The Best Rest, Revisited

A comparison of differing regulatory efforts to control pilot fatigue.

By David Hellerström, Hans Eriksson, Emma Romig and Tomas Klemets

In the June 2010 issue of AeroSafety World *(p. 40), the authors analyzed several flight and duty time limit regulation systems in use around the world for controlling fatigue. Subsequently, the U.S. Federal Aviation Administration published its Notice of Proposed Rulemaking (NPRM) for "Flight Crew Member Duty and Rest Requirements,"[1] which has spurred significant discussion in the United States as pilots unions, airlines and others attempt to understand the implications of the proposed rules. In this article, the authors extend their previous analysis to consider the NPRM, as well as U.K. rules, outlined in Civil Air Publication 371, "The Avoidance of Fatigue in Aircrews."[2] We hope that this analysis, though limited, will inform the discussion.*

In commercial aviation, crew schedules are regulated by duty time limits, flight time limits, minimum rest rules and other constraints. These rules and limits, collectively referred to as flight time limitations (FTLs), were originally conceived to be a simple scheme for limiting fatigue among flight crewmembers.

Over time, FTLs have evolved, driven by industrial pressures or new scientific data, or to cope with changing aircraft capabilities. Today, there are major differences among FTL schemes in different parts of the world affecting crew productivity, crew alertness — and airline competitiveness.

With the results of new research on sleep and work-related fatigue in hand, it becomes useful to compare existing regulations with the new findings.

FTLs are relatively straightforward, and, combined with labor agreements and other safeguards, they do a reasonable job of protecting alertness under most circumstances. Unfortunately, FTLs tend to be extremely rigid and limit operational flexibility and efficiency. But by far the most troublesome aspect of FTLs is the illusion of safety that they create — suggesting that to fly within the limits is inherently safe, while flying outside the limits is inherently unsafe.

In recent years, considerable effort has been directed toward increasing scientific knowledge of fatigue and alertness. By combining new knowledge of fatigue with safety and risk management processes, the concept of the fatigue risk management system (FRMS) was created. In previous work, we have demonstrated that a properly implemented and managed FRMS can be vastly superior to FTLs in managing alertness while maintaining or improving productivity.[3] Whereas FTLs are not feedback-driven and often lack a scientific basis, an FRMS is by definition intended to be a closed-loop, data-driven process. In addition to the stronger scientific basis of an FRMS, an added benefit is increased operational flexibility.

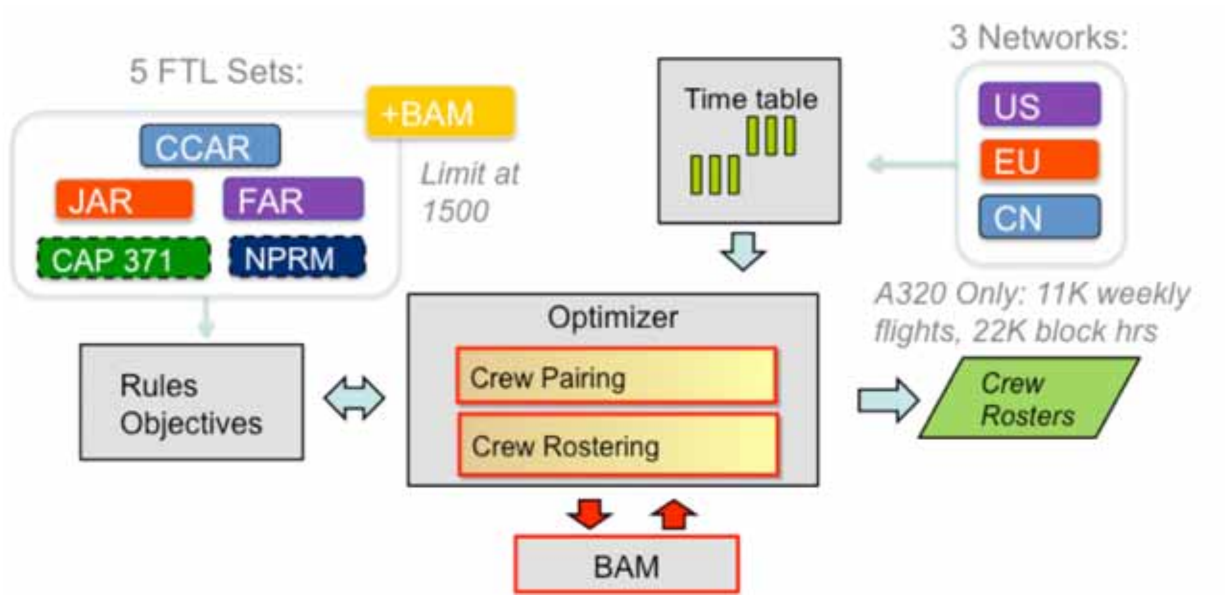
FRMSs have been used, to one degree or another, by a number of airlines around the world — with positive outcomes — for several years. They are built around predictive tools including, but not necessarily limited to, mathematical models of fatigue and alertness. Models predict crew alertness from planned and actual schedules and inferred sleep and wake history. Models also consider known physiological phenomena, such as circadian rhythms and sleep propensity, and make predictions based on these considerations. Unfortunately, while models have been developed and validated in a laboratory environment, more work is required to validate the models in a commercial aviation environment.

Thus, we are faced with a dilemma. FTLs are imperfect but well understood and easy to apply. An FRMS is better for managing fatigue-related risk but must be developed and validated to be trusted. Until FRMSs are widely proved and implemented, the goal must be to refine FTLs to be as close as possible to an FRMS-based approach. A refined FTL should strive to guarantee an equivalent or better level of flight safety while allowing airlines to efficiently and flexibly operate their businesses.

For this updated article, we analyzed five different sets of FTLs for productivity and alertness. We compared these regulatory formulations to a model-based FRMS. The analysis used a fatigue model within crew scheduling optimization software on the timetables of three short-haul airline fleets. Finally, we demonstrated our suggested alternative for improving FTLs.

Analytical Methods

To build the schedules for comparing FTLs, we used the system illustrated in Figure 1, p 3. Our system centers on an "optimizer," which considers an airline's timetable and a set of rules and objectives to build crew schedules. In each of our FTL comparisons, we created a schedule using one airline's timetable and one of the FTL sets as a constraint. To simulate an FRMS, we created schedules without the constraint of an FTL set, instead using the predictions of our alertness model.



BAM = Boeing Alertness Model
CCAR = China Civil Aviation Regulations
JAR = Joint Aviation Requirements (refers to EU-Ops with sub-part Q)
FAR = Federal Aviation Regulations
CAP 371 = Civil Aviation Policy 371 (UK Flight & Duty Time Limits)
NPRM = Notice of Proposed Rulemaking (refers to proposed changes to the FARs)

Figure 1

In the original analysis, the FTL sets used were the EU-OPS with Subpart Q, abbreviated as Joint Aviation Requirements (JARs); the U.S. Federal Aviation Regulations (FARs) Part 121; and China Civil Aviation Regulations (CCARs) 121 Rev 3. In this second round of analysis, we included the FAA NPRM and U.K. Civil Aviation Publication (CAP) 371, which includes flight and duty time limitations.

In addition to the five FTL sets, we created an "FRMS" rule set based on a model's predicted alertness. The rule set was created using the Boeing Alertness Model (BAM), a bio-mathematical model of alertness, and the results are identified with the BAM notation.[4,5] In this rule set, there were no rules on flight time, duty time or rest time; instead, a lowest-allowable limit was established for predicted alertness, as measured on the Common Alertness Scale — defined by the Common Alertness Prediction Interface, which specifies the interaction between the alertness model and crew scheduling software — on a scale from zero (least alert) to 10,000 (most alert).

For this analysis, we set the parameters for briefing time to 45 minutes before active duty and 30 minutes before passive duty.[6] Debriefing time was set to 15 minutes. CCARs define "rest at rest location" as being rest at a hotel, rather than at an airport; therefore, 20 minutes at each end of the rest interval were used for local transport and not regarded as valid rest.

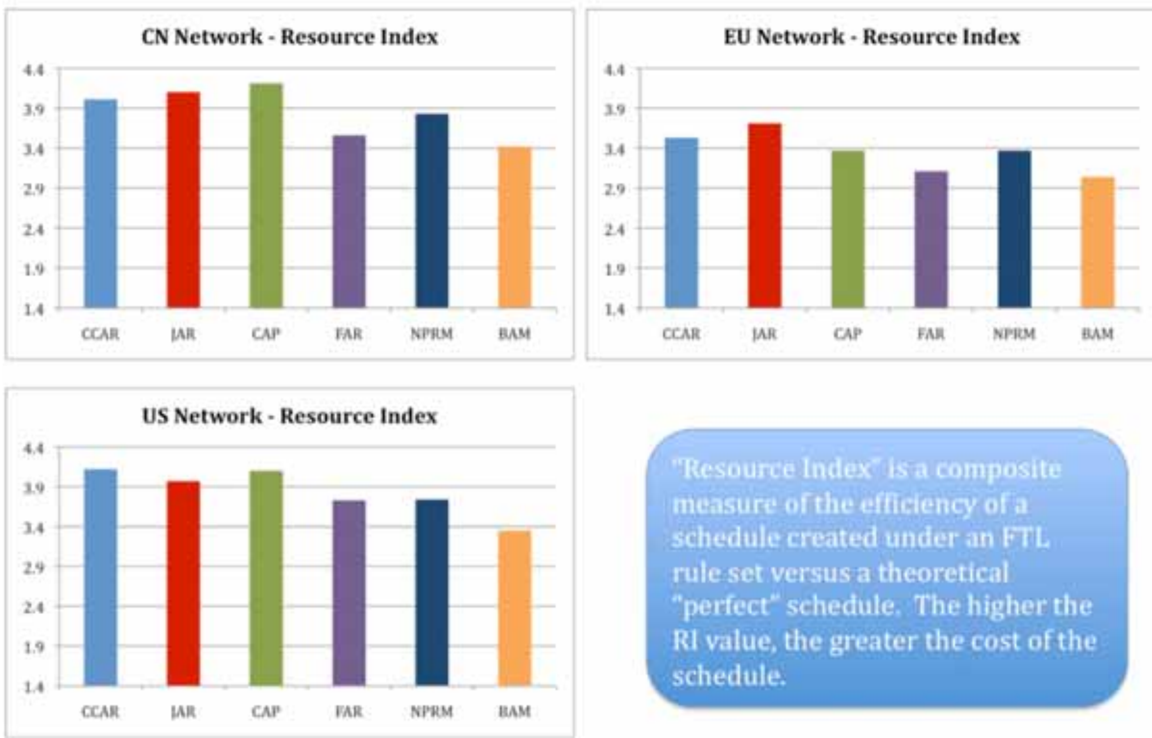
Data Sets

Three large data sets — derived from publicly available flight timetables of China Southern Airlines, Lufthansa and Northwest Airlines — were used to compare the properties of all of the FTLs with respect to productivity and alertness. All flights were two-pilot operations in Airbus A320 aircraft. Average flights in the Europe and China data sets were less than two hours, while flights in the U.S. data set averaged 2.5 hours.

To compare the solutions, we relied on metrics representing the resources needed to implement a solution for each flight and the predicted alertness level of flight crewmembers. A low level of predicted alertness on a flight is associated with higher risk. The alertness properties in the solutions were hard to map to a single descriptive value or statistical measure; therefore, we chose to report and compare the lowest level of predicted alertness, as well as the average alertness experienced on the worst 1 percent, 5 percent and 10 percent of flights.

To quantify the relative productivity of the solutions, we created a composite measure of productivity called the Resource Index (RI). The RI values (Figure 2, p. 5) are a measure of how much less efficient a solution is than a theoretically "perfect" solution. Under all three networks — China, European Union (EU) and U.S. — we observed the same trend in our RI:7 The FARs were the most flexible and most efficient of the FTL schemes, followed by the CCARs and finally the JARs. From the perspective of the resource index, the CAP 371 rules were the least efficient under two of the three network structures. Under the NPRM and the U.S. network conditions, the resource index is slightly lower; however, under the other network conditions, RI increases. The flexibility of the FARs comes primarily from the lack of duty-time limits and the possibility of a rest period as short as eight hours. Notably, the BAM solution outperformed all three FTL sets in terms of the resource index.

Figure 2 Resource Index

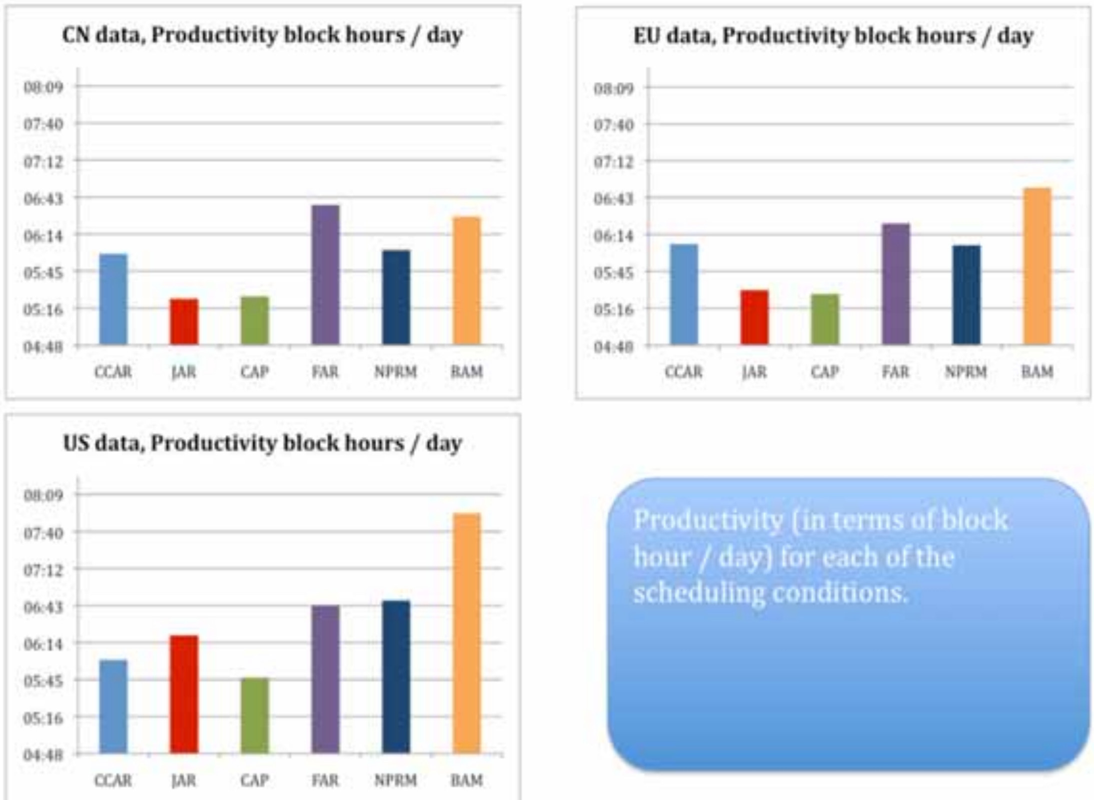


When we considered average block time per duty day — another measure of productivity — we saw similar performance on predicted alertness from BAM and the three original FTLs. Only when applied to the Chinese data set, the FARs generated a solution more efficient than that created by BAM.

Figure 3, p. 6, offers a second measure of the productivity achieved with the five FTL schemes and BAM — indicated by the average number of block hours per day.

Under the U.S. airline operating conditions, with relatively fewer legs and legs of longer duration, the JARs outperformed the CCARs in terms of crew productivity per day; in all other cases, the JARs were the least efficient of the FTL schemes. The performance shortfall on the other FTL sets probably stemmed from the reduction in duty-time limits for many sectors under the JARs.

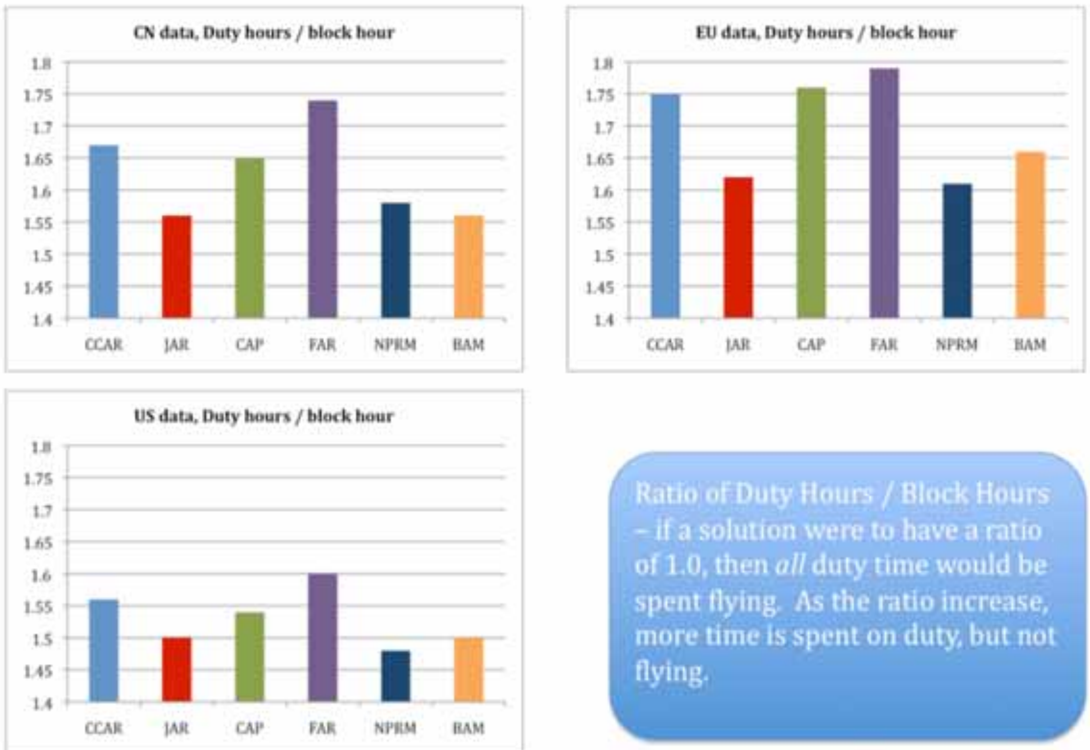
Figure 3 Productivity in Block Hours / day



We also noted that the FARs — without any real duty-time limit — consumed much more duty time than the other FTL schemes to achieve higher levels of productivity. The NPRM, where duty time is a much more significant focus, results in a duty-time/block-time ratio similar to the other flight duty period–focused (FDP–focused) FTLs. In the United States, where pilot pay is mainly based on flight time, this ratio is particularly significant, and this ratio can be thought of as one measure of the "quality of life" associated with the schedules created under each rule set. This ratio is illustrated in Figure 4, p. 7.

Figure 5 plots the level of alertness predicted under "normal" scheduling conditions, with typical airline objectives. It shows that the level of predicted alertness is highly dependent on the airlines' timetable because the legs scheduled very early or very late always caused low alertness. As shown in the figure, the FARs provided the least protection against fatigue; the CCARs and JARs were comparable to each other, but the JARs provided somewhat better protection. Under the U.S. conditions, CAP 371 functions similarly to the CCARs. The NPRM performs at a similar level, bringing — for this timetable and set of objectives — the FARs up to

Figure 4 Ratio of Duty Time to Block Time



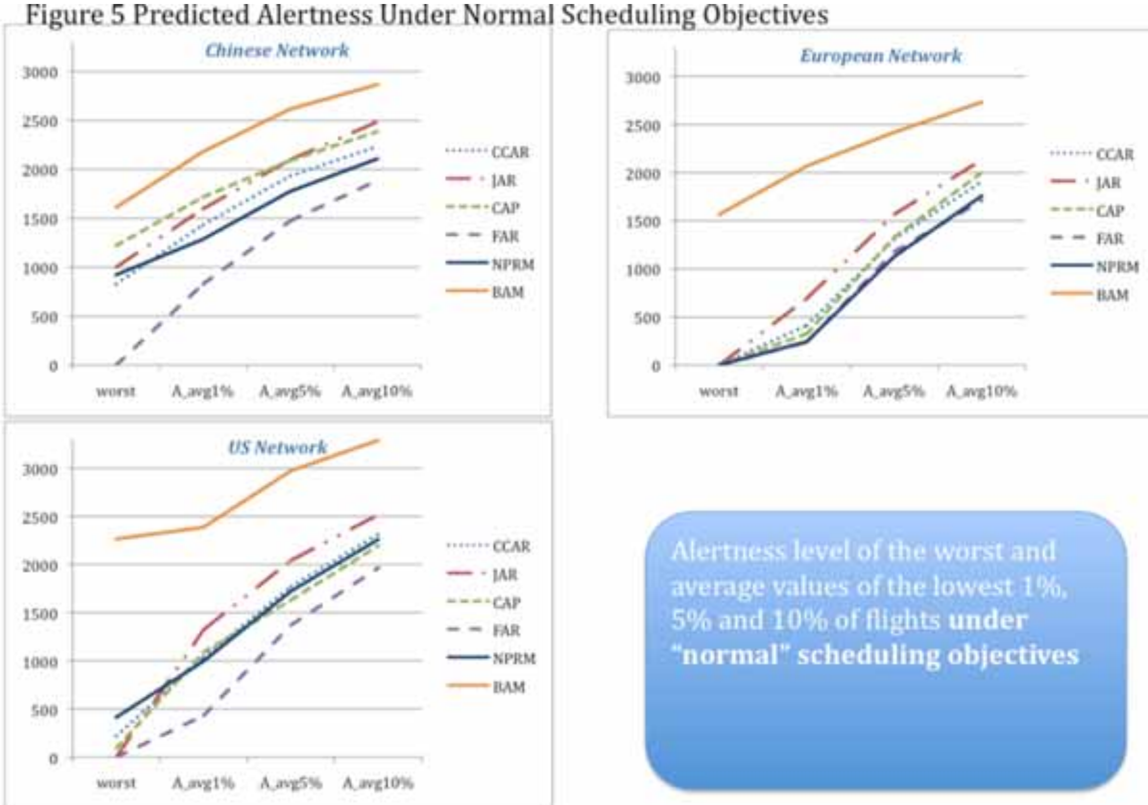
a level similar to the other prescriptive rules. The application of the NPRM rules to the EU network is concerning because under these conditions the schedules, and therefore the predicted alertness, created under the NPRM are virtually identical to those created under the FARs conditions. Under all network conditions, the solutions produced by BAM were better at protecting against fatigue — not surprising, because with BAM, the predicted alertness is a primary objective when constructing the schedules. The BAM solutions were interesting because they showed that it is possible to build solutions that protect against fatigue without sacrificing productivity.

Worth noting is that many of the FTL-permissible flights associated with low alertness would not be allowed under BAM-based rules.

Improving the Rules

The tools used for this productivity and alertness comparison can be extended into a framework to improve prescriptive rules, such as an FTL scheme, to help the FTL scheme provide better protection against low alertness while also maintaining or improving productivity. In this application, the optimizer can be used to analyze the properties, including productivity and

Figure 5 Predicted Alertness Under Normal Scheduling Objectives



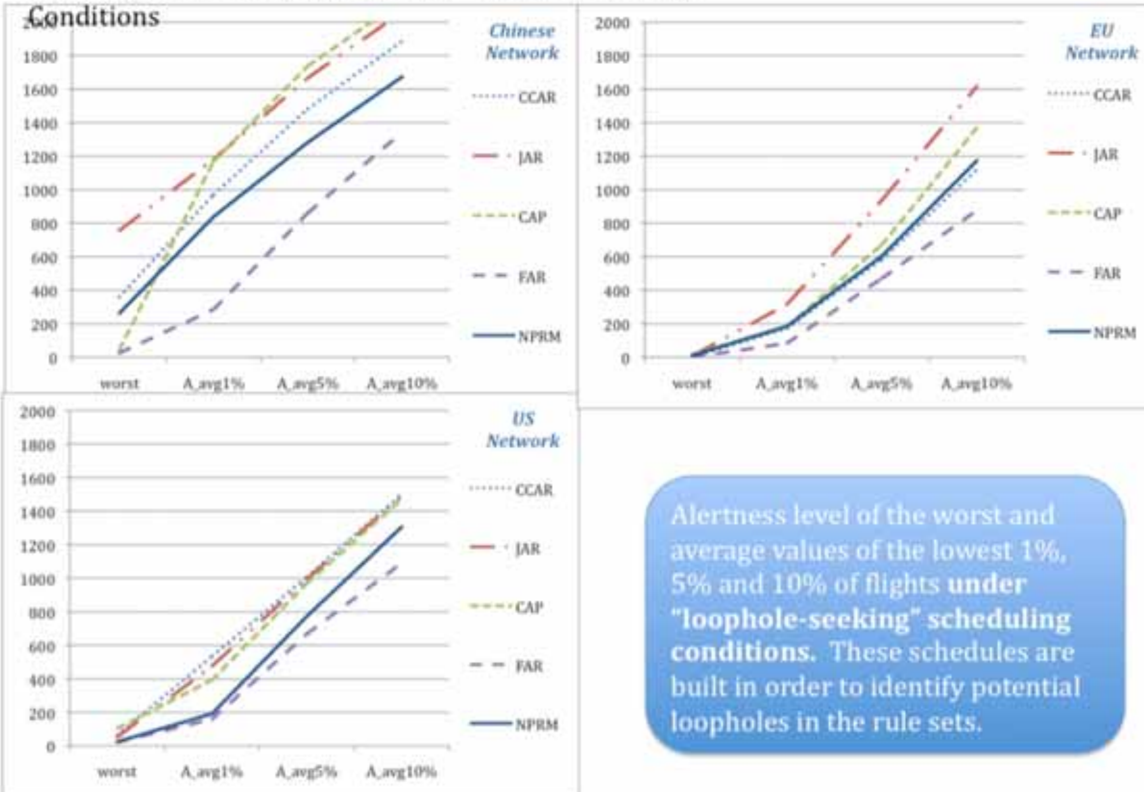
alertness, of an evolving rule set. The method identifies overly restrictive rules and loopholes in the existing rule set.

The improvement begins with the creation of three reference solutions. One solution is based solely on the alertness model with no other limiting rules. The second solution is based on the limits in the prescriptive rules. The third solution is a stress test solution, also based on the limits in the prescriptive rules. In the stress test, the researcher activates an incentive so that the optimizer will produce the most tiring solutions allowed under an FTL. As shown in Figure 6 (p. 9), all of the prescriptive formulations allow flights with very low alertness to be created.

From the first two solutions, we can identify the productivity and protected level of alertness of our original rule set, and the maximum productivity and protected level of alertness. In the third solution, bad patterns of productivity and alertness are easy to identify.

With each iteration, researchers must decide if they want to tighten the rules to improve on alertness, or relax an overly restrictive rule to increase productivity. When the increased productivity option is selected, the revised rule set also changes the alertness outcome —

Figure 6 Predicted Alertness Under "Loophole-seeking" Conditions



probably for the worse. Likewise, when alertness is improved, the rule set usually causes loss of productivity. Changes that improve productivity or alertness — without one affecting the other — are ideal.

Improving Alertness

In our effort to improve the protected level of alertness of an FTL set, we compared the crew schedules produced by the optimizer with the best version of the prescriptive rules. Crew schedules were sorted according to the worst alertness, and flights with low crew alertness were highlighted. In the crew schedules leading up to the flights with low alertness, we identified a combination of duty and sleep opportunities that created a fatiguing pattern.

By looking at these fatiguing patterns, we proposed a few rules designed to prevent these patterns from occurring. The proposed rules were implemented and their impact was estimated by analyzing the number of rules violations they created in the solution. Ideally, the rule should trigger at the fatiguing patterns and nowhere else. The impact of the newly proposed rules was also evaluated on the reference solution created based on the alertness model. Because this solution is considered good from an alertness perspective, the new rules should not give rise to many rule violations in this solution. The proposed rules were adjusted to behave as wanted.

The final impact of new rules was then analyzed by generating a set of new solutions from the prescriptive rule set and the newly proposed rules. One new solution for each added rule and a few solutions using combinations of new rules were generated. The productivity and level of alertness were analyzed for each solution, and the data were plotted on a chart. One rule, or a few rules, that improved alertness were chosen to move forward.

Improving Productivity

To examine possibilities for improving productivity, the BAM reference solution became the starting point. As noted, this solution had no constraints other than maintaining a protected level of alertness. Theoretically, then, it should be the most productive solution possible, unless protection of alertness is sacrificed. In our system, it was possible to turn on the prescriptive rule set and apply it to the BAM reference solution. This resulted in “flags” of rule violations in the BAM solution. We compiled statistics for the number of violations of each rule, the typical limit of violated rules and the typical exceedance of violated rules, from which we were able to gain insight into the most limiting rules in terms of productivity.

From these insights, new FTL-scheme solutions were created from the prescriptive rules, with the proposed relaxations added. One new solution was created for each relaxed rule, as well as a few solutions in which combinations of rules were relaxed. The productivity of the new solutions, as well as protected level of alertness, then could be analyzed. One or several of the best candidates for a new rule set then could be chosen for further refinement.

The research has validated the methodology by applying it to the CCARs rule set and the data set representing the Chinese airline.

In three iterations, nine rule changes were tried and five rule changes were introduced. The final result was a rule set in which the average block time per day was increased by 6 percent from 5 hours 59 minutes to 6 hours 21 minutes and alertness was improved between 250 and 700 points on the Common Alertness Scale. Differences in alertness are compared in Figure 3 (p. 6), where the new rule set is named CCARs+. The new solution’s resource index also dropped 8.5 percent.

The following rule changes were introduced:

- Prohibiting pilots from being asked to report for duty more than once in a 24-hour day;
- Reducing the maximum duty time for duty periods that fall partly within 2300 to 0330;
- Relaxing the rule governing maximum block time in a duty period;
- Relaxing the rule governing minimum rest after duty; and,
- Adding a complementary rule for maximum duty time after short rest periods — rest periods that became legal when the original minimum rest-after-duty rule was relaxed.

The parameter changes tested in the case study were large and had a large impact on productivity and alertness. More refined parameter changes could be tested to find a better trade-off between alertness and productivity.

The final rule set was stress-tested. The test showed that the protected level of alertness had increased by 250 to 450 points on the Common Alertness Scale.

Conclusions

These results and conclusions are offered with a major caveat. It is not possible to say that other fleets — regional, long haul, ultra-long haul, cargo, etc. — will necessarily exhibit the same characteristics. These results should not be generalized without further analysis.

Of the five tested FTL schemes, none completely protected against low alertness in the crew schedules. The most concerning patterns encountered in the FTL-controlled crew schedules were the planning of unusable rest during daytime periods, when it would be difficult for the pilots to sleep, and duty periods of maximum length ending close to midnight. These situations are legal, and appeared in solutions generated from all FTL schemes, including the NPRM.

The JARs and CCARs rule sets are comparable in many aspects, both in productivity and in the protection against low alertness. The JARs FTLs are slightly better at protecting against fatigue but less productive if there are many legs in the average duty.

In light of the recent rulemaking, the comparison between the original FARs and the NPRM are worthy of closer consideration. The FARs FTL rules were the most efficient of the original three FTLs studied — but at the cost of very long duty times. Compared with all other rule sets, the FARs FTLs offered the least protection against low alertness.

In this limited analysis, the NPRM performs slightly better from the perspective of productivity then the other prescriptive rules, with the exception of the FARs. When compared with the FARs directly under the three network conditions detailed in Table 1, the NPRM can be expected to result in a moderate decrease in productivity and higher resource demand but appears to favor a lower ratio of duty to block hours.

Table 1: Productivity & Quality of Life Measures under US Network Conditions

Network	Rule set	Productivity (Block hrs/ day)	Ratio of Duty hours/block hours	Resource Index
US	FAR	6:43	1.6	3.73
US	NPRM	6:47	1.48	3.74
EU	FAR	6:23	1.79	3.11
EU	NPRM	6:06	1.61	3.37
CN	FAR	6:37	1.74	3.56
CN	NPRM	6:02	1.58	3.83

From the perspective of alertness, the NPRM seems to offer some improvements over the old FARs rules but contains some of the same loopholes. This is illustrated most clearly in the case of the EU network conditions, where the NPRM solutions and the FARs solutions for the worst-case alertness flights are virtually identical but the productivity of the solution is lowered.

While the prescriptive aspects of the NPRM are a mixed bag, the inclusion of FRM in the NPRM language represents a potentially major step forward in flight and duty time regulation.

The levels of alertness predicted by BAM for the FTLs should be viewed with caution because the model is not yet fully validated in airline operations. When the model is shown to be valid, the safety and business case for FRMS will be further strengthened. Our results indicate that FTLs do not appear to protect well against low alertness — and within an airline, FRMS model-based scheduling should be both safer and more productive.

In the meantime, assuming that current FTL schemes are to be moved toward FRMS, we have described a method for improving an existing FTL scheme to better protect against low alertness while improving or maintaining flexibility and productivity. Finally, we note that the methodology used in this study, for analysis and improvement of rules, can just as well be applied by an operator in scheduling as an essential part of an FRMS.

David Hellerström and Hans Eriksson are business analysts with Jeppesen Systems. Emma Romig is principal investigator, flight deck research and development, with Boeing Commercial Airplanes. Tomas Klemets is product manager with Jeppesen Systems.

Notes

1. FAA. Flightcrew Member Duty and Rest Requirements. Docket No. FAA–2009–1093. <www.faa.gov/regulations_policies/rulemaking/recently_published/media/FAA_2010_22626.pdf>.

2. U.K. Civil Aviation Authority. CAP 371: *The Avoidance of Fatigue In Aircrews*. Fourth edition, 2004. <www.caa.co.uk/docs/33/CAP371.PDF>.

3. Romig, Emma; Klemets, Tomas. "Fatigue Risk Management in Flight Crew Scheduling." *Aviation, Space, and Environmental Medicine* Volume 80 (December 2009): 1073–1074(2).

4. Åkerstedt, T.; Folkard, S. "The Three-Process Model of Alertness and its Extension to Performance, Sleep Latency, and Sleep Length." *Chronobiology International* 14(2), 115–123, 1997.

5. Åkerstedt, T.; Folkard, S.; Portin, C. "Predictions From the Three-Process Model of Alertness." *Aviation, Space, and Environmental Medicine* 2004; 75(3, Suppl.): A75-83.

6. Passive duty is a duty during which a crewmember flies as a passenger to be positioned for further duty.

7. An RI of 1.0 would be achieved if a schedule was staffed entirely without deadheads or passive flights, and with all flight and duty time assigned to productive flying.