Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

On correct evaluation techniques of brightness enhancement effect measurement data

Leoš Kukačka
Pascal Dupuis
Hideki Motomura
Jiří Rozkovec
Milan Kolář
Georges Zissis
Masafumi Jinno

SPIE.

Leoš Kukačka, Pascal Dupuis, Hideki Motomura, Jiří Rozkovec, Milan Kolář, Georges Zissis, Masafumi Jinno, "On correct evaluation techniques of brightness enhancement effect measurement data," *Opt. Eng.* **56**(11), 114103 (2017), doi: 10.1117/1.OE.56.11.114103.

Optical Engineering 56(11), 114103 (November 2017)

On correct evaluation techniques of brightness enhancement effect measurement data

Leoš Kukačka,[a,b,c,*] Pascal Dupuis,[b] Hideki Motomura,[a] Jiří Rozkovec,[d] Milan Kolář,[c] Georges Zissis,[b] and Masafumi Jinno[a]
[a]Ehime University, Department of Electrical Engineering, Matsuyama, Japan
[b]Université de Toulouse, LAPLACE, UMR 5213 (CNRS, INPT, UPS), Toulouse, France
[c]Technical University of Liberec, Institute of Mechatronics and Computer Engineering, Liberec, Czechia
[d]Technical University of Liberec, Faculty of Economics, Liberec, Czechia

Abstract. This paper aims to establish confidence intervals of the quantification of brightness enhancement effects resulting from the use of pulsing bright light. It is found that the methods used so far may yield significant bias in the published results, overestimating or underestimating the enhancement effect. The authors propose to use a linear algebra method called the total least squares. Upon an example dataset, it is shown that this method does not yield biased results. The statistical significance of the results is also computed. It is concluded over an observation set that the currently used linear algebra methods present many patterns of noise sensitivity. Changing algorithm details leads to inconsistent results. It is thus recommended to use the method with the lowest noise sensitivity. Moreover, it is shown that this method also permits one to obtain an estimate of the confidence interval. This paper neither aims to publish results about a particular experiment nor to draw any particular conclusion about existence or nonexistence of the brightness enhancement effect. [DOI: 10.1117/1.OE.56.11.114103]

Keywords: brightness enhancement; Broca–Sulzer; Talbot–Plateau; pulsed light; total least squares.

Paper 171010 received Jun. 30, 2017; accepted for publication Oct. 23, 2017; published online Nov. 14, 2017.

1 Introduction

Using a pulsing bright light to achieve brightness enhancement has been subject to research for many years. The first experiments with continuously pulsing light were performed as early as in the nineteenth century,[1] and it was concluded that light pulsing at frequency higher than critical fusion frequency (CFF) appears to the observer at the same brightness level as steady light with luminance level equal to the average level of the pulses. This is known as the Talbot–Plateau law. However, the techniques used by that time involving traditional light sources and a shutter did not allow for very steep rising and falling slope of the pulses.

By the beginning of the twentieth century, it was shown that a single light pulse (with steep rising and falling edge) at scotopic luminance level appears brighter than is the actual luminance of the light source, as reported in Ref. 2. This is known as the Broca–Sulzer effect.

With the dawn of the solid-state lighting technology, there has been an effort to revise the Talbot–Plateau law using continuously pulsing light with very steep rising and falling slope of the pulses.[3–11] In these works, the brightness enhancement effect is studied for monochromatic light of various wavelength, for white light, and for pulses of various frequency and various duty ratio.

Human retina was shown to transfer frequencies up to roughly 160 (or even up to 200) Hz.[12,13] Such an invisible flicker may become a cause of headaches. Especially with the rising usage of LED lighting systems where the common method of dimming is a pulse-width modulation operated at hundreds of hertz, pulsing light may be encountered ever more often. Thus, correctly identifying the brightness enhancement effect of fast pulsing light (above CFF) may help to understand how pulsing light is perceived by the human eye.

Results reported by the aforementioned works about brightness enhancement effect are inconsistent, as shown by a meta-analysis of the results provided in this paper. The result inconsistency discussion was first opened in Ref. 10 where comparison with Ref. 8 was given. This paper does not focus on finding and reporting a magnitude of the brightness enhancement gained by pulsed operation. It focuses instead upon reviewing the methods used for evaluating the experiment data. The motivation for doing this is seeking an explanation of the observed discrepancies among past results from other works.

The numerical processing approaches used in the cited works are summarized in this paper. A key requirement is proposed, which the authors think should be satisfied by any unbiased evaluation method. It is found that neither of the cited works satisfies this condition. Further in this paper, recommendations are given about using a correct method that would yield unambiguous results.

This paper is organized as follows: Sec. 2 summarizes research results from referenced works. A brief description of the two most common approaches used in the cited experiments is given in Sec. 3. In Sec. 4, four distinct statistical methods are described. The results of the methods are compared upon an example data set in Sec. 5. Significance of the results and other possible ways to process the data are computed in Sec. 6. Correct usage of the proposed TLS approach is given in Sec. 7, followed by the conclusions.

*Address all correspondence to: Leoš Kukačka, E-mail: leos.kukacka@tul.cz

0091-3286/2017/$25.00 © 2017 SPIE

2 Research Results Meta-Analysis

This section summarizes the reported results from several published experiments. From each experiment, only tests with 10% duty ratio pulses are picked. Some reports describe experiments with monochromatic light; red, green, and blue color tests were picked in these cases, see Table 1. Other reports describe experiments with white light, see Table 2.

For the purpose of this paper, it is crucial to note the evaluation approach chosen within each cited report. Methods in this summary are marked as A and B, choice I or II; in the sections that follow it is described what the methods A and B, or choices I and II, are.

In Ref. 3, monochromatic (464, 520, and 633 nm) light pulses at 60 Hz are tested. Brightness enhancement reported there is summarized in Table 1 (results are given for two observers). In order to obtain g_g, the authors employed method B, choice II, in this paper. In Ref. 4 the light of the same wavelength was tested. Results are shown in Table 1. The authors again used method B, choice II.

In Ref. 5, the authors analyzed white light produced by parallel function of four monochromatic LEDs of different wavelengths, pulsing at 60 Hz and 10% duty ratio. The method used to obtain g_g was method A, choice II. The result ($d_g = 3\%$) was found to be insignificant.

In Ref. 6, the authors tested brightness enhancement of a white LED with the primary emission peak (blue region) filtered out. The reported enhancement effect (g_g) varies between 1.025 and 1.14 ($d_g = 2.5 - 14\%$) in dependence on the duty ratio, the frequency being fixed at 100 Hz. The authors employed method B, choice I, in order to obtain the g_g coefficient.

In Ref. 9, monochromatic lights of wavelengths 430, 460, 490, 520, 580, 610 and 640 nm, pulsing at 100 Hz, were tested. Reported enhancement effect (Ref. 9, p. 7, Fig. 4) varies between $g_g = 1$ and 1.8 ($d_g = 0 - 80\%$), depending on the duty ratio and tested wavelength. Some of the tested wavelengths are identical (or very close) to those tested in Refs. 3 and 4. Results for those wavelengths are shown in Table 1. Method employed here was method A, choice I.

In Ref. 10, white light with various pulse shapes was tested for brightness enhancement effect. For the ideal pulse shape, the reported brightness enhancement effect was $d_g = 5\%$. The method employed here was A, choice II.

In Ref. 11, white LEDs were used for identifying the brightness enhancement effect. Two cases (diffusers used or not used) are described, yielding brightness enhancement effect $d_g = 7$ or 13%. The method employed here was A, choice I.

It is apparent that there are significant differences among the results (d_g). The rest of this paper focuses on the evaluation methods used in the cited works and tries to explain the differences by analyzing the different approaches taken.

Table 1 Monochromatic light brightness enhancement effect.

(a) Results reported by Ref. 3, used method was B-II

Wavelength (nm)	464	520	633
Reported g_g	0.68	0.48	1.29
	0.67	0.62	1.09
Reported d_g (%)	47.06	108.33	−22.48
	49.25	61.29	−8.26

(b) Results reported by Ref. 4, used method was B-II

Wavelength (nm)	464	520	633
Reported g_g	0.65	0.48	0.84
	0.79	0.6	0.79
Reported d_g (%)	53.85	108.33	19.05
	26.58	66.67	26.58

(c) Brightness enhancement reported by Ref. 9 for 10% duty ratio (numbers read from Fig. 4 of Ref. 9), used method was A-I

Wavelength (nm)	460	520	640
Reported g_g	1	1	1.15
Reported d_g (%)	0	0	15

Table 2 White light brightness enhancement effect.

Report	g_g	d_g (%)	Method
5	0.97	3	A-I
8	0.765	30	B-II
10	1.049	4.9	A-II
11	1.07	7	A-I
11	1.13	13	A-I

3 Procedures for Detecting the Brightness Enhancement Effect

The approaches used for experiments are usually based upon finding a combination of two light levels (one continuous and the other pulsing) that appear to a human observer as having the same luminance. One approach is to choose the two light levels randomly and let the observer judge which one of them appears brighter or they have the same brightness. Particularly in Refs. 3, 4, 8, 10, and 11, the experiment procedure is as follows:

1. light level for steady light is chosen randomly within a predefined interval with uniform distribution; the same random draw is performed for the pulsing light;
2. the light levels are applied in parallel onto the measurement setup within a dark room where the observer sits; the observer sees the two lights next to each other at the same time;
3. the observer makes his/her decision about relative brightness of the two lights, not knowing which of the two is steady and which is pulsing;
4. after 5 s of darkness for eye rest the procedure is repeated;

5. for further evaluation only those trials are taken, which were judged as having the same brightness; others are discarded.

This approach is problematic because of point (5). Sometimes the trials leading to "same level judgment" are very close to trials leading to "different level judgments" or even identical. This means that the tested subjects answer ambiguously for a given light level combination. Care needs to be taken when discarding the trials—also "same judgment trials" falling into the ambiguity zone should be discarded. With this issue taken into consideration, the described approach is deemed proper for the purpose of the discussed experiments.

Another possible way to perform the experiment involving the observer's intervention was chosen in Refs. 5–7 and 9. The procedure is as follows:

1. the light level is chosen randomly both for steady and pulsing light;
2. the light levels are applied into the setup;
3. the observer is asked to tune the brightness of one of the lights until the effective brightness matches the other one;
4. the procedure is repeated;
5. all the trials are taken into account for the statistical processing.

To the authors' knowledge, the choice of either approach should not affect the experiment result in any way.

3.1 *Experiment Data*

Whichever way is chosen for the experiment, it results in a set of data pairs $E = [E_{\text{steady}}; E_{\text{pulse}}]$, where the quantities E_{steady} and E_{pulse} are the luminance values (objective light levels) for continuous and pulsing light, respectively. The ratio

$$\alpha = \frac{E_{\text{steady}}}{E_{\text{pulse}}} \tag{1}$$

evaluates the brightness enhancement (or loss) of the pulsing light against the steady light. Because α is a rational quantity, the true nature of the quantities E (be they the luminance, illuminance, luminous flux) does not matter. The quantity

$$\beta = (\alpha - 1) \cdot 100\% \tag{2}$$

gives the enhancement effect in per cent. Thus, for example, $\alpha = 1.3$ of a given data pair means that for achieving the same subjective brightness, 1.3 times more luminance is necessary for the continuous light than for the pulsing light. This further means that $\beta = 30\%$ is positive for a positive enhancement effect gained by using the pulsing light.

More generally these quantities can be defined for any pair of lighting waveforms (WF, numbered 1 and 2), and so

$$E = [E_1; E_2], \tag{3}$$

$$\gamma = \frac{E_1}{E_2}. \tag{4}$$

In this way, it is not defined which one of the waveforms is continuous light or pulsing light, and so it can be no longer stated that γ higher than 1 yields positive detection of an enhancement effect. In order to reflect this ambiguity, the δ can be defined in two ways:

$$\delta = \begin{cases} (\gamma - 1) \cdot 100\%, & \text{for choice I} \\ \left(\frac{1}{\gamma} - 1\right) \cdot 100\% & \text{for choice II} \end{cases}, \tag{5}$$

so the δ truly represents the enhancement effect in the desired direction. The authors believe this is a justified approach as neither the pulsed operation nor the steady power supply should be favored over the other.

3.2 *Evaluating the Data, Swap Indifference Requirement*

The two quantities (γ and δ) can be found for each data pair separately, but the main goal is to find a single global value of γ_g and δ_g for all acquired data points. So far there is no evidence that the enhancement effect should be dependent on the light level; therefore, this approach is deemed acceptable for a broader range of objective light levels.

An estimate of the true value γ_g acquired from statistical evaluation is denoted as g_g; an estimate of δ_g is d_g. Possible methods for finding g_g and d_g are analyzed in this paper.

Because in the basic experiment, we have the steady and pulsing light, we have two equivalent ways of choosing waveforms 1 and 2:

- Choice I: WF1 = steady, WF2 = pulsing;
- Choice II: WF1 = pulsing, WF2 = steady;

the main requirement laid upon the method is that it should yield mutually reciprocal results in respect to the choice of WF1 and WF2. This means that

$$g_{g:\text{I}} = \frac{1}{g_{g:\text{II}}}, \tag{6}$$

where indices I and II each denote one of the options of choice of WF1 and WF2. In other words, the reported enhancement effect

$$d_{g:\text{I}} = d_{g:\text{II}} \tag{7}$$

should be the same if the quantities per WF1 and WF2 are swapped in the calculation and afterward the result is reciprocated. Failure to this condition means that from one set of data points it is possible to get two distinct values of g_g and d_g and then it is up to the researcher whichever result they choose for publication. This may yield a bias in the published results.

Further in the paper, it will be shown that the methods used so far in the conducted research do not satisfy this condition. The choices in the following text will always be numbered the same way as in this section (Roman numbers).

3.3 *Scatter Diagrams*

Placing the resulting data pairs upon a two-dimensional plane results in a scatter plot. This way the value γ_i represents

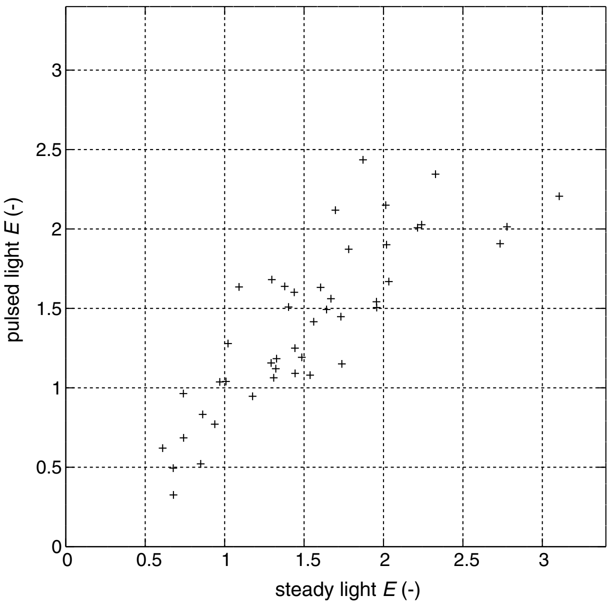

Fig. 1 A scatter plot of the example data used in this paper.

the slope of a line drawn from the origin toward one of the data points (i'th data point), and γ_g represents the slope of a single line drawn from the origin toward the cluster of all the data points.

Figure 1 shows an example of such scatter diagram; in this paper the example dataset will be used for numerical demonstrations. The example data set comes from Ref. 11 and was chosen for easy demonstration of the analyzed methods. The example data set contains $N = 45$ observations. The data set was acquired using the procedure where the observer does not have the possibility to manipulate the brightness and only judges the relative brightness of pulsed and steady light.

4 Examined Methods

In this section, we describe several methods available for statistical evaluation of the experiment data. All these methods allow for calculating the general coefficient estimate g_g from the original data set.

The assumptions necessary for solving the problem in Sec. 3.2 are that the points are scattered along a line with zero intercept; thus the measurements can be hypothetically split into additive zero mean noise and unnoised values

$$E_1 = E_{1,0} + \epsilon, \quad E_2 = E_{2,0} + \epsilon', \tag{8}$$

$$\mathrm{E}[\epsilon] = 0, \quad \mathrm{E}[\epsilon'] = 0, \tag{9}$$

the index 0 indicates unnoised values. Equation (9) means that the expected value of the noise is zero. There are no assumptions about the distribution or variance of the noise. The origin of the noise is either acquisition or uncertainty of the human observer.

For the unnoised values, there is linear relationship

$$E_{1,0} = \gamma_g E_{2,0}, \tag{10}$$

where γ_g is the true physical value that is to be estimated. From Eq. (10) one can arrive to

$$\gamma_g = \frac{\mu_1}{\mu_2}, \tag{11}$$

where μ_1 and μ_2 are mean values of $E_{1,0}$ and $E_{2,0}$, respectively.

4.1 *Method A: Slope Mean Value*

The first of the examined methods is in principal finding the mean value of the ratios Eq. (4)

$$g_{gA} = \frac{1}{N}\sum_i \gamma_i = \frac{1}{N}\sum_i \frac{E_{1,i}}{E_{2,i}}. \tag{12}$$

Histogram of the ratios for the example data set is in Fig. 2. Swapping the variables and finding a reciprocal value in Eq. (12) yields

$$\frac{N}{\sum_i \frac{E_{2,i}}{E_{1,i}}}, \tag{13}$$

which is not equal to Eq. (12) and the swap indifference requirement is not fulfilled.

The expected value of this estimate can be found (using second-order Taylor expansion) as

$$\mathrm{E}[g_{gA}] = \mathrm{E}\left[\frac{E_1}{E_2}\right] \approx \frac{\mathrm{E}[E_1]}{\mathrm{E}[E_2]} - \frac{1}{\mathrm{E}[E_2]^2}\mathrm{Cov}[E_1, E_2] + \frac{\mathrm{E}[E_1]}{\mathrm{E}[E_2]^3}\mathrm{Var}[E_2], \tag{14}$$

which can be simplified into

$$\mathrm{E}[g_{gA}] = \frac{\mu_1}{\mu_2} - \frac{1}{\mu_2^2}\mathrm{Cov}[E_1, E_2] + \frac{\mu_1}{\mu_2^3}\sigma_2^2. \tag{15}$$

The last two terms represent an error of the estimate. Because they will always be nonzero, they also represent a bias and thus prove the estimate is not consistent.

This method is used in Refs. 5 and 9–11.

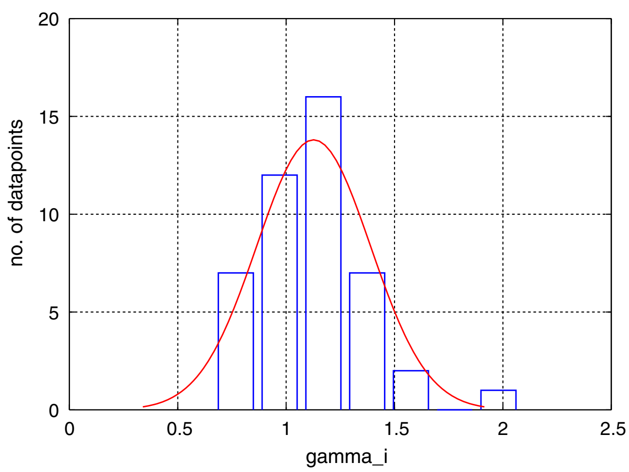

Fig. 2 Example data set: a histogram of the calculated γ_i per each data point.

4.2 *Method B: Ordinary Least Squares Method*

The ordinary least squares method (OLS) is a linear regression method for finding a fit between a predicted variable and explanatory variables. In this case, the fit will be a line with zero intercept

$$E_{2,0} + \epsilon' = \gamma_{\mathrm{gB}} E_{1,0}. \tag{16}$$

The OLS minimizes the sum of squared error ϵ_i of all data points $E_i = [E_{1,i}; E_{2,i}]$. The estimate g_{gB} is computed analytically by

$$g_{\mathrm{gB}} = \frac{\sum_i (E_{1,i} E_{2,i})}{\sum_i E_{1,i}^2}. \tag{17}$$

Swapping the variables and finding a reciprocal value in Eq. (17) yield

$$\frac{\sum_i E_{2,i}^2}{\sum_i (E_{1,i} E_{2,i})}, \tag{18}$$

which is not equal to Eq. (17) and the swap indifference requirement is not fulfilled.

This is because the error is assumed to be within the variable E_2, whereas the quantity E_1 is assumed error-free, as can be seen in Eq. (16). This is not consistent with the assumptions stated at the beginning of this section. Ignoring noise in either variable may introduce bias. Under the assumptions stated above, it can be shown [(Ref. 14, chapter 8, Eqs. (8.8) and (8.9)] that the OLS estimate is biased and does not converge to γ_{g} (the estimate is inconsistent). This method is used in Refs. 3, 4, 6, and 8.

4.3 *Method C: Total Least Squares Method*

The major difference between the OLS and the total least squares (TLS) method is that the error is assumed both within variables E_1 and E_2.[14] This way both variables are treated equally. It is not the scope of this paper to give much detail about this method. The model used in this case is

$$\mathbf{E}_1 = \begin{bmatrix} E_{1,1} \\ E_{1,2} \\ \vdots \\ E_{1,N} \end{bmatrix}, \mathbf{E}_2 = \begin{bmatrix} E_{2,1} \\ E_{2,2} \\ \vdots \\ E_{2,N} \end{bmatrix}, \tag{19}$$

$$[\mathbf{E}_2 \ \mathbf{E}_1] = [g_{\mathrm{gC}} \mathbf{E}_{1,0} \ \mathbf{E}_{1,0}] + [\epsilon' \epsilon], \tag{20}$$

$$\mathrm{Cov}[[\epsilon' \epsilon]] = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{21}$$

The original observation matrix is split in an unnoised matrix of rank 1 and a rank 2 matrix containing two noise vectors, uncorrelated and of the same magnitude. The first term of the RHS of Eq. (20) minimizes the norm of the difference between the original data set and its approximation. In practice, the TLS problem can be solved using the singular value decomposition. The coefficient estimate g_{gC} is then found as a ratio of the norms of the two colinear vectors

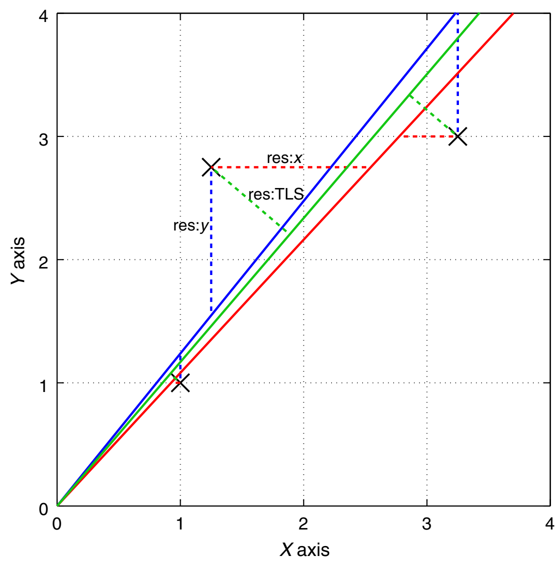

Fig. 3 Demonstration of the approaches of OLS (red, blue) and TLS (green line).

$$g_{\mathrm{gC}} = \frac{\sqrt{\sum_i E_{1,0,i}^2}}{\sqrt{\sum_i E_{2,0,i}^2}}. \tag{22}$$

During the process, the variables are interchangeable; therefore, swapping the variables and finding a reciprocal value in Eq. (22) yields the same formula and the swap indifference requirement is fulfilled. It can be shown (Ref. 14, chapter 8) that under the assumptions stated above this method converges to the true value γ_{g} with no bias (the estimate is consistent).

A graphical example of OLS and TLS fits is given in Fig. 3. The figure gives an example of three points on a plane, each approach resulting in a different fit; dashed lines are the residuals. The blue line is an OLS fit under assumption of noised Y and noiseless X; red line is an OLS fit under assumption of noised X and noiseless Y; green line is the TLS fit, assuming noised both X and Y. The figure shows which noise component is minimized by three different approaches. For the TLS, the residuals are taken perpendicular to the fit, which better describes the way the data are acquired. The TLS method has not been used in any of the hereby cited reports.

4.4 *Method D: Slope of the Centroid*

This approach finds the centroid of the analyzed data set $E_{\mathrm{c}} = [\frac{1}{N}\sum_i E_{1,i}; \frac{1}{N}\sum_i E_{2,i}]$ and determines the slope of the centroid

$$g_{\mathrm{gD}} = \frac{\sum_i E_{1,i}}{\sum_i E_{2,i}}. \tag{23}$$

After swapping the variables and finding a reciprocal value, we receive the identical formula. Therefore, the swap indifference requirement is fulfilled. This is possible because the noise is canceled via summation prior to finding the reciprocal value of one of the variables.

The expected value of the estimate can be expressed similarly to Sec. 4.1 as

$$\mathrm{E}[g_{\mathrm{gD}}]=\mathrm{E}\left[\frac{\mathrm{E}[E_1]}{\mathrm{E}[E_2]}\right]=\mathrm{E}\left[\frac{A}{B}\right] \approx \frac{\mathrm{E}[A]}{\mathrm{E}[B]}-\frac{1}{\mathrm{E}[B]^2}\mathrm{Cov}[A,B]+\frac{\mathrm{E}[A]}{\mathrm{E}[B]^3}\mathrm{Var}[B], \tag{24}$$

substituting

$$A=\mathrm{E}[E_1], \quad B=\mathrm{E}[E_2]. \tag{25}$$

Because

$$\mathrm{E}[A]=\mathrm{E}[\mathrm{E}[E_1]]=\mathrm{E}[\mu_1]=\mu_1, \quad \mathrm{E}[B]=\mu_2, \tag{26}$$

$$\mathrm{Var}[A]=\mathrm{Var}[\mathrm{E}[E_1]]=\frac{\sigma_1^2}{N}, \quad \mathrm{Var}[B]=\frac{\sigma_2^2}{N}, \tag{27}$$

$$\mathrm{Cov}[A,B]=\frac{1}{N}\mathrm{Cov}[E_1,E_2], \tag{28}$$

we will finally receive

$$\mathrm{E}[g_{\mathrm{gD}}]=\frac{\mu_1}{\mu_2}-\frac{1}{\mu_2^2}\frac{\mathrm{Cov}[E_1,E_2]}{N}+\frac{\mu_1}{\mu_2^3}\frac{\sigma_2^2}{N}. \tag{29}$$

The last two terms represent an error, which will decrease with increasing N. This indicates that method D converges to μ_1/μ_2 and the estimate is asymptotically unbiased (i.e., consistent).

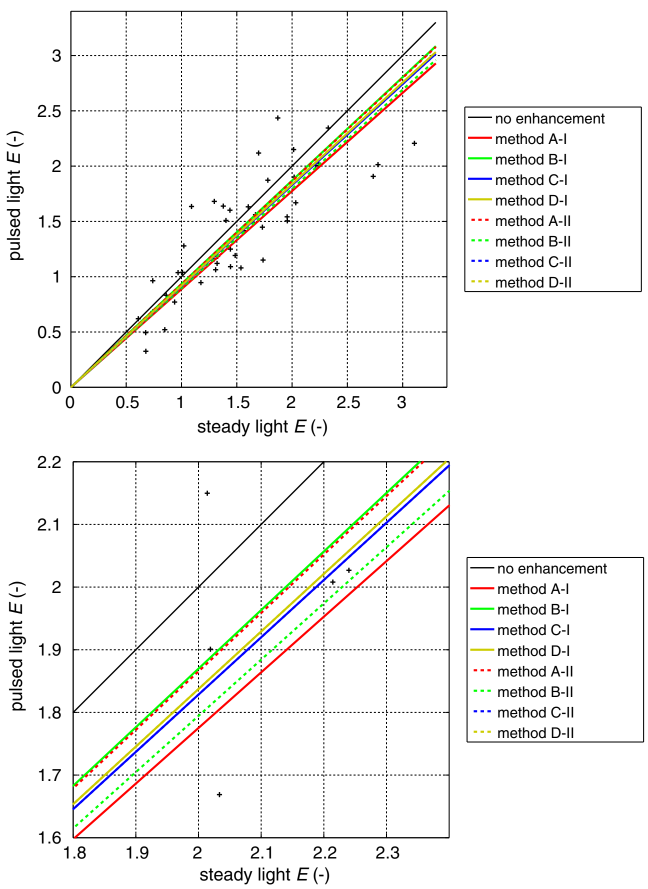

Fig. 4 A scatter plot of the example data used in this paper with lines fitted using several methods; lower subfigure is a detailed view where it is seen that the fits gained by methods C-I and C-II (resp. D-I and D-II) perfectly overlap.

5 Methods Comparison

For numerical comparison of the above-mentioned methods, an example data set was used (see Fig. 4) and the calculations for each of the methods were carried out. Each method was used twice; once with choice I and once with choice II.

Table 3 contains summarized results. Each row contains results for a given method and a particular choice of the WF1 and WF2. For the option I, the choice was: WF1 steady light, WF2 pulsing light (choice I). For the option II, the choice was swapped (choice II) and the brightness enhancement d_g was calculated using a reciprocal value of g_g. In the last column are the values of d_g, which is the enhancement effect magnitude reported by that particular approach.

It can be seen that for the methods A and B it is possible to obtain two distinct values of the same significance. One of them is overestimating the result while the other is underestimating. For the choice I, the table shows that method A overestimates the result while the method B underestimates. For the choice II it is the opposite.

The methods C (the TLS) and D (slope of the centroid) yields identical results for both approaches. This means these methods satisfy the requirement of swap indifference.

Table 3 Results of the comparison upon example data; bold values indicate those from which the d_g was calculated.

Method	WF choice		g_g	$1/g_g$	d_g (%)
(A) Mean value	A-I	WF1 steady WF2 pulsing	**1.127**	0.888	12.7
	A-II	WF1 pulsing WF2 steady	0.933	**1.072**	7.2
(B) OLS	B-I	WF1 steady WF2 pulsing	**1.070**	0.935	7.0
	B-II	WF1 pulsing WF2 steady	0.897	**1.115**	11.5
(C) TLS	C-I	WF1 steady WF2 pulsing	**1.094**	0.914	9.4
	C-II	WF1 pulsing WF2 steady	0.914	**1.094**	9.4
(D) Center slope	D-I	WF1 steady WF2 pulsing	**1.089**	0.919	8.9
	D-II	WF1 pulsing WF2 steady	0.919	**1.089**	8.9

Table 4 Summary of the 95% CI (two-sided) and *p*-values (one-sided) obtained from all methods (methods C and D need not to distinguish between choice I and II); the approach B-I+bootstrap would accept the null hypothesis as the CI covers 0.

Method	Estimate d_g (%)	95% CI (%)	*p*-value
A-I	12.7	±7.9	1.2×10^{-3}
A-I + bootstrap	12.7	±7.5	1.2×10^{-3}
A-II	7.2	±7.2	4.3×10^{-2}
A-II + bootstrap	7.4	±7.0	2×10^{-2}
B-I	7.0	±2.8	3.8×10^{-6}
B-I + bootstrap	7.1	±7.8	3.8×10^{-2}
B-II	11.5	±2.6	3.2×10^{-13}
B-II + bootstrap	11.4	±7.9	2.2×10^{-2}
C	9.4	±7.2	1.2×10^{-2}
C + bootstrap	9.4	±8.0	1×10^{-2}
D	8.9	±6.9	5.6×10^{-3}
D + bootstrap	9.0	±6.9	5.6×10^{-3}

6 Statistical Significance of the Results, Confidence Intervals

This chapter looks into details of checking the significance of the acquired coefficient g_g. In all following numerical examples and figures, the choice of waveforms will always be WF1 steady and WF2 pulsing (choice I) except for Table 4 where the results are summarized for both choices.

6.1 *Method A: Slope Mean Value*

In Ref. 5, the authors calculated the standard deviation of the γ_i

$$s_\gamma = \sqrt{\frac{1}{N}\sum_i (\gamma_i - g_{gA})^2} \tag{30}$$

per each tested subject. An aggregated value for all tested subjects was then used to get a confidence interval $\mathrm{CI} = (g_{gA} \pm s_\gamma)$. The check of the statistical significance lies in checking if the CI contains 1.

There are several problems with this approach. First, the authors do not mention the significance level of this CI and thus the power of the check. Second, the standard deviation in Eq. (30) is for particular observations γ_i. When analyzing the mean value g_{gA}, the standard deviation is correctly identified as $s_\gamma/\sqrt{N}$. Consequentially without such adjustment, this approach is very strict and tends to reject smaller values of the brightness enhancement effect.

The right way to express the CI with significance level a for mean estimate of a population with unknown standard deviation should be calculated using Student's t distribution with $N-1$ degrees of freedom:

$$\mathrm{CI} = \left[g_{gA} \pm t_{N-1}\left(\frac{a}{2}\right)\frac{s_\gamma}{\sqrt{N}} \right]. \tag{31}$$

For our example dataset, the correctly calculated two-sided 95% CI for g_{gA} according to Eq. (31) is $\mathrm{CI} = (1.048; 1.206)$. A p-value of the test can be calculated either one-sided or two-sided. The p-value is the significance level of the test. It indicates the probability of having false-positive detection. The lower p-value, the more significant the result is. The upper limit $p = 0.05$ is a commonly recognized limit for statistical checks. For our case, more proper is one-sided p-value. It can be calculated as

$$p = 1 - t_{N-1}^{-1}(T), \tag{32}$$

where T is a test statistic

$$T = \frac{g_{gA} - 1}{\frac{s_\gamma}{\sqrt{N}}}. \tag{33}$$

For our example data set, the p-value was $p = 1.2 \times 10^{-3}$.

6.2 *Method B: Ordinary Least Squares*

Finding a fit using the OLS makes no assumptions about the input data. Testing the results for significance assumes the residuals follow normal distribution with zero mean and that the variance of the E_2 is constant (the assumption of homoskedasticity). Figure 5 shows that the residuals distribution can be assumed normal. The normality can be checked using several methods; in this paper, we used the Chi-square goodness-of-fit. For our sample, the hypothesis of normality was accepted.

Heteroskedasticity is often encountered though, just as it is in our example data. Several ways of performing a significance test for heteroskedastic data are described in Ref. 15.

In general, a statistical test for null hypothesis $g_{gB} = 1$ (i.e., no enhancement effect) looks like this: the test statistic

$$T = \frac{g_{gB} - 1}{\sqrt{s^2 v_{11}}}, \tag{34}$$

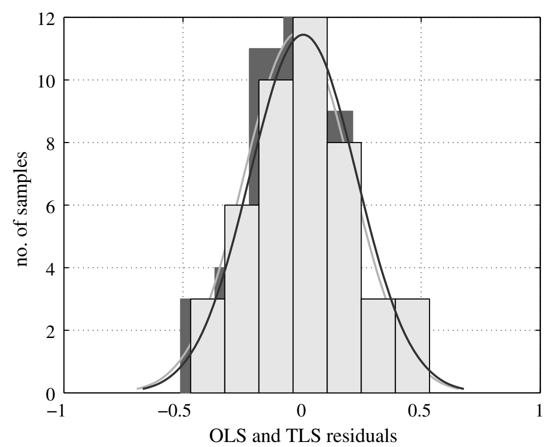

Fig. 5 Histogram of the OLS (light) and TLS (dark) residuals.

where s^2 is the estimate of the parameter variance and v_{11} is a diagonal element of the covariance matrix [in case of using the model (16) the covariance is a scalar], is consistent with the Student's t-distribution of $N-1$ degrees of freedom (N is the sample size) when null hypothesis is valid. The covariance matrix can either be calculated for homoskedastic data or estimated for heteroskedastic data (refer to Ref. 15). For our example data set, the test statistic is $T = 5.0746$.

The one-sided p-value of the test can be calculated identically to Eq. (32). In our example data, the p-value of the t-test is $p = 3.8 \times 10^{-6}$.

A two-sided 95% confidence interval (indicating the result is within given boundaries with 95% probability) for the OLS result g_{gB} can be calculated

$$\mathrm{CI} = [g_{gB} \pm t_{N-1}(0.975)\sqrt{s^2 v_{11}}]. \tag{35}$$

For our example data, the confidence interval is $\mathrm{CI} = (1.042; 1.097)$.

6.3 *Method C: Total Least Squares*

Histogram of the TLS residuals for our example data is in Fig. 5. The Chi-square goodness-of-fit test performed upon the residuals confirmed that the residuals come from normal distribution with zero mean.

One of possible ways of estimating the covariance matrix with heteroskedastic data for the TLS is described in Ref. 16. A test statistic, p-value and confidence interval can be constructed in the same way as in the previous paragraph. For our example data, the test statistic is $T = 2.6339$, p-value is $p = 0.012$ and 95% confidence interval is $\mathrm{CI} = (1.022; 1.166)$.

6.4 *Method D: Slope of the Centroid*

The variance of the estimate g_{gD} can be expressed as

$$\mathrm{Var}[g_{gD}] = \mathrm{Var}\left[\frac{\mathrm{E}(E_1)}{\mathrm{E}(E_2)}\right] = \mathrm{Var}\left[\frac{A}{B}\right] \approx \frac{\mathrm{Var}[A]}{\mathrm{E}[B]^2} - 2\frac{\mathrm{E}[A]}{\mathrm{E}[B]^3}\mathrm{Cov}[A,B] + \frac{\mathrm{E}[A]^2}{\mathrm{E}[B]^4}\mathrm{Var}[B], \tag{36}$$

using similar approach as in Sec. 4. This will simplify to

$$\mathrm{Var}[g_{gD}] = \frac{\sigma_1^2}{N\mu_2^2} - 2\frac{\mu_1}{\mu_2^3}\frac{1}{N}\mathrm{Cov}[E_1, E_2] + \frac{\mu_1^2\sigma_2^2}{N\mu_2^4}. \tag{37}$$

Calculation involves estimating σ_1 and σ_2 by sample standard deviations s_1 and s_2, and μ_1 and μ_2 by sample mean of E_1 and E_2, respectively. Calculations performed for the example data set reveal that $\mathrm{Var}[g_{gD}] = 1.2473 \times 10^{-3}$. A 95% CI can be constructed as

$$\mathrm{CI} = (g_{gD} \pm z_{0.975}\sqrt{\mathrm{Var}[g_{gD}]}), \tag{38}$$

where $z_{0.975}$ is the normal distribution quantile. Calculations show that $\mathrm{CI} = (1.0195; 1.1575)$.

6.5 *Bootstrapping*

The bootstrap is yet another technique for inferring the accuracy of the fitted results.[17] It is very versatile and can be used even when there is no other way of determining the accuracy of a computational result and no knowledge about the noise distribution in the data. The basic idea is to perform a "repetition" of the original experiment yielding similar noise distribution. To this end, another set is created from the original data set, using simple random sampling with replacement. From this new data set, a new g_g estimate is calculated.

This process is repeated B times, producing a set of B new estimates $b_1, b_2 \ldots b_B$. From these we can get the approximation of the distribution of the estimates. The mean value

$$\bar{b} = \frac{1}{B}\sum_{i=1}^{B} b_i \tag{39}$$

can be seen as a new estimate of γ_g.

The bootstrapping technique also allows for calculating the confidence interval. Figure 6 shows a histogram of the bootstrapped estimates of γ_g according to methods A, B, C, and D; the distribution is usually very close to normal. A confidence interval from the bootstrapped results can be constructed (see Ref. 18)

$$\mathrm{CI} = [\bar{b} \pm F(0.975)s(b)], \tag{40}$$

where $s(b)$ is the standard deviation of the estimates b_i

$$s(b) = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(b_i - \bar{b})^2}, \tag{41}$$

and F is either the empirical cumulative distribution function (CDF) of b_i, normal CDF or for N smaller than 30 the CDF of Student's t distribution with N degrees of freedom.

Bootstrapped results and their CIs per each method are summarized together with intrinsic CIs in Table 4. For the method B, the intrinsic CI and bootstrapped CI are very

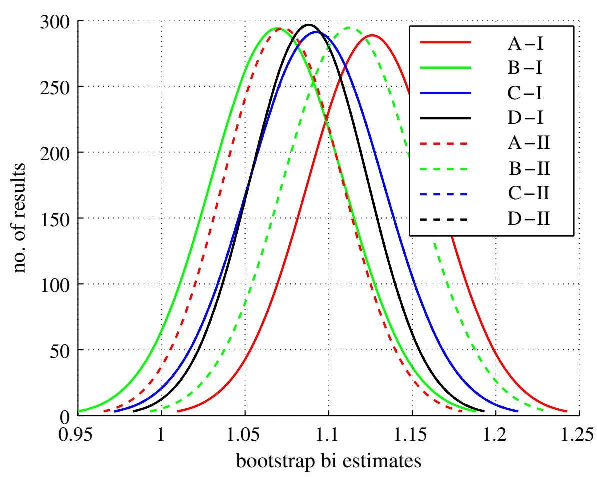

Fig. 6 Distribution of the estimates b_i according to four methods, choice I and II, $B = 10{,}000$. The bootstrap method assumes the (unknown) distribution of estimates g_g can be approximated by the distribution of estimates b_i. All results calculated from identical bootstrap draws; for methods C and D the choices I and II overlap.

different; this indicates that there might be a problem with noise treatment in the method B. Without the bootstrap, the significance of the result is grossly overestimated. For the method D, the intrinsic uncertainty equals the bootstrapped uncertainty almost exactly.

6.6 *Removing Outliers*

Outlier is a term used in statistics for a data point which for some reason does not correspond with the rest of the processed data. The outliers distort the result significantly even though their relative significance to other observations is low. Therefore, every evaluated data set should be cleared of outliers prior to processing.

For the TLS, it is possible to detect the outliers analyzing the noise $E_{n,i} = [\epsilon_i; \epsilon_i']$. The square of the normalized noise should follow Chi-square distribution with N degrees of freedom. Those observations where either ϵ or ϵ' is too large (with more than 95% probability that they do not come from the Chi-square distribution) can be deemed as outliers and should be removed from the data set. In our example data set, this approach did not reveal any outliers.

7 Correct Usage of the TLS Method

In this section, the authors would like to give recommendations about correct usage of the method C (TLS) for evaluating the brightness enhancement effect experimental data.

First, the TLS should be calculated using all data points and using the preliminary results, any detected outliers (see Sec. 6.6) should be removed. The calculation should be repeated on the outliers-free data set until all remaining residuals are compatible with the global standard deviation.

Second, a t-test should be performed confirming that the resulting coefficient g_g is significantly different from 1. The test (Sec. 6.3) should be performed while assuming heteroskedasticity in the data,[16] if it is the case. A p-value of the t-test should be calculated; p-value smaller than 0.05 is commonly recognized as a statistically significant result. The p-value 0.05 indicates 5% chance of false-positive detection.

Third, a CI for g_g can be constructed using Eq. (35). Usually, 95% CIs are used. Alternatively, the confidence interval can be estimated using the bootstrap (Sec. 6.5).

In dependence on the choice of WF1 and WF2 (choice I or II), the final enhancement effect d_g should be calculated using one of the two options in Eq. (5) from the resulting g_g. The endpoints of the CI—if reciprocation is necessary—should be recalculated using this approach

$$\mathrm{CI}_{\mathrm{new}} = \frac{2}{g_g} - \frac{1}{g_g^2}\mathrm{CI}_{\mathrm{orig}}, \tag{42}$$

which is a linearized transform.

The conclusion for our example data set would be that there is an enhancement effect detected of the magnitude $d_g = 9.3 \pm 7.2\%$ (Table 4).

8 Conclusions

The paper evaluated the bias, standard deviation, confidence interval, and p-value of the methods used to quantify the brightness enhancement effect of pulsing light. To the best of the authors' knowledge, those aspects were not fully considered in the reported studies.

The authors have formulated a condition of swap indifference, which should necessarily be satisfied by the employed method. It is shown that the previously used methods do not satisfy this condition and thus may yield a bias in the reported results. The authors come up with the idea of employing the TLS method for data evaluation. It is established that this method (noted here as method C) satisfies said condition of swap indifference as it treats both variables from the experiment data the same way and thus yields unbiased results. The authors also come up with another very simple method (method D), which is swap-indifferent and yields consistent estimates. The authors would like to discourage other researchers from using other methods.

The formulated statements are verified upon an example data set and a numerical example is given for each method. Summary is shown in Table 3.

It is concluded that the method D can be used for data evaluation but performs better for larger sample sets because the noise is canceled out only via summation. On the other hand, the method C can be used even for smaller data sets as it is designed to minimize the residuals directly.

Because the methods A and B (Sec. 4) yield biased (systematically over- or underestimated) results, their usage is strongly discouraged, although the method B yields more significant results (lower p-value) than method C. It is shown that this results from noise underestimation in method B.

The noise level in the example data set is very high. Figure 6 shows that all eight ways to estimate the g_g are consistent with each other (the distributions overlap) and, numerically, it is impossible to make significant difference among the results. Even so, the swap indifference requirement should still be fulfilled by the employed method in order to reduce noise sensitivity.

Considering the findings, it can be said that at least some of the discrepancies shown in Sec. 2 can be explained by improper noise treatment. All the results shown in Table 1 may all suffer from overestimation, and so it is advisable to look for other source of discrepancy.

However, Table 2 shows that the discrepancies may be explained by improper evaluation approaches. Overestimated results may be Refs. 8 and 11 (Table 2, row two, four and five), whereas in Refs. 5 and 10 (Table 2, row one and three) the results may be underestimated.

Disclosures

There are no known conflicts of interests to declare by the authors.

Acknowledgments

This work was supported by the Students Grant Competition at the Technical University of Liberec.

References

1. J. A. F. Plateau, "Betrachtungen über ein von Hrn Talbot vorgeschlagenes photometrisches Princip," *Bulletin de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles* **2**, 52 (1835).
2. A. Broca and D. Sulzer, "La Sensation Lumineuse en Function du Temps," *Comptes Rendus de l'Académie des Sciences (Paris)* **134**, 831–834 (1902).
3. M. Jinno et al., "Beyond the physical limit: energy saving lighting and illumination by using repetitive intense and fast pulsed light sources and the effect on human eyes," *J. Light Visual Environ.* **32**(2), 170–176 (2008).

4. M. Jinno et al., "Effective illuminance improvement of a light source by using pulse modulation and its psychophysical effect on the human eye," *J. Light Visual Environ.* **32**(2), 161–169 (2008).
5. I. Fryc, W. Davis, and Y. Ohno, "Experiment on visual perception of pulsed LED lighting: can it save energy for lighting?" in *CIE 2010 Lighting Quality and Energy Efficiency* (2010).
6. G. Xin, "Pulse modulation effect of light-emitting diodes on human perception enhancement," *Opt. Eng.* **51**, 073608 (2012).
7. S. Fan et al., "Human perception on pulsed red and green lights," *Opt. Eng.* **53**(6), 065105 (2014).
8. H. Motomura, Y. Ikeda, and M. Jinno, "Evaluation of visual perception enhancement effect by pulsed operation of LEDs," in *14th Int. Symp. On the Science and Technology of Lighting*, Italy, p. 14 (2014).
9. S. Fan et al., "Influence of pulse width on luminous efficiency for a two-degree field," *Lighting Research and Technology* **49**(3), 357–369 (2016).
10. L. Kukačka et al., "Broca-Sulzer effect detection over critical fusion frequency for pulse operated white LEDs with varied pulse shape," in *15th Int. Symp. On the Science and Technology of Lighting*, Kyoto, Japan, p. 15 (2016).
11. C. Lassfolk et al., "Brightness enhancement by pulsed operation of LEDs," in *15th Int. Symp. On the Science and Technology of Lighting*, Kyoto, Japan, p. 15 (2016).
12. S. Berman et al., "Human electroretinogram responses to video displays, fluorescent lighting, and other high-frequency sources," *Optom. Vision Sci.* **68**, 645–662 (1991).
13. IEEE, "IEEE 1789-2015 IEEE recommended practices for modulating current in high-brightness LEDs for mitigating health risks to viewers," IEEE Std 1789-2015, pp. 1–80 (2015).
14. S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem*, SIAM, Philadelphia, Pennsylvania, pp. 19104–2688 (1991).
15. A. F. Hayes and L. Cai, "Using hetroskedasticity-consistent standard error estimators in OLS regression: an introduction and software implementation," *Behav. Res. Methods* **39**(4), 709–722 (2007).
16. R. L. Branham, "A covariance matrix for total least squares with heteroscedastic data," *Astron. J.* **117**, 1942–1948 (1999).
17. B. Efron, "Bootstrap methods: another look at the jackknife," *Ann. Stat.* **7**, 1–26 (1979).
18. S. Lei and M. Smith, "Evaluation of several Efron bootstrap methods to estimate error measures for software metrics," in *Canadian Conf. on Electrical and Computer Engineering, 2002. IEEE CCECE 2002*, Vol. **2**, pp. 703–708 vol.2 (2002).

Leoš Kukačka received his bachelor's and and master's degrees in automatic control systems from the Technical University in Liberec in 2011 and 2013, respectively. He is pursuing his PhD under joint thesis supervision of Technical University in Liberec, Czechia, and Université Toulouse III–Paul Sabatier, France. His main field of interest is flicker and power quality within DC grids with modern lighting systems.

Pascal Dupuis received his MS degree in electrical engineering and his PhD in applied sciences from Louvain School of Engineering, Louvain-la-Neuve, Belgium, in 1989 and 2001, respectively. He worked as a postdoctoral researcher at Katholieke Universiteit Leuven, Leuven, Belgium, from 2002 to 2006, and from 2006 to 2012 as a senior researcher at Louvain School of Engineering. He joined Paul Sabatier University, Toulouse, France, as a development engineer in 2012. His background is about measurements in a wide sense.

Jiří Rozkovec studied mathematics and chemistry at Charles University in Prague (1990-1995, MSc.), Faculty of Science. Then he studied financial mathematics at Faculty of Mathematics and Physics of Charles University (1995-2001, MSc.), Department of Probability and Mathematical Statistics. At present he studies in the PhD program of applied mathematics at the Technical University of Liberec, Faculty of Science, Humanities, and Education. His field of research is statistics, probability, and optimizations.

Milan Kolář was born in Liberec, Czechia, in 1963. He studied microelectronics at Czech Technical University in Prague, Faculty of Electrical Engineering. From the same faculty subsequently, he earned his PhD in 1992. He is working at the Technical University of Liberec, currently as head of Institute of Mechatronics and Computer Engineering. He is engaged in the analog and digital electronic design.

Georges Zissis graduated in 1986 from the Physics Department of the University of Crete in general physics. He received his MSc and PhD degrees in plasma science in 1987 and 1990 from Toulouse 3 University, France. His primary area of work is in the field of light sources science and technology. He is interested in the physics of electrical discharges used as light sources; system and metrology issues for SSL systems; normalization and quality issues for light sources; impact of lighting to energy, environment, health, and security. He is a full professor Toulouse 3 University (France). He is director of the Light & Matter research group of LAPLACE.

Biographies for the other authors are not available.