J. Field Ornithol., 66(3):385–390

ERRONEOUS PARTY-HOUR DATA AND A PROPOSED METHOD OF CORRECTING OBSERVER EFFORT IN CHRISTMAS BIRD COUNTS

ALAN P. PETERSON
P.O. Box 1999
Walla Walla, Washington 99362-0999 USA

Abstract.—Published Christmas Bird Count (CBC) data contain values for Party Hours (PH) of effort that frequently are improbable and occasionally impossible. PH values are widely used to adjust for observer effort when analyzing CBC data. Uncritical use of PH values for observer effort correction may result in misleading results. Unlikely PH values are most likely to result from inadvertantly calculating observer hours rather than party hours, but other explanations are likely and are discussed. A method is proposed for revision of unlikely values. Briefly, the trimmed mean is used to calculate a probable number of party-hours per party for a given count, and this number is multiplied times the number of parties during the year in question to estimate a probable value for that year.

DATOS ERRÓNEOS SOBRE GRUPOS-HORA Y UN MÉTODO PROPUESTO PARA CORREGIR EL ESFUERZO DE OBSERVACIÓN EN LOS CONTEOS DE NAVIDAD

Sinopsis.—Los datos publicados en los Conteos de Navidad contienen valores para el esfuerzo de grupos-hora (GH) que frecuentemente son improbables y en ocasiones imposibles. Los valores de GH son ampliamente utilizados para ajustar el esfuerzo del observador cuando se analizan los datos de los Conteos de Navidad. El uso indiscriminado de los valores de GH para corregir el esfuerzo del observador puede dar origen a la malinterpretación de resultados. Valores improbables de GH tienen una alta probabilidad de ocurrir de calcular inadvertidamente las horas de observación en vez de horas por parte del grupo, pero otras explicaciones son posibles y se discuten en este trabajo. Se propone un método para revisar valores improbables. Se utiliza un promedio equilibrado (trimmed mean) para calcular un número probable de grupos-horas por grupo para un conteo particular. Luego, este número se multiplica por el número de grupos durante ese año de trabajo para estimar un valor probable para ese año.

Widespread geographic coverage and long duration make Christmas Bird Count (CBC) data an obvious candidate for monitoring bird populations. Variability in the counts requires that corrections be made before comparing results across space and time. Reliable estimates of relative abundance from CBCs depend, in part, on appropriate standardization for observer effort (Butcher et al. 1990). The use of party hours (PH) as part of the correction is widespread (Bock and Root 1981, Butcher et al. 1990). This correction method divides the number of birds of each species by the PH recorded during the count (Bock and Root 1981). As a CBC compiler I noticed that parties often err when calculating their number of PH. Even experienced, careful birders, possibly influenced by enthusiasm for their earnest efforts, sometimes multiply their hours in the field by the number of observers, producing a number representing observer hours rather than PH. This error might not be picked up by the count compiler. The compiling staff at American Birds recognizes the potential for this error, as recent Christmas Bird Count Instructions (1993) state: "It would be unrealistic to report any more party-hours than the number of parties times 10, since 10 is the approximate number of daylight hours in December and January where the vast majority of CBCs are conducted" (National Audubon Society 1993: [P. 2]).

As PH is important as a standardizing factor, I became interested in the frequency of erroneous PH data in published CBCs. Obviously, besides the calculation error suggested above, data entry mistakes, transcription errors and the host of problems associated with numerical data can also produce erroneous values. A faulty PH value, from any cause, will throw off standardization for that count, and affect the utility of the data in detecting trends.

METHODS AND RESULTS

To investigate the frequency of faulty PH estimations, I looked at data from CBCs from 1972 to 1989, in Washington, Oregon, and southern British Columbia in regions reporting Bewick's Wrens (*Thryomanes bewickii*). This investigation was done as part of a study of changing Bewick's Wren distribution. For these counts I calculated and tabulated the number of party hours per party (PH/P).

Figure 1 is a stem-and-leaf plot of PH/P for these counts (n = 1030). The distribution has quite a long tail on the high end; it does not appear Gaussian. The Lilliefors test for normality gives a d value of 0.0512, rejecting the null hypothesis at $P < 0.0001$ (Dallal 1989). Some values are obviously ridiculous. For example, the high value of 36 PH/P is not possible in a 24-h day. When the calculated PH/P data are tabulated by frequency of values, 423 response cells result, ranging from 2.833 to 36 PH/P. The most frequent result is 8 PH/P (n = 86), followed by 8.5 and 9 PH/P (both with n = 27), 7 PH/P (n = 21), 6 and 7.5 (both with n = 17), and 9.5 (n = 11). These results suggest that, as expected, some parties and counts use rough estimates of time spent in the field, whereas others record time with more care.

Is it possible to improve the published data? Many values are above what is considered a reasonable number; 4% are greater than 10 PH/P. Throughout the rest of the range of PH values, accurate numbers undoubtedly mingle with erroneous ones and values that seem to be reasonable may include the kinds of errors producing the extreme values; transcription errors, or faulty calculations can produce numbers that appear reasonable. It would be a shame to alter accurate data, but it seems that obviously incorrect data must either be excluded or improved. In my work, if I excluded all data from counts having unreasonable PH/P values, I lost much geographically critical data. This prompted me to calculate reasonable estimates for unreasonable values.

I examined counts having PH/P greater than 11. During the CBC period in this region there are generally 10 h of daylight from morning to evening twilight (United States Naval Observatory 1990). As owling hours are recorded separately from other birding hours, I felt safe in assuming that the number of PH that each party spent should be less than 11. Apart from owling, birding in the dark is rarely done and is even more rarely

```
  2    8
  3    1233368999
  4    00011
***OUTSIDE VALUES***
  4    1223333
  4    5555566777889
  5    00011112222223333333444
  5    5555555555555556666666666666777777788888888889999999999
  6    0000000000000000000000000011111111111222222222223333333333 3*
  6 H  5555555555555555555555555555666666666666666666666666 7777777*
  7    0000000000000000000000000000001111111111111111111111111 1222*
  7 M  55555555555555555555555555555555555666666666666666666666666 6*
  8    00000000000000000000000000000000000000000000000000000000000*
  8 H  55555555555555555555555555555555555555555666666666666666666*
  9    0000000000000000000000000000000111111111111111 1222222222222*
  9    55555555555555666777788899
 10    000000000011111122222344
 10    5555666677
 11    0000
***OUTSIDE VALUES***
 11    2345
 12    022334557
 13    0455
 15    022
 16    068
 17    07
 18    36
 19    1
 20    8
 21    1
 23    23
 36    0
```

FIGURE 1. Stem-and-leaf plot of PH/P 1972-1989. A stem-and-leaf plot is like a histogram tipped on its side. The "stem", in this instance, is formed by the tens component of the PH/P value; the "leaves" are composed of the units component of each datum. Thus the first line indicates a single value of 28, the next line indicates 31, 32, 33, 33, 33, 36, 38, 39, 39, and 39. Where a data line has been truncated to produce a manageable plot, an asterisk denotes values not shown. The "M" indicates the line containing the median, "H" shows the line where the "hinge" occurs. The "hinge", sometimes called quarter or fourth, is the datum in the middle of each half of the data. A summary of these measures of spread is given as: Minimum: 2.83; Lower hinge: 6.76; Median: 7.78; Upper hinge: 8.50; Maximum: 36.00. Outside values are determined here as values falling more than 1.5 times the hinge spread beyond the nearest hinge.

productive. Eleven was chosen: (1) because it seemed a value unlikely to be correct due to day-length limitations, and (2) it was identified as a limit for possible outliers in this non-Gaussian distribution because it was greater than 1.5 times the hinge spread (Emerson and Strenio 1983, SYSTAT, Inc 1992). Hinges (i.e., quartiles) are the values midway between the median and the extreme values in a ranked distribution; the hinge spread is the distance between these values.

For PH/P values greater than 11 I calculated a correction based on the trimmed mean (TM) (Rosenberger and Gasko 1983) of the number of PH/P for all the counts at that location. The TM multiplied by the number of parties (P) recorded for the year in question at the count with the apparently erroneous value produced a corrected PH (CPH).

TABLE 1. Sample calculation of corrected party hours[1].

Year	PH	P	PH/P
1989	118.50	14	8.46
1988	121.75	18	6.76
1987	124.00	21	5.90
1986	118.00	18	6.56
1985	122.00	18	6.78
1984	117.75	18	6.54
1983	112.00	14	8.00
1982	130.00	20	6.50
1981	101.50	15	6.77
1980	132.00	19	6.95
1979	94.25	20	4.71
1978	97.00	14	6.93
1977	93.00	12	7.75
1976	109.00	15	7.27
1975	123.00	17	7.24
1974	88.00	11	8.00
1973	98.00	12	8.17
1972	338.00	16	21.13

[1] PH, P, and calculated PH/P for Medford, Oregon. The trimmed mean of PH/P is 7.10 (trim = 25%). 7.10 PH/P × 16 P = 113.60 CPH for the 1972 count.

$$CPH_{count,yr} = TM_{PH/P_{count}} \times P_{count,yr}$$

I calculated the TM assuming a heavy tailed distribution as a model for the data. Rosenberger and Gasko (1983) suggest removing 25% of the values from each tail in such a situation. Then the mean of the remaining values is calculated. The resulting TM excludes the obvious outlier values as well as some believable values. Exclusion of the believable values, however, has only a minor effect on the calculated value. It seems reasonable to use this robust estimator of central location on non-Gaussian data already shown to contain error. In subsequent calculations I substituted CPH for PH.

As an example, the 1972 Medford Oregon count lists 338 PH, 16 P (American Birds 26(2):496), which imply 21.13 PH/P. This cannot be right. Table 1 shows a sample correction for the Medford, Oregon count.

A frequency distribution of the CPH/P derived from the 1030 values in Figure 1 is seen in Figure 2. Erroneous values undoubtedly remain even after this correction scheme, but the tail of high values has been much reduced, with impossible values mostly eliminated.

DISCUSSION

PH correction for observer effort has been used widely, but I show here that many values corrected by PH must be erroneous. An argument can be made for eliminating any such erroneous values. The correction scheme I propose has the advantage of not introducing more temporal and geographic gaps than already exist in the data set. One disadvantage

```
  2    8
  3    1233368999
  4    000111
***OUTSIDE VALUES***
  4    223333
  4    5555566777889
  5    0001111222222333333333444
  5    5555555555555555666666666666777777777888888888889999999999
  6    00000000000000000000000001111111111111222222222222233333333333*
  6 H  55555555555555555555555555555666666666666666666666666666777777*
  7    00000000000000000000000000000000000011111111111111111111111111*
  7 M  55555555555555555555555555555555555555555666666666666666666666*
  8 H  00000000000000000000000000000000000000000000000000000000000000*
  8    55555555555555555555555555555555555555555566666666666666666666*
  9    00000000000000000000000000000000011111111111111111222222222222*
  9    55555555555555555566677778899
 10    000000000001111112222344
 10    5555666677
***OUTSIDE VALUES***
 11    00002345
 12    0
```

FIGURE 2. Stem-and-leaf plot of CPH/P 1972–1989. A summary of these measures of spread is as: Minimum: 2.83; Lower hinge: 6.75; Median: 7.69; Upper hinge: 8.43; Maximum: 12.00. Outside values are determined again as values falling more than 1.5 times the hinge spread beyond the nearest hinge.

of this method is that the commercially available compilation of the CBC data (Cornell Laboratory of Ornithology) does not include P, thus from their data alone, PH/P cannot be calculated and its distribution inspected for outliers. Manual retrieval of P from published CBCs is tedious and itself prone to error. Critical inspection of raw PH distributions, however, frequently will reveal questionable values prompting further investigation and possible correction before they are used as a standardizing factor.

Another disadvantage is that even with this correction scheme, some clearly erroneous values remain. Perhaps the most prudent approach would be to exclude any values from counts having suspicious-looking PH values. In my experience, this eliminated so much CBC data that any attempt at analysis for time and geographic trends was futile. Another approach would be to set the cut-off value at 10, instead of 11. The choice of a particular cut-off value is clearly arbitrary, and should be partly dependant on latitude, as it determines length of daylight. I was reluctant to go as low as 10 for a cut-off value because it necessitated altering so many results. I am suspicious that one CBC "hour" is only loosely coupled to a clock "hour," so I want to be liberal in accepting values. I also was convinced that many values below 10 were erroneous estimates as well, but I had no way of identifying or correcting them.

Values for P are also subject to error. Arbib (1981:31) discussed the effect of unrecorded party splitting on estimates of PH totals and birds/PH, suggesting that such unreported party splitting may produce results that are in "substantial error, on the low side for party-hour totals, high for birds/party-hour." These errors in P will affect the PH data, and any data corrected for observer effort.

Given the widespread use of PH for observer-effort standardization, and the continued use of CBC data to track bird abundance, I believe it is useful to be aware of the character of the data and its frequent errors. In evaluating this one variable (PH) at least 4% of the values appear to be clearly wrong. Many erroneous values will never be recognizable as incorrect. The instructions asking compilers to check the plausibility of PHs submitted by individual parties are helpful, but I warn investigators to use caution when using PHs as a standardizing term for abundance estimation.

ACKNOWLEDGMENTS

I thank the enthusiastic and careful participants of the Walla Walla CBC for helping alert me to this problem. Missy Peterson made valuable suggestions at all phases of the work. Early versions of the manuscript benefitted substantially from comments by S. D. West and G. S. Butcher. L. J. Goodrich made many useful suggestions on later versions of the manuscript; Ken Yasukawa and two reviewers also contributed considerably.

LITERATURE CITED

AMERICAN BIRDS. 1972. The seventy-second Christmas Bird Count. 26(2):496.

ARBIB, R. S. 1981. The Christmas Bird Count: constructing an "Ideal Model." Pp. 30–33, *in* C. J. Ralph and J. M. Scott, eds. Estimating numbers of terrestrial birds. Studies in Avian Biol. 6, Cooper Ornithol. Soc., Allen Press, Lawrence, Kansas.

BOCK, C. E., AND T. L. ROOT. 1981. The Christmas Bird Count and avian ecology. Pp. 17–23, *in* C. J. Ralph and J. M. Scott, eds. Estimating numbers of terrestrial birds. Studies in Avian Biol. 6, Cooper Ornithol. Soc., Allen Press, Lawrence, Kansas.

BUTCHER, G. S., M. R. FULLER, L. S. MCALLISTER, AND P. H. GEISSLER. 1990. An evaluation of the Christmas Bird Count for monitoring population trends of selected species. Wildl. Soc. Bull. 18:129–134.

DALLAL, G. E. 1989. ODDJOB: a collection of miscellaneous statistical techniques. Am. Statistician 43:270.

EMERSON, J. D., AND J. STRENIO. 1983. Boxplots and batch comparison. Pp. 59–62, *in* D. C. Hoaglin, F. Mosteller, and J. W. Tukey, eds. Understanding robust and exploratory data analysis. John Wiley and Sons, New York, New York.

NATIONAL AUDUBON SOCIETY. 1993. National Audubon Society-Leica Christmas Bird Count instructions. National Audubon Society, New York, New York. 2 Pp.

ROSENBERGER, J. L., AND M. GASKO. 1983. Comparing location estimators: trimmed means, medians, and trimean. Pp. 297–336, *in* D. C. Hoaglin, F. Mosteller, and J. W. Tukey, eds. Understanding robust and exploratory data analysis. John Wiley and Sons, Inc., New York, New York.

SYSTAT, INC. 1992. Systat for Windows: graphics, Version 5. Ed Evanston, Illinois.

TUKEY, J. W. 1977. Exploratory data analysis. Pp. 27–53, Addison-Wesley Publishing Co., Reading, Massachusetts.

UNITED STATES NAVAL OBSERVATORY. 1990. The astronomical almanac for the year 1990. Pp. A21, A23. U.S. Government Printing Office, Washington, D.C.

Received 19 Sep. 1994; accepted 28 Nov. 1994.