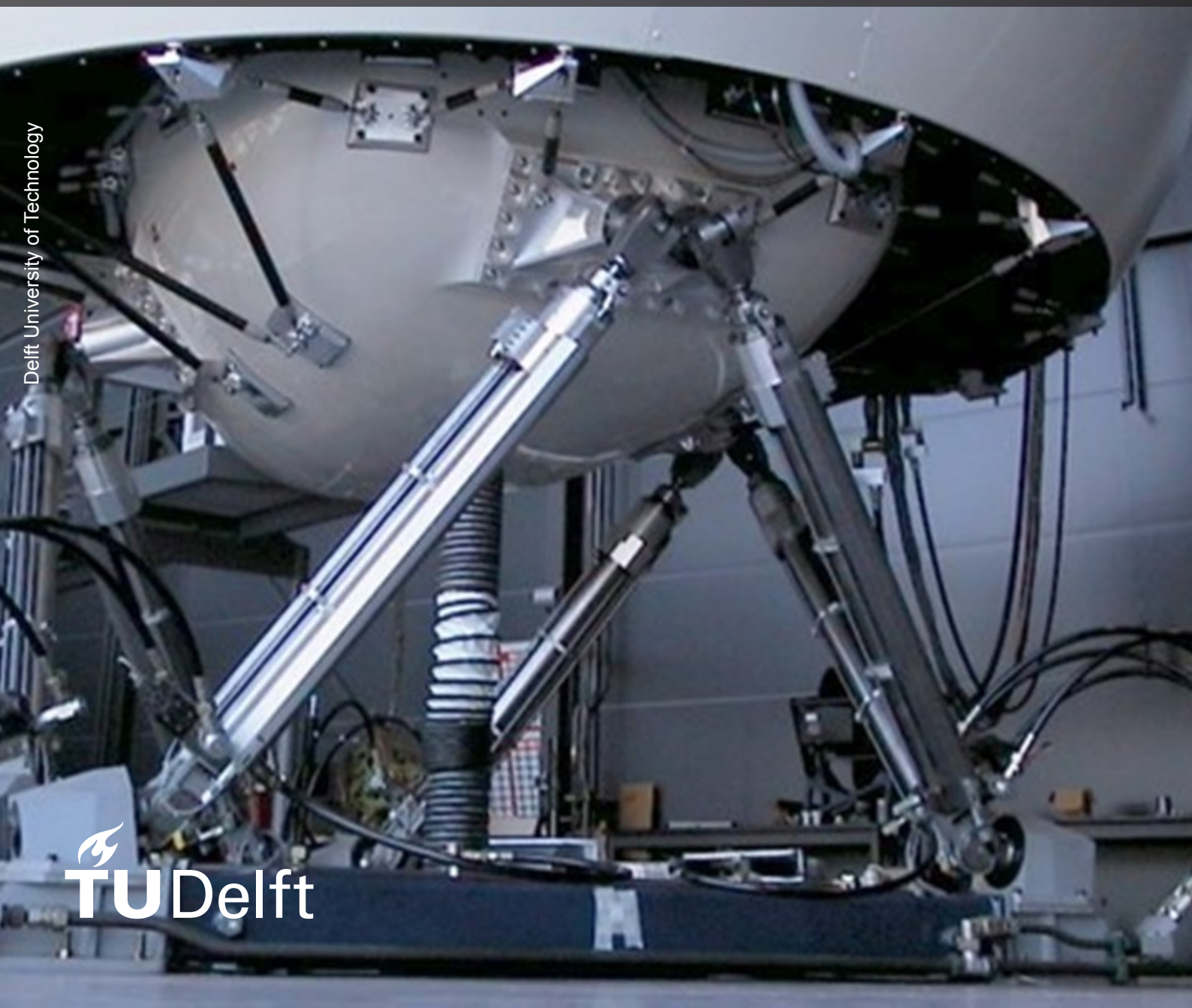
Amplitude Dynamics of Motion Sickness

Master Thesis

Varun Kotian



Delft University of Technology

Amplitude Dynamics of Motion Sickness

By

Varun Kotian

Student Number : 4980611

in partial fulfilment of the requirements for the degree of

Master of Science

in Mechanical Engineering

Track : Vehicle Engineering

at Delft University of Technology,

to be defended publicly on Tuesday September 28, 2021.

Thesis Committee:	Dr. Ir. Riender Happee	Chair	3ME, TU Delft
	Prof. Dr. Jelte Bos	External Member	TNO & VU, Amsterdam
	Dr. Ir. Daan M. Pool	Co-supervisor	LR, TU Delft
	Dr. Ksander de Winkel	Co-supervisor	3ME, TU Delft
	Ir. Tuğrul Irmak	Co-supervisor	3ME, TU Delft

TU Delft

Amplitude Dynamics of Motion Sickness

Varun Kotian

Abstract

Increased susceptibility to motion sickness, due to the transition away from driving, will be one of the major hurdles in the widespread use of automated vehicles. Sustained exposure to motion sickness can lead to the disuse of automated vehicle technology among users. Thus, there is a need to mitigate motion sickness. To do so, a robust model is desired which can predict motion sickness levels while also accounting for individual differences in susceptibility. Studies have been carried out to study the different dynamics of motion sickness and its development. However, the effect of motion amplitude has not yet been quantified.

This study investigated the amplitude dynamics of motion sickness. This was done by perturbing 17 participants along the fore-aft direction with acceleration amplitudes of 1, 1.5, 2 and 2.5 ms^{-2} in four separate sessions. In the experiment, both subjective sickness scores and Galvanic Skin Response (GSR) was recorded. Using the subjective sickness scores, we explored variations of the Oman model of nausea that could capture the time-amplitude dynamics that were observed. Along with this, a neural network prediction model was proposed to predict motion sickness levels using GSR as a predictor, which could then be used as an objective measurement of motion sickness. Lastly, to better understand the MISC as a measure of motion sickness, we determined the functional mapping between subjective discomfort and the MISC scale.

Our findings show a monotonous increase in the rate of motion sickness development, but this increase is not linear, with a sudden change in slope after 1.5 ms^{-2}. This nonlinear increase was captured best by applying a power law to the input conflict vector instead of at the output as proposed in the original Oman model (Oman, 1990). Further, it is found that GSR can indeed be used as a predictor for motion sickness with an accuracy of 77% in the training, 66% in the validation and 62% in the holdout set. Finally, it is found that the subject discomfort has a power-law relation with MISC with a mean exponent of 1.28.

The developed prediction model as well as the variation of the Oman model can be used in other experiments to control the level of sickness to the desired trajectory. The developed models can enable adaptive control algorithms for path and motion planning to mitigate motion sickness. This is will in turn lead to a significant improvement in the experience in commuting with an automated vehicle.

1 Introduction

Automated vehicles are the future. In the next decade, self-driving cars (SAE level 4-5) are expected to be available to the public with a 20-40% market share by 2040 (Litman, 2020). Their advent is hoped to improve safety and comfort for their passengers. Although they have many advantages, they still have to pass an important hurdle. This being the increased prevalence of motion sickness (MS) experienced due to a transition away from active driving, in which control of the vehicle confers immunity from sickness, to passively riding as passengers, which does not (Diels and Bos, 2016). Motion sickness is a malady caused by external motion stimuli which trigger autonomic responses such as salivation, dizziness, headaches, panting, hot/cold flushes, stomach awareness, nausea and vomiting (Bertolini and Straumann, 2016). Sustained exposure to sickening motions can induce sopite syndrome, which is characterized by lethargy, fatigue and drowsiness (Matsangas et al., 2014). This can discourage some people from using automated vehicle technology. Instead, they may be inclined to drive the vehicle manually, thereby making the automated vehicle technology redundant.

Reason and Brand (1975) have concluded that all people with intact and functioning vestibular systems can be made motion sick. Thus, there is a need to study, model and minimize MS and its effects. Various methods are being employed to reduce MS. One way of doing so is through route and path optimization. Another way is by using active suspension to attenuate the frequencies causing MS. All these methods require motion sickness development to be modelled such that the subsequent sickness predictions of this model can be used.

Reason (1978) argued that motion sickness was due to a mismatch between sensed and expected sensory signals by the brain. These expectations originate from an internal model, which is also referred to as the neural store. It was hypothesized that this mismatch causes the internal model to adapt and, consequently, get habituated to the sickening stimuli. Oman (1982) compared this hypothesized model to the nature of a Luenberger Observer (LO). The LO consists of an internal model of the system (body) and sensor dynamics. Due to the noisy nature of the sensory signals, the sensory afferents cannot be used directly for motion perception and state estimation. The true states are therefore estimated by integrating sensory information using an internal model of the system itself. These are then used for task planning and execution. To calculate the conflict, the estimated states are input through the internal model to produce expected sensory afferents, which are then compared to the actual sensory afferents. This conflict is then used to guide the estimated body motions towards the true state, and also to adapt the parameters of the internal model, which in turn can allow for better state estimation. It is hypothesized that the magnitude of the conflict and the duration of exposure then leads to the consequent symptoms of motion sickness.

There are a number of model formulations used to simulate motion sickness (Oman, 1990; Bos and Bles, 1998; Kufver and Förstberg, 1999; **?**; McCauley et al., 1976; Lawther and Griffin, 1988). However, each has a different perspective on how MS development is dependent on the sensory conflict. O'Hanlon and McCauley (1974), Lawther and Griffin (1988) and Kufver and Förstberg (1999) infer the sensory conflict to be a linear function of translational and rotational motions. While the models by Oman (1990), Bos and Bles (1998) and **?** estimate the sensory conflict as a nonlinear function of the true conflict. Furthermore, Oman (1990) and Kufver and Förstberg (1999) also predict the MS development through a nausea magnitude estimate in time instead of the Motion Sickness Incidence (MSI), which is the percentage of people who have vomited from a given stimulus. These have been studied for the effect of various factors such as frequency and exposure time on motion sickness. However, one component that is not yet studied in sufficient detail, is how sickness develops over time and how this development relates to the amplitude of the sickening motions endured.

The study of such amplitude dynamics involves the evaluation of the system response to differing amplitudes of stimuli. However, the relation of how the amplitude scales the motion sickness development is not yet clear. For simple motions, such as a single degree of freedom vertical, or horizontal accelerations, the MS development is taken to be proportional to the stimulus intensity as explained by Lawther and Griffin (1988). However, the amplitude of accelerations was only up to 0.7 ms^{-2}. At such small accelerations, it is possible for the group dynamics to be approximately linear. This claim is strengthened when taking a look at the data by O'Hanlon and McCauley (1974) where a sigmoid relationship between acceleration amplitude and MSI seems to exist. Furthermore, a nonlinear relationship was also found by Bos and Bles (1998) wherein a nonlinear Hill function was used to scale the conflict after rectifying the conflict vector which was then integrated using a second order function to fit the MSI observations, from O'Hanlon and McCauley (1974).

The approach of using two different mechanisms for conflict generation, which is due to the spatial orientation and state estimation, and conflict integration, which leads to MS, is unique. The only drawback here is that these studies use models that predict MSI, which is the percentage of people who will get motion sick, which is only good when studying the group as a whole, rather than the individual sickness responses. For a physiologically robust model, motion sickness should be quantified by individual ratings and not group averaged ones.

An example of a model capturing the individual motion sickness responses is the model developed by Oman (1990). Here, the input of the model is the magnitude of the rectified conflict and the output is sickness, which can be mapped onto any sickness rating scale. The model has two paths, one fast and one slow. The addition of both gives the output of the model. The benefit of the two paths, is that it helps describe the phenomenon of hypersensitivity Oman (1990). Hypersensitivity is the phenomena in

which humans exhibit heightened sensitivity to a new motion stimulus after getting exposed to sickening motions. This is usually observed during the rest or recovery phase after the sickening motions, and may last for minutes or even hours.

In a previous work by Irmak et al. (2020), the Oman model was validated in the context of simple motion sickness generated by slalom manoeuvres performed by a car at 0.2 Hz and 4 ms^{-2} for up to 30 minutes. Here, it was seen that the Oman model provided a good fit to subjective sickness scores as measured on the MIsery rating SCale (MISC) during both the initial motion exposure and the hypersensitivity trials. The MISC is defined on an 11 point scale from 0 to 10, where 0 is no symptoms and 10 is vomiting (Bos et al., 2005). Moreover, using the Oman model parameters governing the trajectory of motion sickness in this paradigm was effectively used to predict responses in a re-exposure to the same paradigm. This indicated a high degree of repeatability in the measured sickness response.

However, it is not clear whether the amplitude to sickness relationship proposed by the Oman model holds true. It may only be able to predict responses to different amplitudes by tuning of the parameters, rather than in a rule based fashion. Literature on subjective sickness dynamics of individuals shows a complex relation between amplitude and motion sickness.

Studies have also been carried out which observed the time response of each individual to a given stimulus. Graybiel (1969) used a slow rotation room combined with intentional head movements (from shoulder to shoulder, around the fore-aft axis) to induce motion sickness. After every 1000 head movements, the rotation speed was increased by 1 RPM. The habituation and susceptibility of each subject were studied. Miller and Graybiel (1969) observed that the symptoms worsen rapidly when the rpm is too high for an individual, and may not even develop if the rpm is too low. It was concluded that motion sickness depended on the individual susceptibility and the rotational velocity of the room. Bock and Oman (1982) attempted to use Stevens magnitude estimation method Stevens (1987) to determine whether subjective discomfort in motion sickness could in fact be governed by Stevens Psychophysical Law. They, too, studied the effect of head movements on the time course of motion sickness on individuals. In their experiments, the temporal summation of head movements made during the training period brought subjective discomfort above the threshold, and a subsequent sequence of head movements triggered an "avalanche" effect. It was suggested that subjective discomfort is related to the conflict stimulus by a power law with an exponent greater than one.

In a previous work by Irmak et al. (2020), with experiments in a car following a slalom route, it was seen that removal of external visual cues led to qualitatively different sickness dynamics. Here, for the same participant, the internal vision condition created a more "avalanching" increase in sickness severity than the external vision condition, which was more gradual. It is not clear whether such switching behaviour is mediated by just the amplitude of the sickening motions, or a combination of both amplitude and duration of effect of these sickening motions. As in the experiments of Oman (1982), the conflict could not be quantified. Thus, the effect of amplitude, that is the amplitude dynamics, on MS development was not quantified.

Literature on subjective sickness dynamics of individuals is indicative of a complex relation between amplitude and motion sickness. However, this relationship has not been quantified. By doing so, we hope to both extend our understanding of motion sickness development, but also develop ways to control sickness.

In addition to a model that can simulate the time course of motion sickness, a model to predict motion sickness is also needed to be able to control sickness levels in real time. For this, a specific indicator is needed which can even consider individual differences. It is well known that sweating is a symptom of motion sickness (Reason and Brand, 1975). Because the electrodermal activity is associated with sweating, many researchers have used it as a physiological indicator of motion sickness. Studies by Parker (1971) showed that an increase in skin conductance is a predictor of seasickness. Cowings et al. (1986) observed that subjects who reported severe symptoms of motion sickness also displayed a significant decrease in skin resistance during rotational stimulation. Many other studies like by Warwick-Evans et al. (1987) and Hu et al. (1991) also found a high correlation between skin conductance and motion sickness.

Thus, further improving upon this, we aim to predict the motion sickness level of an individual using

GSR. This will help in controlling the motion sickness levels of the participants to required levels in experiments or in real case scenarios like in automatic vehicles. This has important methodological implications for motion sickness research. Currently, the researcher must fix their stimulus prior to the experiment and hope that participants become sick, but do not terminate the experiment prematurely. Active control will allow for setting the desired level and variance of motion sickness in an arbitrary manner, this will increase the statistical quality of data collected.

Out of the various scales for measuring sickness, MIsery SCale (MISC) seems to be preferred, as used by Lie Hok Lien (2019), Irmak et al. (2021) and Reuten et al. (2020). This is because it is symptom based, which allows for the anchoring of subjective sickness intensity to a set of symptoms that are similar between participants. However, the MISC is on an ordinal scale, and its values do not rise linearly with respect to the severity of discomfort experienced by the participants. This makes it difficult to directly interpret MISC predictions. Hence, there is a need to identify the mapping between the MISC and the magnitude of the subjective discomfort.

In the following sections, a description will be given about the experiment methods, their results, models describing motion sickness generation and time evolution, and finally conclusions drawn from it.

Based on the literature, we hypothesize that:

1. With increased amplitude of motion stimulus, the subjects show,
    - Earlier onset of symptoms
    - Increase in the slope of the time course of motion sickness development
2. Skin Conductance Level is a good predictor of the level of Motion Sickness.
3. A power law relation exists between subjective discomfort and the MISC Scale.

Building onto the proposed hypothesis, our objective is to:

1. Develop a variation of Oman model that can generalize well in its predictions of the time course of sickness for different amplitude conditions
2. Develop a prediction model using skin conductance as a predictor for motion sickness
3. Find the functional relationship between the MISC and subjective discomfort

In the present study, we assessed the relationship between conflict magnitude (using acceleration amplitude as a proxy) and the time course of motion sickness symptoms at an individual level. We did this by having participants exposed to sinusoidal fore-aft motions of four different acceleration amplitudes. The motions were generated using the SIMONA Research Simulator at the Aerospace Engineering faculty of TU Delft (Stroosma et al., 2003; Berkouwer et al., 2005). In the subsequent analyses, we relate the time course of measured sickness to the amplitude of the fore-aft acceleration, which is taken as a proxy for the amplitude of sensory conflict. Based on this, we extended Oman's model of nausea to model the development of sickness in mixed acceleration environments. Furthermore, the effect of MS on Skin Conductance Level (SCL) will also be studied. This will then be used to study the efficacy of using SCL as an indicator of MS levels.

2 Experiment Method

2.1 Ethics Statement

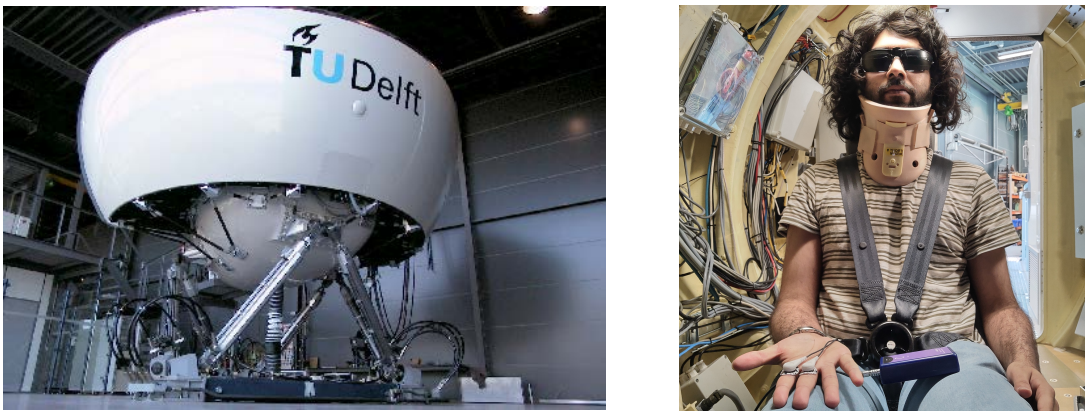
All participants provided written informed consent prior to participation. The experimental protocol was approved by the ethical committee of the Human Research Ethics Committee of TU Delft, The Netherlands, under application number 1425.

2.2 Participants

A within-subject experiment was conducted on a total of 17 participants (15 males and 2 females). Participants were screened based on their motion sickness susceptibility and proper functioning of their vestibular system. Participants who mentioned that they had never experienced motion sickness were not selected for the experiment. The participants were between 22-31 years old ($\mu = 25.3$, $\sigma = 2.6$). Participants were either staff or students at Delft University of Technology. The participants were reimbursed with a voucher of €10 for their participation. The four conditions were presented to the subjects in a random order to reduce order effects. A balanced Latin square was used to randomize the order for the first 16 participants. For the last participant, the order of the first participant was repeated.

2.3 Apparatus

The experiment was conducted using the SIMONA Research Simulator (Figure 1a) at the faculty of Aerospace Engineering of TU Delft. Participants (Figure 1b) were seated in the simulator, facing forwards with their eyes blinded using blackened glasses, but were asked to keep their eyes open, and restrained to the seat using a five-point harness and a neck restraint. To measure skin conductance, Mind Media's Nexus 4 was used and connected using BioTrace+ software. This device is capable of measuring skin conductance at a sampling rate of 32 Hz. The electrodes used to measure skin conductance were placed on the index and middle fingers. In addition to these, the cabin lights were switched off to remove any reference for orientation and ear plugs were used to reduce the perception of motion through the noise of the simulator actuators. Two-way communication was established via an intercom system through which the participants communicated with the experimenter.



(a) SIMONA Research Simulator (b) Participant

Figure 1: Experiment Setup

2.4 Design

Low-frequency horizontal oscillation is a cause of motion sickness in many forms of land transport. Griffin and Mills (2002a,b) have found no significant difference in motion sickness due to fore-aft and lateral oscillations. Furthermore, there was more stroke available in the simulator in the longitudinal direction than in the lateral direction(Koekebakker, 2001). Thus, a longitudinal(fore-aft) motion signal was used as the stimulus in the experiment.

The frequency of the excitation signal was chosen based on the studies by Donohew and Griffin (2004). In this study, a frequency of 0.03 - 0.3 Hz proved to be the frequencies of high sensitivity, that is the frequency range for which the susceptibility to MS is highest. Due to the constraints of the simulator and to allow a wider range of accelerations, the frequency used in the experiment is 0.3 Hz which is a bit higher than the frequencies of high sensitivity, but enough to elicit MS.

The amplitude of the motion signal in the experiment was constrained by the stroke and velocity of the simulator. The velocity and acceleration depends on the frequency of the sinusoidal wave. Thus, by varying the frequency and using all the available stroke, the maximum acceleration can that can be achieved was found to be 2.5 ms^{-2}. As the maximum value of amplitude that can be employed in the simulator, for our given excitation signal, limited by the stroke of the simulator, is 2.5 ms^{-2}, amplitudes of 1.0, 1.5, 2.0, and, 2.5 ms^{-2} were chosen as the levels for the acceleration amplitude. These were chosen to allow a large range of amplitudes to be tested. These selected conditions allow to investigate the differences in progression of motion sickness from 1 to 2.5 ms^{-2}.

An experiment session was divided into four phases, with two excitation (Motion 1 and 2) phases and two rest phases (Rest 1 and 2) as shown in Figure 2. The duration of the 2 excitation phases was 60 and 30 minutes, and the duration of the rest phases was 10 and 5 minutes. The stimulus used in the excitation phase was a sinusoidal motion in the fore-aft direction with a frequency of 0.3 Hz and amplitude of accelerations of 1.0, 1.5, 2.0 and 2.5 ms^{-2}, respectively. The order of conditions was balanced using a Latin square to mitigate order effects.

Using the intercom system, the participants were asked to give out MISC scores at a beep sound, which was programmed to be at every 30 seconds. The participants were also informed that they could also give a score any other time if they felt their MISC had changed. The skin conductance level (SCL) was also measured by placing two electrodes on the participant's left-hand index and middle fingers. Figure 2 shows the four phases of an experiment session along with the MISC response, SCL and the input signal for one of the participants.

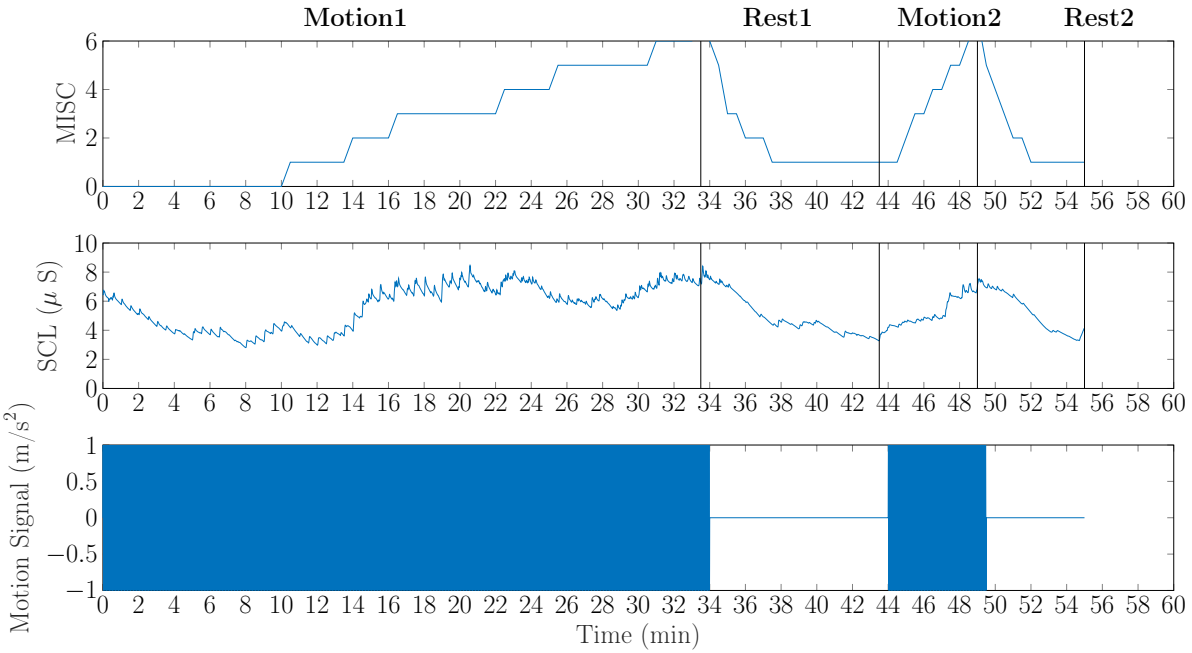


Figure 2: Phases of the Experiment

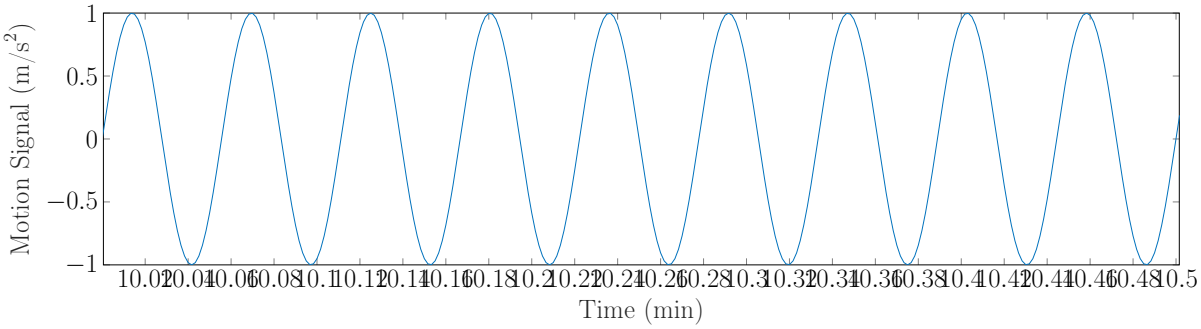


Figure 3: Zoomed Input Motion Signal

Also, shown in Figure 3 is the zoomed in plot of the motion signal. This plot shows the sinusoidal input signal with frequency of 0.3 Hz and amplitude of 1 ms^{-2}.

2.5 Procedure

Each condition was tested on participants with a rest of at least 1 week ($\mu = 22.8$days, $\sigma = 19.8$days) in between two test conditions. The rest days are given to avoid habituation to input motion, as observed by McCauley et al. (1976). The participants were asked to fill MSSQ-Short ($\mu = 16.19$, $\sigma = 10.05$) (Golding, 2006) prior to the experiment. This was done to measure each participant's sickness susceptibility to know whether they are above or below the average susceptibility. This was also used to screen for participants who do not get motion sick. The mean MSSQ-Short score was 16.19 which gives a percentile of 66 which means that the average sickness susceptibility was above 66% of the population. Before the experiment, a briefing about the MISC and the procedure was also be given.

The participants were subjected to the motion exposure while recording MISC at intervals of 30 seconds or whenever the participant felt a change in MISC score. The motion was followed by a rest period, and then the second excitation phase was started. The motion was stopped before the end time of the excitation phase if the participant either reached a maximum MISC score of 6, or wanted to end the experiment.

Finally, after the experiment, the participants were asked to map the subjective discomfort recorded on the MISC, onto the subjective discomfort. They were asked to rate their subjective discomfort by assigning a reference value of 100 to the subjective discomfort they experienced when at a MISC score of 4. If the participant did not reach a MISC score of 4, then the peak sickness level reached would be taken as the reference instead. Furthermore, they were asked to rate a few Verbal Quantifiers, ranging from "Excellent" to "Terrible". The scores were given based on the magnitude estimate reported while mapping the MISC onto the subjective discomfort. This procedure was repeated for all four conditions.

2.6 Data Analysis

2.6.1 Sickness Metrics

Various metrics have been calculated to quantify sickness. For initial assessment of sickness levels, the mean MISC, which is the average MISC across all participants at each moment of time, is used. However, there were not enough data points near the last time instances due to many participants terminating the experiment early. This resulted in skewed mean MISC near the end time. To fix this, the MISC values for such participants were extrapolated to a value of 6 until the end time of the session. This extrapolated mean MISC gives a better perspective of MS development in the experiment for all four amplitudes. The only disadvantage of this approach is that the mean MISC will converge to 6 for all the four amplitudes and reduce the discrimination due to the amplitudes at the end of the motion exposure, which might not be the case in a real scenario. Followed by this, dropout rates, that is the number of participants who terminated the experiment at a given time, were calculated to understand the severity of each amplitude. Additionally, to capture the rate of MS development, average MISC gradient was used, which was calculated as the mean slope of MISC response with time at each data point. Furthermore, the absolute slope of MISC with time was calculated to give the MISC rate. This was calculated to find out the difference in severity of sickness at the end of each motion exposure. Besides this, the time to reach each level of MISC for each of the motion exposures was calculated.

2.6.2 Oman Model Fitting

For fitting the Oman Model, the input signal is sampled at 20 Hz, which is much higher than the sampling frequency of 0.0333 Hz for MISC readings. This done so that the sinusoidal signal is captured properly. To match this sampling rate, the MISC ratings were linearly interpolated.

Furthermore, for each individual participant, data from all 4 sessions, of different motion amplitudes, were combined. Both the two motion and the two rest phases were included to capture the hypersensitivity and recovery. This combined data is then used to fit the Oman model parameters for each individual participant.

2.6.3 SCL Data

The skin conductance data was analysed in MATLAB via Ledalab V.3.4.9 (www.ledalab.de) using the Continuous Decomposition Analysis (Benedek and Kaernbach, 2010). Here, the skin conductance is returned as a continuous measure of tonic and phasic electrodermal activity at a sampling frequency of 16 Hz. Using this, the correlation of MISC response to each of the components of GSR is found. Furthermore, the viability of GSR as a measure of MS level was tested by fitting a neural network model. To reduce the computation power and time, the electrodermal activity data was down-sampled to 0.1 Hz. The MISC ratings were sampled every 30 seconds. To match the sampling frequency of MISC rating to the EDA data, the MISC rating was interpolated. This was only done for fitting a model to predict sickness from GSR data.

2.6.4 Participant Exclusions

For participant 9, the MISC score jumped nearly instantaneously from MISC 0 to 6 (see Figure 4), suggesting that either the symptoms did not develop or the time taken to reach nausea was not enough for the symptoms to be perceived. However, the SCL was in correlation to the MISC scores given by the participant. ruling out the possibility of an error in giving score by the participant. In addition to this, data during the rest phases in the first motion condition (motion amplitude of $1ms^{-2}$) is missing due to the participant temporarily adjourning the session.

Due to the above-mentioned reasons, participant 9 was excluded for the Oman model parameter fitting. Moreover, the model too was not able to capture the sudden change in the MS level after long periods of having no symptoms.

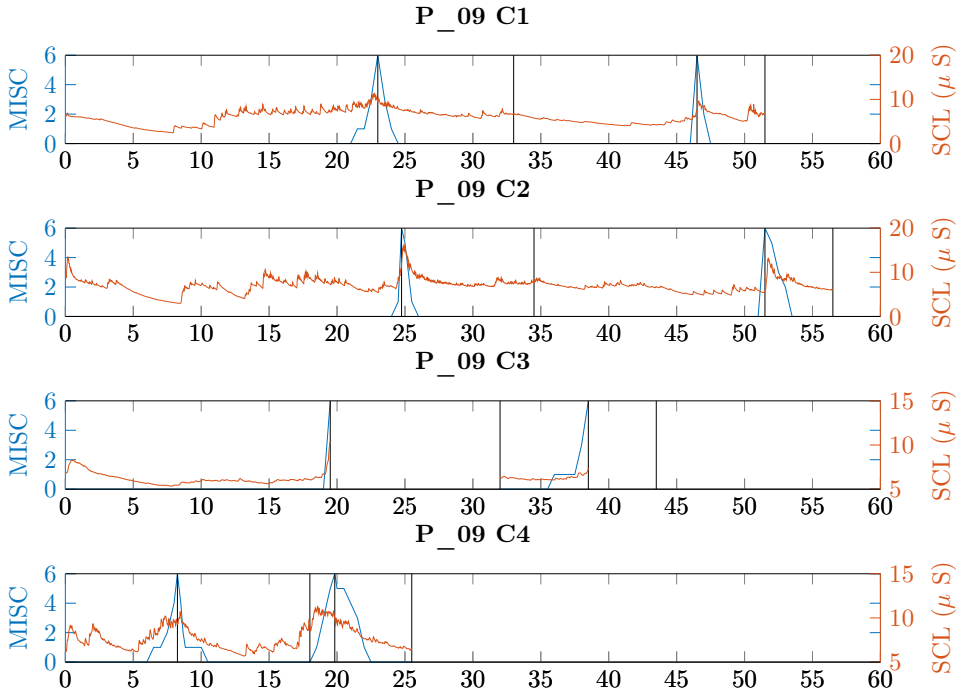


Figure 4: MISC and SCL response for Participant 9

On the other hand, Participant 11 showed a normal MISC and SCL responses. However, the magnitude

estimates of the MISC scores onto the Steven's rating scale was a few orders higher than those of the other participants. Hence, this participant was excluded in the analysis of the data pertaining to the Steven's rating scale.

2.6.5 Statistical Analysis

As multiple measurements of the same variables have been taken for the same participants, but under different amplitude conditions, this results in a repeated measures design study. To analyze such a study, a repeated measures model needs to be used. In such a model, the within-subjects factor are the different amplitudes in this study, while the between-subject factor are the participants. This was analyzed using Repeated Measures Analysis Of Variance (ANOVA) to find out if a significant effect exists between conditions. These test the hypothesis that each column of data is draw from a population with the same mean against the alternative hypothesis that the population means are different. A p-value less than 0.05 indicates significant effect. To test that the assumptions for RANOVA are true, Kolmogorov-Smirnov test for normality and Mauchly's test for sphericity was used. To compare the Oman model fits, AICc (Akaike information criterion corrected) was used. The steps used to calculate this are shown in Appendix E.

3 Results

MISC and SCL responses for all participants are shown in Appendix A. All participants completed at least three sessions, with 15 participants completing all four sessions. One participant terminated the first session after the first motion exposure due to emesis, but still returned for the other three sessions and completed them. Another participant stopped the experiment temporarily during the two rest phases due to retching at both sessions. All but one participant reached a MISC of 6 in at least one session. In all, 2 participants experienced emesis. The mean MISC increased with time and is shown in Figure 5 and Figure 6 during the course of the first motion exposure for all four amplitude conditions. In Figure 6, the MISC values are extrapolated to a maximum of 6 for participants who dropped out early. As seen in both the figures, the MISC at a given point of time increases with amplitude, thus proving the expected positive relation between amplitude and MISC. Figure 7 shows the dropout rate during the first motion exposure for all four amplitudes. It is noticed that the highest dropout rate is for the amplitude of 2.5 ms^{-2}, followed by 2.0, 1.5 and 1.0 ms^{-2}, which is as expected. Another interesting observation in Figure 6 and Figure 7 is how close the values for 2.5, 2.0 and 1.5 ms^{-2} are at 60 minutes, while the value at 1.0 ms^{-2} is at a considerable gap from the others. This suggests a higher change in the rate of the MS development from 1.0 to 1.5 ms^{-2}. Additionally, it is observed that the average rate of increase (gradient) in sickness increases with the motion amplitude. This is clearly seen in Figure 8 where the box plot of the mean rate of increase of MISC with time is plotted across the four amplitudes. Using this average gradient in a repeated measures model and carrying out repeated measures analysis of variance, it was found that there exists a significant difference ($p = 0.01$, ANOVA) between the four motion conditions. Furthermore, in the same way, the order effects were also studied using the average MISC gradient. It is found that no significant difference ($p = 0.71$, ANOVA) exists due to the order in which the different motions were presented. This proves that there is no habituation between sessions.

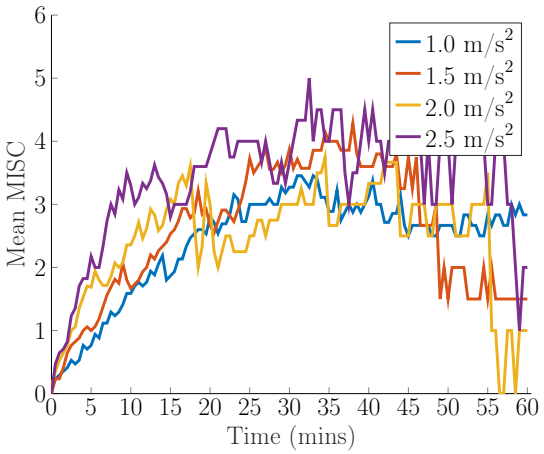


Figure 5: Mean MISC vs Time

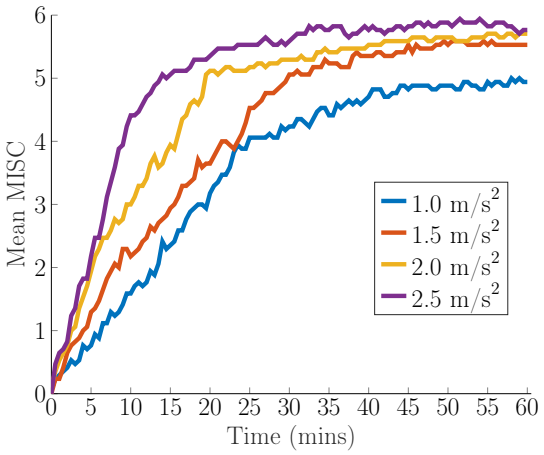


Figure 6: Mean MISC with extrapolated scores vs Time

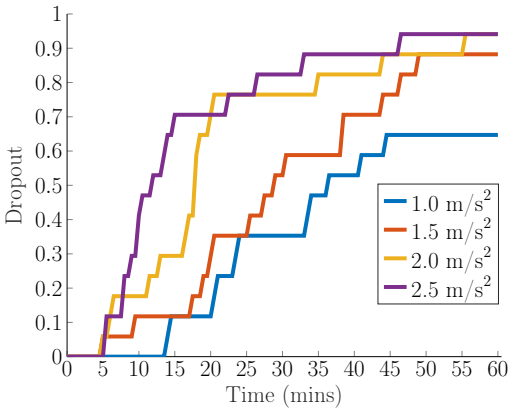


Figure 7: Dropout Rate vs Time

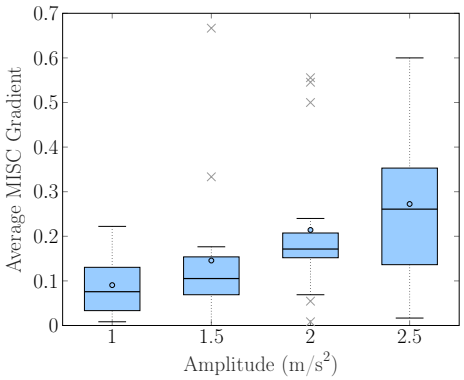


Figure 8: MISC Gradient vs Amplitudes

3.1 MISC Rate

The MISC rate was computed to better understand the severity of motion sickness at the end of each amplitude session. This is given by the MISC score at the end of the first motion phase, divided by the time taken to reach the end of the first motion phase. Further, to remove the effects of the individual differences in susceptibility of the participants, the MISC rate was normalized using a 2-norm. This way, the Figure 9 shows the increase in MISC rate with amplitude. This shows that even though the participants may reach a MISC of 6 in all motion conditions, the rate of increase was higher at higher amplitudes. The mean MISC rates during the first motion exposure for the four amplitudes were 0.16, 0.21, 0.35 and 0.52. It was observed that there was a sudden change in increase in mean MISC rates, from 0.05 to 0.15, after 1.5 ms^{-2}. This could indicate a sigmoidal or a power law relationship between the sickness response and the amplitude of the motion signal.

The same trend also holds during the hypersensitivity phase, see the light blue bars in Figure 9. The MISC rates show a similar trend as shown during the first motion exposure. However, the magnitude of MISC rates is much higher during the second motion exposure phase, proving that hypersensitivity does occur. This is seen more predominantly in Figure 10.

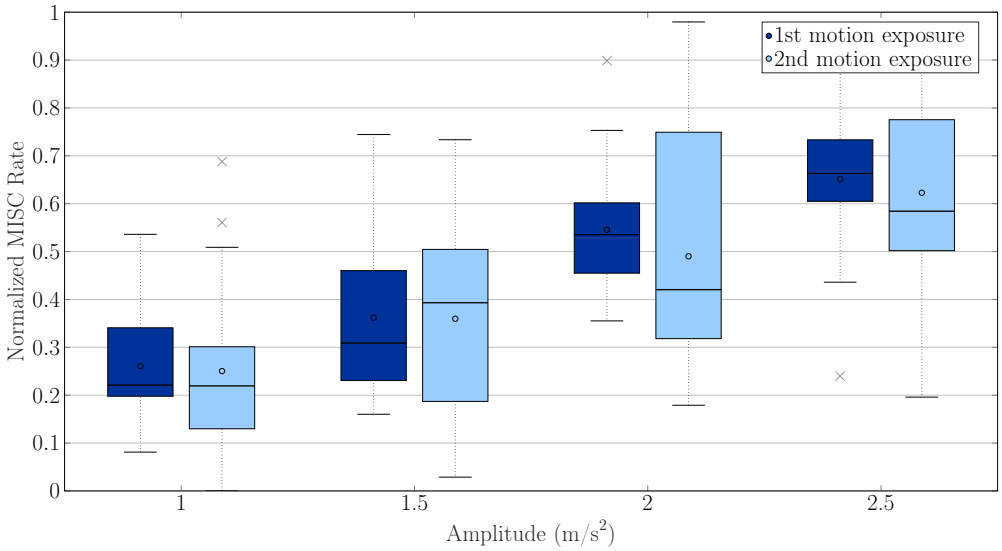


Figure 9: Normalized MISC Rate

3.2 Time to Sickness

With increasing amplitude, a monotonous increase in MISC levels and a monotonous decrease in the time it took to reach a certain MISC level was observed. This can be clearly seen in Figure 10, where the lightest colour is time to reach MISC 1, and the darkest colour is the time to reach MISC 6. This figure shows the distribution of time taken to reach a particular level of MISC for different amplitudes for all participants during the two motion phases. For participants who did not reach a MISC level, maximum time was allotted to them at that particular MISC. The box covers the 25th and 75th percentiles, with the horizontal line inside the box showing the median and the circle showing the mean of the times. Furthermore, the plot also shows that the time to reach a particular MISC level also increases with the respective MISC levels. Hence, it is concluded that the development of severe symptoms always precedes the start of nausea (MISC 6).

Hypersensitivity is also observed during the second motion exposure, followed after the 10-minute rest, see Figure 2. The Figure 10 also shows the second motion phase as well in red. The decrease, by an average of 61%, in time distinctly shows the presence of hypersensitivity. Hence, implying an increase in sensitivity during the second exposure, as expected and observed by Oman (1990).

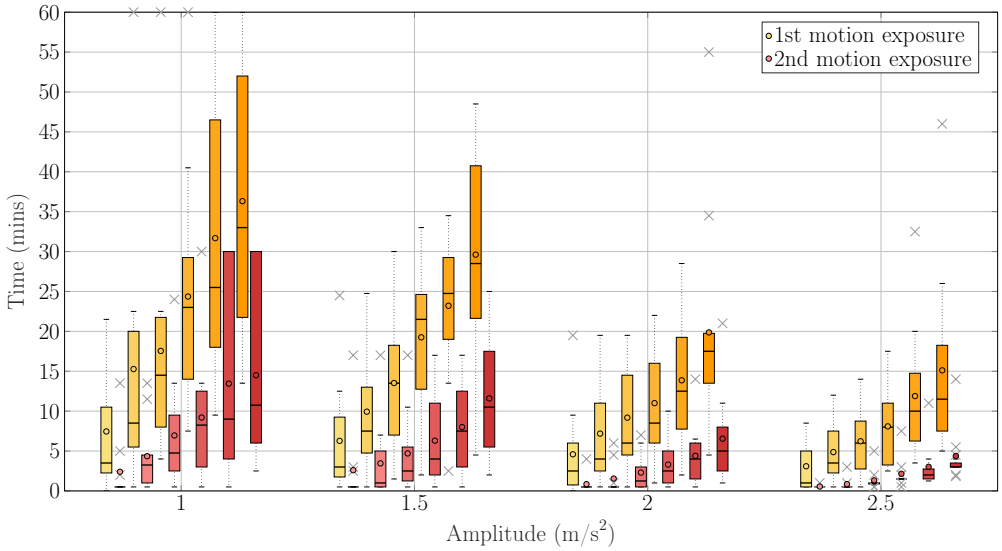


Figure 10: Time to different levels of MISC

3.3 Modelling of MS Development

A model is needed which can reliably predict the amplitude dynamics of motion sickness for a given participant. Hence, in the following subsections, the models developed will be the ones that can be generalized across all amplitudes. In all the models, the amplitude of the motion signal is taken as the proxy for the conflict magnitude. First, the MISC data was fit on a simplified Oman model, and then variations of the model were tested to try and fit the data better.

3.3.1 Oman Model

As discussed earlier in section 1, the Oman model (Oman, 1990) has two paths, fast and slow, whose addition gives the sickness level. This sickness level can be mapped onto any scale like vomiting incidence, nausea magnitude, or in this case, to MISC. The reasoning behind the two paths is that it helps capture the phenomenon of hypersensitivity and also the slow recovery after exposure to sickening motion.

A reduced form of the Oman Model (Oman, 1990) was used to model the time course development of motion sickness. In the reduced form, the threshold term was removed to reduce the number of parameters to be estimated. The Oman model was simplified, to reduce the computation time and the power needed, to have only the two time constants (β_1 and β_2) and two gains (K_1 and K_2) and the power law term. The threshold term was omitted in all the analyses. This was done because the threshold term only scales down the output of the dynamic system in the model and thus will not effect the dynamics (time constants and the gain) of MS development significantly. In addition to this, Oman (1990) has not stated any value for this threshold and thus leading to the speculation of it being used effectively in the Oman model itself. Figure 11 shows the simplified model used for fitting. Moreover, to further reduce the model, a relation between the two time constants was found in section 3.3.1.

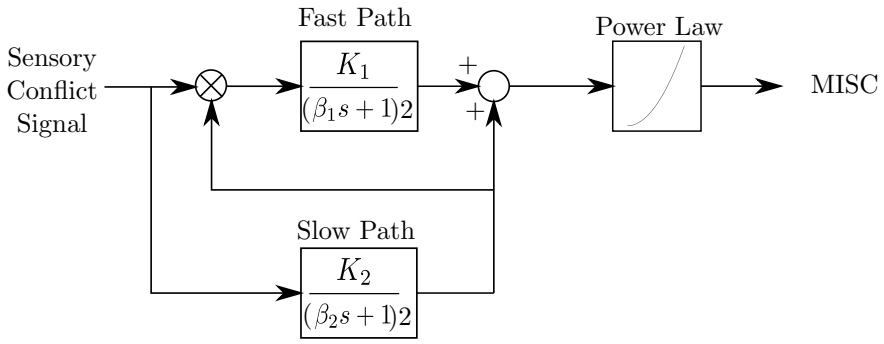


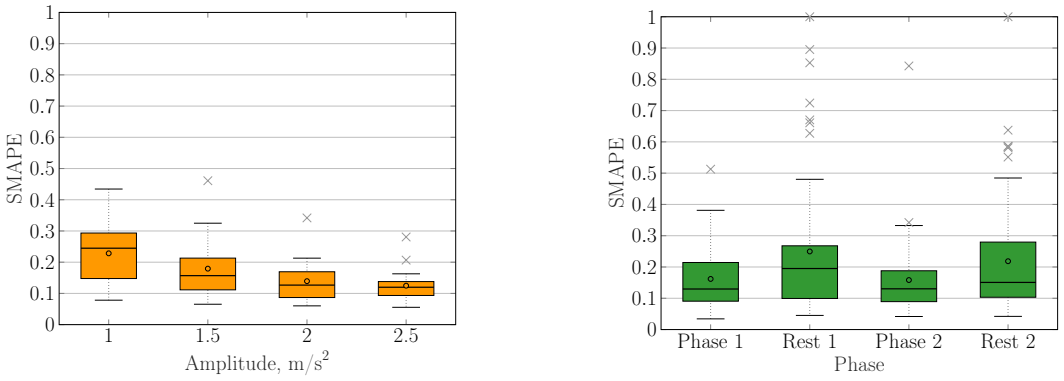
Figure 11: Oman Model

A constrained minimization problem was formulated for fitting the model. The absolute value of the motion signal is assumed as the sensory conflict, which is the input to the model. This assumption was made by considering the works of O'Hanlon and McCauley (1974) and Lawther and Griffin (1988) where a similar relationship is used. The output is the MISC score reported by the participants. The value to be minimized was the Symmetric Mean Absolute Percentage Error (SMAPE) of the actual MISC rating and the simulated rating from the model across the four motion conditions. The detailed structure of the formulated optimization problem is given in Appendix C.The SMAPE is defined as follows:

$$\text{SMAPE} = \frac{\sum_{t=1}^{n} |F_t - A_t|}{\sum_{t=1}^{n} (A_t + F_t)}$$

Where, F_t is the forecast value, and A_t is the actual value.

The advantage of using SMAPE is that the errors have fixed lower (of 0) and upper (of 1) bounds. SMAPE when multiplied by 100 will give the error in a percentage which is more intuitive to understand. However, when the actual and the forecast value tend to zero, it can induce instability. Also, if either the forecast value or the actual value is zero, the error will hit the upper bound. But this will only be the case if the participant does not get sick for a long duration of time. Nonetheless, SMAPE was used in the optimization problem due to its ability to evaluate models better as it gives a relative error.



(a) across different amplitudes

(b) across different phases

Figure 12: Error box plots for Oman Model

Figure 12 shows the errors for the fitted Oman model. As seen, the error increases for the lower amplitudes. This is because the model underestimates the MISC ratings at lower amplitudes. This is observed in all the models which have been proposed throughout this study and can be clearly seen in Figure 23. This may be because of the lower number of data points in the higher amplitudes, which result in an early convergence of parameters and get refined over the time the lower amplitudes fit. Additionally, as the duration of the experiment is longer at the lower amplitudes, there could be higher variance in MISC scores, which may be absent at higher amplitudes, and thus producing a monotonous increase in MISC values. In addition to this, the model is not able to describe the rest phases. Here, not only the errors are large, but the variance in errors is also high, which can be seen for participant 1 and 2 in Figure 13.

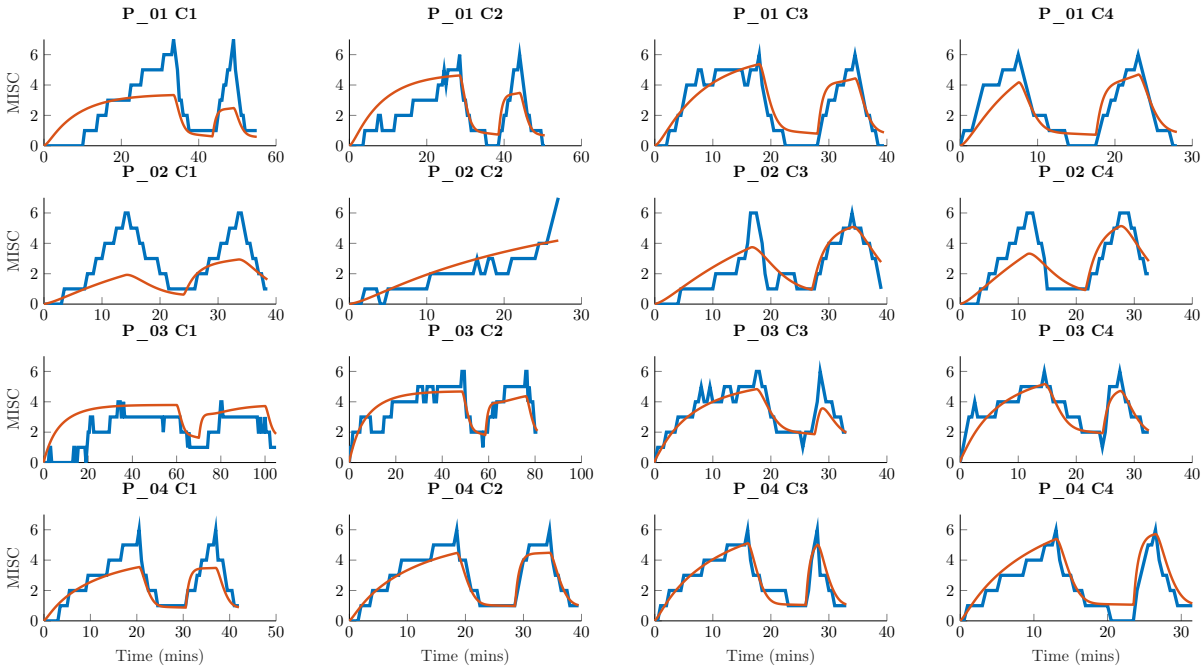
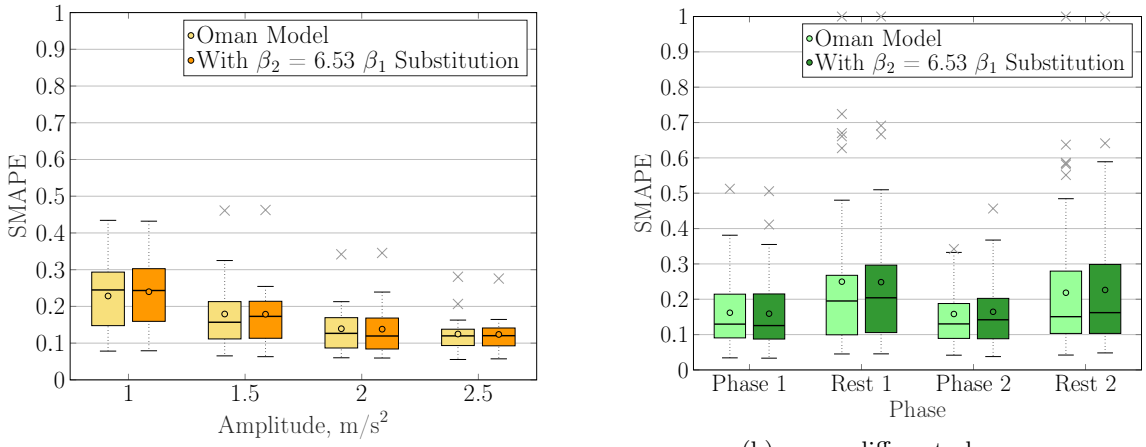


Figure 13: Predicted(in red) and actual(in blue) MISC scores for Participants 1 to 4 by the Oman model

Relationship between Time Constants

On closely observing the fitted model parameters mentioned in Appendix D, it is observed that there is a strong correlation ($\rho = 0.68$, $p < 0.01$) between the time constants of the slow and fast paths, with $\beta_2 = 6.5325\beta_1$. Therefore, in the subsequent models, this relation will be used to simplify the model further. This relation was also seen in an earlier experiment carried out by Irmak et al. (2020), where a gain of 7 was found, which is very close to the gain of 6.53 found in this study.

Figure 14 and Figure 15 compares the model, with the new substitution ($\beta_2 = 6.5325\beta_1$), with the Oman model. This substitution greatly simplifies the model, with no significant difference in fits when the SMAPE are compared. AICc values also favours the model with substitution (AICc = 11.84) over the one without the use of substitution (AICc = 16.19).



(a) across different amplitudes

(b) across different phases

Figure 14: Error box plots for Oman Model and with the substitution of the relation between time constants

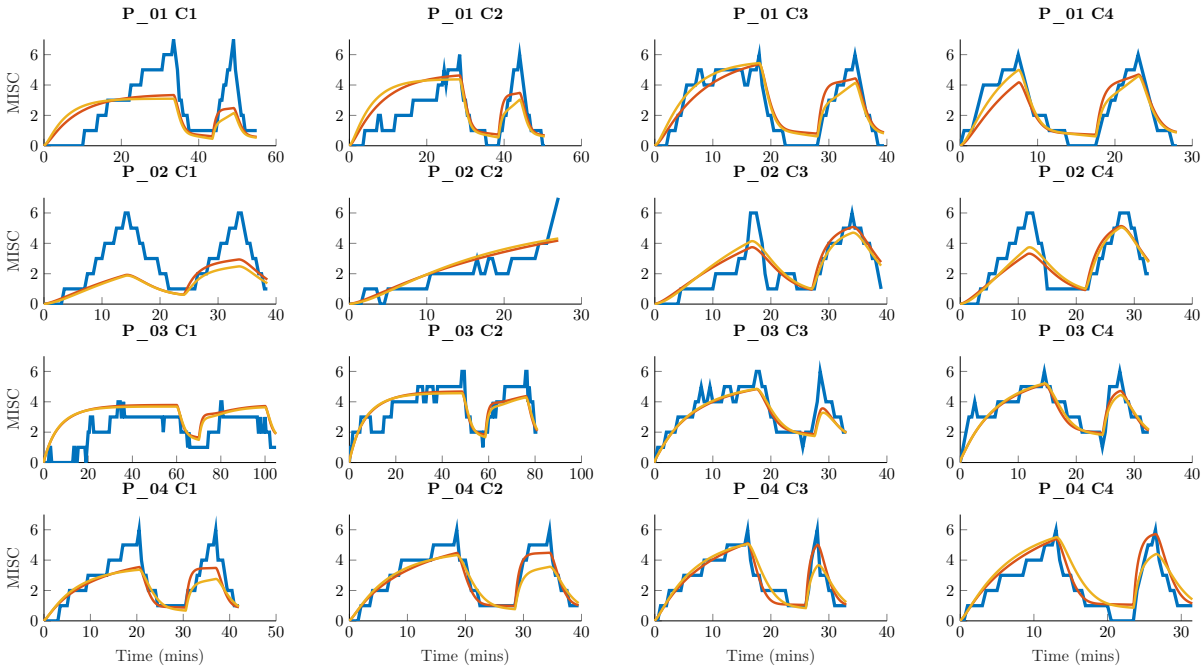


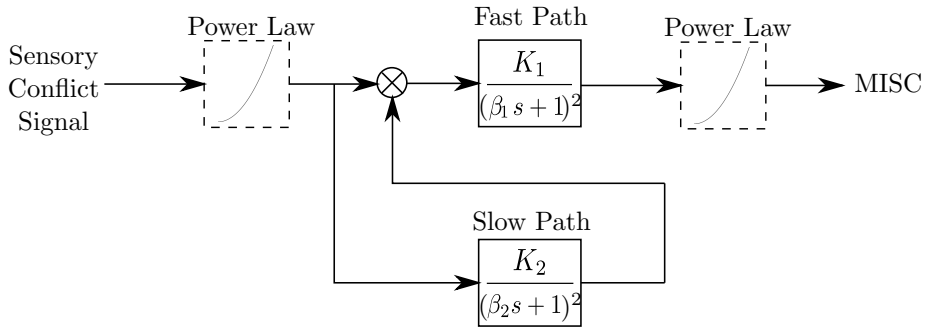
Figure 15: Predicted and actual(in blue) MISC scores for Participants 1 to 4, with(in red) and without(in yellow) the substitution of the relation between time constants

3.3.2 Modified Oman Model

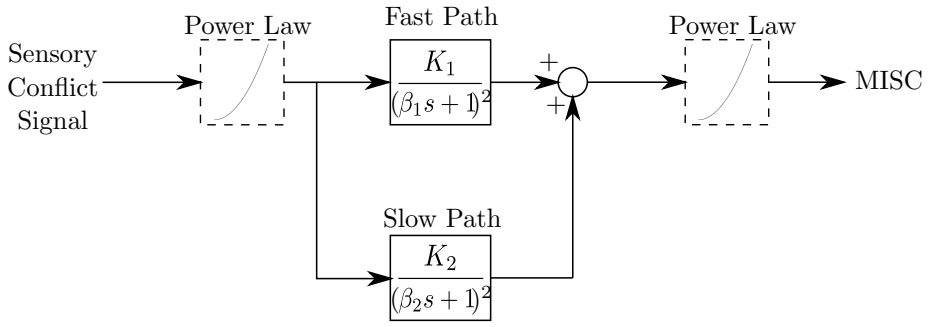
The previous section showed that the Oman Model cannot entirely capture the dynamics of the varying magnitude of the sensory conflict. Hence, a modification is needed to improve the fits. Two modified Oman models are proposed, see Figure 16.

1. **Oman Model Without Additions :** The time course of MS showed a fast recovery during the rest phase for most participants. This suggested that the fast path regulates the rate of change of MS when no stimulus is provided. Hence, the fast path solely must behave as a regulator to scale the slow path output and thus govern the MS development and recovery rate. Thus, in this variation for Oman model, the addition of the slow and fast paths is omitted.

2. **Oman Model Without Multiplication :** The next proposed model is inspired by the model presented by Oman (1982). Here, the output of the slow path is not multiplied by the fast path. Instead, a direct addition of the two paths is taken as the final output.



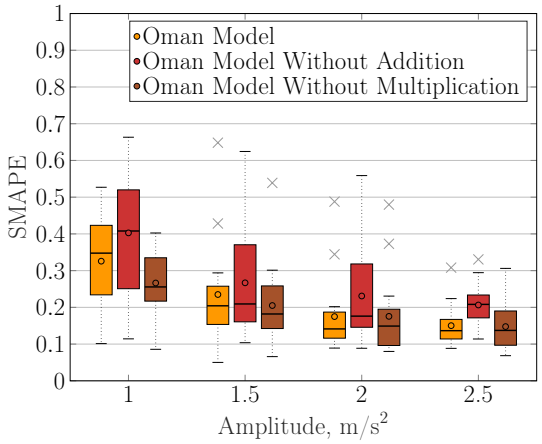
(a) Without Addition



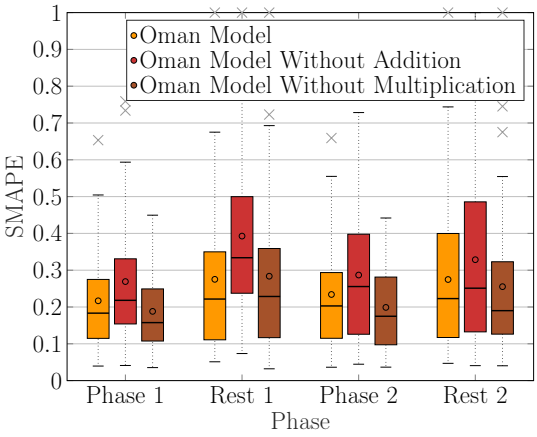
(b) Without Multiplication

Figure 16: Proposed Models

The earlier mentioned constrained minimization problem was used for fitting these models. To save computation time and power, the power law was, see Figure 16, initially left out from all three models. Figure 17 shows the comparison of the two models with the Oman model. A comparison during the different phases of the experiment is also shown in Figure 17. It can be concluded from the plots that the model without multiplication outperforms the other models in terms of both mean and variance of SMAPE. However, the AICc values for the three models are 8.1, 5.1 and 8.3, which suggests that the model without addition is preferred. This is due to the lower number of parameters present in this model, that the metric, AICc, prefers. Figure 18 shows the fitted models on four of the participants. Along with the fitted line, SMAPE values for these fits are also given to better visualize the metric, SMAPE.



(a) across different amplitudes



(b) across different phases

Figure 17: SMAPE for different models for simulating MS

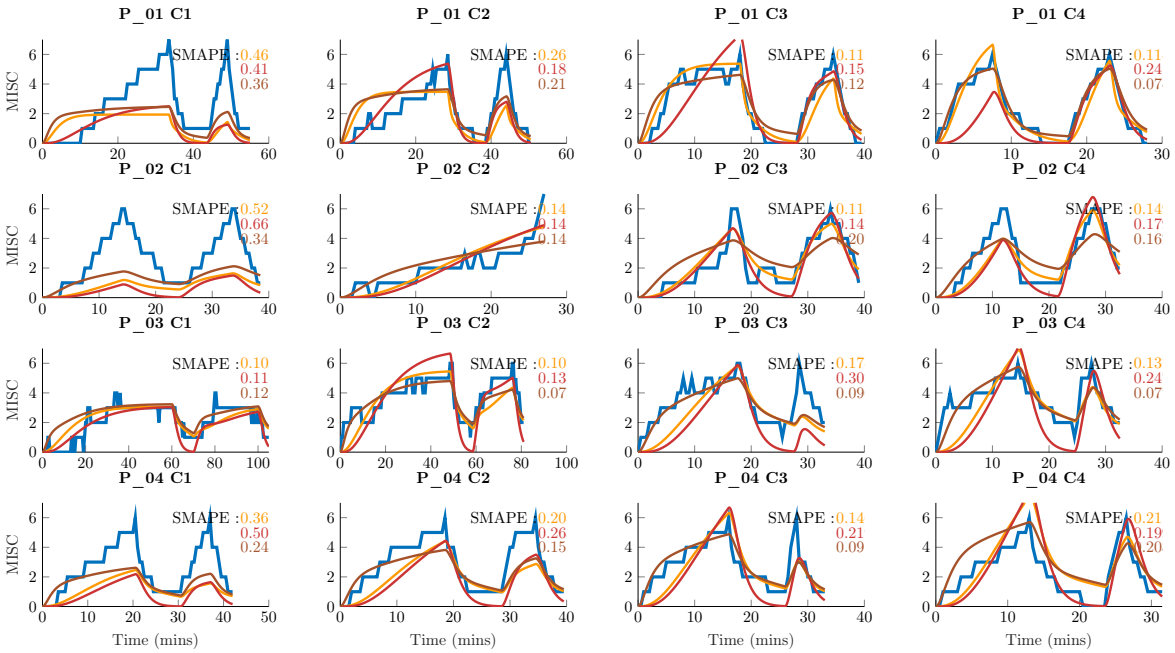


Figure 18: Predicted and actual(in blue) MISC scores for Participants 1 to 4, for the Oman Model(in yellow), without the Addition(in red) and without the Multiplication(in brown)

To further improve upon this, modification to the inputs and the outputs was also tested. The power law which is proposed in the original Oman model (Oman, 1990) is used in all three models. Furthermore, the power law term is also tested by placing it on the input instead of the output. The motive to do this is that the power law should capture the scaling the input conflict vector with amplitude, instead of it being at the output, where it will scale to match the rating scale used to measure MS.

The Figure 22 shows the error plots of all the models for different amplitudes. The circles in the plot show the mean, and the horizontal line between the rectangles shows the median of the errors. It can be said unanimously that the models with the power term are superior to those without it.

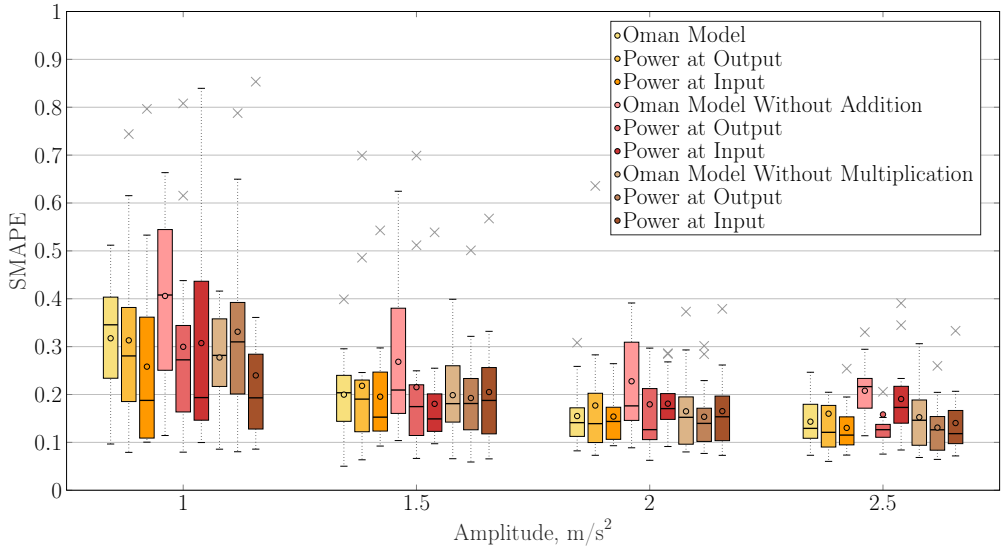


Figure 19: SMAPE for different models for simulating MS across differnet amplitudes

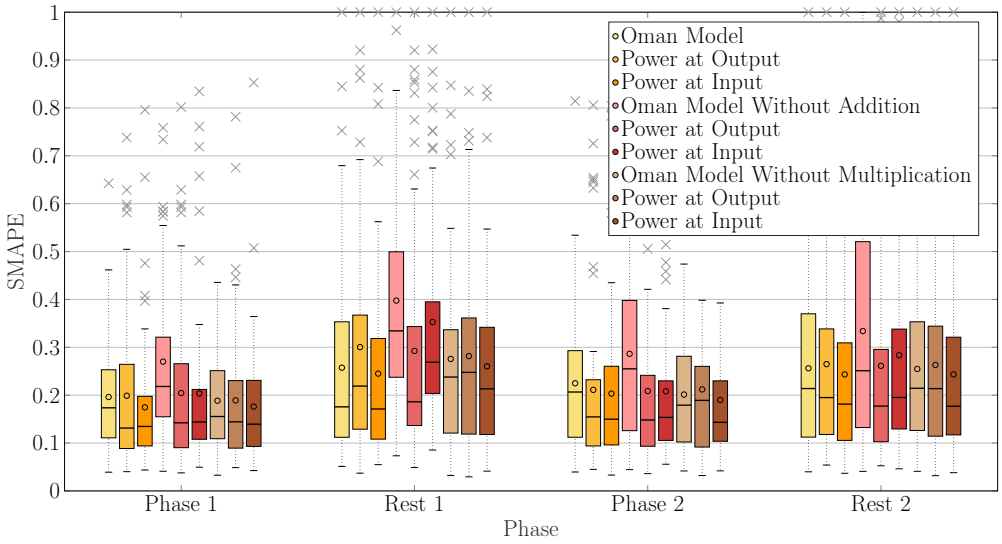


Figure 20: SMAPE for different models for simulating MS across different phases of the experiment

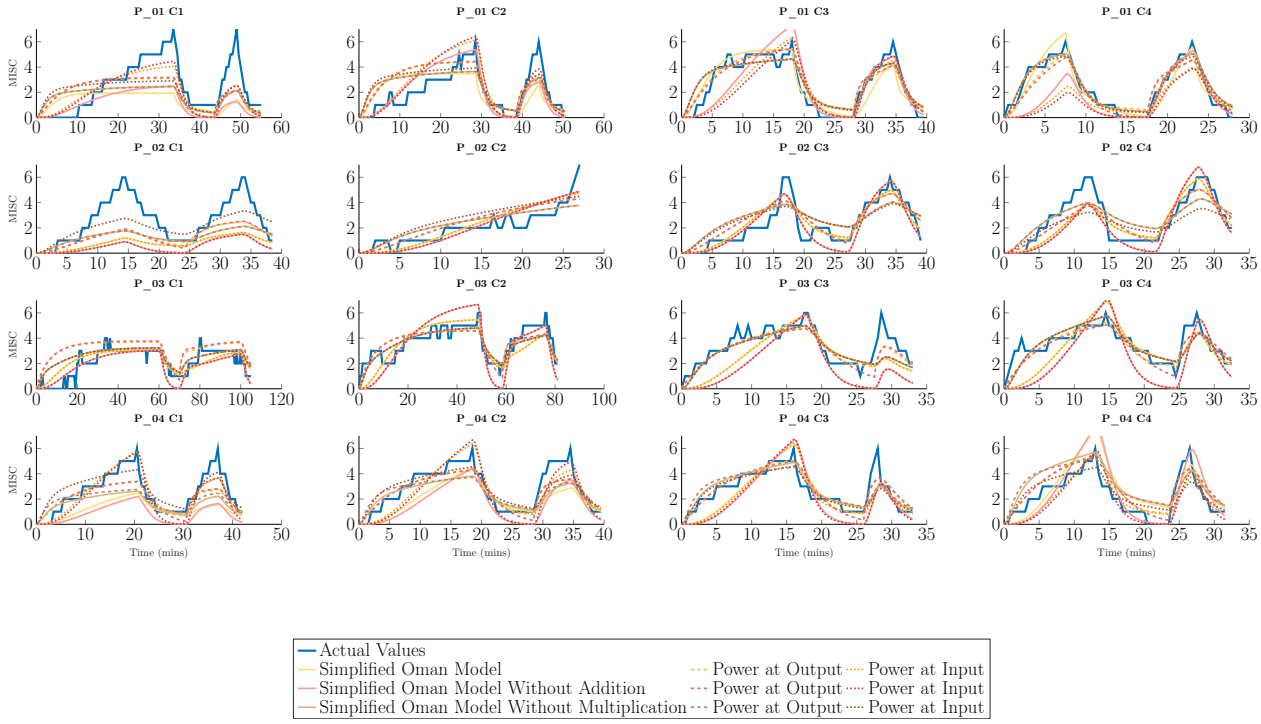


Figure 21: Predicted and actual MISC scores for Participants 1 to 4, for all the nine variants of the model

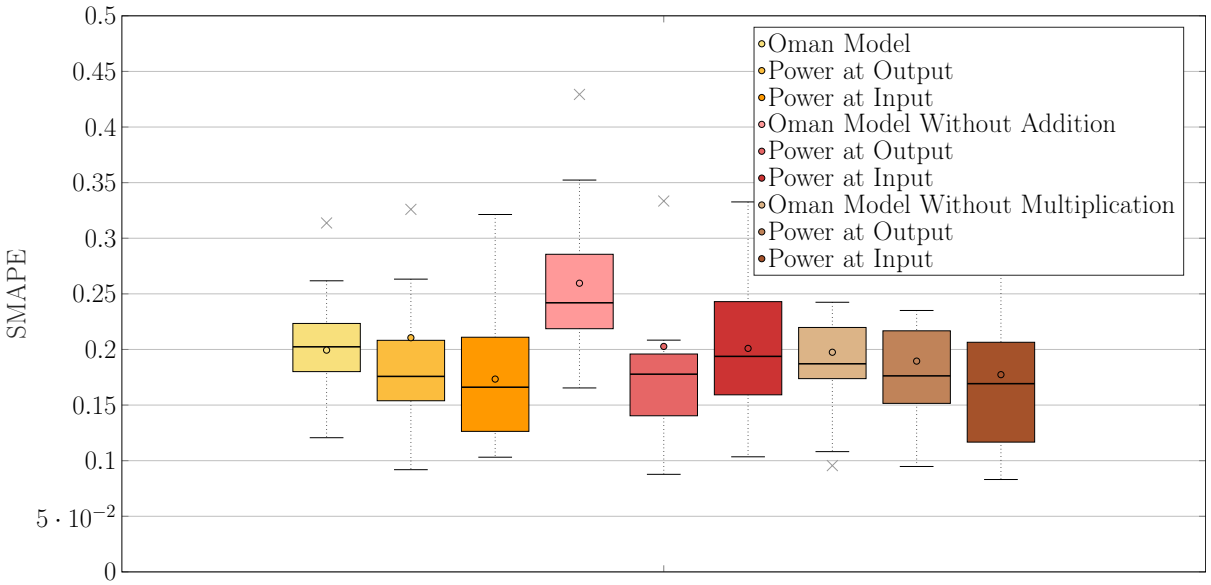


Figure 22: Combined SMAPE for different models for simulating MS

Concerning the placement of the power term in the model, Figure 19, Figure 20 and Figure 22 shows that having the power term at the input reduces the SMAPE for all models. This is especially true at lower amplitudes, see Figure 19. It is observed that power law is preferred at the output only in the model without the addition, that too only at higher amplitudes of 2 and 2.5 ms^{-2}. This could be due to the inability of this model to capture the recovery phases, which becomes a larger part of the experiment at higher amplitudes due to the higher dropout rates, which means shorter exposure, and constant rest times.

Model	No. of Parameters	No. of Observations	AIC	AICc	SMAPE
Oman Model	3	16	6.11	8.11	0.1980
Power at Output	4	16	8.20	11.84	0.1759
Power at Input	**4**	**16**	**8.05**	**11.69**	**0.1677**
Oman Model Without Addition	**2**	**16**	**4.15**	**5.07**	**0.2426**
Power at Output	3	16	6.08	8.08	0.1757
Power at Input	3	16	6.20	8.20	0.1930
Oman Model Without Multiplication	3	16	6.29	8.29	0.1875
Power at Output	4	16	8.14	11.78	0.1773
Power at Input	4	16	8.25	11.88	0.1721

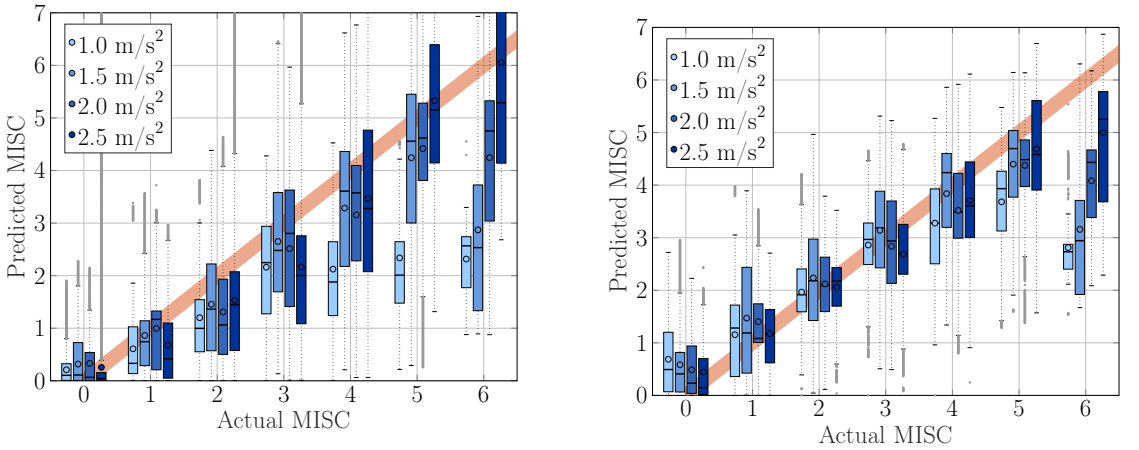
Table 1: Information Criteria for all models

From a broader perspective, when comparing the AICc values in Table 1, it is clear that the model without the addition is the best. This is due to the low number of parameters (two parameters, the gain and the time constant). Also, when the power law is added to this, the performance is still better when comparing with models having the same number of parameters.

However, when comparing raw performance scores, like the SMAPE from Table 1, it is seen that the Oman Model with the power at the output has the lowest errors.

Figure 23 shows the predicted MISC vs the Actual MISC scores for both these models along with a orange line where actual MISC is equal to the predicted MISC (1:1 line). This figure shows a boxplot for the predicted values of MISC vs the actual expected values of MISC. This will help in understanding how the model predicts each MISC score. It can be seen that the model without the addition underestimates MISC scores up-to 5, while the Oman model with the power term at the input has a better response.

Both models fail to predict MISC score of 6 properly. This is due to the small number of data points with MISC 6. Furthermore, it is observed in both the models that at the lowest amplitude of 1.0ms^{-2}, the model is unable to capture the MISC response at higher MISC levels of 3 to 5. This also appears in the amplitude of 1.5ms^{-2} at the MISC of 6. But this may be due to the lack of data at MISC 6, as mentioned earlier. This may also be due to the data point of MISC 6 followed by a MISC of low value, like MISC 3 or 4, which leads to the MISC 6 being ignored to fit the rest of the recovery phase better. This is also seen in Figure 23 where the accuracy of MISC 6 is lower than MISC 5. This underestimation at the lowest amplitude could be due to the change in slope of MISC rate from the amplitude of 1.5ms^{-2}, which may result in the model being generalized to only the top three amplitudes and not the lowest one.



(a) Oman model without Addition without the power law

(b) Oman model with the power law at the input

Figure 23: Predicted vs Actual MISC values from 1 m/s^2(light blue) to 2.5 m/s^2(dark blue) along with the 1:1 line (orange)

3.4 Effect on Skin Conductance

Motion sickness can induce changes in Electrodermal Activity, resulting in changes in Galvanic Skin Response (GSR), both in tonic and phasic GSR. These effects have been previously quantified by Irmak et al. (2020). It was found that out of the 17 participants, 14 participants showed a positive correlation of MISC to both tonic and phasic GSR. Additionally, these effects could also be a function of time in addition to the level of MS, as shown by Irmak et al. (2020). To verify this claim and to see if the relations found in Irmak et al. (2020) hold true in this study as well, a mixed effect model was used with the following structure,

$$\begin{aligned} S_{\text{tonic}} &= \alpha_0 + \alpha_1 \text{MISC} + \alpha_2 t \\ S_{\text{phasic}} &= \gamma_0 + \gamma_1 \text{MISC} + \gamma_2 t, \end{aligned}$$

where, S_{tonic} is the tonic component of GSR, S_{phasic} is the phasic component of GSR, t is the time, and $\alpha_0, \alpha_1, \alpha_2, \gamma_0, \gamma_1, \gamma_2$ are the intercepts and coefficients.

The phasic and tonic data obtained after Continuous Decomposition Analysis (CDA), which was sampled at 16 Hz, was used in the following analysis. The MISC scores, which were sampled at 0.033 Hz, were linearly interpolated to match the sampling frequency of the phasic and tonic data. Using these equations, a linear mixed-effects model was fitted, and the parameters were estimated with the help of the *fitlme* function in MATLAB. The estimated parameters were then tested for significance using a t-test.

For tonic GSR, the effect of MISC is significant, with a mean coefficient (α_1) of 0.00326 ($p < 0.01$, t-test). This means that when MISC rises from 0 to 7, the tonic GSR increases by 0.2283 μS, increasing 66.58%. There is also a significant effect ($p < 0.01$, t-test) of time with a mean coefficient of 0.000012.

For phasic GSR, the effect of MISC is also significant, with a mean coefficient (γ_1) of 0.0135 ($p < 0.01$, t-test). This means that when MISC rises from 0 to 7, the phasic GSR increases by 0.0945 μS, which is an increase of 95.06 %. There is also a significant effect ($p < 0.01$, t-test) of time with a mean coefficient of -0.0000055.

These percentage increase calculated are in line with the results obtained by Irmak et al. (2020). Thus, confirming our previous findings.

3.4.1 Prediction Model

GSR, measured in micro Siemens (μS), thus correlates well with the MISC scores provided by the participants. Out of the 64 sessions (4 sessions each for 17 participants with 4 sessions removed due to faults in device failure), skin conductance levels from 45 sessions showed a positive correlation. This suggests that GSR could be used as an predictor for MS level.

To use GSR as an indicator for MS level, a model is formulated with GSR as input and MISC as output. The relationship between GSR and MISC, in addition to being time dependent, is quite complex. GSR itself depends on many other factors that are difficult to control, like the body's temperature, water and food consumption, perspiration rate, psychological factors and many others. There are two parts of electrodermal activity, phasic (high frequency) and tonic (low frequency). Each of these has a different relationship with MS. Furthermore, the latency of response of the neurons and the sweat glands can differ between individuals. Due to these reasons, it was speculated that classical methods like the nonlinear mixed effects model, ARX, polynomial fitting would not produce precise results. This was confirmed when an attempt was made to fit the collected data using these methods and the prediction accuracy achieved was very low for the fits from all of these models. Hence, to capture this dynamic relation, neural network model is proposed. A neural network is best suited for this because of the large number of data points available due to sampling at 0.1 Hz.

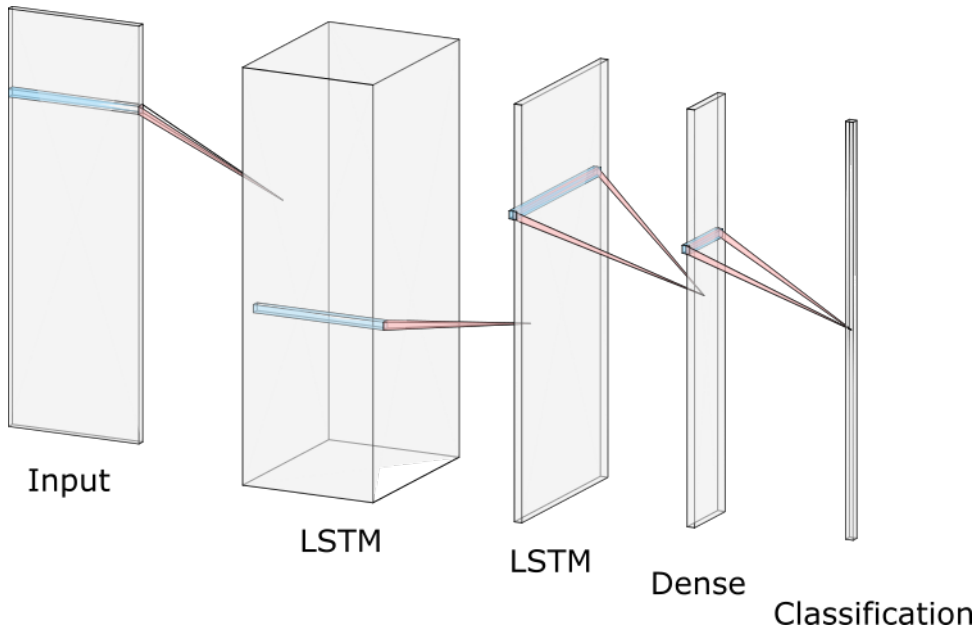


Figure 24: Neural Network Diagram

Two layers of bidirectional Long Short-Term Memory (LSTM) cells with 50 cells were used to capture the time dynamics. This was followed by a fully connected layer of size 8, which was then connected to the classification layer. The use of LSTM was inspired from Liao et al. (2020) and Wang et al. (2019) where multiple LSTM layers were used followed by fully connected layers. The main advantage of using LSTM over other neural networks like multilayer perceptrons (MLP) and convolutional neural networks (CNN) is that LSTM cells have an internal memory while others process inputs independent of the previous

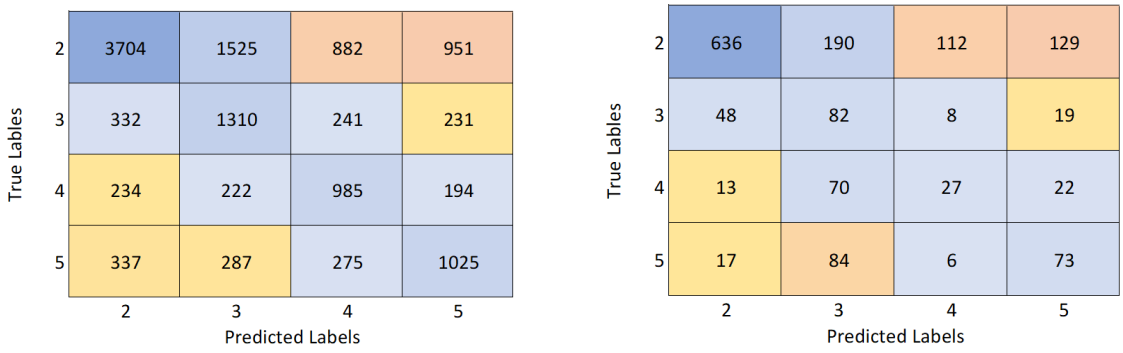
states. Thus, a model with LSTM cells can better process long term dependencies in sequential data. These reasons motivate the choice of LSTM in the model. The schematic of the neural network is shown in Figure 24.

The GSR time series data was split into sections of 13 data points, representing previous 2 minutes of GSR, sampled every 10 seconds. A GSR time series data was created at each MISC data point. Hence, the input size is 13, and the output is MISC value with size 1. Additionally, the MISC scores were bounded between the values of 2 and 5. MISC of 6 was removed as there were very few data point with a MISC value of 6. For most participants only one or two datapoints had a MISC of 6 as the motion is stopped as soon as the participant reached the MISC of 6. This introduced unwanted discrepancy during the training process. Thus, the data from MISC 6 was merged into the class of MISC 5. Besides this, MISC scores below 2 were truncated as there was not much difference in GSR values for MISC of 0, 1 and 2. Hence, these were merged into the class of MISC 2.

10 fold Cross-validation was used, where 10% of the data was excluded from the training and used as a validation set. In addition to this, data from Irmak et al. (2020) was also used an additional validation, also called hereafter holdout set as it will not be used in the cross validation. Class weights were used to balance the uneven distribution of the MISC scores. This neural network is trained for 600 epochs and the results of this are shown in Figure 25a, Figure 25b and Figure 25c.

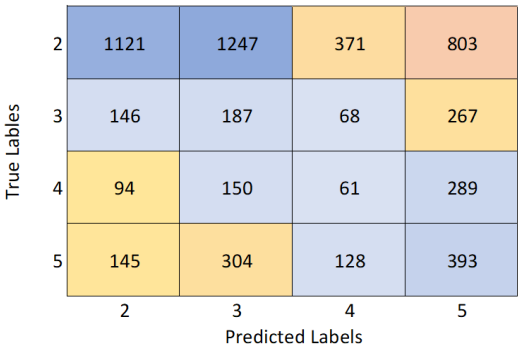
The scale used to measure the MS level of participants, the MIsery SCale, is a subjective rating scale based on symptom development. Thus, it is possible that participants may not give the similar scores for a same sickness level. They may overestimate or underestimate their sickness level. This may become more prominent when the participants reach higher levels of MS (MISC 3 to 5), thereby increasing the probability of the participants making an error in their judgement. Hence, a maximum allowed margin of error ± 1 MISC is set. This way, even if the participant made a mistake of ± 1 MISC, it will not affect the model's training.

Figure 25a, Figure 25b and Figure 25c shows the confusion matrix for the training, validation and holdout set, respectively. Due to the ± 1 MISC allowed margin of error, the values adjacent to the true value have also been considered correct. The correct predictions are shown in shades of blue, while wrong predictions are shown in shades of yellow.



(a) Training Set
(Accuracy:77%, Weighted F1 score: 77.5%, MAE: 0.78)

(b) Validation Set
(Accuracy: 66%, Weighted F1 score: 76.5%, MAE: 0.8)



(c) Holdout Set from Irmak et al. (2020)
(Accuracy:62%, Weighted F1 score: 67%, MAE: 1.2)

Figure 25: Confusion matrix for prediction of MISC from GSR

Furthermore, Accuracy (ACC), Weighted F1 score and Mean Absolute Error (MAE) were calculated for each dataset as shown below,

$$\begin{aligned} \text{ACC} &= \frac{\text{TP}}{\text{P}+\text{N}} \\ \text{F1 Score} &= \frac{\sum_{i=2}^{5} \frac{2\text{TP}_\text{i}(\text{P}_\text{i}+\text{N}_\text{i})}{2\text{TP}_\text{i}+\text{FP}_\text{i}+\text{FN}_\text{i}}}{\sum_{i=2}^{5} \text{P}_\text{i}+\text{N}_\text{i}} \\ \text{MAE} &= \sum |\ \text{Predicted label} - \text{True label}\ | \end{aligned}$$

Where, P stands for the number of real positive cases in the data
N stands for the number of real negative cases in the data
TP stands for true positive
TN stands for true negative
FP stands for false positive
FN stands for false negative
i stands for each class, which varies from 2 to 5

The training set, as expected, had mean accuracy of around 77% (Weighted F1 score: 77.5%, MAE: 0.78), with a standard deviation of 1.8, followed by the validation set having an accuracy of 66% (Weighted F1 score: 76.5%, MAE: 0.8), with a standard deviation of 7.4, and then followed by the holdout set with an accuracy of 62% (Weighted F1 score: 67%, MAE: 1.2), with a standard deviation of 2.3. The low accuracy of the holdout set could be due to the difference in the equipment used to measure GSR and

the different sampling frequency (32Hz in this study compared to 40Hz in the the study by Irmak et al. (2020)).

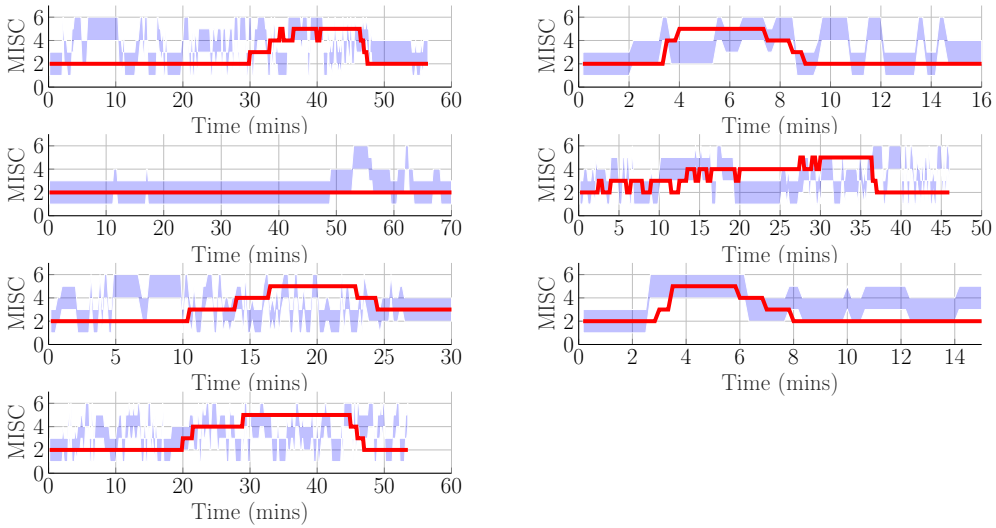


Figure 26: MISC prediction from GSR on Validation Set

Figure 26 shows a few examples of the prediction on the test set with the confidence interval (shaded in blue) and the true value (in red). As can be seen, the prediction is well within the confidence limits of ± 1 MISC most of the time. Prediction on all the data is shown in Appendix F. These predictions hence prove that GSR can indeed be used as an predictor for MS.

3.5 Subjective Discomfort Mapping

After the experiment, participants were asked to give a magnitude estimate of the MISC scores to map the MISC scores to the subjective discomfort. Furthermore, this magnitude estimate was then used to rate a few verbal quantifiers, namely, "Excellent", "Very Good", "Good", "So-so", "Bad", "Very Bad", and "Terrible". Figure 27 shows the magnitude estimates compared to the respective MISC scores for all 4 motion conditions. Additionally, The verbal quantifiers are also shown marked through a horizontal black line followed by the respective verbal feeling. Finally, a fitted power model is shown in red, which shows a power law relationship between the subjective discomfort and the MISC scores with mean exponent of 1.38 withe a standard deviation of 0.5.

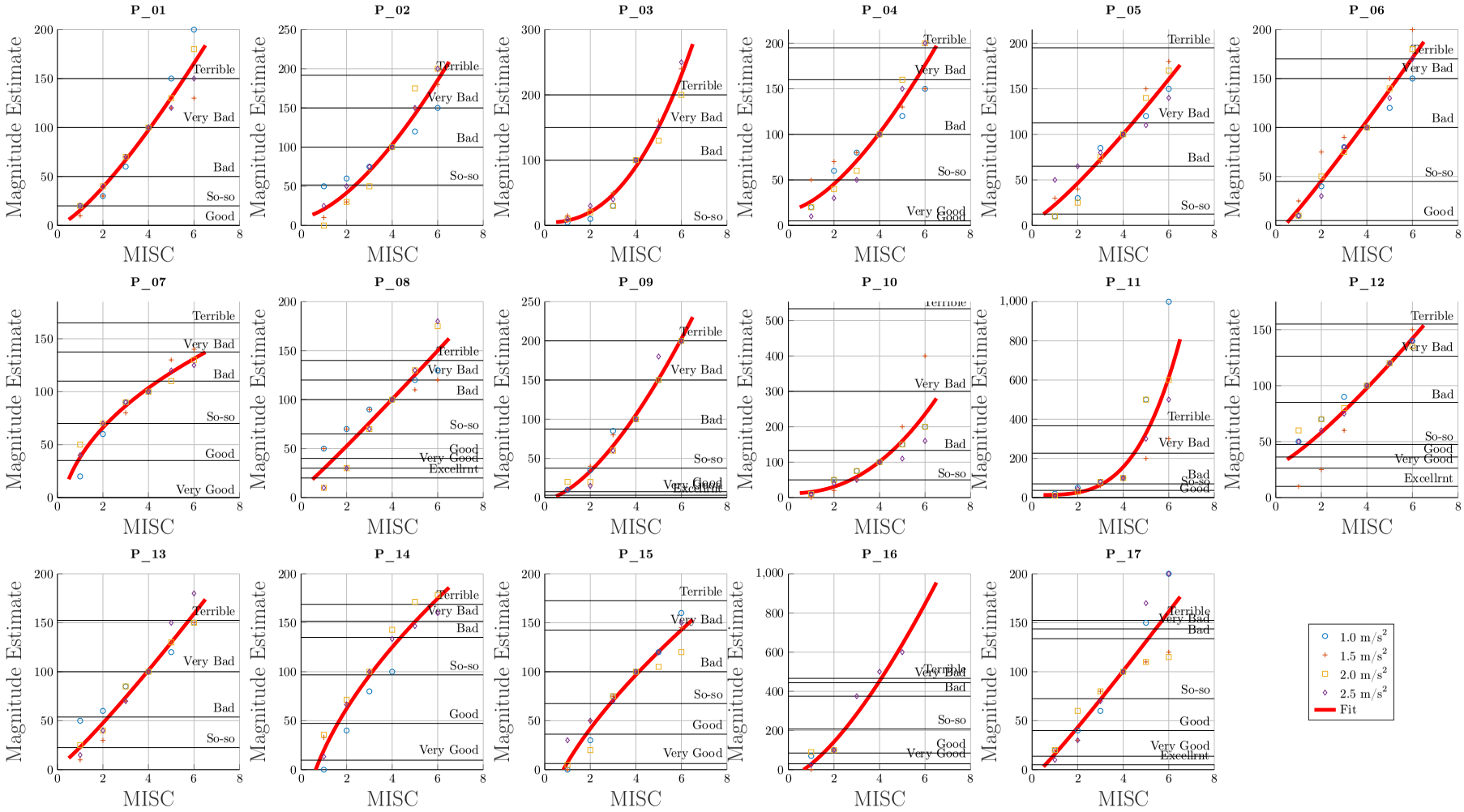


Figure 27: Magnitude Estimate vs MISC with Verbal Quantifiers

4 Discussion

This study investigated the amplitude dynamics of motion sickness development on seated participants due to fore-aft accelerations. This helped in understanding the amplitude dependency of an individual. Participants underwent a translational sinusoidal motion of different amplitudes in the fore-aft direction. The motion sickness level, through MISC, and skin response, through GSR, was measured at each of the four amplitudes. Using this collected data, a correlation between the GSR and MISC was found. Furthermore, using this correlation, an effort was made to create a model to predict MISC levels using GSR. Additionally, modifications were tried on the Oman model to obtain a generalized model for simulating and predicting MS development across all four amplitudes together. Finally, the magnitude estimates for subjective discomfort for the MISC are studied in conjunction with verbal quantifiers. In the following, these will be discussed along with findings in relation to the hypotheses and the methodological issues in the study.

4.1 Group sickness sensitivity

A monotonous increase in sickness sensitivity with amplitude was found. This is concluded from the finding that the time taken for onset of sickness decreased with increasing amplitude. This finding is in agreement with Hypothesis 1.1. Moreover, the time to reach a particular level of MS also increases with the respective MS level. That is, the time to reach MISC 6 is found to be always greater than the time to reach 5 for all participants.

Additionally, the rate of MS development increased with increasing amplitude. This was proved by calculating the MISC rates, which showed a clear increase with increasing amplitude. This aligns with the Hypothesis 1.2. However, the increase is not found to be linear. The rate of change of MISC rate across amplitudes increases, from 0.05 to 0.15, from when the amplitude changes from 1 to 1.5 m/s^2 to when the amplitude changes from 1.5 to 2 m/s^2 as discussed in subsection 3.1. Thus, suggesting a sigmoid or power law relation, which could have been more prominent if more amplitudes in between the ones used in the study had been also studied. This suggested relation could be used either at the input applied to the conflict or directly at the output of the Oman model. Another observation on the MISC rate is that the absolute values of MISC rates are higher during the second motion exposure, proving that individuals experience hypersensitivity after being made motion sick.

Likewise, Hypothesis 1.1 and 1.2 also hold true during the second exposure. During this phase, the participants experience hypersensitivity, where the MS sensitivity is heightened, as also noticed by Oman (1990). The only difference between the two exposures was found to be the reduction in time taken to reach a particular level of MS during the hypersensitive phase, which is again due to the increased MS sensitivity.

4.2 Modelling MS Development

A model to capture MS development with as few parameters as possible and without the need for parameter adjustments due to stimulus amplitude. Hence, the Oman model was considered, without the threshold. To reduce the model, a relation between the two time constants is found and used in further model fitting. Irmak et al. (2020) had already found such a relation. By doing this, the presence of this relationship is verified. Using this relation, a model can be obtained which can fit the individual MISC responses in the best possible way while still having a low number of parameters. Along with this, a few different variations of the Oman model were also tried, one by removing the addition of slow and fast paths and the other by removing the multiplication of the slow path output to the input fed to the fast path. On each of these models, the power law was either not present or at the input or at the output. This way nine different models were obtained.

It is observed that the model without the addition was not able to capture the recovery during rest phases of the experiment accurately (Mean SMAPE of around 0.3). This is being due to the absence of any output when the motion stimulus stops and the output of the fast path goes to zero, thereby reducing the final output to zero as well.

Out of the nine models, two of them were short-listed based on the two different metrics. Using the SMAPE, it is found that the Oman model with the power term at the input is the best. This model has the closest fits to the responses of the individuals. However, SMAPE does not consider the number of parameters used in the model. AICc was computed to take into account the effect of the number of parameters in the model. Using AICc it is found that the Oman model without Addition term is preferred. Oman model without Addition model has only 2 parameters, as compared to 4 for the Oman model with Power at Input. Nonetheless, both model are good depending upon the manner in which they will be used. If the requirement is to be as accurate as possible, definitely a model with more parameters will be preferred. However, when a model which needs to fit quickly with a low computation power requirement, such as in real time adaptive control, a model with the least number of parameters providing a sufficiently good accuracy is preferred. With this, a variation of Oman model has been developed which can generalize well in its prediction of MISC for all four amplitudes, and hence leading to the successful completion of Objective 1.

Finally, it is to be noted that none of the models are able to predict the MISC of 6 accurately, see Figure 23. This is attributed to the lack of data points, with a MISC score of 6. Additionally, the data point of MISC 6 is immediately followed by a sudden decrease in MISC scores, due to the termination of motion signal, for most participants. Thus, the models sacrifices capturing the data point at MISC of 6 in preference to better model the recovery phase. However, if the motion had been continued, the model may not have been able to forecast the MISC above 6. The reason for this is that most of the outputs of fits are converging to a MISC of 5 or 6. This could be solved by modifying the model to include a nonlinear term which can improve the forecast ability. More research needs to be done to know the actual nature of this non-linearity.

Lastly, the model developed to capture the individual dynamics of MS development is only applicable for fore-aft motion. This possibility of this model holding true in other directions is not known. Further, the effect of multi-directional motion is also not quantified. A host of mathematical operations may have to be performed on this input signal, or a sensory conflict generation model needs to be used to accurately calculate the sensory conflict.

A large variation in individual parameters (gains and time constants) for the fitted models was observed. This was because of the variation of susceptibility to MS between participants, which quantified by mean, of 16.19, and standard deviation, of 10.05, of the MSSQ-Short.

4.3 MISC prediction

GSR showed a positive correlation(average correlation coefficient, $\mu = 0.37$, with standard deviation, $\sigma = 0.22$) with the MS levels in all but three participants. However, as the sensitivity and the baseline value varies across the participants, a linear mixed effects model was used to quantify the effect of MS on GSR.

Motion sickness induces a varied response from the skin, which includes sweating. This could be the main reason for the increase in GSR observed with MISC. This is seen in both the components of GSR, tonic and phasic. However, the effect on phasic GSR (95% increase from MISC 0 to 7) was found to be higher than the effect on tonic GSR (66.5% increase from MISC 0 to 7). The increase in phasic activity could also be associated with an increase in sympathetic activity. This definite increase verified a similar claim made by Irmak et al. (2020). These observed effects lead to the acceptance of Hypothesis 2.

The large effect size of MISC on GSR suggested that the GSR could perhaps be used as a real time indicator of MS level. To test this, an attempt was made to build a model to predict MISC scores using GSR. However, the relation between the MISC and GSR is highly non-linear. There are two components, the low frequency (tonic) and high frequency (phasic), which when combined give rise to the final GSR. Both of these are independently related to MS level. Moreover, it is observed that the frequency of

change in GSR also changes with MS level. To capture all these effects, a neural network model was built with the inputs being GSR in micro Siemens and the output being the MISC score in the range of 2 to 5.

An error margin of ± 1 MISC was allowed to be made by the model, as discussed in **??**. The generated model had an accuracy of 77% (Weighted F1 score: 77.5%, MAE: 0.78) on the training set, 77% (Weighted F1 score: 76.5%, MAE: 0.8) on the validation set. This shows that GSR may indeed be used as a reliable indicator of MS level. To further prove this, the same model was used on the data from Irmak et al. (2020), where the participants were made sick in a vehicle driven around a slalom. The results showed an accuracy of 66% (Weighted F1 score: 67%, MAE: 1.2). This increase in error is alleged to be due to the difference in the sampling frequency, as well as the inherent difference in the device used to measure GSR in both studies. Another potential reason for this could be the significant time effects, which were dominant in Irmak et al. (2020). In addition to this, a car was used to define the motion signal in the study, which would not have been as accurate as a simulator, used in this study. Hence, the prediction model was developed and tested not only on a dataset generated within the experiment, but was able to perform well on a dataset generated from an entirely different experiment. The high accuracy of predictions with GSR as a predictor in the validation set and the repeatability achieved in the holdout set fulfills Objective 2.

Further, improvements can be made to this model by increasing the complexity and collecting more data. In this study, the GSR was sampled at 0.1 Hz, due to which the phasic component is not used for the prediction, which was found to be a better indicator by the linear mixed-effects model. Hence, in the further improvements, a higher sampling frequency could be used to make use of the phasic component of GSR.

4.4 Subjective Discomfort

From the magnitude estimates for subjective discomfort given to each MISC score by the participants, it is found that a power law relation exists between the MISC and the subjective discomfort. This is due to the severity of discomfort and thereby the symptoms increase rapidly as the absolute MS increases. This relation is useful because it provides a simple, direct means of determining the subjective magnitude of discomfort and with that the MS. It is also important to note that the fitted power models for all but three participants had an exponent greater than one, with a mean of 1.28 and a standard deviation of 0.5. The deviations are presumed to be due to the variations of perception of discomfort between participants. The positive exponent for participants results in accepting Hypothesis 3. Additionally, Objective 3 is accomplished due to the finding of the relation between MISC and subjective discomfort.

Another interesting conclusion from this study is that all but 4 participants did not feel "Terrible", but all participants felt "Very bad" at the end of the motion exposure at the MISC of 6. Furthermore, all participants reported a monotonous increase in worsening of the feeling with increasing MISC. This is in contradiction with what has been observed in Reuten et al. (2020) where a dip in subjective illness rating was observed at MISC 6. This may be because of how the participants interpret and use the MISC to report MS levels during the experiment. When comparing different symptoms, the participant may experience the symptoms of various but severe symptoms (MISC 5) as worse than a little nausea (MISC 6). However, during the experiment, while being perturbed by a motion signal, the discomfort may be higher when severe symptoms are followed by onset of nausea. Now, the combined effect of severe symptoms and nausea do seem to be worse than only the severe symptoms. This could explain the discrepancy in the results between the studies.

The relation found between MISC and subjective discomfort is very useful. It could be used to substitute the MISC by the subjective discomfort in the models presented in this study. Using subjective discomfort as the output would most likely give a better fit as the scale is linear instead of being ordinal, as was the case with MISC. Additionally, the power term earlier had to be estimated taking into consideration the inherent non linearity in the MISC but with the use of subjective discomfort, it will only need to estimate the amplitude dependency, thus improving its precision.

4.5 Limitations

A limitation of the current experiment is that the range of amplitudes used was limited to a range between 1 and 2.5 ms^{-2}. Moreover, only one frequency (0.3 Hz) condition was tested. It is possible that a few participants would be more sensitive to other frequencies than the one used Irmak et al. (2021), and hence did not reach higher levels of sickness. This would have reduced the significance of the correlations found.

The choice of frequency and the amplitudes was a compromise between the peak sensitivity frequency from the literature and the limitations of the simulator used for the experiment. To reduce these side effects, in the future a wider range of amplitudes and frequencies or a multi-sinusoidal signal could also be used.

Statistical power could have been improved by carrying out repetitions. However, this consumes a lot of time for both the parties, experimenter and participant. This is because each session is required to be performed after a rest period of at least one week after another session. Additionally, it introduces additional dropouts in case the participant decides to not continue the experiment further.

This study was designed to study the amplitude dynamics of motions sickness. However, other factors such as the changes in amplitude sensitivity due to changes in frequency have not been studied here. In future works, a frequency dependent amplitude dependency of the MS development could be studied.

5 Conclusion

The study investigated the amplitude dynamics of motion sickness for fore-aft motion on 17 participants. A significant increase in average MISC gradient rate from 0.2 to 0.3 was observed when acceleration of the motion was increased from 1 to 2.5 m/s^2. This is due to the amplification of the sensory conflict due to the increase in acceleration. The magnitude of this amplification has a complex relationship with the acceleration of the motion.

Additionally, it is shown that the time taken to reach a particular MISC score decreases with increasing acceleration. This also holds true during the hypersensitivity phase (second motion exposure). Moreover, there is a monotonous increase in MISC at all amplitudes, suggesting that the development of symptoms always precedes the onset of nausea.

A wide variation in the dynamics of sickness development over time is seen among the participants. Hence, there is also a wide variation in the amplitude dependency of the amplification. To capture this dependency, an attempt to fit a single model that can fit all four motion conditions was made. The original Oman model with a few modifications were tried. It was deduced that moving the power term to the input will provide the best quality of fits. However, when considering the number of parameters required to make these fits, the model without addition is preferred.

GSR responses showed a significant effect due to MS level. 14 out of 17 participants showed a positive correlation to MISC scores. Thus, the use of GSR as an indicator of MS level was investigated. A neural network model was developed to make MISC score predictions. The developed model had an accuracy of 77%, proving that GSR can be used as an indicator of MS.

The magnitude estimates of subjective discomfort showed a power law relationship with MISC, which can help in translating various other scales to a different one using subjective discomfort as the base. This relation also proved that there exists a monotonous increase in the severity of symptoms mentioned in MISC. Thus, symptoms at higher MISC levels are always found to be worse than those at lower levels. However, the rate of increase in severity of these symptoms increase with the corresponding MISC levels. This means that the difference in discomfort between MISC 1 and 2 is much smaller than the difference in discomfort between MISC 2 and 3.

In all, the models developed could be used in an adaptive control architecture where, it will minimize, maximize MS or hold MS to a required level. Secondly, the prediction model developed, with further

improvements, could be used as a replacement of objective methods of MS measurement. Additionally, these can be very beneficial for researchers to carry out MS experiments. Instead of arbitrarily setting exposure times and magnitudes, they can use adaptive control algorithms to make the participants reach the required level of sickness. This way the experiments could be conducted in a very short amount of time and also ensures that the participant will surely get sick. Using these models, the experience in travelling in an automated vehicles could be greatly improved. By making use of these models, and running model predictive control algorithms, the route and path planning can be smoothed to avoid increase in MS. Moreover, these models can be fine tuned according to the susceptibility of the user.

References

Benedek, M. and Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190:80–91.

Berkouwer, W., Stroosma, O., van Paassen, R., Mulder, M., and Mulder, B. (2005). Measuring the Performance of the SIMONA Research Simulator's Motion System. In *AIAA Modeling and Simulation Technologies Conference and Exhibit*, Guidance, Navigation, and Control and Co-located Conferences. American Institute of Aeronautics and Astronautics.

Bertolini, G. and Straumann, D. (2016). Moving in a moving world: A review on vestibular motion sickness.

Bock, O. L. and Oman, C. M. (1982). Dynamics of subjective discomfort in motion sickness as measured with a magnitude estimation method. *Aviation, space, and environmental medicine*, 53(8):773–777.

Bos, J., Mackinnon, S., and Patterson, A. (2005). Motion sickness symptoms in a ship motion simulator: effects of inside, outside, and no view. *Aviation, space, and environmental medicine*, 76 12:1111–1118.

Bos, J. E. and Bles, W. (1998). Modelling motion sickness and subjective vertical mismatch detailed for vertical motions. *Brain Research Bulletin*, 47(5):537–542.

Cowings, P. S., Suter, S., Toscano, W. B., Kamiya, J., and Naifeh, K. (1986). General autonomic components of motion sickness. *Psychophysiology*, 23(5):542–551.

Diels, C. and Bos, J. E. (2016). Self-driving carsickness. *Applied Ergonomics*, 53:374–382.

Donohew, B. and Griffin, M. (2004). Motion sickness: Effect of the frequency of lateral oscillation. *Aviation, space, and environmental medicine*, 75:649–656.

Golding, J. F. (2006). Predicting individual differences in motion sickness susceptibility by questionnaire. *Personality and Individual Differences*, 41(2):237–248.

Graybiel, A. (1969). Structural elements in the concept of motion sickness. *Aerospace medicine*, 40 4:351–367.

Griffin, M. and Mills, K. (2002a). Effect of magnitude and direction of horizontal oscillation on motion sickness. *Aviation, space, and environmental medicine*, 73:640–646.

Griffin, M. J. and Mills, K. L. (2002b). Effect of frequency and direction of horizontal oscillation on motion sickness. *Aviation, space, and environmental medicine*, 73(6):537–543.

Hu, S., Grant, W. F., Stern, R. M., and Koch, K. L. (1991). Motion sickness severity and physiological correlates during repeated exposures to a rotating optokinetic drum. *Aviation, space, and environmental medicine*, 62(4):308–314.

Irmak, T., De Winkel, K. N., Pool, D. M., Bülthoff, H. H., and Happee, R. (2021). Individual motion perception parameters and motion sickness frequency sensitivity in fore-aft motion. *Experimental Brain Research*.

Irmak, T., Pool, D. M., and Happee, R. (2020). Objective and subjective responses to motion sickness: the group and the individual. *Experimental Brain Research*.

Koekebakker, S. H. (2001). Model Based Control of a Flight Simulator Motion System. In *Doctoral Thesis*, pages 1–264.

Kufver, B. and Förstberg, J. (1999). A net dose model for development of nausea.

Lawther, A. and Griffin, M. J. (1988). Motion sickness and motion characteristics of vessels at sea. *Ergonomics*, 31(10):1373–1394.

Liao, C.-Y., Tai, S.-K., Chen, R.-C., and Hendry, H. (2020). Using EEG and Deep Learning to Predict Motion Sickness Under Wearing a Virtual Reality Device. *IEEE Access*, 8:126784–126796.

Lie Hok Lien, K. (2019). Heart rate, heart rate variability, and GSR in relation to motion sickness caused by lateral vehicle motion.

Litman, T. (2020). Autonomous Vehicle Implementation Predictions: Implications for Transport Planning.

Matsangas, P., McCauley, M. E., and Becker, W. (2014). The effect of mild motion sickness and sopite syndrome on multitasking cognitive performance. *Human Factors*, 56(6):1124–1135.

McCauley, M., Royal, J., Wylie, C., O'Hanlon, J., and Mackie, R. (1976). Motion Sickness Incidence: Exploratory Studies of Habituation, Pitch and Roll, and the Refinement of a Mathematical Model. page 63.

Miller, E. F. and Graybiel, A. (1969). A STANDARDIZED LABORATORY MEANS OF DETERMINING SUSCEPTIBILITY TO CORIOLIS (MOTION) SICKNESS.

O'Hanlon, J. F. and McCauley, M. E. (1974). Motion sickness incidence as a function of the frequency and acceleration of vertical sinusoidal motion. *Aerospace medicine*, 45(4):366–369.

Oman, C. M. (1982). A Heuristic Mathematical Model for the Dynamics of Sensory Conflict and Motion Sickness Hearing in Classical Musicians. *Acta Oto-Laryngologica*, 94(sup392):4–44.

Oman, C. M. (1990). Motion sickness: a synthesis and evaluation of the sensory conflict theory. *Canadian Journal of Physiology and Pharmacology*, 68(2):294–303.

Parker, D. M. (1971). A Psychophysiological Test for Motionsickness Susceptibility. *The Journal of General Psychology*, 85(1):87–92.

Reason, J. T. (1978). Motion sickness adaptation: a neural mismatch model. *Journal of the Royal Society of Medicine*, 71(11):819–829.

Reason, J. T. and Brand, J. J. (1975). *Motion sickness*. Academic Press, Oxford, England.

Reuten, A. J. C., Bos, J., and Smeets, J. B. J. (2020). *Proceedings of the Driving Simulation Conference 2020 Europe VR*. Antibes, France.

Stevens, S. S. (1987). The direct estimation of sensory magnitudes–loudness. By S. S. Stevens, 1956. *The American journal of psychology*, 100(3-4):664–689.

Stroosma, O., van Paassen, M. M. R., and Mulder, M. (2003). Using the SIMONA Research Simulator for Human-machine Interaction Research. In *AIAA Modeling and Simulation Technologies Conference and Exhibit*, Guidance, Navigation, and Control and Co-located Conferences. American Institute of Aeronautics and Astronautics.

Wang, Y., Chardonnet, J.-R., and Merienne, F. (2019). VR Sickness Prediction for Navigation in Immersive Virtual Environments using a Deep Long Short Term Memory Model. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1874–1881. IEEE.

Warwick-Evans, L. A., Church, R. E., Hancock, C., Jochim, D., Morris, P. H., and Ward, F. (1987). Electrodermal activity as an index of motion sickness. *Aviation, space, and environmental medicine*, 58(5):417–423.

A Experiment Data

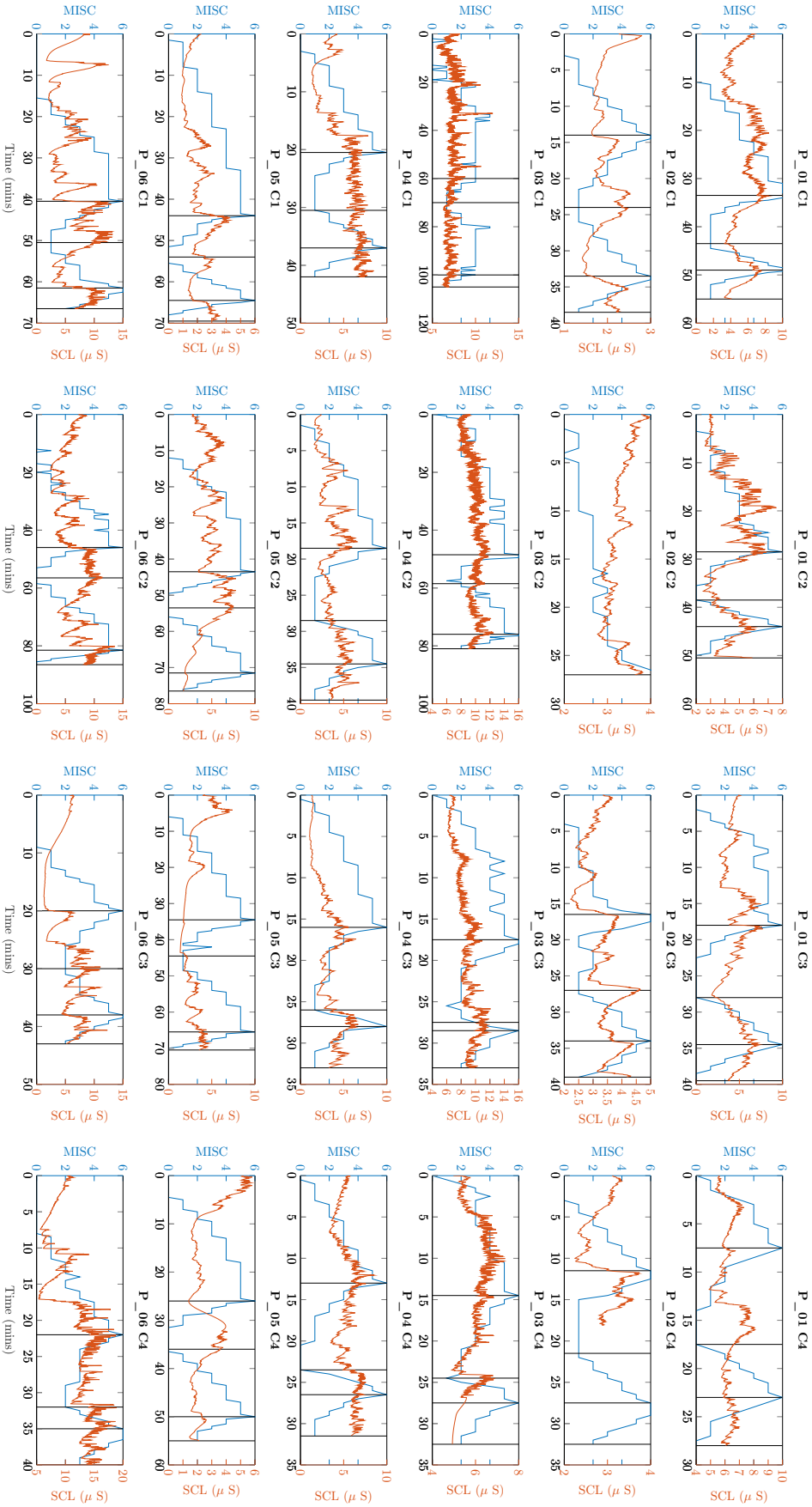


Figure 28: MISC and SCL Data for first 6 participants

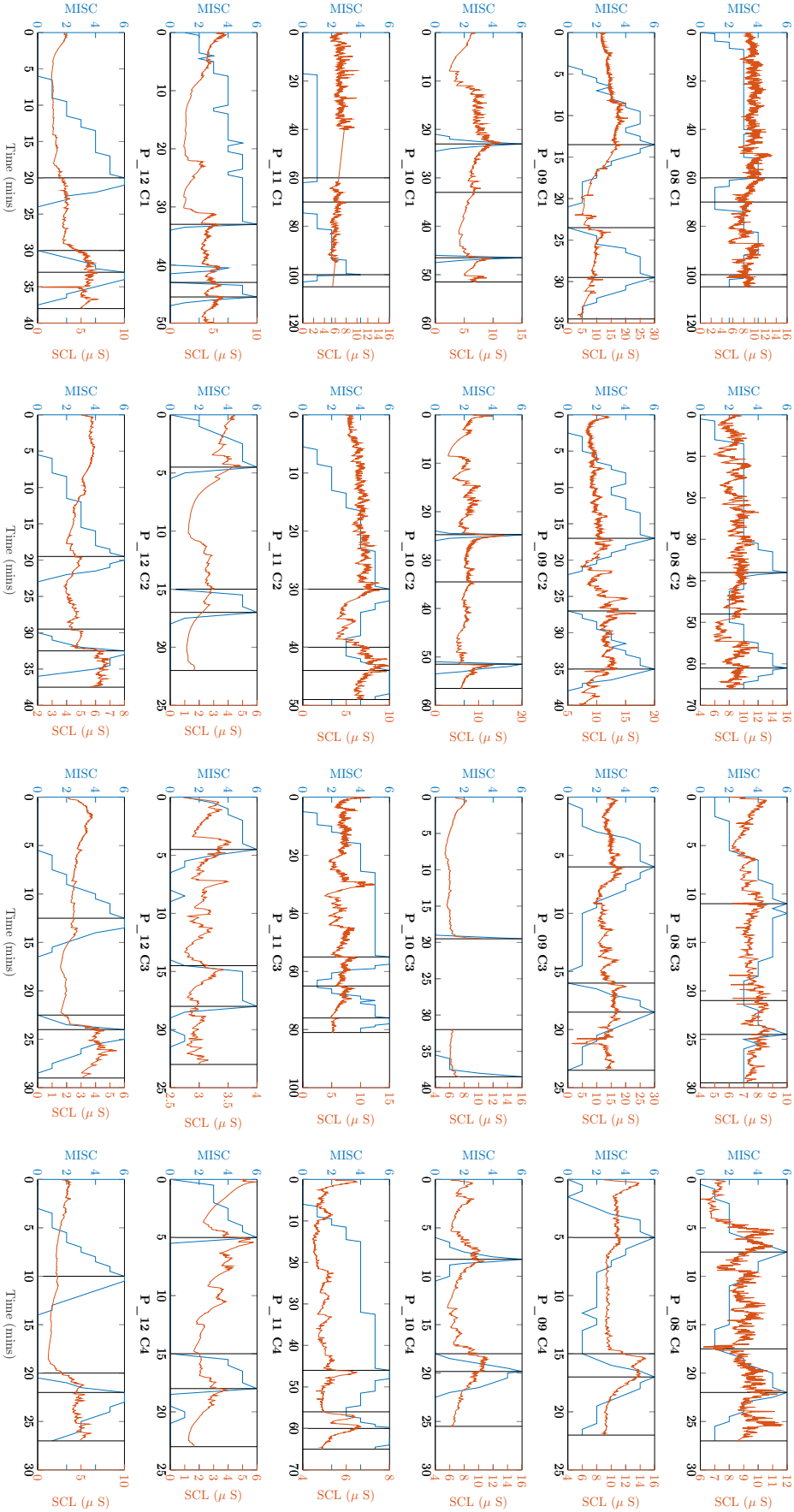


Figure 29: MISC and SCL Data for participants 7 to 12

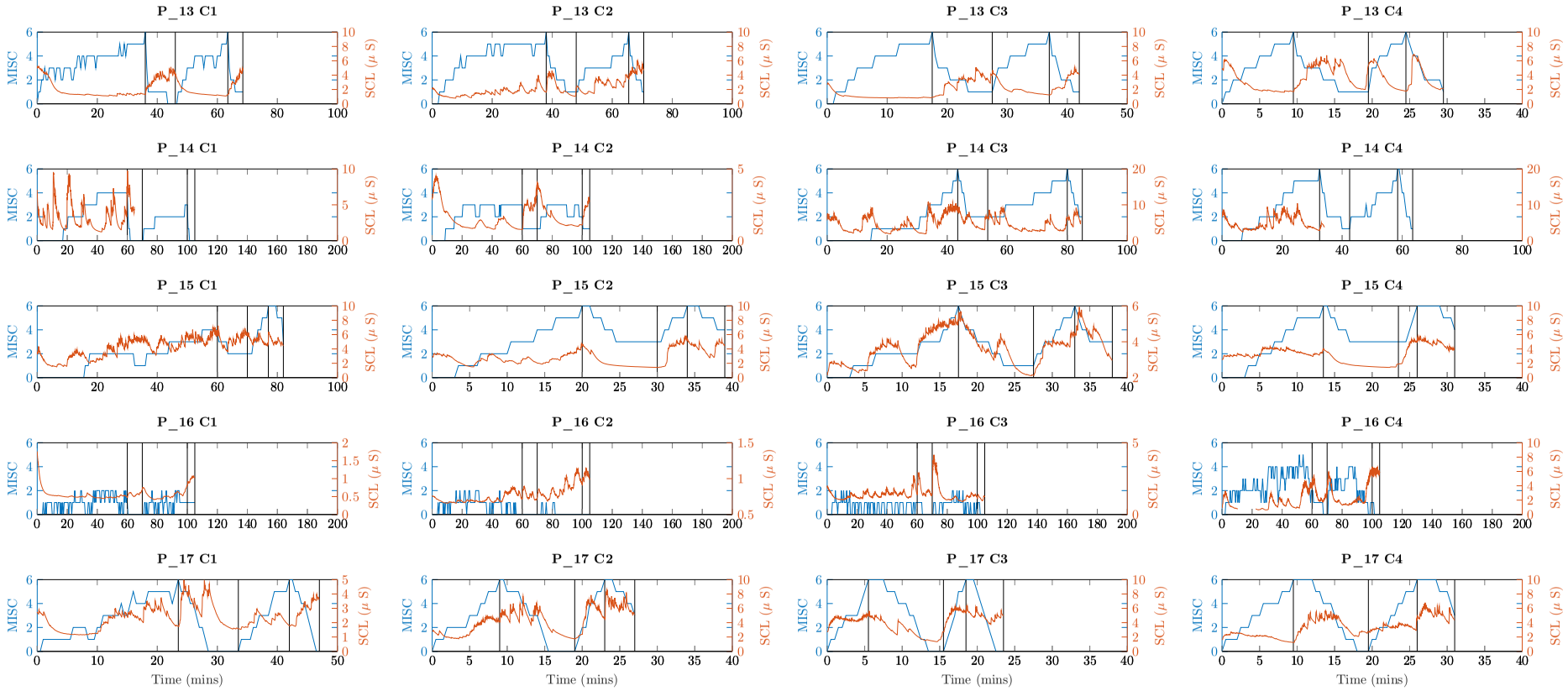


Figure 30: MISC and SCL Data for last 5 participants

B Model Fitting

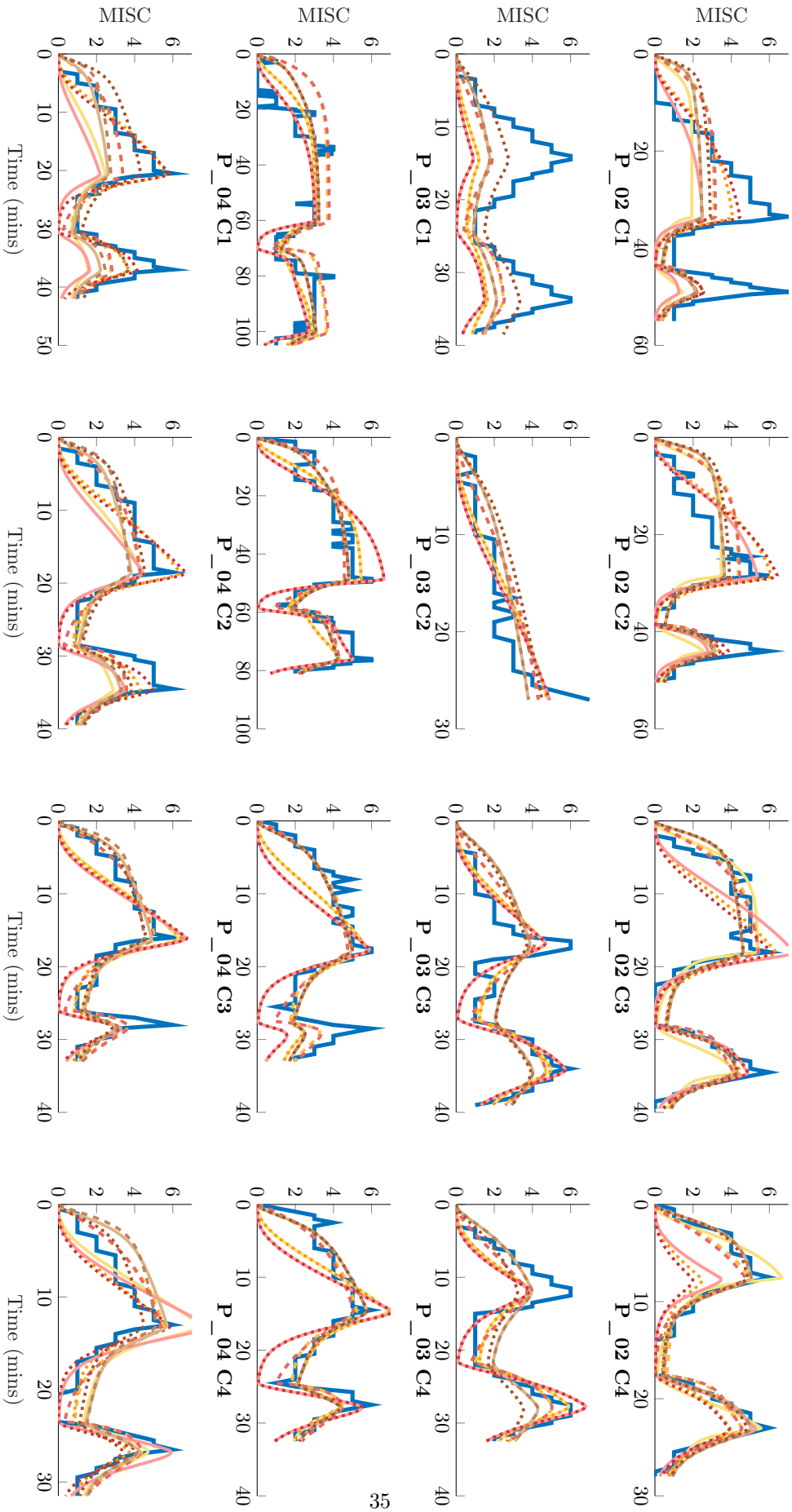


Figure 31: Model Fit 1/4

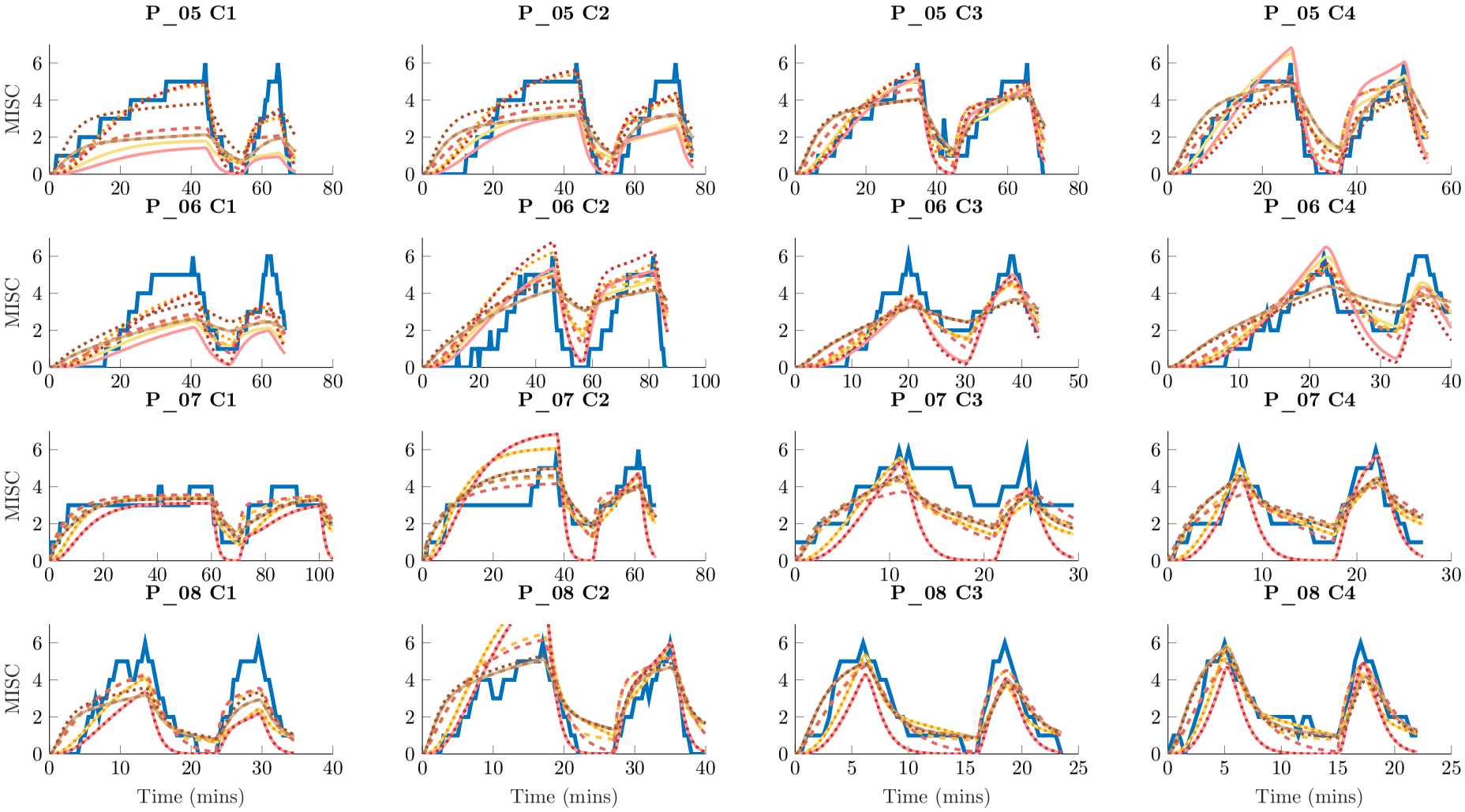


Figure 32: Model Fit 2/4

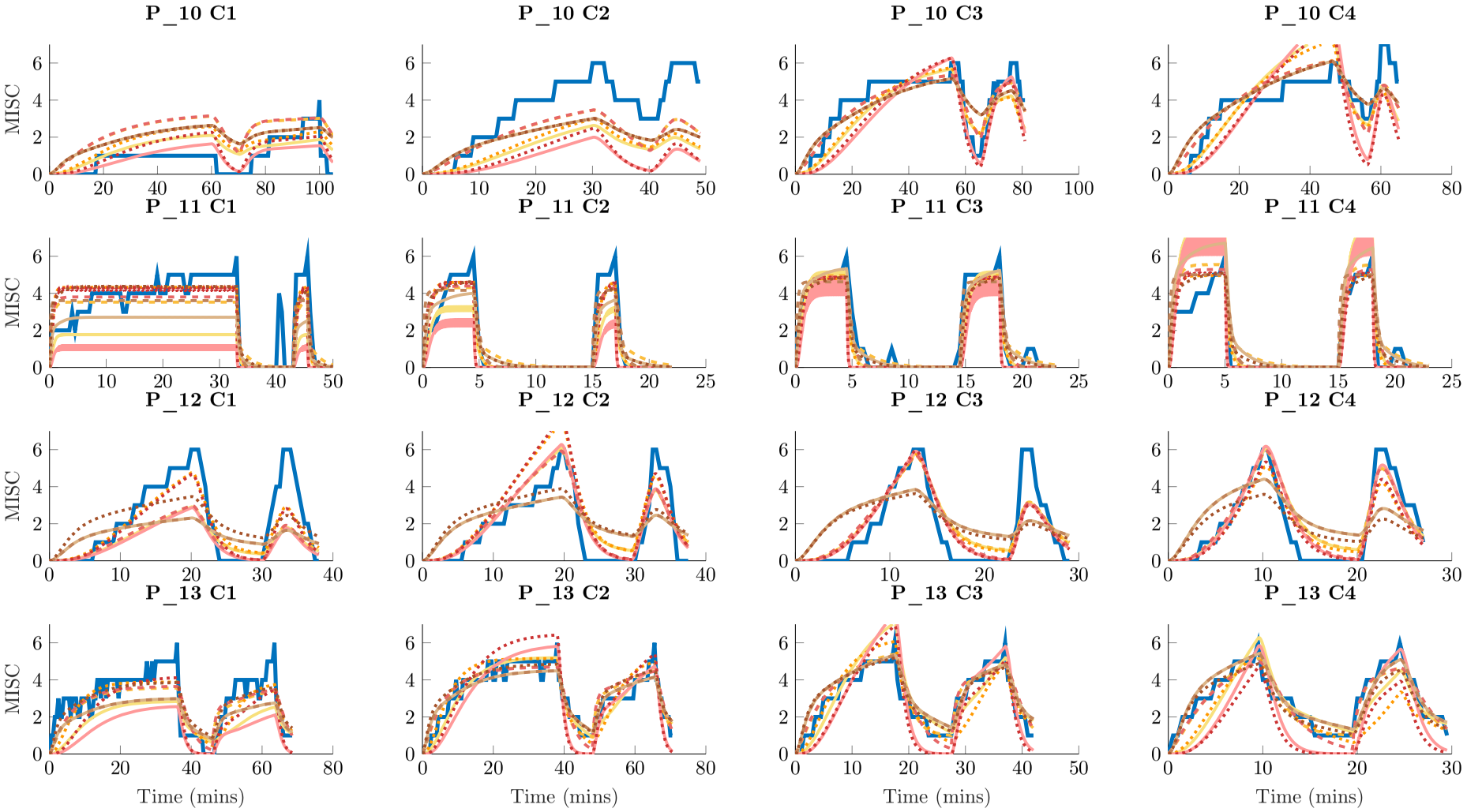


Figure 33: Model Fit 3/4

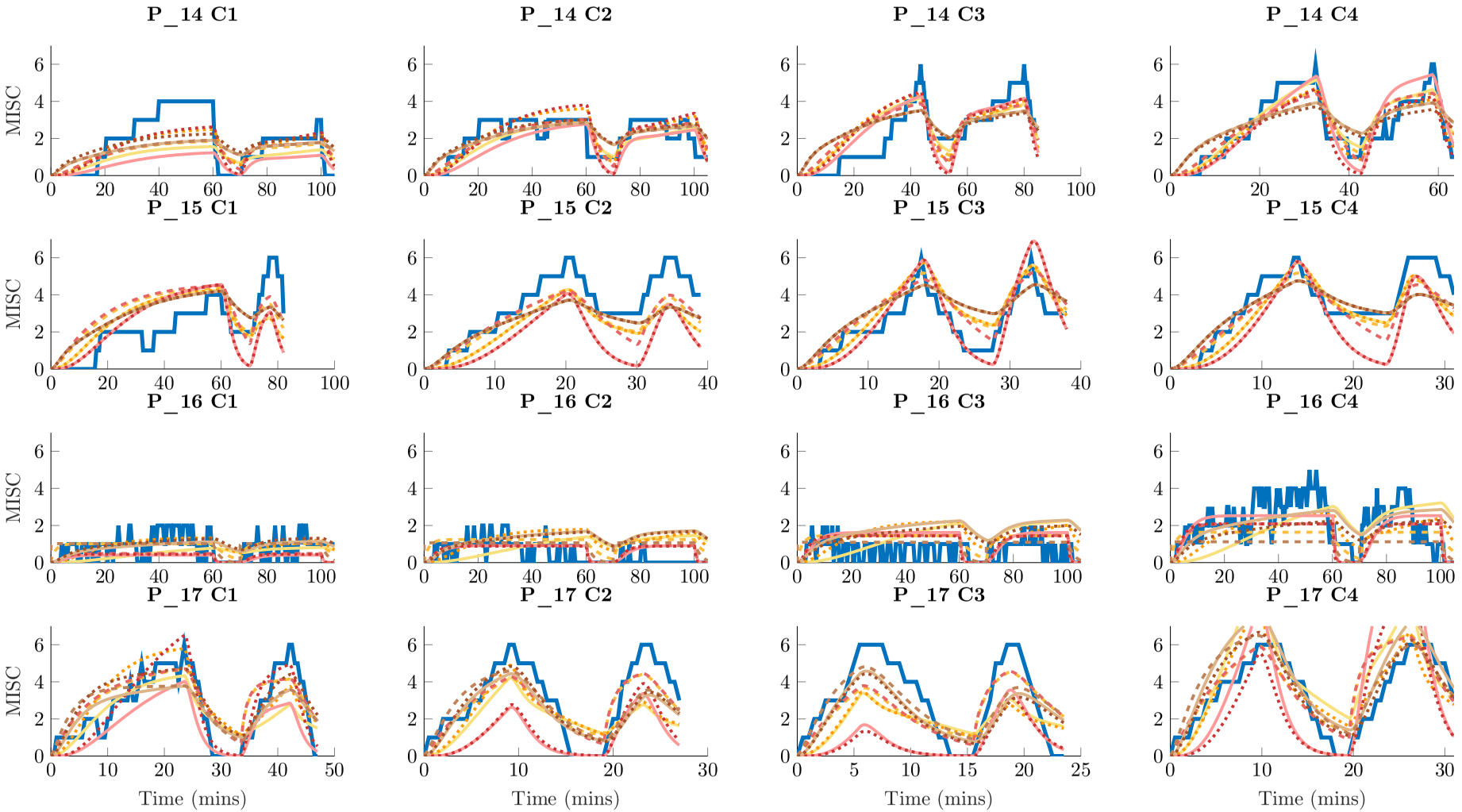


Figure 34: Model Fit 4/4

C Optimization Problem for Model Fitting

A constrained minimization problem was solved using the SQP algorithm through the *fmincon* MATLAB function. 18 multistart points were used to find multiple local solutions to a problem by starting from various points. This way a global solution could be obtained. The function to be minimized was the combined SMAPE for the model fitted on all four motion conditions together. This function and the constraints are shown below.

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \sum_{i=1.0}^{2.5} \frac{\mathrm{N_i SMAPE_i(X)}}{\sum_{i=1.0}^{2.5} \mathrm{N_i}} \\ \text{subject to} \quad & 0 \leq \mathrm{X(1)} \leq 100 \\ & \mathrm{X(1)} \leq \mathrm{X(2)} \leq 100 \\ & 0 \leq \mathrm{X(3)} \leq 200 \\ & 0 \leq \mathrm{X(4)} \leq 10 \end{aligned}$$

$\mathrm{N_i}$ is the number of data points at the each amplitude

X(1) and X(2) are the gains of the model. X(2) is skipped in the models without the addition as the gains only get multiplied so an additional gain is redundant.

X(3) is the time constant for the fast path. The time constant for the slow path is not needed as an relationship has already been established.

X(4) is the exponent in the power term of the model.

D Parameters for the fitted Model

Below shown are the parameters of the fitted model for the two best models obtained from two different metrics.

Participant #	Gain $\mathrm{K_1}$	Time Constant β_1
1	6.48	61.48
2	8.50	102.40
3	7.61	85.51
4	8.30	75.61
5	3.74	85.18
6	7.65	137.82
7	7.67	58.61
8	12.87	50.02
10	4.75	150.33
11	2.64	2.97
12	12.93	87.11
13	6.50	57.28
14	3.28	120.86
15	11.98	116.88
16	0.82	15.08
17	13.12	72.21

Table 2: Oman Model Without Addition

Participant #	Gain (Fast Path) K_1	Gain (Slow Path) K_2	Time Constant β_1	Exponent of Power term
1	0.08	3.62	12.43	0.72
2	2.16	2.16	90.93	1.06
3	0.96	2.98	65.40	1.04
4	2.57	2.57	60.88	0.26
5	1.91	1.91	67.83	0.16
6	1.93	2.72	114.30	0.55
7	0.46	4.02	46.07	1.09
8	2.40	2.40	44.15	1.50
10	0.08	3.63	88.24	1.03
11	0.00	4.64	1.99	0.12
12	2.85	2.85	74.00	0.55
13	0.20	4.07	19.35	0.37
14	1.56	1.56	94.13	0.58
15	1.34	3.54	100.70	1.17
16	0.25	0.26	14.51	2.49
17	0.00	6.29	23.61	0.56

Table 3: Oman Model with the power term at Input

E AICc Calculation

The Akaike information criterion (AIC) estimates the prediction error and therby gives the reltive quality of models. Thus, AIC provides a means for model selection. However, when the number of onservations (sample size) is low, AIC tends to overfit and will select models having too many parameters. To resolve this, AICc is used, which applies a correction to AIC. The equations used to calculate this are as follows,

$$\text{AICc} = \text{AIC} + \frac{2k^2 + 2k}{n - k - 1}$$

$$\text{AIC} = 2k - 2\ln(\hat{L})$$

$$\hat{L} = \frac{\sum_{i=2}^{5} O_i \ln\left(\frac{E_i}{O_i}\right)}{N}$$

$$N = \sum_{i=2}^{5} O_i = \sum_{i=2}^{5} E_i$$

Where, k is the number of parameters
n is the number of observations
O is the observed count
E is the expected count
i takes the value of classes 2 to 5

F MISC Prediction

F.1 Training

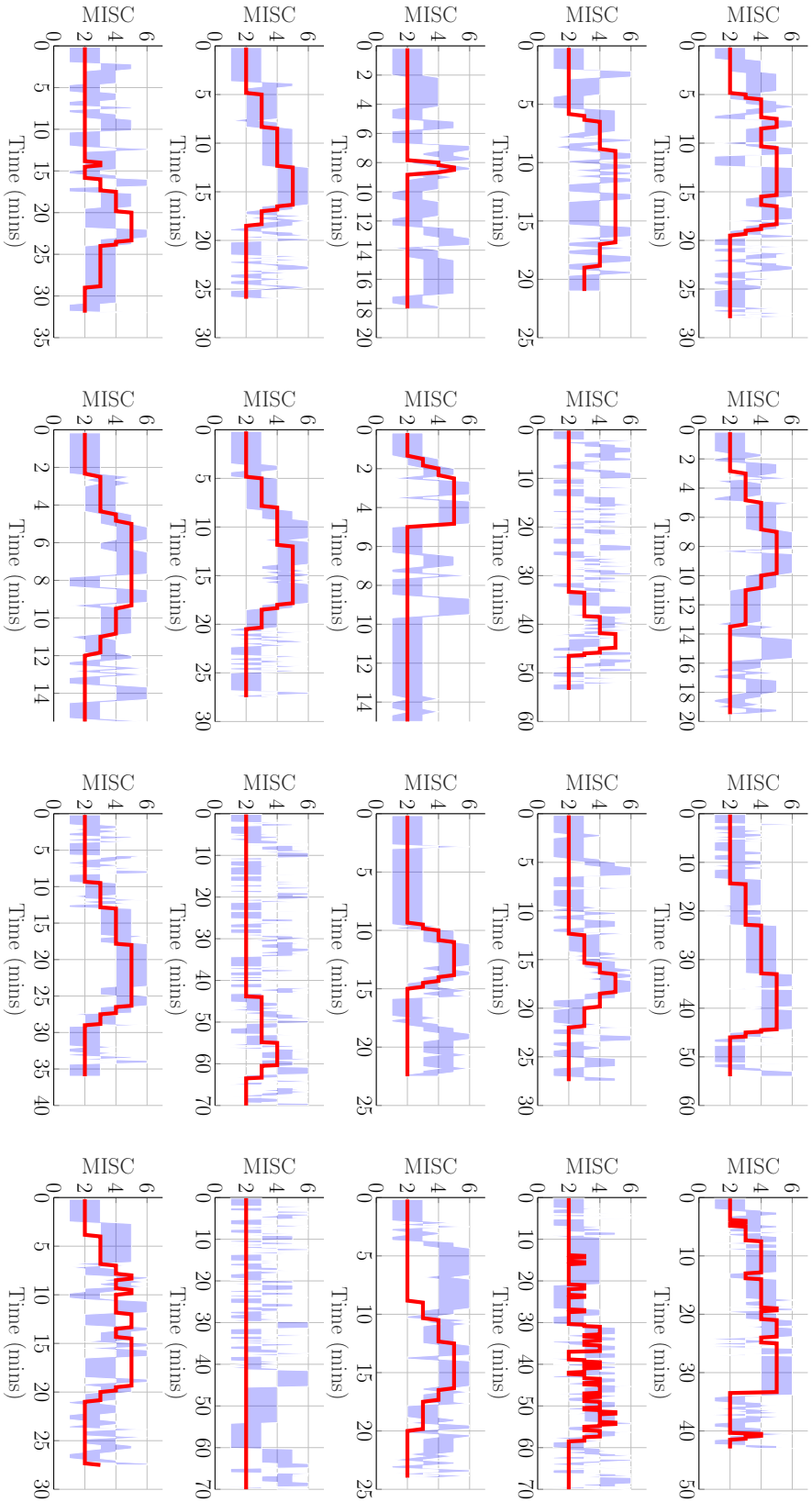


Figure 35: MISC prediction from GSR on Training Set 1/3

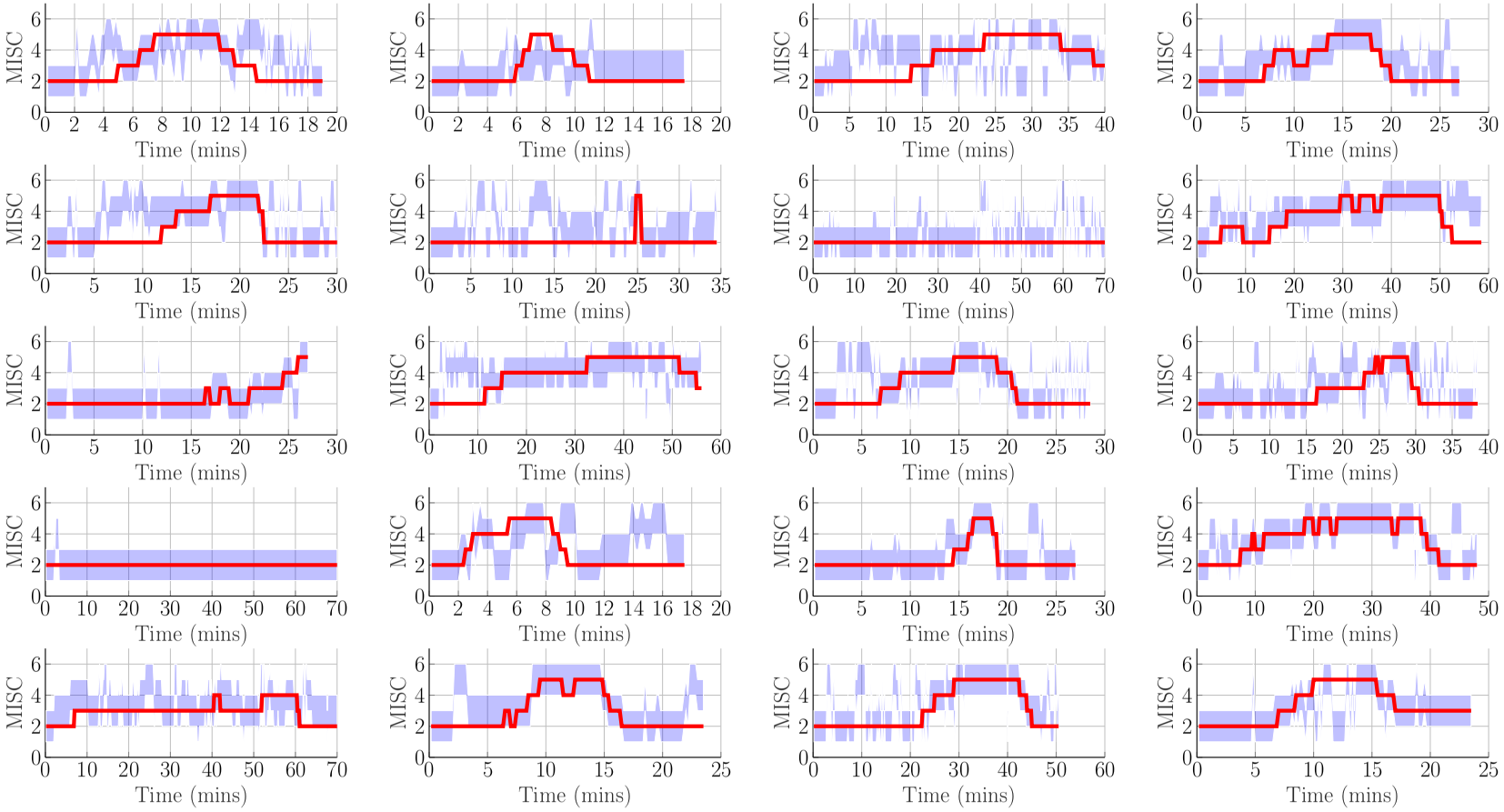


Figure 36: MISC prediction from GSR on Training Set 2/3

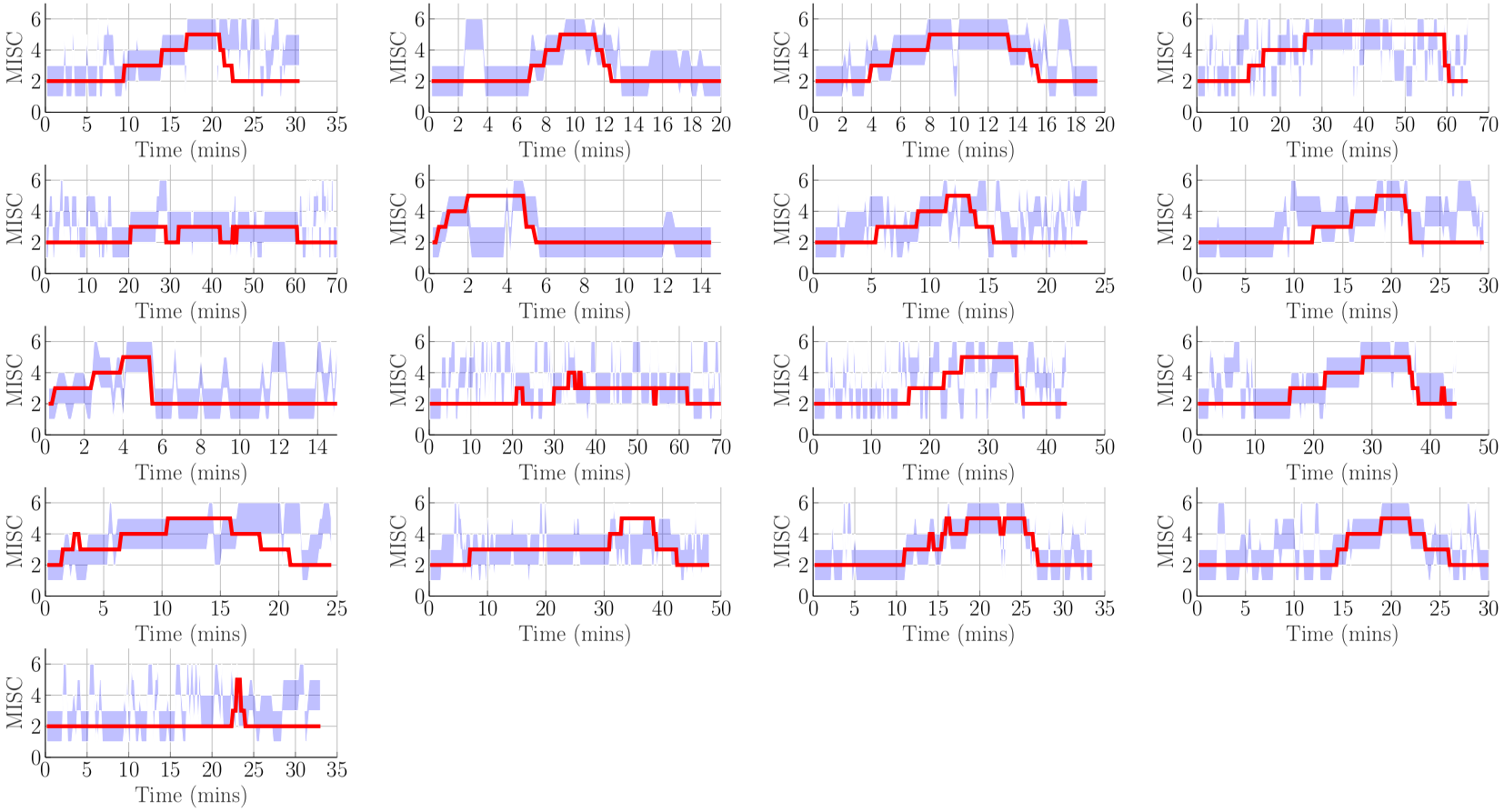


Figure 37: MISC prediction from GSR on Training Set 3/3

F.2 Testing

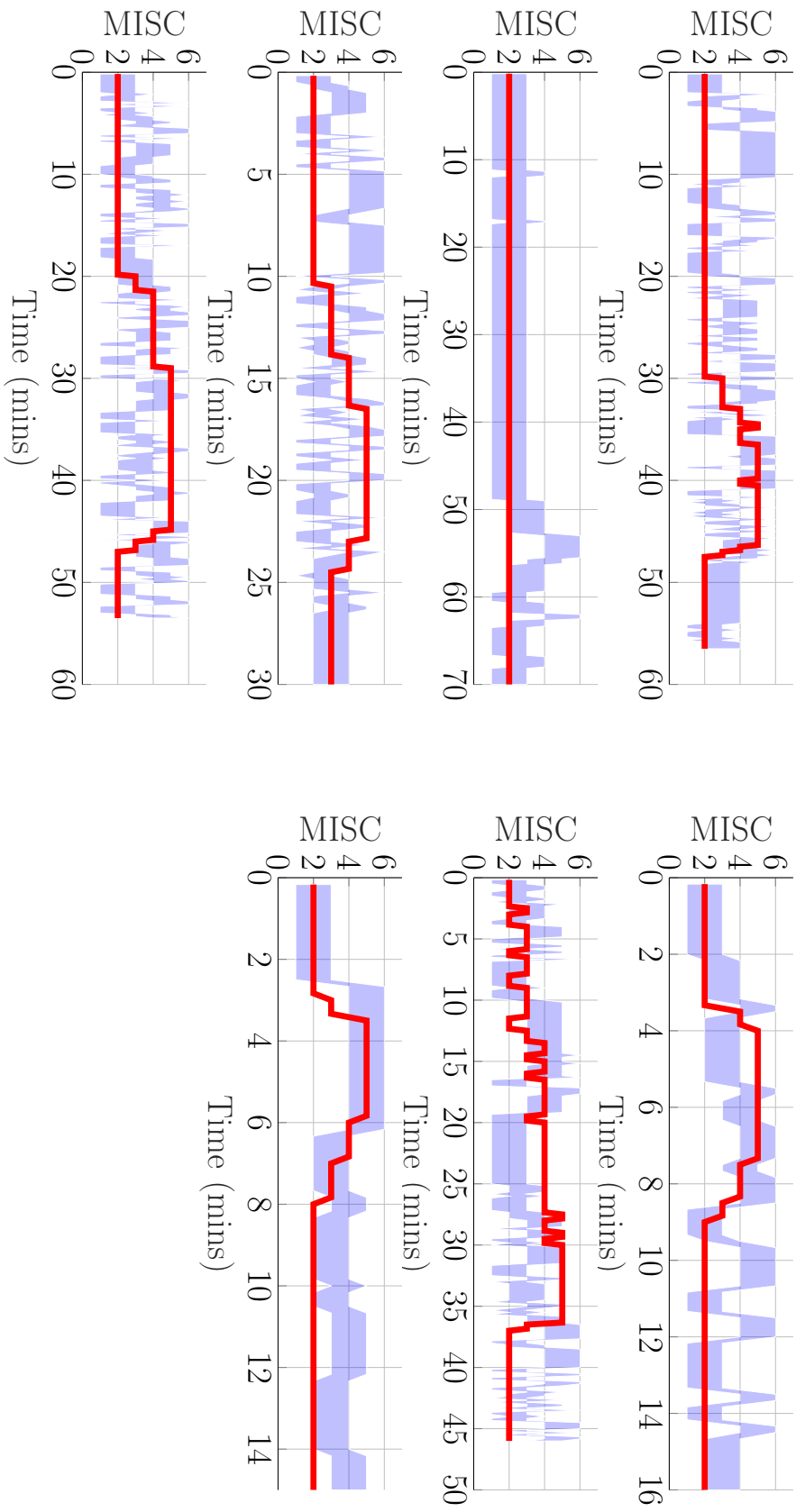


Figure 38: MISC prediction from GSR on Test Set

F.3 Validation

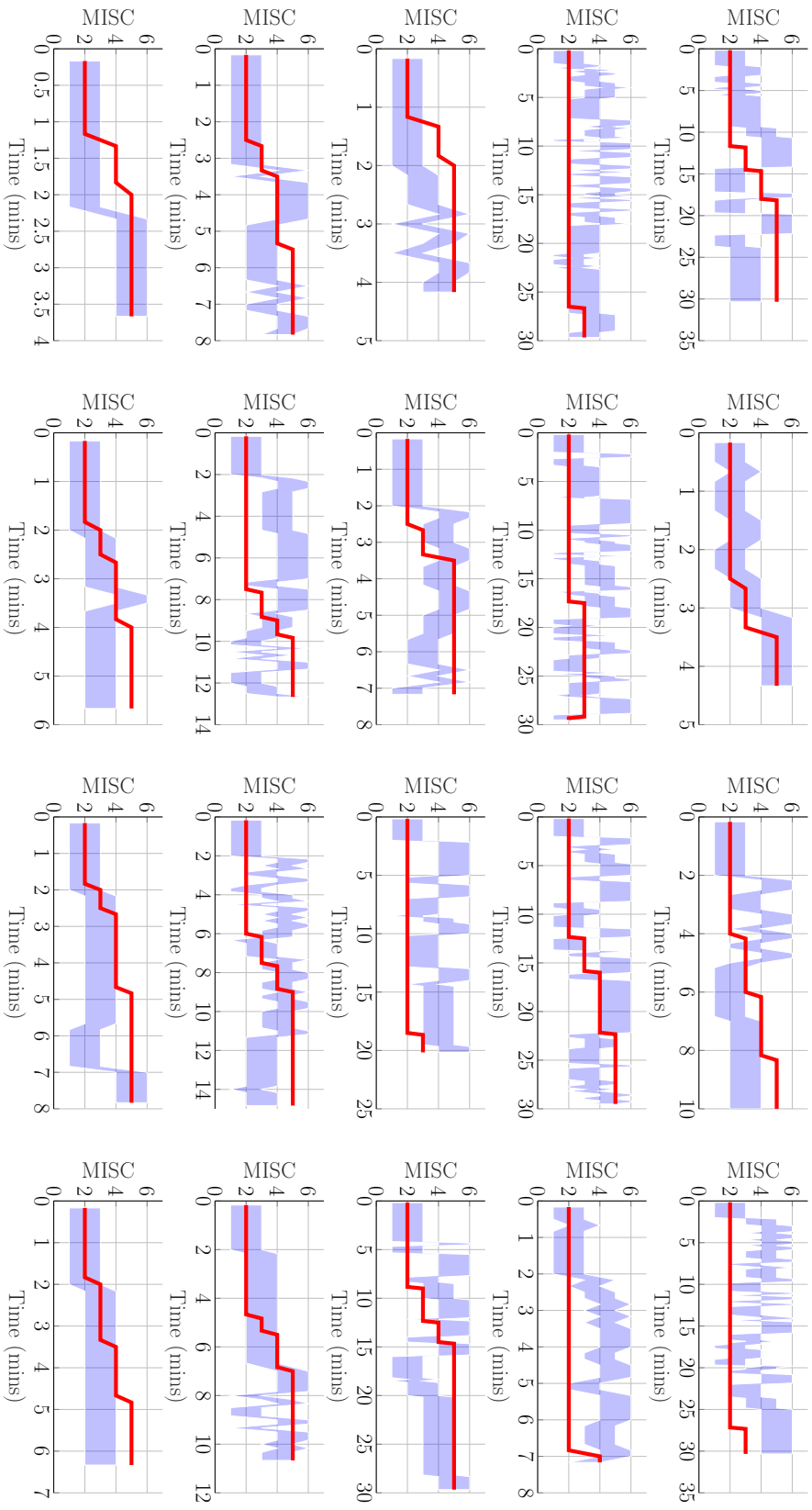


Figure 39: MISC prediction from GSR on Validation Set 1/3

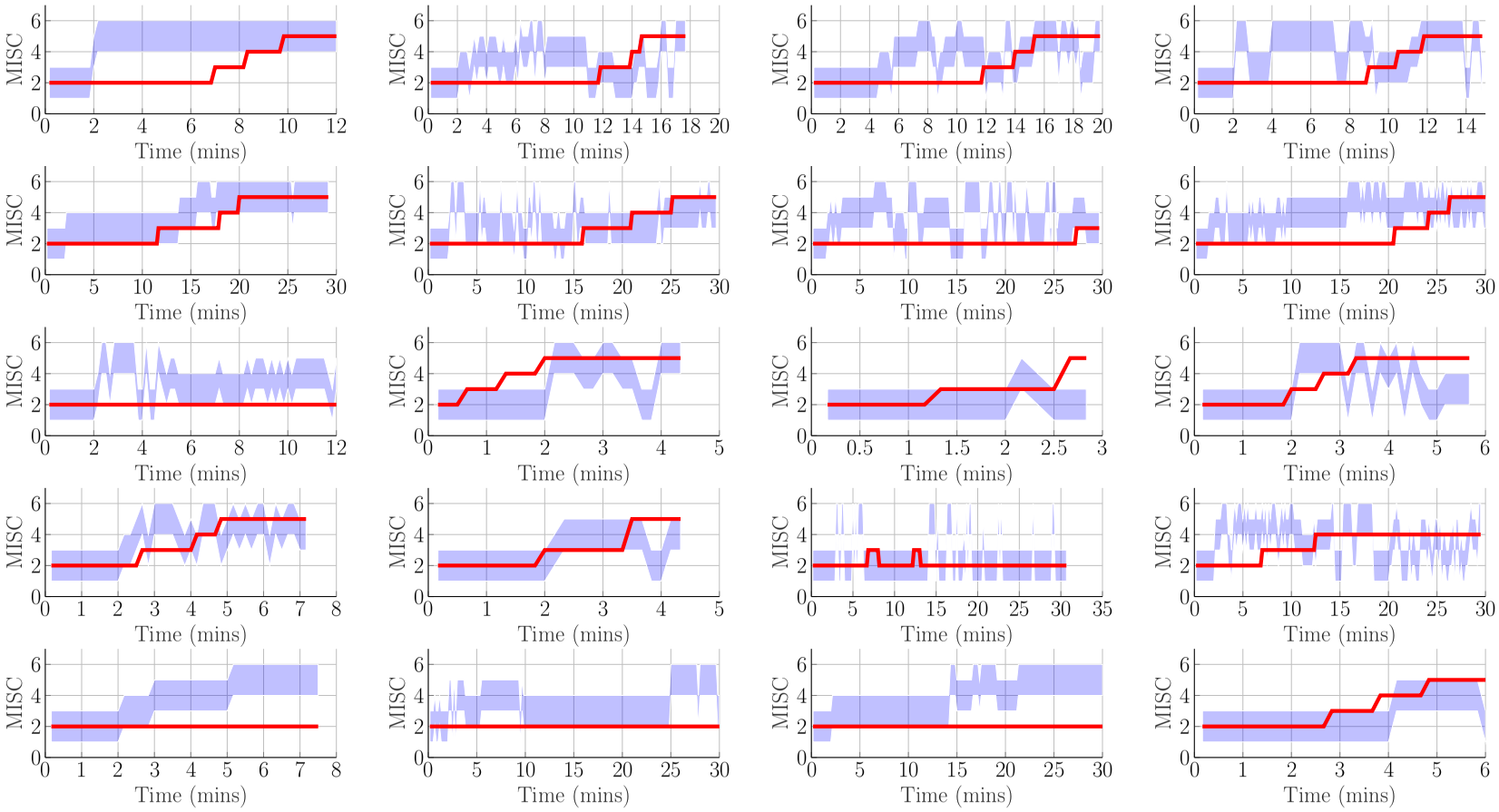


Figure 40: MISC prediction from GSR on Validation Set 2/3

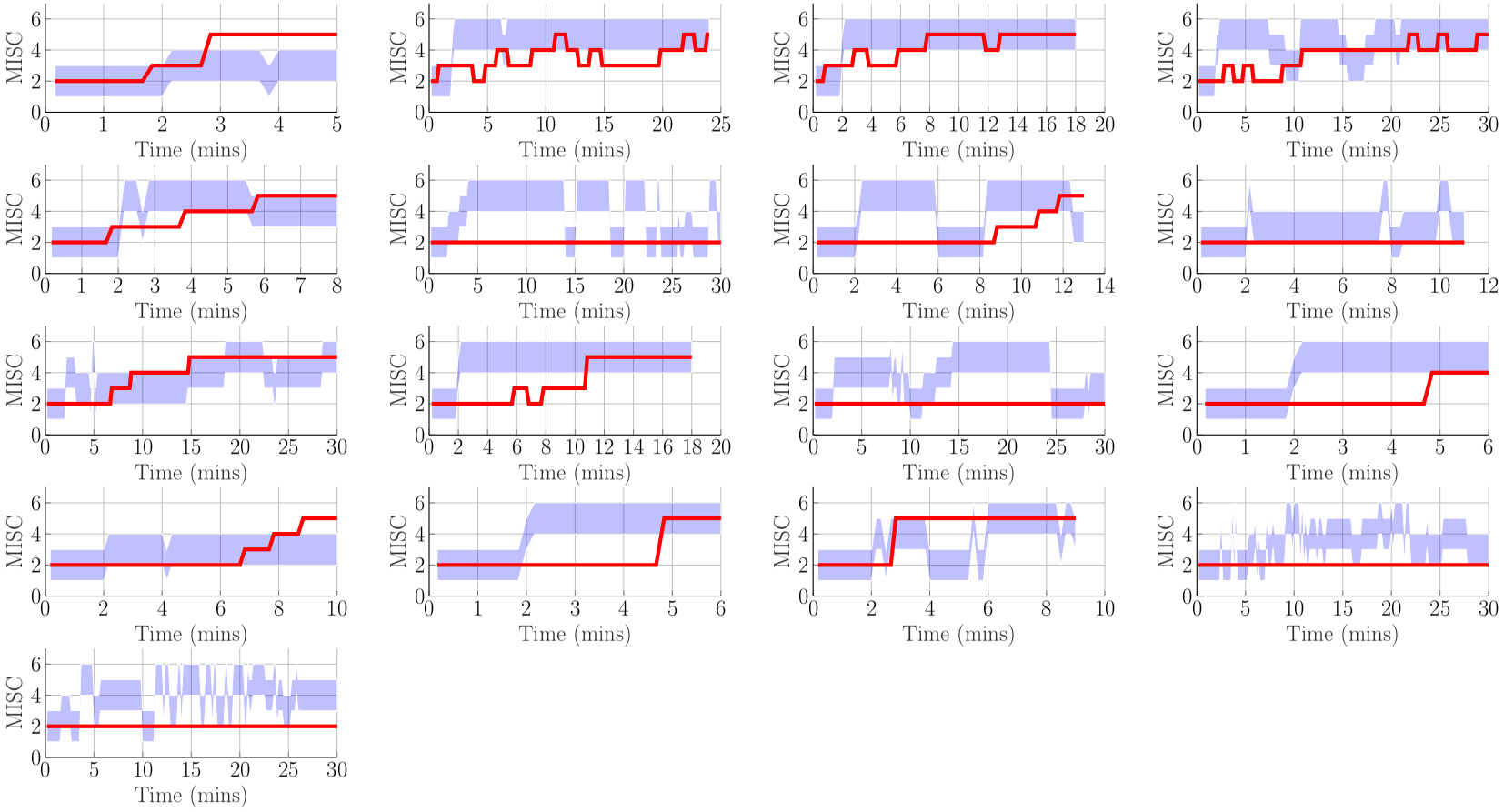
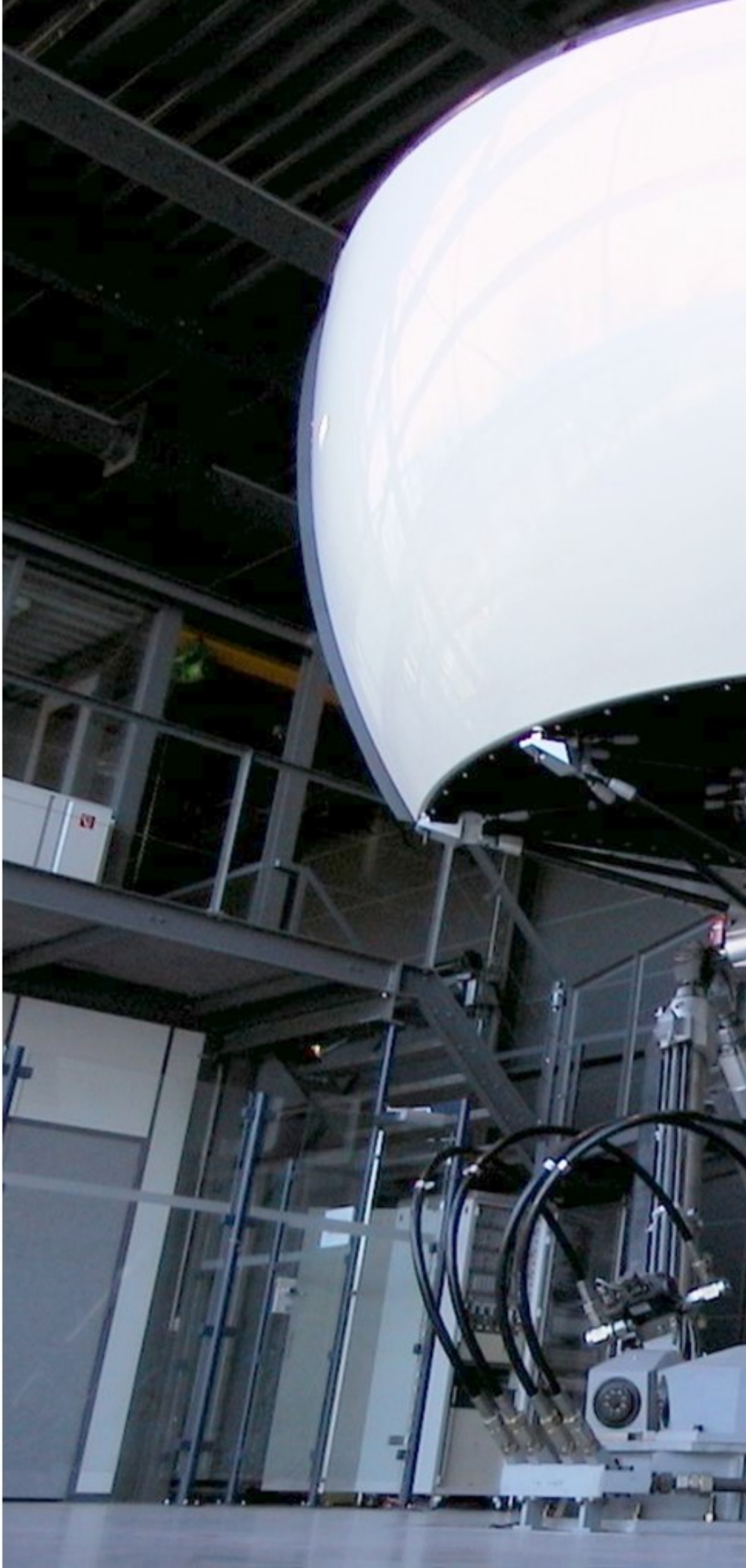


Figure 41: MISC prediction from GSR on Validation Set 3/3

Delft University of Technology



TU Delft