Machine learning protocol from ultrasound data for monitoring, predicting, and supporting the analysis of dam slopes

Werickson F C Rocha
Directorate of Scientific Metrology and Technology (Dimci)
National Institute of Metrology, Quality and Technology (Inmetro)
Duque de Caxias, Brazil
wfrocha@inmetro.gov.br

Antônio U Lucena
Directorate of Scientific Metrology and Technology (Dimci)
National Institute of Metrology, Quality and Technology (Inmetro)
Duque de Caxias, Brazil
antoniounias@gmail.com

Gabriel F Sarmanho
Directorate of Scientific Metrology and Technology (Dimci)
National Institute of Metrology, Quality and Technology (Inmetro)
Duque de Caxias, Brazil
gfsarmanho@inmetro.gov.br

Rodrigo C Félix
Directorate of Scientific Metrology and Technology (Dimci)
National Institute of Metrology, Quality and Technology (Inmetro)
Duque de Caxias, Brazil
rpfelix@inmetro.gov.br

Sandro A Miqueleti
Directorate of Scientific Metrology and Technology (Dimci)
National Institute of Metrology, Quality and Technology (Inmetro)
Duque de Caxias, Brazil
samiqueleti@inmetro.gov.br

Tiago C Dourado
Directorate of Scientific Metrology and Technology (Dimci)
National Institute of Metrology, Quality and Technology (Inmetro)
Duque de Caxias, Brazil
tcdourado@inmetro.gov.br

2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA) | 978-1-6654-6283-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICMLA55696.2022.00084

***Abstract*—Dam monitoring can be used as an important indicator for dam risk management. In this study, a methodology based on machine learning and ultrasound for dam safety monitoring is presented. First, a prototype dam was built to simulate different environmental conditions. Second, ultrasound images were acquired in different areas of a prototype dam. Finally, various machine learning algorithms were applied to distinguish the different regions observed in the prototype dam. The results show that it is possible to distinguish the dam regions, which is of great value for dam safety monitoring and operation.**

***Index Terms*—dam monitoring, ultrasound, machine learning**

I. Introduction

Dams are one of the most critical infrastructures for water resources management. The activity of monitoring dams is vital for society because, in addition to promoting social development, it enhances the environmental security and the safety of the population surrounding these work areas [1], [2]. Dam monitoring can be affected by the extensive impacts of dams with high hazard potential, such as extreme external stresses, design and construction defects, material aging, and intense human activity, and has caused significant societal concern in recent years. One example was the collapse of two dams in the northeastern Brazilian state of Bahia after weeks of heavy rains, which prompted authorities to pay more attention to dam monitoring and protection [6]. The risk factors affecting dams are generally uncertain, diverse, and interrelated, increasing dam risk analysis's complexity. Therefore, measurement and instrumentation systems are needed to identify the most important risk factors and their impacts on dam risk analysis and management.

Measurement and instrumentation systems are critical for providing dam safety monitoring information. Numerous methods have been developed for nondestructive dam safety evaluation, such as infrared thermography [7]–[9], acoustic emission [10]–[12], and ultrasonic testing [13]. The use of ultrasound has been reported in many different fields [14], and it can be used by systems monitoring dam structure, and conditions [15]. According to IEC 60050-802:2011 [16], ultrasound is defined as ”acoustic oscillation whose frequency is above the high-frequency limit of audible sound (about 20 kHz)”. Devices built with ultrasound can collect data containing relevant information about the physical properties of dams. This way, these data can be used to monitor them by applying machine learning (ML) techniques. ML algorithms are usually classified into classes, depending on the goal and available information: the well-known supervised and unsupervised learning [17], and reinforcement learning [18], a more general domain. Other approaches, such as hybrid or semi-supervised learning, combine supervised and unsupervised learning.

In supervised learning, both predictive attributes/variables and targets/outputs are considered for (usually) predictive tasks; unsupervised learning, on the other hand, is (usually) used for descriptive tasks where no information about targets/outputs is considered/available. Finally, reinforcement

This work was partially sponsored by the Norte Energia SA (NESA), grant P&D-S004/2021 - Project P&D ANEEL PD-07427-0421/2021.

978-1-6654-6283-9/22/$31.00 ©2022 IEEE
DOI 10.1109/ICMLA55696.2022.00084

learning is a general framework for sequential decision-making, where an agent learns to respond to an environment in a trial-and-error dynamic to maximize its rewards.

In this work, we propose a methodology using ML and ultrasound to predict and monitor risk factors that may occur in a dam. Our study case is an ultrasound-based dam health monitoring using some piezoelectric transducers to monitor risk factors. Thus, our main contributions are:

- We build a prototype dam to study structure health monitoring.
- We describe a very straightforward methodology to apply machine learning methods to interpret the acquired dam database.
- We describe the construction of a web application based on R programming language [19] to implement machine learning algorithms on this ongoing project.

II. Machine Learning methods

A. Machine learning for dam monitoring

The literature review [20] summarizes the relevant application of ML methods for dam safety assessment in a few categories: (a) predictive models based on monitoring data to interpret and predict the dam behavior and detect anomalies; (b) classification models that look for potential failures by identifying behavioral patterns; (c) pattern recognition in high-dimensional situations; (d) quantification of uncertainties through dimensionality reduction and complex numerical simulation.

In addition to dam conditions monitoring, applications of ML algorithms in the dam environment have been identified to include water level, massif fragility (cracking), safety analysis, operational status, erosion, and displacement prediction, and monitoring [21].

In supervised learning, linear regression is the conventional data-based model to analyze the structural behavior of dams, even though support vector machines (SVM), artificial neural networks (ANN), random forest (RF) and boosted regression trees (BRT) have also been frequently used to test and monitor dams' capacity to satisfy design requirements and avoid accidents and incidents [22]. For example, some of these techniques' prediction accuracy were compared for different parameters to evaluate dam safety assessment in terms of dam displacements and leakage [23]. Applications of deep neural networks are contrasted to usual time series models, for example, to forecast reservoir water levels in short and long terms [24] and to predict single-point and multipoint concrete dam deformation [25].

In an unsupervised learning context, different – often more complex - data-driven models are considered, for instance, to detect temporal and spatial anomalies for anomaly detection and health monitoring of arch dams employing neural networks [26]; and to detect structural damage in different scenarios using a density-based approach based on kernel density maximum entropy and Bayesian optimization [27]. Unsupervised clustering algorithms are also popular techniques used, for example, to detect erosion events in the dam soil and levee passive seismic [28].

Reinforcement learning methods have also been implemented in dam operation contexts, for example, using generalized least square (GLS) and dynamic linear model (DLM) to build offline simulators to efficiently model the inflow dynamics in the upstream [[29]; or a tree-based regression in a reinforcement learning context for optimal water reservoir operation compared to traditional Stochastic Dynamic Programming [30].

B. Graphical User Interface under development to visualize the machine learning methods and their results

A graphical user interface (GUI) for monitoring dam behavior through ultrasound signals and machine learning data analysis methods is under development. It is mainly written in the open-source R [19] programming language using the Shiny [31] library – a powerful framework that facilitates web application development without requiring knowledge of HTML, CSS, or JavaScript. Shiny also takes advantage of R environment since a large number of libraries are freely available to perform data manipulation [32] and heavy computations of ML algorithms.

The GUI supports user-friendly interactive visualization of real-time ultrasound raw data signals and provides graphical and tabular data summaries of the resulting statistical data analysis from the implemented ML models. With this simple and quick tool, the users can significantly reduce processing time and take more efficient decisions using a real-time diagnosis of the dam behavior. Fig. 1 shows a screenshot of the GUI used to monitor the dam behavior using ultrasonic signals.

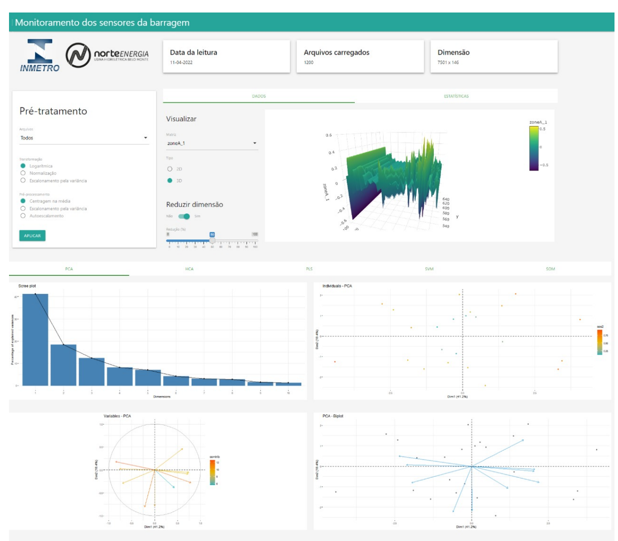

Fig. 1. Screenshot of the graphical user interface.

C. Machine learning methodology used for data analysis

In this work, we used the following machine learning methods, included in the developed web application, to study the dam monitoring state: Multiway Principal Component Analysis and Clustering methods. All of these algorithms were used to confirm that the different algorithms' results have a high-reliability level in the statistical analysis of the dataset and validate our strategy as a promising solution for application in the real dam.

The methodology used in the paper is described below. The typical 2D data matrix can be thought of as a two-way matrix where the absolute value of each element represents the amplitude measured by the system from the ultrasonic response of the applied frequency. As for the spacing between neighboring elements of the matrix, these can be sequenced in cardinal order or converted into the dimensional quantity of interest. In general, the horizontal coordinate is expressed in the magnitude of space, mm for the prototype dam, and the vertical one in time (s).

To understand the organization of the measured data, let us address the analysis by measuring just one region of the dam at a given number of discretization instants. This measurement is repeated several times, considering each measurement as an individual ultrasound spectrum in the dataset obtained in different positions inside a specific dam region. The measured dataset is arranged as follows:

$$\boldsymbol{X} = \begin{bmatrix} X_{1,1} X_{1,2} & \cdots & X_{1,I} \\ X_{2,1} X_{2,2} & \cdots & X_{2,I} \\ \vdots & \ddots & \vdots \\ X_{J,1} X_{J,2} & \cdots & X_{J,I} \end{bmatrix} \tag{1}$$

Be $\boldsymbol{X}$ a matrix with I rows and J columns that represents the organized dataset. Each row represents the measurements from the sensor at a specific position (target distance, mm). In the same way, each column represents the measurements at a particular position (side scan, mm) in the whole set of experiment trials.

In this application, extending this scheme to multiway arrays is necessary because we have different dam regions that are measured at different instants. MPCA is equivalent to performing classical Principal Component Analysis (PCA) on an unfolded version of the original multiway array. To utilize the MPCA model, the data must be a three-way array, a cube of data.

According to Linear Algebra [33], matrices (two-way arrays) are denoted (I × J), whereas three-way arrays are denoted (I × J × K). The ijk-th element is denoted $X_{i,j,k}$ where the indices run as follows: i = 1,. . ., I, j = 1,. . ., J and k = 1,. . ., K. In the three-way data (data structure), each dam region is represented by the two-way data matrix (I × J) and thus K regions can be arranged in a three-way array (K x I × J) where K is the number of regions of the dam, I is the number of variables on the second mode, in this case, the target distance, and J is the number of variables on the third mode, in this case, the side scan.

Fig. 2 shows the methodology used for data analysis. Part A illustrates the data structure obtained by the ultrasound technique. The horizontal axis represents the position (side scan), and the vertical axis represents the target distance where the ultrasound spectrum, used for investigating the prototype dam, was obtained. Item B shows the unfolding procedure of the 3-way array into a two-way array (a matrix). After this step, the traditional PCA method is performed (Item C) producing a score matrix, shown in item D. The last stage is the application of the scores obtained from PCA to cluster analysis (Item E).

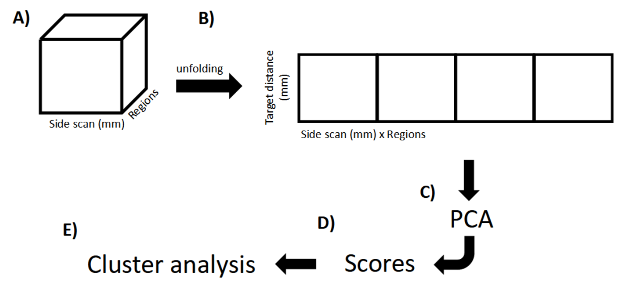

Fig. 2. Workflow of the described method: A) Data structure representation; B) Data unfolding (3-way array) into a matrix; C) PCA execution; D) Scores representation; E) Realization of cluster analysis using the scores obtained by the PCA method.

Cluster analysis (CA) is a statistical method for creating groups of similar samples [34]. CA searches for samples that are close in variable space, depending on the definition of "close". The algorithm starts with each sample representing an individual cluster. These clusters are then joined one by one according to their proximity in variable space, which is a measure of their similarity. The algorithm CA proceeds iteratively, merging the two most similar clusters at each stage until only one cluster remains. The result of this process can be represented in a diagram called a dendrogram. There are several methods for searching for clusters. In this work, Ward's method, which uses the analysis of variance approach to evaluate the distances between clusters, was used [20]. The resulting scores from PCA are then used as input variables for CA, shown in Fig. 2E.

III. Experimental Procedure

Brazil's National Institute of Metrology, Quality and Technology is leading an ongoing project to implement health monitoring of dams using ultrasound-based instrumentation.

Fig. 3 shows the steps involved in assembling the prototype dam. The filling of the acrylic tank with soil and stones is shown in Fig. 3A. The installation of the ultrasonic measuring device (hydrophone) is shown in Fig. 3B. To obtain the data, six different areas were defined, from an area with soil to another one with stones, all under water, simulating different areas of a real dam. The final prototype, developed in another tank, is shown in Fig. 3D.

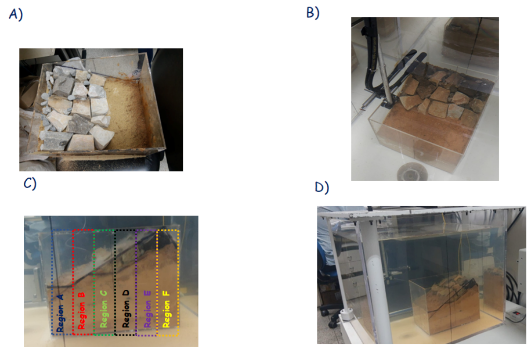

Fig. 3. Assembly of the prototype dam. A) Filling the tank with soil and stones; B) Installation of hydrophones; C) Determination of the regions for data collection; D) Overview of the built prototype dam.

IV. Results and Discussion

The first step to be investigated was the data visualization from the different dam regions. Fig. 4 shows the data from the six studied regions.

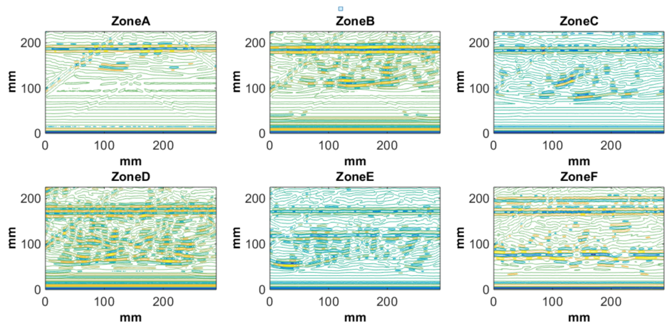

Fig. 4. Images obtained of the different dam regions using ultrasound technique.

After that, MPCA was applied to the dataset. The data were loaded into the PLS-toolbox software as a 7501x146x60 array, representing the 7501 side scan, the 146 target distance, and the 60 images (10 images from each region of the studied dam). For constructing the MPCA model, the number of principal components was chosen to capture at least 80% of the explained variance. Other methods have been proposed, but they generally do not produce very different results.

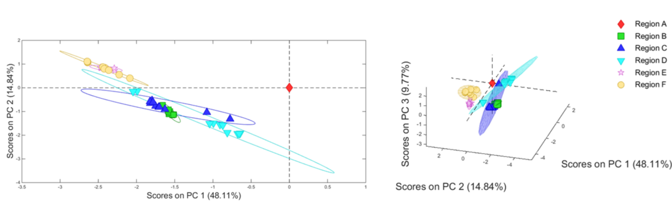

Fig. 5. Graphic scores for the developed MPCA model.

In Fig. 5, it is possible to visualize several images obtained from the six regions studied. Region A is grouped around a single point, while the other regions show a separation trend along the PC1-PC2 scores. In Fig. 4B, displaying the results in 3D dimension, it is possible to see the trend separation of regions B to F. While the project is ongoing, new images will be acquired to improve the model, but preliminary results already show that the methodology is capable of identifying different regions of the prototype dam.

The final step of the work was to create the clusters. Fig. 6 shows the clusters for each studied region. This figure can illustrate the same grouping obtained by the evaluation diagram of the developed MPCA models, which confirms the obtained results.

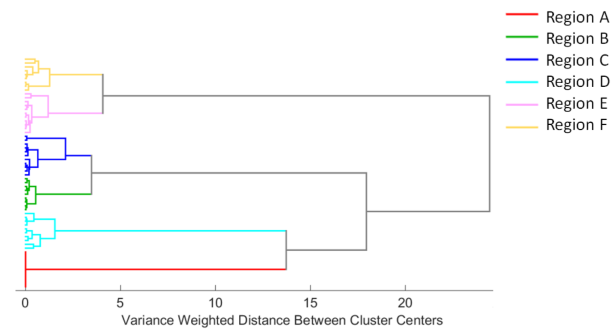

Fig. 6. Dendrogram obtained from studied region dataset.

V. Conclusions

This work presented a study based on a comprehensive approach for getting ultrasonic fingerprints in an experiment that simulates the conditions observed in a health monitoring system of a real dam structure. By combining multiway principal component analysis (MPCA), clustering, and ultrasound techniques, it was possible to understand the similarities and differences of the six different regions of the prototype dam under study. This methodology can be used to build a rational model to monitor the dam's heath and safety and is an effective way to analyze and evaluate the dam's condition.

Acknowledgements

This work was partially sponsored by the Norte Energia SA (NESA), grant P&D-S004/2021 - Project P&D ANEEL PD-07427-0421/2021.

References

[1] W. Ge et al.An innovative methodology for establishing societal life risk criteria for dams: a case study to reservoir dam failure events in ChinaInt. J. Disaster Risk Reduc. (2020)
[2] E. Mavhura Dam-induced displacement and resettlement: reflections from Tokwe-Mukorsi flood disaster, Zimbabwe Int. J. Disaster Risk Reduc. (2020)
[3] E. Ruijters et al. Fault tree analysis: a survey of the state-of-the-art in modeling, analysis and tools Comput. Sci. Rev.(2015)
[4] J. Pearl Fusion, propagation, and structuring in belief networks Artif. Intell. (1986)
[5] N. Khakzad et al. Safety analysis in process facilities: comparison of fault tree and Bayesian network approaches Reliab. Eng. Syst. Saf. (2011)
[6] L. Benassatto and A. Perobelli, Dams burst in northeastern Brazil as region hit by floods, 2021. [Online]. Available: https://www.reuters.com/markets/commodities/dam-collapses-northeastern-brazil-after-torrential-rain-2021-12-26/. [Accessed: 25-Jul- 2022].

[7] R. Zhou, H. Su, and Z Wen, Experimental study on leakage detection of grassed earth dam by passive infrared thermography, NDT & E International, 126.
[8] Y. He, B. Deng, H. Wang, L. Cheng, K. Zhou, S. Cai, and F. Ciampa, Infrared machine vision and infrared thermography with deep learning: A review, Infrared Physics & Technology, 116.
[9] K.Jang, N. Kim, and Y.-K. An, Deep learning–based autonomous concrete crack evaluation through hybrid image scanning, Structural Health Monitoring, 18(5-6):p. 1722-1737.
[10] S. Li, X. Chen, and J. Zhang, Acoustic emission characteristics in deterioration behavior of dam concrete under post-peak cyclic test, Construction and Building Materials, Volume 292, 19 July 2021, 123324.
[11] S. Li, X. Fan, X. Chen, S. Liu, and Y. Guo, Development of fracture process zone in full-graded dam concrete under three-point bending by DIC and acoustic emission, Engineering Fracture Mechanics, Volume 230, 1 May 2020, 106972
[12] S. Wang, Y.Liu, H. Zhou, Y. Zhang, Z. Wu, Q. Yang, Experimental study on failure process of arch dam based on acoustic emission technique Engineering Failure Analysis, Volume 97, March 2019, Pages 128-144
[13] M. Rucka, K. Wilde, Experimental study on ultrasonic monitoring of splitting failure in reinforced concrete, J Nondestruct Eval (2013) 32:372-382
[14] P. A. Oliveira, R. M. Baesso, G. C. Moares, A. V. Alvarenga, and R. P. B. Costa-Félix, "Ultrasound Methods for Biodiesel Production and Analysis,"in Biofuels-State of Development, 2018.
[15] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[16] International Electrotechnical Commission (IEC). IEC 60050-802. International Electrotechnical Vocabulary. Part 802: Ultrasonics. Ed 1.0. Geneva: IEC.
[17] Comparison of Chemometric Problems in Food Analysis using Non-Linear Methods Molecules 2020, 25(13), 3025; https://doi.org/10.3390/molecules25133025
[18] Y. Li. Deep reinforcement learning: An overview. arXiv preprint (2017). arXiv:1701.07274
[19] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ (2021)
[20] F. Salazar, B. M. Crookston M. A. Hariri-Ardebili, D. J. Vicente, and A. Conde. Combination of advanced numerical methods and machine learning for dam safety assessment. (2021) https://www.scipedia.com/public/Salazar_et_al_2019e
[21] F. Yahya, B. Fazli, H. Sallehudin, and M. Jaya. Machine learning in dam water research: An overview of applications and approaches. Int. J. Adv. Trends Comput. Sci. Eng, 9(2), 1268-1274. (2020)
[22] J. Mata, F. Salazar, J. Barateiro and A. Antunes. Validation of Machine Learning Models for Structural Dam Behavior Interpretation and Prediction. Water 13.19: 2717. (2021) https://doi.org/10.3390/w13192717
[23] F. Salazar, M. A. Toledo, E. Oñate, and R. Morán. An empirical comparison of machine learning techniques for dam behaviour modelling. Structural Safety, 56, 9-17. (2015)
[24] S. Mustapha, C. T. Ng, Y. Lu and P. Malinowski. Sensor networks for structural health monitoring. Journal of Sensors, 2020.
[25] S. C., Ibañez, C. V. G. Dajac, M. P. Liponhay, E. F. T. Legara, J. M. H. Esteban and C. P. Monterola, Forecasting reservoir water levels using deep neural networks: a case study of Angat Dam in the Philippines. Water 14.1 (2021): 34. https://doi.org/10.3390/w14010034
[26] X. Shu, T. Bao, Y. Zhou, R. Xu, Y. Li, and K. Zhang. Unsupervised dam anomaly detection with spatial–temporal variational autoencoder. Structural Health Monitoring, p. 14759217211073301. (2022)
[27] K. A. Eltouny, and X. Liang. Bayesian-optimized unsupervised learning approach for structural damage detection. Computer-Aided Civil and Infrastructure Engineering, 36(10), 1249-1269. (2021)
[28] W. Belcher, T. Camp and V. V. Krzhizhanovskaya. Detecting erosion events in earth dam and levee passive seismic data with clustering. In 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA) (pp. 903-910). IEEE. (2015, December)
[29] X. Wang, T. Nair, H. Li, Y. S. R. Wong, N. Kelkar, S. Vaidyanathan and M. Tambe. Efficient Reservoir Management through Deep Reinforcement Learning. (2020) arXiv preprint arXiv:2012.03822.
[30] A. Castelletti, S. Galelli, M. Restelli, and R. Soncini-Sessa. Tree-based reinforcement learning for optimal water reservoir operation. Water Resources Research, 46(9). 2010.
[31] W. Chang, J. Cheng, JJ Allaire, C. Sievert, B. Schloerke, Y. Xie, J.Allen, J. McPherson, A. Dipert and B. Borges. shiny: Web Application Framework for R. R package version 1.7.1. (2021) https://CRAN.R-project.org/package=shiny
[32] H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. A. McGowan, R. François, and H. Yutani. Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, (2019) https://doi.org/10.21105/joss.01686
[33] Linear Algebra and its Applications, Strang Cengage Learning; 4th ed. Belmont, CA: Thomson, Brooks/Cole (2006)
[34] The Journal of Allergy and Clinical Immunology: In Practice Volume 9, Issue 7, July 2021, Pages 2792-2801.e4