Early-Warning Signals for Disease Activity in Patients Diagnosed with Multiple Sclerosis based on Keystroke Dynamics

J. Twose,[1, a)] G. Licitra,[1, b)] H. McConchie,[1, c)] K.H. Lam,[2, d)] and J. Killestein[2, e)]
[1)]*Neurocast B.V., Amsterdam 1097DN, The Netherlands*
[2)]*Department of Neurology, Amsterdam University Medical Centers, Amsterdam 1105AZ, The Netherlands*

(Dated: 7 October 2020)

Within data gathered through passive monitoring of patients with Multiple Sclerosis (MS), there is a clear necessity for improved methodological approaches to match the emergence of continuous, objective, measuring technologies. As most gold standards measure infrequently and require clinician presence, fluctuations in the daily progression are not accounted for. Due to the underlying conditions of homogeneity and stationarity (the main tenets of ergodicity) not being met for the majority of the statistical methods employed in the clinical setting, alternative approaches should be investigated. A solution is to use a Non-Linear Time Series Analysis (NLTSA) approach. Here Early Warning Signals (EWS) in the form of critical fluctuations in Keystroke Dynamics (KD), collected using participant's smart phones, are investigated as indicators for clinical change in three groups. These are: patients with MS and changes in Magnetic Resonance Imaging (MRI), patients with MS but without changes in MRI, and healthy controls (HCs). Here, we report examples of EWS and changes in KD coinciding with clinically relevant changes in outcome measures in both patients with and without differences in the amount of MRI enhancing lesions. We also report no clinically relevant changes in EWS in the HC population. This study is a first promising step towards using EWS to identify periods of instability as measured by a continuous objective measure as a proxy for outcome measures in the field of MS.

The monitoring of disease activity in chronic diseases such as Multiple Sclerosis (MS) is integral to assist clinicians in tracking the progress of the disease, and any treatments or interventions. Non-Linear Time Series Analysis (NLTSA) can be used to quantify change in time series data, enabling the identification of clinically relevant changes in Keystroke Dynamics (KD). In the current study the research question focuses on the feasibility of using the continuous objective measure of KD in conjunction with NLTSA. Hence, we show examples of change in KD which coincide with clinically relevant changes in outcome measures in patients with MS. To contrast this, we also show no clinically relevant change in a HC population. This study may be seen as a promising first step towards a NLTSA approach being applied to data from a continuous, objective measure that can monitor change in status within patients with MS. Future research should look at improving the statistical methodology used here, quantifying the predictive nature of the change in KD as a proxy for the outcome measures, and broadening the use of this approach out to other diseases.

I. INTRODUCTION

Biometric data are physical or behavioural features[1] that are unique to each person and can be used as a means of individual validation for security purposes. Physical biometric features can include, for example, finger prints, iris scans and face scans which are unlikely to change except through physical damage. Behavioural features are able to change based on the variable state of a person and can include KD, gait, voice, and other personal identifiers from which an overall personal profile can be developed[2]. The biometrics of KD are based on the assumption that different people have different typing mannerisms, and that these neuro-physiological factors are reflected in the data of that individual, leading to a "typing signature" of a person at any given time.

Within this study we recruited a population of individuals with MS. Given the nature and subtypes of MS, the disease activity and disease progression can vary greatly over time both between and within individuals[3]. Typically, clinical visits occur every 3-12 months for patients with MS, and increasing clinical visits to track the variation that occurs between these visits would introduce burden into a population already hampered by their chronic disease management[4]. The current gold standard methods[5] for measuring changes in disease activity are lesion changes gathered through MRI[6,7], the Expanded Disability Status Scale (EDSS)[8], Nine Hole Peg Test (9HPT)[9], and the Timed 25 Foot Walk (T25FW)[10]. Many of these measures are subjective, infrequently sampled, require clinician presence, and can be burdensome. Additionally, given the nature of the data (infrequently sampled and therefore in low abundance) this data is analysed on a group level. This aggregation irons out variation on the individual level. A more patient centric, and individualised approach could increase the data quality and quantity while improving the lives of individuals living with chronic diseases[11,12]. KD are an example of one solution to this need.

Here, KD data are collected passively by the *Neurokeys*[13] App designed by the Dutch company Neurocast B.V.[14]. The

[a)]james@neurocast.nlNeurocast B.V., Amsterdam 1097DN, The Netherlands
[b)]Electronic mail: giovanni@neurocast.nl; https://neurokeys.app/
[c)]https://neurokeys.app/
[d)]https://www.vumc.com/
[e)]https:/www.vumc.com/

Neurokeys App is a smart-keyboard which replaces a user's native keyboard and tracks how, but not what the user is typing. Compared to classic methods for gathering clinical data in MS, this offers the ability to collect objective data with high frequency of sampling without increasing patient burden. KD have already been successfully implemented for cryptography applications[15], user authentication[16,17] and group based analyses in both MS[18], Parkinson's disease (PD)[19] and Depressive tendency (DT)[20].

To our knowledge KD have not been implemented previously within an MS population outside of single timepoint correlational analysis[18] and short-term responsiveness analysis[21] produced from this same study cohort. Other technological solutions have been used within a MS population to attempt to solve the limitations of classic methods. Aside from digitalized Patient Reported Outcomes (PRO) there are also digitalized finger and foot tapping tasks which show higher sensitivity and specificity to detect clinical disease progression than the 9HPT and T25FW[22]. Another digital solution to measure change in disease outcomes is through wearables such as actigraphs which are designed to measure gait or movement metrics. One such study found good correlations between mobility-stance, turning angle velocity and postural sway, and the EDSS and multiple sclerosis functional composite[23]. Additionally, correlations between REM sleep based on digitally collected pulse rate variability and disease severity scores were found. These solutions are promising, however, digitalization of existing gold standard measurements retain limitations of infrequent sampling and active data collection. Wearables also introduce difficulties related to comfort, distribution, and accessibility, which is unideal for longitudinal assessment of a patient's state[24].

KD have been utilized within PD[19] and DT[20] populations. PD is a movement disorder where disability is shown through characteristic motor symptoms such as tremor, rigidity, and bradykinesia. This is a logical technological fit for KD, measuring symptomatic changes in fine motor skills. A comparable study within PD recently used smart-phone derived timing and pressure-based features to classify people with or without PD. These data were collected in-person during transcription of prescribed texts using a test Android phone.

Within a DT population[20], KD also find a technological fit. In this population symptoms such as sadness, loss of energy, and increased fatigue can cause different psychomotor behaviours. This psychomotor retardation is reflected in changes in routine such as interaction with a smart device. Gold standards in this population involve in-person clinical diagnosis using the Diagnostic Statical Manual of Mental Disorders as well as other standard rating scales. Questionnaires such as these can be biased by subjectivity and impacted by disease stigma and population motivation. In one study, a Random Forest Classifier was used to distinguish a HC population from participants with DT, through natural typing on their own (Android) smartphones in a real-world environment.

In both of the aforementioned studies[19,20], a binary classification pipeline is applied to distinguish whether a user has symptoms of PD or DT based on averaged predicted probabilities from their keystroke data. This group based analysis aimed to detect presence of disease based on typing behaviour. There are two key differences between these examples and the analysis shown here. The presented analysis identifies changes in disease course through an individual approach, and the methods to do so have a focus on days when a user is most changeable in order to amend the timing of that user's treatment[25]. In the current study the problem is conceptualised in a Non-Linear Time Series (NLTSA) framework. Each KD feature is considered a time series and the identification of EWS via Dynamic Complexity, is used to quantify the pathology related status changes within the complex adaptive system, that is each user.

This non-linear methodology has recently been applied in research looking to identify change in self-report high frequency data of a single patient with depression[26]. In this study continuous self-report was found to have the properties of a complex dynamical system given by 1) memory - that the current state is dependent on previous states - exhibited by long range temporal correlations, 2) non-stationarity - that the mean changes over time, often with multiple change points, 3) sensitive dependence on initial conditions - indicated by the limited predictive horizon, that forecasting upcoming values is nearly impossible due to the structure of the data. To address these challenges, the Adaptive DynAmic Pattern Theory (ADAPT) of psychopathology was introduced[25]. Here the authors postulate that all observable phenomena of the body and mind arise from a complex adaptive system - the individual in their environment. They therefore theorise that pathology is a self-organising emergent property of this system and claim that changes can be described using general principles of pattern formation in complex adaptive systems.

Prior to this unifying theory, ADAPT has been applied in different forms in the literature. In neurology, the Lyapunov exponent of Electroencephalogram (EEG) data has been used to identify a decrease in brain complexity of patients with Alzheimer's disease (AD)[27]. The sample entropy of resting state functional MRI data has been used to quantify the difference in brain complexity between individuals with and without attention deficit hyperactivity disorder on a group level[28]. Recurrence Quantification Analysis (RQA) summary statistics applied to Autonomic Nervous System signals, for example heart rate, have been used to correlate with gold standards in the field of anxiety[29]. A method of quantifying peak complexity similar to the current study was used to predict treatment outcomes in mood disorders[30,31]. This study used changes in complexity of a daily self rating by patients undergoing therapy for mood disorders to predict each patient's therapeutic outcome.

In this paper we combine the use of KD biometrics collected via smartphone interactions and NLTSA within an MS population. As a matter of comparison, a set of HC is considered. We hypothesize that we will see individual change within KD where there is a change in outcome measures within the clinical population, and expect no change within the HC population. Our proposed approach enabled detection of individual patient status changes in outcome measures through the use of a non-linear framework and shows promise for utilization in precision medicine within MS.

II. STUDY OVERVIEW

The current section will address the study design: a description of the populations and how and when the data were collected; the clinical outcomes: a summary of the clinical outcomes used in this study to assess disease activity in MS; keystroke data acquisition: an introduction to the *Neurokeys* technology; and the ethical approval of the study.

A. Study design

The data have been collected through an observational cohort study carried out at Amsterdam University Medical Center, location VU University Medical Centre. This study occurred over five clinical visits with three-month intervals for a total duration of 12 months. Keyboard interaction data was remotely collected throughout the study. Participants with MS and HC were included from August 2018 to December 2019.

Study inclusion criteria were: regular use of a smartphone with Android or iOS, age between 18 and 65 years, and definite diagnosis of MS. The exclusion criteria were: EDSS score of 7.5 or higher, clinical disease activity or changes in disease modifying drugs in the past two months, significant visual or upper extremity deficits affecting the ability to type on a smartphone, and clinically significant mood, sleep or behavioural disorders judged by a screening physician. For further information regarding the study design refer to[18].

Patient reported outcomes were collected at Baseline ($m00$), 2-week follow up ($m002$), and then at 3 month intervals following baseline ($m03$, $m06$, $m09$, $m12$) for participants with MS, and $m00$, $m002$ and $m03$ for HC participants. MRIs were carried out at $m00$, $m03$, $m06$, $m09$ and $m12$ for participants with MS, and not carried out for HC participants. Clinical outcome measures were collected at $m00$, $m03$, $m06$, $m09$, $m12$ for MS participants, and $m00$, $m03$ for HC participants. To measure disease activity, data were gathered via MRI of the brain and included T1 weighted images after administration of gadolinium (Gd). These images were assessed by a neuroradiologist to detect presence of Gd-enhancing demyelinating lesions as an indication of inflammatory activity[32].

B. Clinical outcomes

Outcome measures relating to disease activity and clinical disability, fatigue and quality of life were collected and assessed. These include: MRI[32], EDSS[8], 9HPT[9,10], T25FW[33], Ambulation, Arm Function in Multiple Sclerosis Questionnaire (AMSQ)[34], Symbol Digit Modalities Test (SDMT)[35], Fatigue Severity Scale (FSS)[36], Checklist Individual Strength - 20 Revised (CIS-20R)[37], Modified Fatigue Impact Scale (MFIS)[36], Global Rate of Change Fatigue (GRC-fatigue)[18], Global Rate of Change QoL (GRC-QoL)[18].

C. Keystroke data acquisition

Keystroke data were collected from the participants using their mobile device following the installation of *Neurokeys*. This app consists of a customized software keyboard with identical layouts and similar capabilities compared to the default keyboards in iOS and Android, such as auto-correction and word prediction. Once the participants installed Neurokeys, data from each of their typing sessions were gathered. The raw data consisted of timestamps of key presses and releases augmented with typing metadata such as the number of typing events and backspaces. Note that in order to guarantee the privacy of the participants neither letters nor the (X,Y) coordinates relative to the keys pressed were captured at any point by Neurokeys. All data gathered by the application was temporarily stored locally on the mobile device and subsequently sent in batches to a secure cloud storage whenever an internet connection was available.

D. Study approval and ethics approval

The study protocol was approved by the local institutional ethics review board (reference 2017.576) and conformed to the General Data Protection Regulation (GDPR). In compliance with Dutch legislation regarding clinical research involving medical devices, the *Dutch Health and Youth Care Inspectorate* were notified of the study (reference VGR2006948). Written informed consent was obtained from all participants. The study was registered at `trialregister.nl` (NL7070).

III. METHOD

This section will address the feature engineering: the steps taken to go from keystroke related timestamps to meaningful KD features; the feature selection: a description of the methodology used to select the keystroke features that were used in the analysis. This section also provides an explanation of which keystroke features were chosen; and the construction of complexity measures: the mathematical formulation of the complexity measures used in this study to evaluate change in KD.

A. Feature Engineering

In order to construct features coming from KD, we first defined hold time HT_n as the difference between the press and release of a key, and flight time FT_n = the time between a key release and a key press, that is:

$$\mathrm{HT}_n = t_n^r - t_n^p \qquad n = 1,2,\ldots,\mathrm{N} \tag{1a}$$

$$\mathrm{FT}_n = t_{n+1}^p - t_n^r \qquad n = 1,2,\ldots,\mathrm{N}-1 \tag{1b}$$

where t_n^p and t_n^r denote the timestamp sequences relative to the key press and release events, respectively, with N equal to the

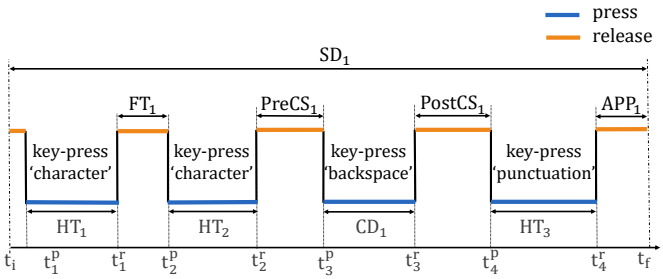


FIG. 1. Depicts a graphical representation of keystroke variables collected within a typing session. Here duration $\mathrm{SD}_1 = t_f - t_i$ where t_i and t_f describes the time when the keyboard is on-screen and off-screen, respectively. The *hold time* keystroke variable $\mathrm{HT}_{\{1,2,3\}}$ represents the time for which a key is pressed, whereas *correction duration* CD_1 denotes the *hold time* relative to the backspace key. The *Flight Time* keystroke variable (FT_1) is the time between two key presses whereas *Pre Correction Slowing* (PreCS_1), *Post Correction Slowing* (PostCS_1) and *After Punctuation Pause* (APP_1) are special cases of flight time relative to the time before and after a backspace and the time after a punctuation, respectively.

total number of keys pressed during a specific interval, for example, daily, hourly or session intervals. Additional keystroke variables were constructed as special cases of FT_n and HT_n, namely[38–40]:

- After Punctuation Pause (APP_n) - the flight time after a punctuation mark such as a question mark "?" or an exclamation point "!";
- Post Correction Slowing (PostCS_n) - the flight time after a backspace keystroke;
- Pre Correction Slowing (PreCS_n) - the flight time prior to a backspace keystroke;
- Correction Duration (CD_n) - the hold time relative to a backspace keystroke;
- Session Duration (SD_n) - the time when the keyboard is on-screen.

Fig. 1 graphically summarizes the keystroke variables introduced above.

In order to avoid outliers coming from edge cases, such as cases when the keyboard is on-screen without any typing activity or when special characters are required, FT_n and HT_n were opportunely filtered prior to any further mathematical operation. Once sequences of keystroke variables were available, it is common practice to compute summary statistics within a specific time interval, for instance daily or hourly intervals. In the current study, a daily interval has been chosen. These aggregations were defined as a linear or non-linear function $\Gamma(\mathbf{x}^p) : \Re^{\mathrm{N}} \rightarrow \Re$ where $\mathbf{x}^p \in \Re^{\mathrm{N}}$ represents a sequence of keystroke variables containing N data points relative to the participant p.

B. Feature Selection

In order to retain the features with the most amount of complexity information, a feature selection procedure was carried out in a non-linear fashion via RQA[41,42]. In short, RQA aims to identify time series which exhibit recurrent patterns by analysing the *Recurrence Plot* which is formally defined as a matrix $\mathbf{R}(\varepsilon, \mathbf{y}) \in \Re^{\mathrm{N_R} \times \mathrm{N_R}}$ where $\mathrm{N_R}$ is the length of a time series $\mathbf{y}$ and where each element of the matrix is given by:

$$r_{i,j} = \begin{cases} 1 & if \; \|y_l - y_m\|_1 \leq \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $\varepsilon \in \Re$ denotes the *threshold distance* and $(l,m) = 1,2,\ldots,\mathrm{N_R}$. In the current study, an explicit threshold distance was not defined, instead a required Recurrence Rate (RR) was specified. This method was chosen in order to make the comparison of the Recurrence Plots (RPs) and the resulting summary statistics of each of the features per user possible. Note that prior to RQA missing values present in the time series, (e.g. insufficient use of the keyboard during the day) were imputed using the Classification and Regression Trees (CART) method[43] implemented in the MICE[44] package. The MICE package creates multiple imputations based on Fully Conditional Specification[45], where each time series with missing values is imputed by a separate model. The amount of missing values varied per keystroke feature and user, despite the median percentage of imputed values being roughly 1 %. RQA was conducted within a PYTHON environment and using the package PYUNICORN[42] with user defined parameters specified in Table I. An example of a RP relative to the participant with user ID = 389 is shown in Fig. 2.

TABLE I. RQA parameter setting.

RQA parameters	Symbol	Value
embedding dimension	m	3.0
embedding delay	τ	2.0
recurrence rate	RR_τ	0.05
line length minimum	l_{min}	2.0
vertical length minimum	v_{min}	2.0

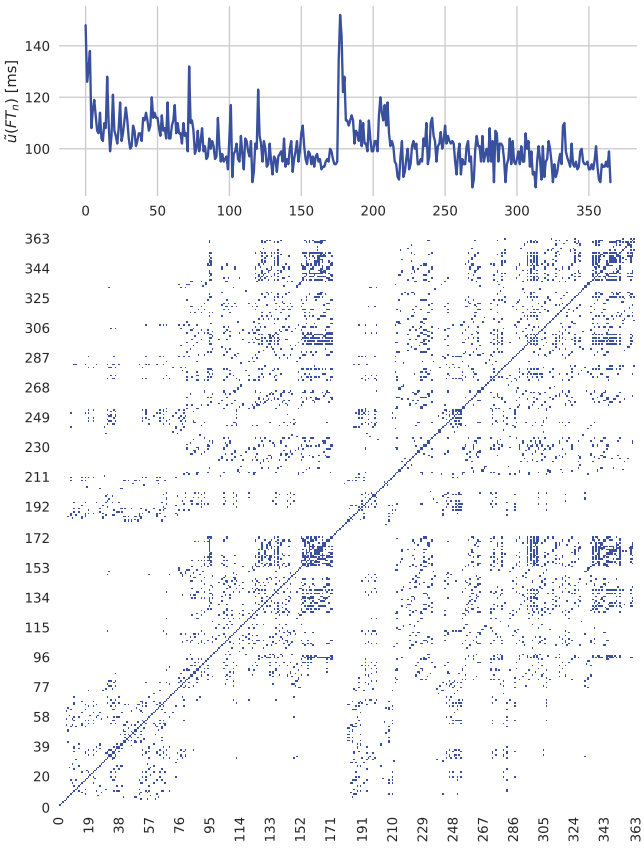


FIG. 2. Depicts the daily median value of flight time $\tilde{\mu}\,(\mathrm{FT}_n)$ relative to the participant with user ID = 390 (Top plot) with corresponding RP (Bottom plot). From the square structures seen in the RP it can be seen that there are periods of high recurrence with abrupt transitions in the middle of the time series (day 175 - 225).

By means of $\mathbf{R}(\varepsilon,\mathbf{y})$ the *RQA summary statistics*[41,46] were extracted, namely *Shannon Entropy*, *Determinism* and *Laminarity* for each participant and for each KD feature.

For each summary statistic the 10 features with the highest median values were selected. This lead to 30 features selected, any duplicates were then dropped. Of these features the correlation of percent change was calculated to avoid redundant information in the feature set. The procedure mentioned above led to a feature subset aggregated as follows:

- Mean value of a central approximation of the second derivative equal to:

$$\mu''(\mathbf{x}) = \frac{1}{2(\mathrm{N}-2)} \sum_{i=1,\ldots,\mathrm{N}-1} \frac{1}{2}\left(x_{i+2} - 2\cdot x_{i+1} + x_i\right)$$

  The following keystroke variables combined with the $\mu''(\mathbf{x})$ operator were selected: PreCS_n , PostCS_n, CD_n, APP_n, SD_n.

- Mean change, that is the mean over the differences between subsequent time series values given by:

$$\mu_c(\mathbf{x}) = \frac{1}{\mathrm{N}-1} \sum_{i=1,\ldots,n-1} x_{i+1} - x_i$$

  For this type of aggregation only APP_n was selected.

- Mean absolute change, namely the mean over the absolute differences between subsequent time series values, written as:

$$\mu_{|c|}(\mathbf{x}) = \frac{1}{\mathrm{N}} \sum_{i=1,\ldots,\mathrm{N}-1} |x_{i+1} - x_i|$$

  For this type of aggregation only the keystroke variable FT_n was selected.

- Partial autocorrelation with lag k which for a time series refers to the partial autocorrelation of x_i with x_{i-k}, conditional on $x_c := \{x_{i-1},\ldots,x_{i-k+1}\}$, defined as[47]:

$$\alpha_k(\mathbf{x}) = \frac{\mathrm{Cov}\,(x_i, x_{i-k}|x_c)}{\mathrm{Var}\,(x_i|x_c)\cdot \mathrm{Var}\,(x_{i-k}|x_c)}$$

  For this type of aggregation PostCS_n was the only keystroke variable selected using lag $k = 1$.

Furthermore, the standard deviation of hold time $\sigma\,(\mathrm{HT}_n)$, the median of flight time $\tilde{\mu}\,(\mathrm{FT}_n)$, the maximum value of flight time $max\,(\mathrm{FT}_n)$ and the skewness of hold time $s\,(\mathrm{HT}_n)$ were also selected. A selection of the chosen features are shown in Fig. 3 for three different users.

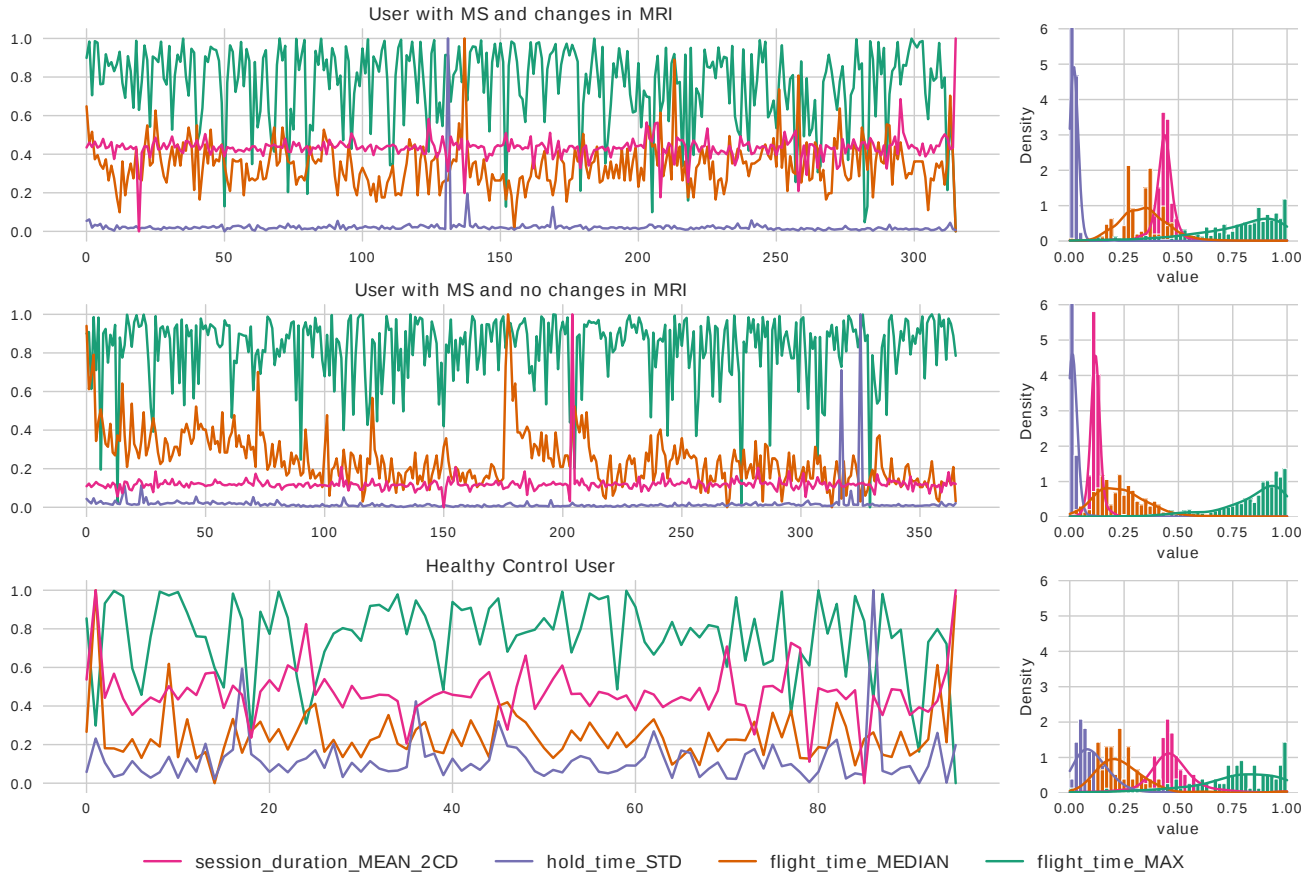


FIG. 3. Depicts examples of day-to-day variation of the selected keystroke features of a patient with MS with change in MRI, a patient with MS with no change in MRI, and a HC subject. For plot legibility a random subset of the features has been chosen. Interestingly there are distinct differences in amount of change between the 2 users with MS, and the HC user. Furthermore, it can be seen that for a certain user one feature may have pertinent change information, but not in another user. For example, flight time median exhibits a period of instability for the MS user with no change in MRI, whilst this is not the case for the HC user. For a more in depth investigation of the features in these users see the supplemental material.

C. Construction of complexity measures

Human change processes are characterized by both non-linear and non-stationary dynamics which can lead to chaotic behaviour with discontinuous phases accompanied by critical instabilities that precede and enable such transitions[48,49]. According to Schiepek and Strunk[50], non-stationary phenomena and critical instabilities can be identified via the *Fluctuation Intensity* and the *Distribution Uniformity* calculated within a moving window with m_w measurement points.

The Fluctuation Intensity is a measure sensitive to the amplitude and frequency of changes which occur in the time series $\mathbf{x} \in \Re^{\mathrm{N}}$ and is given by:

$$\mathrm{F}(\mathbf{x}, m_w) = \frac{1}{(x_{\mathrm{M}} - x_{\mathrm{m}})(m_w - 1)} \sum_{k=1}^{m_w - 1} \frac{|x_{k+1} - x_k|}{n_{k+1} - n_k}, \; \mathrm{F} \in [0, 1] \tag{3}$$

where the index k refers to the *points of return*, specifically, the number of changes in slope of the time series, whereas x_{M} and x_{m} denote the maximum and minimum value of $\mathbf{x}$, respectively. Conversely, the Distribution Uniformity D quantifies the irregularities of $\mathbf{x}$ by comparing it with an ideal distribution produced by another time series $\mathbf{y}$ with equal number of time points N, and is defined as:

$$\mathrm{D}(\mathbf{x}, \mathbf{y}, m_w) = 1 - \sum_{c=1}^{m_w - 1} \sum_{d=c+1}^{m_w} \sum_{a=c}^{d-1} \sum_{b=a+1}^{d} \frac{\Delta_{ba} \Theta(\Delta_{ba})}{y_b - y_a}, \; \mathrm{D} \in [0, 1] \tag{4}$$

with $\Delta_{ba} = (y_b - x_b) - (y_a - x_a)$ that quantifies the aberration of $\mathbf{x}$ with respect to $\mathbf{y}$, and $\Theta(\cdot)$ the Heaviside step function. The two outer sums are permutations of all combinations of c and d within the window, whereas the inner sums with index a and b are representing all combinations of positions within the interval given by c and d. In the current study, it was considered a moving window with $m_w = 28$ days to account for the slow changes in disease activity in MS.

For each feature, both F and D are computed, and subse-

quently merged via an element-wise multiplication in order to construct the *Dynamic Complexity*[50] denoted as C. The dynamic complexity contains information regarding the distribution, amplitude, and frequency changes of the keystroke feature in one single time series of length $N - m_w$. Finally, the EWS are quantified by the Cumulative Complexity Peaks (CCP)[51] which in this case are carried out whenever the standardised keystroke feature's Dynamic Complexity (C) is bigger than the 99^{th} quantile. The C and CCP were calculated within an R environment using the CASNET package[52], and illustration of them is provided in Fig. 4 relative to the user 390.

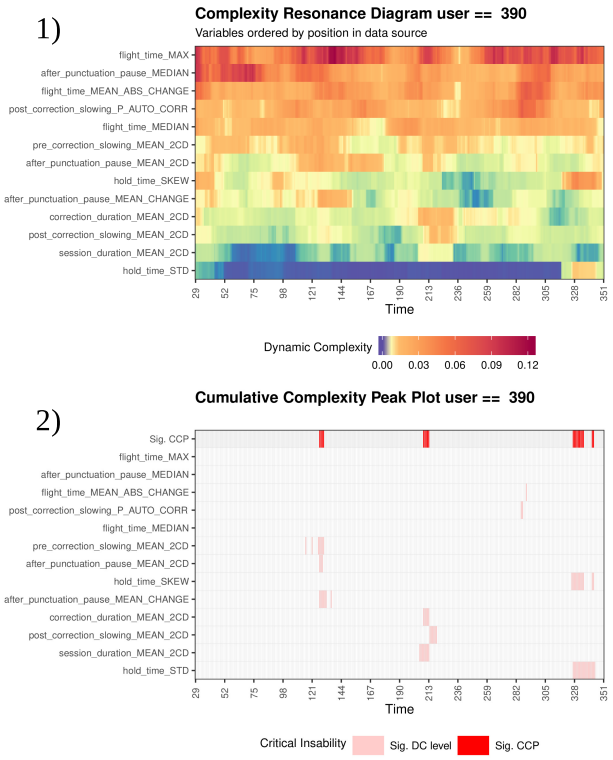


FIG. 4. Displays a graphical example of the Dynamic Complexity C (plot 1) with the corresponding Cumulative Complexity Peaks (plot 2) relative to user 390. For each feature selected using the approach described in section III B, the C is computed by multiplying element-wise the Fluctuation Intensity F and the Distribution Uniformity D. The top plot is known as *Complexity Resonance Diagram* and it highlights the number of changes that occur in the keystroke dynamics. For instance, one can observe that the standard deviation of flight time has predominantly no changes except for the period of time from day 328 onwards. Such abrupt change is seen as critical instability which is visualized in the bottom plot in pink. In plot 2, one can also observe that the cumulative complexity peaks are triggered and flagged in red (top row named SIG. CCP) when critical instabilities occur simultaneously from different sources.

IV. RESULTS

The NLTSA approach was applied to 13 MS users with change in Gd enhancing lesions, 21 MS users with no change in Gd enhancing lesions and 24 HC. Fig. 5, summarizes the entire data processing pipeline, including the collection from the keyboard, the creation of the aggregate features, the imputation of missing values, the feature selection via RQA, the scaling, and the steps taken to create the C and CCP.

Fig. 2 shows the method used to select features, the RP, here the FT_n median is shown. Of note in this plot are the square structures formed by the recurrent points along the time series. These are apparent between days 0 - 76, days 80 - 171, and days 210 - 363. Pertinent to this analysis is the clear destabilization between days 171 - 210 indicated by a comparably small amount of recurrent points (very few short term and some long term points), and the corresponding period in the line plot.

Following feature selection, an inspection of the chosen features in a subset of the users was undertaken. Fig. 3 shows a subset of the selected features for a single user from each group. Of note in this plot is the differences between the two users with MS and the HC user. This is most prominent in the $\mu''(\mathbf{x})$ of the session duration and the hold time standard deviation. Here you can see there is a small amount of variance for the two users with MS and a gaussian distribution for the HC user. Furthermore, the FT_n median is more indicative of state change for the users with MS, as there are periods in the time series which are elevated in comparison with the HC user. The FT_n median of the HC user fluctuates around a relatively stable mean.

To visualize change in the selected features over time Fig. 4 shows the *Complexity Resonance Diagram* (figure 4 - 1), a visualization of the C for each keystroke feature, and the *Cumulative Complexity Peaks Plot* (Fig. 4 - 2), a visualization of the CCPs relative to user 390. Here one can see that there are features which are inherently highly changeable, for example the FT_n maximum. There are also features which do not change often, for example the HT_n standard deviation. Interestingly, there is a period of comparable high change which is quantified by the CCPs. To highlight this concept, the $\mu''(\mathbf{x})$ of the SD_n, $PostCS_n$ and CD_n are reasonably changeable across the whole period. However, for the days surrounding and including day 213, there is comparatively higher complexity than the rest of the time. This indicates a period of destabilisation and based on ADAPT[25] signifies this as the best period to apply an intervention. This higher complexity is highlighted in the CCP plot and is used in the final results to compare change in keystroke features to the outcome measures.

Figure 6 shows the sum of the dynamic complexity (C), the cumulative complexity peaks (CCPs), and the clinical outcome measures that were considered to have clinically relevant change, per user, between at least two time points according to literature[8–10,32–37] for an appropriate sample size. Here it can be seen that in all the examples of patients with MS, changes in C and the occurrence of CCPs coincided with clinically relevant change in outcome measures.

In a-d) there are clear changes (increases and decreases) in the clinical outcome measures which coincide with the CCPs. For example, in b) there is a distinct increase in the amount of Gd enhancing lesions which coincides with a CCP. Whilst there was a certain amount of change in the C in the HC population (e) and f)) there were no CCPs (other than at the beginning and end of the study), and little to no clinically relevant changes in the outcome measures (for the results of all users see supplemental material).

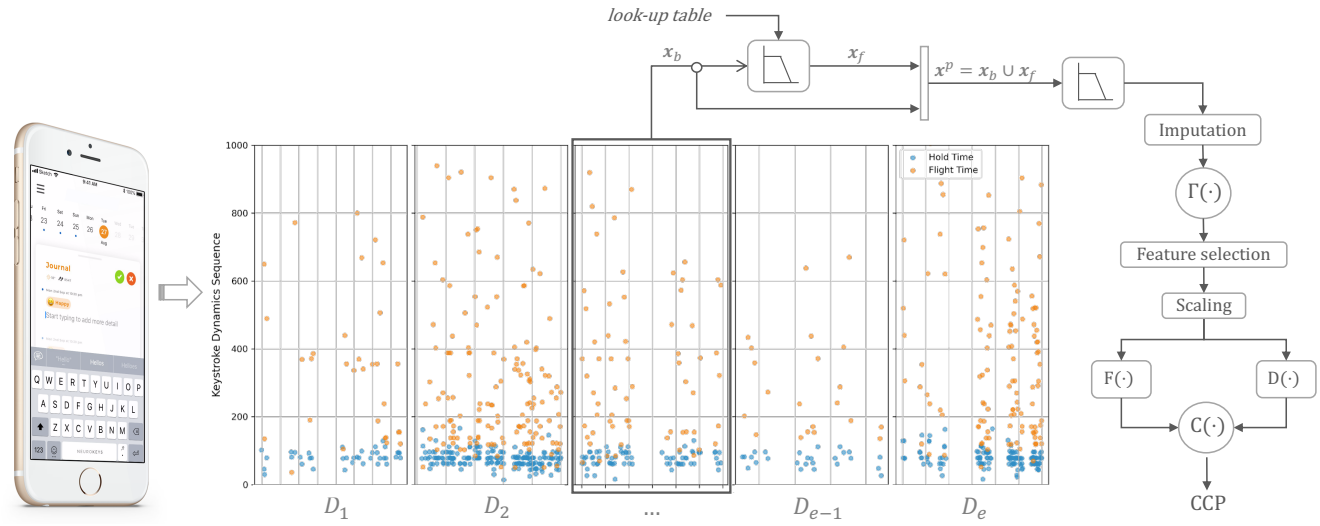


FIG. 5. Keystroke feature construction and non-linear time series pipeline. With respect to keystroke dynamics, for each participant p the flight time FT_n, hold time FT_n and session duration SD_n are constructed and stored in the array $\mathbf{x}_b$. The array $\mathbf{x}_f$ which contains the remaining keystroke variables, namely, after punctuation pause APP_n, correction duration CD_n, pre and post correction slowing PostCS_n, PreCS_n are derived by opportunely filtering FT_n and HT_n via specific typing events such as backspace and selection of punctuation. Subsequently, an outlier removal procedure is carried out and missing values are imputed using the CART algorithm[53]. Keystroke features are then obtained by aggregating several linear and non-linear functions $\Gamma(\cdot)$ such as partial autocorrelation α_k and absolute mean absolute change $\mu_{|c|}(\mathbf{x})$ within a daily $D_i, i = 1,2,\ldots,e$ time interval. The keystroke features are normalised between 0 and 1, to avoid over representation of a feature, and both the Distribution Uniformity $D(\cdot)$ and Fluctuation Intensity $F(\cdot)$ are computed for each keystroke feature. Finally, the Dynamic Complexity $C(\cdot)$ is calculated as the product of $F(\cdot)$ and $D(\cdot)$ and the Cumulative Complexity Peaks (CCPs) are obtained.

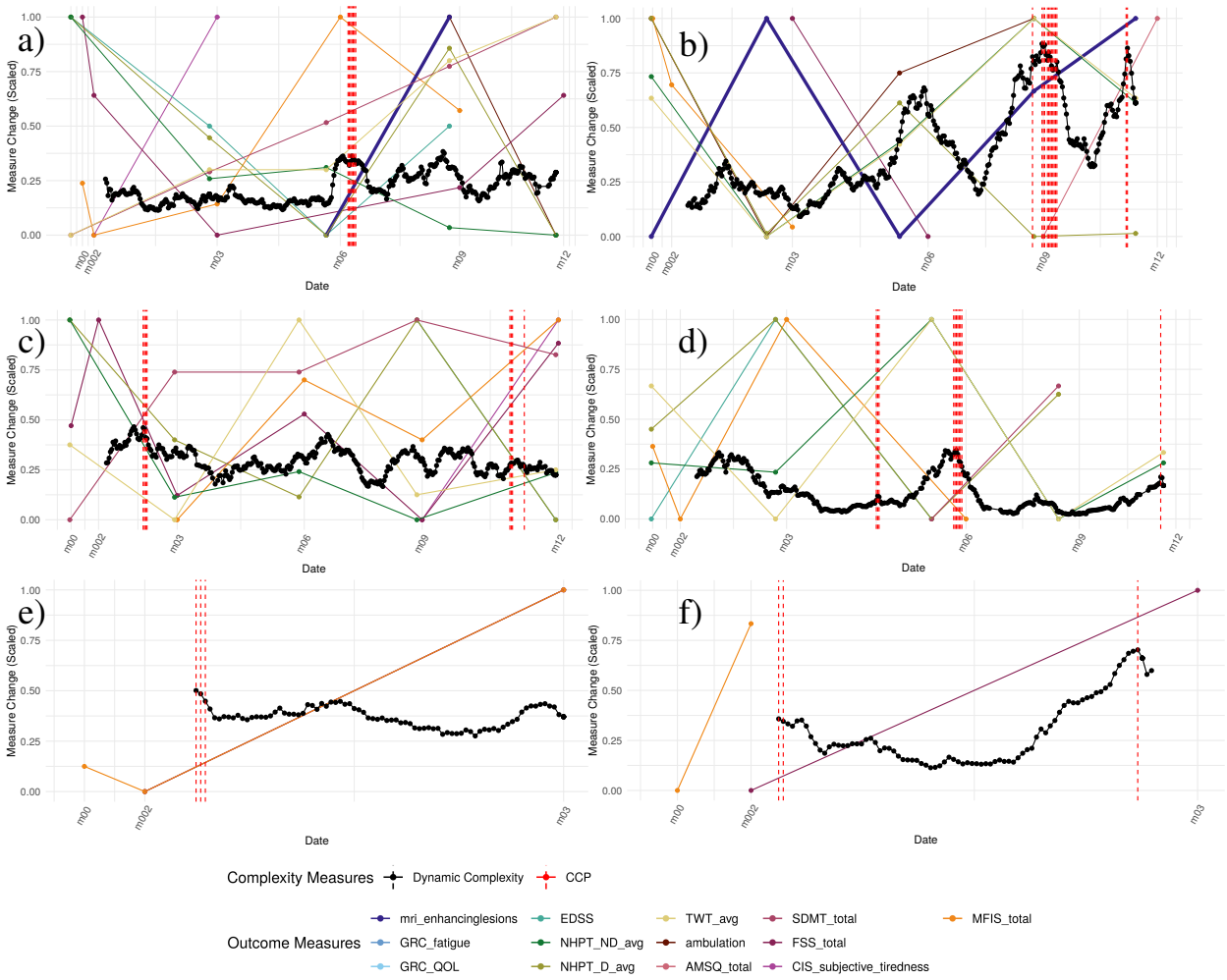


FIG. 6. Depicts the sum of the dynamic complexity on a daily basis (C; shown as black dots), the cumulative complexity peaks (CCPs; shown as vertical dashed red lines) and corresponding clinical outcome measures for a selection of users. Only clinical outcomes considered to have clinically relevant change based on clinically used cut offs from literature are included in the figure. The change in of Gd enhancing lesions is highlighted in bold. a) and b) are patients with MS which also have changes in the amount of Gd enhancing lesions. c) and d) are patients with MS which do not have changes in amount of Gd enhancing lesions. e) and f) are examples from the HC population. It is noticeable that there are no data for the sum of C or CCPs in the beginning of each plot. This is due to the size of the sliding window used to calculate these values being set to $m_w = 28$ days.

V. DISCUSSION

This study aimed to test the feasibility of using NLTSA within KD derived behavioural biometrics. As hypothesized the results showed daily change in KD for all users, but only clinically relevant changes in the population with MS, both with and without Gd enhancing lesions. Furthermore, these clinically relevant changes often coincided with changes in the outcome measures. The clinically relevant changes in KD can be seen as EWS for changes in disease activity of the patient prior to the change occurring. The results from this study are preliminary evidence for the use of the quantification of change in KD, a continuous, objective, non invasive measure, as a proxy for change in disease activity and disease status of patients with MS. The striking differences between the groups is a clear visual indication of how the change profiles of different users can be indicative of the disease activity of MS over time.

A finding that was not accounted for in our hypotheses was the clinically relevant changes in KD in the HC population at the beginning and the end of the study (as indicated by the CCPs in figure 6). A plausible explanation for these CCPs are the inherent perturbations to the system that is introduced by each user when transferring keyboards from their default keyboard to the Neurokeys keyboard, and back again. The reason for why this is apparent in the HC population but not in the participants with MS is likely due to larger changes in the KD of these users masking the initial perturbations. Further research is necessary to confirm this hypothesis.

An additional finding was the difference in destabilisation profiles in specific KD features per user. The flight time median was a good example of this as there was a distinct moment of destabilisation in user 390, but such a moment was not clear in other users for this feature. Based on ADAPT it is argued that this point in time is a transition between states for the user. This period would be the time to implement an intervention for the best possible outcome for user 390, but not for the other users per se.

While interpreting the findings of this study, some limitations need to be considered. In order to better match the changes in C of the KD to the clinical outcome measures, a better approach would have been to administer the outcome measures at more regular intervals than once every three months. For the measures administered during the clinical visits (EDSS, 9HPT, T25FW, SDMT) and the MRI, this would not have been feasible. However a higher sampling rate for the self report measures (AMSQ, FSS, CIS-20R, MFIS, GRC-fatigue, GRC-QoL) would be possible, and would allow for clearer insights into the change profiles of the clinically validated outcome measures, leading to better matching with the KD change profiles. Furthermore, due to the preliminary nature of this exploratory study, there is currently not a quantification of how well the C and the CCPs coincide with the changes in the outcome measures. Building on the current analysis, one could investigate the capabilities of both the C and the CCP in predicting changes in the outcome measures, using time series based predictive modelling[54]. The dynamics of complex systems approach to KD would also benefit from future research investigating other methods of feature selection. The current study uses RQA summary statistics, however there are a host of NLTSA approaches which could be used to select promising keystroke features, for instance *Detrended Fluctuation Analysis*[55]. Moreover, more in depth analysis into an appropriate sliding window size for the calculation of F and D of KD in MS is necessary. In this study we chose a sliding window of 28 days due to the changes in disease activity in MS being markedly slow, whilst still maximizing the amount of data being shown (as the larger the sliding window, the larger the lead in of the F and D). Lastly, the keystroke feature selection was conducted on a group level across all the users, however the results of this analysis would likely be improved by doing individualised feature selection specific to each user. Feature selection on an individual level would likely improve the results as each keystroke feature could hold more complexity information to one user over another. For example, one user may have typing sessions that are very consistent and therefore have low complexity, whereas another user may usually type quickly and have short session durations, apart from when they have a worsening of symptoms.

VI. CONCLUSION

The application of non-linear methods to identify Early Warning Signals in Keystroke Dynamics in order to better monitor change in the course of a MS patient's disease state is compelling despite the exploratory nature of the study. Continued research in this area is impactful given its potential to reduce patient burden and allow for improved remote monitoring of patient state. Furthermore, because this approach is on an individual level, using high frequency data, possible predictions would be specific to each user, and as such would be a first step in the direction of precision medicine. In the future, the next steps of predictive modelling could be achieved using the change in Dynamic complexity and Cumulative Complexity Peaks as predictors. This ultimately could facilitate more timely interventions and positively impact disease course.

SUPPLEMENTAL MATERIAL

See supplemental material for complete R markdown scripts for the data pipeline and final results.

ACKNOWLEDGMENTS

This study is funded by TKI Life Sciences & Health (LSHM16060-SGF), Stichting MS Research (16-946 MS), and Biogen (unrestricted funding). F.B. is supported by the NIHR biomedical research center at UCLH. The authors would like to thank all the participants who participated in the study, and Ton Schweigmann and his colleagues for performing the MRIs.

DATA SHARING POLICY

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due privacy restrictions.

[1]M. Ulinskas, R. Damaševičius, R. Maskeliūnas, and M. Woźniak, "Recognition of human daytime fatigue using keystroke data," Procedia computer science **130**, 947–952 (2018).

[2]F. Monrose and A. D. Rubin, "Keystroke dynamics as a biometric for authentication," Future Generation computer systems **16**, 351–359 (2000).

[3]J. H. Noseworthy, "Progress in determining the causes and treatment of multiple sclerosis," Nature **399**, A40–A47 (1999).

[4]M. H. Sutliff, "Contribution of impaired mobility to patient burden in multiple sclerosis," Current medical research and opinion **26**, 109–119 (2010).

[5]H. Inojosa, D. Schriefer, and T. Ziemssen, "Clinical outcome measures in multiple sclerosis: a review," Autoimmunity Reviews , 102512 (2020).

[6]G. Giovannoni, R. Bermel, T. Phillips, and R. Rudick, "A brief history of neda," Multiple Sclerosis and Related Disorders **20**, 228–230 (2018).

[7]F. D. Lublin, S. C. Reingold, J. A. Cohen, G. R. Cutter, P. S. Sørensen, A. J. Thompson, J. S. Wolinsky, L. J. Balcer, B. Banwell, F. Barkhof, *et al.*, "Defining the clinical course of multiple sclerosis: the 2013 revisions," Neurology **83**, 278–286 (2014).

[8]J. Kragt, J. Nielsen, F. A. van der Linden, B. M. Uitdehaag, and C. Polman, "How similar are commonly combined criteria for edss progression in multiple sclerosis?" Multiple Sclerosis Journal **12**, 782–786 (2006).

[9]P. Feys, I. Lamers, G. Francis, R. Benedict, G. Phillips, N. LaRocca, L. D. Hudson, R. Rudick, and M. S. O. A. Consortium, "The nine-hole peg test as a manual dexterity performance measure for multiple sclerosis," Multiple Sclerosis Journal **23**, 711–720 (2017).

[10]J. Kragt, F. A. van der Linden, J. Nielsen, B. M. Uitdehaag, and C. Polman, "Clinical impact of 20% worsening on timed 25-foot walk and 9-hole peg test in multiple sclerosis," Multiple Sclerosis Journal **12**, 594–598 (2006).

[11]D. A. Robbins, F. A. Curro, and C. H. Fox, "Defining patient-centricity: opportunities, challenges, and implications for clinical care and research," Therapeutic Innovation & Regulatory Science **47**, 349–355 (2013).

[12]N. S. Sharma, "Patient centric approach for clinical trials: Current trend and new opportunities," Perspectives in clinical research **6**, 134 (2015).

[13]"Neurokeys," `http://neurokeys.app/` (2016), accessed: 2020-06-20.

[14]"Neurocast B.V." `https://www.neurocast.nl/` (2016), accessed: 2020-06-20.

[15]T.-Y. Chang, "Dynamically generate a long-lived private key based on password keystroke features and neural network," Information Sciences **211**, 36–47 (2012).

[16]T. Anusas-amornkul and K. Wangsuk, "A comparison of keystroke dynamics techniques for user authentication," in *2015 International Computer Science and Engineering Conference (ICSEC)* (IEEE, 2015) pp. 1–5.

[17]A. Darabseh and A. S. Namin, "Effective user authentications using keystroke dynamics based on feature selections," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (IEEE, 2015) pp. 307–312.

[18]K.-H. Lam, K. Meijer, F. Loonstra, E. Coerver, J. Twose, E. Redeman, B. Moraal, F. Barkhof, V. de Groot, B. Uitdehaag, and J. Killestein, "Real-world keystroke dynamics are a potentially valid biomarker for clinical disability in multiple sclerosis," Multiple Sclerosis Journal (2020).

[19]D. Iakovakis, S. Hadjidimitriou, V. Charisis, S. Bostantzopoulou, Z. Katsarou, and L. J. Hadjileontiadis, "Touchscreen typing-pattern analysis for detecting fine motor skills decline in early-stage parkinson's disease," Scientific reports **8**, 1–13 (2018).

[20]R.-E. Mastoras, D. Iakovakis, S. Hadjidimitriou, V. Charisis, S. Kassie, T. Alsaadi, A. Khandoker, and L. J. Hadjileontiadis, "Touchscreen typing pattern analysis for remote detection of the depressive tendency," Scientific reports **9**, 1–12 (2019).

[21]K.-H. Lam, J. Twose, H. McConchie, G. Licitra, B. Moraal, F. Barkhof, B. Uitdehaag, V. de Groot, and J. Killestein, "Smartphone keystroke dynamics are sensitive to changes in disease activity and clinical disability measures in multiple sclerosis," in *European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS)*.

[22]M. Tanigawa, J. Stein, J. Park, P. Kosa, I. Cortese, and B. Bielekova, "Finger and foot tapping as alternative outcomes of upper and lower extremity function in multiple sclerosis," Multiple Sclerosis Journal–Experimental, Translational and Clinical **3**, 2055217316688930 (2017).

[23]T. Chitnis, B. I. Glanz, C. Gonzalez, B. C. Healy, T. J. Saraceno, N. Sattarnezhad, C. Diaz-Cruz, M. Polgar-Turcsanyi, S. Tummala, R. Bakshi, *et al.*, "Quantifying neurologic disease using biosensor measurements in-clinic and in free-living settings in multiple sclerosis," NPJ digital medicine **2**, 1–8 (2019).

[24]A. Habibipour, A. M. Padyab, and A. Ståhlbröst, "Social, ethical and ecological issues in wearable technologies," in *AMCIS 2019, Twenty-fifth Americas Conference on Information Systems, Cancun, México, Augusti 15-17 2019*. (2019) pp. 1–10.

[25]M. Olthof, F. Hasselman, F. O. Maatman, A. Bosman, and A. Lichtwarck-Aschoff, "Adaptive dynamic pattern theory (adapt) of psychopathology," (2020).

[26]M. Olthof, F. Hasselman, and A. Lichtwarck-Aschoff, "Complexity in psychological self-ratings: Implications for research and practice," (2020).

[27]J. Jeong, "Eeg dynamics in patients with alzheimer's disease," Clinical neurophysiology **115**, 1490–1505 (2004).

[28]M. O. Sokunbi, W. Fung, V. Sawlani, S. Choppin, D. E. Linden, and J. Thome, "Resting state fmri entropy probes complexity of brain activity in adults with adhd," Psychiatry Research: Neuroimaging **214**, 341–348 (2013).

[29]M. Nardelli, G. Valenza, I. A. Cristea, C. Gentili, C. Cotet, D. David, A. Lanata, and E. P. Scilingo, "Characterizing psychological dimensions in non-pathological subjects through autonomic nervous system dynamics," Frontiers in computational neuroscience **9**, 37 (2015).

[30]M. Olthof, F. Hasselman, G. Strunk, B. Aas, G. Schiepek, and A. Lichtwarck-Aschoff, "Destabilization in self-ratings of the psychotherapeutic process is associated with better treatment outcome in patients with mood disorders," Psychotherapy Research **30**, 520–531 (2020).

[31]M. Olthof, F. Hasselman, G. Strunk, M. van Rooij, B. Aas, M. A. Helmich, G. Schiepek, and A. Lichtwarck-Aschoff, "Critical fluctuations as an early-warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders," Clinical Psychological Science **8**, 25–35 (2020).

[32]U. W. Kaunzner and S. A. Gauthier, "Mri in the assessment and monitoring of multiple sclerosis: an update on best practice," Therapeutic advances in neurological disorders **10**, 247–261 (2017).

[33]C. I. Coleman, D. M. Sobieraj, and L. N. Marinucci, "Minimally important clinical difference of the timed 25-foot walk test: results from a randomized controlled trial in patients with multiple sclerosis," Current medical research and opinion **28**, 49–56 (2012).

[34]C. E. van Munster, L. Kaya, M. Obura, N. F. Kalkers, and B. M. Uitdehaag, "Minimal clinically important difference of improvement on the arm function in multiple sclerosis questionnaire (amsq)," Multiple Sclerosis Journal **26**, 505–508 (2020).

[35]L. Strober, J. DeLuca, R. H. Benedict, A. Jacobs, J. A. Cohen, N. Chiaravalloti, L. D. Hudson, R. A. Rudick, N. G. LaRocca, and M. S. O. A. C. (MSOAC), "Symbol digit modalities test: a valid clinical trial endpoint for measuring cognition in multiple sclerosis," Multiple Sclerosis Journal **25**, 1781–1790 (2019).

[36]S. Rooney, A. McFadyen, L. Wood, F. Moffat, and L. Paul, "Minimally important difference of the fatigue severity scale and modified fatigue impact scale in people with multiple sclerosis," Multiple sclerosis and related disorders **35**, 158–163 (2019).

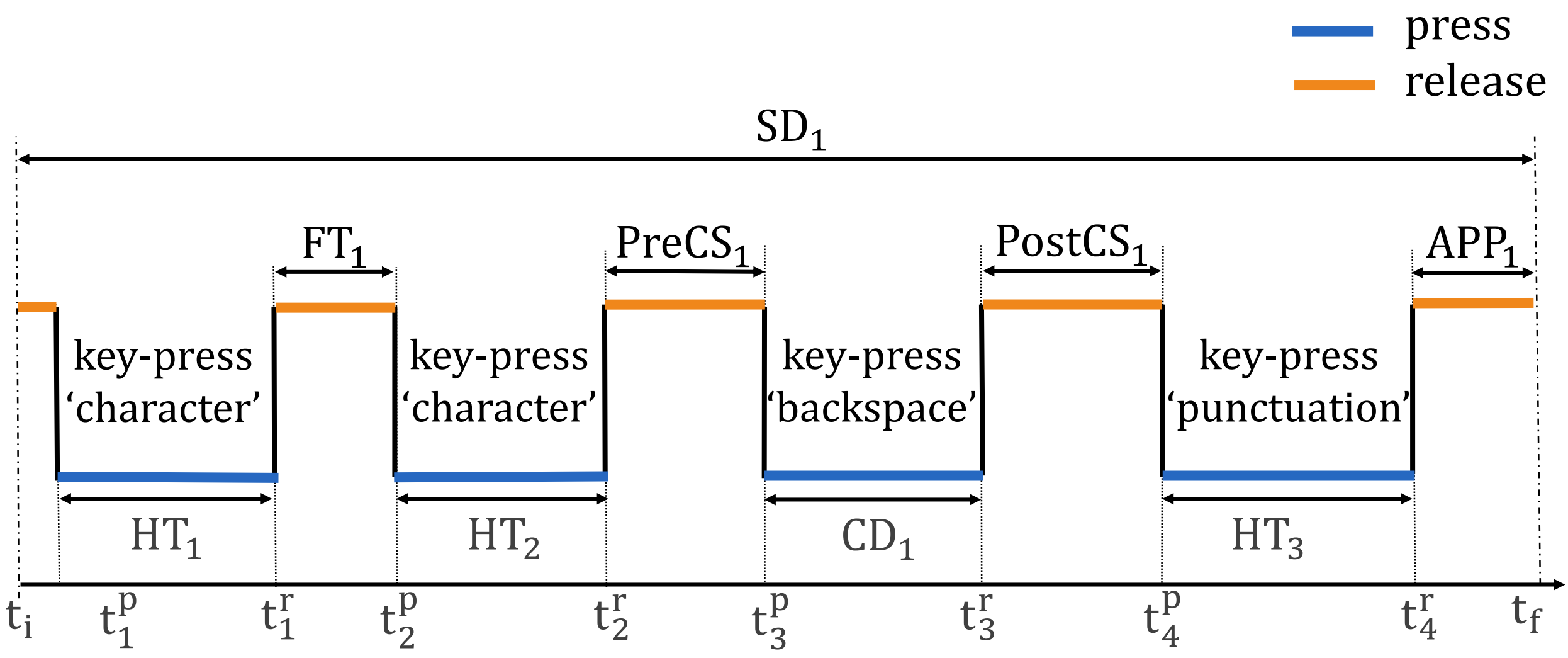
[37]J. B. Peters, Y. F. Heijdra, L. Daudey, L. M. Boer, J. Molema, P. R. Dekhuijzen, T. R. Schermer, and J. H. Vercoulen, "Course of normal and abnormal fatigue in patients with chronic obstructive pulmonary disease, and its relationship with domains of health status," Patient education and counseling **85**, 281–285 (2011).

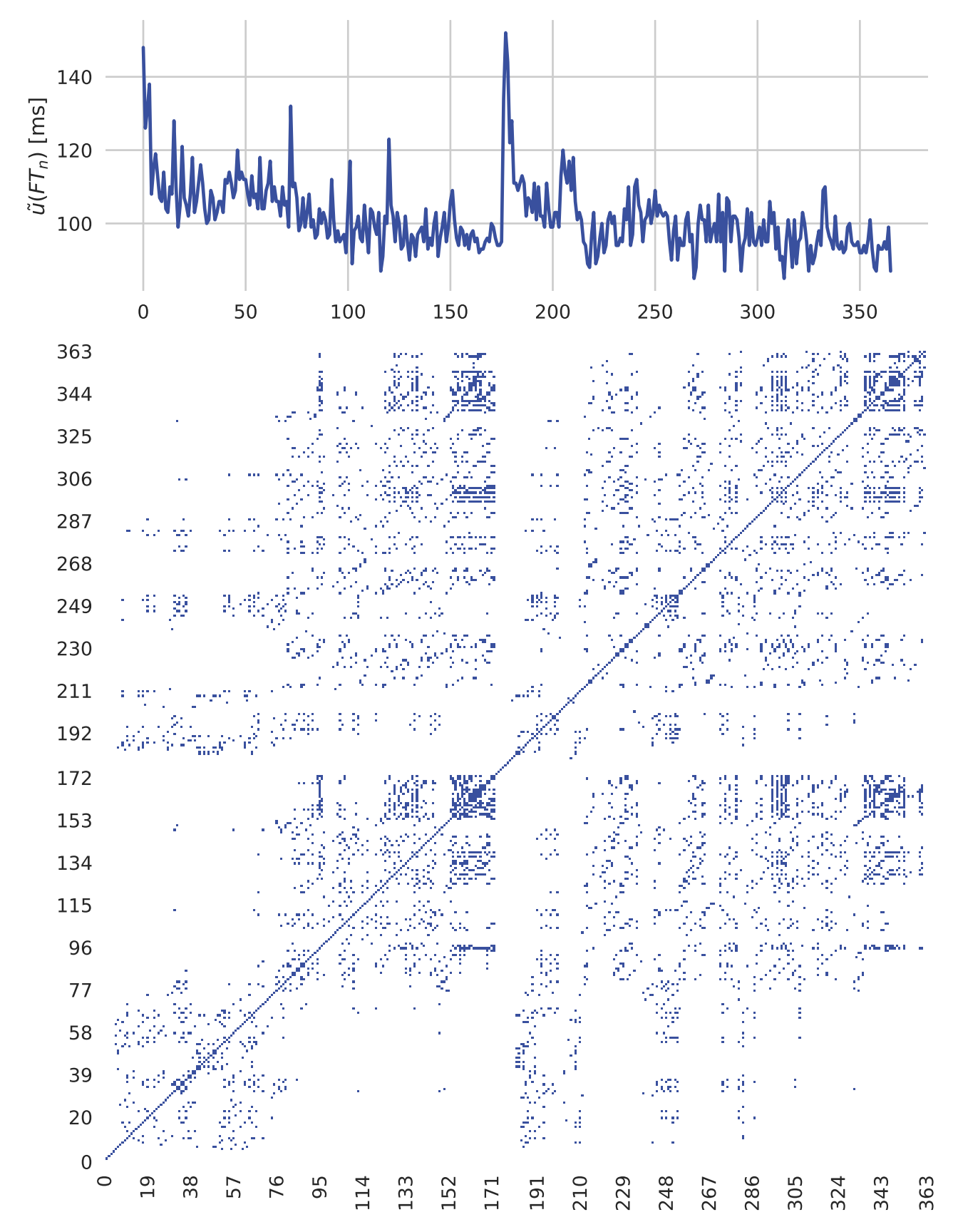
[38]L. M. Vizer, L. Zhou, and A. Sears, "Automated stress detection using keystroke and linguistic features: An exploratory study," International Journal of Human-Computer Studies **67**, 870–886 (2009).

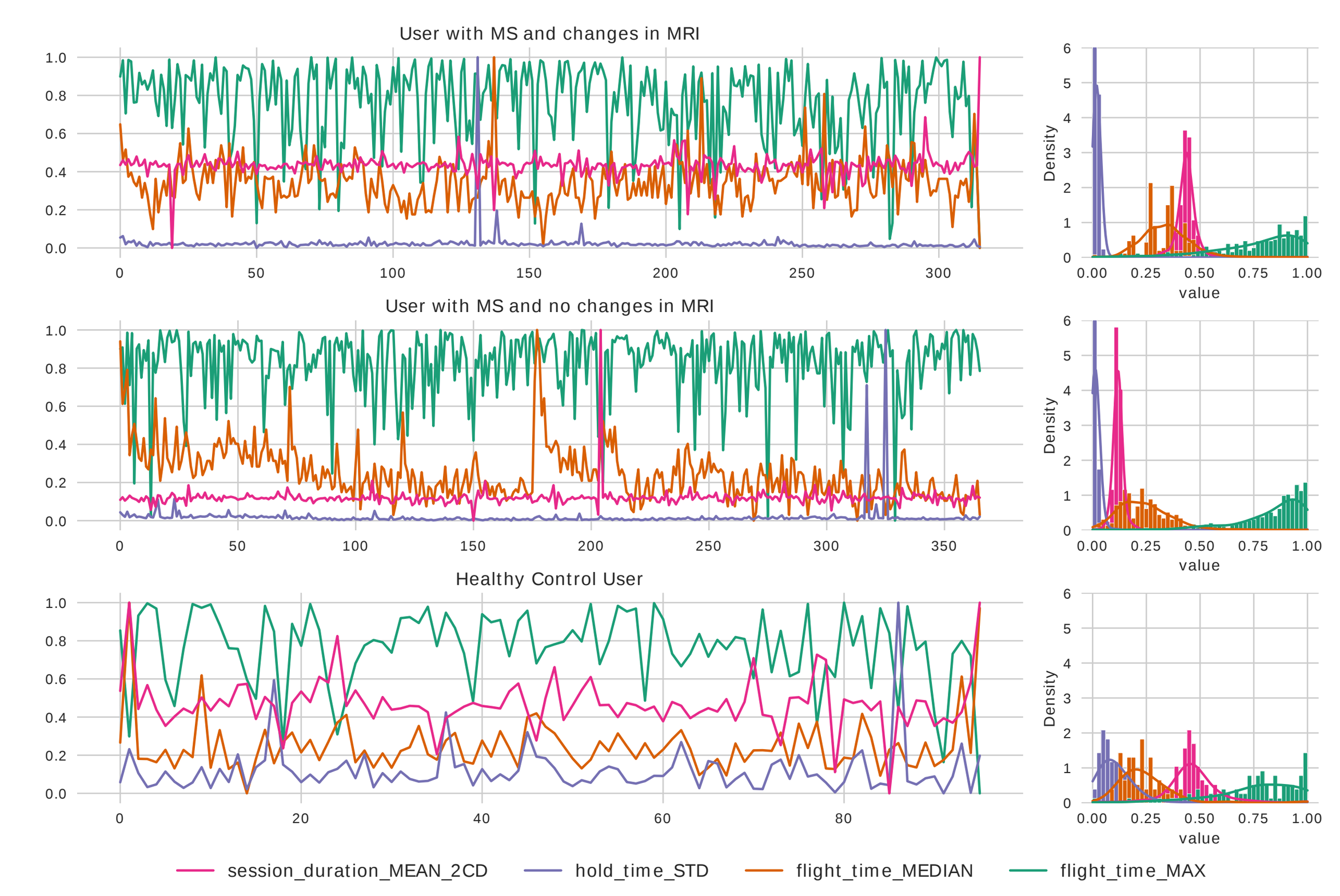
[39]L. M. Vizer and A. Sears, "Classifying text-based computer interactions for health monitoring," IEEE Pervasive Computing **14**, 64–71 (2015).

[40]L. M. Vizer and A. Sears, "Efficacy of personalized models in discriminating high cognitive demand conditions using text-based interactions," International Journal of Human-Computer Studies **104**, 80–96 (2017).

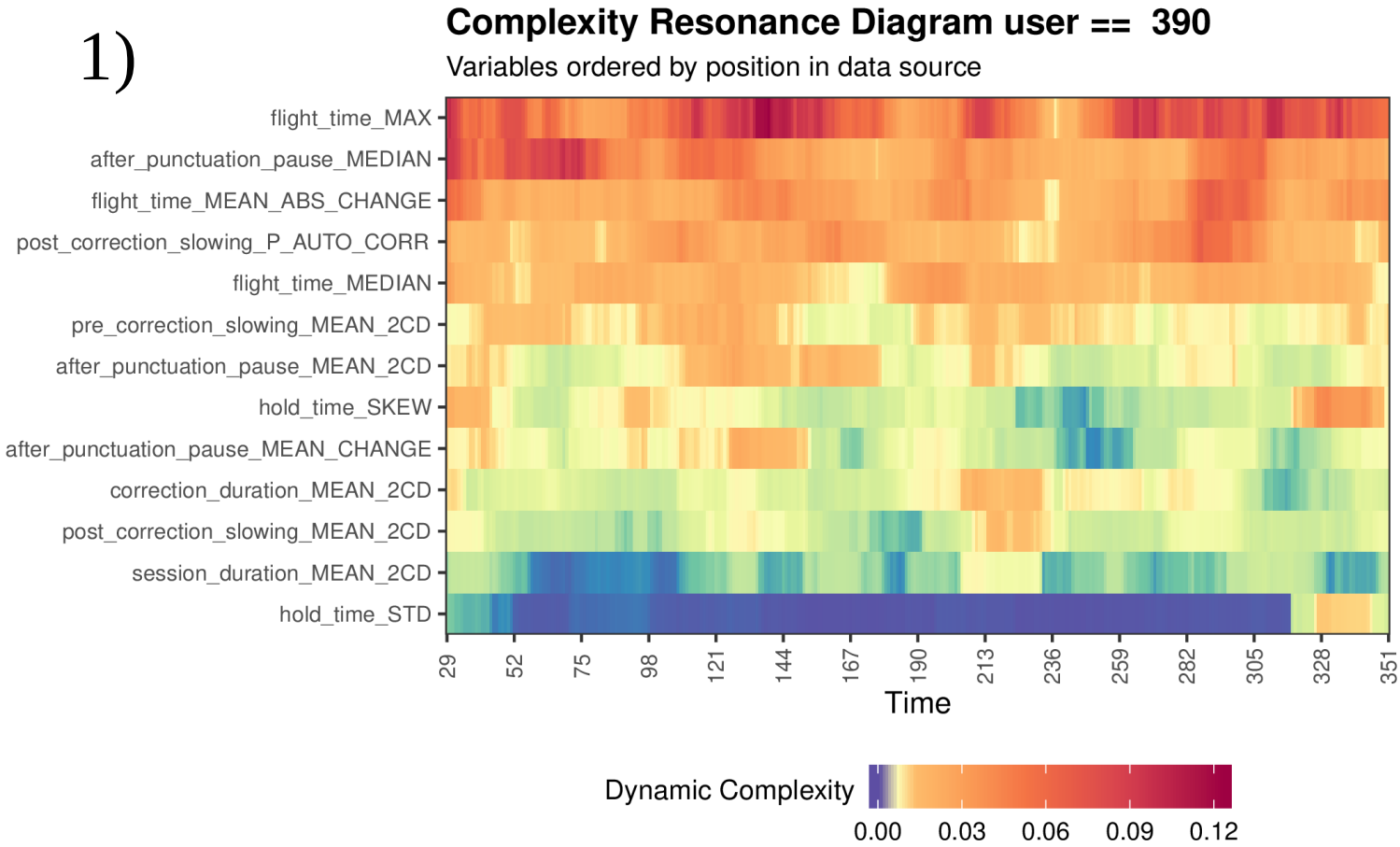
[41] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," Physics reports **438**, 237–329 (2007).
[42] J. F. Donges, J. Heitzig, B. Beronov, M. Wiedermann, J. Runge, Q. Y. Feng, L. Tupikina, V. Stolbova, R. V. Donner, N. Marwan, *et al.*, "Unified functional network and nonlinear time series analysis for complex systems science: The pyunicorn package," Chaos: An Interdisciplinary Journal of Nonlinear Science **25**, 113101 (2015).
[43] S. Van Buuren, *Flexible imputation of missing data* (CRC press, 2018).
[44] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in r," Journal of Statistical Software **45**, 1–67 (2011).
[45] S. V. Buuren, J. P. Brand, C. G. Groothuis-Oudshoorn, and D. B. Rubin, "Fully conditional specification in multivariate imputation," Journal of Statistical Computation and Simulation **76**, 1049–1064 (2006), https://doi.org/10.1080/10629360600810434.
[46] C. Webber and N. Marwan, "Recurrence quantification analysis," Theory and Best Practices (2015).
[47] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control* (John Wiley & Sons, 2015).
[48] U. an der Heiden, "Chaos in health and disease—phenomenology and theory," in *Self-organization and clinical psychology* (Springer, 1992) pp. 55–87.
[49] H. Hermann and J. Haken, "Synergetics," Introduction and Advanced Topics. Berlin–Heidelberg–New York (2004).
[50] G. Schiepek and G. Strunk, "The identification of critical fluctuations and phase transitions in short term and coarse-grained time series—a method for the real-time monitoring of human change processes," Biological cybernetics **102**, 197–207 (2010).
[51] G. K. Schiepek, I. Tominschek, and S. Heinzel, "Self-organization in psychotherapy: testing the synergetic model of change processes," Frontiers in psychology **5**, 1089 (2014).
[52] F. Hasselman, "casnet, a toolbox for studying complex adaptive systems and networks," R package version 0.1 **3** (2018).
[53] L. L. Doove, S. Van Buuren, and E. Dusseldorp, "Recursive partitioning for missing data imputation in the presence of interaction effects," Computational Statistics & Data Analysis **72**, 92–104 (2014).
[54] G. Schiepek, B. Aas, and K. Viol, "The mathematics of psychotherapy–a nonlinear model of change dynamics," Nonlinear Dyn. Psychol. Life Sci **20**, 369–399 (2016).
[55] F. Hasselman, "When the blind curve is finite: dimension estimation and model inference based on empirical waveforms," Frontiers in Physiology **4**, 75 (2013).


press
release
SD_1
FT_1
$PreCS_1$
$PostCS_1$
APP_1
key-press 'character'
key-press 'character'
key-press 'backspace'
key-press 'punctuation'
HT_1
HT_2
CD_1
HT_3
t_i
t_1^p
t_1^r
t_2^p
t_2^r
t_3^p
t_3^r
t_4^p
t_4^r
t_f

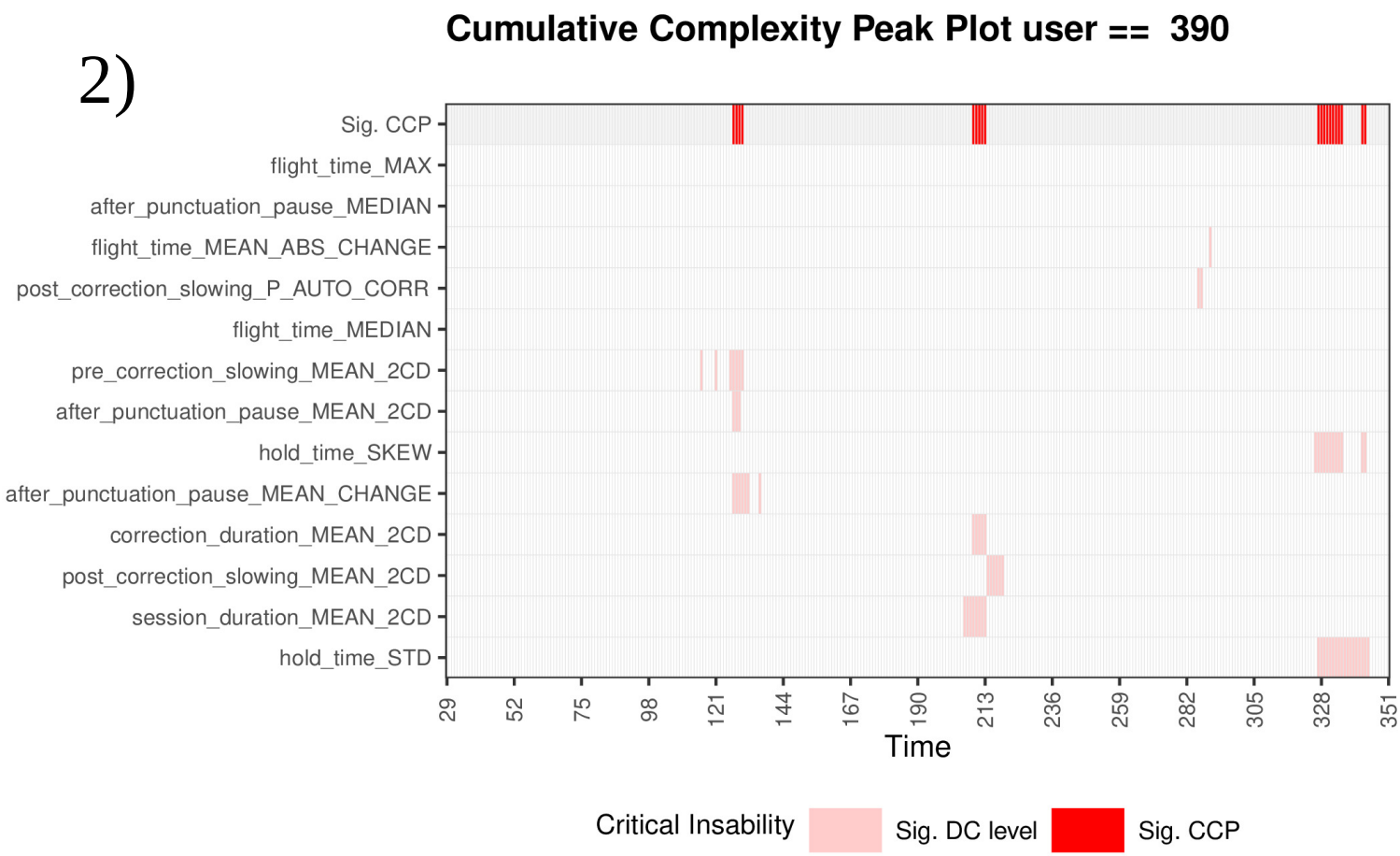

$\tilde{u}(FT_n)$ [ms]
140
120
100
0
50
100
150
200
250
300
350
363
344
325
306
287
268
249
230
211
192
172
153
134
115
96
77
58
39
20
0
0
19
38
57
76
95
114
133
152
171
191
210
229
248
267
286
305
324
343
363

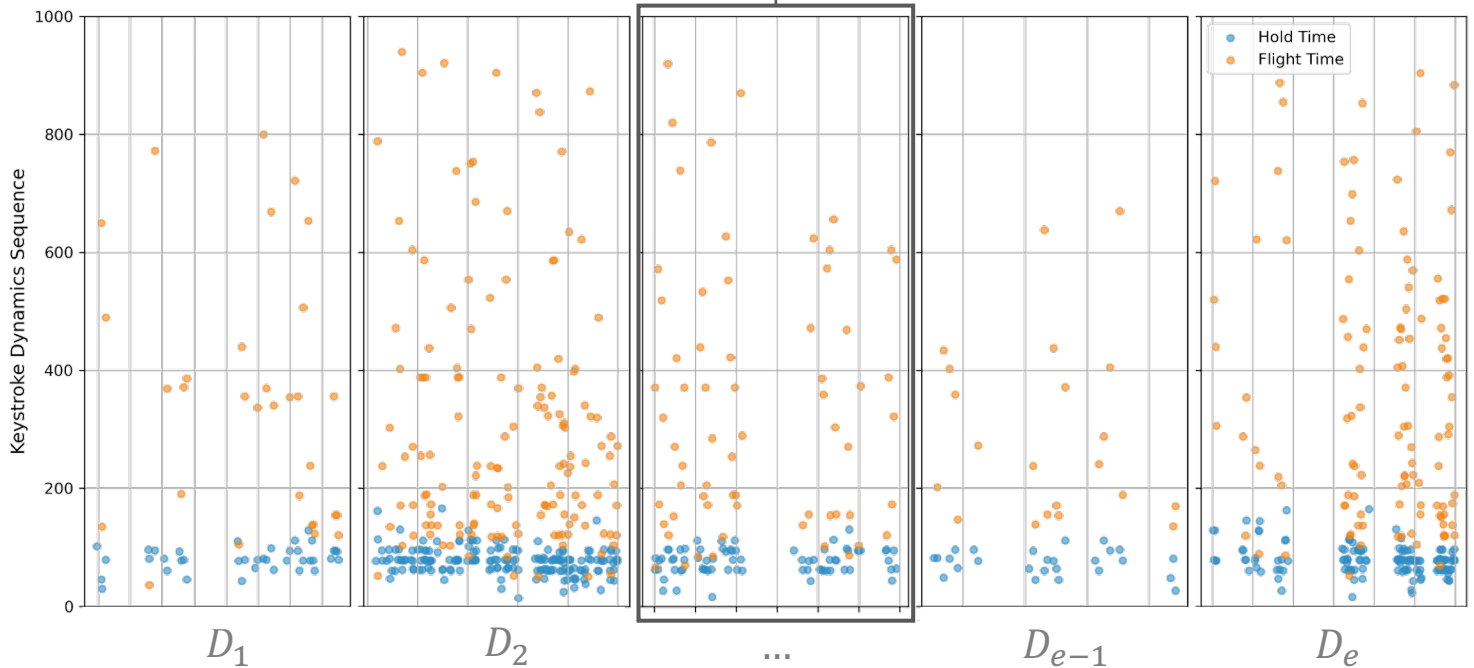

User with MS and changes in MRI
User with MS and no changes in MRI
Healthy Control User
Density
value
session_duration_MEAN_2CD
hold_time_STD
flight_time_MEDIAN
flight_time_MAX

1)

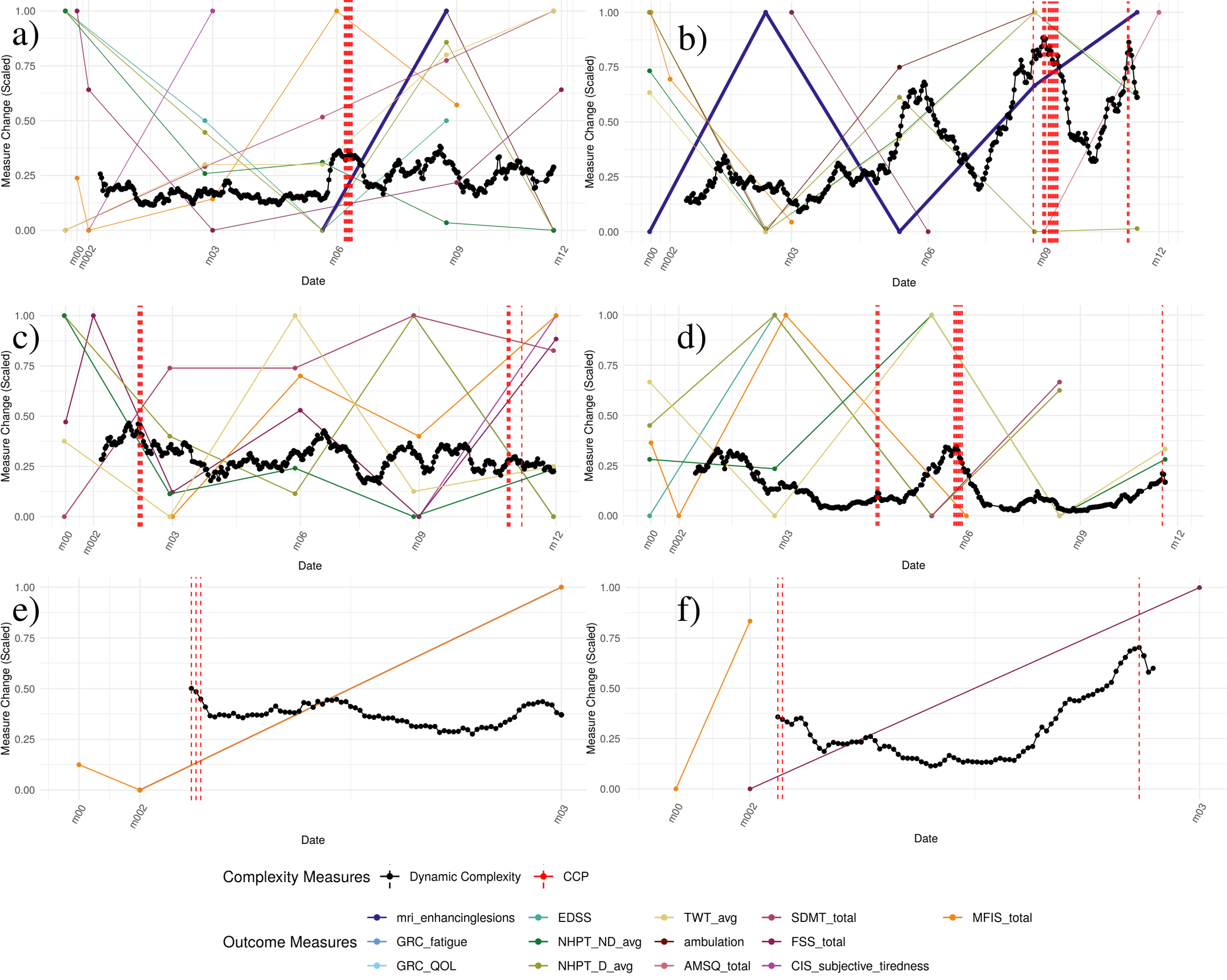

Complexity Resonance Diagram user == 390
Variables ordered by position in data source
flight_time_MAX
after_punctuation_pause_MEDIAN
flight_time_MEAN_ABS_CHANGE
post_correction_slowing_P_AUTO_CORR
flight_time_MEDIAN
pre_correction_slowing_MEAN_2CD
after_punctuation_pause_MEAN_2CD
hold_time_SKEW
after_punctuation_pause_MEAN_CHANGE
correction_duration_MEAN_2CD
post_correction_slowing_MEAN_2CD
session_duration_MEAN_2CD
hold_time_STD
29
52
75
98
121
144
167
190
213
236
259
282
305
328
351
Time
Dynamic Complexity
0.00
0.03
0.06
0.09
0.12

2)


Cumulative Complexity Peak Plot user == 390
Sig. CCP
flight_time_MAX
after_punctuation_pause_MEDIAN
flight_time_MEAN_ABS_CHANGE
post_correction_slowing_P_AUTO_CORR
flight_time_MEDIAN
pre_correction_slowing_MEAN_2CD
after_punctuation_pause_MEAN_2CD
hold_time_SKEW
after_punctuation_pause_MEAN_CHANGE
correction_duration_MEAN_2CD
post_correction_slowing_MEAN_2CD
session_duration_MEAN_2CD
hold_time_STD
29
52
75
98
121
144
167
190
213
236
259
282
305
328
351
Time
Critical Insability
Sig. DC level
Sig. CCP


Keystroke Dynamics Sequence
1000
800
600
400
200
0
Hold Time
Flight Time
D_1
D_2
...
D_{e-1}
D_e

look-up table
x_b
x_f
$x^p = x_b \cup x_f$
Imputation
Γ(·)
Feature selection
Scaling
F(·)
D(·)
C(·)
CCP


a)
b)
c)
d)
e)
f)
Measure Change (Scaled)
1.00
0.75
0.50
0.25
0.00
m00
m002
m03
m06
m09
m12
Date
Complexity Measures
Dynamic Complexity
CCP
Outcome Measures
mri_enhancinglesions
GRC_fatigue
GRC_QOL
EDSS
NHPT_ND_avg
NHPT_D_avg
TWT_avg
ambulation
AMSQ_total
SDMT_total
FSS_total
CIS_subjective_tiredness
MFIS_total