Comparision of Different Data-Driven Models on Prediction of Useful Daylight Illuminance (UDI)

İlkim Canlı, Orçun Koral İşeri and İpek Gürsel Dino

August 26, 2021

Comparision of Different Data-Driven Models on Prediction of Useful Daylight Illuminance (UDI)

I.Canlı[1], O. K. Işeri[1] and I. G. Dino[1]

[1] Department of Architecture, Middle East Technical University, Turkey, Ankara
*ilkim.canli@metu.edu.tr

Please indicate your preference for attending the symposium	☒ online	☐ in-person	☐ not decided yet

Abstract

Energy consumption of the built environment worldwide is increasing at a faster rate than the population. With building energy consumption, the amount of greenhouse gas released into the atmosphere also increases, further enhancing global warming. In this context, lighting is seen as one of the most critical issues associated with energy consumption. Despite many studies related to the evaluation of daylighting in specific buildings, few studies examine daylighting in the urban context. Therefore, daylighting in dense urban areas should be analyzed and predicted for optimization. The predictions using surrogate models based on machine learning approaches have the potential to predict daylighting results and reduce the environmental impact of the buildings. This research aims to enhance the applications of machine learning approaches in daylighting predictions in the urban context. As a case study, a small urban area in Turkey, Ankara, is simulated, and different machine learning models (i.e., Multiple Linear Regression, Artificial Neural Network, Random Forest) are presented. Performances of prediction models will be compared.

Keywords: Daylighting, Daylighting Metrics, Machine Learning, Prediction Models

1.Introduction

Today, every building constructed is a serious sustainability problem from energy consumption which causes climate change. In Europe and the other developed countries, almost 40% of total energy is used for buildings energy consumption (Allouhi et al., 2015; USDOE, 2019). One of the most crucial factors in the energy consumption of buildings is lighting, which is one of the most used areas of electrical energy. Lighting accounts for approximately %15 of the total electric energy consumption (Ryckaert et al., 2010). At this point, usage of daylighting in buildings saves energy and costs significantly. The integration of daylighting with buildings does not only affect electricity usage, but also it reduces the internal heat gain by reducing lighting equipment (Pan et al., 2007). Moreover, it reduces the air conditioning requirements by reducing internal heat gain. Efficient indoor daylight can drastically reduce the need for artificial lighting, thereby reducing building energy use.

Moreover, appropriate daylighting in buildings increases the well-being, visual comfort, and productivity of building occupants. It is essential to understand and analyze daylighting metrics and design parameters affecting daylighting usage in buildings. In addition to that, people who live in urban areas and buildings in urban increased with an uncontrollable speed by causing enormous energy demands in urban areas. However, increasing urban density leads to a conflict between space-use efficiency and daylight access (Dogan & Park, 2019). Today, the distribution of daylighting to the inside of the building is challenging because of context buildings in dense urban areas. The effectiveness of daylighting can be reduced in dense urban settings, as the surrounding buildings act as context shading and reduce the reach of direct solar illumination to the indoor spaces. This study aims to create a prediction model with different data-driven models for daylighting performance metrics in the urban context.

2.Literature Review

The most common methods to analyze daylighting metrics are analytical formulas, computer simulations, and machine learning techniques. Analytical formulas are traditional ones to estimate mostly illuminance in buildings (Kazanasmaz et al., 2009). Metrics such as daylight factor, sky component, internally and externally reflected components could be calculated by using formulas. For manual calculation, the Lumen Method was introduced by Frühling to approximate DF by mathematical formulas (Ayoub, 2019), Treganza proposed modification of split-flux formula for daylight factor and internal reflected component with large external obstructions (Treganza, 1989), and formulas were modified in years for different kind of spaces. Another method to evaluate daylighting metrics is a computer simulation. Various computer simulation programs have been used to calculate daylight, such as DIVA (Jakubiec & Reinhart, 2011), RADIANCE (Ward, 1994), DAYSIM (Reinhart & Walkenhorst, 2001), Energy Plus (Crawley et al., 2000). Although these tools allow detailed analysis in daylight assessment, only experienced users can obtain accurate results. Moreover, as the simulation model and inputs expand, it can take hours to get detailed results, limiting designers' use of the tool in the early design phase. Machine learning applications are the most up-to-date method, fast and detailed tool in daylight evaluation.

Recently, machine learning approaches in daylight studies have increased considerably in the literature and many researched studies evaluated daylighting metrics and parameters: Lee et al. (2019) tried to analyze and predict the impact of building design parameters on daylighting metrics using statistical learning techniques. Ngarambe et al. (2020) predicted indoor daylight illuminances by comparing different machine learning techniques. Maltais & Gosselin (2017) tried to evaluate buildings' performance based on daylighting using sensitivity analysis, metamodel, and Pareto Front methods. Kazanasmaz & Günaydın studied a prediction model developed to determine daylight illuminance for office buildings using artificial neural networks (Kazanasmaz et al., 2009). Although there are many studies related to machine learning techniques to calculate daylighting metrics, just a few consider the urban context that affects daylighting. However, the urban environment influences daylighting results to such an extent that they cannot be neglected. Surrounding building heights, their material characteristics, and urban canyon characteristics contribute to the amount of direct sunlight and diffuse daylight that a building facade and adjacent street receive (Saratsis et al., 2017). There are a few comprehensive studies that comprehend the urban environment. Reinhart et al. (2013) studied an urban modeling interface to calculate annual daylight availability for each building.

For the decision of evaluation metrics, their pros and cons are evaluated. Daylight Factor indicates the ratio of the light level inside a structure to the light level outside the building. However, it does not consider different sky conditions and building locations. As a climate-based metric, Daylight Autonomy indicates how often the minimum illuminance threshold is met or exceeded during occupancy hours (Dogan & Park, 2019). Maximum Daylight Autonomy indicates the percentage of the annual occupied timesteps when the illuminance level exceeds ten times a predefined threshold (Rogers & Goldman, 2006).

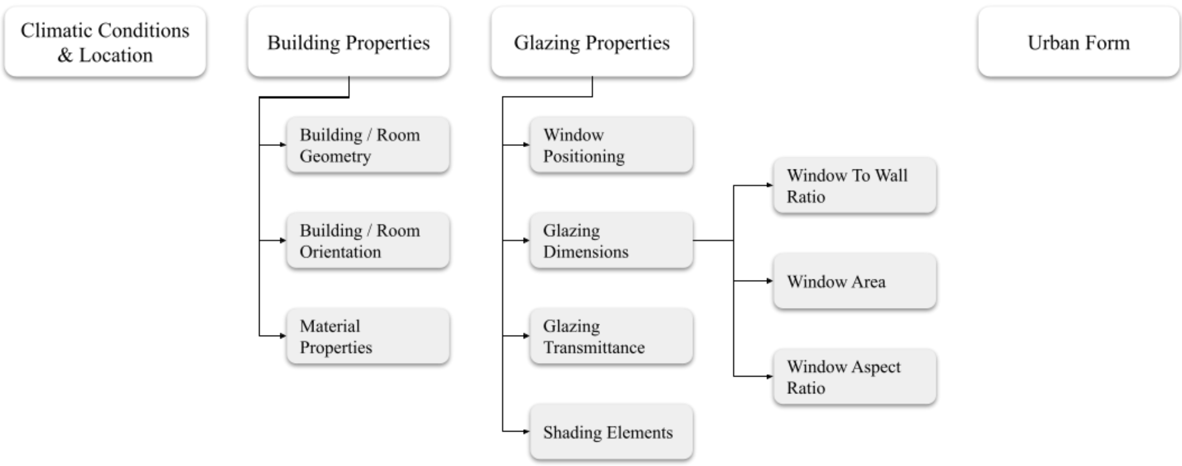

Figure 1. Design parameters affect the UDI

Useful Daylight Illuminance (*UDI*) measures the percentage of occupancy hours during which illuminances fall within a range of 100 lx and 3000 lx (*UDI* combined) (Dogan & Park, 2019). Compared to Daylight Autonomy metrics, UDI offers more information to assess visual and thermal discomfort (Ayoub, 2019). Moreover, as a dynamic and climate-based metric, Useful Daylight Illuminance (*UDI*) provides knowledge for different conditions by dividing the annual illuminance distribution into three bins (Nabil & Mardaljevic, 2006). By indicating lower and upper thresholds, it shows the percentage of occupancy hours in which oversupply of daylight happens and insufficient daylight that occur (Yu & Su, 2015). It gives an idea about excessive heat gain, visual and thermal discomfort of occupants. There are also glare indices metrics that can help to evaluate visual comfort; however, calculating these metrics takes a long time and long-term studies. Therefore, in the scope of this study, only *UDI* is determined as a performance metric for prediction models. **Figure 1** shows the design parameters that affect the UDI value.

3.Methodology

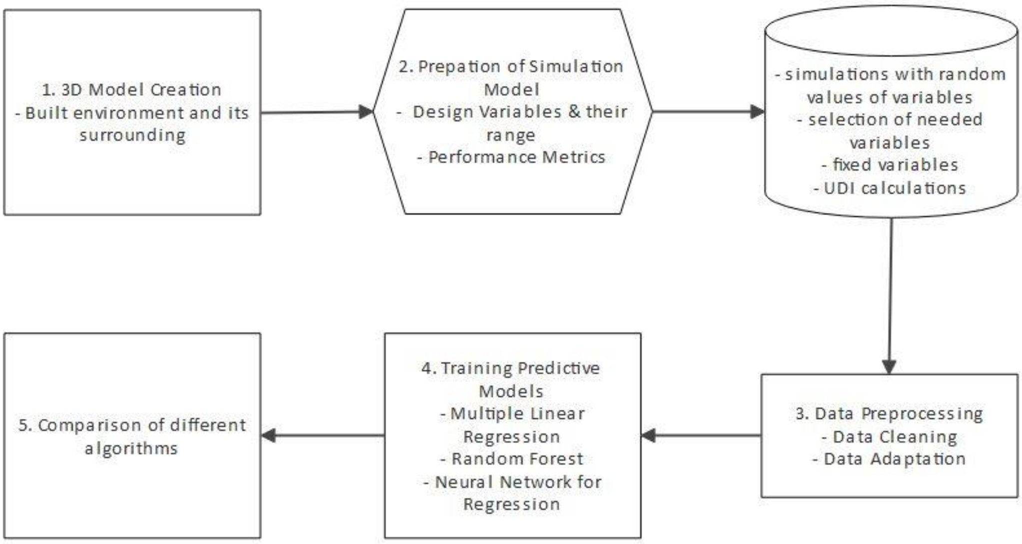

Figure 2. The flow of the process

3.1.Creation of 3D Model and dataset

In this research, as Bahçelievler reflects one of the densest urban areas in Ankara, a small urban context in Bahçelievler was studied. The 3D model was created in Rhino. The studied building has four floors, and each floor was divided into four units. To be able to discuss needed daylighting for different spaces, each unit was divided into three: one is a living room, and the other two are bedrooms. While there is one glazing in bedrooms, the living room has two glazing faces in the east and south direction. As an urban context, exterior buildings were represented as surfaces instead of 3D modeling to reduce computational cost. Surfaces that only affect the amount of light falling on the studied building are modeled as surfaces. Trees were detected according to Google Maps and modeled in an abstract way. They were raised 3m to reflect the shadow of trees in a proper way. However, after trying different shape scenarios, it was observed that tree shapes did not affect the results so much, and they were represented as a decagon. **Figure 3** shows the 3D model and studied area's location, and **Figure 4** shows the division of each unit.

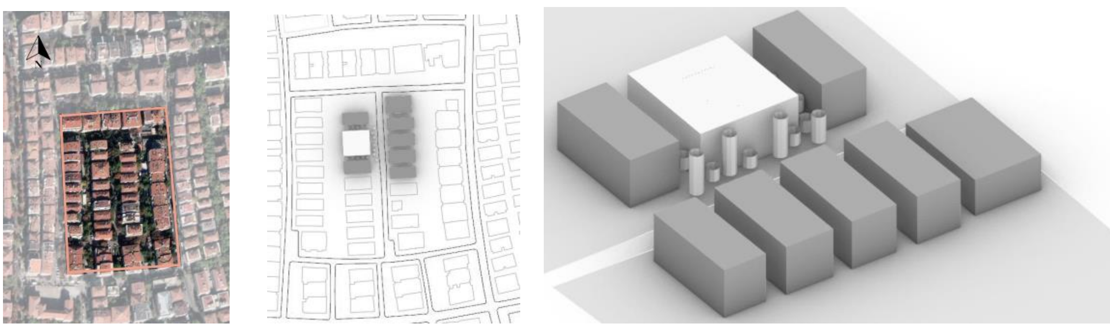

Figure 3. 3D Model and studied area's location

Figure 4. The division of each unit

3.1.1.Database Generation

In this study, Honeybee and Ladybug plugins were used to simulate the model. According to the literature view, *UDI* was selected as a performance metric, and inputs were investigated, affecting the *UDI* value. Initially, nine design parameters were selected as input: x_1: glazing width, x_2: glazing height, x_3: transmittance value, x_4: interior wall reflectance, x_5: interior floor reflectance, x_6: interior ceiling reflectance, x_7: tree reflectance, x_8: exterior ground reflectance, and x_9: context building reflectance. Weather data was obtained, and the model was simulated. The modeled area was divided into five grids, and the average *UDI* value falling on these five points was calculated.

At first, nine inputs were given to the model with only the maximum and minimum extreme values of 0.1 and 0.9 to make feature selection with the parameters that affect the results the most. The results showed that glazing width, glazing height, transmittance value, and context building reflectance are the most significant parameters that affect the average *UDI*. **Figure 5.** shows the correlation matrix between inputs and the average *UDI* value.

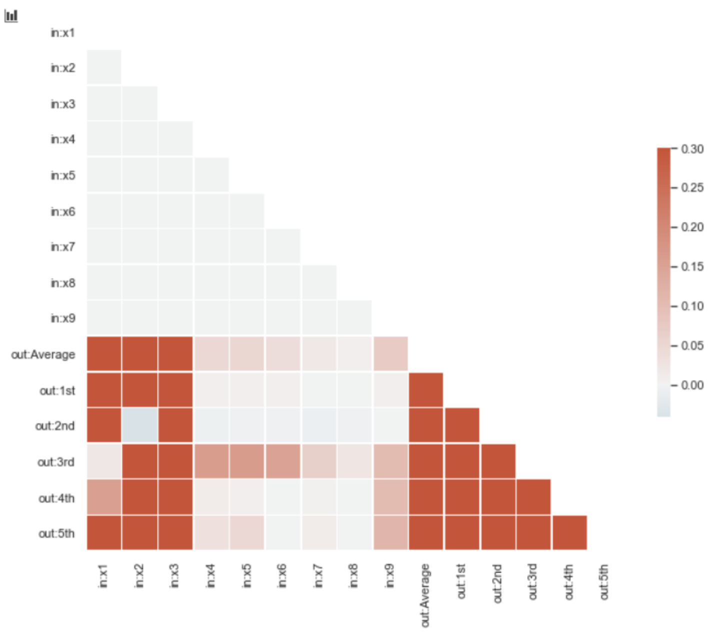

Figure 5. Correlation matrix between inputs and UDI value

After selecting needed inputs, fixed values were entered in other inputs by scanning the literature. **Table 1** shows the design parameters and references for fixed values. Four inputs with values between 0.1 and 0.9 were given as input to the model with random values and simulated.

Design Parameters	Values
X_1: glazing width	0.1< <0.9
X_2: glazing height	0.1< <0.9
X_3: transmittance value	0.1< <0.9
X_4: Interior wall reflectance (fixed)	0.5 (IESNA Handbook (2011))
X_5: Interior floor reflectance (fixed)	0.2 (IESNA Handbook (2011))
X_6: Interior ceiling reflectance (fixed)	0.8 (IESNA Handbook (2011))
X_7: tree reflectance (fixed)	0.2 (Sielachowska et al., 2020)
X_8: exterior ground reflectance (fixed)	0.3 (Page & Lebens, 1986)
X_9: context building reflectance	0.1< <0.9

Table 1. Design parameters and references for fixed values

3.2.Methods

3.2.1.Multiple Linear Regression

Multiple Linear Regression represents the relationship between the dependent variable and n number of explanatory variables (Grégoire, 2015).

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \dots\dots\dots + \beta nXn + \mathcal{E}\,(1)$$

Given equation (1) shows the relationship between the dependent variable (Y) and independent variables ($X1, X2, \dots\dots Xn$). $\beta 0$ is the constant term, $\beta 1 \dots\dots\dots \beta n$ represents the coefficient factor of independent variables, and Ɛ represents the notation for the model deviations.
Multiple linear regression models were used to predict daylighting performance metrics and design parameters in various studies (Kim et al., 2007; Turan et al.,2020).

3.2.2.Random Forest

Random Forest (*RF*) is a supervised learning algorithm that consists of decision trees and can be used for both regression and classification tasks (Breiman, 2001). The predictions are based on average results of decision tree classifiers on various sub-samples on data.
RF is one of the most used supervised learning algorithms in building energy consumption predictions and daylighting problems due to its prediction based on average results of an ensemble of decision trees that increase the accuracy of prediction results (Ahmad et al., 2017; Ayoub, 2020).

3.2.3.Neural Network for Regression

Neural Networks (*NN*) are algorithms that are inspired by information on how the human brain learns and its process (Wang, 2003). Basically, *NN* is composed of three main layers: input layer, hidden layer, and output layer, and each input has different weights that affect the prediction results. *NN*s are trained to make predictions by applying training sets, and its result is verified by using testing sets. *NN*s adjust themselves based on the error.
Neural Network algorithms have been widely used in building energy forecasting models, energy consumption predictions, retrofit scenarios, and daylighting forecasts in recent years (Robinson et al., 2017; Wang et al., 2019).

The Performance Evaluation

To be able to compare performances of different learning algorithms, four main performance evaluation metric was calculated in this study: Mean Absolute Error (*MAE*) (2), Mean Squared Error (*MSE*) (3), Root Mean Squared Error (*RMSE*) (4) and Coefficient of Determination (*R2*) (5).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}| \tag{2}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2 \tag{3}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2} \tag{4}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} \tag{5}$$

where $\hat{y}$ is predicted value of y and $\bar{y}$ is the mean value of y.

4.Results

4.1.Results of Multiple Linear Regression

Results for multiple linear regression were tested with different sizes of training sets. **Table 2** shows the results of performance evaluation metrics with different sizes of the dataset.

Table 2. The evaluation of Multi-Linear Regression results with different sizes of testing data

Y	*MAE*	*MSE*	*RMSE*	*R2*	*Testing Data Size*
Average *UDI*	1.829	5.122	2.263	0.932	10%
Average *UDI*	1.832	5.281	2.298	0.931	20%
Average *UDI*	1.841	5.272	2.296	0.926	30%

Among the algorithms used in the study, it was the fastest algorithm with the least computational cost. Since it is known that the relationship between the input parameters and the output is linear, it is predicted that the R^2 value of the model will be high, and the results of the study have proven this. As the R^2 value increases, model predictions get closer to real results, the accuracy of the model increases. With the lowest error rate and highest R^2 value, the regression model that made the closest estimation to reality was the model in which 10% of the test data was used.

4.2.Results of Random Forest

Results for Random Forest were tested with different sizes of training sets. **Table 3** shows the results of performance evaluation metrics with different sizes of the dataset.

Table 3. The evaluation of Random Forest results with different sizes of testing data

Y	*MAE*	*MSE*	*RMSE*	*R2*	*Testing Data Size*
Average *UDI*	0.518	0.501	0.708	0.992	10%
Average *UDI*	0.534	0.536	0.732	0.992	20%
Average *UDI*	0.662	0.784	0.885	0.991	30%

Random Forest algorithm was successful at prediction with the highest R^2 value compared with other algorithms. The number of testing data sizes did not affect the R^2 results dramatically, but when testing data size increased, *MAE*, *MSE*, and *RMSE* values were also increased, which means the model deviated from accuracy.

4.3.Results of Neural Network for Regression

Results for Neural Network for Regression were tested with different sizes of training sets. **Table 4** shows the results of performance evaluation metrics with different sizes of the dataset.

Table 4. The evaluation of ANN results with different sizes of testing data

Y	*MAE*	*MSE*	*RMSE*	*R2*	*Testing Data Size*
Average *UDI*	0.084	0.011	0.105	0.987	10%
Average *UDI*	0.075	0.009	0.095	0.991	20%
Average *UDI*	0.09	0.013	0.115	0.986	30%

Neural Network algorithm showed the best performance with a high R^2 value and low *MAE*, *MSE*, and *RMSE* values. Changing the size of test data did not affect the results much, but the model showed the best performance with a 20% testing size.
In Neural Network, Hyperparameter tuning was used to train the model. Different batch size and epoch size alternatives were tested with manual grid search, and the model was retrained with the values that increased the accuracy of the model the most. The computational cost was higher compared to other algorithms due to the Hyperparameter tuning stage.

5.Discussion

In this research, performances of three different prediction algorithms which have been using in daylighting were evaluated. The building in Bahçelievler is divided into three rooms, two of which are simulated as a bedroom and the other as a living room. At the same time, the building was divided into grids, and *UDI* values on five points formed because of partitioning were calculated. Much lighter fell on the living room unit, which has openings in two different directions. The east-facing bedroom receives much more light compared to the south-facing bedroom because there is a surrounding building blocking the light next to the south-facing bedroom. The dataset used was limited as it only contained four inputs. It can be enlarged in future studies to enhance algorithms' performances and develop prediction models. Also, only one daylighting metric (*UDI*) was estimated against four inputs given to the models. More daylighting performance metrics should be added, such as daylight glare probability, annual sunlight exposure, and spatial daylight autonomy, to create a comprehensive urban daylighting model. Future studies aim to analyze the effect of the urban environment on glare and use prediction models in which glare metrics are included. The striking point in the study is that the four parameters found to be the most effective on daylighting sometimes show incorrect results in the extreme values they can take. Still, *NN* for regression and Random Forest were successful at prediction models with high R^2 scores, and *NN* for regression showed the best result based on *MAE*, *MSE*, and *RMSE*.

6.Conclusion

In this study, the relationship between building design parameters and daylighting metric (*UDI*) was analyzed, and prediction models were created based on three different algorithms: multiple linear regression, random forest, and *ANN* for regression. At first, the model was simulated with nine different design parameters. Then to reduce the computational cost, it has been re-simulated with four parameters that affect the *UDI* value the most. It has been observed that the surrounding buildings greatly affect the daylight falling on the building. The aim of this research was to show and compare the performance results of different algorithms which have been using in daylighting. To evaluate performances of algorithms, each algorithm was trained with different sizes of training sets. According to the results, random forest and *NN* showed better results in terms of R^2 score. In more detail, *NN* showed better performance with less *MAE*, *MSE*, *RMSE* values. In building base and including urban context, these predictive models help the designer better understand the relationship between daylighting in building and its urban context. However, more design parameters and more metrics can be added to predictive models to increase data and algorithm prediction accuracy.

References

Ahmad, M. W., Reynolds, J., & Rezgui, Y. (2017). Random Forests and Artificial Neural Network for Predicting Daylight Illuminance and Energy Consumption BRE Centre for Sustainable Engineering, School of Engineering, AhmadM3@Cardiff.ac.uk ; 2 HippolyteJ@Cardiff.ac.uk ; Related work. Building Simulation, May, 1–7.

Allouhi, A., El Fouih, Y., Kousksou, T., Jamil, A., Zeraouli, Y., & Mourad, Y. (2015). Energy consumption and efficiency in buildings: Current status and future trends. *Journal of Cleaner Production*, *109*, 118–130. https://doi.org/10.1016/j.jclepro.2015.05.139

Ayoub, M. (2019). 100 Years of daylighting: A chronological review of daylight prediction and calculation methods. Solar Energy, 194(August), 360–390. https://doi.org/10.1016/j.solener.2019.10.072

Ayoub, M. (2020). A review on machine learning algorithms to predict daylighting inside buildings. Solar Energy, 202(April), 249–275. https://doi.org/10.1016/j.solener.2020.03.104

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.

Crawley, B. D. B., Pedersen, C. O., Lawrie, L. K., & Winkelmann, F. C. (2000). EnergyPlus : Energy Simulation Program. ASHRAE Journal, (April), 49–56. https://doi.org/10.1.1.122.6852

DiLaura, D. L. (2011). The lighting handbook: reference and application. Illuminating Engineering Society of North America.

Dogan, T., & Park, Y. C. (2019). A critical review of daylighting metrics for residential architecture and a new metric for cold and temperate climates. Lighting Research and Technology, 51(2), 206–230. https://doi.org/10.1177/1477153518755561

Grégoire, G. (2015). Multiple linear regression. EAS Publications Series, 66, 45–72. https://doi.org/10.1051/eas/1466005

Jakubiec, J. A., & Reinhart, C. F. (2011). DIVA 2.0: Integrating daylight and thermal simulations using rhinoceros 3D, DAYSIM and EnergyPlus. Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association, 2202–2209.

Kazanasmaz, T., Günaydin, M., & Binol, S. (2009). Artificial neural networks to predict daylight illuminance in office buildings. Building and Environment, 44(8), 1751–1757. https://doi.org/10.1016/j.buildenv.2008.11.012

Kim, S. Y., & Kim, J. J. (2007). The impact of daylight fluctuation on a daylight dimming control system in a small office. Energy and Buildings, 39(8), 935–944. https://doi.org/10.1016/j.enbuild.2006.10.009

Lee, J., Boubekri, M., & Liang, F. (2019). Impact of building design parameters on daylighting metrics using an analysis, prediction, and optimization approach based on statistical learning technique. Sustainability (Switzerland), 11(5). https://doi.org/10.3390/su11051474

Maltais, L. G., & Gosselin, L. (2017). Daylighting 'energy and comfort' performance in office buildings: Sensitivity analysis, metamodel and pareto front. Journal of Building Engineering, 14(September), 61–72. https://doi.org/10.1016/j.jobe.2017.09.012

Nabil, A., & Mardaljevic, J. (2006). Useful daylight illuminances: A replacement for daylight factors. Energy and Buildings, 38(7), 905–913. https://doi.org/10.1016/j.enbuild.2006.03.013

Ngarambe, J., Irakoze, A., Yun, G. Y., & Kim, G. (2020). Comparative performance of machine learning algorithms in the prediction of indoor daylight illuminances. Sustainability (Switzerland), 12(11), 1–23. https://doi.org/10.3390/su12114471

Page, J. & Lebens, R. (eds) Climate in the United Kingdom. A handbook of solar radiation, temperature and other data for thirteen principal cities and towns. London: HMSO, 1986

Pan, Y., Huang, Z., & Wu, G. (2007). Calibrated building energy simulation and its application in a high-rise commercial building in Shanghai. Energy and Buildings, 39(6), 651–657. https://doi.org/10.1016/j.enbuild.2006.09.013

Reinhart, C. F., Dogan, T., Jakubiec, J. A., Rakha, T., & Sang, A. (2013). UMI - An urban simulation environment for building energy use, daylighting and walkability. Proceedings of BS 2013: 13th Conference of the International Building Performance Simulation Association, 476–483.

Reinhart, C. F., & Walkenhorst, O. (2001). Validation of dynamic RADIANCE-based daylight simulations for a test office with external blinds. Energy and Buildings, 33(7), 683–697. https://doi.org/10.1016/S0378-7788(01)00058-5

Robinson, C., Dilkina, B., Hubbs, J., Zhang, W., Guhathakurta, S., Brown, M. A., & Pendyala, R. M. (2017). Machine learning approaches for estimating commercial building energy consumption. Applied Energy, 208(May), 889–904. https://doi.org/10.1016/j.apenergy.2017.09.060

Rogers, Z., Goldman, D., 2006. Daylighting Metric Development Using Daylight Autonomy Calculations in the Sensor Placement Optimization Tool. Boulder, Colorado, USA.

Ryckaert, W. R., Lootens, C., Geldof, J., & Hanselaer, P. (2010). Criteria for energy efficient lighting in buildings. Energy and Buildings, 42(3), 341–347. https://doi.org/10.1016/j.enbuild.2009.09.012

Saratsis, E., Dogan, T., & Reinhart, C. F. (2017). Simulation-based daylighting analysis procedure for developing urban zoning rules. Building Research and Information, 45(5), 478–491. https://doi.org/10.1080/09613218.2016.1159850

Sielachowska, M., & Zajkowski, M. (2020). Assessment of Light Pollution Based on the Analysis of Luminous Flux Distribution in Sports Facilities. Mechanisms and Machine Science, 70(January), 139–150. https://doi.org/10.1007/978-3-030-13321-4_12

Treganza, 1989. Modification of split-flux formulae for mean daylight factor and internal reflected component with large external obstructions. Lighting Res. Technol. 21(3), 125-128

Turan, I., Chegut, A., Fink, D., & Reinhart, C. (2020). The value of daylight in office spaces. Building and Environment, 168(October 2019), 106503. https://doi.org/10.1016/j.buildenv.2019.106503

USDOE, 2019. 2017 renewable energy data book. (NREL) National Renewable Energy Laboratory

Wang, S.-C. (2003). Artificial Neural Network. Interdisciplinary Computing in Java Programming, 81–100. doi:10.1007/978-1-4615-0377-4_5

Wang, W., Hong, T., Xu, X., Chen, J., Liu, Z., & Xu, N. (2019). Forecasting district-scale energy dynamics through integrating building network and long short-term memory learning algorithm. Applied Energy, 248(March), 217–230. https://doi.org/10.1016/j.apenergy.2019.04.085

Ward, Gregory J., "The RADIANCE Lighting Simulation and Rendering System," Computer Graphics (Proceedings of '94 SIGGRAPH conference), July 1994, pp. 459-72.

Yu, X., & Su, Y. (2015). Daylight availability assessment and its potential energy saving estimation -A literature review. Renewable and Sustainable Energy Reviews, 52, 494–503. https://doi.org/10.1016/j.rser.2015.07.142