How to Predict Lightness Variations from One Illuminant to Another?

Steven Le Moan
School of Engineering and Advanced Technology
Massey University
Albany, New Zealand
Email: s.lemoan@massey.ac.nz

***Abstract*—The notion of subjective image quality (SIQ) is particularly difficult to define as it depends on a large variety of physical attributes and psycho-physical factors. One important attribute which SIQ depends on is the illumination under which visual stimuli are observed. Recent studies have shown that achromatic fidelity (i.e. perceived resemblance in lightness) is significantly less dependent on illumination than other perceptual attributes such as chroma and hue. These findings were used to develop the 5-dimensional extended colour space LabAB, which uses a single lightness channel to represent 31-dimensional multispectral (reflectance) data for efficient, perception-driven processing in spectral reproduction workflows. This channel, which is derived from the perceptually uniform LAB2000HL colour space with CIE D50 white point, is meant to approximate the perceived lightness under most light sources. However, little is known as to how accurate this approximation actually is. For instance, if a spectral cross-media reproduction application requires a high quality of reproduction under both daylight and LED light, will LabAB and its daylight-based lightness channel be a good enough representation to capture and convey all attributes of SIQ throughout the reproduction workflow? A metric that would allow to evaluate the reliability of using a single lightness dimension to represent spectral data for a given application is desirable. Here, I investigate how several spectral and colour difference measures perform at estimating this reliability.**

I. Introduction

Imagine looking at two seemingly identical paintings under the light of day. One is an original, the other is a printed reproduction. Chances are that, although they produce the same colour sensation under that particular light, there exists other illuminants under which they would appear different to the human eye. This is because the reflectance of the original painting's pigments is nearly impossible to reproduce exactly with common printing technologies, and they can therefore only be approximated by exploiting metamerism. The question is then: What is a good approximation? In the case of art reproduction, it is likely that the answer to such question has to do with how the original and reproduction differ in terms of human visual perception, as opposed to e.g. in terms of the chemical content of their respective pigments. However, traditional (colourimetric) printing techniques account only for perceptual match under daylight (typically CIE D50) while a variety of applications involve more than just one illuminant (replication of artwork, security printing, furniture catalogues, colour checkers for camera calibration,...). Hence the rationale behind *spectral* printing technologies, that is to reproduce not only colour but reflectance spectra, which can be used to estimate colour under any arbitrary illuminant. Therefore, *spectral* reproduction generalises *colourimetric* reproduction. However, this comes at a cost: larger datasets and higher computational loads. In the case of very high dimensional spectral images, another drawback can be encountered, that is the *curse of dimensionality* [1]. In this study however, I will concentrate exclusively on multispectral data acquired in the visible range of wavelength, which typically contain only 31 channels.

Early work on *spectral* reproduction relied on a purely computational measure such as the Mean Square Error (MSE) to measure spectral fidelity [2]. However, if the reproduced painting needs only to *look* like the original, the aim is then to mimic human perception. In that case, the MSE conveys very little meaning and a more perceptually-driven approach is needed [3].

In this study, I will focus on the problem of formatting spectral data (reflectance spectra) for a perceptually-driven spectral reproduction workflow. More specifically, I will follow up on recent studies on the five-dimensional LabAB extended colour space[1] [4], which allows to convey most of the underlying perceptual "information" from 31-dimensional spectral data into 5 dimensions. It permits in particular the encoding of device transforms in look-up tables and enables a high run-time performance. One of the key principles upon which the design of LabAB is based is that lightness is a redundant colour feature across illuminants, especially in comparison to chromatic features even when considering chromatic adaptation, i.e. the ability of our visual system to compensate for changes in illumination chromaticity [3]. Therefore, a single channel is used to convey lightness information from reflectance. Though it was proven that this simplification yields an acceptable tradeoff between loss of spectral "information" and reduction of computational effort, a more thorough investigation of this reliability is needed. Consider the following situation: a spectral printing workflow based on LabAB with a lightness channel corresponding to CIE D50 is available, but a product needs to be printed so that its perceived quality is optimal under daylight *and*

[1]Can also be referred to as *interim connection space* or *profile connection space*.

978-1-5090-2748-4/16/$31.00

fluorescent light. Chances are that the workflow will ensure a good quality under the former (due to the lightness channel being directly estimated from the daylight CIE D50), but what about the latter? In other words: is considering only one lightness channel in a low-dimensional representation of spectral reflectance suitable for a given application? Or will it engender noticeable colour distortions?

In order to answer these questions, one can compute all the corresponding colours for CIE D50 as well as for the fluorescent lights under consideration and compare them, but that would be time consuming. Alternatively, one should be able to predict, by analysing the shapes of the illuminants' spectral power distribution (SPD), just how well they correlate lightness-wise for any given reflectance. In this paper, I investigate means to perform such a prediction with simple spectral-divergence or colour-difference metrics.

II. The LabAB extended colour space

Multi- or hyper-spectral pixels are usually represented as high-dimensional tensors. For many applications, not all of these dimensions are actually necessary, and a few values are enough to measure and/or process the pixels with a very good accuracy. Extended colour spaces such as LabPQR [5], XYZXYZ [6] or LabAB [4] were designed for that purpose. They consist of a limited number of dimensions that can convey most of the information from the raw spectral data for efficient spectral colour management. Note that, although these extended colour spaces were primarily designed for cross-media communication, they can be used for any application that involve perceptually-driven multispectral image reproduction such as compression. In this study, I will focus on LabAB and provide new results to help understanding its reliability.

LabAB has four perceptual dimensions: an achromatic one (L) and four chromatic ones (abAB). The first three dimensions (Lab) represent the colour of the input spectral reflectance under the standard illuminant CIE D50 and converted to the hue-linear and perceptually uniform LAB2000HL colour space [7]. It is particularly important to use CIE D50 in order to ensure that a spectral workflow based on LabAB provides at least the same level of performance as a colourimetric workflow (which is typically based on D50). The last two dimensions of LabAB are meant to convey chromatic variations under a set of illuminants $\mathbf{\Theta}$. This set is dependent on the application under consideration (e.g. tungsten lights, fluorescent lights...) and a representative SPD is extracted from it by principal component analysis or simple averaging. The colour of the input spectral data is then computed for this representative illuminant but only the chromatic components $\mathbf{a}_{00\mathrm{HL}}$ and $\mathbf{b}_{00\mathrm{HL}}$ (from LAB2000HL) are kept on account of the fact that variations of lightness from one illuminant to another for a given spectral reflectance are generally relatively small. These components are then mapped to the last two dimensions of LabAB.

Again, the main underlying assumption here is that the perceived lightness of an object varies very little when it is observed under daylight, fluorescent light or most other types of lights.

III. Experiments and results

A. Data

Experiments were performed on the representative Standard Object Colour Spectra (SOCS) database [8] as test data (53490 reflectance spectra). This set will be noted $\mathbf{R}$ (for reflectance).

Additionally, I used a collection of 88 illuminants' SPDs noted $\mathbf{\Theta}$, made of 20 daylights, 17 tungsten lights, 18 fluorescent lights and 33 LED lights. These illuminants were selected from the National Gallery's set [9] as well as from the University of Eastern Finland's daylights set [10] and the CIE standard illuminants. They were all normalised to have a maximal value of 1. Note that CIE D50 was excluded from $\mathbf{\Theta}_{\mathrm{All}}$. Figure 1 depicts the SPDs of the whole set.

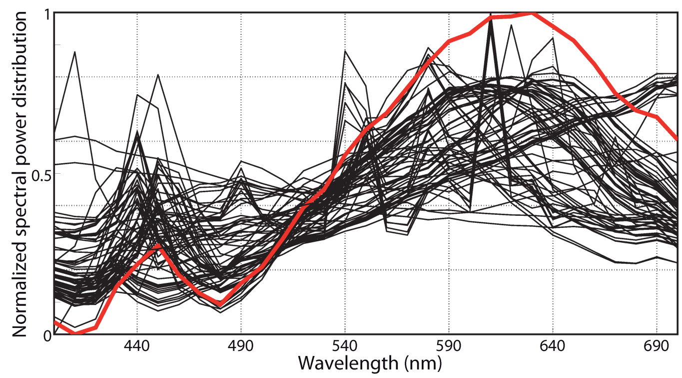

Fig. 1. Spectral power distributions of the illuminants used in this study. In red: normalised first principal component of the set. (Figure reproduced from [11])

B. Reference measures of rendered lightness difference

Since we are interested in the ability for the standard illuminant CIE D50 (noted $\mathbf{i}_1$) to predict the lightness rendered by another arbitrary illuminant $\mathbf{i}_2$, one option would be to take a set of spectra, render them for both $\mathbf{i}_1$ and $\mathbf{i}_2$ and measure $r_{\mathrm{ref}}(\mathbf{i}_1, \mathbf{i}_2) = r\left(\mathbf{L}_{\mathbf{R}}^{\mathbf{i}_1}, \mathbf{L}_{\mathbf{R}}^{\mathbf{i}_2}\right)$, the Pearson's correlation coefficient between the two sets of rendered lightness values for each reflectance spectrum $\mathbf{r}$ from $\mathbf{R}$ ($\mathbf{L}_{\mathbf{R}}^{\mathbf{i}_1} = \left\{\mathrm{L}_{\mathbf{r}}^{\mathbf{i}_1} | \forall \mathbf{r} \in \mathbf{R}\right\}$, and respectively for $\mathrm{L}_{\mathbf{R}}^{\mathbf{i}_2}$).

However, in order to fully account for the end goal of spectral reproduction, i.e. visual fidelity, a more perceptually-driven measure should also be considered. Additionally, as pointed out in [4], these correlation values are usually very close to 1 and the difference between them is difficult to evaluate in terms of significance. This is mainly due to the fact that luminance is always normalised to the range $[0 \ldots 1]$.

For these reasons, I also introduce $\Delta E_{00\mathrm{HL}}\left(\left\{\mathbf{L}_{\mathbf{r}}^{\mathbf{i}_2}, \mathrm{a}_{\mathbf{r}}^{\mathbf{i}_2}, \mathrm{b}_{\mathbf{r}}^{\mathbf{i}_2}\right\}; \left\{\mathbf{L}_{\mathbf{r}}^{\mathbf{i}_1}, \mathrm{a}_{\mathbf{r}}^{\mathbf{i}_2}, \mathrm{b}_{\mathbf{r}}^{\mathbf{i}_2}\right\}\right)$, the colour difference engendered by replacing one stimulus' lightness by another's, measured in LAB2000HL. I then propose to compute the maximal value of this difference over the set of test reflectance spectra: $\Delta_{\text{ref}}(\mathbf{i}_1,\mathbf{i}_2) = \max_{\mathbf{R}} \left\{ \Delta E_{00\text{HL}} \left(\left\{ \mathbf{L}_{\mathbf{r}}^{\mathbf{i}_2}, \text{a}_{\mathbf{r}}^{\mathbf{i}_2}, \text{b}_{\mathbf{r}}^{\mathbf{i}_2} \right\} ; \left\{ \mathbf{L}_{\mathbf{r}}^{\mathbf{i}_1}, \text{a}_{\mathbf{r}}^{\mathbf{i}_2}, \text{b}_{\mathbf{r}}^{\mathbf{i}_2} \right\} \right) \middle| \forall \mathbf{r} \in \mathbf{R} \right\}$.

Note that these two measures do require to render all test spectra for each illuminant, which is tedious and precisely what this study aims at making unnecessary. If a single spectral or colour difference measure that solely takes as an input the SPDs of $\mathbf{i}_1$ and $\mathbf{i}_2$ can approximate these two reference measures well, it can then be used to quickly and easily assess the reliability of using LabAB for a particular application of cross-media spectral reproduction.

Over the full set of experimental data, it was evaluated that these two measures exhibit a Pearson correlation coefficient of 0.953, meaning that they correlate to a large extent. Table I gives some the average, minimal and maximal values evaluated by each of them.

TABLE I
STATISTICS OF THE REFERENCE MEASURES OVER THE FULL SET OF EXPERIMENTAL DATA, ALWAYS WITH CIE D50 AS REFERENCE ILLUMINANT.

	mean	min	max
r_{ref}	0.994	0.982	1.000
Δ_{ref}	2.79	0.18	5.40

C. Test measures of estimated rendered lightness difference

Aiming at estimating the reference measurements given by r_{ref} and Δ_{ref}, I considered the following benchmark of simple spectral and colour difference measures:

- RMSE$(\mathbf{i}_1,\mathbf{i}_2)$: The root mean square error between SPDs.
- GFC$(\mathbf{i}_1,\mathbf{i}_2)$: The goodness-of-fit coefficient between SPDs.
- $\Delta E_{\text{XYZ}}(\mathbf{i}_1,\mathbf{i}_2)$: The colour difference between white points in CIE XYZ. Note that a white point has by definition a Y value (luminance) of 1.
- $\Delta E_{\text{xyY}}(\mathbf{i}_1,\mathbf{i}_2)$: The chromatic difference between white points in the CIE xyY chromaticity diagram.
- $\Delta E_{00\text{HL}}(\mathbf{i}_1,\mathbf{i}_2)$: The colour difference between white points in the hue-linear perceptually uniform LAB2000HL colour space (normalised for CIED50's white point).
- $\Delta E_{\text{C}}(\mathbf{i}_1,\mathbf{i}_2)$ and $\Delta E_{\text{H}}(\mathbf{i}_1,\mathbf{i}_2)$: Respectively the Chroma and Hue differences between white points in the corresponding channels of the cylindrical representation of LAB2000HL, LCh00HL, where $\text{C}_{00\text{HL}} = \sqrt{\text{a}_{00\text{HL}}^2 + \text{b}_{00\text{HL}}^2}$ and $\text{h}_{00\text{HL}} = \tan^{-1}\left(\frac{\text{b}_{00\text{HL}}}{\text{a}_{00\text{HL}}}\right)$. Note that, because values of ρ are so large, meaning that the difference of lightness $\Delta E_{\text{L}}(\mathbf{i}_1,\mathbf{i}_2)$ is close to zero, we obtain that $\Delta E_{00\text{HL}}(\mathbf{i}_1,\mathbf{i}_2) \approx \Delta E_{\text{C}}(\mathbf{i}_1,\mathbf{i}_2)$.

The ability of each of these measures to predict r_{ref} and Δ_{ref} is estimated by means of the Pearson Correlation Coefficient (PCC) and Kendall Rank Order Correlation Coefficient (KROCC) for each illuminant from $\mathbf{\Theta}$.

D. Significance analysis

In order to assess whether two Pearson correlation coefficients are significantly different from each other, I used Steiger's Z-test [12] with a p-value of 0.05. It consists of first converting the coefficients to z-scores via Fisher's r-to-z transform and then computing the asymptotic covariances of the estimates. This test accounts for the fact that the bivariate correlation coefficients to compare are not independent in that they share one variable. Indeed, we are looking at how two different test metrics correlate with a same reference metric. Note also that the distributions of r_{ref} and Δ_{ref} over the test data were found to be normal as their respective kurtosis values were both slightly above 2.

The significance of Kendall rank order correlation coefficients was assessed by first converting them to PCC (see [13], page 126) and then applying the same procedure.

E. Results

Results are given in Tables II and III.

It can be seen that ΔE_{xyY} is overall the best performing measure, with PCC with r_{ref} and Δ_{ref} of 0.940 and 0.945 as well as KROCC of 0.852 and 0.784, respectively. Figure 2 helps visualising the high PCC of ΔE_{xyY} with Δ_{ref} on the CIE chromaticity diagram.

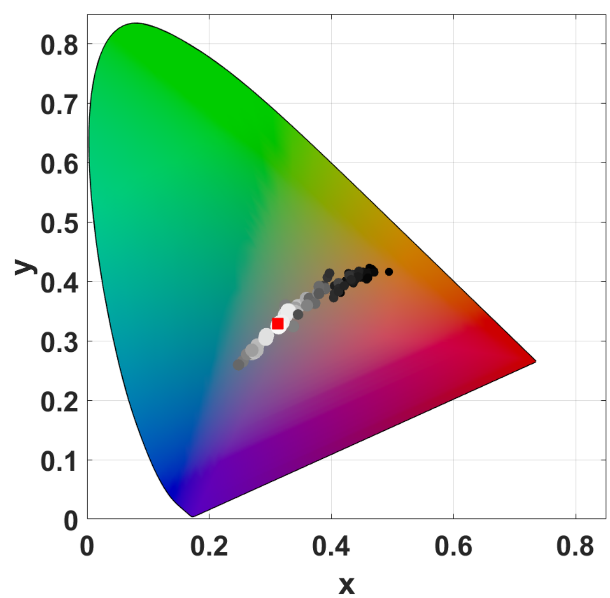

Fig. 2. Chromaticity diagram where each dot represents one illuminant from $\mathbf{\Theta}$. The size and lightness of each dot are proportional to the Δ_{ref} value for the corresponding illuminant (lighter and bigger means smaller Δ_{ref}). The red square represents CIE D50. It can be seen that there is a correlation between the distance $\Delta E_{\text{xyY}}(\mathbf{i}_1,\mathbf{i}_2)$ and the values of the reference measure Δ_{ref}.

We can also observe the very poor performance of ΔE_H at predicting their similarities in rendering lightness, as opposed to ΔE_C, which on the other hand achieves the same results as $\Delta E_{00\text{HL}}$ and better ones than RMSE and GFC.

TABLE II

PEARSON AND KENDALL RANK ORDER CORRELATION COEFFICIENTS (PCC AND KROCC) BETWEEN MEASURES OF SPECTRAL DIVERGENCE AND r_{REF}. THE BEST RESULTS ARE SHOWN IN BOLD FONT (NOT SIGNIFICANTLY DIFFERENT FROM EACH OTHER, ROW-WISE).

	RMSE	1-GFC	ΔE_{XYZ}	ΔE_{xyY}	ΔE_{00HL}	ΔE_{C}	ΔE_{H}
PCC	0.717	0.750	0.846	**0.940**	0.899	0.899	0.135
KROCC	0.640	0.720	0.767	**0.852**	**0.875**	**0.875**	-0.180

TABLE III

PEARSON AND KENDALL RANK ORDER CORRELATION COEFFICIENTS (PCC AND KROCC) BETWEEN MEASURES OF SPECTRAL DIVERGENCE AND Δ_{REF}. THE BEST RESULTS ARE SHOWN IN BOLD FONT (NOT SIGNIFICANTLY DIFFERENT FROM EACH OTHER, ROW-WISE).

	RMSE	1-GFC	ΔE_{XYZ}	ΔE_{xyY}	ΔE_{00HL}	ΔE_{C}	ΔE_{H}
PCC	0.840	0.825	0.868	**0.945**	**0.935**	**0.935**	0.218
KROCC	0.656	**0.733**	0.709	**0.784**	**0.797**	**0.797**	-0.187

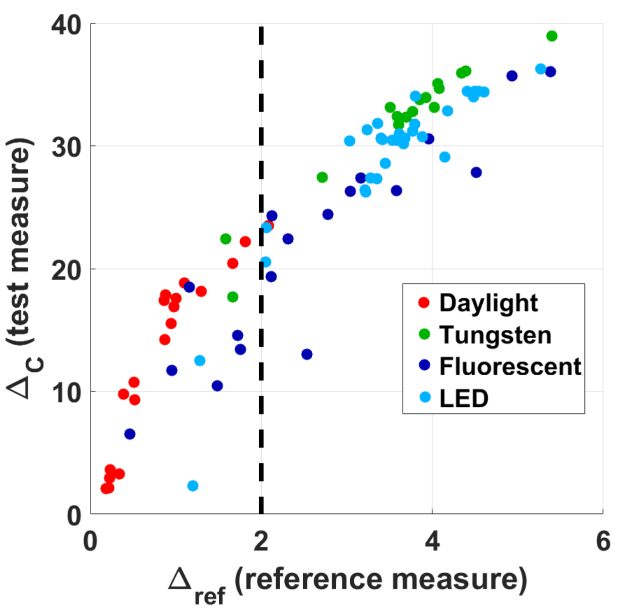

Fig. 3. Graph showing the correlation between ΔE_{C} and Δ_{ref}. The dotted line represents the threshold of just noticeable difference for Δ_{ref}.

Incidentally, this confirms the intuitive idea that the difference of lightness between two illuminants' white points in a colour space normalised for one of them conveys less information regarding the similarity of the lightness values that they can render than the difference of chromaticity. These results also point out that chroma is a considerably more informative channel than hue in that regard.

Having established which test measures are the most reliable, it would be useful to know what is an acceptable range of values, i.e. to find a threshold of difference between the two illuminant beyond which a conversion to LabAB is likely to engender noticeable colour distortions.

Since ΔE_{00HL} and ΔE_{C} convey significantly more perceptual uniformity than ΔE_{xyY} and are therefore more intuitive to deal with, I will only consider them for this part. And because they are approximately equivalent, I will narrow the benchmark down to solely ΔE_{C}, the simplest one.

From Figure 3, it can be seen that there is no clear ΔE_{C} threshold separating the whole subset of $\mathbf{\Theta}$ that yield Δ_{ref} values smaller than a just noticeable difference (which I will assume to be equal to 2 ΔE_{00HL} units, though this is an approximation, the JND depending on a variety of factors not considered here) from its dual subset. This threshold is however likely to be in the range of 13 to 22. Any value $\Delta E_{\text{C}}(\mathbf{i}_1, \mathbf{i}_2)$ below 13 means that the lightness of CIE D50 is fit to represent that of $\mathbf{i}_2$ without perceptible difference. On the other hand any value larger than 22 will very likely result in artifacts.

Incidentally, this plot also shows that for all of the tungsten light considered here but two, using LabAB leads to perceptible colour distortions, while all daylights but one lead to no such artifacts.

IV. CONCLUSION

In order to better understand the reliability of the LabAB extended colour space, I challenged its underlying assumption – perceived lightness under CIE D50 can predict that under any other light source – and investigated means to predict how suitable LabAB is for a given application. For this purpose, I looked at the performance of several measures of spectral divergence (including colour difference metrics). Interestingly, experimental results suggest that the Euclidean distance in CIE xyY as well as the normalised chroma difference between two illuminants white points ΔE_{C}, correlate to a very large extent to the perceived effect of the assumption upon which LabAB is based. These results are to be considered as preliminary guidelines rather than strict rules. Further investigations are needed to make more solid conclusions, and in particular to check the possibility of finding a single ΔE_{C} threshold for an easier and more intuitive use of LabAB.

REFERENCES

[1] LO Jimenez and DA Landgrebe, "Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 28, no. 1, pp. 39–54, 1998.

[2] A. M. Bakke, I. Farup, and J. Y. Hardeberg, “Multispectral Gamut Mapping and Visualization - A First Attempt,” 2005, vol. 5667, pp. 193–200, SPIE.

[3] Steven Le Moan and Philipp Urban, “Image-difference prediction: From color to spectral,” *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2058–2068, 2014.

[4] Steven Le Moan, Jana Blahová, Philipp Urban, and Ole Norberg, “Five dimensions for spectral colour management,” *(to be published in) Journal of Imaging Science and Technology*, vol. 60, 2016.

[5] Maxim Derhak and Mitchell Rosen, “Spectral colorimetry using LabPQR: an interim connection space,” *Journal of Imaging Science and Technology*, vol. 50, no. 1, pp. 53–63, 2006.

[6] X. Zhang, Q. Wang, J. Li, P. Yang, and J. Yu, “The interim connection space based on human color vision for spectral color reproduction,” *Journal of Optical Society of America A*, vol. 29, no. 6, pp. 1027–1034, 2012.

[7] I. Lissner and P. Urban, “Toward a unified color space for perception-based image processing,” *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1153–1168, 2012.

[8] Technical Committee ISO/TC 130, Graphic technology, “ISO/TR 16066: Standard object colour spectra database for colour reproduction evaluation (SOCS),” Tech. Rep., 2003.

[9] “Spectral power distribution curves, the national gallery: http://research.ng-london.org.uk/scientific/spd/,” Accessed on April. 23rd 2016.

[10] “Daylight spectra, university of eastern finland: http://cs.joensuu.fi/{$\sim$}spectral/databases/download/daylight.htm,” Accessed on April. 23rd 2016.

[11] Steven Le Moan and Philipp Urban, “A new connection space for low-dimensional spectral color management,” in *Measuring, Modeling, and Reproducing Material Appearance*, Maria V. Ortiz Segovia, Philipp Urban, and Jan P. Allebach, Eds., San Francisco, CA, February 2014, vol. 9018, pp. 9018–12, SPIE.

[12] James H Steiger, “Tests for comparing elements of a correlation matrix.,” *Psychological bulletin*, vol. 87, no. 2, pp. 245, 1980.

[13] Maurice George Kendall, “Rank correlation methods.,” 1948.