Poor Reproducibility and Inference in Hydrogen-Stable-Isotope Studies of Avian Movement: A Reply to Wunder et al. (2009)

Authors: Smith, Adam D., Lott, Casey A., Smith, Jeff P., Donohue, Kara C., Wittenberg, Sara, et al.

different labs (initial analysis at the EC lab and repeated analysis at the Colorado Plateau Stable Isotope Laboratory in Flagstaff, Arizona [hereafter "CPSI lab"]), providing ancillary documentation of poor measurement reproducibility between labs. Measurement precision decreased outside the calibration range of keratin standards (greater than –100‰), compared with measurements inside this range (–190‰ to –100‰) (Smith et al. 2009: fig. 2).

In their letter, Wunder et al. (2009) nicely summarize some of the complexities of δD_f analysis, most importantly (1) the lack of internationally accepted reference standards of a material comparable to feathers, (2) the pressing need for additional keratin working standards with high δD values that would more thoroughly bracket the range of natural δD values in bird feathers, and (3) the need for standardized analytical protocols among isotopic laboratories. We agree completely with Wunder et al. (2009) that researchers should be cognizant of these complexities, inform themselves of the analytical protocols used by the laboratory analyzing their samples (e.g., the types and number of standards), and carefully interpret δD_f values outside the keratin standard calibration range.

Despite this common ground, Wunder et al. (2009) make three broad criticisms of Smith et al. (2009) with which we disagree: (1) that we provided analytical detail insufficient for study replication and failed to engage the laboratories that analyzed our samples, (2) that our design failed to appropriately consider two important sources of variation (i.e., intra-feather variation and the presence of δD_f measurements outside the calibration range of keratin standards), and (3) that our results contradict a substantive body of work regarding δD_f measurement error. Here, we hope to clarify the main points of Smith et al. (2009) and respond to Wunder et al.'s (2009) primary criticisms, which do not lead us to alter our original conclusion of poor δD_f measurement reproducibility. We discuss the effect that poor reproducibility has on inference in stable-isotope studies in the context of a recently advanced probabilistic framework for geographic assignment (Wunder and Norris 2008a, b; Wunder 2010). We also suggest some avenues toward a potential solution to the problem of poor reproducibility and encourage future practitioners in this field to more carefully consider this problem when designing studies and interacting with labs. Finally, we advise that researchers more carefully qualify their claims of the value of information that stable-isotope studies provide regarding migratory origins and connectivity at spatial scales relevant to conservation or management, because predicted origins may be biased (Smith et al. 2009) or have low geographic specificity (Meehan et al. 2001, Kelly et al. 2002, Wunder et al. 2009: fig. 1).

Why the reproducibility of keratin standards may not be equivalent to that of feathers.—In Smith et al. (2009), we described measurement error with the metric of reproducibility, the difference between repeated measurements of the same feather when one or more analytical conditions have changed (i.e., independent analysis events). We reported reproducibility for a group of samples with summary statistics where the mean (of differences) described the average systematic shift in δD_f from an initial to a repeated measurement and the standard deviation (of differences) described the variability in the magnitude of this shift (Smith et al. 2009: fig. 1). Poor reproducibility was characterized by considerable systematic error (represented by a large mean) or random

Poor reproducibility and inference in hydrogen-stable-isotope studies of avian movement: A reply to Wunder et al. (2009).—In Smith et al. (2009), we tested the assumption that measurements of hydrogen stable isotope ratios in feather samples (δD_f) are reproducible among independent analysis events in which feathers are equilibrated and analyzed concurrently with keratin standards (Wassenaar and Hobson 2003). For nine independent sample groups of raptor body feathers, we documented poor measurement reproducibility, with systematic error (i.e., bias) of often large magnitude and variable direction, as well as considerable random error (i.e., imprecision) in paired measurements of adjacent subsamples from a single feather. As we reported, for eight of these sample groups, initial and repeated analyses occurred at a single lab (Environment Canada's Stable Isotope Hydrology and Ecology Laboratory in Saskatoon, Saskatchewn; hereafter "EC lab"), providing robust documentation of poor δD_f measurement reproducibility within a lab. A ninth group comprised samples for which initial and repeated analyses occurred at

error (represented by a large SD), or both. Systematic and random errors have different implications for inferences of migratory connectivity that rely on measurements of δD_f. Systematic error shifts the entire spatial distribution of predicted origins, whereas random error reduces the geographic specificity of predictions.

In most work to date, δD_f measurement error has been described by the precision (e.g., SD) of homogenized keratin standards within a single analysis (i.e., repeatability) or asymptotically over time (i.e., reproducibility). However, keratin standards typically are developed from materials that have been homogenized to exhibit high reproducibility. By contrast, nonhomogenized feather samples from wild birds lack this desirable quality. Therefore, the precision of δD measurements from keratin standards likely underestimates that of feather δD measurements. Although standard repeatability and reproducibility represent important metrics for quality assurance and quality control, they do not necessarily describe the reproducibility of nonhomogenized feather materials accurately, which is a separate metric that must be assessed independently. Thus, satisfactory repeatability, or even reproducibility, of standards does not dismiss the poor reproducibility of feather measurements documented in Smith et al. (2009), because geographic assignments are made from measurements of feathers, not standards.

Analytical disclosure and communication.—Wunder et al. (2009) asserted that we failed to provide sufficient methodological detail for study replication. We agree that some ambiguity existed in the analytical details presented in Smith et al. (2009). We suggest that much of this ambiguity resulted from the editorial removal of lab names from Smith et al. (2009), a point on which we also strongly disagreed with *The Auk*'s editors. Although we clearly indicated that samples were analyzed by only two labs, with one lab analyzing most of the samples, we did not explicitly quantify this division. In fact, 95% (402 of 422) of the δD_f measurements in Smith et al. (2009) were completed at the EC lab (including all measurements resummarized below), with the remaining 20 measurements (the repeat analysis from group "NA2") completed at the CPSI lab using the same published protocols and keratin standards as the EC lab, as we were informed by the contributors of those data (see Acknowledgments in Smith et al. 2009). With lab identities and the distribution of samples between labs now disclosed, readers should find sufficient detail for replication in our original manuscript, for two reasons. First, the discussion of reproducibility in Smith et al. (2009) focused nearly exclusively on results from the eight sample groups analyzed only at the EC lab and the comparison between laboratories was only a marginal consideration. Second, because Smith et al. (2009) primarily assessed δD_f measurement reproducibility at the EC lab, our reference to the two publications of Wassenaar and Hobson (2003, 2006) that detail the exact laboratory procedures and three keratin standards used to measure δD_f at this lab accords with Wunder et al.'s (2009) statement that referencing published laboratory techniques is sufficient when the work is carried out by a single lab. Our description of laboratory methods is comparable to such descriptions in recent manuscripts involving the co-authors of Wunder et al.'s (2009) letter that used the EC lab for δD_f measurements (e.g., Hobson et al. 2009, Langin et al. 2009, Paritte and Kelly 2009).

More troubling is Wunder et al.'s (2009) claim that we failed to communicate two important lines of information to the laboratories involved in Smith et al. (2009): that we believed they were producing "questionable" data and that the data were to be used in a publication related to reproducibility. We acknowledge that we did not have contact with CPSI lab personnel, for three reasons: (1) the 20 samples (<5% of all samples) analyzed there represented only a marginal component of our analysis and discussion, (2) data from CPSI were contributed independently of our analyses at the EC lab by an outside party that was in close contact with the CPSI lab, and (3) the CPSI lab has used Wassenaar and Hobson (2003, 2006) as primary references for laboratory protocols (e.g., Paxton et al. 2007). By contrast, however, we communicated regularly, directly, and honestly with the EC lab director, Len Wassenaar, and Keith Hobson, a long-time associate of this lab, about sample preparation, the use of analytical standards, results, reanalyses, and problems with reproducibility. This communication began at the outset of the study, continued through the analysis of data, and included the disclosure of our intent to pursue publication and the results of preliminary analyses indicating poor reproducibility. Given this history, Wunder et al.'s (2009) claim that we failed to interact with laboratory personnel to understand or interpret our results seems disingenuous.

Study design and presentation of results.—Wunder et al. (2009) suggested that the results of Smith et al. (2009) are ambiguous because our study design did not account for (1) systematic δD_f changes along the length of a single feather and (2) imprecise measurement of δD_f outside the calibration range of keratin standards. As pointed out by Smith et al. (2009) and reiterated by Wunder et al. (2009), true replicate measurement of nonhomogenized feather samples is impossible, because feather material is destroyed during analysis. Given this physical reality, differences between replicate measurements of biological samples could result from either measurement error or real biological variation within samples. Thus, biological variation must be accounted for to adequately assess reproducibility. Wunder et al. (2009) claim that we failed to acknowledge that biological intra-feather variation confounds estimates of reproducibility, ignoring our work on this topic (Smith et al. 2008). On the contrary, the complication of intra-feather variation was the preeminent consideration in the discussion of Smith et al. (2009). Below, we provide further clarification that intra-feather variation is minor compared with the poor reproducibility observed in Smith et al. (2009).

Wunder et al. (2009) incorrectly contend that we defined poor reproducibility as the widening pattern of residuals outside the calibration range of keratin standards (Smith et al. 2009: fig. 2). We agree with Wunder et al. (2009) that not distinguishing between samples inside and outside the calibration range was an oversight on our part. However, we clearly identified poor reproducibility as the considerable systematic and random error present in our entire data set (Smith et al. 2009: fig. 1), and not simply the decrease in precision we observed outside the calibration range of keratin standards (Smith et al. 2009: fig. 2). The decrease in measurement precision outside the calibration range was a secondary result and an unsurprising consequence of applying a normalizing equation to δD_f values outside the range of values on which the calibration regression was based (i.e., –190‰ to –100‰). Likewise, our suggestion to expand the isotopic range of keratin standards to include more positive values was an obvious solution to the problem, although doing so is not a trivial task, as Wunder

et al. (2009) explicate. More importantly, expanding the isotopic range of keratin standards would decrease only the random error of δD_f measurements currently outside the calibration range to the relatively imprecise levels observed inside the calibration range; it would have no effect on the larger problem of systematic error, which occurred both inside and outside the keratin standard calibration range.

In a previous publication (Smith et al. 2008), we estimated the biological magnitude of intra-feather variation for the three most common species represented in Smith et al. (2009), independent of the confounding effect of measurement reproducibility. Specifically, all samples in the previous study were run in a continuous laboratory-analysis event, with the additional safeguard of random interspersion of samples. Smith et al. (2008) documented consistent differences in δD_f between adjacent longitudinal subsamples of body feathers, the magnitude of which varied to some extent among species. Feathers of Merlins (*Falco columbarius*) and Sharp-shinned Hawks (*Accipiter striatus*) showed similar differences, with more negative δD_f values in distal feather material than in proximal feather material (least squares mean ± SE, combined for the two species: –9.68 ± 1.08‰; $n = 29$). The same pattern appeared in Red-tailed Hawk (*Buteo jamaicensis*) feathers, but the difference was less pronounced (–3.00 ± 1.41‰; $n = 17$). The magnitude and direction of these differences serve as an expectation against which reproducibility can be assessed for samples from these three species in Smith et al.'s (2009) data set. That is, if intrafeather variation in δD_f were driving the poor reproducibility observed in the δD_f measurement of equivalent feather subsamples in Smith et al. (2009), differences in repeated δD_f measurements in these three species should average near zero once adjusted for the intra-feather variation described above. This is not the case (Fig. 1), which suggests that some factor other than the magnitude of intra-feather variation observed in raptor body feathers is responsible for poor reproducibility. To compare reproducibility inside and outside the calibration range of keratin standards, we simply distinguished, also in Figure 1, between samples that were inside (i.e., average δD_f of repeated measurements greater than or equal to –100‰) and outside (i.e., average δD_f of repeated measurements greater than –100‰) this range. In general, poor reproducibility of nonhomogenized feather material existed whether the analysis occurred inside or outside the calibration range (Fig. 1). Systematic error was present and often severe regardless of how the data were summarized, with mean adjusted differences between original and repeated analyses ranging from –6.14‰ to 16.78‰ within the calibration range and from –8.93‰ to 37.41‰ outside of the calibration range. Random error was larger outside the calibration range (Fig. 1), as we noted previously (Smith et al. 2009: fig. 2).

Ignoring a substantive body of work.—Wunder et al. (2009) claim that our work ignores a substantive body of work on δD_f measurement error. We disagree, for there is currently a paucity of literature concerning the reproducibility of nonhomogenized feather material. Certainly, the repeatability and reproducibility of homogenized keratin standards have been well reported (Wunder and Norris 2008b, Wunder et al. 2009), but the extent to which nonhomogenized δD_f measurements exhibit this same reproducibility remains largely untested outside of Smith et al. (2009). Intra-feather variation within a single analysis, which should reflect real biological variation in nonhomogenized
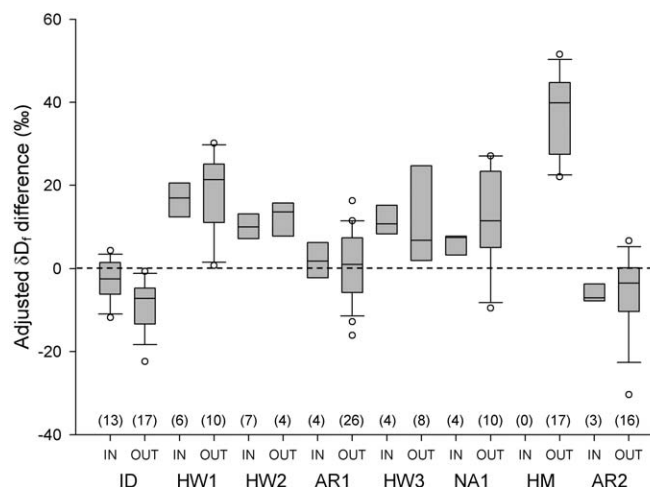

FIG. 1. Differences between two δD_f measurements from adjacent locations of a single raptor contour feather for eight independent sample groups measured during independent analysis events at Environment Canada's Stable Isotope Hydrology and Ecology Laboratory in Saskatoon, Canada. Poor reproducibility is present for all sample groups. Substantial systematic error is present whether samples are "in" or "out" of the calibration range of keratin standards used for δD_f value normalization (see text), whereas random error is large in both cases but larger "out" of the calibration range. Sample groups comprise feathers from three species of raptors (Merlin, Sharp-shinned Hawk, and Red-tailed Hawk) drawn from the data sets originally described in Smith et al. (2009). Differences in δD_f measurements are calculated as initial measurement minus repeated measurement, adjusted for natural intra-feather variation in these species (Smith et al. 2008; see text). The broken horizontal line at 0‰ indicates no systematic error (i.e., bias) between initial and repeated measurements. Box plots indicate the median and the 25th and 75th percentiles of the difference, and whiskers indicate the 10th and 90th percentiles. Circles indicate outliers. Sample sizes are indicated in parentheses. The abbreviations along the abscissa correspond to those given in table 1 and figure 1 of Smith et al. (2009). Sample group species composition: ID (30 Red-tailed Hawks), HW1 (16 Sharp-shinned Hawks), HW2 (6 Merlins and 5 Sharp-shinned Hawks), AR1 (30 Merlins), HW3 (12 Sharp-shinned Hawks), NA1 (7 Merlins and 7 Sharp-shinned Hawks), HM (17 Sharp-shinned Hawks), and AR2 (19 Merlins).

feather material that is not confounded by the problem of reproducibility, has been studied to a limited extent (e.g., Wassenaar and Hobson 2006, Smith et al. 2008). However, these studies provide no information about measurement reproducibility among independent laboratory events. Thus, aside from Smith et al. (2009), we know of only a single, small inter-laboratory comparison of 18 nonhomogenized passerine feathers (Wassenaar 2008) that permits an assessment of δD_f measurement reproducibility. Although Wassenaar (2008) did not quantify reproducibility, he presented a graph (fig. 2.5) from which he inferred good comparability of δD_f measurements among labs. Nonetheless, a close inspection of this figure indicates consistent systematic differences in δD_f measurements between some laboratories, ~5‰ on average and up to ~15‰, despite an apparent lack of intra-feather variation in passerine feathers (Mazerolle et al. 2005, Wassenaar

and Hobson 2006, Langin et al. 2007). Thus, although we do not imply that researchers have failed to consider isotopic variation or δD measurement error, we suggest that ours (Smith et al. 2009) is the first published experiment to adequately address reproducibility in measurements of nonhomogenized feather material using the analysis protocol employed by many isotope laboratories (i.e., Wassenaar and Hobson 2003).

Evaluating reproducibility within the probabilistic framework of Wunder (2010).—Recently, Wunder and colleagues (Wunder and Norris 2008a, b; Wunder 2010) have taken great strides to put the problem of geographic assignment into a probabilistic framework that is well suited to partition and propagate independent sources of uncertainty on predictions. We find this framework a vast improvement over previous approaches for predicting the origins of migratory animals. However, Wunder et al. advocate using the reproducibility of keratin standards as an estimate of total measurement error (e.g., Wunder and Norris 2008b, Wunder 2010). Because geographic assignment is based on measurements of feathers, not standards, and homogenized standard reproducibility is not necessarily representative of nonhomogenized feather reproducibility, we suggest that the reproducibility of nonhomogenized feather material must be estimated directly and incorporated into this probabilistic framework concurrently with, but separately from, the reproducibility of keratin standards. The large magnitudes and variable directions of systematic error documented by Smith et al (2009) complicate the specification of a probability distribution for the reproducibility of nonhomogenized feather material within the probabilistic framework, particularly without an understanding of the mechanism(s) driving systematic error between analysis events.

Alternatively, the effect of random error in δD_f measurements on predictions of origin might be assessed using the probabilistic framework, despite systematic error, by considering the variation (e.g., SD) around the average of repeated δD_f measurements from the same feather, after adjusting for intra-feather variation. Such an approach is comparable to how Wunder and Norris (2008b) modeled the reproducibility of keratin standards and how Hobson et al. (2009) modeled "within-population" SD. If "replicates" from independent analysis events are available from a large number of individuals, a distribution of standard deviations can be generated (*sensu* Wunder and Norris 2008b) to represent feather δD reproducibility. However, this approach will fail to address the problem of systematic error among analytical events. Regardless of how the reproducibility of nonhomogenized feather material is modeled, the negative effects of δD_f measurement error on the reliability or specificity of inferences regarding migratory origins and connectivity will be larger than what has previously been reported once this error has been addressed. For this reason, we question the conclusion that δD_f measurement error has minor effects on the specificity of geographic assignment from studies that based the input of measurement error on the relatively precise and, by definition, unbiased reproducibility of keratin standards (Wunder and Norris 2008b, Wunder 2010).

Suggestions for improving and demonstrating δD_f measurement reproducibility.—An important point on which Smith et al. (2009) and Wunder et al. (2009) agree, and which we hope has been clarified during this exchange, is that the widespread practice of analyzing a single subsample of nonhomogenized feather

to represent the isotopic identity of an individual requires that assessments of reproducibility account for real biological variation within feathers (as we have done with Fig. 1). If nonhomogenized feather persists as the material on which inferences of migratory connectivity are based, documenting measurement reproducibility for nonhomogenized feathers will require correcting for intra-feather variation on a case-by-case basis. After accounting for the relatively minor effects of intra-feather variation on Smith et al.'s (2009) results, we conclude that Figure 1 provides robust documentation of the failure of current protocols for feather sampling and analysis to produce acceptably reproducible results for the measurement of δD in nonhomogenized feather material.

Where does this leave us? Given that homogenized keratin standards demonstrate adequate reproducibility to limit the effects of keratin standard measurement error on geographic assignment (e.g., Wunder and Norris 2008b), we suggest that feather-sample homogenization may be an advisable next step toward improving δD_f measurement reproducibility, despite the likely increase in processing time and per-sample costs. Homogenization offers several potential benefits. (1) Feather material is treated identically to keratin standards, in full accordance with the principle of identical treatment for stable-isotope analyses (Werner and Brand 2001). (2) True replicate samples are obtainable from the same feather (or multiple feathers from the same individual), given that homogenization would control for intra-feather or inter-feather variation. (3) It is thus possible to quantify the true reproducibility of δD_f measurements independent of biological variation in feathers, which would allow measurement error to be fully modeled within Wunder's (2010) framework. Regardless of how feathers are ultimately selected, prepared, and analyzed, we believe that demonstrably improved measurement reproducibility is necessary before we can fully understand the confidence with which we can infer migratory origins and connectivity using stable-hydrogen isotopes. Additionally, we suggest that although it may be convenient to assess reproducibility in analysis events that are in temporal proximity but are technically independent (e.g., after shutting down the furnace to replace the tube in the pyrolysis column), δD_f measurement reproducibility also should be demonstrated when replicate measurements occur at different times of the year under widely different ambient δD water-vapor conditions, before and after major modifications to laboratory equipment, and between different laboratories (using identical protocols and standards).

Pattern, prediction, and probabilistic assignment.—Wunder et al. (2009: 924) point to the r^2 value of 0.61 for a regression model of δD_f versus predicted δD in growing-season precipitation (their fig. 1) as evidence that these patterns "provide an excellent basis for productive approaches to the geographic assignment of individuals." This model, essentially identical to the model published in Lott and Smith (2006: fig. 2), serves as a sobering reminder that pattern must not be confused with prediction. Assessments of predictive accuracy based on inverse prediction intervals from calibration data sets, such as that presented in Lott and Smith (2006), have suggested for years that the spatial resolution of this technique is quite poor, rarely capable of assigning an individual migrant with confidence to a narrow portion of its range (Meehan et al. 2001, Kelly et al. 2002). This is problematic because it is the predicted origins of individuals that must ultimately inform population-level inferences of migratory origins and connectivity (Wunder 2010).

The probabilistic framework advocated by Wunder et al. (2009) has the potential to transparently illustrate the low geographic specificity that typifies transfer functions relating stable isotopes in feathers to precipitation. When (1) studies of migratory origins or connectivity are framed at *a priori* spatial scales that might inform conservation or management (i.e., states, provinces, bird conservation regions, or ecoregions) and (2) uncertainty is propagated fully into predictions using Wunder's (2010) probabilistic framework, stable-isotope studies have little power to reliably infer source regions at these scales for individuals or populations (by summing assignments across individuals; e.g., Wunder 2010). This is not the result of any shortcomings in Wunder's (2010) probability framework, which is extremely useful, but follows from our present understanding of the variability that is inherent to the feather–precipitation isotope system. Wunder's (2010) framework will be useful in addressing the problems with reproducibility that were identified in Smith et al. (2009) and discussed further in this exchange. However, it is our opinion that the fundamental variability of stable isotopes in precipitation (Dansgaard 1964, Rozanski et al. 1993, Farmer et al. 2008, Bowen 2010) and limitations to our current ability to construct reliable and case-specific transfer functions that describe how feathers incorporate this variability (reviewed in Lott and Smith 2006, Wunder and Norris 2008a, Wunder 2010) will continue to result in low specificity of geographic assignments using stable isotopes once the potentially tractable issue of reproducibility has been resolved. Although *The Auk*'s editors decided that presentation of the novel analyses required to substantiate this claim was beyond the scope of this rebuttal, we encourage interested readers to check this claim with their own data using the methods outlined in Wunder (2010) to propagate all of the error that is inherent to transfer functions describing the relationship between isotopes in feathers and precipitation (e.g., the ~ 25‰ residual variation inherent to the data set of Lott and Smith [2006: fig. 2, replicated as fig. 1 in Wunder et al. 2009]).

Moving forward with tempered expectations.—We disagree with Wunder et al.'s (2009) claims that Smith et al. (2009) was counterproductive, inappropriately alarmist, or a call for inaction. On the contrary, we hope that Smith et al. (2009) and this exchange generate greater interest in studies that seek to explain the sources of variation that most limit the specificity with which migratory origins can be predicted. We agree that compartmentalizing the variation in δD_f associated with different sources (e.g., Wunder 2010) is an important step in identifying those sources that are most limiting. However, since the foundational publications of Hobson and Wassenaar (1997) and Chamberlain et al. (1997), applied hydrogen–stable-isotope studies (which we think have routinely overstated their inferences, given considerable uncertainty) have greatly outnumbered studies designed to improve our understanding of basic specificity-limiting patterns such as (1) sources of variation in δD_f measurements, (2) spatial and temporal variability in the distribution of δD in precipitation (δD_p), and (3) the mechanisms, and case-specific variation, of δD_p incorporation into feathers. Even if problems of δD_f measurement reproducibility are resolved, probabilistic assignment of an individual's origin will still have remarkably low geographic specificity given our current understanding of the relationship between δD_f and δD_p, particularly when assignments are made with a reasonable level of credibility. Thus, we believe that researchers should discuss their inferences of migratory connectivity with the level of confidence (or skepticism) that might reasonably characterize a field in which independent validation of results is rarely possible, and that other fundamental issues that likely limit the specificity of predicted origins, described above, remain largely unexplored within the probabilistic framework of Wunder (2010).

Adam D. Smith, *Department of Natural Resources Science, 105 Coastal Institute, University of Rhode Island, Kingston, Rhode Island 02881, USA* (*e-mail: raptorbio@hotmail.com*); Casey A. Lott and Jeff P. Smith, *HawkWatch International, 2240 South 900 East, Salt Lake City, Utah 84106, USA;* Kara C. Donohue, *Department of Biology, 1910 University Drive, Boise State University, Boise, Idaho 83725, USA;* Sara Wittenberg and Kimberly G. Smith, *Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas 72701, USA;* and Laurie Goodrich, *Hawk Mountain Sanctuary, 410 Summer Valley Road, Acopian Center for Conservation Learning, Orwigsburg, Pennsylvania 17961, USA.*

Literature Cited

Bowen, G. J. 2010. Statistical and geostatistical mapping of precipitation water isotope ratios. *In* Isoscapes: Understanding Movement, Pattern, and Process on Earth through Isotope Mapping (J. B. West, G. J. Bowen, T. E. Dawson, and K. P. Tu, Eds.). Springer, New York. In press.

Chamberlain, C. P., J. D. Blum, R. T. Holmes, X. Feng, T. W. Sherry, and G. R. Graves. 1997. The use of isotope tracers for identifying populations of migratory birds. Oecologia 109: 132–141.

Dansgaard, W. 1964. Stable isotopes in precipitation. Tellus 16:436–468.

Farmer, A., B. S. Cade, and J. Torres-Dowdall. 2008. Fundamental limits to the accuracy of deuterium isotopes for identifying the spatial origin of migratory animals. Oecologia 158:183–192.

Hobson, K. A., H. Lormée, S. L. Van Wilgenburg, L. I. Wassenaar, and J. M. Boutin. 2009. Stable isotopes (δD) delineate the origins and migratory connectivity of harvested animals: The case of European woodpigeons. Journal of Applied Ecology 46:572–581.

Hobson, K. A., and L. I. Wassenaar. 1997. Linking breeding and wintering grounds of Neotropical migrant songbirds using stable hydrogen isotopic analysis of feathers. Oecologia 109:142–148.

Kelly, J. F., V. Atudorei, Z. D. Sharp, and D. M. Finch. 2002. Insights into Wilson's Warbler migration from analyses of hydrogen stable-isotope ratios. Oecologia 130:216–221.

Langin, K. M., P. P. Marra, Z. Németh, F. R. Moore, T. K. Kyser, and L. M. Ratcliffe. 2009. Breeding latitude and timing of spring migration in songbirds crossing the Gulf of Mexico. Journal of Avian Biology 40:309–316.

Langin, K. M., M. W. Reudink, P. P. Marra, D. R. Norris, T. K. Kyser, and L. M. Ratcliffe. 2007. Hydrogen isotopic variation in migratory bird tissues of known origin: Implications for geographic assignment. Oecologia 152:449–457.

Lott, C. A., and J. P. Smith. 2006. A geographic-information-system approach to estimating the origin of migratory raptors in North America using hydrogen stable isotope ratios in feathers. Auk 123:822–835.

Mazerolle, D. F., K. A. Hobson, and L. I. Wassenaar. 2005. Stable isotope and band-encounter analyses delineate migratory patterns and catchment areas of White-throated Sparrows at a migration monitoring station. Oecologia 144:541–549.

Meehan, T. D., C. A. Lott, Z. D. Sharp, R. B. Smith, R. N. Rosenfield, A. C. Stewart, and R. K. Murphy. 2001. Using hydrogen isotope geochemistry to estimate the natal latitudes of immature Cooper's Hawks migrating through the Florida Keys. Condor 103:11–20.

Paritte, J. M., and J. F. Kelly. 2009. Effect of cleaning regime on stable-isotope ratios of feathers in Japanese Quail (*Coturnix japonica*). Auk 126:165–174.

Paxton, K. L., C. Van Riper III, T. C. Theimer, and E. H. Paxton. 2007. Spatial and temporal migration patterns of Wilson's Warbler (*Wilsonia pusilla*) in the Southwest as revealed by stable isotopes. Auk 124:162–175.

Rozanski, K., L. Araguás-Araguás, and R. Gonfiantini. 1993. Isotopic patterns in modern global precipitation. Geophysical Monographs 78:1–36.

Smith, A. D., K. Donohue, and A. M. Dufty, Jr. 2008. Intrafeather and intraindividual variation in the stable-hydrogen isotope (δD) content of raptor feathers. Condor 110:500–506.

Smith, A. D., C. A. Lott, J. P. Smith, K. C. Donohue, S. Wittenberg, K. G. Smith, and L. Goodrich. 2009. Deuterium measurements of raptor feathers: Does a lack of reproducibility compromise geographic assignment? Auk 126:41–46.

Wassenaar, L. I. 2008. An introduction to light stable isotopes for use in terrestrial animal migration studies. Pages 21–44 *in* Tracking Animal Migration with Stable Isotopes (K. A. Hobson and L. I. Wassenaar, Eds.). Academic Press, London.

Wassenaar, L. I., and K. A. Hobson. 2003. Comparative equilibration and online technique for determination of nonexchangeable hydrogen of keratins for use in animal migration studies. Isotopes in Environmental Health Studies 39:211–217.

Wassenaar, L. I., and K. A. Hobson. 2006. Stable-hydrogen isotope heterogeneity in keratinous materials: Mass spectrometry and migratory wildlife tissue subsampling strategies. Rapid Communications in Mass Spectrometry 20:2505–2510.

Werner, R. A., and W. A. Brand. 2001. Referencing strategies and techniques in stable isotope ratio analysis. Rapid Communications in Mass Spectrometry 15:501–519.

Wunder, M. B. 2010. Using isoscapes to model probability surfaces for determining geographic origins. *In* Isoscapes: Understanding Movement, Pattern, and Process on Earth through Isotope Mapping (J. B. West, G. J. Bowen, T. E. Dawson, and K. P. Tu, Eds.). Springer, New York. In press.

Wunder, M. B., K. A. Hobson, J. Kelly, P. P. Marra, L. I. Wassenaar, C. A. Stricker, and R. R. Doucett. 2009. Does a lack of design and repeatability compromise scientific criticism? A response to Smith et al. (2009). Auk 126:922–926.

Wunder, M. B., and D. R. Norris. 2008a. Analysis and design for isotope-based studies of migratory animals. Pages 107–128 *in* Tracking Animal Migration with Stable Isotopes (K. A. Hobson and L. I. Wassenaar, Eds.). Academic Press, London.

Wunder, M. B., and D. R. Norris. 2008b. Improved estimates of certainty in stable-isotope-based methods for tracking migratory animals. Ecological Applications 18:549–559.