Building and Environment 206 (2021) 108335

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

Comparing performance of discomfort glare metrics in high and low adaptation levels

Geraldine Quek [a,*], Jan Wienold [a], Mandana Sarey Khanie [b], Evyatar Erell [c], Eran Kaftan [c], Athanasios Tzempelikos [d], Iason Konstantzos [e], Jens Christoffersen [f], Tilmann Kuhn [g], Marilyne Andersen [a]

[a] LIPID, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[b] ICIEE, Technical University of Denmark, Kongens Lyngby, Denmark
[c] BIDR, Ben-Gurion University of the Negev, Sde Boqer Campus, Israel
[d] School of Civil Engineering, Purdue University, West Lafayette, IN, USA
[e] The Durham School, University of Nebraska, Lincoln, NE, USA
[f] VELUX A/S, Hoersholm, Denmark
[g] Fraunhofer Institute for Solar Energy Systems ISE, Freiburg, Germany

ARTICLE INFO

Keywords:
Glare
Discomfort glare
Glare metrics
Saturation
Contrast

ABSTRACT

Current discomfort glare prediction metrics usually account for at least one of the two categories of effects that induce discomfort glare – the saturation and contrast effects. Saturation-driven metrics (overall illuminance on the eye) are suited for brightly lit scenes in general. On the contrary, contrast-driven metrics (luminance ratio in the field of view) usually perform better in high contrast conditions such as with small-sized bright glare sources. Only a few existing metrics consider both effects, such as the Daylight Glare Probability (DGP). However, even these "hybrid" metrics may underperform in conditions other than those considered when they were developed, such as in dim scenes with high contrast glare. This paper investigates the ability of current glare indicators to predict perceived discomfort glare in user-evaluated scenes depending on two different adaptation levels. Towards this end, we used a composite dataset of six laboratory studies performed previously and separately in various parts of the world. According to Receiver Operator Characteristics (ROC) findings and complementary statistical research, the hybrid metrics DGP and Eccologit perform best in both investigated ranges (dimmer and brighter scenes). For the single-effect metrics, the contrast-driven metrics appear to perform better than saturation-driven metrics in lower adaptation levels (dimmer scenes), while the reverse is seen in higher adaptation levels (brighter scenes). As a result, metrics that only consider one effect should be used with caution. Although hybrid metrics continue to perform well in the investigated scenes, further research is needed to extend their applicability to a larger variety of lighting conditions that may be observed in work environments.

1. Introduction

Visual comfort in indoor workspaces is one of the key components that contribute to occupant satisfaction, along with other indoor environmental qualities such as thermal conditions, air quality, and the acoustic environment [1]. The presence of windows and views to the outside have been strongly linked to occupants' satisfaction, stress, and productivity [2,3]. In addition to providing illumination, the role of daylight in indoor spaces also extends to physiological aspects, such as entraining the human body's circadian clock and inducing alertness through a non-image forming pathway [4–6]. In other words, one could say that daylight in buildings has much to offer as the primary source of illumination [7,8], from human health and wellbeing to energy efficiency. However, in addition to overheating risks associated with excessive solar gains, poorly managed daylight has other liabilities, such as visual or thermal discomfort. Discomfort glare [9] and disability glare [10], are indeed amongst the primary visual comfort concerns. Disability glare impairs vision through intraocular scatter [11] or by reduced contrast on visual displays. Discomfort glare, which is much more common in workspaces, is defined, by the International

* Corresponding author. École Polytechnique Fédérale de Lausanne (EPFL), EPFL ENAC IA LIPID, LE 1 113, Bâtiment LE, Station 18, 1014, Lausanne, Switzerland.
E-mail address: geraldine.quek@epfl.ch (G. Quek).

https://doi.org/10.1016/j.buildenv.2021.108335
Received 9 June 2021; Received in revised form 6 September 2021; Accepted 7 September 2021
Available online 16 September 2021

Commission on Illumination (CIE) as "glare that causes discomfort without necessarily impairing the vision of objects" [12]. Empirical metrics have been developed to predict subjective discomfort glare in one's visual field based on photometric properties, but their performance is typically limited to the range of scenarios present in the dataset they were developed from. Most metrics were developed from laboratory setups in typical office-like rooms subject to relatively high levels of illuminance; hence it is not entirely surprising that such metrics have poor prediction performance for discomfort glare in conditions where illuminance falls in lower levels [13–15]. A previous cross-validation study evaluated the performance of twenty-two glare metrics based on a dataset of six laboratory user studies, that had been conducted in five different countries, and which were not used for the development of the metrics themselves [16]. The lighting conditions experienced by the participants in these user studies covered a range of adaptation levels, from quite bright to quite dim scenarios, and the previous cross-validation study evaluated the ability of the 22 glare metrics to predict discomfort glare as reported by the participants. This paper revisits the same dataset with the goal of assessing the performance of discomfort glare metrics in dim and bright adaptation levels.

2. Background

Discomfort glare can occur due to excessive brightness, also known as the saturation effect, or due to an excessive range of luminances in the field of view, logically termed the contrast effect. The four main factors influencing discomfort glare when considering these effects are, on one hand, the luminance, the size, and the position of the glare source, and on the other hand, the adaptation level or overall illumination [17]. Glare metrics usually account for at least one effect that explains discomfort glare - the contrast effect, or the saturation effect - but may incorporate both. The contrast effect is usually represented by a ratio between the luminance and size of the glare source and the adaptation level and position of the glare source in the field of view. The saturation effect is usually represented by the overall photopic illuminance at the eye, to check for excessively bright sources of light. For example, Daylight Glare Probability (DGP) [18], a glare metric in the European Standard "Daylight in Buildings" EN17037 [19], has two main terms in its equation which corresponds to the saturation effect and contrast effect respectively:

$$DGP = 5.87 \cdot 10^{-5} E_v + 9.18 \cdot 10^{-2} \log_{10}\left(1 + \sum_{i=1}^{n} \frac{L_{s,i}^2 \omega_{s,i}}{E_v^{1.87} P_{s,i}^2}\right) + 0.16 \quad (1)$$

where E_v refers to the vertical illuminance (lux), $L_{s,i}$ refers to the luminance of the glare source (cd/m^2), $\omega_{s,i}$ refers to the solid angle of the glare source in steradians, and $P_{s,i}$ refers to the position index of the glare source for the i-th glare source.

Such glare metrics are calculated from luminance maps, either in the form of simulation-based renderings or as high dynamic range (HDR) images [20–22]. Experimental data pertaining to the luminous environment, such as calibrated fisheye HDR images, are typically taken at different points in time during user assessments to generate luminance maps of the visual environment that the user experiences, calibrated using spot luminance measurements. Simulation-based renderings can be computed annually for glare evaluation using calibrated daylighting models [23,24], subject to accuracy and computational power. Glare sources are then identified from the luminance maps through various approaches. The glare detection methods have an impact on the final calculations of glare metrics and should be executed meticulously according to recommendations [22]. For example, in task-driven situations, areas in the field of view are identified where the luminance exceeds five times that of the task area, or are above a fixed luminance threshold of 2000 cd/m^2. Thresholds or cut-off points, specific to each glare metric are then used to predict if there is glare in the scene. For example, if the calculated DGP of a luminance map is equal to or greater than 0.38, disturbing glare is predicted [16]. As a hybrid metric, DGP performed best in a cross-validation study conducted [16], as it also considers a secondary contrast effect term in addition to the main saturation effect term. This makes it quite adaptable to a range of scenarios, such as where the contrast effect may explain discomfort glare perception better. Despite this, glare prediction by DGP might still be unsatisfactory in scenarios that differ significantly from the user-evaluated scenes under which this metric was developed.

In the dataset of user studies analyzed during the development of the DGP (DE-DK-Ecco), the evaluated scenes were generally brightly lit, E_v with a mean of 3426 lux and ranging from 153 to 11298 lux. However, DGP may not predict glare well in scenes that differ from its development dataset due to extrapolation, particularly the relative weighting of the saturation and contrast effect terms. Some researchers have found that DGP does not perform well in predicting discomfort glare in post-occupancy evaluations (POE) of some office spaces: This is likely because deep open-plan offices are characterized by lower vertical illuminance [14] (with a mean E_v of 445 lux) [24–27]. In these POE scenes found in office buildings and also in laboratory studies [28], glare frequently occurs due to the contrast effect between the luminance of the window [29] and an overall low background level to which the eye adapts to. Here, E_v frequently correlates poorly with subjective discomfort glare responses, as the saturation effect is not the primary cause of glare, and instead the contrast effect dominates. In such cases, the formulation of the DGP equation generally places more emphasis on the saturation effect term, E_v, and its accompanying coefficient after accounting for the order of magnitudes of E_v (usually around $\cdot 10^3$ to $\cdot 10^4$ lux, assuming that the vertical illuminance is around 1000–10000 lux) and of the contrast term in DGP, log_gc (usually between $\cdot 10^{-2}$ to $\cdot 10^1$):

$$\log\_gc = \log_{10}\left(1 + \sum_{i=1}^{n} \frac{L_{s,i}^2 \omega_{s,i}}{E_v^{1.87} P_{s,i}^2}\right) \quad (2)$$

Thus, in general office environments, DGP may perform less well in predicting glare in lower light situations (in the photopic range) where the contrast effect plays a bigger role in causing glare. However, it should also be noted that for dim scenarios when the sun disk is visible through the façade, the contrast term in DGP may dominate despite the low emphasis in the current equation [30].

Overall, determining appropriate discomfort glare metrics in high and low adaptation levels, as well as whether contrast or saturation effects are dominant under each lighting scenario, remains a challenge. It is likely that for DGP to predict glare accurately in a wider range of scenarios, a more nuanced balance between the coefficients of the saturation and contrast terms might be necessary, depending on which effect is the primary cause of glare for various scenarios. This paper aims to evaluate and compare the predictive performance of three categories of 15 glare metrics (hybrid, contrast-driven, and saturation-driven), in scenes considered as either "bright" or "dim", and provide recommendations for future research.

3. Methodology

The purpose of this paper is to assess the ability of existing glare metrics to predict perceived discomfort, with a specific focus on different brightness levels. We hypothesize that, while hybrid metrics may perform well in most scenarios, contrast-driven metrics are better at predicting glare in dimmer scenes and saturation-driven metrics are better at predicting glare in brighter scenes, depending on the eye's level of adaptation. The evaluation is based on a composite dataset of six user studies conducted in controlled daylit rooms in office-like settings [28, 31–35]. All six experiments were deemed eligible for this and previous assessments because they avoided reflections on computer screens and concentrated mostly on discomfort glare rather than disability glare [16]. Because E_v was found to be a good predictor of pupil size [36], this analysis employs it as a discriminator for the level of adaptation to

which the eye adapts.

3.1. Glare metrics

Table 1 contains a list of the 15 discomfort glare metrics that we evaluate in this paper, categorized by one or both effects they mainly account for (hybrid/contrast/saturation), as well as their abbreviations and references. Due to the monotonic nature of the statistical tests used in this study, two pairs of metrics with the same empirical expression were merged for simplicity; UGR and UGP (Metric 9), as well as E_v and the Simplified DGP (DGPs) (Metric 15). In addition to common glare metrics developed for daylit scenes such as DGP, Daylight Glare Index (DGI), and the Predicted Glare Sensation Vote (PGSV), metrics developed for electric lighting, such as Unified Glare Rating (UGR) and the CIE Glare Index (CGI), the recently proposed metrics emerging from post-occupancy evaluations, such as Unified Glare Probability (UGP), are also evaluated in this study. To investigate how a logistic regression version of DGP would perform in predicting glare in comparison to other linear models, we included "Eccologit" because it is a probability-based model with multiple variables, bound from 0 to 1. It was generated using logistic regression with two input variables, E_v and "log_gc" as in the DGP equation (Equations (1) and (2)) The model coefficients were derived from the original dataset used to develop DGP (DE-DK-Ecco, conducted between 2003 and 2005). From this, we defined "Eccologit" as follows:

$$Eccologit = \frac{e^{a+b\cdot E_v+c\cdot log\_gc}}{1 + e^{a+b\cdot E_v+c\cdot log\_gc}} \quad (3)$$

where the input variables E_v (lux) and log_gc ($cd^2/m^4\cdot$ lux) corresponds to the terms used in the DGP equation (Equation (1)) and the coefficients a, b, and c were derived as follows: $a = -1.506$, $b = 2.685\cdot 10^{-5}$, $c = 3.232\cdot 10^{-1}$.

Table 1
List of discomfort glare metrics evaluated in this paper, classified according to whether they account for saturation, contrast, or both.

Type	No.	Name of metric	Abbreviation	Reference
Saturation and contrast based (Hybrid)	1	Daylight Glare Probability	DGP	[18]
	2	Predicted Glare Sensation Vote	PGSV	[35,37]
	3	Experimental Unified Glare Rating	UGR_{exp}	[38]
	4	Logistic Regression model (using DGP development dataset)	Eccologit	–
Contrast based only	5	CIE Glare Index	CGI	[39,40]
	6	Daylight Glare Index	DGI	[41,42]
	7	Modified Daylight Glare Index	DGI_{mod}	[38]
	8	Predicted Glare Sensation Vote (contrast)	$PGSV_{con}$	[35]
	9	Unified Glare Probability/Unified Glare Rating	UGP/UGR	[14,40]
	10	Logarithmic Contrast (from DGP)	log_gc	[18]
	11	Daylight Glare Rating	DGR	[43]
	12	Visual Comfort Probability	VCP	[43]
Saturation based only	13	Average Luminance of Image	L_{avg}	[44]
	14	Predicted Glare Sensation Vote (saturation)	$PGSV_{sat}$	[35]
	15	Vertical Illuminance/ Simplified DGP	E_v/DGPs	[45]

3.2. Datasets

To analyze the performance of discomfort glare metrics in bright and dim scenes, we applied performance analyses on subgroups based on a compiled dataset from six previously published glare user assessment studies, leading to a total sample size of n = 800. The experiments for these studies had been conducted in controlled daylit office-like environments, carried out in Israel, Japan, Germany, Argentina, and the United States between 2008 and 2016, leading to individual datasets called, respectively, IL-DayVICE, JP-Office, DE-Gaze, AR-DEO, US-Fabric. The sixth dataset, which is abbreviated DE-Quanta, is directly derived from the cross-validation paper [16]. A detailed presentation of the selection procedure for all 6 datasets to ensure that the retained datasets are of high quality and reliable can be found in the supplementary material of that paper. This is briefly discussed further below. The resulting combined dataset is quite diverse, ranging over different climates and cultural backgrounds of subjects. This dataset does not include any data used to develop any of the metrics. The data that was used for the development of DGP is used in this study only to derive the parameters for a new logistic regression model called Eccologit, but not for any performance analysis. Each data point includes a validated HDR image taken from the user's point of view, which is paired with a subjective response on glare perception based on a questionnaire using the four-point scale introduced by Osterhaus and Bailey [46]. In the tests, participants were asked to rate the level of discomfort caused by glare on a scale of "imperceptible," "noticeable," "disturbing," and "intolerable." A mapping of glare scale responses to the four-point scale was necessary for IL-DayVICE and JP-Office, as they used a five-point Likert scale and a linear scale respectively [16]. The HDR images were checked for pixel overflow by checking illuminance calculated from the HDR image to the measured illuminance at the camera lens. Data points with discrepancies of more than 25% were omitted from the dataset. Metrics were computed using evalglare [47,48] and glare sources were detected via the task-driven method [18] with a multiplier of 5 times the task area applied with the task position adjusted for each of the scenes. Fig. 1 illustrates four examples of HDR images with varying levels of vertical illuminance and contrast from the combined dataset.

3.3. Splitting of the dataset into high and low adaptation levels by using E_v

To be able to discriminate between cases in high and low adaptation levels, a suitable criterion must be chosen. As E_v was found to be a good predictor of pupil size due to its high correlation [36], we use it to discriminate between the high and low adaptation levels. A 3000 lux E_v threshold was chosen to divide the dataset into high and low adaptation levels, based on two criteria, the DGP threshold of noticeable glare (0.34) and when there is no contrast in the scene (log_gc = 0). This is then used to test our hypothesis whether contrast-driven metrics describe glare better in lower adaptation levels. The relationship between log_gc and E_v, which are both terms used in the DGP equation, is depicted in scatter plots of the 800 data points of the combined dataset, colored by study and by subjective responses (Fig. 2). The DGP thresholds for "disturbing" glare are marked as a linear relationship between log_gc and E_v, derived from the DGP equation (Equation (1)) assuming that DGP is 0.34 and 0.38. The binary classification into 'glare' and 'no glare' conditions is drawn from subjects' responses: glare is reported if the subjects indicated on the four-point scale that conditions were "disturbing" or "intolerable", and not reported if the response was "noticeable" or "imperceptible". As the data contains only laboratory studies in which subjects were mostly seated close to the façade, with a range of shading devices employed, the resulting vertical illuminance ranges widely from as low as 73 lux to a maximum of 13749 lux. The median and mean of the combined dataset are 2648 and 3012 lux, respectively. The distribution of vertical illuminances per study is shown in Fig. 3. Fig. 4 shows the splitting of the data points by 3000 lux E_v thresholds, where the lower range (lr_3000) has user-evaluated scenes

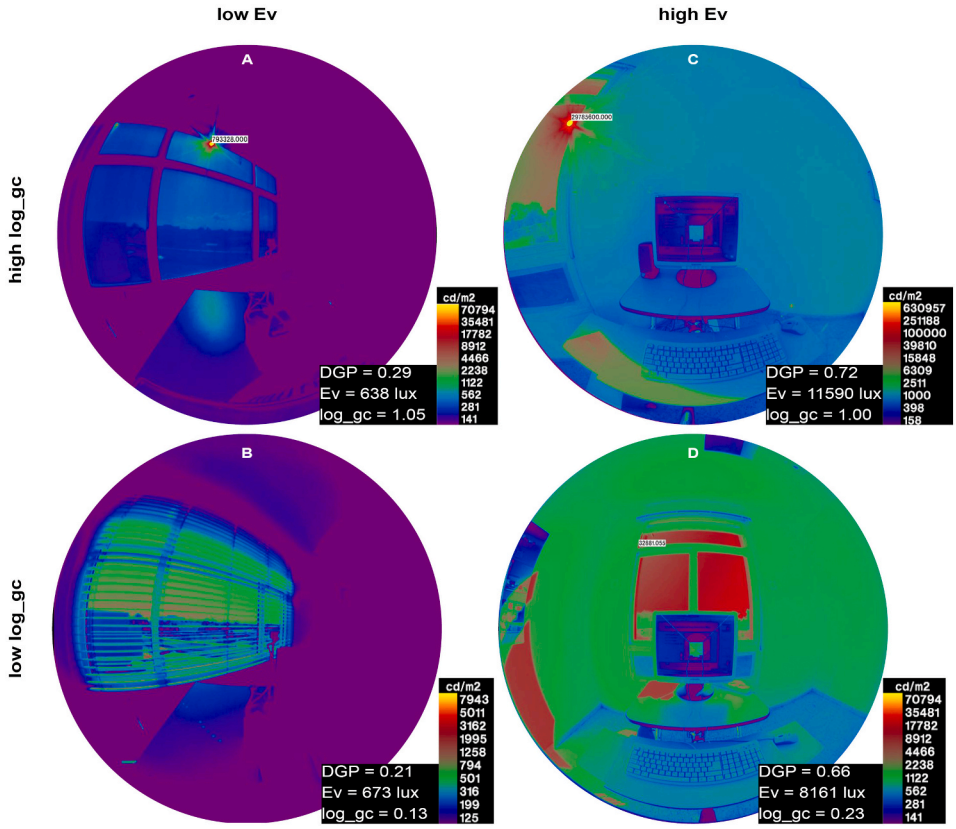

Fig. 1. Four example HDR images with varying levels of vertical illuminance (E_v) and contrast (log_gc), two terms used in the DGP equation to represent saturation and contrast respectively.

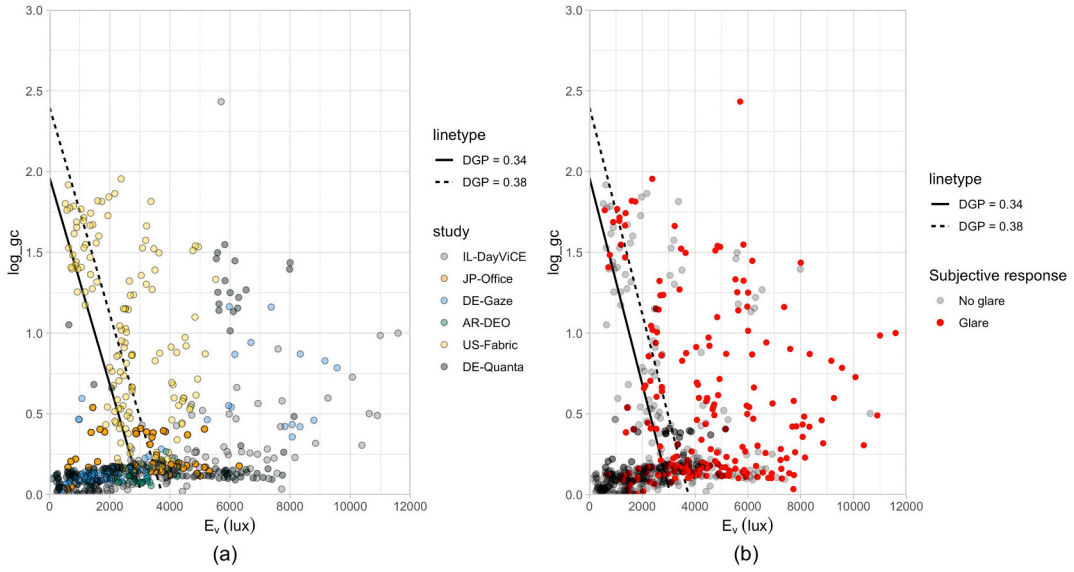

Fig. 2. Scatter plot of the 800 data points (log_gc against Ev), colored by (a) study and (b) subjective glare. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

with $E_v < 3000$ lux and the upper range (ur_3000) has scenes with $E_v \geq 3000$ lux.

3.4. Statistical methods

The study's main aim is to determine which glare metrics better predict perceived discomfort glare for scenes with lower and higher ranges of adaptation levels. The key statistical methods used in this paper were thus chosen based on their ability to determine the predictive power of the 15 discomfort glare metrics, which contributed to the selection of two main statistical tests - Receiver Operating Characteristics (ROC) analysis and Spearman's rank correlation. A Delong's test was conducted on the ROC-curves to evaluate if the difference between the AUC values of the models is significant.

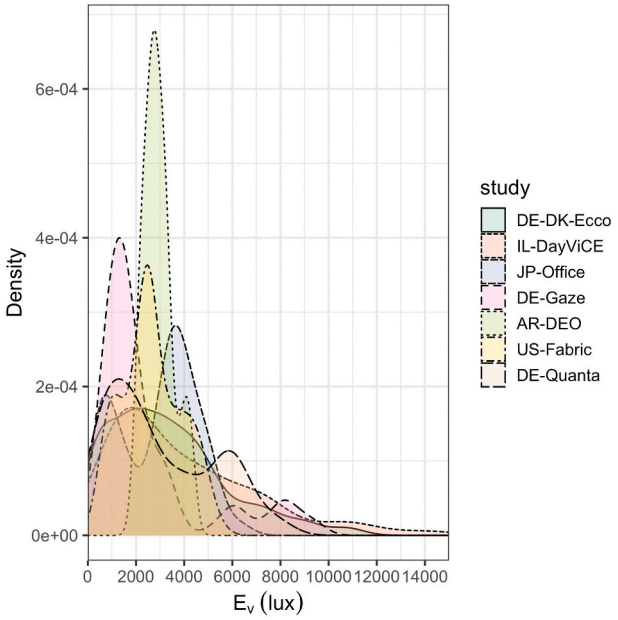

Fig. 3. Probability density plot of vertical illuminances of each of the six laboratory studies.

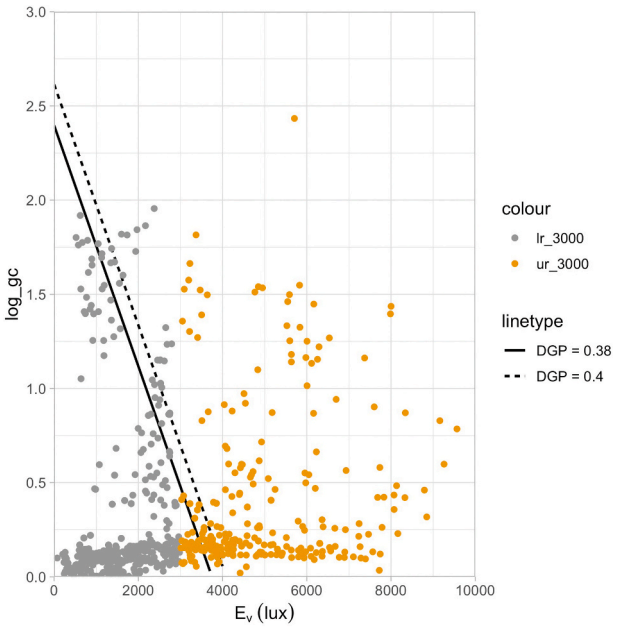

Fig. 4. Splitting of data by a 3000 lux E_v threshold, resulting in lr_3000 (n = 442) and ur_3000 (n = 358).

3.4.1. Receiver operating characteristic (ROC) analysis

ROC analysis is used in various fields, starting from its first usage in warfare when assessing prediction accuracy of radar receiver operators [49], to psychophysics, medical research, epidemiology, and machine learning applications for performance diagnostics [50]. It is a statistical method that is used to assess the diagnostic ability of a binary classifier system. ROC analysis was recently performed to predict glare from daylight [31] and it was also employed extensively in the comprehensive cross-validation paper on daylight glare metrics mentioned above [16]. Here we apply ROC analysis to assess the predictive performance of glare metrics in the low and high illuminance range subsets of our data (lr_3000 and ur_3000). Subjective glare responses on the four-point scale are categorized into binary bins of disturbing and non-disturbing glare as previously described.

A ROC curve is plotted by computing the true positive rates (TPR) and true negative rates (TNR) as the discrimination threshold is varied. True positive rate, also known as sensitivity, is the ratio of the number of true positive cases to the total number of positive predictions by the metric at a certain discrimination threshold (Equation (4)); if the glare metric output is greater than the particular discrimination threshold, glare is predicted. Likewise, the true negative rate, also known as specificity, is the ratio of the number of true negative cases to the total number of negative predictions (Equation (5)). In an ideal scenario where the metric predicts glare perfectly, TPR and TNR would be both maximized with a value of 1. The ROC curve is obtained by computing the TPR and TNR while varying the discrimination thresholds possible for each glare metric. The area under the curve (AUC) summarizes the ROC curve and is commonly used to compare the overall prediction accuracy of the metric in question [51]. Absolute AUC values can be used to compare diagnostic accuracies to some extent, and a value of 0.7 and above is considered as acceptable accuracy, while a value of 0.8 and above is excellent [52]. In this paper, a package named pROC [53] based on R was used to perform the ROC analysis to investigate the full range of discrimination thresholds possible in intervals of up to 0.001. It is crucial not to compare ROC curves across different datasets, as the distribution of responses changes, thus biasing the comparison. Therefore in this paper, we will only compare ROC curves within each of the subgroups.

$$\mathrm{TPR} = \frac{TP}{TP + FN} \tag{4}$$

where TP refers to number of true positives and FN refers to number of false negatives

$$\mathrm{TNR} = \frac{TN}{TN + FP} \tag{5}$$

where TN refers to number of true negatives and FP refers to number of false positives.

While comparing AUCs by their absolute values can be one method to differentiate the prediction performance of two ROC curves resulting from two prediction models, the correlational nature of the ROC curves should also be accounted for. Delong's test [54] between two ROC curves is a non-parametric statistical approach based on the Mann-Whitney U-statistic to analyze the area under correlated ROC curves arising from two ROC analyses on the same dataset. The test shows whether the difference between two correlated ROC curves is statistically significant based on the resulting p-value.

3.4.2. Spearman's rank correlation

Spearman's rank correlation is a non-parametric statistical test that measures the strength of a monotonic relationship between paired variables. Unlike Pearson's correlation, Spearman's rank correlation can be applied to data that may not have a normal distribution. The data can be ordinal but is not required to be of equidistant ordinal nature [55]. Since the glare responses are of categorical, ordinal nature, we analyzed Spearman's rank correlation between the raw responses on the four-point scale and the predicted values from the glare metrics in question. This can tell us how well the glare metrics generally describe the users' glare responses on the four-point scale without categorizing the responses into binary bins. Like the ROC AUCs, results of the Spearman's rank correlation should only be compared within each subgroup. They cannot be used across subgroups because one subgroup might have a wider range of data than the other, which can bias the correlational strength.

4. Results

The combined dataset, which excludes data used to develop any of

the tested discomfort glare metrics, has a total of 800 data points assessed by 309 participants in 6 independent lab experiments. ROC analysis was used to test the predictive performance of 16 existing glare metrics, along with Delong's test to determine if the difference in predictive performance is significant. Spearman's rank correlations were also performed to test for correlational strength between glare metrics and 4-point glare ratings.

4.1. Receiver operating characteristic (ROC) analysis results

The performance of the 15 selected glare metrics (see Table 1) was tested across lr_3000 and ur_3000. Because ROC analysis is specific to the distribution of the data, the area under the ROC curves (AUCs) associated with the glare metrics must be compared within each group and not across any two groups. E_v was omitted in the ROC analysis because a previously proposed metric, DGPs, is already included in the analysis and is a linear function of E_v. Contrast-driven metrics are colored in blue in the bar graph, saturation-driven metrics are colored in orange and hybrid metrics that include both saturation and contrast terms are colored purple.

The predictive performances of the 15 glare metrics was evaluated on the subgroups lr_3000 (n = 442) and ur_3000 (n = 358). Based on the AUCs, the hybrid models DGP and Eccologit emerged as the best-performing metrics for predicting glare. The hybrid and the contrast-driven metrics outperform saturation-driven metrics in scenarios where vertical illuminance is less than 3000 lux (lr_3000), while the hybrid and the saturation-driven metrics significantly outperform contrast-driven metrics in high-illuminance conditions with higher adaptation levels (ur_3000). Despite being a hybrid model, UGR_{exp} had the lowest AUCs in both ranges. Fig. 5 shows the ROC curves and a comparison of AUCs in the lr_3000 and ur_3000 datasets.

To evaluate if a difference between two metrics' AUC values is significant or not, Delong's test is employed Delong's test compares the difference in the ROC curves while accounting for the correlational nature of the curves. The alternative hypothesis of Delong's test is that the difference between the two ROC curves is not equal to 0. The result of this analysis is shown in Table 2 and Table 3, where the light grey colored field shows pairs of significantly different AUC-values. While hybrid metrics such as DGP and Eccologit had the highest AUCs in lr_3000, their AUCs were not significantly different from DGI_{mod}, CGI, UGR/UGP, Log_gc and DGI according to the Delong's test. Similarly, for the upper range data ur_3000, L_{avg}, Eccologit, DGP, PGSV, $PGSV_{sat}$, and

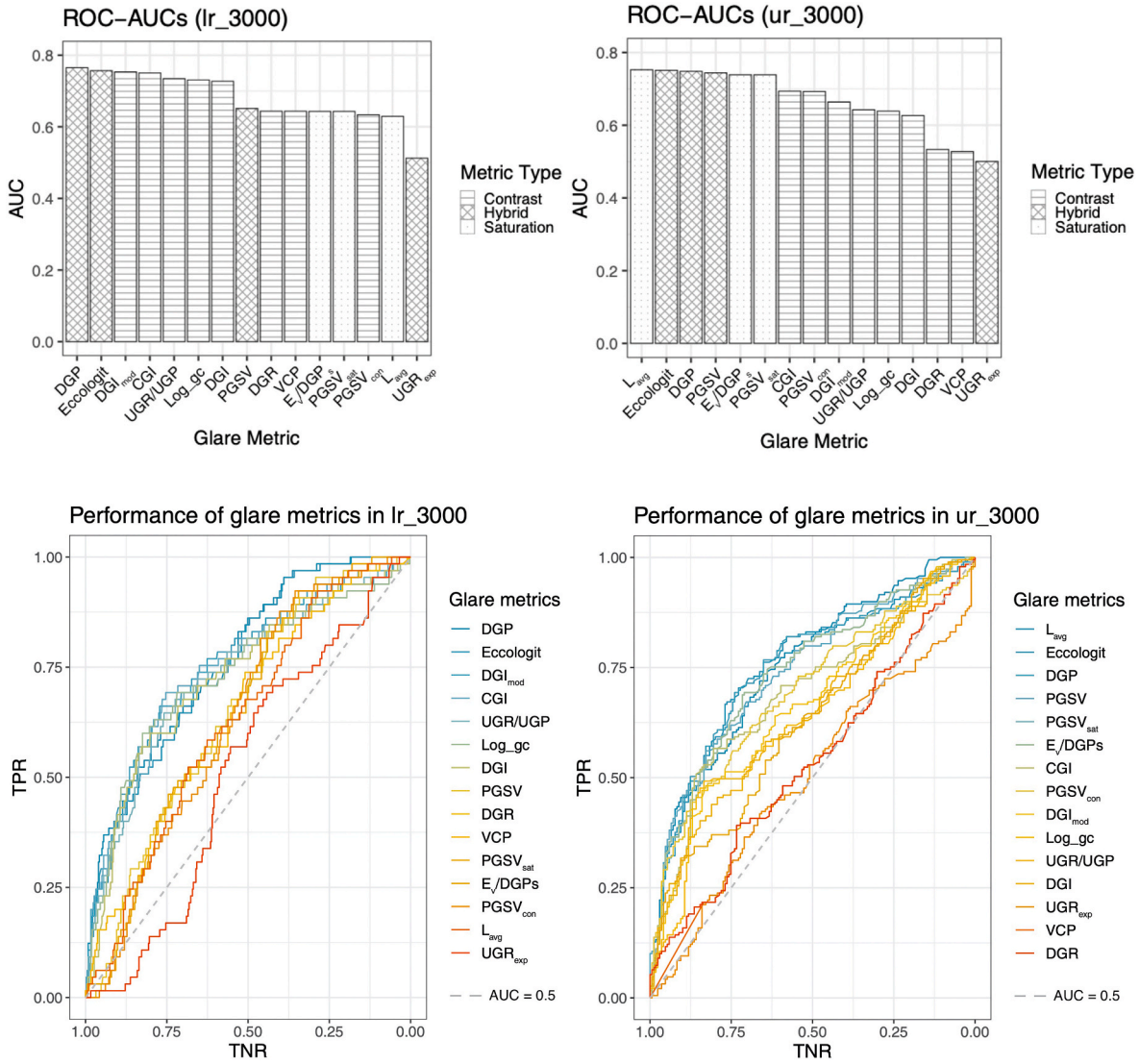

Fig. 5. Comparison of ROC area under curves (AUCs) in a bar-plot (Top) and the ROC curves of evaluated glare metrics for lr_3000 and ur_3000, sorted and color-coded by predictive performance within each group (Bottom). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
p-values of Delong's test comparing the difference between ROC curves of 15 glare metrics in lr_3000. Light grey values are statistically significant at a 0.05 significance level, dark grey values indicate that a bootstrap test was performed but is not statistically significant. White values are not statistically significant. In cases where the ROC curves have different directions and Delong's method cannot be used, the bootstrap test is used instead. This occurred only for the inverted direction of the ROC curve of UGR_{exp}, with an AUC value close to 0.5 implying very poor prediction performance (and that a random function would deliver better results).

	DGP	Eccologit	DGI_{mod}	CGI	UGR/UGP	Log_gc	DGI	PGSV	DGR	VCP	$PGSV_{sat}$	E_v/DGPs	$PGSV_{con}$	L_{avg}	UGR_{exp}
DGP	-	7.94E-02	5.36E-01	3.57E-01	1.17E-01	1.38E-01	7.83E-02	3.34E-04	7.37E-07	7.23E-07	1.97E-04	1.97E-04	1.84E-05	6.41E-04	4.63E-08
Eccologit	-	-	8.70E-01	7.31E-01	3.14E-01	3.13E-01	2.06E-01	2.43E-04	2.50E-05	2.47E-05	9.25E-05	9.25E-05	1.34E-05	5.00E-04	4.55E-08
DGI_{mod}	-	-	-	7.47E-01	9.86E-02	4.33E-02	6.63E-03	9.99E-03	8.08E-06	7.99E-06	8.83E-03	8.83E-03	1.24E-03	1.33E-02	4.76E-06
CGI	-	-	-	-	2.72E-03	6.82E-02	9.78E-03	9.89E-03	9.72E-08	9.56E-08	1.03E-02	1.03E-02	1.17E-03	1.52E-02	2.40E-06
UGR/UGP	-	-	-	-	-	7.32E-01	3.07E-01	4.09E-02	1.67E-06	1.65E-06	3.96E-02	3.96E-02	7.35E-03	4.59E-02	3.92E-05
Log_gc	-	-	-	-	-	-	7.72E-01	6.72E-02	3.02E-04	3.00E-04	6.49E-02	6.49E-02	1.75E-02	6.54E-02	4.31E-05
DGI	-	-	-	-	-	-	-	6.36E-02	1.01E-04	1.00E-04	5.48E-02	5.48E-02	1.35E-02	6.27E-02	1.01E-04
PGSV	-	-	-	-	-	-	-	-	8.62E-01	8.61E-01	7.06E-01	7.06E-01	1.55E-01	3.63E-01	1.06E-02
DGR	-	-	-	-	-	-	-	-	-	6.17E-01	9.92E-01	9.92E-01	7.97E-01	7.86E-01	1.43E-02
VCP	-	-	-	-	-	-	-	-	-	-	9.93E-01	9.93E-01	7.99E-01	7.87E-01	1.39E-02
$PGSV_{sat}$	-	-	-	-	-	-	-	-	-	-	-	1.00E+00	6.87E-01	3.42E-01	1.48E-02
E_v/DGPs	-	-	-	-	-	-	-	-	-	-	-	-	6.87E-01	3.42E-01	1.46E-02
$PGSV_{con}$	-	-	-	-	-	-	-	-	-	-	-	-	-	8.75E-01	2.54E-02
L_{avg}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3.06E-02
UGR_{exp}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 3
p-values of Delong's test comparing the difference between ROC curves of 15 glare metrics in ur_3000.

	L_{avg}	Eccologit	DGP	PGSV	$PGSV_{sat}$	E_v/DGPs	CGI	$PGSV_{con}$	DGI_{mod}	UGR/UGP	Log_gc	DGI	DGR	VCP	UGR_{exp}
L_{avg}	-	9.11E-01	7.91E-01	6.08E-01	2.91E-01	2.91E-01	2.81E-02	1.90E-02	2.14E-03	7.30E-04	1.09E-03	1.89E-04	1.73E-10	3.23E-10	8.37E-09
Eccologit	-	-	1.98E-01	6.34E-01	1.78E-01	1.78E-01	1.33E-03	8.98E-03	1.35E-04	8.72E-05	2.60E-05	3.30E-06	2.56E-12	4.92E-12	2.33E-08
DGP	-	-	-	7.81E-01	3.71E-01	3.71E-01	9.98E-04	1.41E-02	9.96E-05	6.76E-05	2.02E-05	2.63E-06	2.98E-12	5.80E-12	4.11E-08
PGSV	-	-	-	-	6.73E-01	6.73E-01	4.79E-02	2.90E-03	7.17E-03	2.22E-03	7.48E-04	1.62E-04	2.07E-09	3.68E-09	1.75E-07
$PGSV_{sat}$	-	-	-	-	-	1.00E+00	5.80E-02	3.51E-02	7.67E-03	2.86E-03	1.50E-03	2.72E-04	4.23E-10	7.87E-10	5.19E-07
E_v/DGPs	-	-	-	-	-	-	5.80E-02	3.51E-02	7.67E-03	2.86E-03	1.50E-03	2.72E-04	4.23E-10	7.87E-10	3.89E-07
CGI	-	-	-	-	-	-	-	9.75E-01	3.48E-03	2.01E-04	3.14E-05	2.17E-06	6.80E-08	1.24E-07	4.38E-05
$PGSV_{con}$	-	-	-	-	-	-	-	-	3.82E-01	1.33E-01	9.29E-02	4.04E-02	1.24E-05	1.99E-05	5.98E-05
DGI_{mod}	-	-	-	-	-	-	-	-	-	2.85E-02	1.76E-01	4.29E-03	8.99E-06	1.62E-05	3.32E-04
UGR/UGP	-	-	-	-	-	-	-	-	-	-	8.04E-01	4.57E-01	4.59E-04	7.60E-04	1.97E-03
Log_gc	-	-	-	-	-	-	-	-	-	-	-	1.03E-01	3.39E-04	5.40E-04	4.11E-03
DGI	-	-	-	-	-	-	-	-	-	-	-	-	1.90E-03	2.92E-03	1.51E-02
DGR	-	-	-	-	-	-	-	-	-	-	-	-	-	2.88E-02	5.70E-01
VCP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5.01E-01
UGR_{exp}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Ev/DGP_s had the highest AUC values and Delong's test revealed no significant differences in their AUC values. In lr_3000, although DGI_{mod} had the highest AUC out of the contrast-driven metrics, its AUC value was not significantly different from CGI and UGR/UGP. CGI generally outperforms all contrast-driven metrics in both lr_3000 and ur_3000 in predicting; its higher AUC value was significantly different to all other contrast-dominated metrics except for $PGSV_{con}$.

4.2. Spearman's rank correlation results

Although Spearman's rank correlation is an entirely independent test from the above analysis, it shows similar tendencies to the ROC analysis when applied to glare metrics and the 4-point glare ratings from participants. Compared to a Bonferroni-adjusted significance level for 15 comparisons ($3.33{\cdot}10^{-3}$), the p-values of the correlations were statistically significant for most metrics except DGR, VCP, and UGR_{exp} in ur_3000 and UGR_{exp} in lr_3000 (A full table of p-values and rho can be found in the appendix). The rho value tells us the strength of the correlation between the metric values and the glare rating on Osterhaus's 4-point scale [46]. The results for the single-effect metrics show that the contrast-driven models correlate stronger to the 4-point glare ratings in dim scenes (lr_3000) than in the bright scenes (ur_3000), and the opposite is true for saturation-driven metrics (Fig. 6). This further confirms the finding that contrast-driven metrics correlate better with glare

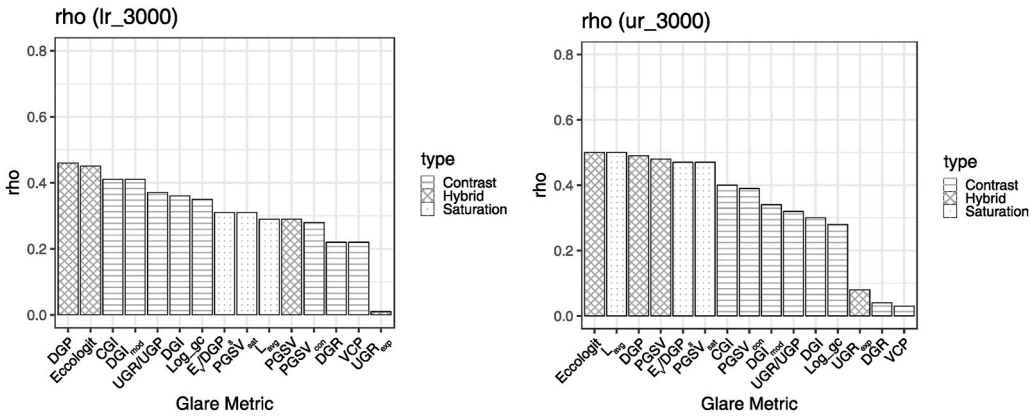

Fig. 6. Spearman's rank correlation rho in lr_3000 and ur_3000 between glare metrics and subjective glare responses on a 4-point scale.

ratings in lower light scenes ($E_v < 3000$ lux) than in brighter scenes ($E_v \geq 3000$ lux). The two main terms in the DGP equation, E_v and Log_gc, which represent saturation and contrast effects respectively, also show a vast difference in correlation with the glare ratings. Log_gc correlates with glare ratings in the dimmer conditions while E_v has a higher correlation with glare ratings in brighter scenes. The hybrid models DGP and Eccologit show a high correlation in both regions and therefore outperform the other metrics based on only one effect.

5. Discussion

The combined dataset assembled from six separate glare studies allowed us to identify the difference in predictive performance between the three categories of glare metrics (contrast based, saturation based and hybrid) in high and low adaptation levels (bright and dim) scenarios. Confirming our hypothesis, contrast-driven glare metrics perform better in dim scenarios with lower adaptation levels, while saturation-driven glare metrics perform better in bright scenarios with higher adaptation levels. This also confirms that the contrast effect is the driver for glare in dimmer scenes (i.e. with low vertical illuminances). Of the contrast-driven metrics, CGI performs best in both dim and bright scenes. This could be due to the use of the sum of direct vertical illuminance, E_d, and indirect vertical illuminance, E_i, which is equivalent to E_v, as the adaptation level term in the CGI equation. The analysis also shows that the hybrid DGP and Eccologit metrics perform well in both categories of adaptation levels for the investigated scenarios and outperform the single effect metrics.

It should be noted that in dimmer scenarios, the adaptation level is lower and the contrast effect causes glare more than the saturation effect, so the contrast effect should not be overlooked. There are currently a limited number of data points in the overall dataset with log_gc values ranging 0.25 to 1.25 (Fig. 2). These points are most likely to represent scenes in open-plan offices where there is a medium to large-sized glare source, such that glare may be better explained by the contrast effect as it is more prevalent annually [29]. However, existing hybrid models like DGP are found in recent studies to be not sensitive enough [14]. Such scenarios should be investigated in depth and models adjusted to improve the hybrid glare prediction models' applicability to a wider range of scenarios. Despite this, it should be noted that in dim scenarios with very high contrast (e.g. the use of electrochromic glazing or dark fabrics where the sun disk is visible, with log_gc values approximately between 1.3 and 2.6), DGP was found to be sensitive enough and performed reasonably well in predicting the user's perception of glare across the scenes [56].

Therefore, there is room for improvement for hybrid metrics that already account for both effects, such as the DGP. Although it outperforms all available thresholds in the scenarios investigated here, with a ROC-AUC of around 0.75 for both categories of high and low adaptation levels, it is necessary to investigate such scenarios in future research where development data is lacking and where previous studies indicated a lack of sensitivity ($E_v < 3000$ and log gc values between 0.25 and 1.25). That would improve the glare prediction in dimmer scenarios and help identify the recommended thresholds. In future research, instead of changing DGP thresholds to account for this shortcoming of the metric, it may be better to change the weight of the contrast term or to add a second contrast term in the DGP equation. To ascertain this, additional user-evaluation data in the above-mentioned range is necessary and is an ongoing effort [57].

One of the main technical takeaways from this study is that for any fair comparisons of performance diagnostics of glare metrics, the two groups that are studied should always have the same range of glare stimuli, and absolute statistics should not be compared between groups with different stimuli. When comparing the predictive performance of glare metrics in laboratory studies versus POE studies, for example, differences in the resulting performance of glare metrics may be attributed to the difference in the range of glare stimuli; POE studies in active workspaces tend to include systematically dimmer scenarios with lower glare stimuli than typical laboratory studies. Furthermore, a fair and systematic overall comparison of daylight discomfort glare metrics across studies also requires consistency in measurement techniques, temporal factors, lighting conditions, and space-occupant configuration, which is difficult to achieve for practical reasons. It is perhaps desirable to standardize for a minimum exposure duration before asking participants to evaluate glare, agree upon reliable questionnaire items and define gatekeeping criteria for consolidating datasets [58,59].

The experimental logistic regression model, Eccologit, derived from the developing dataset of DGP, is that it is a true mathematical probability model bound from 0 to 1. Its performance is similar to DGP as well, which may mean that logistic regression models are suitable candidates for formulating future glare metrics. Note that Eccologit is distinct from the DGP low light correction function implemented in evalglare [60]. The low light correction in evalglare was meant to mathematically extend DGP values below 0.16 using logistic regression when $E_v < 300$ lux but there were no datapoints present for those conditions in the dataset and is thus not a validated correction. On the other hand, Eccologit is a logistic regression version of the DGP model, derived using the same variables and the original dataset of DGP, which is validated for DGP >0.2.

Overall, the results allow us to extend our understanding of the performance of existing glare metrics in dim and bright adaptation levels, in particular that the selection of glare metrics should be carefully considered to avoid oversimplification. Simplified glare metrics that rely too heavily on the saturation effect via an illuminance-based proxy term (e.g., simplified DGPs or simply using E_v) were initially proposed to reduce computation time [45]. As computer processing power and luminous simulation efficiencies improve [61], researchers and

practitioners will be able to consider contrast effects of glare alongside saturation effects via faster simulations of hourly luminance maps, while also opening up new possibilities for annual glare metrics. For example, a recent study that simulated annual saturation and contrast glare across an open-plan office space using a more efficient simulation method showed that only 6% of view directions near the perimeter of the façade frequently experiences saturation glare, while around 38% of view directions frequently experiences contrast glare [29].

6. Conclusion

Glare prediction metrics have the potential to assist building designers and managers in achieving mostly glare-free and well-lit working environments, as well as in making informed decisions about different shading products. In this paper, we found that although hybrid glare metrics outperform single-effect metrics overall (with ROC-AUCs around 0.75), contrast-driven metrics outperform saturation-drive metrics in dim environments with lower adaptation levels. On the other hand, saturation-driven metrics outperform contrast-driven metrics in bright environments with higher adaptation levels. These results make it clear that one should be aware of the dangers of relying solely on over-simplified illuminance-based metrics for glare prediction. "Eccologit", a logistic regression model of the terms of the DGP equation, but which is mathematically bound between 0 and 1 (in contrast to the original DGP equation) was tested and showed comparable performance to DGP. There is also room for improvement in the performance of other established hybrid metrics, which account for both saturation and contrast effects. As of now, the applicability of hybrid metrics like DGP is still limited by the scope of their development dataset, which may not represent the entire spectrum of scenarios encountered in real-world situations. More research is needed, in particular, in dim environments with low adaptation levels and uncomfortable contrast, such as those found in large open-plan offices where the visible sky causes glare. Future research should bridge this gap, and glare metrics should be more inclusive of illumination levels found in various real-life scenarios. Moving forward, striving for a better understanding of discomfort glare can hopefully improve the accuracy and flexibility of glare prediction metrics used in early design simulations, with the ultimate goal of ensuring higher levels of visual comfort in indoor work environments.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Swiss National Science Foundation (SNSF) as part of the ongoing research projects, "Visual comfort without borders: Interactions on discomfort glare" (SNSF #182151). Geraldine Quek is a recipient of the Graduate Merit Scholarship from the Singapore University of Technology and Design (SUTD). We thank Toshie Iwata, RG Rodriguez, and JA Yamin Garreton for contributing their experiment datasets to this research. We also thank Marc-Olivier and Jonathan Koh for their advice in statistics.

Appendix

Spearman's rank correlation (additional results):
a. p-values and rho of Spearman's rank correlations between glare metrics and glare ratings on the Osterhaus' 4-point scale (lr_3000 and ur_3000).

DATASET	lr3000	ur3000	lr_3000	ur_3000
	p-value	*p-value*	*rho*	*rho*
DGP	***< 2.2e-16***	***< 2.2e-16***	0.46	0.49
Eccologit	***< 2.2e-16***	***< 2.2e-16***	0.45	0.50
CGI	***< 2.2e-16***	***2.65E-15***	0.41	0.40
DGI_{mod}	***< 2.2e-16***	***2.55E-11***	0.41	0.34
UGR/UGP	***5.29E-16***	***1.08E-09***	0.37	0.32
DGI	***4.66E-15***	***7.71E-09***	0.36	0.30
Log_gc	***2.10E-14***	***4.67E-08***	0.35	0.28
E_v/DGPs	***1.62E-11***	***< 2.2e-16***	0.31	0.47
$PGSV_{sat}$	***1.62E-11***	***< 2.2e-16***	0.31	0.47
L_{avg}	***5.42E-10***	***< 2.2e-16***	0.29	0.50
PGSV	***5.49E-10***	***< 2.2e-16***	0.29	0.48
$PGSV_{con}$	***1.18E-09***	***3.87E-14***	0.28	0.39
DGR	***1.86E-06***	0.4595	0.22	*0.04*
VCP	***1.92E-06***	0.5979	0.22	*0.03*
UGR_{exp}	9.05E-01	0.1347	*0.01*	*0.08*

References

[1] S. Abbaszadeh, L. Zagreus, D. Lehrer, C. Huizenga, Occupant satisfaction with indoor environmental quality in green buildings. Proc. Healthy Build., Portugal, Lisbon, 2006, pp. 365–370, 2006, https://escholarship.org/uc/item/9rf7p4bs. (Accessed 12 February 2019). accessed.

[2] J. Brand, C. Donnelly, J. Veitch, M. Aries, K. Charles, Linking indoor environment conditions to job satisfaction: a field study AU - Newsham, Guy, Build. Res. Inf. 37 (2009) 129–147, https://doi.org/10.1080/09613210802710298.

[3] P. Leather, M. Pyrgas, D. Beale, C. Lawrence, Windows in the workplace: sunlight, view, and occupational stress, Environ. Behav. 30 (1998) 739–762, https://doi.org/10.1177/001391659803000601.

[4] M.L. Amundadottir, S.F. Rockcastle, M. Sarey Khanie, M. Andersen, A human-centric approach to assess daylight in buildings for non-visual health potential, visual interest and gaze behavior, Build. Environ. 113 (2017) 5–21, https://doi.org/10.1016/j.buildenv.2016.09.033.

[5] M. Inanici, M. Brennan, E. Clark, Spectral daylighting simulations: computing circadian light, Hyderabad, India, 2015. Proc. BS2015 14th Conf. Int. Build. Perform. Simul. Assoc.

[6] M. Rea, M. Figueiro, Light as a circadian stimulus for architectural lighting, Light. Res. Technol. 50 (2018) 497–510, https://doi.org/10.1177/1477153516682368.

[7] M. Knoop, O. Stefani, B. Bueno, B. Matusiak, R. Hobday, A. Wirz-Justice, K. Martiny, T. Kantermann, M. Aarts, N. Zemmouri, S. Appelt, B. Norton, Daylight: what makes the difference? Light. Res. Technol. 52 (2020) 423–442, https://doi.org/10.1177/1477153519869758.

[8] M. Aries, M. Aarts, J. van Hoof, Daylight and health: a review of the evidence and consequences for the built environment, Light. Res. Technol. 47 (2015) 6–27, https://doi.org/10.1177/1477153513509258.

[9] P. Petherbridge, R.G. Hopkinson, Discomfort glare and the lighting of buildings, Trans. Illum. Eng. Soc. 15 (1950) 39–79, https://doi.org/10.1177/147715355001500201.

[10] J. Jakubiec, C. Reinhart, Assessing disability glare potential of reflections from new construction, Transp. Res. Rec. J. Transp. Res. Board. 2449 (2014) 114–122, https://doi.org/10.3141/2449-13.

[11] T.J.T.P. Van Den Berg, On the relation between glare and straylight, Doc. Ophthalmol. 78 (1991) 177–181, https://doi.org/10.1007/BF00165678.

[12] Discomfort glare | eilv (n.d.), http://eilv.cie.co.at/term/333. (Accessed 28 July 2020). accessed.

[13] K.V.D. Wymelenberg, M. Inanici, Evaluating a new suite of luminance-based design metrics for predicting human visual comfort in offices with daylight, Leukos 12 (2016) 113–138, https://doi.org/10.1080/15502724.2015.1062392.

[14] M.B. Hirning, G.L. Isoardi, I. Cowling, Discomfort glare in open plan green buildings, Energy Build. 70 (2014) 427–440, https://doi.org/10.1016/j.enbuild.2013.11.053.

[15] M.B. Hirning, G.L. Isoardi, V.R. Garcia-Hansen, Prediction of discomfort glare from windows under tropical skies, Build, Environ. Times 113 (2017) 107–120, https://doi.org/10.1016/j.buildenv.2016.08.005.

[16] J. Wienold, T. Iwata, M. Sarey Khanie, E. Erell, E. Kaftan, R. Rodriguez, J. Yamin Garreton, T. Tzempelikos, I. Konstantzos, J. Christoffersen, T. Kuhn, C. Pierson, M. Andersen, Cross-validation and robustness of daylight glare metrics, Light. Res. Technol. 51 (2019) 983–1013, https://doi.org/10.1177/1477153519826003.

[17] C. Pierson, J. Wienold, M. Bodart, Review of factors influencing discomfort glare perception from daylight, Leukos (2018) 1–38, https://doi.org/10.1080/15502724.2018.1428617.

[18] J. Wienold, J. Christoffersen, Evaluation methods and development of a new glare prediction model for daylight environments with the use of CCD cameras, Energy Build. 38 (2006) 743–757, https://doi.org/10.1016/j.enbuild.2006.03.017.

[19] E. Standards, DIN EN 17037, Httpswwwen-Stand. (n.d.), https://www.en-standard.eu/din-en-17037-daylight-in-buildings/. (Accessed 3 August 2020). accessed.

[20] M. Inanici, Evaluation of high dynamic range photography as a luminance data acquisition system, Light. Res. Technol. 38 (2006) 123–134, https://doi.org/10.1191/1365782806li164oa.

[21] J.A. Jakubiec, K.G. Van Den Wymelenberg, M.N. Inanici, A. Mahic, Accurate measurement of daylit interior scenes using high dynamic range photography. Proc. CIE Light. Qual., Energy Effic., 2016.

[22] C. Pierson, C. Cauwerts, M. Bodart, J. Wienold, Tutorial: luminance maps for daylighting studies from high dynamic range photography, Leukos (2020) 1–30, https://doi.org/10.1080/15502724.2019.1684319.

[23] G. Quek, J. Jakubiec, Calibration and Validation of Climate-Based Daylighting Models Based on One-Time Field Measurements: Office Buildings in the Tropics, LEUKOS, 2019, https://doi.org/10.1080/15502724.2019.1570852.

[24] J. Jakubiec, G. Quek, T. Srisamranrungruang, Long-term visual quality evaluations correlate with climate-based daylighting metrics in tropical offices – a field study, Light. Res. Technol. (2020), https://doi.org/10.1177/1477153520926528, 1477153520926528.

[25] C. Pierson, M. Sarey Khanie, M. Bodart, J. Wienold, Discomfort glare cut-off values from field and laboratory studies. Proc. 29th Quadrenn. Sess. CIE, 2019, pp. 295–305, https://doi.org/10.25039/x46.2019.OP41. International Commission on Illumination, CIE, Washington DC, USA.

[26] K. Konis, Predicting visual comfort in side-lit open-plan core zones: results of a field study pairing high dynamic range images with subjective responses, Energy Build. 77 (2014) 67–79, https://doi.org/10.1016/j.enbuild.2014.03.035.

[27] R.A. Mangkuto, K.A. Kurnia, D.N. Azizah, R.T. Atmodipoero, F.X.N. Soelami, Determination of discomfort glare criteria for daylit space in Indonesia, Sol. Energy 149 (2017) 151–163, https://doi.org/10.1016/j.solener.2017.04.010.

[28] I. Konstantzos, A. Tzempelikos, Daylight glare evaluation with the sun in the field of view through window shades, Build. Environ. 113 (2017) 65–77, https://doi.org/10.1016/j.buildenv.2016.09.009.

[29] G. Quek, S. Wasilewski, J. Wienold, M. Andersen, Spatial evaluation of potential saturation and contrast effects of discomfort glare in an open-plan office, Bruges, Belgium, 2021, p. BS2021.

[30] I. Konstantzos, A. Tzempelikos, Y.-C. Chan, Experimental and simulation analysis of daylight glare probability in offices with dynamic window shades, Build. Environ. 87 (2015) 244–254, https://doi.org/10.1016/j.buildenv.2015.02.007.

[31] R.G. Rodriguez, J.A. Yamín Garretón, A.E. Pattini, An epidemiological approach to daylight discomfort glare, Build. Environ. 113 (2017) 39–48, https://doi.org/10.1016/j.buildenv.2016.09.028.

[32] C. Moosmann, J. Wienold, A. Wagner, V. Wittwer, Ermittlung relevanter Einflussgrößen auf die subjektive Bewertung von Tageslicht zur Bewertung des visuellen Komforts in Büroräumen, Abschlussbericht, 2012, https://doi.org/10.5445/IR/1000034968.

[33] M. Sarey Khanie, J. Stoll, W. Einhäuser, J. Wienold, M. Andersen, Gaze and discomfort glare, Part 1: development of a gaze-driven photometry, Light. Res. Technol. 49 (2017) 845–865, https://doi.org/10.1177/1477153516649016.

[34] E. Erell, E. Kaftan, Y. Garb, Daylighting for visual comfort and energy conservation in offices in sunny regions. Proc. 30th PLEA Int. Conf. – Sustain. Habitat, Dev. Soc., Ahmedabad, 2014.

[35] T. Iwata, W. Osterhaus, Assessment of Discomfort Glare in Daylit Offices Using Luminance Distribution Images, 2010, pp. 174–179. Vienna, Austria.

[36] A. Bay, E.L. Nielsen, M. Fontoynont, A. Lumbye, W. Osterhaus, L. Kessler, Copenhagen september, 2017, p. 118, 2017.

[37] M. Tokura, T. Iwata, M. Shukuya, Experimental study on discomfort glare caused by windows, part 3: development of a method for evaluating discomfort glare from a large light source, J. Archit. Plan. Environ. Eng. (1996) 17–25.

[38] K. Fisekis, M. Davies, M. Kolokotroni, P. Langford, Prediction of discomfort glare from windows, Light. Res. Technol. 35 (2003) 360–369, https://doi.org/10.1191/1365782803li095oa.

[39] H.D. Einhorn, Discomfort glare: a formula to bridge differences, Light. Res. Technol. 11 (1979) 90–94, https://doi.org/10.1177/14771535790110020401.

[40] CIE, Discomfort Glare in Interior Lighting, Commission Internationale de l'Eclairage, Vienna, Austria, 1995.

[41] R.G. Hopkinson, Glare from daylighting in buildings, Appl. Ergon. 3 (1972) 206–215, https://doi.org/10.1016/0003-6870(72)90102-0.

[42] P. Chauvel, J.B. Collins, R. Dogniaux, J. Longmore, Glare from windows: current views of the problem, Light. Res. Technol. 14 (1982) 31–46, https://doi.org/10.1177/096032718201400103.

[43] M.S. Rea, The IESNA Lighting Handbook: Reference & Application, 2000.

[44] K. Van Den Wymelenberg, M. Inanici, A critical investigation of common lighting design metrics for predicting human visual comfort in offices with daylight, Leukos 10 (2014) 145–164, https://doi.org/10.1080/15502724.2014.881720.

[45] J. Wienold, Dynamic daylight glare evaluation, Proc. Build. Simul. (2009) 944–951.

[46] W.K.E. Osterhaus, I.L. Bailey, Large area glare sources and their effect on visual discomfort and visual performance at computer workstations, Conf. Rec. 1992 IEEE Ind. Appl. Soc. Annu. Meet. 2 (1992) 1825–1829, https://doi.org/10.1109/IAS.1992.244537.

[47] J. Wienold, Evalglare–A new RADIANCE-based tool to evaluate daylight glare in office spaces, in: 3rd Int. RADIANCE Workshop, 2004, 2004.

[48] J. Wienold, M. Andersen, Evalglare 2.0: new features, faster and more robust HDR-image evaluation. 3rd Int. Radiance Workshop, 2016. Padova, Italy, https://www.radiance-online.org/community/workshops/2016-padua/presentations/211-Wienold-Evalgaare2.0.pdf.

[49] J.P. Egan, J.P. Egan, Signal Detection Theory and ROC-Analysis, Academic press, 1975.

[50] K.A. Spackman, Signal detection theory: valuable tools for evaluating inductive learning, in: A.M. Segre (Ed.), Proc. Sixth Int. Workshop Mach. Learn., Morgan Kaufmann, San Francisco (CA), 1989, pp. 160–163, https://doi.org/10.1016/B978-1-55860-036-2.50047-3.

[51] J.A. Swets, R.M. Pickett, Evaluation of Diagnostic Systems: Methods from Signal Detection Theory, Academic Press, New York, NY, 1982 n.d.

[52] D.W. Hosmer, S. Lemeshow, Area under the ROC curve, Appl. Logist. Regres. 160 (2000) 164.

[53] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, M. Müller, pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinf. 12 (2011) 77, https://doi.org/10.1186/1471-2105-12-77.

[54] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (1988) 837–845, https://doi.org/10.2307/2531595.

[55] S. Siegel, Nonparametric Statistics for the Behavioral Sciences, 1956.

[56] Sneha Jain, Caroline Karmann, Jan Wienold, Behind electrochromic glazing: Assessing user's perception of glare from the sun in a controlled environment, Energy and Building, 2021. Submitted for publication.

[57] G. Quek, J. Wienold, M. Andersen, User evaluations of contrast-dominant discomfort glare in dim daylit scenarios: preliminary findings. CIE 2021 Midterm Meet. Conf., Kuala Lumpur, Malaysia, 2021.

[58] S. Fotios, Correspondence: new methods for the evaluation of discomfort glare, Light. Res. Technol. 50 (2018) 489–491, https://doi.org/10.1177/1477153518773577.

[59] J.A. Veitch, S.A. Fotios, K.W. Houser, Judging the scientific quality of applied lighting research, Leukos 15 (2019) 97–114, https://doi.org/10.1080/15502724.2018.1550365.

[60] J. Wienold, New Features of Evalglare, 2012.

[61] Raytraverse v1.0.4 S. Wasilewski, Zenodo, https://doi.org/10.5281/zenodo.4278918, 2020.