A multiple Linear Regression Model to predict indoor temperature trend in historic buildings for book conservation: the case study of “Sala del Dottorato” in Palazzo Murena, Italy.

Elisa Moretti¹, Maria Giulia Proietti² and Ettore Stamponi²

¹ Department of Engineering – University of Perugia, Via G. Duranti 93 Perugia, Italy.
² CIRIAF (Interuniversity Research Centre on Pollution and Environment “Mauro Felli”) – University of Perugia, Via G. Duranti 67 Perugia, Italy

3 Author to whom any correspondence should be addressed: elisa.moretti@unipg.it

Abstract. The indoor climate of historic buildings is governed by the desire to preserve them, their interiors and to ensure human comfort. For preservation of cultural heritage and libraries, relative humidity and temperature are very important parameters, including their amplitudes and changes rate in time. In the present study an experimental campaign of thermo-hygrometric parameters inside of “Sala del Dottorato”, located in Palazzo Murena (Perugia), is carried out. In this room a great number of rare and ancient books are preserved. The paper deals with the study and the evaluation of the correlation between outdoor and indoor microclimate conditions in the room, to ensure the proper conservation of the books; it is aimed at understanding how the two parameters follow outdoor variations and how the hygrothermal inertia of the building can mitigate these variations. This is done, specifically for temperature, which is the most critical aspect. Thanks to a continuous monitoring system for indoor and outdoor thermo-hygrometric parameters, a Multiple Linear Regression model is developed in order to predict and analyse the indoor temperature trend. This model allows to estimate a future forecast of this parameter and to predict in advance critical conditions for correct conservation.

1. Introduction
Preventive conservation is fundamental to limit and to slow down the natural damage processes and the subsequent restoration work of book heritage [1-2]. Indoor micrometric conditions play a key role in the conservation and maintenance of library collections, especially in historic buildings [3-5]. For this reason, the hygrothermal conditions must respect precise ranges of temperature and relative humidity, according to literature and technical standards [6-11]. Anyway, the conservation of book collections requires the evaluation of several factors, including the knowledge of their original microclimate to evaluate if the current one is correct, and also an ability to anticipate the future. In this context, a continuous monitoring system for indoor thermo-hygrometric parameters is essential and predicting indoor microclimate trend is important to detect critical situations. However, predicting indoor microclimate trend, such as the temperature one, is a complicated task, especially in historical building...
because it depends on multiple variables such as building characteristics (of which we have little knowledge), weather conditions, occupancy rate and use of indoor space [13].

Regression analysis is one of the methods most commonly used to measure the relationship between two or more variables. In particular, Multiple Linear Regression is often utilized to predict the behavior of the dependent variable according to more independent variables such as weather data and heating system of the building [12-13]. The paper deals with the development of a Multiple Linear Regression (MLR) model to predict and analyze daily indoor air temperature trend of a historical building. The case study was carried out in the “Sala del Dottorato” in Perugia. This work can become a first step towards a more complex analysis concerning also indoor relative humidity and lighting parameters in order to investigate and to predict the level of artificial and natural lighting, thus to mitigate the damage caused by UV radiation.

2. The case study
The Sala del Dottorato (Hall of Graduate) is located on the first floor of Palazzo Murena, which is the headquarters of the University of Perugia since 1810. Originally, Palazzo Murena was an Olivetan Orders’ male monastery, it was built by the architects Carlo Murena and Luigi Vanvitelli in 1739.

The Sala del Dottorato is one of the University’s most magnificent rooms, by its decorations of arched-ceiling and collection of rare and ancient books it stores: the library collection of “Fondo Antico”. Originally it was the monastery library, then, from the Sixties and up to a few years ago, the hall hosted the discussion of thesis, while nowadays it is used occasionally for important events. A smaller atrium precedes the main hall, which has a long inner balcony with a balustrade (mezzanine). Both of levels of the room are covered by wooden bookcases, which hold about 10500 rare and ancient books of the University, printed from the 15th to the 19th century. The collection is composed of a series of organic materials such as paper, tissue, animal hair and adhesive substances, which may alter and perish in time.

3. Methodology
In the present study, multi-linear regression analysis was carried out in order to predict and analyze daily indoor temperature trend in the investigated historic hall.

Multiple linear regression is one of the methods to measure the relationship between two or more variables by fitting a linear equation to observed data. In this way, it is possible to predict the behavior of a dependent variable, according to one or more independent ones [13-18].

Generally, this relationship between the dependent and the independent variables is given as presented in Equation (1):

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \tag{1}
\]

where:
- \( y \) is the predicting daily indoor temperature (°C);
- \( x_i \) is the value of the chosen parameter (independent variables);
- \( \beta_i \) is the corresponding regression coefficient;
- \( \varepsilon \) is the statistical error.

The linear regression coefficients are obtained by the method of least squares, which tries to minimize the sum of the squares of the error terms.

In general, the accuracy of the models is evaluated by the coefficient of determination (R²), however in the case of multiple linear regression is more correct to consider the adjusted coefficient of determination (R²_adj), a statistical index that provides information about the goodness of fit of a model. This parameter can take values from 0 to 1: more R²_adj is close to 1 and more the corresponding variable is good for the model. Instead, the significance of the MLR model is tested by using the Fisher-Snedecor test with p-value (F): if the p-value is less than 0,05 the model is statistically significant [15].
The validation phase is an important step in developing a model, in fact it is essential for demonstrating its accuracy. In this case, the dataset was divided into two groups: yearly data (from October 2019 to October 2020) were used in order to develop the regression model, whereas the data of the last two months of 2020 (winter season) were used for testing the model, taking into account the range of accuracy.

4. Results and discussion

4.1. Data analysis
The data required to develop the model were collected every day over a period of 12 months from October 2019 to October 2020. The monitoring campaign of indoor air temperature concerned a specific point of the hall, K0 (Figure 1 and Figure 2): the first letter specifies the bookcase, while the number specifies the level of the room (0 for the floor level). This measurement point was chosen for its central position, in front of the wall with the windows thus to receive also the solar light. However, it can’t be considered totally representative of the indoor temperature of the whole hall, because of its large dimensions, especially in height (about 9 meters); in fact, a previous study [5] showed that there is a certain temperature gradient of about 0.5°C (in some cases also 1°C) between the level floor and mezzanine and between some points of the same level floor. This is an important aspect to take into account during the validation model.

The books conserved in the room are made of different organic material such as paper, wood, leather, parchment and some of them are stored in the same place since 1810.

The standard of conservation [6-11] suggests different ranges for temperature depending on the material, so the optimal range cannot be strictly defined. However, it is possible to define a general interval for this microclimate parameter which can be considered good for most materials: 14 – 22°C with a maximum daily variation of 3°C to prevent physical damage, which may happen if the temperature variations are associated with relative humidity ones. Indeed, temperature and relative humidity are closely linked and they require a simultaneous monitoring. The temperature range stated above is the narrowest one, which allows the damages to be minimized. It includes most of organic materials such as paper, parchment, wood and leather.

In Figure 3 the indoor air temperature trend and the outdoor one is reported for the above mentioned period in the bookcase K0. The data were recorded every 10 minutes and they were processed on a daily basis. Obviously, the indoor trend is influenced by the outdoor one, especially during the summer, but with a certain attenuation and a thermal lag, up to three days in some periods, due to the hygrothermal inertia of the building with its massive walls (the wall thickness is about 1 meter). However, during the summer, the temperature reaches a high value (about 29°C) and it is a problem, especially if associated
with a high value of direct solar lighting, because the degradation process of the organic materials can be accelerated. Instead, in winter period, is visible the effect of the heating made by the radiators, which causes daily fluctuations over the references limit, especially when the heating remains off for more consecutive days (weekends and holiday days).

![Figure 3](image-url)  
Figure 3. Daily Indoor air temperature (measured in K0) and outdoor air temperature.

4.2. Model validation
Multiple linear regression was used to model the relationship between the explanatory variables and the daily indoor air temperature which is the response variable by fitting a linear regression.

Several models were tested, taking into account the most representative independent variables to characterize the indoor temperature trend, such as:

- Maximum and minimum daily outdoor air temperature (°C);
- Average daily running outdoor air temperature, also referred to the previous days (°C);
- Outdoor temperature variation between the previous days (°C);
- Average daily solar radiation (W/m²);
- Heating system mode, also referred to the previous days (-);
- Type of season (-);
- Occupancy rate of the indoor space (-);

The outdoor parameters (air temperature and solar radiation) were measured by a microclimate station located on the roof of the Department of Engineering building, not far away from Palazzo Murena; also in this case, the data were recorded every 10 minutes before being processed on a daily basis.

Eventually, it was found that the eight input variables provided the most appropriate solution, with a good value of $R^2_{adj}$; in fact a higher number of inputs would have made the model too complicated and difficult to be used, while with less inputs the errors would have been higher. From the tested models it emerged that the outdoor temperature of the previous three days is more significant than the temperature of the previous day and the temperature of the previous two days; this aspect is demonstrated by the value of $R^2$ which is the biggest of the three ones. In Figure 4 and 5 are shown the relationship between the average daily indoor air temperature with the maximum outdoor one and the average daily indoor temperature with the average outdoor one of the previous three days. Future developments of the study could include the running mean temperature in the model. Even if the season type seems to be a similar information to the heating system mode one, its regression coefficient passed the significance test, so it was considered in the model [17-18]. The input variables for the final model have been chosen for their significance degree from Fisher-Snedecor test and they are shown in Table 1.
Table 1. Input variables.

| Variable | Values range | p-value (F) |
|----------|--------------|-------------|
| $x_1$ = Maximum outdoor air temperature (°C) | 4 - 37 | $2,612 \cdot 10^{-25}$ |
| $x_2$ = Average outdoor air temperature referred to previous three days (°C) | 1.5 - 29 | $3,827 \cdot 10^{-42}$ |
| $x_3$ = Average solar radiation (W/m²) | 8.45 - 399 | $1,586 \cdot 10^{-11}$ |
| $x_4$ = Outdoor temperature variation type between the same day and third previous one | 0: $\Delta T$ with absolute value between 0 and 1°C 5: $\Delta T$ with absolute value between 1 and 3°C 10: $\Delta T$ with absolute value more of 3°C | 2,604 $\cdot 10^{-5}$ |
| $x_5$ = Number of consecutive previous days with heating off | 0 - 188 | $1,552 \cdot 10^{-31}$ |
| $x_6$ = Heating system day type | 0: heating system is off 1: heating time 8 a.m. - 1 p.m. 2: heating time 7 a.m. - 6 p.m. 3: heating time 6 a.m. - 6 p.m. 4: heating time 5 a.m. - 6 p.m. | $3,031 \cdot 10^{-10}$ |
| $x_7$ = Heating system day type of previous two days | 0: heating system is off 1: heating time 8 a.m. - 1 p.m. 2: heating time 7 a.m. - 6 p.m. 3: heating time 6 a.m. - 6 p.m. 4: heating time 5 a.m. - 6 p.m. | $3,257 \cdot 10^{-7}$ |
| $x_8$ = Type of season | 0.5: transition season 0: summer 1: winter | 0.015 |

These eight input variables are the best solution to represent the indoor air temperature and the output MLR model is an equation which accounts all of them. Figure 6 compares the measured daily indoor temperature and regression model for the whole year: the analysis highlights that the indoor temperature trend is mostly influenced by the outdoor one, especially during summer where the variations of...
temperature between more consecutive days are marked in the model. This aspect can be explained because in the real case the indoor microclimate is affected by the hygrometric state of the previous days, as the hall is characterized by poor ventilation. Instead, the model results devoid of this information, which is difficult to replicate. However, the model can be considered acceptable with a mean absolute percentage error less than 1.5%. The model error statistics are:

- $R^2_{adj} = 0.92$
- Standard error = 1.19
- $F$-value = 530

The model may estimate the maximum temperature that can be reached in the hall, but regarding the variations, it should be improved because they are a little bit different from the real ones. In winter period are slightly visible the effects of the heating system during the weeks, but also in this case the outdoor effects are predominant, even if less than summer period.

4.3. Indoor air temperature trend forecast

The application of the regression model to November and December of 2020 is plotted in Figure 7.

There are some data with great differences between the predicted value and the real one. In particular, the model overestimates the daily indoor temperature for the first week of November and in some days of December (during holiday days), in which the outdoor temperature has risen and the heating schedule may has changed significantly. Despite this aspect, for most days the real indoor temperature trend remains inside the confidence range of the model (except for the above mentioned periods); this is a
positive aspect which confirms the reliability of the model. The confidence range depends on the standard error (in this case 1.19), which represents the differences between the observed response and the response predicted by the regression model [14].

4.4. Future developments

Although the model could be improved, it could be a useful tool to analyse, in a standard year, in how many days and what of them the microclimate indoor conditions are out of the range for a good conservation of the library collection. The climate data of a standard year for Perugia is made available by CTI (Comitato Termotecnico Italiano), which collects and processes the data of a great monitoring period [12]. In Figure 8 is shown the first attempt to model the indoor air temperature for a standard year: the outdoor conditions, especially for the air temperature, are a little bit different from the 2020’s ones and this aspect influences the indoor trend, especially in October (significant temperature decrease). However, also in this case, it is clear that the most critical period is summer, in which the indoor temperature reaches high values.

![Figure 8. Regression model for a standard year of Perugia.](image)

5. Conclusions

In this study eight variables were identified as inputs in the regression model and its adjusted coefficient of determination $R^2_{adj}$ is about 0.92, which means that 92% of variation in daily indoor temperature for the hall can be explained by change eight parameters. This model is a good representation of the real hall temperature trend, even if is too much influenced by the outdoor one, especially during the summer, when, in some weeks, the variations of temperature are considerable. For this reason, it could be improved dividing the dataset in two groups based on the season type or on a temperature limit. However, the developed model is a good tool to predict the indoor temperature trend and to evaluate its effects in the library collections, especially in terms of maximum and minimum values which the trend reaches. In historical buildings, without mechanical ventilation system, the indoor microclimate is determined by the topographic site, the orientation of the walls, location of the opening and construction materials [19]. Even if in some case the model not adequately represent the real temperature trend, it shows that the hygrothermal inertia of the building is not sufficient to mitigate temperature fluctuations and it is a critical aspect for the conservation of library collection. In this case, a conventional HVAC system couldn’t be investigated because the hall is in a listed building by the Superintendence for its high historical and artistic value. However, recently, a continuous monitoring system of indoor thermo-hygrometric parameters was installed in more points of the hall, and so it is possible, in the future, to design an MVHR system (Mechanical Ventilation Heat Recovery System) to mitigate the indoor microclimate, without strictly changing it. This ventilation system could be implemented in the regression model to evaluate its effects.
Acknowledgements
The authors wish to thank Monica Fiore (Responsible for “Fondo Antico - Centro dei Servizi Bibliotecari”, University of Perugia) for her willingness to provide useful information and to grant access to the Library during the experimental campaign.

References
[1] Sciurpi F, Carletti C and Pierangioli L 2018 IOP Conf. Ser.: Mater. Sci. Eng. 364 012023 Assessment of thermo-hygrometric indicators for preventive conservation inside museum: In-field monitoring and passive microclimatic control strategies applied to “la Specola” museum of Florence.
[2] Sciurpi F, Carletti C and Pierangioli L 2015 Energy Build Environmentale monitoring and microclimatic control strategies in 2la Specola” museum of Florence
[3] Balocco C, et al 2016 J. Cult. Herit. Indoor microclimatic study of Cultural Heritage protection and preventive conservation in the Palatina Library
[4] Sahin CD, Coşkun T, Arsan ZD and Gökcen Akkurt G 2017 Sustain. Cities Soc. Investigation of indoor microclimate of historic libraries for preventive conservation of manuscripts. Case Study: Tire Necip Paşa Library, İzmir-Turkey
[5] Moretti E, Sciurpi F, Stamponi E and Fiore M 2020 IOP Conf. Ser.: Mater. Sci. Eng. 949 012098 Microclimatic monitoring for book heritage preservation in historic buildings: preliminary investigation on “Sala del Dottorato” in Palazzo Murena, Perugia, Italy
[6] Adcock EP 1998 Ifla Principles for the Care and Handling of Library Material Preservation
[7] ASHRAE. ASHRAE Handbook - HVAC Applications (SI) - Chapter: Museums, Galleries, Archives and Libraries Ashrae (2011)
[8] UNI 10829:1999 Beni di interesse storico-artistico. Condizioni ambientali di conservazione. Misurazione ed analisi (Italian)
[9] UNI-EN 15757:2010 Conservation of Cultural Property - Specifications for temperature and relative humidity to limit climate-induced mechanical damage in organic hygroscopic materials
[10] D.M. 10 maggio 2001 Atto di indirizzo sui criteri tecnico- scientifici e sugli standard di funzionamento e sviluppo dei musei (Art. 150, comma 6, del D.Les. n. 112 del 1998) (Italian)
[11] UNI-EN 15759-2:2018 Conservation of cultural heritage - Indoor climate - Part 2: Ventilation management for the protection of cultural heritage buildings and collections
[12] Comitato Termotecnico Italiano https://try.cti2000.it/
[13] Moretti E, Nassuato L and Bordoni GP 2019 Italian Journal of Engineerign Science Development of regression Models to Predict Energy Consumption in Industrial Sites: The Case Study of a Manufacturing Company in the Central Italy
[14] Tso GKF and Yau KKW 2007 Energy Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks
[15] Mottahedi M, Mohammadpour A, et al. 2015 Procedia Eng Multi-linear Regression Models to Predict the Annual Energy Consumption of an Office Building with Different Shapes
[16] Fang T, and Lahdelma R 2016 Appl Energy Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system
[17] Prieto AJ, Silva A, de Brito J, et al. FJ 2017 J. Cult. Herit. Multiple linear regression and fuzzy logic models applied to the functional service life prediction of cultural heritage
[18] Freire RZ, Oliveira GHC and Mendes N 2008 Energy Build Development of regression equations for predicting energy and hygrothermal performance of buildings
[19] Lucero-Gómez P, Balliana E, Izzo FC and Zendri E 2020 Build Environ A new methodology to characterize indoor variations of temperature and relative humidity in historical museum buildings for conservation purposes