

Article

# Impact of Building Design Parameters on Daylighting Metrics Using an Analysis, Prediction, and Optimization Approach Based on Statistical Learning Technique

Jaewook Lee <sup>1,\*</sup>, Mohamed Boubekri <sup>1</sup> and Feng Liang <sup>2</sup>

<sup>1</sup> Illinois School of Architecture, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA; boubekri@illinois.edu

<sup>2</sup> Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA; liangf@illinois.edu

\* Correspondence: jlee764@illinois.edu; Tel.: +1-217-402-4335

Received: 22 January 2019; Accepted: 5 March 2019; Published: 10 March 2019



**Abstract:** Daylighting metrics are used to predict the daylight availability within a building and assess the performance of a fenestration solution. In this process, building design parameters are inseparable from these metrics; therefore, we need to know which parameters are truly important and how they impact performance. The purpose of this study is to explore the relationship between building design attributes and existing daylighting metrics based on a new methodology we are proposing. This methodology involves statistical learning. It is an emerging methodology that helps us to analyze a large quantity of output data and the impact of a large number of design variables. In particular, we can use these statistical methodologies to analyze which features are important, which ones are not, and the type of relationships they have. Using these techniques, statistical models may be created to predict daylighting metric values for different building types and design solutions. In this article we will outline how this methodology works, and analyze the building design features that have the strongest impact on daylighting performance.

**Keywords:** daylighting; architectural design; building simulation; machine learning; data mining

## 1. Introduction

The benefits of daylight extend beyond building energy savings, and the importance of daylight has attracted the attention of building designers as well as researchers [1]. This leads us to the question of what constitutes good building design or a good daylighting performance. From a daylighting standpoint, this question is not easily answerable, because there is a need to maximize daylight inside a building to reduce electric energy consumption and improve the wellbeing of building occupants; on the other hand, there are issues of visual comfort that are not always compatible with the simple idea of maximizing daylight [2–6]. In order to optimize the design solution, we need to consider specific characteristics of daylight and their relationships to building design [7–9]. Given the large number of design parameters, this can be a daunting and very time-consuming task.

With the development of computer simulation protocols, analyzing how much daylight enters a building has become a sophisticated process. Computer programs such as Radiance and others can analyze how much daylight enters a building with an error rate significantly similar to that of measurements taken by a hand-held light meter [10]. In recent years, performance-driven computer simulation approaches have often been used to estimate the impact of one or a few design variables. Based on this, building design parameters can be categorized, and with the use of a novel statistical

approach called statistical learning techniques (SLT), one can analyze a range of variable and large data sets. The main goal of statistical learning theory is to provide a framework for studying the problem of inference, of gaining knowledge, making predictions, and making decisions or constructing models from a set of data. This is studied in a statistical framework, by making assumptions of a statistical nature about the underlying phenomena, given the nature and the trend seen in the data being generated.

Thus, the integration of the statistical learning techniques and daylight design in a building can be helpful in developing building design solutions. For example, these statistical learning techniques may provide mainly a system for analyzing correlations, identifying associations, and predicting new alternatives based on existing data. They may be used primarily to describe relationships between features based on input and output data, and to create and evaluate predictive models [11,12] as shown in Figure 1.

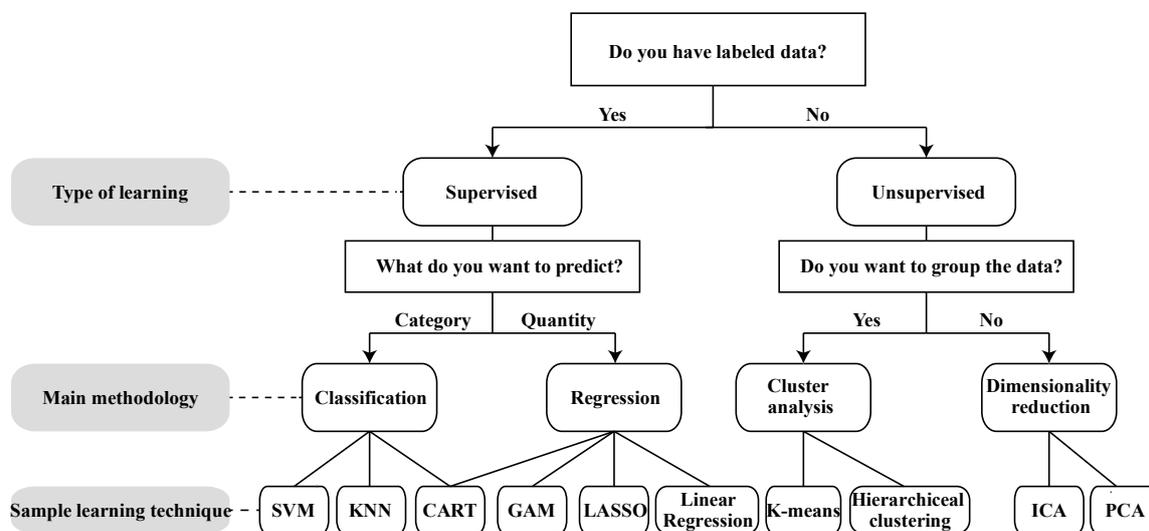


Figure 1. Diagram of statistical learning techniques [11].

There are supervised and unsupervised learning techniques. The goal of supervised learning is to use the input from the training data to map and generalize the new predicted data based on given trends between the input and out variables. The main methodologies used in supervised SLT are classification and regression; Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Classification and Regression Tree (CART), Generalized Additive Model (GAM), Least Absolute Shrinkage and Selection Operator (LASSO), and Linear regression. Conversely, the goal of unsupervised learning is to discover interesting patterns or properties of the data and generate features to feed into a supervised model. Likewise, the main methodologies for this type of learning are cluster analysis and dimensionality reduction; K-means, Hierarchical clustering, Independent Component Analysis (ICA), and Principal Component Analysis (PCA).

These various methodologies can be applied to various research purposes, including for example applications in building environmental performance studies. Multi-objective optimization is based on statistical analysis and it is often used to find optimal alternatives within a given range [13]. Blischke, W. R., et al. [14] described the prediction and optimization through modeling for reliability analysis. Studies were conducted to predict and optimize building energy related fields based on this basic statistical learning methodology. With development, in terms of the application of statistical learning on the built environment, the ant colony algorithm is a probabilistic technique which is inspired by the behavior of real ants, and has been successfully applied for building energy optimization, for example [15]. Furthermore, Support Vector Machine (SVM) which analyze data used for classification and regression analysis, has been applied for rooftop photovoltaic potential [16] or for

electric demand [17], and cluster analysis (K-means) has been applied for determining key variables for building energy consumption [18]. However, the majority of these studies applying these novel statistical paradigms have mostly been used in the field of building energy performance and not in daylighting studies.

### 1.1. Daylighting Metrics

As the issues of improving human comfort inside buildings become increasingly important, various daylighting metrics have been developed to evaluate how a particular room or a building will perform under a given fenestration design. These metrics, described in detail below, show major differences in how they measure and evaluate daylighting performance [19].

(i) Daylight Autonomy (DA) [10]: is a metric defined as the fraction of occupied time of a day, week, month, or a year, that the daylight levels exceed a specified target illuminance, 300 lux. DA gives an intuitive look at how well daylight will penetrate into the space and allows the designer to estimate, with relative ease, the related electric energy savings. A main disadvantage is that, since DA does not have an upper limit of the daylight illuminance, it does not take into consideration issues related to visual comfort that may be caused by excessive sunlight, for instance.

(ii) Spatial Daylight Autonomy (sDA) [20]: sDA is an upgraded version of DA and it describes the percentage of area that is above 300 lux for 50% of the occupied hours. Even though sDA does not incorporate glare or direct sun exposure, it has been verified to reliably predict occupant satisfaction using a single number for the space [21]. sDA values can range from zero to 100% of the floor area. For example, an sDA value of 75% indicates a space in which daylighting is “preferred” by occupants; that is, occupants would be able to work comfortably there without the use of any electric lights and find the daylight levels to be sufficient. An sDA value between 55% and 74% indicates a space in which daylighting is “nominally accepted” by occupants. Architects or lighting designers should therefore aim to achieve sDA values of 75% or higher in regularly occupied spaces, such as an open-plan office or classroom, and at least 55% in less-occupied areas where people need less daylight.

(iii) Continuous Daylight Autonomy (cDA) [22] is an upgraded version of DA, as is sDA, and includes partial contributions from the daylight hours. To explain, cDA awards partial credit in a linear fashion to values below the user defined threshold of 300 lux. Thus, it has high correlation with a control system that dims the electric lighting. For example, if a threshold value is 300 lux and a room has 150 lux at a specific point 100% of operating hours, the cDA for that point would be 0.500.

(iv) Useful Daylight Illuminance (UDI) [23]: UDI is a set of three indicators for every point in the space. These indicators are the percentage of time that a point is below a minimum threshold, between a useful minimum and maximum value. 100 lux is often used as the lower bound of useful illuminance, and 2000 lux as the upper bound.

(v) Annual Sunlight Exposure (aSE) [20]: aSE represents the number of hours per year at a given point where direct sun is incident on the surface, potentially causing discomfort, glare, or increased cooling loads. To explain, aSE is defined as the percentage of square footage that has direct sunlight (at least 1000 lux) for more than 250 h a year. According to Leed v4, it is recommended to design a space with an aSE value of less than 10% [24].

(vi) Daylight Availability (D.A.) [25]: D.A. is defined as ‘the percentage of the occupied hours of the year when a minimum illuminance threshold is met by daylight alone’ which tries to combine DA and UDI. In terms of evaluation, any number between 49–100% represents ‘day-lit’ nodes.

(vii) Daylight Factor (DF) [26]: DF is the ratio between the internal illuminance level at a specific point to exterior horizontal illuminance, under a CIE overcast sky. For instance, if we have 1000 fc exterior, and 20 fc at a given location inside the room, that is a 2% DF. This is the earliest standard of daylight, developed as a legal basis in 19th-century Britain for determining when a new structure would intrude on the daylight of another. A DF greater than 2% is considered adequate and 2–5% is considered well day-lit. As a major flaw, DF shows the same results independent of orientation, time of day, and climate.

(viii) Mean Hourly Illuminance (MHI) [27]: The mean hourly illuminance (MHI) represents the average illuminance level calculated at a given location, or as an average based on several prescribed locations inside a room. It represents the mean value of the total hourly illuminance levels in a room calculated at one or several preselected locations throughout a day, a month, or a year [27]. This is a computer-based calculation whereby the mean is a statistically derived number of illuminance values based on the weather files available for that geographic location, using probabilistic occurrences of all sky conditions. Unlike the DF method, the MHI accounts for all sky conditions, as well as for direct and diffuse illuminance inside the room, but does not have any considerations of visual comfort issues.

(ix) Energy Use Intensity (EUI) is a metric that possibly, though not necessarily, is related to daylighting. It is used to evaluate the energy performance of buildings, as it represents the energy per square meter or foot per year, and it is calculated by dividing the total yearly energy consumption of the building by its total gross floor area.

(x) Uniformity Ratio (UR) [28]: This is a metric related to a qualitative aspect of an electric lighting as a daylighting design scenario. UR represents the ratio between the minimum and the average illuminance levels within a room. UR is a metric solely used, and more related to visual comfort and visual interest in a room rather than any other type of daylighting performance studies.

Researchers have carried out various studies using these metrics. Early research in this area was carried out to describe the characteristics of various metrics and to compare their design outcomes. Reinhart, C.F., et al. [29] compared four such metrics, namely the DF, DA, cDA, and UDI under different locations, shading systems, and shading controls, including façade orientation and space function. They highlighted the shortcomings of the static metrics compared to the dynamic ones. Similarly, Boubekri, M., et al. [30] compared the four daylighting metrics (DF, DA, UDI, and MHI) according to the three different shading designs, explaining the characteristics of each metric. A novelty of Boubekri's research was the use of the projection factor when it comes to sizing shading devices in relation to the window aperture. Their conclusion was that the UDI would lead to different results than the other metrics because of its intrinsic definition, which takes into consideration the elimination of illuminance levels below a certain minimum and above a maximum. Lee, K.S., et al. [31] compared DA and UDI in variation of shading type and design as well. The result illustrated the impact of shading on DA and UDI. In particular, the additional installation of the shading system showed a tendency that the DA value decreases and the UDI value increases, to a certain extent. Additionally, various daylighting metrics were integrated and combined with the building design to derive the best alternative. Mangkuto, R.A., et al. [32] examined optimization of window size and orientation design based on various daylighting metrics. A total of six different metrics (DF, average uniformity, DA, UDI, DGP, and lighting energy consumption) were used for this optimization. For data analysis, sensitive analysis and multi-objective optimization were used. As a reference office module, they used the International Energy Agency (IEA) Task 27 reference office [33], namely, an office module measuring 5.4 m in length, 3.5 m in width, and 2.7 m in height, with 20%, 80%, and 50% reflectance of the floor, ceiling, and walls, respectively. The window was equipped with a double-pane Low-E glazing. Unlike many previous studies that have used what one might refer to as conventional statistical techniques, the uniqueness of our methodology is rather statistical learning methodologies involving database creation, analysis, prediction, and optimization. As a result, three optimal solutions based on daylight and energy performance were presented.

### *1.2. Limitations of Current Daylighting Metrics and Proposed Approach*

The review above of all currently available daylighting metrics points to the following deficiencies or limitations of daylighting metrics. First, although various daylighting metrics have been proposed, however, no one metric is suitable to analyze all design conditions while considering quantitative and qualitative visual performance parameters, such as glare. In addition, existing studies did not consider all the possible various building designs, nor did they analyze in detail which design parameters have the strongest impact on the daylighting performance. Additionally, existing studies have not analyzed

the daylighting metrics based on a large dataset of all possible design variables. Most examples use either one size or are resized based on a small dataset based on typical shading devices, etc. Thus, based on the relative characteristics of the various metrics, clustering or categorization of each metric is not possible. We propose to overcome these limitations of current daylighting metrics through an approach based on the concept of statistical learning techniques (SLT).

## 2. Methods

In this paper, database creation, analysis, prediction, and optimization processes were applied to show the relationship between building design parameters and daylighting metrics. To be specific, the objectives include (i) analyzing the correlation and characteristics of input features (design parameters) and output features (daylighting metrics) statistically; (ii) building the predictive models that can digitize building design parameters to explain the relationship between building design parameters and daylighting metrics; (iii) presenting the design characteristics of optimal daylight design condition among existing buildings; (iv) investigating the feasibility of statistical learning methodology on daylight and building design processes.

The analysis is presented in steps as shown in Figure 2:

(1-1) Select target buildings and rooms and digitize the design parameters as input features.

(1-2) Perform simulation to obtain the daylighting metric values as output features.

(2-1) Analyze and explain the characteristics of the input and output features through correlation analysis.

(2-2) Analyze and explain the relationship between input and output features based on statistical learning models.

(3) Investigate and discuss optimal design characteristics and daylighting metrics through optimization methodology.

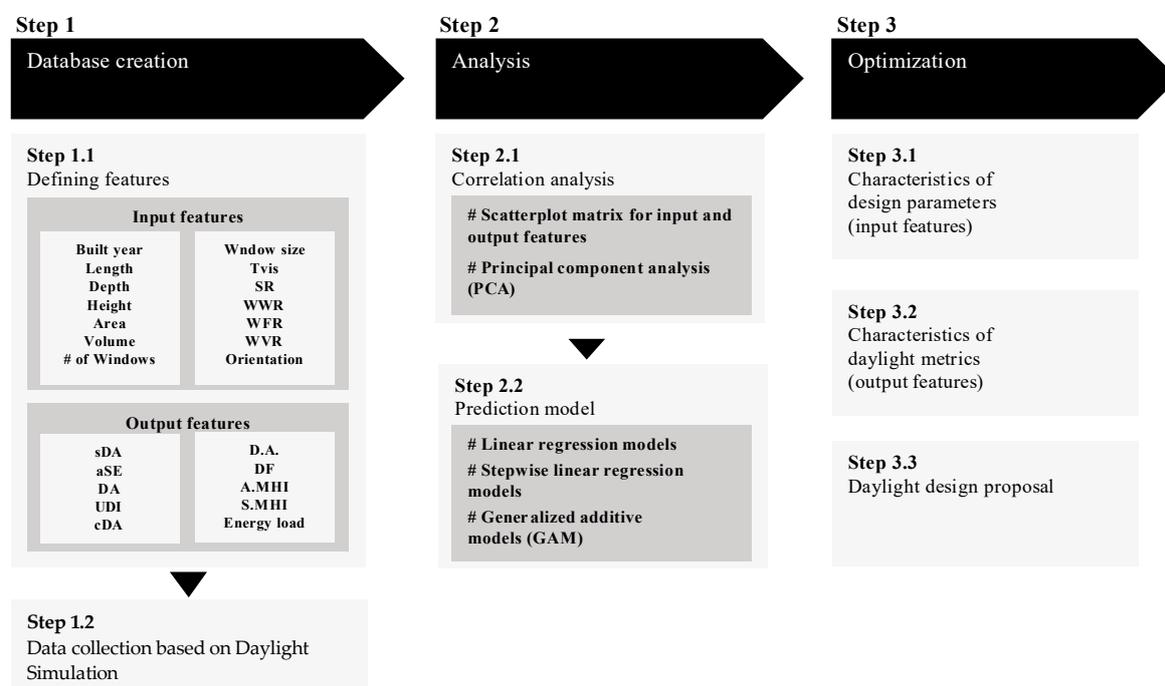


Figure 2. Research diagram.

### 2.1. Database Creation

Several buildings at the University of Illinois were used in this study. The choice of these buildings was solely based on some other objectives in this study which were related to health and wellbeing assessments of office workers. Our choice of buildings was simply wherever those volunteers were

working. These buildings were comprised of 70 buildings scattered across the campus of the University of Illinois at Urbana-Champaign, having different styles, typologies, etc. Based on this, 300 rooms of office spaces were randomly selected for the database.

Thirteen different types of building design parameters were digitized and used as input features, including: Length, Depth, Height, Area, Volume (Vol), Number of windows (NoW), Total size of window (TAoW), Visible transmittance (Tvis), Slenderness ratio (SR), Window Wall Ratio (WWR), Window Floor Ratio (WFR) [34], Window Volume Ratio (WVR), and Orientation. As output features, nine different types of daylighting metrics were used: sDA, aSE, DA, UDI, cDA, D.A., DF, Average of MHI (A.MHI), Standard deviation of MHI (S.MHI), and lighting energy per square meter (Light).

The minimum (Min), maximum (Max), median, and mean values of 300 inputs and output features are shown in Tables 1 and 2. In terms of the application properties in the input features, SR has a range of 0.3 to 3 and an average value of 1.2. When SR is greater than 1, it means that the depth, which does not contain the window, is relatively large compared to the length which contains the window. Thus, on average, the depth is slightly longer than the length, since SR is defined as the ratio of the length to the depth of the space. For WWR and WFR, the ranges are from 0.06 to 0.92 and from 0.03 to 0.69, with mean values of 0.34 and 0.24, respectively. Thus, the relative size of the windows in each room, on average, accounts for 34% of the area containing the windows and 24% of the floor area. The East, West, South, and North appear as 59 (20%), 59 (20%), 107 (35%), and 75 (25%), respectively, of the total 300 features.

**Table 1.** Input features (design parameters).

Type	Internal Cognitions					Window Properties				Design Factors		
	Length (m)	Depth (m)	Height (m)	Area (m <sup>2</sup> )	Volume (m <sup>3</sup> )	NoW	TAoW (m <sup>2</sup> )	Tvis	SR	WWR	WFR	WVR
Max	23.1	11.1	4.2	138	509	4.0	24.7	0.88	3.0	0.92	0.69	0.24
Min	1.7	2.2	2.3	6.9	20.0	1.0	0.5	0.65	0.3	0.06	0.03	0.01
Mean	4.8	4.9	3.2	25.4	80.5	1.6	5.1	0.80	1.2	0.34	0.24	0.08
Median	4.0	4.5	3.3	17.0	54.8	1.0	4.5	0.80	1.1	0.31	0.21	0.07

**Table 2.** Output features (daylighting metrics).

Type	sDA (%)	aSE (%)	DA (%)	UDI (%)	cDA (%)	D.A. (%)	DF (%)	A.MHI (lux)	S.MHI	Light (kWh)
Max	100.0	100.0	94.2	91.8	95.3	84.1	14.5	1744.9	1037.8	37.9
Min	12.1	0.0	13.6	30.8	37.0	−47.8	0.8	88.5	110.1	4.3
Ave	76.4	40.1	68.9	69.5	82.9	12.7	4.9	608.8	579.0	10.6
Median	84.8	42.3	73.0	70.8	86.8	10.6	4.4	541.3	615.1	7.8

These output features were obtained through daylight simulation. For daylighting metric simulation, Grasshopper and Design Iterate Validate Adapt (DIVA) were used [35]. DIVA is one of Grasshopper's add-ons based on Radiance and Daysim simulation codes, which are effective in simulating daylighting performance [10]. This study assumes the default materials with typical surface reflectance, namely 20% for the floor, 70% for the ceiling, and 35% for the walls. Single-pane clear at 88%, double-pane clear at 80%, and double-pane Low-E at 65% transmittance were selected for the windows. In the case of grid setting, a 20-points grid system was used for each model. An example of the basic model is shown in Figure 3. The average value of each analysis surface was used as the result of each output features. For occupancy schedules, 8:00 A.M.–6:00 P.M. daily was set as occupied hours in a standard year. The ambient parameters for daylighting simulation, were kept at the default levels as shown in Table 3.

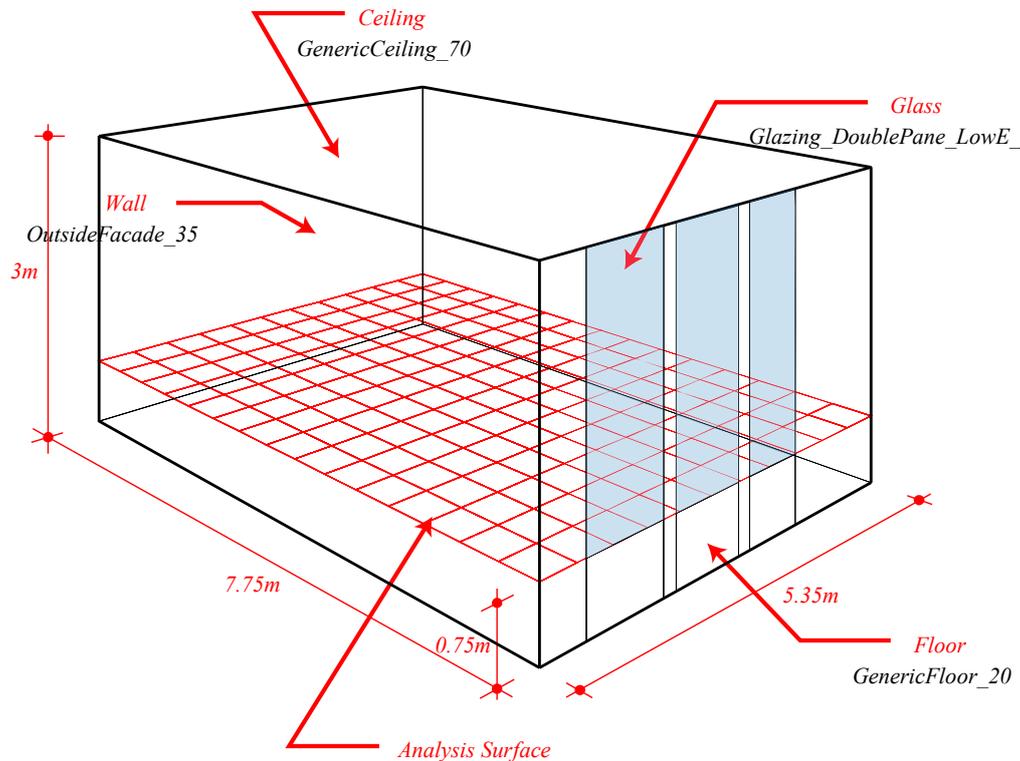


Figure 3. An example of a daylight simulation model.

Table 3. Radiance ambient parameters.

Parameters	Description	Setting
ab	Ambient bounces	2
aa	Ambient accuracy	0.15
ar	Ambient resolution	256
ad	Ambient divisions	512
as	Ambient super-samples	128

The weather data of the city of Champaign-Urbana, United States (40'1" N, 88'2" W) was supported by the Energyplus default climate file [36]. The settings for each daylighting metric were selected based on the default values used for each metric. As such, we used 300 lux as target illuminance in our DA assessment, from January 1 to December 31, from 8:00 A.M. to 5:00 P.M. For the UDI computation, we used the illuminance range between 100 and 2000 lux. A modified methodology was used for sDA, since shading design was not applied in calculating sDA values by IES LM-83-12 [20].

## 2.2. Analysis, Prediction, and Optimization

### 2.2.1. Analysis

Two statistical techniques were used: correlation analysis and Principal Component Analysis (PCA). Correlation analysis is a statistical method of analyzing the linear relationship between two features in probability theory and statistics [37]. A scatterplot, a type of data display that shows the relationship between two numerical features, can be used to explain correlation. When the y variable in the y-axis tends to increase as the x variable increases in the x-axis, we say there is a positive correlation between the features. Thus, in the correlation analysis, the number of Pearson correlations  $\rho$  is used as a unit for expressing the degree of correlation.

PCA is a technique that summarizes and quantifies the various features into a new feature, called ‘Principal Component’, which is a linear combination of several highly correlated features [38]. Thus, PCA helps identify relationship patterns between the independent and dependent variables and expresses the data in such a way to highlight their similarities and differences. The first component is chosen to best explain the overall variability. The second principal component is not correlated with the first principal component, so a linear combination of variables is created to account for the remaining variability that the first principal component cannot account for without loss of information. In this study, data analysis, prediction, and optimization were performed with the statistical software R [39].

### 2.2.2. Prediction

Prediction relies mainly on statistical learning models. Therefore, a model or algorithm can be developed based on the relationship between the input and output data. In this study, simple linear regression, stepwise linear regression, and Generalized Additive Models (GAM) were used.

Simple linear regression is a regression technique that models the linear correlation of the output features,  $y$  with one or more input features,  $x$  [40]. Linear regression models the regression equation using a linear prediction function, and unknown parameters are estimated from the data. This regression equation is called a linear model. Linear regression has been extensively studied and widely used in building science [41,42], mainly because it is easier to build models with linear relationships between unknown parameters than with nonlinear relationships. With linear models, the algebraic equation would take the form

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad (1)$$

where  $y_i$  is the output features of input features,  $x_i$ ,  $\beta$  represents a parameter vector, and  $\beta_0$  is the intercept term.

Stepwise regression is a method of fitting regression models in which the choice of input features is carried out by an semi-automatic procedure and it is a combination of the forward and backward selection techniques [43]. In each automatic step, an input feature is considered for addition to or subtraction from the set of output features based on the specified tolerance level of the t-statistics of the estimated coefficients. In this process, if a nonsignificant variable is found, it is removed from the model. Thus, the stepwise approach is relatively faster, less prone to overfitting the data, and we often learn something by watching the order in which variables are removed or added.

GAM is expressed by combining an additive model based on generalized linear model [12,44]. The basic formula for GAM is as follows:

$$y_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \dots + f_p(x_{ip}) + \varepsilon_i, \quad (2)$$

where  $y_i$  is the output features of input features  $x_i$ ,  $\beta_0$  is the intercept,  $\varepsilon_i$  is the residual, and  $f_p$  is a smooth non-linear function. Therefore, GAM can use the nonlinear function to model the missing part of the linear regression. Also, the accuracy can be significantly increased when the input and output features tend to have non-linear relationships.

In the process of evaluation of statistical models, the root mean square error (RMSE) is used. RMSE is a measure commonly used when dealing with the difference between the estimated value or the predicted value of the model and the observed value [45]. It is suitable for expressing precision and measuring the accuracy of statistical learning models.

### 2.2.3. Optimization

Optimization was performed to evaluate which daylight design may be considered as a better solution based on daylighting computer simulation analyses. The optimization was based on selecting the building design solution that more closely meets the daylighting metric criteria as closely as

possible. This selection was based on the results of 300 simulation outputs. The design parameters of a room that simultaneously satisfy the criteria of each metric is explained below:

- sDA:  $\geq 55\%$  [46]
- aSE:  $\leq 10\%$  [46]
- DA:  $\geq 50\%$  [46]
- UDI:  $\geq 50\%$  [32]
- DF:  $\geq 2\%$  [47]

It is useful to note that some of the metrics do not have their own criteria for measuring optimization of daylighting and building design. In these cases, we set up our own criteria to select the top best 30% of the database results. For example, cDA, D.A. A.MHI, S.MHI, and light energy considered the top 30 percent of each 300 samples for the optimization process.

### 3. Results and Discussion

#### 3.1. Analysis Stage

In this section, internal relationship analyses of the input and output features were carried out, with four window orientations, namely East, West, North, and South, being used as the main breakdown point. Through this process, the scatterplot matrix graphs based on 300 data sets were used in the correlation analyses. Furthermore, PCA analysis was applied to provide a compact description of data variability.

##### 3.1.1. Correlation Analysis of Input Features

The fourteen input features are mainly comprised of design features and the scatterplot matrix of input features is shown in Figure 4. In the scatterplot, the upper right corner shows the correlation value between each feature, and the axis in the middle shows the histogram of each feature based on the four orientations. The lower left part shows the pair plot graph. Thus, the most visible is the correlation between each input features. The independent variables indicating correlation values of 0.7 or more are as follows; Length and Size, Length and Volume, Length and SR, Depth and Size, Depth and Volume, Size and Volume, WWR and WFR, WWR and WVR, and WFR and WVR. These nine relationships can be expected to correlate with each other, indicating a statistical relationship between two features. Since the height of the interior of the building is mostly within a certain range, the dimensions of the building and its volume (Length and Size, Length and Volume, Depth and Size, Depth and Volume, Size and Volume) or WWR, WFR, and WVR are closely related. In addition, there is no difference in the input features due to orientation.

The scatterplots allow us to analyze the relationship between each input and output features based on the four design factors; SR, WWR, WFR, and WVR. The biggest difference between SR and the other variables is that SR only contains information on the overall proportional dimensions of the building, and the variables contain information about the window size in relation to other properties of the room, such as the walls or the floor areas. So far, in the case of WWR, it is currently the most frequently used variable in daylighting or energy analyses [48]. However, the correlation is relatively low compared to WFR and WVR (correlation: 0.708 and 0.773, respectively), because WWR only accounts for the percentage of the window area to the wall within which it is contained only. It may be more effective to compare the area of the window to the floor area. In our case, and using this criterion, the correlation coefficient was much higher (correlation: 0.937).

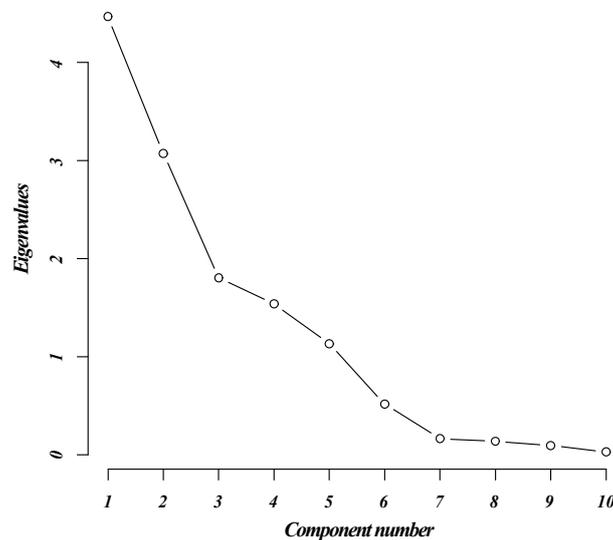


Figure 4. Scatterplot of input features.

### 3.1.2. PCA of Input Features

Through PCA, we can find design features that are strongly correlated with the various input features and can therefore explain most of the overall change and variability of the daylighting performance metrics. A PCA plot of cumulative portion of variance based on 13 input features is shown in Figure 5. PCA analysis shows that 3 of the 13 design variables can explain 72% of the variance of the model and 7 variables out of the 13 can explain 97.7% of the variance in the data. Therefore, the number of input data can be significantly reduced to explain which are the input variables that have the strongest impact on performance. To examine the relationship of the principal components, and after rotation, according to the principal components, a coefficient of eigenvector appears (Table 4). For each of the nine principal component items displayed on the table, the important parameters are written in bold colors. Where more than one parameter per item is marked, a high correlation exists. The table shows that the first critical factors are the area and the volume, which are the most critical. According to the data, we can substitute either the area or the volume because the volume is expressed as the product of the area and the height of the interior of the building. Therefore, since the height of a room exists within a limited range, the correlation between the area and the volume is quite high.

Furthermore, the second most important factors are WWR and WVR, the third factor is year and Tvis, the fourth factor is the depth and SR, the fifth factor is the height, the sixth factor is NoW, and the seventh factor is TAoW, respectively. Therefore, both the correlation analysis and PCA method can identify parameters that have a high correlation among the input features. In addition, NoW or TAoW should be used as a separate input feature because there is no correlation with other input features.



**Figure 5.** Plot of cumulative portion (Eigenvalue vs. the number of components in the principal component analysis).

**Table 4.** Coefficient of eigenvector after rotation according to the principal components.

Types	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
BuiltYear	0.013	−0.150	−0.674 *	0.153	0.062	0.068	−0.223	0.634 *	−0.196
Length	−0.440	0.058	−0.094	−0.195	0.077	−0.144	−0.028	0.120	0.428
Depth	−0.320	−0.131	0.160	0.505 *	0.070	0.076	−0.111	−0.172	−0.544 *
Height	−0.066	0.088	0.122	0.063	−0.895 *	−0.123	0.012	0.192	−0.139
Area	−0.461 *	−0.015	−0.009	0.082	0.100	−0.146	−0.258	−0.102	0.064
Volume	−0.459 *	−0.015	0.002	0.073	−0.091	−0.093	−0.376	−0.106	0.128
NoW	−0.284	0.226	−0.080	−0.153	−0.110	0.900 *	0.100	−0.038	0.013
TAoW	−0.327	0.365	−0.003	0.082	0.054	−0.234	0.549 *	0.311	0.057
Tvis	−0.013	0.084	0.646 *	−0.255	0.230	0.076	−0.290	0.577 *	−0.181
SR	0.170	−0.139	0.232	0.648 *	−0.037	0.219	−0.061	0.205	0.611
WWR	0.023	0.471 *	0.013	0.388	0.228	−0.034	0.236	0.015	−0.170
WFR	0.166	0.511 *	−0.061	0.030	−0.178	−0.058	−0.350	−0.023	0.037
WVR	0.177	0.505 *	−0.124	0.031	0.113	−0.025	−0.389	−0.149	0.053

\* Highly correlated coefficients in each principal component.

### 3.1.3. Correlation Analysis between DF and A.MHI

Since its introduction in 1895, DF has long been used to determine daylighting performance [49]. However, due to the problem of climate and orientation, it is currently not used frequently [50], and instead the annual hourly mean illuminance (A.MHI) metric is more used nowadays. A.MHI is a dynamic metric, simply expressed as the average value of the total illuminance entering the interior of the building [27,30]. Nevertheless, A.MHI, as shown in Figure 6, contains most of the characteristics of DF, but, unlike DF values, the A.MHI indicates the importance and impact of orientations as well [30]. In other words, the overall correlation is 0.989, and the correlation according to each orientation has a value of 1. In this regard, assuming that a DF between 2% and 5% is an appropriate condition, a corresponding range of A.MHI may be between 200 and 500 lux.

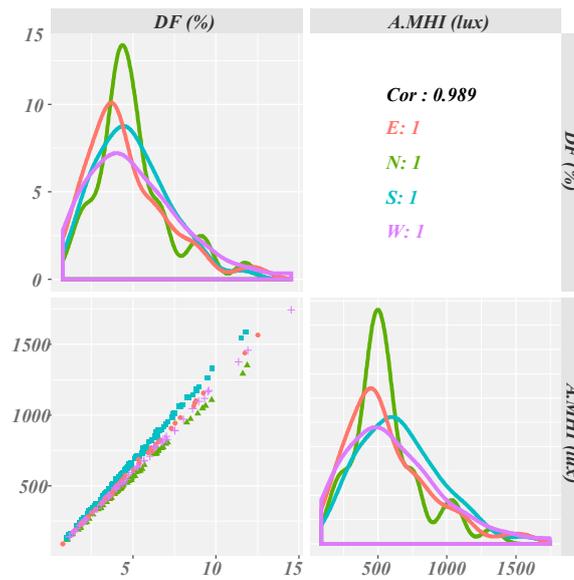


Figure 6. Scatterplot between the Daylight Factor (DF) and the Annual Mean Hourly Illuminance (A.MHI).

3.1.4. Correlation Analysis of DA Related Metrics

DA has been used as an alternative to DF since its introduction in 2002 [51]. In addition, cDA (2006) [29] has been introduced while changing the algorithm slightly, and sDA (2012) [20] has been developed based on changing calculation method and considering shading conditions. Of course, with the introduction of evolved metrics, detailed values have been further refined, but these relationships are still tightly linked. There are some differences according to Figure 7; DA and cDA have convex relationships, and cDA and sDA have concave relationships. Correlation analysis shows that the correlation between them is 0.95 to 0.98, and the relationship between them is quite similar. Compared to other daylighting metrics, the correlation of these three metrics is fairly high. Therefore, in our subsequent analyses, the most recent sDA value among these three metrics is selected as representative value.

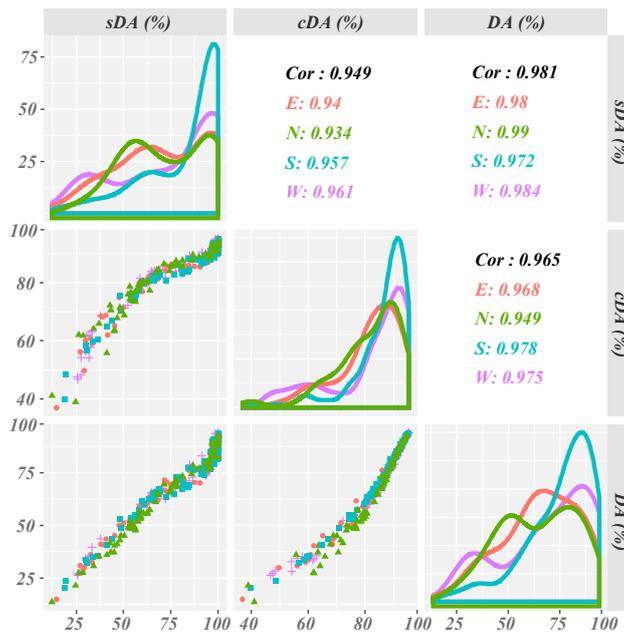


Figure 7. Scatterplot in terms of the Spatial Daylight Autonomy (sDA), the Continuous Daylight Autonomy (cDA), and the Daylight Autonomy (DA).

### 3.1.5. Correlation Analysis of Output Features

Figure 8 is a graph of the correlation of one input feature and six output features. First, in terms of the relationship between WFR and output features, A.MHI is most closely related to WFR based on correlation of 0.866. A.MHI is more closely linked to the characteristics of WFR because it calculates the average value of total daylight entering the room. Next, sDA is more correlated with WFR based on a correlation of 0.633. Since the basic concept of the sDA is based on determining the percentage of time a given illuminance threshold is met, this metric is highly correlated with the WFR variable.

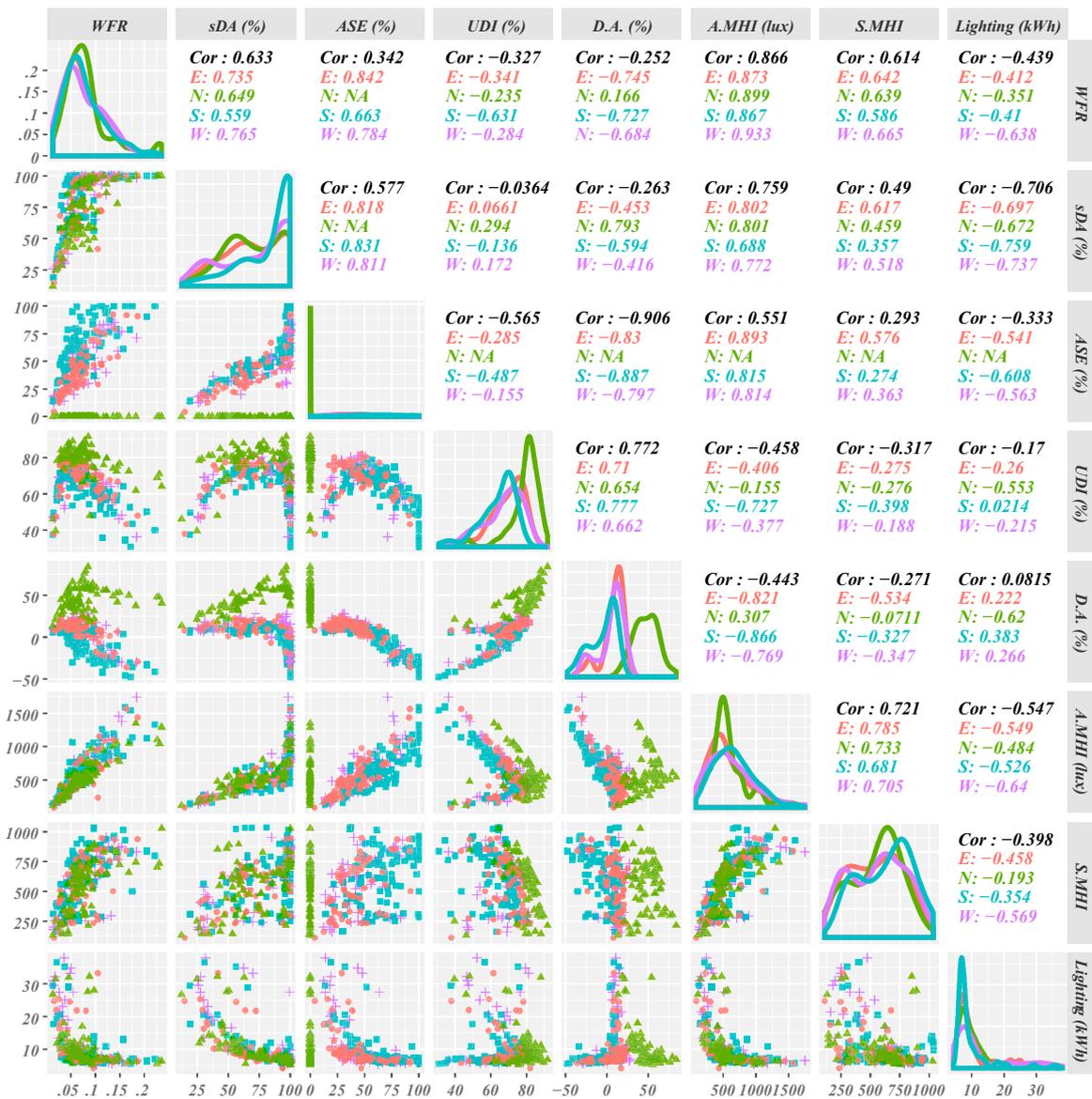


Figure 8. Scatterplot of input and output features.

Secondly, we further analyzed how the results of the daylighting metrics change of the window as the orientation changes. From the North orientation, all aSEs have a value of zero. This means that in the North, light above 1000 lux is hardly coming into the room. The distribution of Useful Daylight Illuminance (UDI) and D.A. is also relatively large. Therefore, the probability of occurrence of glare and light discomfort in the North is considerably low, but the overall level of light (illuminance) is relatively small. In the South, the distribution of sDA is high, and the distribution is relatively similar in the East and West. The main commonalities between UDI and WVR, sDA, and A.MHI is that in the

case of South, East, and West orientations, their values increase. However, the value of UDI increases up to a certain point and then decreases sharply after a certain point. This is because the UDI only takes into account light levels within a specific range, from 100 to 2000 lux. Therefore, the UDI values decrease in cases where the total window size is relatively large, or the light level is high.

Thirdly, Spatial Daylight Autonomy (sDA) shows a strong correlation with A.MHI (correlation value: 0.759), and with aSE (correlation value: 0.820), excluding the case of the north orientation. The UDI is closely related to D.A. (correlation value: 0.772), whereas the correlation with the A.MHI is only moderate and negative (correlation value:  $-0.458$ ), S.MHI also shows a strong association with A.MHI (correlation value: 0.721), but no significant association with the metrics. Our analyses indicate that the daylighting metrics that are more reliable at predicting performance include sDA, A.MHI, and aSE. Conversely, UDI, D.A., and S.MHI may have difficulty in predicting because they indicated weaker correlations with the other design variable and are more impacted by window orientations.

### 3.2. Prediction Stage

In the daylighting prediction stage, sDA, UDI, A.MHI, S.MHI, and lighting were selected as the output features, and RMSE was used as the main evaluation item. The lower the RMSE, the higher the fitness for the model. The basic model was created using simple linear regression and the other two models used logarithmic values exceeding a threshold of 0.75 for skewness for some input and output data [52]. Thus, a total of 12 features (Length, Depth, Size, Volume, NoW, TAoW, SR, WWR, WFR, WVR, A.MHI, and Lighting) were taken as logarithmic values for two advanced model predictions.

#### 3.2.1. Simple Linear Regression and Stepwise Linear Regression

Stepwise linear regression was used to generate the model, which was compared with the RMSE values in the simple linear regression, as shown in Table 5.

**Table 5.** Root Mean Square Error (RMSE) of regression models.

Type	sDA	UDI	A.MHI	S.MHI	Lighting
Selected features of stepwise linear regression	BuiltYear *		Length *		BuiltYear
	Length *	BuiltYear ***	Depth *	BuiltYear ***	Depth
	Depth **	Length *	Size	Length ***	Height *
	Height ***	Size ***	Volume *	Size ***	Size ***
	Size	Volume ***	TAoW ***	Volume ***	Volume **
	NoW **	TAoW ***	Tvis ***	NoW ***	TAoW ***
	TAoW ***	SR *	WWR *	TAoW ***	SR *
	Tvis **	WFR ***	WVR		WFR ***
WFR ***					
RMSE of stepwise linear regression	0.147	0.100	0.017	0.119	0.101
RMSE of linear regression	0.160	0.101	0.119	0.125	0.412

\*  $p \leq 0.05$ , \*\*  $p \leq 0.01$ , \*\*\*  $p \leq 0.001$ .

Compared to the base case model, the stepwise linear regression reduces the number of input data from 14 to 6 to 9, and all the RMSE values show improved results. In particular, this model shows that the prediction of the lighting metric is greatly increased. In terms of feature selection, the common feature in the stepwise linear regression model is TAoW. Therefore, the total window size is considered to be one of the most essential items for predicting the daylighting metric. In addition, WWR, WFR, and WVR, which represent the ratio of window size to a particular space, were used as important input parameters for most daylighting metrics. However, SR is not very suitable to stepwise linear regression models. The ratio of the length of the building to the depth does not seem to make a significant contribution to forecasting the daylighting metric. Therefore, the basic dimensionality

design parameters (length, depth, size, or volume) seem to replace SR's role. A.MHI is the most predictable metric in terms of RMSE since it is the most intuitive metric, as previously explained in the scatterplot matrix in Figure 8.

### 3.2.2. Generalized Additive Models (GAM)

GAM was used to generate a third model based on six degrees of freedom with smoothing splines. The GAM model was compared with the RMSE and Anova (analysis of variance) for parametric and nonparametric effects, as shown in Table 6. Compared to the other two models, the RMSE value was improved on all five outputs. Specifically, we conclude that TAoW, BuiltYear, and height are major factors in common. While other dimension-related components may be exchanged for alternative features, the three above are highly influential on the model and difficult to replace with other variable elements. In addition, there is very clear evidence that a non-linear term is required based on Anova for nonparametric effects.

**Table 6.** RMSE and Anova of generalized additive models (GAM).

Type	sDA	UDI	A.MHI	S.MHI	Lighting
RMSE	0.104	0.068	0.013	0.098	0.074
Anova for Parametric Effects	BuiltYear ***				
	Depth ***	Height ***	Length ***	Height ***	Depth ***
	Height ***	Size ***	Depth ***	Volume ***	Height ***
	Size ***	NoW **	Height ***	TAoW ***	Volume ***
	Volume ***	TAoW ***	Size **	TAoW ***	TAoW ***
	NoW ***	Tvis ***	NoW ***	Tvis ***	SR *
	TAoW ***	SR ***	TAoW ***	WVR ***	Orien **
	WFR **	WVR ***	Tvis ***	Orien **	
	WVR *	Orien ***	Orien ***		
	Orien ***				
Anova for Nonparametric Effects		Depth ***	BuiltYear ***	BuiltYear ***	
		Height **	Height **	Length *	
	Size *	TAoW **	TAoW ***	Depth **	BuiltYear **
	WVR ***	Tvis **	Tvis ***	Height ***	Depth *
		WVR *	SR *	Volume *	WVR *
		WVR **	WVR *	TAoW ***	
				Tvis ***	

\*  $p \leq 0.05$ , \*\*  $p \leq 0.01$ , \*\*\*  $p \leq 0.001$ .

Compared to the three models, the A.MHI is generally the most predictable, while the most unpredictable output is represented by the sDA. Although it is easy to know the approximate distribution of sDA (RMSE of GAM: 0.104), it is relatively difficult to predict the exact value compared to other metrics. This is because the definition of sDA is the rate at which more than 300 lux of light enters the interior of a building by more than 50%, which makes predicting its exact value difficult. In terms of the feature selection, not all input parameters need to be used for the prediction and only a few of the input features are described as necessary. It is possible, in particular, to describe the model through only a few of the dimension and the applied variables.

### 3.3. Optimization Stage

The optimization of daylighting metrics was performed based on the desired criteria. The number of cases used to determine whether the results are satisfactory or not represented 3% of the total number of simulation cases, in this case, 300 cases. More specifically, according to the orientation, the number of rooms satisfying the criteria is one case each in the South and East, two cases in the West, and six cases in the North. The input and output features of these 10 cases are shown in Tables 7 and 8.

Table 7. Input features of 10 best daylighting metrics performance.

id.	BuiltYear	Length	Depth	Height	Size	Volume	NoW	TAoW	Tvis	SR	WWR	WFR	WVR	Orien
1	1987	2.13	4.88	3.33	10.41	34.62	1	2.67	0.80	2.29	0.38	0.26	0.077	E
2	2001	3.79	3.28	3.80	12.45	47.34	1	4.46	0.65	0.86	0.31	0.36	0.094	N
3	2006	3.06	2.67	3.66	8.16	29.86	1	2.31	0.65	0.87	0.21	0.28	0.077	N
4	1986	3.98	3.70	3.68	14.72	54.22	1	4.02	0.80	0.93	0.27	0.27	0.074	N
5	1992	4.41	3.40	3.56	14.97	53.24	2	4.53	0.80	0.77	0.29	0.30	0.085	N
6	2006	3.84	2.18	3.66	8.38	30.66	1	2.31	0.65	0.57	0.16	0.28	0.075	N
7	1999	7.21	4.81	3.38	34.64	117.03	2	9.16	0.70	0.67	0.38	0.26	0.078	N
8	1904	4.72	2.92	2.46	13.80	33.90	1	2.01	0.88	0.62	0.17	0.15	0.059	S
9	1910	2.61	3.94	3.61	10.30	37.14	1	3.17	0.88	1.51	0.34	0.31	0.085	W
10	1904	4.53	3.11	2.46	14.11	34.65	1	1.96	0.88	0.69	0.18	0.14	0.057	W

Table 8. Output features of 10 best daylighting metrics performance.

id.	sDA	aSE	DA	UDI	D.A.	DF	A.MHI	S.MHI	Lighting
1	75.00	32.50	69.30	77.65	20.70	4.10	493.60	391.14	6.72
2	97.62	0.00	86.24	82.45	64.33	6.25	725.90	518.31	5.85
3	93.33	0.00	79.77	89.67	76.50	4.43	507.20	296.51	6.07
4	94.64	0.00	79.87	87.37	70.75	4.83	541.07	352.11	6.56
5	96.83	0.00	84.11	91.78	84.11	4.91	541.33	213.15	6.88
6	81.25	0.00	73.00	86.38	69.91	4.26	487.63	316.47	6.91
7	96.03	0.00	80.51	84.14	65.30	5.35	604.95	527.61	7.14
8	81.48	44.44	69.52	71.72	15.91	3.74	506.59	713.86	6.42
9	97.50	30.00	83.15	76.13	27.25	5.64	675.40	733.97	6.50
10	72.22	31.48	62.52	76.44	28.37	3.61	433.56	626.67	7.03

### 3.3.1. Characteristics of Design Parameters

Six cases of the best 10 designs with the best daylight conditions were selected, as shown in Figure 9. The color expressed in the grid is shown based on the average value of DA. In the color segment, the closer to dark red, the higher the daylight level, and the closer to blue, the lower the daylight level. In particular, the UR metrics range between 0.79 and 0.82, indicating that the daylighting designs were well developed not only in the quantitative, but also in the qualitative aspects.

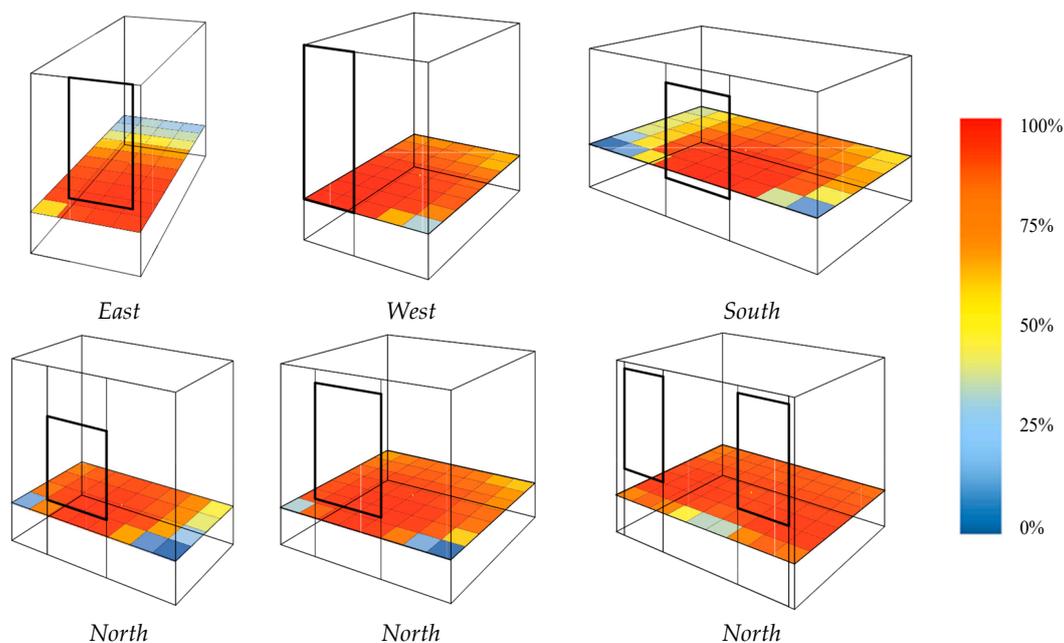


Figure 9. 3-D drawing of six optimized models based on daylighting metrics with DA grids distribution.

In the case of orientation, a relatively large number of optimal cases are found in the case of the North orientation. Given this orientation, the daylight level is generally lower compared to the other orientations. So, consequently, it shows particular strength when the aSE metric is used. Conversely, in the case of South, West, and East orientations, there are few optimal cases. Thus, a great deal of attention should be paid to the aSE metric in South, West, and East, which cannot be easily solved by manipulating the building shape alone. It is necessary to examine other design factors, such as shading design, interior finishes of the rooms, as well as the size and location of windows.

The basic important characteristics of the design parameters of the building are length, depth, and size of window. Among them, the design factors SR, WWR, WFR, and WVR are constructed by multiplying or dividing some of the above internal cognitions. These design factors are more effective in explaining daylight design characteristics, since these values can explain the output features relatively well. Essentially, the smaller the SR (the longer the length is), the more favorable the availability of daylight. To explain, the shorter the depth than the length is, the more useful the light comes in considering the altitude of the sun, since daylight does not have to go deep into the room. In the North direction especially, the optimal design range of SR is relatively low, from 0.57 to 0.93 compared to the average value (1.2). In the case of WFR, the optimal design range is 0.26 to 0.36, which is slightly higher than the average value of 0.24, indicating that having a window area of 26% to 36% of the floor area is effective for daylighting optimization.

### 3.3.2. Characteristics of Daylighting Metrics

Among the daylighting metrics, the most difficult metrics to satisfy all the optimal values were aSE and S.MHI. In the case of aSE, the problem can be solved by a proper shading system, internal or external shading design, since the large value of aSE indicates high probability of an excessive light and potential glare. However, when analyzing the output features in North facing, we do not need to install any shading design and we can optimize daylight distribution by building shape design only. For example, in this study, six out of about 75 samples in North satisfy all the output criteria.

In the case of S.MHI, it would be easier to find and apply the optimal value if we apply a limited range of actual floor available inside the building. There is a need for further research and development of daylighting metrics for areas that can account for daylight uniformity. Conversely, there are a relatively large number of cases that meet the minimum threshold setting criteria for sDA, DA, DF, and A.MHI. Thus, in terms of current building design, there is a fairly high level of daylight coming in at a certain level of light, especially in the South, West, and East.

In other words, a daylighting metric can be roughly divided into three categories. At first, sDA, DA, CDA, DF, and A.MHI can be grouped into one category, so-called the sDA-related metrics. The higher the value of these metrics, the higher the probability of having a higher lighting level. The second most indicative are UDI and D.A., which show relatively large differences compared to sDA-related metrics. Where the daylight level is too high, the values of UDI and D.A. tend to decrease, and they are relatively unpredictable with respect to these metrics. Finally, lighting energy shows a tendency to be inversely proportional to sDA-related metrics.

## 4. Conclusions

The relationship between building design parameters and daylighting metrics was compared through database creation. In this process, a statistical learning method was used, which was followed by analysis, prediction, and optimization. The main innovations of the study are: (i) the use of the statistical learning paradigm, applied to daylighting problems; and (ii) to create a rank list of the most influencing parameters on the daylighting in a space, which would be of great help to building practitioners and scientists. As a limitation, more samples would be needed in the future to support these conclusions for more robust analysis. Key findings are summarized by the following categories:

### *Data establishment*

- In the conventional computer simulation, a 3-D model is used to analyze daylight. Because of the computer simulation, this conventional method requires modelling techniques, which can have difficulty analyzing how each design parameter affects daylight availability. Therefore, this cubic model was deconstructed and classified based on design parameters and used as an input feature. Also, nine well-known daylighting metrics were selected as output features. For the target building, 300 rooms were randomly selected based on a total of 70 university buildings.

### *Correlation analysis*

- The input parameters showing relatively high correlations are (1) WWR, WFR, and WVR, (2) size and volume. Since the height of the room does not change significantly within a certain range, the above three results are obtained. In addition, it is more effective to use WFR or WVR than to use WWR, which is the most commonly used parameter to describe the percentage of windows occupied in a building or room.
- According to PCA, our analyses indicated that only 7 room attributes (Length, Depth, Volume, TAoW, Tvis, SR, and WFR) out of 13 input attributes can predict the internal change to about 97.7%. Therefore, the number of input features can be relatively reduced, which can greatly enhance the interpretability of the models. The overall tendency of the results of PCA is similar to that of the correlation analysis.
- As an alternative to the DF metric, A.MHI may be used, since it contains most of the characteristics of DF.
- DA related metrics, DA, sDA, and cDA are highly correlated, and there is no major problem in using the most recently proposed sDA as a representative value.

### *Prediction and statistical learning model*

- In the statistical prediction process, the simple linear regression model was analyzed as a basic model.
- Log transformation was performed on some skewed data, and a stepwise linear regression model was created and compared with the basic model. Based on RMSE, the prediction accuracy has increased in all output features. Also, based on feature selection, the input features were reduced and the interpretability was increased.
- GAM model shows that the model's predictive power is significantly better than other models based on RMSE. The variable selection also shows a similar tendency to the stepwise linear regression model. In particular, some input and output features have non-linear relationships, making the GAM model highly suited to current data.

### *Optimization and daylight design guide*

- Our study has shown that using one single metric may not be a panacea to predicting performance in all scenarios. and that using one single metric is unlikely to result in a better daylight environment.
- The daylighting metric used as the output feature can be classified into three categories; (1) sDA, DA, CDA, DF, and A.MHI, (2) UDI and D.A, and (3) lighting according to the correlation between the measurement purpose and the result value.
- Based on the correlation results, we have proposed which combination of metrics best represents the daylighting condition. For example, it is necessary to use one of the sDA-related metrics and one of the UDI-related metrics. Also, a supplemental use of aSE (especially in the north) is necessary because it can appear to vary greatly depending on orientation.
- Among the design parameters, the WFR or WVR value has a large influence on the output features. However, it is difficult to explain the optimal daylight design by WFR alone. SR is an additional

design parameter that can be supplemented. Thus, according to the SR, the relative window size of building floor area, that is, the WFR value, must be reconsidered. For daylight optimization, as SR becomes smaller, WFR should be relatively smaller. On the contrary, when SR becomes larger, a relatively large WFR is advantageous.

- The South, West, and East require proper shading design. The sDA and A.MHI are relatively large and enough light enters. However, even though the aSE should be small, in most cases the aSE value is large, indicating the high probability of excessive light and glare. This problem can be solved by installing shading inside or outside of a window.
- The absence of shading design in the North may be advantageous in most cases.

**Author Contributions:** The work presented in this article is the result of a collaboration of all authors. J.L. contributed to writing the manuscript and edited the document. M.B. and F.L. critically reviewed the article. All authors have read and approved the final manuscript.

**Funding:** This research was funded by the Research Board of the University of Illinois at Urbana-Champaign.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Boubekri, M.; Cheung, I.N.; Reid, K.J.; Wang, C.-H.; Zee, P.C. Impact of windows and daylight exposure on overall health and sleep quality of office workers: A case-control pilot study. *J. Clin. Sleep Med. Jcsm Off. Publ. Am. Acad. Sleep Med.* **2014**, *10*, 603. [[CrossRef](#)] [[PubMed](#)]
2. Boubekri, M. *Daylighting, Architecture and Health*; Routledge: London, UK, 2008.
3. Sahin, L.; Wood, B.M.; Plitnick, B.; Figueiro, M.G. Daytime light exposure: Effects on biomarkers, measures of alertness, and performance. *Behav. Brain Res.* **2014**, *274*, 176–185. [[CrossRef](#)] [[PubMed](#)]
4. Boyce, P.R. The impact of light in buildings on human health. *Indoor Built Environ.* **2010**, *19*, 8–20. [[CrossRef](#)]
5. Rea, M.; Figueiro, M. Light as a circadian stimulus for architectural lighting. *Light. Res. Technol.* **2018**, *50*, 497–510. [[CrossRef](#)]
6. Leccese, F.; Salvadori, G.; Casini, M.; Bertozzi, M. Analysis and measurements of artificial optical radiation (AOR) emitted by lighting sources found in offices. *Sustainability* **2014**, *6*, 5941–5954. [[CrossRef](#)]
7. Baker, N.; Steemers, K. *Daylight Design of Buildings: A Handbook for Architects and Engineers*; Routledge: London, UK, 2014.
8. Boubekri, M. *Daylighting Design: Planning Strategies and Best Practice Solutions*; Birkhäuser: Basel, Switzerland, 2014.
9. Leccese, F.; Salvadori, G.; Casini, M.; Bertozzi, M. Lighting of indoor work places: Risk assessment procedure. *Wit Trans. Inf. Commun. Technol.* **2012**, *44*, 89–101.
10. Reinhart, C.F.; Walkenhorst, O. Validation of dynamic RADIANCE-based daylight simulations for a test office with external blinds. *Energy Build.* **2001**, *33*, 683–697. [[CrossRef](#)]
11. Friedman, J.; Hastie, T.; Tibshirani, R. *The Elements of Statistical Learning*; Springer: New York, NY, USA, 2001; Volume 1.
12. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: New York, NY, USA, 2013; Volume 112.
13. Deb, K. *Multi-Objective Optimization Using Evolutionary Algorithms*; John Wiley & Sons: Hoboken, NJ, USA, 2001; Volume 16.
14. Blischke, W.R.; Murthy, D.P. *Reliability: Modeling, Prediction, and Optimization*; John Wiley & Sons: Hoboken, NJ, USA, 2011; Volume 767.
15. Bamdad, K.; Cholette, M.E.; Guan, L.; Bell, J. Ant colony algorithm for building energy optimisation problems and comparison with benchmark algorithms. *Energy Build.* **2017**, *154*, 404–414. [[CrossRef](#)]
16. Assouline, D.; Mohajeri, N.; Scartezzini, J.-L. Quantifying rooftop photovoltaic solar energy potential: A machine learning approach. *Sol. Energy* **2017**, *141*, 278–296. [[CrossRef](#)]
17. Chen, Y.; Tan, H. Short-term prediction of electric demand in building sector via hybrid support vector regression. *Appl. Energy* **2017**, *204*, 1363–1374. [[CrossRef](#)]

18. Deb, C.; Lee, S.E. Determining key variables influencing energy consumption in office buildings through cluster analysis of pre-and post-retrofit building data. *Energy Build.* **2018**, *159*, 228–245. [CrossRef]
19. Costanzo, V.; Evola, G.; Marletta, L.; Pistone Nascone, F. Application of Climate Based Daylight Modelling to the Refurbishment of a School Building in Sicily. *Sustainability* **2018**, *10*, 2653. [CrossRef]
20. IESNA. *LM-83-12 IES Spatial Daylight Autonomy (sDA) and Annual Sunlight Exposure (ASE)*; Lesna Lighting Meas: New York, NY, USA, 2012.
21. Mardaljevic, J.; Heschong, L.; Lee, E. Daylight metrics and energy savings. *Light. Res. Technol.* **2009**, *41*, 261–283. [CrossRef]
22. Rogers, Z. *Daylighting Metric Development Using Daylight Autonomy Calculations in the Sensor Placement Optimization Tool*; Archit. Energy Corp.: Boulder, CO, USA, 2006; Available online: [http://www.daylightinginnovations.com/system/public\\_assets/original/SPOT\\_Daylight%20Autonomy%20Report.pdf](http://www.daylightinginnovations.com/system/public_assets/original/SPOT_Daylight%20Autonomy%20Report.pdf) (accessed on 8 March 2019).
23. Nabil, A.; Mardaljevic, J. Useful daylight illuminances: A replacement for daylight factors. *Energy Build.* **2006**, *38*, 905–913. [CrossRef]
24. Green Building Council. *LEED v4 User Guide*; Green Building Council: Washington, DC, USA, 2013; Volume 12, p. 2014.
25. Reinhart, C.F.; Wienold, J. The daylighting dashboard—A simulation-based design analysis for daylight spaces. *Build. Environ.* **2011**, *46*, 386–396. [CrossRef]
26. Moon, P. Illumination from a non-uniform sky. *Illum. Eng.* **1942**, *37*, 707–726.
27. Tregenza, P. Mean daylight illuminance in rooms facing sunlit streets. *Build. Environ.* **1995**, *30*, 83–89. [CrossRef]
28. Nocera, F.; Lo Faro, A.; Costanzo, V.; Raciti, C. Daylight Performance of Classrooms in a Mediterranean School Heritage Building. *Sustainability* **2018**, *10*, 3705. [CrossRef]
29. Reinhart, C.F.; Mardaljevic, J.; Rogers, Z. Dynamic daylight performance metrics for sustainable building design. *Leukos* **2006**, *3*, 7–31.
30. Boubekri, M.; Lee, J. A comparison of four daylighting metrics in assessing the daylighting performance of three shading systems. *J. Green Build.* **2017**, *12*, 39–53. [CrossRef]
31. Lee, K.S.; Han, K.J.; Lee, J.W. The Impact of Shading Type and Azimuth Orientation on the Daylighting in a Classroom—Focusing on Effectiveness of Façade Shading, Comparing the Results of DA and UDI. *Energies* **2017**, *10*, 635. [CrossRef]
32. Mangkuto, R.A.; Rohmah, M.; Asri, A.D. Design optimisation for window size, orientation, and wall reflectance with regard to various daylight metrics and lighting energy demand: A case study of buildings in the tropics. *Appl. Energy* **2016**, *164*, 211–219. [CrossRef]
33. Van Dijk, D.; Platzer, W. *Reference Office for Thermal, Solar and Lighting Calculations*; IEA-Shc Task: Delft, The Netherlands, 2001; Volume 27.
34. Tahmasebi, M.M.; Banihashemi, S.; Hassanabadi, M.S. Assessment of the variation impacts of window on energy consumption and carbon footprint. *Procedia Eng.* **2011**, *21*, 820–828. [CrossRef]
35. Jakubiec, J.A.; Reinhart, C.F. DIVA 2.0: Integrating daylight and thermal simulations using Rhinoceros 3D, Daysim and EnergyPlus. In Proceedings of the Building Simulation, Sydney, Australia, 14–16 November 2011; pp. 2202–2209.
36. Crawley, D.B.; Lawrie, L.K.; Winkelmann, F.C.; Buhl, W.F.; Huang, Y.J.; Pedersen, C.O.; Strand, R.K.; Liesen, R.J.; Fisher, D.E.; Witte, M.J. EnergyPlus: Creating a new-generation building energy simulation program. *Energy Build.* **2001**, *33*, 319–331. [CrossRef]
37. Cohen, J.; Cohen, P.; West, S.G.; Aiken, L.S. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*; Routledge: London, UK, 2013.
38. Tipping, M.E.; Bishop, C.M. Probabilistic principal component analysis. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **1999**, *61*, 611–622.
39. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2013.
40. Neter, J.; Kutner, M.H.; Nachtsheim, C.J.; Wasserman, W. *Applied Linear Statistical Models*; Irwin: Chicago, IL, USA, 1996; Volume 4.
41. Asadi, S.; Amiri, S.S.; Mottahedi, M. On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design. *Energy Build.* **2014**, *85*, 246–255. [CrossRef]

42. Ghiaus, C. Experimental estimation of building energy performance by robust regression. *Energy Build.* **2006**, *38*, 582–587. [[CrossRef](#)]
43. Wilkinson, L. Tests of significance in stepwise regression. *Psychol. Bull.* **1979**, *86*, 168. [[CrossRef](#)]
44. Hastie, T.J. Generalized additive models. In *Statistical Models in S*; Routledge: London, UK, 2017; pp. 249–307.
45. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [[CrossRef](#)]
46. Heschong, L.; Wymelenberg, V.D.; Andersen, M.; Digert, N.; Fernandes, L.; Keller, A.; Loveland, J.; McKay, H.; Mistrick, R.; Mosher, B. *Approved Method: IES Spatial Daylight Autonomy (sDA) and Annual Sunlight Exposure (ASE)*; IES-Illuminating Engineering Society: New York, NY, USA, 2012.
47. Raynham, P. *BS 8206-2: 2008 Lighting for Buildings—Part 2 Code of Practice for Daylighting*; British Standards Institute: London, UK, 2008.
48. Lee, J.; Jung, H.; Park, J.; Lee, J.; Yoon, Y. Optimization of building window system in Asian regions by analyzing solar heat gain and daylighting elements. *Renew. Energy* **2013**, *50*, 522–531. [[CrossRef](#)]
49. Love, J.A. The evolution of performance indicators for the evaluation of daylighting systems. In Proceedings of the IEEE Industry Applications Society Annual Meeting, Houston, TX, USA, 4–9 October 1992; pp. 1830–1836.
50. Tregenza, P.; Mardaljevic, J. Daylighting buildings: Standards and the needs of the designer. *Light. Res. Technol.* **2018**, *50*, 63–79. [[CrossRef](#)]
51. Reinhart, C.F. Lightswitch-2002: A model for manual and automated control of electric lighting and blinds. *Sol. Energy* **2004**, *77*, 15–28. [[CrossRef](#)]
52. Changyong, F.; Hongyue, W.; Naiji, L.; Tian, C.; Hua, H.; Ying, L. Log-transformation and its implications for data analysis. *Shanghai Arch. Psychiatry* **2014**, *26*, 105.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).