Systematic assessment of intrinsic factors influencing visual attention performances in air traffic control via clustering algorithm and statistical inference

Jing-Qiang Li[1,2]*, Hong-Yan Zhang[3¤]*, Yan Zhang[3], Hai-Tao Liu[4]

1 Research Institute of Civil Aviation Safety, Civil Aviation University of China, Tianjin 300300, P. R. China, **2** National Key Laboratory of Air Traffic Operation Safety Technology, Civil Aviation University of China, Tianjin 300300, P. R. China, **3** Sino-European Institute of Aviation Engineering, Civil Aviation University of China, Tianjin 300300, P. R. China, **4** College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, P. R. China

¤ Current address: Department of Computer Science, School of Information Science and Technology, Hainan Normal University, Haikou 571158, P. R. China
* jingqiang_li@foxmail.com (JQL); hongyan.rd@foxmail.com (HYZ)

Abstract

The intrinsic factors (IF) influencing visual attention performance (VAP) might cause potential human errors, such as "error/mistake", "forgetting" and "omission". It is a key issue to develop a systematic assessment of IF in order to distinguish the levels of VAP. Motivated by the Stimulus-Response (S-R) model, we take an interactive cancellation test—Neuron Type Test (NTT)—to explore the IF and present the corresponding systematic assessment. The main contributions of this work include three elements: a) modeling the IF on account of attention span, attention stability, distribution-shift of attention with measurable parameters by combining the psychological and statistical concepts; b) proposing quantitative analysis methods for assessing the IF via its computational representation—intrinsic qualities (IQ)—in the sense of computational model; and c) clustering the IQ of air traffic control (ATC) students in the feature space of interest. The response sequences of participants collected with the NTT system are characterized by three parameters: Hurst exponent, normalized number of decisions (NNoD) and error rate of decisions (ERD). The K-means clustering is applied to partition the feature space constructed from practical data of VAP. For the distinguishable clusters, the statistical inference is utilized to refine the assessment of IF. Our comprehensive analysis shows that the IQ can be classified into four levels, i.e., *excellent, good, moderate* and *unqualified*, which has a potential application in selecting air traffic controllers subject to reducing the risk of the inadequacy of attention performances in aviation safety management.

Nomenclature

For the convience of writing and reading, we summarize the abbreviations and the corresponding full names adopted in this paper in Table 1.

Introduction

Backgrounds and motivations

Human error is a major source of air traffic control (ATC) accidents or incidents [1–6]. Luo [7] showed that 74.3% of accidents related to ATC of China are caused by air traffic controllers, or controllers for simplicity, instead of ATC equipments. The Aviation Safety Reporting System (ASRS) indicated that the aviation accidents are closely related with the *visual attention performance* (VAP) [8]. The objective of the controller is to resolve flight conflicts and ensure flight safety management subject to dealing with a large amount of visual and auditory information and making decisions *accurately*, *rapidly* and *stably* [9–11]. When there is hazardous weather, communication failures and/or aircraft mechanical failures, controllers will commit unsafe acts which lead to human errors, like "error/mistake", "forgetting" and "omission". It is necessary for controllers to concentrate and allocate attention so as to avoid flight accidents or incidents [12, 13]. The history of ATC shows that controllers' misidentification may cause severe aviation accidents or incidents. On 11th October 2016, a serious runway incursion incident during takeoff occurred because of the controller's erroneous operations in Shanghai Hongqiao International Airport of China [14, 15]. This incident was due to controller's forgetting runway state and commanding mistake. It should be pointed out that there are two main factors for this incident: *lapse of attention* and *lack of expertise as a novice in ATC*.

A key issue for the cause of operational errors by novice controllers is the inadequate VAP [16]. According to the applied attention theory [17], visual misidentification may occur due to individual factors and attentional slips. Visual misidentification usually engenders errors in visual detecting and searching. There are quite a number of references about human perception processing [18–23], visual searching and cognition mechanism [24–27], which emphasize the mechanism of attention and play a fundamental role for human perception, cognition and action. There are two categories of factors influencing VAP: *extrinsic factors* (EF), which refer to the impersonal factors that affect attention or the factors that are not related to individual and affect attention, and *intrinsic factors* (IF), which refer to the personal-related factors that affect attention. Our emphasis is put on the IF influencing VAP and their systematic assessment. The inadequacy of VAP may cause serious human errors in civil aviation. The descriptions of the EF influencing VAP are abundant such as budget of search, complexity of task, noise or dark environment [21, 22]. However, the studies on the relation of human errors and

Table 1. Nomenclature: Abbreviations vs. full names.

Abbr.	Full Name	Abbr.	Full Name
ATC	Air Traffic Control	NoD	Number of Decisions
ASRS	Aviation Safety Reporting System	NNoD	Normalized NoD
EF	Extrinsic Factors	NTT	Neuron Type Test
ERD	Error Rate of Decisions	QAR	Quick Access Recorder
EVD	EigenValue Decomposition	R/S	Range/Standard deviation
IF	Intrinsic Factors	S-R	Stimulus-Response
IQ	Intrinsic Qualities	SSE	Sum of the Squared Errors
I/O	Input/Output	TLD	TrLogDet
ITM	Information Theoretic Metric	VAP	Visual Attention Performance

https://doi.org/10.1371/journal.pone.0205334.t001

the IF influencing VAP, such as attention span, attention stability, and distribution-shift of attention [20, 28–30], are deficient. Moreover, these IF are conceptual and there is no quantitative method to analyze them mathematically. It is necessary to analyze the VAP quantitatively by converting the conceptual IF into the computational counterparts, i.e., *intrinsic qualities* (IQ).

Intuitively and empirically, the controllers can be classified reasonably into four levels, i.e., excellent, good, moderate and unqualified. Theoretically, this classification can be implemented by clustering the features constructed from the IQ and evaluated with statistical inference. The determination of specific level which the control belongs to has a potential application in selecting air traffic controllers subject to detecting human errors and reducing the operation risk.

Visual attention performances and clustering

It is feasible to analyze the VAP quantitatively with the help of our Neuron Type Test (NTT) system [31] and data collected (see S1 Appendix). Once the response sequences of ATC students are collected with NTT, we can define, compute and analyze the VAP by modeling the IF with mathematical methods. Technically, we take a three-step strategy to attain this objective:

1. **Modeling**: Based-on the response sequences, the feature space is constructed with measurable parameters proposed for IF with psychological experiments, mathematical and statistical theories.

2. **Clustering**: The ATC students characterized by the 2-dim or 3-dim features are classified via the *K*-means clustering algorithm automatically.

3. **Assessing**: Different clusters are evaluated by statistical inference in order to control the risk of human errors in aviation safety management by selecting potential professionals of ATC.

In civil aviation, the clustering method has been successfully applied in intelligent classification of air traffic controllers [32] and aeronautical information network intrusion detection [33]. On account of worldwide aircraft accidents between 1950s and 2010s, Christopher [34] and Olja [35] adopted the clustering method to predict the warning level of aircraft components and constructed an accident predictive models respectively. Normal and abnormal operations can be identified by trajectory clustering [36], which helps to surveillance the airspace efficiently. In 2012, main factors about approaching and the quick access recorder (QAR) data were analyzed by Zhou [37] with gray clustering method for assessing aircrew and improving flight safety. Vitali [38] studied the statistical regularities in ATM. In 2013, Monechi [39] analyzed the interrelation of safety data with flight trajectories and network metrics. In 2017, anomalous flight operations are identified from energy-based metrics and clustering [40] and Li et al. [41–43] studied the clustering and classification of IQ influencing the VAP of controllers.

With the purpose of controlling human errors and enhancing the safety performance in aviation operations, not only should we monitor the abnormal operations in time, but also assess the attention of operators efficaciously at the same time. The history of aviation safety shows that it is vital to pay more attention on operators instead of equipments [44]. In this paper, we propose systematic assessment on individual's IF via the computational IQ by analyzing the response sequence with R/S analysis [31], *K*-means clustering algorithm [45] and statistic inference so as to identify the relation of human error and information processing

level specified by the IF. We deem that the predictive quantification assessment for human errors caused by inadequacy of attention qualities could be used for selecting controller candidates and risk management.

Roadmap, contents and organization

For the purpose of quantifying the essential differences among controllers, a group of participants—air traffic control students—are invited to take two consecutive 12 minutes selective attention test with the "Ruleout-Ignore-Eliminate" rules. The experiments were carried with the help of the NTT system. The participants' IF are coded in the response sequences collected in the experiment. Motivated by psychology and data science, we defined 2-dim and 3-dim feature spaces clustering the participants and assessing the IF with K-means clustering algorithm and statistical inference. The main ideas, logics and contents framework of this work are depicted in Fig 1.

The contents of this paper are organized as follows: the mathematical preliminaries are presented briefly in Section 2; the computational representation of IF in the sense of attention theory and data science are discussed in Section 3; the interpretation of the experiments about intrinsic factors with NTT are explained in Section 4; the intrinsic qualities are explained with the help of 2-dim and 3-dim clustering results in Section 5; finally the conclusions are given in Section 6. In the appendices, we give some introductions to the NTT system and K-means clustering algorithm.

Mathematical preliminaries

R/S analysis and Hurst exponent

In [31], the R/S analysis is used to analyze the Hurst exponent [46–48] of the reaction time data collected from the NTT system. For a discrete reaction time sequence $\{X_k\}_{k=0}^{L-1}$ with length $\ell = d \times n$, it can be segmented into d subsequences in which the m-th subsequence is $\left\{X_i^{(m)}\right\}_{i=0}^{n-1}$ where $m \in \{0, \cdots, d-1\}$ and

$$X_i^{(m)} = X_{mn+i}, \quad m \in \{0, \cdots, d-1\}. \tag{1}$$

We can compute the local average E_m, bias $Y_i^{(m)}$, cumulative deviate $Z_i^{(m)}$, range $\mathcal{R}_n^{(m)}$ and standard deviation $\mathcal{S}_n^{(m)}$ respectively as follows:

$$E_m = \frac{1}{n} \sum_{i=0}^{n-1} X_i^{(m)}, \quad m \in \{0, \cdots, d-1\} \tag{2}$$

$$Y_i^{(m)} = X_i^{(m)} - E_m, \quad i \in \{0, \cdots, n-1\} \tag{3}$$

$$Z_i^{(m)} = \sum_{j=0}^{i} Y_j^{(m)}, \quad i \in \{0, \cdots, n-1\} \tag{4}$$

$$\mathcal{R}_n^{(m)} = \max_{0 \leq i \leq n-1} Z_i^{(m)} - \min_{0 \leq i \leq n-1} Z_i^{(m)} \tag{5}$$

$$\mathcal{S}_n^{(m)} = \sqrt{\frac{1}{n-1} \sum_{i=0}^{n-1} \left[X_i^{(m)} - E_m\right]^2} \tag{6}$$
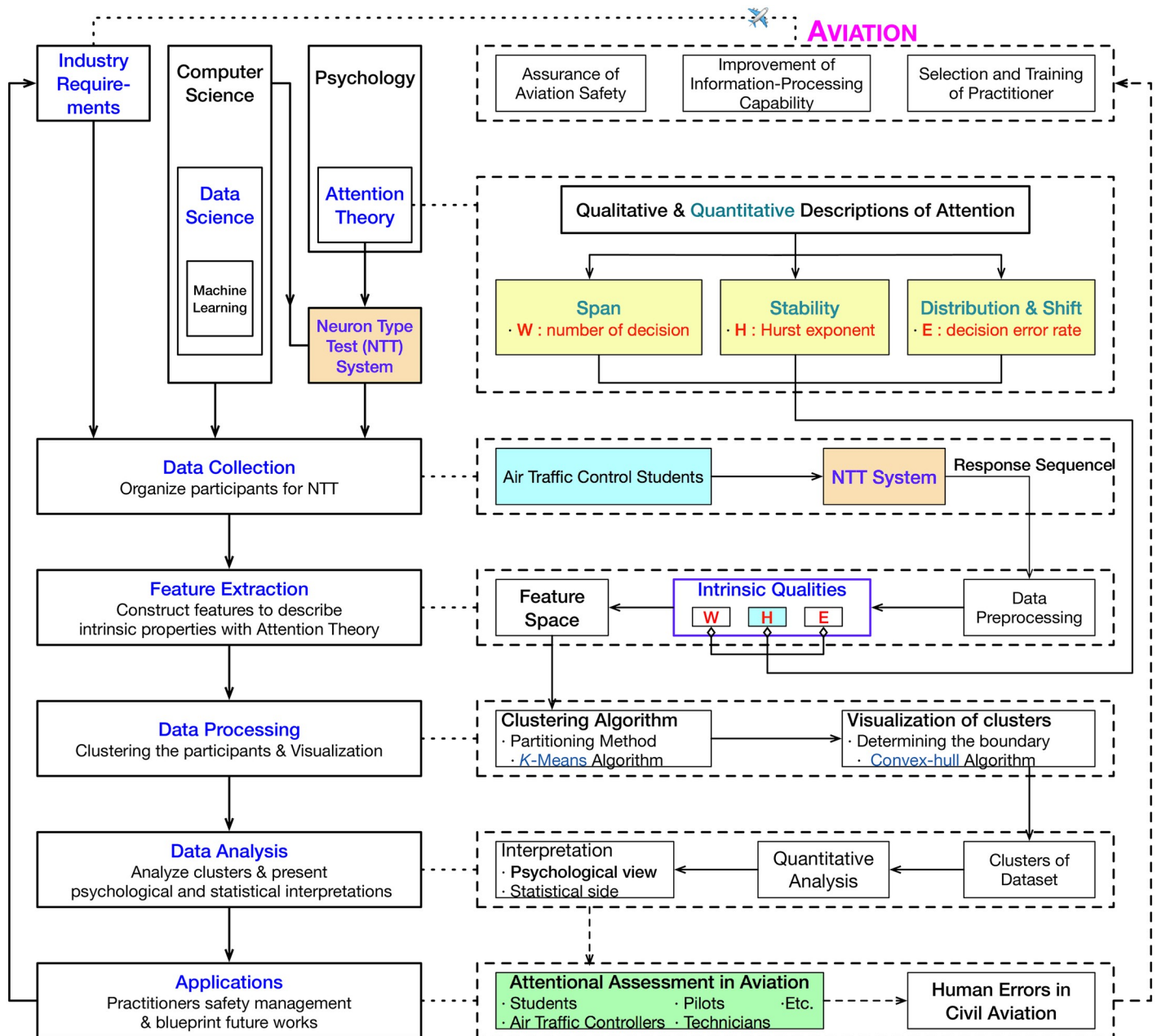
Fig 1. Aviation safety, visual attention theory and clustering algorithm: Logical hierarchy and technical implementation strategy.

https://doi.org/10.1371/journal.pone.0205334.g001

Thus the average of the rescaled range for all subsequences of length *n* will be

$$(R/S)_n = \frac{1}{d}\sum_{m=0}^{d-1} \frac{\mathcal{R}_n^{(m)}}{\mathcal{S}_n^{(m)}} \tag{7}$$

In [47], Mandelbrot showed that the R/S statistics asymptotically follows the relation:

$$(R/S)_n \sim c \cdot n^H. \tag{8}$$

Hence the Hurst exponent H can be estimated via least square approach according to (8). Usually, $H \in (0, 1)$. For the reaction time sequences in NTT experiment, [31] shows that $H > 0.5$ means the process is persistent, $H = 0.5$ means that the process is near random, and $H < 0.5$ implies that the process is anti-persistent.

Features and clusters

n-dim feature space and partition. The n-dim feature space is denoted by

$$S^n = \left\{ \boldsymbol{x} = [x^1, \cdots, x^n]^\top : x^i \in R_i \subset \mathbb{R} \right\} \tag{9}$$

where the component x^i lies in the set R_i for each feature $\boldsymbol{x}$, and each R_i may be discrete or continuous. For any two features $\boldsymbol{x}$ and $\boldsymbol{y}$ in S^n, we use the Euclidean distance $\|\boldsymbol{x} - \boldsymbol{y}\|$ to measure their difference although there are plenty of candidates of metric.

If the magnitudes of components for a feature point span a large scale, then we normalize each component in order to balance the contribution of each component. In this paper, we take the following min-max normalization step:

$$x_{new}^j = \frac{x^j - x_{min}}{x_{max}^j - x_{min}^j}, \quad x_{new}^j \in [0, 1], x^j \in R_j \tag{10}$$

where

$$x_{min}^j = \min_{x^j \in R_j} x^j, \quad x_{max}^j = \max_{x^j \in R_j} x^j. \tag{11}$$

For simplicity, we will use the notation $\boldsymbol{x}$ for both the primitive feature $\boldsymbol{x}$ and its normalized version $\boldsymbol{x}_{new}$. For the normalized feature space, we have $R_i = [0, 1]$ and each feature is located in an n-dim unit cube

$$\mathcal{C}^n = \underbrace{[0, 1] \times \cdots \times [0, 1]}_{n \text{ terms}} = \left\{ \boldsymbol{x} = [x^1, \cdots, x^n]^\top : 0 \leq x^i \leq 1 \right\}. \tag{12}$$

For the index set $\{1, \cdots, N\}$, it can be partioned into K disjoint subsets $\Phi_1, \cdots, \Phi_K$ such that $\{1, 2, \cdots, N\} = \cup_{k=1}^{K} \Phi_k$ and $\Phi_j \cap \Phi_i = \emptyset$ for $i \neq j$. Let $|A|$ be the cardinality of set A and $|\Phi_k| = N_k$, then we have $|\cup_{k=1}^{K} \Phi_k| = N$. Similarly, for the data set

$$S^n = \left\{ \boldsymbol{x}_k = \left[x_k^1, \cdots, x_k^j, \cdots, x_k^n \right]^\top : k \in \{1, \cdots, N\} \right\} = \{\boldsymbol{x}_k\}_{k=1}^N, \tag{13}$$

which consists of N discrete n-dim features, we can divided it into K disjoint subsets $S_1^n, \cdots, S_K^n$ as follows

$$S_k^n = \left\{ \boldsymbol{x}_{i_k} = \left[x_{i_k}^1, \cdots, x_{i_k}^n \right]^\top \in S^n : i_k \in \Phi_k \right\}, \quad k \in \{1, \cdots, K\}. \tag{14}$$

In consequence,

$$S^n = \bigcup_{k=1}^{K} S_k^n, \quad S_j^n \cap S_k^n = \emptyset \quad \text{for} \quad j \neq k. \tag{15}$$

and $|S| = \sum_{k=1}^{K} |S_k^n| = N$. For the purpose of clustering, we use the disjoint sets $S_1^n, \cdots, S_K^n$ to denote the K different clusters.

For each feature point $\boldsymbol{x}_j$ in $S^n = \bigcup_k S_k^n$, we use the binary indicator variable I_{kj}, known as the 1-of-K coding scheme [49],

$$I_{jk} = I_{S_k^n}(\boldsymbol{x}_j) = \begin{cases} 1, & \boldsymbol{x}_j \in S_k^n; \\ 0, & \boldsymbol{x}_j \notin S_k^n \end{cases} \tag{16}$$

to label how to assign the point $\boldsymbol{x}_j$ into the cluster of interest. If $\boldsymbol{x}_j$ is assigned to the k-th cluster S_k^n, i.e., $\boldsymbol{x}_j \in S_k^n$, then $I_{jk} = 1$, otherwise $I_{jk} = 0$.

Statistics and metrics for clusters. The centroid of a given cluster S_k^n is the average of the features in the cluster S_k^n, i.e.,

$$\boldsymbol{\mu}_k = \boldsymbol{\mu}(S_k^n) = \frac{1}{|S_k^n|} \sum_{\boldsymbol{x} \in S_k^n} \boldsymbol{x}. \tag{17}$$

For the features $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ in feature space S^n, their dissimilarity is measured by their Euclidean distance. viz.,

$$\mathrm{dist}_E(\boldsymbol{x}_i, \boldsymbol{x}_j) = \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \tag{18}$$

For clusters S_k^n and S_i^n, we use the Euclidean distance of their corresponding centroids to represent their distance

$$\mathrm{dist}_E(S_k^n, S_i^n) = \|\boldsymbol{\mu}_k - \boldsymbol{\mu}_i\|, \quad \forall i, k \in \{1, \cdots, N\}. \tag{19}$$

Without doubt, the distance of clusters is an inter-class metric which describes the degree of separation of two clusters. For each cluster S_k^n, the standard deviation of features is defined by

$$\sigma_k = \sigma(S_k^n) = \sqrt{\frac{1}{|S_k^n| - 1} \sum_{\boldsymbol{x} \in S_k^n} \|\boldsymbol{x} - \boldsymbol{\mu}_k\|^2}. \tag{20}$$

Note that we use the unbiased estimation of variation since $N_k = |\Phi_k|$ may not be efficiently large. Intuitively, the smaller the standard deviation is, the better the cluster is. The standard deviation per dimension is defined by

$$\bar{\sigma}_k = \frac{\sigma(S_k^n)}{n}, \tag{21}$$

which can be used to compare the clusters with different dimensions when the impact of the dimension of features is considered in the practical problems.

Although the standard deviation is simple and useful, it is difficult to describe the orientation of the features in a cluster. It is necessary for us to consider the (unbiased) estimation of the covariation matrix of S_k^n, i.e.,

$$\boldsymbol{\Sigma}_k^n = \mathrm{Cov}(\boldsymbol{x}_i I_{S_k^n}, \boldsymbol{x}_j I_{S_k^n}) = \frac{1}{|S_k^n| - 1} \sum_{\boldsymbol{x} \in S_k^n} (\boldsymbol{x} - \boldsymbol{\mu}_k)(\boldsymbol{x} - \boldsymbol{\mu}_k)^\top. \tag{22}$$

For the features $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ in S^n, their correlation coefficient is defined by

$$\mathrm{Cr}\left(\boldsymbol{x}_i I_{S_k^n}, \boldsymbol{x}_j I_{S_k^n}\right) = \frac{\mathrm{Cov}(\boldsymbol{x}_i I_{S_k^n}, \boldsymbol{x}_j I_{S_k^n})}{\sqrt{\mathrm{Cov}(\boldsymbol{x}_i I_{S_k^n}, \boldsymbol{x}_i I_{S_k^n})} \cdot \sqrt{\mathrm{Cov}(\boldsymbol{x}_j I_{S_k^n}, \boldsymbol{x}_j I_{S_k^n})}}. \tag{23}$$

For the purpose of comparing the difference of clusters, we define the *TrLogDet* (TLD) divergence for the clusters S_k^n and S_i^n by

$$\text{div}_T(S_k^n, S_i^n) = \frac{1}{n}\left[\text{tr}\left(\mathbf{\Sigma}_k \mathbf{\Sigma}_i^{-1}\right) - \log \det\left(\mathbf{\Sigma}_k \mathbf{\Sigma}_i^{-1}\right) - n\right]. \tag{24}$$

Then we introduce the *TrLogDet* (TLD) distance

$$\text{dist}_T(S_k^n, S_i^n) = \frac{1}{2}\left[\text{div}_T(S_k^n, S_i^n) + \text{div}_T(S_i^n, S_k^n)\right]. \tag{25}$$

Note that the expression for TLD in (24) is different from LogDet in [50, 51] since the factor $1/n$ is introduced for the purpose of eliminating the impact of the dimension of the data and the trace operation is emphasized here. It is easy to show that the TrLogDet distance satisfies the axioms of distance:

- (non-negative) $\text{dist}_T(S_k^n, S_i^n) \geq 0$ and $\text{dist}_T(S_k^n, S_k^n) = 0$;

- (symmetric) $\text{dist}_T(S_k^n, S_i^n) = \text{dist}_T(S_i^n, S_k^n)$;

- (triangle inequality) $\text{dist}_T(S_k^n, S_i^n) + \text{dist}_T(S_i^n, S_j^n) \geq \text{dist}_T(S_k^n, S_j^n)$.

The TLD distance measures the dissimilarity of two clusters and it is a metric of inter-clusters.

For the clusters S_k^n and S_i^n, when $\mathbf{\Sigma}_i = \text{Cov}(S_i^n) = \mathbf{I}$ is the identity matrix, namely an "ideal" or "standard" cluster, we immediately have

$$\rho_k = \rho(S_k^n) = \frac{1}{n}\left[\text{tr}\left(\mathbf{\Sigma}_k\right) - \log \det\left(\mathbf{\Sigma}_k\right) - n\right] = -1 + \frac{1}{n}\sum_{j=1}^{n}(\lambda_k^j - \log\lambda_k^j). \tag{26}$$

In this case, $\rho(S_k^n)$ is the so called *Information-Theoretic Metric* (ITM) [50, 51] per dimension, which is a positive, strictly convex and spectral function. Particularly, if $\mathbf{\Sigma}_k = \mathbf{I} = \text{diag}(1, \cdots, 1)$, then $\rho_k = 0$. Usually, the ITM ρ_k can be used as a divergence metric for the cluster S_k^n, just as the standard deviation σ_k does.

The choice of K in the K-means algorithm is a key issue in clustering. The value of K has direct impact on the clustering results, which is related to the *sum of the squared errors* (SSE) for K clusters:

$$\text{SSE} = \sum_{k=1}^{K}\text{SSE}\left(S_k^n\right) = \sum_{k=1}^{K}\sum_{\mathbf{x}\in S_k^n}\|\mathbf{x} - \mathbf{\mu}_k\|^2 \tag{27}$$

where $\text{SSE}\left(S_k^n\right)$ is the sum of squared errors for the k-th cluster S_k^n.

Modeling the intrinsic factors influencing visual attention performances

Conceptions

Logically, there are three key issues to be done for quantitatively assessing the IF influencing VAP:

1. proposing an abstract description for IF by mathematical modeling;

2. developing an interactive system to collect data which preserve the intrinsic and invariant properties interested naturally;

3. finding proper ways to analyze the data collected and give reliable, effective and practical assessment methods for the IF.

Conceptually, a reasonable data collecting system for analyzing the IF should have the following elements:

- STRUCTURE: the stimulation paradigm should be an Stimulus-Response (S-R) model, which can be represented by an Input/Output (I/O) model theoretically;

  - INPUT: non-semantic symbols serve as visual stimulus;

  - OUTPUT: the response of participant is the decision which includes reaction time and the symbolized description of correctness;

- FUNCTION: the data collected with the system can be used to analyze the attention span, stability of attention, distribution-shift of attention for individuals;

- INTERFACE: the human-machine interaction interface should be friendly and easy to be used for the users.

According to these elements, the NTT system has been well developed [31] and used to collect the response sequence of individuals, in which the i-th response $\langle y_i, \tau_i \rangle_p = \langle y_i^p, \tau_i^p \rangle$ for the p-th individual consists of two components:

- symbolic decision, denoted by y_i such that $y_i \in \{Y, N, I\}$ where Y, N and I stand for "Yes", "No" and "Ignore" respectively;

- reaction time, denoted by τ_i such that $\tau_i > 0$.

The length of the response sequence refers to the count of responses in a time interval specified by the experimental setup. Usually, the length is different for different individuals and it can be regarded as a random variable mathematically.

Modeling the intrinsic factors: Qualitative and quantitative descriptions

The performance of attention is affected by a variety of factors, which can be divided into two fundamental categories [29, 30]: extrinsic factors and intrinsic factors. Extrinsic factors include mission objectives, plans, operating procedures and so on. Alternatively, the intrinsic factors include knowledge, skills, experience, personal attention level and so on.

In terms of human errors in aviation, the *rapidness*, *accuracy*, and *stability* of the behaviors of operators are essential for avoiding incidents or accidents in case of emergencies. There are two famous cases in civil aviation in the recent decade: the forced landing of New York National Airline on Hudson River on 15 January 2009 (engine failed) [52] and the abdomen landing of Polish Airline (landing gear failed to lay down) on 1 November 2011. These three characteristics are closely related to the IF influencing VAP for aircraft operators.

Currently, there is a lack of a complete test approach for attention performances, thus we can only assess some aspects of human behaviors characterized by some components of attention performances [53]. In the sense of psychology, the interests of the assessment of IF in this work consist of three aspects: attention span, attention stability, distribution-shift of attention. With the response sequence collected by the NTT system, it is feasible to measure and assess the IF with observable data.

For the convience of understanding the ideas and methods for modeling the IF influencing VAP and the corresponding IQ of interest, we figure out a block diagram for VAP, see Fig 2.
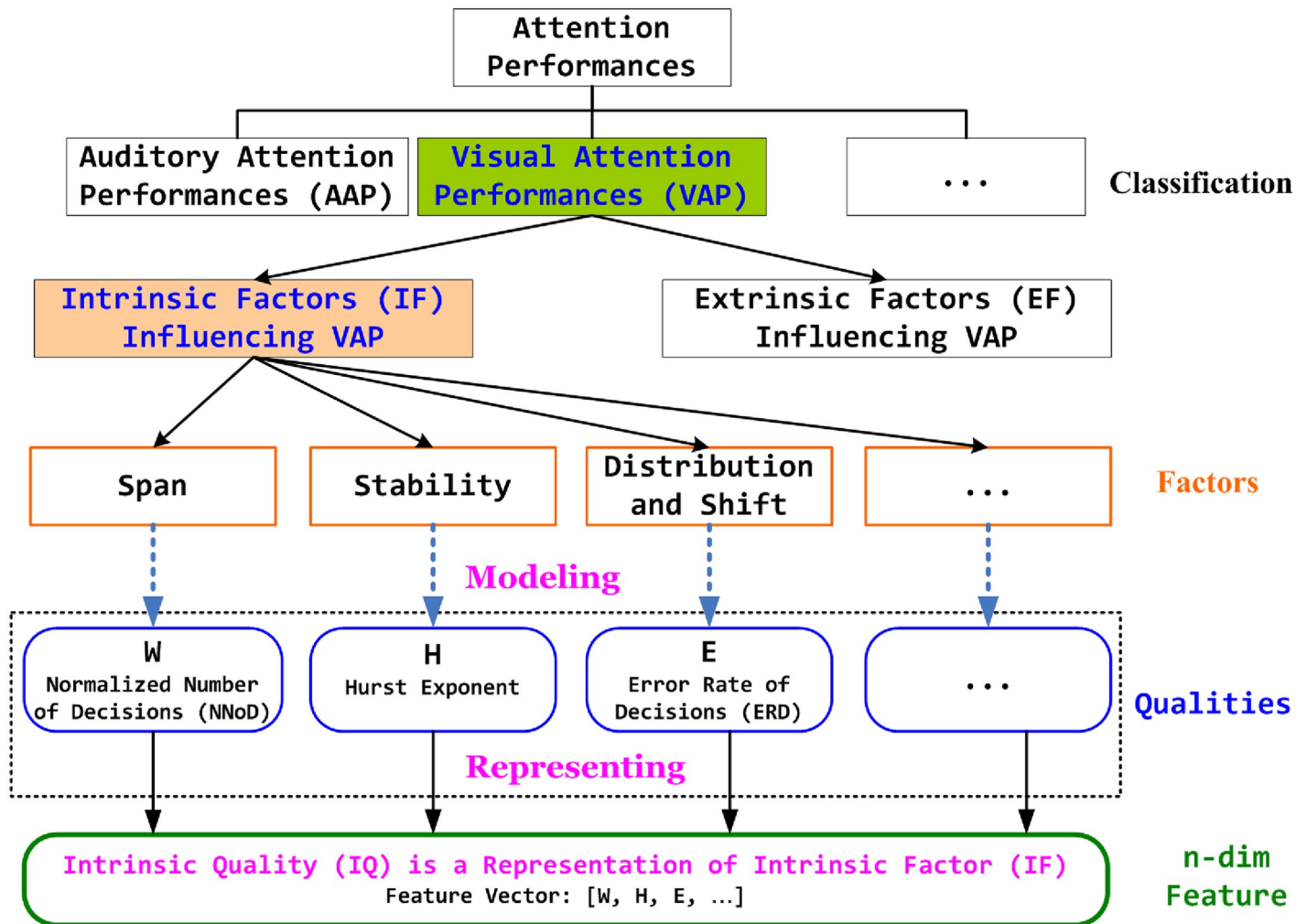
Fig 2. Visual attention performances: From concepts to feature vectors.

Attention span—Number of decisions and its normalization. Attention span is the amount of concentrated time one can spend on a task without becoming distracted. If the concentrated time interval is fixed, then attention span refers to the amount of perceived or recognized objects during a given period for an individual. Intuitively, the more decisions the participant makes in each test via NTT System, the faster the participant responses in a given time interval ($2 \times 12 = 24$ minutes). Therefore, the *number of decisions* (NoD) made in the fixed time can be used as a data feature to describe the operational capacity of a single participant. Attention span relates to the individual's information processing intensity, which determines the number of the objects perceived. NoD maps the number of percepts and furthermore reflect the degree of rapidness which serves as the assessment index for attention span.

Mathematically, the NoD is denoted by I and its value must be a positive integer, i.e., $I \in \mathbb{N}$. In our NTT test, we have $I \in [200, 800]$ according to the setup of the practical experiments of NTT. For the sake of data processing and pre-conditioning in data science, we take the following *normalized number of decisions* (NNoD) with min-max normalization step according to

(10) and (11)

$$W = \frac{I - I_{min}}{I_{max} - I_{min}} \in [0, 1]. \tag{28}$$

Attention stability—Hurst exponent. Stable attention relates to the goal of the activity, the level of understanding of the task, the interest on the object, the degree of motivation to engage in the activity, emotional state, activeness of thought and healthy state. Attention stability describes the individual's quality of keeping attention on a certain object or an activity for a long duration [54]. In civil aviation, the attention stability reflects ATM controllers' ability for tracking aircrafts among sectors and/or flights en rout continuously and smoothly.

Based on the authors' previous work [31], the Hurst exponent [55], denoted by H, is a good feature for the reaction time data such that the latent information and long term correlation can be characterized efficiently. *The Hurst exponent H is a natural and intrinsic parameter for representing the level of attention stability.* In a heuristic way, there are two aspects which should be noted:

- The available methods for describing attention stability are based on average, range of reaction time and the standard deviation independently [56]. On the contrary, the Hurst exponent H integrates all of these factors into a single and simple statistical invariant in the sense of affine transformation by its definition.

- This statement is a fundamental hypothesis and its effectiveness should be verified and validated by practical cases and possible applications. In the sense of mathematical modeling of IF, it is an abstraction for describing the attention stability with Hurst exponent H. For our response sequences obtained by NTT system based on Stimulus-Response model, this hypothesis is acceptable and the abstraction leads to interesting and reasonable results when clustering with experimental data.

Distribution-shift of attention—Error rate of decisions. Distribution of attention indicates distributing attention on distinct objects simultaneously, which reflects the capability of processing multiple tasks parallelly. Shift of attention refers to transferring the attention intentionally from one object, or an activity, or an operation, to another.

"Error" is usually a conscious behavior of individual. It is due to the lack of practitioners' ability, behavioral fault, unfamiliarity with the rules of operation, the lack of relevant knowledge and other factors caused by operational errors or fault. The job of air traffic controller is to monitor, control and service the flight activities of aircrafts by using communication, navigation, monitoring techniques and methods to prevent aircraft collision and guarantee the safety, which covers the different stages of perception of information, analysis, decision-making and implementation. During this information processing, the controller's memory and attention play the dominate roles in the accuracy of his operation. *Error*, *omission*, and *forgetting*—three typical forms of human mistakes—are caused by the lapse of distribution-shift of attention. We use the *error rate of decisions* (ERD), denoted by E, to represent the correctness of decisions which indicates the distribution-shift of attention. In the sense of aviation operation, ERD is of significance for the air traffic controllers. For the tests with NTT System, it is defined by

$$E = \frac{\text{Number of Incorrect Decisions}}{\text{Number of Decisions}} \times 100\% \in [0, 1]. \tag{29}$$

Representation of IF and feature space

The representation of IF is IQ, which leads to the feature space with high dimensional features generally, please see Fig 2.

2-dim feature space. In practical problems, some features of data may not be easily acquired. Hence, the analysis for low dimensional sample space may be indispensable. On the basis of the modeling of IF, we have obtained three features and constructed interested 2-dim and 3-dim feature spaces.

For the 2-dim feature space, we use the Hurst exponent H and the NNoD W to construct the feature space which consists of points $[W, H]^{\mathsf{T}}$ such that $H \in [0, 1]$ according to [31] and $W \in [0, 1]$ by normalization. We should remark it is not a wise choice if we use the NoD I instead of the NNoD W. Actually, the order of magnitudes of H and I are 10^{-1} and 10^2 respectively, and the weight for the two components of each feature vector $[I, H]^{\mathsf{T}}$ are the same. In consequence, the Euclidean distance of the feature vectors is dominated by the NoD I. To avoid such a result, it is necessary to use the normalized version W rather than the primitive I as the suggestion proposed in [57].

Consequently, the p-th feature vector $\boldsymbol{x}_p^2 = [H_p, W_p]^{\mathsf{T}}$, which corresponds to the p-th individual/participant and is constructed by Hurst exponent and NNoD, are located in the normalized feature space $\mathcal{C}^2$ (a unit square), i.e.,

$$\boldsymbol{x}_p^2 \in \mathcal{C}^2 = \left\{ [W, H]^{\mathsf{T}} : 0 \leq W \leq 1, 0 \leq H \leq 1 \right\}.$$

For the p-th participant and the corresponding response sequence $\langle y_i^p, \tau_i^p \rangle_{i \in \mathbb{N}}$, we can find that:

- I_p is the number of decisions determined by the reaction time sequence $\{\tau_i^p\}_{i \in \mathbb{N}}$;

- W_p can be calculated from I_p with the help of Eqs (10) and (11);

- H_p can be computed from the reaction time sequence $\{\tau_i^p\}_{i \in \mathbb{N}}$.

3-dim feature space. Meanwhile, it is not sufficient if only 2-dim clustering is taken for analyzing our problem. For the sake of more deep understanding of the problem, we also constructed 3-dim feature space by adding the component E, i.e., the ERD.

Hence, the p-th 2-dim feature $\boldsymbol{x}_f^2$ now turns into a 3-dim vector $\boldsymbol{x}_p^3 = [W_p, H_p, E_p]^{\mathsf{T}}$, which is constructed by NNoD W, Hurst exponent H and ERD E. Obviously, each $\boldsymbol{x}_p^3$ is located in the 3-dim normalized feature space $\mathcal{C}^3$ (a unit cube), i.e.,

$$\boldsymbol{x}_p^3 \in \mathcal{C}^3 = \left\{ [W, H, E]^{\mathsf{T}} : 0 \leq W \leq 1, 0 \leq H \leq 1, 0 \leq E \leq 1 \right\}.$$

For the the p-th participant and its decision sequence $\{y_i^p\}_{i \in \mathbb{N}}$, the ERD E_p can be determined easily.

Eigen features, measurability and operation risk. Mathematically, the selection of the basis is not unique for the n-dim feature space $\mathcal{C}^n$. However, different bases are equivalently and they can be related with an invertible transition matrix. For the general features, it is sufficient for us if we take any form of the basis. However, if the clusters are considered, we need the concept of eigen-feature and eigen space.

For the k-th covariance matrix $\mathbf{\Sigma}_k^n$ of cluster S_k^n defined by (22) in $\mathbb{C}^n$, we denote its eigenvalue decomposition (EVD) as

$$\mathbf{\Sigma}_k^n = \sum_{i=1}^n \lambda_i^{(n,k)} e_i^{(n,k)} (e_i^{(n,k)})^\intercal, \quad 1 \leq k \leq K \tag{30}$$

where $\lambda_1^{(n,k)}, \lambda_2^{(n,k)}, \cdots, \lambda_n^{(n,k)} \in \mathbb{R}$ are the eigenvalues and $e_1^{(n,k)}, e_2^{(n,k)}, \cdots, e_n^{(n,k)} \in \mathbb{R}^{n \times 1}$ are the *eigen-features* of the cluster S_k^n. Usually, the eigenvalues are sorted with decreasing order, i.e., $\lambda_1^{(n,k)} \geq \lambda_2^{(n,k)} \geq \cdots \geq \lambda_n^{(n,k)}$. The subspace spanned by the eigen-feature $e_i^{(n,k)}$ is called the i-th eigen space if the eigenvalue $\lambda_i^{(n,k)}$ is not degenerate. For simplicity, we take the following notation

$$\mathrm{Eig}(\mathbf{\Sigma}_k^n) = \left\{ \lambda_1^{(n,k)}, \lambda_2^{(n,k)}, \cdots, \lambda_n^{(n,k)} \right\}. \tag{31}$$

It should be noted that all of the eigenvalues are positive since the covariance matrix is positive.

If $e_i^{(n,k)} = [0, 0, \cdots, 1, \cdots, 0]^\intercal$ where only the i-th position is 1, then the eigen-feature is standard. This fact is very important for our problem in which W, H and E are three components of the features concerned. In the sense of attention theory in psychology, the IF or equivalently IQ are not coupled and the i-th component of the IQ feature is *directly measurable* since W, H and E can be computed directly with the data collected in the NTT experiments. Generally, the eigen-feature $e_i^{(n,k)}$ is not standard and there are more than one factors/components should be considered simultaneously. On the other hand, this implies that each cluster has its own eigen-features and manifests its characteristics, thus the corresponding operators of civil aviation may have specific operation risk. Logically, we have the chain "cluster—eigen-features—operation-risk" for the aviation safety about air traffic control.

Levels of intrinsic qualities

Here we define four levels—*High, Moderate, Low and Bottom*—which are sequencing the intrinsic visual attention quality from "great" to "trivial". We now give some remarks for the levels of IQ:

- The bottom W indicates that the participants' reaction is very slow and it costs much time for making a decision. Experimentally, the process for making symbolic decision $y_i \in \{Y, N, I\}$ is long and the speed is very slow in the sense of statistics. Psychologically, it implies that the attention span of the participants is small. On the contrary, high W means fast reaction and large attention span.

- The moderate H, say $H \sim 0.6$, shows that there is a weak positive correlation in the sense of long term time [31]. Therefore, the attention stability of the participants is considerable.

Testing experiments and implementation of clustering algorithm
Setup of experiments

We seriously declare that the Ethics Committee of Research Institute of Civil Aviation Safety approved this research and the consent procedure. Given that the present study only measures simple behavior responses, it does not elicit adverse physiological and psychological reactions from the participants. Thus, the institutional ethics committee waived the need for written

informed consent from the participants. After all procedures had been explained by one of the researchers, those participants who decided to join the study could get an invitation code to access the program, which was the representation of their oral consent. Participants have the right to withdraw from the study at any time, and the researcher counted four participants who quited from the study in total. The participants are college students and they are required to read some simple symbols on the computer screen and make some harmless decisions by clicking the mouse [31].

In this work, the 168 participants involved are undergraduate and graduate students ($20 \sim 25$ years old) selected from the College of Air Traffic Management in Civil Aviation University of China. In other words, the data collected for the systematic assessment of IF are based on air traffic control students. It should be pointed that the testing conditions of experiments are considered seriously:

- The participants involved in the NTT test are in a good mental state, i.e., with no sickness or being lack of sleep in the past 48 hours.

- The test environment is quiet and isolated from any observable natural and/or artificial interruptions.

In each NTT experiment, each participant is required to finish 2 successive testing tasks [31] within 2×12 minutes. In each task, each participant is guided to read lots of positive/negative non-semantic symbols and is asked to make decisions simultaneously. Without doubt, different participant will have different response sequence, and also different NoD in the same time interval due to the different reaction time in each decision operation. After data cleaning, the responses of 143 participants are kept and 25 records are abandoned. For the final data set, please see S1 Appendix.

Constructing features for intrinsic qualities

The selected features and the number of features will affect the clustering result. For a prototype modeling, the data feature extraction shall be described firstly. After the pre-processing of data, the features of interested are extracted from the reaction time sequence of each participant. As a companion of Fig 2, the pseudo-code for the feature extraction for particilants in NTT is given in Algorithm 1.

Algorithm 1: n-dim Features Extraction for Participants

Input: Dimension $n \in \{2, 3\}$, response sequences $\langle y_i^p, \tau_i^p \rangle_{i \in \mathbb{N}}$ of N participants.

Output: Feature vectors $\{x_p^n\}_{p=1}^N$, where $x_p^n = [x_{1,p}^n, \cdots, x_{i,p}^n, \cdots, x_{n,p}^n]^\top$ of IQ.

```
 1: Preconditioning of the input response sequences.
 2: for p ∈ {1, ···, N} do
 3:    Compute W_p^n by Eq (28) with the decision sequence {y_i^p}_{i∈ℕ}.
 4:    Set the first component of x_p^n: x_{2,p}^n = W_p^n
 5:    Compute H_p^n by Eqs (1)~(8) with the reaction time sequence {τ_i^p}_{i∈ℕ}.
 6:    Set the second component of x_p^n: x_{1,p}^n = H_p^n.
 7:    if n == 3 then
 8:       Calculate E_p^n by Eq (29) with the decision sequence {y_i^p}_{i∈ℕ}.
 9:       Set the third component of x_p^n: x_{3,p}^n = E_p^n
10:    end if
11: end for
12: return Feature vectors {x_p^n}_{p=1}^N.
```

Implementation of *K*-means clustering algorithm

Applicability of *K*-means clustering algorithm. The advantage of *K*-means lies in its simple interpretation due to the intuitive and reliable clustering results. However, there are also some disadvantages for this algorithm: the possibility of convergence to local optimal and high computational complexity for large scale sample spaces. In our experiment, the data set collected from the NTT system is a small scale, thus the *K*-means algorithm works well.

Choice of *K*. For the implementation of *K*-means clustering algorithm, the *K* value can not be too large. The cross-validation method is used to select the *K* value. The majority of cases of selection are relying on the prior belief or intrinsic characteristics of the data. According to the characteristics of the NTT data set, the data set is divided into four categories which aims at classifying all of the participants into four levels: excellent, good, moderate and unqualified. The objective of clustering the feature points corresponding to the controllers (or air traffic control students), includes three folds:

- exploring the potential regulation for classifying controllers into different levels;

- predicting the lapse of attention in the practical operations in air traffic control by quantitative analysis;

- improving the IQ of controllers via professional training.

In our recent studies [41–43], the characteristics of attention quality of controllers are described and the differences between controllers and ATC students are revealed according to the types of attention quality. The results obtained show that:

- both controllers and ATC students could be classified into four types according to attention quality;

- the qualified controllers concerned show large attention span, strong attention stability, rational attention switching and improper attention distribution;

- there is a significant difference between full time controllers and ATC students.

Pseudocode for clustering algorithm. For the details of *K*-means algorithm, please refers to [45]. The 2-dim and 3-dim feature spaces of interest have been obtained from the feature extraction, and serve as the input of the *K*-means algorithm. Mathematically, we have $\boldsymbol{x}_j = \boldsymbol{x}_k^2$ or $\boldsymbol{x}_j = \boldsymbol{x}_k^3$ for the *j*-th feature when clustering.

Clustering with experimental data from ATC students

The data obtained from the NTT System are analyzed carefully and systematically. The distribution of clusters can be visualized intuitively with 2-dim and 3-dim clustering. It is necessary for us to explain why both 2-dim and 3-dim clustering are valuable for classifying air traffic control problem. The key issue lies in the level of information processing of ATC students in the sense of psychology:

- the 2-dim clustering analysis with *K*-means algorithm is mainly based on the Hurst exponent and the workload, which reflects the speed and stability of information processing of the ATC students (see Fig 3);

- the 3-dim clustering analysis also reflects the accuracy of information processing of the ATC students (see Fig 4) if compared with the 2-dim clustering analysis;
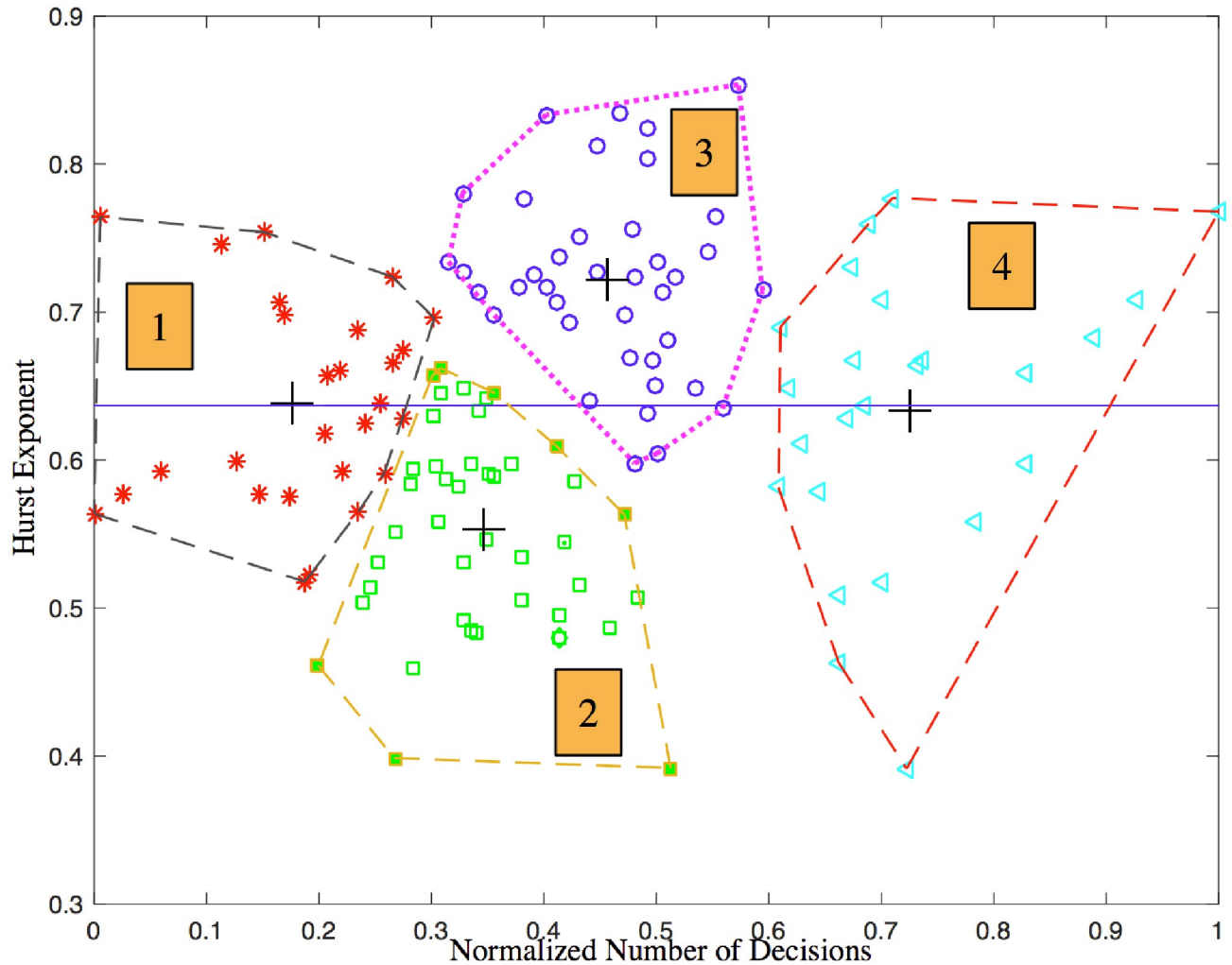
Fig 3. Clusters of 2-dim features $[W, H]^\top$ with boundaries generated by convex hull algorithm.

- in general, the 3-dim clustering analysis is more comprehensive than the 2-dim clustering analysis for describing and explaining the level of information processing of the ATC students.

2-dim clustering. The 2-dim feature space is taken for the 2-dim clustering. Eq (14) shows that the set of 2-dim feature vectors

$$S^2 = \{\boldsymbol{x}_i \in \mathbb{C}^2 : 1 \leq i \leq N\} \tag{32}$$

can be partioned into $K = 4$ disjoint subsets S_1^2, S_2^2, S_3^2 and S_4^2, where

$$S_k^2 = \left\{ \boldsymbol{x}_{i_k} = [W_{i_k}, H_{i_k}]^\top : i_k \in \Phi_k \right\}, \quad 1 \leq k \leq K. $$

With the purpose of partitioning different clusters with a distinct visualization, here we use Graham's Scan [58] with a quasi-linear time computational complexity to find the convex hull of the feature points in our problem, i.e. the contours of clusters. This procedure simplifies the step of separating different clusters and avoid the complex methods for determining the border
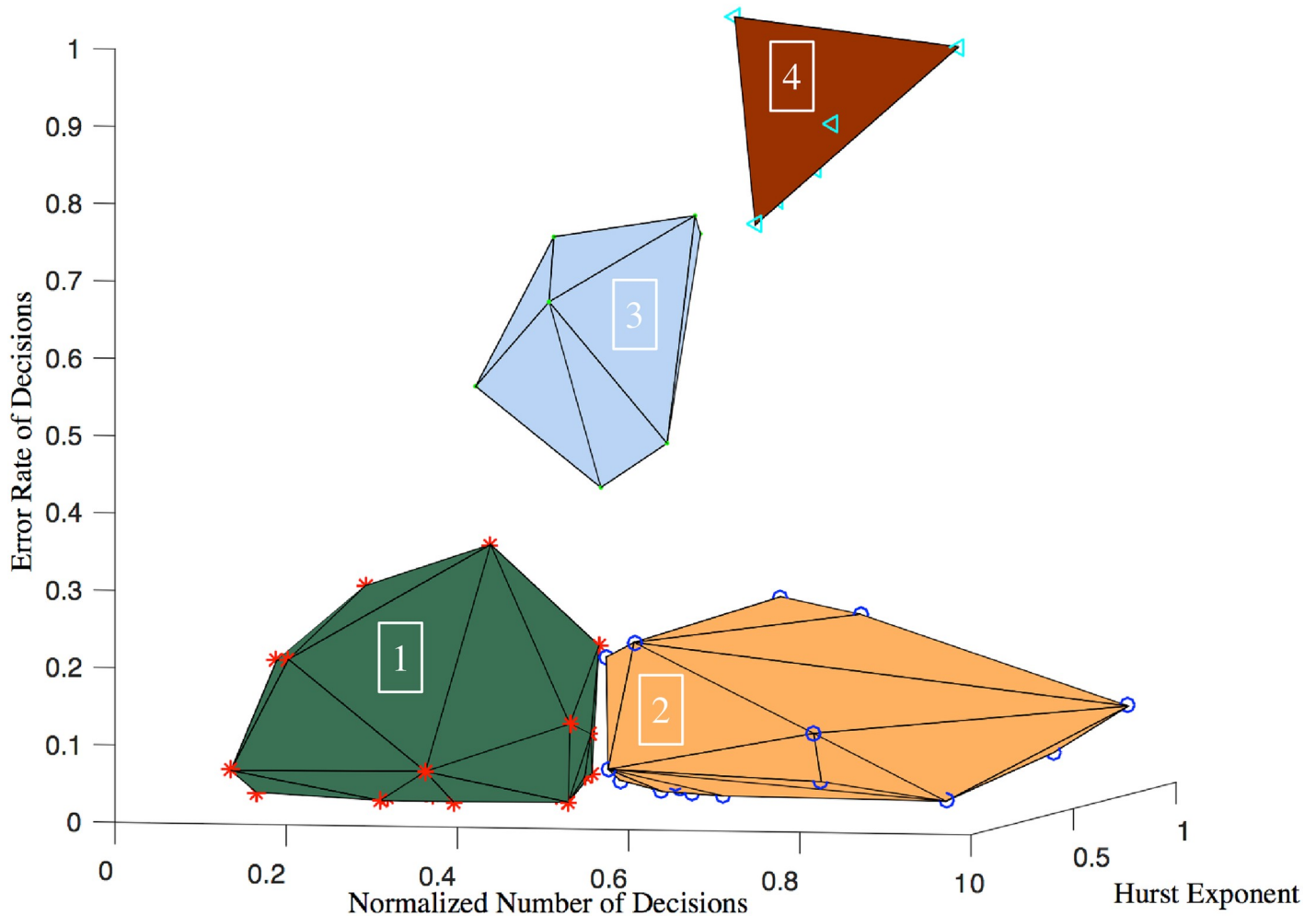
Fig 4. Clusters of 3-dim features $[W, H, E]^\top$ with contours.

of clusters in clustering algorithms. Intuitively, Fig 3 shows that the third cluster is the best when the 2-dim clustering is taken because of the high stability of attention (large H) and good attention span (acceptable W).

3-dim clustering. The 3-dim feature space is taken for 3-dim clustering. Eq (14) shows that the set of 3-dim discrete feature vectors

$$S^3 = \{\boldsymbol{x}_i^3\}_{i=1}^N \subset \mathcal{C}^3 \tag{33}$$

can be partitioned into $K = 4$ disjoint subsets $S_1^3, S_2^3, S_3^3, S_4^3$ where

$$S_k^3 = \left\{ \boldsymbol{x}_{i_k}^3 = [W_{i_k}, H_{i_k}, E_{i_k}]^\top : i_k \in \Phi_k \right\}, \quad 1 \leq k \leq K.$$

Fig 4 shows that the second cluster is the most expected for 3-dim clustering obviously because of three characteristics: low error rate of decisions (samll ERD), high workload (large NNoD) and stable information processing (good Hurst exponent). The distances between clusters in 3-dim feature space are greater than their counterparts in 2-dim feature space and separation of clusters is much more clear, as shown in Fig 4. By comparison with Figs 3 and 4,

we will find that a higher-dimensional leads to a better clustering result, which will be proved in the following analysis.

Assessing the attention qualities by statistical inference

Comparison of 2-dim and 3-dim clustering

Centroid. According to the data collected by the NTT system for the ATC students, the centroids for the clusters are obtained, see Table 2. The centroids are computed by the following expressions:

$$\boldsymbol{\mu}(S_k^2) = \frac{1}{N_k} \sum_{\boldsymbol{x}_{i_k} \in S_k^2} \boldsymbol{x}_{i_k} = \frac{1}{N_k} \sum_{i_k=1}^{N_k} [H_{i_k}, W_{i_k}]^\mathsf{T}, \tag{34}$$

$$\boldsymbol{\mu}(S_k^3) = \frac{1}{N_k} \sum_{\boldsymbol{x}_{i_k} \in S_k^3} \boldsymbol{x}_{i_k} = \frac{1}{N_k} \sum_{i_k=1}^{N_k} [H_{i_k}, W_{i_k}, E_{i_k}]^\mathsf{T}. \tag{35}$$

It clearly shows that the clusters respectively possess different centroids, which demonstrates the validation of the clustering algorithm. The centroids hold the role of the most representative points of each cluster.

Euclidean distance. Figs 5 and 6 are two heat-maps about the matrix of Euclidean distance $\mathrm{dist}_E(S_k^n, S_i^n)$ between every two centroids by (18). The distance ranges for 2-dim and 3-dim clusters are $[0, \sqrt{2}]$ and $[0, \sqrt{3}]$ respectively. The Euclidean distance gives an intuitive clustering result. By comparison, we find that 3-dim clustering separates the sample space with more prominent distinctness. If the sample space is better partitioned into four clusters, then more distinct characteristics will be obtained. Consequently, clustering on a higher dimensional sample space could represent more profound IF hidden in the participants.

Second order statistics. For each cluster, either 2-dim or 3-dim clustering, the values of second order statistics are satisfactory for statistical inference:

- Table 3 shows that the variances and standard deviations of 2-dim and 3-dim clusters are small by Eq (20);

- Tables 4 and 5 show that the covariance matrices are diagonally dominant for each 2-dim cluster S_k^2 and 3-dim cluster S_k^3.

Obviously, this means good inter-cluster statistical characteristics. In other words, the validation of clustering algorithm has been proved once again.

Once two clusters are considered at the same time, we can find the intra-cluster characteristics with the covariance matrix by (22). Covariance matrix represents the relevance among different features and its eigenvalues indicate the contribution of the corresponding feature to the characteristic of an individual (participant of NTT).

Table 2. Centroids of 2-dim and 3-dim clusters.

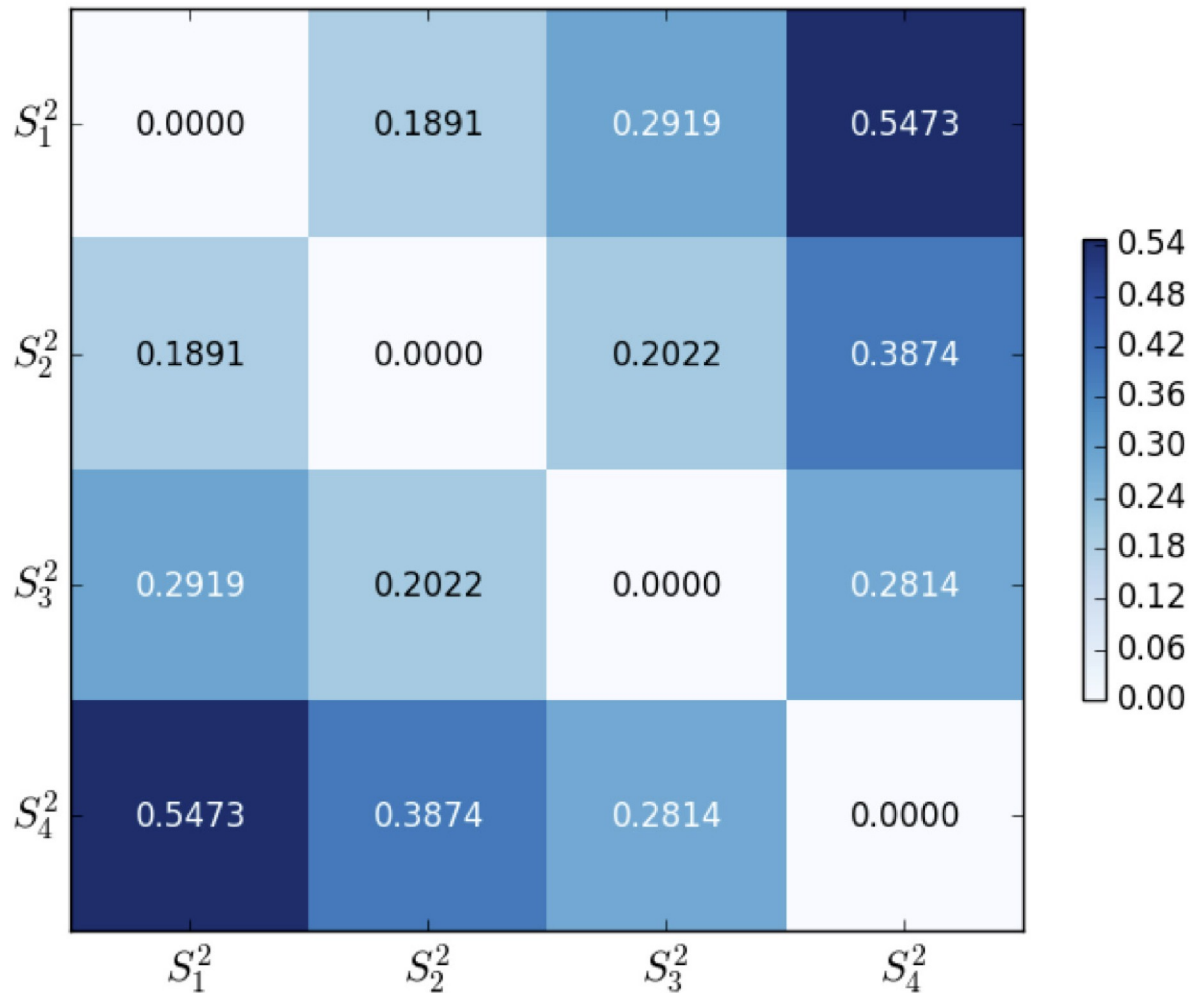| Cluster S_k^2	Centroid $\boldsymbol{\mu}(S_k^2)$	Cluster S_k^3	Centroid $\boldsymbol{\mu}(S_k^3)$
S_1^2	$[0.1771, 0.6390]^\mathsf{T}$	S_1^3	$[0.2714, 0.6032, 0.0510]^\mathsf{T}$
S_2^2	$[0.3455, 0.5530]^\mathsf{T}$	S_2^3	$[0.5789, 0.6675, 0.0680]^\mathsf{T}$
S_3^2	$[0.4571, 0.7216]^\mathsf{T}$	S_3^3	$[0.3980, 0.6683, 0.5880]^\mathsf{T}$
S_4^2	$[0.7244, 0.6335]^\mathsf{T}$	S_4^3	$[0.6590, 0.6490, 0.8619]^\mathsf{T}$

Fig 5. 2-dim Euclidean distances of clusters $\{S_k^2\}_{k=1}^4$.

With the help of Table 6 for the 2-dim clustering, we find that the features are weakly dependent due to the small correlation coefficients $Cr_{ij,i\neq j} < 0.3$ by (23). However, Table 6 also shows that for the 3-dim clustering the dependency among three features is higher with the similar calculation for the correlation coefficients. In other words, the analysis explored on the 3-dim feature space is more complex than its counterpart in 2-dim space. The higher the dependency of the features, the stronger the cross-impact of the IQ is. Intuitively, the attention span, stability of attention, distribution-shift of attention may interplay in some sense. Fortunately, both correlation coefficient of these clusters own a weak dependency among different features, i.e., the features of the IQ corresponding to an excellent participant are nearly independent.

TrLogDet distance. The TrLogDet distance $\mathrm{dist}_T(S_k^n, S_i^n)$ for each two clusters are computed according to Eq (25), see Figs 7 and 8. We find that TrLogDet values for the 2-dim clusters are smaller than their counterparts in 3-dim space. For 2-dim clustering, the clusters have similar profiles where the samples locate within a neighborhood. By contrast, for 3-dim clustering, the samples spread in a wide range. Although the clustering in 3-dim sample space has
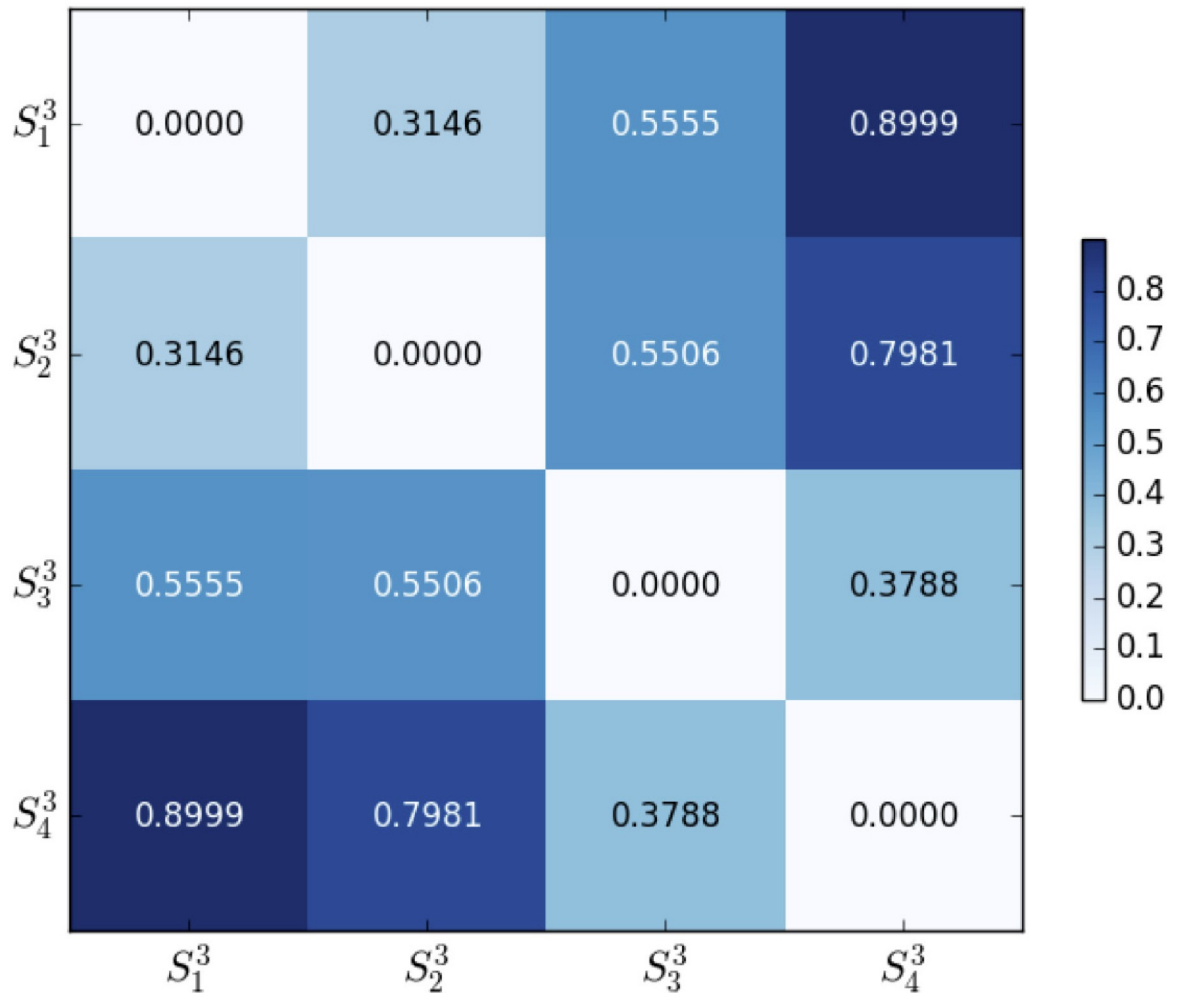
Fig 6. Euclidean distance of 3-dim clusters $\{S_k^3\}_{k=1}^4$.

https://doi.org/10.1371/journal.pone.0205334.g006

higher complexity and the assessments for 3-dim objects is more complicated, the 3-dim result is more persuasive in general since it contains more information.

Information theoretic metric. The information theoretic metric (ITM) is calculated via (26), see Table 7. The ITMs for the four 2-dim clusters have no significant difference, which means their profiles are similar. However, for the 3-dim clusters, the difference of ITMs are much larger than those in 2-dim space. This implies two facts:

Table 3. Variances and standard deviations of clusters: 2-dim vs. 3-dim.

S_k^2	$[\sigma(S_k^2)]^2$	$\sigma(S_k^2)$	$\frac{1}{2}\sigma(S_k^2)$	S_k^3	$[\sigma(S_k^3)]^2$	$\sigma(S_k^3)$	$\frac{1}{3}\sigma(S_k^3)$
S_1^2	0.0595	0.2439	0.1220	S_1^3	0.0597	0.2443	0.0814
S_2^2	0.0150	0.1225	0.0613	S_2^3	0.0799	0.2867	0.0956
S_3^2	0.0223	0.1493	0.0747	S_3^3	0.0223	0.1526	0.0509
S_4^2	0.0118	0.1086	0.0542	S_4^3	0.0169	0.1300	0.0433

https://doi.org/10.1371/journal.pone.0205334.t003

Table 4. Covariance matrix and eigenvalues for 2-dim clusters.

Cluster S_k^2	Covariance Matrix Σ_k^2	Eigenvalues Eig(Σ_k^2)
S_1^2	$\begin{bmatrix} +0.0076 & +0.0005 \\ +0.0005 & +0.0047 \end{bmatrix}$	{0.0076, 0.0046}
S_2^2	$\begin{bmatrix} +0.0047 & -0.0007 \\ -0.0007 & +0.0043 \end{bmatrix}$	{0.0053, 0.0038}
S_3^2	$\begin{bmatrix} +0.0049 & -0.0008 \\ -0.0008 & +0.0045 \end{bmatrix}$	{0.0055, 0.0039}
S_4^2	$\begin{bmatrix} +0.0106 & +0.0029 \\ +0.0029 & +0.0093 \end{bmatrix}$	{0.0129, 0.0070}

Table 5. 3-dim clusters, covariance matrix and eigenvalues.

Cluster S_k^3	Covariance Matrix Σ_k^3	eigenvalues Eig(Σ_k^3)
S_1^3	$\begin{bmatrix} +0.0105 & -0.0015 & -0.0009 \\ -0.0015 & +0.0073 & +0.0019 \\ -0.0009 & +0.0019 & +0.0039 \end{bmatrix}$	{0.0114, 0.0073, 0.0030}
S_2^3	$\begin{bmatrix} +0.0208 & -0.0015 & +0.0008 \\ -0.0015 & +0.0123 & -0.0005 \\ +0.0008 & -0.0005 & +0.0061 \end{bmatrix}$	{0.0211, 0.0121, 0.0060}
S_3^3	$\begin{bmatrix} +0.0065 & +0.0035 & +0.0019 \\ +0.0035 & +0.0081 & +0.0044 \\ +0.0019 & +0.0044 & +0.0101 \end{bmatrix}$	{0.0152, 0.0062, 0.0032}
S_4^3	$\begin{bmatrix} +0.0092 & -0.0011 & +0.0025 \\ -0.0011 & +0.0027 & +0.0046 \\ +0.0025 & +0.0046 & +0.0112 \end{bmatrix}$	{0.0003, 0.0090, 0.0139}

Table 6. Pairwise correlation coefficient of W, H and E for specified clusters.

| Cluster	$\mathbf{Cr}\big[W(S_k^2), H(S_k^2)\big]\}$	Cluster	$\mathbf{Cr}\big[W(S_k^3), H(S_k^3), E(S_k^3)\big]$
S_1^2	$\begin{bmatrix} +1.0000 & +0.0776 \\ +0.0776 & +1.0000 \end{bmatrix}$	S_1^3	$\begin{bmatrix} +1.0000 & -0.1672 & -0.1345 \\ -0.1672 & +1.0000 & +0.3577 \\ -0.1345 & +0.3577 & +1.0000 \end{bmatrix}$
S_2^2	$\begin{bmatrix} +1.0000 & -0.1631 \\ -0.1631 & +1.0000 \end{bmatrix}$	S_2^3	$\begin{bmatrix} +1.0000 & -0.0925 & +0.0681 \\ -0.0925 & +1.0000 & -0.0529 \\ +0.0681 & -0.0529 & +1.0000 \end{bmatrix}$
S_3^2	$\begin{bmatrix} +1.0000 & -0.1616 \\ -0.1616 & +1.0000 \end{bmatrix}$	S_3^3	$\begin{bmatrix} +1.0000 & +0.4913 & +0.2362 \\ +0.4913 & +1.0000 & +0.4871 \\ +0.2362 & +0.4871 & +1.0000 \end{bmatrix}$
S_4^2	$\begin{bmatrix} +1.0000 & +0.2932 \\ +0.2932 & +1.0000 \end{bmatrix}$	S_4^3	$\begin{bmatrix} +1.0000 & -0.2288 & +0.2447 \\ -0.2288 & +1.0000 & +0.8244 \\ +0.2447 & +0.8244 & +1.0000 \end{bmatrix}$
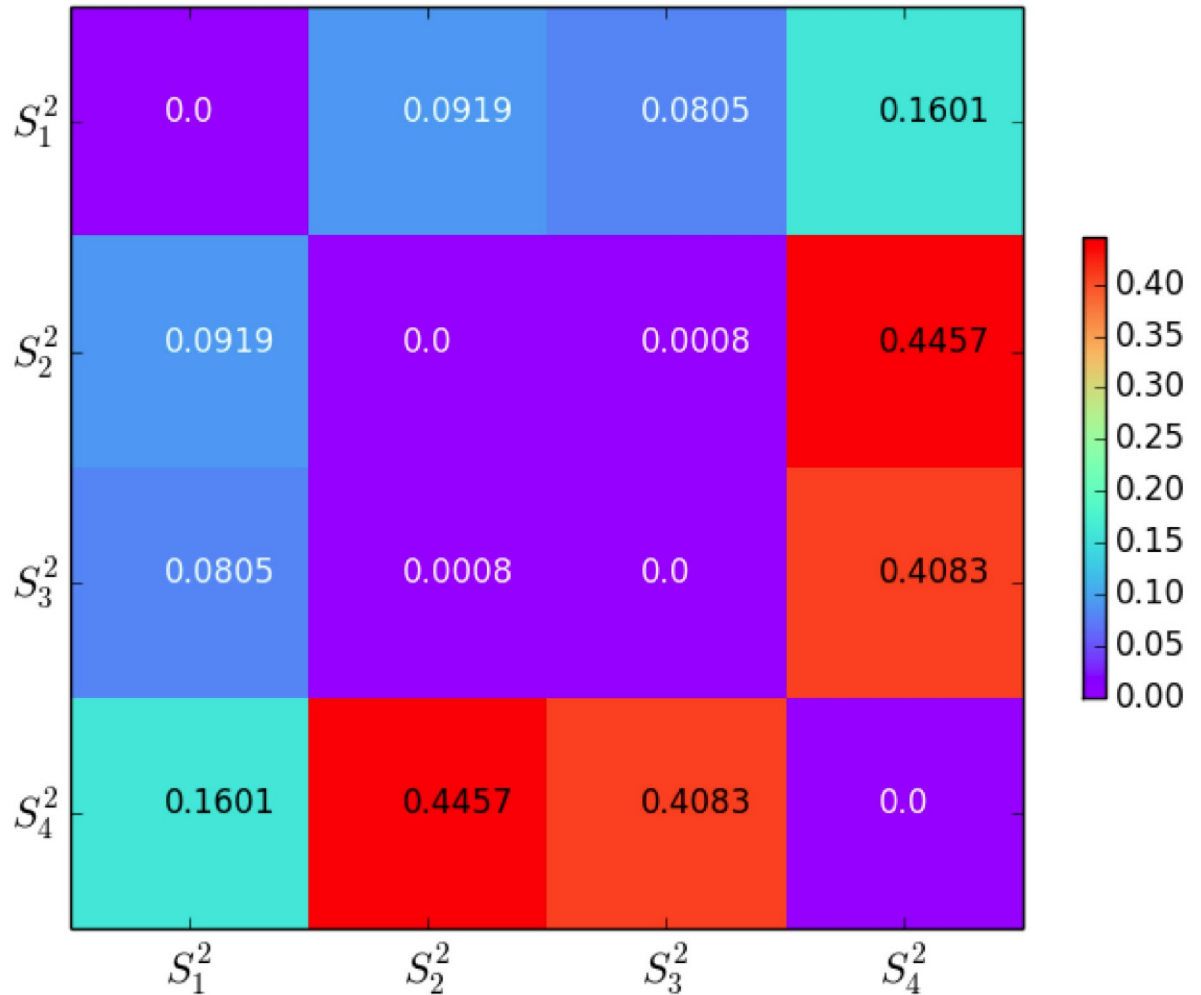
Fig 7. 2-dim clusters and TrLogDet distance per dimension.

- the feature points can be distinguished well with ITM if the dimension can be set up properly;

- the performance of 3-dim clustering is better than 2-dim clustering since more informaton is involved in for the purpose of distinguishing feature points.

Sum of squared errors. According to Eq (27) for SSE, we obtain Table 8. For the SSE of n-dim clustering, a small value of SSE per sample reflects satisfactory clustering outcome for a given K. A good choice of K depends on the characteristics of the data set, experiences, and objective of the data analysis. In our study, we set $K = 4$.

Clusters and performances of attention

2-dim clusters of features $[W, H]^\top$. Fig 3 illustrates the distinct clusters with K-means algorithm and $K = 4$. It is clear that the centroids of the four clusters are separated reasonably. For the 2-dim clusters demonstrated in Fig 3, the corresponding box-plot is shown in Fig 9. With the four levels (high, moderate, low and bottom) mentioned above, we now propose some interpretations as following:
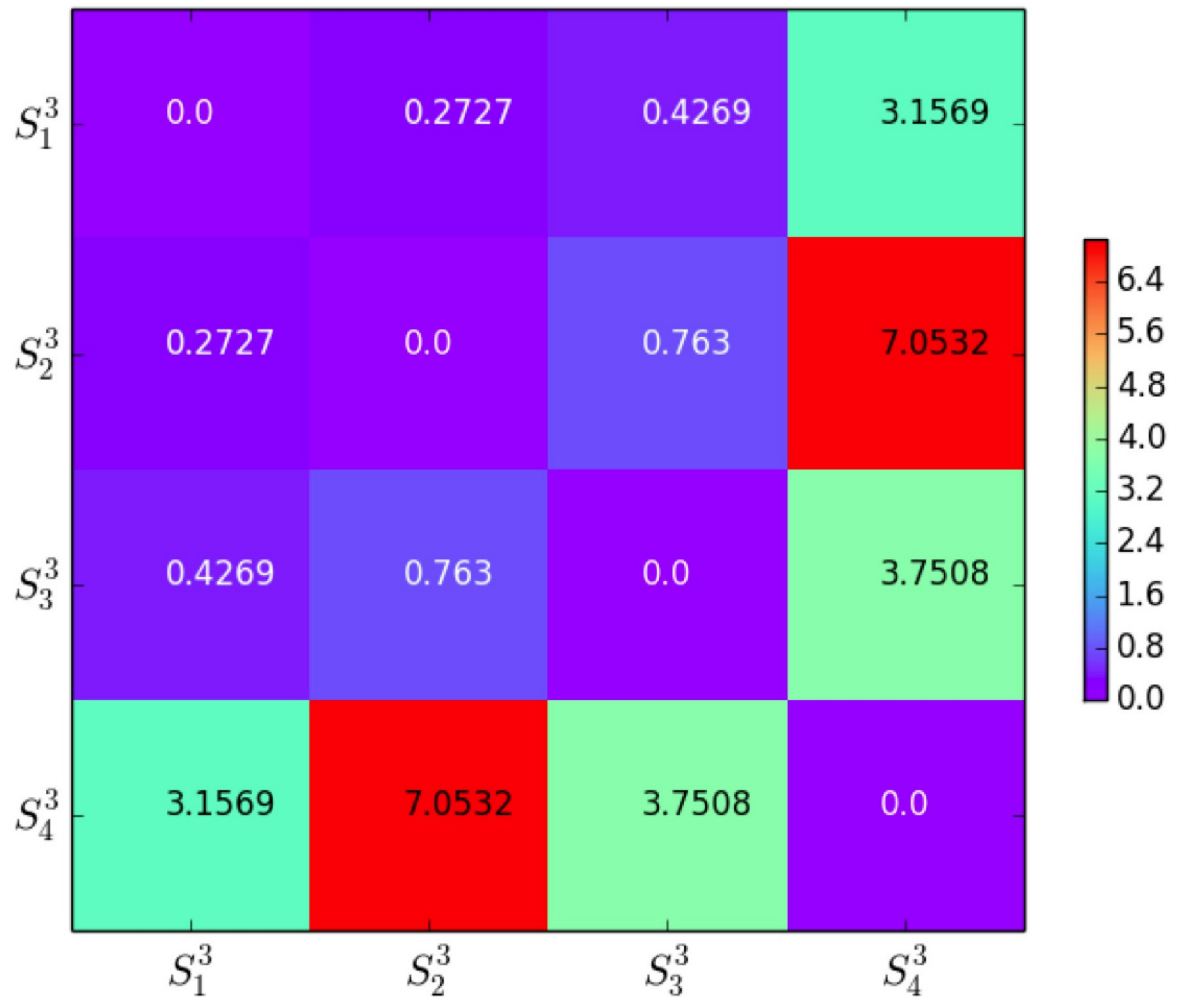
Fig 8. 3-dim clusters and TrLogDet distance per dimension.

Table 7. ITM per dimension for 2-dim and 3-dim clusters.

k	1	2	3	4
$\rho(S_k^2)$	8.2740	8.8435	8.7624	7.3384
$\rho(S_k^3)$	12.2218	10.4285	12.0250	14.2618

Table 8. SSE for 2-dim and 3-dim clusters.

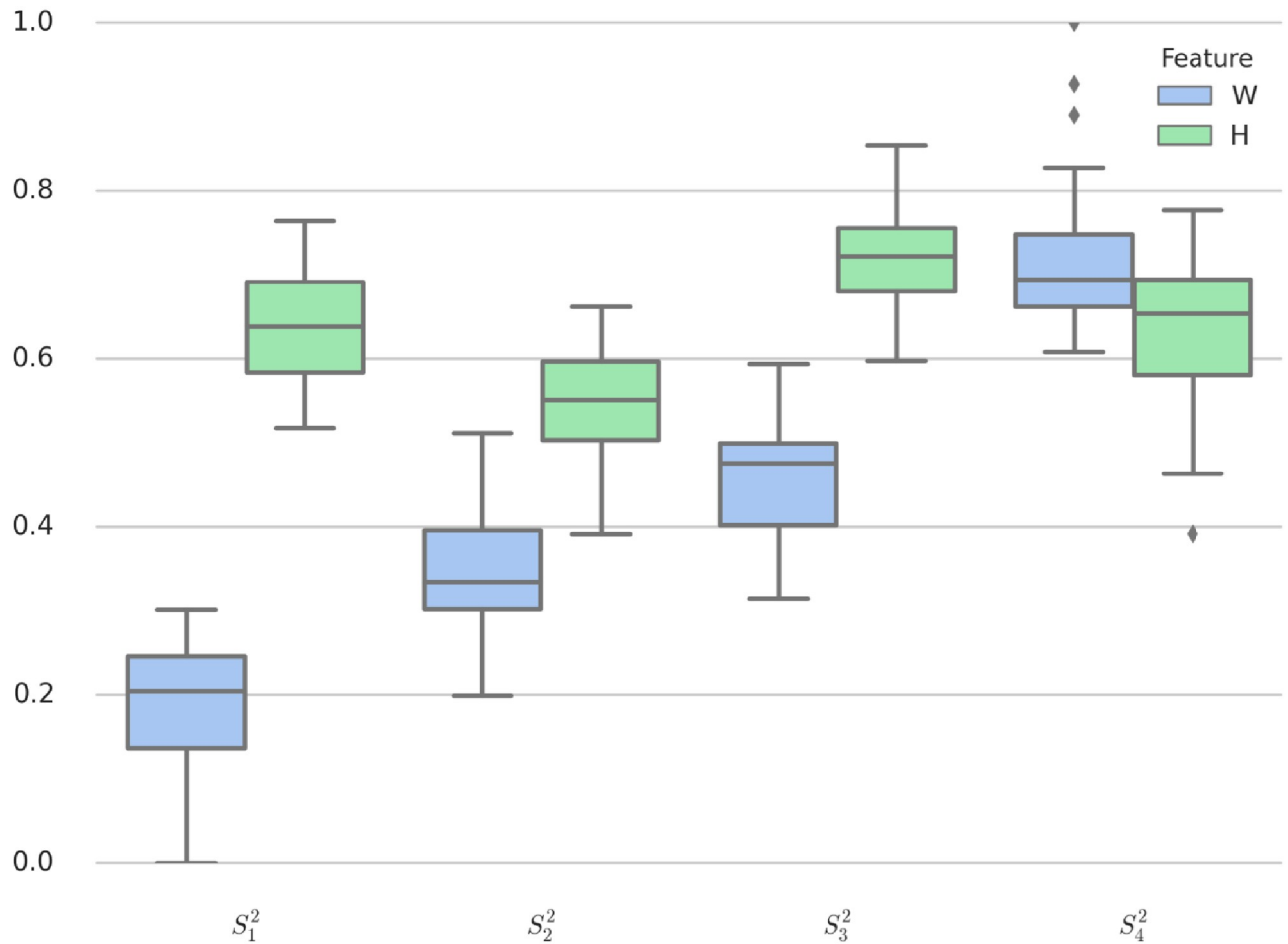| k	1	2	3	4	Amounts
$SSE(S_k^2)$	0.1063	0.0104	0.0200	0.0032	0.1399
$SSE(S_k^2)/|S_k^2|$	0.0034	0.0002	0.0005	0.0001	0.0042
$SSE(S_k^3)$	0.2563	0.3131	0.0055	0.0020	0.6291
$SSE(S_k^3)/|S_k^3|$	0.0034	0.0061	0.0005	0.0003	0.0103

Fig 9. Box-plot of 2-dim clustering data.

- S_1^2 with bottom W and moderate H

  - The average W is the smallest among the four clusters, which indicates that the participants' reaction is very slow and they spend much time to make a decision. Experimentally, the process for making the symbolic decision $y_i \in \{Y, N, I\}$ is slow in the sense of statistics. Psychologically, it implies that the attention span of the participants is small.

  - The average H is about 0.64 which shows a good positive correlation in the sense of long term time [31]. Therefore, the attention stability of the participants is acceptable.

  - This type of controllers are not appropriate for busy airports with large traffic flow. On the contrary, only small airports with fewer flights could be considered for these controllers. For the sake of low operation risk, the participants should be allocated with low workload.

- S_2^2 with low W and bottom H

  - H floats up and down at 0.5, which indicates that the participants are most unstable during decision-making since $H = 0.5$ means strong randomness and poor correlation [31]. Therefore, the attention of the participants is not stable.

- Low W shows that more time is needed for decision-making, thus the attention span is small.

- The participants included in S_2^2 are not appropriate for being selected as controllers since neither the attention span nor the attention stability is good in the sense of controlling the operation risk.

- S_3^2 with moderate W and high H

  - For a given period, the NoD of participants in the 3rd cluster is stable. H above 0.6 shows that their response sequences are of the following characteristics: better long-range correlation and respondent continuity as well as positive correlation.

  - Moderate W indicates an ordinary level of decision-making of participants and good attention span.

  - Samples in cluster S_3^2 and cluster S_2^2 are at opposite positions and also with opposite characteristics. It proves that the clustering algorithm is capable of partitioning the samples with significant differences.

  - S_3^2 includes the best participants separated automatically via 2-dim clustering process.

- S_4^2 with high W and low H

  - High W means high attention span while low H means low attention stability.

  - The corresponding participants could work with those in S_3^2 when the workload of controllers is large in the cases of peak air traffic flow at busy airports.

3-dim clusters of features $[W, H, E]^\top$. Fig 4 illustrates the result of clustering via K-means algorithm with $K = 4$ for the 3-dim features $\boldsymbol{x}_k^3 = [W_k, H_k, E_k]^\top$ and the corresponding box-pot is given in Fig 10. It is clear that the centroids of the four clusters are separated more distinctly than those in 2-dim space.

For 3-dim clustering, the rank of H in Table 9 has no distinct difference because of $H \in [0.6, 0.7]$ for four different clusters (which implies good positive long-range correlations and attention stability). We now give some interpretations for these clusters as follows:

- S_1^3 with bottom W, low E and poor H

  - The process of decision-making of participants in cluster S_1^3 needs a long time to make decisions, thus the attention span must be small. Meanwhile, a sufficient thinking time will lead to high accuracy or equivalently low E.

  - The values of H spread in a wide range which covers the critical point $H = 0.5$, thus their long-range correlation is relatively poor. We can infer that the occurrence of mistakes fluctuates rapidly and randomly. In consequence, the attention stability is poor.

- S_2^3 with moderate W, low E and a wide range of H

  - Moderate W of the participants in cluster S_2^3 indicates a moderate level of information processing and attention span. Particularly, 25% of the participants in S_2^3 have excellent attention span since they have large W values when compared with other clusters.

  - Low E shows the high accuracy during decision-making. Therefore, the distribution-shift of attention is good and acceptable.
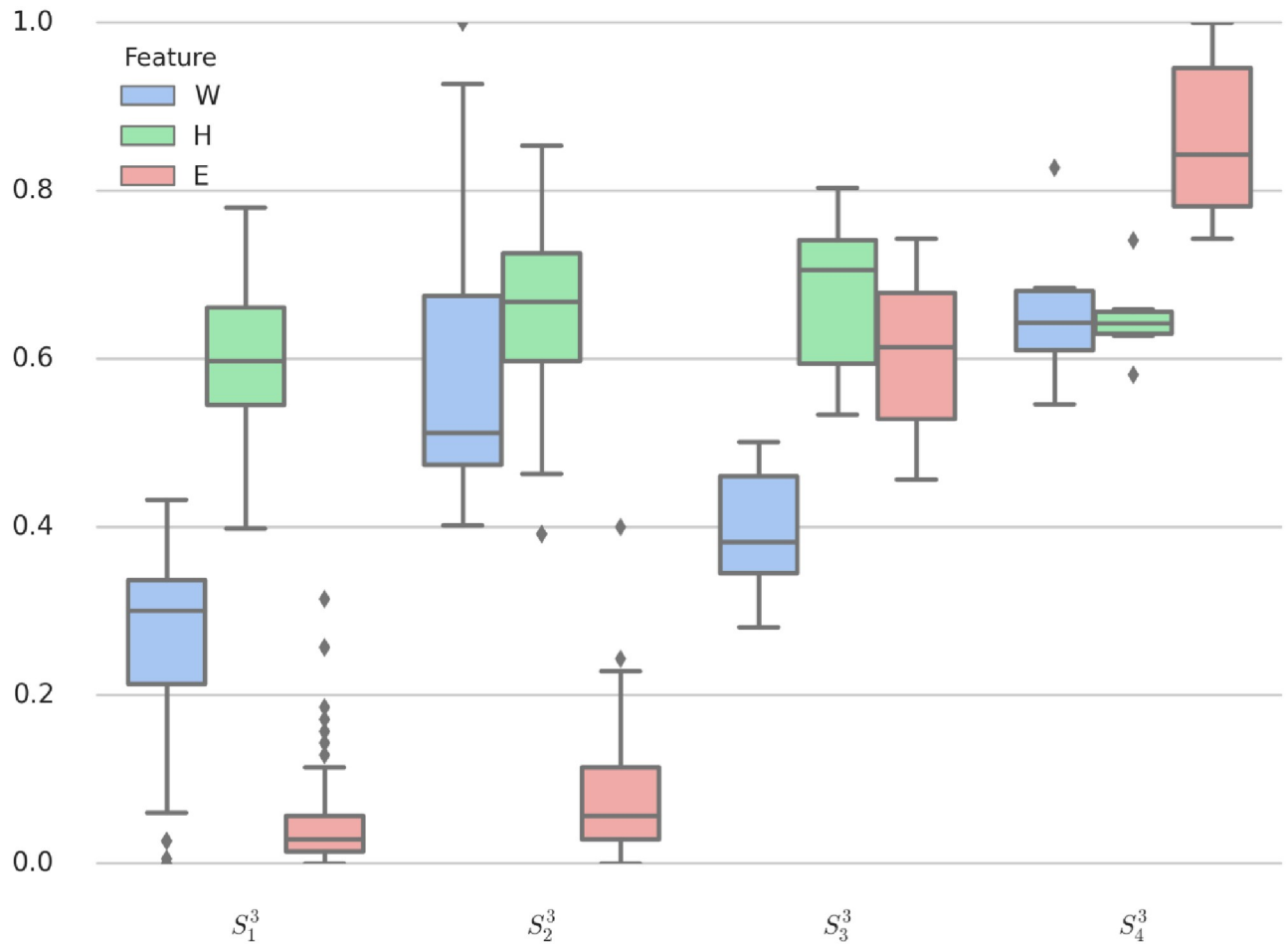
Fig 10. Box-plot of 3-dim clustering data.

- The value of H has a wide range, which implies the long-range correlation hidden in the response sequence is complex. However, for the participants in S_2^3, the box-plot also shows that more 70% of them have good attention stability since $H \geq 0.6$ and 25% of them have excellent attention stability since $H \geq 0.7$.

- In general, S_2^3 is the best cluster for 3-dim clustering and the participants of interest can be selected objectively since the IQ are the best.

- S_3^3 with low W, moderate E and high H

- About 75% of the participants have good attention stability since $H > 0.6$ and 25% of them have moderate H. Moreover, it is clear that the W is low and E is high since $W \leq 0.5$ and $E \geq 50\%$.

Table 9. Rank the IQ for 3-dim clustering result.

Cluster	Span	Stability	distribution-shift
S_1^3	Bottom	Bottom	**High**
S_2^3	Moderate	Moderate	Moderate
S_3^3	Low	**High**	Low
S_4^3	**High**	Low	Bottom

Table 10. Expert evaluation on IF influencing VAP.

IF of Attention	IQ	1 ∼ 3 Points	4 ∼ 6 Points	7 ∼ 9 Points
Span	W	Narrow	Medium	Wide
Stability	H	Bad	Medium	Good
Distribution-shift	E	Slow & Inaccurate	Medium	Rapid & Exact

- Obviously, the participants characterized by S_3^3 have poor work performances: low attention span, low distribution-shift of attention and strong stability. In other words, they response slowly, make decisions with high error rate, and it is difficult for them to improve their work efficiency and correctness of decisions.

- S_4^3 with high W, low H, and high E

  - Both the W and the H are moderate, thus the attention span and stability of the participants in S_4^3 are acceptable.

  - The distribution-shift of attention is very poor since $E > 70\%$ for each participant and $E \approx$ 100% for participants. This implies that the participants should be re-evaluated seriously and they may not be selected as air traffic controllers.

Levels of intrinsic qualities for ATC students and controllers

According to the results of 2-dim and 3-dim clustering analysis that are shown in Figs 3 and 4 respectively, we invited five experts of ATC with more than 10 years work experiences to evaluate attention span, attention stability and distribution-shift of attention of the controllers (grade from 1 to 9 points). The higher the score is, the better the attention features are. In the end, we get three intervals: 1 ∼ 3 points, 4 ∼ 6 points and 7 ∼ 9 points, see Table 10.

Based on the assessment of IQ for 2-dim and 3-dim clusters, the ATC students can be divided into four classes, from which the classification could be generalized to the assessment of air traffic controllers, see Table 11. Consequently, we have

- *Level-1* air traffic controllers—Concentrating attention rapidly, distributing attention reasonably, and finishing the specific task with high accuracy subject to high workload. It is feasible to allocate complex air traffic control tasks for this kind of controllers.

- *Level-2* air traffic controllers: Processing information swiftly and stably with high error rate. In other words, they may suffer great difficulties if multi-tasks or the task with multi-objectives are concerned. This kind of controllers may make lots of mistakes due to high air traffic flow. Their low performance of distribution-shift of attention might cause omission of traffic conflicts, which may increase the operation risk.

- *Level-3* air traffic controllers: The distribution-shift of attention is swift. However, neither the attention span is sufficient nor the attention stability is acceptable. Apparently, the

Table 11. Assessment of air traffic controllers' IQ influencing VAP: A generalization.

Level	Span	Stability	Distribution-Shift	Volatility
1	Wide	Good	Rapid & Exact	Quasi-Stationary
2	Wide	Good	Slow & Inaccurate	Small-Amplitude
3	Narrow	Bad	Rapid & Exact	Large-Amplitude
4	Narrow	Bad	Slow & Inaccurate	Small-Amplitude

"error/mistake","forgetting" and "omission" phenomenon appear frequently if the controllers' workload is high. As a suggestion, it would be better for this kind of controllers to work in the period of an average air traffic flow intensity or in a small airport with low air traffic flow intensity.

- *Level-4* air traffic controllers: Characterized by a non-timely reaction and high probability of making mistakes. Generally, the inferior IQ of Level-4 controllers may cause serious operation risk.

Conclusions

Key results

Aviation safety is closely related with the IF influencing the VAP of air traffic controllers. However, there is a lack of mathematical descriptions and effective analysis of IF based on observable data. We presented the IQ—a mathematical representation for the IF—and constructed the interested 2-dim and 3-dim feature spaces with the computational parameters NNoD, Hurst exponent and ERD for characterizing the response sequences. The *K*-means clustering algorithm is applied for the IQ and the systematic assessment is performed for the clusters in 2-dim and 3-dim through statistical inference through computing and comparing the geometric metrics and statistics defined.

By the NTT system and the response sequence collected from 143 ATC students, the statistical analysis for the clusters obtained demonstrates some interesting results:

- The data set has been well divided into four clusters which coincides with the practical observations for the air traffic controllers.

- The eigen-features of the clusters and the correlation coefficients for the directly measurable IQ, i.e., NNoD, Hurst exponent and ERD, show that each cluster has its own eigen-features and manifests its unique characteristics, thus the corresponding controllers or professionals may have specific operation risk.

- The participants can be divided into four levels: Excellent, Good, Medium, Unqualified, which implies:

  - the potential rules for selecting controllers for specific tasks in different air traffic flow intensity in order to satisfy the demand for ensuring safe operations in the civil aviation;

  - the controllers in different clusters should be allocated with different workload when the type of airports and the air traffic flow intensity are considered.

In the sense of methodology, our work illustrates a novel chain of logic for analyzing and reducing the operation risk in civil aviation: professionals → VAP → IF → IQ → clustering → statistical inference → IQ levels → human errors. We could hope that the method developed in this work will be helpful for selecting professionals and allocating workload more reasonably subject to operation risk allowed in lots of fields.

Future work

Generally, there are plenty of forms of human factors which have impact on aviation safety. The "error/mistake", "forgetting", "omission" are three kinds of top factors which can be captured by NTT system. Statistical inference itself shows that we should design experiments for collecting data carefully and objectively as possible as we can. In future, we will pay more attention to the following aspects:

- Explore new IF influencing VAP which can be captured by the NTT system. We just take three intrinsic factors and the corresponding IQ in this work. It is still an open problem that seeking measurable VAP and their representations (IQ).

- Invite left-handers to do the tests with NTT system. Currently, all of the ATC students invited are right-handers. Thus, the visual attention is mainly focused on the left brain in the sense of neurophysiology and cognitive psychology. In the future, we will also collect the data from the left-handers.

- Increase the number of participants, from which we can utilize the statistical theory of a large number of samples. The more participants, the more confidential and convincing the levels of IQ for air traffic controllers.

- Determine the parameter K automatically. In our paper, the parameter K for clustering is set as $K = 4$ according to our experiences and knowledge of classifying air traffic controllers, which has potential impact on the statistical inference. In the future, we will design algorithms to select the parameter K automatically and find a trade-off between the complexity of clustering algorithm and difficulty of interpreting the clusters obtained subject to the facts and requirements of ATC.

Supporting information

S1 Appendix. Data of NTT. The data used in this paper is uploaded. For more questions, please contact with the first author Dr. Jing-Qiang Li (jingqiang_li@foxmail.com). (XLSX)

Acknowledgments

We really appreciate the earlier work about NTT done by the graduate students Ning Zhao and Ming-Cui Kang. Our thanks also go to Dr. Man Liang of Ecole Nationale de l'Aviation Civile and Prof. Jie-Ning Wang of Civil Aviation University of China due to their suggestions for this paper. The corresponding author's gratitude goes foremost to Dr./Professor Daniel Delahaye of Ecole Nationale de l'Aviation Civile for his constant encouragement, guidance and suggestions for doing researches in ATM and applied mathematics. The corresponding author is also greatly indebted to Professor Fei Liu of Civil Aviation University of China for her checking the revision of this paper.

Author Contributions

Conceptualization: Jing-Qiang Li, Hong-Yan Zhang.

Data curation: Jing-Qiang Li, Hong-Yan Zhang, Yan Zhang.

Formal analysis: Hong-Yan Zhang.

Funding acquisition: Jing-Qiang Li, Hong-Yan Zhang.

Investigation: Jing-Qiang Li, Hong-Yan Zhang.

Methodology: Jing-Qiang Li, Hong-Yan Zhang, Hai-Tao Liu.

Project administration: Jing-Qiang Li, Hong-Yan Zhang, Hai-Tao Liu.

Resources: Jing-Qiang Li, Hong-Yan Zhang, Yan Zhang.

Software: Hong-Yan Zhang, Yan Zhang.

Supervision: Jing-Qiang Li, Hong-Yan Zhang, Hai-Tao Liu.

Validation: Jing-Qiang Li, Hong-Yan Zhang, Yan Zhang.

Visualization: Hong-Yan Zhang, Yan Zhang.

Writing – original draft: Jing-Qiang Li, Hong-Yan Zhang, Yan Zhang.

Writing – review & editing: Jing-Qiang Li, Hong-Yan Zhang, Yan Zhang.

References

1. Wiegmann DA, Shappell SA. Human error perspectives in aviation. The International Journal of Aviation Psychology. 2001; 36 (2):341–357. https://doi.org/10.1207/S15327108IJAP1104_2

2. Shorrock ST, Kirwan B. Development and application of a human error identification tool for air traffic control, Applied Ergonomics. 2002; 33 (4): 319–336. https://doi.org/10.1016/S0003-6870(02)00010-8 PMID: 12160336

3. Isaac A, Shorrock ST, Kirwan B. Human error in european air traffic management: the HERA project. Reliability Engineering and System Safety. 2002; 75 (2): 257–272. https://doi.org/10.1016/S0951-8320 (01)00099-0

4. Scarborough A, Bailey L, Pounds J. Examining atc operational errors using the human factors analysis and classification system. Tech. rep., Civil Aerospace Medical Institute, Fedaral Aviaiton Administration. December 2005. Availavle from: http://libraryonline.erau.edu/online-full-text/faa-aviation-medicine-reports/AM05-25.pdf

5. Ohare D, Wiggins M, Batt R, Morrison D, Cognitive failure analysis for aircraft accident investigation. Ergonomics. 2007; 37 (11):1855–1869. https://doi.org/10.1080/00140139408964954

6. Statistical summary of commercial jet airplane accidents: Worldwide operations 1959–2016, Tech. rep., Boeing Commercial Airplanes, Seattle, July 2017. Available from: http://www.boeing.com/resources/boeingdotcom/company/about_bca/pdf/statsum.pdf

7. Luo XL. Study on categorized statistics of 152 cases of insufficient separation event of China civil aviation in 1990–2003, China Safety Science Journal. 2014; 14 (12): 26–32. in Chinese.

8. Dismukes K, Loukopoulos L, Jobe K. The challenge of managing interruptions, distractions, and deferred tasks, in: Proceedings of the 11th International Symposium on Aviation Psychology, Ohio State University Columbus, OH, 2001. Available from: https://interruptions.net/literature/Dismukes-ISAP01.pdf?origin=publication_detail.

9. Nolan MS. Fundamentals of Air Traffic Control, 5th Edition, Delmar Cengage Learning; 2010.

10. Isaac AR, Ruitenberg B. Air Traffic Control: Human Performance Factors. Routledge; 2017.

11. Hilburn B. Cognitive complexity in air traffic control: A literature review. EEC Note 4 (04). Available from: http://www.eurocontrol.int/eec/gallery/content/public/document/eec/report/2004/007_Cognitive_Complexity_in_ATC.pdf

12. Averty P, Collet C, Dittmar A, Athènes S, Vernet-Maury E. Mental workload in air traffic control: An index constructed from field tests. Aviation, Space and Environmental Medicine. 2004; 75 (4):333–341.

13. Loft S, Sanderson P, Neal A, Mooij M. Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. Human Factors: The Journal of the Human Factors and Ergonomics Society. 2007; 49 (3):376–399. https://doi.org/10.1518/001872007X197017

14. Findings on airport near miss. Oct. 2016. Available from: https://www.shine.cn/archive/metro/society/Findings-on-airport-near-miss/shdaily.shtml

15. Controllers punished over near collision incident at Shanghai Airport, China. Oct. 2016. Available from: https://news.aviation-safety.net/2016/10/23/5512/

16. Endsley MR, Rodgers MD. Distribution of attention, situation awareness and workload in a passive air traffic control task: Implications for operational errors and automation. Air Traffic Control Quarterly. July 1997; 6(1):25. Available from: https://www.researchgate.net/publication/235105328_Distribution_of_Attention_Situation_Awareness_and_Workload_in_a_Passive_Air_Traffic_Control_Task_Implications_for_Operational_Errors_and_Automation

17. Isaac A, Shorrock ST, Kennedy R, Kirwan B, Andersen H, Bove T. Technical review of human performance models and taxonomies of human error in ATM (HERA). 2003; pp. 30–33.

18. Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998; 20 (11):1254–1259. https://doi.org/10.1109/34.730558

19. Borji A, Itti L. State-of-the-art in visual attention modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013; 35 (1):185–207. https://doi.org/10.1109/TPAMI.2012.89 PMID: 22487985

20. Itti L, Koch C. Computational modelling of visual attention. Nature Reviews Neuroscience. 2001; 2 (3):194–203. https://doi.org/10.1038/35058500 PMID: 11256080

21. Frintrop S, Rome E, Christensen HI. Computational visual attention systems and their cognitive foundations: A survey. ACM Transactions on Applied Perception. 2010; 7 (1): 1–39. https://doi.org/10.1145/1658349.1658355

22. Tsotsos JK. A Computational Perspective on Visual Attention. The MIT Press, 2011.

23. Sun RS, Wang N. Visual features of pilot based on Markov chain under flight simulator. Journal of Beijing University of Aeronautics and Astronautics. 2013; 39 (7):897–901. in Chinese.

24. Deco G, Pollatos O, Zihl J. The time course of selective visual attention: Theory and experiments. Vision Research. 2002; 42 (27):2925–2945. https://doi.org/10.1016/S0042-6989(02)00358-9 PMID: 12450503

25. Wickens CD, McCarley JS. Applied Attention Theory. CRC press. 2007.

26. Bao S-R, Ma H-M, Li W-Y. ThuPIS: A new affective image system for psychological analysis. IEEE International Symposium on Bioelectronics and Bioinformatics, 2014. pp. 1–4.

27. Wang Y-D, Ma H-M. Identification differences among people under context of complex images. IEEE International Symposium on Bioelectronics and Bioinformatics, 2015. pp. 164–167.

28. Heinke D, Humphreys GW. Computational models of visual selective attention: A review. Connectionist Models in Psychology. 2014; 1 (4):273–312.

29. Robinson LJ, Stevens L, Threapleton CJD, Vainiute J, McAllister-Williams RH, Gallagher P. Effects of intrinsic and extrinsic motivation on attention and memory. Acta Psychologica. 2012; 141 (2):243–249. https://doi.org/10.1016/j.actpsy.2012.05.012 PMID: 22738789

30. Staugaard CF, Petersen A, Vangkilde S. Eccentricity effects in vision and attention. Neuropsychologia. November 2016; 92:69–78. https://doi.org/10.1016/j.neuropsychologia.2016.06.020 PMID: 27342258

31. Zhang H-Y, Kang M-C, Li J-Q, Liu H-T. R/S analysis of reaction time in Neuron Type Test for human activity in civil aviation. Physica A: Statistical Mechanics and Its Applications. 2017; 469 (3): 859–870. https://doi.org/10.1016/j.physa.2016.11.054

32. Chiang WY, Liang GS, Yahalom S. The fuzzy clustering method: Applications in the air transport market in Taiwan. Journal of Database Marketing & Customer Strategy Management. 2003; 11(2):149–158. https://doi.org/10.1057/palgrave.dbm.3240215

33. Portnoy L, Eskin E, Stolfo S. Intrusion detection with unlabeled data using clustering. Proceedings of ACM CSS Workshop on Data Mining Applied to Security. 2001. pp. 5–8.

34. Christopher ABA, Balamurugan SAA. Performance of different clustering methods and classification algorithms for prediction of warning level in aircraft accidents: An empirical study. International Conference on Computing, Communications and Networking Technologies (ICCCNT). IEEE Computer Society, 2013. pp.1–7.

35. Olja Č, Mario DL, Gianluca D, Aircraft safety analysis using clustering algorithms. Journal of Risk Research. 2014; 17 (10):1325–1340. https://doi.org/10.1080/13669877.2013.879493

36. Gariel M, Srivastava AN, Feron E. Trajectory clustering and an application to airspace monitoring. IEEE Transactions on Intelligent Transportation Systems. 2011; 12 (4):1511–1524. https://doi.org/10.1109/TITS.2011.2160628

37. Zhou CC, Hu DD. Safety assessment of aircraft during approach-landing stage based on grey clustering method. Journal of Safety Science and Technology. 2012; 8 (7):99–102. in Chinese.

38. Vitali S, Cipolla M, Micciche S, Mantegna RN, Gurtner G, Lillo F, et al., Statistical Regularities in ATM: network properties, trajectory deviations and delays. Second SESAR innovation days. 2012. pp. 27–29. Available from: https://www.researchgate.net/publication/234091549_Statistical_Regularities_in_ATM_network_properties_trajectory_deviations_and_delays

39. Monechi B, Ducci M, Cipolla M, Vitali S, Micciché S, Mantegna RN, et al. Exploratory analysis of safety data and their interrelation with flight trajectories and network metric. Air Transportation System Conferences. 2013. Available from: https://www.researchgate.net/publication/263782490_Exploratory_analysis_of_safety_data_and_their_interrelation_with_flight_trajectories_and_network_metric

40. Puranik TG, Jimenez H, Mavris DN. Utilizing energy metrics and clustering techniques to identify anomalous general aviation operations. AIAA Information Systems—AIAA Infotech Aerospace, 2017. Available from https://www.researchgate.net/publication/312110594_Utilizing_Energy_Metrics_and_Clustering_Techniques_to_Identify_Anomalous_General_Aviation_Operations

41. Li J-Q, Li H, Wang Y. Research on human error probability quantification method for controllers based on information processing process. Science Technology and Engineering. 2017; 17(10): 68–73. in Chinese.

42. Li J-Q, Wang B, Zhao N, Wang Y. On characteristic of attention quality of controller based on K-means cluster. China Safety Science Journal. 2017; 27(6): 13–18. in Chinese.

43. Li J-Q, Li H, Wang Y, Zhao N. Research on correlations between human errors of raw controllers and characteristics of their nervous system. China Safety Science Journal. 2017; 27(3): 13–18. in Chinese.

44. Shappell S, Wiegmann D. A Methodology for Assessing Safety Programs Targeting Human Error in Aviation. The International Journal of Aviation Psychology. 2009; 19 (3):252–269. https://doi.org/10.1080/10508410902983904

45. Friedman J, Hastie T, Tibshirani R. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Edition. Springer series in statistics, Springer, Berlin. 2009.

46. Hurst HE. Long term storage capacity of reservoirs. Transactions of the American Society of Civil Engineers. 1951; 116 (12):776–808.

47. Mandelbrot BB, Wallis JR. Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence, Water Resources Research. 1969; 5 (5):967–988. https://doi.org/10.1029/WR005i005p00967

48. Weron R. Estimating long-range dependence: Finite sample properties and confidence intervals. Physica A: Statistical Mechanics and Its Applications. 2002; 312 (1–2): 285–299. arXiv:cond-mat/0103510v2 [cond-mat.stat-mech]. https://doi.org/10.1016/S0378-4371(02)00961-5

49. Bishop CM, Nasrabadi NM. Pattern Recognition and Machine Learning. Academic Press. 2006.

50. Kulis B. Low-Rank Kernel Learning with Bregman Matrix Divergences. Journal of Machine Learning Research. 2009; 10 (1):341–376.

51. Kulis B. Metric learning: A survey. Foundations and Trends in Machine Learning. 2012; 5 (4): 287–364. https://doi.org/10.1561/2200000019

52. Airplane crash-lands into Hudson River: All aboard reported safe. Jan 2009. Available from: http://edition.cnn.com/2009/US/01/15/new.york.plane.crash/

53. van Zomeren AH, Brouwer WH. Assessment of attention. A handbook of neuropsychological assessment. Lawrence Erlbaum Associates, Ltd., 1992. pp. 241–266.

54. Mathot S, Theeuwes J. Visual attention and stability. Philosophical Transactions of the Royal Society B. 2011; 366 (1564):516–527. https://doi.org/10.1098/rstb.2010.0187

55. Mandelbrot BB, Wallis JR. Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. Water Resources Research. 1969; 5 (5):967–988. https://doi.org/10.1029/WR005i005p00967

56. Lim J, Ebstein R, Tse CY, Monakhov M, Lai PS, Dinges DF, et al. Dopaminergic polymorphisms associated with time-on-task declines and fatigue in the Psychomotor Vigilance Test. PLOS ONE. 2012; 7(3): e33767. https://doi.org/10.1371/journal.pone.0033767 PMID: 22438994

57. Visalakshi NK, Thangavel K. Impact of normalization in distributed K-means clustering. International Journal of Soft Computing. 2009; 4 (4): 168–172.

58. Graham RL. An Efficient Algorithm for Determining the Convex Hull of a Finite Planar Set. Information processing letters. 1972; 1 (4):132–133. https://doi.org/10.1016/0020-0190(72)90045-2