

The Society of
Light and Lighting

Cross-validation and robustness of daylight glare metrics

J Wienold Dr-Ing^a , **T Iwata** Dr Eng^b, **M Sarey Khanie** PhD^c, **E Erell** PhD^d , **E Kaftan** PhD^d,
RG Rodriguez PhD^e, **JA Yamin Garreton** PhD^e, **T Tzempelikos** PhD^f, **I Konstantzos** PhD^g,
J Christoffersen PhD^h, **TE Kuhn** Drⁱ, **C Pierson** MScEng^j  and **M Andersen** PhD^a

^aLIPID, École Polytechnique Fédérale de Lausanne, Switzerland

^bDepartment of Architecture and Building Engineering, Tokai University, Japan

^cICIEE, Technical University of Denmark, Kongens Lyngby, Denmark

^dBIDR, Ben-Gurion University of the Negev, Sde Boqer Campus, Israel

^eINAHE, CONICET, Mendoza, Argentina

^fSchool of Civil Engineering, Purdue University, West Lafayette, Indiana, USA

^gThe Durham School, University of Nebraska, Lincoln, Nebraska, USA

^hVELUX A/S, Hoersholm, Denmark

ⁱFraunhofer Institute for Solar Energy Systems ISE, Freiburg, Germany

^jUniversité Catholique de Louvain, Louvain-la Neuve, Belgium

Received 11 August 2018; Revised 4 December 2018; Accepted 14 December 2018

This study evaluates the performance and robustness of 22 established and newly proposed glare prediction metrics. Experimental datasets of daylight-dominated workplaces in office-like test rooms were collected from studies by seven research groups in six different locations (Argentina, Denmark, Germany, Israel, Japan and the United States). The variability in experimental setups, locations and research teams allowed reliable evaluation of the performance and robustness of glare metrics for daylight-dominated workplaces. Independent statistical methods were applied to individual datasets and also to one combined dataset to evaluate the performance and robustness of the 22 glare metrics. As performance and robustness are not established in literature, we defined performance as: (1) the ability of the metric value to describe the glare scale (evaluated by Spearman rank correlation), and (2) the ability of the metric to distinguish between disturbing and non-disturbing situations (evaluated by diagnostic receiver operating characteristic curve analysis tests). Furthermore, we defined robustness as the ability of a metric to deliver meaningful results when applied to different datasets and to fail as few as possible statistical tests. Average Spearman rank correlation coefficients in the range of 0.55–0.60 as well as average prediction rates to distinguish between disturbing and non-disturbing glare of 70–75% for several of the metrics indicate their reliability. The results also show that metrics considering the saturation effect as a main input in their equation perform better and are more robust in daylight-dominated workplaces than purely contrast-based metrics or purely empirical metrics. In this study, the daylight glare probability (DGP) delivered the highest performance amongst the tested metrics and was also found to be the most robust. Future research should aim to optimise the terms of glare equations which combine contrast and saturation effects, such as DGP, PGSV or UGR_{exp} , to achieve metrics that also perform reliably in dimmer lighting conditions than the ones explored in this study.

Address for correspondence: J Wienold, Laboratory of Integrated Performance in Design LIPID, ENAC, EPFL, LE 1 111, Station 18, CH-1015, Lausanne, Switzerland.
E-mail: jan.wienold@epfl.ch