Feasibility of Performance-Based Training Programs for Combat Aircraft Pilots

Eirik Løhaugen Fjærbu, Guro Kristin Svendsen
Norwegian Defence Research Establishment (FFI)
NORWAY

Eirik-Lohaugen.Fjarbu@ffi.no, Guro.Svendsen@ffi.no

Jelke van der Pal
Royal Netherlands Aerospace Centre (NLR)
THE NETHERLANDS

pal@nlr.nl

ABSTRACT

Performance-based training (PBT) provides a possibility to adapt training to individual training needs. The PBT approach is particularly well suited for simulation-based training, where scenarios and training elements are easier to control than in live training. Training programs tailored to individual needs can increase the training value per resource spent on training. This will, however, place high demands on the flexibility of the training system.

It is challenging to schedule training when strict competency requirements are combined with strong interdependency between the trainees, as for combat aircraft pilots. Individually tailored training programs may further amplify this challenge. We consider continuation training for combat aircraft pilots with an extensive use of simulators, and explore possibilities and challenges of a performance-based training program.

The analysis combines two approaches. NLR use a PBT-approach for defining continuation training programs based on competencies. FFI has a simulation tool (TREFF) to evaluate the feasibility of a currency-based training program with specific resources and constraints, taking stochastic events into account.

As a first step in this joint approach, we consider a test-case with a training program optimised with respect to a set of competency requirements. TREFF simulation experiments were conducted to explore the feasibility of the requirements. Our results give a good indication of the ability of the training system to maintain proficiency among the pilots, and the corresponding challenges. For a full analysis of a real training system, a new competency-oriented scheduler would be required.

The test case may also serve as a baseline to show the benefits of more adaptive PBT systems, or identify aspects that should be incorporated in their design. The methods used in the test case represent a starting point from which to explore scheduling principles and feasibility of PBT.

1.0 BACKGROUND AND METHODS

The current continuation training philosophy for the Royal Netherlands Air Force (RNLAF) and Royal Norwegian Air Force (RNoAF) is that each pilot must have completed a set of training missions within the last 12 months in order to stay combat ready. These training missions are described in the Annual Training Program (ATP), and have been developed over the years to provide the essential skills and competencies for the average pilot. However, the ATP does not take into consideration the fact that pilots are different individuals, and that some pilots need more training in some areas, and less in other areas. If the training program could be tailored to the individual pilot needs, the pilot skill level would increase — and the hours spent performing unnecessary training missions could be reduced — thus increasing the capability of the pilot, and decreasing the resources spent on pilot training.

NLR is working on a competency-oriented, performance-based training (PBT) approach for defining continuation training programs. PBT allows for the training to be adapted to individual needs, but requires a flexible training system and represents a challenge when scheduling pilots to training missions. FFI has a simulation tool (TREFF) for analysis of the feasibility of executing a training program with a given set of resources and constraints, taking stochastic events and scheduling into account. Combining the two approaches would enable NLR and RNLAF to test the feasibility of proposed training programs, and provide a means for FFI and RNoAF to analyse the composition of the continuation training program.

An introduction to the training-analysis approaches of NLR and FFI is given in this section 1. In chapter 2, we define a set of competency requirements and a training system. In chapter 3, we present the results from TREFF simulations of the training system, including the proficiencies of the simulated pilots. We discuss the results and conclude in chapter 4.

1.1 Performance-Based Continuation Training

There may be several perspectives on Performance-Based Training (PBT) for maintaining proficiency of qualified personnel[1]. A common and central feature is the departure from a frequency-based training approach, in which the professional, in this paper the fighter pilot, would require refresher training in a predetermined schedule. This could, for example, be twice a year for a specific training objective, which may relate to a task, a condition, or specific skills or knowledge. While these predetermined intervals are based on experience, and therefore empirically grounded, they are only rough indications and do not reflect individual differences in training needs. There are several possible variants of PBT, which can be grouped into a series of steps; see figure 1. The first step, with predetermined intervals, is referred to as Competency-Based Training. PBT focuses on the desired proficiences, and aims to use actual performances to determine the optimal amount and timing of the training. Therefore, PBT might also stand for *Proficiency*-Based Training.

Another feature of many PBT perspectives is its data-driven base, the notion that proficiency should be measured as objectively as possible. If objective data cannot be obtained, qualitative data (instructor or self ratings) should be sufficiently discriminative (not just pass/fail), and the rating process should be standardised and calibrated. In civil aviation this objectification of instructor ratings is part of what they label as Evidence-Based Training (EBT), and is formalised in ICAO guidelines and EASA regulations. Civil-EBT is rooted in the concepts of resilient operations, and aims to develop pilot proficiency in several core competencies instead of training a narrow, familiar, and repetitive set of tasks. It features the combination of evidence (data) driven training and competency-based training. Civil-EBT generates competency data, and uses this data to adapt the operator's recurrent training to focus on actual competency training needs.

[1] In military aviation, proficiency is often seen as related to (instrument) flight-specific skills, which are subject to regular proficiency tests for regulatory safety reasons. In PBT, all skills, flight as well as tactical, are to be monitored for proficiency, both for training efficiency reasons as well as safety reasons.

To our perspective civil-EBT is a combination of Competency-Based Training and PBT, including evidence from operational performance, where the PBT component strongly depends on instructor ratings (although improved and standardised). As such, this is still a first step towards full-scale, objective PBT. In our framework of training concepts, we distinguish EBT from PBT. Here, EBT relates to the performance data measured in actual operations being used to improve or optimise training. In training practice, however, EBT should be used together with PBT.

With sufficient amount of performance data from training and/or operations, the statistical power (or its equivalent in artificial intelligence techniques) for reliable predictions on (fading) performances may be strong enough, even on a personal level. Such performance/evidence-based personalised training will require highly flexible training schemes.

A final development of PBT may be a real-time adjustment of the level of difficulty, or injections of specific events, during a (simulated) training session based on the performance measured and/or the effort expended (cognitive load) within a training session. Such real-time, automated, personalised adaptation would be the ultimate end state of PBT evolution. The different stages of PBT are illustrated in figure 1.

Figure 1: An evolutionary concept of training theories

While any of the above-mentioned training theories may be applied without a competency-based perspective or using performance data — for example, good instructors may achieve adaptive personalised training on their own merit — we anticipate great benefits from a systematic build-up of these applications. This way, an evolutionary concept of training theories is formed with competency-based training as a start, to which performance-based models are added as well as evidence-based refinements. Only then, a firm basis will be obtained for automated personalised training within a larger training plan, or automated adaptive training within a single training session.

In an ideal world, where personalised training has been optimised, there is little need for adaptive training within a training session. The most appropriate scenarios to restore loss of proficiency should have been selected at the most optimal timing. Some adaptation may still be effective in case of uncontrolled conditions, such as personal fitness or distractions. The system may, for example, detect unforeseen loss of performance and lower the difficulty level. The focus should therefore be given to optimising personalised training first, and only when that is successfully implemented, adaptive training should be added. Performance-based training based on competencies obviously requires an overview of the competencies required for the job-related missions and tasks. A mission or task will require a composition of competencies, i.e. an integration of higher-level and lower-level skills, as well as knowledge and attitudes. A mission therefore relates to a specific integration of competencies, and needs to be trained on that integrated level as well, even if certain component skills are already trained in other missions and a large transfer of training from the same or similar tasks in other missions can be expected.

Equally important is an understanding of the conditions that make task performance more or less difficult, such as weather conditions or the threat level of the opponent. These conditions are complexity factors, and performance needs to be seen in the context of these factors. Training of the tasks, the competencies, and the complexity factors may never be achieved such that all possible combinations are trained in a certain timeframe. This would likely require completion of a 20 by 20 by 20 matrix of training events. A well-defined set of combinations that fits the operational needs is sufficient.

Personalised PBT will only be effective if performance can be predicted over time. This requires extensive collection of data. For continuation training, the primary data to collect relates to the loss of skill that can be expected after a period of non-use, also known as skill decay. In recent years, the tendency to rephrase the notion in a more positive way is increasing; the flipside of the coin is called the retention of skill, the amount of performance retained after a period of non-use. An abundance of empirical data on skill decay is available for elementary skills or knowledge items [1]. It has been found that recently acquired skills and knowledge decay according to a power-law curve. The slope of the curve differs between tasks (and between associated skills). For example, the psychomotor skills for physical tasks are retained much longer than the knowledge components of procedural tasks. The approach to and quality of training both have an effect on retention as well. Distribution of practice over time (spaced practice as opposed to massed practice) and high engagement is very beneficial for the retention of skills. There are personal factors as well. Cognitive ability and motivation will influence the level of retention. Several further influencing factors have been reported. Full overview can be found in the final report of the NATO RTG HFM-292 [2].

However, little is known about interaction effects of these factors, and how retention works for more complex skill sets where several elementary skills have been integrated to perform a complex task such as a basic fighter manoeuvre. Skill decay for more experienced pilots on such complex skills may not follow a power-law curve, but has been found to follow a much slower S-curve decay pattern [3], which is in line with our professional experiences. Also, some pieces of knowledge can be retained forever, and certain skills will remain intact for a very long time as well (e.g. riding a bicycle). The S-curve for such skills may stretch over very long periods of time.

The process of change in the decay curve with respect to expertise building, and how this differs between persons and tasks, can only be understood by collecting more detailed data than we do today. Attempts to model retention computationally are provided by [4][5].

1.2 Simulation-Based Analysis of Training

Discrete-event simulations (DES) [6] is a widely-used method of analysis in the military domain. We have used DES to support decisions regarding design or redesign of training systems. Obtaining sufficient training depends on making the best use of limited resources. Time is often a key aspect. From the perspective of an individual trainee, he or she should ideally receive training according to a certain progression. This will require access to a training medium with instructors. In addition, trainees often need to train together with other trainees, and possibly also with someone acting as the enemy forces. The training system consists of instructors, supporting actors, training media, a training program, and intended users. This system must be designed to give sufficient training for reasonable utilisation of resources. In our analyses, we also include stochastic elements to see how vulnerable the system is for unwanted effects like grounded aircraft or difficult weather.

The TREFF-tool [7] has been used to advise RNoAF on combat aircraft pilot training. Training for combat aircraft pilots is a complex task from a resource-allocation perspective. The training requires a set of training media, for example aircraft or high-fidelity simulators. Further, the pilots train in formations. Optimally, the mission under training should fit the needs of all trainees. In reality, this is often a difficult scheduling problem, especially for squadrons emphasising tactics in larger formations. TREFF was designed and used to evaluate the ability of alternative training systems to provide sufficient training for a group of pilots. A currency-based annual training program represented the requirements for pilot training. The main question addressed by TREFF was whether this training could be completed within one year, and if so, how many days margin was there?

TREFF simulates the training sorties flown by the individual pilots, based on stochastic factors like weather and the availability of aircraft and simulators. Training is scheduled according to a training program consisting of a set of missions with associated yearly currency requirements. In addition, the pilots must contribute as supporting actors by manning red-air formations. The model includes generic daily training schedules providing constraints for which training slots can be combined for a given pilot.

As the main question of TREFF is whether the training system has sufficient capacity to allow the pilot to complete a yearly training program in one year, TREFF has less focus on the details of how this is achieved. The scheduler therefore attempts to complete as much training as possible as fast as possible. In reality, one would rather distribute the training as evenly as possible over the year, and also schedule the different missions in an order supporting the best training progression for the pilot. In addition, TREFF models average pilots. Constraints on which pilots can fill which seats in a formation are therefore not included, and the training requirements are based on an average over requirements for experienced and unexperienced pilots.

In order to answer more detailed questions regarding how to conduct the training, a new simulation tool was developed – TREFF 2 [8]. TREFF 2 includes a more sophisticated scheduler and more details regarding the training, such as qualifications for different roles within formations. As the work described in this report describes a first step towards evaluating the potential of joining the approaches of FFI and NLR, we have opted to use the original TREFF in this work. As it is more lightweight, it provides a more rapid overview of the potential. For real-system analyses, TREFF 2 will likely be a better option.

2.0 TRAINING PROGRAM AND SYSTEM

To demonstrate our approach, we created a test case consisting of a basic set of competency-based training requirements based on the NLR approach, which were evaluated using TREFF simulations. The test case is derived from a generic analysis of fighter-pilot requirements, but we use fictitious data. First, in section 2.1, we describe the competency-based training requirements. Then, in section 2.2, we construct a training program suitable for TREFF simulations, based on the requirements of section 2.1. Finally, in section 2.3, we describe the squadron itself as well as the rest of the training system.

2.1 Competency-Based Training Requirements

The missions, competencies, complexity factors, and retention intervals are the main constituents of our competency-based training requirements. We use acronyms and abbreviations for their names; the full names are listed in appendix A. Two types of missions are included: part-task missions (PT1 to PT7) focused on specific tasks and tactical missions (TM1 to TM6) that more closely resemble full (deployed) missions.

The missions defined in the test case are general classes of missions: there is a wide range of possible variations within each class. In the test case, the missions can be modified with respect to complexity factors, the number of pilots involved, and a few of the competencies. Pilots can take part both as friendly forces (blue side) and as opponents (red side); see figure 2 (a).

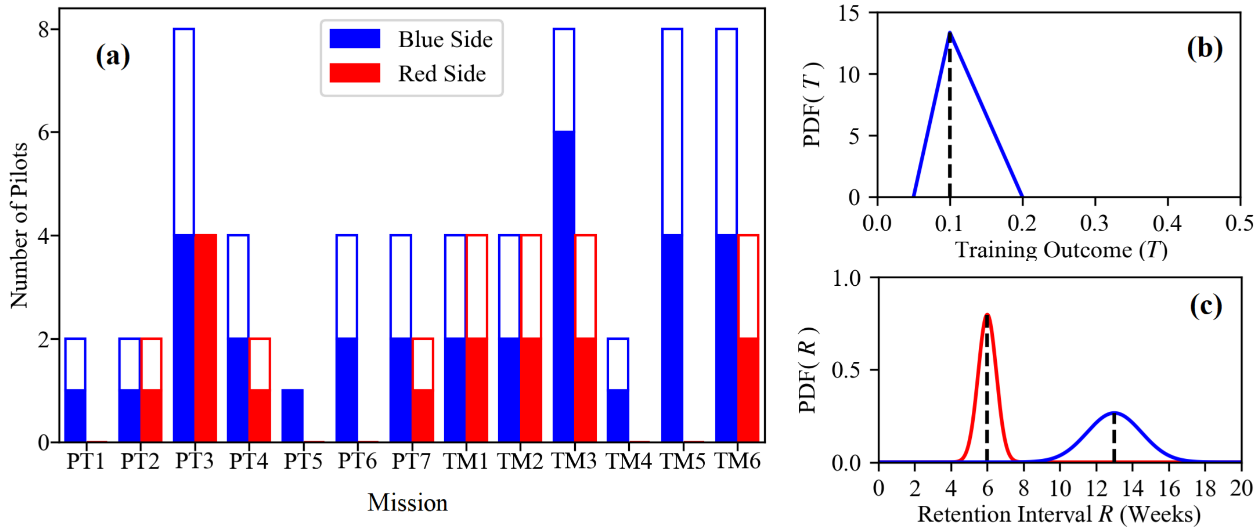

Figure 2: (a) Number of pilots taking part in the missions on the blue and red sides. The solid bars represent the minimum requirements and the outlines indicate the ideal (maximum) numbers. (b) Probability distribution (triangular) for the training outcome *T*. (c) Probability distributions for the retention intervals. The retention intervals have Gaussian distributions where the mean values and standard deviations depend on the competency.

The test case includes 20 competencies, some of which are very comprehensive and complex, and involve a wide range of enabling competencies, whereas others are somewhat more specific. More specific competencies tend to require more frequent training, so they have a significant impact on the total need for training. Most competencies are either inherent to a mission or not included, but a few of the competencies are optional; see figure 3 (a).

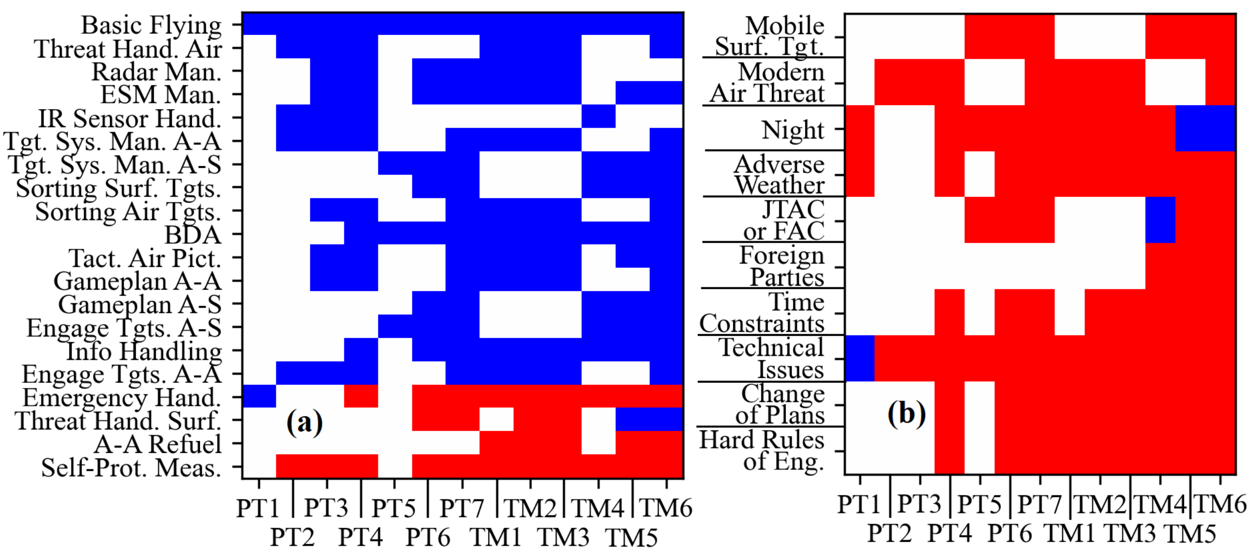

Figure 3: (a) Mapping showing which competencies (vertical axis) the pilots train for during each mission (horizontal axis). (b) Mapping of complexity factors (vertical axis) and missions (horizontal axis). The blue blocks indicate that the competency/complexity factor is inherent to the mission, and red blocks indicate that it is optional.

The complexity factors represent elements that can be modified within each mission, and each complexity factor has two possible states; see appendix A. The complexity factors affect the difficulty of the mission, and they allow for more variation in the training. The mapping showing which complexity factors are relevant to each mission is shown in figure 3 (b).

While skill-decay models are known to be non-linear, we will start testing a competency-based TREFF model by implementing a linear skill-decay model for the competencies. Here, a proficiency of P=100% is the maximum achievable, and the pilots should retain a proficiency higher than 90% to ensure combat readiness. If the pilots do not carry out any training, the proficiencies will decay linearly from 100% to 90% over the retention interval R. We draw retention intervals from Gaussian distributions; see figure 2 (c). We use different mean values and standard deviations for the different competencies; see table 1. We draw independent values for each pilot and each competency, so that each pilot gets a unique competency retention profile.

Each time the pilots carry out a training mission, they gain proficiency for all the competencies relevant to the mission. We refer to the increase in proficiency as the training outcome T. In the test case, only pilots on the blue side get any training outcome. The training outcome is drawn from a triangular distribution shown in figure 2 (b). We draw independent values of the training outcome for each time a mission is completed, but use the same values of T for all the pilots in the same formation and all competencies involved.

Table 1: Expectation values and standard deviations for the retention intervals of the competencies (in weeks).

Competency	Mean Value (w)	Std. Dev. (w)	Competency	Mean Value (w)	Std. Dev. (w)
Basic Flying	6.5	.75	Tact. Air. Pict.	6.5	.75
Threat Hand. Air	4	.5	Gameplan A-A	13	1.5
Radar Man.	3	.25	Gameplan A-S	13	1.5
ESM Man.	3	.25	Engage Tgts. A-S	6.5	.75
IR Sensor Hand.	3	.25	Info Handling	6.5	.75
Tgt. Sys. Man. A-A	3	.25	Engage Tgts. A-A	6.5	.75
Tgt. Sys. Man. A-S	3	.25	Emergency Hand.	4	.5
Sorting Surf. Tgts.	6.5	.75	Threat Hand. Surf.	4	.5
Sorting Air Tgts.	6.5	.75	A-A Refuel	26	3
BDA	6.5	.75	Self-Prot. Meas.	4	.5

We also include retention intervals for each complexity factor and each mission. This ensures that the pilots practise all the missions, and different variants and conditions within each mission. These retention intervals can be interpreted as extensions of the competency retention, where the pilots are required to practise the competencies within the specific context of a mission or with a specific complexity factor. The retention intervals used for the missions and complexity factors are given in appendix B.

2.2 Currency-Based Training Program

The competency-based training requirements of the previous section are defined in terms of (generic) missions, competencies, complexity factors, and retention intervals. By contrast, the currency-based training programs used in TREFF are based on ordered lists of more specific missions. In addition to the type of (generic) mission, the currency-based training programs specify the complexity factors, optional competencies, and training medium for each (specific) mission.

We have developed a tool named COMFORT to create the currency-based training program [9]. This tool takes competency-based training requirements on the form given in section 2.1 as input, and generates a list of specific missions suitable for use with the TREFF simulations. The user is required to specify a cost function and a planning period. COMFORT works by formulating two Constrained Optimisation Problems (COPs), which are solved using OR-Tools [10].

We consider a planning period of half a year, and then divide by the (mean) retention intervals to get the corresponding minimum number of repetitions for each competency, mission, and complexity factor. These requirements are only sufficient if the pilots train every competency at exactly the right time and the training outcome is perfect. Since that is not possible in practice, we scale all the competency requirements by a redundancy factor C. As an example, we use a value of $C = 2.5$. We use the simulations to evaluate whether the redundancy is sufficient; see chapter 3. Additionally, we require that at least half of the exercises are live, and that Air-to-Air Refuelling is only possible in live training.

We then define a cost function and use COMFORT to minimise it; see appendix C. The resulting (optimal) training program contains 43 sorties, and is summarised in figure 4. Some of the missions are repeated more often than others; see figure 4 (b). This is due to our choice of cost function, and highlights some limitations of our basic set of competencies. As expected, there are only a few excess repetitions of the complexity factors, see figure 4 (c), since these are optional to most missions. The JTAC or FAC complexity factor is repeated 7 times because it is inherent to the TM4 mission.

The training program is broken down in figure 4 (a), showing how the different missions add up to fulfil the competency requirements. For most of the inherent competencies, the pilots get significantly more repetitions than the minimum requirements. The exceptions are the competencies with the shortest retention intervals (Radar Man., ESM Man., IR Sensor Hand., and Tgt. Sys. Man. A-A/A-S). There is no excess training of the optional competencies.

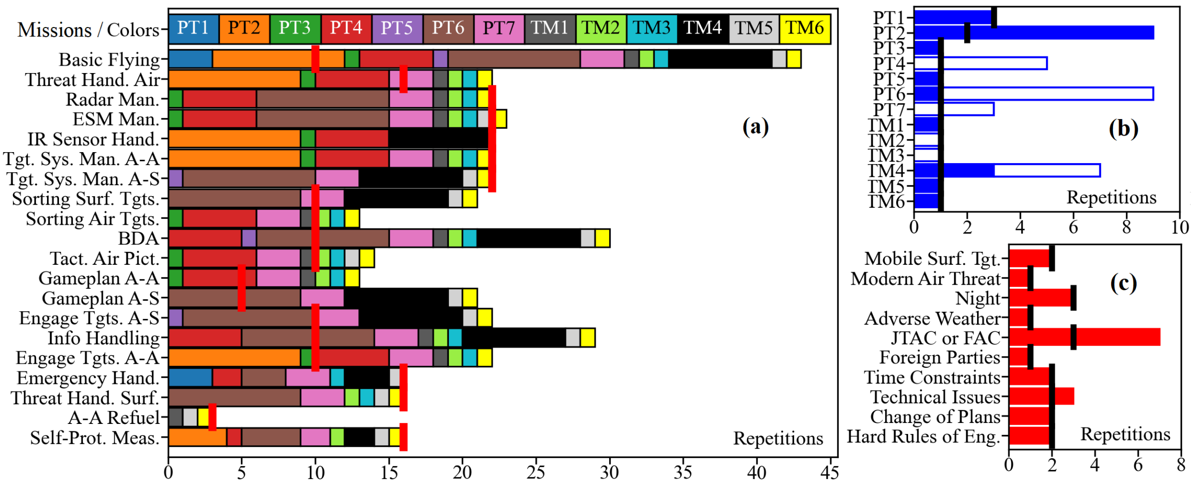

Figure 4: Half-year currency-based training program. (b) Number of repetitions of each mission. The full blue bars represent live training, and the blue outlines represent simulator usage. (c) Number of repetitions of the complexity factors. (a) Breakdown showing which missions contribute to each competency. The colours indicate which mission contributes. The black/red (vertical) lines indicate the minimum requirements.

After we find the optimal set of sorties, COMFORT will shuffle them in order to distribute the training for each competency as evenly as possible. This ensures that the pilots train each competency as close to the end of their retention intervals as possible. The shuffling is implemented as a separate COP where we minimise a total cost summed over the competencies. The cost of a competency is proportional to the maximum number of missions in a row without that competency times the required number of repetitions. To reduce the complexity of the problem, we exclude the mission and complexity-factor requirements from the shuffling. The solution to the shuffling is not guaranteed to be optimal, but COMFORT finds a relatively even distribution.

2.3 Training System Specification

In addition to the training program, TREFF allows us to specify a number of properties of the squadron, their training resources, their available time, as well as their working conditions. These are specified using a spreadsheet.

Among the most central parameters are the number of pilots, aircraft, and simulators available. In the test case, we consider one squadron of 18 pilots with one available aircraft time slot per day. The number of aircraft available is given by a truncated Gaussian distribution with 8 aircraft available on average and a maximum of 10 aircraft available. There are two training slots in the simulator, 6 simulators available, and each simulator has a 98% chance of being operational during any given training slot. Note that for live training, some pilots have to participate as opposing (red) forces, whereas in the simulator the opponents are computer generated and/or controlled by instructors.

We also specify how many days the pilots have available for training. By default, the pilots work 5-day weeks, but TREFF can be configured to reserve time for other activities. We use this option to reserve one day a week for common briefings, administrative issues, and physical training. We include three different types of absence: four-week holidays, single-week leaves, and one-day absences. They represent holidays, sick days, non-flying training courses, and more.

The absence is modelled using stochastic variables. Each pilot has one four-week holiday each year within the period from May to August. At most two are on holiday each week. At the start of each week, every pilot has a 15% chance of going on a one-week leave. Each pilot can have up to 4 such leaves each year. Finally, at any given weekday, the pilots have a 5% chance of not being able to perform flying-related training, for example due to an illness.

3.0 SIMULATION RESULTS

We configured TREFF according to section 2.3, and added additional output channels allowing us to track and analyse data related to CBT. We made a full-year training program consisting of two repetitions of the half-year program from section 2.2. The pilots complete the full-year training program 100 times consecutively without resetting the proficiencies in between. Both at the start of a year and after the first half, the pilots have to wait for everyone to complete the (half-year) program before they can start the next repetition. This remedy is needed because the test-case scheduler does not distribute the training sufficiently well across the year otherwise. In section 3.1, we break down how much time and how many sorties the pilots need to complete the (full-year) training program. We emphasize again that the focus of this work is on the method used, and that the specific test case is not intended to represent a realistic training system.

One of the main purposes of the simulation is to assess whether it is feasible to maintain sufficient proficiency across the squadron over time with the given training resources. Requiring that no proficiencies ever drop below 90% at all is likely too strict, and would require a disproportionate amount of training resources to maintain. For squadrons converting to performance-based training, it will be important to consider requirements in such a way that a workable and flexible training regime is achieved. In section 3.2, we show how many days per year and per pilot that the proficiencies fall below the requirement of 90% in the simulations.

3.1 Simulation Run Length and Working Days

The simulation results showed that the average time used to complete the yearly training program was 355 days; see figure 5 (a). The training program therefore seems to be rather well matched with the available resources. Note that as we reserve one day a week for common briefings, administrative issues, and physical training, there is some margin that can be exploited. However, the amount of training required will depend strongly on the strictness of the (squadron-wide) proficiency requirements. Changes to these requirements can be taken into account by adjusting the redundancy factor *C*.

Figure 5 (b) shows how many times the pilots repeat each mission (on the blue side). The scheduler in TREFF prioritises simulator training and live training differently, and there are two timeslots each day for the simulators and only one for the aircraft. The pilots often get ideal formation sizes in the simulators, but as a result some of them get more repetitions than needed. The pilots spend more time in the air than in the simulator, see figure 5 (c), but they get more training output from the simulator, since red-air sorties do not give any training output in the test case.

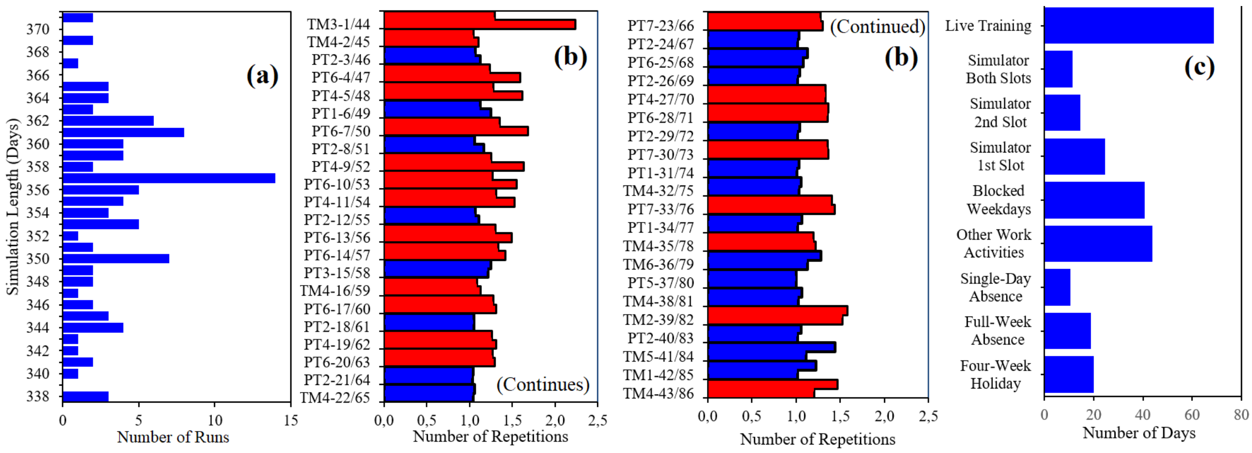

Figure 5: Simulation results: (a) Time needed for all pilots to complete the full-year training program. (b) Average repetition numbers for each mission, split in two halves. The red and blue bars indicate simulator use and live training, respectively. The top half of each outline represents the first repetition of the training program, and the bottom half the second. (c) Average number of days by activity. Red-air sorties are included in (c), but not in (b).

3.2 Proficiencies

TREFF generates plots of the proficiencies during the simulation, giving immediate feedback about the training status of the pilots. To highlight pilots that get insufficient training, we calculate the lowest proficiency among all the pilots for each competency. Figure 6 shows the behaviour of these minimum proficiencies over a one-year period.

The pilots start new repetitions of the half-year training program around January and July in figure 6. Around these dates, several of the proficiencies fall significantly below the 90% requirement. This is because some pilots have to wait for others to complete the training program. All the pilots have to complete the entire training program before anyone can start the next repetition. During summer, some of the pilots are also on 4-week holidays. The proficiencies are not updated when a pilot is on leave, but they are corrected for the entire period when the pilot resumes training. This leads to some flat sections and large jumps in the minimum proficiencies.

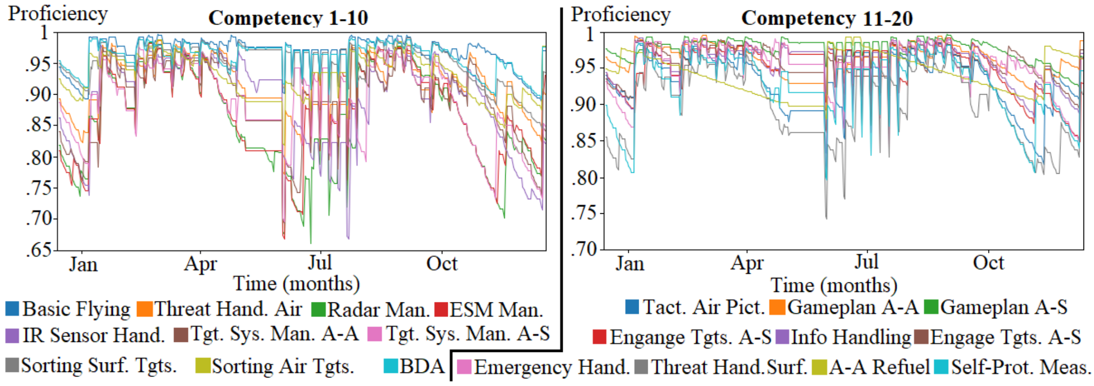

Figure 6: Minimum proficiencies among all the pilots as a function of simulation time, for an arbitrarily chosen year. The pilots start back at the beginning of the training program around January and July. The 4-week holidays occur between May and August. The plots were directly generated in TREFF, which is implemented in Anylogic [11]. Please note truncated vertical axis.

To indicate whether the pilots get enough training to remain proficient, we count the number of days that their proficiencies fall below the limit of 90%; see figure 7. As we saw in figure 6, some of the pilots fall well below the 90% limit, but figure 7 shows that most of the pilots are proficient most of the time. If the deficiencies shown in figure 7 were distributed evenly across the pilots and throughout the year, none of the pilots would ever be sufficiently proficient in all competencies. However, as we saw in figure 6, it is more likely that the pilots lose proficiency in many competencies at the same time during periods of low activity.

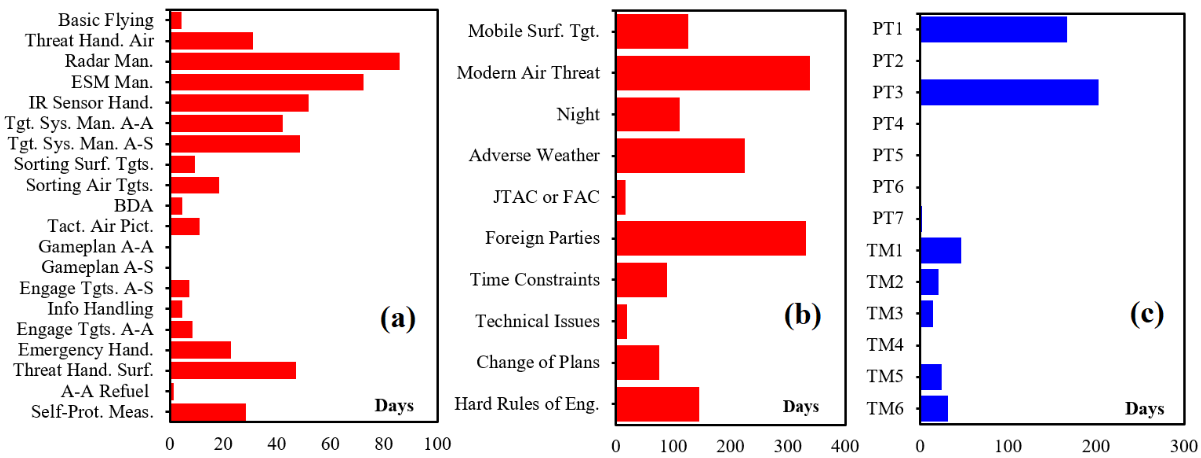

Figure 7: Average number of days per year where each pilot has proficiency lower than 90% for each competency (a), complexity factor (b), and mission (c).

As one would expect, the competencies where there is the least excess training are also the ones that fall below the proficiency requirements most often; compare figure 4 and 7. The proficiencies associated with the missions and complexity factors are not given the same focus as the competencies. These requirements can be seen as a means to improve the training of the competencies rather than being necessary by themselves. For two of the missions, as well as three of the complexity factors, the proficiencies fall below the requirements more than half of the time. No redundancy was added for these requirements, and they were not taken into account in the shuffling, so it is not surprising to see that the pilots get too little training.

4.0 DISCUSSION AND CONCLUSIONS

Performance-based training offers an alternative to the currency-based training approach currently used in the RNoAF and RNLAF. PBT allows increased flexibility, individual tailoring, and continuous adjustments based on objective and subjective performance data. However, this flexibility leads to a large state space of possible scenarios to consider when we want to evaluate the training system. Discrete-event simulations may be used to explore this space of possible scenarios.

We created a test case based on the NLR approach to performance-based training and the TREFF simulations developed by FFI, to demonstrate the potential of a joint approach. The test case consists of an example set of competency-based training requirements for continuation training, and a TREFF simulation to evaluate whether the requirements are feasible.

The scheduling problem in such a competency-based pilot training system is challenging, not only because of the great flexibility, but also due to limited training resources and high co-dependency among the pilots. For the test case, we constructed a static currency-based training program and used the existing scheduler from TREFF simulations. The COMFORT tool ensures that the training program gives sufficient training to maintain proficiency, including a redundancy factor.

In the NLR approach to competency-based training, high-level competencies are broken down into a large number of enabling competencies. In the test case, we include mostly high-level competencies. By including more enabling competencies, the competency-based training requirements can be made more complete and realistic. In the test case, only a few missions are needed to cover all the competencies. A more complete set of competencies would likely reveal a need to practise a greater number of different scenarios, although in general it may be assumed that more specific competencies could also be practised on PC-based training software with simplified part tasks, requiring less training time than regular training sorties. More specific competencies, which are likely more prone to decay as well, will pose challenges for the scheduling. Further analysis, discussions with operational pilots, and actual experiments are crucial in order to determine the most important training requirements.

For a full analysis of a real training system, the scheduler should use the competency requirements directly rather than through a currency-based training program (as in the test case). However, the test-case results should still give a good indication of whether our requirements are feasible, and may help us identify some of the aspects that should be incorporated in such a scheduler. The test case may also serve as a baseline to show the benefit of more adaptive performance-based training systems.

Our objective to maintain every pilot combat ready throughout the year by minimising the time in which the performance levels of competencies are less than 90% may not reflect the actual needs and the policy of an air force. For instance, the requirements may be specified in terms of how many pilots should be combat ready for a specific mission, or how many can be trained to combat-ready status within a given time frame. Such requirements will be central to a future competency-based scheduler. The redundancy factor shows that the pilots have to train significantly more than if they could train what they need at exactly at the right time. Therefore, one cannot determine whether a set of training requirements can be maintained by looking at the requirements alone. Each pilot only completes around two sorties per week on average, and because some of the retention intervals are only three weeks, relatively short delays can cause significant drops in proficiency.

The TREFF simulation provides a natural way to illustrate the workings of the test-case training system, and enable us to explore a large variety of alternative training systems. For the basic test case, we used a linear skill decay model. More sophisticated models should be used for an evaluation of a real system, but a simple model is useful to understand the potential of our approach. Even with improved retention models and more realistic input data, one should be careful not to pay too much attention to the exact values of the proficiencies. A full, accurate model would involve a lot of details and parameters, and thorough verification through testing would be necessary.

The decay functions for the selected competencies in our simulation are examples used to demonstrate our methods; more data and more specific data is needed to define validated decay functions. True performance-based training is only possible if one is able to actually measure the proficiencies in a detailed and consistent way, which goes beyond the capability and available time of instructor pilots or supervisors. Including such measurements in the simulations is an interesting possibility for future work.

While our simulation does not directly prove that an actual training schedule based on PBT is feasible, it does explore the potential of PBT. PBT may provide insight and better control of the mechanisms that generate or hamper combat readiness, be it for an individual pilot, a given mission, a squadron, or the complete air force. In this exploratory work, we developed a method that represents a starting point for further refinements of the scheduling principles, models, and simulations for performance-based training. This may be necessary to make PBT a success in the daily training practice of squadrons, and to relieve the challenging job of the scheduling officer.

REFERENCES

[1] W. Arthur, W. Bennett, P. L. Stanush, and T. L. McNelly, "Factors that influence skill decay and retention: A quantitative review and analysis". *Human Performance, 11*(1)*,* 57-101, 1998.

[2] NATO Research Task Group HFM-292, "Understanding and Reducing Skill Decay", (in prep)

[3] J. I. Vlasblom, H. J. Pennings, J. van der Pal, and E. A. Oprins, "Competence retention in safety-critical professions: A systematic literature review". *Educational Research Review*, 100330, 2020.

[4] T. S. Jastrzembski, K. A. Gluck, and G. Gunzelmann, "Mathematical models of spacing effects in memory: Applications in education and training". In Proceedings of the Society for Mathematical Psychology Annual Meeting. Vancouver, Canada, 2006

[5] F. Sense, T. Jastrzembski, M. C. Mozer, M. Krusmark, and H. van Rijn, "Perspectives on computational models of learning and forgetting". In Proceedings of the ICCM 2019: 17th International Conference on Cognitive Modeling. T.C. Steward, (ed.) Applied Cognitive Science Lab, p.216-221

[6] A. M. Law, "Simulation Modeling and Analysis 4th edition", McGraw-Hill, New York, 2007

[7] G. Skogsrud and O. M. Mevassvik, "Simulation of a combat training program for Norwegian pilots", in RTO meeting proceedings of the NATO Modelling and Simulation Group, Bern, Switzerland, 2011.

[8] G. K. Svendsen, I. U. Haugstuen, O. M. Mevassvik, and H. Rorvik, "Optimized Pilot Training for Combat Aircraft", In Proceedings of Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2017, No 17199

[9] E. L. Fjærbu, "Optimization of training programs for competency-based training", Norwegian Defence Research Establishment (FFI), FFI-report 20/00002, 2020.

[10] L. Perron and V. Furnon, OR-Tools, https://developers.google.com/optimization/ (Retrieved August 27 2020)

[11] The Anylogic Company, Anylogic, https://www.anylogic.com (Retrieved September 8 2020)

APPENDIX

A Acronyms and Abbreviations

Acronym	Explanation
BDA	Battle Damage Assessment
CBT	Competency-Based Training
COP	Constrained Optimisation Problem
COMFORT	COmpetency-Mission Frequency Optimizer for Readiness Training
EASA	European union Aviation Safety Agency
EBT	Evidence-Based Training
FAC	Forward Air Controller
ICAO	International Civil Aviation Organization
JTAC	Joint Terminal Attack Controller
PBT	Performance-Based Training
PT	Part-Task mission
RNoAF	Royal Norwegian Air Force
RNLAF	Royal Netherlands Air Force
TM	Tactical Mission
TREFF	TREning av Flygere for F-35 (Pilot Training for F-35)

Abbreviation	Competency Name	Abbreviation	Competency Name
Basic Flying	Basic Flying Skills	BDA	Battle Damage Assessment
Threat Hand. Air	Threat Handling Air	Tact. Air Pict.	Building Tactical Air Picture
Radar Man.	Managing Radar	Gameplan A-A	Gameplan Execution Air-to-Air
ESM Man.	Managing Electronic Support Measures	Gameplan A-S	Gameplan Execution Air-to-Surface
IR Sensor Hand.	Handling Infrared Sensor System	Engage Tgts. A-S	Engage Targets Air-to-Surface
Tgt. Sys. Man. A-A	Managing Targeting System Air-to-Air	Info Handling	Information Handling
		Engage Tgts. A-A	Engage Targets Air-to-Air
Tgt. Sys. Man. A-S	Managing Targeting System Air-to-Surface	Emergency Hand.	Emergency Handling
		Threat Hand. Surf.	Threat Handling Surface
Sorting Surf. Tgts.	Sorting Surface Targets	A-A Refuel	Air-to-Air Refueling
Sorting Air Tgts.	Sorting Airborne Targets	Self-Prot. Meas.	Use of Self-Protection Measures

Complexity Factor Abbreviation	Complexity Factor On-State/Off-State
Mobile Surf. Tgt.	Mobile/Fixed Surface Target
Modern Air Threat	Modern/Old Air Threat
Adverse Weather	Adverse/Normal Weather
JTAC or FAC	With/Without JTAC or FAC
Foreign Parties	With/Without Foreign Parties
Time Constraints	With/Without Time Constraints
Technical Issues	With/Without Technical Issues
Change of Plans	With/Without Change of Plans
Hard Rules of Eng.	Hard/Easy Rules of Engagement

B Retention Intervals

The mission retention intervals can be interpreted as extensions of the competency retention: requiring that the pilots practise each competency within the specific context of a mission. This could be achieved by defining retention intervals for each combination of an inherent competency and a mission. Since most of the competencies are inherent, the set of competencies is mostly the same for all the variants of a mission in the training program. Therefore, it is sufficient to include only the shortest combination-of-competency-and-mission retention interval for each mission. The retention intervals of the missions as well as the complexity factors are listed in table 2.

Table 2: Retention intervals of the missions and complexity factors (in weeks).

Part-Task Mission	Interval (w)	Tactical Mission	Interval (w)	Complexity Factor	Interval (w)
PT1	8	TM1	26	Mobile Surf. Tgt.	13
PT2	16	TM2	26	Modern Air Threat	20
PT3	20	TM3	26	Night	8
PT4	26	TM4	26	Adverse Weather	20
PT5	52	TM5	26	JTAC or FAC	8
PT6	26	TM6	26	Foreign Parties	20
PT7	26			Time Constraints	13
				Technical Issues	13
				Change of Plans	13
				Hard Rules of Eng.	13

C Cost function

We define a cost for each mission, complexity factor, and optional competency; see figure 8. The cost of a training program is given by the sum of the costs of all the missions, complexity factors, and optional competencies. We also add an additional cost (2 units) for live exercises relative to simulated ones.

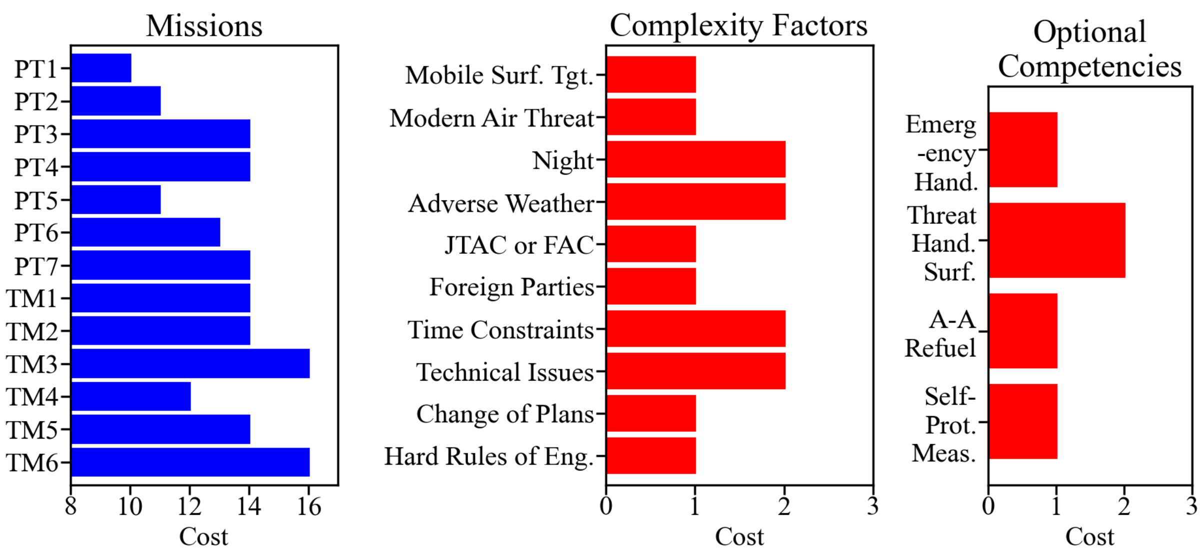

Figure 8: Costs for missions, complexity factors and optional competencies optimised by COMFORT. The cost of a sortie is given by the sum of the inherent mission costs, complexity factor costs, and the costs of the optional competencies.