5-2021

Human Fatigue Predictions in Complex Aviation Crew Operational Impact Conditions

Suresh Rangan
srangan@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Suresh Rangan entitled "Human Fatigue Predictions in Complex Aviation Crew Operational Impact Conditions." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial Engineering.

Dr. Mingzhou Jin, Major Professor

We have read this dissertation and recommend its acceptance:

Dr. Andrew Yu, Dr. Anahita Khojandi, Dr. Wenjun Zhou

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Human Fatigue Predictions in Complex Aviation Crew Operational Impact Conditions

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Suresh Rangan

May 2021

DEDICATION

I dedicate this work to my parents Rangan, Kamala, my wife Sridevi, kids Manish, Sahithi, siblings and friends.

ACKNOWLEDGEMENTS

I wanted to take this extreme pleasure to acknowledge several individuals and organizations who were instrumental for completion of my Ph.D. research.

My deepest gratitude first goes Dr. Mingzhou Jin, for his guidance throughout my graduate education and truly encouraged me with strong support and excitement. He always gave his ears and patience to all the crazy ideas brought in front of him. Without him I wouldn't have been able to complete this milestone.

Next, Dr. Hans Van Dongen educated this complex human fatigue science and contributed to many of my journal papers. He always assisted me to evolve the ideas and model them to a research character. He has been a true inspiration to me and many others in my team.

I would like to acknowledge all my committee members Dr. Anahita Khojandi, Dr. Andrew Yu and Dr. Wenjun Zhou for their very needed and helpful suggestions and found it to be a prized opportunity to work with them through the last 12 months. I would also like to thank the Department of Industrial & Systems Engineering at the University of Tennessee, my fellow graduate students in the department and all administrative staffs for their accessibility and support throughout my life at UTK.

Finally, this work wouldn't have been possible without the problem space and research opportunity provided by Flight Operations and special thanks and gratitude to its thought leaders John Maxwell, Jim Bowman and Don Dillman. Words cannot describe to the amount of freedom, flexibility and support provided by them to contribute to the world of fatigue science.

I also would like to thank Dr. Ying Zhang, Sri Teja, and Scott Zheng for providing academic and modeling support.

My deepest appreciation belongs to my parents, Kamala and Rangan, wife Sridevi Jagadeesan, kids Manish, Sahithi, siblings and many friends and relatives who always supported me to achieve this accomplishment.

Thank you all very much

ABSTRACT

In this last decade, several regulatory frameworks across the world in all modes of transportation had brought fatigue and its risk management in operations to the forefront. Of all transportation modes air travel has been the safest means of transportation. Still as part of continuous improvement efforts, regulators are insisting the operators to adopt strong fatigue science and its foundational principles to reinforce safety risk assessment and management. Fatigue risk management is a data driven system that finds a realistic balance between safety and productivity in an organization. This work discusses the effects of mathematical modeling of fatigue and its quantification in the context of fatigue risk management for complex global logistics operations. A new concept called Duty DNA is designed within the system that helps to predict and forecast sleep, duty deformations and fatigue. The need for a robust structure of elements to house the components to measure and manage fatigue risk in operations is also debated. By operating on the principles of fatigue management, new science-based predictive, proactive and reactive approaches were designed for an industry leading fatigue risk management program

Accurately predicting sleep is very critical to predicting fatigue and alertness. Mathematical models are being developed to track the biological processes quantitatively and predicting temporal profile of fatigue given a person's sleep history, planned work schedule including night and day exposure. As these models are being continuously worked to improve, a new limited deep learning machine learning based approach is attempted to predict fatigue for a duty in isolation without knowing much of work schedule history. The model within also predicts the duty disruptions and predicted fatigue at the end state of duty.

PREFACE

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
SCIENCE BASED APPROACH TO FATIGUE RISK MANAGEMENT

A version of this chapter was originally published by Suresh Rangan

Rangan, Suresh & Bowman, James & Hauser, William & McDonald, William & Lewis, Richard & Dongen, Hans. (2013). **Integrated fatigue modeling in crew rostering and operations**. Canadian Aeronautics and Space Journal. 59. 1-6. 10.5589/q13-001

As the primary author, I developed the vision to advance fatigue management framework with a science-based rostering and scheduling software to help airlines manage fatigue. The other authors are part of the operations and Washington State University who helped support the ideas.

Suresh Rangan, Samantha M. Riedy, Rob Bassett, Zachary A. Klinck,
Patrick Hagerty, Ethan Schek, Ying Zhang, Steven R. Hursh & Hans P.A. Van Dongen (2020): **Predictive and proactive fatigue risk management approaches in commercial aviation**, Chronobiology International

As the primary author, I set the strategy to publish the thought leading approaches built and practiced in the areas of fatigue risk management. I also worked on authoring core functional areas. Fatigue Risk Management is a collaborative process with the Company, Association and the Scientists.

1.1 Introduction

1.1.1 Air Cargo Operations

Recent Changes in U.S. hours of service regulations across several modes of transportation have brought to the foreground the question of what is a maximally acceptable level of fatigue risk (P. Gander et al., 2011). Advances in mathematical modeling of fatigue have facilitated systematic investigation of this issue in the context of fatigue risk management (Hursh & Van Dongen 2010).

In the international logistics network, Air transport is a critical component to control the flow of goods and services. Logistics involves geographical repositioning of raw materials, work in progress and finished inventories (Bartsch 2013). The air cargo market refers to transportation of good via air either by commercial airlines and the larger cargo airlines. Global e-commerce has been the primary driver to influence the growth of air cargo market. With much of the passenger flights grounded during COVID-19, It has proven even more essential for the cargo carriers in distributing essential goods all over the world, maintaining global value chains. In 2019 prior to COVID-19, freight volumes reached over 61.3 million metric tons.

1.1.2 Operational Challenges and Risk

World's largest express transportation company, provides fast and reliable delivery services to every US address and to more than 220 countries and territories. Its global air-and-ground network is among the most complex distribution networks in the world. With about 20,000 flights scheduled every month through 10 express global sort hubs, this express air cargo operations manages the schedules of 5000+ pilots in 6 worldwide crew bases across 5 different aircraft fleets. Express air cargo operations functions flights to more than 375 airports around the world under Domestic, Flag, and Supplemental FAA regulations.

Operations combine many types of flying such as repeated backside-of-the-clock flying, extended international operations, in-theater flying, day/night swapping, etc. Operational flexibility and scheduling reliability require flight and crew schedules that change dynamically from month to month.

Given such a frequently changing and challenging operational environment, the express air cargo operation's scheduling practices are evolving to include systems and processes that carefully balance operational needs, contractual rules, and the science of circadian rhythms and sleep to manage fatigue. Thus far, this effort has been accomplished primarily through a rule-based scheduling system in conjunction with an experience-based pilot schedule advisory group (scheduling improvement group; SIG). The schedules and rosters generated by this current process meet the federal aviation regulations (FARs) and the contractual provisions of the collective bargaining agreement (CBA). To supplement this process, a more dynamic, science-based fatigue risk management approach, based on state-of-the-art mathematical fatigue modeling (Achermann 2004), is being pursued.

1.2 Fatigue

1.2.1 Fatigue as Safety Risk

Fatigue is a safety risk in aviation because of performance deficits in flight crews that have been shown to be correlated with fatigue. Per ICAO's definition (ICAO, 2015) Fatigue is *"a physiological state of reduced mental or physical performance capability resulting from sleep loss, extended wakefulness, circadian phase, and/or workload (mental and/or physical activity) that can impair a person's alertness and ability to adequately perform safety-related operational duties*". At this express air cargo, owing to the around-the-clock and across-time-zones nature of its operations, the development of a science-based system for tracking and managing fatigue risk will be critical for safe operations and operational flexibility going forward. To improve upon proactively-implemented fatigue-friendly scheduling policies, this thought leader in express air cargo operations developed science-based rostering and scheduling software (Costa et al, 1990) to further minimize fatigue and enhance alertness in operations. This paper outlines the development process for this software and discusses how that software model will integrate into the company's operational processes. Once validated and implemented, the software model is likely to improve both scheduling performance and increase flight safety.

Current methods of predicting fatigue require information about sleep times and durations that is measured, assumed, or mathematically estimated. However, individuals differ substantially in the timing and duration of their sleep under given circumstances due to neurobiological trait, prior sleep/wake history and circadian state, and/or situational factors such as availability of sleeping facilities or competing demands on their time. This causes potentially considerable inaccuracy in fatigue estimates from one individual to another. In settings where fatigue is of significant concern, such as 24/7 operations or trans meridian travel, this inaccuracy can have substantial adverse consequences for fatigue risk management and for mission safety and success. The present invention overcomes this deficiency by taking a statistical approach to accounting for sleep behavior in estimating fatigue. That is, a distribution of fatigue trajectories over time is derived, using an available mathematical model of fatigue, from the distribution of sleep timings and durations that has been or may be observed under given circumstances. The distribution of fatigue trajectories (D. Darwent et al, 2010) can be readily used to evaluate risk from fatigue, to compare alternative duty schedules to select the least fatiguing option, and to optimize sets of duty schedules to minimize fatigue risk while also maximizing productivity or other schedule objectives given operational scheduling constraints.

There are three types of fatigue: transient, cumulative, and circadian:

- Transient fatigue is acute fatigue brought on by extreme sleep restriction or extended hours awake within 1 or 2 days.

- Cumulative fatigue is fatigue brought on by repeated mild sleep restriction or extended hours awake across a series of days.
- Circadian fatigue refers to the reduced performance during nighttime hours, particularly during an individual's "window of circadian low" (WOCL) (typically between 2:00 a.m. and 05:59 a.m.).

Researches show that the accumulation of "sleep debt" (K. Spiegel et al., 1999), should be a consideration for the person to recover from cumulative fatigue. However, it depends on the amount of debt and time and quality of sleep during the recovery process.

1.2.2 Need for Sleep

1.2.2.1 Types of Sleep

We are meant to spend about a third of our lives asleep. The optimal amount of sleep per night varies between individuals, but most adults require between 7 and 9 hours. Sleep science (R. Ferri et al., 2008) makes it very clear that sleep cannot be sacrificed without consequences. Sleep has multiple functions – the list keeps growing - but it is clear that it has vital roles in memory and learning, in maintaining alertness, performance, and mood, and in overall health and well-being. A complex series of processes is taking place in the brain during sleep. Various methods have been used to look at these processes, from reflecting on dreams to using advanced medical imaging techniques (JM. Gottselig, et al., 2006). Sleep scientists have traditionally looked at sleep by monitoring electrical patterns (T. Åkerstedt, et al., 2002) in brain wave activity, eye movements, and muscle tone. These measures indicate that there are two very different types of sleep as shown in Fig 1.

- Non-rapid eye movement (Non-REM) sleep; and
- Rapid eye movement (REM) sleep.

1.2.2.2 Factors affecting sleep quality

Sleep quality (its restorative value) depends on going through unbroken non-REM/REM cycles (which suggests that both types of sleep are necessary and one is not more important than the other). The more the non-REM/REM cycle (McCarley 2007) is fragmented by waking up, or by arousals that move the brain to a lighter stage of sleep without actually waking up, the less restorative value sleep (T. Roenneberga et al., 2007) has in terms of how you feel and function the next day. Uninterrupted non-REM/REM cycles are the key to good quality sleep, so operators should develop procedures that minimize interruptions to crewmembers' sleep. Rest periods (in flight or on layovers) should include protected blocks of time (sleep opportunities) during which crewmembers are not contacted except in emergencies. These protected sleep opportunities need to be known to crewmembers and all other relevant personnel. For example, calls from crew scheduling should not occur during a rest period as they can be disruptive. The operators should collect data, understand the sleep distributions and establish data driven processes to protect the disruptions.

1.2.3 Fatigue Science

Fatigue resulting from sleep loss and circadian (i.e., 24-hour) rhythm is associated with decreased capacity to perform cognitive tasks and increased variability in performance (Jackson & Van Dongen 2010), leading to greater probability of errors, incidents and accidents (Van Dongen & Hursh 2010) and thus representing a safety risk (Philip & Åkerstedt 2006). Like any other safety risk in aviation, fatigue needs to be managed (MR Rosekind et al., 2002) (Caldwell 2005). For effective fatigue risk management in complex, around-the-clock operations, a scientific understanding of the neurobiological mechanisms underlying fatigue is essential. At the cognitive/behavioral level, these mechanisms are well understood, involving two key physiological processes: a sleep/wake homeostatic process and a circadian process (Borbély & Achermann 1982). The sleep/wake homeostatic process tracks sleep history (P. McCauley et al., 2009) and seeks to balance time spent awake with an appropriate amount of recuperative sleep. The circadian process, driven by the biological clock in the brain, tracks time of day (Waterhouse & DeCoursey 2003) and seeks to place wakefulness during the day and sleep during the night.

The homeostatic and circadian processes normally operate in tandem to provide a stable level of alertness during the day and consolidated sleep during the night (Dijk & Czeisler 1994). However, when deviating from a normal schedule of daytime wakefulness and nighttime sleep, the interaction of the two processes leads to fatigue and performance impairment (Van Dongen & Dinges 2005). During nighttime operations, the homeostatic and circadian processes align to steadily increase fatigue over time of night, while also making it difficult to obtain enough sleep during the day (Åkerstedt 2003). When crossing time zones, the circadian process becomes temporarily desynchronized and takes several days to adjust to the new time zone, which gives rise to the phenomenon of jet lag (T. Reilly et al., 2005). These fatigue challenges are routinely faced by the flight crews.

1.2.4 Fatigue and Alertness Model

1.2.4.1 Fatigue Modeling

In aviation operations involving backside-of-the-clock flying, international travel, day/night swapping, and/or other challenging schedules, it is difficult to foresee at which times the homeostatic and circadian processes produce the greatest level of fatigue. Fortunately, mathematical models have been developed (MM. Mallis et al., 2004) (Hursh & Van Dongen 2010) to track the two processes quantitatively and predict the temporal profile of fatigue given a person's sleep history, planned sleep/wake schedule, and expected day/night exposure (i.e., time zone). Such fatigue models have been applied successfully in aviation (Belyavin

& Spencer 2004) (KJ. Kandelaars et al., 2006), and continue to be improved (P. McCauley et al., 2009) (MA. St. Hilaire et al., 2007) (T. Åkerstedt et al., 2008).

Fatigue modeling can be integrated productively with rostering and scheduling software to produce schedules that manage both operational cost and predicted fatigue (Belenky & Van Dongen 2009) (Romig & Klemets 2009). At the planning side of operations, work schedules can be evaluated to determine if the roster solution is sufficiently fatigue-friendly. At operations control, schedule disruptions, commonly known as Irregular Operations (IROPS), can be reviewed in advance in order to implement fatigue mitigation strategies such as flight augmentation (JA. Caldwell et al., 2009). In more advanced applications (HPA. Van Dongen et al., 2007), flight crews can utilize a personalized fatigue modeling tool to help them proactively manage fatigue, e.g., by planning rest opportunities at sleep-conducive circadian times. Fatigue modeling can also be useful for investigating fatigue-related incidents or other post-schedule analyses.

The current human alertness or fatigue models as shown in figure 2 does one of the following 1) takes input of sleep and work episodes and it will predict fatigue levels. 2) Provide input of work schedule. The sleep predictor algorithms embedded inside first determine the sleep episodes, the sleep and work are then inputted to the model to predict fatigue. Bio mathematical modeling is a means of objectively estimating the potential impacts of fatigue on performance. They focus purely on the Process S which is built on timing and duration of sleep and Process C the circadian system.

Several bio mathematical models have come forward to provide a useful framework for the operations to predict the effects of fatigue on performance. This helps the operations to build fatigue friendly schedules. The newer models also take into account such as sleep inertia (Tassi & Muzet 2000) which is a component that occurs after awakening that can last anywhere from 30 minutes to few hours. Some models are beginning to understand the task related inputs like number of flights in a duty, time zone transitions (KJ. Kandelaars et al., 2006) etc.

1.2.4.2 Current Sleep Model and the Timing of Human Sleep

In humans for many decades, the sleep and wake durations are modeled with exponential and power law models (S. Bernard et al., 1969). This model stages sleep regulating variable (Process S) where it decreases as exponential function when sleep and rises as power law function when awake. The regulative variable simply builds up the sleep pressure during awake and declines during sleep. Fig. 3 above demonstrates the circadian waveform of wake threshold (L) using sleep deprivation results (S. Daan et al., 1984) (P. Achermann et al., 2004). Subscripts of S refer to 30 min units since onset of wakefulness at 7am.

1.2.4.3 Sleep Prediction Model

Sleep timing and duration are primary determinants of alertness. Sleep timing and duration depend on the same biology underlying alertness, but other, non-biological factors also play a role when working consecutive nights or when traversing multiple time zones. Besides prior sleep/wake/work schedule, these factors range from jet-lag, availability of hotel facilities (time for check-in and -out) and store opening hours to communications with family at home, etc. Some of these factors vary substantially from person to person and from duty to duty, such that pre-duty, in-flight, layover and post-duty sleep schedules exhibit statistical distributions (Darwent et al., 2010). This industry leader air cargo airline is developing a distribution-based prediction model for sleep timing and duration (Van Dongen 2004) based not only on the underlying biology but also on observations of real-world in-flight and layover sleep behavior, as a function of time spent away from base, rate of change of time zone transitions, direction of time zone transition, time of day, prior duty schedule, location and the duty following. This will allow for probabilistic predictions of alertness that account for the natural variability in sleep behavior. Fig. 4, 5 shows a simulated example. Data from a recent study of actigraphy (DS. Lauderdale et al., 2008), a wrist-worn rest/activity monitor were used as surrogate data for a scenario involving a duty start time and duration of 0100 and 8h respectively. This was followed by a layover period of ~25h in which sleep was simulated as distributed according to the surrogate data, from which 100 random samples were drawn. For the duty period following the layover with extended wakefulness, alertness predictions were made after 12h of duty begin time (at time of day 00:00). The Fig. 6,7 shows the distribution of alertness predictions based on the distribution of the layover sleep. This alertness distribution can be used to make informed decisions regarding the need for fatigue countermeasures or additional time for sleep.

1.2.4.3 Validation of 2-process model

The original Two Process model has been validated against many diverse disrupted sleep schedules, such as sleep deprivation and sleep restriction. It was confirmed by several experiments in the laboratory (D. Hand et al., 2001). Using subjective alertness data from a number of experiments of altered sleep/wake patterns, it was found that alertness was also predictable. Circadian and a homeostatic component in combination with a component for sleep inertia (Taasi & Muzet 2000). The output of the model has been validated against subjective ratings, performance and electrooculogram (EOG) measures of sleepiness, and has shown considerable accuracy (Folkard & Akerstedt 1991). There are a number of studies that have confirmed the variations (Carrier & Monk 2000) in performance. There are several mathematical models (MB Spencer 1987) (S. Folkard et al., 1997) (P. McCauley et al., 2009) that were developed based on the seminal process.

1.2.4.4 Limitations of the 2-process model

All bio math models predict fatigue for an average person, assuming that individual requires about 8 hours of sleep per night to remain fully rested and has regular circadian rhythms. They help us to set up safer operations with risk assessments. They are good at doing incident or accident investigations when all the variables are known. (A. Shawn, et al., 2011). The model doesn't not also understand the experience and history of the operations (CASA 2014). (For example, the model will not be able to differentiate the differences, mitigations and the learning history between the day and night operations). The model is definitely not recommended for the individuals in a specific operating environment and doesn't necessarily correlate to a safety risk. Chronic sleep deprivation comes out as a big disadvantage and only couple of models have tried to address it. Models also haven't addressed the increase in fatigue levels due to the workload in operation. The number of sectors combined with low duty times has seen increase in objective and subjective levels of fatigue (G. Belenky, FRMS Forum 2014) (D. Powell, et al., 2008) (HPA Van Dongen, et al., 2011).

NASA-TLX developed a model to measure workload. The workload rating is based on the weighted average on size sub scales: 1) mental demand, 2) physical demand, 3) temporal demand, 4) effort, 5) performance and 6) frustration level. The measure combines weighted ratings on the six subscales to provide one integrated workload rating (SG Hart, 2006). For operations more emphasis is given to the errors and risks that the lower alertness exposes and not the alertness itself. Intersection of tasks demands to alertness should be explored and that's the gap. There is a difference between in fatigued state when you are in cruise vs landing in a busy airport.

The models were successful in predicting variations in alertness and performance in both laboratory and field settings. Unfortunately, while these mathematical models might be able to account for the trends in productivity described above, they have great difficulty in accounting for the trends in risk. While the process captures well the seminal two process and predicts the neurobiological functions, the model doesn't understand well the cumulative performance impairments observed over days of chronic sleep restrictions. This was exposed by the recent studies on sleep dose response (G. Belenky et al., 2003) (HPA. Van Dongen et al., 2003). They found out that the existing of an additional process (ML. Johnson et al., 2004) that is modulating the homeostatic process over the long term.

In one case that the McCauley and colleagues in Sleep and Performance research center are able to extend the effort and discovered a modeling solution (P. McCauley et al., 2009). The model only predicts alertness if sleep and wake schedules are provided. In our process, the model is extended to predict sleep and thereby performance. As you could see from the Fig. 8 below the model predictions for daytime averages of performance lapses on a psychomotor vigilance test, expressed a relative to baseline, across three different 4-week rotation schedules. This was captured from the medical residency program.

1.2.5 Measuring Fatigue, Sleep and Circadian

Fatigue is a complex phenomenon and measuring fatigue at an individual level is complicated. There is no single standard measurement that can be applied. A wide variety of fatigue measures are used in measuring human performance. Currently there are two types of measurements. Subjective questioning and objective measures.

Subjective questioning (M. Matousek et al., 1988) helps the researchers to understand the state of the individual during the task or at the point of interest by the researcher. Measurements include understanding sleepiness, fatigue levels, type of person etc., Subjective measurements rely on individual mood, impression and will account for a lot of individual differences. However, they are inexpensive to collect and easier to analyze the data. In the industry settings, these measurements are done in large numbers to account of skew in data. Sleep researchers have standardized these questionnaires to differentiate between the mental and physical fatigue levels. For subjective measurement of sleep length and quality, sleep diaries both paper and digital are widely used in laboratory and field test experiments. For sleepiness and fatigue levels, there are a number of scales that is being used for measurement and the few commonly used are listed below

- Epworth Sleepiness Scale (ESS) developed by Johns (1991) (YW. Cho et al., 2011)
- Fatigue Assessment Scale (FAS) developed by Michielsen, De Vries, Van Heck. (H. J. Michielsen et al., 2003)
- Karolinska Sleepiness Scale (KSS) developed by Akerstedt and Gillberg (T. Akerstedt et al., 1990)
- Samn-Parelli Fatigue Scale (SP) developed by (T. Akerstedt et al., 1990)

Objective measurements include measuring sleep, reaction time, vigilance, circadian body clock cycle etc., Polysomnography is the most reliable technique for measuring sleep (E. Hertenstein et al., 2018) and it involves sticking electrodes to the scalp and face. It measures EEG (brainwaves), EOG (eye movements) and EMG (muscle tone). This method often requires the technicians in the lab and they are well trained. This is often measured in a lab setting and are relatively expensive, obtrusive and time consuming. It's hard to do the same in operational settings with field studies, in some remote cases it's even impossible.

Recently the scientist has explored a new method for measurement of sleep as sleep can also be measured through activity movement. Actigraphs (M. Marino et al., 2013) are wrist worn devices running on accelerometers. Sleep is measured from activity movement through an algorithm that is validated against the polysomnography. The participants will be wearing these devices 24x7 during the test period. Actigraphy measurement (M. Marino et al., 2013) is widely used today due to the ease of data collection and it's not obtrusive and relatively inexpensive.

Since its not directly measuring sleep there is always a possibility of predicting errors due to the algorithm that is translating the movement into sleep.

The other important feature to help measuring cognitive performance is the circadian body clock (MH. Vitaterna et al., 2001). It is one of most difficult measurement of physiology markers. Core body temperatures (MH. Vitaterna et al., 2001) and melatonin (RJ. Reiter et al., 1984) hormones are the two rhythms that are measured. Core body temperatures are measured using ingested temperature pill or using rectal inserted probe. Melatonin is measured from blood, saliva or urine samples at different intervals. All these measurements are heavily challenged by the differences that exist between individuals (HPA. Van Dongen et al., 2007) (HPA. Van Dongen et al., 2005)

1.2.6 Propose Fatigue Model Architecture for Aviation

In 2009, Express Air Cargo Flight Operations began developing a framework for software to evaluate schedule effectiveness from a fatigue risk management perspective, laying a foundation for an improved schedule design and operations process. The software development process includes the implementation of a validated model for predicting fatigue risk in crew schedules across the breadth of its operations. Ideally, the software model and its applications will include the ability to:

1) Perform fatigue predictions for pairings and rosters.

2) Evaluate fatigue in “what if” scenarios during planning and operations.

3) Apply phased fatigue mitigations during schedule building to enhance alertness and manage fatigue risks.

4) Suggest physiologically realistic sleep strategies to mitigate fatigue in challenging operations.

5) Advise the Systems Operations Center (SOC) on operational mitigations that could be applied to manage fatigue during IROPS.

6) Provide a scientific basis for planned fatigue mitigations such as split-duty sleep opportunities, dedicated sleep rooms at hubs, lay-flat crew rest bunks on long-range flights, etc.

7) Serve as a personal fatigue risk management tool for crew members to help them plan optimal, physiologically realistic sleep schedules.

A related aspect of the software model development process will likely include field collection of de-identified human performance and alertness data to validate the effectiveness of current prescriptive scheduling controls and planned fatigue mitigation strategies. This data collection and analysis process is anticipated to be on-going, and should be valuable, in particular, for the development and validation of future improvements to software-based fatigue models throughout the airline industry. Currently, the airline's fatigue model framework will require development and validation of the following software components:

1.2.6.1 Basic Fatigue Model

The model framework will be based upon the seminal two-process model of sleep regulation to predict fatigue (S. Daan et al., 1984). This model utilizes mathematical equations to predict fatigue throughout the sleep/wake homeostatic and circadian processes (Borbély & Achermann 1999). Anticipated improvements to the existing model include the development of modules to predict the cumulative effects of chronic sleep restriction (P. McCauley et al., 2009) and circadian phase shifting (RE. Kronauer et al., 2007).

1.2.6.2 Sleep Solver

The software model framework also calls for the development of a proprietary "sleep solver" component, which will identify all realistic sleep opportunities within a schedule based on the predicted sleep/wake homeostatic and circadian process states, and will include fatigue mitigation prescriptions unique to a particular schedule (i.e., identifying the best sleep opportunities predicted to minimize fatigue within the overall context of rules and regulations, planned fatigue mitigations, augmentations, etc.).

1.2.6.3 Knowledge Manager

This software module will process de-identified alertness data collected in the field to track sleep and fatigue as a function of duty sequence patterns. Based on probabilistic modeling (D. Darwent et al., 2010) (Van Dongen & Hursh 2010), the knowledge manager should be able to improve the predictive capabilities of the sleep solver as increasing amounts of data are collected on actual sleep/wake behavior in the field.

1.2.6.4 Mitigation Manager

Within the model framework, the mitigation manager component would iteratively evaluate available fatigue mitigations, and compare predicted fatigue levels (generated by the basic fatigue model component of the program) both with and without these mitigations applied. Ideally, the mitigation manager would rank pairings on this basis and suggest a fatigue mitigation strategy that the other components of the software predicted to best achieve overall schedule effectiveness and alertness. This component would also estimate required napping opportunities and compute the number of augmentations and hub sleep rooms needed for the proposed schedule.

1.2.6.5 Fatigue Workbench

Within the model framework, the fatigue workbench component would enable ad hoc manipulation of schedules to help schedule designers interactively apply “what if” scenarios and evaluate the predicted fatigue-related benefits of incorporating specific sleep opportunities for the proposed schedule.

1.2.6.6 Fatigue Predictor

The software model will also include parameters for the development of a fatigue predictor component. This component would be a comprehensive reporting system to provide planners and schedulers the ability to compare the predicted fatigue risk associated with different scheduling options. This component would permit the comparison of multiple pairing and roster solutions on the basis of their fatigue-friendliness in the context of other scheduling considerations and constraints. It then would enable planners and schedulers to apply fatigue mitigations and schedule changes in order to resolve high fatigue risk situations and improve the fatigue-friendliness of the overall operation. Other tools for the model framework under consideration for development include modules tracking crew status (HPA. Van Dongen et al., 2007), operating environment operational task demands (Van Dongen & Hursh 2010) and workload (KM. Vitellaro et al., 2003). Future initiatives will address the potential roles of diet and exercise in fatigue risk management.

1.3 Fatigue Prediction

1.3.1 Risk Assessment and Human Factors

Aviation systems are characterized by a huge number of complex interactions and interdependencies. Crewmember fatigue is now acknowledged as a hazard that predictably degrades various types of human performance and can contribute to aviation accidents and incidents. In addressing human factors in risk assessment, performance optimization is attempted to reduce human related failures. It is estimated that up to 90% of all workplace accidents have human error to cause (Feyer & Williamson 1998). Several approaches have evolved over the years to manage fatigue related risk developing defenses based on assessments of risk (V. Socha, et al., 2015). In the world of fatigue risk management two primary type of approaches are followed, a) Prescriptive or Compliance based approach: Where the operations must remain within the prescribed limits established by the regulatory agencies. b) Performance or Risk based approach where a through examinations and protections to avoid risk are placed in the operational setting. Risk based approach manages the operations through predictive, proactive and reactive methods. The predictive process has to determine risk associated with fatigue through proper crew planning control and consideration of known factors affecting sleep and fatigue and their effect on performance. Bio mathematical models (Roberts & Nesthus 2016) (Van Dongen 2004) are used for testing of current understanding of the matter how factors such as sleep deprivation; work load or circadian rhythms affect human performance. The process of modelling starts with simulation of so called" development data set", where factors such as self-evaluation of the fatigue and data collected by fatigue measurement are used. That data is used for prediction of different situations. Then, the modelled predictions are tested using newly acquired data.

1.3.2 Regulatory impact on Fatigue Risk Management

1.3.2.1 Flight Time Limitations

Since the fatigue factor is one of the possible causes in many accidents, in the 1944 Chicago Convention, the civil aviation authorities determined that fatigue in the flight crew is an issue and formed rules and named it as FTL (Flight and Duty time Limitations). This is the rule set that is necessary for the flight crew has to comply so the safety of the flight is not decreased. It limits for how many hours a pilot can work in a day, week or month. This can be complex and is prescribed typically for each state. For Example. European countries should comply with EASA flight time limitations and US based Carriers should comply with FAA Part 117 regulations. The regulators are looking to revise the rules based on scientific findings with respect to sleep, sleep loss, fatigue, circadian rhythms, and performance into the prescriptive rules.

1.3.2.2 Prescriptive Approach

Aviation regulations as dictated by each state regulators are relatively strict rules which must be followed. Typically, no exemptions are allowed. This forces the operators to define the more conservative ruleset so they don't exceed due to operational impacts and constraints. This is characteristically achieved through air operational manual or through the collective bargaining agreement between the pilot's association and the company. These prescriptive rules are very critical because it significantly improves safety and easier to be programmed through the company crew operations systems. Since the aviation network and its operation brings so much of complexity, a novel idea is needed to transition from a regulation-based framework to a performance-based framework. This innovation move was further sponsored by US Federal regulations in Part 117 $117.7 provision. This is a brave alternative to the "one-size fits all" prescriptive approach. Such alternatives have to go through the non-inferiority data analysis to prove that the performance-based criteria is equal or better than the regulator defined criteria.

1.3.2.3 Performance Based Approach

Non-significant (Mascha & Sessler 2011) could be due to the lack of statistical power. And therefore, is not a positive indication of equivalence. In contrast, attempts are made to apply the approach that is widely used in the pharmaceutical industry. The approach is non-inferiority design that can be used to test for equivalence, superiority, and non-inferiority (Walker & Nowacki 2011). Typically, the carriers apply for exemption to the state regulatory body which includes collecting human performance data and analyzing them for non-inferiority testing. The regulator then reviews all the information including operational data and if they are convinced will authorize the operate under the control parameters of the new performance-based regulation.

1.3.3 Probability based Fatigue Risk Prediction

Sleep timing and duration depend on the same biological (homeostatic and circadian) processes as that underlying alertness. However, non-biological factors also play a role, especially when working consecutive nights or when traversing multiple time zones. Aside from the sleep/wake/work schedule itself, these factors range from availability of hotel facilities (check-in/out times) or store opening hours to communications with family at home, etc. Some of these factors vary substantially from person to person and from duty to duty, such that pre-duty, in-flight, layover and post-duty sleep schedules exhibit probabilistic distributions. The air cargo airline is developing a prediction model for sleep timing and duration based not only on the underlying biological processes but also on systematic patterns in observations of real-world in-flight and layover sleep behavior (D. Darwent et al., 2010) (Van Dongen & Hursh 2010), as a function of prior and planned duty schedule, time of day, and location. This will allow for fatigue distribution modeling, which entails making probabilistic predictions of alertness that account for the predictable component of the natural variability in sleep behavior.

Fig. 10 shows a simplified example of fatigue distribution modeling. Data from a recent study of actigraphy (i.e., wrist-worn rest/activity monitor) in 11 pilots were used as surrogate data for a simulated 48h scheduling scenario. The scenario involved a 9h duty period starting at 01:00. This was followed by a 25h layover period, in which sleep was simulated to be distributed according to the surrogate data, from which 100 random samples were drawn (with replacement). Then there was a second, 12h duty period starting at noon. Alertness predictions were made across the 48h scenario using the two-process model (AA. Borbély et al., 1999). Figure 1 shows the distribution of the alertness predictions during the layover and during the second duty period, based on the sampled distribution of the layover sleep. At the end of the second duty period, alertness scores were 0.35±0.05 (mean ± standard deviation). Compare this to the beginning of the scenario, 48h earlier, where the alertness score was 0.67 (assumed to be the same for each pilot in this simplified example). In the airline's model framework, such alertness distribution information can be used to statistically compare the predictions to those of other possible duty schedules, and to make informed decisions regarding the need for alertness-enhancing countermeasures. In Fig. 10, Alertness predictions for a simulated scheduling scenario. Distribution modeling is shown for 100 samples of varying sleep timing and duration in the layover period (some of the predicted alertness curves overlap). The gray bar indicates the layover period; black bars indicate times when sleep occurs in some or all of the sampled cases

1.4 Fatigue Management

1.4.1 Fatigue Management Process

In commercial aviation and other safety-sensitive industries, advances in safety management approaches are expected to be widely shared among stakeholders. This also applies to fatigue risk management (FRM), which is an important facet of safety management in 24/7 operations. In this context, fatigue is operationally defined as a physiological state of reduced mental or physical performance capability resulting from sleep, circadian, or workload factors that can impair the ability to operate safely or perform safety-related duties (IATA. 2011). Fatigue is profoundly influenced by biological processes underlying sleep/wake regulation (Van Dongen et al. 2016), with a "homeostatic process" driving sleepiness as a function of time awake and prior sleep loss, and a "circadian process" driving alertness as a function of time of day. Together these two processes determine the overall level of fatigue in a manner that is predictable and captured in bio mathematical models of fatigue (Hursh et al. 2016).

Manipulating the two biological processes provides a means of managing fatigue risk, but the circadian process is remarkably resilient to manipulation (Smith & Eastman 2012). By contrast, manipulating the homeostatic process merely requires adjustment of the timing and duration of wakefulness and sleep. In practice, this can be accomplished, in part, by strategic scheduling of duty periods and protecting opportunities for sleep. As such, targeted scheduling of duty periods and protected rest breaks is a core aspect of FRM in many operational environments. Modern FRM approaches are multi-faceted, incorporating predictive, proactive, and reactive components based on science and operational experience, which collectively serve to manage fatigue and safety (Rangan et al. 2013). The predictive component of FRM includes duty schedule generation and fatigue risk review processes, along with rostering, which occur in advances of daily operations. The proactive component includes schedule adjustments possibly needed during daily operations, as well as the application of fatigue countermeasures. The reactive component involves data collection, including a non-punitive fatigue reporting system, to continually evaluate and improve FRM processes.

FRM approaches and procedures for commercial aviation are promulgated globally by international organizations (IATA. 2011), but know-how regarding the implementation of FRM in practice is not easily accessible across the spectrum of commercial aviation settings. Here, to help fill this gap, I focus on some science- and practice-based predictive and proactive approaches to FRM currently implemented at a US-based commercial cargo carrier.

1.4.2 Fatigue Evaluation Process in Crew Schedules

1.4.2.1 Predictive process of managing crew schedules

In model-based FRM, a bio mathematical model of fatigue (BMF) is used to help generate duty schedules that meet operational needs while simultaneously mitigating fatigue risk (Van Dongen & Belenky 2012). Large-scale operations may accomplish this by integrating a BMF with a computer-based flight crew schedule optimizer, which can yield schedules that effectively avoid highly fatiguing duty patterns (Romig & Klemets, 2009). However, this approach has the potential to produce duty schedules that meet operational needs by eliminating both highly fatigue-inducing and highly fatigue-avoiding schedules. Thus, the outcome could be a rearrangement rather than a reduction of overall fatigue risk across the operation, which may not achieve an acceptable balance between meeting operational demands and addressing fatigue and safety.

A variation of model-based FRM that avoids this problem and incorporates best-practice principles of FRM (KA. Honn et al., 2019) involves first generating duty schedules without including a BMF in the optimization process. This is followed by an evaluation of proposed duty patterns using a BMF, to identify and address the most fatiguing duty periods through a process involving relevant stakeholder input. Fig. 12 (bottom) illustrates this approach as implemented at a US-based commercial cargo carrier, where the crew scheduling group, in coordination with the pilot union, engages in a predictive FRM process focused on proposed schedules for the upcoming month. There are two primary goals: 1) efficiently utilizing available resources (aircraft and pilots) to meet operational demands; and 2) solving that logistical problem such that pilot fatigue is considered and minimized to achieve high levels of safety.

The process starts with the primary inputs of scheduled flights between airports using available aircraft and available pilots, which are considered by an automated system to solve the logistical problem. The initial solution is constrained by regulatory flight and duty time limits, additional limits potentially included in collective bargaining agreements, and predefined rules for avoiding trips known to be fatiguing based on historical experience. The result is a set of proposed trips that become the basis for the monthly flight schedules. This initial solution is subjected to an evaluation that considers potential fatigue risk factors determined from four sources:

- BMF computations that estimate fatigue from the timing of the flights, opportunities for sleep, and time of day.

- Research data collected previously by the airline that have identified certain flight sequences as causing greater fatigue.
- Fatigue reports from pilots that have experienced similar trips in the past.
- Operational risk factors, such as complex airports, that could exaggerate risk if fatigue is present.

These risk factors are considered, and proposed trips are rank ordered to flag trips that have the highest combination of risk factors. Flagged trips are then assessed to identify the primary fatigue factors involved, and, if needed, a revision process is undertaken. For the revision process, a fatigue management group including representatives from the crew scheduling group and, if applicable, from the pilot union examines the fatigue factors involved and proposes changes to the flagged trips.

Proposed changes may be made manually to the specific trips for which risk factors have been identified, in order to include appropriate mitigations. For example, a schedule might be created that involves a series of four flight segments, all occurring at night. Depending on the destinations and the length of the segments, such a trip might not pose excessive fatigue risk. But, under other circumstances involving relatively long flight segments, little or no opportunities for sleep or rest between segments, and difficult airport approach requirements, such a trip could be split into two separate trip fragments. These fragments might be assigned to other trips containing flights that are shorter and less fatiguing. The original trip could also be revised to include a layover between the first two and the last two segments.

Alternatively, changes may be made by adding additional constraints to the automated system to avoid fatigue factors globally and creating a new set of proposed trips. This new set is again assessed to ensure that the intended fatigue reduction is achieved, without introducing new fatigue factors. There are often multiple ways to change a schedule to achieve equivalent reductions in fatigue risk, allowing for a solution to be selected that does not cause unacceptable cost increases. However, managing fatigue should be an overarching priority with long-term value, so that reasonable short-term costs to reduce fatigue would normally be considered acceptable. When the revised set of trips is approved, the monthly schedule is published.

1.4.2.2 Proactive sleep opportunity management

Proactive planning of sleep periods immediately before and after duty periods (Boivin & Boudreau 2014), along with planned napping during duty periods (Ruggiero & Redeker 2014), can help mitigate fatigue, especially during night operations commonly encountered in cargo flight operations. Duty schedules that are protective of sleep opportunities and known well in advance can facilitate this approach to fatigue mitigation.

The recuperative potential of sleep periods and naps depends in part on the availability and adequacy of sleep facilities. At a US-based commercial cargo carrier, in addition to providing hotel accommodation during layovers, this issue has been addressed as follows:

1. Building quiet, comfortable, secure sleep rooms at departure and arrival airports, and making them available before, during, and after duty periods for pilots to maximize preparatory and recuperative sleep opportunities.
2. Instituting a wake-up call program, which puts the responsibility of waking up a pilot napping before a flight on company personnel (e.g., the station manager in charge of the flight). In the wake-up call program of a US-based commercial cargo carrier, pilots using the wake-up call program are awakened prior to the scheduled report time (while allowing sufficient time for any impairment from sleep inertia to dissipate). If their flight is delayed, the wake-up call is also delayed, and the pilots' sleep opportunity is increased. These proactive strategies are used frequently, especially at night; see Fig. 12 (bottom). Pilots have reported that they reduce anxiety and allow for additional, more restful, sleep.

1.4.2.3 Reactive Process of Managing Fatigue

The predictive process will involve identifying specific pairing designs that create fatigue risk without intervention in the planning process. Here the fatigue model would be used to analyze the schedules generated through cargo carrier's prescriptive rule sets, taking into account planned fatigue mitigations and physiologically realistic sleep estimates.

Fig. 13 demonstrates how an us commercial cargo carrier have adopted the three major fatigue risk identification processes: predictive, proactive, and reactive. CBA = collective bargaining agreement; FARs = federal aviation regulations; SIG = scheduling improvement group; FRMS = fatigue risk management system. The proactive process involves identifying situations that may lead to fatigue during the day of operations. Here the fatigue model would be used to monitor schedule changes in real time and alert schedulers of increased fatigue risk, to apply fatigue-friendly schedule revisions, and to suggest personalized sleep/wake strategies for pilots operating flight pairings at risk for fatigue. Finally, the reactive process involves data-driven identification of flight parings that resulted in fatigue risk during post schedule analysis. Here the model would also incorporate anecdotal data generated from fatigue and incident reports, and would be useful in preventing future fatigue-related incidents.

1.5 Fatigue Data and Analysis

1.5.1 Background

Fatigue is inherent to flight operations that fly through multiple time zones and also on irregular shift patterns (Smith et al, 1998). Many aspects of performance are subject to fatigue-related impairment, and the consequences (G. Belenky et al., 2014) in a workplace can be complex. Operators have challenges to understand the impact of fatigue in their operations. Operators and Regulators have turned towards using fatigue risk management system. A good foundational Fatigue Risk Management offers tools to manage and mitigate organizational fatigue. However, to provide scientific proof and guidance to their operations, a comprehensive robust data collection is necessary. Operational performance data collection helps to identify the practices requiring additional attention and also helps measure the effectiveness of the current risks. Subjective and objective data collection metrics can be used as the key safety performance indicators (SPIs) to help identify the fatigue hazards.

It’s always challenging to collect data during aviation operations because there is not too much of opportunities to perform experimental controls and measurement strategies. However, collating and comparing data collected through a combination of techniques provides the airlines more avenues to help understand their operations.

1.5.2 Types of Data

In the fatigue risk management world, operators on their crew schedule network monitor sleep history, crew subjective feedback and performance measurements to have in-depth information to detect fatigue hazard. Crew will be asked to volunteer and participate in the efforts to help improve the safety of operations. This will be in more focus whenever there is a change to the existing pattern or a new pattern is introduced. The operators will typically compare to something that they felt safe to operate. The study design and measures used need to be provide scientifically defensible answers to the operational questions being asked about fatigue and safety. In the aviation context, recommended criteria for selecting measures include the following: They have been validated, to confirm that they measure what they purport to measure; they do not jeopardize a crew member’s ability to perform operational duties; and they have been widely used in aviation, allowing data to be compared between different types of operations (ICAO 2012)

1.5.2.1 Sleep

Understanding the sleep/wake history is the most valuable information about the fatigue status, because sleep is one the two major factors that affects performance and other is circadian. Polysomnography is the gold standard laboratory measurement that has been used to monitor sleep for short periods of time and recently has been tried in field studies (T. Akerstedt et al., 1991) (RC. Graeber, et al., 1986) (P. Ho et al., 2005) as well, yet the effort to undertake the equipment and analytical measurements is expensive and time-consuming.

An actigraphy, unlike polysomnography is an unobtrusive wristwatch that needs to be worn on the nondominant wrist, which sums movements across specified time period governed by the administrative setting of epoch time. In the 1950s, the scientists started using mechanical sensors to evaluate psychologic disorders as many believe that’s the first medical use of actigraphy. This led to rapid development in the areas of piezoelectric sensors for enhanced accuracy, reliability and storage capacities. Many studies have used Actigraphy as a reasonable alternative to polysomnography especially to objectively monitor sleep in field studies. It has a very good reliability on total measurement of sleep time. The major drawback of this measure is it doesn’t fully identify the sleep stages (TL. Signal, et al., 2005) (KL. Stone et al., 2011) very well. The actigraphy manufactures have developed algorithms that understands the sensitivity of the accelerometer, combined with the epoch reading (TL. Signal, et al., 2005) to determine the state if the wearer is sleep or awake. The algorithms are usually validated against the polysomnography in a lab. Depending on the frequency of the epoch time interval, memory space and the battery holding time the data can be recorded for longer period of time. This is again another reason why actigraphy has gained a lot of interests in the field studies where you are collecting non-intrusive operational data for a longer period of time. In 1995, the American Academy of Sleep Medicine

(AASM) concluded that actigraphy was useful as a research tool for the study of sleep (ASDA 1995).

Volunteers are also asked to keep track of the sleep diaries along with wearing the actigraph watch. This helps to gather the subjective data perspective and are either collected in the form of paper or electronically. The manufacturers have started to add additional sensors to enhance the quality of information. One such is the inclusion of light sensors on the device. For the operators this is useful to better understand the conditions that the crew members are preparing for sleep and also aids the sleep scorers to better compare and validate against the sleep diaries. Fig.14 was generated by internally developed software program that uses the raw actigraphy data and overlapped that with activity information to give the full picture of the sleep wake history. Each row reshows the day of work beginning at midnight to midnight. Each vertical line represents an hour in home base time of the pilot. Work activity in shown in orange, Sleep is shown in blue, intensity of light exposure is shown in yellow, strength of activity is shown as vertical lines

1.5.2.2 Subjective Sleepiness and Fatigue

In the world of aviation fatigue data collection, two prominent scales validated have been recommended (ICAO 2012). The Karolinska Sleepiness Scale (KSS) is the scale from 1 to 9 where 1 = “extremely alert;” to 9 = “extremely sleepy, fighting sleep”; (Akerstedt & Gillberg 1990). This scale has been validated against EEG (M. Kaida, et al., 2006) and other behavioral variables. It is used to measure the subjective level of sleepiness at a particular time of the day. The participants of data collection scores their number that best reflects the psycho-physical state experience in the last 10 minutes. It is the measure of situational sleepiness. The use of KSS as shown in Fig 15 has been used for studies related to shift work, jet lag, attention and performance (Kecklund & Akerstedt 1993).

Samn-Parelli Scale (SP) is the scale (Samn & Parelli 1982) from 1=” fully alert, wide awake;” to 7=” completely exhausted, unable to function effectively;”

This self-assessment scale as shown in Fig 16 was originally developed in 1982 for military airlift operations and recently is popular in many aviation studies. The performance impairment starts to show on score more than 5. The participants of data collection scores their number that best reflects the psycho-physical state experience in the last 10 minutes. It is the scale for an individual's subjective fatigue measures. Although it can be easily administered, the wide range of individual variance limits its efficacy of once-off assessment tool. In the B747-400 simulator study, the crews demonstrated greater sleep loss and higher subjective fatigue ratings associated with slower decision making and tendency to choose lower-risk options (RM. Petrilli et al., 2006)

1.5.2.3 Performance

For measuring performance in field studies, a number of approaches have been developed each with its own strength and weaknesses. With the invention of smart phones, well validated laboratory tasks are now implemented in smart phones that helps administer sleep studies (G Bekenky et al., 2014) (GD. Roach et al., 2006) (DR Thorne et al., 2005). One such task is Psychomotor Vigilance Task (PVT) (Basner & Dinges 2011) (Dinges & Powell 1985) is as shown in Fig 17. This type of measurement of performance involves interrupting the normal flow of work and considered intrusive (TJ. Balkin et al., 2004). Participants have to pause what they are doing and take a test that ranges from 3 minutes to 10 minutes. The more the time the better is the accuracy of data. However, the 10-minute standard duration of PVT is regarded by many as too long for applied, operational settings. Shorter duration PVT versions (A. Samel et al., 1997) (PH. Gander et al., 2013) (PH. Gander et al., 2013) though is better from use perspective, still has challenges on being too short to detect relevant deterioration in vigilant detection. The test involves the participants to react to a stimulus that appears in the computer, tablets or the smart phones. The time from the stimuli and the response from the participant records the reaction time. The test is designed in such a way that the stimulus appears at random intervals so the participant cannot anticipate the impulse. Several lab studies have shown strong correlation between PVT performance and circadian nadir (EJ. Silva et al., 2010) (KP. Wright et al., 2002) and also PVT performance with cumulative sleep-dose-dependence (G. Belenky et al., 2003) (HPA Van Dongen et al., 2003). The form of measurement is recommended for use in FRMSs in aviation and also highly recommended by the regulators (ICAO 2012) when the operators is looking for deviation from the regulations.

1.6 Gaps and Research

In 2012, this industry leading cargo airline committed to scientifically understand its complex operations and its current mitigations and practices on what is making its crew operate safely. These mitigations include fatigue-limiting scheduling policies, world class sleep facilities at the hub, schedule improvement group to review 100% of the schedules, wakeup call program, applying science-based rostering, real-time fatigue predictions, contractual requirements of fatigue risk management, matured fatigue event review process, fatigue reporting policies.

The around-the-clock and across-time-zones nature of its flight operations exceeds the boundaries of accuracy and validity of currently available fatigue and alertness models. Most of the tools to address the needs of the cargo airline to operationally manage fatigue is not in existence. This research would focus on bridging the gap with necessary tools, models and processes. The company plans to collect (de-identified) field data on human performance and alertness as part of its commitment to fatigue risk management. These data will advance our understanding of fatigue in the field, and will be useful to educate the pilots, schedulers and management about fatigue risks. The field data will also support the continued development of modeling software encompassing the scheduling parameters of airline's air distribution network and applicable regulatory requirements. Data collection and research and development of modeling software

are critical aspects of the regulatory environment that will increasingly require airlines to develop science-based fatigue risk management systems. As outlined above, this airline has already taken several proactive steps to develop tools for generating alertness-friendly schedules that maintain operational integrity. The next major step is to further build the knowledge base that will allow for the prediction and advance mitigation of fatigue (and the resulting improvement in overall pilot alertness) across the many facets of airline's air operations world-wide.

In January 2014, the FAA introduced new duty hour regulations (Part 121, Section 117). The changes to work hour rules limit duty hours relative to duty start time and enforce a "hard stop" on crew duty length that leads to a violation when exceeded. Although the work hour rules are intended to minimize the likelihood of a crewmember working at an adverse circadian phase, the rules do not consider issues such as prior work/rest schedule, commuting, sleep in hotel rooms or daytime rest periods as confounders. Furthermore, the regulations do not account for the effects of sleep inertia during ultra-long-haul operations, where pilots are required to perform landing activities shortly after waking from in-flight rest. The change in duty hour regulations has also led to situations where crewmembers have reported experiencing an increase in fatigue. As part of this efforts and continued evolution of the program, I identified the following gaps and engaged in research to explore them for effective solutions,

A. **Streamlined data collection process that will help airlines to help collect human performance data and benchmark with other airlines**
B. **Fatigue Model Estimations and Quantification of Fatigue Risk.**
C. **Operational factors and impact on workload and crew duty delay.**
D. **Forming clusters of work patterns to help predict fatigue calls from crew.**
E. **Models to estimate Real-time fatigue detection without prior knowledge of sleep wake history.**

CHAPTER II
FIELD STUDIES, OPERATIONAL IMPACT AND ALERTNESS PREDICTIONS

Rangan, Suresh & Dongen, Hans. (2013). **Quantifying Fatigue Risk in Model-Based Fatigue Risk Management**. Aviation, Space, and Environmental Medicine. 84. 155-157. 10.3357/ASEM.3455.2013.

As primary author, I came up with identifying the current limitations on the quantitative approaches used in fatigue risk management in commercial aviation. The alternative approach of first order approximation was proposed. Dr Hans who is my mentor and consultant fatigue risk management helped me coauthor and publish this article.

US Patent #10,799,175

Title: **Research performance framework**

Grant Date: Oct 13, 2020

Inventors: Suresh Rangan (Germantown, TN)

Assignee: Federal Express Corporation

Filed: Aug 15, 2014

http://patft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=10799175.PN.&OS=PN/10799175&RS=PN/10799175

US Patent Application Number: US 2020-0290740 A1

Title: **Mitigating Operational Risk in Aircraft**

Inventors: Suresh Rangan (Germantown, TN)

Assignee: Federal Express Corporation

http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20200290740.PGNR.&OS=DN/20200290740&RS=DN/20200290740

2.1 Data Collection Process

2.1.1 Process

Collecting human performance data in the field settings can be complicated especially when comes to collecting data on irregular schedules. Fig. 18 represents the full fatigue data collection cycle. The process of human performance data collection (Rattray & Jones, 2007) (M. Rosekind et al., 2000) in the field is shown in Fig below. To answer the key research questions the operators, have to collect data on certain operational settings. The process involves the committee comprised of pilot's representatives, the pilots, company fatigue risk management team and the scientists who defines protocols and SOPs for the study.

2.1.1.1 Identification

Identification process begins with several triggers. It could be based on systemic risk that was detected by an effective crew reporting feedback system. The crew uses the company's internal reporting system to inform the management about experienced fatiguing situations. These fatigue conditions could be attributed by combination of duty patterns, a hotel room etc,. These reports are reviewed by a committee and looked from a scientific perspective. The other triggers could be new design that is planned to operate. In all the cases above, the review committee if they don't have all answers, will trigger the objective data collection process.

2.1.1.2 Preparation

After the problem is identified the operator prepares for data collection. This is an involved process. The first of the task in this process is engaging with a sleep scientist to design the protocols for the study including (Institutional Review Board) IRB approvals. Depending on the problem and the research questions asked, the study design could carry one or more data collection variables. The design also determines the number of subjects and the length and the patterns of the work schedules. After defining the protocols, the data collection steering committee solicits for volunteers and explains the nature of the study and its protocols to the potential participants. A detailed communication is established to answer important questions on the study. Pilots will have to acknowledge and confirm their participation with a written consent. The data collection committee then selects the volunteers and confirms back the selection. Before beginning the process, the team will DE identify and assign them a pseudo number for further correspondence and analysis.

2.1.1.3 Collection Process

When the volunteers are identified along with their operating schedule design, the team gears up on preparing the study packet. The packet may include fully configured actigraphs, detailed instructions of use, PVT device, sleep and other logs. The team will provide 24x7 in-study support to administer any questions that might arise from the participants. The participants once completing the study will return the packet along with the devices back to the team who is administering the study. The data collection team will document the acceptance of the study packet.

2.1.1.4 Analysis and Implementation

Each of the study packet contents are both manually and electronically transferred in to a database management system. Analytical tools and methods (Braun & Clarke 2012) are applied to answer the research questions. The results are summarized and shared with the business owners. The analysis could also be recommendations to operators, crew or both to help improve on fatigue risk. Once the solutions are implemented, the team may or may not decide on verifying if the solution is working or not.

2.1.2 Data Structure and Taxonomy

Centralized Data Management and Relationship Framework. The taxonomy (D. Grossi et al., 2005) that manages all the access, security, data ingest, visualize, analytics, reporting. Profile Data Base. Each of the profiles sets (Study, Actigraph, PVT, Subjects, Schedules) are individually managed within the system. Data Sets. Data from the subject’s common data sets include static information like airport information, time zone conversions, and other labels.

2.1.3 Database and Analysis

The creation of the fatigue management database system to help monitor and mitigate fatigue-related conditions for the pilots provided a huge leap in accelerating the conduction of the studies. The software allows the fatigue management team to visualize fatigue study information in near real time through the automation of several aspect of the data collection and analysis. The software built on top a human physiology taxonomy as shown in Fig 20 allows for the integration of a number of measurements including, the psychomotor vigilance task (PVT), actigraphy (accelerometer & ambient light), subjective sleepiness scales (Karolinska Sleepiness Scale (KSS) & Samn-Perelli (SP)), and duty schedules. In its present form, the software shows promise as an integration tool for the management and analysis of datasets, but it also has some shortcomings including: the absence of some commonly collected variables, glitches and usability errors, and design concerns. Below are few screens out of the data base management system built to analyze and visualize the data elements.

This display as shown in Fig 21 is used to examine scatter plots of an individual PVT session for a selected subject. Information on the specific actigraph watch is also displayed. The researcher can also see basic descriptive statistics of time of PVT session and how many lapses were present.

The screenshot shown in Fig 22 represents a three-process model output for a single participant created by the software and based on collected data.

This display as shown in Fig 23 represents multiple subjects wake-sleep curves overlaid with one another. This allows a researcher to visually examine general trends among individuals.

This screenshot in Fig 24 shows the variety of additional displays that are executable within a participant's dataset. The arrows denote the windows that open with a click on the buttons along the top of the screen allowing additional data points to be presented.

This display shown in Fig 25 presents the capabilities to examine scatterplots from a participant's overview screen. The boxes the arrows come from denote the number of lapses on a given PVT session, which once clicked, present a PVT scatter plot.

The display shown in Fig 26 presents sleep distributions across multiple studies. The present configuration is displaying the differences between a day and a night time schedule. The pop-out figure displays probability that an individual subject is sleeping or working

2.2 Fatigue Field Study Example - Night Hub Turn Study

2.2.1 Background

In this section, I will go through a real-world study conducted through the process per Section 5 and explain the impacts to fatigue model predictions. US passenger carriers' regulations recently changed from FAR 121 to FAR 117. FAR 117 provides a regulatory framework for the implementation of fatigue risk management systems on a foundation of iterative sleep and fatigue assessment and improvement.

NASA Ames Research Center previously gathered similar data (1987–1988). Findings were published in Gander, Gregory, Connell, Miller, Graeber & Rosekind (1996), “Crew Factors in Flight Operations VII: Psychophysiological Responses to Overnight Cargo Operations.” Much has changed in this cargo airline flight operations, and the data from the NASA study are no longer current. The present, multi-year strategic initiative plans to study all phases of the Cargo Airline system form. By gathering pilots’ sleep data, this cargo airline will be in ambitious position to understand the effects of mitigation and scientifically validate their system form through objective data. With this objective data, the company and the association will be able to accurately address their scheduling problems using fatigue forecasting software. Sleep and fatigue data can be compared from segment to segment and from trip to trip to help improve the system and support fatigue risk management.

In this section, I present how human performance data is collected, interpret the results and promote the findings into prediction modeling. The study focuses on impact of sleep, performance on 4 consecutive nights operations in two different hubs Hub 1 and Hub2 and compare and contrast the differences in work and sleep patterns. These pilots represent a smaller group to study and a more controlled system form. They operate a schedule of 4 consecutive nights mostly from/to the same city each night. They have a layover in a hotel in a specific city every day after each flight duty period. An example of this type of sequence would begin each flight duty period in Oakland, CA (OAK). Each night, for 4 consecutive nights, the pilot flies OAK–Hub1–OAK, then returns to the same hotel in the morning (except for the last flight duty period, after which the pilot goes home).

2.2.2 Study and Pilot Characteristics

Data were available for 64 pilots, which each flew hub turns on 4 consecutive nights. The total data set consisted of 255 hub turn duty. Of the 255 hub turns, 187 were to Hub1, the other 68 were to Hub2. Pilots were scheduled to travel to and from the city where their week of work started by a commercial flight. For the majority of the pilots flying to Hub1, their residence was in the Eastern Time zone (ET), whereas for most of the pilots flying to Hub2, their residence was in either the Central or Pacific time zones (CT or PT). In this report, all times of day are expressed in base time (CT) unless otherwise specified.

Duty start time was 22:01 base time on average for hub turns to Hub1, and 22:27 base time on average for hub turns to Hub2. Block-in at the hub occurred at 23:38 base time on average in Hub1, and at 00:52 base time on average in Hub2. As such, time spent at the hub was longer in Hub1 (3.83 hours on average) than in Hub2 (2.47 hours on average). Duty end time was 04:57 base time on average after returning from Hub1, and 05:59 base time on average after returning from Hub2.

2.2.3 Hub Napping

Out of the 255 hub turns, 176 had nap sleep at the hub as determined by wrist actigraphy. On average, pilots had hub naps 69.1% of the time, with a standard deviation over pilots of 40.3%. Pilots were relatively consistent in whether they napped or not ($ICC=0.75$, $Z=5.18$, $p<0.001$). Napping occurred more frequently in Hub1 (78.6% on average) than in Hub2 (42.6% on average). In those cases when a hub nap was taken, the naps were significantly longer ($F_{1,127}=5.5$, $p=0.020$) in Hub1 (1.46 hours on average) than in Hub2 (0.93 hours on average). In terms of individual differences, pilots were relatively stable in their nap durations ($ICC=0.72$, $Z=4.33$, $p<0.001$).

The difference in nap duration between Hub1 and Huub2 is not surprising, because time spent at the hub was 1.36 hours longer in Hub1 than in Hub2. Indeed, hub time was a significant predictor of hub nap duration ($F_{1,190}=76.4$, $p<0.001$). After the first 0.71 hours of hub time (which was the minimum duration, on average, to have any sleep), hub nap duration increased by an average of 0.48

hours for every hour of hub time (i.e., roughly half of each hour of extra hub time was spent sleeping, on average). Regardless of whether the hub turn was at Hub1 or at Hub2, pilots with residence in ET or PT napped more frequently than pilots with residence in CT or MT ($F_{3,191}$=3.0, p=0.031). The nap duration of pilots with residence in ET or PT was also slightly longer than that of pilots with residence in CT or MT, but the difference was not statistically significant ($F_{3,191}$=2.3, p=0.080).

There was a trend for small but systematic changes in nap duration from the first to the fourth duty night of the hub turn sequence ($F_{3,185}$=2.4, p=0.070). The second night showed the longest nap duration and the last night showed the shortest nap duration, regardless of whether the hub turns were at Hub1 or at Hub2. The average difference between the second and last duty nights was 0.2 hours. Whether pilots napped in a hub room (as assessed by self-report) or in a hotel also made a difference for nap duration, with napping in a hub room adding 0.46 hours of sleep time on average, although the difference with sleeping in a hotel did not reach statistical significance ($F_{1,124}$=3.1, p=0.079). When hub sleep occurred, it took longer from block-in time to the start of sleep during the hub turn in Hub1 (0.91 hours on average) than in Hub2 (0.51 hours on average) – the difference was significant ($F_{1,127}$=7.3, p=0.008). Some of this difference may be attributed to transportation time to a hotel from the Hub1 hub.

The average difference in nap duration between Hub1 and Hub2 is not surprising, because average time spent at the hub was 1.36 hours longer in Hub1 than in Hub2. Overall, hub time was a significant predictor of hub nap duration (F1,190=76.4, p<0.001), explaining 35.1% of the variance. See Fig. 27. After the first 0.71 hours of hub time (which was the minimum duration, on average, to have any sleep), hub nap duration increased by an average of 0.48 hours for every hour of hub time (i.e., roughly half of each hour of extra hub time was spent sleeping, on average).

2.2.4 Further Analysis

A commonly expressed concern is whether sleep at the hub may interfere with sleep after the end of the duty period. There is some evidence to corroborate that concern in the present data set, in that increased hub nap duration significantly predicts reduced post-duty sleep duration ($F_{1,190}$=7.4, p=0.007). However, for each hour of additional hub nap duration, only 0.43 hours of post-duty sleep is lost. Thus, in general, hub napping still constitutes a net gain in sleep obtained. There was no evidence in the present data for the opposite relationship. That is, across each of the duty nights and all of the subjects, sleep in the 24 hours before duty did not significantly affect hub nap duration ($F_{1,190}$=2.14, p=0.15) – see Fig 28. This finding held up when only duty nights with more than zero minutes of hub nap duration were included ($F_{1,126}$=1.27, p=0.26).

It is important to note that the present results for sleep in the 24 hours prior to duty and sleep after duty are not independent of each other – in consecutive duty days, post-duty sleep is also included here in sleep during the 24 hours prior to the start of the next duty period. Subsequent analyses considering whole 4-day pairings at once are needed to tease this apart. These analyses will be pursued in the near future. Finally, to what extent the sleep findings discussed here are relevant for fatigue levels during the hub turns to Hub1 and Hub2 could be explored through biomathematical fatigue modeling (which will be pursued in the foreseeable future) and/or through fatigue measurement during a subsequent study.

2.3 Fatigue Risk Quantification

2.3.1 Shortcomings

The question of what is a maximally acceptable level of fatigue risk is hotly debated in model-based fatigue risk management in commercial aviation and other transportation modes. A quantitative approach to addressing this issue, referred to by the Federal Aviation Administration with regard to its final rule for commercial aviation “Flight crew Member Duty and Rest Requirements,” is to compare predictions from a mathematical fatigue model against a fatigue threshold. While this accounts for duty time spent at elevated fatigue risk (Goode 2003), it does not account for the degree of fatigue risk and may, therefore, result in misleading schedule assessments. I propose an alternative approach based on the first order approximation that fatigue risk is proportional to both the duty time spent below the fatigue threshold and the distance of the fatigue predictions to the threshold — that is, the area under the curve (AUC). The AUC approach is straightforward to implement for schedule assessments in commercial aviation and also provides a useful fatigue metric for evaluating thousands of scheduling options in industrial schedule optimization tools.

2.3.2 Quantifying Fatigue Risk in Model-Based Fatigue Risk Management

RECENT CHANGES in U.S. hours of service regulations across several modes of transportation have brought to the foreground the question of what is a maximally acceptable level of fatigue risk (P. Gander et al., 2011). Advances in mathematical modeling of fatigue have facilitated systematic investigation of this issue in the context of fatigue risk management (SR. Hursh et al., 2010). One way to approach the issue is by comparing model predictions of fatigue against a fatigue threshold to distinguish acceptable from unacceptable (overly fatigue-risky) duty schedules (SR. Hursh et al., 2011) (see Fig. 29).

Comparison of two sleep/wake/duty schedules based on two different fatigue thresholds. The abscissa indicates cumulative clock time (0, 24, and 48 denote midnight). Schedule 1 is shown on the bottom (black) and schedule 2 is shown on the top (gray), where solid bars indicate sleep periods and hatched bars indicate duty periods. The curves show fatigue predictions (expressed as effectiveness percentage) made with the SAFTE model for schedule 1 (thin black curve) and for schedule 2 (thick gray curve). The dashed lines show two different fatigue thresholds, at 77% and 80% effectiveness, which could be used to evaluate the fatigue risk associated with the duty period on the second day in the two schedules. With the 80% effectiveness threshold, in both schedules the fatigue prediction curve falls below the threshold, indicating unacceptable levels of fatigue risk; whereas with the 77% effectiveness threshold, only schedule 1 would be considered unacceptably fatigue-risky. Regardless of which threshold is used, though, schedule 1 is predicted to have the most duty time with unacceptably high fatigue risk. Comparing schedule 1 to schedule 2, this is a consequence of the early awakening to begin an early duty period and the early afternoon placement of the nap on the first day in schedule 1.

In 2011, the Federal Aviation Administration referred to this strategy in its discussion of public comments and final rule for commercial aviation “Flight crew Member Duty and Rest Requirements ” (Federal Aviation Administration 2011). A widely used mathematical model of fatigue called the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE) model (SR. Hursh et al., 2004) was used. This model predicts fatigue based on a homeostatic process tracking sleep/wake history and a circadian process tracking time of day; the model is distinct from most other mathematical models of fatigue in how well it captures the fatiguing effects of chronic sleep restriction and shifted sleep times. SAFTE model predictions are conventionally expressed on a “% effectiveness” scale, and a fatigue threshold of 77% effectiveness has been used to distinguish acceptable from unacceptable duty schedules in commercial aviation. The fatigue threshold approach has been criticized because it does not account quantitatively for the degree of fatigue risk associated with a given schedule (HPA Van Dongen et al., 2012).

Whether or not a schedule is deemed to be associated with an unacceptable level of fatigue depends on where the line is drawn, that is, the selected threshold level to which fatigue predictions are compared (see Fig. 29). This issue has led to debates in the literature, for example about whether or not a threshold level established for one mode of transportation is also applicable for another (SR. Hursh et al., 2010). Even though fatigue thresholds have found their way into real-world operations (D. Dawson et al., 2011), these debates are ongoing.

The issue of where to draw the line in an absolute sense is to some extent irrelevant in the context of schedule optimization, which intrinsically relies on relative comparisons between different schedules in terms of fatigue risk and other operationally relevant factors (HPA Van Dongen et al., 2012). This can be seen in Fig. 29, where regardless of which of two different threshold levels is used, the same schedule would be selected as the one predicted to be the least fatigue-risky. Yet, there is ambiguity in this approach to using fatigue thresholds for comparing schedules. This problem can arise when the most relevant period of a schedule the duty period (or a safety-critical portion thereof) begins or ends with a predicted percent effectiveness that is below the threshold. This is illustrated in Fig. 30, where one schedule involves more duty time spent subthreshold, whereas another schedule involves less duty time spent subthreshold, but with a greater drop of percent effectiveness relative to the threshold. It is a priori unclear which of the two schedules involves the highest fatigue risk while on duty.

We propose an alternative strategy for using fatigue thresholds which helps to overcome this ambiguity. A key ingredient in this approach is the realization that fatigue risk (i.e., risk of fatigue-induced errors, incidents, and accidents) is increased not only by more duty time spent below the threshold (regardless of the threshold level), but also by the degree of fatigue at those times. A reasonable linear approximation of this is that fatigue risk (Goode 2003) is proportional to both the duty time spent below the threshold and the distance of the fatigue predictions to the threshold. In other words, fatigue risk is proportional to the integrated area under the curve (AUC) across the duty period. Thus, we recommend that threshold-based fatigue evaluations of work schedules employ AUC as the primary metric of fatigue risk (see Fig. 31 for an illustration). The AUC approach connects readily with the neurobiology of sleep and fatigue as instantiated in the seminal two-process model of sleep regulation (AA. Borbély et al., 1982) (S. Daan et al., 1984). In this model, a homeostatic process S builds up a pressure for sleep across time awake (and dissipates this pressure during sleep). When the pressure for sleep exceeds a threshold H Hm C, where Hm is a constant and C is a circadian process modulating the threshold level over time of day, sleep is predicted to occur naturally. Staying awake causes, a state of sleep deprivation, yielding fatigue. In the two-process model, fatigue risk can be seen as proportional to both the time during which process S is above the threshold H and how far process S is above the threshold H at each time point. Thus, fatigue risk is proportional to the integrated area above the time varying threshold (see Fig. 32).

Ambiguity in the comparison of duty periods based on fatigue thresholds. Plot elements are the same as in Fig. 29. Three days of two different schedules are shown: schedule 1 is shown on the bottom and in black; schedule 2 is shown on the top and in gray. Regardless of which of the two thresholds drawn is used, the last duty period in schedule 2 is predicted to have the most duty time with unacceptably high fatigue risk. However, due to different sleep/wake history, the last period in schedule 1 is predicted to involve greater reduction of percent effectiveness. This can make the selection of a schedule based solely on duty time spent below the fatigue threshold suboptimal in terms of mitigating fatigue risk.

Magnification showing the AUC below the 77% effectiveness threshold for the two duty periods being compared in Fig. 37. Schedule 1 is shown in near-black and schedule 2 is shown in light gray; overlap is shown in dark gray. (Sleep periods are not shown.) Even though the duty period in schedule 2 clearly involves more duty time spent below the fatigue threshold, the AUC for schedule 1 is estimated to be 1.62 times greater than that for schedule 2. This suggests that schedule 1, not schedule 2, and is associated with the greater fatigue risk.

Equivalent of Fig. 31 in terms of the two-process model of sleep regulation applied to schedule 1. A) Process S is shown in black (thick curve rising during wakefulness and falling during sleep); the threshold H (modulated by the circadian process C) is shown in gray (thin curve). The near-black area shows the integrated area during the duty period when process S is above the threshold H. B) By subtracting the circadian process from the homeostatic process and inverting the scale (thick black curve), the two-process model view reduces to the AUC approach proposed here (with a threshold level of zero, shown in light gray). The near-black area is the neurobiological equivalent of the AUC for fatigue in Fig. 31.

The AUC approach we propose here is no more difficult to implement than the threshold approach currently in use in various operational settings and could thus easily replace the latter. The AUC approach also provides a computationally feasible means of evaluating thousands of possible scheduling options, as will one day be needed to encompass fatigue risk management in industrial schedule optimization algorithms (E. Romig et al., 2009) (HPA Van Dongen et al., 2012).

2.3.3 Fatigue Risk Estimation and Relationship to Errors and Flight Safety

Fatigue model predicts alertness (E. Romig et al., 2009), however these models also started proposing thresholds for the operations to define acceptable and unacceptable schedules. This type of threshold approach has been criticized because it does not account quantitatively for the degree of fatigue risk. Different tasks require a different minimum level of alertness. As illustrated in Fig. 33, the required alertness level (the red line) changes over the duty period. Takeoff and landing have more duty, involve higher risk, and therefore require higher alertness level. Fatigue-related alertness level goes down over time. The shadow area measures the fatigue-related risk, which is more accurate compared to the single value of minimum level of alertness. The model will further be used to improve duty planning and mitigate overall risk. Extending further the AUC approach (S. Rangan et al., 2013) to risk assessment, we intend to develop a comprehensive risk model (S. Rangan Patent 9540118) for all fatigue factors defining logical relationship between the risks. We then would want to develop a dynamic way factor based on the risk on task to define threshold values on the alertness curve.

In the literature of Crew Resource Management (CRM), crew training plays a critical role of mitigating risks caused by threats and errors. Here, threats are external conditions that have the potential to hurt the safety of a flight and include both expected threats and unexpected threats. The CRM literature focuses on how to train pilots to recognize threats. However, a proactive risk management may purse mitigate threats, especially expected ones, through systematic and coordinated planning and adjustment. For example, weather, airport condition, and pilot's familiarity on the airport are often considered common threats. A systematic way may not assign a pilot to an unfamiliar airport when there is a predicted hazardous weather. Errors are defined as crew action or inaction that leads to a deviation from crew or organizational intentions or expectations, including intentional noncompliance errors, procedural errors, communication errors, proficiency errors, and operational decision errors. Most researchers argue that errors are unavoidable and the focus should be on error management. An accident is often caused by a series of errors and CRM training is important for crew to better manage threats and errors. However, we argue that fatigue risk, measured in Fig. 34, could be a major contributor to the happening of initial errors, cause additional errors when crews detect and deal with errors, and exacerbate the error cycles to increase the chance of final accidents. Of course, threats may interact with fatigue risks to cause more risks and finally hurt the overall flight risk. Therefore, we will study the fatigue risk's relationship with errors under the interaction with threats and final relationship with flight safety. The Operations Quality Assurance data and sleep data that have collected will be used to study the relationship that connects the area between the required alertness levels and fatigue-related alert levels and the counts of errors over time.

Biomathematical models produces alertness, fatigue, lapses curve that is two dimensional. At time of the day in x axis, (alertness, fatigue and lapses) are all plotted on the y axis. Operations run the schedules through the model and draw a arbitrary thresholds to flag a particular schedule as fatiguing or non-fatiguing. The assumptions of the model on the homebased time of the individuals flying the schedule, age and other factors are kept static. So, quantification of risk is a big discussion item as the only way is to find the schedule above a threshold. As I discuss in my paper on Quantifying fatigue risk in model-based fatigue risk management. I would like to consider adding more dimensions to the quantification. In addition to the length, the depth of exposure with the integral based AUC approach will be considered. Along with the depth and length, I would like to introduce the risk per minute. This will modify the thresholds to not be a straight line but a curve by itself based on all the factors.

2.4 Crew Duty Scheduling

2.4.1 Crew Planning

Airline Crew scheduling is process where the airlines have to build monthly assignments for the crews which intends to minimize the cost for the operations. It is one of the most challenging planning problems faced by airlines. Although these problems are closely interrelated, they are typically solved sequentially, due to their size and complexity. Airlines usually begin by solving a schedule design problem, in which they determine the flights to be flown during a given time period. In the next step, the fleet assignment problem (Li & Tan 2013), they decide what type of aircraft (such as Boeing 767, 727, etc.) to assign to each flight, as a function of the forecasted demand for that flight.

Once the maintenance planning activities are schedules, the flight schedules are up for generating the crew pairing and rostering which is a twostep sequential process that is the most challenging and compute intensive tasks in the industry. For large fleet sizes where the complexity increases, the planners with the use of optimization software can take multiple days to produce the pairing (sequence of segments as referred below). The runs are done on a monthly basis. The number of pairings is one the target that will be based on the number of available crews and the demand for flight hours for the airline for a given month. Crew rostering is an assignment phase where pairings and other activities like training are assigned to crew members according to their qualification, vacation days and other parameters as defined by their contractual agreement. As part of the rostering process, there is a bidding process that allows the crew to take some control over their allocations. Different airlines around the world does this process differently. For example, majority of the North American carriers bid for predefined anonymous rosters where the senior members have priority over their other junior colleagues. The last step in the process is the day of operational recovery process. This is where the crew schedulers take control on resolving all the damages caused to the plan either due to unavailability of crew, cancellation of flights or pick additional demands. The schedulers with the assistance of the day of operations crew tracking system evaluates combinations to determine the best crew on time.

- Crew is referred as pilot and copilots needed to operate to commercial flight.
- Flight is referred as the flight from a city also known as departure station to another city also called as arrival station at a given time.
- Duty is referred as a period which the crew flies a set of segments without taking a legal rest break.
- A trip is referred as a sequence of segments flown by the crew which typically originates and terminates at the base at which the crew is assigned to. On some cases they don't have to start and stop the trip at base and some airlines also allows to fly their trip through the crew members base.

2.4.2 Schedule DNA

Pilot schedules are governed primarily by the regulatory framework of the respective country and also the negotiated agreements between the company and the pilots the latter being the most restrictive. Schedule also called as pairing consist of a sequence of duty periods where a duty is defined as the set of tasks to be performed by a crew member during a given day. A pairing is a sequence of connectable flight legs, within the same fleet, that starts from and ends at the same crew base, where the crew actually lives. A pairing is sometimes called an itinerary for the crew assigned to this journey. It typically spans from one to five days where some of them can go up to 15 days. Fig. 36 shows, a legal pairing is an alternating sequence of events that are mainly divided into duties and rests. Duties are in orange blocks and rest is the space between the duties. The grey blocks are predicted sleep during the layovers. Sometimes crew members mitigate fatigue by taking nap in sleep room between flights in a duty period. Each duty period represents a daily task segment that can be subdivided into flight leg segments separated by sit connections. Hence duties must be formed before pairings. A duty can become a pairing if it starts from and ends at the same city. In United States, the length of a duty is largely determined by Federal Aviation Regulations (FAR). However, the length can vary from airline to airline. The Federal law only requires that a pilot cannot fly more than eight hours within a 24-hour period, and he/she must also be able to rest for eight hours in the same time span. A sit connection during a duty period mainly consists of the waiting time of the crew for changing planes on to the next flight leg. Finally, an overnight rest (or layover) is a rest period between two consecutive duties. If the overnight rest happens away from the home base of the crew, the airline must pay for their hotel expenses and additional compensations.

2.4.2.1 Duty and Schedule DNA

Flight duty period**.** The allowable length of a flight duty period depends on when the pilot's day begins and the number of flight segments he or she is expected to fly, and ranges from 9-14 hours for two pilot operations. The flight duty period begins when a flight crew member is required to report for duty with the intention of conducting a flight and ends when the aircraft is parked after the last flight. It includes the period of time before a flight or between flights that a pilot is working without an intervening rest period. Flight duty includes deadhead transportation, training in an aircraft or flight simulator, and airport standby or reserve duty if these tasks occur before a flight or between flights without an intervening required rest period. The pairings and duties are subjected to the contractual and state regulations and some of the examples are listed below

i. The total flying time within a duty cannot exceed an upper bound. There is also an upper bound on the total elapsed time within a duty.
ii. There is a lower bound on the sit time which guarantees that the crew has enough time to connect between two consecutive flights within a duty.
iii. The rest time between duties should be greater than or equal to a minimum rest time which ensures that the crew is sufficiently rested between duties.
iv. There is typically an upper limit on the number of duties within a pairing.

A typical duty in pilot schedule has a lot of variables such as duty start time, duty end time, length of the duty, number of flights in a duty, flight time vs total time, risk associated with duty, mitigations and countermeasures, number of time zones transitions etc., I have identified at least 10 variables that can possibly affect sleep and performance some being more prominent over others.

2.4.3 Structure of DNA

Generic DNA coding is being proposed to better cluster the individual schedules of work and sleep. Each of the attribute is coded into a two-digit alpha numeric code. This crew schedule in Fig. 37 has 4 duty periods. First duty from Memphis (MEM) to Anchorage (ANC), second duty from Anchorage to Osaka (KIX), third duty from Osaka to Pudong (PVG) and fourth duty from Pudong to Memphis.
Extracting all the raw data from the schedule, see table 1 with all the attributes of the duty

2.4.4 Coding DNA

I am proposing the coding of the Duty DNA to be based on a threshold values of each of the attributes of the Duty.

For example in Fig. 38, P_L/O for last duty of 19hours and 23 minutes is converted into 2 digit alpha numeric code of LA. 00=start of duty, LS=short layovers upto 14 hours, LA=adjusted layovers up to 20hours, LW=swap layovers up to 28hours, LR=reset layovers up to 36 hours, LX=extra-long layovers above 36 hours. Detailed DNA Code properties for all attributes are available in the dna.properties file. Similarly, two-digit alpha numeric codes are provided for duty start, duty length, duty end, landings, mitigations within duty, total flying time, time zone from base, time zone differential within duty, day in the sequence, base and aircraft.

2.4.5 Sleep DNA Variables influencing the predictions

As seen in Fig. 39, SleepDNAHead, contains all the historical information of the current predictor of sleep. SleepDNABody, contains all the current sleep opportunity attributes that influences the predictions of sleep. SleepDNATail, contains the future schedule opportunity that could influences the predictions of sleep. (eg, the restricted and/or higher risk opportunity. As need basis we could also include other environmental conditions such as global economic, geo political, weather, global events can contribute to the sleep patterns. (Super bowl can affect the sleep times). In Fig. 40, the sleep during Layover1 will be dependent on the sleep predictions of actual/predicted sleep in layover 0 and predicted sleep in layover 2. The layover1 sleep is also based on the predicted sleep in duty 1 and duty 2. Mid duty sleeps are typically mitigation sleep where crew could sleep in flight during long international flights or in between two flights which is part of the same duty. Excess sleep in Layover 0 and Duty 1 can affect the distribution of sleep in Layover 1. Also, there could be situations where the increased risks and reduced sleep opportunities of future duties (Duty 2 and Layover 2) could cause the individuals to bank sleep during layover 1.

2.4.6 Sleep Pressure Prediction based on Pattern of Duty Periods

The sleep pressure level at the end of a duty period could be well predicted through the exiting process S function based on the sleep pressure level at the beginning of the duty period (ie. work schedule of the duty). The sleep pressure S_i at the beginning of a duty period (duty 2) depends on the sleep pressure $S_{(i-1)}$ at the end of the previous duty period (duty 1) and the sleep periods between the two duties at illustrated in Fig. 41. Please note that the alertness level is negatively correlated to the sleep pressure but also depends on the circadian body clock cycle. However, intra and inter-individual variability of sleep patterns between the two consecutive duty periods may exist within the pilots. Depending on the variables associated with the pervious duty, including all DNA information discussed in the previous subsection, and the following duty, pilots may choose different sleep patterns. Numerous data regarding sleep pattern data have been collected from flights. In this task, a prediction model will be built and trained with the historical sleep pattern data. The input of the prediction model will be the DNAs of the previous duty and the following duty of a layover. The output of the model will be the predicted sleep pressure level at the beginning of the following duty, which is equivalent to the sleep pressure level at the end of the following duty under the assume sleep pressure level evolvement function and the DNA of the following duty. The hypothesis for this prediction is that the sleep patterns during the layover period depends on the duties immediately before and after it. The training data include the following historical data

- Previous duty DNA,
- Following duty DNA, and
- Sleep pattern during the layover period.

For each historical layover period, the sleep pressure level at the beginning of the following duty will be calculated based on the pervious duty DNA and then used for the training purpose. Statistical methods will be first applied to identify the major factors and the regression relationship (linear or nonlinear). If the results are not satisfactory, machine learning methods, such as the neural network, will be investigated. Due the large amount of available data, we expect that a rather accurate prediction model will be possible.

2.4.7 Clustering of Duty Periods

2.4.7.1 Background

In this research, I am attempting to understand the sleep structure better and how biology is interacting with the work schedules. Eventually this sleep predictions are applied to the biomathematical model and predict fatigue risks. These risks are further applied to the prediction of overall flight risk

2.4.7.2 Duty and Sleep Plots

Fig. 42 represents how the observed sleep patterns on a work schedule. The schedule of the DNA has variables that represents the structure of the DNA. These variables include but not limited to duty start time, end time, length of duty, number of flights etc., aligning all the schedules on the same scale ignoring the day start of the sequence and aligning them as if they all start on the same day but not the same time. The duties can be represented better visually to understand if there is any homogeneity within the sub group. These characteristics include how the crew members would act within a cluster. (Sleep determination within the subgroup).

As a method to group the data to distinct subgroups, we can employ multiple algorithms. For the literature review, I am looking at some of the below mentioned clustering algorithms.

2.4.7.3 Partition based clustering selection

This is one of the most popular class of clustering algorithms. Partition based clustering (RO Duda et al., 2001) (J. Han et al., 2001) iteratively relocates data points between cluster until an optimal partition is attained. One of the disadvantages of this algorithm is, it only converges local and the global optimal solution is not guaranteed. In our problem we have to answer how to classify the variables that differs by minute. In the aviation scheduling practices, the lowest denominator captured and analyzed in minute. The times like arrival, scheduled layover, duty time all is marked by the minute. One way to approach is to bin based on fixed threshold-based classification of splitting the arrival times into six, 4-hour periods. [0000-0359, 0400-0759, 0800-1159, 1200-1559. 1600-1959, 2000-2359]. Currently in my work so far, I am using discrete classification of these identifiers to structure the dutyDNA. However, to avoid the loss of precision, In the future research I am planning to use suitable partition-based clustering methods to construct and populate the bins. This will find natural clusters instead of artificial cluster based on discrete threshold ranges. Loss of precision could lead to incorrect predictions

2.4.7.4 Time Series Analysis

Time series is a sequence of data points, measured typically at successive time points. This has been the focus in the data mining community for long period time (Roddick & Spiliopoulou 2002) It’s a collection of observations Xt, each one being recorded at time t. (Time t can either be discrete or continuous t>0). Time series data (Cowpertwait & Metcalfe 2009) is in a series of particular time periods or intervals. Time series analysis could be used for many applications in the areas of data compression, explanatory, signal processing, predictions etc, at the simplest it’s a set of time-ordered observations where interval between the observations remain constant.

2.4.7.5 Component of Time Series

m(t) – trend component (R. McCleary et al., 1980) (Hyndman & Athanasopoulos 2014) that changing in time. The overall movement and general direction of the data. Both the direction and slope (rate of change) of a trend may remain constant or change throughout the course of the series.

c(t) – cyclical component (R. McCleary et al., 1980) (Hyndman & Athanasopoulos 2014) that is typically available in data (patterns of oscillations) with long historical data. A cyclical component in a time series is conceptually similar to a seasonal component: It is a pattern of fluctuation (i.e., increase or decrease) that reoccurs across periods of time. It takes many years to play out.

s(t) – seasonal component (R. McCleary et al., 1980) (Hyndman & Athanasopoulos 2014) (Bell & Hillmer 1984) for known periods (either minute, hourly, daily, monthly, yearly). Unlike the trend component, the seasonal component of a series is a repeating pattern of increase and decrease in the series that occurs consistently throughout its duration. Although its underlying pattern remains fixed, the magnitude of a seasonal effect may vary across periods.

Y(t) – random noise that comes with every data. Component such as unknown frequency that is uncontrolled. While the previous three components represented three systematic types of time series variability, the irregular component represents statistical noise and is analogous to the error terms included in various types of statistical models.

When the magnitude of the trend-cycle and seasonal components remain constant, we will use additive decomposition model as follows

$$Xt = m(t) + c(t) + s(t) + Y(t)$$

When the magnitude of the trend-cycle and seasonal component varies but still appears proportional over time, the series is better represented by multiplicative decomposition model.

$$Xt = m(t) * c(t) * s(t) * Y(t)$$

2.4.8 Clustering Time Series for Duty Sequences

This is one of the most frequently used analysis of time series is time series clustering (M. Halkidi, et al., 2000). The clustering is broadly classified into two categories. Whole clustering is very similar to conventional clustering of discrete data objects. The time series data is grouped similar into same cluster. Subsequence clustering (E. Keogh, et al., 2003) where a single time series as extracted with sliding window and then clustering is performed within the sliding window (Golding & Kanellakis 1995). This is commonly used in rule discovery (P. Das, et al., 1993), classification (P. Cotofrei et al., 2002), prediction (C. Schittenkopf, et al., 2000) and anomaly detection (P. Tino et al., 2000).

2.4.8.1 Subsequence clustering

With duty sequence, alertness curves are generated using fatigue model equations. To demonstrate the subsequence clustering, 10 alertness curves are generated for different work sequences (see fig. 43). The orange line indicates work patterns and purple lines indicate sleep patterns. If we try to cluster these time series, subsequence clustering is one method that could be used. In a given time series T of length m, a subsequence Cp of T is a sampling of length w<m of contiguous positions from T, that is C=tp….tp+w-1 for l<= p <= m-w+1. We will use a concept of sliding window to achieve this.

Sliding windows of length w that are user defined sequences that are extracted from the time series T of length m. A matrix S of all possible subsequences can be built across T and placing subsequence Cp in the pth row of S. the size of the matrix S is (m-w+1) by w. If all the 10 different alertness curves are overlapped one each other with the same starting time. There are different types of w that can be created. w fixed to 24hours (see fig. 44 window) and it will be (duty_end_time-16hours) to (duty_end_time+8h). This gives us opportunities to understand sleep preparation for duty and also recovery from the duty. Duty length is that is where the most difficult point in the duty with respect to workload and fatigue. w can also be created with 24 hours before and after the duty_end_time that will keep the cluster center to the point to measure. w can be of varying length. In the example illustration we have created 6 different subsequences extracted by a sliding window. In our example similar steps will be iterated for all the 10 examples. A total of 52 subsequences if we use the sliding window as mentioned in (Fig 44a).

2.4.9 K-Means

2.4.9.1 Definition

This is one of the common algorithms used to partition a dataset into pre-defined subgroups or clusters. These subgroups are typically non overlapping (ie in our paper the data point or the duty can only be within one subgroup or cluster). The goal of k-means (M. Haikidi, et al., 2000) will to group the duties as similar as possible but also create a farther distance between the next clusters of duties (Bradley & Fayyad 1998). Centroids are formed for each of the cluster and the sum of the squared distance between the data points and the subgroup (cluster's) centroid is kept at minimum. In a 2-dimensional space, if x axis are the duty period start time and y axis is the length of the duty period. Then we should see a scatter plot of all the duties as clusters. K Means will help resolve the different clusters.

2.4.9.2 Disadvantages

There are several disadvantages. One of the main one is that the number of clusters needs to be pre-defined. It assumes spherical shapes of clusters. Due to centroiding, the radius is assumed to be equal distance between the centroid and the furthest data point. Good example where this won't work is if the clusters are of different shapes like elliptical.

2.4.9.3 Algorithm

- Decide on a value for k.
- Initialize the k cluster centers (randomly, if necessary).
- Decide the class memberships of the N objects by assigning them to the nearest cluster center.
- Re-estimate the k cluster centers, by assuming the memberships found above are correct.
- If none of the N objects changed m

2.4.9.4 Results

Java program is written to parse 365 dutyDNA with three attributes (previous layover, duty start time and duty length) into 5 clusters as shown in Fig 45. It's written for three-dimension clusters. The data is attached in the excel sheet. See attached the java programs.

2.4.10 Hierarchical Clustering

2.4.10.1 Definition

This is the most widely used clustering approaches and that is due to the visualization effects (Mantegna 1999). It groups the objects and produces a nested hierarchy according to the pairwise distance matrix. One big advantage compared to k-means is it generalizes and the user doesn't have to provide the number of clusters. The disadvantages to this would be the size of data set due to quadratic computational complexity. However, when the dimensions are higher the clustering becomes meaningless because of the nearest neighbor and average neighbor becomes one. (R. Agrawal, et al., 1993)

2.4.10.2 Algorithm

- Calculate the distance between all objects. Store the results in a distance matrix.

- Search through the distance matrix and find the two most similar clusters/objects.
- Join the two clusters/objects to produce a cluster that now has at least 2 objects.
- Update the matrix by calculating the distances between this new cluster and all other clusters.
- Repeat step 2 until all cases are in one cluster

In this example we are illustrating with 10 random clusters picked from the total of 52 subsequences if length w generated from the cluster of timeseries T. Fig. 46 shows how the time series are converted to k partitional clustering by sliding and the dendrograms. In this figure it shows we will have three types of clusters of subsequences. With the same timeseries clustering, we can also detect anomalies. The best example would be. Once we defined a cluster to exhibit certain sleep patterns based on observed training data. In the operations if the crew calls in and if he mentioned that he had a different sleep pattern then we can consider that as an anomaly and manage that risk in operations.

2.4.11 Clustering DNA String

2.4.11.1 Problem

Creating the discrete range/bin I thought is the best way to analyze the data for predictive modeling. However, I do understand the loss of precision in creating random size bins with random thresholds. Collecting this continuous data by categories is also not a smart way and is well explained by (Good & Hardin 2006) That’s why I am being watchful in making sure we employ proper clustering methods (MR Anderberg 1973) to categorize the bins. (Wainer et al., 2006) argue that if a large enough sample is drawn from two uncorrelated variables, it is possible to group the variables one way so that the binned means show an increasing trend, and another way so that they show a decreasing trend. They conclude that if the original data are available, one should look at the scatterplot rather than at binned data.

2.4.11.2 Solutions considered

Using partition-based clustering principles to find a natural cluster of data points is a better option instead of an artificial threshold-based on range of values. Based on scatter plots, we came up with the right thresholds to define the boundaries rather than fixed bins. The same process is used for all the 25 attributes. See DNACode.properties file for the derived bins based on scatter plots and clusters. Once the clusters are defined for each of the attribute, the full DNA code is generated for a schedule depending on the cluster each of the attribute belongs. Full string DNA coding is done for each of the duty DNA, iterate and perform the same for all the duties. Sample DNA code will be as below. The challenge was to cluster these string duty DNA sequences. For that we need to find similarity between the sequences. One way to find the similarity of the DNA is to find the distance between the DNA string, see below the java program that ran through 15000 duty DNA samples and found the closest and farthest string. Similar approaches could be made to cluster the DNA sequences.

For clustering string fields, there are two well-known methods

a) Hamming Distance

   The Hamming distance H (Hamming 1950) is defined only for strings of the same length. For two strings s and t, H (s, t) is the number of places in which the two string differ, i.e., have different characters.

b) Levenshtein Distance

The Levenshtein (or *edit*) distance (Levenshtein 1966) is more sophisticated. It's defined for strings of arbitrary length. It counts the differences between two strings, where we would count a difference not only when strings have different characters but also when one has a character whereas the other does not.

For this exercise, I am using Hamming Distance where two equal sized strings are compared and it tells all the bits that's compared different. Hamming distance is much, much faster than Levenshtein as a distance metric for sequences of longer length. Hamming distance can be considered in our program for comparing the two DNA sequences for a order-biased similarity metric rather than the absolute minimal number of moves to match the sequences. The program below as shown in figure 55 also finds the farthest and nearest point in the list of over 15550 dutyDNASamples.

2.4.12 Feature Selection Process

Isolating the variables that would improve the predictions is a very important process in machine learning. There are several methods and algorithms that were explored. The key is to select the right method that will fit to solve the problem. It is clear, that we have a multivariate dataset of many variables (LM Bermingham et al., 2015). To understand the relationship between variables, would like to bring a scatter plot of two variables to the discussion here. We are looking at understanding the relation between length of Layover1 and length of Duty2, the sample scatter plot between the two is shown in Fig. 49 below. It is visually noticeable that this dataset is cluster able. You could see there are different types of layovers and duty length combination that exist in the system. This example is taken from a subset of the system that the company operates. Similar patterns will exist in other forms of operations. To determine the number of clusters for each relation, a heuristic Elbow method (Thorndike 1953) was used (Fig. 50 and 51). It plots the explained variation as a function of the number of clusters. In this example, we demonstrate the use of elbow method to determine the number of clusters in the duty start. We are determining between 4-6 is the right cluster.

2.4.12.1 What variables needs to be used?

Not all of the listed attributes are equal. Filtering those attributes is very critical and important to a successful prediction. The process of identifying the more prominent data to the research questions that is being asked is Feature Selection. In the sleepDNA attributes in Fig. 39, we will choose features. Any redundant and irrelevant variables will be removed. The classifiers performance may decrease if the dimensions increase without enough training samples. Feature selection methods was further explored to see what was available to experiment. There are two main types of feature selection algorithms. A) Wrapper methods and b) Filter methods. I am planning to attempt a new method called Recursive feature elimination algorithm which is a wrapper type of feature selection process. Recursive feature elimination algorithm (PM. Granitto et al., 2006) (I. Guyon et al., 2002) which is a wrapper method type of feature selection process. This method considers the selection of selected features as a search problem where different features are compared to other combinations. As shown below we will start computing all the sleepDNA features and generate the feature importance in random forest (Ho 1998). Then the least important feature will be eliminated from the feature set. This is repeated until the highest performance is reached in the model.

2.4.12.2 Recursive feature elimination algorithm

- Initial: training set T
    - set of all features F
- For iteration i in F
    - compute feature importance with Random Forest
    - rank the feature set
    - find the last ranked feature f* in F
- Compute the performance with F-f*
    - if performance is improved:
    - update: F = F-f*
    - go back to Step 2 with new F
    - else break

Steps to be performed for the feature selection.
Exploring all the attributes to be used
Preprocessing sleep and work schedule data
Running algorithms and select attributes.

2.4.12.3 Classification of variables and clustering techniques

Clustering technique helps us to analyze the internal structure of a complex data. (MacQueen 1967) stated that clustering applications are considered more as an aid for investigators to obtain qualitative and quantitative understanding of a large amount of multivariate data than only a computational process that finds some unique and definitive grouping for the data. In the recent years many clustering algorithms (Han & Kambler 2001) (D. Hand et al., 2001) were developed to support large complex and unsupervised data. According to (Anderberg 1973), these are the major elements in the clustering analysis study.

Clustering methods helps to define the data in a cluster with internal cohesion and external isolation (Duda & Hart 1973). All clusters are defined with certain properties such as density, variance, dimensions, shape and separation (Aldenderfer & Blashfield 1984). We are looking for clusters to be created with tight and compact high-density region of data points when compared to other space. With fuzzy clusters there could be overlapping clusters where data points could belong to 2 clusters. In this research I am proposing we use distinct clusters with traditional partition clustering methods such as K-Means and Hierarchical methods.

2.5 Operational Impact on Duty DNA Structure

2.5.1 Flight and Crew Duty Impact Variables

Even before the pandemic, the delays in flight operations costed tens of billions of dollars annually to the world economy. The data from Department of transportation (DOT) data show that flight delays and cancellations have generally increased over the last decade. Since 1998 the flight delays have increased 62% nationwide. In 2007 alone the flight delay estimated was over $30bn (Ball et al., 2010). For the time between 2011 and 2015, over 30% of the gets delayed over 15 minutes and about 2% of the US domestic flights got cancelled. Since aviation network is a complex system, determining the causes has always been a challenging and considerable interesting discussion for the researchers and policy makers. The US Department of transportation classifies the delays into five main categories.

i. Air Carrier Delays
ii. Late Arriving Aircraft Delays
iii. National Aviation System Delays
iv. Extreme Weather Delays
v. Security Delays

In order to account for the delays, the airlines are introducing buffers or slack times that are distributed across the crew schedules. This can reduce extreme delays. There is always a trade-off between the goals of reducing delays and disruptions while not being overly conservative in buffer placement.

2.5.1.1 Air Cargo Operations Delay Impact

Cargo market has been consistently growing at the rate of 4-5% for the next 20 years (Boeing 2018). There is not much literature that are focusing on estimating the cost of flight delay in the cargo operations. From a recent study, the estimates of flight delay ranges from $8000 to $38000 per flight hour, with an average of about $20,000. This number is roughly three times the cost of average block hour in this industry, the cost is due to the large late package delivery commitments and the cost of delay is higher than the cost to passengers. The delay due to the unpunctual deliveries often directly impacts the operating cost such as on ground delivery cost and indirectly on degradation of services to shippers or recipients.

2.5.2 COVID Impact to Airlines

The global pandemic COVID-19 has disrupted the entire commercial industry that the losses in revenues are expected to be around 250B$. (IATA 2020) This resulted in a huge disruption to global supply chains as the freight carriers and commercial airlines are looking to bridge the gap in consumer spending.

In contrast to the dramatic drop in passenger demand, air cargo operations are surging to respond to calls to move essential supplies. It's primarily due to the loss of belly freight in passenger aircrafts which is about 40% of annual global air cargo. On the humanitarian side, air carriers had to think creatively to offer time sensitive relief shipments with less resources. The pandemic has created new dimensions to the operational impacts. The increased historical volumes are due to the supply chain demands, COVID protocols related to social distancing, application of disinfectants to machines and the shortage of workers contributes to major delays at the hubs. Outside of these challenges the freight flight operations have to deal with the challenges related to overfly regulations, operational curfews of airports, ever changing border restrictions. When the crew arrives at a destination, he is subjected to dynamic changing requirements for testing and quarantine where some requiring as much as 14-day quarantine for the entire crew. Not only the operators losing the crew for that long days but also the essential supplies of cargo need to be moved.

2.5.3 COVID Duty Delays and Fatigue

Operational disruptions in a complex aviation operation are inevitable. When the regulators and the union agreements prescribe limits, they include flexibility for day of operations extensions to allow the operator to manage on-the-day operational disruptions. On the other side it also allows reducing the minimum rest during the operational disruptions. The ability to use these duty extensions and/or rest reductions should depend on the crew member's assessment that they are fit to continue. Where such "flexibility" limits are prescribed, the airline operator should manage the frequency of their use as part of their normal Safety Management System (SMS) processes. Alternatively, the State may require the use of variations to allow the airline operator flexibility to manage operational disruptions on the day. Addressing unexpected operational circumstances and risks is discussed further in Comparing data on planned versus actual work periods can be used to identify times when fatigue might have been higher than expected. For example, an operator might track how often each month:

• Flight duty periods end at least 30 minutes later than scheduled;

• The maximum scheduled duty day is exceeded
(e.g., duty days longer than 13 hours);

• Flight duty periods start or end within the window of circadian low (WOCL); or

• Reserve crew are called out on particular flights, at a particular crew base etc.

These kinds of metrics point to possible mitigations if needed, for example changes to scheduled flight times or increasing the number of crew members at a given base. As part of routine SMS processes, the data need to be monitored regularly to evaluate whether the hazards identified warrant additional action.

2.6 Modeling

2.6.1 Biomathematical model parameters

The equations of the original two process model defined by (Daan et al., 1984) are

Process S

$$S_t = \begin{Bmatrix} d\,S_{t-1}\,; d = e^{\frac{-\Delta t}{\tau_d}}\ (sleep) \\ 1 - r(1 - S_{t-1}); r = e^{\frac{-\Delta t}{\tau_r}}\ (wake) \end{Bmatrix}$$

Process C

$$C = A\left\{\begin{array}{l} 0.97\sin\left[\omega(t-t_0)\right] + 0.22\sin\left[2\omega(t-t_0)\right] + 0.07\sin\left[3\omega(t-t_0)\right] \\ +0.03\sin\left[4\omega(t-t_0)\right] + 0.001\sin\left[5\omega(t-t_0)\right] \end{array}\right\}$$

$$\omega = \frac{2\pi}{\tau}$$

Process S is the homeostatic process that is an increasing exponential function for being awake and decreasing exponential function during sleeping.

d and r are the decay factor and rising factor that is used for S

τ_r - Time Constants for rising factor = 18.2h

τ_d - Time Constants for decay factor = 4.2h

Δt = increment step defined at 0.5h

As shown in Figure 4, Process S is generated through the recursive iteration with time indices t, t-1.

Process C is a skewed sinusoidal equation that is independent of sleep and waking.

A = the sinusoidal sine wave's Amplitude and the direction of skewing is determined by the sign

t = time

τ = period of the circadian process at 24h

t0 = phase of the circadian at the beginning of the simulation = 8.6h

Hm and Lm are the upper and lower threshold of the circadian waveforms as shown in Figure 4 (Orange and purple sine curves).

Daan also suggested that the parameters could require changes depending on the applications and the type of simulations.

2.6.2 Dynamic Circadian Modulation in a Biomathematical Models

Building on these equations on the two process seminal models of sleep and regulation formulated a new class of models in terms of first order ordinary differential equations (ODEs).

(1a) $$\begin{bmatrix} \frac{dp(t)}{dt} \\ \frac{du(t)}{dt} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ 0 & \alpha_{22} \end{bmatrix} \begin{bmatrix} p(t) \\ u(t) \end{bmatrix} + \begin{bmatrix} \kappa\, c(t-\phi) + \mu \\ 0 \end{bmatrix}$$ during wakefulness

(1b) $$\begin{bmatrix} \frac{dp(t)}{dt} \\ \frac{du(t)}{dt} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ 0 & \sigma_{22} \end{bmatrix} \begin{bmatrix} p(t) \\ u(t) \end{bmatrix} + \begin{bmatrix} \kappa\, c(t-\phi) + \mu \\ 0 \end{bmatrix}$$ during sleep

(2a) $$\frac{dp(t)}{dt} = \alpha_w\left[p(t) + \beta_w u(t)\right] + g_w(t)$$ during wakefulness

(2b) $$\frac{dp(t)}{dt} = \alpha_s\left[p(t) + \beta_s u(t)\right] + g_s(t)$$ during sleep

(3a) $$\frac{du(t)}{dt} = \eta_w u(t)$$ during wakefulness

(3b) $$\frac{du(t)}{dt} = \eta_s u(t) + 1$$ during sleep

(4) $$\eta_s = \frac{A_c}{A_c - 1}\eta_w$$

(5a) $$g_w(t) = \kappa(t)\left[c(t) + \mu_w\right]$$ during wakefulness

(5b) $$g_s(t) = \kappa(t)\left[c(t) + \mu_s\right] + \frac{\alpha_s \beta_s}{\eta_s}$$ during sleep

(6) $$c(t) = \sin\left(2\pi \frac{t-\phi}{\tau}\right)$$

(7a) $$\frac{d\kappa(t)}{dt} = \lambda_w \kappa(t)\left(1 - \frac{\kappa(t)}{\xi}\right)$$ during wakefulness

(7b) $$\frac{d\kappa(t)}{dt} = \lambda_s \kappa(t)$$ during sleep

FOOTNOTE: In Eqs. (1) in the table above,

$$c(t-\phi) = A\sum_{k=1}^{5} a_k \sin\left(\frac{2k\pi}{\tau}(t-\phi)\right) = A\sum_{k=1}^{5} a_k \sin\left(\frac{2k\pi}{\tau}(t-\theta-t_0)\right)$$

The researchers at Washington State University used scheduled time in bed as input and calibrated with large datasets of neuro behavioral performance from laboratory-based dose response studies of sleep loss. The mathematical framework was focused on the temporal dynamics between sleep/wake cycles across multiple days. The model presents the nonlinear interaction between the homeostatic process and the circadian process. The new set of differential order equations are presented by (McCauley et al., 2009) and (McCauley et al., 2013). where time is represented by t (expressed in hours); τ = 24h; the values of A and a1 through a5; θ = 12.7h and t0 = 8.6 h is taken from the two-process model. The ODE system equations describe two state variables p and u. p represents the primary outcome variable, which denotes predicted performance impairment. u represents the slow dynamic process that causes modulation of p over days. The parameters in figure α and σ governs the homeostatic changes in p and u respectively. Function c in (6) as represented in the two process model equations earlier is a 5-harmonic skewed sinusoidal oscillator function (Borbély & Achermann 1999). The function c is scaled by K, offset by μ and shifted by Φ. Java programs are written for sleep predictions, two process oscillators and the ODE model for temporal dynamics and are found in appendix. I have used these to do validate assumptions, generate data and produce simulations,

2.6.3 Predicting Fatigue without knowing the history of schedule

2.6.3.1 Problem

During the crew planning process, the schedules are constructed using the optimization models and operations have begun to use predictive biomathematical models to estimate fatigue levels for each of the duty planned. Further to the estimation of fatigue exposure, the operators runs through the complete fatigue risk evaluation process as provided in section 1.4.2.1 (Predictive process of managing crew schedules). Earlier in the sections we went through the disruptions, impact and how operators manage the disruptions. Owing to the disruptions, deformed schedules are repaired by fixing part of duties, duty itself or combination of duties. Sometimes these repaired portions of the duties will be flown by a different crew either reserve or adding the portion of the schedule to an existing crew who has more rest time in their schedule. During the process of duty repair or rebuilding process, more often we won't know the history of which crew the repairs will impact to involve the biomathematical process. As mentioned in section 2.4, prediction of fatigue using biomathematical model requires the previous work history before the current duty in question at least to 4-5 days. We needed a prediction mechanism that will get us a fatigue prediction score with any random duty built in isolation which does not have any context of the previous duty period sequences. The approach we are taken is to come up with model framework that will deploy combinations of ML models to predict the future state of duties and its predicted fatigue variables.

2.6.3.2 Proposed Approach

Two ML models are constructed for this approach. One to predict the future end state of the duty based on what we know before the duty. Second is to predict the fatigue variables based on the new state of duty without any knowledge about the history before the duty. It's more important to predict the fatigue variables without knowing the 96-hour time window before the duty. In the world of operations and recovery, we will know very little on the history because as we put together the duties the final sequence of multiples of duties are less known.

2.6.3.3 Brief Explanation of the proposed model

1.0 The flights schedules are imported and sequenced into a trip. This is a continuous sequence of flights generated through the optimization model. Refer 2.4.1 and 2.4.2 on crew planning, duty and schedule DNA.

1.1 The flights in a schedule sequence are grouped into duty periods where the duties are separated by legal rest periods. Depending on the type of duty periods legal rest can be anywhere from 9 hours to 12 hours. Schedule will have 1 or more duty periods connected through legal rest periods between them (Refer 2.4.1).

(1.1.1, 1.1.2, 1.1.3) once the duties are assembled in a schedule, the schedule is passed through the sleep prediction algorithms where the model predicts minute by minute set of sleep states (0=wake,1=sleep). Sleep states are then converted into sleep segments and patterns that comprises of start time, end time and the location of sleep periods.

(1.1.4, 1.1.5): Once sleep patterns are generated from work patterns, both the segments are fed into the biomathematical model to produce fatigue variables.

1.1.6. We repeat the process for the last 10 months of data generating over 200k duty periods to preparing for the ML model.

1.1.7. Do the exercise of data prepare, clean, model choosing, split, train and test ML model.

1.1.8. Develop the model to predict fatigue variables for any random duty period DNA with all known variables.

1.2.1. To help predict the duty deformation. (For the scope of this project, we are only considering duty extension as the deformation variable). Collect the data values as flown. To measure the extension of the duty, we will capture the actual duty start and duty end and compute the actual duty length and compare it with the scheduled duty length. For Example (if the duty scheduled to start at 10am and end at 2pm, the length of duty is 4 hours. If the duty actually started on time at 10am but ended at 3pm due to delays, then the duty extension variable is set to 1 hour).

1.2.2. Repeat extracting these delays with 10 months of historical information generating over 200k records.

1.2.3 Do the exercise of data prepare, clean, model choosing, split, train and test ML model. 1.2.4 Develop the model to predict the duty end state. For the purposes of this research is to predict the duty delay values.

2.0 The new predicted duty end state for any random duty period is now passed into the ML model (1.1.8) and predict the fatigue variables.

1.2.1, the new values of duties are populated as duty start and duty end and sent to the model 1.1.8 to predict the fatigue variable.

2.6.4 Machine Learning Approach to predict duty deformation.

As discussed in section 2.5, the operational impact on a duty sometimes lead into decreased alertness. The operators try to deploy controls to keep the schedule in tact most of the times. However, beyond their control, sometimes due the reason as mentioned before could cause deformation and delays in a duty. This section describes methods to be able to accurately predict those situations and predict decreased alertness for proactively manage fatigue risk. In this schematic representation as shown in Fig 54, we will use combination of machine learning algorithms to predict fatigue at the end state of the duty as it would be operated given all the known variables of duty at the beginning of duty before operating. More on the approach and models are mentioned below.

2.6.4.1 Business Problem

The Fig. 55 below is a representation of a duty for explanation purposes only. The trip in Fig. 55 is scheduled with three duty periods. Duty1 with one segment MEM-EWR, Duty2 with three segments EWR-PIT-IND-DFW, Duty3 with one segment DFW-MEM. In this representation I have highlighted how the trip and its duty has changed its structure when it operated due to several disruptions. Duty 1 as operated took a delay in Memphis, MEM of an hour so it arrived late. The crew reported to duty on time and the flight took a delay of an hour possibly due to any of the combinations of reasons (maintenance, volumes, ATC clearance, weather etc.). This caused the duty to arrive in Newark, EWR an hour late. This is a simple deformation of a duty from originally planned. The second duty went through some complex changes, instead of crew going to Indianapolis IND from Pittsburg PIT, he was revised to go to Alliance AFW and from there to Atlanta ATL. Due to these changes and other factors on turning the aircraft the duty was substantially extended to as much as 2 hours. The third duty had a change in route but the total duty time was the same as planned. In order to predict the futuristic state of risk in a crew, we need to predict the fatigue variables for not what was planned but to what the duty will transform to. In this section we will discuss the approach of predicting these transformations. To keep the scope limited we will be taking on the delay time in a duty from originally planned.

2.6.4.2 Data Cleaning

As part of this exercise, we collected data from multiple sources, joined them and looked for empty or null values and cleaned them. In total we generated 65000 duty periods with about 27 features (‘Base’, ’LBT Tm’, ‘Int’, ‘Type’, ‘Op in Critical’, ‘D St’, ’D End’ etc.,) as shown in the table 2 below and 1 label ‘Duty Variance’ to train the machine learning model. The whole process can be divided into two part, encoding and training. Also, we introduced new fields like COVID impact to understand the current situation.

2.6.4.3 Features and Encoding

The data was further cleaned and coded where ever it is necessary. For example, the int field had to be coded to TRUE/FALSE, B-S field which represents the total flying time in that duty period was in HHmm time format, we had to convert that to the double value with 2 decimals. Duty variance that was originally computed as a difference between duty scheduled and duty actuals, was also coded into finite set of labels. Scored probabilities and labels are computed fields by the model (Table 2).

2.6.4.4 Attempted ML Models

This problem of predicting delay is characteristic of a classification problem. We attempted to categorize the type of delay. No Delay, Small Delay and Large

Delay coded as 0,1,2 respectively in the last column duty variance. Here are few examples of how the variables were used to train.

In this task we are predicting a class label of a duty delay based on the input data in table 3. From modeling perspective this classification requires a training dataset with the input features. After studying a number of machine learning algorithms to solve classification problems, we zeroed in on first using the binary classification to first understand the baseline. Binary classification (Duda et al., 2001) is the task of classifying the elements of this dataset into two groups. Did the duty encounter a delay or not? Based on the prediction rule, the individual data set can be put in one of these categories. Based on the understanding of data in binary classification, we can explore doing multiple labels with multi classification algorithms gradient boosting and also neural network classification to improve the accuracy further.

2.6.4.5 Models and Baselines approach

When looking at the 65k data spread over 3 covid months, in Figure below, we found not overly imbalanced but a fair distribution of the delay/duty variance. The first look at codifying these absolute delay values in minutes to a binary classification label gave a 70/30 split between 0 and 1. Any delay that is less than or equal to 30 minutes was considered a “No Delay”. Delays greater than 30 minutes was considered a “Delay”. Duty variance field was adjusted to define 0 and 1 values. On the baseline we used a 90/10 split between training and testing data with the 3 months of data. All these data are collected during Covid times. For the baseline we chose Gradient boosting (Breiman 1997) model. Gradient boosting is a technique used for both regression and classification problem where models are built on a stage wise boosting methods as an optimization algorithm on a suitable cost function. As part of the baseline of learning obtained from the 3 months of data, we wanted to validate on a completely different month that was not part of the initial 3-month data. Also, we wanted to apply this baseline model and see how it compared with the combination of covid and non covid impacted months. For the improvements, we wanted to expand the labels from binary to multiple labels to distinguish between shorter and longer delays. To do so we have to change the labels to 0, 1, 2 that represents less than 30 minutes delay, 30-60 minutes delay and excess of 60 minutes delay

2.6.5.7 Evaluation Metrics

In the classification models, confusion matrix (Stehman 1997) is often used to describe the performance on the test and training data where true values are known. It is also known as error matrix. The table layout allows the visualizing of the algorithm as in Fig 57.

Recall: Recall provides a measurement of all the positive classes, how much we predicted correctly. It should be high as possible. RC=TP/(TP+FN).

Precision: Of all the positive classes we have predicted correctly, how many are truly positive. TP/(TP+TN).

Accuracy: of all the classes, how much we predicted correctly, which will be (TP+TN)/Total.

F-measure: is a measure we should use when we have low precision and high recall or vice versa. F-Score or measure helps to measure recall and precision at the same time. it uses the arithmetic mean. = 2*Recall*Precision/ (Recall + Precision).

2.6.5 Machine Learning Approach to predict fatigue variables

In order to predict fatigue on the end state of a duty, we need to know two things a) to know how the end state of duty will look like b) to know the state of fatigue at the beginning of the duty. In section 2.6.4, we discussed the approach to the baseline on duty deformation. Using the duty variables, we will look to predict the fatigue at the start of duty and then use both these predicted variables to predict the end state of the duty. To predict fatigue on duty, start and duty end, I am proposing to use ML approach to predict the fatigue variables. To do this we will first identify the various fatigue variables. Fig. 57 below is a representation of a biomathematical model output with both predicted sleep and predicted fatigue variables. The curve represents the predicted fatigue scores plotted minute by minute for a schedule with three duty periods. Duty 1 with one flight (MEM-LCK), Duty 2 with three flights (LCK-BNA, BNA-EWR, EWR-BOS) and Duty 3 with one flight (BOS-MEM). All the three duty periods are separated by a legal rest period. The orange overlay on the curve represents the flight segments. the purple overlay on the curve represents the predicted sleep segments per 1.1.3. Fig. 58 above represents the step by step functioning of this multi phased approach to solve duty end state fatigue predictions. Step 1 contains all the known duty variables at the duty start or prior to duty start. Step 2a computes the duty delay to predict the duty delay attribute. Step 2b which is independent of Step 2a and can be modeled separately. Step 2b is predicting the fatigue state of the duty period at the beginning of the duty. Combining these two predictions and feed into the third model is Step 3. The figure represents how the missing values are predicted using combinations of models

2.6.5.1 Fatigue Variables

For both Step 2b and Step 3 above, set of fatigue variables are computed. These five key fatigue variables/indicators will be the target variable for predictions in the Machine Learning model. these five variables are computed separately at the duty state in the beginning of duty period and the duty state at the end of the duty period.

a. Fatigue Levels (KSS) at the beginning of the duty period
b. Fatigue levels (KSS) at the end of duty period
c. Sleep in the previous 24 hours at start of duty
d. Sleep in the previous 24 hours at end of duty
e. Sleep during the duty period

2.6.5.2 Data Generation

For the first iteration, I wanted to run the model for subset of schedule by computing the 5 fatigue variables using the biomathematical model. A sample of about 20k duty records were chosen for the dataset. Along with fatigue variables, the duty DNA attributes were also captured.

Data Generation

- Extract schedule is in Table 4 for each pilot for the entire month
- Generate duty periods in sequence
- Construct Duty DNA with all parameters
- Use sleep prediction algorithms to generate predicted sleep for entire set of duty sequences for the monthly period.
- Use biomathematical model to compute predicted KSS
- For each of the duty period, bind the 5 variables to the duty DNA
- Repeat the same for all duty periods for the pilot
- Repeat the same for all pilots in the month.
- The output will be a list of all duties operated by all pilots in the month

2.6.5.3 Data Cleaning

As part of this exercise, we generated data from multiple sources, joined them and looks for empty or null values and cleaned it accordingly. Also all date and time formats were converted into numeric values and string based Boolean were converted into “true, false”. We also removed variables that didn’t add values to the model. We converted the date and time field of duty local base time start to “Day of week”, “month”, “year”, “hour”, “mins into day”. We understood that there are several operational factors that influenced these variables. For example, there could be additional volumes on a particular departure and hub turn city on certain day of the week that could cause additional delays in duty starts. Also, availability of crew resources could vary by week or month. Also, we introduced new fields like COVID impact and if the turn city is operating through the hub or not.

2.6.5.4 Features comparison and selection

I ran the distribution plots and heatmaps for all the variables and their relationship with each other. The data we used was for a subset extracted as in Table 5 for a given month. In this I was looking to target 5 fatigue variables. These values for all 23k records were generated using biomathematical models as shown in Fig 61.

a) Fatigue score KSS value at the beginning of duty period
b) Fatigue score KSS value at the end of duty period
c) Sleep count in the last 24 hours from beginning of duty period
d) Sleep count in the last 24 hours from end of the duty period
e) Sleep count within the duty period

The distribution of those 5 values as computed by the biomathematical model is shown below. for the detailed analysis purposes and model improvements, we will keep the most significant fatigue feature (i.e. fatigue levels as scored at the end of the duty period). The reason this was selected over the KSS score at the start of the duty period was that the assumption that pilot will be adequately rested to start the duty period.

2.6.5.5 Machine Learning Model

For this model building we will be considering the machine learning technique (Mitchell 1997) where the algorithm learns through data to solve without programming the rules. in the classical algorithm, rules are explicitly given to the computer to perform a task. In the machine learning, a parameterized model that defines a family of possible rules is given to the computer along with whole bunch of data and strategy to find the better rules among the possible ones. We will be considering all the supervised machine learning techniques. In this technique we will have set of input variables which will be all the features of the duty DNA and the output variable will be a target variable. If we are setting dutyEndKSS then $Y_{dutyEndKSS} = f(X_{dutyDNA})$, similarly we can derive the output for all the 5 fatigue target variables as mentioned above. The goal will be the approximate the function so output variable is predicted based on the input variable.

Since we will be setting up algorithms from the training or historical datasets, this will fall under the category of supervised machine learning (Russell & Norvig 2010). the algorithms are made to learn till we get it to the acceptable level of performance. This depends on the end user determining how the target variables needs to be approach and evaluated. The choice of the algorithms are largely dependent on several factors such as bias-variance and its tradeoff, complexity of the classifier or regression functions, amount of training data that will be used, higher dimensions of input features (Guyon & Elisseeff 2003) and noise of the output values with overfitting and underfitting (Cai 2016). For both the models in Step 2b and Step 3, we are considering regression type over classification because the target variable will be a real value (for example dutystartKSS and dutyendKSS will be a double value from 1 to 9. Classification approaches are best suited for predicting categorical targets using training data while the regression is used to predict the continuous values.

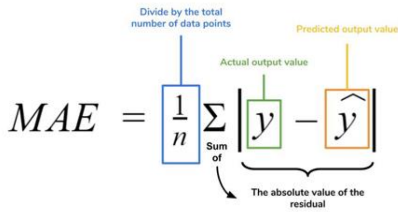
2.6.5.6 Models and Baselines approach

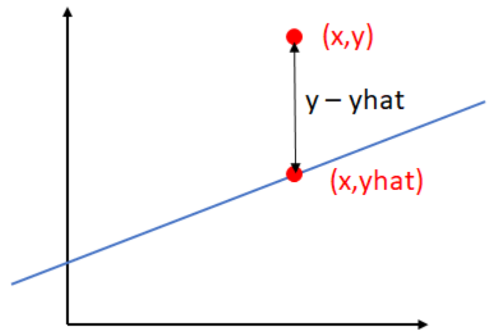
In doing some background research for a suitable decision tree regression, we found gradient boosted regression trees (Elith 2008) (Friedman 2003) are the most effective machine learning model for predictive analytics also called the industrial workhorse of machine learning. The model uses the technique where the prediction model is built in the form of an ensemble of weak prediction models typically decision trees.

For the model to predict duty end state, the goal was to generate one baseline with this decision tree approach. We will call this run as A1. The second baseline was developed with the linear regression (Freedman 2009) technique. Since most of the values are numerical, this linear regression method was considered as another baseline to find the best fit line between the independent and the dependent variables. Our goal in this technique was to predict the fatigue variables based on the generated datasets which are our explanatory variables. We will call this run B1. The third baseline was developed with a neural network deep learning technique (Schmidhuber 2015). This method is a much more complex and comprehensive linear regressions with a much better performance against nonlinear fitting. It tries to mimic the working of neuron in human brain for learning. At first it is unstable and after certain iteration of data it adjust itself such that it's accuracy increases. Basically, you can apply any know function using neural network. We will call this run as C1. For the model to predict the fatigue levels at the duty start at the beginning of the duty period, we wanted to do a deep learning approach with neural network regression model. Since the scope was more leaning towards improving accuracy on the duty end state, we took a simple straight forward approach and left more room for further research and improvements. We took the baseline approach with 106k records.

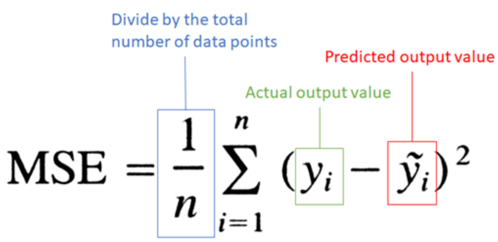
2.6.5.7 Evaluation Metrics

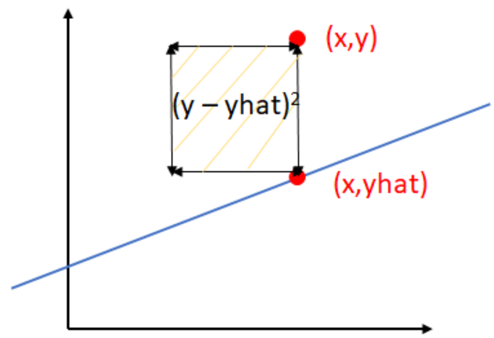
There are metrics that helps us to evaluate the model performance. In our results we will be referring to these metrics and how they are performing. These metrics are computed from the generated line to the real data points. Mean absolute error (MAE), is the average of the absolute difference between real data points and the predicted outcome.





Mean square error (MSE), is the average of the squared difference between the real data points and predicted outcome (Hyndman & Koehler 2005).

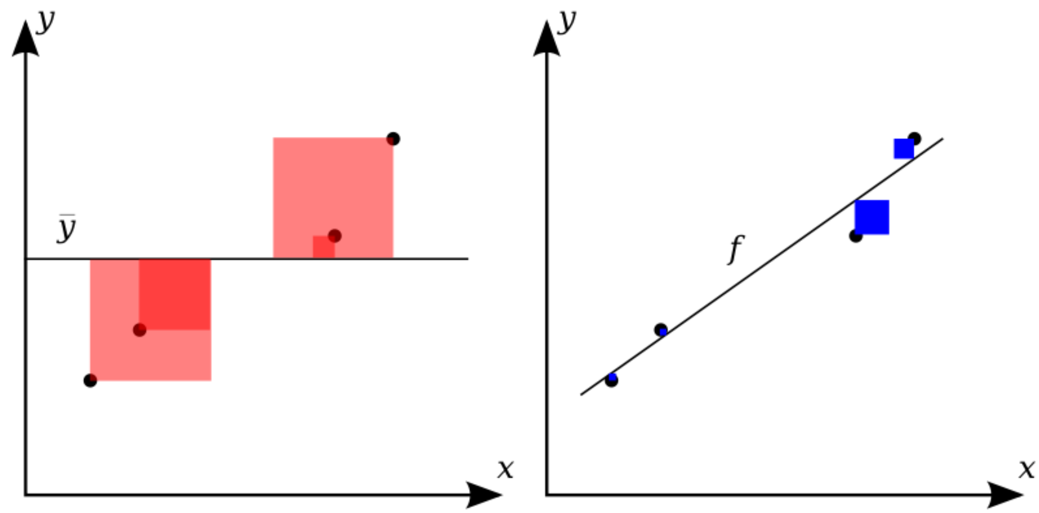




Root Mean squared error or RMSE, is the root of the man of the squared errors and is the most popular in determining the performance of regression models.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Coefficient of Determination or R^2, is the proportion of the variance in the dependent variable that is predicable from independent variables (Glantz & Slinker 1990).



$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

Total sum of squares SS_{tot} is proportional to the variance of the data whereas the SS_{res} is the sum of the squares of residuals as described to the right of the picture above. If the modeled values are exactly matching the observed values then the value of SS_{res} wil be 0 so R^2 will be 1. Anything that is closer to 1 is considered good prediction model.

Relative Squared Error (RSE), is relative to what it would have been if a simple predictor had been used. More specifically, this simple predictor is just the average of the actual values. Thus, the relative squared error takes the total squared error and normalizes it by dividing by the total squared error of the simple predictor.

$$E_i = \frac{\sum_{j=1}^{n} \left(P_{(ij)} - T_j\right)^2}{\sum_{j=1}^{n} \left(T_j - \bar{T}\right)^2}$$

Relative Absolute Error (RAE). is very similar to the RSE (Thiel 1966) in the sense that it is also relative to a simple predictor, which is just the average of the actual values. In this case, though, the error is just the total absolute error instead of the total squared error. Thus, the relative absolute error takes the total absolute error and normalizes it by dividing by the total absolute error of the simple predictor.

$$E_i = \frac{\sum_{j=1}^{n} \left| P_{(ij)} - T_j \right|}{\sum_{j=1}^{n} \left| T_j - \overline{T} \right|}$$

2.6.5.8 Azure Machine Learning Studio

Azure Machine Learning Studio is a GUI-based integrated development environment (Fig 62) for constructing and operationalizing Machine Learning workflow on Azure. For this exercise we used the combination of python programming and Microsoft azure portal for building, train, test and validate different models. The framework helps to formulate the entire machine learning development life cycle from pre-processing data, preparing data, preparing baseline ML models, evaluating to consuming ML model.

CHAPTER III
RESULTS, CONCLUSIONS AND FUTURE WORK

3.1 Duty Delay Prediction Results

The first of the algorithm is predicting the duty delay. The input variables will be the duty attributes that are known at the start of the duty period as shown in Fig 54.

3.1.1 Initial Baseline Results – Run D1, D2 and D3

The initial run was started with a simple binary classification approach. The initial dataset used was 3 months of data captured during covid times when the operations experienced more delays compared to non covid times. The data set size was about 65k records of duty periods. The duty variance which is the difference between the actual duty length to schedule duty length were classified with 0 for delays <= 30 minutes and 1 for delays exceeding 30 minutes. The data had about 68% of 0 and 32% of 1 value. For training 90/10 split model was used.

3.1.1.1 Run D1 – Gradient Boosting Model – 65k records - Table 6

The ML model used for the first of the initial baseline run was a well-known gradient boosting binary classification model. The model used 17 of 26 features that was provided and the most prominent features are the ones that is listed in reverse order in figure above. Sequence of tail numbers shows a higher correlation to the cause of delay compared to the others. This could be because of maintenance issues, the routing of a particular aircraft in a specific region.
Day of the week is also another prominent feature. This could be because of the additional cargo volume on certain days of the week flowing through the system. The model results on test data showed an accuracy of 82% (See Table 7), that it predicted either delay happened (1337 times) and on time duty prediction of (3972 times). Since the business objective is to predict delays, the recall percentage of delays were about 64%. This means that out of all the true values of delays, the model predicted 64% of times right. To further investigate, different true/predicted values of 0s and 1s were explored on the scored probabilities. The rule that is

applied for a Boolean 1 or 0 is a 50% mark. Currently as baseline we determined anything great than or equal to 50% probability on >30 mins delay will be a 1 and anything less than 50% will be 0. The above graph represents the plots of the different probabilities. If the goal of the problem is to predict risk, the emphasis should be more on the duties that predicts delay on actual delayed duties. Assuming that the risk on the ontime duties are known at duty construction phase and the operations accepted certain risk at that point. If the focus is on the Actual 1 in Fig 63, then we have about 149 of the test data scored probabilities between 0.4 to 0.5. As an improvement, AUC approach can be used to define the right threshold of where the segregation should be.

3.1.1.2 Run D2 – Gradient Boosting Model – Validation 21k records – Table 8

ML Model Run D2 used the same classification model that was trained with 65k records but for validation we used a new month data of 21k records. This was to understand if the model was good enough to predict something for the values that aren't known. The attempt was also made to see the validity of the model due to the ever-changing system form. The overall results came to good 80% for the entire month data used for validation. The model used 17 of 26 features that was provided and the most prominent features are the ones that is listed in reverse order in figure above. Sequence of tail numbers shows a higher correlation to the cause of delay compared to the others. This could be because of maintenance issues, the routing of a particular aircraft in a specific region. Day of the week is also another prominent feature. This could be because of the additional cargo volume on certain days of the week flowing through the system.

The model results on test data showed an accuracy of 80% (Table 9), that it predicted either delay happened (10173 times) and on time duty prediction of (4780 times). Since the business objective is to predict delays, the recall percentage of delays were about 48%. This means that out of all the true values of delays, the model predicted 48% of times right. This could be because of the imbalance in data and the characteristic of the delay couldn't be answered by the model

3.1.1.3 Run D3 – Gradient Boosting Model – 215k records – Table 10

For the third of the baseline, ML Run D3 also used the same classification model. D3 used a larger dataset covering upto 10 months with 5 months during covid and 5 months pre covid. In total the model had 215k records and used a 90/10 split for training.

The overall results came to about 83% (Table 11). The Recall numbers of predicting the delay value of > 30 minutes was still at 50%. (ie we were able to predict only 50% of the times for all duties that had a delay of 30 minutes or more). One reason for low performance on recall is the fact that in the 215k records, only 22% of the duties that had a delay of greater than 30 minutes. This is much lesser in percentage when compared to the covid months (in D1 set) that was about 32% of delayed duties.

3.1.2 Improvements made to Duty Delay Prediction

For improvements, there were two options considered. One to improve the predictions further by optimization or rebalancing the data. This can be done either by undersampling the higher ontime duty performance data and/or oversampling the lower number of delayed duty data. The second option is to consider increasing the number of labels and try improve the accuracy with additional labels. Due to the time constraints and the number of experiments, decision was made to further explore on increasing the number of labels. Classifying the delays further will provide more operational use on either removing the pilots from the trip if their predicted risk is higher over others. The data was prepared with new labels of 0,1,2 for delay less than or equal to 30 minutes, between 30 and 60 minutes and excess of 60 minutes respectively. We call these delays as no delay, small delay and large delays. The data in Fig 64 shows a 68%,18%,14% in 0,1,2 labels respectively.

For improvements in modeling, I wanted to select one non-deep learning and one deep learning methods. Multi class boosted decision tree and multi class neural network classification model was considered. Boosted decision tree uses the ensemble methods to combine predictions from many individual trees whereas the deep learning neural network uses hidden layers between inputs and outputs. Other reasons for deep learning are that data generation for training models is not a problem. For this exercise, the same data of 65k records that was used for Run D1 is used.

3.1.3 Final Results – Run D4 and D5

The results from the improved classification model are documented below

3.1.3.1 Multi Class Boosted Decision Tree Classification Model – Run D4

In the boosted decision tree with 65k data, random split of 80/20 is used for training. Overall, the system was able to predict on time over 94% and small delays of 32% and large delays over 53% giving an overall accuracy of 78%. The recall for the delayed records showed increase to 60% from the baseline data.
This result in Table 12 and 13 is a good progress because accuracy was expected to compromise with the increase in the number of labels.

3.1.3.2 Multi Class Neural Network Classification Model – Run D5 – Table 14

The second model that was used for improvement is the deep learning neural network classification model with the same 65k data. The model also used the same random split of 80/20 for training. A satisfactory performance was noticed on the overall accuracy given the fact that we added additional labels. Overall, the system was able to predict on time over 93% and small delays of 60% and large delays over 63% giving an overall accuracy of 83% and recall for any delays increased noticeably to 72% from baseline. The neural network classification showed in Table 15 is a very good improvement on both the small and large delay predictions

3.1.4 Results Comparison – All Duty Delay Prediction Models

Comparing all the models for the delay predictions in Table 16, multi label neural network classification model came out as the best. This model can be further improved by rebalancing and improved optimization. Also, more duty records could be used to help improve deep learning models. The future work should also consider looking into other attributes like weather history, curfew, and maintenance records. These features has possibility of improving the predictions more.

For further explorations on the binary delay or on time predictions approach, I will be considering the balancing options (He & Garcia 2009) with over and under sampling. Also, I will be look into using Area under the Curve (AUC) before rebalancing and after rebalancing to determine the threshold value to separate the Boolean value.

3.2 Duty Begin State Fatigue Prediction Results

The second of the algorithm is predicting the fatigue state of duty before duty starts. Predicting KSS at duty start and predicting duty delay are independent process and can be computed in parallel. The input variables will be the duty attributes that are known at the start of the duty period as shown in Fig 54.

3.2.1 Baseline Results

Predicting the duty end state fatigue required the predicted state of fatigue in the duty begin state. To help fast forward the evaluation and with over 100k records available, an attempt was made to check the accuracy with the deep learning neural network regression model. Regression was used because the values were predicted between 1 to 9. A total of 106k records was used as input and 80/20 split was used to train the model. The model used about 17 features. Figure 65, 66 shows the graphical flow of how the neural network was model was built and executed. The frequencies are plotted on the y axis. The blue line is the data generated from the original biomathematical model or the target values. The orange line is the data scored by the neural network regression model. This model showed better predictions on higher KSS values and inaccuracies in the lower KSS values. The overall fit is at 88.3%.

3.2.2 Future Improvements

The above histogram plot in Fig 67 and Table 17 shows only 7% (1495 out of 21259) of the duties that fell outside of +/- 0.5 KSS. Out of the 7%, 3% of the duties had the target KSS of less than 3. This is considered extremely safe duties and are very fatigue friendly. If we are considering from risk perspective, not predicting this right might not have adverse safety impact to the operations. Since the model was already predicting close to 88% fit and also considering all the other scenarios above, the accuracy can be assumed even more. Further improvements to model can be achieved by running the model with additional data and exploring other models.

3.3 Duty End State Fatigue Prediction Results

The third and the last of the algorithm is predicting the fatigue state of end of duty or when the duty actually ends. This model will use the prior prediction model outputs along with the other duty variables as shown in Fig 54.

3.3.1 Initial Baseline Results

With the initial data set of 23k duty period, predictions were run for all the five variables. For defining the scope of this research, analysis is being limited only to one of the five variables. KSS score at the end of the duty being the most important of all was selected as the target variable to demonstrate improvements and accuracy.

These are the runs attempted for baseline results

Run A1 – Boosted Decision Tree Regression
Run B1 – Linear Regression
Run C1 – Neural Network Regression

3.3.1.1 Results from Boosted Decision Tree Regression – A1 – Table 18

The model graph is shown in Fig 68. The graph describes the lifecycle of the machine learning model. It starts with extracting the data from traditional crew system data factory and formulate them into an excel/csv sheet. Data is then imported into Azure data platform. Inside the platform, column selections and data cleansing methods are deployed to get the data ready for model. Once the data is ready, the right model is chosen along with splitting the data for train/test. For this baseline approach the 80/20 split was used. Other splits could be checked for further improvements as necessary. Score model component helps to score the training and testing datasets. Scored output is then extracted for review and iterative improvements.

The original KSS values at duty end computed from the biomathematical model is the target variable and is compared against the boosted decision tree regression model for all of 4.5k test records (Fig 69). The data was aggregated into bins of 0.5 between 1 and 9 on the x-axis. 1 being the least fatigued and 9 being extremely fatigued. The frequencies are plotted on the y axis. The blue line is the data generated from the original biomathematical model. The orange line is the data scored by the boosted decision tree. The model showed larger differences in the original values between 3.5 to 5.5. For values greater than 5.5, the model predictions were better. The histogram plot in Fig 70 is the differences between the original target value and scored value, it follows the nice normal distribution with majority of the duties are scored within +/- 0.5 of KSS. The output of the model shows a 91.4% fit with the following evaluation metrics.

3.3.1.2 Results from Linear Regression – B1 – Table 19

Similar to A1, the life cycle of B1 in Fig 71 is also the same except the model used is linear regression. The model was created with the same 80/20 split. The original KSS values at duty end computed from the biomathematical model is compared against the linear regression model for all of 4.5k test records (Fig 72). The data was aggregated into bins of 0.5 between 1 and 9 on the x-axis. 1 being the least fatigued and 9 being extremely fatigued. The frequencies are plotted on the y axis. The blue line is the data generated from the original biomathematical model. The orange line is the data scored by the linear regression. The model showed larger differences in the original values between 2.0 to 5.5. The above histogram plot in Fig 73 is the differences between the original value and scored value, it follows the normal distribution with over 500 duties were off by +/-1 KSS. The output of the model shows a simple 76% fit with the following evaluation metrics.

3.3.1.3 Results from Deep Learning Neural Network Regression – C1 – Table 20

Similar to A1 and B1, the life cycle of C1 in Fig 74 is also the same except the model used is neural network regression. Same 80/20 split was used; the trained model was scored on the 20% of the data. The original KSS values at duty end computed from the biomathematical model is compared against the neural network regression model for all of 4.5k test records (Fig 75). The data was aggregated into bins of 0.5 between 1 and 9 on the x-axis. 1 being the least fatigued and 9 being extremely fatigued. The frequencies are plotted on the y axis. The blue line is the data generated from the original biomathematical model. The orange line is the data scored by the neural network regression model. This model showed a much closer prediction to the other two models.

The histogram plot in Fig 76 is the differences between the original value and scored value, it follows the normal distribution with over 500 duties were off by +/-1 KSS. The output of the model shows a mere 94% fit and less than 100 duties have a difference of +/- 1 with the following evaluation metrics.

3.3.2 Improvements

All the three baseline models are put side by side for comparison (Table 21). Keeping the model features and dataset size the same, the deep learning neural network regression performed the best at 94% overall fit.
To improve further on the deep learning neural network regression, one of the options considered was to increase the number of records to 106k records. same exercise of the baseline models was repeated for the data generation, selecting columns, cleaning the dataset and coding the data set. The following three experiments were attempted for improvements.

C2 – use the model and other model parameters the same but just increase the size of data to 106k records. train the model with this 106k records with 80/20 split.

C3 – use model C2, but validate against the a larger dataset of 242k records

C4 – retrain the model with a larger dataset of 242k records with 80/20 split.
Further improvements can be considered by changing the split, improving the data cleansing and maybe add additional features if needed.

3.3.3 Final Results

Model was built and experimented for runs C2, C3 and C4. See below their results and comparison

3.3.3.1 Neural network regression with 106k records – C2 – Table 22

In this model, a total of 106k records was used with about 85k records for training and 21k records for testing. Similar to the baseline models A1, B1 and C1, the life cycle of C2 as shown in Fig 77 is also the same except the model used larger number of records. The experiment used the same 80/20 split for training the model. The original KSS values at duty end computed from the biomathematical model is compared against the neural network regression model for all of 21k test records. the data was aggregated into bins of 0.5 between 1 and 9 on the x-axis. 1 being the least fatigued and 9 being extremely fatigued. The frequencies are plotted on the y axis. The blue line is the data generated from the original biomathematical model. the orange line is the data scored by the neural network regression model (Fig 78). This model showed a much better improvement in predictions as compared to C1 which was the best of the baseline models. The histogram plot in Fig 79 is the differences between the original value and scored value, it follows the normal distribution with just over 250 duties out of 21000 duties were off by +/-1 KSS. The output of the model shows a mere 95.8% with the following evaluation metrics.

3.3.3.2 Neural network regression – C3 – Table 23

Same C2 model was used but validated against a new larger set of 242k records. this was to ensure for the prediction scaling to a larger variety of real-world duties. Life cycle of C3 (Fig 80) is slightly different that 242k records was added to the flow. In the previous run, the testing was done with 21k records and for this run, validation is at 242k records. The picture below you could see that the validation part is mentioned within a dotted circle, cleaned and scored against the same neural network regression model.

The original KSS values at duty end computed from the biomathematical model is compared against the neural network regression model for all of 21k test records (Fig 81). the data was aggregated into bins of 0.5 between 1 and 9 on the xaxis. 1 being the least fatigued and 9 being extremely fatigued. The frequencies are plotted on the y axis. The blue line is the data generated from the original biomathematical model. the orange line is the data scored by the neural network regression model. This model showed a much better improvement in predictions as compared to C1 which was the best of the baseline models.

The histogram plot in Fig 82 is the differences between the original value and scored value, it follows the normal distribution with just over 3.8k duties out of 242k duties were off by +/-1 KSS. This account to just 1.5% of the duties that were incorrectly predicted by +/-1 KSS. The output of the model shows a strong accuracy of 94.5% with the following evaluation metrics. Even thou the overall fit was slightly down from C2, the amount of duties and the varieties was 10 times (21k vs 242k records) more than what was tested in C2. The model showed much more stability with larger variety of dataset.

3.3.3.3 Neural network regression with 242k records – C4 – Table 24

For the Run C4, a dataset that was used for validation is used to build and train the model. The new larger dataset that was extracted using the similar exercise to C2 is used to train the model. the model contains 242k records and 194k records were used to train the model and 48k record were used to test. Similar to the baseline models and C2, the life cycle of C4 in Fig 83 is also the same except the model used much larger number of records. The original KSS values at duty end computed from the biomathematical model is compared against the neural network regression model for all of 48k test records. the data was aggregated into bins of 0.5 between 1 and 9 on the x-axis. 1 being the least fatigued and 9 being extremely fatigued.

The frequencies are plotted on the y axis. The blue line is the data generated from the original biomathematical model. the orange line is the data scored by the C4 neural network regression model (Fig 84). This model showed by far the best improvement in predictions as compared to all the baseline and previous C2 and C3 models.

The above histogram plot in Fig 85 is the differences between the original value and scored value, it follows the normal distribution with just 500 duties out of 48000 duties were off by +/-1 KSS. This is about 1% of the duties are off by +/- 1 KSS. The output of the model shows a mere 96.3% with the following evaluation metrics.

3.3.4 Results Comparison

The full comparison of all the models and its evaluation metrics are shown in table 25. Choosing the deep learning neural network from the baseline and adding more training data helped the model to improve dramatically to 96.3%. By running some additional parallel tests, this model can be taken to production and used.

3.3 Contributions and Future Work

In this dissertation, three open areas were focused. First chapter entirely focused on the basics of fatigue science, models and innovative ways to manage fatigue risk management in operations. Second chapter focused on innovating the new fatigue data collection taxonomy, quantification of fatigue and modeling the risk. The third chapter extends itself to a novel idea of predicting fatigue risk in operations without the use of bio mathematical model that requires work history to predict fatigue score. The paper discusses the use of combination of classification and deep learning machine learning techniques to predict the work schedule deformation, initial fatigue state of the duty and thereby predicting its fatigue risk of an end state of duty. The ML models are trained with biomathematical score outputs.

Key contributions made but not limited to

- Introduced the new model architecture to manage fatigue in complex logistic operations. Published a journal paper.
- Probability based fatigue risk prediction. Patent applied.
- Introduction of several innovative fatigue management processes. Published a paper in leading journal.
- Introducing the first of the kind data collection taxonomy to harmonize and tabulate data collection across the operation. The common taxonomy also allows airlines to share and understand safety data. Patent Approved 2020.
- Conducted a human physiology field study and analysis in complex airline operations. Paper presented at Fatigue Conference, Fremantle Australia
- Coding and Structure the new concept of Duty DNA
- Clustering duty periods with K-Means and time series clustering with duty sequences.

- Predicting crew duty deformation using classification machine learning models. To be published.
- Predicting fatigue variables to a duty end state using deep learning regression techniques. To be published.

Further work

The focus of this work is to lay the foundation and keep the scope limited to predicting an end state. The work involved several years of background research, process improvements and tool building. The modeling work involved applying the combination of clustering, classification and deep learning techniques to solve the prediction of fatigue during operational stress. While looking at duty deformation, the scope was kept only to duty delays. Also, when predicting the fatigue variables, the scope was limited to KSS values at the end of duty.

Future work should look into but not limited to

- Be able to predict further deformations like duty structural changes (dna sequence, multiple legs, origin delay, duty start changes, previous layover changes etc.).
- Apply elbow methods and determine the clusters for all features not just select features. Once clusters are formed, assignment outcomes should be predicted on a duty period. for example predicting a sick call on a duty DNA. Other applications could be likeability of a dutyDNA by a segment of crew.
- Bring additional features external to duty dna such as weather, maintenance, IT issues, cargo volumes, manpower planning, geopolitical sates etc.,
- Be able to predict the remaining 4 fatigue variables (sleep in duty, sleep in 24 before duty strt, sleep in 24 before duty end, kss at duty start, max kss within duty.

- Use AUC for determining delay thresholds in binary tree classifications. Rebalancing and Optimization should be considered for all the models.
- Be able to combine that with flight risk to predict the real risk on flight. Combinations of risk quantification and the predictions at different states.

List of References

Achermann, P. (2004). The two-process model of sleep regulation revisited. Aviat. Space Environ. Med., 75, A37–A43.

Agrawal, R., Imielinski, T. and Swami, A., (1993). Mining Association Rules Between Sets of Items in Large Databases. In proceedings of the 1993 ACM SIGMOD Int'l Conference on Management of Data. Washington, D.C., May 26-28. pp. 207-216.

Åkerstedt T (2003). Shift work and disturbed sleep/wakefulness. Occupational Medicine 53: 89-94.

Åkerstedt T and Folkard S (1995) Validation of the S and C components of the three-process model of alertness regulation. Sleep 18:1-6.

Akerstedt T, Gillberg M. Subjective and objective sleepiness in the active individual. Int J Neurosci (1990);52(1–2):29–37.

Åkerstedt T, Ingre M, Kecklund G, Folkard S, Axelsson J (2008). Accounting for partial sleep deprivation and cumulative sleepiness in the three-process model of alertness regulation. Chronobiology International 25: 309-319.

Akerstedt T, Kecklund G, Knutsson A (1991). Spectral analysis of sleep electro-encephalography in rotating three-shift work. Scand J Work Environ Health;17(5):330–6

Åkerstedt Torbjorn; Billiard Michel; Bonnet Michael; Ficca Gianluca; Garma Lucile; Mariotti Maurizio; Salzarulo Piero; Schulz Hartmut (2002). "Awakening from Sleep". Sleep Medicine Reviews. 6 (4): 267–286.

Altman, Naomi S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression". The American Statistician. 46 (3): 175–185.

Balkin TJ, Bliese PD, Belenky G (2004) Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. J Sleep Res;13(3):219–27

Ball, Michael & Barnhart, Cynthia & Dresner, Martin & Hansen, Mark & Neels, Kevin & Odoni, Amedeo & Peterson, Everett & Sherry, Lance & Trani, Antonio & Zou, Bo & Britto, Rodrigo & Fearing, Doug & Swaroop, Prem & Uman, Nitish & Vaze, Vikrant & Voltes, Augusto. (2010). Total Delay Impact Study: A Comprehensive Assessment of the Costs and Impacts of Flight Delay in the United States.

Bartsch, Butsri (2013). "Air freight - it could not be faster!". BB Handel. Archived from the original on 2013-06-07. Retrieved 29 May 2013.

Basner M, Dinges DF (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. Sleep;34(5):581.

Belenky G, Lamp A, Hemp A, Zaslona JL (2014). Fatigue in the workplace. In: Bianchi MT, editor. Sleep deprivation and disease: effects on the body, brain and behaviour. New York: Springer;

Belenky G, Van Dongen HPA (2009). Computer implemented scheduling systems and associated methods. Patent Application No. 12/253,703; Publication No. US 2009/0132332 A1.

Belenky G, Wesensten NJ, Thorne DR (2003) Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. J Sleep Res 2003;12(1):1–12.

Belenky, G., Hemp, A., Hoeg, L., Lamp, A., Edwards, A., & Gregory, K. (n.d.). Island Hopper – The Effect of Long-Duration, Multi-Segment Flights on Pilot Sleep and Performance. In FRMS Forum. United Airlines

Bell W. R., Hillmer S. C. (1984). Issues involved with the seasonal adjustment of time series. J. Bus. Econ. Stat. 2, 291–320. 10.2307/1391266

Belyavin AJ, Spencer MB (2004). Modelling performance and alertness: the QinetiQ approach. Aviation, Space, and Environmental Medicine 75: A93-A103.

Bermingham, Mairead L.; Pong-Wong, Ricardo; Spiliopoulou, Athina; Hayward, Caroline; Rudan, Igor; Campbell, Harry; Wright, Alan F.; Wilson, James F.; Agakov, Felix; Navarro, Pau; Haley, Chris S. (2015). "Application of high-dimensional feature selection: evaluation for genomic prediction in man".

Bernard S. Pasternack and Rosamond R. Gianutsos (1969), Application of the Exponential and Power Functions to the Study of Allometric Growth, with Particular Reference to Doryline Ants, The American Naturalist, Vol. 103, No. 931, pp. 225-234
Boeing (2018) http://www.boeing.com/commercial/market/cargo-forecast/
Boivin DB, Boudreau P (2014). Impacts of shift work on sleep and circadian rhythms. Pathologie Biologie. 62:292–301.
Borbély AA (1982). A two process model of sleep regulation. Human Neurobiology 1: 195-204.
Borbély AA, Achermann P (1999). Sleep homeostasis and models of sleep regulation. Journal of Biological Rhythms 14: 557-568.
Bradley, P. S. and Fayyad, U. M., 1998. Refining Initial Points for K-Means Clustering. In proceedings of the 15th Int'l Conference on Machine Learning. Madison, WI, July 24-27. pp. 91-99.
Braun V, Clarke V. Thematic Analysis. In: Cooper H, editor. APA Handbook of Research Methods in Psychology. 2. Research Designs. Washington DC: American Psychological Association; 2012. p. 57–71.
Breiman, L. (June 1997). "Arcing The Edge" (PDF). Technical Report 486. Statistics Department, University of California, Berkeley.
Cai, Eric (2014-03-20). "Machine Learning Lesson of the Day – Overfitting and Underfitting". StatBlogs. Archived from the original on 2016-12-29. Retrieved 2016-12-29.
Caldwell JA (2005). Fatigue in aviation. Travel Medicine and Infectious Disease 3: 85-96.
Caldwell JA, Mallis MM, Caldwell JL, Paul MA, Miller JC, Neri DF (2009). Fatigue countermeasures in aviation. Aviation, Space, and Environmental Medicine 80: 29-59.
Carrier J, Monk TH. Circadian rhythms in performance: new trends. Chronobiol Int 2000;17:719–732.
CASA, 2014. Biomathematical fatigue models guidance. https://www.casa.gov.au/files/biomathematical-fatigue-models-guidance
Cho YW, Lee JH, Son HK, Lee SH, Shin C, Johns MW. The reliability and validity of the Korean version of the Epworth Sleepiness Scale. Sleep Breath, 2011; 15: 377-384
Costa, G., Cesna, G., Kogi, K. and Wedderburn, A (Eds.) (1990) Shiftwork: health, sleep and performance. Frankfurt am Main: Verlag Peter Lang.
Cotofrei, P., 2002. Statistical Temporal Rules. In proceedings of the 15th Conference on Computational Statistics - Short Communications and Posters. Berlin, Germany, Aug 24-28
Cowpertwait P. S., Metcalfe A. (2009). Introductory Time Series with R. New York, NY: Springer-Verlag.
Daan S, Beersma DGM, Borbély AA (1984). Timing of human sleep: recovery process gated by a circadian pacemaker. American Journal of Physiology 246: R161-R178.
Darwent D, Dawson D, Roach GD (2010). Prediction of probabilistic sleep distributions following travel across multiple time zones. Sleep 33: 185-195.
David A. Freedman (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 26.
Dawson D, Noy YI, Härmä M, Åkerstedt T, Belenky G . Modelling fatigue and the use of fatigue models in work settings . Accid Anal Prev 2011 ; 43 : 549 – 64 .
Dijk DJ, Czeisler CA (1994). Paradoxical timing of the circadian rhythm of sleep propensity serves to consolidate sleep and wakefulness in humans. Neuroscience Letters 166: 63-68.
Dinges DF, Powell JW. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behav Res Methods Instrum Comput 1985;17(6):652–5
Elith, Jane (2008). "A working guide to boosted regression trees". Journal of Animal Ecology. 77 (4): 802–813. doi:10.1111/j.1365-2656.2008.01390.x. PMID 18397250.
Federal Aviation Administration. 14 CFR, Parts 117, 119, and 121. Flightcrew member duty and rest requirements. Docket no. FAA-2009-1093; Amendment nos. 117-1, 119-16, 121-357; RIN 2120-AJ58. Washington, DC: Federal Aviation Administration; 2011 .
Ferri R., Manconi M., Plazzi G., Bruni O., Vandi S., Montagna P., Zucconi M. (2008). "A quantitative statistical analysis of the submentalis muscle EMG amplitude during sleep in normal

controls and patients with REM sleep behavior disorder". Journal of Sleep Research. 17 (1): 89–100
Feyer, A.M. & Williamson, A.M. (1998): Human factors in accident modelling. In: Stellman, J.M. (Ed.), Encyclopaedia of Occupational Health and Safety, Fourth Edition. Geneva: International Labour Organisation.
Folkard S, Åkerstedt T, Macdonald I, Tucker P, Spencer M. Beyond the three-process model of alertness: estimating phase, time on shift and successive night effects. J Biol Rhythms 1999;14:577–587.
Folkard S, Akerstedt T. A three process model of the regulation of alertness and sleepiness. In: Ogilvie R, Broughton R, eds. Sleep. arousal and performance: problems and promises. Boston: Birkhiiuser, 1991:11-26.
Friedman, Jerome (2003). "Multiple Additive Regression Trees with Application in Epidemiology". Statistics in Medicine. 22 (9): 1365–1381.
Gander P, Hartley L, Powell D, Cabon P, Hitchcock E, et al. Fatigue risk management: organizational factors at the regulatory and industry/company level . Accid Anal Prev 2011 ; 43 : 573 – 90
Gander P, van den Berg M, Mulrine H, et al. Circadian adaptation of airline pilots during extended duration operations between the USA and Asia. Chronobiol Int 2013;30(8):963–72.
Gander PH, Signal TL, van den Berg MJ, et al. In-flight sleep, pilot fatigue and Psychomotor Vigilance Task performance on ultra-long range versus long range flights. J Sleep Res 2013;22(6):697–706.
Glantz, Stanton A.; Slinker, B. K. (1990). Primer of Applied Regression and Analysis of Variance. McGraw-Hill. ISBN 978-0-07-023407-9.
Goldin D.Q , P.C. Kanellakis. On Similarity Queries for Time-Series Data: Constraint Specification and Implementation. 1st Int'l Conf. on the Principles and Practice of Constraint Programming, LNCS 976, pp. 137-153, Cassis France, Sep.1995.
Good and Hardin (2006) Common Errors in Statistics, pp. 28 - 29.
Goode, Jeffrey H. (March 27, 2003). "Are pilots at risk of accidents due to fatigue?". Journal of Safety Research. 34 (3): 309–313
Gottselig JM, Adam M, Rétey JV, Khatami R, Achermann P, Landolt HP (March 2006). "Random number generation during sleep deprivation: effects of caffeine on response maintenance and stereotypy". Journal of Sleep Research. 15 (1): 31–40.
Graeber RC, Lauber JK, Connell LJ, Gander PH (1986). International aircrew sleep and wakefulness after multiple time zone flights: a cooperative study. Aviat Space Environ Med;57(12 Pt 2):B3–9.27.
Granitto, P. M., Furlanello, C., Biasioli, F., and Gasperi, F (2006) “Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products,” Chemometrics and Intelligent Laboratory Systems, Vol. 83, No. 2, 2006, pp. 83–90.
Grossi, Davide, Frank Dignum and John-Jules Charles Meyer. (2005). "Contextual Taxonomies" in Computational Logic in Multi-Agent Systems, pp. 33-51
Guyon I.; Weston J.; Barnhill S.; Vapnik V. (2002). "Gene selection for cancer classification using support vector machines". Machine Learning. 46 (1–3): 389–422. doi:10.1023/A:1012487302797
Guyon, Isabelle; Elisseeff, André (2003). "An Introduction to Variable and Feature Selection". JMLR.
H. He and E. A. Garcia, “Learning from imbalanced data,” IEEE Trans. Knowl. Data Eng., vol. 21, no. 9, pp. 1263–1284, Sep. 2009
Halkidi, M., Vazirgiannis, M., and Batistakis, I. (2000). Quality Scheme Assessment in the Clustering Process. In Proceedings of PKDD, Lyon, France.
Hamming, R. W. (April 1950). "Error detecting and error correcting codes" (PDF). The Bell System Technical Journal. 29 (2): 147–160.
Hand. D, H. Mannila, And P. Smyth, Principles Of Data Mining, Mit Press, 2001

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. In Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting (pp. 904-908). Santa Monica, CA: Human Fac-tors & Ergonomics Society
Hertenstein E, Gabryelska A, Spiegelhalder K, Nissen C, Johann AF, Umarova R, Riemann D, Baglioni C, Feige B. Reference data for polysomnography-measured and subjective sleep in healthy adults. J Clin Sleep Med. 2018;14(4):523–532.
Ho P, Landsberger S, Signal L, et al. The Singapore experience: task force studies scientifc data to assess flight. Flight Saf Dig 2005;26:20–40
Ho TK (1998). "The Random Subspace Method for Constructing Decision Forests" (PDF). IEEE Transactions on Pattern Analysis and Machine Intelligence. 20 (8): 832–844. doi:10.1109/34.709601.
Ho, Tin Kam (1995). Random Decision Forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282.
Honn KA, Van Dongen HPA, Dawson D. 2019. Working Time Society consensus statements: prescriptive rule sets and risk management-based approaches for the management of fatiguerelated risk in working time arrangements. Ind Health. 57:264–280.
http://patft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=9540118.PN.&OS=PN/9540118&RS=PN/9540118
Hursh SR . U.S. DOT/FAA – Part 117 SAFTE FAST scientifi ctechnical report. Docket no. FAA-2009-1093-2483; RIN 2120-AJ58. Washington, DC: Federal Aviation Administration ;2011
Hursh SR, Balkin TJ, Van Dongen HPA. 2016. Sleep and performance prediction modeling. In Kryger MH, Roth T, Dement WC, eds. Principles and Practice of Sleep Medicine. 6th edition. Philadelphia: Elsevier, p. 689–696.
Hursh SR, Redmond DP, Johnson ML, Thorne DR, Belenky G,et al. Fatigue models for applied research in warfi ghting . Aviat Space Environ Med 2004 ; 75 : A44 – 53 .
Hursh SR, Van Dongen HPA (2010). Fatigue and performance modeling. In Kryger MH, Roth T, Dement WC (Eds.), Principles and Practice of Sleep Medicine (5th ed.). Elsevier Saunders, Philadelphia: 745-752.
Hyndman R. J., Athanasopoulos G. (2014). Forecasting: Principles and Practice.
Hyndman, R. and Koehler A. (2005). "Another look at measures of forecast accuracy"
IATA 2020 https://airlines.iata.org/news/the-impact-of-covid-19-on-aviation
IATA, ICAO, IFALPA. 2011. Fatigue Risk Management Systems Implementation Guide for Operators, 1st edition.
International Civil Aviation Organization. Fatigue risk management systems: manual for regulators. Montreal: International Civil Aviation Organization; 2012 [Doc 9966]
International Civil Aviation Organisation. Fatigue Management Guide for Airline Operators, 2nd edition Montreal, Canada: International Civil Aviation Organisation Montreal, Canada: International Civil Aviation Organisation; 2015
J. Han, M. Kamber, And A. K. H. Tung, Spatial Clustering Methods in Data Mining: A Survey, Taylor and Francis, 1 ed., December 2001, ch. 8, pp. 188–217.
J. Macqueen, Some Methods For Classification And Analysis Of Multivariate Observations, In Proc. Of The Fifth Berkeley Symposium On Mathematical Statistics And Probability, 1967, Pp. 281–297
Jackson ML, Van Dongen HPA (2010). Cognitive effects of sleepiness. In Thorpy MJ, Billiard M (Eds.), Sleepiness: Causes, Consequences and Treatment. Cambridge University Press, Cambridge, in press.
Johnson ML, Belenky G, Redmond DP, Thorne DR, Williams JD, Hursh SR, Balkin TJ. Modulating the homeostatic process to predict performance during chronic sleep restriction. Aviat. Space Environ. Med. 2004;75:A141-A146.
Kaida M, Takahashi T, Åkerstedt A, Nakata Y, Otsuka T, Haratani K, et al. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. Clinical Neurophysiology, 117, 1574–81.

Kandelaars KJ, Fletcher A, Eitzen GE, Roach GD, Dawson D (2006). Layover sleep prediction for cockpit crews during transmeridian flight patterns. Aviation, Space, and Environmental Medicine 77: 145-150.
Kecklund G, Akerstedt T. (1993). Sleepiness in long distance truck driving: an ambulatory EEG study of night driving. Ergonomics, 36, 1007–17.
Keogh, E., Lin, J., Truppel, W. Clustering of Time Series is meaningless. In Proc.IEEE Conf. on Data Mining, IEEE Computer Society (2003) 115-122
Kronauer RE, Gunzelmann G, Van Dongen HPA, Doyle III FJ, Klerman EB (2007). Uncovering physiologic mechanisms of circadian rhythms and sleep/wake regulation through mathematical modeling. Journal of Biological Rhythms 22: 233-245.
Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ (2008). "Self-Reported and Measured Sleep Duration: How Similar Are They?". Epidemiology. 19 (6): 838–845.
Levenshtein, Vladimir I. (February 1966). "Binary codes capable of correcting deletions, insertions, and reversals". Soviet Physics Doklady. 10 (8): 707–710
Li, Yaohua & Tan, Na. (2013). Study on Fleet Assignment Problem Model and Algorithm. Mathematical Problems in Engineering. 2013. 10.1155/2013/581586.
M. R. Anderberg, Cluster Analysis For Applications, Academic Press, Inc., London, 1973.
M. S. Aldenderfer and R. K. Blashfield, Cluster Analysis, Sage Publications, London, England, 1984
Mallis MM, Mejdal S, Nguyen TT, Dinges DF (2004). Summary of the key features of seven biomathematical models of human fatigue and performance. Aviation, Space, and Environmental Medicine 75: A4-A14.
Mantegna, R. N., 1999. Hierarchical Structure in Financial Markets. European Physical Journal, vol. B11. pp. 193-197.
Marino M, Li Y, Rueschman MN, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep 2013; 36:1747
Mascha EJ, Sessler DI. Equivalence and noninferiority testing in regression models and repeated-measures designs. Anesth Analg. 2011 Mar;112(3):678-87. doi: 10.1213/ANE.0b013e318206f872. Epub 2011 Feb 8. PMID: 21304155.
Matousek M: Correlation between EEG and psychological variables in volunteers. Res Commun Psychol Psychiatr Behav. 1988, 13: 77-82.
McCarley Robert W. (2007). "Neurobiology of REM and NREM sleep". Sleep Medicine. 8 (4): 302–330.
McCauley P, Kalachev LV, Smith AD, Belenky G, Dinges DF, Van Dongen HPA. A new mathematical model for the homeostatic effects of sleep loss on neurobehavioral performance. J. Theor. Biol. 2009;256:227-239.
McCleary R., Hay R. A., Meidinger E. E., McDowall D. (1980). Applied Time Series Analysis for the Social Sciences. Beverly Hills, CA: Sage
Michielsen, H. J., De Vries, J., & Van Heck, G. L. (2003). Psychometric qualities of a brief self-rated fatigue measure the fatigue assessment scale. Journal of Psychosomatic Research, 54, 345–352
Mitchell, Tom (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892.
Petrilli RM, Thgomas MJ, Dawson D, Roach GD. The decision-making of commercial airline crews following an international pattern. Proceedings of the seventh international Australian Aviation Psychology Association symposium. Manly (N.S.W., Australia): 2006
Philip P, Åkerstedt T (2006). Transport and industrial safety, how are they affected by sleepiness and sleep restriction? Sleep Medicine Reviews 10: 347-356.
Powell, D., Spencer, M. B., Holland, D., & Petrie, K. J. (2008). Fatigue in Two-Pilot Operations: Implications for Flight and Duty Time Limitations. Aviation, Space, and Environmental Medicine, 79(11), 1047–1050. doi:10.3357/ASEM.2362.2008
Practice parameters for the use of actigraphy in the clinical assessment of sleep disorders. American Sleep Disorders Association. Sleep. 1995 May; 18(4):285-7.

R. O. Duda, P. E. Hart, And D. G. Stork, Pattern Classification, John Wiley & Sons, Inc., 2001
Rangan S, Bowman JL, Hauser WJ, McDonald WW, Lewis RA, Van Dongen HPA. 2013. Integrated fatigue modeling in crew rostering and operations. Cnd Aeron Space J. 59:1–6.
Rangan S, Van Dongen HPA. Quantifying fatigue risk in model-based fatigue risk management. Aviat Space Environ Med 2013;84:155-7
Rattray, J., and Jones, M.C. (2007): Essential elements of questionnaire design and development. Journal of Clinical Nursing 16: 234 – 243.
Reilly T, Waterhouse J, Edwards B (2005). Jet lag and air travel: implications for performance. Clinics in Sports Medicine 24: 367-380.
Reiter RJ. The melatonin rhythm: Both a clock and a calendar. Experientia 1993;49:654-64.
Roach GD, Dawson D, Lamond N. Can a shorter psychomotor vigilance task be usedas a reasonable substitute for the ten-minute psychomotor vigilance task? Chronobiol Int 2006;23(6):1379–87.
Robert L. Thorndike (December 1953). "Who Belongs in the Family?". Psychometrika. 18 (4): 267–276
Roberts, D. E.; Nesthus, T. E. FRMP and FRMS Overview, 2011. MegaSlides. http://megaslides.com/doc/3717151/frmp-and-frms-overview (accessed June 07, 2016).
Roddick, J. F. and Spiliopoulou, M., 2002. A Survey of Temporal Knowledge Discovery Paradigms and Methods. Transactions on Data Engineering, vol. 14 (4). pp. 750-767
Roenneberga Till; Kuehnlea Tim; Judaa Myriam; Kantermanna Thomas; Allebrandta Karla; Gordijnb Marijke; Merrow Martha (2007). "Epidemiology of the human circadian clock" (PDF). Sleep Medicine Reviews. 11 (6): 429–438.
Rokach, Lior; Maimon, O. (2008). Data mining with decision trees: theory and applications. World Scientific Pub Co Inc. ISBN 978-9812771711.
Romig E, Klemets T (2009). Fatigue risk management in flight crew scheduling. Aviation, Space, and Environmental Medicine 80: 1073-1074.
Rosekind MR, Boyd JN, Gregory KB, Glotzbach SF, Blank RC (2002). Alertness management in 24/7 settings: lessons from aviation. Occupational Medicine: State of the Art Reviews 17: 247-259.
Rosekind, M., Gregory, K., Co E., Miller, L. and Dinges, D. (2000): Crew Factors in Flight Operations XII: S Survey of Sleep Quantity and Quality in On-Board Rest Facilities. NASA Ames Research Center. Moffett Field, CA. NASA/TM – 2000-209611
Ruggiero JS, Redeker NS. 2014. Effects of napping on sleepiness and sleep-related performance deficits in night-shift workers: a systematic review. Biol Res Nurs. 16:134–142.
Samel A, Wegmann HM, Vejvoda M. Aircrew fatigue in long-haul operations. Accid Anal Prev 1997;29(4):439–52.
Samn SW, Perelli LP. Estimating aircrew fatigue: a technique with application to airlift operations. Brooks Air Force Base. San Anonio (Tex.): USAF School of Aerospace Medicine; 1982 <www.dtic.mil/dtic>; DTIC document.
Schittenkopf, C., Tino, P. and Dorffner, G., 2000. The Benefit of Information Reduction for Trading Strategies. Report Series for Adaptive Information Systems and Management in Economics and Management Society. July. Report# 45.
Schmidhuber, J. (2015). "Deep Learning in Neural Networks: An Overview". Neural Networks. 61: 85–117.
Shawn A. Pruchnicki, Lora J. Wu, Gregory Belenky (2011). An exploration of the utility of mathematical modeling predicting fatigue from sleep/wake history and circadian phase applied in accident analysis and prevention: The crash of Comair Flight 5191. Accident Analysis & Prevention
Signal TL, Gale J, Gander PH (2005) Sleep measurement in flight crew: comparing actigraphic and subjective estimates to polysomnography. Aviat Space Environ Med;76(11):1058–63
Silva EJ, Wang W, Ronda JM (2010). Circadian and wake-dependent influences on subjective sleepiness, cognitive throughput, and reaction time performance in older and young adults. Sleep ;33(4): 481–90

Smith MR, Eastman CI (2012). Shift work: health, performance and safety problems, traditional countermeasures, and innovative management strategies to reduce circadian misalignment. Nat Sci Sleep. Sep 27;4:111-32. doi: 10.2147/NSS.S10372. PMID: 23620685; PMCID: PMC3630978.
Smith, P.A., Wright, B.M., Mackay, R.W., Milsop, H.W. and Yates, S.C. (1998) Results from a study of participative roster design: a change from slowly rotating 8-hour shifts to rapidly rotating 8- and 12-hour shifts. Scandinavian Journal of Work, Environment and Health, 24, 3, 55-61.
Socha, V., Schlenker, J., Kalavksý, P., Kutílek, P., Socha, L., Szabo, S., & Smrčka, P. (2015). Effect of the change of flight, navigation and motor data visualization on psychophysiological state of pilots. 339-344. doi:10.1109/SAMI.2015.7061900
Spencer MB (1987). The influence of irregularity of rest and activity on performance: a model based on time since sleep and time of day. Ergonomics 1987;30:1275–1286.
Spiegel K, Leproult R, Van Cauter E (October 1999). "Impact of sleep debt on metabolic and endocrine function". Lancet. 354 (9188): 1435–9.
St. Hilaire MA, Klerman EB, Khalsa SBS, Wright Jr KP, Czeisler CA, Kronauer RE (2007). Addition of a non-photic component to the light-based mathematical model of the human circadian pacemaker. Journal of Theoretical Biology 247: 583-599.
Stehman, Stephen V. (1997). "Selecting and interpreting measures of thematic classification accuracy". Remote Sensing of Environment. 62 (1): 77–89.
Stone KL, Ancoli-Israel S. Actigraphy. In: Kryger MH, Roth T, Dement WC (2011). Principles and practice of sleep medicine. 5th ed. St. Louis (Mo.): Saunders; 2011
Stuart J. Russell, Peter Norvig (2010) Artificial Intelligence: A Modern Approach, Third Edition, Prentice Hall ISBN 9780136042594.
Tassi, P. and Muzet, A. (2000). Sleep inertia. Sleep Med. Rev., 4 (4), 341–353
Thiel, H. (1966), Applied Economic Forecasting. Chicago: Rand McNally.
Thorne DR, Johnson DE, Redmond DP (2005) The Walter Reed palmheld psychomotor vigilance test. Behav Res Methods 2005;37(1)
Tino, P., Schittenkopf, C. and Dorffner, G., 2000. Temporal Pattern Recognition in Noisy Non-stationary Time Series Based on Quantization into Symbolic Streams: Lessons Learned from Financial Volatility Trading. Report Series for Adaptive Information Systems and Management in Economics and Management Sciences. July. Report# 46.
Van Dongen HP, Maislin G, Mullington JM, Dinges DF (2003). The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep;26(2):117–26
Van Dongen HPA (2004). Comparison of mathematical model predictions to experimental data of fatigue and performance. Aviation, Space, and Environmental Medicine 75: A15-A36.
Van Dongen HPA (2010). Predicting sleep/wake behavior for model-based fatigue risk management. Sleep 33: 144-145.
Van Dongen HPA, Balkin TJ, Hursh SR. (2016). Performance deficits during sleep loss and their operational consequences. In Kryger MH, Roth T, Dement WC. eds. Principles and Practice of Sleep Medicine, 6th edition. Philadelphia: Elsevier, p. 682–688. 11
Van Dongen HPA, Belenky G, Krueger JM (2010). Investigating the temporal dynamics and underlying mechanisms of cognitive fatigue. In Ackerman PL (Ed.), Cognitive Fatigue. American Psychological Association, Washington, D.C.: 127-147.
Van Dongen HPA, Belenky G. 2012. Model-based fatigue risk management. In Matthews G, Desmond PA, Neubauer C, Hancock PA, eds. The Handbook of Operator Fatigue. Farnham, UK: Ashgate, p. 487–506.
Van Dongen HPA, Dinges DF (2005). Sleep, circadian rhythms, and psychomotor vigilance performance. Clinics in Sports Medicine 24: 237-249.
Van Dongen HPA, Hans & M Vitellaro, Kristen & Dinges, David. (2005). Individual Differences in Adult Human Sleep and Wakefulness: Leitmotif for a Research Agenda. Sleep. 28. 479-96. 10.1093/sleep/28.4.479

Van Dongen HPA, Hursh SR (2010). Fatigue, performance, errors and accidents. In Kryger MH, Roth T, Dement WC (Eds.), Principles and Practice of Sleep Medicine (5th ed.). Elsevier Saunders, Philadelphia: 753-759.
Van Dongen HPA, Maislin G, Mullington JM, Dinges DF (2003). The cumu¬lative cost of additional wakefulness: dose-response effects on neurobe¬havioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep;26:117-126.
Van Dongen HPA, Mott CG, Huang JK, Mollicone DJ, McKenzie FD, Dinges DF (2007). Optimization of biomathematical model predictions for cognitive performance impairment in individuals: Accounting for unknown traits and uncertain states in homeostatic and circadian processes. Sleep 30: 1129-1143.
Van Dongen, H. P. A., & Hursh, S. R. (2011). Chapter 67 - Fatigue, Performance, Errors, and Accidents. In M. H. K. R. C. B. T.-P. and P. of S. M. (Fifth E. Dement (Ed.), (pp. 753– 759). Philadelphia: W.B. Saunders. doi:http://dx.doi.org/10.1016/B978-1-4160-6645- 3.00067-0
Vitaterna MH, Takahashi JS, Turek FW (2001). "Overview of circadian rhythms". Alcohol Research & Health. 25 (2): 85–93. PMID 11584554.
Vitellaro KM, Levin AA, Ballas C, Dinges DF, Van Dongen HPA (2003). Neurobehavioral performance under varying workload conditions during repeated exposure to sleep deprivation. In Van Bemmel AL, Beersma DGM, Hofman WF, Ruigt GSF, Vos PJE (Eds.), Sleep/Wake Research in the Netherlands (vol. 14). Dutch Society for Sleep-Wake Research: 106-109.
Wainer, Gessaroli, and Verdi (2006). Finding What Is Not There through the Unfortunate Binning of Results: The Mendel Effect, Chance Magazine, Vol 19, No.1, pp. 49 -52
Walker, E., & Nowacki, A. S. (2011). Understanding equivalence and noninferiority testing. Journal of general internal medicine, 26(2), 192–196. https://doi.org/10.1007/s11606-010-1513-8
Waterhouse JM, DeCoursey PJ (2003). Human circadian organization. In Dunlap JC, Loros JJ, DeCoursey PJ (Eds.), Chronobiology: Biological Timekeeping. Sinauer, Sunderland: 291-323.
Wright KP Jr, Hull JT, Czeisler CA (2002). Relationship between alertness, performance, and body temperature in humans. Am J Physiol Regul Integr Comp Physiol;283(6):R1370–7

Appendix - Figures

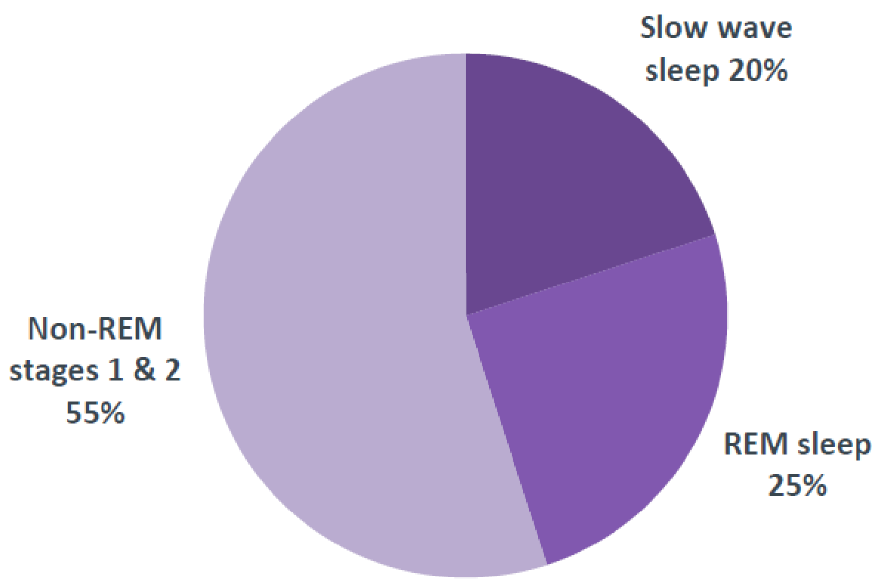


Figure 1. Types of Sleep.

Proportion of the night spent in each types of sleep, for a young adult

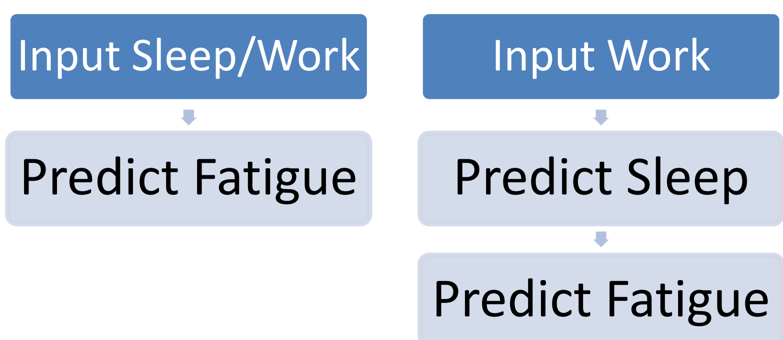


Figure 2. Types of biomathematical models.

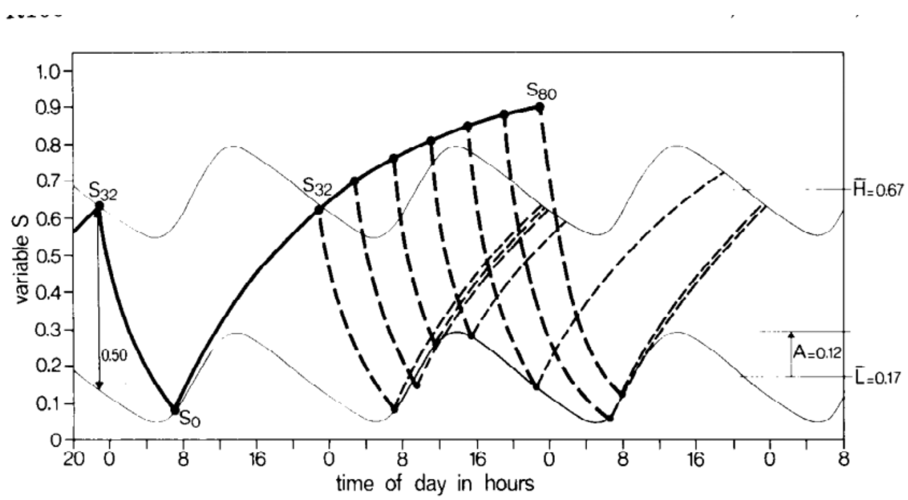


Figure 3. Evaluation of circadian waveform of wake threshold.

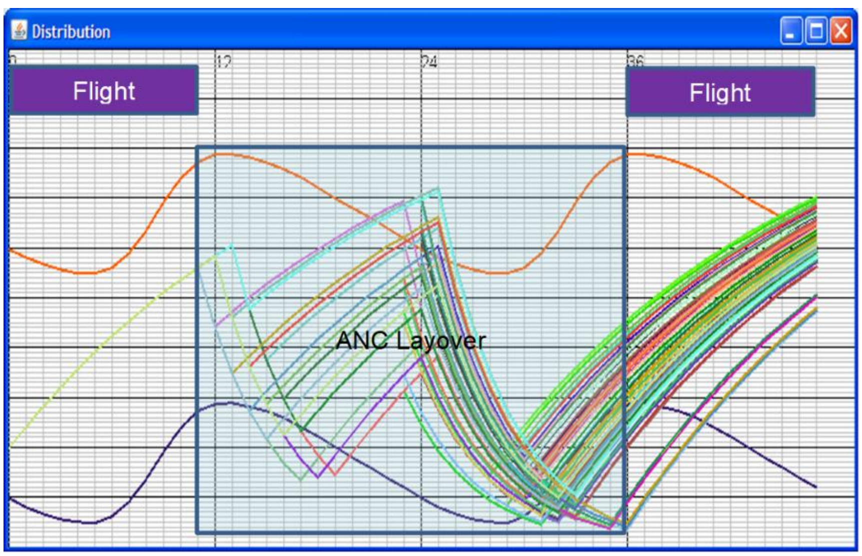


Figure 4. Biphasic homeostatic sleep curves with circadian waveform

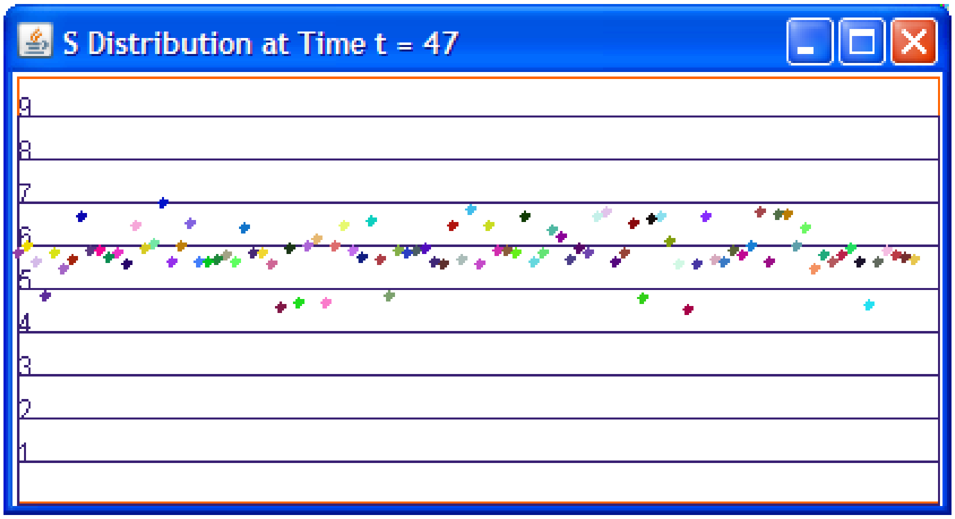


Figure 5. Shows distribution of S value at t=48

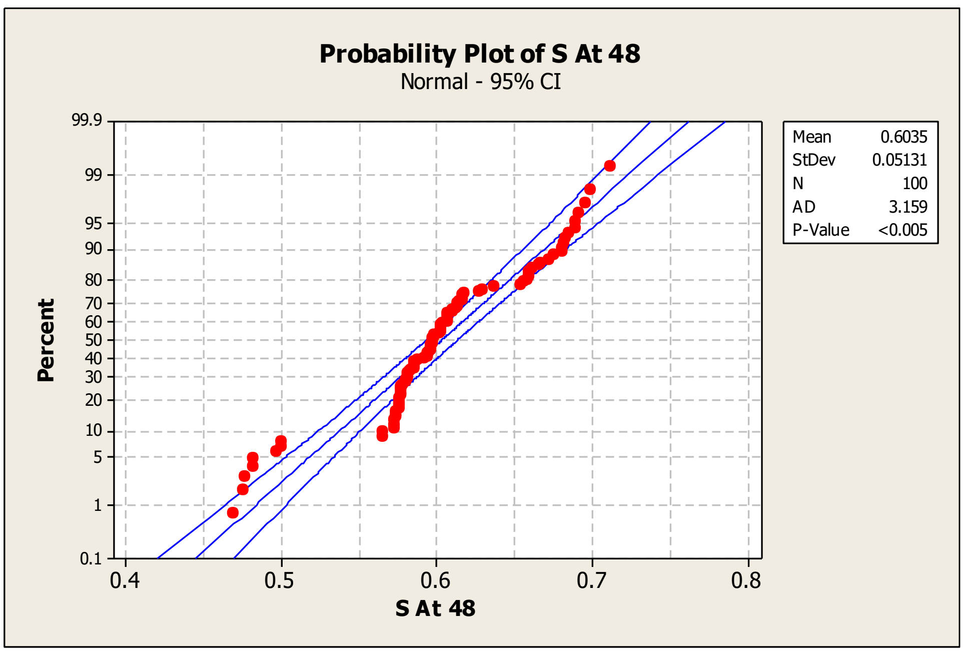


Figure 6: Probability plot of S Values at 12hrs after duty following layover.

μ of 0.6035 and σ=0.05131

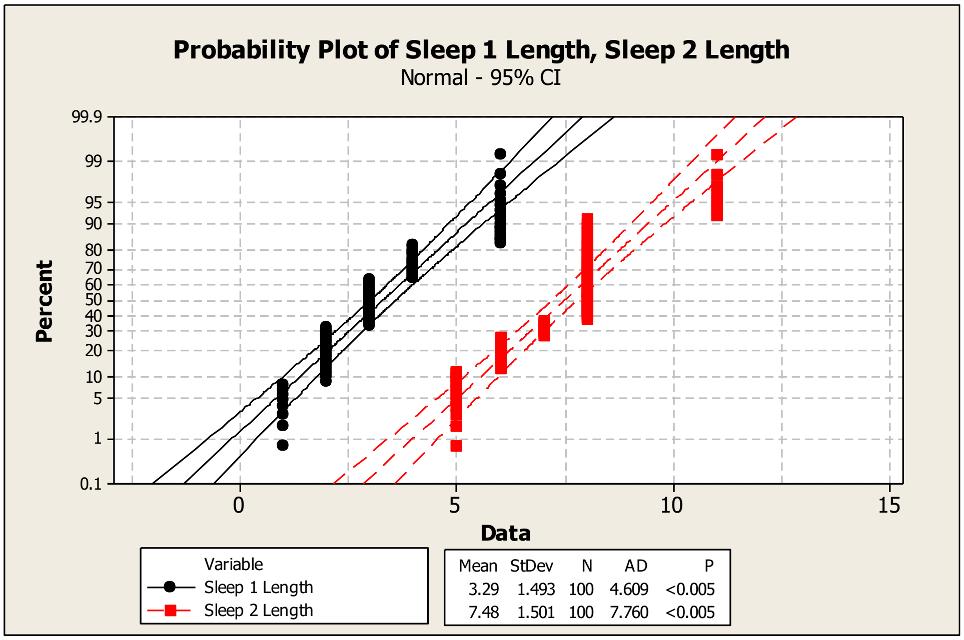


Figure 7: Probability plots of both the sleep period in the sample layover.

A short long sleep pattern is observed during the layover

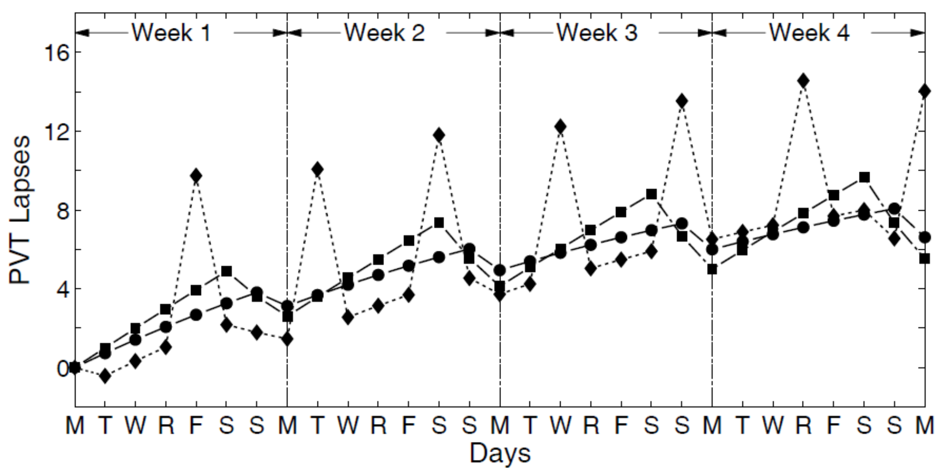


Figure 8: McCauley Model for the homeostatic effects of sleep loss

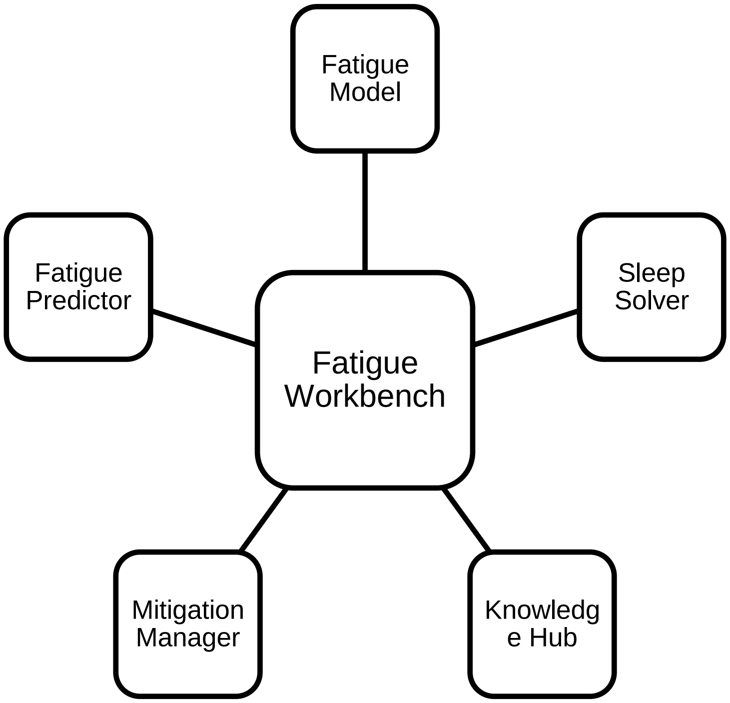


Figure 9: Proposed Fatigue Workbench for Airline

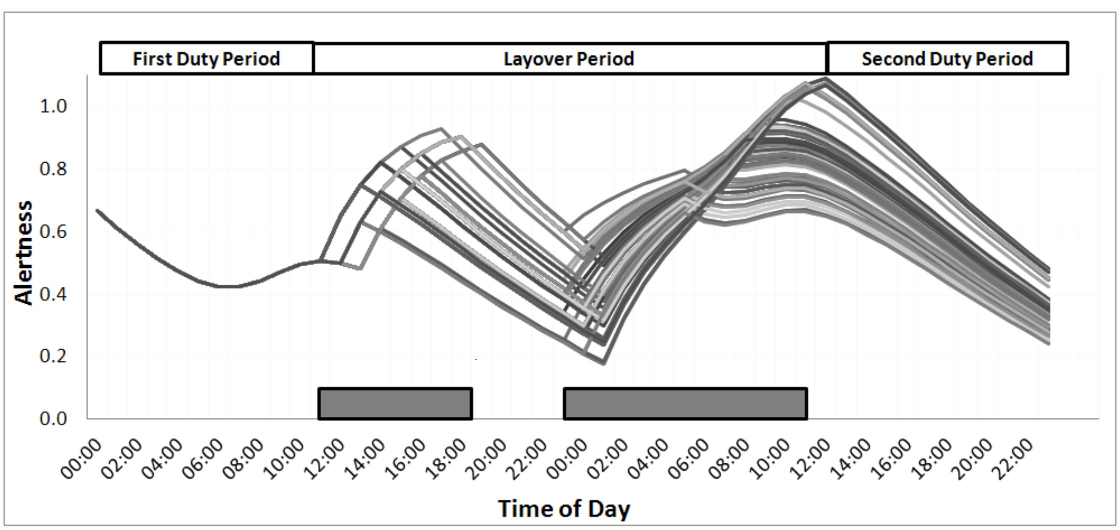


Figure 10: Alertness predictions for a simulated scheduling scenario

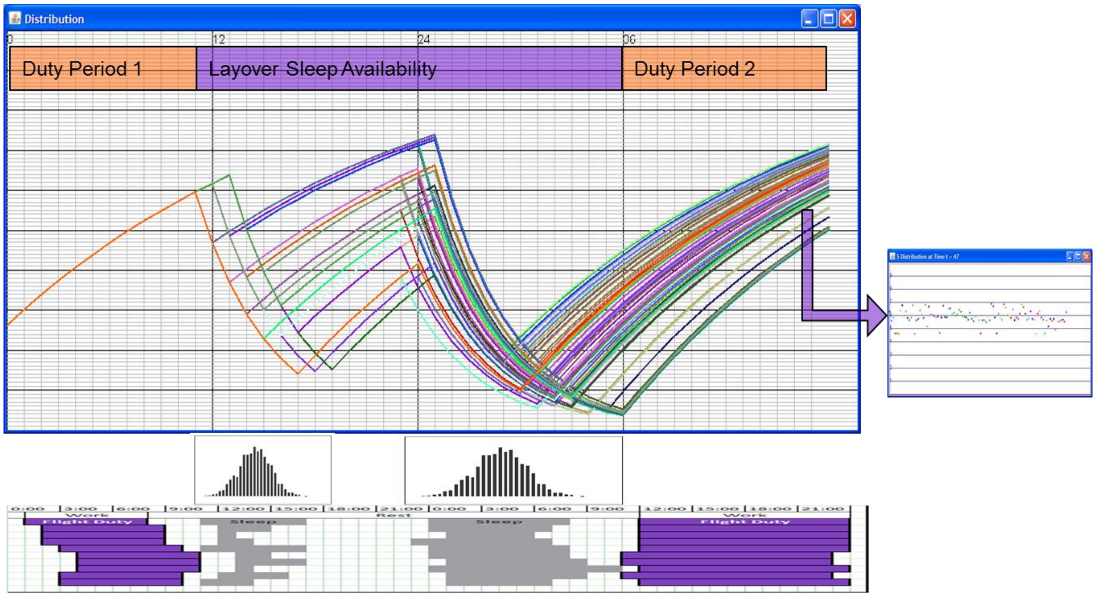


Figure 11: Sleep Pressure Curves generated using the two-process model

(top) Sleep Pressure Curves generated using the two-process model (AA. Borbély et al., 1999) based on actual sleep data.
(bottom) The horizonyal grey bars at the bottom represent sleep data captured from wristwatch actigraphy, horizontal purple bars represents length of work segment. Each row represents an individual subject data for two work periods.

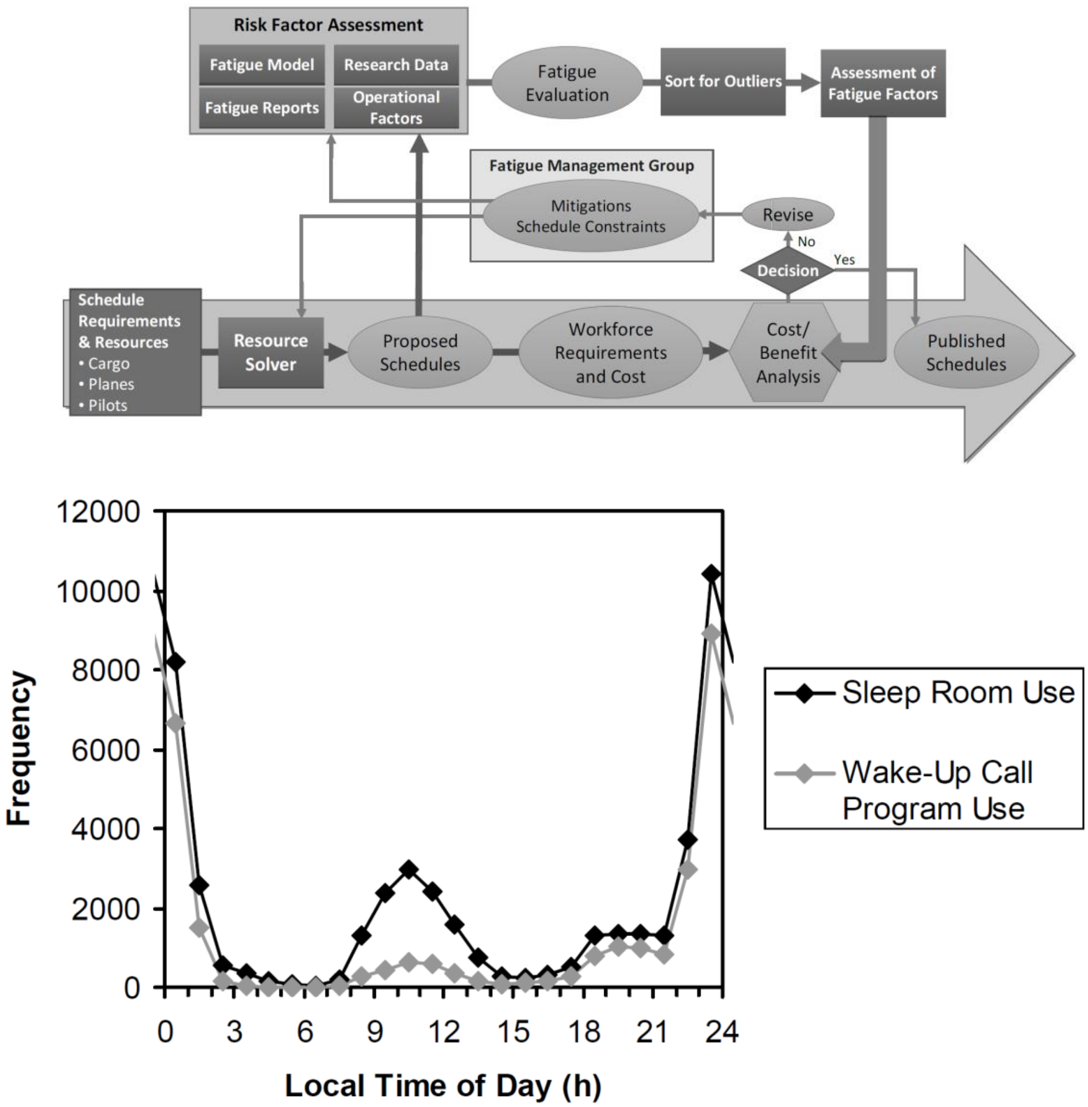


Figure 12: Predictive duty fatigue management and sleep room use

Predictive duty schedule fatigue management (top) and proactive sleep opportunity management (bottom) in a US-based cargo flight operation. The bottom graph shows data from four bases over a 6-month period in 2019.

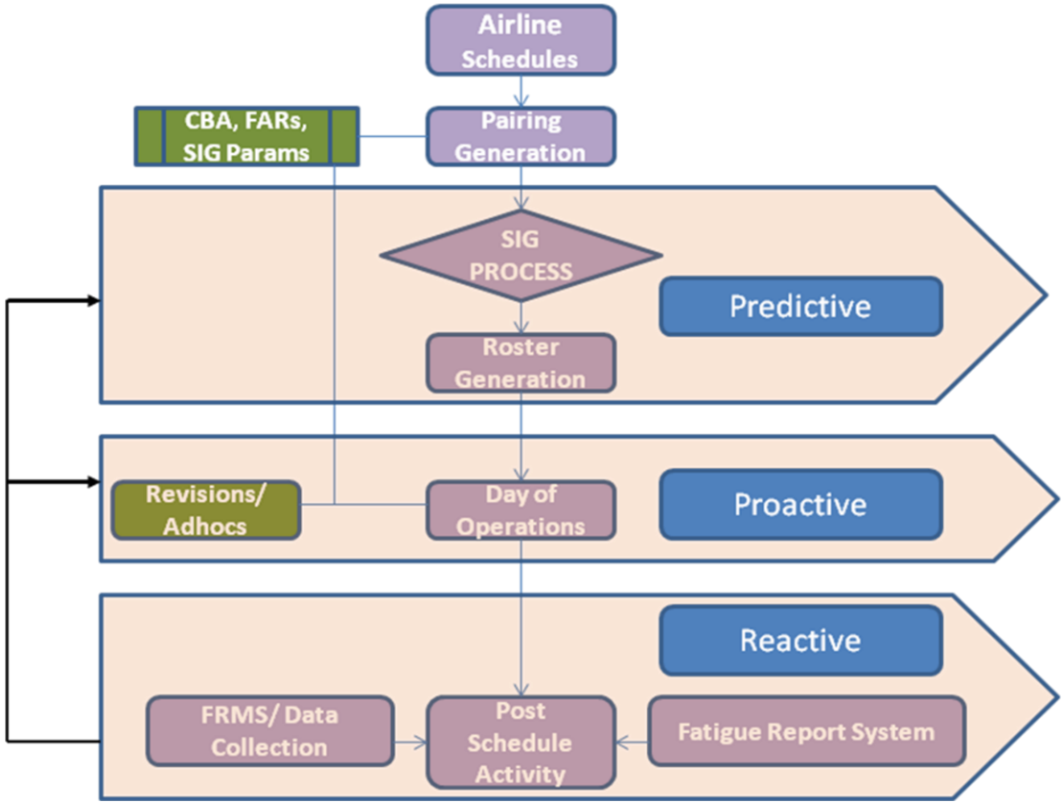


Figure 13: Fatigue risk identification processes.

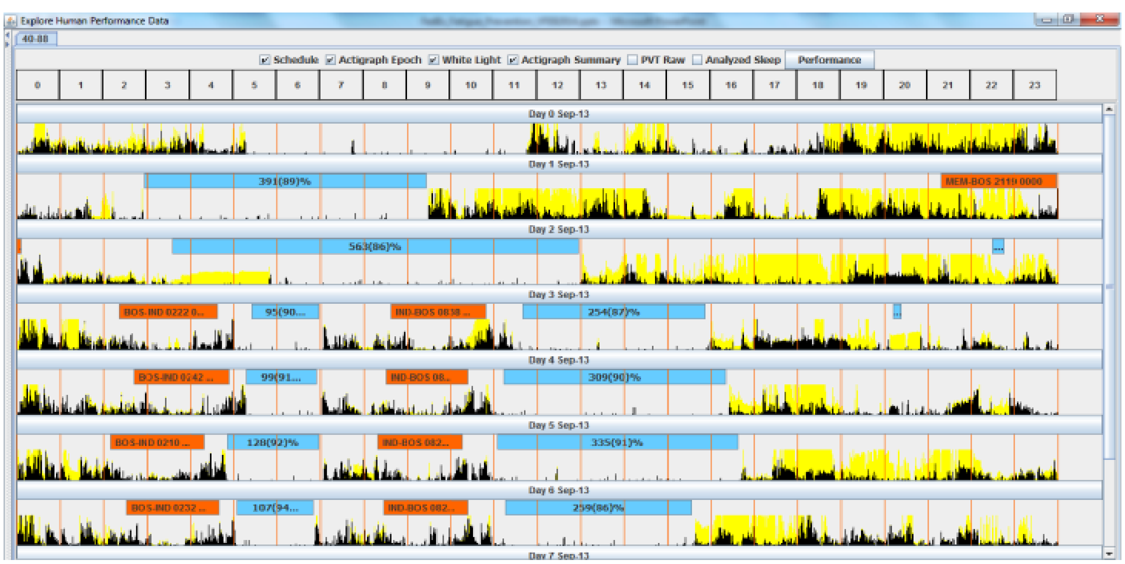


Figure 14: Sleep Work Visualization – Airline's software program

1 = extremely alert
2
3 = alert
4
5 = neither sleepy nor alert
6
7 = sleepy, but no difficulty remaining awake
8
9 = extremely sleepy, fighting sleep

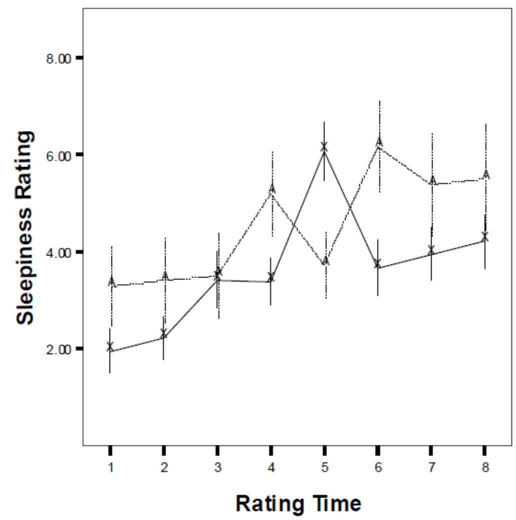


Figure 15: KSS Sleepiness Ratings and Sample plot

KSS sleepiness ratings on flights from a study. Singapore-Los Angeles (solid line from the command crew, dotted line from relief crew)

1 = fully alert, wide awake
2 = very lively, responsive, but not at peak
3 = okay, somewhat fresh
4 = a little tired, less than fresh
5 = moderately tired, let down
6 = extremely tired, very difficult to concentrate
7 = completely exhausted, unable to function effectively

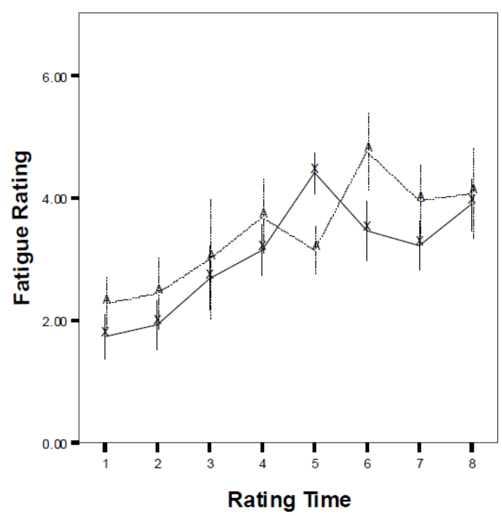


Figure 16: Samn Parelli Fatigue Ratings and Sample Plot

Samn-Parelli fatigue ratings on flights from a study. Singapore-Los Angeles (solid line from the command crew, dotted line from relief crew)

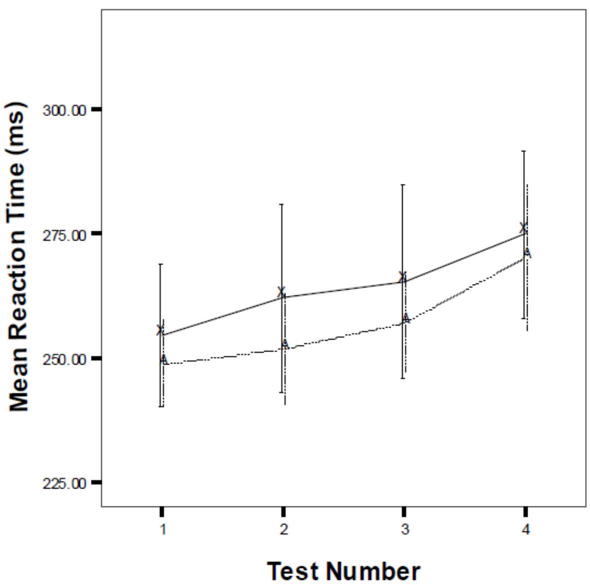


Figure 17: PVT Plot – Sample data shows the mean reaction time.
Graph is plotted at various test times on a study conducted between Singapore to Los Angeles. (Solid line from the command crew, dotted line from relief crew). Test 1 through 4 represents top of climb, start of inflight rest, close to top of descent and post flight respectively

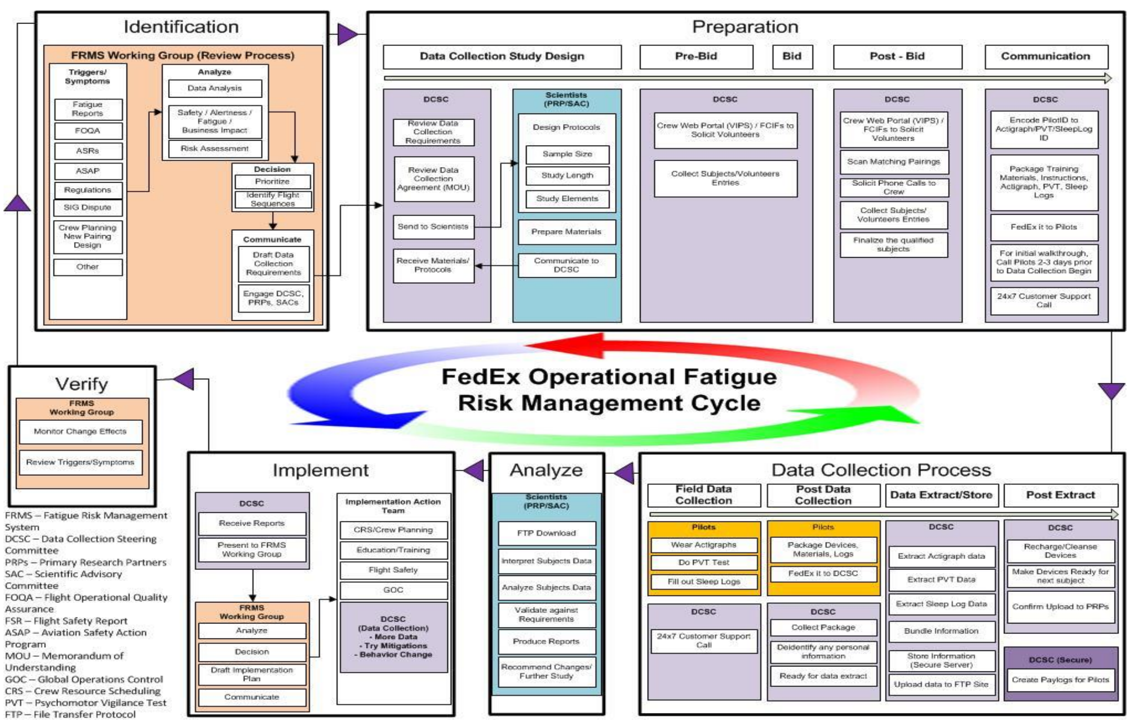


Figure 18: Data Collection Iterative Process and Continuous Improvement

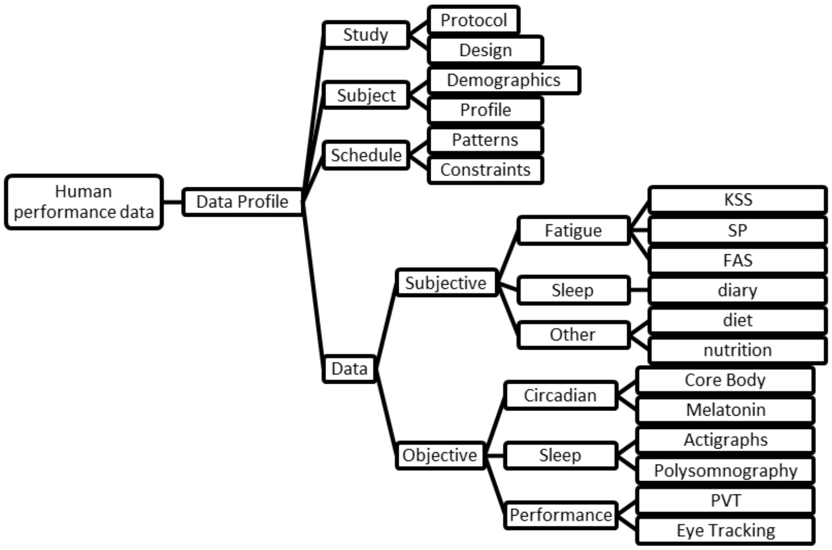


Figure 19: Structure of individual study

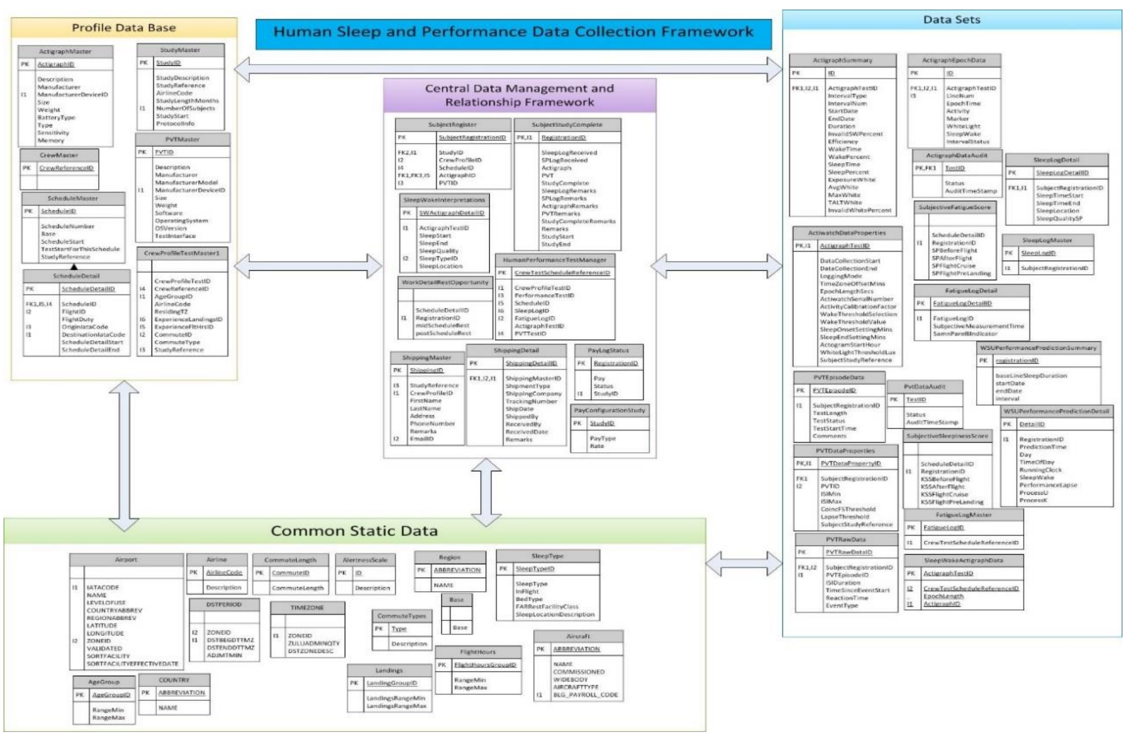


Figure 20: Common Taxonomy for human physiology database

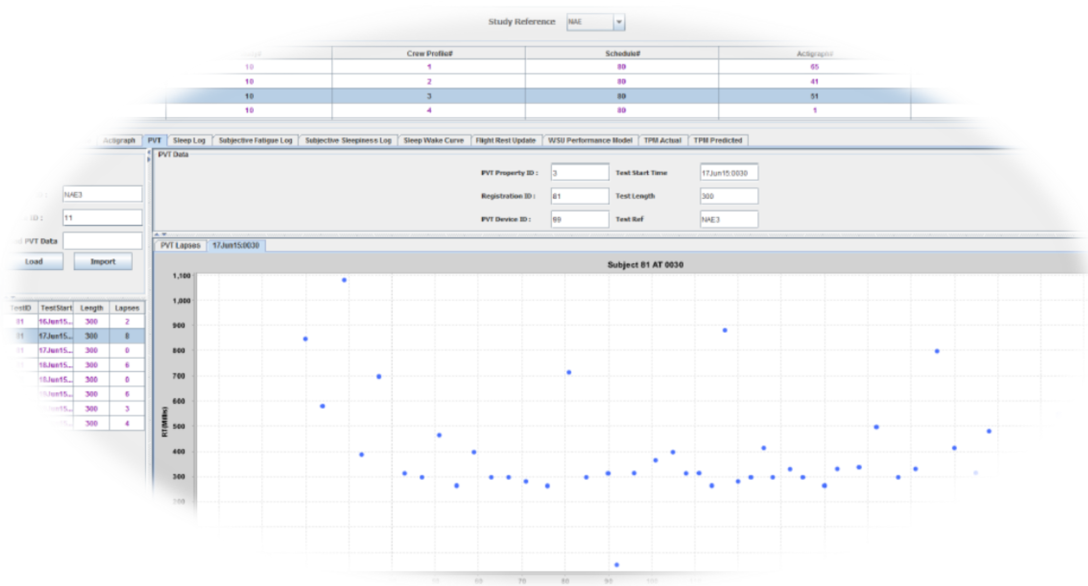


Figure 21: Data Upload: PVT QA Check

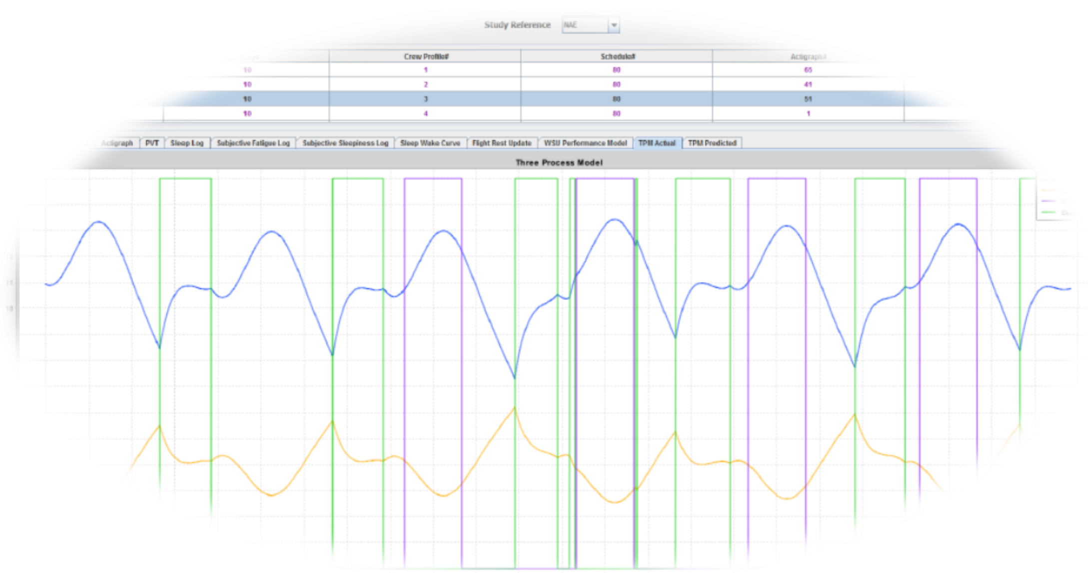


Figure 22: Data Upload: Three Process Model

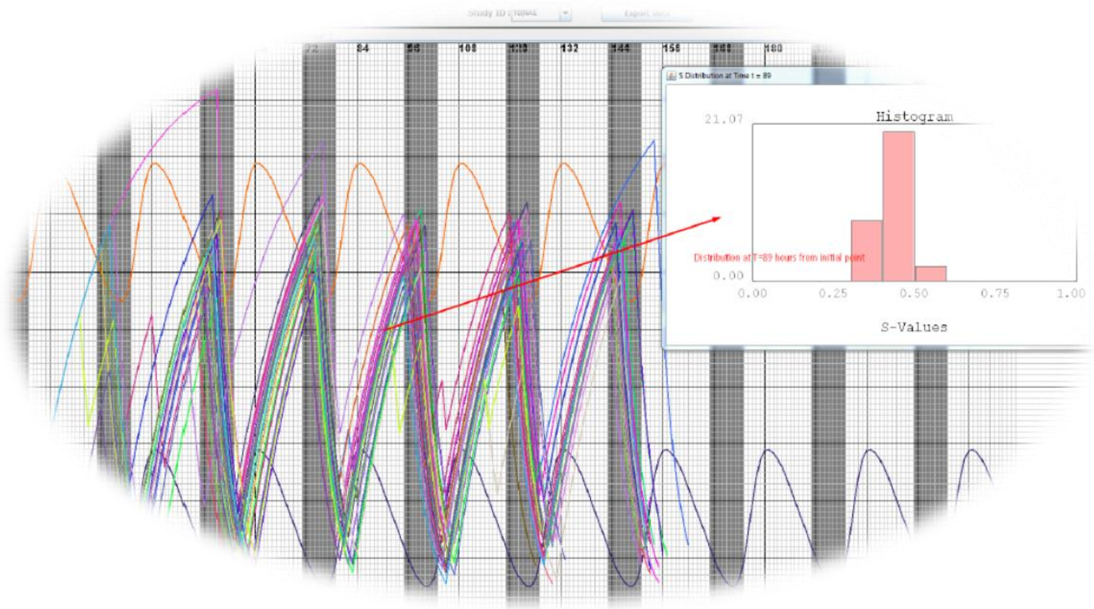


Figure 23: Group Analysis – Homeostatic distribution

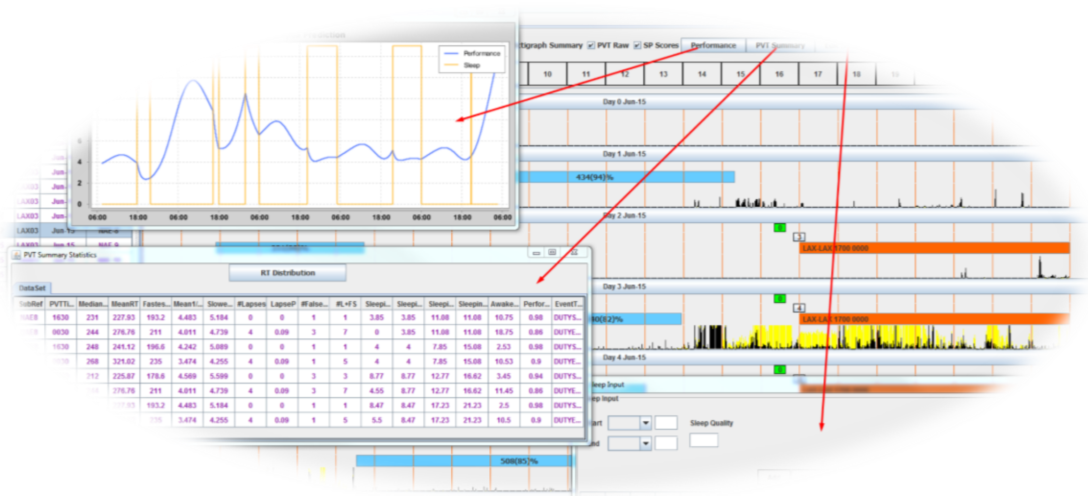


Figure 24: Individual data Investigate Detail

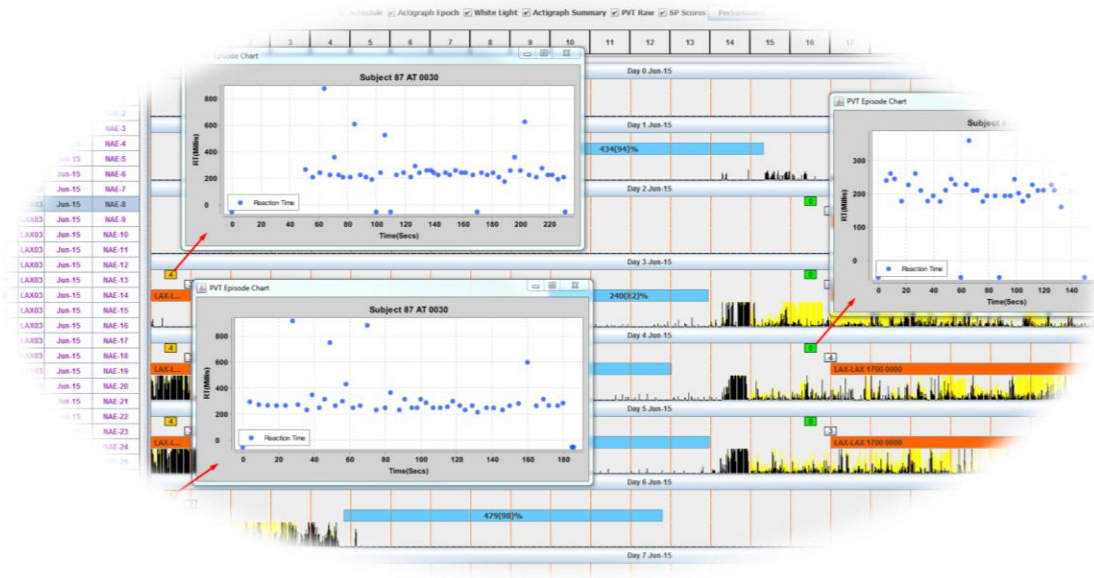


Figure 25: Individual data – Investigate Detail

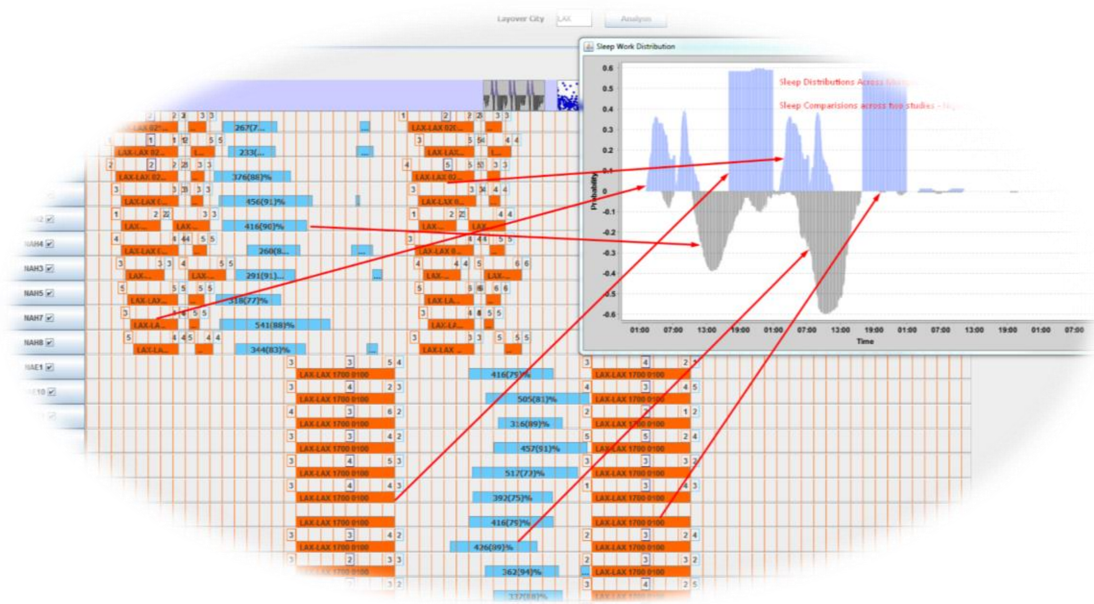


Figure 26: Sleep Work Distribution across Multiple Studies

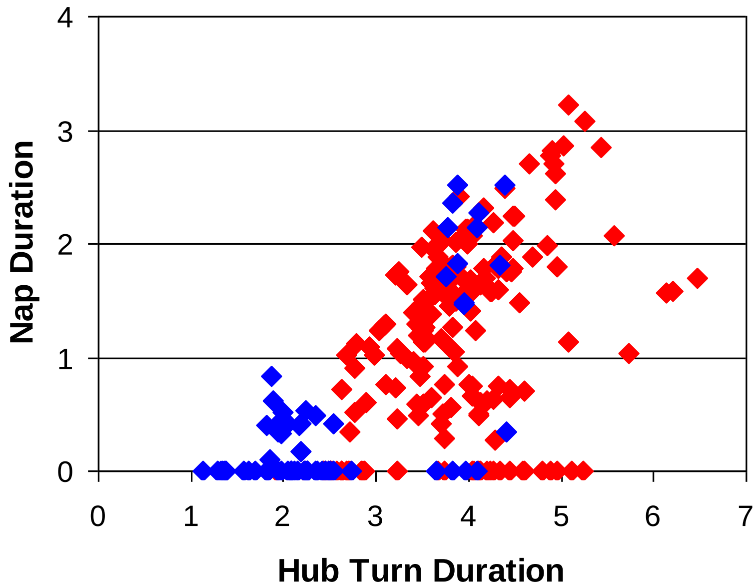


Figure 27: Nap duration (in hours) plotted against hub turn duration (in hours)
Hub1 and Hub2, for each individual pilot and duty night.

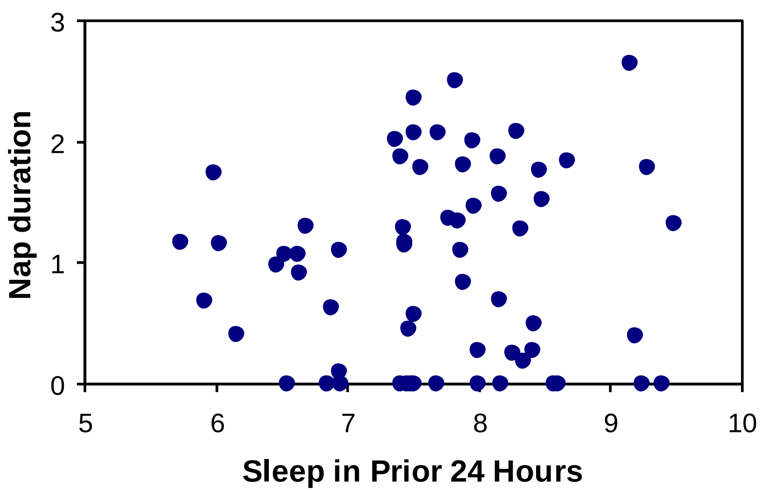


Figure 28: Sleep Work Distribution across Multiple Studies
Scatter plot of nap duration at the hub (in hours) versus sleep duration in the 24 hours before duty (in hours). Although analyses were performed on the raw data, for the purpose of clarity the plot shows individual subjects' averages across the 4 duty nights.

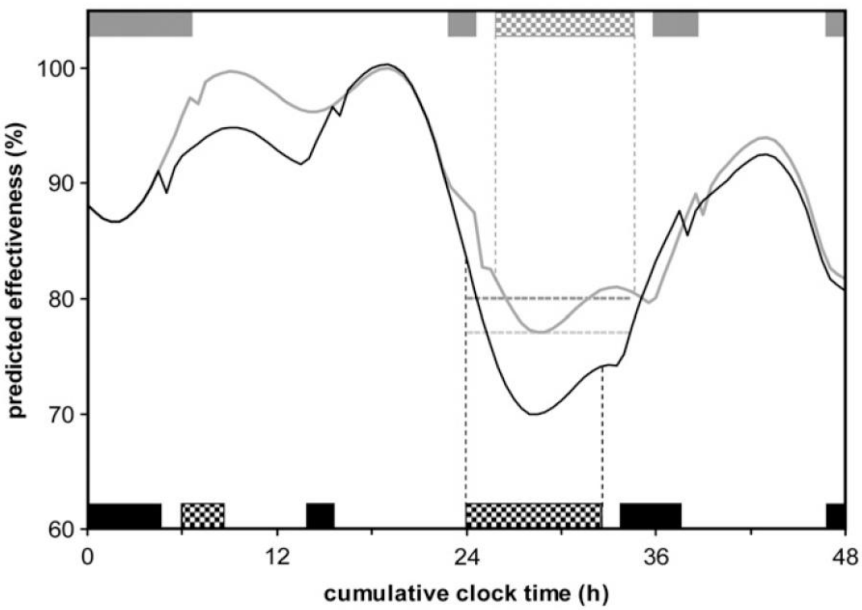


Figure 29: Comparison of two sleep/wake/duty schedules based on different thresholds

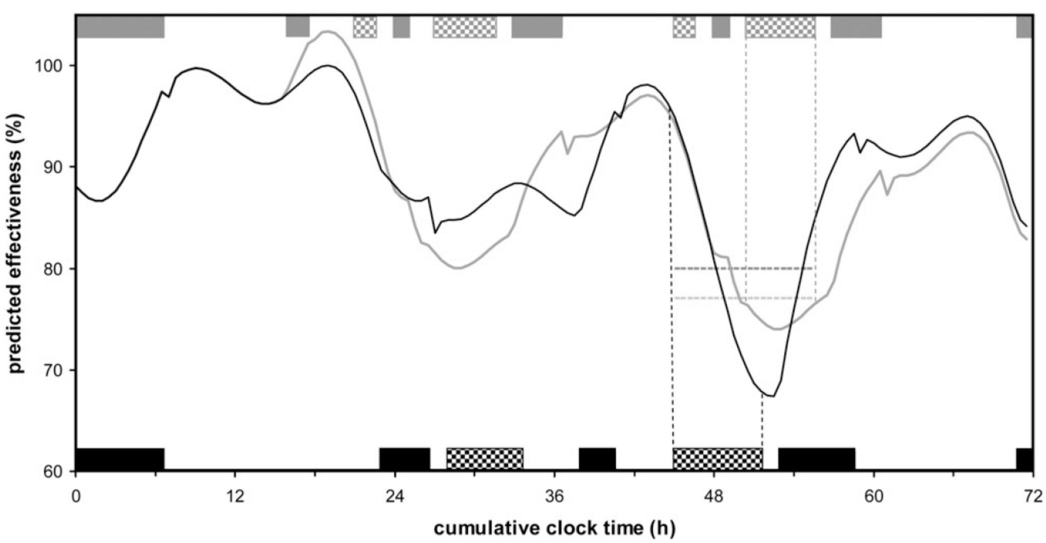


Figure 30: Ambiguity in the comparison of duty periods based on fatigue thresholds.

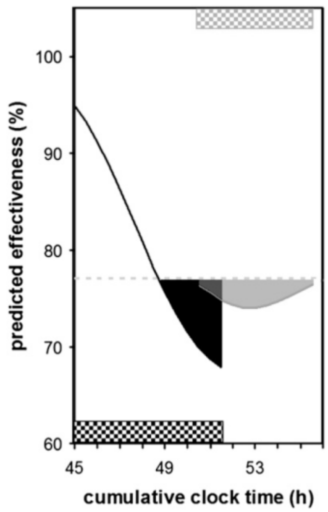


Figure 31: AUC magnification to compare schedules

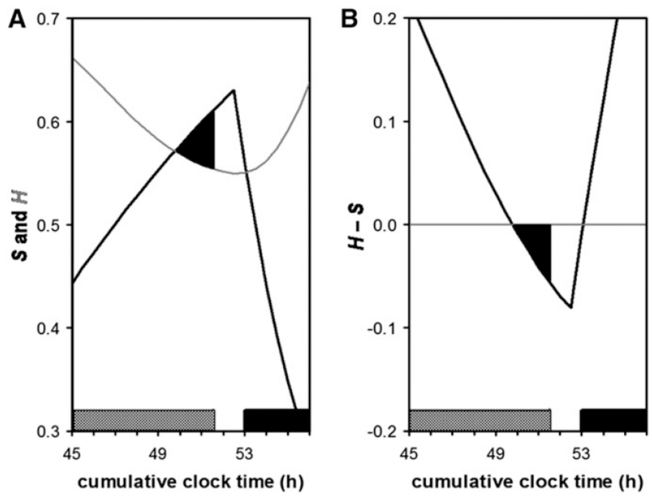


Figure 32: Process S and Process C Interactions

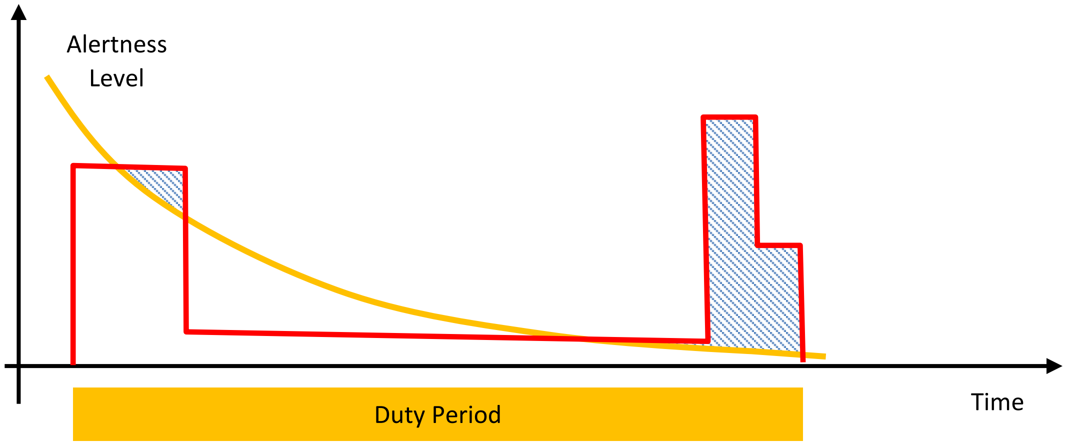
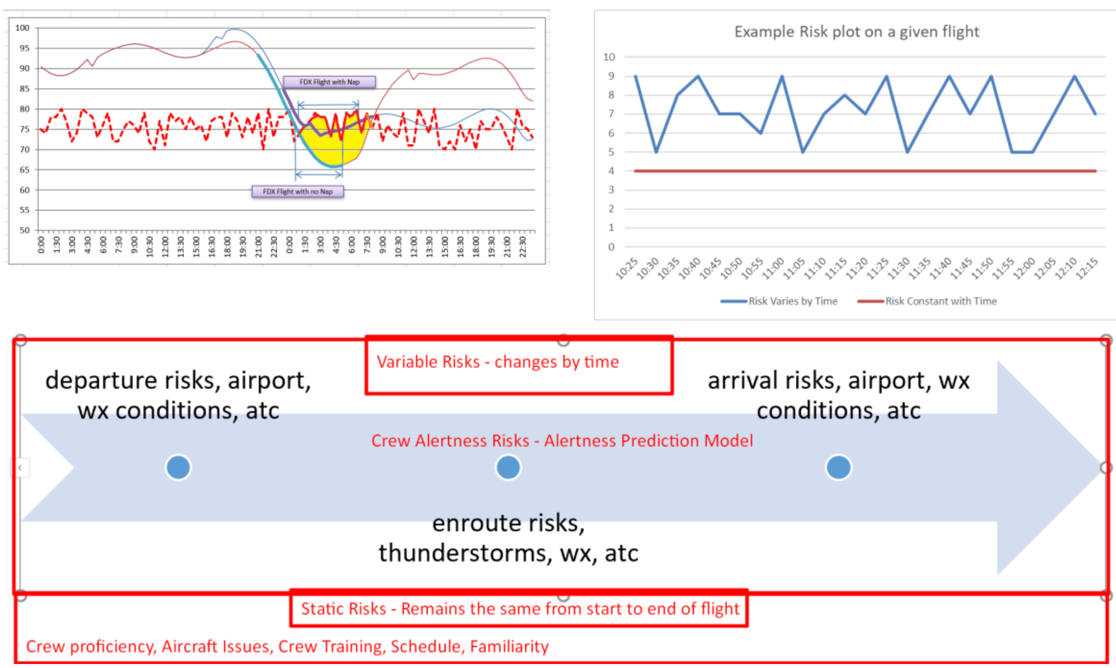


Figure 33: New method to evaluate fatigue risk associated with required alertness levels



Total risk = Static Risk + Variable at each minute of X.

Figure 34: Representation of Three-dimensional quantification of fatigue risk

Figure explains how the third-dimension quantification to fatigue risk is added based on computed risk score. Instead of the straight line of defining threshold, now a curve will be dynamically created based on the risks.

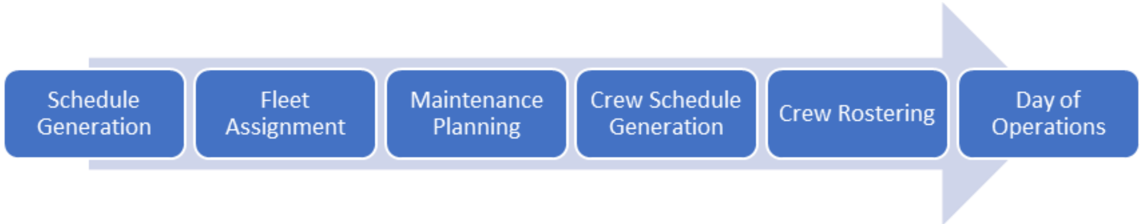
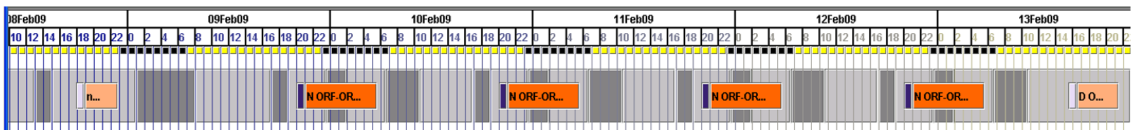
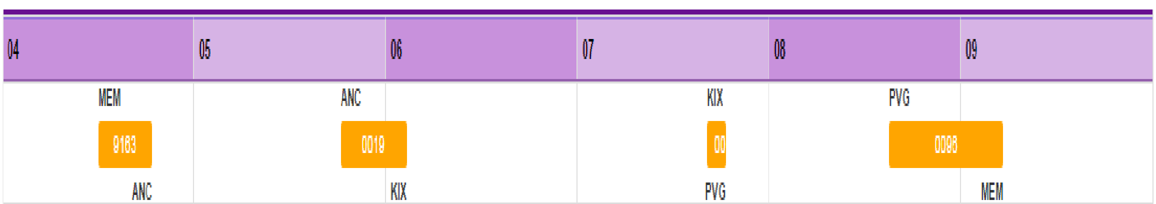


Figure 35: Sequential crew operations process





Base	A/C	B_Lat	B_Long	Day#	LBT_S	LBT_E	P_L/O	Dep	D_Lat	D_Long	TZD-Base	LDepT	Arr	A_Lat	A_Long	TZD-Base	LArrT	TZD	D_Mins	B_Mins	MaxTurn	MaxTurn	#Crews	#Flts	DOW
MEM	77	35N	-89W	0	0500	1242		MEM	35N	-89W	0.0	0500	ANC	61N	-149W	-3.0	0942	-3.0	07:42	06:42		-	2	1	Tue
MEM	77	35N	-89W	1	1125	2035	22:43	ANC	61N	-149W	-3.0	0825	KIX	34N	135E	-9.0	1135	-6.0	09:10	08:10		-	2	1	Wed
MEM	77	35N	-89W	3	0910	1237	36:35	KIX	34N	135E	-9.0	0010	PVG	31N	121E	-10.0	0237	-1.0	03:27	02:27		-	2	1	Fri
MEM	77	35N	-89W	4	0800	2320	19:23	PVG	31N	121E	-10.0	2200	MEM	35N	-89W	0.0	2320	-14.0	15:20	14:20		-	2	1	Sat

Figure 37: sample crew schedule and duty attributes.

Key	Layover	Duty Start	Duty Len	Duty End	Block	Mitigation	Landings	TZ BASE	TZ Duty	Position	DutyDay	BASE	AC
10-04Apr17	00	SC	DM	ED	BM	MN	L1	ta	tA	PS	0	MEM	77
10-04Apr17	LW	SD	DM	EN	BL	M1	L1	ta	tB	PS	1	MEM	77
10-04Apr17	LX	SD	DS	ED	BS	MN	L1	tb	tA	PS	3	MEM	77
10-04Apr17	LA	SD	DX	EN	BX	M2	L1	tb	tB	PS	4	MEM	77

Figure 38: sample crew schedule and dna coding.

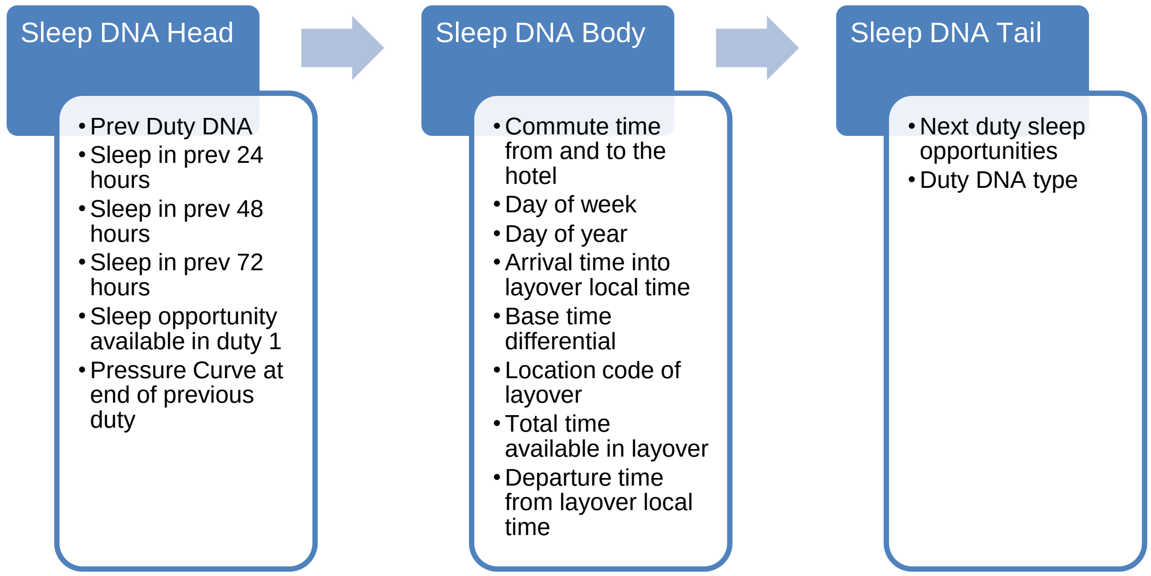


Figure 39: structure of sleep DNA as DNAHead, DNABody, DNATail.

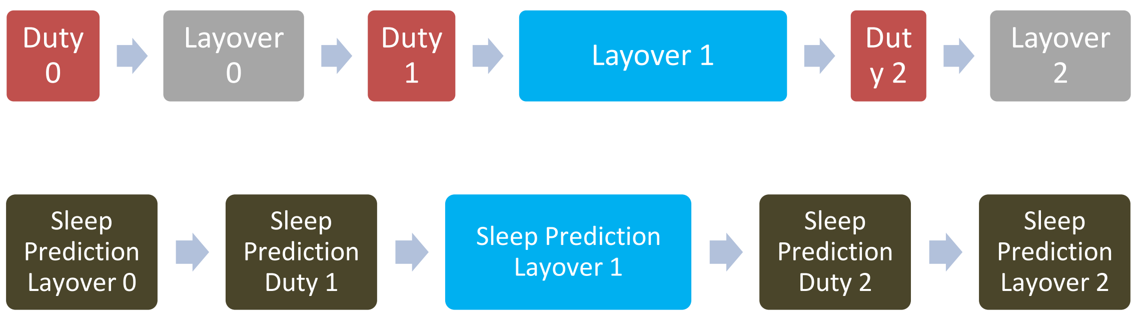


Figure 40: schematic of the duty sequence where predictions are made for layover.
Top: Schematic of the duty sequence where predictions are made for Layover 1 which is in between duty DNA 1 and duty DNA 2. Bottom: These sleep predictions are connected.

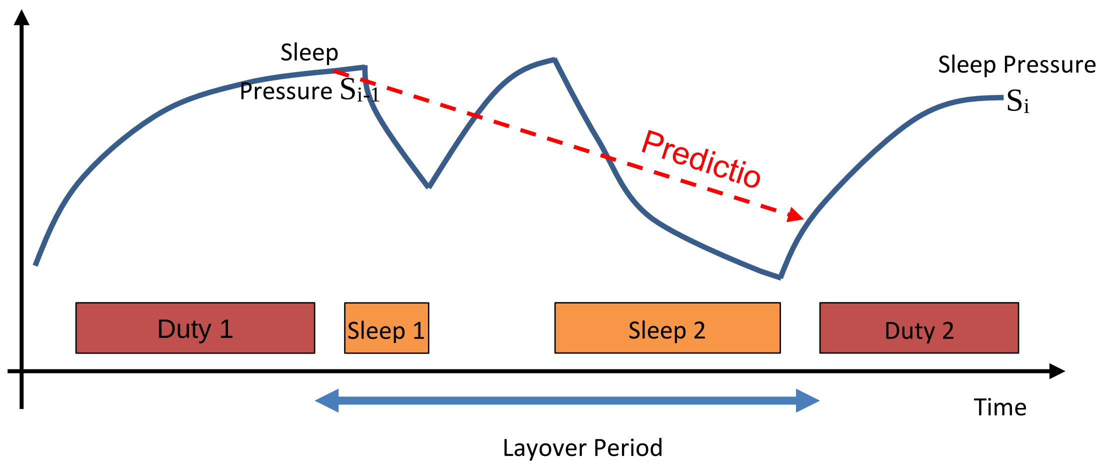


Figure 41: sleep pressure curves generated using two-process model
Sleep pressure curves generated using the two-process model based on DNA of two duty periods.

Figure 42: sleep and duty plots as observed, synchronized to reference time scale
Sample duty and sleep plots observed. Blue indicates sleep and Orange indicates flight segments. each line represents one subjects sleep and work schedule observed over a period of time.

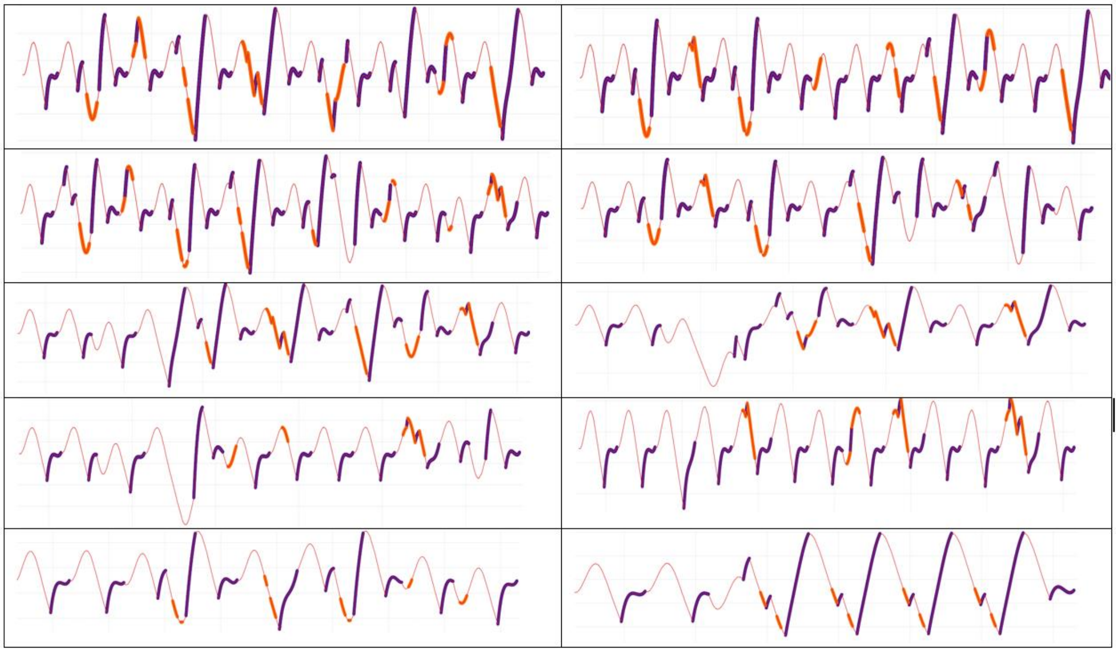


Figure 43: 10 sample alertness curves generated from different schedules.

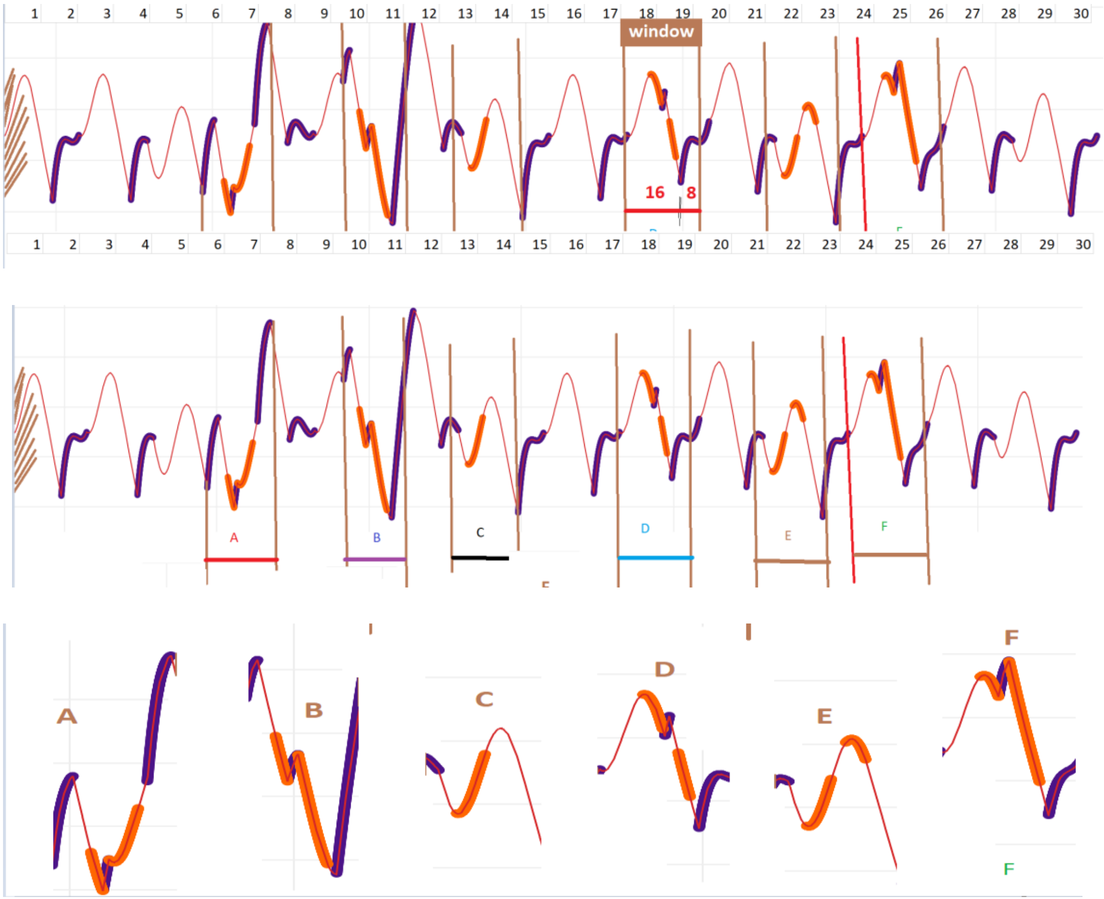
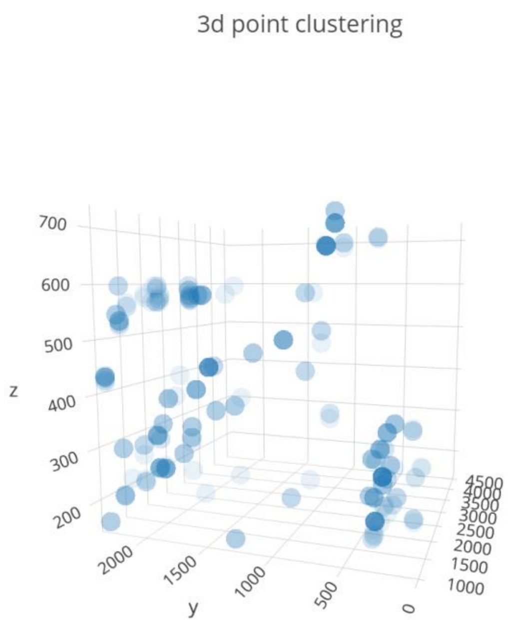


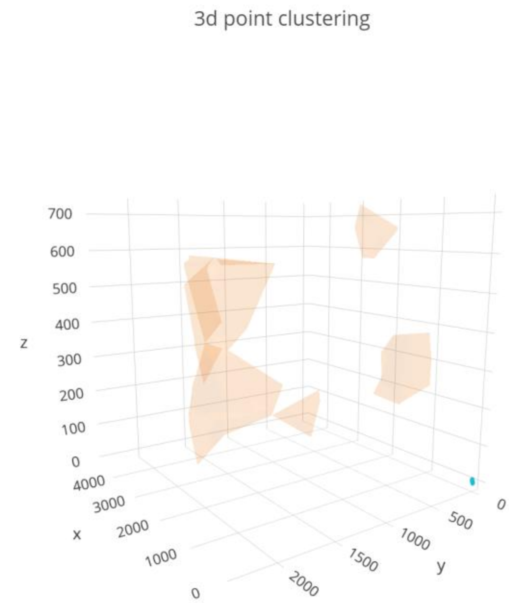
Figure 44: generation of subsequences from alertness curves.

Figure 44 (top) is an example illustration of timeseries T of length 300 hours creating a subsequence of length w.

Figure 44 (middle). six subsequences of length 24 is created by the sliding window.

Figure 44 (bottom). the extracted sequences of performance curves.





Pre clustering in 3d scatter plots plots Before clustering. 365 duty dnas.

Post clustering. The clusters are grouped as 157,66,65,42.36 data points

Figure 45: K-Means clustering of duty dna – pre and post clustering

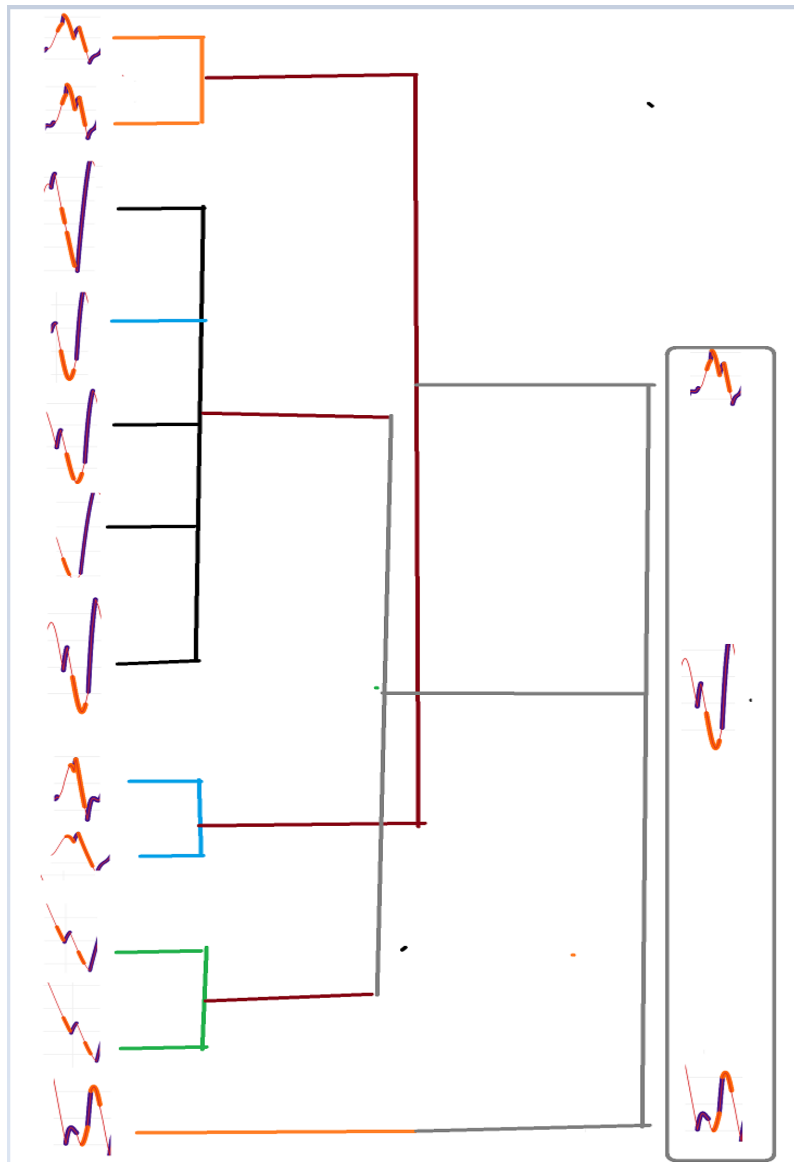


Figure 46: K-Means clustering of duty dna – pre and post clustering

In the illustration of figure 44a, 44b, 44c, the values of subsequences are clustered.

LRSCDLENBLM1L1tatBPS01ME77Jun18
LXSNDXEDBXM2L1tbtBPS04ME77Jun18
LXSDDXENBXM2L1tbtBPM08ME77Jun18
00SEDXENBXM2L1tatBPS00ME77Jun18
LXSNDXEDBXM2L1tbtBPS04ME77Jun18
LWSDDLENBLM1L1TatBPM06ME77Jun18
00SEDXENBXM2L1tatBPS00ME77Jun18
LXSDDLENBLM1L1tbtBPS05ME77Jun18
LXSCDLENBLM1L1tatBPM07ME77Jun18
LWSNDXEDBXM2L1tbtBPM10ME77Jun18
00SEDXENBXM2L1tatBPS00ME77Jun18
LXSNDXEDBXM2L1tbtBPS03ME77Jun18
LRSNDMEEBLM1L1TaTAPS05ME77Jun18
LXSDDXENBXM2L1tbtBPM09ME77Jun18
00SEDXENBXM2L1tatBPS00ME77Jun18
LXSNDXEDBXM2L1tbtBPS03ME77Jun18
LRSNDMEEBLM1L1TaTAPS05ME77Jun18
LXSDDXENBXM2L1tbtBPM09ME77Jun18
LASCDMEDBLM1L1TbTAPS00ME77Jun18

Figure 47: Full String dutyDNA with all variables.
List shows the list of 20 sample dutyDNA Sequences.

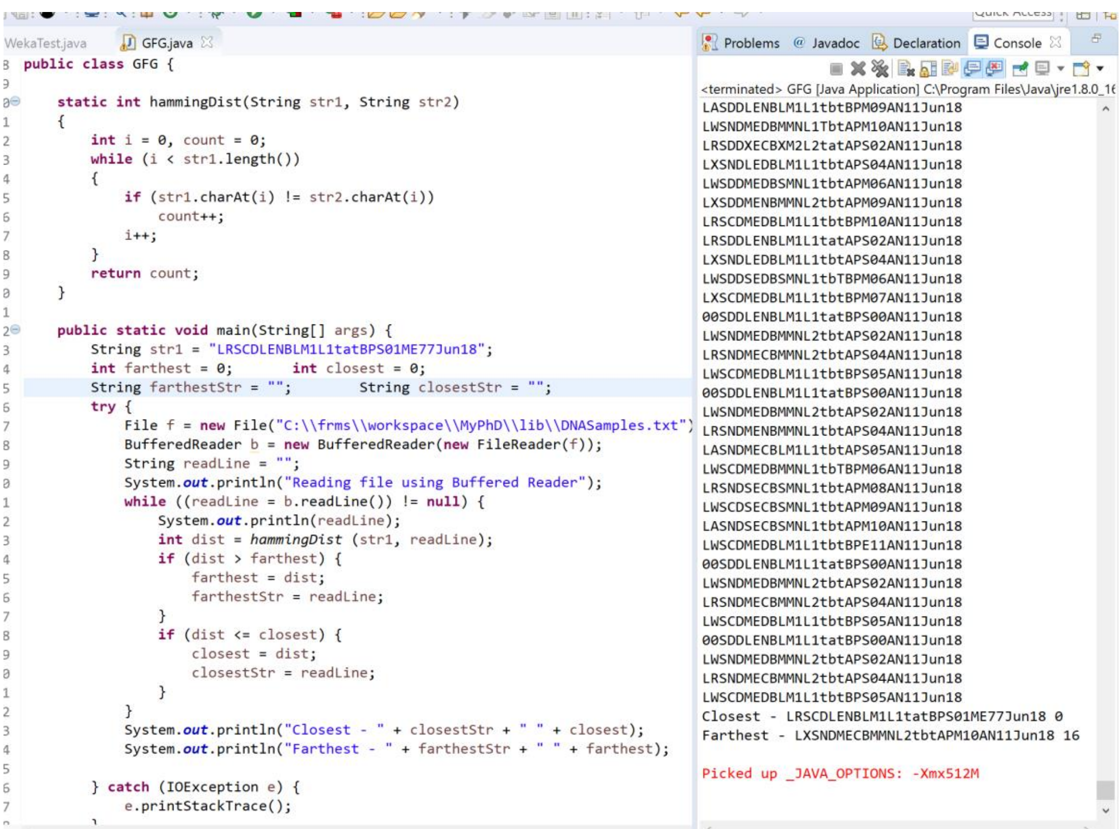


Figure 48: Java code written to measure the hamming distance from a given DNA to the rest of DNAs.

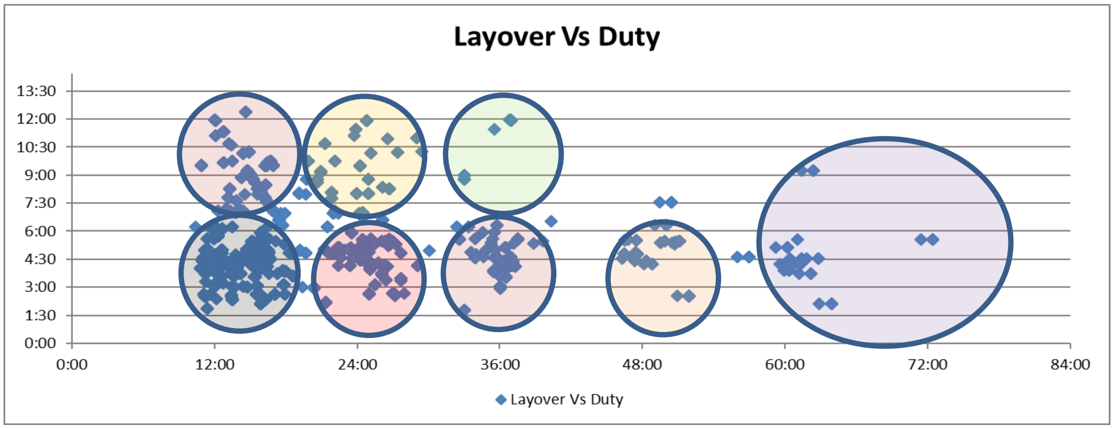


Figure 49: A scatter plot of 2 influencing variables in a duty.

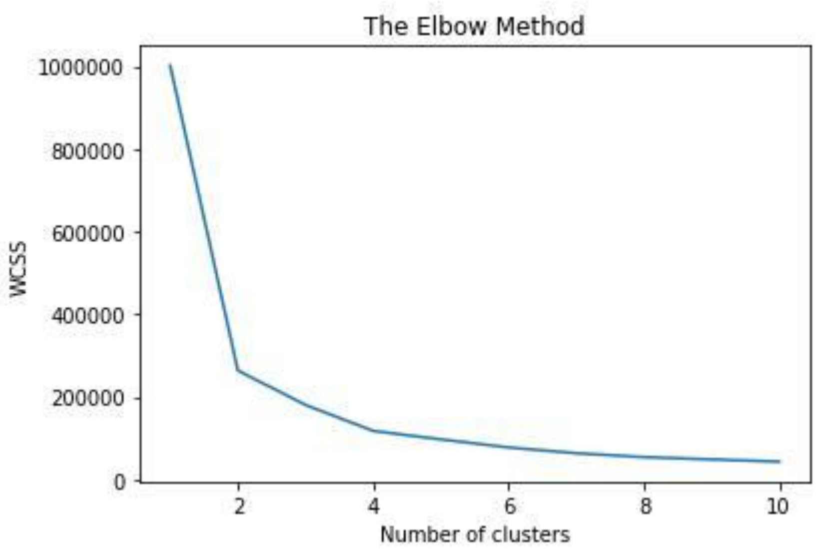


Figure 50: the elbow indicated when the curve straightens. The number of clusters is 4 above.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

# Importing the dataset
dataset = pd.read_csv('AllKSS-Sri.csv')
wh1 = dataset[['D-S','kss_duty_start']]

# Using the elbow method to find the optimal number of
clusters
from sklearn.cluster import KMeans
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++',
random_state = 42)
    kmeans.fit(wh1)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 11), wcss)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```

Figure 51: python program to compute elbow method.

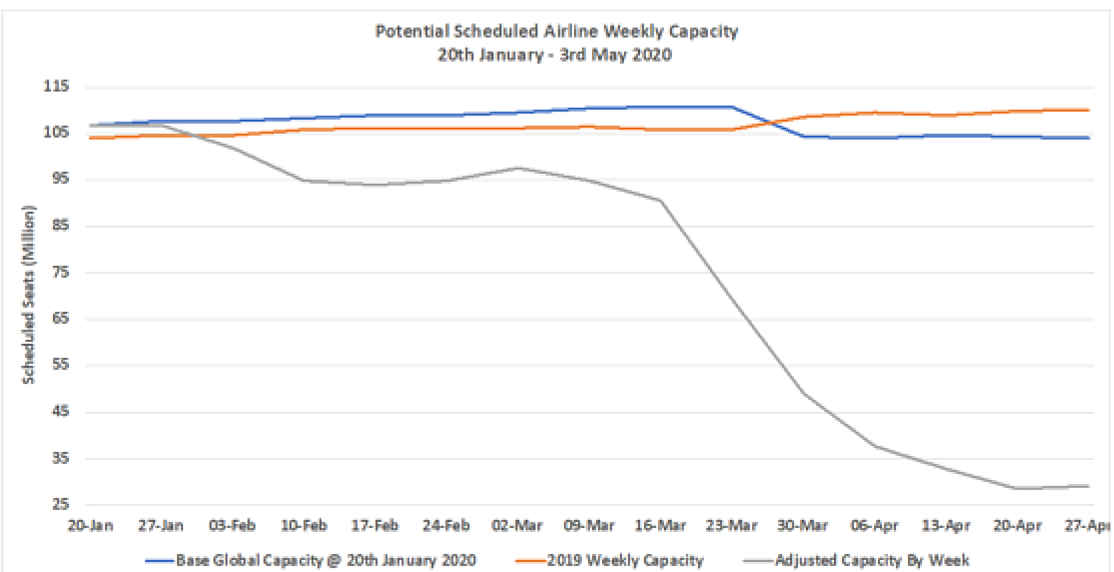


Figure 52: Scheduled Airline Capacity by Week
Scheduled Airline Capacity by Week Compared to Schedules Filed on 20th January 2020

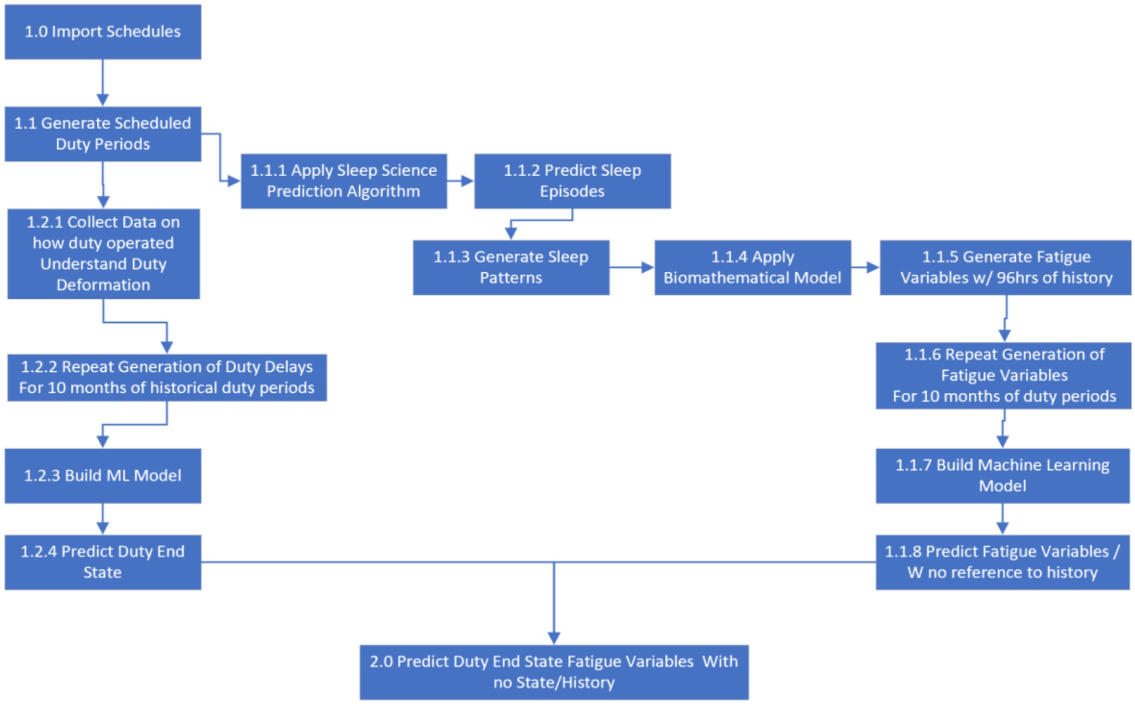


Figure 53: Proposed interconnected machine learning models to predict fatigue on duty end state

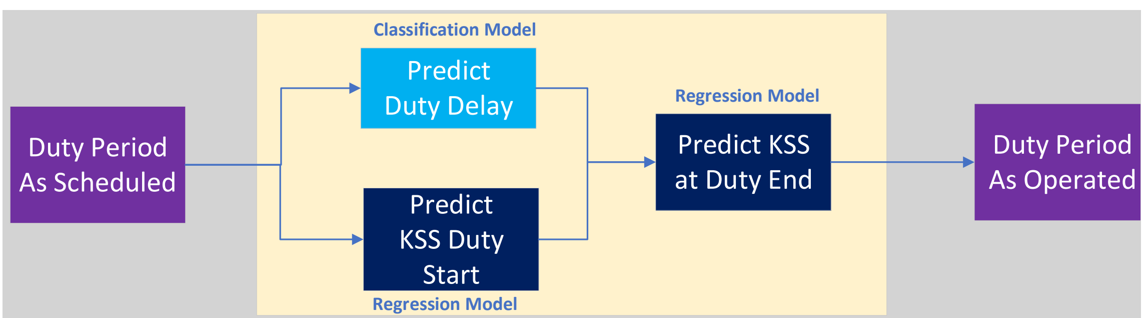


Figure 54: schematic representation of the approach to solve the business problem.

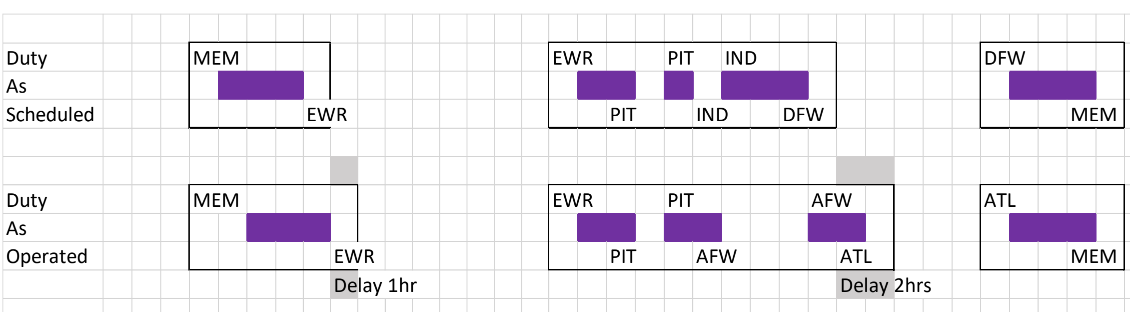


Figure 55: schematic representation of a crew member schedule with duty periods.

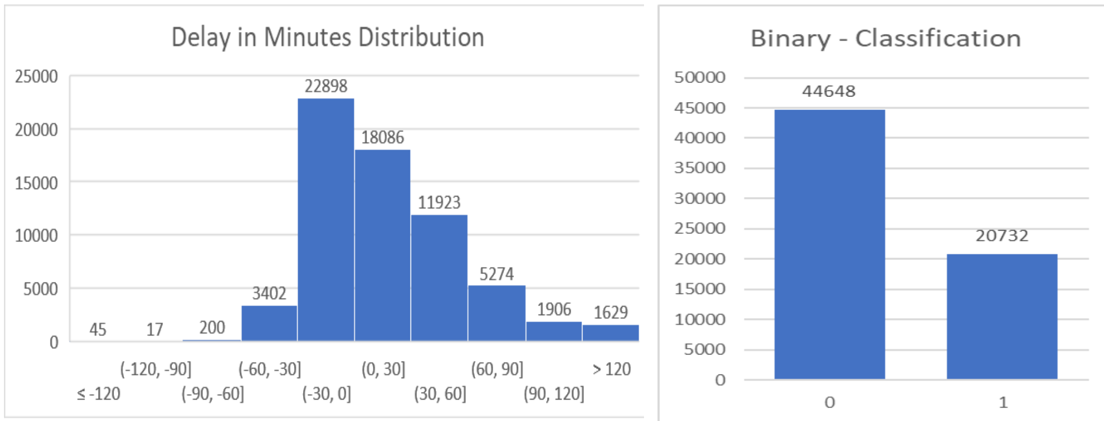
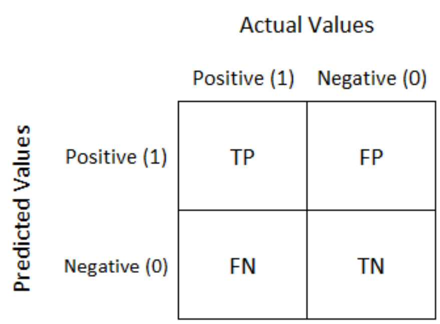


Figure 56: duty delay variance



TP – you predicted positive and its true

TN – you predicted negative and it true

FP – you predicted positive and its false

FN – you predicted negative and its false

Figure 57: confusion matrix evaluation criteria for a classification

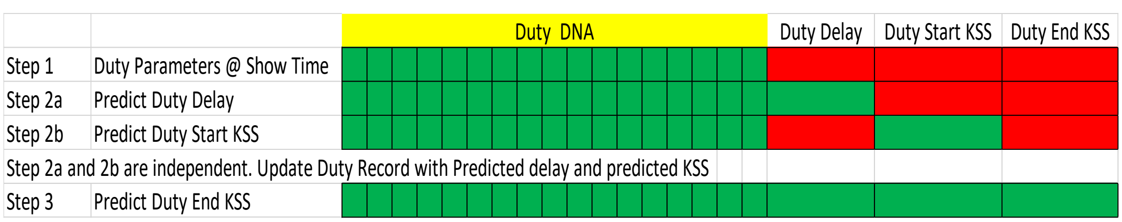


Figure 58: schematic representation of variables prediction

In the multi machine learning model approach, the schematic representation above explains how different variables are predicted to feed to the next model.

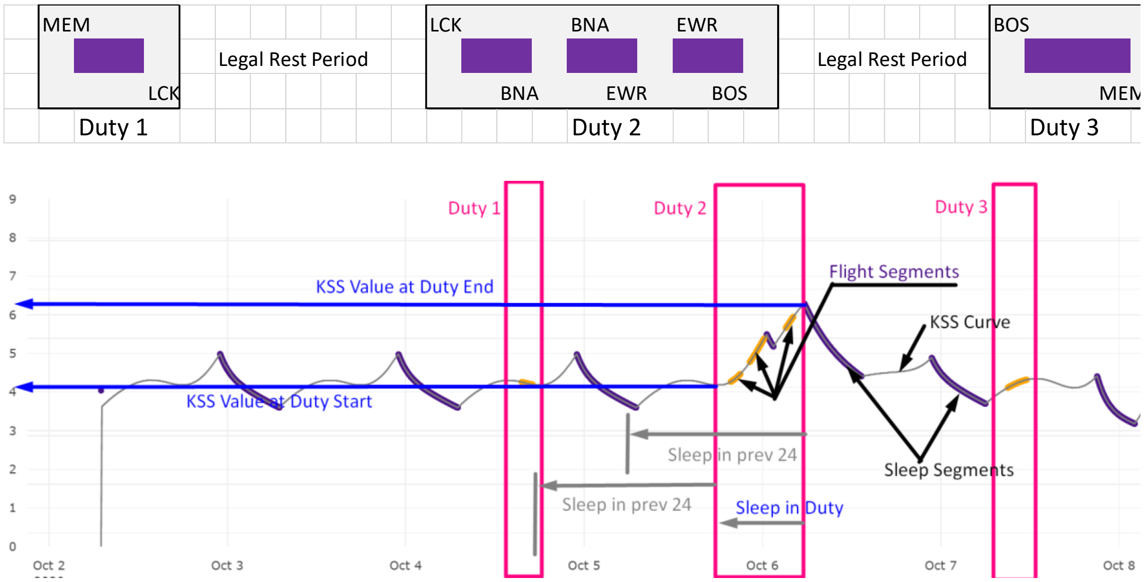


Figure 59: Sequence of Duty and the computed biomathematical model.
also presented the five fatigue variables captured for each of the duty periods. Y-Axis is the KSS score 1-9 and x-Axis is time.

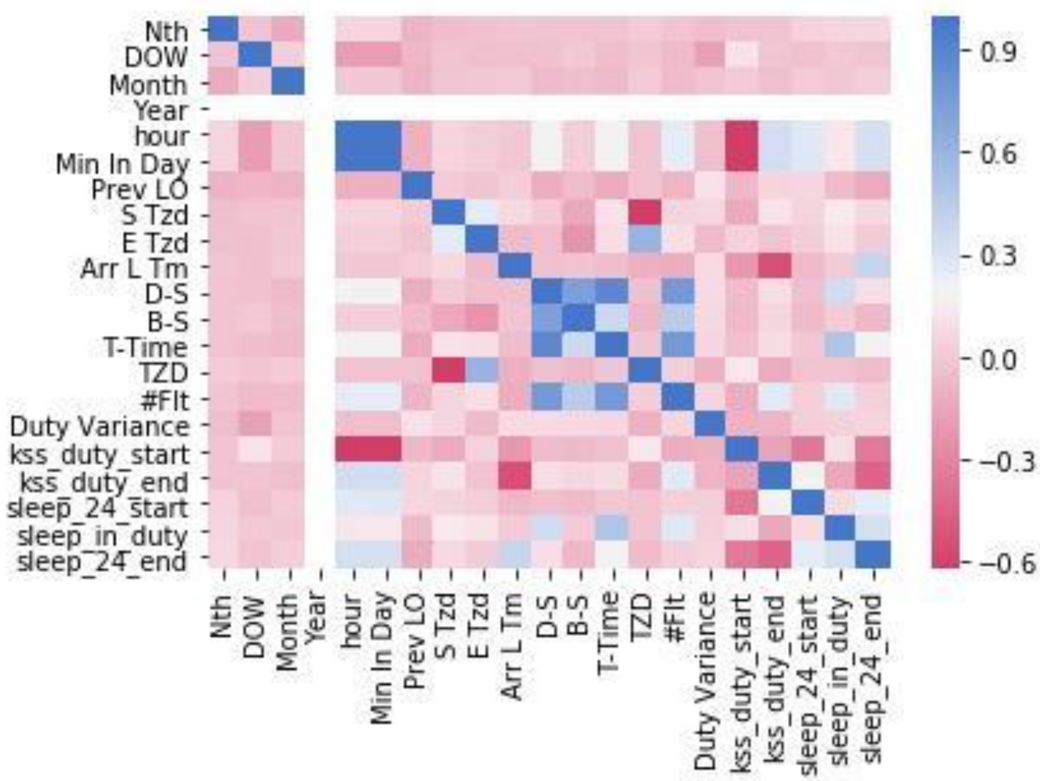


Figure 60 top: Plots representing the feature importance, correlation and collinearity with each other. The stronger the color the stronger the correlation.

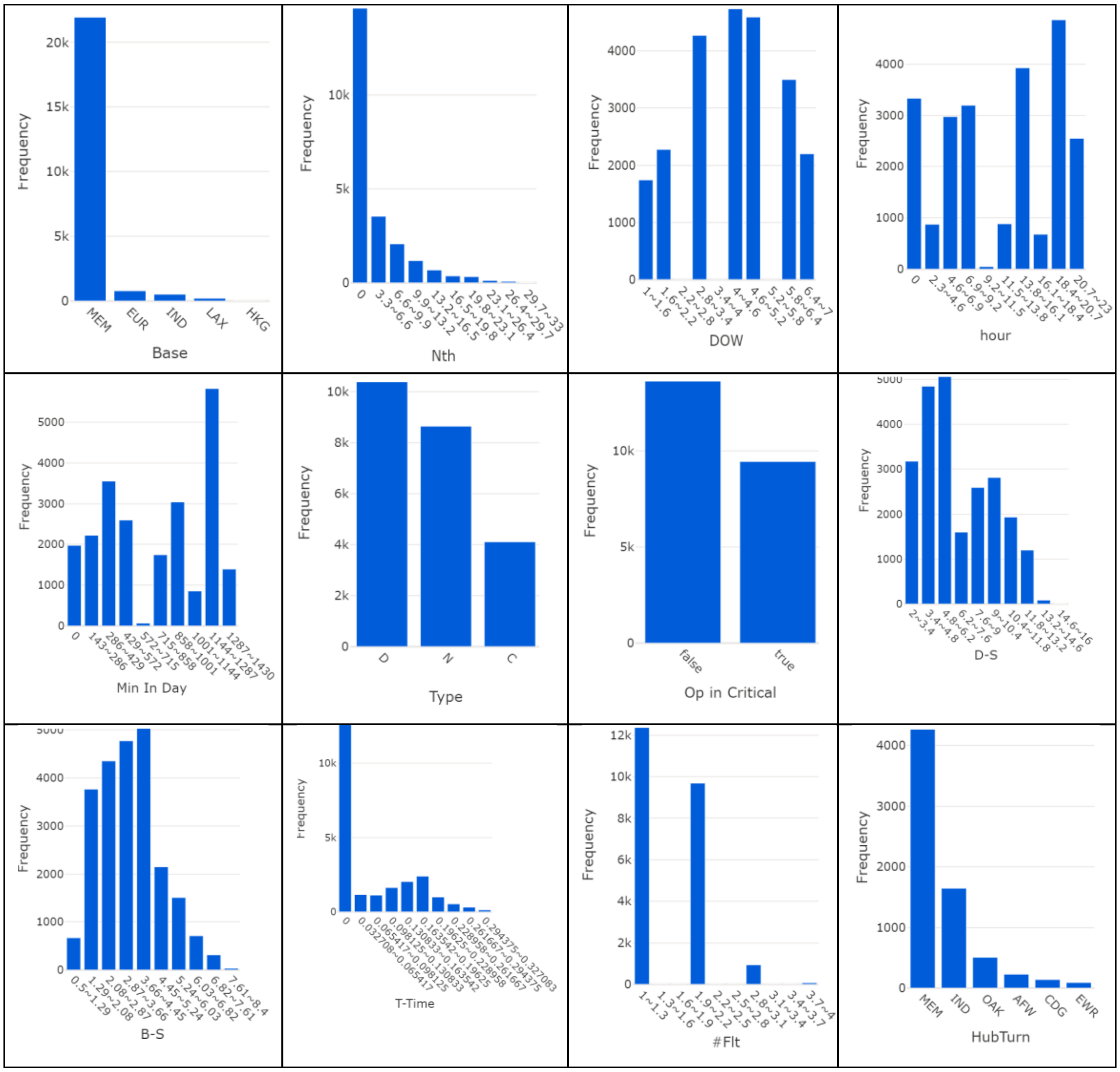


Figure 60: DNA Features, relationships and their histogram plots.
Figure 60 middle: Plot of histogram for few features representing their distribution

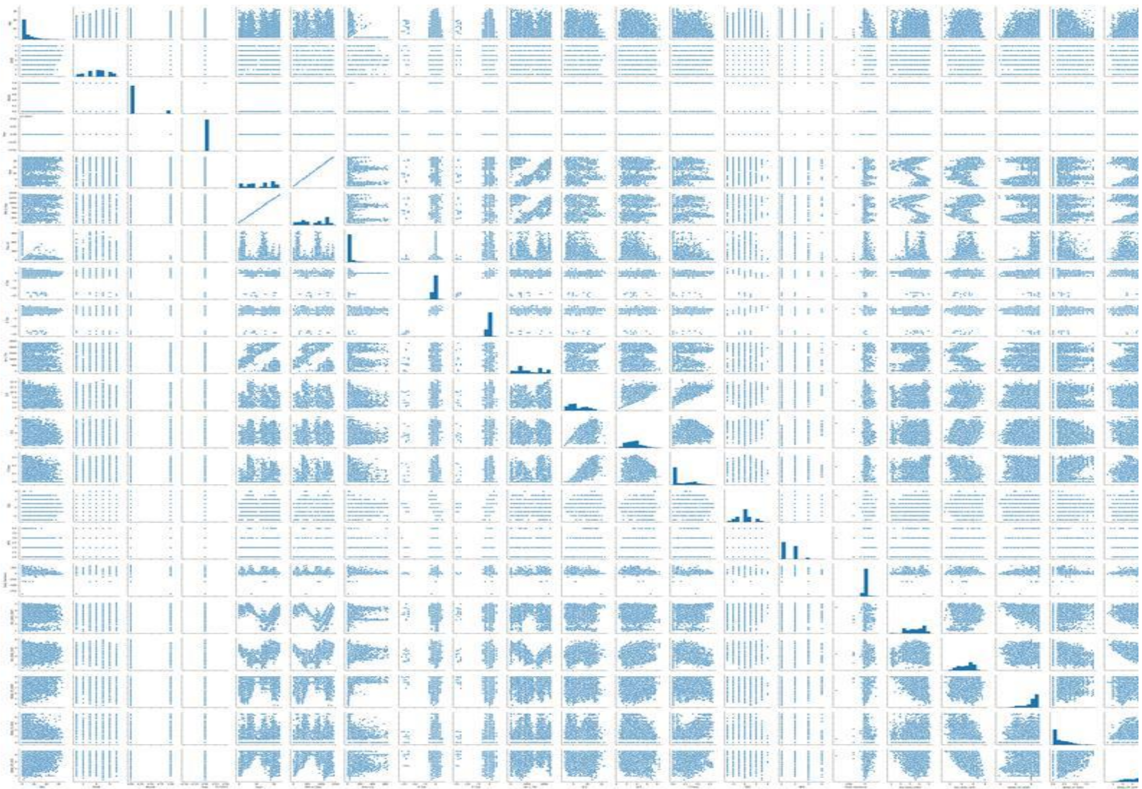


Figure 60 bottom: Plot of histogram of all features representing their correlations

Figure 61: Fatigue Variables as scored by the biomathematical models in histogram plots

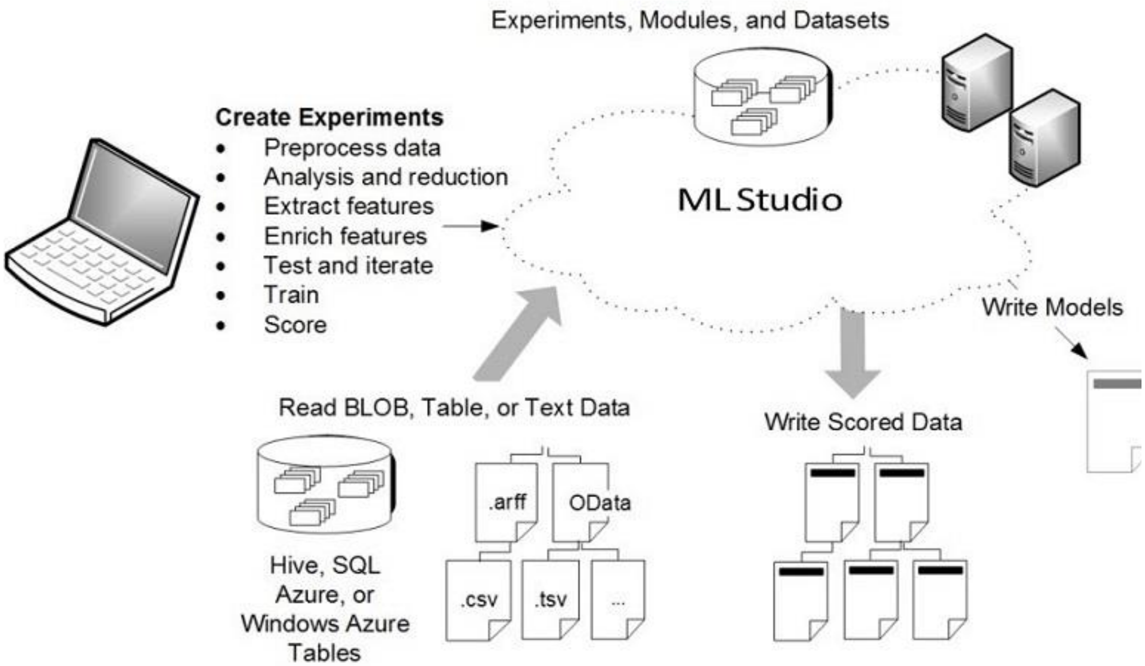


Figure 62: Azure data platform architecture

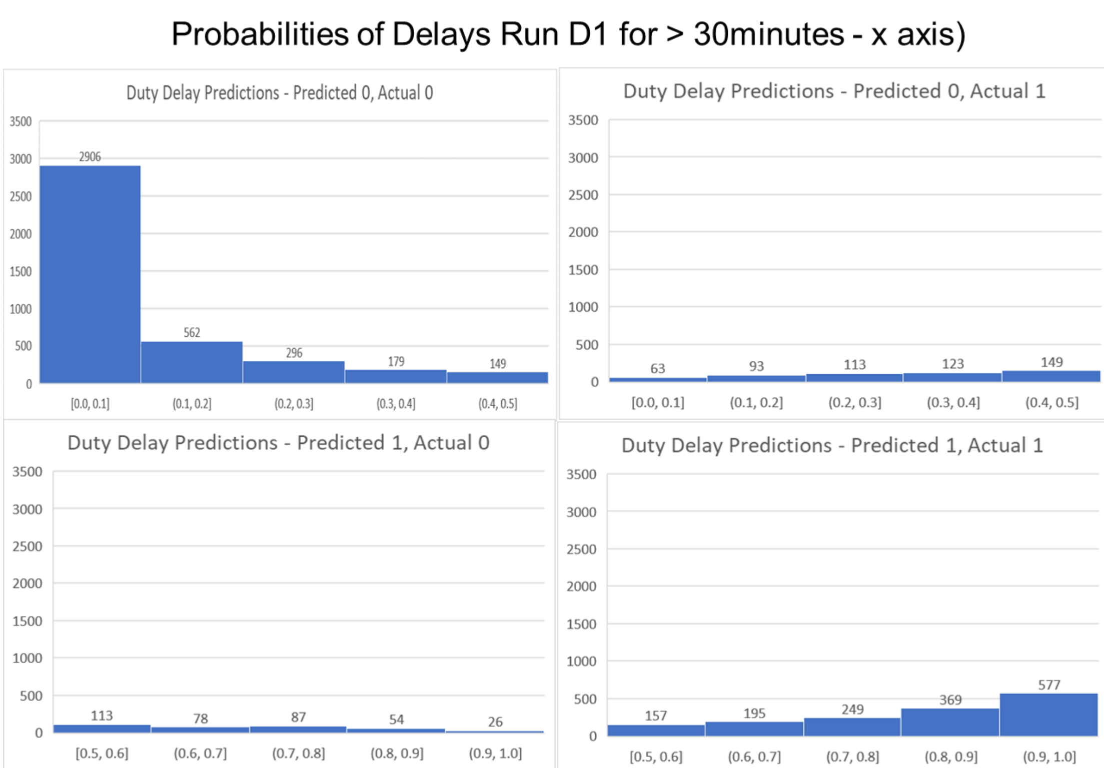


Figure 63: probabilities of delays – Run D1 for > 30 minutes

$$Label = \begin{cases} 0, & \text{if } \; Duty\ Variance \le 30 \\ 1, & \text{if } \; 30 < Duty\ Variance \le 60 \\ 2, & \text{else} \end{cases}$$

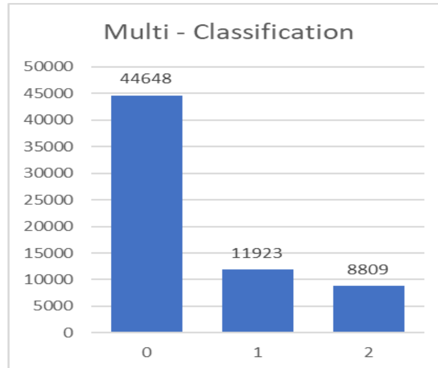


Figure 64: multi classification labels and population data

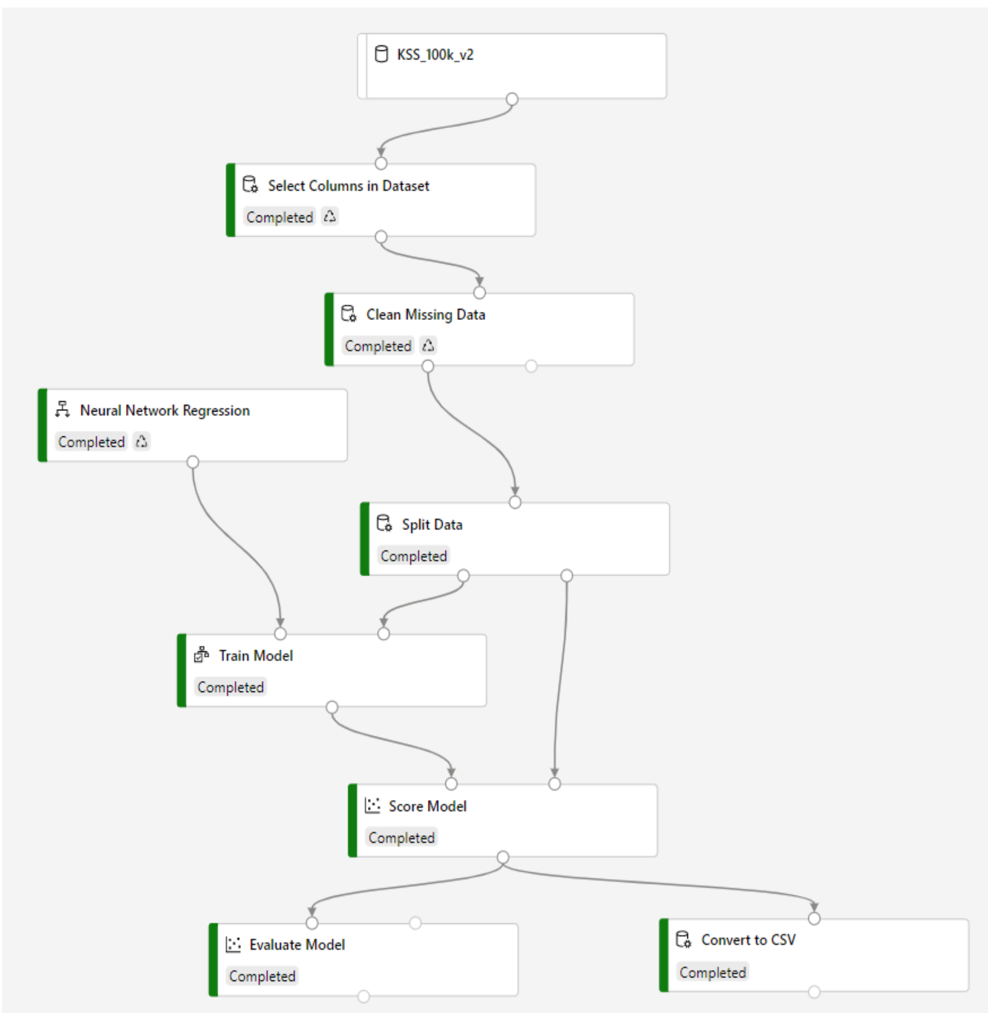


Figure 65: Run K1 – logical flow of the deep learning neural network regression

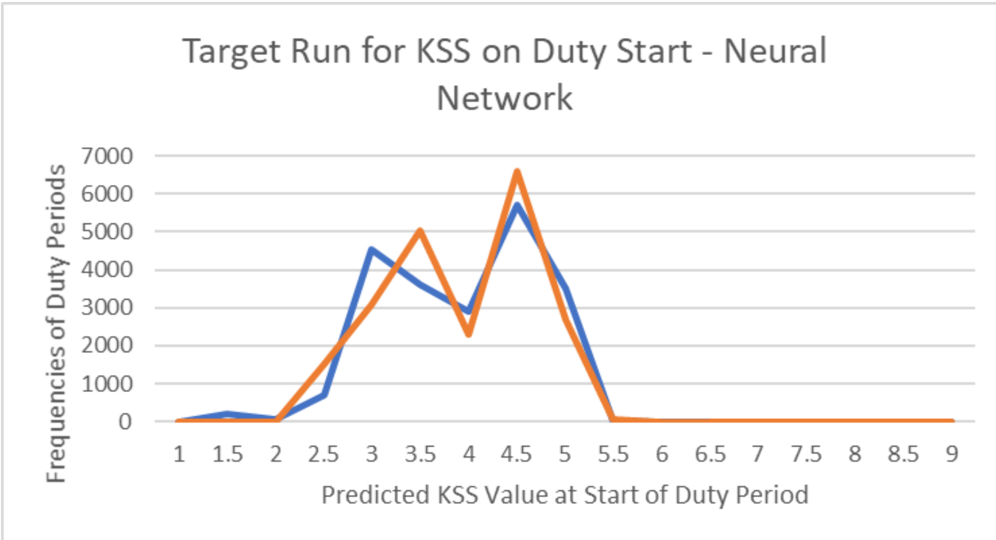


Figure 66: Run K1 – comparing the target for Run K1 with scored values

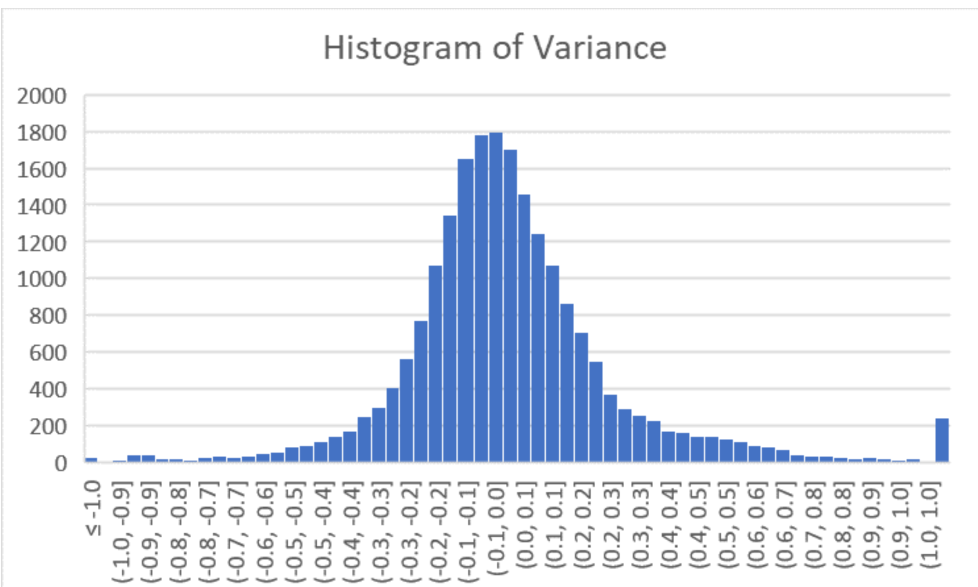


Figure 67: Run K1 – histogram of variance

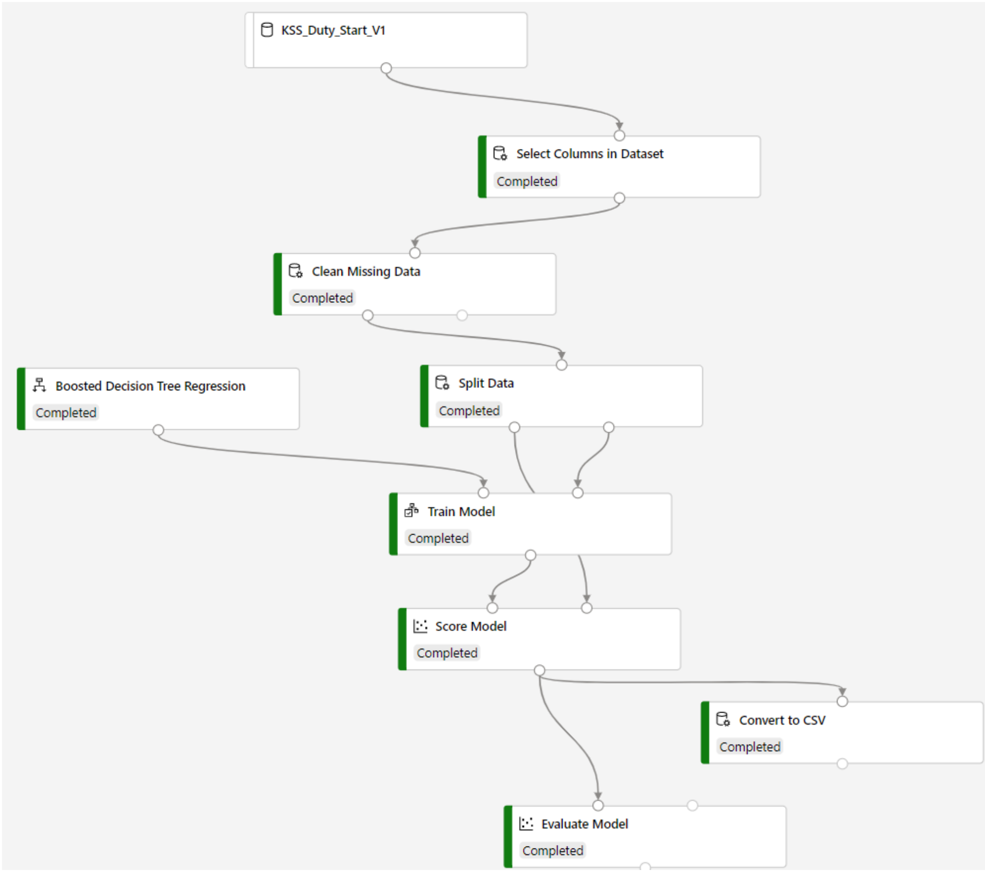


Figure 68: A1 Boosted Decision Tree Regression flow chart

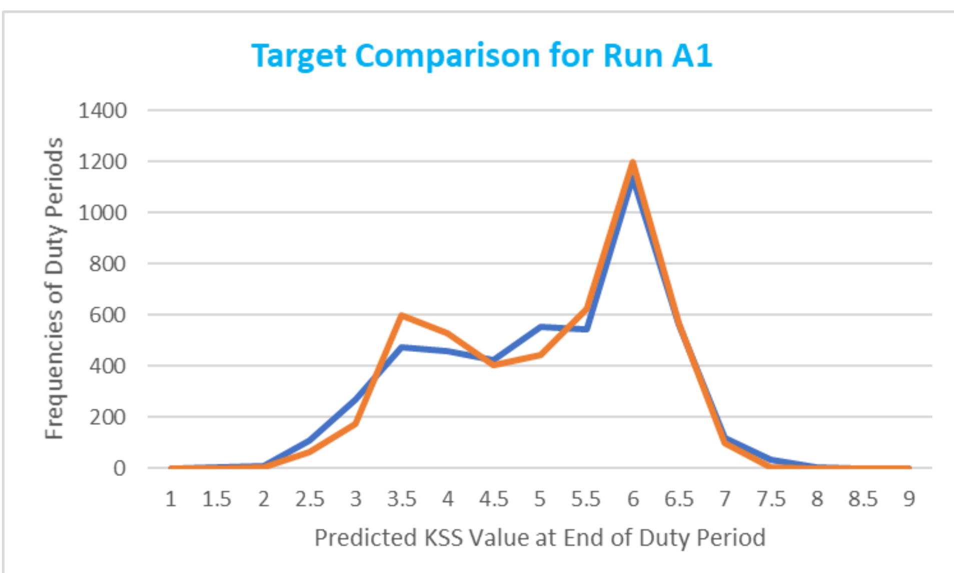


Figure 69: Target vs observed value comparison – A1

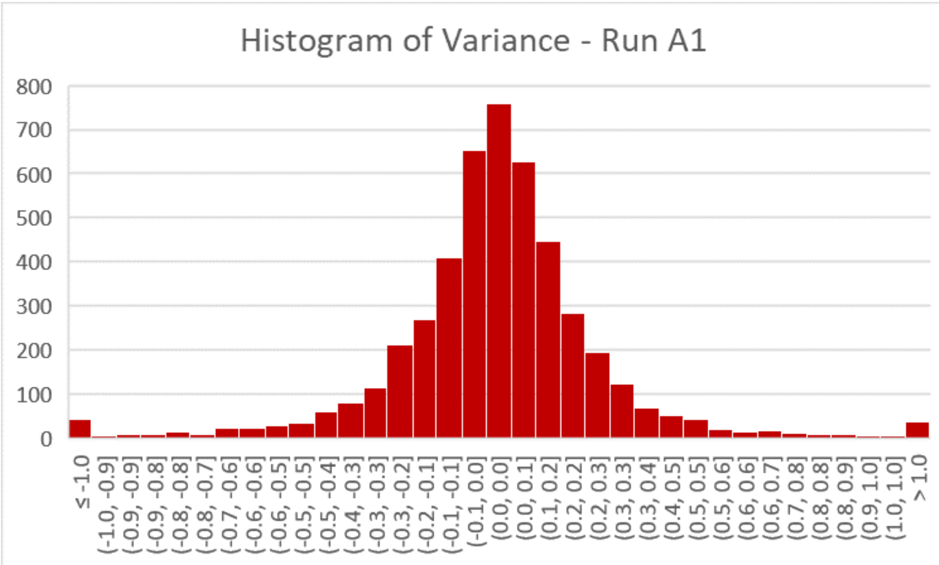


Figure 70: histogram plot of the variance. A1

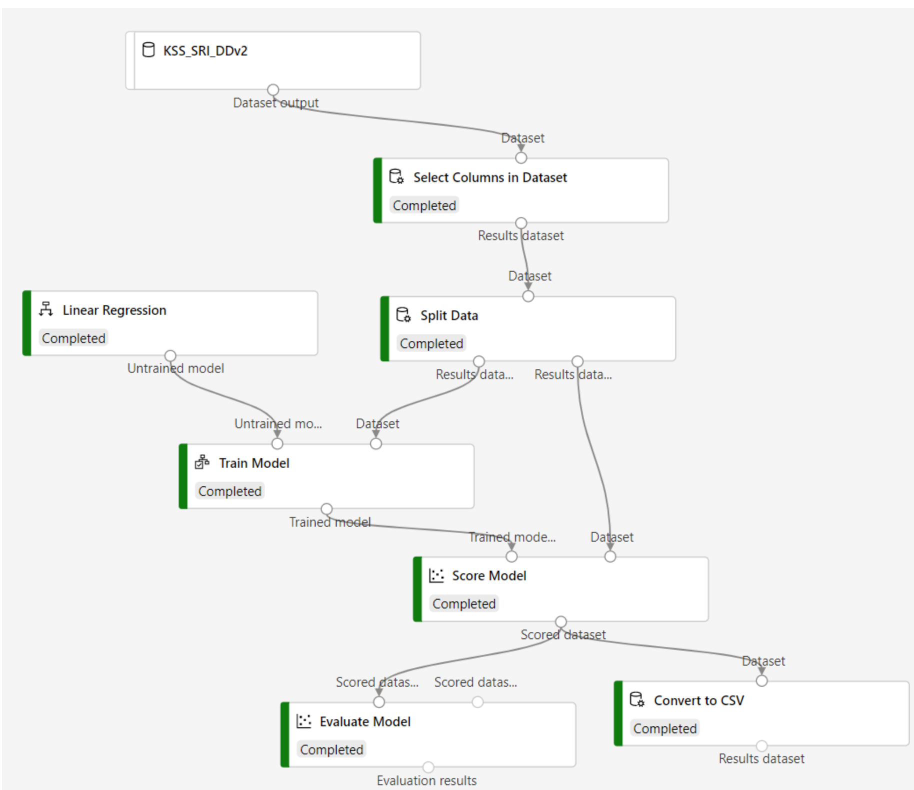


Figure 71: B1 Linear Regression flow chart – 23k records

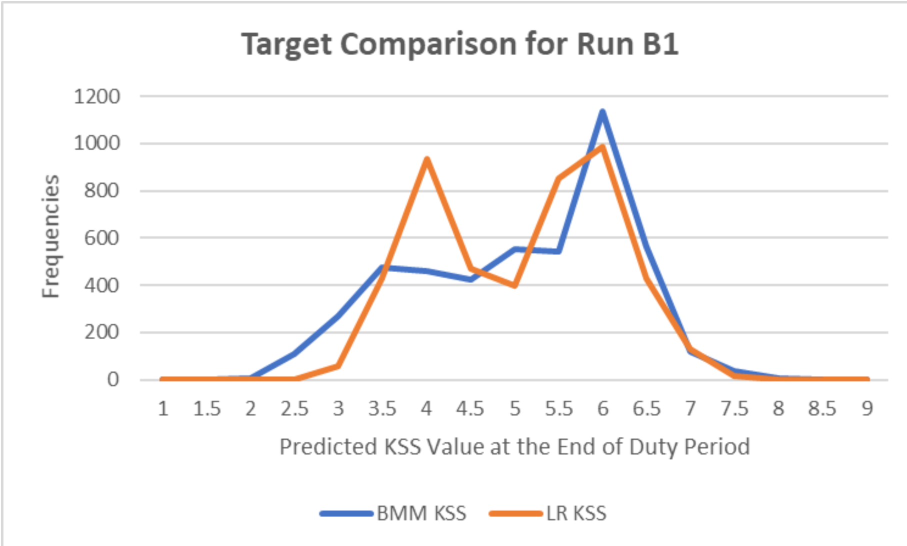


Figure 72: Target vs observed value comparison – B1

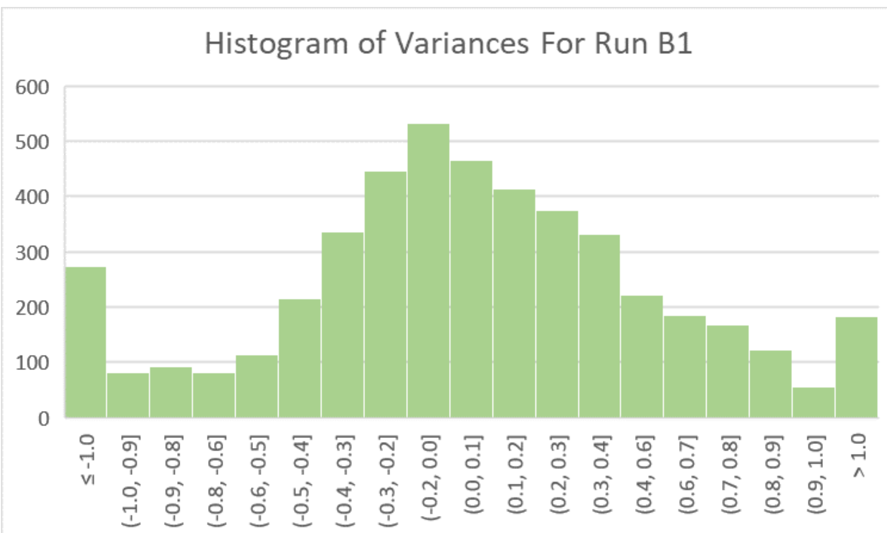


Figure 73: histogram plot of the variance. B1

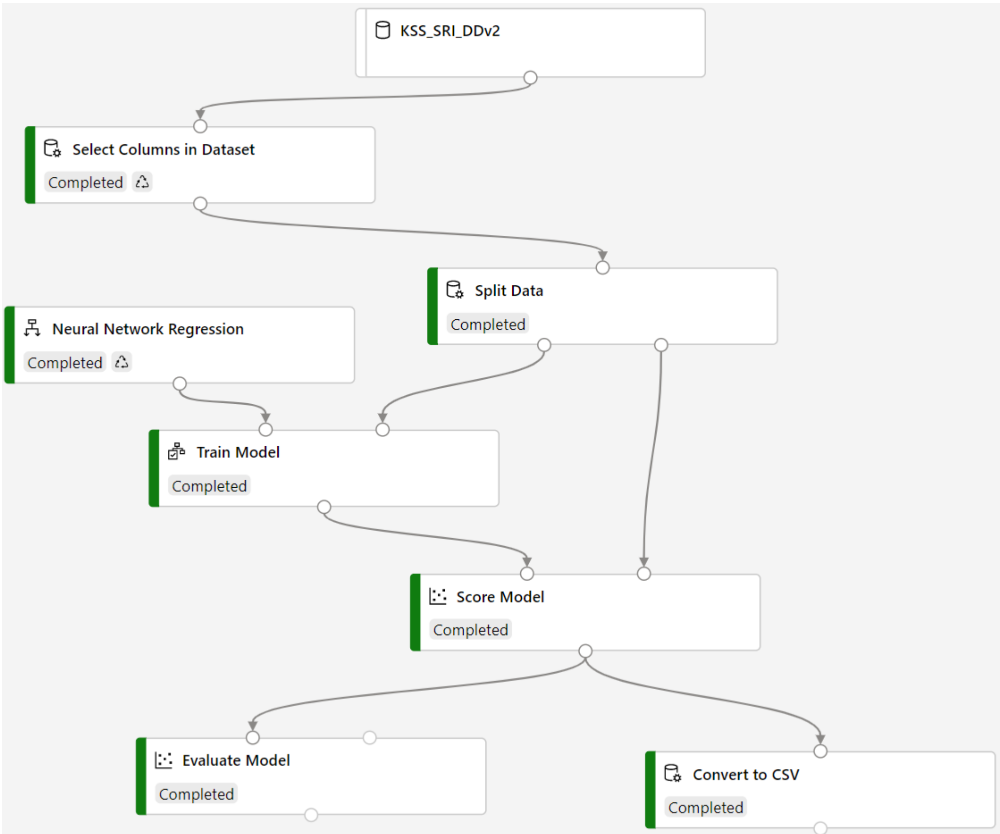


Figure 74: C1 Neural Network Regression flow chart – 23k records

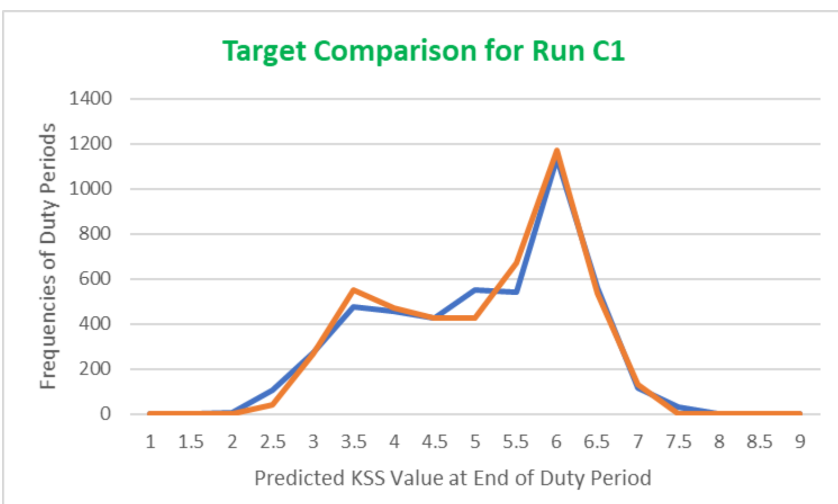


Figure 75: Target vs observed value comparison – C1

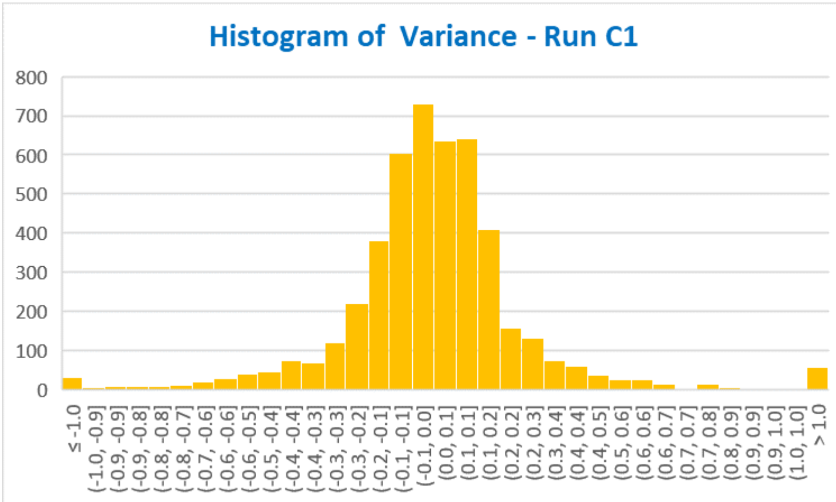


Figure 76: histogram plot of the variance. C1

Figure 77: C2 Neural Network Regression flow chart – 106k records

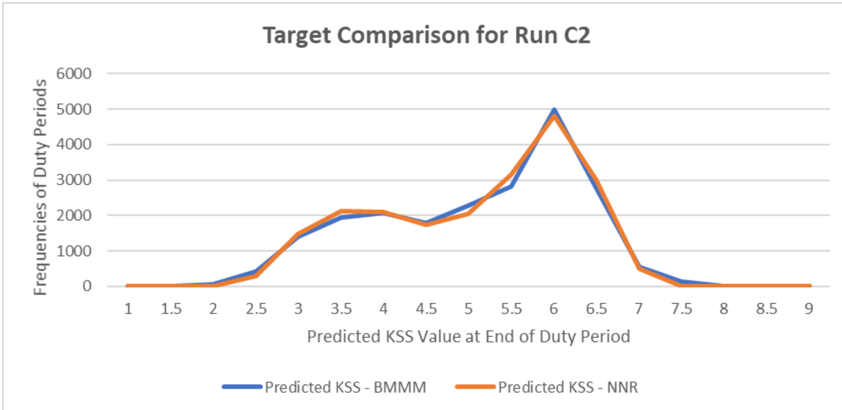


Figure 78: Target vs observed value comparison – C2

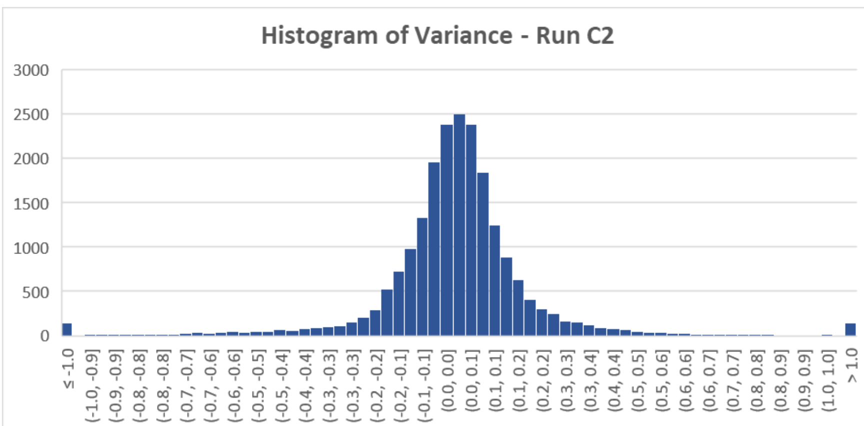


Figure 79: histogram plot of the variance. C2

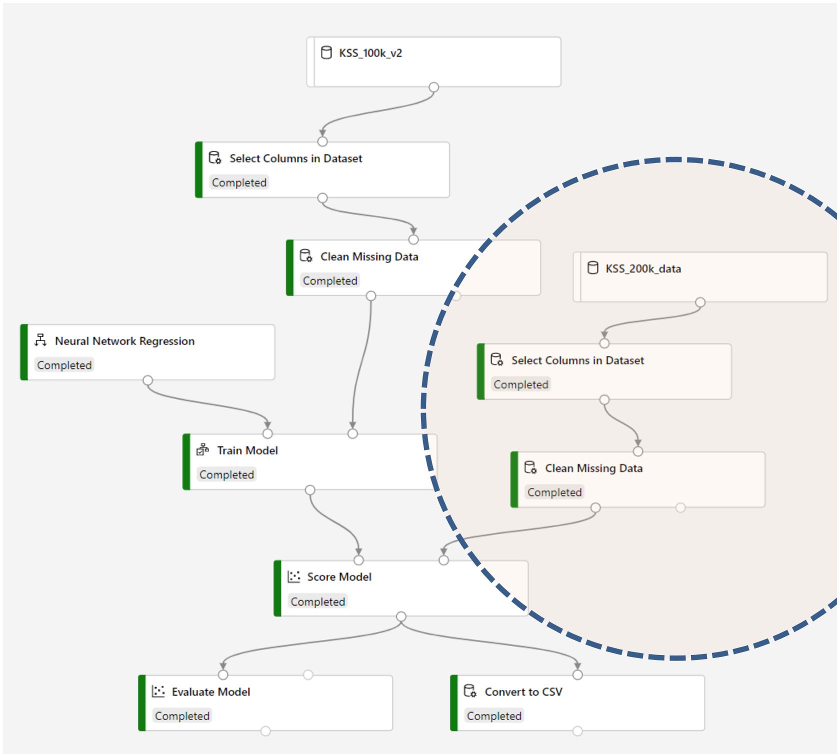


Figure 80: C3 Neural Network Regression flow chart – Model C2 with 242k records for validation

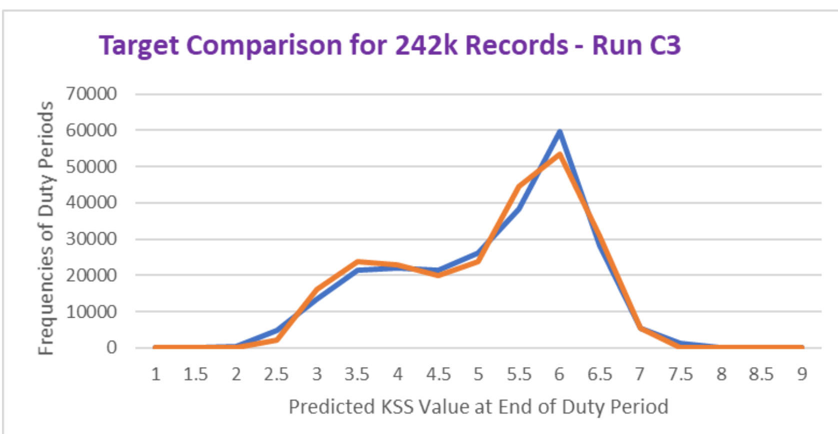


Figure 81: Target vs observed value comparison – C3

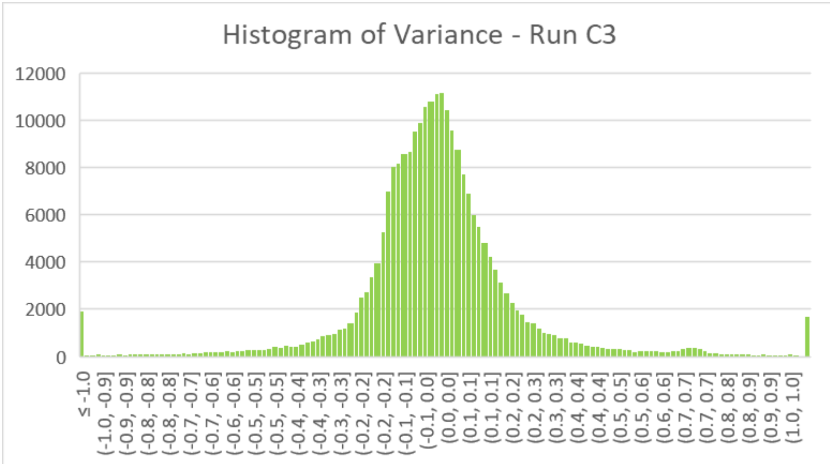


Figure 82: histogram plot of the variance. C3

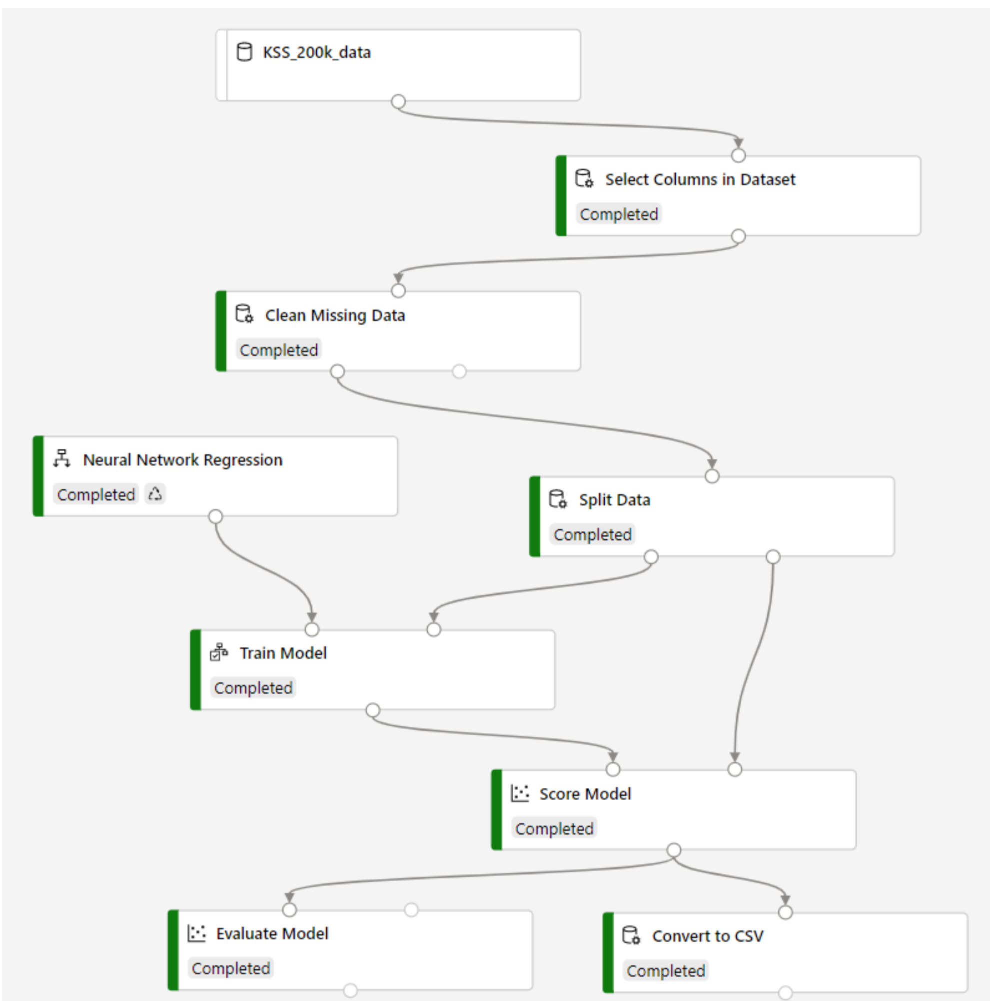


Figure 83: C4 Neural Network Regression flow chart – 242k records

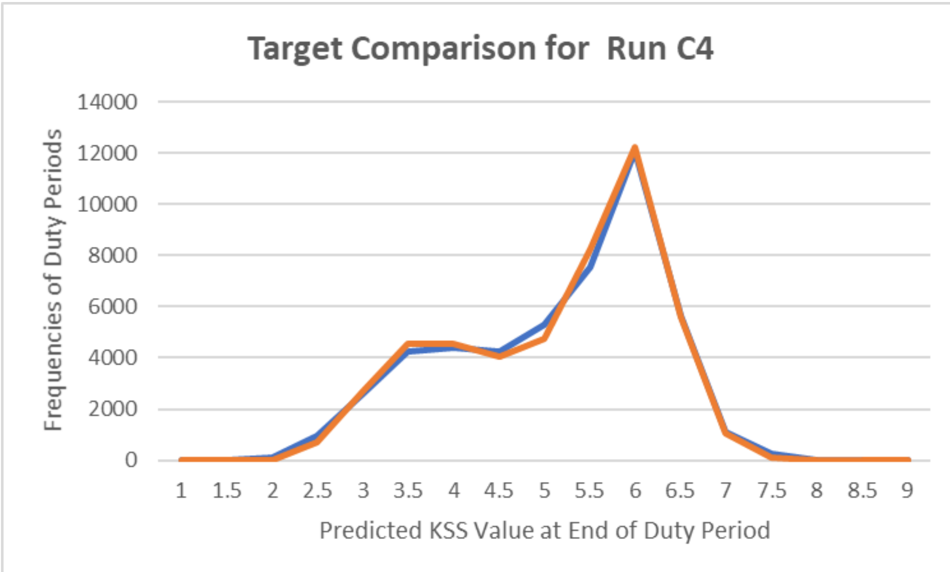


Figure 84: Target vs observed value comparison – C4

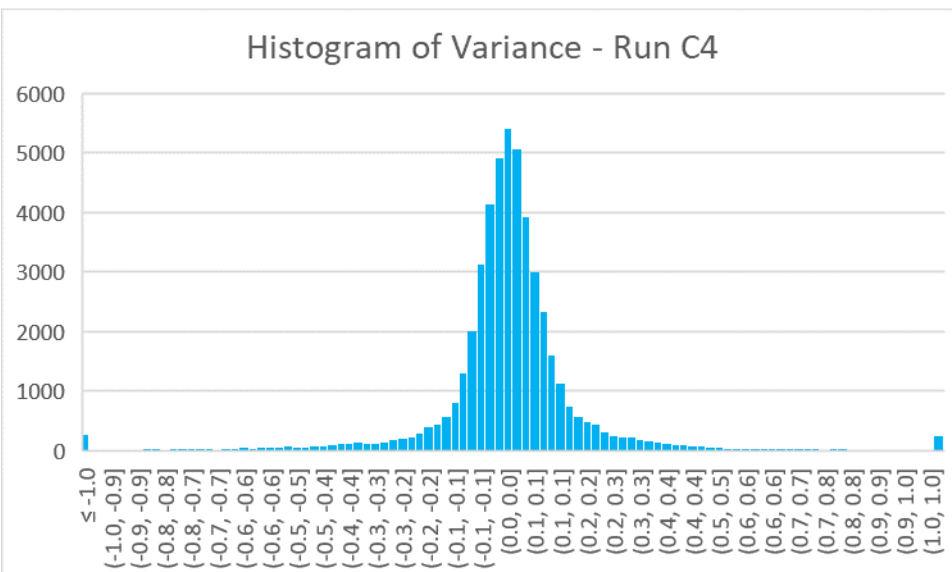


Figure 85: histogram plot of the variance. C4

Appendix - Tables

Table 1. Duty DNA.

Base	domicile of the pilot, three letter airport code
A/C	Aircraft type flown
B_Lat	latitude of the base
B_Long	longitude of the base
Day#	days away from original start time of pilot schedule
LBT_S	Time of duty start with reference to pilot home time.
LBT_E	Time of duty end with reference to pilot home time.
P_L/O	rest time (layover) prior to the duty
Dep	three letter airport code where duty begins
D_Lat	latitude of the departure airport
D_Long	longitude of the departure airport
TZD-Base	time zone differential from base time before beginning the duty
LDepT	local time when the duty start
Arr	three letter airport code where duty ends
A_Lat	latitude of the arrival airport
A_long	longitude of the arrival airport
TZD	Time zone differential between start of duty and end of duty
D_Mins	total minutes from duty start to duty end
B_Mins	total flying minutes from duty start to duty end
MaxTurnCd	airport code where there was turn from one flight to another
MaxTurnTm	amount of time spent between one flight to another within duty.
#Crews	number of crews flown together in the duty
#flts	total number of flights in the duty
DOW	day of the week this duty
Key	combination of number and the start time of the full schedule

Table 2. Features used for Duty Delay Prediction

Column name	Column Description	Example
Base	Base of the pilot	MEM
Nth	days away from base	7
DOW	day of the week	4
Month	month of the year	4
Year	year	2020
hour	hour of the day	14
Min In Day	mins in the day	870
Int	international=y	FALSE
Type	day=D, night=N, critical=C	D
Op in Critical	is duty operaing in C period	FALSE
Prev LO	previous rest time before duty	28
D St	duty start location	MEM
S Tzd	time zone from base at duty start	0
Arr L Tm	arrival time in local mins	1758
D-S	duty scheduled start	5
B-S	flight times in duty period	3.5
T-Time	turn time iwthin duty	0.00
T-Loca	turn location	-
TZD	time zone difference within duty	-1
#Flt	number of flights	1
Opt	optional duty y/n	FALSE
City Seq	sequence of city in the duty	MEMGTF$
Flt Num Seq	sequence of flight numbers in the duty	938$
Tail SeQ	sequence of aircraft tails	0542$
Covid Impact	covid impact month y/n	TRUE
HubTurn	hub turn	0
Duty_variance1	duty variance - delay as measured - train and target. Values 0, 1, 2	1
Scored Probabilities_0	Probabilities of scoring a 0	0.29
Scored Probabilities_1	probabilities of scoring a 1	0.44
Scored Probabilities_2	probabilities of scoring a 2	0.27
Scored Labels	final label that was scored by ML	1

Table 3. Sample records after cleaning and preparing

Base	Nth	DOW	Month	Year	hour	Min In Day	Int	Type	Op in Critic	Prev LO	D St	S Tzd	Arr L Tm	D-S	B-S	T-Time	T-Loca	TZD	#Flt	Opt	City Seq	Flt Num Se	Tail SeQ	Covid Impa	HubTurn	Duty_varia
MEM	0	4	4	2020	8	480	FALSE	D	FALSE	15	BFM	0	1655	9.4	2.5	0.23	MEM	0	2	FALSE	BFMMEM$	662$662$	0860$0746	TRUE	MEM	1
MEM	7	3	4	2020	20	1224	FALSE	N	FALSE	13	GTF	-1	41	4.8	3.3	0.00	-	1	1	FALSE	GTFMEM$	985$	1348$	TRUE	0	0
MEM	0	6	4	2020	2	138	FALSE	C	TRUE	0	MEM	0	625	3.6	2.1	0.00	-	1	1	FALSE	MEMROC$	176$	1528$	TRUE	0	0
MEM	1	5	4	2020	14	845	FALSE	D	FALSE	22	IND	1	1725	5.8	4.3	0.00	-	-3	1	TRUE	INDSAN$	134$	3713$	TRUE	0	0
MEM	0	4	4	2020	2	138	FALSE	C	TRUE	0	MEM	0	601	3.2	1.7	0.00	-	1	1	FALSE	MEMFNT$	723$	1469$	TRUE	0	0
MEM	0	4	4	2020	13	815	FALSE	D	FALSE	0	MEM	0	1719	3.2	1.7	0.00	-	1	1	FALSE	MEMDTW$	321$	0348$	TRUE	0	1
MEM	12	7	4	2020	8	480	FALSE	D	FALSE	26	DEN	-1	1703	9.6	4.2	0.16	MEM	1	2	FALSE	DENMEM$	318$307$	0853$0379	TRUE	MEM	0
MEM	0	4	4	2020	2	168	FALSE	C	TRUE	0	MEM	0	521	5.1	3.6	0.00	-	-2	1	FALSE	MEMLAS$	105$	1440$	TRUE	0	1
MEM	0	3	4	2020	14	885	FALSE	D	FALSE	0	MEM	0	1838	3.4	1.9	0.00	-	1	1	FALSE	MEMRIC$	667$	0794$	TRUE	0	2
MEM	6	4	4	2020	14	895	FALSE	D	FALSE	16	MEM	0	2251	8.4	4.3	0.11	ORF	0	2	FALSE	MEMORF$	724$724$	0399$1338	TRUE	0	0
MEM	4	6	5	2020	18	1089	FALSE	N	FALSE	34	YOW	1	8	6.5	3.4	0.07	BUF	-1	2	FALSE	YOWBUF$	740$740$	0153$0153	TRUE	0	0
MEM	4	1	4	2020	15	900	FALSE	D	FALSE	62	MEM	0	2013	7.7	5.5	0.03	PHX	-2	2	FALSE	MEMPHX$	383$383$	0982$0982	TRUE	0	2
MEM	4	5	4	2020	19	1172	FALSE	N	FALSE	24	PDX	-2	39	5.6	4.1	0.00	-	2	1	FALSE	PDXMEM$	643$	1216$	TRUE	0	0
MEM	0	3	4	2020	21	1274	FALSE	N	FALSE	16	AUS	0	2358	3.2	1.7	0.00	-	0	1	FALSE	AUSMEM$	396$	1227$	TRUE	0	0
MEM	0	3	4	2020	3	186	FALSE	C	TRUE	159	MEM	0	606	3.5	2	0.00	-	0	1	FALSE	MEMHRL$	138$	1537$	TRUE	0	0

Table 4. Data Extraction of Duty DNA for Fatigue Prediction.

Extracted Fields and Duty Construction			Selected and Used for Training
BASE	Base of the pilot	MEM	Y
B_Lat	base coordinates - Latitude	35N	N
B_Long	base coordinates - Longitude	-89W	N
Nth	Days away from base	0	Y
LBT_Dt	Date of Duty	12-May-20	N
LBT_Tm	HHmm - Hours Minutes	1325	N
Int	International = Y	N	Y
Type	Day - D, Night - N, Critical - C	D	Y
OpinCritical	Is Duty Operating in Critical Period that is (0100 - 0500)	N	Y
PrevLO	Rest time before Duty - If 0:00 means its first duty	0:00	Y
D_Org	Duty Start Location	MEM	Y
DO_TB	Time Zone Difference from Base	0	N
D_Dest	Duty End Location	EWR	Y
DD_TB	Time Zone Difference from Base	1	N
Dend_LT	Local End Time wherever the duty is ending.	1801	N
D-S	Total Scheduled Duty Minutes	7:12	Y
B-S	Total Scheduled Flying Time within Duty	3:11	Y
D-A	Total Actual Duty Minutes	8:01	N
B-A	Total Actual Flying Time within Duty	3:18	N
T-Time	Max Turn Time within Duty	1:18	Y
T-Loc	Turning City	ATL	N
TZD	Total Time Zone Difference within Duty	1	N
#Flt	Total Number of Flight Segments in Duty	2	Y
Opt	Is Duty Having Any Optional Duty Segments	N	N
Seq_Loc	Sequence of Flights (OriginDestination) in duty	MEMATLEWR$	N
Seq_Tail	Sequence of Tail Numbers Within duty	787$234	N
Seq_FltNum	Sequence of Flight Numbers within duty	0524$0163$	N
Computed Fields			
DOW	using LBT_Dt, compute this field for day of the week	3	Y
Month	using LBT_Dt, extract month of the year	5	Y
Year	using LBT_Dt, extract year	2020	Y
Hour	using LBT_Tm, extract hour	13	Y
MinInDay	using LBT_Dt, min in the day	805	Y
CovidImpact	if the month is after Apr 2020, we set this flag to Y	Y	Y
HubTurn	if the T_loc goes through a sort facility, then set this field		Y

Table 5. Features, description and sample data for fatigue prediction

Feature	Description	Examples		
Base	Base of the pilot who will work this duty	MEM	MEM	MEM
Nth	Nth day ths duty is operated in sequence	0	0	0
Int	domestic or international duty	FALSE	FALSE	FALSE
Type	C = (0100-0459), D = (0500-1559), Else N	D	D	N
Op in Critical	if duty is flown during (0100-0459)	FALSE	FALSE	TRUE
Prev LO	0 if this is 1st duty, otherwise in hrs	0	0	16
D-S	duty in hours	4.2	6.23	8.92
B-S	flight time in duty in hours	1.95	3.48	3.07
T-Time	total turn time within duty	0.8	1.3	4.3
#Flt	number of flights	2	2	3
Covid Impact	is this duty in covid times	FALSE	FALSE	FALSE
HubTurn	is this duty going through one of hubs	0	0	MEM
DOW	day of the week	3	3	3
Month	month of the year	10	10	10
Year	year	2019	2019	2019
hour	hour of the day	14	14	20
Min In Day	exact minutes in the day this duty starts	840	850	1238
kss_duty_start	kss fatigue score at duty start	2.59	2.58	3.48
kss_duty_end	kss fatigue score at duty end. training var	3.54	4.50	6.23
Scored Labels	target variable for prediction	3.62	4.53	5.88
Diff	variances between scored vs duty end kss	0.08	0.03	-0.35

Table 6. confusion matrix and prominent features for Run D1

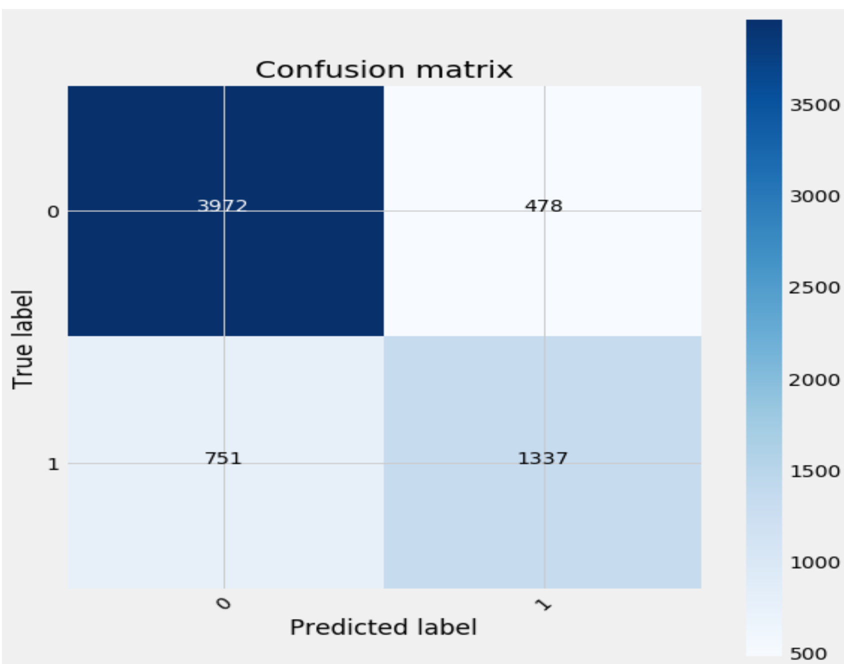


Table 7. Gradient Boosting Run – D1 - Summary

Business Objectives	Duty Delay Predictions		
Run#	D1	D1:D2	D3
Overall Records - 1000s	65		
Train/Test Split	90/10		
Training Data - 1000s	58.5		
Testing Data - 1000s	6.5		
Number of Features	17		
Approach	Binary Classification Approach		
Model	Gradient Boosting		
Overall Accuracy	82		
Recall >30 mins delay	64%		

```
param_grid=
{'n_estimators':range(20),
'max_depth':range(5),
'min_samples_split':range(200),
'min_samples_leaf':range(30)}
```

Table 8. confusion matrix and prominent feature for D2.

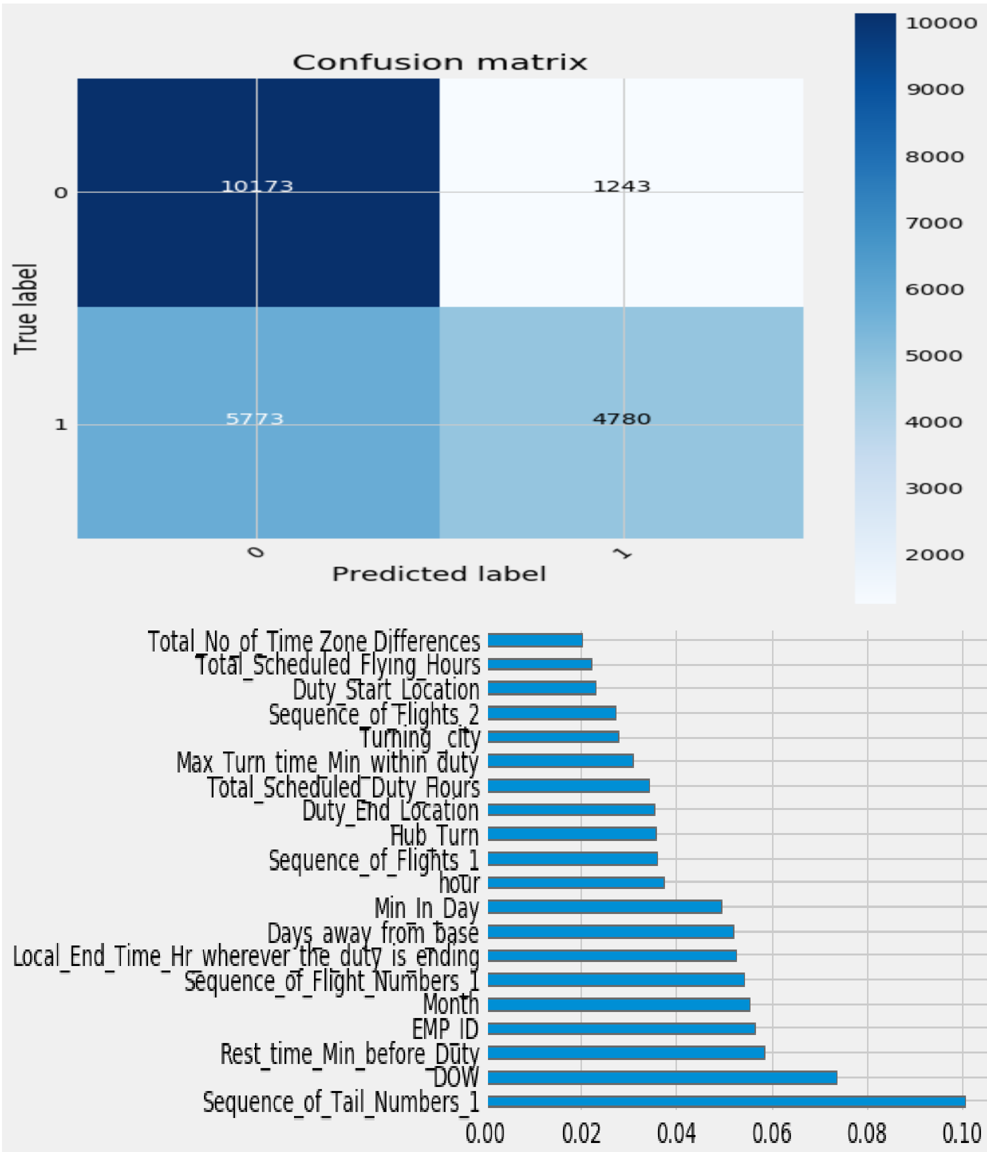


Table 9. Gradient Boosting Run – D2 - Summary

Business Objectives	Duty Delay Predictions		
Run#	D1	D1:D2	D3
Overall Records - 1000s	65	Same as D1	
Train/Test Split	90/10	Same as D1	
Training Data - 1000s	58.5	Same as D1	
Testing Data - 1000s	6.5	21	
Number of Features	17	17	
Approach	Binary Classification Approach		
Model	Gradient Boosting		
Overall Accuracy	82	80	
Recall >30 mins delay	64%	48%	

Table 10. confusion matrix and prominent feature for D3. 215k records

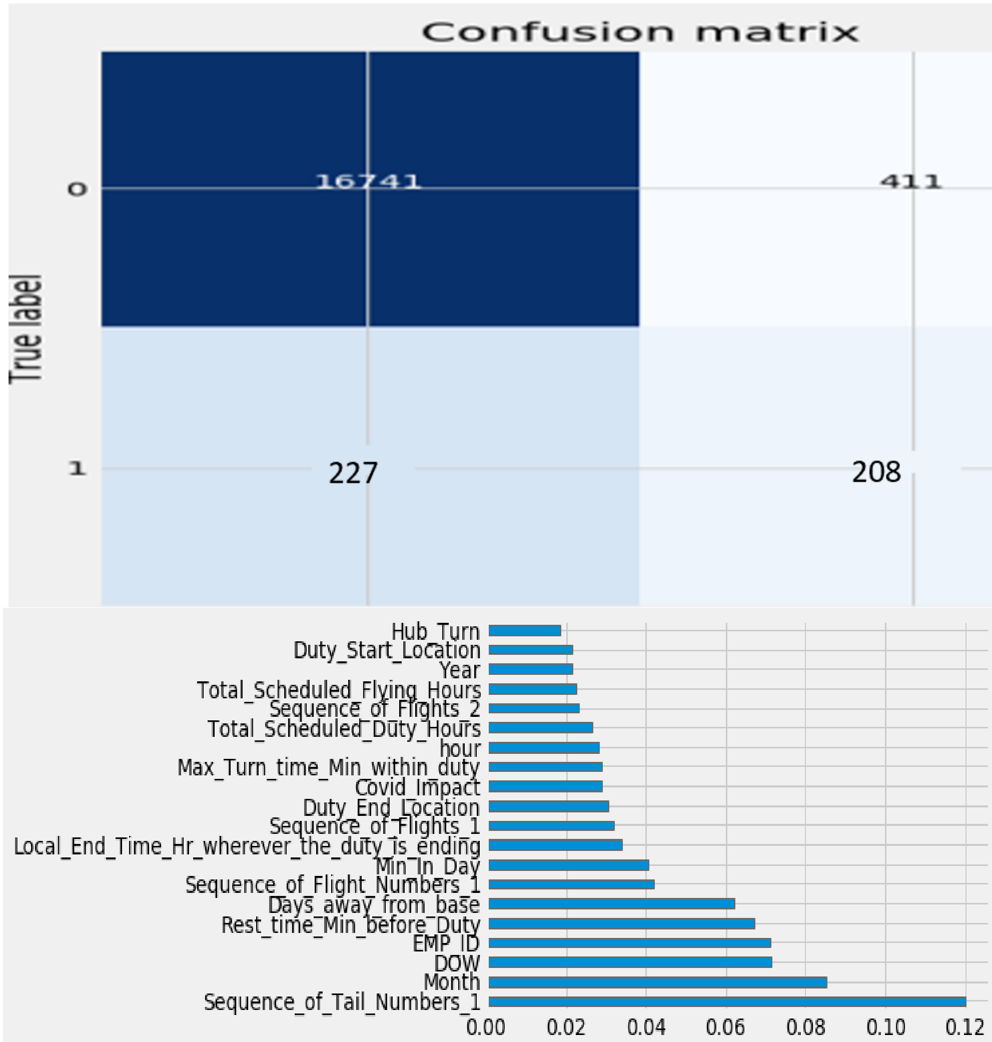


Table 11. Model Run D3 and results

Business Objectives	Duty Delay Predictions		
Run#	D1	D1:D2	D3
Overall Records - 1000s	65	Same as D1	215
Train/Test Split	90/10	Same as D1	90/10
Training Data - 1000s	58.5	Same as D1	193
Testing Data - 1000s	6.5	21	21
Number of Features	17	17	17
Approach	Binary Classification Approach		
Model	Gradient Boosting		
Overall Accuracy	82	80	83
Recall >30 mins delay	64%	48%	48%

Table 12. Run D4 confusion matrix for multi class boosted decision tree classification

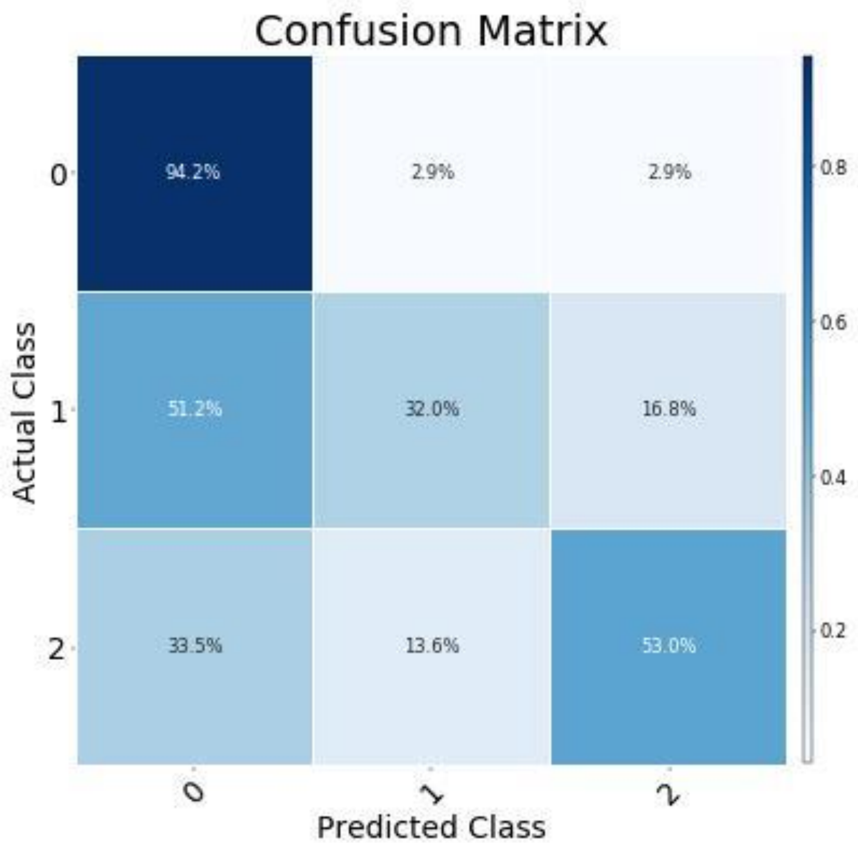


Table 13. D4 run results – multi class boosted decision tree

Business Objectives	Duty Delay Predictions	
Run#	D4	
Overall Records - 1000s	65	
Train/Test Split	80/20	
Training Data - 1000s	58.5	
Testing Data - 1000s	6.5	
Number of Features	17	
Approach	MultiClass Classification	
Model	Boosted Decision Tree	
Overall Accuracy	78%	
Recall >30 mins delay	60%	

Table 14. confusion matrix for multi class neural network classification

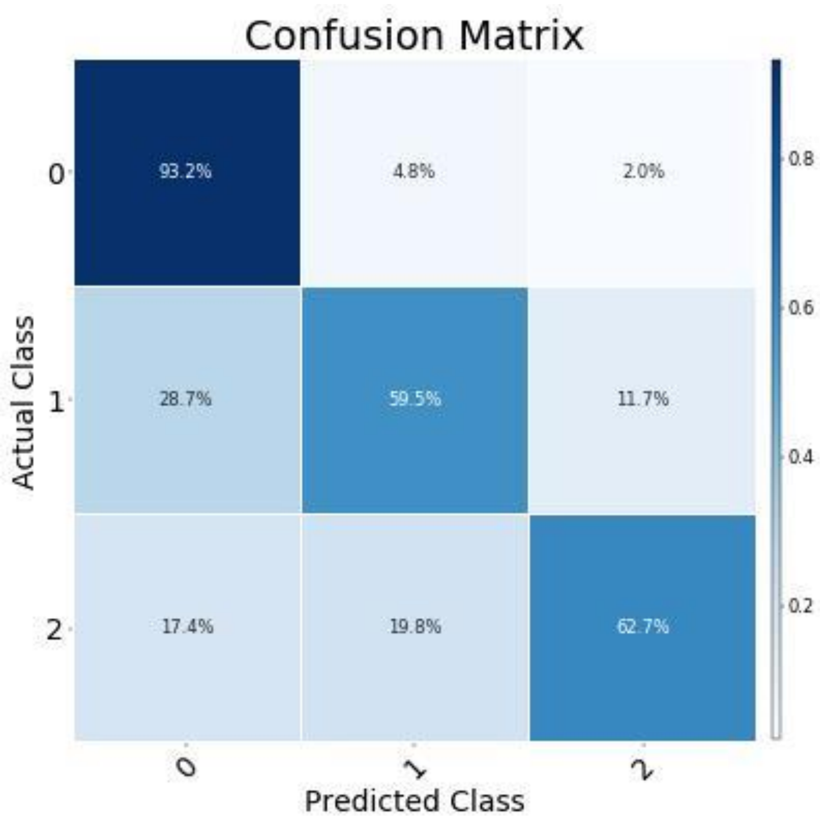


Table 15. D5 run results – multi class neural network decision tree results

Business Objectives	Duty Delay Predictions	
Run#	D4	D5
Overall Records - 1000s	65	65
Train/Test Split	80/20	90/10
Training Data - 1000s	58.5	58.5
Testing Data - 1000s	6.5	6.5
Number of Features	17	17
Approach	MultiClass Classification	
Model	Boosted Decision Tree	Multi Class Neural Network
Overall Accuracy	78%	82.9%
Recall >30 mins delay	60%	71.8%

Table 16. Comparison of all the Delay Models

Business Objectives	Duty Delay Predictions			Duty Delay Predictions	
Run#	D1	D1:D2	D3	D4	D5
Overall Records - 1000s	65	Same as D1	215	65	65
Train/Test Split	90/10	Same as D1	90/10	80/20	90/10
Training Data - 1000s	58.5	Same as D1	193	58.5	58.5
Testing Data - 1000s	6.5	21	21	6.5	6.5
Number of Features	17	17	17	17	17
Approach	Binary Classification Approach			MultiClass Classification	
Model	Gradient Boosting			Boosted Decision Tree	Multi Class Neural Network
Overall Accuracy	82	80	83	78%	82.9%
Recall >30 mins delay	64%	48%	48%	60%	71.8%

Table 17. KSS at Duty Start Predictions Summary for Run K1

Model	Neural Network Regression
Overall FIT	88.3%
Coefficient of Determination	0.8834
Relative Abs Error	0.2599
Relative Squared Error	0.1165
Root Mean Squared Error	0.276
Mean Absolute Error	0.1834

Table 18. A1 evaluation metrics.

Model	Booster Decision Tree
Overall FIT	91.4%
Coefficient of Determination	0.9137
Relative Abs Error	0.2024
Relative Squared Error	0.0862
Root Mean Squared Error	0.2481
Mean Absolute Error	0.1481

Table 19. B1 evaluation metrics.

Model	Linear Regression
Overall FIT	76.7%
Coefficient of Determination	0.7674
Relative Abs Error	0.4206
Relative Squared Error	0.2325
Root Mean Squared Error	0.5727
Mean Absolute Error	0.4303

Table 20. C1 evaluation metrics.

Overall FIT	94.0%
Coefficient of Determination	0.9395
Relative Abs Error	0.1741
Relative Squared Error	0.0634
Root Mean Squared Error	0.299
Mean Absolute Error	0.1781

Table 21. comparison of baseline model parameters and evaluation metrics

Business Objectives	KSS Predictions at End of Duty					
Run#	A1	B1	C1			
Overall Records - 1000s	23	23	23			
Train/Test Split	80/20	80/20	80/20			
Training Data - 1000s	18.5	18.5	18.5			
Testing Data - 1000s	4.5	4.5	4.5			
Number of Features	18	18	18			
Approach	Decision Tree	Regression	Regression - Deep Learning			
Model	Booster Decision Tree	Linear Regression	Neural Network Regression			
Overall FIT	91.4%	76.7%	94.0%			
Coefficient of Determination	0.9137	0.7674	0.9395			
Relative Abs Error	0.2024	0.4206	0.1741			
Relative Squared Error	0.0862	0.2325	0.0634			
Root Mean Squared Error	0.2481	0.5727	0.299			
Mean Absolute Error	0.1481	0.4303	0.1781			

Table 22. C2 evaluation metrics.

Overall FIT	95.8%
Coefficient of Determination	0.9575
Relative Abs Error	0.1304
Relative Squared Error	0.0424
Root Mean Squared Error	0.2463
Mean Absolute Error	0.135

Table 23. C3 evaluation metrics.

Overall FIT	94.5%
Coefficient of Determination	0.9454
Relative Abs Error	0.1641
Relative Squared Error	0.0546
Root Mean Squared Error	0.268613
Mean Absolute Error	0.1603

Table 24. C4 evaluation metrics.

Overall FIT	96.3%
Coefficient of Determination	0.9627
Relative Abs Error	0.1162
Relative Squared Error	0.0372
Root Mean Squared Error	0.2213
Mean Absolute Error	0.1132

Table 25. Comparison of all model parameters and evaluation metrics.

Business Objectives	KSS Predictions at End of Duty					
Run#	A1	B1	C1	C2	C3	C4
Overall Records - 1000s	23	23	23	106	Use Model B	242
Train/Test Split	80/20	80/20	80/20	80/20		80/20
Training Data - 1000s	18.5	18.5	18.5	85		192
Testing Data - 1000s	4.5	4.5	4.5	21	200	30
Number of Features	18	18	18	18	18	18
Approach	Decision Tree	Regression	Regression - Deep Learning			
Model	Booster Decision Tree	Linear Regression	Neural Network Regression			
Overall FIT	91.4%	76.7%	94.0%	95.8%	94.5%	96.3%
Coefficient of Determination	0.9137	0.7674	0.9395	0.9575	0.9454	0.9627
Relative Abs Error	0.2024	0.4206	0.1741	0.1304	0.1641	0.1162
Relative Squared Error	0.0862	0.2325	0.0634	0.0424	0.0546	0.0372
Root Mean Squared Error	0.2481	0.5727	0.299	0.2463	0.268613	0.2213
Mean Absolute Error	0.1481	0.4303	0.1781	0.135	0.1603	0.1132

VITA

Suresh Rangan was born in southern coastal part of India, Chennai. He got his Bachelor Degree in Mechanical Engineering in 1995 from one of the oldest and reputed institution, College of Engineering, Anna University, India. In 1997, he left his mother land and worked in several countries in the Middle East before walking into the land of opportunity in 2001. He came into USA to join FedEx and build airline technology systems. In 2006 he completed his Masters in computer applications course. In 2013, He started working with Dr Mingzhou Jin on the PhD programme. He is expected to complete his PhD in spring 2021. He is always passionate about aviation and his decades of experience in IT and airline crew domain helped him to uniquely position in this crew fatigue research.