July 7, 2004

Comments on CEC report
Provided by Wally Erickson and Dale Strickland, WEST INC.

We appreciate the opportunity to comment on the report entitled “Developing Methods to Reduce Avian Mortality at the Altamont Pass Wind Resource Area”. Please feel free to call us to discuss these comments. Shawn has discussed some of these comments over some recent email exchanges. Hopefully our comments are helpful. We feel that a thorough review of this report by a statistician(s) is important, given the complexity and amount of data. We believe this review is a good starting point to that thorough review, but may fall short given our recent schedules (including vacation time), and the time frame provided for the review.

General

This is an extremely important report, and the PI’s should be commended for the amount of data that has been collected, the number of hypotheses that were addressed, and the amount of work that went in to producing this mammoth document. We also acknowledge the importance of a brief executive summary and conclusions chapter that attempts to synthesize an enormous amount of information. Given the size of this document, many readers will probably read only the Executive Summary and Conclusions Chapter. We have already seen some select lines of text taken from your work that we believe are misinterpreted or not put into proper context. The Executive Summary and Conclusions sections, as they are currently written, leave the reader with the idea that most of the conclusions and associations are clearly defined, while the other chapters seem to indicate that most of the results are not so clearly defined. We think it is paramount that enough detail is provided in the Executive Summary and Conclusions so readers do not misinterpret the basis for conclusions. Specifically, there should be a paragraph in the Executive Summary, describing the type of study (mensurative), types of analyses (correlations and associations, not causation) the number of variables considered, many of which are confounded.

Response: On the top of page xxxiii of the Executive Summary, we stated that we used tests for association. That the study is mensurative is unremarkable and warrants no additional mention. Mensurative studies are not necessarily inferior to manipulative studies. Also, it is normal for studies to experience confounding among measured variables, and so there is nothing remarkable about this phenomenon and does not merit a paragraph of discussion in the already lengthy Executive Summary.

We think it is also important to clearly state in the Executive Summary and the Conclusions Chapter that the range of turbines sampled does not represent the typically turbines currently being installed.

Response: We disagree that this is an important statement to make in the Executive Summary or Conclusions chapter. This fact is made clear enough in the report, and one does not have to read very far into it to learn that the wind turbines we studied are relatively old.

Causality

Causality should not be discussed in this document at all, other than that causality cannot be statistically inferred from a mensurative experiment like this one.

Response: We disagree that causality should not be discussed, because this is the point of performing the study in the first place. Also, it is simply not true that causality cannot be inferred statistically from a mensurative study. The purpose of using statistics is to draw inference from the test results about underlying causal relationships. This is why these statistical tests are called *inferential statistics*. Also, it is acceptable and expected that a model should guide the researcher's interpretations in terms of ecological relationships they wish to report or suggest. We feel obligated, and justified, to draw inferences from the model if they appear to us to be consistent with our overall observations, interpretations, and understanding of the ecological relationships at play in the APWRA.

It should be clearly stated that the basis for conclusions etc. is from associations between fatalities, behavior and use with physical, biological and wind turbine characteristics.

Response: See pages 5-6 and 180-184 as examples of clear statements of the tests we used and how we drew conclusions from them.

It should also be clearly stated that there is confounding and correlation among variables and that can affect these apparent univariate assocations. This is discussed in several sections, but should also be discussed in the Executive Summary, and Conclusions Section.

Response: Confounding and inter-variable correlation is common to most if not all research studies involving the measurement of multiple variables. This is widely understood and why reports typically do not make a special effort to discuss these problems in the Executive Summary.

Predictive Modeling

Conducting univariate tests is a reasonable start to developing a list of candidate variables for a "predictive model". The approach to combine results from univariate tests into a scoring system, that does not account for confounding of variables, correlation of variables and interaction of variables is fairly uncommon and is often criticized as data dredging.

Response: The definition of data dredging suggested by WEST is incorrect. Data dredging is the attempt to make too much of the data collected.

The more commonly accepted practice would be to use logistic (presence/absence of fatalities) or poisson regression (counts of fatalities), with the ability to model interactions of factors, treat variables as continuous (for example slope), etc. The influence of a variable could conveniently be portrayed as an odds ratio, and each turbine could be assigned a probability of occurrence of a fatality during a fixed time frame, or an expected number of fatalities in a fixed time frame.

Response: As explained in the report, our fatality data were collected using a differential search effort, which precluded proper use of logistic regression.

This would hopefully provide a better map of the "predicted highly dangerous turbines". In a univariate approach, the results of logistic regression and chi-square would be similar, with logistic regression or poisson regression using the turbine and not the individual fatality or individual bird minute as the unit of replication. Furthermore, with logistic or poisson regression, interactions could be tested.

Response: We disagree. The fundamental problem with whatever test is used is the sample size problem, along with differential search effort. Using logistic regression to identify supposed interaction effects would be an example of data dredging. See page 237.

The authors have said they do not think logistic regression and other types of multivariate analyses are appropriate. You have effectively developed a multivariate predictive model by combining results of univariate tests into a "scoring" or "ranking" system. You would want to limit the number of factors to consider, and number of interactions, but there are likely some very important interactions that should probably be discussed (e.g., topography and canyons). The models you have developed would be more defensible if they were corroborated with more standard approaches. The unequal sampling effort among sampled turbines (set 1, 2 and 3) can be accounted for in these multivariate approaches, using similar approaches to the adjustments you used in your chi-square analyses.

Response: Whether or not these univariate methods can accommodate unequal sampling effort does not matter, because one cannot reliably test interaction effects with small sample sizes, no matter whether logistic regression is used. If we cannot reliably test for interaction effects, then using logistic regression provides us with no advantage over the univariate tests we perfomed. See page 237.

The omnibus chi-square tests (Tables 7-4 etc) does say anything about what levels of factors are significantly different. It could mean that one level is different than one other level, that a combination of two levels are different than a combination of another two levels etc. This is another reason why fatality rates and the chi-square tests should be portrayed for the factors considered in Chapter 7.

Response: We don't understand what is being suggested here.

The method of "testing" adequacy of model is problematic and should be acknowledged. Testing models with the same data used to build the models will yield an overestimate of model fit.

Response: WEST's comment is misleading because we did not state in the report that we "tested" the adequacy of our simple models. What we said was that we compared actual fatalities to predicted levels of threat in order to assess the effectiveness of the models. This step is pretty simple and straightforward, and yes, it is prone to slight inflation, but we were aware of more significant shortfalls in our models, and so the slight inflation typical of post-hoc comparisons of frequencies was not that important.

Pseudoreplication

There is no discussion of the fact that the chi-square analyses conducted are based on assumptions of statistical independence of the experimental units. The individual fatality (Chapter 6) or minutes of bird use (Chapter 8) are considered the experimental unit in most of the analyses, and not the turbine or turbine string.

Response: This characterization of our study unit is inaccurate. The turbine and the turbine string were central to our study units. Our search effort was quantified per wind turbine, and was factored into chi-square tests accordingly, and used to generate expected values per wind turbine (or turbine string). The observed frequencies of fatalities or bird behaviors at these turbines were then compared to expected values. We recommend that WEST revisit the methods sections of our chapters 7 and 8. Also, we assume WEST made a mistake when it referred to Chapter 6 regarding this matter.

This appears to be an inappropriate experimental and results in "pseudoreplication," resulting an overestimate of the precision of estimates. For the "predictive model" analyses and most individual species, the issue is not that big of a deal, since there are very few turbines that had more than one fatality of an individual species. It is likely a bigger issue for combined groups like all birds and all raptors. That may be why you are seeing so many "significant" differences in those categories (e.g., Table 7-1).

Response: No, it is simply the law of large numbers in action.

The number of turbines sampled is the true sample size, not the individual fatalities or individual minutes of bird use (Table 8-12).

Response: This is not true. Whereas our study unit was indeed the wind turbine in terms of search effort applied, the sample size that really counts is the number of wind turbines with fatalities, or in even simpler terms, the number of fatalities.

We recommend using fatality rates or use rates and tests using the appropriate experimental unit.

Response: We strongly disagree. See our Appendix A. Converting fatality data to mortality is fine for making impact comparisons across large groups of wind turbines or through time, but inappropriate for testing the significance of main and interaction effects of measured variables. When mortality is calculated for a turbine string, the sample size of turbines and accompanying fatalities is much too small and results in many values near or at zero and a few values that are much larger than zero. This is because mortality is an inverse power function of the sampling effort used to generate the mortality estimates, influenced in part by the number of wind turbines in the string. The error terms will be so large that the lower value of the confidence interval is much less than zero, which is absurd. Using these mortality values and their inflated error terms in multivariate tests would be inappropriate, and we just won't do that.

This is likely why there were so many "statistically significant" tests. Many of those tests would likely not be "significant" if the appropriate experimental unit and error term were used. The logistic regression approach or poisson regression approach described above would use the turbine or turbine string as the unit of replication (experimental unit). Furthermore, for some factors such as tower type, rotor diameter, turbine model, fatality rates on a per turbine and per MW basis would likely help the reader have a better understanding of the differences when the fatality data are adjusted by turbine nameplate output. Larger rotor diameters were identified as more risky, but in fact, if you account for output or RSA differences, it will likely go the other way.

Response: We proposed just this pattern in our report (see Appendix A), which is why we incorporated rotor swept area into the search effort term applied to turbine strings (see page 182).

The same might be true for lattice and tubular. You advocate per MW calculations in discussing species and group fatality rates for the overall wind plant, and this should also be addressed in Chapter 7 when discussing differences in fatality rates among turbine types and turbine models.

Response: This suggestion is feasible. We could do this for the wind turbine models that were searched over sufficiently long periods. However, little additional insight would be gained because the chi-square approach we used actually uses more of the information than is used in comparing rates.

Confounding of Variables
There should be a much bigger discussion of the correlation and confounding among the predictor variables. Many of the independent variables are correlated with one another.

Response: This is true of most every field study with multiple measured variables.

Position within a string is likely correlated with slope and steepness and degree of lateral edge. Tower type is somewhat confounded with the canyon variable. For example, apparently there were no lattice turbines in canyons in the first sampling effort, while there were 142 of the 405 tubular tower turbines in canyons the same set.

Response: Not true. There were 19 lattice towers in the first set. But besides the inaccuracy of WEST's numbers, it is irrelevant that the first set of turbines had only 19 lattice towers in canyons because our analysis in this report was based on all the turbines, not just the first set. Including all the turbines, which is what we did, the number of lattice towers in canyons increases to 296.

This should be acknowledged. Could this be a plausible explanation for why tubular are considered worse than lattice?

Response: In an effort to infer interaction effects between canyons and tower type, we split the chi-square tests for association of fatalities with tower type according to whether the towers were inside or outside canyons (see the Table below). Splitting up the data this way resulted in reduced sample sizes and mostly insignificant test results, which demonstrates the effect of the law of large numbers that we discussed previously. Some of the tests should not have been performed because too few expected cell frequencies were greater or equal to 5. Despite the lack of significance in test results in some cases, an examination of the numbers in the table indicate that there may be slight interaction effects for red-tailed hawk, American kestrel, and all raptors combined, indicating slight confounding between variables in the direction suggested by WEST. On the other hand, vertical axis or tubular tower outside of canyons killed disproportionately greater numbers of burrowing owl, mallard, mourning dove, western meadowlark, horned lark, rock dove, and all birds combined. Thus, the suggestion that tower type is confounded with canyon is not compelling; confounding occurred, but varied among species and appeared more often to be strongest in the direction opposite of that suggested by WEST.

We also reported that all of our focal raptor species flew disproportionately closer to tubular towers and disproportionately farther away from lattice towers. These behavior results also contributed to our overall conclusion that tubular towers are no less dangerous to birds than are lattice towers. We made the case that tubular towers on the landscape may appear less busy or dangerous to birds as compared to lattice towers, which may result in birds deciding more often to fly through turbine fields supported by tubular towers. From our split analyses in the table below, it does not appear that canyons factor dramatically into the effect of tower type on fatalities.

Species/Group Tower type	**Outside Canyons**			**Inside Canyons**		
	χ^2	**Fatalities**	**Observed ÷ Expected no.**	χ^2	**Fatalities**	**Observed ÷ Expected no.**
Golden eagle	2.08			1.91		
Vertical axis		0	0.00		0	0.00
Tubular		11	0.95		11	0.86
Lattice		28	1.09		4	1.85
Red-tailed hawk	2.38			4.32		
Vertical axis		3	0.43		0	0.00
Tubular		45	1.03		50	0.91
Lattice		100	1.03		15	1.60
American kestrel	1.30			6.61*		
Vertical axis		1	0.42		0	0.00
Tubular		13	0.88		5	0.65
Lattice		36	1.10		4	3.09
Burrowing owl	34.60**			0.84		
Vertical axis		10	3.93		0	0.00
Tubular		25	1.57		14	1.10
Lattice		19	0.54		1	0.46
Barn owl	4.85[t]			1.92		
Vertical axis		3	2.19		0	0.00
Tubular		4	0.47		15	0.88
Lattice		22	1.15		5	1.74
Great horned owl	1.31			0.18		

Vertical axis		0	0.00		0	0.00
Tubular		4	0.80		1	1.18
Lattice		13	1.16		0	0.00
Mallard	6.26*			1.14		
Vertical axis		2	2.65		0	0.00
Tubular		8	1.69		16	1.11
Lattice		6	0.57		1	0.41
Mourning dove	4.76[t]			0.49		
Vertical axis		0	0.00		0	0.00
Tubular		12	1.62		7	0.92
Lattice		13	0.79		2	1.54
Horned lark	5.36[t]			0.35		
Vertical axis		1	1.01		0	0.00
Tubular		11	1.77		2	1.18
Lattice		9	0.65		0	0.00
Western meadowlark	9.40*			3.08		
Vertical axis		8	2.53		0	0.00
Tubular		23	1.16		28	1.14
Lattice		36	0.82		1	0.24
Rock dove	6.88*			0.24		
Vertical axis		2	0.25		0	0.00
Tubular		61	1.21		21	1.03
Lattice		108	0.96		3	0.87
European starling	1.04			0.15		
Vertical axis		4	1.46		0	0.00
Tubular		19	1.11		8	1.05
Lattice		35	0.92		1	0.77
All birds	9.00*			2.54		
Vertical axis		43	1.01		0	0.00
Tubular		307	1.15		213	0.98
Lattice		551	0.93		42	1.14
All hawks	2.55			3.41		
Vertical axis		4	0.49		0	0.00
Tubular		56	1.09		54	0.92
Lattice		114	1.00		15	1.51
All raptors	1.04			6.26*		
Vertical axis		18	1.00		0	0.00
Tubular		122	1.08		106	0.92
Lattice		242	0.96		29	1.49

Because of confounding, actual effect sizes could be larger (or smaller). Some many other factors could be confounded as well. Based on the maps, there would appear to be confounding among a higher percentage of tubular towers in canyons in the intermediate rodent control area. The areas of no control in the northwest portion of the project area has historically been associated with higher golden eagle mortality (Orloff and Flannery 1992, Hunt 2002) and could also affect interpretation.

Response: And we discuss this in the report. See Appendix B, page 267.

Does it make sense to test some of the important effects within certain subgroups of turbines, to see if the pattern of effects of physical attributes of the locations are consistent. For example, what about looking at the effects of certain variables using only the 56-100 turbines? They are distributed throughout the wind project, and turbine characteristics would not be so confounded.

Response: This could be done, but with a smaller sample size and the associated limitations.

Adjustments in the Chi-Square Analyses

Were adjustments made in the chi-square analyses for search effort of individual turbines, or for search effort as a whole for turbine sample set 1 and set 2 turbines? We believe the Seawest turbines, which are part of your set 1 turbines (?we believe), were added after you started your NREL study as well. Was this differential sampling effort for individual turbine strings specifically accounted for, or was it accounted for more generally (all set 1 turbines adjusted the same, regardless of sampling effort).

Response: Please examine pages 181 to 182. WEST misread or misinterpreted the description of our methods. The search effort applied to wind turbines was tallied per individual turbine, and not grossly across broad groups of turbines considered members of Set 1, Set 2 or so on.

Was the effort (# searches/year or # searchs/6 months) considered the same for all turbines in set 1 and set 2.

Response: See above. Read pages 181-182.

Since there was an increasing interval between searches, was that considered in the development of "predictive models"?

Response: Yes.

Inference

The conclusions and executive summary do not acknowledge the data mining and data dredging aspects of these analyses.

Response: Because we disagree on the definitions of data dredging.

Were all these variables and hypotheses decided upon a priori? With the number of variables and bird groups considered, there are expected to be many "significant associations" that are not real associations would be expected.

Response: We acknowledge this in the report.

If variables were truly independent and pseudoreplication was not an issue, you would expect 10% of the tests to be significant by chance alone, using alpha=0.10. Since pseudoreplication is an issue, you would expect even more tests indicating significant effects by chance alone. Reiterating, we believe you should acknowledge the limits of the study in the executive summary and conclusions including the pseudoreplication issue, the fact the associations do not imply causation, the multiple testing issue, and the confounding of variables.

Response: We will have to disagree on very fundamental issues regarding the use and interpretation of statistical tests, and the standard of reporting of standard analytical shortfalls of studies like this one and every other one. Some level of multicollinearity is inevitable with any study analyzing multiple measured variables, and the same is true with pseudoreplication.

Higher Use near Turbines

Two of the outcomes relate to raptor use and behavior are:

"Inter-specific variation in mortality among species could not be explained by variation in the number of flights within close proximity to wind turbines."

"some bird species spent more time flying within 50 m of wind turbines than expected, and they spent less time within 51-100 m or 101 – 300 m, which indicated that those species were attracted to the areas near the wind turbines."

Note: the first outcome contains redundant text, inter-specific variation is variation among species.

Response: Noted, and change was made to the text.

It is pretty well known that raptors use slopes along ridges and other topographic features, taking advantage of updrafts, and patterns in flight path data we have gathered at many wind projects prior to turbines being built have supported this. A great example would be a Foote Creek Rim, where we documented much higher use on the upwind site of the rim edge. Do you have data on ridges at Altamont Pass that do not have turbines?

Response: Yes, but they were not used specifically in this study for this report. We might tap them at a later date.

Could some of the "turbine attraction" be based on raptor behavior in association with updrafts? Do you have any flight path data?

Response: We suggested this as a possibility. We noted that raptors may be using the same declivity winds as used by the wind turbine owners in the APWRA.

Is it possible that siting turbines as far to the leeward side of the ridge as possible might reduce mortality? We recommend this is discussed.

Response: It is possible, but this is purely speculation. Our recommendations are based on empirically founded patterns described in our report.

Rock Piles

There are several conclusions drawn that do not appear to be supported by the data. For example, presence of rock piles are said to increase fatalities in some areas of the APWRA although the univariate tests (Table 7-1) are not significant for the 4 target species (BUOW, AMKE, RTHA and GOEA). The only group that is significant is for all birds as a group, and as pointed out above, pseudoreplication caused by treating each fatality as an independent observation is a large problem with the all bird analyses. Based on more recent conversations with the authors we understand that there was an apparent relationship with rock piles in the earlier NREL study. Some discussion of this would be important so people understand why it was not shown in the larger sample of turbines and why the effect "went away" when data were pooled.

Response: See page 210 of the report.

We recommend putting in a map of rock piles (at least string categories), like you have done for other variables. An explanation supported by data would be helpful so people understand why this particular effect (rock piles) was only observed in the earlier NREL study and not when data were pooled. We do not believe it is sufficient to say it is because it only occurred at the turbines initially studied. What is the "biological" reason for this inconsistency.

Response: The biological reason for this "inconsistency" is that the effect of artificial rock piles will occur where rock piles occur, but not where they don't occur. For example, we cannot detect a canyon effect at turbines only outside of canyons. As you increase the sample size of turbines with fatalities outside areas where artificial rock piles occur, then you dampen the effect that can be detected in the subsequent test.

It would also appear to be the case with canyons. Some initial analyses we conducted of the effects of canyons are not as obvious in the 3rd sampling effort (set 2 turbines), based on some analyses we conducted.

Response: We cannot comment on any analysis WEST conducted on our data set because we cannot know whether WEST factored in sampling effort, or whether WEST used the same standards that we did in deciding whether the sample sizes warranted a statistical test.

This is not intuitive. The variable definition is somewhat arbitrary. We do acknowledge longer sampling at the 700-1500 turbines sampled in set 1. But 6 months of sampling at 2500 turbines is a big effort (similar effort to 400 turbines sampled for three years). One would hope that patterns observed in the first sample set were also observed in the 2nd and 3rd sample set.

Response: The sets of turbines compared are from different parts of the APWRA. WEST was just alleging that we pseudoreplicated our experiment and statistical tests, and then claims that the same patterns ought to be detected between sets of turbines from different locations in the APWRA. We obviously disagree.

The outcome regarding canyon effects appears to be supported by most of the data. A clear definition of a canyon turbine should be identified. The effect may be larger or smaller when considering confounding effects such as tower type. No lattice towers were located in "canyons" at the turbines sampled for the longest period (Group 1) (see confounding above).

Response: This is incorrect. See above.

We do not believe there is an association with canyons when you only look at the Group 3 data (2500 turbines x 6 months = 1200 turbine years of effort). Is there an explanation for this? We recommend that this be discussed.

Response: Yes, the Group 3 turbines were located in a different part of the APWRA and were sampled for a much shorter time. Sample sizes per species are smaller than at Groups 1 and 2, and so small that we might not test for a canyon effect with this subset of the data considered alone.

We did not test for associations within each group of wind turbines. But to address WEST's suspicion that there may be no canyon effect in group 3 data, we went ahead and tested for canyon effect among wind turbines in Group 3. Sample sizes were adequate for performing chi-square tests only for two species, red-tailed hawk and barn owl. Turbines in canyons killed red-tailed hawks 2.6 times more often than expected by chance ($P < 0.005$), barn owls 3 times more often than expected by chance ($P < 0.01$), all raptors 1.7 times more often than expected by chance ($P < 0.005$), and all birds 1.3 times more often than expected by chance ($P < 0.05$).

Appendix A

The discussion of how search effort and fatalities/turbine/year or fatalities/MW/year is not easily understood and we believe will be miss-interpreted. We agree that the rate of a particular turbine string may not be stable after one year, but average rates of more than one turbine string would stabilize sooner than 3 years.

Response: We demonstrated with our data that the average rate derived from more than one turbine string definitely did not stabilize before 3 years.

If interest is in an average fatality rate for a wind project, and not an individual turbine string, rates would stabilize sooner.

Response: This conclusion is wrong, as demonstrated by Figure 6 in Appendix A.

Several turbine models are nearly identical in tower height, as well as in other turbine characteristics, although there are great differences in effects. Any explanation? To get a more general relationship, does it make sense to combine similar turbine models

Response: We would need more information to interpret this comment and to respond.

Only ~500 fatalities out of ~1100 "wind turbine fatalities" were listed on page 25. Are the remaining unknowns?

Response: On page 23, where we reference the Figure on page 25, we stated the following: "However, many of the carcasses showed signs of multiple injuries, and these are not represented in Figure 2-3."

Were there any trials done to test the "days since death" estimates. We believe estimation of time of death can be pretty problematic.

Response: Whether we estimated a bird's death at 22 days, but it died 35 days before, or whether we estimated a bird's death at 300 days ago, but it died at 400 days ago, really is not very important. Such errors would have made little difference to our mortality estimates and absolutely none to our fatality associations.

Standardization by Effort

Several graphs display the total number of fatalities by levels of factors. For example, winter and summer are described as being the periods of highest fatalities, and a reference is given (Figure 2-5). We recommend standardizing the data by effort for seasonal comparisons. We believe winter may have been sampled with more effort, since the last sampling effort (2500 turbines) occurred during primarily the winter (November 2002 – April 2003). Figure 2-6 might be interpreted to indicate a certain model is more risky. We believe strongly that fatality rates should be graphed, not total numbers, or have at least both.

Response: We strongly disagree, because such a graph does not inform the viewer of sampling effort, and can therefore easily mislead the viewer.

We recommend more detail on your conclusions that Altamont Pass is not an anomaly. We believe your basis for this is that raptor fatalities compared to your estimates of raptor use was not very different than other wind sites.

Response: Correct, so from that perspective it is not an anomaly.

Our analysis appears to suggest an anomaly or uniqueness of some sort. Where in the report do you show these relationships?

Response: in Chapter 4.

Our comparisons suggest 2-4 times the use and 10-20 times the fatalities when compared to Foote Creek Rim, Buffalo Ridge, and Stateline wind plants. The largest per MW rate using the more intensive fatality methods that we have used (searches more frequently and adjustments for scavenging and searcher efficiency) is around 0.10 raptor fatalities per turbine per year, while your unadjusted estimates appear to be 1 raptor/MW/year.

Response: We need more information to address this comment. We do no understand the point being made.

Furthermore, if you included TUVU in your use estimates, use differences would be inflated even more between APWRA and other areas, since TUVU use is high at the APWRA

.

We think you would agree that fatality data for "all birds combined" are not very comparable to other studies, due to the high uncertainty from wide search intervals and scavenging and searcher efficiency biases.

Response: No, we disagree.

We wonder if the methods used previously for documenting use in the APWRA (Orloff and Flannery 1992, 1996) were the same methods that you used. When you documented interactions of turbines and birds in your behavior studies, were you focused on a 360 degree area around the observer, or were observers focused on a smaller more focused area in front of the observer. This would be important in describing use including changes from your studies to the previous studies.

Response: We and Orloff and Flannery explained our methods. We did use 360 degrees.

Make clear the adjusted fatality estimates reported in executive summary and conclusions is based primarily on experimental bias studies in Oregon and Washington.

Response: Referencing other studies is not typically done in an Executive Summary.

Conflicting Results

Some results, on face value, seem to conflict with each other. It is stated that repowering with the tallest of towers should reduce mortality, but in the first bullet, it is suggested that turbines on taller towers are worse. Most people reading this would think those two results conflict with one another.

Response: The tallest towers in our study are shorter than the towers proposed for repowering.

Also, is the outcome that most flights occur below rotor plane of new generation turbines for all birds, raptors, all diurnal birds etc. What about nocturnal migrants and bats?. Most of our data on flight altitudes at other wind projects put a large percentage of raptors in the rotor plane. The difference could be behavior.

Response: True, we focused mostly on raptors when recommending that taller towers be used. Other species might get killed more often, such as gulls. We did not address bats, since our study was focused on birds.

KVS-33 turbines are discussed as some of the most dangerous turbines in the Executive Summary, but this is not mentioned in Table 7-4 except for one case (AMKE). Does not seem like a strong statistical basis for such a strong statement?

Response: We contend that associations with turbine attributes tell more of the story than do the associations with turbine model. The KVS-33 turbine has the largest rotor diameter, for example, and shows up repeatedly for species further down Table 7-4 from the associations with turbine model.

It should be noted that the KVS-33 fatality rate appears to be approximately twice as high for AMKE compared to 56-100, but the nameplate MW is 4 times higher, suggesting the per MW fatality rate is one-half that of 56-100. Replacing 4 56-100's with 1 KVS-33 would reduce mortality by 50% based on your data for AMKE. This was the only species where KVS-33 were listed in Table 7-4, so it would appear the fatality rates for these turbines would likely be quite a bit lower than the other smaller turbines for most other bird groups. This is another good example of why the data should be portrayed in Chapter 7 and in other chapters in terms of fatality rates by turbine type etc. It is easier to understand than "the accountability mortality %" for most readers.

Response: Fatality rates may be easier for some to understand, but they are inappropriate in cases of uneven sampling effort due to an inverse power function between the rates and the sampling effort used to estimate those rates.

Relocation and Shutting Down Turbines

Page 237, end of page. For golden eagles, red-tailed hawks and American kestrels, it was stated that elimination of most turbines may be the only way to "substantially reduce mortality". What is the definition of a substantial reduction in mortality, and is this considering all management measures, or only relocation?

Response: We repaired this sentence for another reviewer.

What about repowering, what about other factors such as painting, and range management that have not been tested. What if effects are not additive, what about interactions?

Response: We agree that there are a lot of 'what if's?' But we still believe that measures should be taken to reduce mortality to the extent feasible. Measures with greater uncertainty of their effectiveness should also be monitored, so that we can make changes as needed.