The importance of cross-validation, accuracy, and precision for measuring plumage color: A comment on Vaquero-Alba et al. (2016)

Author(s): Joanna K. Hubbard , Amanda K. Hund , Iris I. Levin , Kevin J. McGraw , Matthew R. Wilkins , and Rebecca J. Safran

Source: The Auk, 134(1):34-38.
Published By: American Ornithological Society
DOI: http://dx.doi.org/10.1642/AUK-16-99.1
URL: http://www.bioone.org/doi/full/10.1642/AUK-16-99.1

The Auk
Ornithological Advances

AmericanOrnithology.org

Volume 134, 2017, pp. 34–38
DOI: 10.1642/AUK-16-99.1

COMMENTARY

The importance of cross-validation, accuracy, and precision for measuring plumage color: A comment on Vaquero-Alba et al. (2016)

Joanna K. Hubbard,[1]* Amanda K. Hund,[2] Iris I. Levin,[2] Kevin J. McGraw,[3] Matthew R. Wilkins,[1] and Rebecca J. Safran[2]

[1] School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, Nebraska, USA
[2] Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, Colorado, USA
[3] School of Life Sciences Arizona State University, Tempe, Arizona, USA
* Corresponding author: jkhubbard@unl.edu

Submitted May 24, 2016; Accepted July 31, 2016; Published October 19, 2016

ABSTRACT

Vaquero-Alba and colleagues published a study in *The Auk: Ornithological Advances* comparing objective color measurements of plumage taken in the field directly on a bird's body to those taken in the lab on collected feathers arranged to emulate the appearance of a bird's natural plumage. Although the field measures of plumage color were less repeatable than lab measures, the authors concluded that measurements taken in the field were more representative of a bird's "true color." Accordingly, they recommend that researchers should bring spectrophotometers into the field to measure color on live birds. We question the assumption that their field measurements represent true color and highlight concerns regarding their experimental design and methodology. Because they did not measure color of live birds in the lab or the color of plucked feathers in the field, they cannot directly test whether the assessment of color in the field on a live bird is superior. Also, rather than assume field measures are the most accurate or precise way to assess plumage color, we suggest cross-validation with other methodologies, such as digital photography, pigment biochemistry, or measures of a known color standard in both environments. Importantly, researchers should be aware of the limitations and advantages of various methods for measuring plumage color so they can use the method most appropriate for their study.

Keywords: plumage coloration, melanin, spectrophotometry, reflectance, Barn Swallow, *Hirundo rustica*

La importancia de la validación cruzada, la exactitud y la precisión al medir el color del plumaje: un comentario sobre Vaquero-Alba et al. (2016)

RESUMEN

Vaquero-Alba y sus colaboradores publicaron un estudio en *The Auk* comparando medidas objetivas del color del plumaje tomadas en el campo directamente en el cuerpo del ave con medidas tomadas en el laboratorio en plumas recolectadas y organizadas para emular la apariencia natural del plumaje. Aunque las medidas de campo del color del plumaje fueron menos repetibles que las de laboratorio, los autores concluyeron que las medidas tomadas en el campo fueron más representativas del "color verdadero" de un ave. En consecuencia, recomendaron que los investigadores deben llevar espectrofotómetros a los sitios de campo para medir el color en aves vivas. Cuestionamos la suposición de que sus mediciones de campo representan el "color verdadero" y resaltamos nuestras preocupaciones con respecto a su diseño experimental y metodología. Debido a que ellos no midieron el color de las aves vivas en el laboratorio ni el color de las plumas sueltas en el campo, no pueden evaluar directamente si la evaluación del color en el campo en un ave viva es superior. También, en vez de asumir que las medidas de campo son la forma más exacta o precisa de determinar el color del plumaje, sugerimos que se haga una validación cruzada con otras metodologías como la fotografía digital, la bioquímica de los pigmentos o las medidas de un estándar de un color conocido en ambos ambientes. Es importante que los investigadores tengan en cuenta las limitaciones y avances en varios métodos para medir el color del plumaje para que puedan usar el método más apropiado para su estudio.

Palabras clave: coloración del plumaje, espectrofotometría, *Hirundo rustica*, melanina, reflectancia

Obtaining objective measures of coloration is central to understanding the control, function, and evolution of plumage color variation. Numerous objective methods for quantifying color exist (e.g., digital photography, spectrophotometry), and it is important to compare the accuracy, precision, and feasibility of these methods so that researchers can make informed decisions when designing their studies and seeking to optimally analyze color differences among individuals or populations (Andersson and Prager 2006, White et al. 2015). In the April issue of

The Auk: Ornithological Advances, Vaquero-Alba et al. (2016) published a study comparing field- and lab-based plumage color measurements of the Barn Swallow (*Hirundo rustica*) taken with an ultraviolet–visible (UV/VIS) spectrophotometer (USB2000, Ocean Optics, Dunedin, Florida, USA). They found that both field- and lab-based methods were highly repeatable within a sample, with lab-based methods being somewhat more repeatable. In contrast to repeatability measures within the same location (field or lab), the individual-level repeatability between the field and lab measures was much lower and often nonsignificant. From these results, Vaquero-Alba et al. (2016) concluded that the lab-based method does not reliably reflect the "true color" of Barn Swallow plumage and that the field-based method is more accurate and realistic.

We agree with the general premise that it is critical to ground-truth various types of color measurements for both precision and accuracy to ensure that the obtained measurements capture the appropriate level of variation. We question, however, the authors' assumption that field measurements taken on the body of a bird represent the true color better than lab measurements taken on a sample of collected feathers. True color implies some knowledge about the accuracy (representation of reality) and precision (degree of variance surrounding the measurement), including how it is perceived by conspecifics, as this is the functional context of this trait (Endler 1990, Endler and Mielke 2005, Stoddard and Prum 2008). Here, we argue that Vaquero-Alba et al. (2016) were not able to infer the accuracy of their measurements and that their estimates of precision actually indicate the superiority of a lab-based measurement of color.

Vaquero-Alba et al. (2016) used a sexual selection framework to explore Barn Swallow color; although there is no empirical evidence that ventral coloration is the target of sexual selection in the European Barn Swallow subspecies (*H. r. erythrogaster*; Scordato and Safran 2014), in this framework, true color can be defined by how it is perceived by potential mates. Further, they only provided data on how much light is directly reflected by the feather surface (i.e. surface reflectance), which alone does not provide reasonable insight into how a colorful signal is perceived by conspecifics. Such insights would require knowledge of environmental light (e.g., illumination and background scene against which the trait is perceived) and receiver visual sensitivity (Endler and Basolo 1998) or evidence of behavioral responses to color variation by the receiver (Bennett et al. 1994). Although the authors quantified color variables using an analytical method that accounts for oscine visual sensitivity (Endler and Mielke 2005), information on environmental light was neither captured nor incorporated into their spectral calculations. Therefore, Vaquero-Alba et al.'s (2016) results cannot demonstrate whether field-based measures better capture true color variation as it might be perceived by potential mates.

An alternative interpretation of Vaquero-Alba et al.'s (2016) use of the term true color could be that measurements taken on a bird better capture the within-individual variation or color heterogeneity. However, color estimates for a variable patch from spectral measurements would require systematic sampling throughout the patch, and such a sampling scheme would sacrifice precision, as demonstrated by the first year of field-based measures presented in Vaquero-Alba et al. (2016). Although a single average color measure can be calculated from these point samples, it does not provide a more accurate representation of true color. Rather, researchers could use a measure of within-individual variation or digital photography, which can capture red-green-blue (RGB) color values for the entire patch as well as measure patch heterogeneity (Oh and Badyaev 2010, Vortman 2013, Troscianko and Stevens 2015), and UV-sensitive cameras can capture variation in the UV range (Stoddard et al. 2012, Troscianko and Stevens 2015).

Skin and the arrangement of feathers on a bird's body can add to color heterogeneity and influence spectral measurements such that field-based measures may be a more accurate representation of color; however, Vaquero-Alba et al. (2016) did not present convincing evidence to support this explanation for differences among their methods. Critically, differences in the conditions under which Vaquero-Alba et al.'s (2016) field and lab measurements were taken could account for the low repeatability between methods. To definitively test whether or not taking measurements in the field on a bird's body differs from measurements in the lab on collected feathers, a more thorough experiment is needed. We suggest a full factorial design in which reflectance is measured in 4 conditions: (1) on the bird in the lab, (2) on the bird in the field, (3) on a plumage sample in the lab, and (4) on a plumage sample in the field. Only conditions 2 and 3 were tested by Vaquero-Alba et al. (2016); consequently, the lack of repeatability between methods could be the result of one method more accurately capturing plumage color, or it could be the result of environmental differences in which field and lab measurements were taken. Therefore, we argue there is no basis to infer that spectral field measurements are a better representation of true color than lab measurements, and, by making this assumption, Vaquero-Alba et al. (2016) limited their possible conclusions to lab measurements being either equal or inferior relative to field measurements.

Because most studies of animal coloration focus on among-individual or among-population differences, it is important to mitigate the effects of these environmental factors to ensure the precision of color measurements and accuracy of detected differences. Precise, standardized measurements are more feasible in the lab because researchers have better control of the environment, thus reducing possible variation due to temperature (Ocean Optics 2012) and light (Park et al. 2006, Sarafianou et al. 2012), which are known to influence spectral equipment and impact measurements. Moreover, plumage color measurements taken on a bird's body may introduce noise caused by movement of live animals during measurement or feather ruffling between measurements; additionally, the curvature of a bird's body may prevent probe holders or non-reflective sleeves from adequately blocking out ambient light. The generalized linear mixed model (GLMM) repeatability estimates reported by Vaquero-Alba et al. (2016) were higher for lab measurements compared to field measurements for 17 of 20 color variables, demonstrating that lab measurements are indeed more precise. Despite the more controlled environment of a lab, ambient light, if not completely eliminated, can affect the accuracy of reflectance measurements taken with a spectrophotometer. Consequently, we encourage researchers working in either the field or the lab to reduce the effects of environmental light as much as possible because devices designed to block out light, such as probe holders and non-reflective sleeves, can be imperfect (KJM, personal observation). Blocking light can be achieved by taking spectral measurements in a consistent light environment (e.g., Hubbard et al. 2015), a dark room (e.g., Siitari et al. 2007, Juola et al. 2008), or by blocking out ambient light by placing a cardboard box around the spectrophotometer and light source (White et al. 2015).

The lack of correspondence between field and lab measures in Vaquero-Alba et al. (2016) may also be explained by several aspects of their sampling and measurement methodology. Vaquero-Alba et al. (2016) reported the collection of 5–10 feathers from each ventral region; however, Quesada and Senar (2006) already demonstrated that carotenoid-based color measurements taken on <10–15 plucked feathers are expected to differ from measurements taken on live birds in the field. Although the spectral properties of carotenoid- and melanin-based colors differ such that fewer feathers may be needed to recapitulate melanin-based color as it appears on a bird, it seems unlikely to be the case for pale plumages such as that of the European Barn Swallow. Second, in both lab and field, spectral measurements were taken with the light source and reflection probe held at a 45° angle to the plumage surface (coincident oblique [CO] geometry). Although this arrangement is recommended for glossy surfaces such as iridescent plumages (Andersson and Prager 2006, Verdes et al. 2015), for pigment-based plumages, such as the melanin-based ventral color of Barn Swallows, the reflectance probe should be held at a 90° angle (coincident normal [CN] geometry; Andersson and Prager 2006). Plumage color measurements taken with CO geometry are known to be more difficult to standardize because there will likely be (1) a larger area of the feather surface illuminated, (2) variation in the rotation angle of the probe, and (3) inconsistent pressure against the plumage between measurements (Andersson and Prager 2006, White et al. 2015). Points 2 and 3 are likely to exaggerate differences between measurements taken in a controlled lab environment and measurements taken on a potentially moving animal in the hand in the field. And finally, we question the decision to use a single spectral scan per measurement because most studies measuring color with spectrophotometers average several scans (e.g., 10–20) per measurement (Siefferman and Hill 2003, Gunderson et al. 2009, Dakin and Montgomerie 2014, Hubbard et al. 2015, White et al. 2015). Use of a single scan per measurement increases the signal-to-noise ratio and may explain the unusual but seemingly significant UV reflectance seen in Vaquero-Alba et al. (2016), which is inconsistent with studies of European (*H. r. rustica;* McGraw et al. 2004, Saino et al. 2013), North American (*H. r. erythrogaster*; McGraw et al. 2005), and Eastern Mediterranean (*H. r. transitiva;* Vortman et al. 2011) Barn Swallows that have demonstrated little to no reflectance in the UV range.

The trait studied by Vaquero-Alba et al. (2016), melanin-based ventral plumage color of the Barn Swallow, has been the focus of numerous studies across several populations of Barn Swallow, including other European populations (e.g., McGraw et al. 2005, Eikenaar et al. 2011, Vortman et al. 2011, Hasegawa and Arai 2013, Saino et al. 2013, Hubbard et al. 2015). Consequently, Barn Swallow color has been measured with various methods, many of which have been cross-validated, as well as compared to the underlying biochemical properties of the plumage. For example, in the North American subspecies, McGraw et al. (2005) compared reflectance measurements taken with a Colortron II reflectance spectrophotometer (Light Source Inc., San Rafael, CA, USA) to melanin pigment concentrations quantified with biochemical techniques. They found a strong correlation between color and total melanin concentration, demonstrating that surface reflectance of North American Barn Swallow plumage (measured from feathers mounted on an index card) indicates the underlying pigment concentration (McGraw et al. 2005). In another validation approach, Vortman et al. (2011) used digital photography to calculate RGB color scores for plumage samples collected from the Eastern Mediterranean Barn Swallow. These scores were highly repeatable when compared to lab-based measurements of the same plumage samples taken with a UV/VIS spectrophotometer. Taken together, these results suggest that surface reflectance measured with a spectrophotometer in a lab setting accurately captures important color variation among individuals, and we argue that validation across methods, such as these examples, are more informative than comparing the same method in different environmental conditions.

Vaquero-Alba et al. (2016) concluded that, because field measurements of melanin-based color in the Barn Swallow capture true color, researchers measuring color should carry a spectrophotometer into the field to take color measurements. Although certain situations require field measurements, for example when temperature affects ectotherm coloration (King et al. 1994) or humidity affects iridescent feathers (Shawkey et al. 2011), additional evidence is needed to fully support this conclusion. Researchers should use a feasible method that allows them to collect high-quality data; our concern is that the firm recommendation of carrying a spectrophotometer to the site made by Vaquero-Alba et al. (2016) might discourage researchers from pursuing particular research questions, when in fact collecting plumage samples is often a viable option. For example, many biologists study natural populations of birds in remote or hard-to-access locations that may or may not have electrical access to run the light source and laptop required for measurements with a spectrophotometer. Moreover, many field studies employ a network of research assistants or sites, which would require multiple expensive pieces of equipment to allow each team to take field measurements. Additionally, the practice of collecting plumage samples, which can be done in addition to field measurements, provides researchers the opportunity to reanalyze samples at later points in time in the case of measurement error, if new methodologies become available, or to analyze additional aspects of the feather (e.g., pigment biochemistry, microstructure). A final consideration is the effect of field measurements on animal handling time; increased handling time can lead to increased stress of the animal, which may result in researchers feeling rushed and being less consistent when taking measurements. These issues can be compounded if the curvature or movement of a bird's body leads to probe placement that allows ambient light to reach the probe, thus introducing additional error into color measurements.

We propose a more balanced recommendation for researchers interested in collecting precise and accurate measures of plumage color, keeping in mind the advantages and limitations of various methods. From both Quesada and Senar (2006) and Vaquero-Alba et al. (2016), it is clear that lab- and field-based methods can each be highly repeatable, but researchers should avoid using a mix of the 2 methods in a single study. Moreover, because lab-based measurements seem more repeatable than field-based methods, we suggest researchers opt for a lab-based method when possible, which contradicts the suggestion made by Vaquero-Alba et al. (2016), although it is supported by their own repeatability results. We also encourage researchers to cross-validate their methods to ensure they are capturing the appropriate aspects of color variation for their specific questions (see Cal et al. 2006 for an example in which methods sometimes disagree). Such cross-validation can be done, as explained above, with biochemical analyses, digital photography, measurement of a known standard, or assessing whether measured differences elicit distinct behavioral responses in receivers. With more studies and improved technology, our understanding of the role of colorful traits in birds and other animals will continue to advance. The overall objective of this commentary is to encourage ornithologists to be aware of the advantages and limitations of different color measurement methodologies and design their studies accordingly.

ACKNOWLEDGMENTS

We would like to thank 2 anonymous reviewers for comments on a previous draft of this manuscript.

Funding statement: AKH was supported by a National Science Foundation Graduate Research Fellowship (No. 1000124290). ILL was supported by a National Science Foundation Postdoctoral Fellowship (No. 1306059). MRW was supported by an SBS POE Postdoctoral Fellowship from the University of Nebraska-Lincoln. None of our funders had any influence on the content of the submitted or published manuscript. None of our funders require approval of the final manuscript to be published.

Author contributions: All authors contributed to the ideas and writing of this paper.

LITERATURE CITED

Andersson, S., and M. Prager (2006). Quantifying colors. In Bird Coloration, volume I: Mechanisms and Measurement (G. E. Hill and K. J. McGraw, Editors). Harvard University Press, Cambridge, MA, USA. pp. 41–89.

Bennett, A. T. D., I. C. Cuthill, and K. J. Norris (1994). Sexual selection and the mismeasure of color. The American Naturalist 144:848–860.

Cal, E., P. Güneri, and T. Kose (2006). Comparison of digital and spectrophotometric measurements of colour shade guides. Journal of Oral Rehabilitation 33:221–228.

Dakin, R., and R. Montgomerie (2014). Deceptive copulation calls attract female visitors to peacock leks. The American Naturalist 183:558–564.

Eikenaar, C., M. Whitham, J. Komdeur, M. van der Velde, and I. T. Moore (2011). Testosterone, plumage colouration and extra-pair paternity in male North-American Barn Swallows. PLOS One 6:e23288. doi:10.1371/journal.pone.0023288

Endler, J. A. (1990). On the measurement and classification of colour in studies of animal colour patterns. Biological Journal of the Linnean Society 41:315–352.

Endler, J. A., and A. L. Basolo (1998). Sensory ecology, receiver biases and sexual selection. Trends in Ecology & Evolution 13: 415–420.

Endler, J. A., and P. W. Mielke (2005). Comparing entire colour patterns as birds see them. Biological Journal of the Linnean Society 86:405–431.

Gunderson, A. R., M. H. Forsyth, and J. P. Swaddle (2009). Evidence that plumage bacteria influence feather coloration and body condition of Eastern Bluebirds *Sialia sialis*. Journal of Avian Biology 40:440–447.

Hasegawa, M., and E. Arai (2013). Differential female access to males with large throat patches in the Asian Barn Swallow *Hirundo rustica gutturalis*. Zoological Science 30:913–918.

Hubbard, J. K., B. R. Jenkins, and R. J. Safran (2015). Quantitative genetics of plumage color: Lifetime effects of early nest environment on a colorful sexual signal. Ecology and Evolution 5:3436–3449.

Juola, F. A., K. McGraw, and D. C. Dearborn (2008). Carotenoids and throat pouch coloration in the Great Frigatebird (*Fregata minor*). Comparative Biochemistry and Physiology B - Biochemistry and Molecular Biology 149:370–377.

King, R. B., S. Hauff, J. B. Phillips, R. B. King, S. Hauff, and J. B. Phillips (1994). Physiological color change in the green treefrog: Responses to background brightness and temperature. Copeia 1994:422–432.

McGraw, K. J., R. J. Safran, M. R. Evans, and K. Wakamatsu (2004). European Barn Swallows use melanin pigments to color their feathers brown. Behavioral Ecology 15:889–891.

McGraw, K. J., R. J. Safran, and K. Wakamatsu (2005). How feather colour reflects its melanin content. Functional Ecology 19: 816–821.

Ocean Optics (2012). Thermally stabalizing your spectrometer with the USB-TC. http://oceanoptics.com/wp-content/uploads/Thermally-Stabilizing-Your-Spectrometer-with-the-USB-TC.pdf

Oh, K. P., and A. V. Badyaev (2010). Structure of social networks in a passerine bird: Consequences for sexual selection and the evolution of mating strategies. The American Naturalist 176:80–89.

Park, J. H., Y. K. Lee, and B. S. Lim (2006). Influence of illuminants on the color distribution of shade guides. Journal of Prosthetic Dentistry 96:402–411.

Quesada, J., and J. C. Senar (2006). Comparing plumage colour measurements obtained directly from live birds and from collected feathers: The case of the Great Tit *Parus major*. Journal of Avian Biology 37:609–616.

Saino, N., M. Romano, D. Rubolini, C. Teplitsky, R. Ambrosini, M. Caprioli, L. Canova, and K. Wakamatsu (2013). Sexual dimorphism in melanin pigmentation, feather coloration and its heritability in the Barn Swallow (*Hirundo rustica*). PLOS One 8:e58024. doi:10.1371/journal.pone.0058024

Sarafianou, A., P. Kamposiora, G. Papavasiliou, and H. Goula (2012). Matching repeatability and interdevice agreement of 2 intraoral spectrophotometers. Journal of Prosthetic Dentistry 107:178–185.

Scordato, E. S. C., and R. J. Safran (2014). Geographic variation in sexual selection and implications for speciation in the Barn Swallow. Avian Research 5:1–13.

Shawkey, M. D., L. D'Alba, J. Wozny, C. Eliason, J. A. H. Koop, and L. Jia (2011). Structural color change following hydration and dehydration of iridescent Mourning Dove (*Zenaida macroura*) feathers. Zoology 114:59–68.

Siefferman, L., and G. E. Hill (2003). Structural and melanin coloration indicate parental effort and reproductive success in male Eastern Bluebirds. Behavioral Ecology 14:855–861.

Siitari, H., R. V Alatalo, P. Halme, K. L. Buchanan, and J. Kilpimaa (2007). Color signals in the Black Grouse (*Tetrao tetrix*): Signal properties and their condition dependency. The American Naturalist 169:S81–S92.

Stoddard, M. C., A. L. Fayet, R. M. Kilner, and C. A. Hinde (2012). Egg speckling patterns do not advertise offspring quality or influence male provisioning in Great Tits. PLOS One 7. doi:10.1371/journal.pone.0040211

Stoddard, M. C., and R. O. Prum (2008). Evolution of avian plumage color in a tetrahedral color space: A phylogenetic analysis of new world buntings. The American Naturalist 171: 755–776.

Troscianko, J., and M. Stevens (2015). Image calibration and analysis toolbox - a free software suite for objectively measuring reflectance, colour and pattern. Methods in Ecology and Evolution 6:1320–1331.

Vaquero-Alba, I., A. McGowan, D. Pincheira-Donoso, M. R. Evans, and S. R. X. Dall (2016). A quantitative analysis of objective feather colour assessment: Measurements in the lab are more reliable than in the field. The Auk: Ornithological Advances 133:325–337.

Verdes, A., W. Cho, M. Hossain, P. L. R. Brennan, D. Hanley, T. Grim, M. E. Hauber, and M. Holford (2015). Nature's palette: Characterization of shared pigments in colorful avian and mollusk shells. PLOS One 10:1–13. doi:10.1371/journal.pone.0143545

Vortman, Y. (2013). The evolution of multiple sexual signals in the East-Mediterranean Barn Swallow (*Hirundo rustica transitiva*). PhD dissertation, Tel Aviv University, Department of Zoology, Tel Aviv, Israel.

Vortman, Y., A. Lotem, R. Dor, I. J. Lovette, and R. J. Safran (2011). The sexual signals of the East-Mediterranean Barn Swallow: A different swallow tale. Behavioral Ecology 22:1344–1352.

White, T. E., R. L. Dalrymple, D. W. A. Noble, J. C. O'Hanlon, D. B. Zurek, and K. D. L. Umbers (2015). Reproducible research in the study of biological coloration. Animal Behaviour 106:51–57.