Data Mining the Effects of Testing Conditions on Brain Biomechanical Properties

Folly Crawford[1,2], Osama AbuOmar[3], Mike Jones[4], Roger King[2], and R. Prabhu[1,2]

Abstract *– Traumatic brain injury is highly prevalent in the United States. However, despite its frequency and significance there is little understanding of how the brain responds during injurious loading. A confounding problem is that because testing conditions vary between assessment methods, brain biomechanics cannot be fully understood. Data mining techniques, which are commonly used to determine patterns in large data sets, were applied to illicit how changes in testing conditions affect the mechanical response of the brain to high strain rate testing. High strain rate data was collected from published literature and Fuzzy C-means clustering applied to show that the data cluster based on strain rate, confirming that the brain is indeed strain rate-dependent. Self-organizing maps were used to conduct a sensitivity analysis to provide the following ranking of parameters by significance: (1) Age, (2) Strain rate, (3) Matter, (4) Testing temperature, (5) Storage temperature, (6) Diameter, (7) Thickness, and (8) Post-mortem preservation time.*

Key words *– Traumatic brain injury, principal component analysis, fuzzy C-means clustering, self-organizing maps*

I. Introduction

Traumatic brain injury (TBI) sent about 2.5 million people to the emergency room in the United States in 2013[1]. Of these people, 56,000 died and 280,000 were hospitalized [1]. TBI is most frequently caused by falls, blunt trauma, and motor vehicle accidents. TBI can cause a variety of long- and short-term health effects such as impaired memory, balance, and communication, as well as increased depression and anxiety. Further, TBI increases the risk of Alzheimer's disease and other neurological disorders, approximately 5.3 million Americans live with a TBI-related disability. Such disabilities affect individuals' relationships, productivity, and everyday living. The economic cost of TBI was estimated to be $76.5 billion in 2010, with the vast majority of this amount coming from fatal TBIs and TBIs resulting in hospitalization. It is apparent that TBI has a substantial impact on our society.

Understanding the biomechanics of traumatic brain injury mechanisms is an imperative if effective protective countermeasures are to be established. Numerous *in vitro* studies have been conducted in an attempt to improve understanding; however, the results of these studies vary by orders of magnitude in terms of the material's stress state. This can be attributed to a number of reasons, such as *in vitro* specimen age, storage and testing temperature, sample aspect ratio, material heterogeneity (white or gray brain matter, or a combination) [2-4]. This is further compounded by significant inconsistencies in the brain tissue biomechanical testing protocol. Though the ultimate goal of most of these tests is to obtain a uniaxial stress-strain response for the brain at quasi-static, intermediate, and high strain rates, the influence of the above factors on the data has not been yet quantified.

In response, unsupervised learning techniques were applied to determine how changes in brain mechanical properties relate to changes in testing conditions. One such technique, Principal Component Analysis, was used to reduce the dimensionality of the data from 10 dimensions to two without significant loss in data complexity. Fuzzy C-means clustering with a Gustafson-Kessel distance measure was used to determine what, if any, patterns exist in the data. Finally, self-organizing maps were used to conduct a sensitivity analysis on the data to determine which parameters were most important.

II. Materials and Methods

Data was gathered from four high strain rate brain compression studies [4-7]. The plot digitizer developed by Ankit Rohatgi was used to extract data from plots in each journal article [8]. The data from Prabhu et al. (2011) was "in-house" and did not require digitization. All parameters were converted to SI units and stress and strain were converted to true stress and strain, as required, for consistency.

Funding for this research was provided by the Center for Advanced Vehicular Systems at Mississippi State University, Mississippi State, MS, USA.

F. Crawford is with the Department of Agricultural and Biological Engineering and with the Center for Advanced Vehicular Systems, Mississippi State University, Mississippi State, MS 39759 USA (email: fdc33@cavs.msstate.edu).

O. AbuOmar is with the Department of Computing Sciences, Coastal Carolina University, Conway, SC 29528 USA (email: oabuomar@coastal.edu).

M. Jones is with the Department of Medical Engineering at Cardiff University, Cardiff, Wales, UK (email: JonesMD1@cardiff.ac.uk).

R. King is with the Center for Advanced Vehicular Systems at Mississippi State University, Mississippi State, MS 39759 USA (email: rking@icres.msstate.edu).

R. Prabhu is with the Department of Agricultural and Biological Engineering and with the Center for Advanced Vehicular Systems, Mississippi State University, Mississippi State, MS 39759 USA (phone: 662-325-3282; email: rprabhu@abe.msstate.edu).

Tension, compression, shear, and rotational biomechanical tests are all performed on brain tissue in order to characterize its deformation. In the case of compression, a small sample of brain matter is placed between the top and bottom loading plates in the testing apparatus and compressed uniaxially at a specific constant displacement rate while the force and displacement are measured. The brain samples are cylindrical and typically cut out of the brain with a cylindrical die. Typically, a physiologically conducive solution, such as phosphate-buffered saline (PBS), is used to immerse specimens during transportation and testing to prevent loss of moisture, which might affect the sample's stress response.

The input parameters of interest were: age of the individual from which samples were taken, sample diameter and thickness, sample storage temperature prior to testing, sample mechanical testing temperature, post-mortem preservation time, and brain matter composition. The parameter of species difference was excluded because prior work has shown that there is no significant difference in brain properties between species [6]. Brain matter composition is a categorical variable and is therefore represented numerically. Strain rate was also recorded. Because the strain rate is set by the researcher, it is normally considered an input parameter. Stress and strain were considered the output parameters of the data set. The final data set, therefore, included eight input parameters and two output responses, with approximately 3,400 data points. The data was normalized with the peak value of the corresponding variable so that all values shown in figures are between zero and one.

III. Theory and Calculations

The data mining procedure used here to identify patterns in the data was: (1) Apply PCA to reduce the number of dimensions from ten to two; (2) Analyze the dimensionally reduced data using FCM clustering; and (3) Utilize SOMs to conduct a sensitivity analysis on the data to determine which parameters are most significant.

A. Principal Component Analysis

It is difficult to visually represent and analyze a data set's patterns in high-dimensional space. As such, a technique like Principal Component Analysis (PCA) can be used to determine patterns in the data and represent it in an easier to comprehend format. The number of dimensions can be reduced without losing any data complexity. The procedure for PCA was: (1) Calculate the mean across each parameter; (2) Subtract this mean from each parameter; (3) Find the covariance matrix and its eigenvectors and eigenvalues; and (4) Determine the principal components making up the dimensionally reduced data sets using the eigenvectors and eigenvalues.

B. Fuzzy C-means Clustering

After the dimensionality reduction with PCA, the fuzzy C-means (FCM) algorithm [9] was applied to find patterns in the stress state data. Clustering tends to involve a $C \times N$ membership matrix U, where C is the number of clusters and N is the number of data points. Each element in U represents the degree of membership of a data point to a cluster:

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1N} \\ u_{21} & u_{22} & \cdots & u_{21} \\ \vdots & \vdots & \ddots & \vdots \\ u_{C1} & u_{C2} & \cdots & u_{CN} \end{bmatrix} \quad (1)$$

For a hard partitioning of the stress state data into C clusters, each membership must be zero or one. Clustering can be achieved by optimizing a cost function, and then iteratively alternating estimates of the vectors in the cost function.

FCM is then an objective function-based clustering method, where $V = \{\boldsymbol{v_1}, \dots, \boldsymbol{v_C}\}$ with the initial value $\boldsymbol{v_i}$ being the prototype for cluster i, set randomly, and

$$\sum_{i=1}^{C} u_{ik} = 1, \ \forall\, k = 1, \dots, N, \quad (2)$$

Meaning the memberships of each data vector must sum to one. The cost function for FCM can be written as,

$$J(U,V) = \sum_{i=1}^{C} \sum_{k=1}^{N} u_{ik}^{Q} d(\boldsymbol{x_k}, \boldsymbol{v_i}), \quad (3)$$

Where Q is the fuzzifier, or weighting exponent ($1 \leq Q < \infty$), and $d(\boldsymbol{x_k}, \boldsymbol{v_i})$ is the distance metric between data vector $\boldsymbol{x_k}$ and cluster center $\boldsymbol{v_i}$.

By taking the partial derivative of the Lagrangian of the cost function with respect to a specific membership value u_{rs}, it can be shown that

$$u_{rs} = \frac{1}{\sum_{i=1}^{C} \left(\frac{d(x_s, v_r)}{d(x_s, v_i)} \right)^{\frac{1}{Q-1}}}. \quad (4)$$

When using a Gustafson-Kessel distance measure,

$$d_{ik} = \sqrt{|\boldsymbol{\Sigma_i}|^{\frac{1}{D}} ((\boldsymbol{x_k} - \boldsymbol{v_i})^T \boldsymbol{\Sigma_i}^{-1} (\boldsymbol{x_k} - \boldsymbol{v_i}))}, \quad (5)$$

Where the distance is scaled by a hyper-volume approximation $|\boldsymbol{\Sigma_i}|^{\frac{1}{D}}$, and $\boldsymbol{\Sigma_i}$ is the covariance matrix of class i,

$$\frac{\partial d_{ik}^2}{\partial \boldsymbol{v_i}} = -2|\boldsymbol{\Sigma_i}|^{\frac{1}{D}} \boldsymbol{\Sigma_i}^{-1} (\boldsymbol{x_k} - \boldsymbol{v_i}). \quad (6)$$

Thus,

$$\boldsymbol{v_i} = \frac{\sum_{k=1}^{N} u_{ik}^{Q} \boldsymbol{x_k}}{\sum_{k=1}^{N} u_{ik}^{Q}}. \quad (7)$$

The GK distance measure is used here because it uses covariance matrices for each cluster, allowing the distance measure to capture the statistical features of each cluster, meaning more information is gleaned from each cluster.

C. Self-organizing maps

Kohonen maps, or self-organizing maps (SOMs), are useful for visualizing patterns in high-dimensional data in a 2-D or 3-D array [10]. The inputs for the SOM are the dimensions of the data set to be analyzed. Each input element connects to each neuron through a weight vector; after training, the SOM will create a mapping between the input space and the 2-D neuron map. The nonlinear SOM mapping uses a technique, such that vectors which are close together in the higher dimensional space, are also close together on the map.

SOM training is usually conducted on a 2-D neuron array, along with a method of data compression such that the data is more convenient to handle; however, no data complexity is lost during compression. The neuron array has

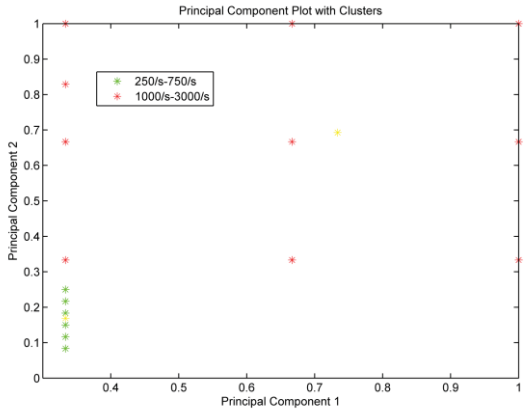
spatially defined neighborhoods, and Euclidean distance is frequently used to determine the nearest neighbors.

Using spatial neighborhoods allows for determining the similarity between the input vector and the vector of weights between the inputs and neurons. Prior to training, weights are chosen randomly and an initial learning rate and neighborhood size are chosen. When a training vector comes in, the neuron with the closest weight is found, and the winning neuron's weights are adjusted to make them even closer to the training input vector. This is repeated until convergence. The neurons in the winning neighborhood (around the winning neuron) change according to the decaying learning rate, while other neighborhoods stay the same.

IV. Results and Discussion

PCA was applied to the data, reducing the dimensionality from ten to two, with each data vector plotted in a 2-D (principal components 1 and 2 for the x and y axes, respectively) space, such that similar vectors are plotted together. PCA is necessary here to determine patterns in the stress state data which can explain the biomechanical behavior of the brain.

(a)



(b)

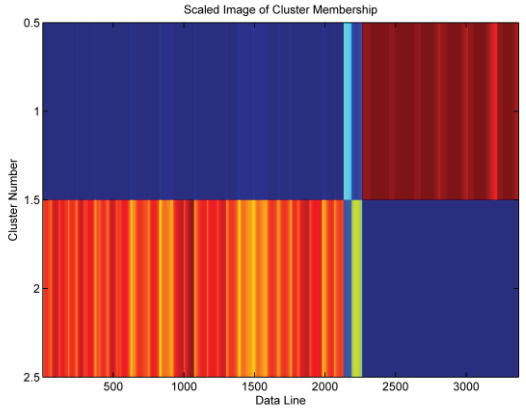


Fig. 1. (a) Clustering plot after applying FCM clustering when C=2, with included strain rate labels. (b) Plot of scaled data with two bands showing two clusters.

Following PCA, FCM was run on the stress state data using a GK distance measure. In Figure 1a, the PCA and FCM results are shown, with each of the two clusters represented by a different color. The brain samples tested at a strain rates of 1000/s, 2000/s, and 3000/s formed one cluster, with maximum strain values of 0.55, 0.54, and 0.53, and maximum stresses of 2.15 MPa, 2.33 MPa, and 3.73 MPa. Samples tested at 250/s, 350/s, 450/s, 550/s, 650/s, and 750/s formed the second cluster. The maximum strain values for these rates were 0.23, 0.30, 0.40, 0.42, 0.58, and 0.42, respectively. The maximum stresses were 0.042 MPa, 0.042 MPa, 0.391 MPa, 0.198 MPa, 0.195 MPa, and 0.313 MPa, respectively. The clustering behavior of the data confirms the notion of strain rate dependency of the brain.

In Figure 1b, the scaled data image, the three clusters are represented by three horizontal bands. Each band displays the data density of each cluster, along with the distances between data points in Figure 1a. Bluer colors represent low membership in that cluster whereas redder colors represent higher membership. These two plots demonstrate that strain rate is a key determinant in the brain's stress response.

In Figure 2, a 10 × 10 SOM from the stress state data set is shown. Values of 0.33, 0.67, and 1.00 correspond with strain rates of 1000/s, 2000/s, and 3000/s, respectively. 0.08, 0.12, 0.15, 0.22, and 0.22 reflect strain rates of 250/s, 350/s, 450/s, 650/s, and 750/s. Brain samples tested at similar strain rates tend to cluster together. Very high strain rate data (1000/s, 2000/s, and 3000/s) form clusters at the bottom, whereas the lower range of strain rates (250/s – 750/s) cluster neatly at the top.

Figure 3 shows the SOM with strain labels, used for comparison to determine the importance of each of the eight input parameters. There is some clustering with respect to strain, but there are fewer similar values clustered together than with strain rate. Figure 4a-h show the SOM with strain labels, but with one of each input parameter removed from the data set. If removing one parameter does not significantly change the clustering tendency of the output response, then it can be concluded that the input parameter is not significant.

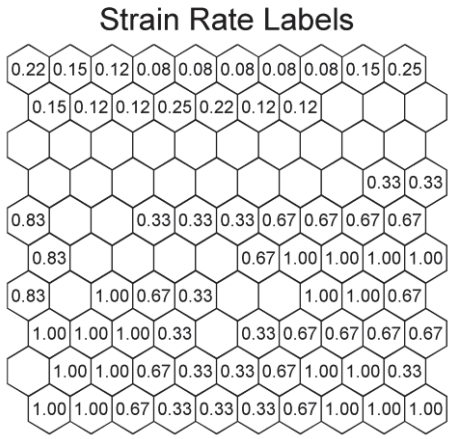


Fig. 2. A 10 × 10 SOM with respect to strain rate. There is strong clustering evident in the top with the lower range of high strain rates (250/s-750/s), with strain rates of 1000/s and 3000/s forming additional clusters at the bottom.

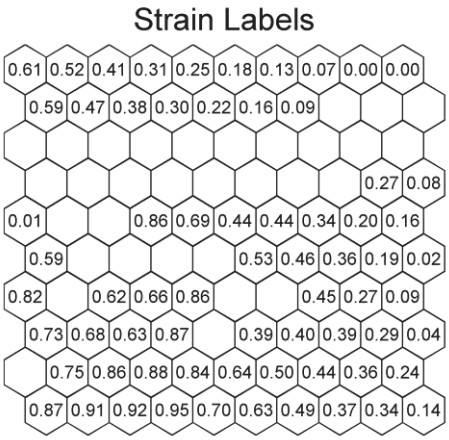


Fig. 3. A 10 × 10 SOM with respect to strain. There is some clustering, but to a lesser degree compared to strain rate.

The SOMs in Figures 4(a)-4(h) are listed in order, from most to least important. It is easy to see the significant number of changes in Figures 4(a)-4(c) compared to Figure 3, hence age, strain rate, and brain matter composition are the three most important testing parameters. Previous research has suggested that brain white matter exhibits greater stiffness than gray matter [2], and an apparent difference between the two in clustering tendency confirms this. Additionally, previous studies have suggested that the brains of older individuals are stiffer than younger brains [3]; this observation is confirmed in the present study. As described previously, the brain is strain rate dependent.

Testing temperature, diameter, and storage temperature all demonstrate a similar amount of difference, but thickness and post-mortem preservation time showed very few differences, leading to the conclusion that these two parameters make little difference in the stress state of the brain. Some have suggested that brain samples tested from several hours to almost a week have no statistically significant differences in stress response [11], though this conclusion has been disputed [12]. There are conflicting results, too, about how storage and testing temperature may affect results, with some showing a stiffer response at higher storage or testing temperatures and some at lower storage or testing temperatures [4].

V. Conclusion

Several data mining techniques were applied to a brain stress state data set; namely Principal Component Analysis, Fuzzy C-means Clustering, and Self-organizing Maps. The data set of interest consisted of multiple input parameters that may affect the stress response of the brain: age, strain rate, diameter, thickness, storage and testing temperatures, post-mortem preservation time, and brain matter composition. Strain and stress were considered the output responses to the eight input features. PCA was run on this data to reduce the number of dimensions from ten to two, followed by FCM to find patterns in the data. A sensitivity analysis was conducted using SOMs with respect to strain.

Based on the FCM results, the data tended to cluster based on strain rate and strain, with stronger clustering tendency in strain rate than in strain. These results confirm that the brain is strain rate dependent. The sensitivity analysis revealed a ranking of the testing condition parameters in the following order: age, strain rate, brain matter composition (heterogeneity), testing temperature, storage temperature, diameter, thickness, and post-mortem preservation time.

The results from these data mining techniques contribute to a greater understanding of brain tissue biomechanics. Since mechanical testing conditions can vary greatly from study to study, the results from each may be difficult to compare and may cause confusion about what stresses the brain is truly experiencing during TBI. By determining relationships between the many variables, a classifier for the data can be built so that future researchers can evaluate the stress response observed during testing and compare it to the predicted results. Further, the relationships determined here can improve the computational modeling of TBI.

Applying the proposed clustering techniques, the wide-ranging applications of data mining have been demonstrated in the context of biomechanical engineering. It is anticipated that these methods will have wider relevance to the biomedical research community.

References

[1] M. Faul, L. Xu, M.M. Wald, V.G. Coronado. *Traumatic Brain Injury in the United States: Emergency Department Visits, Hospitalizations and Deaths 2002–2006*. Atlanta (GA): Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; 2010.

[2] M.T. Prange, D.E. Meaney, S.S. Margulis. "Defining brain mechanical properties: effects of region, direction and species," *Stapp Car Crash J*. vol. 44, 205–213, 2000.

[3] S. Chatelin, J. Vappou, S. Roth, J.S. Raul, R. Willinger. **"**Towards child versus adult brain mechanical properties," *J Mech Behavior Biomed Mat*, vol. 6, pp.166-173, 2012.

[4] J. Zhang, N. Yogananda, F.A. Pintar, Y. Guan, B. Shender, G. Paskoff, P. Laud. "Effects of tissue preservation temperature on high strain-rate material properties of brain," *J Biomech*, vol. 44, pp. 391-396, 2011.

[5] F. Pervin and W. Chen. "Dynamic mechanical response of bovine gray matter and white matter brain tissues under compression," *J Biomech*, vol. 42, pp. 731-735, 2009.

[6] F. Pervin and W. Chen. "Effect of inter-species, gender, and breeding on the mechanical behavior of brain tissue," *NeuroImage*, vol. 54 supp. 1, pp. S98-S102, 2011.

[7] R. Prabhu, M.F. Horstemeyer, M.T. Tucker, E.B. Marin, J.L. Bouvard, J.A. Sherburn, J. Liao, L.N. Williams. "Coupled experiment/finite element analysis on the mechanical response of porcine brain under high strain rates," *J Mech Behavior Biomed Mat*, vol. 4, pp. 1067-1080, 2011.

[8] *A. Rohatgi. "WebPlotDigitizer." Internet:* http://arohatgi.info/WebPlotDigitizer/app/*, 2016.*

[9] J.C. Bezdek, R. Ehrlich, "FCM: the fuzzy c-means clustering algorithm," *Comp & Geosci*, vol. 10, pp. 191–203, 1984.

[10] T. Kohonen, *Self-Organization and Associative Memory*, Verlag Berlin Heidelberg: Springer, 1988.

[11] F. Shen, T.E. Tay, J.Z. Li, S. Nigen, P.V. Lee, H.K. Chan. "Modified Bilston nonlinear viscoelastic model for finite element head injury studies" *J Biomech Eng*, vol. 128 issue 5, pp. 797–801, 2006.

[12] A. Garo, M. Hrapko, J.A.W. van Dommelen, G.W.M. Peters. "Towards a reliable characterisation of the mechanical behavior of brain tissue: the effects of post-mortem time and sample preparation," *Biorheology* vol. 44 issue 1, pp. 51–58, 2007.

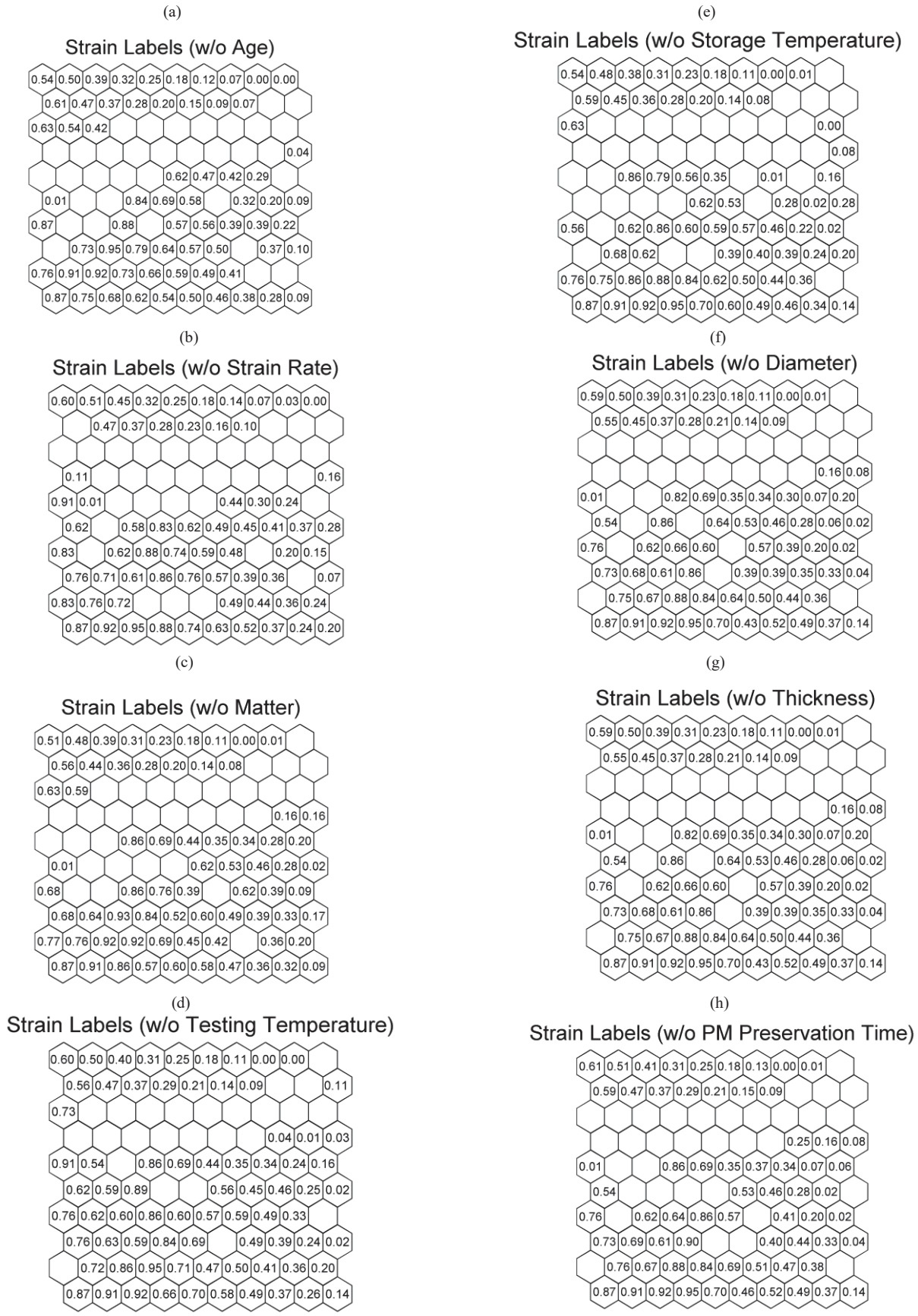


Fig. 4. 10 × 10 SOMs with respect to strain, in order of parameter importance. (a) SOM for strain with age removed. (b) Strain rate removed. (c) Matter removed. (d) Testing temperature removed. (e) Storage temperature removed. (f) Diameter removed. (g) Thickness removed. (h) Post-mortem preservation time removed.