Short Term Memory* cell to account for information embedded in a sequence of measurements [10, 11]. The most relevant reinforcement learning (hyper-)parameters applied for training the control policy for stall prevention are listed in Table 1.

The reward function for this problem depends on the control input u (also called "action"), which is the normalized spoiler angle $\bar{\gamma}$, and the critical angle of attack. The state s is mapped on the reward function $R(\cdot)$ as follows:

$$R : \mathscr{S} \to \mathbb{R},\ s \mapsto r = 1 - \frac{1}{2}\bar{\gamma} \tag{8}$$

The reward function penalizes nonzero spoiler commands to prevent unnecessarily large spoiler deflections which perturb the aircraft characteristics. Transgression of the critical AoA is not penalized by a negative reward but since the agent is only rewarded as long as the stop criterion ($\alpha < 12°$) is not

[1]The source code of the parent reinforcement learning framework is publicly available on *GitHub*: `https://github.com/ifrunistuttgart/RL_Integrated-Updraft-Exploitation`

Table 1 Reinforcement learning (hyper-)parameters

Parameter	Value
Batch size	4096
Sequence length	256
Learning rate, actor	1e−5
Learning rate, critic	1e−4
Discount factor	0.99
PPO clipping value	0.2

met, the agent tries to keep the AoA under the stop criterion. Due to the structure of the RL algorithm, the action, i.e. the spoiler command, varies between 0 and 1. 0 corresponds to zero deflection and 1 represents full spoiler deflection.

The environment was converted from *Simulink* to a *Python Gym* environment [12], which is then evaluated by the RL algorithm to train the agent. During 40 s long test episodes, the agent must prevent a stall. Preliminary results for the trained agent subject to the RL algorithm proposed are provided in the next section 4.3.

4.3 Preliminary Results

Fig. 15 shows the behavior of the trained agent after 1500 policy updates. Both the evolution of the angle of attack and the spoiler deflection are depicted. The control policy resulting from reinforcement learning deflects the spoiler three times to prevent the glider from stalling within the time span of 40s. The trained agent reacts to an increase in the angle of attack when the value is on a high level already. Unlike a PD-command, however, the proposed control policy is nonlinear and predictive. Earlier results showed, that a continuous spoiler command leads to earlier deflections and hence disturb the normal flight characteristics.

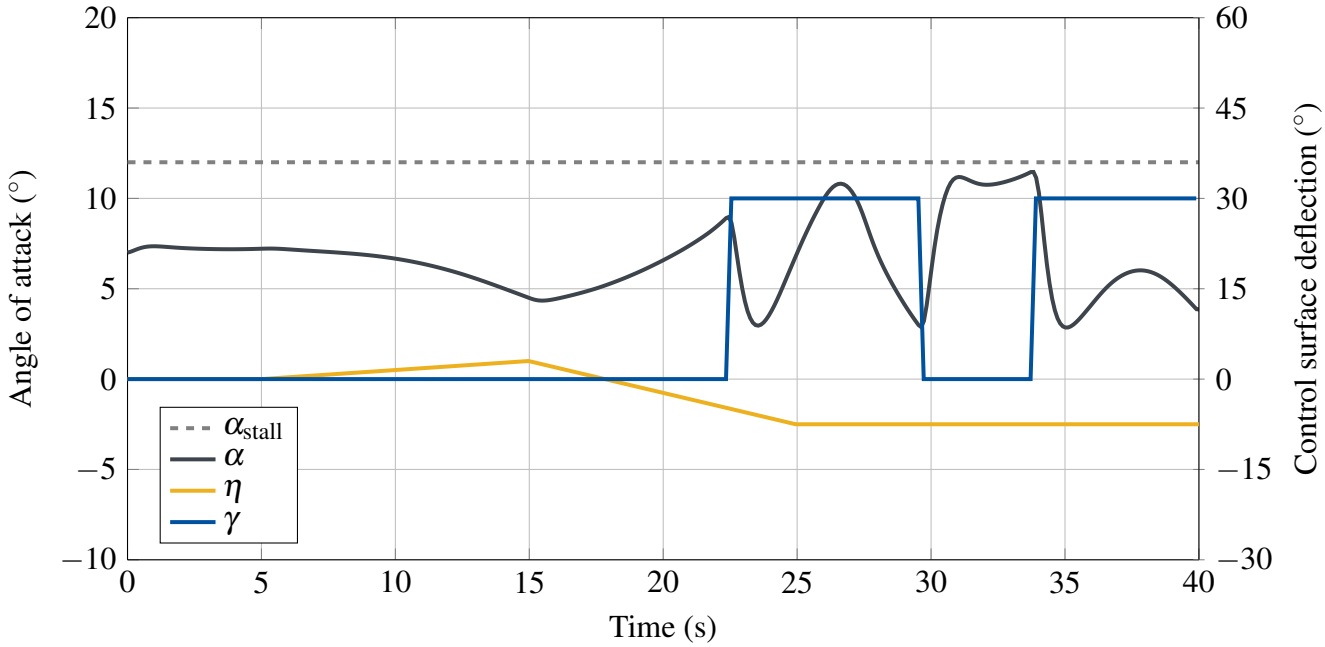

Fig. 15 RL approach: Evolution of the angle of attack and spoiler deflection commanded by the trained agent for the same reference elevator input as in Fig. 11. The spoiler deflection prevents the glider from stalling.

The trained policy is capable of preventing the glider from stalling for the given, critical pilot elevator input. Additionally, it does only deflect right before a stall situation. In order to prevent an unreasonably early spoiler deflection, the LSTM hidden state is "burned-in" for a few seconds after initialization.

It must be taken into account, that the given flight trajectory is a very difficult one. As this situation can be handled by the RL algorithm, easier ones can be, too. During the learning process, we noticed a major challenge for solving this optimization problem: The margin between the best solution of intervening as late as possible to prevent the stall and a fatal stall situation is extremely small. The discount factor must be chosen small enough to make the agent not pushing the nose down way too early. If the discount factor is chosen too small, on the other hand, the trained agent does not intervene early enough to prevent the stall.

5 Conclusion

In this paper, two new methods for detecting the critical angle of attack of a glider plane were derived. For the method using embedded wall shear sensors (EWSS), the measured pressure drop was connected to the corresponding angle of attack for different flight configurations. For the other method, a set of equations was derived to estimate the angle of attack of a glider plane. With the help of flight data, both methods were validated and can be used further for the detection of an impending stall. The method using the EWSSs is valid for different types of airplanes whereas the second method is only valid for glider aircraft. The flight test data and wind tunnel data concerning the EWSSs only coincide for an angle of attack between $8° - 13°$. Future flight tests will reveal whether one of the methods is superior for estimating the angle of attack.

The paper proposes a reinforcement learning (RL) approach to solve the highly nonlinear stall prevention problem. Preliminary results show that the RL approach offers a possible solution to the problem. We showcased that a recurrent, deep artificial neural network (ANN) control policy prevents the glider from stalling in a particularly critical flight situation while also not deflecting the spoiler unnecessarily early. Future work will refine the RL approach through domain randomization to target the sim-to-real transfer. The objective is to reduce the spoiler deflection in different flight scenarios while safely preventing the glider from stalling. The real-time evaluation of the ANN control policy is feasible on low-cost off-the-shelf embedded hardware. The authors are working towards flight test demonstrations of the approach proposed.

Acknowledgments

The topic presented has been investigated within the LuFo V-3 project *ASASys*, funded by the German Federal Ministry for Economic Affairs and Energy (BMWi). The financial support is gratefully acknowledged.

References

[1] European Aviation Safety Agency. *Annual safety review 2021*. Publications Office, 2021. DOI: 10.2822/071257.

[2] Tahsin Sejat Saniat, Tahiat Goni, and Shaikat M. Galib. LSTM recurrent neural network assisted aircraft stall prediction for enhanced situational awareness, 2020. DOI: 10.48550/ARXIV.2012.04876.

[3] J.A. Stoop and J.L. de Kroes. Stall shield devices, an innovative approach to stall prevention? In *Air Transport and Operations*, pages 55–68, jan 2021. DOI: 10.3233/978-1-61499-119-9-55.

[4] Matthys Michaelse Basson. *Stall prevention control of fixed-wing unmanned aerial vehicles*. PhD thesis, Stellenbosch: University of Stellenbosch, 2010.

[5] Werner Scholz, Sebatian Leis, Valentin Petters, Werner Würz, Jan Axthelm, and Walter Fichter. Asasys – anti-stall assistant system for sailplanes. In *XXXV OSTIV Congress*, pages 19–22, jul 2021.

[6] Stefan Notter, Markus Zürn, Pascal Groß, and Walter Fichter. Reinforced Learning to Cross-Country Soar in the Vertical Plane of Motion. In *AIAA Scitech 2019 Forum*, 2019. DOI: 10.2514/6.2019-1420.

[7] Stefan Notter, Fabian Schimpf, and Walter Fichter. Hierarchical Reinforcement Learning Approach Towards Autonomous Cross-Country Soaring. In *AIAA Scitech 2021 Forum*, 2021. DOI: 10.2514/6.2021-2010.

[8] Stefan Notter, Gregor Müller, and Walter Fichter. Integrated Updraft Localization and Exploitation: End-to-End Type Reinforcement Learning Approach. In *6th CEAS Conference on Guidance, Navigation and Control*, may 2022.

[9] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms, 2017.

[10] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, November 1997. DOI: 10.1162/neco.1997.9.8.1735.

[11] F. A. Gers, J. Schmidhuber, and F. Cummins. Learning to forget: continual prediction with LSTM. In *1999 Ninth International Conference on Artificial Neural Networks ICANN 99*, volume 2, pages 850–855, 1999. DOI: 10.1049/cp:19991218.

[12] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. OpenAI Gym. `https://arxiv.org/pdf/1606.01540.pdf`, 2016.