A MULTI-FACTORIAL ANALYSIS OF HUMAN PERFORMANCE DURING A 9-DAY SEA TRIAL

P. Valk[1], M. Grech[2] and J. Bos[1]

[1] TNO Human Factors, Soesterberg, The Netherlands, pierre.valk@tno.nl, jelte.bos@tno.nl
[2] Maritime Platforms Division, DSTO, Melbourne, Australia, Michelle.Grech@dsto.defence.gov.au

ABSTRACT
A multi-national sea trial on the effects of ship motions on human performance was performed on Canadian Forces Auxiliary Vessel Quest, off the coast of Nova Scotia, Canada, in February and March of 2007. The primary goal of these experiments was to obtain subjective and objective measures for human task performance, possibly affected by real ship motion. Based on the measurements from 12 subjects collected during 9 days at sea, a data base was constructed that was used for this research. Four types of variables were categorized; independent (e.g. ship motion), intermediate (alertness, sleep, fatigue), subjective task performance (cognitive, physical, workload), and objective task performance (dynamic vision, vigilance and tracking, multiple task performance, reaction time). Multi-factorial and regression analysis techniques were used to determine the effect of the so called intermediate variables on cognitive task performance, sea sickness and dynamic vision. Results indicate that increased feelings of misery led to impaired visual and cognitive performance. Furthermore, impaired sleep, high levels of fatigue and sleepiness affected cognitive performance. It is concluded that human performance at sea seems to be affected more by indirect effects of sickness, fatigue and impaired sleep, rather than ship motion per se.
Keywords: ship motion, sea sickness, dynamic vision, cognitive performance, sleep

1. INTRODUCTION

Given the trend towards reduction in crew numbers on navy vessels, the Royal Netherlands Navy has expressed the need for knowledge and models on the effects of ship motion on crew well being and performance. To that end, TNO Human factors was commissioned to conduct further analyses of the data gathered during the joint sea trial on the Canadian research vessel Quest 303 on the Atlantic Ocean in 2007. The Quest trial itself, as well as the analyses involving TNO data has already been described by Bos et al (2008), and will not be repeated here. Further analyses will focus on data collected by other ABCD partners (in particular the Australian Defence Science & Technology Organisation (DSTO), the Defence Research & Development Canada, and the Memorial University of Newfoundland, Canada), and will include such variables as sleep quality and sleep duration, alertness, fatigue, and postural stability. In addition to these parameters a larger meta-analysis on all data gathered is currently being realized within the framework of the ABCD WG on HP@Sea.

2. QUEST DATA ANALYSES

Data were available for 13 days, given an index and distributed over the days. Following extensive discussions among the partners who participated in the Quest trial, a basic set of data was assembled. A prerequisite for a meta-analysis comparing all the different variables is a complete data set, i.e., in each comparison, an equal number of observations should be present. In the dataset available, the number of observations per subject per day varied between variables. Hence, in order to manage these differences, we calculated the daily means of all parameters. Although we are aware of the possible fact that subtle differences over the day are obscured this way, it was felt that for the purpose of the meta-analyses, this was the most straightforward way of dealing with an incomplete set of data.

In the analysis process it was imperative to discern the following four basic categories between the variables as listed in Table 1.

Table 1. Variable categories.

Category	Description
1. ***Independent***	Variables that could not have been affected by any of the variables from the other categories (e.g. ship motion).
2. ***Intermediate***	Variables that could have been affected by the independent variables, but which are not performance measures *per se* (e.g., sickness).
3. ***Subjective performance***	Variables quantifying performance based on subjective ratings, typically obtained from questionnaires (e.g., task effort).
4. ***Objective performance***	Variables quantifying performance based on objective ratings, typically obtained from charting task results (e.g., reaction time).

Within each of these categories, variables showing a high correlation will not add to the information obtained from the analysis; hence, these have been excluded from further analyses. *Between* categories, high correlations may be indicative of relationships of interest. The idea behind this categorization has also been exemplified by the flow chart shown in Fig. 1.

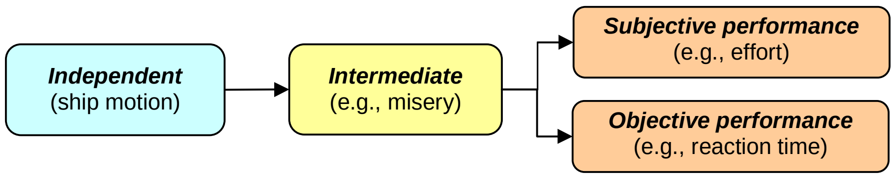


Figure 1. Flow of information from independent via intermediate to performance variables.

Furthermore, variables showing hardly any variability were deemed to be of no interest. This has been tested by determining the ratio between observed variability relative to the possible range. Variables showing none or only limited variability have been excluded from further analysis. This left the variables listed in Table 2.

Lastly, not all parameters left, although of possible interest with respect to questions not addressed in the project description, needed to be analysed. In this regards a subset of variables assumed to be suited for the purpose of answering the questions posed in the project description were selected for use in the analyses to be performed for the HPAS2010 conference.

Table 2. Final set of variables of interest with the subset used for the current paper indicated by italics.

1. Independent	***2. Intemediate***	***3. Subjective performance***	***4. Objective performance***
Lat. acc.	*CSS average workload* *CSS fatigue* *MISC misery* *Groningen Sleep Quality Scale* *Stanford Sleepiness Scale* *Number of awakenings* *Hours slept*		*DVA stat* *DVA RS3* *MAT rt* *MAT rms* *VigTrack rms* *VigTrack false* *VigTrack incorrect*
Vert. acc.	PAQ sum-sleep CSS max workload daily food PANAS pos. PANAS neg. CogFail	PAQ sum-cog PAQ sum-phys PAQ sum-fail TLX summed	DVA SP15 Romberg CFF SART rt SART hit SART reject

CSS= Crew Status Survey, MISC = Miscery Scale, PAQ = Performance Assessment Questionnaire, PANAS = Positive and Negative Affect Schedule, CogFail = Cognitive Failures Questionnaire, TLX = NASA Task Load Index, DVA (stat/RS3/SP15) = Dynamic Visual Acuity (no motion/retinal slip/smooth pursuit), MAT = Multi-attribute Task battery, VigTrack = Vigilance & Tracking test, CFF = Critical Flicker Fusion, SART = Sustained Attention to Response Test

3. RESULTS

First all parameters were plotted as a function of time, i.e., the trial day-index. This was done in order to gain a better perspective of the nature of these parameters. In addition, the vertical and lateral ship acceleration was added to each graph due to the important nature assumed to be linked to these two parameters. Appendix A provides an overview of these graphs for all variables under consideration.

First we conducted simple analyses, i.e., that linear correlation and *t*-tests, looking for main-effects only at this stage. Further analysis with input from the DSTO, was conducted involving higher level linear regression analyses.

Accordingly, correlations between variables of the categories 1 onto 2 (see Table 2) were looked at, i.e., ship motion on the intermediate variables. Table 3 lists the results thereof, and shows that increased ship motions lead to impaired sleep, increased sleepiness and higher feelings of misery.

Table 3: correlation matrix for independent and intermediate variables; ++ = *p*<0.01, + = *p*<0.05, +/-=*p*<0.1.

	average workload	fatigue	sleepiness	sleep quality	awakenings	total sleep time	misery
Lateral acc.				++	+/-		++
Vertical acc.			+/-	++	+/-		++

Table 4 shows the significance of the relationships between the objective performance measures on the one hand and ship motion and intermediate variables on the other. The distribution of the scores on the number of awakenings and the misery scale however, appeared to be skewed, hence these two variables were divided into 2 categories ('low' vs. 'high'). For both variables the 'low' category was scored if the subject reported the absence of awakenings or feelings of misery. The 'high' category contained all other scores (i.e. >0). The relationship for these two intermediate variables and performance variables were computed using a t-test.

Table 4 also shows that increased ship motions lead to increased workload, impaired sleep, increased sleepiness, increased feelings of fatigue and misery lead to impaired visual and cognitive performance (especially perceptual-motor-skills).

Table 4. Correlation and HLM model significance / t-test (*) for dependent (performance) and independent/intermediate variables; +++=p<0.001++ = p<0.01, + = p<0.05, +/-=p<0.1.

	DVA static		DVA RS-3		MAT rt		MAT rms		VigTrack rms		VigTrack false		VigTrack incorrect	
	corr	HLM	corr	HLM	corr	HLM	corr	HLM	corr	HLM	corr	HLM	corr	HLM
Lateral acc.														
Av. workload				+++			+++		+/-		+/-			
Fatigue							+++		+++					+++
Sleepiness	+	+/-	+++				+++	+/-	+++			+	+	++
Sleep quality			+/-						+					
Awakenings*	+++		++											
Sleep duration									+	+/-				+
Misery*		+++	+++				+++	+++	+/-			+/-		+
Day				+++		+++		+				+		+
Quadratic day						+++						+		
Fatigue x Misery										+/-				
Fatigue x Day														+

To obtain an insight into the data, we next performed a series of factorial analyses using multi level modelling technique. Hierarchical Linear Modeling (HLM) was used because it allows an examination of how data vary systematically across different levels of analysis. This is relevant to the present research as data varies across three levels of analysis; between-session (four sessions per day), between-days (across 13 days), and between-persons (12 individuals). Due to the nature of the data and the amount of missing data at the session level it was decided to focus on the day level and ignored the session level, so a two level model was used instead. At level 1, the within-person variables were measured across the study period. These included such factors as ship motion, performance, fatigue, workload, sleep quality, sleepiness, seasickness, sleep duration and number of awakenings (see Table 4). Each of these within person variables (Level 1) is nested within the 12 participants (within person level 1 variables). No between-person predictors were included in these analyses.

The dependent variables used for these analyses included the objective performance measures, namely the title row in Table 4. To examine whether any of the variables listed under the first column in Table 4, had linear or non linear (quadratic) effects on performance, both linear and quadratic variables were entered into the models. Quadratic variables were computed by creating a cross-product of the relevant centered linear variables. Single level interaction parameters were computed by creating a cross-product of the relevant centered variables. The significance of all variables was assessed via t-tests at $p < .05$. Equation [1] shows an example of the full simplified model with all relevant variables within the model[1]. For each performance measure under examination the model shown in equation [1] was used.

[1] Note that only one quadratic effect is shown in the model (quadratic day). All other variables were tested for quadratic effects but none were found, hence these were excluded from the final model. Similarly, a combination of interaction effects was tested with few evident, namely the fatigue by seasickness interaction as shown in the results table. Hence, this was retained within the final models.

Equation [1]:

Level 1: Performance = Intercept + Day + Quadratic Day + Fatigue + Workload + Sleep Quality + Sleep Duration + No. of Awakenings + Seasickness + Sleepiness + (Fatigue x Seasickness) +error

Table 5 shows the results of the empty models run for each of the dependent variables. These models indicated that the percentage amount of variance in performance shown in column (1) of Table 5 was within persons and that shown in column (2) on Table 5 was between persons. The random effects for the intercept at the between person level was significant with good reliability (≥0.85) for all performance scores. Thus, it was appropriate to specify the Level 1 variables as predictors of performance scores.

Table 5. Percentage variance and random effects for all performance variables at each level (within and between person); +++=p<0.001++ = p<0.01, + = p<0.05.

Performance	Variance in performance		Random Effects	
	Within-Person (1)	Between-Person (2)	Effect	Reliability
DVA Static	27.70	72.30	$\chi^2_{(11)}$ = 222.99, +++	0.95
DVA RS-3	27.75	72.25	$\chi^2_{(11)}$ = 204.31, +++	0.95
MAT rt	65.05	34.95	$\chi^2_{(11)}$ = 72.89, +++	0.85
MAT rms	33.25	66.75	$\chi^2_{(11)}$ = 247.68, +++	0.96
VigTrack rms	57.41	42.59	$\chi^2_{(11)}$ = 99.27, +++	0.89
VigTrack false	31.28	68.72	$\chi^2_{(11)}$ = 275.82, +++	0.96
Vigtrack Incorrect	20.10	70.90	$\chi^2_{(11)}$ = 301.78, +++	0.96

Next the level 1 growth curve variables were entered into each model to test how these affect performance scores as specified in equation [1]. As indicated a separate model was created for each performance score.

As shown in Table 4 misery was a positive significant predictor of DVA Static (B = 0.061, t(64) = 5.035, p = .000). In this regards the higher the misery value the worse the DVA Static. In addition sleepiness was a negative marginally significant predictor of DVA Static (B = -2.031, t(64) = -1.985, p = .051). There were no other significant Level 1 predictors of DVA Static.

The findings also indicate that time across days was a negative significant predictor of DVA-RS3 (B = -0.028, t(62) = -3.997, p = .000). This is also evident is Appendix A graph for DVA-RS3 showing a decrease in this value across days. Average workload was also found to be a significant negative predictor of DVA-RS3 (B =- 0.064, t(62) = -4.880, p = .000), with higher workload leading to lower DVA-RS3 scores and vice versa.

Within this model, the linear (B = -1.107, t(75) = -4.551, p = .000) and the quadratic trend of day (B = 0.660, t(75) = 4.666, p = .000) were significant. This result is also evident in the MAT-RT by day graph shown in Appendix A, showing a decrease in the MAT-RT scores across days. This increased slightly from day 10 onwards, hence the quadratic effect becoming significant. Similar to the correlation results shown in Table 4 there were no other significant predictors of MAT-RT.

Similar to the DVA Static, misery was a positive significant predictor of MAT-RMS (B = 1.377, t(75) = 7.354, p = .000). This effect was also evident in the correlation result shown in Table 4. There was a significant linear trend of day (B = -2.138, t(75) = -2.244, p = .028), as the MAT-RMS score decreased across days, which is also evident in the MAT-RMS by Day graph shown in Appendix A. Sleepiness approached significance (B = -0.485, t(75) = -0.478, p = .084) with a negative linear trend.

Sleep duration was a significant predictor of VigTrack-RMS (B = -0.514, t(76) = -1.685, p = .096). This indicates that participants performed worse on the VigTrack-RMS test when their sleep duration was lower. In addition, a marginally significant interaction effect was evident between fatigue and misery (B = 1.466, t(76) = 1.683, p = .096). For the high misery group, the effect of fatigue on VigTrack-RMS was positive with the opposite evident in the low misery group.

Similar to the MAT-RT model, the linear (B = -7.773, t(75) = -2.054, p = .043) and the quadratic trend of day (B = 3.896, t(75) = 2.211, p = .030) were significant predictors of VigTrack-False. The graph in Appendix A for the changes in Vigtrack-False across days shows this quadratic trend with this performance measures getting better across days, then from day 11 onwards. The effect of sleepiness on VigTrack-False was significant (B =

2.982, t(75) = 2.189, p = .032) showing a linear positive trend. No other effects were evident for VigTrack-False.

Fatigue was a significant positive predictor of VigTrack-Incorrect (B = 8.413, t(75) = 4.467, p = .000) with high levels of fatigue leading to high levels of VigTrack-Incorrect scores. There was also a significant interaction between fatigue and day on VigTrack-Incorrect (B = -6.068, t(75) = -2.614, p = .011) showing a positive interaction between fatigue and VigTrack-Incorrect of the first day of the trial, with this effect weakening across the trial. Misery was also a positive significant predictor of VigTrack-Incorrect (B = 1.183, t(75) = 2.291, p = .025). Sleep duration was a significant negative predictor of VigTrack-Incorrect (B = -1.579, t(75) = -2.605, p = .011) with participants performing worse when their sleep duration was low. Sleepiness was also a negative significant predictor of VigTrack-Incorrect (B = -4.460, t(75) = -2.605, p = .011).

4. DISCUSSION & CONCLUSIONS

With respect to the meta-analysis of the data gathered by different ABCD partners in the 2007 Quest 303 trial the main focus of this analysis was on seasickness, sleep quality, sleepiness and sleep duration, alertness, fatigue, and postural stability.

Our over-all conclusion in this respect confirms the one stated before in Bos et al. (2008), i.e., that human performance at sea is less affected by ship motion per se, but more by its intermediate effects, i.e., by seasickness, fatigue, and impaired sleep. This conclusion is based on the fact that less significant and smaller effects were observed in Tables 3 and 4 of ship motion on objective task performance per se. The most, highest significant and largest effects were found in the intermediate factors on objective task performance. These observations however, do not allow a firm conclusion to be drawn. Theoretically, a number of effects may be ascribed to learning or other (more) complex phenomena that could have been accounted for by performing the same tests without motion. These control conditions however, are lacking. Yet, we are confident about the validity of these conclusions, because several aspects are hard to explain otherwise, such as the fact that motion does affect sickness in a clear and understandable way (see Appendix A), it does not affect the MAT rms that clear, while yet there is a highly significant effect of sickness on the MAT RMS.

In this paper, we made a distinction between independent, intermediate, subjective performance and objective performance variables. Within the final models there were a number of factors that impacted on performance. These were similar to the trends obtained in the initial analyses, namely:

- increased feelings of misery led to impaired visual and cognitive performance (DVA Static, MAT-RMS and VigTrack-Incorrect);
- decreased sleep duration led to impaired cognitive performance (VigTrack-RMS and VigTrack-Incorrect);
- high levels of fatigue impaired cognitive performance (VigTrack-Incorrect);
- fatigue impaired cognitive performance when misery was high (VigTrack-RMS);
- high levels of sleepiness affected cognitive performance (VigTrack-False) negatively;
- a learning effect was evident for some of the cognitive performance measures (MAT-RT and VigTrack-False) with improvement in performance evident across the first few days of the trial. However, performance did level out and become worse during the last few days of the trial as evident by the significant quadratic effects.

The detrimental effects of increased workload and impaired sleep quality on cognitive performance were more pronounced during the second week of the trial probably due to cumulative effects.

Regarding the dynamic visual acuity test, the findings indicate a difference between the different sub-tests depending on misery/sickness based on the daily averages, which we did not find previously based on the individual tests (Bos et al, 2008). This adds to the value of using dynamic vision to quantify human performance, rather than using the static vision only.

For the cognitive performance tests (MAT and VigTrack) the root mean square tracking error (RMS) and the incorrect responses (VigTrack) appeared to be the most sensitive parameters for performance impairment as a result of increased workload, fatigue, misery and impaired sleep. This finding is in concordance with other fatigue and sleep deprivation studies, and other studies investigating sedative effects of drugs and alcohol.

The relationship between postural stability and seasickness (not addressed in this paper) remains of special interest with respect to the model developed at TNO to predict seasickness. This model is based on a hypothetical mechanisms controlling body motion, including an error signal assumed to be correlated with motion sickness severity (see e.g., Bos & Bles, 1998). Although this model has been validated in several ways already, a positive correlation between postural instability and sickness severity would add to the validity of using the mechanism to predict motion sickness in general.

The basic conclusion drawn here thus concerns the observation that human performance at sea seems to be affected more by indirect effects of sickness, fatigue and impaired sleep, rather than ship motion per se. This implies that ***it does not make sense to reduce ship motion to improve human performance without further considering sickness, fatigue and sleep quality.*** Accounting for better sleep conditions aboard (improved bunks and/or work rest schedules) rather than focusing on the more costly ride control system, for example, may improve the crew's morale and performance.

5. REFERENCES

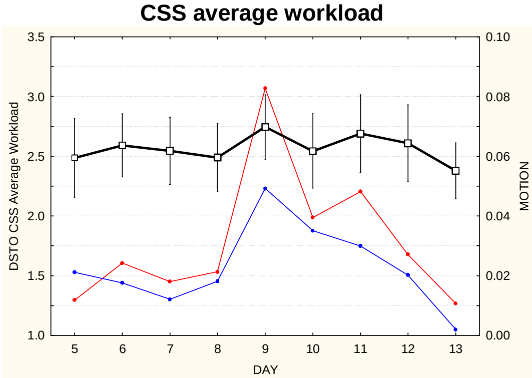
Bos JE, Bles W. (1998) Modelling motion sickness and subjective vertical mismatch detailed for vertical motions. Brain Research Bulletin 47:537-542.

Bos JE, Valk PJL, Hogervorst MA, Munnoch K, Perrault D, Colwell JL. (2008) TNO contribution to the Quest 303 trial - Human performance assessed by a vigilance and tracking test, a multi-attribute task, and by dynamic visual acuity. TNO report, TNO Human Factors, Soesterberg, the Netherlands TNO-DV 2008 A267.

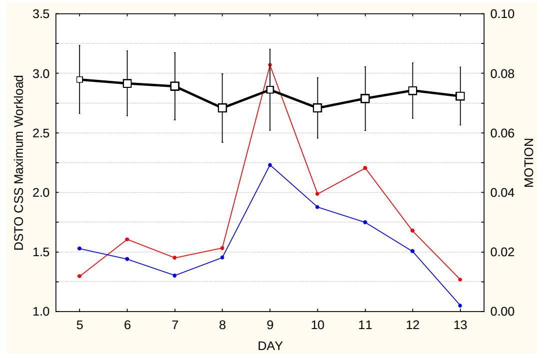
Appendix A: Basic variables as a function of time

Intermediate variables

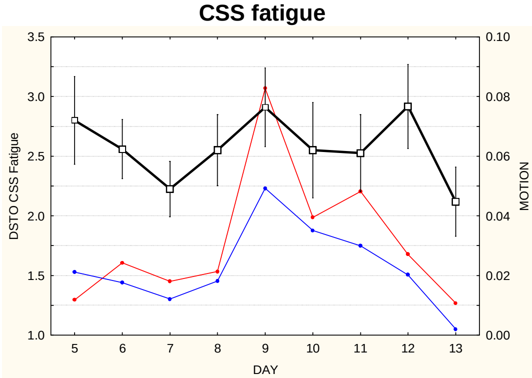
CSS average workload



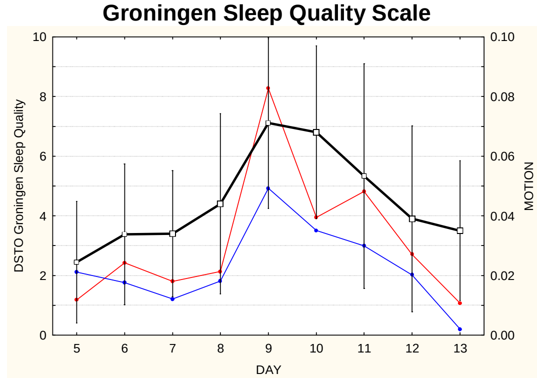
CSS maximum workload



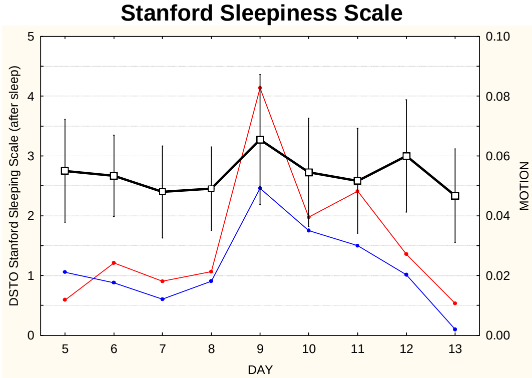
CSS fatigue



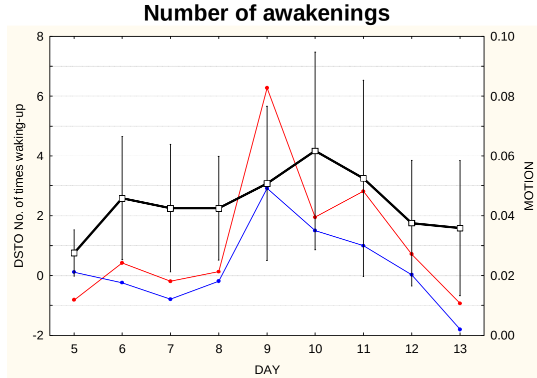
Groningen Sleep Quality Scale



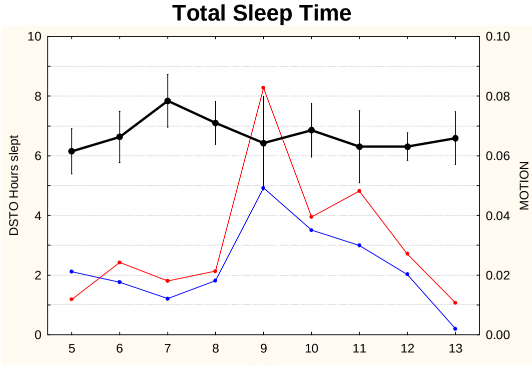
Stanford Sleepiness Scale



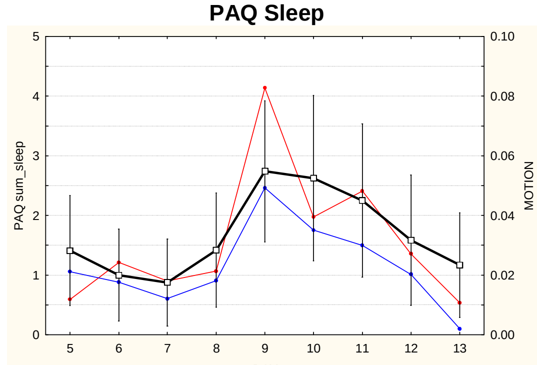
Number of awakenings



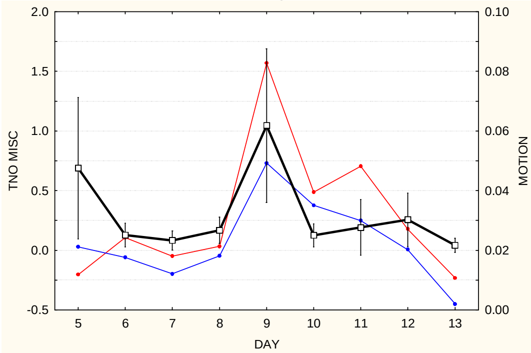
Total Sleep Time



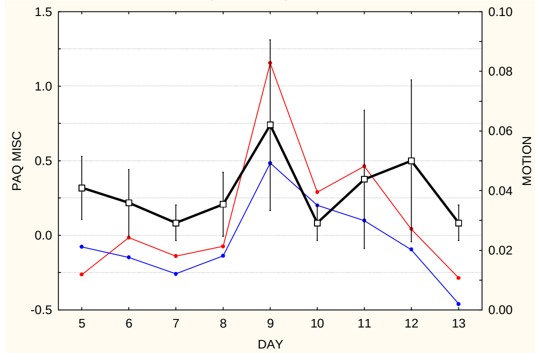
PAQ Sleep



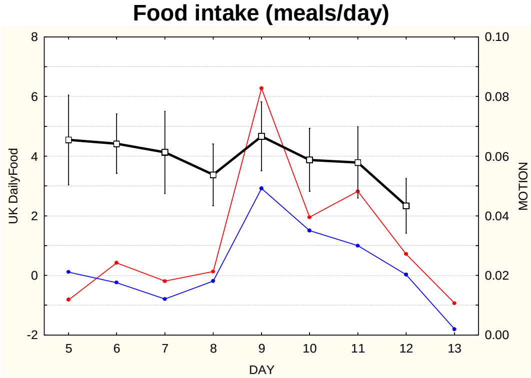
TNO Misery Scale



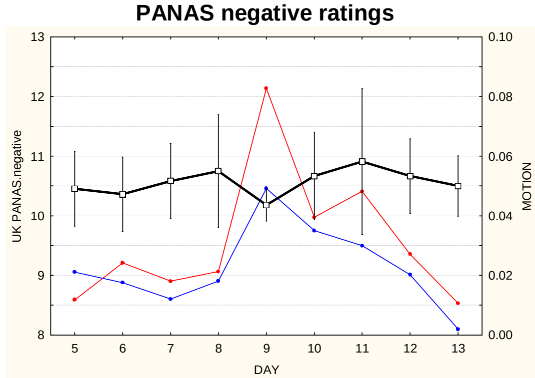
PAQ Misery Scale



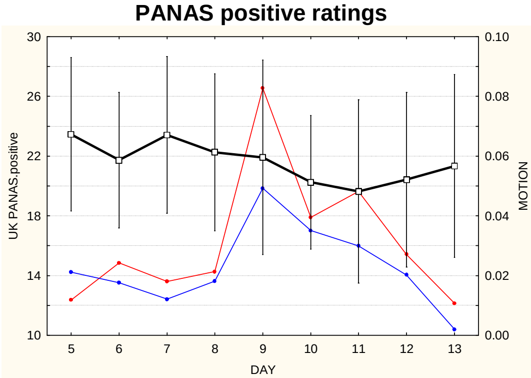
Food intake (meals/day)



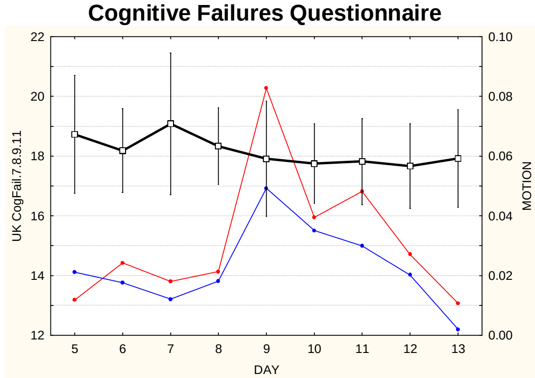
PANAS negative ratings



PANAS positive ratings

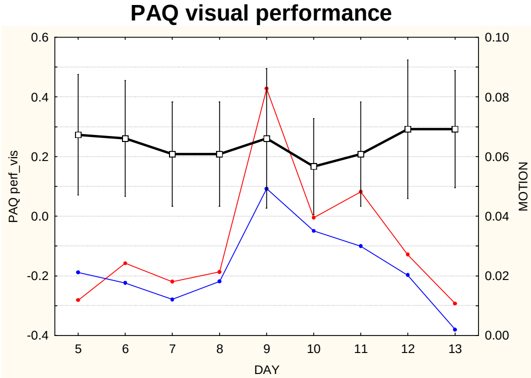


Cognitive Failures Questionnaire

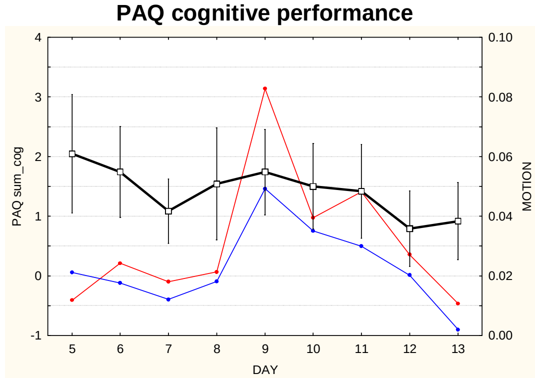


Subjective performance

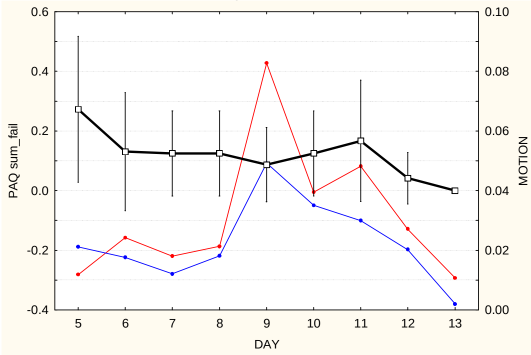
PAQ visual performance



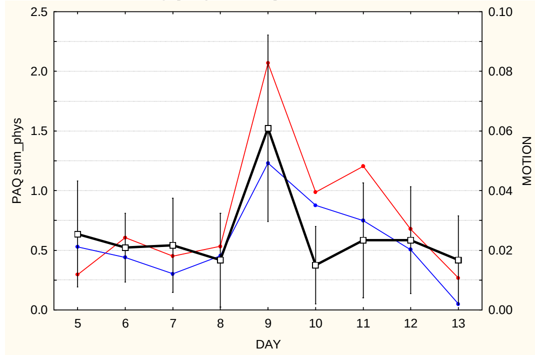
PAQ cognitive performance



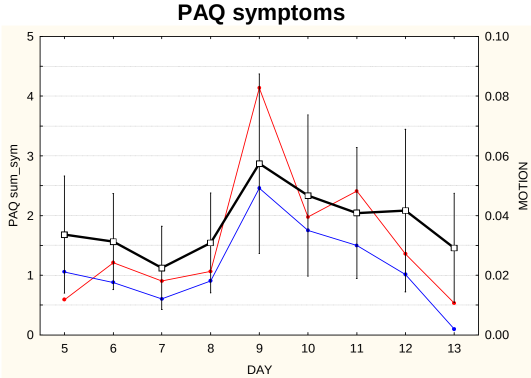
PAQ failures



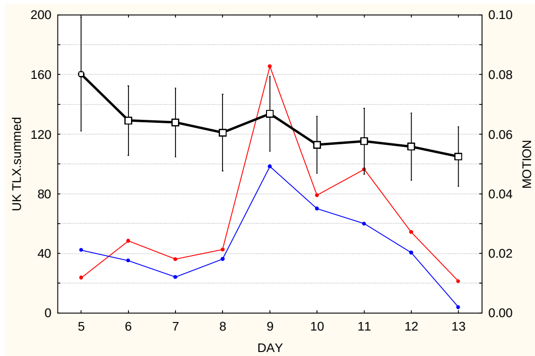
PAQ physical performance



PAQ symptoms

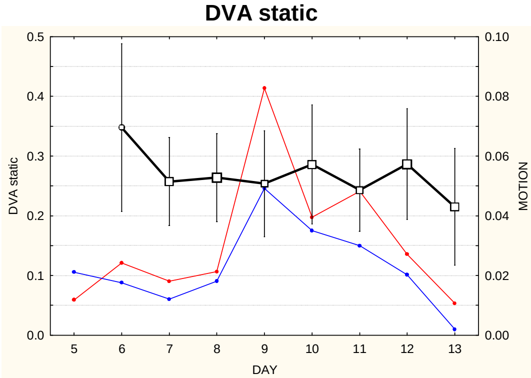


NASA TLX workload

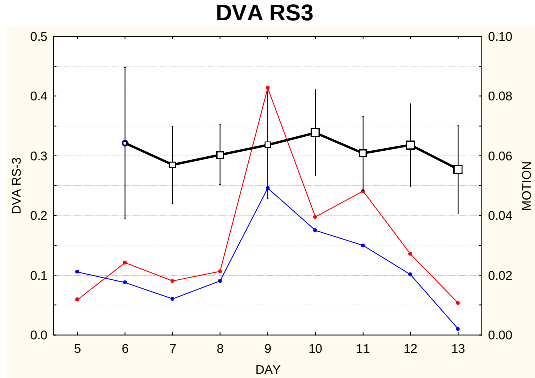


Objective performance

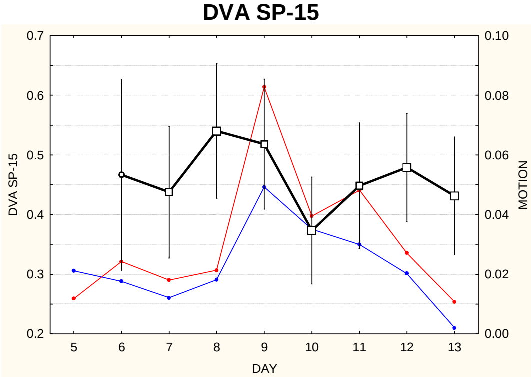
DVA static



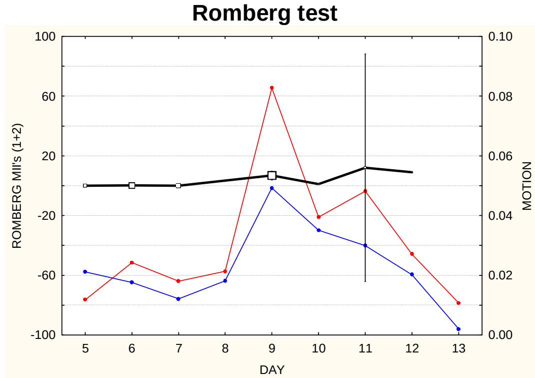
DVA RS3

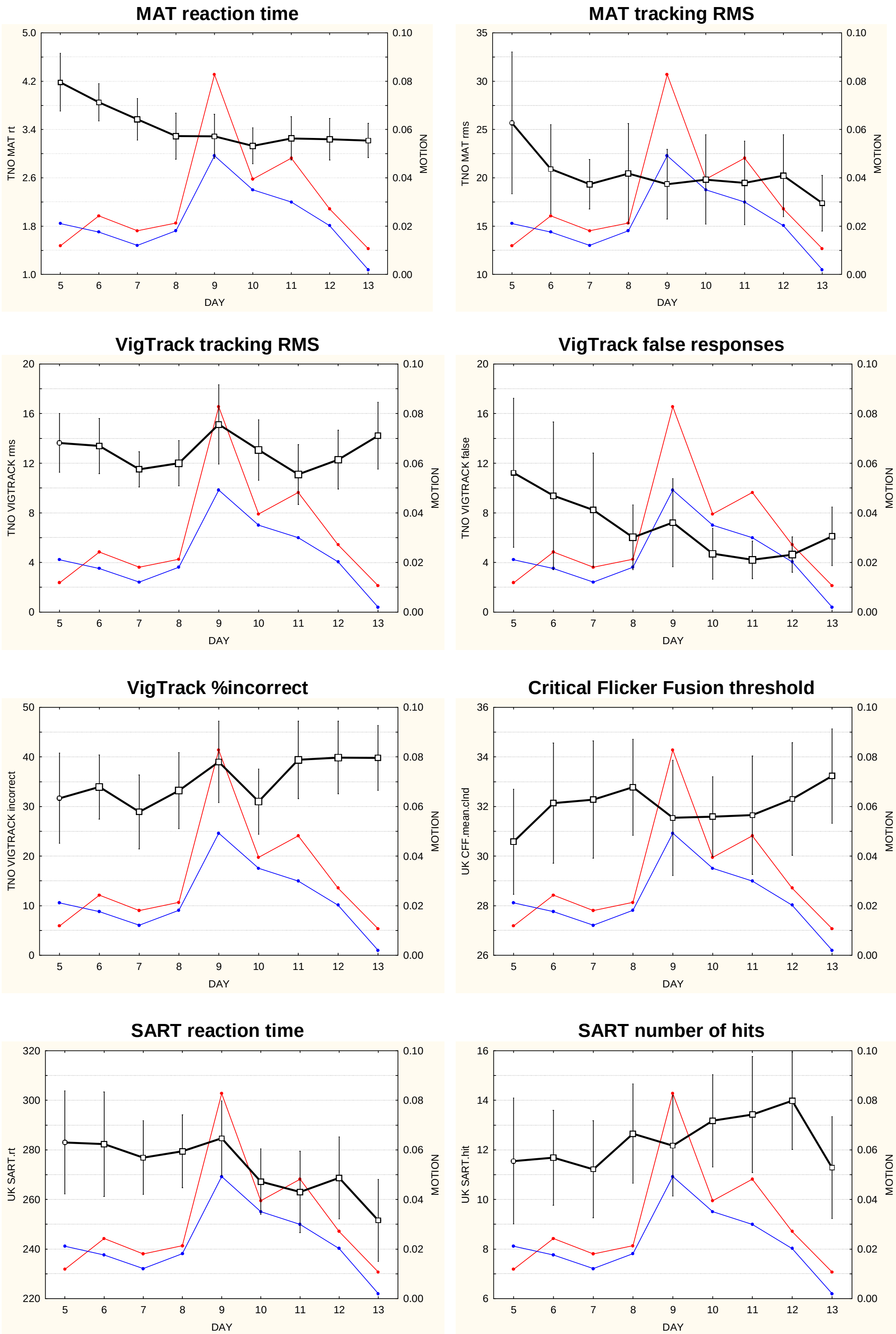


DVA SP-15



Romberg test





CSS= Crew Status Survey, MISC = Miscery Scale, PAQ = Performance Assessment Questionnaire, PANAS = Positive and Negative Affect Schedule, CogFail = Cognitive Failures Questionnaire, TLX = NASA Task Load Index, DVA (stat/RS3/SP15) = Dynamic Visual Acuity (no motion/retinal slip/smooth pursuit), MAT = Multi-attribute Task battery, VigTrack = Vigilance & Tracking test, CFF = Critical Flicker Fusion, SART = Sustained Attention to Response Test