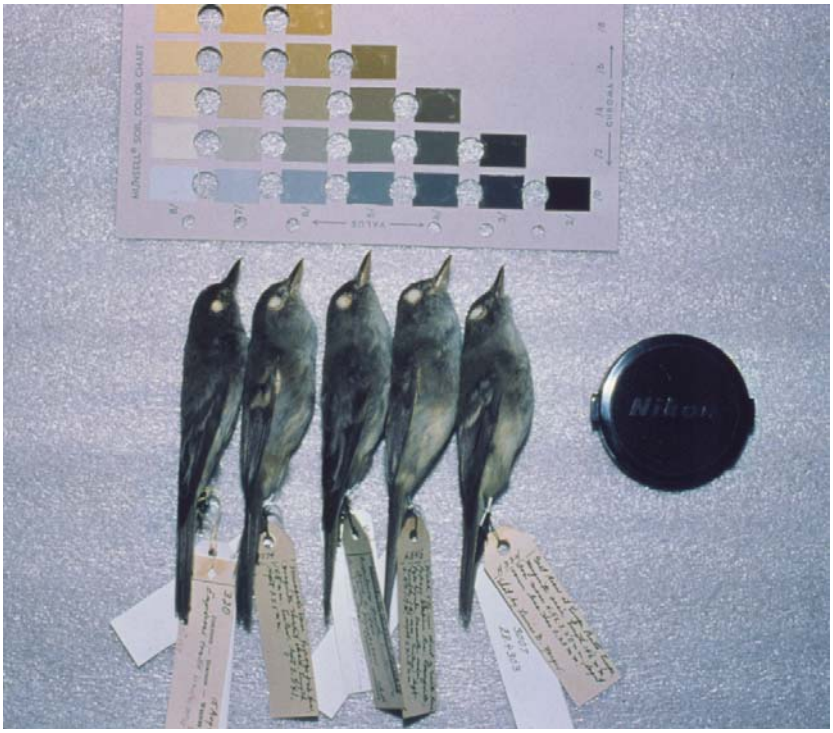
Assessing Variation of Plumage Coloration within the Willow Flycatcher: A Preliminary Analysis



Eben H. Paxton [1], Caroline F. Causey [1], Tom J. Koronkiewicz [2], Mark K. Sogge [1], Matt J. Johnson [1], Mary Anne McLeod [2], Phil Unitt [3], and Mary J. Whitfield [4]

[1]U.S. Geological Survey, Southwest Biological Science Center, Colorado Plateau Research Station, Flagstaff, Arizona
[2]SWCA Environmental Consultants, Inc., Flagstaff, Arizona
[3]San Diego Natural History Museum, San Diego, California
[4]Southern Sierra Research Station, Weldon, California

Recommended citation: Paxton, E.H., C. F. Causey, T. J. Koronkiewicz, M. K. Sogge, M. J. Johnson, M. A. McCloud, P. Unitt, and M. J. Whitfield. 2005. Assessing variation of plumage coloration within the Willow Flycatcher: a preliminary analysis. U.S. Geological Survey Report.

Executive Summary

We present a preliminary analysis evaluating the potential use of a Minolta Colorimeter to quantify plumage coloration variation in the Willow Flycatcher. A colorimeter is a device that measures the color of an object, such as a bird's plumage, and produces a standardized value that can be analyzed statistically. Over the 2004 breeding season, we captured and measured 93 Willow Flycatchers of three subspecies at seven sites, and measured the plumage coloration on the head and the back. Although the resulting dataset was limited in terms of geographic distribution and sample size, preliminary analysis revealed that the colorimeter can detect substantial plumage variation within the Willow Flycatcher subspecies, and significant differences among the subspecies. Furthermore, preliminary modeling suggests colorimeters have the potential to be a powerful tool in assigning subspecies status to individuals of unknown origin (i.e., migrants, wintering flycatchers), but additional sampling is needed before it can be used for this purpose.

Introduction

The Willow Flycatcher (*Empidonax traillii*) is a Neotropical migrant that breeds across much of the conterminous United States and southern Canada, and winters from central Mexico south to northern South America. The Willow Flycatcher is a polytypic species, with four subspecies commonly recognized: *E. t. traillii*, ranging east of the northern Rocky Mountains; *E. t. adastus*, ranging across the northern Rocky Mountains and Great Basin; *E. t. brewsteri*, found west of the Sierra Nevada and Cascade Mountains along the Pacific Slope; and *E. t. extimus*, the Southwestern Willow Flycatcher, which breeds across the southwest (Fig. 1). The division of the flycatcher into subspecies reflects variation in morphology partitioned within specific geographical areas. In particular, variation in plumage coloration is the most reliable method for distinguishing the four subspecies (Unit 1987). However, these plumage coloration differences are subtle (Hubbard 1987, 1999), and identifying the subspecies of a flycatcher requires a skilled taxonomists carefully comparing unknown individuals with voucher specimens from each subspecies.

In 1995 the Southwestern Willow Flycatcher was declared an endangered species (USFWS 1995), increasing interest in the ability to identify Willow Flycatchers to subspecies. Methods for identifying subspecies in a non-lethal manner are needed to help refine the geographic boundary of the endangered subspecies, identify individuals of unknown origin found on the wintering grounds, and identify important migration routes of the southwestern subspecies. To date, research on genetic (Paxton 2000), song (Sedgwick 2001), and biometric (Paxton, unpub. data) variation has confirmed the morphologically-derived subspecies designations, and these approaches show varying levels of promise in assigning individuals of unknown origin to a particular subspecies. However, none of these techniques is a panacea in terms of identifying the subspecies of all individuals of unknown origin, and there is some disagreement among techniques. Therefore, another tool is needed that can either

provide better results or can be used in conjunction with the existing techniques to improve classification accuracy.

A new tool for avian research, the electronic colorimeter, has recently been described as a reliable method for characterizing plumage coloration differences in birds (Figuerola et al. 1999). A colorimeter measures the color and lightness (lightness, saturation, and hue) of an object (e.g., feathers), and provides a standardized value of that color that can be compared across individuals. Figuerola et al. (1999) concluded that the colorimeter was able to discriminate Blue Tit (*Parus caeruleus*) plumage differences by age, sex, and location. More recently, Unitt (unpub. data) found that a colorimeter could differentiate among museum specimens of the Willow Flycatcher's subspecies. Our current study extends Unitt's initial work by evaluating variation in the plumage coloration of living flycatchers measured in the field. Measurements were collected opportunistically in 2004, and the results presented in this report represent a preliminary analysis to assess the colorimeter's performance in the field. In particular, we present an analysis of the variation across three of the four subspecies, explore the degree of differences, evaluate the power of the colorimeter to predict subspecies, and discuss future research needed to evaluate the full potential of this technique.

Methods

Study sites and field methods

In 2004, we captured and measured plumage color of resident breeding flycatchers at six sites within the ranges of three subspecies (Fig. 1, Table 1). In addition, migrants were captured and measured along the Lower Colorado River at Imperial NWR, Arizona (Fig. 1, Table 1).

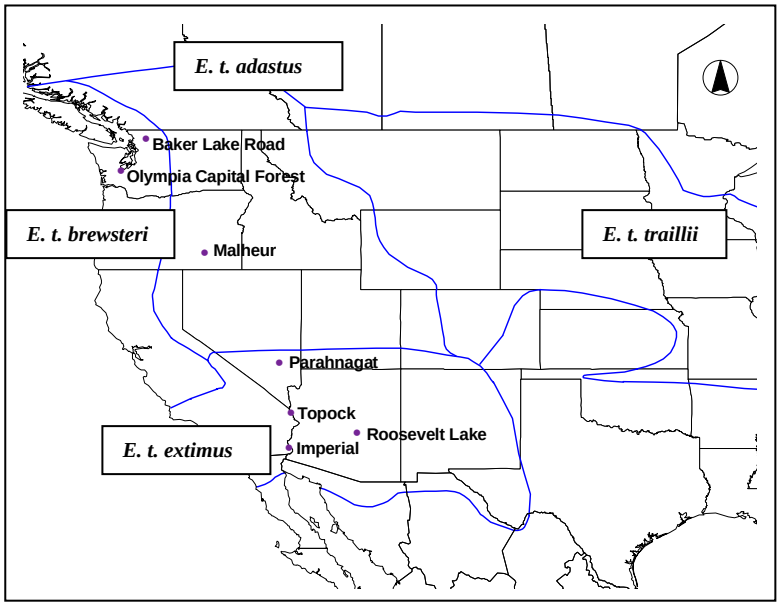


Figure 1: Range of Willow Flycatcher subspecies (blue lines) and location of sites where flycatchers were sampled for this study.

All flycatchers were caught in mist nets, either passively or via target netting, as part of ongoing studies (see Johnson et al. 2002, Koronkiewicz et al. 2005, Newell et al. 2005). With the flycatcher held securely in the hand, the colorimeter was placed firmly against the bird on the crown and a measurement of the plumage coloration recorded. After each measurement, the colorimeter was lifted away from the bird, then placed back onto the same spot for a total of eight independent measurements. This process was repeated to gather eight similar readings from the back (the interscapular region between the wings) of each individual. The entire process, including banding, took less than five minutes.

We used a Konica Minolta Chroma Meter CR-400 colorimeter. This instrument is ideal for field work as the device is easily controlled by one hand (see Fig. 2). The colorimeter measures differences in chromaticity and lightness, which are represented in CIELAB (Commission Internationale de L'Eclairages) 3-dimensional color space and denoted as "L*", "a*", and "b*." A color has three components: lightness, saturation, and hue. The value "L*" denotes how light or dark the color is (lightness), while "a*" and "b*" together indicate color directions (saturation and hue) in two-dimensional space. An increase in a* indicates more red, while a decrease indicates a movement to green; an increase in b* is an increase in yellow, while a decrease in b* indicates a move toward blue.

Figure 2: Measuring a Willow Flycatcher's back color with a Minolta Colorimeter.

Statistical Analysis

Every individual flycatcher had 16 measurements taken: eight replicate measurements of the crown, and eight replicate measurements of the back. We first checked for, and removed, obvious misreadings, which were defined as color values four or more standard deviations from the majority of readings. Misreadings were rare and did not result in the exclusion of any individuals from the analysis. We then chose the five (of eight) color measurements closest to their averaged value; these remaining five readings for each body location (crown and back) were averaged for L*, a*, and b*, and those averages used in the subsequent analysis. Thus, for each flycatcher measured we had six variables: the three color values (L*, a*, B*) for both the crown and back.

Willow Flycatchers molt once (possibly twice) on their wintering grounds (Unitt 1987, Pyle 1997, USGS, unpub. data). Feather fading, primarily from exposure to ultraviolet radiation, changes plumage coloration and lightness over the course of a year. To address whether feather fading and wear had an effect on L*, a*, and b* values, we regressed each variable against measurement date to evaluate whether there was a significant change in color values over time.

We used a MANOVA to test for differences among subspecies, considering all six variables simultaneously, with Bonferroni post-hoc multiple comparison tests to evaluate the level of effects for all pairwise comparisons. Separate ANOVA's were used to evaluate the power of each of the six variables to distinguish the subspecies.

We used a Canonical Discriminant Analysis to represent the linear relationship of the colorimeter variables by subspecies in multivariate space. We also evaluated a non-linear model, Neural Network analysis, as a different approach for predicting subspecies of individual flycatchers. Neural Network Analysis is a powerful method for looking at non-linear, additive relationships, with a high degree of interconnection between values (Mi et al. 2004). We excluded two randomly chosen individuals from each subspecies prior to building the model; these were withheld to later test the predictive model.

We used a standard power analysis to evaluate the sample size needed to detect significant differences between subspecies with alpha = 0.05 and beta > 0.8. To determine the sample size needed to detect most of the plumage coloration variation within a breeding site, we used a resampling (with replacement) simulation program to assess the mean variation captured with a varying sample size of n (n = 3 to 25); as sample size increases, variation should stabilize. The sample size where the variance begins to stabilize was selected as the target sample size for future sampling at breeding sites. The values collected in 2004 from each of the three subspecies were used to evaluate target sample size.

We used SPSS (v. 12) for the MANOVA analysis and JMP (v. 5) for all other analyses. The resampling simulation program was written using the Excel spreadsheet add-in Poptools. Statistical significance was accepted at $P < 0.05$.

Results

Ninety-three adult Willow Flycatchers were captured and measured between May 15 and August 6, 2004 (Table 1). One to three breeding sites per subspecies were sampled, but in general only one site from each subspecies had a relatively large sample size.

Table 1: Summary of flycatchers measured with the Minolta Colorimeter, showing subspecies, site, state, sample size (number of birds) and period that sampling was conducted.

SUBSPECIES	SITE	STATE	SAMPLE SIZE	RANGE OF DATES
E. t. adastus	Malheur NWR	OR	25	June 8 - June 11
E. t. brewsteri	Baker Lake Road	WA	20	June 15 - June 17
	Olympia State Forest	WA	5	June 14
E. t. extimus	Roosevelt Lake	AZ	31	May 21 - August 1
	Topock Marsh	AZ	5	June 23 - August 6
	Pahranagat NWR	NV	3	May 15 - May 16
migrants	Imperial NWR	AZ	4	June 10 - June 11

Feather fading

We measured different flycatchers throughout the entire season at only one site, Roosevelt Lake (Table 1), and evaluated feather fading at this site only. Three color values showed significant change over time: the "a*" values for both the crown and back and the "b*" value for the crown (a* crown: R-squared = 0.456, P<0.001; a* back: R-squared = 0.183, P = 0.016; b* crown: R-squared = 0.151, P = 0.031; Fig. 3). The statistically significant changes are relatively small and the overall color remains essentially the same. However, because of this evidence that the colors change detectably over time, we excluded individuals from all sites measured after July 1 from any further analysis; re-running the regressions with the truncated data indicated non-significant change from May 21 to July 1 at Roosevelt Lake, although the resulting sample size is small.

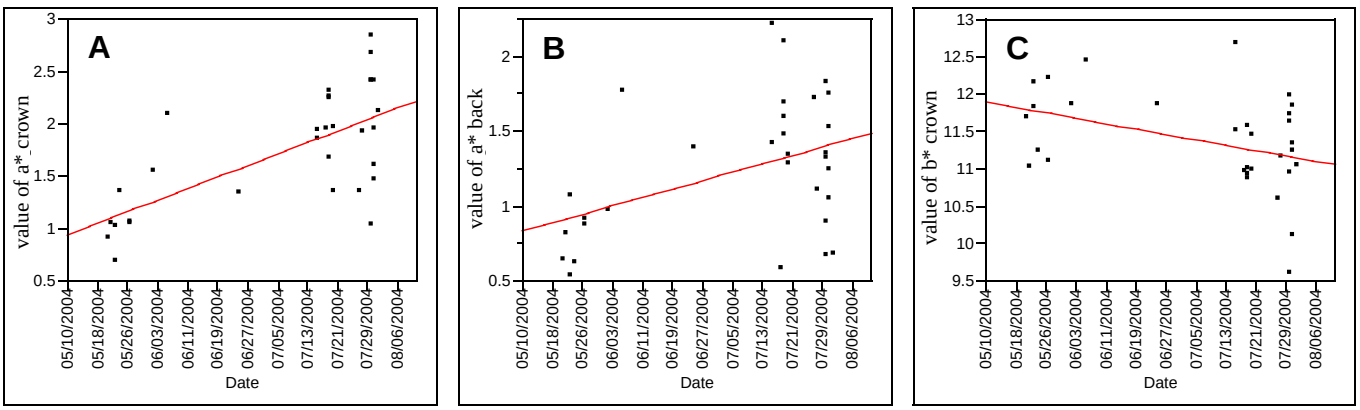


Figure 3: Plumage coloration significantly changes over time for three of the six values measured on Willow Flycatchers at Roosevelt Lake. Thirty-one individuals measured from May 21st to August 1st (X-axis) showed significant change for (A) "a*" on the crown, (B) “a*” on the back, and (C) "b*" on the crown.

Differences among subspecies

Using a MANOVA Pillai's Trace test, we tested for differences among the mean colorimeter values of the three subspecies (Table 2), after assuring normality and homogeneity of variance. The MANOVA indicated large differences among the mean values of the three subspecies ($F_{12,118} = 9.427$, $P<0.001$), and subsequent ANOVA's run on each variable suggest that all but "b* back" are significantly different among the three subspecies (Table 3). Although five of the six variables differed among subspecies, there was considerable overlap (as indicated by the low R-square values; Table 3); in most cases a single variable could not discriminate amongst all three subspecies (Fig. 4). This indicates that the most powerful method for analyzing these data is the simultaneous consideration of all variables via multivariate analyses.

Table 2: Observed colorimeter values for each variable by subspecies, including sample size, mean value, 95% confidence interval, and minimum and maximum values recorded.

Measurement		Subspecies	N *	Mean	95% C. I.	Minimum – Maximum
Crown	L*	*E. t. adastus*	25	26.86	26.09-27.64	23.61-34.36
		E. t. brewsteri	25	24.06	23.28-24.83	21.45-26.11
		E. t. extimus	16	27.27	26.29-28.24	23.74-29.27
	a*	*E. t. adastus*	25	1.07	0.91-1.23	-0.10-1.75
		E. t. brewsteri	25	1.41	1.26-1.57	0.60-1.90
		E. t. extimus	16	1.38	1.18-1.57	0.69-2.83
	b*	*E. t. adastus*	25	10.45	10.02-10.87	8.41-14.71
		E. t. brewsteri	25	10.92	10.50-11.34	9.28-12.60
		E. t. extimus	16	11.47	10.94-12.0	9.61-12.68
Back	L*	*E. t. adastus*	25	31.40	30.75-32.05	27.06-35.64
		E. t. brewsteri	25	29.81	29.16-30.46	26.87-32.12
		E. t. extimus	16	32.74	31.93-33.56	30.28-37.95
	a*	*E. t. adastus*	25	0.80	0.68-0.93	-0.12-1.51
		E. t. brewsteri	25	1.03	0.91-1.16	0.59-1.54
		E. t. extimus	16	1.02	0.87-1.18	0.53-2.38
	b*	*E. t. adastus*	25	13.50	12.90-14.10	8.98-15.73
		E. t. brewsteri	25	13.80	13.20-14.40	11.05-16.41
		E. t. extimus	16	14.37	13.62-15.13	7.75-16.92

* 23 *E. t. extimus* individuals measured after July 1 were excluded from general analysis due to issues of feather pigment fading (see text).

Table 3: R-square and P-values for each variable from ANOVA tests of significance among subspecies

Variable	crown		back	
	R^2	P Value	R^2	P Value
L*	0.37	P < 0.001	0.34	P < 0.001
a*	0.15	P = 0.006	0.12	P = 0.02
b*	0.13	P = 0.013	0.05	P = 0.201

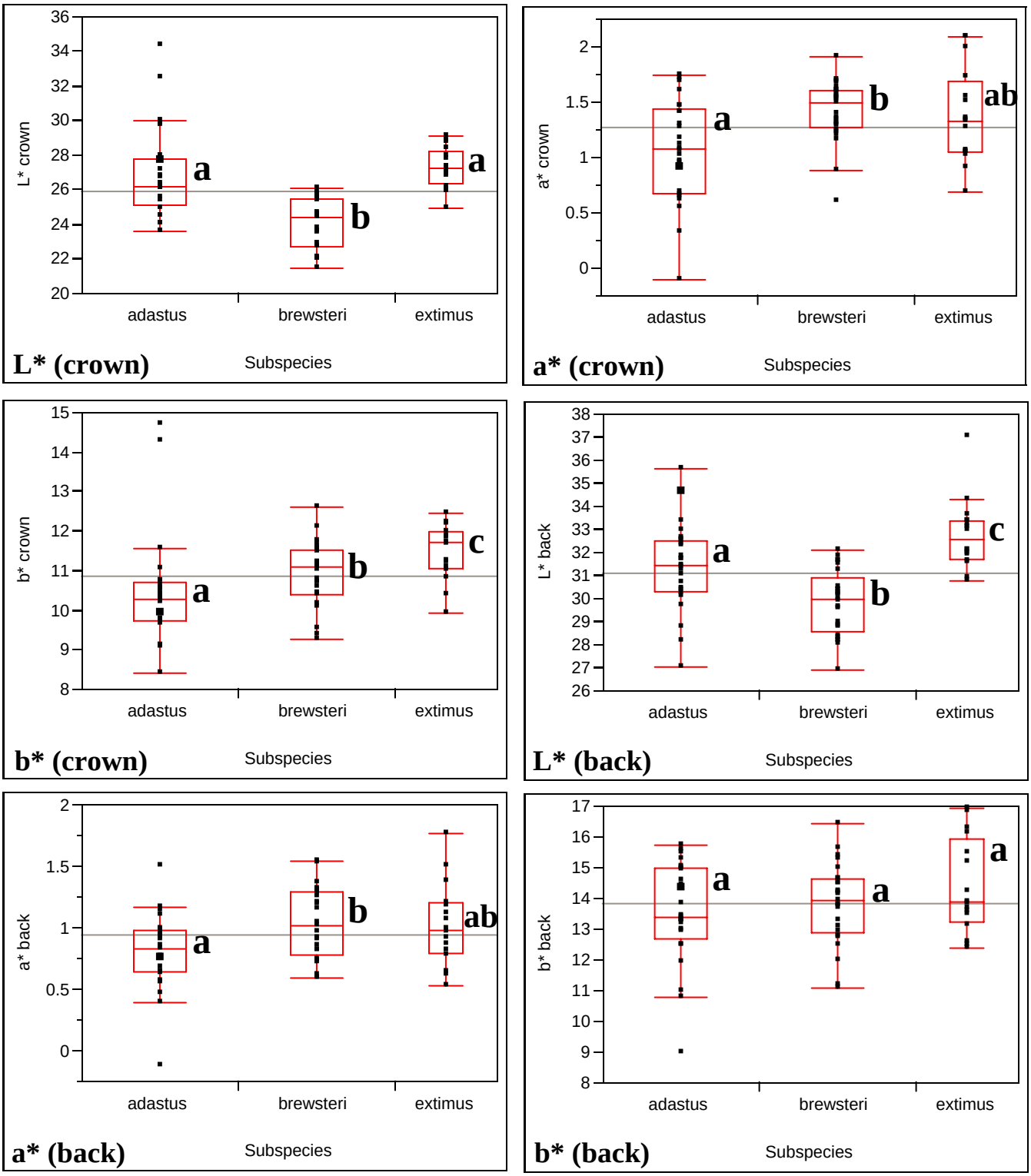


Figure 4: Box plots showing the range of values for each color variable (L*, a*, b*) measured from the crown and back of Willow Flycatchers, grouped by subspecies. We used a Bonferonni post-hoc test to detect differences among the subspecies for each value denoted by the letters. Boxplots for each value with different letters indicate differences between the subspecies at the P<0.05 level.

Predictive models

To examine the power of the colorimeter to predict subspecies status of Willow Flycatchers of unknown subspecies origin, we considered two multivariate models: a linear model (Canonical Discriminant Analysis) and a non-linear model (Neural Network Analysis). The Discriminant Analysis indicated strong separation of the three subspecies (Figure 5), but with a misclassification rate of 17% indicative of overlap among individuals of the subspecies.

Because relationships among variables may not be linear, we assessed the power of a non-linear model, the Neural Network Analysis. We developed a simple four-node Neural Network model based on 91% of the data (excluding two individuals randomly chosen from each of the three subspecies for testing the model). The analysis developed a model with an R-square value of 0.997 and a misclassification rate of zero. However, when we applied the six individuals withheld to test the model,

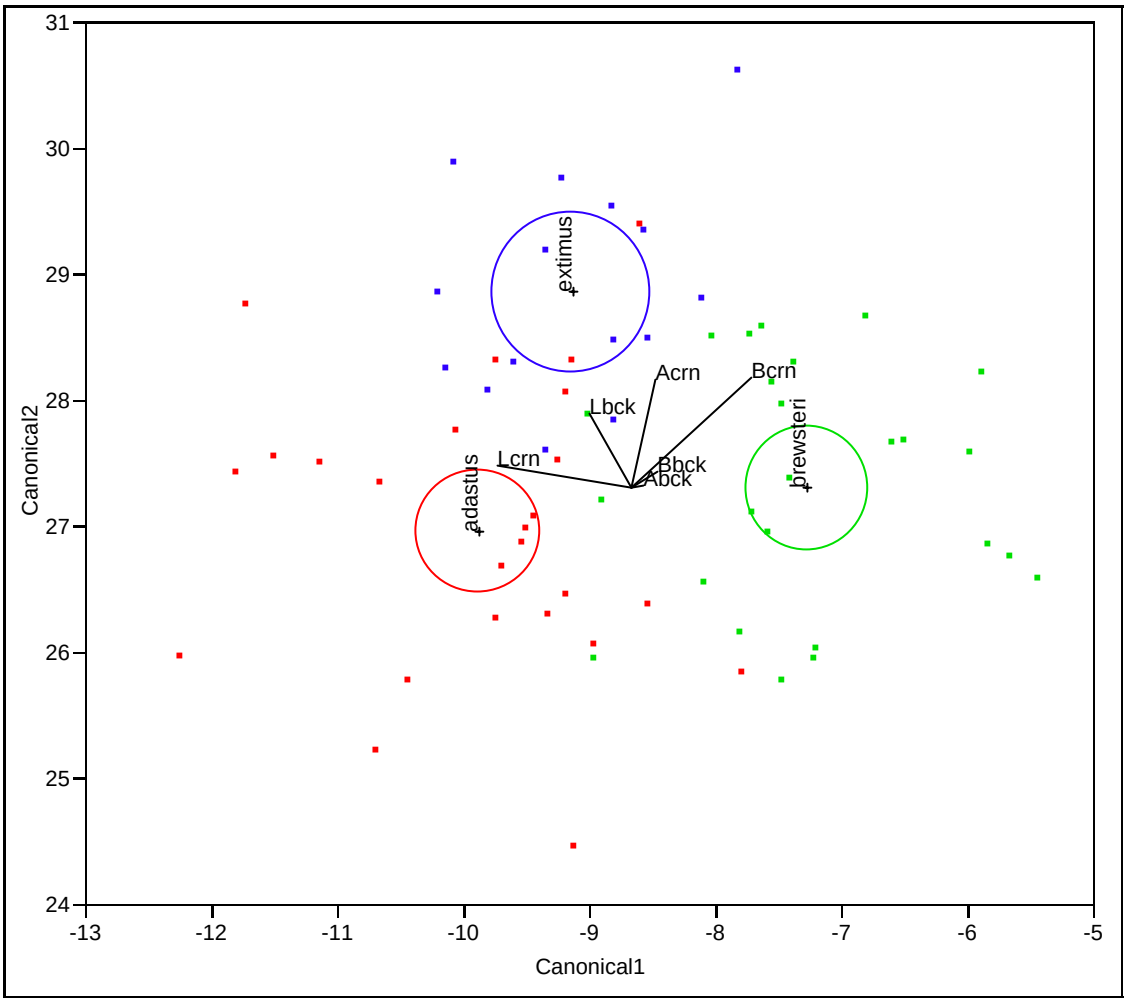


Figure 5: Canonical Discriminant Analysis showing the significant differences among the three subspecies in ordination space. The "+" indicates mean for each subspecies; colored circles represent 95% confidence intervals; dots represent individual flycatchers color coded to represent which subspecies it was sampled from.

one was incorrectly classified, and another could not be placed within a subspecies with high confidence, resulting in a misclassification rate of 33%. This poor performance is not surprising given that the model was built on a very small data set, and the results suggest we could expect a much more robust model with a larger sample size. Although the model requires more extensive development, we ran the values from the four migrants captured at Imperial NWR through the model. The results tentatively suggest that two were from the range of *E. t. adastus*, and two were from the range of *E. t. brewsteri*.

Power analysis

We used the variance in the 2004 colorimeter data to estimate (a) how many flycatchers we need to sample from a breeding site to capture most of the coloration variation present at the site, and (b) how many individuals from each subspecies we need to sample to detect differences in plumage coloration among subspecies. Our resampling simulations suggest that samples of 15-20 individuals captured most of the variation within a site, and future sampling efforts should use this number as the minimum sample size target for breeding sites. To differentiate among subspecies, a power analysis indicated that a sample size of 40 individuals from each subspecies is the maximum size needed to detect significant differences if they exist.

Discussion

Although limited and somewhat opportunistic, our colorimeter sampling of flycatchers in 2004 was adequate to allow us to evaluate the feasibility of the colorimeter to discriminate among subspecies and predict subspecies status of unknown individuals. However, the size and geographic scope of this dataset is far from complete, and more sampling is needed before the technique can be widely employed in flycatcher research, management, and conservation efforts.

Feather Fading

The confounding effects of feather fading may present a limitation in the use of this technique, although those variables that showed significant fade contributed the least to separating the subspecies. One solution is to limit sampling to the beginning of the season, as we did in this study. However, this may not be practical in all cases, and would limit present application of this model to samples collected at other key times of the annual cycle, such as fall migration and wintering grounds. Another option would be to stratify the analysis over the course of the breeding season, but that may entail problematic reductions in sample size. A preferable solution is to develop a correction equation for fading that would be applied to all individuals depending on when they are captured and measured. While this could easily be derived for the Roosevelt Lake population, we do not know whether fading occurs at differing rates for different subspecies/populations. For example, Pacific Northwest populations may be exposed to substantially less ultraviolet radiation than southwestern populations, because of differing weather patterns, and may fade more slowly than southwest populations. On the other hand, more northern-breeding flycatchers are exposed to more ultra-violet radiation due to the longer day lengths in the summer. Ultimately, there is a need to evaluate fading in flycatchers across the breeding range to find a correction for these changes.

Differences among the subspecies

Even with relatively small sample sizes, we found large differences among the three subspecies in the mean values of five of the six colorimetery variables. Although this demonstrates the strength of the technique, additional thorough and carefully planned sampling is needed. One weakness of the current data set is that only one site from each subspecies' range had a sample size larger than five. Thus, we are essentially evaluating the difference among the three large sites to determine subspecies differences, and we therefore cannot clearly distinguish site vs. subspecies effects. For the full utility of this technique to be realized, and for it to be used with confidence, more sampling is needed. Multiple breeding sites from each subspecies need to be sampled so that we can assess how much of the variation is due to individuals, how much is explained by different breeding populations, and how much is the result of differences among the subspecies. We believe there should be at least three populations from each subspecies, specifically targeted to have the largest geographic scope possible to adequately describe within-subspecies variation.

Assigning subspecies to unknown individuals using predictive models

Although we found strong differences among the mean colorimeter values of each subspecies, there is enough overlap in values among subspecies to make building predictive models difficult. Powerful linear models, such as the Canonical Discriminant Analysis that we used, showed strong differences but incorrectly "predicted" 15% of individuals used to build the model. While the result of the Discriminant Analysis is better than random, it does not provide high-confidence accuracy. The accuracy of this Discriminant analysis probably would not improve appreciably with additional samples as the degree of overlap among individuals would stay the same or even increase. However, Discriminant analysis can only evaluate linear relationships among the colorimeter values, and it is not known whether the relationships are linear. A powerful, non-linear alternative model is Neural Networks. Using our simple Neural Network model, most individuals could be assigned to a subspecies. This model was trained on a very small data set, and will become more powerful and accurate with additional data. Unfortunately, Neural Networks are a "black box" analysis, and the relationships that form the predictive model are difficult to determine. For this reason, it is essential that Neural Network models be validated with a robust sub-sample to fully challenge the model. Nonetheless, Neural Networks hold the promise of highly accurate assignment of individuals of unknown origin, and should improve as sample size increases.

Despite the encouraging results that a colorimeter could be used to predict subspecies status of individuals, it is critical to note that this preliminary model building was based on a very small, geographically-limited data set. Many more individuals must be sampled and models refined before colorimetry-based predictions can be made reliably on migrants and wintering birds, and for evaluating subspecies boundaries.

Conclusions and recommendations for future work

Analysis of the 2004 data strongly indicates that the colorimeter is a promising tool for separating subspecies of the Willow Flycatcher. Therefore, we believe that our results warrant an expanded effort to collect the measurements needed to fully evaluate the utility of the colorimeter. Specifically, we recommend the following ***minimum*** sampling effort for 2005:

1) Differences among subspecies – Each subspecies ranges over a large geographical area, so it is likely that there are differences in the plumage coloration within the range of each subspecies. An essential part of evaluating the degree of differences *among* the subspecies is to assess the degree of variation *within* each subspecies. To accomplish this, we recommend obtaining samples from ***at least*** three breeding sites within each subspecies' core range, selected to have the largest geographic spread possible yet avoiding boundary edges. Power analyses suggest that a target of 15-20 individuals from each breeding location (or clusters of geographically-proximate sites) is needed to capture the variance in plumage coloration present at those populations. Such sampling should provide statistical power to detect differences among the subspecies, and possibly population-level differences within a subspecies.

2) Plumage coloration fading – To avoid limiting all sampling efforts to May and June, we need to sample several geographically distant populations over the course of a season to derive a correction factor to account for fading. Because a limited sampling period would greatly limit the utility of this tool, we strongly urge that an effort be made to account for fading. We suggest sampling at least one population from each of three to four regions (Southwest, Pacific Northwest, Great Basin region, eastern U.S.) at least three times within a season (early-May, mid-June, late-July).

3) Predictive models – The utility of a predictive model depends on the data used to construct the model, and colorimetery samples from many flycatcher breeding sites are needed to maximize the applicability of predictive models. Therefore, we emphasize the need for wide geographic sampling so that we have the greatest opportunity to capture the full variation in plumage coloration. If the above sampling suggestions are met or exceeded, we believe that a model could be constructed to predict subspecies status of unknown individuals with reasonable accuracy.

Acknowledgements

This work was funded by the U.S. Geological Survey, U.S. Fish and Wildlife Service, and the U.S. Bureau of Reclamation. The hard work of many members of the SWCA and USGS banding crews made this study possible. Particular thanks to Gary Slater for sampling sites in Oregon and Washington. Thanks to Kimball Garrett and Barbara Kus for helpful comments on the report.

Literature Cited:

Figuerola, J., J.C. Senar, and J. Pascual. 1999. The use of a colorimeter in field studies of Blue tit *Parus caeruleus* coloration. Ardea 87:269-275.

Hubbard, J.P. 1987. The status of the Willow Flycatcher in New Mexico. Report to the New Mexico Department of Game and Fish. 29 pages.

Hubbard, J.P. 1999. A critique of Wang Yong and Finch's field-identifications of Willow Flycatcher subspecies in New Mexico. Wilson Bulletin 11:585-588.

Johnson, M., J. Kelly, S. Langridge, J. Holmes, and M. Whitfield. 2002. Linking breeding and wintering distribution of Southwestern Willow Flycatchers with stable isotopes, 2004. USGS annual progress report to the USFWS. 25 pp.

Koronkiewicz, T.J., M.A. McLoud, B.T. Brown, and S.W. Carothers. 2005. Southwestern Willow Flycatcher surveys, demography, and ecology along the lower Colorado River and tributaries. 2004. Annual report submitted to U.S. Bureau of Reclamation, Boulder City, NV by SWCA Environmental Consultants, Flagstaff, AZ. 125 pp.

Newell, P.J., C. Causey, M. Pollock, E.H. Paxton, and M.K. Sogge. 2005. Survivorship and movements of Southwestern Willow Flycatchers at Roosevelt Lake, Arizona – 2004. U.S. Geological Survey report to the U.S. Bureau of Reclamation, Phoenix.

Paxton, E.H. 2000. Molecular Genetic structuring and demographic history of the Willow Flycatcher (*Empidonax traillii*). Ms thesis, Northern Arizona University.

Pyle, P. 1997. Identification guide to North American Birds. Part 1. Slate Creek Press, Bolinas, CA.. 730 pp.

Sedgwick, J.A. 2001. Geographic variation in the song of Willow Flycatchers: Differentiation between *Empidonax traillii adastus* and *E. t. extimus*. *Auk*, 118:366-379.

Unitt, P. 1987. *Empidonax traillii extimus*: an endangered subspecies. *Western Birds*, 18:137-162.

United States Fish and Wildlife Service (1995) Final rule determining endangered status for the southwestern Willow Flycatcher. *Federal Register*, 60:10694-10715.

.