2023 IMEKO TC-4 International Conference on
Metrology for Archaeology and Cultural Heritage
Rome, Italy, October 19-21, 2023

Cluster Analysis to identify Microclimate Patterns in a Multi-room Film Archive

Lisa Vergelli[1], Francesca Frasca[2], Chiara Bertolin[3], Gabriele Favero[4], Anna Maria Siani[2]

[1]*Dept. of Earth Science, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy,* lisa.vergelli@uniroma1.it
[2]*Dept. of Physics Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy,* f.frasca@uniroma1.it, annamaria.siani@uniroma1.it
[3]*Dept. of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Richard Birkelands vei 2B, Gløshaugen, Trondheim, Norway,* chiara.bertolin@ntnu.no
[4]*Dept. of Environmental Biology, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy,* gabriele.favero@uniroma1.it

Abstract – **Inspections of the state of preservation of each film stored in multi-room archives can be time-consuming and expensive. For this reason, the common approach is to perform inspections on films only in rooms selected randomly by the archive curators, that might be not accurately representative of different conservation conditions. This paper presents an objective methodology to select rooms in film archives for inspection, based on their microclimate conditions. A multivariate statistical approach (cluster analysis) was used to objectively identify clusters of rooms with specific microclimate patterns. The proposed methodology was applied on a multi-room archive housing several cellulose acetate films located in the Mediterranean area. The outputs of this work led to a reduction from 26 rooms to 4 clusters representing 4 homogeneous microclimate patterns. These allowed to target specific rooms where to assess the likely occurrence of the vinegar syndrome for cellulose acetate films triggered by thermo-hygrometric storage conditions.**

I. INTRODUCTION

In the case of cinematographic archives, it is pivotal to combine the air quality control with the monitoring of thermo-hygrometric conditions, as both atmospheric pollutants and microclimate can affect the preservation of film materials [1]. For example, the Vinegar Syndrome (VS) is the first sign of deterioration of cellulose triacetate (CTA) films. It is a spontaneous and autocatalytic process triggered by unfavourable microclimate conditions [2]. It involves the depolymerization of cellulose and results in the emission of an organic pollutant, the acetic acid (AA), into the air [3]. It should be noted that VS can also be triggered by other co-factors acting synergistically (e.g., the chemical composition of the film, the year of production, the manufacturer, previous storage conditions, etc.), so the principal driver is often difficult to determine.

Nowadays, it is common practice in film archives to perform periodic inspections using passive samplers for AA detection [4] (e.g., Acid Detection - AD - strips® devised by the Image Permanence Institute - IPI - in Rochester, USA [5]) and to continuously measure the temperature (T) and relative humidity (RH) indoors. To prevent the deterioration of CTA films, cold and dry storage is internationally recommended [6], but in practice it is strictly challenging to keep the suggested indoor climate conditions. Obsolete climate control systems may not be sufficiently energy-efficient, while high-performance systems may not be cost-effective. In addition, periodic air quality inspections in randomly selected rooms are not accurate and representative of the different storage conditions occurring in a multi-room archive, and the assessment of the air quality in all archive rooms is expensive and time-consuming.

Multivariate statistical methods have been applied in several studies to characterize the microclimate patterns. To cite but a few: to optimise the number of thermo-hygrometric sensors [7], and/or their deployment [8,9], and/or their maintenance costs and redundancy in data analysis [10]; but the case studies investigated were museums or historical buildings, never cinematographic archives equipped with climate control systems.

The overall objective of this research concerns the application of a multivariate statistical approach, i.e., the cluster analysis (CA), to identify microclimate patterns in multi-room film archives, in order to objectively select a specific number of representative rooms (one per pattern) in which to perform inspections to assess the occurrence of VS affecting CTA films by means of AA passive samplers.

The paper is structured into the following sections:

- *Materials and methods* outline the features of the multi-room CTA film archive selected as a case study (A), the collection of microclimate data at the case study (B), the statistical approach of the *k-means* CA (C) applied to identify microclimate patterns among rooms using long time series of thermo-hygrometric data collected *in situ*, the type of passive sampling of gaseous acetic acid employed and the variety of CTA films inspected (D).
- *Results and conclusions* discuss the outcomes of

applying CA to thermo-hygrometric data and its possible implications. Finally, these sections present a preliminary investigation of the occurrence of VS affecting CTA films by means of AA passive sampling performed in rooms representative of the identified microclimate patterns.

II. MATERIALS AND METHODS

A. The case study

In compliance with the policy agreement, the site that generously provided the data of T and RH should remain anonymous. The CTA film archive building is located in the Mediterranean area. It consists of several rooms below ground level. All rooms open onto a corridor. There are no windows, and there are four entrances from outdoors and one from indoors. Each of the rooms is equipped with an independent air conditioning and dehumidification system. Visitors are not allowed, except in rare cases, only staff may access for collection management and conservation practices, and rooms' cleaning is intentionally minimized.

B. The microclimate data collection

The indoor monitoring system consists of one thermo-hygrometer per room, installed in 26 climate-controlled rooms in August 2022. The microclimate monitoring is still in operation: both T and RH have been measured systematically and continuously since September 2022. Technical details of the instruments are given in Table 1. The set sampling rate (i.e., time between two consecutive observations) was set to 10 minutes, and the position of the thermo-hygrometers was designed not to interfere with the ease of staff access to the CTA films (i.e., ensuring *compactus* sliding and film handling). In this study, we analysed data collected over the period from September 2022 to March 2023.

Table 1. Technical specifications of thermo-hygrometric sensors employed at the multi-room CTA film archive.

Microclimate variable	**Sensor specifications**	**Range and uncertainty**
Temperature (T)	Platinum (Pt) 100 thermistor	-40 to +60°C ± 0.1°C
Relative humidity (RH)	Thin-film hygroscopic condenser	0 to 100% ± 1.5%

C. Statistical analysis of microclimate data

The multivariate statistical method of CA was applied to the microclimate data recorded *in situ* within the case study archive's rooms. The aim was to search for a discrete set of microclimate patterns, i.e., to classify the 26 storage rooms into a number of clusters. All collected data of both T and RH were aggregated into k clusters using the "*k-means*" clustering analysis (i.e., an iterative algorithm, schematised in Figure 1).

The Silhouette index (S) was used to assess the quality of clustering and to identify the optimal number of clusters (k). S assesses how similar a room is to the cluster to which it belongs in comparison to the other identified clusters. The value of S ranges from -1 to +1: a value close to +1 denotes a high similarity of the room to its cluster and a low similarity to the other clusters. The lower threshold of S was preset to 0.5 according to [8], considering that if $S > 0.5$ the clustering configuration would be appropriate, whereas if $S < 0.5$ it would result in an inappropriate k (too many or too few clusters). Starting with $k = 2$ (assuming at least two subsets of rooms with significantly different T and RH observations), by an iterative process k was increased (setting $k <$ total number of rooms as the upper limit, otherwise no clustering would have been achieved).

The output for each cluster is a centroid, and for each room it is a Euclidean distance (i.e., the distance of the room's T and RH data from their mean value represented by the centroid of the cluster to which the room belongs). The T and RH dataset of a room was aggregated to the centroid of a cluster if the mean values of the room's T and RH data had a minimum anomaly from the value of the centroid of that cluster. CA was applied to weekly and monthly averages of the indoor microclimate observations, as hourly or daily scaling was not required since clustering clearly emerged with the average values.

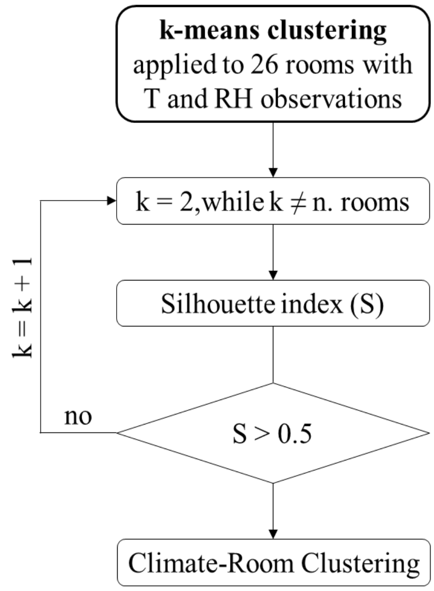

Fig. 1. Workflow of the k-means clustering, figure reproduced from [8], with the permission of the authors under CC BY license, adapted to this contribution.

D. Acetic acid passive sampling

For each identified cluster of rooms, we planned to inspect the cans of 25 films (all 35mm format, from different manufacturer, both colour and black and white, dated from 1950 to 1991) in order to understand the relationship between the conservative T and RH and the presence of AA possibly released from single CTA films. The higher the concentration of AA within the film can, the more advanced the film's VS stage.

Danchek® passive samplers (i.e., coloured patches, similar to the conventionally employed IPI's AD strips®, developed and distributed by Dancan Cinema Services

S.r.l., Denmark) were used. These are paper strips, soaked with a pH indicator, allowing ranking the film degradation level from good to critical according to their shift in colour, after a 24-hour exposure at room T within film cans [11]. Their colours change from blue to shades of green and yellow in the presence of increasing amount of AA accumulated within the CTA film cans due to the film VS.

III. RESULTS AND DISCUSSION

It is common practice to analyse a dataset covering at least a period of 1 year for a microclimate characterization. However, for the purpose of this study, since the archive building is T-conditioned and dehumidified, and therefore it is less affected by seasonal variability and external climate forcing, a data collection period of 7 months is assumed to be sufficient.

The CA was applied to the weekly and monthly averages of T and RH data, defining the microclimate patterns for the 26 rooms. Four multi-room subsets were identified, based on a Silhouette index above the set threshold (i.e., $S > 0.5$ for $k = 4$). The *k-means* algorithm was iterated 3 times, in order to refine the rooms' grouping, by excluding after each iteration the room that emerged as a cluster of a single room (one room was excluded after iteration I, and another room after iteration II). The output of the iteration III is thus a clustering referring to 24 rooms (i.e., 26 - 2 = 24). The number of rooms belonging to each cluster is reported in Table 2, along with details on the T and RH values of the centroids for the four clusters. Looking at the T outputs of the CA, a small difference among the four clusters can be noticed (the lowest T value of centroid is 10.9°C for cluster 4, while the highest T value of centroid is 15.0°C for cluster 2). This is quite expected due to the thermal control system of the rooms which results to be homogeneous. Considering the RH outputs of the CA, the values of the centroids indicate: dry conditions for cluster 1, medium RH conditions for cluster 2, humid conditions for cluster 3, and very humid conditions for cluster 4.

Table 2. Centroid T and RH values for the microclimate clusters identified among the rooms of the film archive.

Cluster	Number of rooms	Centroid T value (°C)	Centroid RH value (%)
1	9	12.4	15.6
2	5	15.0	57.4
3	7	11.6	72.3
4	3	10.9	80.0

The average monthly values for each cluster show the temporal behaviour of the four microclimate patterns corresponding to the four clusters of rooms, both for T (Figure 2) and for RH (Figure 3) data. Lower T monthly averages were observed in the winter months, i.e., from November to February, than in the autumn and spring months, revealing that the rooms might be slightly influenced by the seasonal evolution of the outdoor T. On the other hand, RH monthly averages were significantly different from cluster to cluster, while remaining stable over time for each of them.

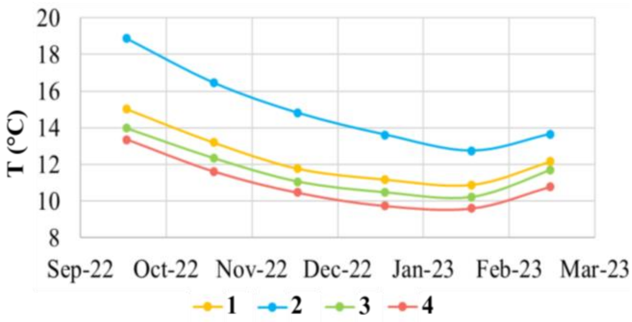

Fig. 2. Temporal trend for four clusters' monthly averages of temperature observations. Data from September 2022 to March 2023.

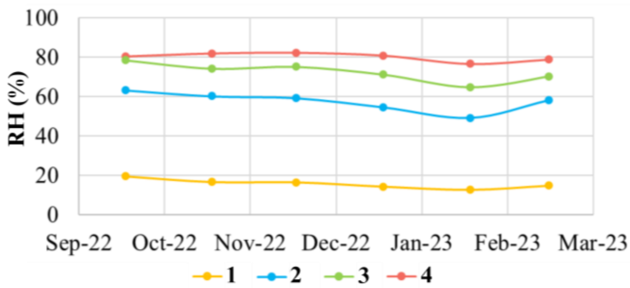

Fig. 3. Temporal trend for four clusters' monthly averages of relative humidity observations. Data from September 2022 to March 2023.

For a complete thermo-hygrometric characterization, a common univariate statistical method for microclimate analysis (i.e., the Box-and-Whisker plot) was applied to the raw T and RH data recorded from September 2022 to March 2023 in the rooms of the film archive. This type of graph provides a synthetic and easy-to-read rendering of the T and RH data per room, identifying outliers and comparing medians, as well as assessing microclimate similarities and differences among rooms. The median, 25th, and 75th percentile values correspond, respectively, to the inner, lower, and upper line of the box. The whiskers represent the expected data variation, and their length corresponds to the minimum or maximum value of the dataset when there are no outliers (these latter are the values beyond 1.5 times the interquartile range, i.e., the difference between 25th and 75th percentile, below the 25th percentile or above the 75th percentile). Otherwise, if there are outliers, the length of the whiskers is set to 1.5 times the interquartile range from the top and bottom of the box.

For the sake of brevity, the Box-and-Whisker plots are presented for four rooms only. These rooms are the ones selected to perform the objective and accurate inspection of the occurrence of VS affecting CTA films by means of AA passive samplers. Each of these rooms, indeed, represents one of the clusters, hence one of the thermo-hygrometric patterns identified. Box-and-Whisker plots are shown for T (Figure 4) and for RH (Figure 5), respectively. Considering that the outliers are the values above the mean value plus 3 standard deviations or below the mean value minus 3 standard deviations, 6.2% of the T data for cluster 4 are outliers, as well as 3.5% of the RH data for cluster 1. Outliers are drawn graphically as circles. The data reveal an overlap in the T boxes (for both the whiskers and median values), whereas a greater variability occurs for the RH boxes. It can be inferred that T control is more effective than RH control in these rooms, and that the clustering is more driven by RH than by T.

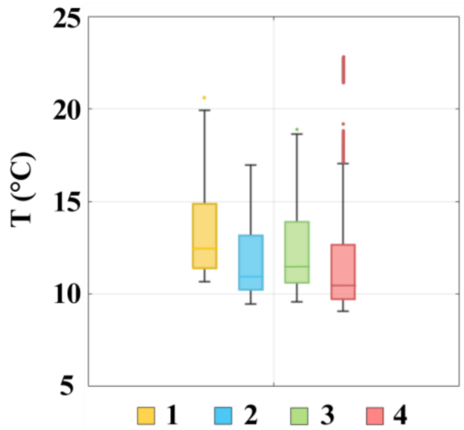

Fig. 4. Box-and-Whisker plots, comparison of T values in four rooms representative of the identified clusters. Data from September 2022 to March 2023. The legend refers to the clusters to which the single rooms belong.

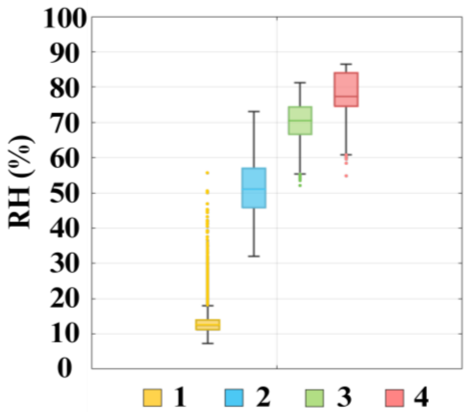

Fig. 5. Box-and-Whisker plots, comparison of RH values in four rooms representative of the identified clusters. Data from September 2022 to March 2023. The legend refers to the clusters to which the single rooms belong.

In addition to the characterisation of microclimate patterns, the concentration of AA was monitored to assess whether (and if so, the amount) it was released from the CTA films in VS (at different stages). A preliminary evaluation was carried out through Dancheck® AA passive samplers deployed within 25 film cans per each room representative for each microclimate pattern, for a total number of 100 film inspected. Results are reported in Table 3, showing that most films are not affected by VS.

No further conclusions can be drawn, as it seems that the most affected films belong to the rooms constituting cluster 1. Presumably, the intent was to store the films that appeared most vulnerable when entering the archive under the least risky long-term storage conservation conditions.

Table 3. Percentages of Dancheck® passive samplers' colour-changes from blue (good film level) to yellow (critical film level). 25 CTA film cans per cluster surveyed.

Cluster					
1	52%	4%	13%	17%	13%
2	48%	52%	0%	0%	0%
3	92%	8%	0%	0%	0%
4	28%	64%	8%	0%	0%

IV. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, it was demonstrated that a new application of multivariate statistical method in the field of cultural heritage management provides a successful and accurate assessment of indoor microclimate patterns in a multi-room building intended for the preservation of climate-sensitive materials. In this context, a cinematographic archive housing CTA films was chosen as a case study.

A continuous monitoring campaign was conducted in 26 rooms of this archive from September 2022 to March 2023 (a data collection considered adequate since the building is air-conditioned, dehumidified, and located underground). The thermo-hygrometric data collected *in situ* were analysed using the *k-means* CA statistical method. The proposed approach provides an effective and solid procedure for efficiently grouping physically consistent climate environments, with similar thermo-hygrometric features, into a number of clusters (in our case, $k = 4$). However, it should be noted that reducing the number of clusters may not capture the indoor thermo-hygrometric variability of multi-room buildings, particularly when the S value is close to the set threshold (in our case, 0.5). Therefore, it is advisable that the interpretation of the CA outputs should be performed according to the physical mechanism under investigation. In general, the CA reveals the relationships between the T and RH variables and their temporal evolution, and it is effective in the case of long monitoring periods because, once the CA has been performed, further time-consuming computations of large amounts of microclimate data are reduced by removing redundant information. In the case of the CTA film archive analysed in this contribution, the clustering of the storage rooms highlighted a higher thermal homogeneity (although with slight seasonal variations, coherent among clusters), and a lower hygrometric homogeneity (but coherent over time).

The microclimate characterization was complemented with the use of a univariate statistical method (i.e., Box-and-Whisker plots) providing a direct comparison of the data recorded by the thermo-hygrometers, one per room, in terms of median values, interquartile ranges, and outliers, separately for T and RH. The focus was on the rooms selected as representative for the assessment of the climate-induced triggering of the CTA films' degradation. It should be emphasised that, while achieving Box-and-Whisker plots is easy through user-friendly statistical programs (no advanced expertise required), CA is more complex and requires the user to have advanced expertise in statistical data processing (referring to the creation of an *ad hoc* script in a statistical program, to the workflow of the iterative clustering algorithm, and to the calculation of the Silhouette index). On the other hand, the impossibility of a univariate statistical method to combine the T and RH contributions, as the two datasets can only be processed separately, is overcome by the multivariate CA, which identifies aggregations of rooms on the basis of their microclimate features, also providing the temporal evolution of the monthly averages of T and RH for each microclimate pattern.

The detailed knowledge achieved on the microclimate differences between the rooms of a multi-room film archive provides promising basis for the implementation of room-specific preventive conservation strategies, focusing on the conservation priorities and potential criticalities for each identified cluster. Based on the outcomes obtained in the multi-room cinematographic archive, it becomes possible to integrate the microclimate characterization with the evaluation of the indoor air pollutants. Indeed, since the release of AA, i.e., the VS, is the first sign of climate-induced degradation observed for CTA films, it appears significant to investigate how the occurrence of VS can be triggered by the microclimate patterns outlined at the film archive. Inspections of VS occurrence may be performed on the basis of the clusters identified by the CA method, saving economic resources and time. In fact, it results to be sufficient to inspect only a number of rooms equal to the number of clusters identified (in our case, four), thereby representative of the different film storage microclimates. The individual room is then chosen not among all the monitored rooms (in our case, 26), but among the rooms objectively grouped under each identified cluster.

In the future, the proposed approach can be tested on other film archives in the same Mediterranean area or in different climatic areas to refine the multivariate statistical method in order to understand how specific microclimate patterns can affect degradation of the film materials under investigation. Variations in the outdoor climate conditions should be also taken into account to maintain specific indoor microclimate conditions, while both minimizing energy costs and maximizing the energy efficiency of the storage building in terms of thermo-hygrometric control. Finally, further research will be devoted to integrating other variables (e.g., AA concentration) into the adopted multivariate statistical method (i.e., CA) in order to relate the macroscopic measurable features of the CTA films (e.g., shrinkage, embrittlement, colour change) to environmental agents. Thus, both physical and chemical parameters will be considered.

ACKNOWLEDGEMENTS

Vergelli L. and Frasca F. acknowledge fellowship funding from MUR (Ministero dell'Università e della Ricerca) under PON "Ricerca e Innovazione" 2014-2020 (ex D.M. 1062/2021 and ex D.M. 1062/2021, respectively). Authors are grateful to PhD Alessia Caratelli for her support in microclimate analysis and for her useful suggestions.

REFERENCES

[1] A. del Amo Garcia, J.S. de Oliveira, B. Pritchard, D. Walsh, "Film preservation", International Federation of Film Archives (FIAF) - Technical Commission, 2004.

[2] N.S. Allen, M. Edge, T.S. Jewitt, C.V. Horie, "Initiation of the Degradation of Cellulose Triacetate Base Motion Picture Film", on Journal of Photographic Science, vol. 38, 1990, pp. 54-59.

[3] N.S. Allen, M. Edge, J.H. Appleyard, T.S. Jewitt, "Degradation of historic cellulose triacetate cinematographic film: the Vinegar Syndrome", on Polym. Degrad. Stab., vol. 19, 1987, pp. 379-387.

[4] J. L. Bigourdan, "Preservation of acetate base motion-picture film: from stability studies to film preservation in practice" in "The vinegar syndrome: a handbook, prevention, remedies, and the use of new technologies", Bologna, The GAMMA Group, 2000.

[5] IPI - Image Permanence Institute, User's Guide for AD strips® – film base deterioration monitors, 2022, (assessed on August 2023) https://s3.cad.rit.edu/ipi-assets/publications/adstrips_instructions.pdf

[6] ISO 18934:2011, "Imaging materials - Multiple media archives - Storage environment", International Organization for Standardization (ISO), 2011.

[7] A.M. Siani, F. Frasca, M. Di Michele, V. Bonacquisti, E. Fazio, "Cluster analysis of microclimate data to optimize the number of sensors for the assessment of indoor environment within museums", on Environ. Sci. Pollut. Res., vol. 25, 2018, pp. 28787-28797.

[8] F. Frasca, E. Verticchio, P. Merello, M. Zarzo, A. Grinde, E. Fazio, F.J. García-Diego, A. M. Siani, "A Statistical Approach for A-Posteriori Deployment of Microclimate Sensors in Museums: A Case Study" on Sensors - Special Issue: Sensors for Cultural Heritage Monitoring 2022-2023, vol. 22, No. 12, 4547, 2022.

[9] N. Bucarelli, N. EL-Gohary, "Consensus-based clustering for indoor sensor deployment and indoor condition monitoring," on Building and Environment, 2023, https:// doi.org/10.1016/j.buildenv.2023.110550.

[10] S. Ramirez, M. Zarzo, F.J. García-Diego, "Multivariate Time Series Analysis of Temperatures in the Archaeological Museum of L'Almoina, Valencia, Spain", on Sensors - Sensors and Data Analysis Applied in Environmental Monitoring, vol. 21, No. 13, 4377, 2021.

[11] S. Hackney, "Colour measurement of acid-detector strips for the quantification of volatile organic acids in storage conditions" on Studies in Conservation, vol. 61 (sup1), pp. 55-69, 2016.