Manuscript title

Measuring workload weak-resilience-signals (WRS) at a rail control post

Occupational Applications

This paper describes an observational study at a rail control post to measure workload weak-resilience-signals (WRS). A weak resilience signal indicates a possible degradation of a system's resilience, which is defined as the ability of a complex socio-technical system to cope with unexpected and unforeseen disruptions. A method, based upon a WRS framework, introduces a new metric, Stretch, to measure the signals. Stretch is a subjective or an objective reaction of the system to an external cluster-event, and is an operationalization of variables in an earlier Stress-Strain model. The Stretch-ratio between the subjective and objective Stretch are used to identify workload WRS. WRSs identified during real-time operation revealed obstacles that influence the resilience state and enabled actions to anticipate and mitigate changes, to maintain the resilience of the system.

Technical Abstract

Background:

Continuous performance improvement of a complex socio-technical system may result in a reduced ability to cope with unexpected and unforeseen disruptions. As with technical and biological systems, these socio-technical systems may become “robust, yet fragile”. Resilience engineering examines the ability of a socio-technical system to reorganize and adapt to the unexpected and unforeseen. However, the resilience doctrine is not yet sufficiently well developed for designing and achieving those goals, and metrics are needed to identify resilience change.

Purpose:

We explored a new approach to identify changes in the resilience of a rail system around the workload boundary, to anticipate those changes during normal operations, and hence to improve the ability to cope with unexpected and unforeseen disruptions.

Method:

We developed a weak-resilience-signal (WRS) framework with a resilience state model for a railway system, resulting in a generic, quantifiable, WRS model. Two workload measurements (i.e., External Cognitive Task Load and Integrated Workload Scale) were combined into a new metric called Stretch. Heart rate variability was used for correlation and validation. An observational study was used to measure workload WRS, through workload quantification, at an operational rail control post.

Results:

A theoretical resilience state model for a railway system was developed, and used to generate a generic quantifiable WRS model. These theoretical models form a WRS framework, which is the basis for a method to measure workload WRS through a new metric called Stretch with three variations: objective Stretch, subjective Stretch, and Stretch-ratio. A component of the subjective Stretch is the Integrated Workload Scale (IWS), for which a real-time tool was developed for measuring and monitoring. Workload WRSs identified at a rail control post triggered analysis to reveal to-be-anticipated obstacles.

Conclusion:

A resilience state model of a rail system can be used to quantify workload WRSs. Stretch-ratio differences represent changes of the workload state used to measure workload WRSs, which aid in revealing obstacles jeopardizing the resilience state.

Keywords: Stretch, weak resilience signal, WRS, resilience, workload, rail operation, rail control post

1. Introduction

The continuous performance improvement of a complex socio-technical system may necessarily result in a more limited ability to cope with unexpected and unforeseen disruptions. Just as found with technical and biological systems, these socio-technical systems may become “robust, yet fragile” (Alderson & Doyle, 2010, p. 839). Resilience engineering investigates, among other aspects, the ability of a socio-technical system to reorganize and adapt to the unexpected and unforeseen (Hollnagel, Woods, & Leveson, 2006). However, the resilience doctrine is not yet sufficiently well developed for designing and achieving these goals (Madni & Jackson, 2009). An important step to account for the resilience of a system is information on its resilience state. The resilience state has been described through theoretical models but so far lacks solid quantification. Woods, Schenk, & Allen (2009) describe some of these models and compare them with each other. The Ball and Cup model (Scheffer, Hosper, Meijer, Moss, & Jeppesen, 1993), for example, is aimed at the system steady state that presents boundaries after which another steady state or system break-down occurs. However, this model does not have the ability to explain potential adaptations that may occur around the boundaries.

In another approach, the Stress-Strain (S-S) model (Woods & Wreathall, 2006) takes its analogy from materials sciences, by mapping the external demand onto the material’s stress and the system behavior onto the material’s strain. The S-S model focuses on behavior near the boundaries explaining system degradation, system restructuring, and system transitions, which are potentials that need to be managed during challenging stress events. Woods, Chan, et al. (2013) extended the Stress-Strain model further to operationalize four cornerstones postulated to be essential to resilience: anticipating, monitoring, responding, and learning (Hollnagel, 2009), and introduced regions for base and extra adaptive capacity. The region for base capacity represents the “normal” functioning of the system to external events. The region for extra adaptive capacity represents the potential for adaptive shortfalls to arise where responses cannot match the demands of challenging events that fall near or beyond the boundary area of the base envelope. These regions explain the behavior of the system beyond the base envelope, however they do not provide a means to measure the properties in the extra adaptive region. Furthermore, the behavior in the extra adaptive region is a hidden capacity to react to unforeseen disturbances. An objective of this paper was thus to develop a method to measure properties in the base capacity region, which signal changes of properties in the extra adaptive region. This objective makes quantification possible, and provides clues that can be analyzed and interpreted by human operators about aspects of the hidden capacity.

We introduce the concept of weak-resilience-signal (WRS), which we will use to quantify changes of the resilience state. We define WRS as signals indicating a possible degradation of the socio-technical system's resilience that can be traced to its original cause. We contrast a weak resilience signal with a strong resilience signal, the latter being a clear signal that the resilience of the system has degraded and which should be considered as an alarm triggering a relevant action. This comparison also emphasizes that a WRS is not an alarm but rather a trigger of interesting information about the system state. A weak signal in this context can be seen as analogous to a human feeling some chest pains during daily activities. When investigating this signal, he may conclude that this is just a spasm or a serious problem with the heart that would only be evident at the time of a large effort.

A weak signal measuring a minor issue during nominal operation may be a crucial factor of failure. Dekker (2011) goes even further, theorizing that the accumulation of an unnoticed set of events is the main cause of the incubation of and surprise at failure. The weak signal can also be explained through the Stress-Strain model (Woods et al., 2014), in which changes occur in the base adaptive capacity such as a change in the Young's modulus slope (Woods & Wreathall, 2006), the linear relation between stress and strain. A slope change in the base region indicates a creeping failure to be exposed at a large stress. Only collecting many detailed weak signals would not necessarily result in a corrective action in response to a specific signal. It may cause fatigue or vigilance (Davis & Parasuraman, 1982), and due to many irrelevant weak signals, which do not need any action, it could cause a "cry wolf" (Breznitz, 1984) effect. Therefore, the WRS needs an extra set of properties to account for the above. First, it needs to be an aggregation of a lower/detailed weak signal set, to lower the number of signals, and second, the aggregation needs to be of interest to the operators to understand the behavior of the system beyond resilience. These are "sending" properties of the WRS. Yet, a "receiving" property of the rail sector is also needed to expand its culture from "working by virtue of many rules and formal agreements" (Top & Steenhuisen, 2009) to an inquisitive one of understanding, tracking, and anticipating the relevant weak resilience signals.

In this paper, we focus on a framework for rail weak-resilience-signal (WRS) modelling and we emphasize one main area - workload - for which we develop a specific method to measure a workload WRS at a rail control post. We verify and validate this method in real operations through an observational study during a reorganization of a rail control post. Our research questions were twofold: 1) How can a weak resilience signal (WRS) be modeled to enable its quantification and be demonstrated in the area of workload in real operations? 2) How can workload WRS be measured and utilized at a rail control post?

The remainder of this paper is structured as follows. In section 2, we develop a framework for rail WRS modelling and describe mathematically its generic quantification. In section 3, we describe a method to measure workload WRS at a rail control post. Section 4 describes the observational study we carried out during two separate weeks at the rail control post. We conclude the paper with the results of the observational study (section 5) and a discussion (section 6).

2. A framework for rail weak-resilience-signal (WRS) modelling

2.1. Theoretical resilience state model for a railway system

A theoretical model describing the resilience state of a railway system is needed to: (1) better understand in which areas weak resilience signals (WRS) are to be sought; and (2) provide a foundation upon which a quantitative model of a WRS can be built. Rasmussen's (1997) safe operating envelope was used as a starting point since it uses three boundaries – performance, economy and workload – to describe the envelope of a generic socio-technical system operating in an economic environment. That model described the various pressures on the Operating State (OS) that may result in crossing one of the borders or readjusting the border to create a new steady state. This readjustment is actually resilience, which is defined by the capacity to adapt to unforeseen events (Hollnagel et al., 2006). In Rasmussen's framework, the performance boundary is directly linked to safety culture pressure, the economic boundary is linked to efficiency pressure and the workload boundary is linked to least effort pressure. In our adaptation of Rasmussen's model, we have introduced some changes to reflect the nature of a railway system. First, we separated performance from safety to reflect their independent nature, while their mutual influence on the operating state is made explicit in the new model by upgrading safety to a boundary entity, which creates safety pressure. Second, we moved the economic boundary backwards, thereby creating efficiency pressure on the performance boundary, which in turn creates a performance pressure. This change is justified by the fact that in rail systems, economic considerations play a more prominent role in the long run than in daily decisions. However, the performance pressure, created by capacity growth and punctuality to deliver the planned schedule, plays a major role in daily considerations. The workload boundary stays intact, reflecting the human importance within a socio-technical rail system, and the result of these changes is shown in Figure 1 (section I).

The above model is considered useful when reasoning about resilience. For example, Cook & Rasmussen (2005) use different areas in the model to explain the stability of a system: unstable, low-risk stable, and high-risk stable. The fact that the boundaries put pressure on the Operating State (OS) is indicated textually with the term "gradient", and grey areas show the OS jump domain that is due to shallow gradients. These gradients are of interest, since they represent the internal pressure on the OS, and may be indirectly measured and can help explain the resilience of the system when the OS is located at any position between the boundaries. When a gradient is steep, it represents system resilience against external perturbations, while shallowness represents brittleness. As described by Woods et al. (2009), who related the work of Walker & Holling (2004) to that of Rasmussen (1997), this gradient can be made explicit by adding a depth dimension to Rasmussen's model as if it were viewed from above in a landscape of valleys. The slope (α) of the valley (see Figure 1 section II) describes the internal force gradient (or Resilience Engineering as in Walking and Holling, 2004) acting on the OS. The vector $\vec{d}$ describes the external perturbations acting on the OS, while $d_P = d \cdot \mathrm{Cos}\alpha_P$ represents the pressure of boundary B_P. This third dimension with the valley slope is important to understand the level of resilience when moving towards one of the boundaries. A shallow slope is analogous to a small hurdle, representing brittleness, to approach the boundary, while a steep slope represents resilience. As an example, Figure 1 section III shows an OS that is moving towards the marginal boundary, a boundary to guard the safety boundary. There are two options to reflect the change of the internal state. When only the capacity of the system is increased and no safety measures are taken, this will result in a brittle state, option a, in which the marginal boundary risks being crossed. However, when measures are taken to also enlarge the safety hurdle, as in option b, it may result in a deeper valley, thereby maintaining the resilience engineered to cope with a higher capacity. This theoretical model will be used in the following subsection to model quantifiable weak resilience signals (WRS) through pressure change acting on the OS near the boundaries.

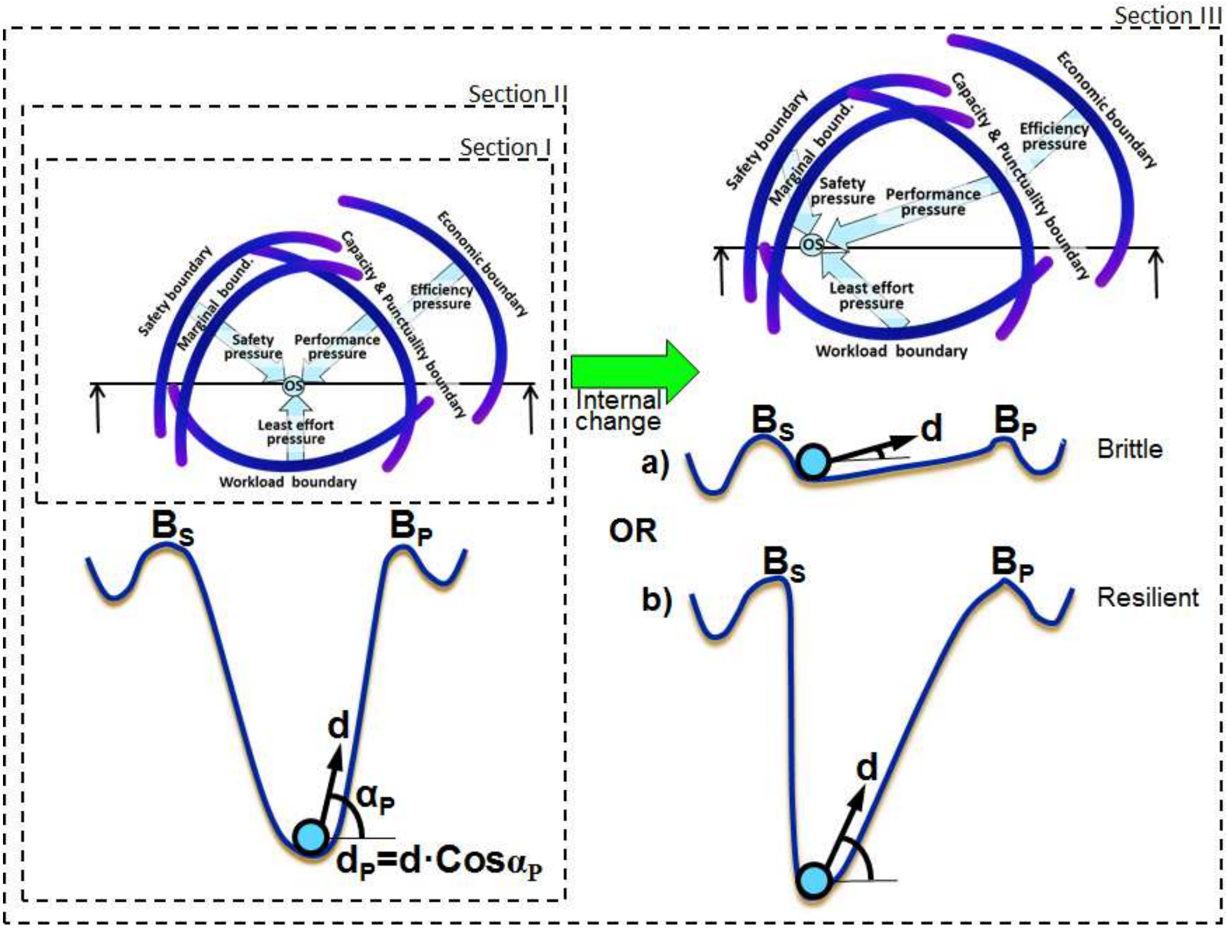

Figure 1 : Resilience state model for a railway system

Section I: Rail-sector boundaries putting pressure on the Operating State (OS)

Section II: Rail-sector boundaries with resilience slope α_P, causing pressure d_P

Section III: OS move caused by internal change, a or b, influencing system resilience

2.2. Generic quantifiable weak-resilience-signal (WRS) model

Assuming an internal pressure α_B on boundary B, caused by a certain phenomenon described through a function f_B of n measurable parameters, P_{iB}, can be expressed mathematically as:

$$\alpha_B = f_B(P_{iB}, i = 1, \dots, n) \quad (1)$$

When assuming small changes, the pressure change $\Delta\alpha_B$ can be estimated by the cumulative weighted changes of the function parameters P_{iB}:

$$\Delta\alpha_B = \sum_{i=1}^{n}(K_{iB} \cdot \Delta P_{iB}), i = 1, \dots, n \quad (2)$$

Or, as the change of two moments in time, t_1 and t_2:

$$\Delta\alpha_B = \sum_{i=1}^{n} K_{iB} P_{iB}(t_1) - \sum_{i=1}^{n} K_{iB} P_{iB}(t_2) \;, i = 1, \dots, n \quad (3)$$

A weak resilience signal WRS_B is created when it is smaller than a $Threshold\text{-}WRS_B$, which is a negative value since by definition a larger α_B represents a growing resilience (as in fig. 1):

$$\mathrm{WRS_B}: \Delta\alpha_B < \mathrm{Threshold\text{-}WRS_B} < 0 \quad (4)$$

where the weights $K_{iB}\,;\, i=1,\dots,n$ and $Threshold\text{-}WRS_B$ are defined by empirical investigation in which K_{iB} is used to set the relative proportion of influence among the parameters on the pressure α_B, and may be set initially to 1. $Threshold\text{-}WRS_B$ is a way to search for a level at which attention is needed for deeper analysis. A possibility to define $Threshold\text{-}WRS_B$ is the added standard deviation of the measurements at t_1 and t_2 to make the difference significant, or it may be set to a value reducing the number of WRS_B's to the most significant ones. It may be possible that instead of a hard threshold, a graphical representation, such as a continuous graph, will be chosen for monitoring by the rail controller. However, the crux of this model is choosing the phenomenon that is described by f_B. As explained in the Introduction, this phenomenon needs to cover many possible WRSs and must be chosen in such a way that it is of interest to the controllers independently of the signals occurring. The following section gives an example of such a phenomenon worked out with respect to the workload boundary. We assume that passing the workload boundary with a certain threshold implies a possible degradation of the system resilience. This is in line with Woods & Patterson (2000), who claimed that unexpected events produce an escalation of cognitive demands. When cognitive workload change is significant and identified, it is a signal that the resilience of the system is reduced, due to the reduction of the spare cognitive capacity, and which may be needed when the unexpected event occurs. There are two period types of passing the boundary. A short period passage is a real-time signal for operations to respond to by an intervention. Passages in a long period indicate a possible structural change to be addressed. With an empirical study we will show the usage of parameter settings and validate the model with the results through observation.

3. A method to measure workload WRS at a rail control post

Workload measurement methods have been studied extensively (Gao, Wang, Song, Li, & Dong, 2013; Pickup, Wilson, Nichols, & Smith, 2005; Pretorius & Cilliers, 2007; Veltman & Gaillard, 1993). Different factors influence mental workload, such as time, mental tasks, physical tasks, and stress (Xie & Salvendy, 2000), which makes it clear that one measurement type will not cover all aspects. Veltman & Gaillard (1996) reason that the measurement of mental workload needs performance, subjective, and physiological data for a complete understanding of workload. We suggest using three different measurements: 1) external cognitive task load (XTL), 2) subjective workload, and 3) heart rate variability to identify arousal created by workload.

To compose the XTL, we built upon Neerincx' (2003) model of cognitive task load (CTL) in three dimensions: task complexity, task duration, and task switching. The XTL is defined specifically to the rail control situation and to parameters that are available in real-time. The real-time aspect, of all the measurement components, provides possibilities to set up experiments to close the loop throughout operations. Rail signalers' task execution can be divided into four main activities (see Figure 2), which are measurable within the system: 1) monitoring (Mon), 2) plan mutations (Plan), 3) manual actions (Man), and 4) communication (Com). Monitoring is keeping track of trains and infrastructure through observation of system displays. Plan mutations refer to activities concerning the logistic plan, which is the basis of train movements on the infrastructure as agreed among all parties and used by system automation. Manual actions are activities performed directly on the infrastructure, like setting a switch, instead of system automation according to the plan. Telephone calls, with external parties, are the main communication task. We assumed that monitoring is in proportion with automated activities executed by the system. This assumption refers to imposed task load, while in reality the rail controller can actually ignore the monitoring task. Monitoring can thus be measured by counting all the automated activities. These activities were counted in 5 minute base-slots, used throughout all the types of measurement for ease of comparison. We normalized these counts by dividing them by the maximum count (Mon_{max}) occurring throughout a test period, causing the measurement to be normalized between 0 and 1. This same idea was applied to normalizing the plan mutations and the manual actions. Each of these were counted within the 5 minute base-slot and divided by the maximum count, $Plan_{max}$ and Man_{max} respectively, throughout a test period. The communication normalization was done differently. Communication was defined by the percentage of verbal exchanges over the phone, which is measureable, during the 5 minute base-slot. A rail signaler talking the whole 5 minutes, results in a 100% communication value.

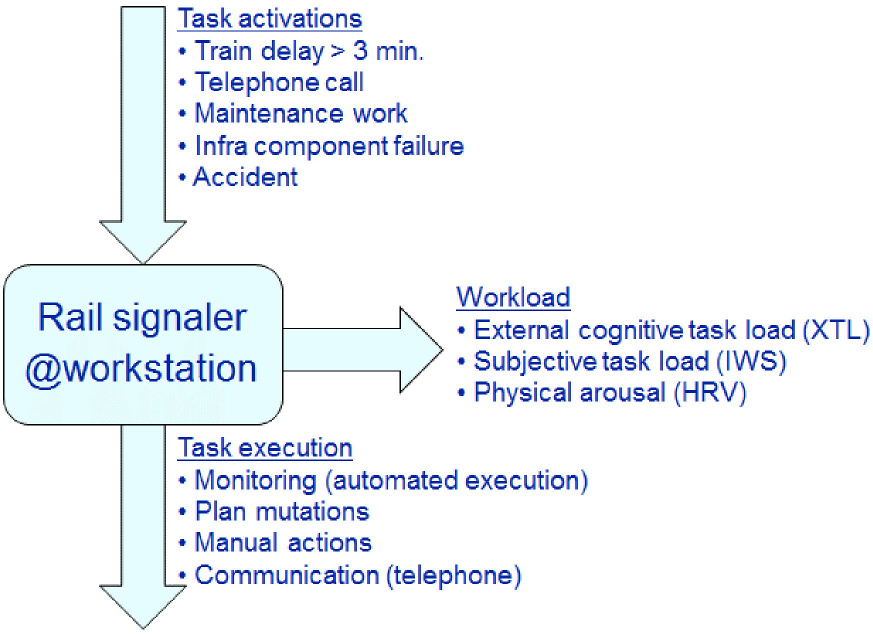

Figure 2 Task flow of a rail signaler at their workstation

The combination of these four normalized activities refers to task complexity as stated by Neerincx (2003). However, Neerincx used the Skill-Rule-Knowledge (SRK) model (Rasmussen, 1997) to express task complexity by rating each task on its SRK cognition load level. Since we do not know the cognitive relationship among the tasks, we multiplied each with their relative task complexity constant (K_{mon}, K_{plan}, K_{man}, and K_{com}), and tracked their identity throughout the whole process. In addition to these activities, task switching and task duration are two extra dimensions amplifying the workload. To estimate the number of task switches, we examined the task activations and counted them in each time slot as long as they were activated, to reflect the task duration. In Figure 2, we list the task activations imposed on a particular workstation. These activations resulted in the activities discussed above and resulted in workload we measured by XTL, IWS and HRV.

Since the analysis is based upon log-data, we can search for the maximum number of activations occurring in the 5 minute base-slots. We divided the number of activations, occurring in the 5 minute base-slot, by the maximum activations occurring throughout the test period to achieve a normalized switching factor between 0 and 1. Task switching and duration are a cognitive add-on to the activity load. With the same activity load, 0 to n parallel task switches can occur, behaving like a cognitive amplifier to the activity load. We added one to the normalized switching factor to act as a cognitive amplifier by becoming a growth multiplier of the activity load. Graphically, the multiplication will show jumps attracting the attention needed for interpretation. Thus, the switching factor becomes:

$$K_{switch} = \frac{number\ of\ activations\ in\ 5\min base\text{-}slot}{maximum\ number\ of\ activations\ in\ 5\min base\text{-}slot} + 1 \quad (5)$$

We calculated the task complexity load with the sum of the four normalized tasks, each multiplied with their relative task complexity constants: K_{mon}, K_{plan}, K_{man}, and K_{com}. These constants are initially set to 1 and may be adjusted proportionally during empirical investigation, but keeping their sum to the initial value of 4 and only changing their interrelationship. We multiplied the task switching factor with the task complexity load to achieve a combined XTL number. This approach creates a number between 0 and 8 to be used as an overall graphical indication on the XTL magnitude and change. Maximum load due to task execution is 4 X 1 = 4, multiplied by a maximum switching factor, 2 X 4 = 8. However, it is important to present all the components and their relationships separately to understand the situation.

The XTL calculations can be performed for workstation WS with its subscripted WS values using:

$$XTL_{WS} = K_{switch\text{-}WS}\left(K_{mon}\frac{Mon_{WS}}{Mon_{max}} + K_{plan}\frac{Plan_{WS}}{Plan_{max}} + K_{man}\frac{Man_{WS}}{Man_{max}} + K_{com}\cdot Com_{WS}\right) (6)$$

Subjective load measurement can be divided into two categories: multidimensional and unidimensional scales. Multidimensional scales, such as the NASA-TLX (Hart & Staveland, 1988), explicitly represent the dimensions of workload and allow ratings to be obtained from each dimension. Unidimensional scales (Muckler & Seven, 1992) represent the concept of workload as one continuum. Hendy, Hamilton, & Landry (1993) claim that a univariate rating is expected to provide a measure that is at least as sensitive to manipulations of task demand as a derived estimate from multivariate data. In addition, a unidimensional scale is easier to use and in our case easier to automate for real-time purposes. Pickup, Wilson, Norris, Mitchell, & Morrisroe (2005) have developed a unidimensional scale specifically for rail signalers, called the Integrated Workload Scale (IWS). They have automated the IWS tool for usage of the trial facilitator for a few-hour period. Our aim was to let the rail signaler assess and enter their own rating for 24 hours a day. We developed a Java-tool that can run within the operational system to be seen as part of their routine work. The rail signaler RS_i, working at work station WS_j, was alerted every 5 minutes by a peripheral blinking rectangle, to rate their subjective workload. They were presented with a 9 scale figure containing the following text (from the original Dutch) (see Figure 3): (1) Not demanding; (2) Minimal effort; (3) Some spare time; (4) Moderate effort; (5) Moderate pressure; (6) Very busy; (7) Extreme effort; (8) Struggling to keep up; and (9) Work too demanding. The rail signaler had the option to add a comment to their rating and received a graphic overview of their scoring.

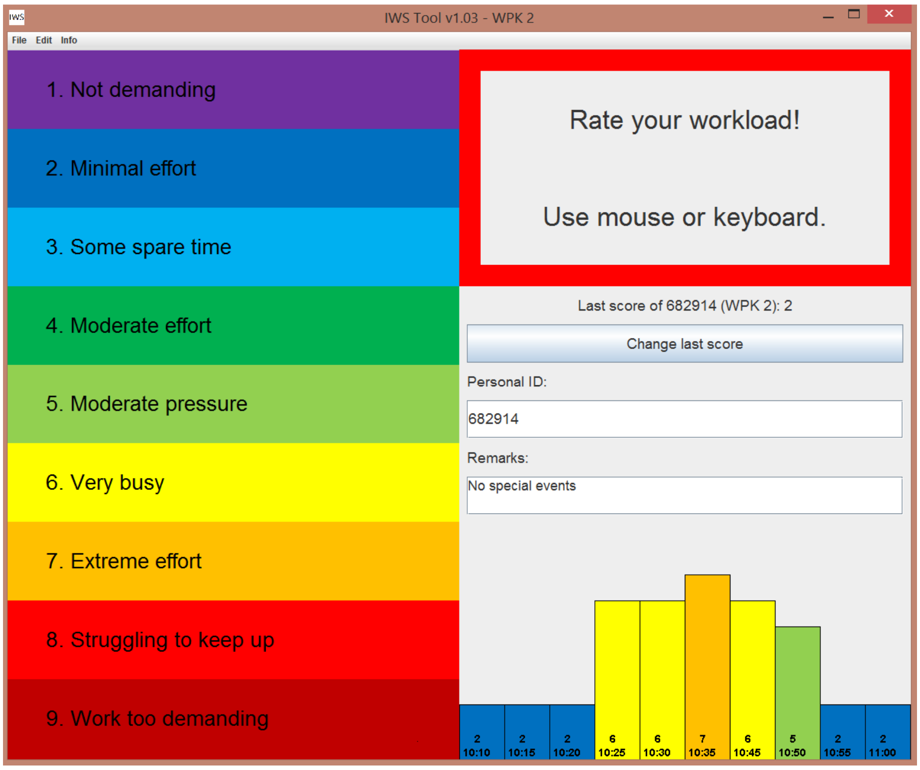

Figure 3 IWS application screenshot translated from Dutch (upper-right red rectangle blinked to draw attention)

The extensively-researched heart rate variability (HRV) was used to identify physiological arousal due to workload change (Billman, 2011; Goedhart, van der Sluis, Houtveen, Willemsen, & de Geus, 2007; Hoover, Singh, Fishel-Brown, & Muth, 2012; Jorna, 1992; Malik, 1996; Togo & Takahashi, 2009). The HRV was mainly used to cross-check the subjective measurement, and will be lower at a higher workload and identify IWS ratings that are given due to other reasons than a higher workload. HRV was measured with a commercial device (Zephyr HxM BT) that was positioned on a chest strap and transferred data to a laptop near each workstation. A signaler wore the device at the start of their work. The device sends continuous strings with recorded R-R intervals in msec. HRV can be calculated in various ways, roughly divided into time domain and frequency domain methods (Malik, 1996). We used the most common occupational health method (Togo & Takahashi, 2009), SDNN, the standard deviation (SD) of all normal-to-normal

(NN) intervals, from the time domain. We calculated the measures in the same 5 minute base-slot used for the calculations of XTL and IWS.

The three measurements described above, XTL, IWS and HRV, are all measured in 5 minute slots. This timeslot enables comparison of the measurements in a timeline, as Pickup, Wilson, Norris, et al. (2005) did to validate IWS. We did this for validation of IWS through the HRV, but it is not sufficient for the analysis of events taking much longer than 5 minutes, which is the case in the rail environment. Serious events take more than half an hour, as can be seen in the results section. To compare the XTL and IWS, they should be referenced to a time frame of events, clustered from and to a steady state. The steady state of a rail control post is the state when the train activities are occurring as planned, without any intervention. In order to relate the IWS and XTL measurements, a new metric was introduced – Stretch (see Figure 4).

Figure 4 Defining Objective and Subjective Stretch from XTL & IWS over time

A Stretch is the cumulative workload effort during a period initially defined by IWS rising from a baseline until it returns to the baseline. The IWS-baseline is defined as the steady state IWS rating before and after a disruption. However, the activity in the system may have started earlier and ended later. Therefore, the starting moment of a Stretch is adjusted to the first XTL-minimum moment before the IWS rising. Similarly, the ending moment of a Stretch is adjusted to the first XTL-minimum moment after the IWS return. In other words, a Stretch is the reaction to an external cluster-event. We use the term cluster-event, since more than one event may occur during a stretch. An Objective Stretch is the name of the area under XTL, since it is objectively measured. We name the area under IWS a Subjective Stretch, due to its subjective IWS rating. The ratio of Subjective Stretch and Objective Stretch is called Stretch-ratio, which is used to identify a workload WRS. These terms are better related, than the measurements, to the Stress-Strain (S-S) model (Woods et al., 2014; Woods & Wreathall, 2006) and the resilience state model, developed in the previous section. The objective Stretch is related to the Stress axis of the S-S model. Stress is the theoretical concept of the demand of the system through Challenge events. The objective Stretch is the operationalization of the Stress concept through measuring the factual reaction of the system. The subjective Stretch is the human perception of the system Strain. The Stretch-ratio relates to α_B of the workload boundary ($\alpha_{workload\text{-}boundary}$), the internal pressure on the workload boundary, of the resilience state model. When a growing change of a Stretch-ratio is identified, larger than a threshold, and the Stretch values are larger than a pre-defined value, a weak resilience signal (WRS) is generated. When comparing two periods, the accumulated standard deviation (SD) of the Stretch-ratio in each period, can function as the threshold, indicating a significant change. However, such a principle needs to be validated in empirical testing. A larger Stretch-ratio during a given period, compared to a baseline period, indicates more subjective workload in response to similar external events. The Objective Stretch is used to identify an absolute workload growth, throughout a specific period like a day or a workweek.

4. Observational study during rail operations

To validate and verify the applicability of the method to measure workload WRS at a rail control post, we applied it throughout the restructuring tryout of a control post to improve its work efficiency. In our specific case, the control post was restructuring only one group around a corridor for a test period of half a year, by: 1) setting focus on a corridor by seating the corridor team together; 2) splitting-up the responsibility of a rail controller's tasks to planning and safety related activities by adding a planner to the team; 3) enforcing standardization through position rotation; and 4) growing their expertise level through training as part of the position rotation. This efficiency step can, however, affect the post's spare, and sometimes hidden, adaptive capacity needed when an unexpected disruption occurs. In addition, this efficiency step can also affect the organization's ability to manage this capacity. As improved work efficiency may conflict with an organization's resilience due to common resource demands, methods are needed to identify this potential conflict, which can be shown by a WRS. A rail control post is responsible for a large area containing railway stations, controlled by rail signalers managing the traffic on the rail infrastructure. The post we studied is active 24/7 with 10 to 20 rail professionals. A rail control post is an example of a socio-technical system due to the critical human-system interaction.

The generic setting is a rail control post with m_{Post} workstations and n_{Post} rail signalers evaluating a new organizational form to increase their performance. Each workstation, WSj, is allocated to a set of railway stations and operated by one rail signaler, RSi, who is responsible for all the workstations' aspects. These aspects are roughly divided into logistics and safety, and the workstations are split into two groups. The first group, G_T, is the target group that will reorganize, as described above, to improve its performance. The second group, G_R, is the reference group that will not reorganize throughout the testing period. All the n_{Post} rail signalers of the control post may be allocated to each of the groups and to each of its workstations. In group G_T there are m_T

workstations, and in group G_R there are m_R workstations. In addition, there is a calamity workstation, WS_{cal}, which is added to give support to the workstation being at the core of a calamity. The calamity workstation, which is not related to the reorganization, can be added to each group, G_T or G_R. The setting is depicted in Figure 5:

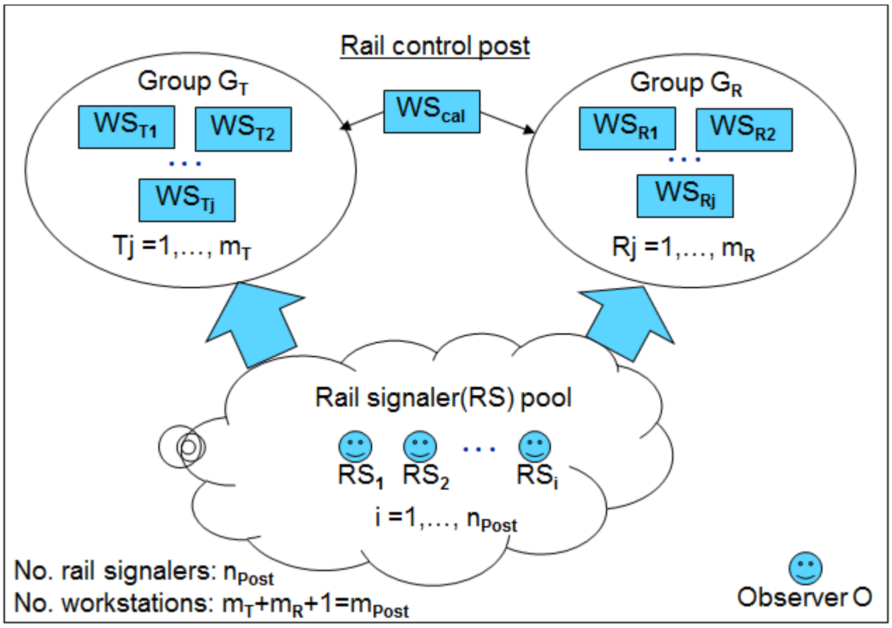

Figure 5 Rail control-post setting with observer O

In our case, we carried out structured observations at a Dutch rail post with 44 participating rail signalers (n_{Post}=44), during two periods of one working week (Monday until Friday). The age of the participants ranged between 23 and 64 years, with a mean of 43.6 years, and the population contained 79.5% males. All of them rated their subjective workload with the IWS tool, though 39% consented to wearing a heart rate sensor during their work. The work experience varied between 0 and 37 years, with mean 17.6. The first measurement period was immediately before the reorganization of the target group, and the second measurement period was two months afterward. In the first period, measurements were recorded in two shifts from 7AM until 9PM with the IWS tool on a separate laptop near each workstation. During the second period, the measurements were recorded continuously, 24 hrs a day, with the IWS tool integrated within the operational system (see Figure 6). Initially, there were three workstations at the target and reference group ($m_T = m_R = 3$). After the reorganization, one workstation was added to the target group (m_T=4), for planning activities of the corridor. The protocol guiding the observations was approved by the ethical committee of the University of Twente, except for its request to obtain written consent by participants, which was replaced by oral consent by each participant at the request of Post management.

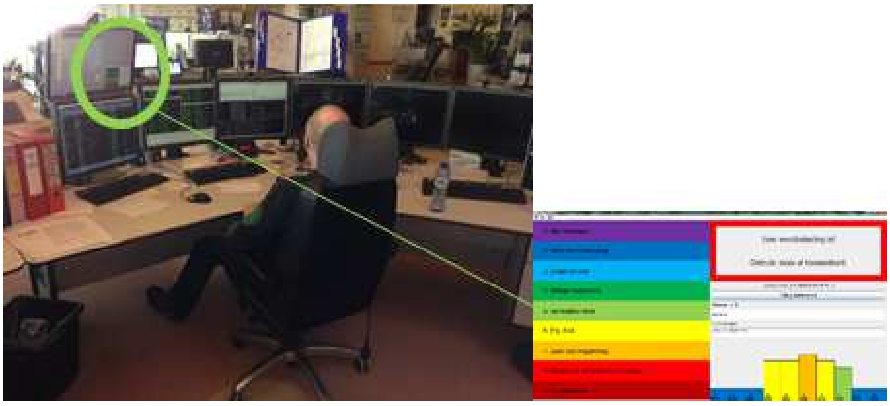

Figure 6 Integration of the IWS tool within operations

5. Results

The quantitative results of the Stretch measurements before and during the reorganization are summarized in Table 1. Before the reorganization, the mean Stretch-ratio of the target group was 5.30 [IWS/XTL] with a standard deviation (SD) of 2.61. The mean Stretch-ratio of the reference group was 5.82 [IWS/XTL] with and SD of 2.55. Since the standard deviations were large, and the means were similar, we may conclude that the Stretch-ratio of both groups were in the same order of magnitude, indicating the similarity of work in both groups. The duration of the Stretch varied substantially. This can be seen clearly by comparing the Stretch with the Stretch divided by its duration (Table 1: SS/Dt and OS/Dt), the latter representing the mean workload throughout the

Stretch. For example, the subjective Stretch of both groups before the reorganization was 21.13 [IWS x min] with a SD of 15.60, whereas subjective Stretch divided by its duration was 3.09 [IWS] with a SD of 0.80.

During the reorganization, a planner was added to the target group. The mean Stretch-ratio of the planner was 11.83 [IWS/XTL] with a SD of 5.54. The reason the planner had a much larger Stretch-ratio than the normal rail signaler is because their XTL was much lower since that individual does less work. The planner had no monitoring task, no manual action task, and fewer phone calls since they do not communicate with the train drivers. In contrast, the planner rated IWS similar to colleagues, causing the Stretch-ratio to become larger. This could be solved by adjusting the relative task complexity constants, which were initially set to 1, and give more relative weight to plan activities. However, more empirical research is needed in this area, causing the existing Stretch-ratio to be valuable for comparison of similar tasks, but not yet suitable to compare between different tasks. For that reason, we have added to the summary table entries where the planner is excluded (Table 1: Target-planner and All-planner). The mean Stretch-ratio of the Target group during the reorganization without the planner was 6.17 [IWS/XTL] with a SD of 2.81. The mean Stretch-ratio of the Reference group during the reorganization was 6.36 [IWS/XTL] with a SD of 1.80. The Stretch-ratio for both groups remained similar, but increased in the measurement week during the reorganization. The reason for the increase can be found in the figures of the objective Stretch, which are lower during the reorganization than before. Deeper investigation shows that fewer phone calls are the cause for the objective Stretch reduction. In summary, in the measurement week during the reorganization, no evidence was found that the reorganization significantly influenced the workload adaptive capacity needed for system resilience.

Table 1 Stretch measurements over one week, both before and during reorganization (cells that are not relevant for the line of argumentation are not filled in)

	Group	## Stretch	Stretch-ratio		Subjective Stretch (SS)				Objective Stretch (OS)			
			Mean [IWS/XTL]	SD	Mean [IWS x min]	SD	Mean(SS/Dt) [IWS]	SD(SS/Dt)	Mean [XTL x min]	SD	Mean(OS/Dt) [XTL]	SD(OS/Dt)
Before reorganization	Target	35	5.30	2.61								
	Reference	107	5.82	2.55								
	All (T & R)	142	5.69	2.57	21.13	15.60	3.09	0.80	4.28	3.58	0.62	0.26
During reorganization	Target	170	7.37	4.24								
	Target - planner	134	6.17	2.81								
	Reference	134	6.36	1.80								
	All (T & R)	304	6.92	3.42	21.17	24.30	2.75	0.59	3.49	3.82	0.47	0.21
	All - planner	268	6.26	2.36	21.18	25.59	2.75	0.59	3.70	4.00	0.50	0.20

Another representation of the measurement results is a plot of the objective Stretch versus the subjective Stretch before and during the reorganization (Figure 7). The two Stretch types are highly correlated, with r (Pearson) = 0.90 before reorganization and 0.88 during reorganization. Most Stretches in both weeks are small. We have drawn a threshold line with a Stretch-ratio of 9 [IWS/XTL], since the mean Stretch-ratio in the first week was 5.69 [IWS/XTL] with a SD of 2.57 (Table 1). A first threshold line would be the rounded sum of the means with one standard deviation above (i.e., 6+3). It is the threshold, as explained in the previous section, which needs to be set empirically to optimize the number of WRSs to handle. With this threshold, two weak resilience signals during the reorganization need further investigation (WRS-1 and WRS-2, labeled “1” and “2” in Figure 7).

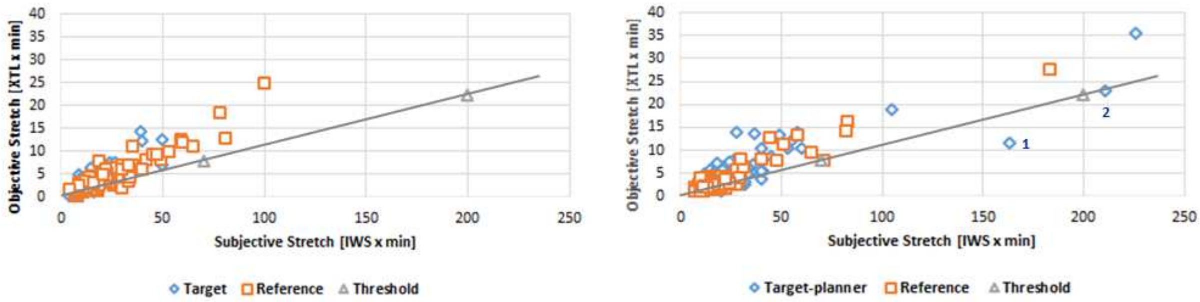

Figure 7 Objective versus Subjective Stretch in one week, both before (left) and during (right) reorganization

WRS-1 has a Stretch-ratio of 14.11 [IWS/XTL] ,with a subjective Stretch of 163 [IWS x min] and an objective Stretch of 11.55 [XTL x min], which are numbers for comparison of Stretches in the given setting The WRS occurred on the first measurement day at Workstation-3, at 7:10 AM with a duration of 195 minutes, while performing shunting of rail material as the main activity. The rail-controller subjectively rated their mean workload during this Stretch as “moderate effort” (4.17), which is higher than the mean IWS-rating (“some spare time” = 2.75) of the whole group during the test week. The higher IWS-rating, combined with the long duration of shunting activities, triggers further investigation or at least causes the tracking of the shunting for a longer period to understand the phenomena and take appropriate actions. This is an example of a WRS causing the identification of an obstacle, which could become a main cause of incubation and surprise at failure, as stated by Dekker (2011).

WRS-2 has a Stretch-ratio of 9.16 [IWS/XTL] with a subjective Stretch of 211 [IWS x min] and an objective Stretch of 23.03 [XTL x min]. The WRS occurred on the second measurement day, at Workstation-3 at 8:40 AM, with a duration of 350 minutes, which again performing mainly shunting of rail material. The rail-controller subjectively rated their mean workload during this Stretch as “some spare time” (3.01). Although the mean IWS-rating was lower than that of WRS-1, the duration was much longer. This recurring shunting activity emphasizes the importance of investigating the reasoning for the long periods. Such an investigation is an example of actions taken as a result of a WRS.

The above results and reasoning give some confidence in the validity of the data, since they correlate with the observations in both weeks. In both weeks, no special events occurred, and both groups were able to cope with daily disturbances. The shunting issues of the WRSs were recorded as well, and were caused by the three train companies, who had had extensive unplanned rail materiel to be treated manually by the rail signalers. The reorganization did not have a visible effect on the average disturbances. To further validate the data, we analyzed the work distribution, based upon the XTL components, and verified it as well with the observations. In Figure 8, we have shown the work distribution of the Target group before and during the reorganization. It is clear from the graphs that the extra workstation (WS 4) does most of the planning, communicates less than the other workstations, and does not perform manual or monitoring activities. These figures are consistent with the observations, where all planning activities that were more than 10 minutes ahead were allocated to WS 4.

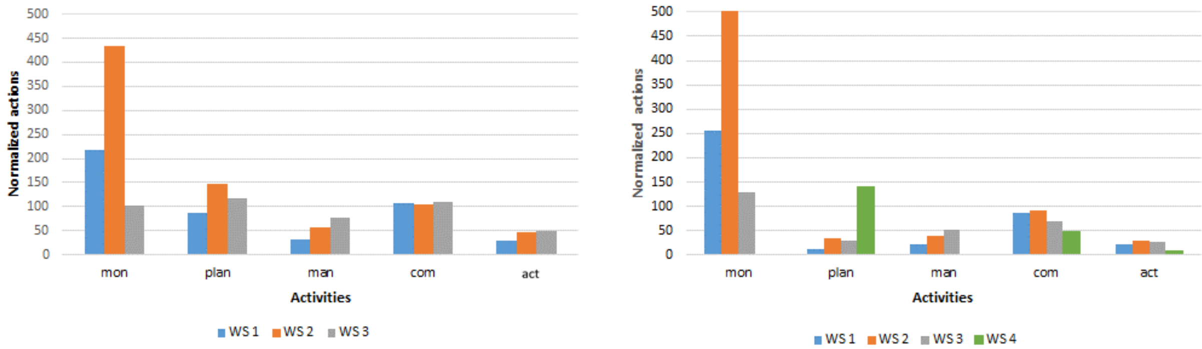

Figure 8 Work distribution of target group before (left) and during (right) reorganization

(mon = monitoring ; plan = plan mutation ; man = manual action ; com = communication ; act = activations)

In addition, HRV was been correlated to the objective Stretch. The following algorithm has been applied to identify a lowering HRV during a Stretch. First, the highest value of the HRV on the boundaries of its Stretch was marked. Then, this value was multiplied by the Stretch-duration and the integral under the HRV throughout the stretch was subtracted. A negative value was assumed to confirm the subjective Stretch by the physiological response. This algorithm was applied to the data available in the week before the reorganization. A lower HRV was recorded during 83% of the subjective Stretches, which is in line with the literature (Togo & Takahashi, 2009). This finding provides an additional means to evaluate Stretches passing the Threshold boundaries.

6. Discussion

There is a need during real-time operations to quantify the system resilience state. Quantification is challenging because, on the one hand, socio-technical systems are complex and non-linear (Doyle & Csete, 2011), while on the other hand resilience is about hidden capacity that is measured only during the response to such disruptions (Woods et al., 2013). Woods et al. (2013) have made some progress in the quantification of resilience parameters by looking at the system boundaries. This paper focused on the area of daily operations, seeking quantifiable weak resilience signals (WRS) around the workload. The aim of this research was to show how a WRS can be modeled, to enable its quantification and to demonstrate this in the area of workload in real-time train operations. In addition, we wanted to determine whether, and how, we can measure workload WRS at a rail control post and demonstrate how it can be utilized.

A WRS framework was developed and used to concretize a workload WRS at a rail control post, specifically for the work of a rail signaler. The modelling was built up from specific types of workload measurements adjusted to the rail context, resulting in three measurements: 1) eXternal Task Load (XTL), 2) Integrated Workload Scale (IWS), and 3) Heart Rate Variability (HRV). The first two measurement results were merged into a new metric, Stretch, describing the efforts during clusters of events occurring at the control Post. HRV measurement was used for validation. The two variations, objective and subjective Stretch, are an operationalization of Woods' Stress-Strain (S-S) model variables (Woods, et al., 2013; Woods & Wreathall, 2006). An objective Stretch is related to the stress on the system and the subjective Stretch is the human response perception related to strain. Stretch-ratio is the relation between both Stretches and relates to the slope of the S-S line. Stretch seemed to describe well the variations of the same task set. However, more research is needed to tune the multiplying constants of the sub-tasks, initially being set to 1 here, to compare with other task sets. For comparison of the groups here, we have excluded the planner, who had a consistently larger Stretch-ratio than the others.

Overall, the Stretch gave a clear picture of the events occurring at the control Post and created two Workload WRSs. These were analyzed and triggered further analysis of the shunting activities engaged in at workstation 3, and which is a concrete example of anticipation driven by a WRS. Beyond this finding, there was no indication of a resilience reduction caused by the reorganization. A longer period, with significant disruptions, is needed to understand the impact of the reorganization on the workload resilience border and resilience as a whole. This longer testing period can also contribute to validation of the workload WRS, since more WRSs will occur that can be analyzed and reveal other obstacles influencing the resilience state. In the current testing, we have validated components of the Stretch against observations.

In summary, the Stretch, which is based upon the WRS theoretical and quantification model, offers the ability to quantify a workload WRS. Such WRSs provide new means to measure the, sometimes creeping, resilience changes. When analyzed during operations, it creates awareness of obstacles that can become a (main) cause of incubation and surprise at failure. This awareness stimulates the anticipation to take actions in the period before the unexpected and unforeseen external event occurs. In such a way, the hidden extra adaptive capacity is maintained and can be utilized through the ability of managing this capacity. This will improve the performance of the controllers. A future research step is to measure for longer periods and extend the specific WRS modeling to the other two boundaries, Safety and Capacity. WRS coverage, the identified percentage of obstacles compromising the resilience state, will be investigated as well. Our aim is eventually to test and validate the contribution of the total WRS concept to managing the resilience of the socio-technical rail system.

Acknowledgement

We are grateful for the hospitality of the ProRail control post at Zwolle, who gave us all the freedom for this research and were willing to use our experimental tooling. We thank Jaldert van der Werf for his development of the IWS and analysis software tooling, and his contribution to the observational study. We appreciate the guidance by Alfons Schaafsma. This research was conducted within the RAILROAD project and was supported by ProRail and the Netherlands organization for scientific research (NWO) (under grant 438-12-306).

References

Alderson, D. L., & Doyle, J. C. (2010). Contrasting Views of Complexity and Their Implications For Network-Centric Infrastructures. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *40*(4), 839–852. doi:10.1109/TSMCA.2010.2048027

Billman, G. E. (2011). Heart rate variability - a historical perspective. *Frontiers in Physiology*, *2*, 86. doi:10.3389/fphys.2011.00086

Breznitz, S. (1984). *Cry wolf : the psychology of false alarms*. Hillsdale N.J.: Lawrence Erlbaum Associates.

Cook, R., & Rasmussen, J. (2005). “Going solid”: a model of system dynamics and consequences for patient safety. *Quality & Safety in Health Care*, *14*(2), 130–134. doi:10.1136/qshc.2003.009530

Davis, D. R., & Parasuraman, R. (1982). *The psychology of vigilance*. New York: Academic Press.

Dekker, S. (2011). *Drift into failure - from hunting broken components to understanding complex systems*. Farnham, Surrey: Ashgate Publishing Limited.

Doyle, J. C., & Csete, M. (2011). Architecture, constraints, and behavior. *Journal of the National Academy of Sciences*, *vol. 108*(suppl. 3), 15624–15630.

Gao, Q., Wang, Y., Song, F., Li, Z., & Dong, X. (2013). Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics*, *56:7*, 1070–1085. doi:10.1080/00140139.2013.790483

Goedhart, A. D., van der Sluis, S., Houtveen, J. H., Willemsen, G., & de Geus, E. J. C. (2007). Comparison of time and frequency domain measures of RSA in ambulatory recordings. *Psychophysiology*, *44*(2), 203–215. doi:10.1111/j.1469-8986.2006.00490.x

Hart, S. S. G. S., & Staveland, L. E. LE. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload*, 139–178. Retrieved from http://humanfactors.arc.nasa.gov/groups/TLX/downloads/NASA-TLXChapter.pdf

Hendy, K. C., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: When is one scale better then many? *Human Factors*, *35*(4), 579–601.

Hollnagel, E. (2009). The four cornerstones of resilience engineering. In C. P. Nemeth, E. Hollnagel, & S. Dekker (Eds.), *Resilience Engineering Perspectives. Volume 2: Preparation and restoration* (pp. 117–134). Surrey: Ashgate Publishing Limited.

Hollnagel, E., Woods, D. D., & Leveson, N. (Eds.). (2006). *Resilience engineering: concepts and percepts*. Hampshire: Ashgate Publishing Limited.

Hoover, A., Singh, A., Fishel-Brown, S., & Muth, E. (2012). Real-time detection of workload changes using heart rate variability. *Biomedical Signal Processing and Control*, *7*(4), 333–341. doi:10.1016/j.bspc.2011.07.004

Jorna, P. G. A. M. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological Psychology*, *34*(2), 237–257.

Madni, a. M., & Jackson, S. (2009). Towards a Conceptual Framework for Resilience Engineering. *IEEE Systems Journal*, *3*(2), 181–191. doi:10.1109/JSYST.2009.2017397

Malik, M. (1996). Heart Rate Variability. *Annals of Noninvasive Electrocardiology*, *1*(2), 151–181. doi:10.1111/j.1542-474X.1996.tb00275.x

Muckler, F. A., & Seven, S. A. (1992). Selecting performance measures: “Objective” versus “subjective” measurement. *Human Factors*, *34*(4), 441–455.

Neerincx, M. A. (2003). Cognitive task load analysis: allocating tasks and designing support. In E. Hollnagel (Ed.), *Handbook of cognitive task design* (Vol. 2003, pp. 283–305). Mahwah, NJ: Lawrence Erlbaum Associates.

Pickup, L., Wilson, J. R., Nichols, S., & Smith, S. (2005). A conceptual framework of mental workload and the development of a self-supporting integrated workoad scale for railway signallers. In J. Wilson, B. J. Norris, T. Clarke, & A. Mills (Eds.), *Rail human factors* (pp. 319–329). Surrey: Ashgate.

Pickup, L., Wilson, J. R., Norris, B. J., Mitchell, L., & Morrisroe, G. (2005). The Integrated Workload Scale (IWS): a new self-report tool to assess railway signaller workload. *Applied Ergonomics*, *36*(6), 681–693. doi:10.1016/j.apergo.2005.05.004

Pretorius, A., & Cilliers, P. J. (2007). Development of a mental workload index: a systems approach. *Ergonomics*, *50*(9), 1503–15. doi:10.1080/00140130701379055

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety Science*, *27*(2/3), 183–213.

Scheffer, M., Hosper, S. H., Meijer, M. L., Moss, B., & Jeppesen, E. (1993). Alternative equilibria in shallow lakes. *Trends in Ecology & Evolution*, *8*(8), 275–9. doi:10.1016/0169-5347(93)90254-M

Togo, F., & Takahashi, M. (2009). Heart rate variability in occupational health -a systematic review. *Industrial Health*, *47*(6), 589–602.

Top, J. Van Den, & Steenhuisen, B. (2009). Understanding ambiguously structured rail traffic control practices. *International Journal of Technology, Policy and Management*, *9*(2), 148–161.

Veltman, J. A., & Gaillard, A. (1993). Indices of mental workload in a complex task environment. *Neuropsychobiology*, *28*, 72–75.

Veltman, J. A., & Gaillard, A. W. K. (1996). Pilot workload evaluated with subjective and physiological measures. In K. Brookhuis, C. Weikert, J. Moraal, & D. Waard de (Eds.), *Human factors and ergonomics society* (pp. 107–128). Haren: University of Groningen.

Walker, B., & Holling, C. (2004). Resilience, adaptability and transformability in social-ecological systems. *Ecology and Society*, *9*(2).

Woods, D. D., Chan, Y. J., & Wreathall, J. (2014). The Stress-Strain Model of Resilience Operationalizes the Four Cornerstones of Resilience Engineering. In *5th Resilience Engineering Symposium in press*.

Woods, D. D., & Patterson, E. S. (2000). How Unexpected Events Produce an Escalation of Cognitive and Coordinative Demands. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue*. Hillsdale N.J.: Lawrence Erlbaum Associates.

Woods, D. D., Schenk, J., & Allen, T. (2009). An initial comparison of selected models of system resilience. In *Resilience engineering perspectives* (pp. 73–94). Surrey: Ashgate Publishing Limited.

Woods, D. D., & Wreathall, J. (2006). Stress-Strain Plots as a Basis for Assessing System Resilience. In E. Hollnagel, C. Nemeth, & S. Dekker (Eds.), *Resilience engineering perspectives, volume 1: Remaining sensitive to the possibility of failure* (pp. 145–161). Aldershot, UK: Ashgate Publishing Limited.

Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. *Work & Stress*, *14*(1), 74–99. doi:10.1080/026783700417249