

Bayesian approach to color-difference models based on threshold and constant-stimuli methods

Fernando Brusola,^{1,*} Ignacio Tortajada,¹ Ismael Lengua,¹ Begoña Jordá,¹ and Guillermo Peris¹

¹Centro de Investigación en Tecnologías Gráficas, Universitat Politècnica de València, Camino de Vera s/n 46022 Valencia, Spain.

*fbrusola@upv.es

Abstract: An alternative approach based on statistical Bayesian inference is presented to deal with the development of color-difference models and the precision of parameter estimation. The approach was applied to simulated data and real data, the latter published by selected authors involved with the development of color-difference formulae using traditional methods. Our results show very good agreement between the Bayesian and classical approaches. Among other benefits, our proposed methodology allows one to determine the marginal posterior distribution of each random individual parameter of the color-difference model. In this manner, it is possible to analyze the effect of individual parameters on the statistical significance calculation of a color-difference equation.

© 2015 Optical Society of America

OCIS codes: (330.0330) Vision, color, and visual optics; (330.1690) Color; (330.1730) Colorimetry.

References and links

1. A. R. Robertson, "CIE guidelines for coordinated research on colour-difference evaluation," *Color Res. Appl.* **3**(3), 149–151 (1978).
2. R. M. Rich, F. W. Billmeyer, Jr., and W. G. Howe, "Method for deriving color-difference-perceptibility ellipses for surface-color samples," *J. Opt. Soc. Am.* **65**(8), 956–959 (1975).
3. R. S. Berns, D. H. Alman, L. Reniff, G. D. Snyder, and M. R. Balonon-Rosen, "Visual determination of suprathreshold color-difference tolerances using probit analysis," *Color Res. Appl.* **16**(5), 297–316 (1991).
4. D. C. Rich and F. W. Billmeyer, "Small and moderate color differences. IV. Color-difference-perceptibility ellipses in surface-color space," *Color Res. Appl.* **8**(1), 31–39 (1983).
5. M. Huang, H. Liu, G. Cui, and M. R. Luo, "Testing uniform colour spaces and colour-difference formulae using printing samples," *Color Res. Appl.* **37**(5), 326–335 (2012).
6. H. Wang, G. Cui, M. R. Luo, and H. Xu, "Evaluation of colour-difference formulae for different colour-difference magnitudes," *Color Res. Appl.* **37**(5), 316–325 (2012).
7. M. R. Luo and B. Rigg, "Chromaticity-discrimination ellipses for surface colours," *Color Res. Appl.* **11**(1), 25–42 (1986).
8. W. R. J. Brown, W. G. Howe, J. E. Jackson, and R. H. Morris, "Multivariate normality of the color-matching process," *J. Opt. Soc. Am.* **46**(1), 46–49 (1956).
9. R. G. Kuehni, "Visual and instrumental determination of small colour differences: a contribution," *J. Soc. Dyers Colour.* **91**(3), 68–71 (1975).
10. L. F. C. Friele, "Fine color metric (FCM)," *Color Res. Appl.* **3**(2), 53–64 (1978).
11. M. Cheung and B. Rigg, "Colour-difference ellipsoids for five CIE colour centers," *Color Res. Appl.* **11**(3), 185–195 (1986).
12. R. S. Berns and B. Hou, "Rit-Dupont supra-threshold color-tolerance individual color-difference pair dataset," *Color Res. Appl.* **35**(4), 274–283 (2010).
13. D. Strocka, A. Brockes, and W. Paffhausen, "Untersuchungen von farbabstands-ellipsoiden-methodik und ergebnisse an einem beispiel," *Farbe Design.* **25**(16), 75–79 (1980).
14. K. Witt and G. Döring, "Parametric variations in a threshold color-difference ellipsoid for green painted samples," *Color Res. Appl.* **8**(3), 153–163 (1983).
15. E. Kirchner and N. Dekker, "Performance measures of color-difference equations: correlation coefficient versus standardized residual sum of squares," *J. Opt. Soc. Am.* **28**(9), 1841–1848 (2011).
16. D. H. Alman, R. S. Berns, G. D. Snyder, and W. A. Larsen, "Performance testing of color difference metrics using a color tolerance dataset," *Color Res. Appl.* **14**(3), 139–151 (1989).

17. P. A. García, R. Huertas, M. Melgosa, and G. Cui, "Measurement of the relationship between perceived and computed color differences," *J. Opt. Soc. Am. A* **24**(7), 1823–1829 (2007).
 18. S. S. Guan and M. R. Luo, "Investigation of parametric effects using small colour differences," *Color Res. Appl.* **24**(5), 331–343 (1999).
 19. I. Lissner and P. Urban, "Upgrading color-difference formulas," *J. Opt. Soc. Am. A* **27**(7), 1620–1629 (2010).
 20. S. Shen and R. S. Berns, "Color-difference formula performance for several datasets of small color differences based on visual uncertainty," *Color Res. Appl.* **36**(1), 15–26 (2011).
 21. G. Cui, M. R. Luo, B. Rigg, and W. Li, "Colour-difference evaluation using CRT colours. Part I: Data gathering and testing colour difference formulae," *Color Res. Appl.* **26**(5), 394–402 (2001).
 22. S. Wen, "A color difference metric based on the chromaticity discrimination ellipses," *Opt. Express* **20**(24), 26441–26447 (2012).
 23. M. Melgosa, J. Martínez-García, L. Gómez-Robledo, E. Perales, F. M. Martínez-Verdú, and T. Dauser, "Measuring color differences in automotive samples with lightness flop: A test of the AUDI2000 color-difference formula," *Opt. Express* **22**(3), 3458–3467 (2014).
 24. A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *J. Am. Stat. Assoc.* **85**(410), 398–409 (1990).
 25. R. Neal, "Slice sampling," *Ann. Stat.* **31**(3), 705–767 (2003).
 26. OpenBUGS License, Release (3.2.2). Available via the INTERNET; <http://www.openbugs.net/> (Accessed 2015 Feb 2).
 27. I. Ntzoufras, *Bayesian Modeling Using WinBUGS*. (John Wiley & Sons, 2009).
-

1. Introduction

Most of the published work regarding color-difference models has been based on the CIE TC-1.3 Guidelines for Coordinated Research on Color-Difference Evaluation [1]. Guidelines for step one, which are related to object colors that consist mainly for the determination of perceptual ellipsoids around different color centers, can be summarized as follows:

1. Prepared samples are well distributed in the neighborhood of a selected color center.
2. Selection of the psychophysical data-collection technique. There are two main techniques: Pass-Fail, with two variants: Threshold and Constant Stimuli Method, and the Gray Scale.

A selected team of observers is asked if pairs of samples, viewed under very controlled conditions, match or not (threshold method), or if the perceived color difference of the sample pairs look less or greater than a standard selected pair (constant stimuli method). The Constant Stimuli and Threshold Methods have been widely used by many authors [2–4].

Alternatively, observers can be asked to give a value (using a reference gray scale), with pairs of patches ordered from lowest to highest, to determine the measured color difference ΔE . The chosen number should indicate that the perceived difference between the corresponding pair of patches in the reference scale and the difference between the observed pairs appear the same (gray scale method [5,6]).

3. Collect visual data depending on the chosen psychophysical method. If the gray scale method is selected, a visual difference, ΔV_i , is assigned to each sample pair based on the average of each observer-selected grade (number of the reference gray scale) transformed to the fitted ΔE_i values of the scale that correspond to ΔV_i .

The relative frequency of acceptances (pass) or rejections (fail) reported by observers, that a pair of samples match or not, if the threshold method is selected, or that the color difference of a pair of samples exceeds or not a standard pair, if constant stimuli method is selected, can be related to the estimated visual difference using a psychometric function.

Several s-shaped, psychometric functions can be used: the cumulative normal distribution, as used by many authors such as Luo & Rigg [7]; and the Gaussian density, as used by Rich et al. [2]. In this way, when used in tandem with the inverse

function (probit, logit, other log functions) it is possible to translate frequency data in a linearly measured color-difference space.

4. Fit visual ΔV_i data to measured ΔE_i data, calculated using Eq. (1) if ΔE_i is evaluated in chromaticity space, Yxy; or by Eq. (2) if ΔE_i is evaluated in CIELAB space. Any equivalent expression can be used when there are small differences in the other color spaces (Brown et al [8]).

$$\Delta E_i^2 = g_1 \Delta x_i^2 + g_2 \Delta y_i^2 + g_3 \Delta Y_i^2 + 2g_4 \Delta x_i \Delta y_i + 2g_5 \Delta x_i \Delta Y_i + 2g_6 \Delta y_i \Delta Y_i \quad (1)$$

$$\Delta E_i^2 = g_1 \Delta a_i^2 + g_2 \Delta b_i^2 + g_3 \Delta L_i^2 + 2g_4 \Delta a_i \Delta b_i + 2g_5 \Delta a_i \Delta L_i + 2g_6 \Delta b_i \Delta L_i \quad (2)$$

5. Calculate the g_j coefficients using the difference between ΔE_i and ΔV_i to minimize

$f = \sum_{i=1}^n (\Delta V_i - \Delta E_i)^2$, or by choosing other goodness-of-fit criteria such as the linear correlation coefficient between ΔE_i and ΔV_i , as used by Kuehni [9]. Other goodness-of-fit criteria should be considered by finding the minima of the other objective f -functions.

Friele [10], Luo & Rigg [6], and Cheung & Rigg [11] proposed the following function:

$$f = \sum_{i=1}^n [\Delta V_i^2 - \Delta E_i^2]^2 W_i \quad (3)$$

Where W_i is a weighting factor chosen to give more influence to data pairs that are closer to the fitted ellipsoid.

Berns & Hou [12] used, for supra-threshold differences, the following weighted function:

$$f = \sum_{i=1}^n [\Delta V_i - \Delta E_i]^2 W_i \quad (4)$$

For threshold and constant stimuli cases, g_j coefficients can be determined using algorithms that find the best fit between the observed frequencies, P_i , and the estimated probabilities, \bar{P}_i . These are obtained by using psychometric (s-shaped) functions, such as the logistic function, shown in Eq. (5), the Gaussian function, shown in Eq. (6) and the normal cumulative distribution function, shown in Eq. (7):

$$\bar{P}_i = 1 / (1 + e^{\alpha + \beta \Delta E_i}) \quad (5)$$

Where α and β are parameters that are optimized during the fit.

$$\bar{P}_i = 1 - (1 - p) e^{-(\Delta E_i)^2 / \sigma^2} \quad (6)$$

Where p (a false-alarm parameter) and σ are parameters that are optimized during the fit.

$$\bar{P}_i = F(\Delta E_i | \alpha, \beta) = \frac{1}{\beta \sqrt{2\pi}} \int_{-\Delta}^{\Delta E_i} e^{-\frac{(t-\alpha)^2}{2\beta^2}} dt \quad (7)$$

Where α and β are parameters that are optimized during the fit.

In such cases, ellipsoid coefficients can be found via methods that minimize the quadratic deviation Eq. (8), or by using alternative weighted methods such as one that uses the difference between the observed and calculated frequencies, e.g. Strocka [13], Witt [14], etc.:

$$\min\left(\sum_{i=1}^n (P_i - \bar{P}_i)^2\right) \quad (8)$$

An alternative method that is also widely used by other authors [15], is to find the coefficients that maximize the likelihood of the samples of observed data.

The likelihood for constant stimuli, or the threshold method, can be calculated using Eq. (9) under the assumption that the absolute frequency of the observers judging a pair passes or fails a color comparison follows a binomial distribution:

$$L = \prod_{i=1}^N \frac{n_i!}{r_i!(n_i - r_i)!} \bar{P}_i^{r_i} (1 - \bar{P}_i)^{n_i - r_i} \quad (9)$$

Where r_i is the number of positive answers, n_i is the number of presentations of pair i , and N is the number of pairs.

As we can see, the aforementioned methods estimate model parameters (treated usually as regression coefficients) via an inference process based on least-squares or the maximum likelihood method. The estimation process is, usually, followed via model checking. The checking procedure can include:

1. Evaluation of the accuracy of the estimated parameters. Find confidence intervals of the parameters, or the standard error.
2. Determine the goodness-of-fit of the model. The overall goodness-of-fit can be checked by evaluating the coefficient of determination (R^2), the Pearson's correlation coefficient (r), as recommended by Kichner [15], or by the chi-squared test (Alman et al. [16]). Alternatively, a more specific statistic related to the color-difference field, such as STRESS (Standardized Residual Error Sum of Squares), as recommended by García et al. [17], or the combined index PF/3 as proposed by Guan & Luo [18], can be used. Most of the recent published work regarding the performance of color-difference formulae [19–23] is based on goodness-of-fit measurements.
3. Check the structural assumptions of the model (independence, normality, or homoscedasticity of errors).
4. Detection of possible outliers, or highly unlikely observations, given the assumed model.

The aforementioned methodology is what we call the standard or “frequentist” approach to color-difference models. This approach includes the notion of frequentist probability as the limit of the relative frequency of the appearance of a possible event in a large number of trials of a random experiment. Often, scientists refer to frequentist probability as a physical or objective probability that is associated with a random process. Under this approach, it is possible to assign a probability to a random process. In contrast, in Bayesian probability (which is often referred as subjectivist probability or evidential probability), a probability is assigned to any process or statement even when random processes are not involved. This approach can be defined as the degree of belief to which a statement or process is supported by evidence.

In the frequentist approach to inference, unknown parameters are often, but not always, treated as fixed, and hence there is no way to associate probabilities to them. In this branch of statistics, parameter estimation, hypothesis testing, and confidence intervals are based on the underlying influence of the parameter to a random component of data. Under this approach, we define a statistic to estimate the parameter that is expected to be random before a measurement has taken place. We can estimate a parameter by maximizing its calculated likelihood, and thus selecting the parameter that maximizes the probability of being observed in the data, and then manage the reliability of the estimation by defining a confidence interval.

This measure of uncertainty is difficult to interpret due to the subtleties involved. For example, a confidence interval of 95% confidence level does not mean that the probability that the parameter belongs to the selected interval is 95%. After a sample has been taken, the parameter is either in or out of the interval, so that the probability can only be either 1 or 0, which is deterministic and not random. A 95% confidence level indicates the probability that the extremes of an interval selected in accordance to a specific procedure from a sequence of samples captures the true population parameter. That is, the relative frequency with which the confidence interval would include the true value in a long sequence of repetitions of the experiment.

In contrast, a Bayesian inference allows probabilities to be associated with the unknown parameters. The Bayesian approach allows us to incorporate prior knowledge regarding the parameters of a given model. The uncertainty of the unknown parameters is quantified using a probability so that the unknown parameters are regarded as random variables. Prior parameter probability distributions can be modified using new data acquired from experiments. The goal is to find the resulting posterior probability distribution of the parameters given new and previous knowledge. The core of this inference method is Bayes' theorem, which relates prior knowledge with a posterior probability:

$$P(\text{parameter}|\text{data}) = P(\text{data}|\text{parameter}) * \frac{P(\text{parameter})}{P(\text{data})} \propto P(\text{data}|\text{parameter}) * P(\text{parameter}) \quad (10)$$

Where \propto indicates that the posterior distribution of parameters is proportional to the likelihood of data, given the parameters, and is multiplied by the prior distribution of the parameters.

$P(\text{data})$ acts as a 'normalizing' constant, independent of parameters, which is necessary to ensure that the posterior distribution integrates to one.

Once we have determined the posterior distribution of the parameters, the problem of estimation is solved and many questions regarding the accuracy of the process can be easily determined. One can calculate the expected value of the parameters from the posterior distributions, and obtain a measure of the accuracy of the process by means of the standard deviation, credible intervals, or via any other measure of dispersion.

Bayesian credible intervals are analogous to frequentist confidence intervals. Unlike frequentist confidence intervals, however, a credible interval of 95% expresses that the estimation of the parameter, through the posterior probability distribution, is between the extremes of the interval with a 95% probability. In this sense, credible intervals are more intuitive and easier to interpret than confidence intervals.

The whole process of running a color-difference model, following both approaches, can be summarized as follows:

a. Frequentist

1. Specify a probability model for the dependent variable. For example, $r_i \sim \text{Binomial}(n_i, p_i)$ for pass-fail methods, or $\Delta V_i \sim \text{Normal}(m_i, \sigma_i)$ for the gray-scale method.
2. Find a maximum likelihood estimator for g_j and its coefficients (\hat{g}_j).
3. Measure the uncertainty or accuracy of the estimation using the standard error or a confidence interval, assuming a distribution of \hat{g}_j (usually a multivariate normal distribution).
4. Check the goodness-of-fit or other structural assumptions of the model.

b. Bayesian

1. Specify a probability model for the dependent variable ($r_i \sim \text{Binomial}(n_i, p_i)$) for pass-fail methods, or $\Delta V_i \sim \text{Normal}(m_i, \sigma_i)$ for the gray-scale method) and specify the prior distribution of g_i and other parameters involved in the model.
2. Find the posterior distributions of the parameters using Bayes' rule and other conditional dependencies, and calculate the posterior mean.
3. Measure the uncertainty using standard deviations, credible intervals, quartiles, etc.
4. Check the goodness-of-fit or other structural assumptions of the model.

In this paper, we present a Bayesian approach for finding ellipsoidal coefficients of pass-fail methods. We show how to define a color-difference model from a Bayesian point of view, and how to implement and simulate the posterior distribution of parameters using the OpenBUGS software. To check the performance of the model, our approach was applied to simulated (aka “theoretically perfect”) data, generated from ΔE_{94} ellipsoids created from hypothetical observed frequencies given by the Gaussian psychometric function Eq. (6), and to real data published by Witt & Döring [14].

2. Method

The color-difference problem can be assimilated to a special case of a generalized linear-regression model by means of a link function, such as a logistic Eq. (5), a Gaussian Eq. (6) or a normally cumulative function Eq. (7). In all the cases, except for Gaussian link function, the problem is essentially non-linear with respect to g_j , due to the square root of ΔE .

As mentioned above, the first step when running a difference color model from a Bayesian point of view is to specify the probability distribution of the dependent variable. In this paper, we applied the Bayesian approach to the threshold and constant stimuli method to fit the observed frequencies to those predicted by the model. In this case, the dependent variable, the frequency, follows a binomial distribution. We state this using Eq. (11):

$$r_i \sim \text{Binomial}(n_i, p_i) \quad (11)$$

Where p_i can be estimated using a psychometric function. If we use the Gaussian psychometric function Eq. (6), then the estimated value of p_i , \bar{P}_i , is given by Eq. (12):

$$\bar{P}_i = 1 - (1 - p) e^{-\frac{(\Delta E_i)^2}{\sigma^2}} = 1 - (1 - p) e^{-\frac{g_1 \Delta a_i^2 + g_2 \Delta b_i^2 + g_3 \Delta l_i^2 + 2g_4 \Delta a_i \Delta b_i + 2g_5 \Delta a_i \Delta l_i + 2g_6 \Delta b_i \Delta l_i}{\sigma^2}} \quad (12)$$

Under the cited hypothesis, the likelihood is given by Eq. (13), assuming that it has no false alarm factor, $p = 0$ (thus pairs with no instrumental color difference are not allowed), and fixing $\sigma = 1$, thus transferring all of the variation to just the g_j coefficients.

$$P(\text{data}|\text{parameter}) = P(r_i|g_j) = L(g_j|r_i) = \prod_{i=1}^N \frac{n_i!}{r_i!(n_i-r_i)!} \bar{P}_i^{r_i} (1-\bar{P}_i)^{n_i-r_i} = \prod_{i=1}^N \text{binomial}(r_i, n_i, \bar{P}_i) = \prod_{i=1}^N \text{binomial}(r_i, n_i, 1 - e^{-(g_1\Delta a_i^2 + g_2\Delta b_i^2 + g_3\Delta L_i^2 + g_4\Delta a_i\Delta b_i + g_5\Delta a_i\Delta L_i + g_6\Delta b_i\Delta L_i)}) \quad (13)$$

If our prior knowledge of the parameters is accumulated using a probability distribution function $P_j(g_j)$, the posterior probability distribution of g_j is given by Eq. (14), assuming that g_j are independent.

$$P(g_j|\text{data}) \propto \left[\prod_{i=1}^N \text{binomial}(n_i, 1 - e^{-(g_1\Delta a_i^2 + g_2\Delta b_i^2 + g_3\Delta L_i^2 + 2g_4\Delta a_i\Delta b_i + 2g_5\Delta a_i\Delta L_i + 2g_6\Delta b_i\Delta L_i)}) \right] \prod_{j=1}^6 P_j(g_j). \quad (14)$$

However, before we can calculate the posterior probability distribution, we must ensure that $\Delta E_i^2 > 0$ for any combination of g_j . This can be achieved using the Cholesky decomposition.

The Cholesky decomposition method states that if G is a symmetric matrix with real entries, and is positive-definite, then G can be decomposed as $G = L * L'$, where L is a lower triangular matrix with strictly positive diagonal entries and L' denotes the transpose of L .

Conversely, the product $L * L'$ of a lower triangular matrix with strictly positive diagonal entries will result in a symmetric positive-definite matrix:

$$G = \begin{bmatrix} g_1 & g_4 & g_5 \\ g_4 & g_2 & g_6 \\ g_5 & g_6 & g_3 \end{bmatrix} = \begin{bmatrix} d_1 & 0 & 0 \\ d_4 & d_2 & 0 \\ d_5 & d_6 & d_3 \end{bmatrix} * \begin{bmatrix} d_1 & d_4 & d_5 \\ 0 & d_2 & d_6 \\ 0 & 0 & d_3 \end{bmatrix}. \quad (15)$$

$$G = \begin{bmatrix} d_1^2 & d_1*d_4 & d_1*d_5 \\ d_1*d_4 & d_2^2 + d_4^2 & d_4*d_5 + d_2*d_6 \\ d_1*d_5 & d_4*d_5 + d_2*d_6 & d_3^2 + d_5^2 + d_6^2 \end{bmatrix} \quad (16)$$

G is positive-definite, if $d_1, d_2,$ and $d_3 > 0$. Thus, before we compute g_j we must first compute d_j , replacing the values of g_j shown in Eq. (14) by the corresponding values shown in matrix Eq. (16) Once the replacement is done, we can compute the posterior distribution of d_j , as shown in Eq. (17):

$$P(d_j|\text{data}) \propto \left[\prod_{i=1}^N \text{binomial}(n_i, 1 - e^{-(g_1\Delta a_i^2 + g_2\Delta b_i^2 + g_3\Delta L_i^2 + 2g_4\Delta a_i\Delta b_i + 2g_5\Delta a_i\Delta L_i + 2g_6\Delta b_i\Delta L_i)}) \right] \prod_{j=1}^6 P_j(d_j). \quad (17)$$

Under the cited assumptions (depending on the priors selected) it is not always possible to analytically compute the posterior distribution of g_j or d_j . However, by using simulation via Monte Carlo methods, one can generate random samples of the parameters, and from them estimate the posterior mean, median, standard deviation, and other probability distribution parameters.

It is possible, using the Gibbs sampling algorithm [24] or the slice-sampling algorithm proposed by Neal [25], to sample from a distribution using an arbitrary density function that is known only up to a constant of proportionality. These algorithms generate Markovian

sequences whose stationary distribution is the target distribution. Gibbs and slice samplers are also referred to as Markov Chain Monte Carlo (MCMC) algorithms. They differ from other MCMC algorithms in that a proposal distribution is not needed.

The posterior distribution of the parameters given by Eq. (17) can be easily defined and sampled using Matlab's "slicesample" function, or by using the OpenBUGS software [26].

OpenBUGS is an open source version of the University of Helsinki's (in Finland) WinBUGS, the first experimental version of BUGS for windows. BUGS is an acronym of *Bayesian inference Using Gibbs Sampling*. BUGS is the ancestor of WinBUGS and more recently OpenBUGS. The MRC Biostatistics Unit initiated the BUGS project in 1989. Latest versions of WinBUGS are developed at the Imperial College School of Medicine at St. Mary's London, UK.

OpenBUGS allows one to build complex statistical Bayesian models. Models can be specified using a simple code, which is similar to the S language, or by means of a graphical interface (aka a DOODLE interface). Graphical models implemented by the BUGS software belong to a class known as Directed Acyclic Graphs (DAGs). The background theory regarding DAGs allows one to break down the analysis into a sequence of relatively simple computations that reveal the conditional dependence of the model variables.

The possibility to specify the conditional dependence easily by OpenBUGS makes this software more efficient and faster than Matlab for building Bayesian models.

The proposed BUGS doodle dag for model given by Eq. (17) is shown in Fig. 1.

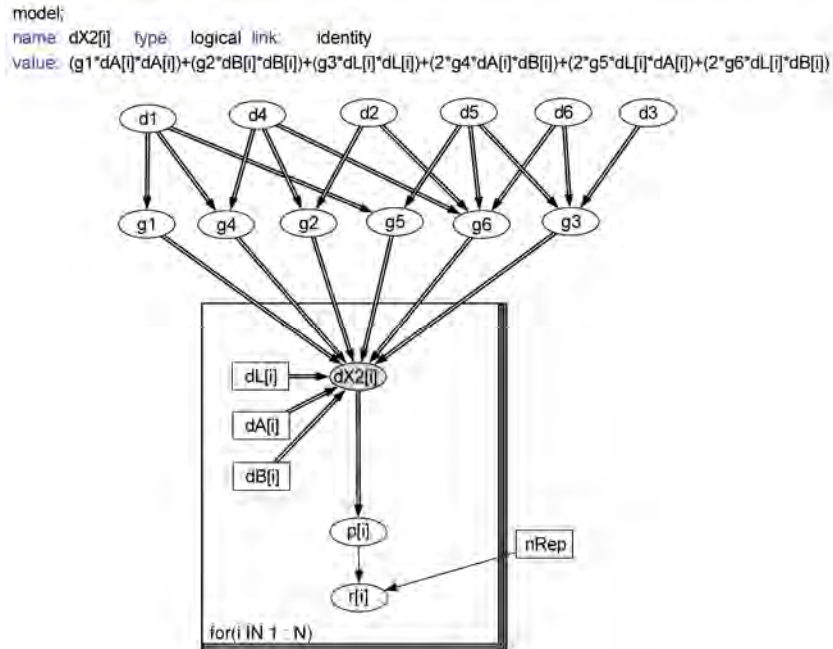


Fig. 1. BUGS doodle of the color-difference model using Cholesky decomposition and the Gaussian psychometric function.

A Doodle has three elements: nodes, edges, and plates.

A node can be stochastic, logical, or constant. Stochastic and logical nodes are elliptically shaped while constant nodes are square shaped.

Stochastic nodes are associated with a probability density function whose parameters depend exclusively on linked parent nodes. Logical nodes are deterministic and therefore are associated with defined non-stochastic functions whose parameters are dependent parent nodes. Constant nodes are related to fixed data and have no parent nodes.

Edges represent direct links. They can also represent either logical (thick hollow arrow) or stochastic (thin solid arrow) dependencies.

Plates are used to represent repeated parts of the graph.

The graph represents the assumption of conditional independence of the stochastic nodes, so that the full joint distribution can be factorized in terms of the conditional distribution of each of the nodes, given its parents.

An equivalent model to that shown in Fig. 1, which instead uses the BUGS formal language, is shown in Fig. 2. A complete description of the model syntax and other details can be consulted in the WinBUGS manual [26]. See Ntzoufras [27] for a very good introduction to the principles of Bayesian modeling, model building, and implementation using WinBUGS.

```

model;
{
d1 ~ dunif(0.0001,1000)
d2 ~ dunif(0.0001,1000)
d3 ~ dunif(0.0001,1000)
d4 ~ dunif(-500,500)
d5 ~ dunif(-500,500)
d6 ~ dunif(-500,500)
g1 <- (d1*d1)
g2 <- (d2*d2) + (d4*d4)
g3 <- (d3*d3) + (d5*d5) + (d6*d6)
g4 <- (d1*d4)
g5 <- (d1*d5)
g6 <- (d4*d5) + (d2*d6)
for(i in 1:N) {
dX2[i] <- (g3 * dL[i] * dL[i]) + (g1 * dA[i] * dA[i]) + (g2 * dB[i] * dB[i]) + ((2 * g4) * dA[i]*dB[i]) +
((2 * g5) * dA[i]*dL[i]) + ((2 * g6) * dB[i]*dL[i])
p[i] <- 1 - exp(-dX2[i])
r[i] ~ dbin(p[i],nRep)
}
}

```

Fig. 2. BUGS model using formal language for color difference model using Cholesky decomposition and Gaussian psychometric function equivalent to doodle model shown in Fig. 1.

Notice that the six lines of code after the initial “{” refer to the specification of the prior probability distributions of the parameters. The next six lines refer to the logical links needed to compute g_j from their Cholesky coefficients d_j , and the code inside the ‘for’ loop refers to the likelihood specification.

After the model definition and implementation, we have to compile and run the model with the data provided. As mentioned above, two data sets have been used to check the performance of our approach. The first is a theoretical one, where the hypothetical frequency of affirmative answers to the question: “there is a noticeable difference”, is applied to color differences that are uniformly distributed around the five-color center, as suggested by Robertson [1]. These were calculated from ΔE_{94} and the psychometric Gaussian function Eq. (6) using no false alarm parameter ($p = 0$) and $\sigma^2 = 1/\log(2)$ so that $\bar{P}_i = 0.5$ if $\Delta E_i = 1$. The second data set corresponds to data series II from Witt & Döring [14], which were obtained using the pair-comparison method.

The distribution of samples around the color centers of the theoretical data set were selected over 24 equally spaced vector directions, similarly to the sample design specified by Alman, Berns, Snyder & Larsen [16]. Each vector direction has a maximum length of $\Delta E = 3$, and samples are distributed uniformly in each direction at $\Delta E = 3/N_s$ intervals (where N_s is the number of samples in each direction, excluding the origin). Figure 3 shows the case when $N_s = 2$.

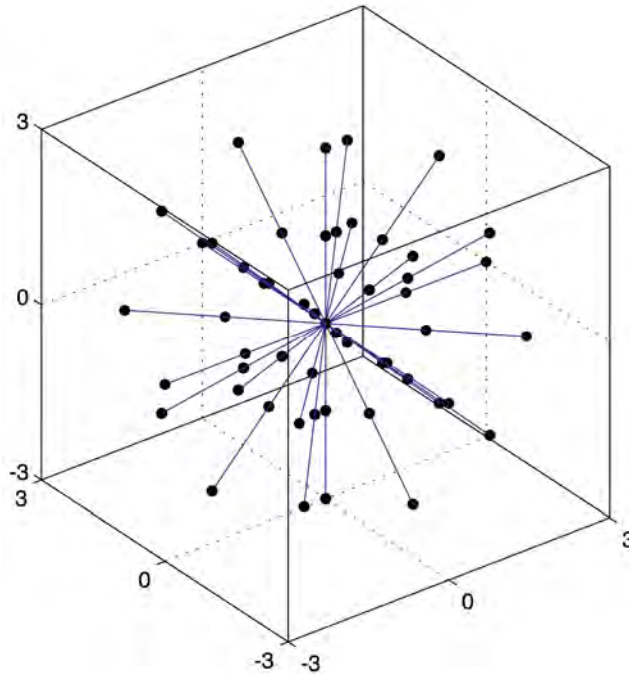


Fig. 3. Theoretical color-difference data set corresponding with $N_s = 2$ samples uniformly distributed in 26 vector directions shown around a color center.

Theoretical absolute-frequency data calculated using the aforementioned procedure for $N_s = 2$, and 40 repeating observer trials for five Robertson's [1] recommended color centers is shown in Table 1. Note that frequency data have been rounded to the nearest integer.

Table 1. Theoretical absolute frequency for each of the samples shown in Fig. 3, around five Robertson's [1] color centers. These values were calculated using ΔE_{94} and the Gaussian link function, and 40 observer repeating trials. The calculated frequency values have been rounded to the nearest integer.

Delta vector			Absolute frequency					Delta vector			Absolute frequency				
da*	db*	dL*	Color Center					da*	db*	dL*	Color Center				
da*	db*	dL*	R	Y	G	B	Gy	da*	db*	dL*	R	Y	G	B	Gy
-0.87	-0.87	-0.87	19	22	23	24	32	1.5	0	0	10	16	9	20	32
-1.73	-1.73	-1.73	37	38	39	39	40	3	0	0	27	35	27	38	40
0	-1.06	-1.06	25	23	27	24	32	-1.06	0	1.06	24	26	24	27	32
0	-2.12	-2.12	39	39	40	39	40	-2.12	0	2.12	39	39	39	40	40
0.87	-0.87	-0.87	23	20	23	22	32	0	0	1.5	32	32	32	32	32
1.73	-1.73	-1.73	39	38	39	38	40	0	0	3	40	40	40	40	40
-1.06	-1.06	0	7	13	15	17	32	1.06	0	1.06	24	26	24	27	32
-2.12	-2.12	0	23	32	34	36	40	2.12	0	2.12	39	39	39	40	40
0	-1.5	0	15	6	20	10	32	-0.87	0.87	-0.87	24	20	23	22	32
0	-3	0	34	20	38	27	40	-1.73	1.73	-1.73	39	38	39	39	40
1.06	-1.06	0	17	10	16	14	32	0	1.06	-1.06	25	23	27	24	32
2.12	-2.12	0	35	28	35	32	40	0	2.12	-2.12	39	39	40	39	40
-0.87	-0.87	0.87	19	22	23	24	32	0.87	0.87	-0.87	19	22	23	24	32
-1.73	-1.73	1.73	37	38	39	39	40	1.73	1.73	-1.73	37	38	39	39	40
0	-1.06	1.06	25	23	27	24	32	-1.06	1.06	0	17	10	15	14	32
0	-2.12	2.12	39	39	40	39	40	-2.12	2.12	0	36	27	34	34	40
0.87	-0.87	0.87	23	20	23	22	32	0	1.5	0	15	6	20	10	32
1.73	-1.73	1.73	39	38	39	38	40	0	3	0	33	19	38	28	40
-1.06	0	-1.06	24	26	24	27	32	1.06	1.06	0	7	13	16	18	32
-2.12	0	-2.12	39	39	39	40	40	2.12	2.12	0	21	31	35	37	40
0	0	-1.5	32	32	32	32	32	-0.87	0.87	0.87	24	20	23	22	32
0	0	-3	40	40	40	40	40	-1.73	1.73	1.73	39	38	39	39	40
1.06	0	-1.06	24	26	24	27	32	0	1.06	1.06	25	23	27	24	32
2.12	0	-2.12	39	39	39	40	40	0	2.12	2.12	39	39	40	39	40
-1.5	0	0	10	16	9	21	32	0.87	0.87	0.87	19	22	23	24	32
-3	0	0	28	35	25	38	40	1.73	1.73	1.73	37	38	39	39	40

The actual experiment was carried out by generating theoretical samples taken from 10 equal steps in each of the 26 specified directions around each color center. Supposedly, each sample should be judged 100 times by all hypothetical observers.

A slight modification to the model shown in Fig. 2 is needed to check the performance of Bayesian approach with respect to the results of Witt & Döring [14], which were calculated using the logistic psychometric function Eq. (5) and the maximum likelihood method Eq. (9). The modified model is shown in Fig. 4.

```

model;
{
d1 ~ dunif(0.0001,1000)
d2 ~ dunif(0.0001,1000)
d3 ~ dunif(0.0001,1000)
d4 ~ dunif(-500,500)
d5 ~ dunif(-500,500)
d6 ~ dunif(-500,500)
alfa ~ dunif(0.1,5)
g1 <- (d1*d1)
g2 <- (d2*d2) + (d4*d4)
g3 <- (d3*d3) + (d5*d5) + (d6*d6)
g4 <- (d1*d4)
g5 <- (d1*d5)
g6 <- (d4*d5) + (d2*d6)
for( i in 1 : N ) {
dX2[i] <- (g3 * dL[i] * dL[i]) + (g1 * dA[i] * dA[i]) + (g2 * dB[i] * dB[i]) + ((2 * g4) * dA[i]*dB[i]) +
((2 * g5) * dA[i]*dL[i]) + ((2 * g6) * dB[i]*dL[i])
dX[i] <- sqrt(dX2[i])
logit(p[i]) <- (dX[i]) - alfa
r[i] ~ dbin(p[i],nRep)
}
}

```

Fig. 4. The BUGS model using formal language for the color-difference model, which includes Cholesky decomposition and the logistic psychometric function.

3. Results

The Bayesian model and theoretical data sets were implemented using OpenBUGS and Matlab's 'slicesample' function. The results obtained from each were very similar. Processed results from generated Markov Chains using OpenBUGS are shown in Table 2. Table 2 shows the 50% ellipsoid coefficients for Robertson's color centers, as predicted by the Bayesian model and computed using ΔE_{94} . The coefficients g'_j shown are the 50% probability ellipsoid coefficients, and were calculated by dividing the g_j coefficients of Eq. (17) by $\log(2)$. As can be seen, the OpenBUGS results are very similar to the theoretical ones. The calculated OpenBUGS g'_j coefficients differ from the theoretical ΔE_{94} coefficients in the fourth decimal position only. Hue angle differences are smaller than a tenth of a degree and the principal axes of the ellipsoids differ by less than a hundredth from the theoretical value.

Table 2. Ellipsoid coefficients pertaining to a 50% probability: theoretical values versus those predicted by the Bayesian model. Theoretical data obtained from 26 vector directions and 10 steps on each direction, deltaE94 color difference formula, 100 theoretical observers judging each pair and Gauss psychometric function. $g'_j = g_j / \log(2)$, to obtain the 50% probability ellipsoid coefficients of Eq. (17), where a, b, c are the principal semi-axis lengths of the ellipsoid and θ is the angle between the projection of the major axis of the ellipsoid with respect to the positive semi-axis a^* . The angle θ of the dE94 ellipsoid corresponds with the hue angle (h_{ab}^*) of the color center.

Color	[L* a* b*]	Model	g'_1	g'_2	g'_3	g'_4	g'_5	g'_6	a	b	c	θ (deg)
Red	[47 37,23]	dE94	0.1842	0.2956	1.0000	-0.1128	0.0000	0.0000	2.96	1.65	1.00	31.87
		Bayesian	0.1843	0.2948	0.9979	-0.1125	0.0005	0.0001	2.96	1.65	1.00	31.92
Yellow	[87 -7 47]	dE94	0.3357	0.1067	1.0000	0.0349	0.0000	0.0000	3.14	1.71	1.00	98.47
		Bayesian	0.3357	0.1069	0.9975	0.0351	0.0000	0.0000	3.13	1.71	1.00	98.52
Green	[56 -32 0]	dE94	0.1680	0.4565	1.0000	0.0000	0.0000	0.0000	2.44	1.48	1.00	180.0 0
		Bayesian	0.1676	0.4553	0.9942	0.0000	0.0000	0.0000	2.44	1.48	1.00	180.0 0
Blue	[36,5-30]	dE94	0.4636	0.1862	1.0000	0.0476	0.0000	0.0000	2.37	1.46	1.00	279.4 6
		Bayesian	0.4624	0.1868	1.0021	0.0477	0.0000	0.0000	2.37	1.46	1.00	279.5 4
Gray	[62 0 0]	dE94	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	1.00	1.00	1.00	0.00
		Bayesian	1.0009	1.0012	1.0009	0.0000	0.0000	0.0000	1.01	1.00	1.00	0.00

Results computed by OpenBUGS from simulated chains of theoretical red-color centers are shown in Table 3. Moreover, shown are the mean, median, standard deviation, Monte Carlo error (MC_error) and 2.5 and 97.5 percentiles computed from the simulated chains of the marginal posterior distribution of the model's parameters. Monte Carlo errors were determined by calculating the standard deviation of the mean of chain batches of a determined size that was controlled by the user (100 in Table 3). MC errors must be low in order to calculate parameters with increased precision.

Table 3. Results from the red-color center using the OpenBUGS model and theoretical data shown in Table 2, which are equivalent to a color-difference experiment with 232 pairs and 100 observers judging once each pair. Sd is the standard deviation, MC_error is the Monte Carlo error, and 2.5 pc and 97.5 pc are the 2.5 and 97.5 percentiles. The sample indicates the number of chain samples that were considered to simulate the posterior distributions. Start indicates the first sample considered after the burning period needed to obtain convergence of the distribution.

	Mean	Sd	Mc_error	2.5 pc	Median	97.5 pc	Start	Sample
g1	0.1277	0.0049	0.0001	0.1181	0.1277	0.1376	1001	10000
g2	0.2043	0.0061	0.0001	0.1927	0.2043	0.2163	1001	10000
g3	0.6918	0.0114	0.0002	0.6696	0.6918	0.7140	1001	10000
g4	-0.0779	0.0040	0.0001	-0.0858	-0.0779	-0.0700	1001	10000
g5	-0.0003	0.0070	0.0002	-0.0139	-0.0003	0.0132	1001	10000
g6	0.0000	0.0072	0.0001	-0.0140	-0.0001	0.0142	1001	10000

The last two columns in Table 3 show the start sample obtained after the burning period, used to get convergence, and the number of iterations (size) of the simulated chain once convergence was reached. As can be seen, all of the samples have been obtained after a burning period of 1000 iterations. All chains have a sample size of 10,000 elements.

Markov chain trace plots of the estimated parameters are shown in Fig. 5. As can be seen, convergence and mixing were achieved after a burning period of 1000 iterations (burning period not shown in the plot). The corresponding density (histogram kernel smoothed) plots are shown in Fig. 6.

Convergence stability can be observed in Figs. 7 & 8. The parameter autocorrelation function (ACF) converged very quickly. For a lag greater than 4, ACF is nearly zero and for a very few iterations the moving average over 100 iterations was very stable.

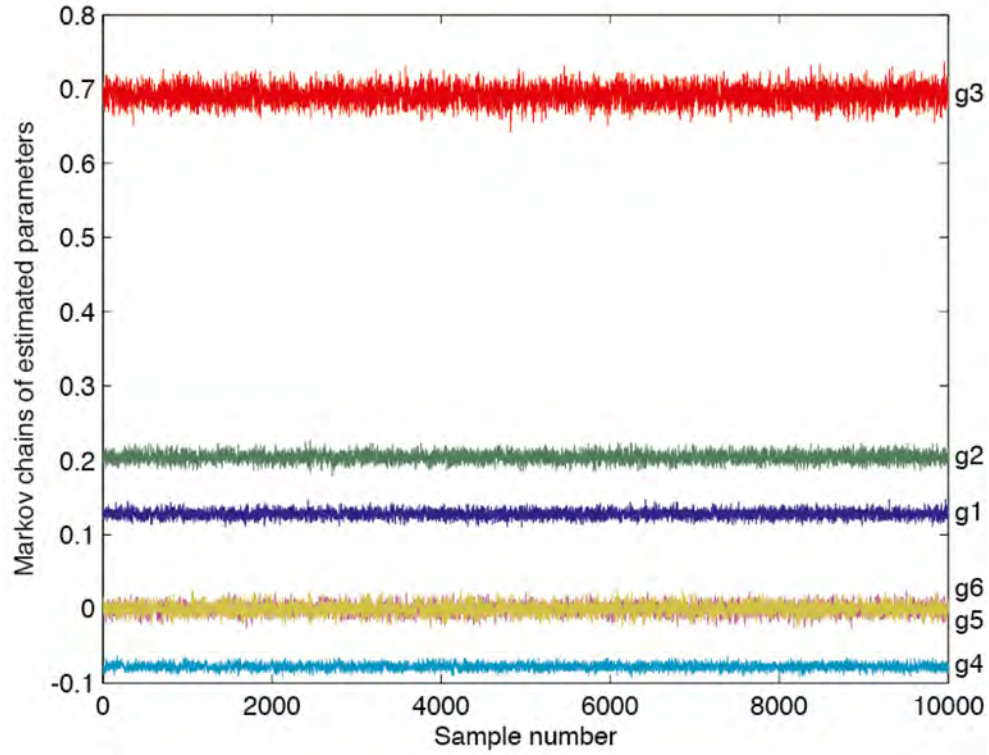


Fig. 5. Trace plots of the posterior parameter distributions after a burning period of 1000 iterations.

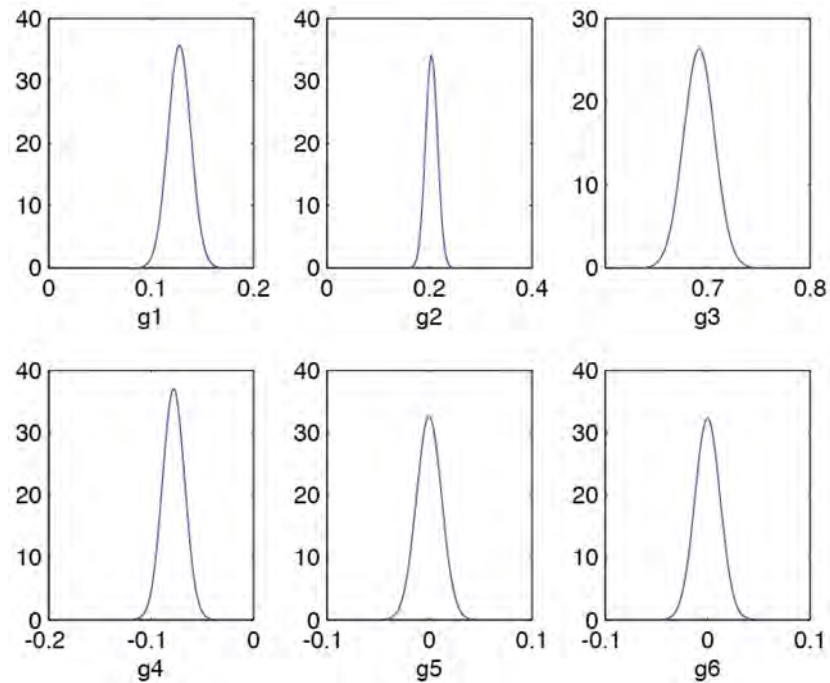


Fig. 6. Density plots of the estimated marginal posterior distributions of the model parameters.

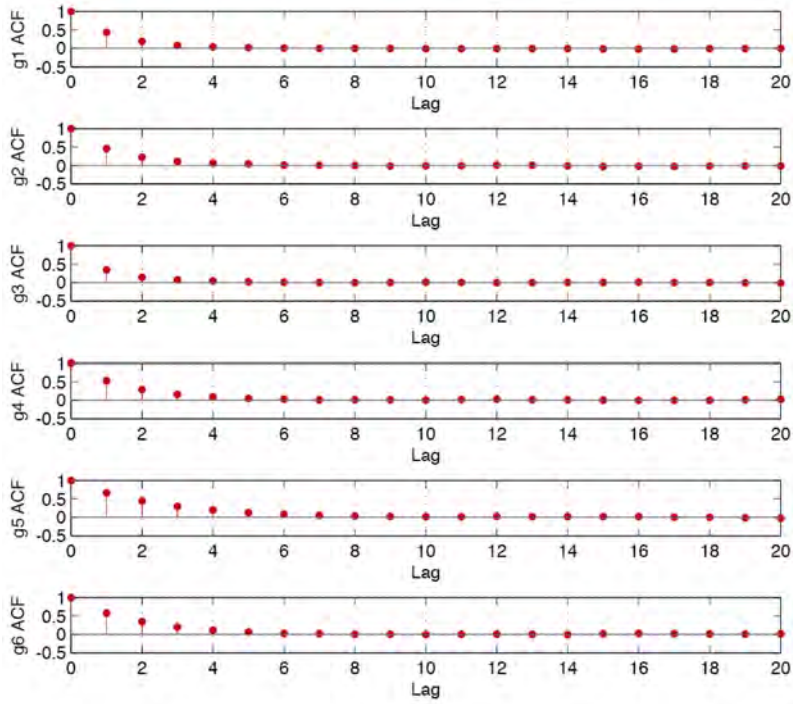


Fig. 7. Sample autocorrelation function (ACF) values show small values as lag increases.

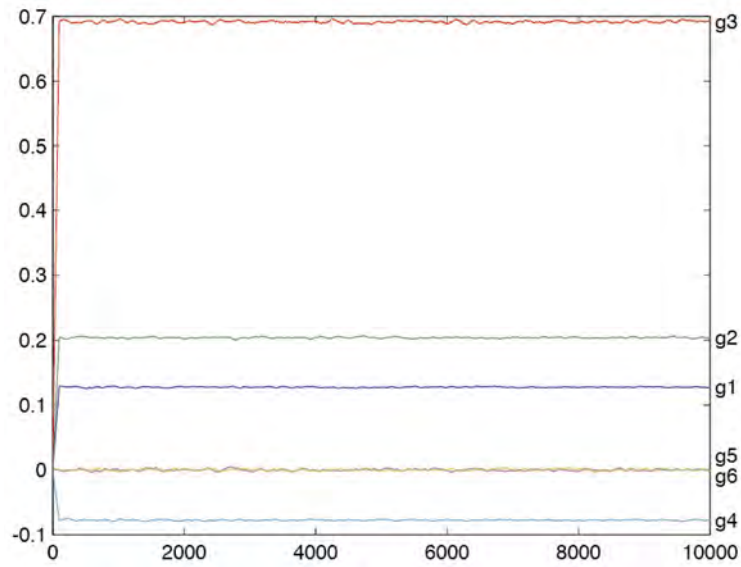


Fig. 8. A moving average over 100 iterations shows that convergence is reached after a few iterations.

Results corresponding to the series II data set from Witt & Döring [14] for the average group of observers are shown in Table 4.

Table 4. Comparing the BUGS results for data Series 2 in xyY color space to those of Witt & Döring [14]. a, b, c are principal semi-axis lengths of the ellipsoid. θ is the angle between the projection of the major axis of the ellipsoid with respect to the positive semi-axis x, and ϕ is the angle between the major axis of the ellipsoid with respect to the horizontal (xy) plane.

Color Center	Green		
Color difference method	Anchor Pair/S 2		
Color space/Coordinates	xyY = [0.2514 0.36382 24.3]		
Source	Witt	BUGS	2σ (BUGS)
g1 (10^{-6})	0.5351	0.5543	0.09
g2 (10^{-6})	0.4279	0.4773	0.08
g3	8.542	8.7660	1.20
g4 (10^{-6})	0.13419	0.1512	0.06
g5 (10^{-3})	0.4126	0.4077	0.24
g6 (10^{-3})	-0.883	-0.1287	0.23
a (10^3)	1.72	1.72	0.15
b (10^3)	1.26	1.23	0.09
c	0.3417	0.34	0.02
θ (deg)	126	130	11
ϕ (deg)	2.1967	2.4393	1.5

Results previously translated to CIELab space, are also shown in Table 5.

Table 5. Comparing the BUGS results for data Series 2 in CIELab color space to those of Witt & Döring [14]. a, b, c are principal semi-axis lengths of the ellipsoid. θ is the angle between the projection of the major axis of the ellipsoid with respect to the positive semi-axis a, and ϕ is the angle between the major axis of the ellipsoid with respect to the horizontal (ab) plane.

Color Center	Green			
Color difference method	Anchor Pair/S 2			
Color space/Coordinates	Lab = [56.4 -31.4 1.2]			
Source	Witt*	BUGS	2σ (BUGS)	Witt**
g1	2.05	2.1125	0.39	2.11
g2	10.03	10.9791	1.54	10.95
g3	9.81	10.1196	1.36	10.10
g4	1.72	1.7309	0.59	1.72
g5	1.78	1.8511	0.57	1.85
g6	1.25	1.0670	1.07	1.08
a	0.8394	0.8236	0.08	0.82
b	0.3397	0.3258	0.02	0.33
c	0.2911	0.29	0.02	0.29
θ (deg)	170.238	170.9051	3	170.93
ϕ (deg)	10.44	10.8502	3.0	10.90

Two methods of calculation were used to translate the Witt ellipsoidal data from xyY to Lab space. Coefficients calculated using the first method, whose results are shown in Table 4 in column labeled 'Witt*', were calculated using a direct translation of the quadratic form xyY space to Lab space, by means of the corresponding transformation matrices, which can be constant for small changes in the coordinates around the color center. Coefficients calculated via the second method, whose results are shown in column labeled 'Witt**' from Table 5, were calculated using the maximum log-likelihood method for the xyY data points translated to Lab space.

As we can see, in the case of the theoretical data, the results agree quite well with the objective values, taking into account the dispersion of the posterior parameter distribution.

The dispersion observed in the theoretically perfect fit corresponds to two causes:

a) Rounding errors in the observed frequencies because the binomial distribution is discrete and not continuous. In fact, when repeating the same experiment using the theoretical samples, the uncertainties of the coefficients were reduced as the number of repetitions increased, as is expected to occur in the frequentist approach.

b) Although the experiment was conducted using continuous samples that were perfectly adjusted to the psychometric curve, the results should have some uncertainty. This is stated by the initial hypothesis, in that the frequency of acceptance by observers that a color pair matches follows a binomial distribution. We must note that the variance of a binomial distribution, in relative terms, is the product $p*(1-p)$, where p is the expected probability. Thus, a variance at 50% of the expected probability is 0.25 multiplied by the number of repetitions.

Results in Table 4 show a good match between the ellipsoid coefficients calculated using Bayesian approach, and those calculated by Witt & Döring [14] in plane 'xy'. However, less agreement is seen for other coefficients, specifically g_6 , when the 'Y' direction is implied. This can be explained by the differences in the distribution of samples around 'x', 'y', and 'Y' axes.

The orientation of the Bayesian ellipsoid differs from that of Witt & Döring [14] by less than 4 degrees. This difference is inside the posterior credible interval of 98% probability ($\approx 2\sigma = 11^\circ$) of the orientation parameter, as can be seen in Table 4. In the same way, principal axis sizes match quite well.

The same applies when the Witt & Döring [14] data are translated to CIELab space, as can be seen in Table 5. It seems that a smaller range of variation of the data along the 'b' and 'L' axes is present, especially when compared to that of the 'a' axis. This implies that greater values of posterior standard deviations of the corresponding parameters are associated with those axes.

Figure 9 shows that the color cloud of the Witt & Döring [14] S2 data series varies over a wider range along the 'a' axis ($\sigma_a = 0.625$) than along the 'b' axis ($\sigma_b = 0.343$) and the 'L' axis ($\sigma_L = 0.423$), which is opposite to the corresponding posterior standard deviation of the 'g' parameters, as can be observed in Table 5 (column 2σ).

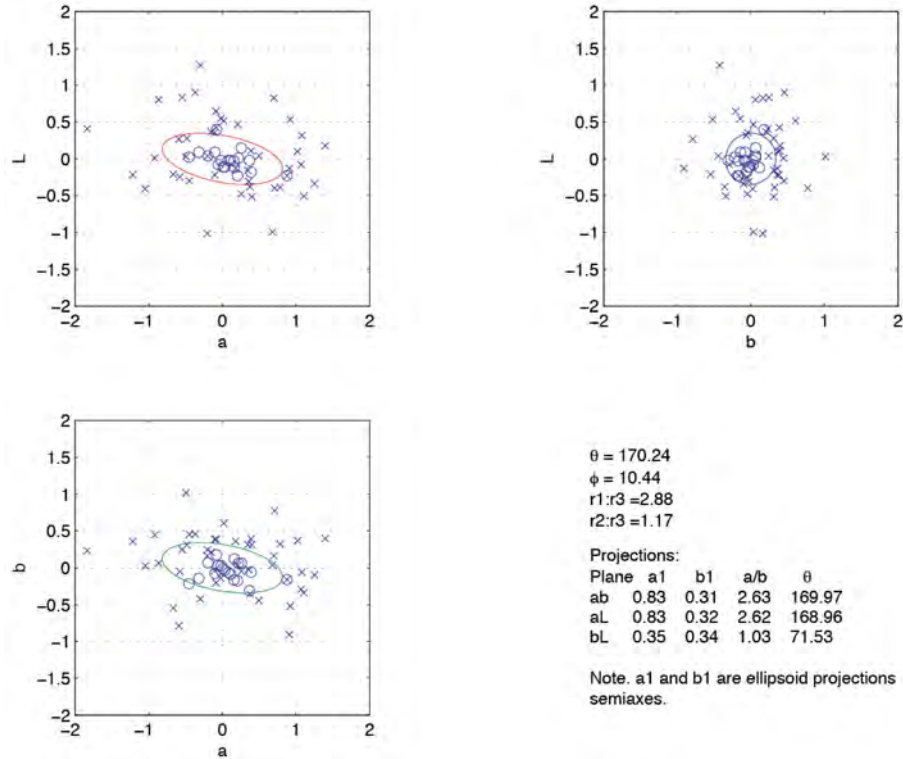


Fig. 9. Witt & Döring [14] S2 data series translated to CIELab space and 50% probability ellipsoid projections into coordinate planes. (o) Considered color pairs that fail the matching test by less than 50% of the judgments. (x) Considered color pairs that fail the matching test by more than 50% of the judgments (repetitions). Standard deviations of the color cloud are: $\sigma_a = 0.625, \sigma_b = 0.343, \sigma_L = 0.423$.

4. Conclusions

The proposed Bayesian approach produces good results with the studied data sets. The mean values of the posterior parameter distributions are consistent with the theoretical values, and those calculated by other authors using different procedures (within the order of accuracy determined by the posterior standard deviation). Different models with different psychometric curves are easily implementable using MATLAB or other more specialized tools such as OpenBUGS.

Convergence is reached very rapidly, before 1000 iterations, using Cholesky decomposition, which always guarantees $\Delta E^2 > 0$ when computed by Eq. (1) if properly selected priors for d_j coefficients are used.

Processing time basically depends on the number of model parameters and the characteristics of the computer and software. OpenBUGS processing is significantly faster than MATLAB; MATLAB computing times can be 30% larger and convergence problems may arise, which usually depend on the configuration for generating Markov chains. OpenBUGS processing times can range from 5 to 20 seconds per 1000 iterations.

The accuracy of parameter estimation depends mainly, as for the frequentist models, on the number of repetitions for which each pair of visual stimuli is judged, and less dependent on the number of the point cloud around the color center.

The advantage of using the Bayesian approach is that the interpretation of the results in relation to the level of confidence and the precision is much more intuitive. Bayesian models

allow one to incorporate previous knowledge regarding the parameters by means of priors, which is not possible in frequentist models.

The influence of low-informative priors on posterior distributions is small for a modest high number of repetitions (around 40), similar to that needed by frequentists to obtain relevant confidence intervals of the estimated parameters.

Another advantage of the Bayesian approach is that the uncertainty of the model can be stated separately via the posterior marginal parameters distribution, which can tell us how to modify data sets to improve estimations.

Determination of confidence intervals using the frequentist approach requires harder simplifying assumptions than those needed by Bayesians. Hypothesis tests are more difficult to be stated and to be justified, many of which are based on hypothetically known distributions that are met when the sample size is very large, which is not required in the Bayesian approach.

In general, Bayesian statistics are able to estimate more complex models than classical approaches.

Acknowledgments

The authors wish to thank Dr. Agustín Blasco, Professor of Animal Breeding and Genetics at Universitat Politècnica de València, for his teaching and advice in the field of Bayesian statistics.