Review Team #1

Review of "Developing Methods to Reduce Bird Mortality in the Altamont Pass Wind Resource Area" by BioResource Consultants

Executive Summary

The report "Developing Methods to Reduce Bird Mortality in the Altamont Pass Wind Resource Area" represents a large effort collecting data about the association of avian mortality with wind turbine type, topography, rodent management, and other variables. Despite the extensive survey effort, flaws in study design and statistical analysis hamper the interpretation of results.[1]

The study has three major statistical flaws. These are:

Pseudoreplication. Statistical inference depends on samples being randomly selected and measured. Pseudoreplication is when samples are not independent, but are treated independently. In the study, turbine strings were sampled as a unit, but individual turbines were then treated as independent samples. This causes results to appear to be more significant than they actually are.

Inappropriate use of Chi-squared analyses with measured variables. Chi-squared analysis must be used with counts, not measurements of time. The authors of the study used Chi-squared analysis to assess significance of timed bird behaviors.

Multiple comparisons with inter-correlated variables without appropriate corrections. Conducting many statistical tests increases the chance that "significant" results will be found that are actually not significant (called a Type I error). The probability of Type I errors is also increased by conducting multiple tests with correlated independent variables. The authors of the study failed to use the appropriate corrections to account for this.

The study design also has several flaws that could compromise the reliability of results.

Nonrandom sampling of turbine strings. Constraints imposed by wind farm operators precluded the authors of the study from implementing a random sampling design. Although they ultimately sampled 75% of the turbine strings, their results cannot be extrapolated to turbine strings that were not sampled and the order turbine strings were added to the surveys could have affected results.

Differences in observer ability were not incorporated. The authors failed to incorporate differences in observer ability, or to demonstrate that they are insignificant.

Differences in scavenger removal were not incorporated. Assessments of avian mortality by locating carcasses should account for carcass removal by scavengers. Because scavenging is distributed unevenly across landscapes in response to vegetation conditions, its influence could confound patterns caused by other variables (e.g., turbine type, topography, etc.). The authors

---

[1] ~~Until these flaws are addressed, the conclusions of the study are premature.~~

assumed that scavenging rates were equal across the whole study area, regardless of local vegetation conditions. This may have affected the results of the study.

Technical Overview

This document reviews the methodological and statistical adequacy of "Developing methods to reduce bird mortality in the Altamont Pass Wind Resource Area" prepared by BioResource Consultants and published in August 2004.

The study has three major statistical flaws: 1) Pseudo-replication; 2) Inappropriate application of Chi-squared analyses to measured variables; 3) High probability of Type I errors resulting from using uncorrected post-hoc comparisons and inter-correlated independent variables.

[2]Individual turbines are not statistically independent *because they were surveyed in strings and turbines were not sufficiently separate to unquestionably assign a carcass to the turbine that caused the death. The appropriate unit of analysis is the turbine string. Consequently*, all analyses at the turbine level must account for this hierarchical structure in the data. Failure to do so, as occurred in this study, artificially inflates error degrees of freedom and F-ratios and makes effects appear more significant than they actually are. All of the ANOVA tests *using turbines as the experimental unit* performed in Chapter 3 suffer from pseudo-replication. Though less obvious, the Chi-squared analyses in both Chapter 7 and Chapter 8 should have also accounted for the structure of the data.

Chi-squared tests are often viewed as designed to compare observed data values to expected values derived from some model. This notion is only partly correct. The data must be counts of sample units possessing a particular attribute. The analyses in Chapter 8[3] apply Chi-squared frequency analyses to a measurement variable (i.e., *minutes of activity, top of p 254*). This is an incorrect application of the Chi-squared test.[4] *Consequently, the results presented in Tables 8-6 (minutes of perching), 8-7 (minutes of flight), 8-8 (mean flight height), 8-9 (mean distance from nearest wind turbine), 8-10 (flight time within 50 meters), 8-11 (minutes of perching, minutes of flying) are not valid. The appropriate analysis would have been to use an Analysis of Variance.*

The study furthermore has three methodological flaws that may alter the conclusions drawn from the study: 1) turbine strings were sampled haphazardly, 2) results were not adjusted for observer ability, and 3) adjustments for scavenger removal relied on other studies and did not account for differences in vegetation type or height.

Investigators added turbines to the study as they were made available by wind farm operators, not according to a pre-determined sampling design. Consequently, turbines were surveyed for different periods and turbines that may have had different characteristics were added to the pool of sampled turbines over time, potentially affecting study results. Although the authors specifically assert that the results of the study cannot be extrapolated to turbines that have not

---

[2] ~~The last paragraph of p. 47 (in Chapter 3) indicates the basic sampling unit was a string of turbines. This sampling scheme imposes a structure on the data where an individual turbine is a subunit of a string. Because of this structure,~~

[3] ~~attempt to~~

[4] ~~and consequently, almost all of the tests reported in Chapter 8 are invalid.~~

been sampled, this still represents a methodological drawback that was caused it seems by the wind farm operators.

The authors assert that it is not necessary to adjust for observer ability in reporting fatalities around turbines because they know that mortality will be underestimated. The study does, however, rely heavily on comparisons of numbers of birds killed at different turbine types. Observer ability could bias fatality rates up or down and consequently could alter conclusions about different turbine types and locations if different observers disproportionately surveyed one type of turbine or landscape position.

The authors assert that it is not necessary to adjust for scavenger removal because it is already known that mortality will be underestimated. The analysis relies on numbers of fatalities that are detected; differential scavenger removal throughout the study area would affect results of all subsequent analyses. If scavenging were uniform across the study area and among turbine types this variable probably would not affect conclusions, but the report contains no information to indicate that this is true.

Finally, the report did not address the existing literature on birds colliding with tall, lighted structures at night. Although raptors and grassland birds are presumably killed most by turbines at Altamont Pass during the day, collisions with migratory birds *may* also occur at night, *especially if taller turbines are installed in the future*. The study does not record whether turbines were lighted, and does not recognize that recommending that turbines be replaced on the tallest possible towers may actually increase mortality of migratory birds.[5]

We note here that the turbine operators at APWRA significantly hindered the design of the research project. Access was only granted to subsets of turbines from the start of the project, limiting the ability of the investigators to conduct random (or stratified random) sampling (or even complete sampling). Furthermore, the investigators reported that staff for turbine operators may have buried or hidden carcasses of birds. These factors must be eliminated to improve any future studies at APWRA. Although the investigators apparently agreed to do the study even under these conditions and therefore bear some responsibility for these limitations themselves, the greater interest of increasing knowledge about this system cannot be served if the turbine operators do not cooperate in research efforts and indeed, the investigators stated that they would not undertake a similar study if full access were not granted.

---

[5] *Regarding this issue, Smallwood and Thelander responded, "Some comments were irrelevant or confused, caused by lack of familiarity. For example, Review Team 1 appeared amazed that we neglected to discuss turbine lighting as an issue in the APWRA, but wind turbines in the APWRA are not lit. The issue of nocturnal migrants colliding with tall towers on the east coast of the U.S. cannot be extrapolated to wind turbines on the west coast, especially these small ones in the APWRA. The wind turbines in the APWRA are nowhere near the heights of communication towers, so citing the literature on collisions with communication towers would be irrelevant." By way of clarification, we suggested that the issue should be discussed given that Smallwood and Thelander recommend taller turbines in the repowering program. Groups of towers more than 200 feet tall would need to be lighted, and taller towers is currently the trend in the wind industry. This research has the potential to be applied elsewhere (despite admonitions by the authors to the contrary) and we therefore thought this issue should be addressed.*

Specific Comments

Further detail about our general comments and specific questions are raised in the following responses to aspects of each Chapter. Where relevant to the discussion, we have quoted text from the BioResource Consultants report (in *italics*) or reprinted figures. Our comments and questions are presented in Roman text. *In the final report, italicized text in our comments indicates revisions since the draft review.*

Chapter 1– Introduction

Page 7, "*In March 1998, the National Renewable Energy Laboratory (NREL) initiated research to address certain complex questions: What is the full extent of bird mortality in the APWRA? What are the underlying causes of the mortality? What role do bird behaviors play in mortality at wind turbines? Is mortality predictable at wind turbines with certain suites of characteristics? If it is, then can management strategies be developed to reduce mortality?*"

The questions raised by authors in paragraph 2 are valid, but could have been expressed as sequential components that, acting in concert, result in mortality of birds. The primary components are:

(**1) physical and operational attributes of turbines** (i.e., variables of: turbine model, turbine size, rotor diameter, tip speed, window, rotor-swept area/sec., tower type, tower height, blade color scheme, perch guard, low reach of blades, high reach of blades — as used in analyses for Table 7-2),

(2) placement of individual turbines (or turbine strings) relative to topography and prevailing wind (i.e., variables of: orientation to wind, derelict turbine, whether in wind wall, position in string, position in farm, turbine congestion, elevation, slope grade, physical relief, whether in canyon, slope aspect),

(3) ecological aspects relative to avian foods that may affect behavior of birds near turbines (e.g., variables of: edge index, rock piles, rodent control, cattle pats-grass, cattle pats-turbines, cottontails-grass, cottontails-turbines, and vegetation) and

(4) behavior and seasonal abundance of avian species near (within 300 m) of turbines in "rotor zone" (e.g., variables of: season of the year, time spent flying (20 behaviors — as in Table 8-2 and Table 8-18), flight height, distance from turbines, dangerous flights, time spent perching (26 perch structures as in Table 8-2).

Inasmuch as Step 1 (susceptibility) and Step 2 (vulnerability) result in Step 3 (impacts, i.e. mortality), the first two terms mean nearly the same thing ("capable of being affected" vs. "capable of or susceptible to" some variable. *Notwithstanding the existence of literature using these terms*, the use of these two terms as meaning different things is jargon that is not familiar to most readers, *even ecologists and wildlife scientists*. The authors should either provide a more detailed explanation of the difference between susceptibility and vulnerability or avoid this usage. Furthermore, although terminology is defined throughout the document, a glossary of terms in an appendix would be useful.

Page 8, first paragraph, last sentence: "*For this report we have combined the data from both research efforts.*"

[6]The non-random addition of study sites *during these two studies* confounds various analyses. This is especially important because of the year-to-year variation in measured fatalities.

Page 9, first paragraph under 1.1.2: "*The **placement and** operation of wind turbines can make birds vulnerable to wind turbine collisions…*"

This sounds as if birds can die at wind turbines (fly into them) even if turbines are not operating (blades not turning). Are deaths in this manner minuscule compared to deaths in moving blades or is this known?

Page 11, Section 1.1.4, first paragraph: *"….then the probability of an individual being killed by a wind turbine occurring on a particular environmental element would equal the proportion of the wind turbines associated with…*"

Not sure of the wording here. Would it ***equal*** the probability or just be associated with it?

Page 12, "*At selected turbines in the APWRA, we compiled data separately for bird behaviors, wind turbine and tower characteristics, fatality searches, fatality search results, maps of rodent burrow systems, and various other physical and biological factors."*

Both here and elsewhere, the authors fail to provide a rationale for arbitrary actions. How were turbines selected? The authors do note that the turbine selection was a result of the operational constraints of APWRA, "*Only about 28% of the APWRA's total wind turbine population was included in the project initially, due to limitations placed on access to turbines*" (also p. 12). Even with such limitations, the approach (although not stated in the report) seems to have been to survey all turbines to which access was granted, rather than selecting either a random sample or a stratified random sample that would have included representation of each turbine type. [7] This limits the conclusions of the study considerably, and, indeed, the authors acknowledge later that their models cannot be extrapolated to turbine strings that were not surveyed. Bird mortality could be higher or lower at turbines never studied. Furthermore, the addition of turbines to the search effort opportunistically creates *potentially* severe problems for the analyses. Because measured fatalities varied from year to year, the addition of a large number of a specific type of turbine during a "low" fatality year would give the false impression that a certain turbine type caused less mortality when data were pooled over multiple years.

Page 19, Table 1-1.

It is of interest that for 2 of the 3 turbine types (i.e., Bonus, Micon) "percent time in operation" is lacking and these two have higher mortality rates than other types except for the Kenetech KCS-

---

[6] ~~Data collected from two different research studies are not as robust as data collected for all variables in the same time period. This circumstance will limit the evaluation of interaction effects among some (maybe even) critical variables.~~

[7] *The authors provided a more detailed explanation of the sampling scheme in response to comments. While we still believe the sampling scheme was flawed (in no small part because of the turbine operators' unwillingness to provide access), this additional detail should be included in the methods section of the report.*

56, which has the highest number of carcasses associated with it in the APWRA (see Fig. 2-6). Table 1-1 lists percent time in operation as only 39% for the Kenetech KCS-56 type. It may be that operating duration was less for the Bonus and Micon turbines, or that the Kenetech KCS-56 just kills more birds because of its unique mechanical attributes. *The failure of turbine operators to provide data such as operation time compromises the ability of researchers to provide guidance to the wind industry.*

Comment deleted.[8]

Page 20, next to last paragraph: "*Within the APWRA study area, we performed focused studies involving smaller areas or select groups of wind turbines.*"

The authors never provide a rationale for how turbines were selected for the focused studies.[9] This description also gives the mistaken impression that turbines were the sampling unit, when the sampling unit was actually the turbine string (p. 47).[10] Neither this text nor the following chapters describes the selection process as being random. If turbine strings for the more focused studies were not selected randomly (that is, not every turbine string had an equal probability of being included in the focused study), then results of the focused studies on rodent burrows and bird activities cannot be extrapolated to the non-sampled turbine strings.

Page 27, last paragraph: "*To uncover and understand the patterns of bird mortality at a wind farm one must first interpret the influences on wildlife ecology that are caused by wind turbines. They are artificial structures installed in an otherwise natural setting that can have a profound influence on how arrays of interrelated landscape components function.*"

These statements suggest that an additional component to the four we presented earlier needs to be included. That is, without before and after turbine installation studies (which authors have acknowledged are needed) some of the ecological aspects are confounded inasmuch as the act of installing the turbines changes the food base that in turn affects bird behavior and may increase exposure to effects of turbines, even if the turbines are not operating.

Chapter 2 – Cause of Death and Locations

Page 28, last paragraph: "*The statistical tests included mostly one-way analysis of variance (ANOVA) and least significant differences (LSD) between groups. All LSD tests reported below were associated with P-values < 0.05. We also estimated Pearson's correlation coefficient for the distance of the carcasses and elevation of the tower base.*"

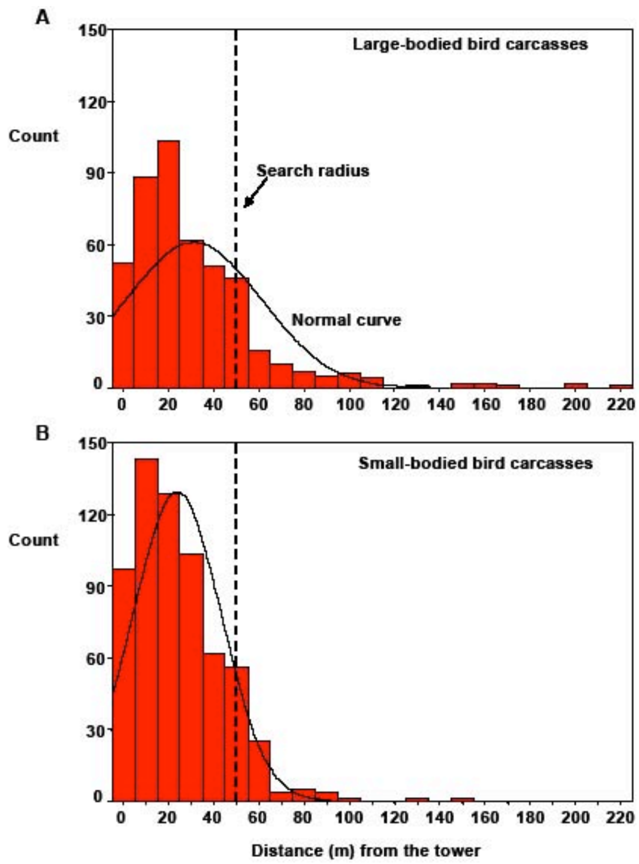
---

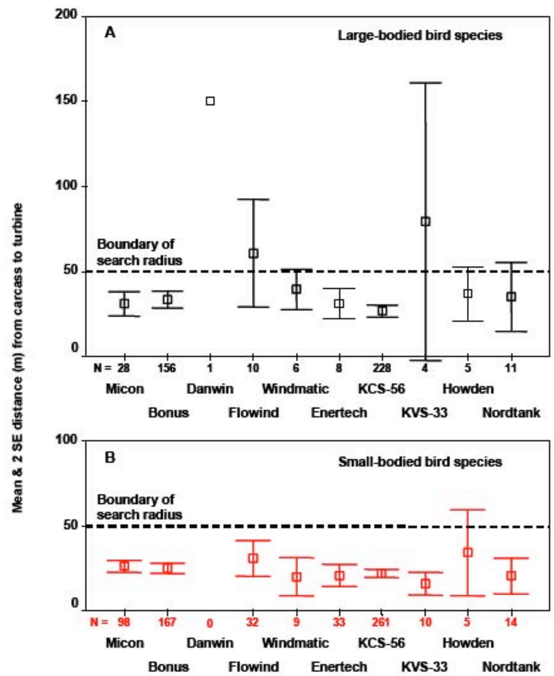
[8] ~~Page 20, first paragraph.~~
~~The authors give a general description of the study area but no vegetation map is provided. Such a map would be useful to show that either turbines are all in similar habitats, or to provide a means to visually assess the degree to which vegetation may affect survey results (e.g., lower detection of carcasses in taller vegetation or increased scavenger abundance near dense cover).~~

[9] *The authors provided more detail about the selection process in their response to these comments. Their description of a "systematic" approach needs to be fully described in the methodology sections of the report.*

[10] *In response to comments the authors claimed that both the turbine and the turbine string were the sampling unit. Both cannot be true. If turbine strings were sampled in a single visit then the turbine string is the sampling unit. This does not preclude analysis of individual turbines, but any tests must incorporate the nested nature of the data.*

The use of distances to the individual birds is pseudo-replication, which invalidates the results of the one-way ANOVA tests. The sampling unit is the string of turbines.[11] Consequently, the individual turbines are sub-samples of the strings. For this situation the appropriate analysis of the distances is a two-factor nested design. Furthermore, the authors use one-way ANOVA seemingly without regard for the underlying assumptions of the procedure, which include normality of error distribution and homogeneity of variance across variable levels. Figures 2-9 (p. 39) and 2-12 (p. 43) (reproduced below) illustrate violations of both assumptions.



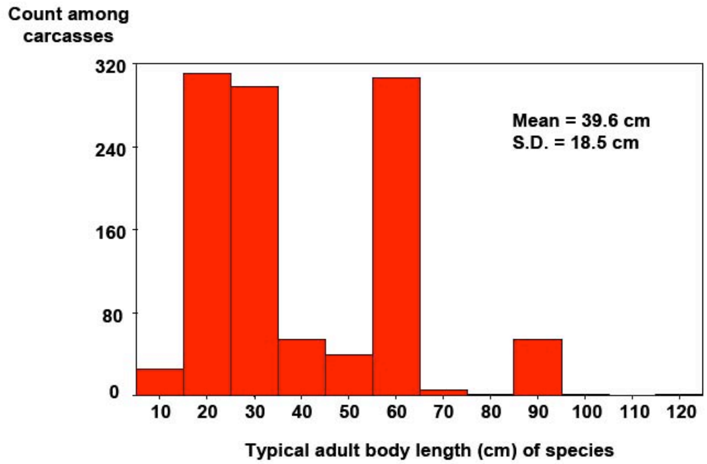


Use of LSD for *post-hoc* multiple correlations dramatically increases the chance of Type I error (i.e., labeling differences as significant when underlying population means are not). For example, in the discussion of blade tip speed (top of p. 42), with 10 categories there would be 45 possible LSD tests, which would lead to a Type I error probability of 90%.

---

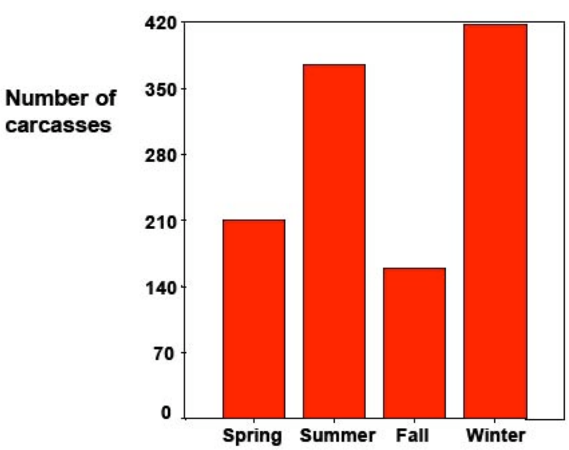
[11] *In response to this comment the authors claim that they can switch from the turbine being the sampling unit to the turbine string being the sampling unit depending on the analysis. This is not true, the data are collected in a nested format that must be accounted for in the analysis. A similar analysis can be conducted, but it cannot presume that each turbine is independently sampled.*

Page 29, Figure 2.1.



The choice of 38 cm as *a "natural beak" for dividing between* large and small body sizes seems arbitrary (see Figure 2-1 reproduced above). *To us the "natural break" occurs closer to* 50 cm. *An alternative could be* the median.[12]

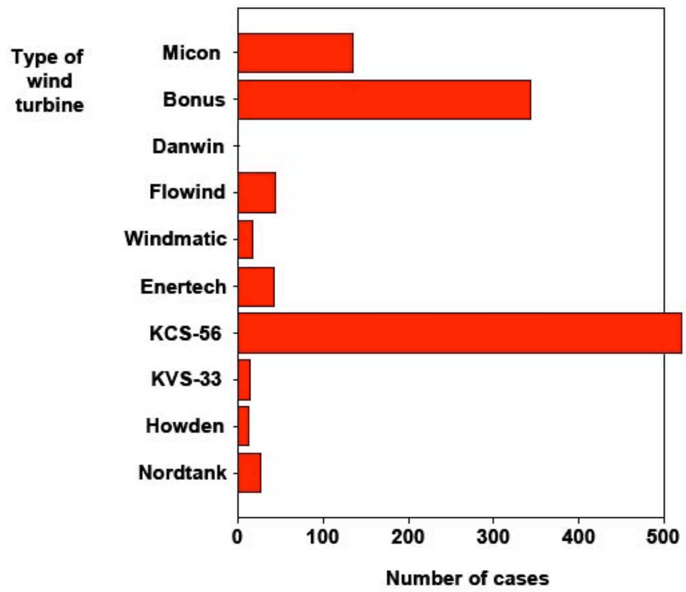
Page 36, Figure 2-5:



These results are presented as uncorrected counts. For comparability, the fatalities need to be expressed as carcasses per search effort, which needs to be clearly defined, e.g., hours or area or a combination, i.e., search effort per unit area. As raw counts, the reader does not know if the seasonal differences result from differences in search effort or seasonal changes in mortality.

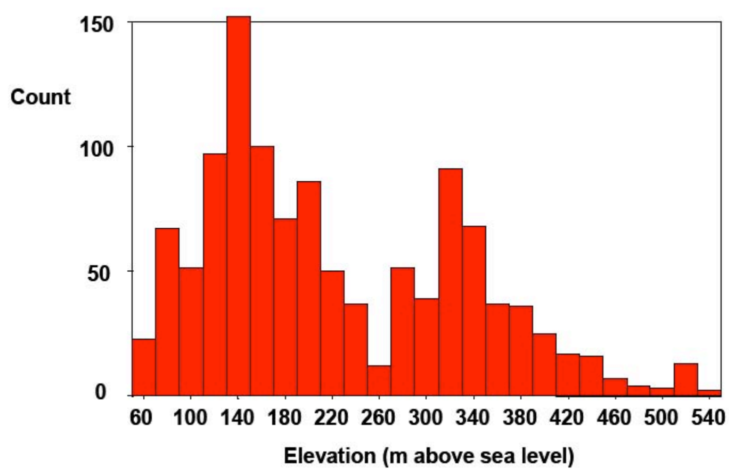
[12] *seem like more reasonable choices*

Page 35, Figure 2-6.



Results in Figure 2-6 should be expressed either as mortality per turbine type per search effort, or the graphic should express mortality as a percentage of the total mortality *and* this or another adjacent graph should depict the percentage represented by each turbine type so that readers can quickly assess whether some turbine types are associated with mortality disproportionate to their prevalence on the landscape. The figure as currently constructed could be misleading.

Page 38, Figure 2-8.

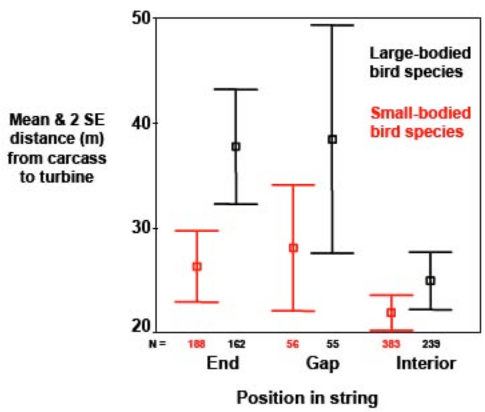


This figure and associated text should be expressed in fatalities per turbine at each altitude, or should express the mortality as a percentage of the total mortality and also should graph the percentage represented by each elevation class. The current figure does not provide much useful information because it is not clear if the pattern results from the elevational distribution of turbines or an inherent elevational pattern in mortality.

Page 41, Figure 2.11.

[13]The LSD tests described on p.38 indicate that the relationship between distance and height is not linear (i.e., the 43-m tower mean is less than the intermediate height towers.) *So the presentation of this figure, and the analysis it represents is meaningless.* In addition, the scatter plots show clear violations of the assumption of constant variation in distance across the tower heights.

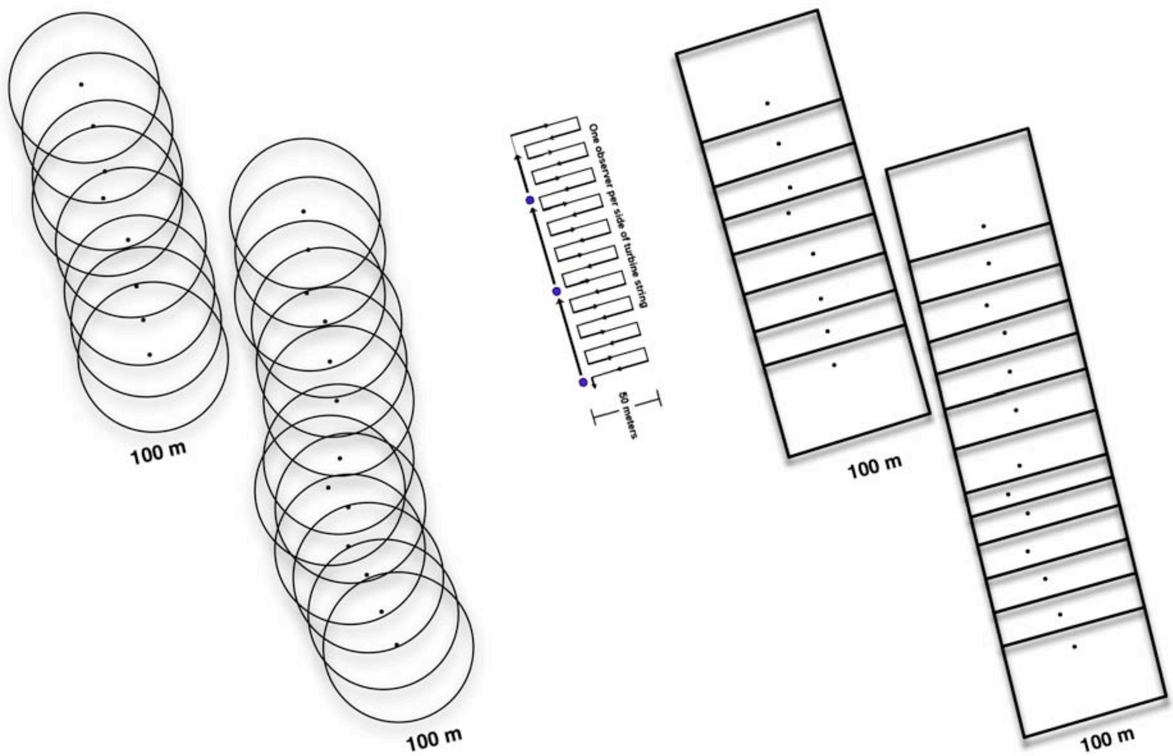
Page 42, second paragraph: "*The distance of carcass locations from the wind turbines differed according to whether the wind turbine was located at the end, at a gap, or in the interior of a string of towers (ANOVA F=11.11; df = 2, 455; P < 0.001), and post-hoc LSD tests found distances to be 13 m greater on average from end and edge of gap turbines, compared to interior turbines.*" See also Figure 2-13:



Assuming for the moment that the methodological problems (i.e., pseudo-replication resulting in inflated degrees of freedom, repeated *post-hoc* tests increasing probability of Type I errors) did not affect results, it is significant for other aspects of the study that dead birds were detected on average farther from end turbines and at gaps. This result *is probably spurious, resulting from*[14] a systematic problem allocating carcasses to turbines (and more importantly to turbine type). When a string is searched, those carcasses found along the string will be allocated to the nearest turbine (see p. 45). However, for end turbines and turbines at gaps, carcasses in a greater area would be allocated to the turbine. *This occurs* because turbines are often located *far less* than 100 m from one another. Therefore, a smaller total area allocated to turbines on the interior of strings and a greater area allocated to those at the ends and at gaps. *To illustrate this problem, we digitized the turbine locations from Figure 6-43 of the report. We used the scale to draw 50-m radius circles around each tower. This 50-m radius overlaps substantially for this set of turbines and we must conclude that the authors allocated mortalities to the closest turbine. To visualize the search area, we drew a rectangle around each turbine string that was 100 m wide and reaching 50 m beyond each end turbine. This represents the total search area for the string and is consistent with the authors' own figure (at left). We then divided up this total search area into rectangles that would be attributed to each turbine. Visual inspection of this figure leads to the conclusion that the search area for end turbines is far greater than for interior turbines*

[13] ~~Both halves of this figure are meaningless (i.e., R^2 =0.01) and inappropriate.~~
[14] ~~suggests that there is~~

(which would lead to both increased mortality estimates and to a greater carcass distance). Furthermore, it leads to the conclusion that because turbines are so closely spaced (at least in this example, which is not unique) attributing avian mortalities to a single turbine could result in many misattributions.



These figures are based on a figure in the original report and rely on the scale in that figure being accurate. The left side shows 50-m radius circles around two strings of turbines and the right side shows the areas presumably attributed to each turbine based on the description of the search methodology. The small inset in the middle is a copy of a figure by Smallwood and Thelander showing the survey methodology.

The extent of this problem could be ascertained by reporting the average distance between turbines in a string. If this distance is less than 100 m, then the conclusions of the turbine level analysis become difficult to justify because of misattribution of mortalities, and the observed greater mortality caused by end turbines could be the result of a larger search area. The authors acknowledged in response to this comment that the area searched at end turbines could be different than interior turbines but thought this difference was small. If our figure above correctly depicts the search areas, this difference could be as much double or triple the size, depending on inter-turbine spacing. The search area for each turbine should be calculated and reported to resolve this question.

[15]The authors *then* should adjust for the different search areas for interior vs. end and gap turbines. Adjusting for these differences may change the conclusion that end and gap turbines

---

[15] ~~End turbines are likely to be situated at the top of slopes (resulting in carcasses falling father away), which the authors use as an explanation for the increased distance to carcasses. However, this pattern is not likely to hold for~~

cause more mortality. It is possible that this observed relationship is merely a result of the greater search area for end and gap turbines, especially because turbines are often spaced closer than 100 m within strings (see e.g., Figure 6-41). This aspect of the methodology could jeopardize all of the turbine-level analyses in the report *because the <100 m distance between turbines will lead to some unknown degree of misattribution of mortality to individual turbines. Furthermore, our figures above show that the search area for turbines varies considerably, even on the interior of strings. As the authors prepare these data for publication in journals, a possible direction might be to use a GIS to depict the location of all avian fatalities, then describe the characteristics of turbine strings within a buffer around each fatality. This would avoid the potential problem of misattribution, especially for carcasses found equidistant from two turbines.*

Comment deleted.[16]

Page 45, first paragraph: "*We found birds beyond the 50-m search radius because the search crew members could sometimes see carcasses at these greater distances as they approached the 50-m termini of their transect segments.*"

Inclusion of these carcasses will result in a higher apparent mortality rate at those turbines where detectability is higher for any reason.[17] Because information about detectability was not gathered, it is not possible to assess the effect of this bias.

Chapter 3 - Bird Mortality

Page 46, Section 3.1 Introduction, paragraph 2.

We are not sure about mortality being expressed relative to megawatts (MW) of rated power generated per year. We can understand why authors chose to express mortality in these terms, but each type of turbine does not have the same relative effect on killing birds because of the inherent attributes of each type of turbine and we know that three different models seem to kill the most birds. Unless deaths per MW / year (or numbers of actual birds killed per year) can be clearly linked with "hours of rotating blades / year" for the particular type of turbine in question, the use of MW / year to associate with mortality may be misleading because rated power MWs do not kill birds, mechanical blades do. We therefore agree with the authors' response to this

---

~~gap turbines, and the often steep ground ("*precipices of very steep hills descending into ravines and canyons*") *could* also result in fewer carcasses being detected. The most logical explanation is that the implementation of the survey protocol, including the inclusion of carcasses located beyond 50 m, resulted in a greater effective search area for end and gap turbines.~~

[16] ~~Page 44, second paragraph: "*Carcass distances from wind turbines differed significantly by season of the year (ANOVA F=3.61; df =3, 630; P=0.013), and post-hoc LSD tests revealed that fatalities in spring were significantly closer to wind turbines (mean = 19.6 m) than were fatalities during summer (mean = 24.8 m), fall (mean = 28.1 m), and winter (mean = 23.5 m).*"~~

~~Assuming that this pattern is real, it suggests that detectability of carcasses differs by season. Because carcasses greater than 50 m from turbines were included only as observed from within the 50 m search radius, their inclusion increases the average distance of carcasses from the turbine. The authors should investigate whether carcasses from > 50 m caused this pattern. That would be logical, because vegetation is usually tallest in the spring in Mediterranean grasslands. If this pattern does result from detectability differences, it would underscore the need to account for detectability in the study design and to account for seasonal variation in search effort.~~

[17] ~~(e.g., vegetation is lower, slopes are not steep, etc.).~~

comment that MW is essentially a proxy for rotor diameter and time in operation. With this additional understanding of the authors' motives, we accept this metric as a second best metric until information on time in operation can be obtained and combined with rotor sweep.

Page 47, last paragraph before Methods: "*Finally, we extrapolated our mortality estimates to the portion of the APWRA not sampled in order to characterize the range of likely project impacts per species and larger taxonomic groups*."

The non-random sampling scheme[18] *requires that* such extrapolation *be supported by evidence that the unsampled portion of the APWRA is well represented by the data that were collected.*

Page 47, Section 3.2 Methods.

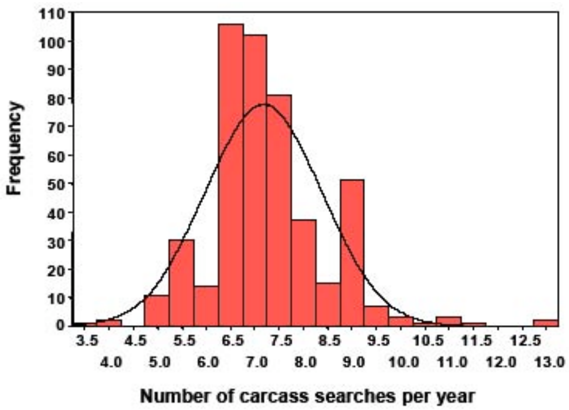
Several details are omitted from the Methods section that directly affect any judgment about the validity of the data collection methods. 1) P. 47 indicates the 1,526 turbines were sampled, but gives no specifics about how the sampled turbines were selected. *If all were searched, this should be stated.* The same criticism applies to the additional comments that note that other groups of turbines were added periodically. How were these selected for inclusion? 2) No mention is made of any efforts to prevent double counting on successive visits. Found carcasses were flagged, but no mention was made whether that flagging was permanent throughout the course of the study. *Details about carcass removal or flagging to prevent double-counting should be included in the report.* 3) No discussion is presented about the search sequence. Were strings searched in the same order throughout the rotation?

The search effort per turbine varied substantially by turbine.[19] How would this difference in effort affect the reliability of estimates of mortalities? Are these mortality estimates more conservative than if the same effort for the first group of turbines had been applied to the larger second group? And did the types of turbines in the strings differ between the two sampling periods? The sampling unit is described as the turbine string. Are turbine strings most always composed of the same type of turbine?

[18] ~~does not support~~

[19] ~~The authors acknowledge the disparity in searches for dead birds between the time periods (March 1998 Sept 2002 with 3 or 4 years for each month around 1,526 turbines) vs. (November 2002 May 2003 with only 1 search per month around 2,548 turbines) and they note that all turbine strings were searched every month.~~

Page 49, Figure 3-1.



Given the range of search effort per turbine per year (Figure 3-1), fatality estimates should be corrected upwards to adjust estimates for turbines searched less frequently. Authors assume that the same number of fatalities would have been found during a given year regardless of whether twelve searches or eight searches were performed. They acknowledge that fewer carcasses would be detected at turbine strings with fewer searches but do not adjust for this factor. What supports the assumption that the influence of search effort on carcass detection would not affect the subsequent analysis?

Page 49, second paragraph: "*Searcher detection and scavenger removal rates were not studied, because it had already been established that mortality in the APWRA is much greater than experienced at other wind energy generating facilities. We were unconcerned with underestimating mortality, and in fact we acknowledge that we did so. We were more concerned with learning the factors related to fatalities so that we can recommend solutions to the wind turbine-caused bird mortality problem. Thus, we put our energy into finding bird carcasses rather than into estimating how many birds we were missing due to variation in physiographic conditions, scavenging, searcher biases, or other actions that may have resulted in carcasses being removed.*"

Searcher detection and scavenger removal rates[20] *could affect* the results of the analyses.[21] Although both search detection ability and scavenger removal would result in underestimates of total mortality, these influences are not constant over space and time. Consequently, detection and scavenging rates would affect the results of all subsequent analyses if they are not constant over space and time. Both detection and scavenging rates are likely also affected by vegetation, which varies over space and time. Given that the remainder of the study involves multiple tests of the number of fatalities and the characteristics of the related turbines, the nonrandom, geographically varying effects of scavenging and detection are indeed centrally important to the study. This effect would be especially profound for turbines that were searched infrequently. Indeed, there could be massive scavenger losses, especially of small birds, even at the average 50-day period between searches.

[20] ~~are not inconsequential to~~
[21] ~~as implied by the authors.~~

Page 51, "*Because we did not perform trials to estimate searcher detection and scavenger removal rates, we relied on published estimates from other studies.*"

This adjustment results in simply inflating fatalities by a constant rate, but it does not incorporate the differences across space and time that *almost* certainly exist. This adjustment therefore does nothing to counteract the nonrandom influence of vegetation on detection and scavenging rates, or on observer detection ability to the extent that observers were not assigned to survey routes randomly. *If the authors believe that scavenging and detection rates are constant across the APWRA, they should provide evidence to support this assumption.*

Page 52, first paragraph: "*based on our experience with raptor carcasses in the APWRA, we did not believe that these scavenger removal rates were accurate for raptors, and we halved the removal rate estimates reported by Erickson et al. (2003).*"

What *specific evidence*[22] led *the authors* to believe that the scavenger removal rates were inaccurate for raptors?

Page 59, Figure 3-15.

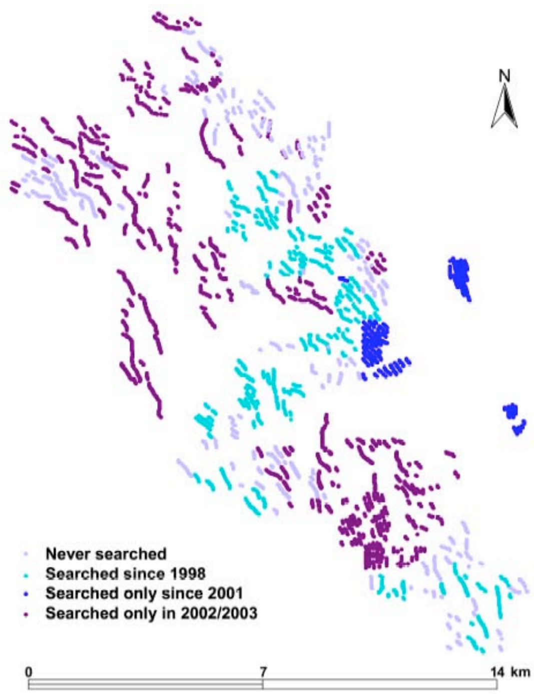


Figure 3-15 shows spatial distribution of survey effort. This figure does not appear to show a random sample. The authors should provide statistics about how the surveyed towers differ from the non-surveyed towers in key parameters (tower type, topography, elevation, turbine manufacturer, etc.). The non-random search pattern may influence other results. For example, elsewhere the authors report results for turbines that were searched four years without highlighting how the characteristics of those turbines differ from non-sampled turbines (e.g., turbine type, elevation, landscape position, etc.).

[22] ~~was the author's experience that~~

Page 64, First paragraph: It is stated that "*The mortality of all bird species combined increased steadily and significantly throughout the study, according to the comparison including all turbine strings searched for a least one year*."

Can some of this result be attributed to the increased familiarity of the investigators with the study areas, especially when areas were studied for 4 years? *If not, to what do the authors attribute this increase?*

Page 67, Table 3-3.

The right column has only turbines searched for 4 years. This is a geographically clustered sample, so it is unclear how results can be compared to the other turbines or to all other turbines at APWRA. The authors disclose that these turbines were within areas of rodent control, but do not describe the other differences from the other sampled turbines or the unsampled turbines.

Page 70, Table 3-9.

This table shows mortality per turbine string for two sets of turbines searched for different time periods. Because neither sample is random, and years of data are pooled (rather than comparing data from one year at one set to the same year at the other set), it is not obvious how the reader is to interpret this information.

It would be of interest to know how many deaths by species per year were associated with the total sum of "hours of operation / year" of all turbines and for each type of turbine in these two groups. Were there about equal proportions of each turbine type in each of these two groups? Because information like this is lacking, it is difficult to draw any conclusions from these data.

Page 75, Table 3-12.

This table provides results on a "per turbine" basis but the sampling unit was a string of turbines. *As we have illustrated above, the search methodology may have resulted in misattribution of mortality to individual turbines.*

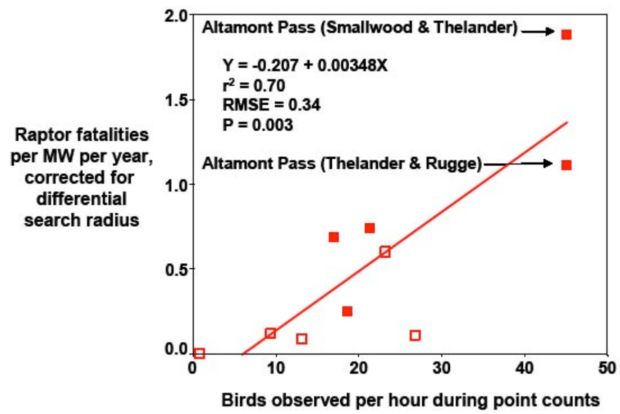
Page 76: Regarding the nonrandom sampling scheme, the authors write "*This shortfall in our study was beyond our control, since the owners of the wind turbines allowed us access to various new groups of turbines at different times during the study.*"

We are sympathetic in that the wind turbine operators did not allow access to turbines uniformly so that designing a random sampling scheme was difficult, if not impossible. This remains, however, a shortcoming of the study. The authors should have restricted all comparisons of mortality rates to turbines that were sampled during the same period and within a random sampling framework. *If a rationale can be presented to support the nonrandom sampling scheme, then results should be compared for similar time periods.*

Chapter 4 – Impacts from Wind Mortality

Herein the authors attempt to compare mortality rates at wind farms as determined in different studies. The authors make many assumptions about scavenging, detection, and search interval that cannot be verified *because they did not collect information about the influence of these factors in their study.*

Page 79, Figure 4-2.



In the regressions of raptor fatalities by birds observed per hour it seems that most of the explanatory power comes from the current study and its precursor at Altamont pass. Furthermore, the two high fatality estimates constitute partial duplication of the same data, because it seems that the data from Thelander and Rugge are incorporated into Smallwood and Thelander.

Page 83 and following, Figures 4-5 through Figure 4-7.

We are not convinced that the mortality rates from the different studies can be compared. Furthermore, the use of "bird observations" as a metric is not particularly useful because it is already apparent from the data that avian species are not all equally vulnerable to collision with turbines.

Chapter 5 – Range Management Issues

This chapter has several methodological problems: 1) The turbine level analyses are pseudo-replication and the analyses using one-way ANOVAs are therefore not valid. 2) The two transect types are paired by turbine string and should be analyzed using an analysis method that accounts for the pairing, e.g. a block or repeated measures design. 3) The use of LSD tests results in a very high probability of Type I errors. The chapter does not contain information about the sampling scheme (e.g., is it randomized, is it stratified by turbine type?).

Aspects in this chapter follow our component framework #3. Even without operating the turbines, their establishment modifies the local environment by changing the food base that may affect the behavior of birds and cause some low-level mortality. The effect on behavior, in turn, may predispose birds to be hit by turbines when they are operating and cause higher levels of avian mortality.

Page 90, Section 5.2 Methods.

Unfortunately, the amounts of lateral edge and vertical edge were characterized as “some”, or “lots”. If we understand the layout correctly, these variables could have been quantified in terms of *x* meters of lateral or vertical edge. Also, please describe *in the report* the difference between ridge crests and ridgelines. Where these topographic classifications made with automated Geographic Information System tools or based on judgments in the field or another technique?

Page 90, Paragraph 5.

How did the authors determine that cottontail pellets were especially abundant?[23] Is there any citation or precedence that connects rabbit pellets with abundance? Fecal abundance as an index of animal abundance is not always reliable.

Page 91, Section 5.2.1, first sentence.

The text “March, 1998” is missing after the word “from”.

Chapter 6 – Rodent Burrows

Page 111.

How did the authors choose the 571 turbines to map rodent burrows? This should be described in terms of turbine strings because strings are still the sampling unit. The choice of turbine stings appears to have been arbitrary, perhaps guided by an idea of a stratified random sample of turbine strings associated with different raptor mortality, physiography, and rodent control. If the sample was, indeed, a stratified random sample this should be stated clearly with a description of how many replicates of turbine strings were associated with the three criteria (i.e., range of raptor mortality, physiographic conditions, and level of rodent control). If not, then the method for choosing these turbine strings should be clearly described.

Page 124, first paragraph: “*Eleven strings of wind turbines were selected for seasonal monitoring purposes…*”

Were these strings selected randomly? The numbers (and types?) of turbines in the strings were widely variable ranging from 3–35 turbines, and 1 to 3 or 4 strings per group. How comparable were these groups?

---

[23] ~~Random transects?~~

Page 140, Fig. 6-25.

This photo suggests that type of tower, at least, was not uniform within groups of strings. Tower type seems important; did inclusion of different types of towers have any effect on results?

Page 149, first paragraph, last sentence.

Did the type of turbine have any measurable effect?

Pages 151 to 161, Fig. 6-34 through 6-44

Again it seems important to recognize the large disparities in numbers of turbines (and perhaps types of turbines?) among these sites. Is it possible that unadjusted mortality of species is related to number and type of turbines and not rodent control treatments? Is there no way to test for interaction effects?

Chapter 7 – Predictive Models

Page 186, Section 7.2.2, Analyses, Paragraph 3.

"Search effort" is defined as m^2 times number of years during which surveys were made.[24] *This definition should also include the number of minutes visited during a year (or the number of standardized visits during a year). Number of years times area is not a complete description of search effort unless each turbine was searched the same amount of time during each year. As it stands, search effort (quantified in terms of hours of searching per unit area) is not presented and perhaps not available. This shortcoming affects the credibility of mortality estimates, inasmuch as any differences in numbers of birds found may be related to search effort and not to differences in other variables (e.g., turbine type).*

Page 186, paragraph 4: "*Figure 7-1B illustrates the inverse power relationship between a fatality rate and search effort, which casts doubt on the reliability of a simple conversion of fatalities to fatality rates (mortality) for inter-string (or inter-site) comparisons and hypothesis testing.*"

---

[24] ~~How and when, does the amount of time spent on transects looking for carcasses (or number of visits per year) factor in?~~

Doesn't fatality rate imply deaths per unit of time? Not unit of area? And even more appropriate may be to express as deaths per hours of turbine operation *(if available)*, because flying into moving turbine blades is the primary cause of bird deaths.

Page 188.

The predictive model is flawed. The variables examined are clearly not independent and so summing the accountable mortality values across variables (p. 188) must necessarily overestimate the predicted impact. All model results are suspect because of this flaw. Furthermore, this is a complex study with many potential confounding factors, yet the development of the predictive model strikes us as simplistic and fails to account for such effects.

Page 189, Figure 7-2 through 7-4 and 7-8 through 7-18.

It is not always evident what the figure caption "count" means in these figures. It seems to be number of turbines, mostly.[25]

The words "search effort" are used in captions for measurements that really *are* the number of years during which searches were made, multiplied by a search area. This measurement ignores the number of visits (or hours) that each area was searched and assumes that there would be no variation in the number of dead birds found with greater or fewer visits during a year. *It does not seem prudent to assume that no variation exists in the number of dead birds found with greater or fewer visits during a year.*

It is also peculiar that in this analysis, the authors use rotor swept area as a measurement of turbine size, rather than MW rating. We prefer the rotor swept area as a method of standardization.

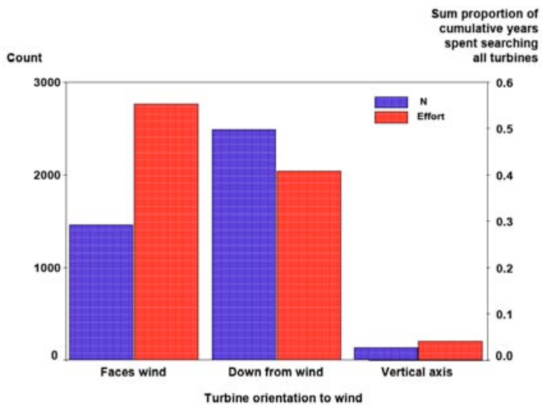
Page 205, Table 7-3.

On what basis were the two groups "Hawks" and "Raptors" segregated? *The AOU checklist (7th ed.), which we consider the standard reference for bird classification, does not identify a group "Raptors". The Order Falconiformes, which is designated "Diurnal Birds of Prey", includes the Family Accipitridae, which is designated "Hawks, Kites, Eagles, and Allies". Owls in the Order Strigiformes include the Family Strigidae that includes an important study species, Burrowing Owl. As currently written the report does not provide sufficient information to know taxonomic designations of the species that comprise "Hawks" and "Raptors".*

Figures 7-2 thru 7-4 and 7-7 thru 7-13.

---

[25] ~~Is that correct? It also seems that~~



These figures are all misleading. The adjacent bars suggest direct comparisons, yet the opposing scales are not comparable. As an example, in Figure 7-8 (above) the left scale (count) maximum is 3,000, which is 74% (=3000/4675) of the total number of turbines, whereas the right hand scale maximum is 60%. This imbalance of scales makes the effort bars taller than they ought to be. *This information can be presented together in the same graph, but the scales should be comparable because it presents a comparison to see whether any particular turbine orientation was over- or under-sampled relative to its incidence.*

Tables 7-1 through 7-3.

The results for a large number of the Chi-squared tests in Tables 7-1 through 7-3 that are suspect because too many of the expected values for individual categories presented in Appendix C are less than 5. The authors mention this fact on p. 206 but present the tests anyway. The test ought not to have been done.

The individual turbines within the same string are not independent and just as in the ANOVAs this fact needs to be accounted for in the Chi-squared analyses. In the analysis of seasonal differences the repeat visits are not independent and that needs to be accounted for also. *This is not to say that analyses cannot be completed, but they must account for the nested nature of the data.*

Comment deleted.[26]

Page 215.

The conclusion about rock piles does not seem to be adjusted for different mortality rates in different years, and for all the other factors that differ between the samples?

Page 222, Second paragraph.

The model was not validated by withholding a subset of data then using those data to check the accuracy of the model. Such validation would be desirable. [27]

---

[26] ~~Page 210, Wind Turbine Attributes, First paragraph.~~
~~What was the rationale for excluding turbine model from the tests?~~

On p. 222, the authors ask the reader to assume "our predictive model are relatively precise" yet provide no justification for the assumption. The authors appear to be ignoring the possibility of false positive predictions. There are two aspects to a predictive model; correctly identifying as "dangerous" turbines where fatalities were found (called sensitivity) and correctly identifying "non-dangerous" turbines where fatalities are not found (called specificity). While the model for Golden Eagles has a sensitivity of 82%, the specificity is 50%. The authors argue that the model identifies a collection of "dangerous" turbines. The model's ability to correctly identify a turbine that actually has an associated fatality depends on both the sensitivity and the specificity.

A calculation, using Bayes Theorem, can be used to answer the question, what is the likelihood that more searches would "*add many more wind turbines to the pool of wind turbines documented to have actually killed members of each species*?" (p. 222, line 4). To perform the calculation, one must assume an average fatality rate. Here is a table of hypothetical fatality rates for and corresponding likelihoods that a "dangerous" turbine will be found to have killed one or more Golden Eagles.

Fatality rate	0.001	0.01	0.05	0.1	0.25	0.5
Likelihood	0.002	0.016	0.079	0.153	0.352	0.619

To interpret this table, consider this example: With an average fatality rate of 5% (.05 in the table), prior to applying the predictive model one would expect about 5% of turbine searches to produce a Golden Eagle fatality. If the searches were restricted to "dangerous" turbines (as identified by the model) then one would expect to find Golden Eagle fatalities in 8% (.079 in the table) of the searches. Thus, the model increases the chance of finding Golden Eagle fatalities from 5% to 8%. *Then* one can conclude about 92% of the turbines identified as "dangerous to Golden Eagles" will **not** have an associated fatality.[28]

Page 223, Table 7-8.

[29]Because the physical attributes of operating turbines manifest the lethal force in bird deaths, it may have been instructive to use only those variables identified in framework component #1 to develop a predictive model with AIC methods. Similarly the same approach may be applied to the other framework components as outlined at the beginning of this review to determine which variable(s) contributed to bird mortalities. From results of the four predictive models, perhaps an overall model could be developed that used the most important variable(s) from each component model.

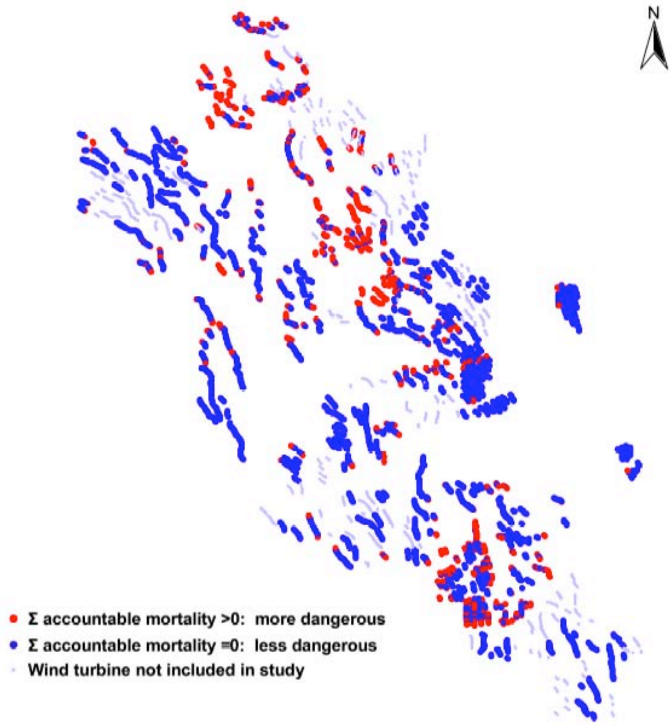
---

[27] ~~It is not clear if a subset of data was withheld from the data that was used to develop the empirical models so that they could be validated. On Page 243 (first paragraph) it is stated that "…. 472 strings that were used for developing the model." This implies that an effort was made to validate the model. Perhaps this could be confirmed and further elucidated.~~

[28] ~~and hence would not be "dangerous"~~

[29] ~~The authors appear to be selective about inclusion of "important" variables. Using the Golden Eagle as an example. The variable 'Part of wind wall' (p < .05) yet the variable 'Tower height' (p < .05) was not. The accountable mortality for 'Position in string' was reported in the table as 19 while in Appendix C it is given as -18. There were several other similar occurrences with other variables in the list.~~

Page 237, Figure 2-27.



The authors conclude that "dangerous" turbines are distributed "relatively narrowly" across the APWRA. The distribution in the maps does not seem narrow to us (see Figure 7-27 above for red-tailed hawks).

Page 244.

A typographic error seems to have resulting in a duplicate discussion of rock piles near the bottom of the page.

Chapter 8 – Bird Behaviors

Comment deleted.[30]

Page 247, paragraph 2.

Was there a random order of choosing which plot was sampled next?

Page 254.

---

[30] ~~Page 246.~~
~~What is spatial distribution of 61 study plots? It seems that they are associated with turbine strings that were chosen arbitrarily, meaning that the behavioral study plots were not selected randomly. Consequently, behaviors from these plots cannot be extrapolated to other areas within the APWRA.~~

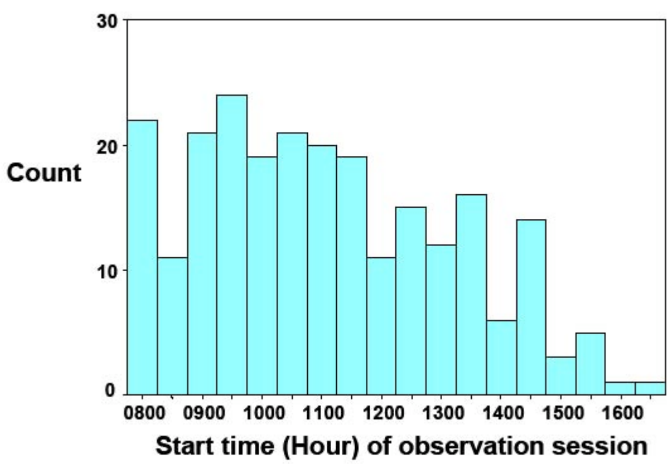
The analysis of a measured variable, such as minutes, using a Chi-squared analysis is invalid. The Chi-squared tests are not invariant to changes of scale, i.e. the results would change if the data were expressed in seconds or in hours. (Using seconds would make the tests more significant and using hours would make the tests less significant.) This invalidates almost all of the tests performed here. *The appropriate test would have been to use an Analysis of Variance.*[31]

Turbine level analysis involves pseudo-replication because turbines were sampled as strings. *A two-factor nested design could allow for investigation of turbines while recognizing their place within turbine strings.*

Page 256, paragraph 3.

Our experiences *(including over forty years conducting observational studies of birds in the field duly documented in the literature)* reinforce the authors' conclusion that the observation time for the sessions was minimal at 30 minutes. Other observational studies with which we are familiar found that 2-hour blocks, randomly assigned throughout the entire period available to observe birds, were adequate to determine reliable patterns of bird activity.

Page 257, Figure 8-3.



The behavioral surveys are biased toward morning observations and include no summer observations (Figure 8-4). This survey pattern may influence results, especially by underestimating behaviors occurring when conditions are hotter (later afternoon and in summer).

Page 331, Section 8.4.5., paragraph 5.

We agree with the authors that BACI study designs will be required to sort out effects of some variables, because the mere presence of the turbines as they are installed affects the environment and in turn affects bird behavior, which is a variable related to mortalities.

Page 332.

---

[31] ~~On p.254 the authors indicate the observed values used in the Chi-squared tests were either minutes or behavioral events. It seems that no Chi-squared tests were based on behavioral events.~~

The authors argue for taller turbines to repower at APWRA, but they seem not to consider how this will influence mortality rates for migratory songbirds. Turbines greater than 200 feet will require obstruction lighting, which is associated with increased mortality of nocturnally migrating birds. *Although mortality of night migrating songbirds at lighted towers is generally less in the West, the conclusions of this report may be applied to other areas of the country (notwithstanding the authors' best efforts to admonish readers that the study is not meant to be extrapolated to other situations).*

Chapter 9 – Recommendations

Unless and until the methodological and statistical problems described above are resolved, the conclusions reached in the report must be considered premature *from a statistical perspective. We note, however, that reanalysis of these data with other statistical methods may indeed result in similar conclusions and that the information in the report is more detailed than previous research efforts. Even as reanalysis is undertaken, the working hypotheses presented by Smallwood and Thelander can be used to implement adaptive management actions (e.g., repowering) that will test these hypotheses while potentially reducing avian mortality from collision.*

Appendix A – Measuring Impacts by MW

Page A-1, first paragraph, first sentence.

The term "confusion" may be correct but the term "complexity" also depicts the situation. It may be that it is inappropriate to try to compare mortalities between wind generating facilities because each facility has unique features for each of the four framework components, thereby preventing any reliable comparison between facilities. Conversely, the individual turbine type (and its attributes) is of utmost importance in how many birds are killed *(according to the data in tables 7-1 to 7-3)*. Variables of the other three framework components (that we outlined at the beginning) can be neutral or either increase or perhaps decrease the predisposition of birds to being killed by the turbines. But, each wind farm site is unique with specific effects of variables that cannot be fully replicated.

Comment deleted.[32]

Page A-6, paragraph 2.

Another reason to question the use of fatalities/ MW/ year is that the MW is a constant (as stated), but that the number of fatalities is variable over time and depends on amount of search effort, so that inadequate search effort in a given year will weaken the reliability of results. Authors further acknowledge (Page A-7) that this is likely that areas around wind turbines that were not searched over a long enough period will not provide a robust estimate of mortality.

---

[32] ~~Page A-2, Section 3.0 Results, paragraph 5.~~
~~This statement reinforces our earlier comments (See comments page 46, Chapter 3) that attempting to standardize by basing number of fatalities/ MW/ year instead of number of fatalities/ turbine/ year does not provide insights about effects of individual turbine types, which is the killing structure. Actually, Figure A-3b, Page A-5 depicts an even more direct metric of what kills birds — the area of rotor-swept / year, which again relates to turbine type, size and blade speed.~~

Page A-11, Section 4.0 Discussion, paragraph 1.

Indeed, it may be more convenient to express mortalities on the basis of MW / year, but information on which type of turbine and supporting structure that kills birds is not emphasized. *The authors have subsequently endorsed the use of fatalities per kWh as a metric. We are more comfortable with this, because it implicitly embodies the mechanical attributes of a turbine and duration of operation.*

Page A-12, paragraph 1.

Alternative ways to express mortalities may be by use of actual physical attributes that are involved in killing birds (i.e., those variables listed in framework component #1).