

Response to questions by the settlement parties
Questions delivered by Sandi Rivera by email on October 3, 2007

Audubon Questions to SRC

Responses are inserted by Julie Yee using Word's Track Changes
October 9, 2007

Note: I have tried to not to repeat responses that I felt were adequately
covered by Shawn Smallwood's responses dated October 6, 2007.

Comments by J. Burger, taking into account both Julie and Shawn's comments

1. In SRC member Yee's 9/14/07 "Guide to September 2007 Analyses on Shutdown Effect," she points to an analysis of the data that shows kestrel fatalities are 76% less during shutdown compared to operation.

Question: Is this the type of analysis of the data that persuaded the SRC to recommend a four month seasonal shutdown? Is it correct to say that the SRC believes there will be a substantial reduction in avian mortality of the four focal raptors if there is a four month seasonal shutdown beginning on November 1, 2007?

See Smallwood 10/6/07 response.

I also agree with Smallwood's responses

Question: Does the SRC believe it is probable that the companies are currently on track to reach a 50% reduction in raptor mortality as defined in the Settlement Agreement? Why or why not?

See Smallwood 10/6/07 response. Also, rough estimates based on the M16b document put the current mortality in the neighborhood of 1100 to 1270. See response to Question 1 on page 7.

The settling parties should understand that statistical approaches are just that, and we can never be absolutely sure or completely accurate in our predictions.

2. SRC member Yee's Guide also states:

For the analysis on page 2 of M16, there were over 15000 turbine survey visits, and the estimated reduction is highly significant (p-value ,0.0001).

Question: Would the SRC explain the significance of this information so a layperson can understand what is meant by "highly significant"?

The term "highly significant" in the context of a statistical test has a very specific meaning defined in terms of probability. In layman's terms, it is very unlikely for the sampled data to show this magnitude of reduction if the shutdown-related reduction was not actually occurring throughout Altamont over the course of the sampling period. Another way to think about this is: It is possible for there to have been no shutdown-related reduction and yet, by some fluke of probability, the data showed a large reduction (i.e. 76.01% reduction on p. 2 of M16); the probability of this is < 0.0001 (p-value). The smaller the p-value, the more statistically significant is the result.

In other words, the more statistical significance, the more likely the mortality will be as predicted. A P value of < 0.0001 is much much better than a p value of 0.01.

Question: How is the "shutdown effect" estimate derived in the M16 analysis? For example, on page 1, from where is the value of parameter "a" (0.000225) or parameter "b" (0.2399) derived?

The meaning of the parameters a and b are defined in P55. The numeric values on page 1 are estimated by SAS software fitting the Poisson model to the data. The software derives the numerical values for a and b that best fits with the data.

Question: Might different statistical tools or models change the analysis? For example, under "Specifications," the M16 report states that "Distribution for Dependent Variable" is "Poisson," and that the "Optimization Technique" is "Dual Quasi-Newton." What do these tools entail? Would other models/tools/techniques change the conclusions supporting seasonal shutdown? Different statistical tools or models typically result in different numerical estimates. When the differences have to do with computational algorithms, such as choosing to use the Dual Quasi-Newton, then the estimates typically do not differ to the extent of changing ones conclusions. Different models can produce widely different estimates; in those situations it is usually because of differences in assumptions that the models require of the data. A model's estimates are only as reliable as the assumptions on which it is based. As an example, the M16 results differ somewhat from Wally Erickson's analysis of the data because he assumed more restrictions needed to be placed on the data. See below for more details about this.

The Dual Quasi-Newton is the default algorithm selected by SAS software to search for and converge to best values of a and b. As for the use of a Poisson distribution, this is a common choice for modeling count data. The Poisson distribution is a probability distribution for a dependent variable (in our case, fatalities counted) that is integer-valued and non-negative. The Poisson distribution is derived from problems in statistical science dealing with counting the numbers of random occurrences of an event, such as a fatality.

It is always the case that scientists and statisticians select the tools they feel are most appropriate to the analysis in question, which has to do with both the data that are available, the type of data, and the questions being asked.

Question: Why do the various units of measurement (MW or Turbine) and the differing sampling units (Turbine or String) change the "fit statistics" in the M16 report? What does it mean to change "fit statistics" based on turbine vs. mega watt units?

These particular fit statistics (-2 log Likelihood, AIC, AICc, and BIC) are relative fit statistics, which means that they can be used to compare the fits of different models only when the models are comparable. In the M16 document, the only models that are comparable are models that feature the same dependent variable (i.e. models of AMKE should be only be compared to other models of AMKE) and the same sampling unit (i.e. per-turbine models cannot be compared to per-string models). This means we cannot say, for example, that the AMKE model fits better or worse than the BUOW model;

similarly, we cannot say that the string sampling model fits better or worse than the turbine sampling model.

As it turned out, the Poisson model approach I used (P55 document) allows the models on fatalities/MW to be compared to the models on fatalities/Turbine. As an example from M16, the fit statistics of the model on pp. 1-2 (AMKE fatalities/turbine) can be compared to the model on pp. 3-4 (AMKE fatalities/MW). Of the statistics shown here, I recommend using the AIC (Burnham and Anderson, 1998, or 2001 2nd edition). A smaller value indicates a better fit. In the example, the per-turbine model has a better fit (AIC=315.6) than the per-MW model (AIC=321.9) and the difference in AIC is 6.3. In this 2-model comparison, the per-turbine model is considered the "best" model and per-MW model is the runner-up with Δ AIC=6.3. Burnham and Anderson (1998) wrote "As a rough rule of thumb, models for which $\Delta \leq 2$ have substantial support...Models having Δ of about 4 to 7 have considerably less support." A common approach to quantifying the level of support for a model is to translate the Δ into a weight of evidence defined as "AIC weight" = $\exp(-\Delta/2)/(1+\exp(-\Delta/2)) = \exp(-6.3/2)/(1+\exp(-6.3/2)) = 0.041$. This means that 4.1% of the evidence favors the per-MW model while the remaining 95.9% of the evidence favors the per-Turbine model. Similarly, most of the other pairs of comparisons in M16 favored the per-Turbine model. In a couple of cases, such as the GOEA (pp. 17-18 vs. 19-20, or pp. 21-22 vs. 23-24), the AIC was lower for the per-MW model, but not by much (Δ AIC \leq 0.6). With Δ this low, the AIC weights are nearly even (57% vs. 43%). I would not recommend using the GOEA analysis to select a single model for use. (Ref: Burnham, K.P., and D.R. Anderson. 1998. Model Selection and Inference: A Practical Information-Theoretic Approach Springer-Verlag New York, Inc., New York, NY).

Question: The M16 analysis reports % mortality reduction in terms of reduction per turbine, per MW, or per string. However, these respective reductions vary. What would be the overall % reduction in mortality from seasonal shutdown?

In situations where there are several models offering different estimates of reduction, analysts will often conduct model averaging. This means an average is calculated across models, with each model weighted by AIC weight. These weighted averages have not yet been calculated, but in many of the M16 comparisons it can be easy to anticipate what they will be. When two models have similar estimates, such as when the AMKE models estimate reductions at approximately 70% (pp. 1-8 in M16), then the average will also be approximately 70%. When two models lead to disparate estimates and have different fits (large Δ), such as the RTHA showing a 37% reduction based on per turbine fatalities vs. a 9.5% reduction based on per MW fatalities (pp. 25-32), then the average will be weighted towards the model with the lower AIC; for the RTHA, the Δ is 17 which leads to weights of 99.98% vs. 0.02% in heavy favor of the 37% reduction based on the per-turbine model.

Note that AIC-based model averages can be used to combine model results only when the models are comparable. I should repeat that models based on per-string sampling cannot be compared to models based on per-turbine sampling, therefore these models cannot be averaged together. Fortunately, M16 shows no large differences between per-turbine and per-string sampling (i.e. model results of AMKE on pp. 1-4 do not differ much from results on pp. 5-8, and same is true for corresponding models of other species).

Question: Why is the M16 report's standard error so high, and the t-Value so low, for Golden Eagles?

The number of fatalities counted for GOEA is relatively low, compared to other focal species. These low counts lead to highly uncertain estimates of the shutdown effect, as the high standard error shows. The t-value = Estimate/StandardError, so a high standard error will lead to a low t-value.

Given the vastly different mortality numbers (and probably the real presence and population estimates for the four focal species), the statistical probabilities and variances will change.

3. In Wally Erickson's September 12, 2007 Updated Seasonal Shutdown Analyses, he states,

For red-tailed hawks, the fatality estimate is 2.6 times higher at the operating turbines compared to the non-operating turbines.

Question: Does the SRC agree with Erickson's analysis of the impact of operating turbines on red-tailed hawks?

I accept Erickson's analysis as an illustrative comparison, but I also consider it to be severely limited in information-content because many of the fatalities had to be eliminated from the tables due to uncertainties about the timing and cause of death. His statement about a 2.6 times increase is consistent with the findings from my analysis that showed a significant shutdown effect of reducing RTHA fatalities. I do not believe the effect is exactly 2.6 (which translates into a 72% reduction), and I don't think Erickson intended for anyone to interpret his statement that way.

There are at least two ways to look at the data. All of the fatalities can be considered (regardless of cause of death, Wally's approach), or best scientific judgement can be used to attribute cause of death (Julie's approach) based on field notes and data

4. While acknowledging that a seasonal shutdown is likely to reduce mortality for red-tailed hawks, Erickson also claims "[t]he data for the other 3 raptor species is too limited to make any inference."

Question: Is Erickson's claim about the adequacy of the data for the three other focal raptor species inaccurate?

I agree with Erickson to the extent that the data is inadequate for the three other species given the approach he was using to analyze it. His approach filtered out too many fatalities that could not be assigned to an operating or non-operating period, or to cause of death. This rendered the remaining data uninformative for three species. My approach used a model in which all fatalities would be allowed into analysis. I consider these fatality counts to be informative even though we can't classify them into operational and non-operational periods. Under this model, if shutdown is reducing fatalities, then the searches at intervals containing more operational days would produce more fatalities than searches at intervals containing fewer operational days. This is why I consider all of the data admissible for analyzing and estimating the shutdown effect.

SRC Member Yee's Guide to the September 2007 Analyses on Shutdown Effect states,

For example, in the first analysis on page 2 of M16, the estimated rates of kestrel fatalities per turbine is 76% less during shutdown compared to during operations with a standard error of 15.7%. Whether the shutdown is one month or two months in winter, the percent reduction in fatalities is predicted to be 76% less.

Question: Are Erickson and Yee looking at the same data? Please explain.
For all intents and purposes, Erickson and I used the same data. He sent me fatality data in early September 2007. For about a week after that, leading up to the September 2007 meeting, WEST continued to review the data for errors, I ran a few diagnostics to identify data entry errors, I conferred with Erickson, and he notified me of corrections. It's possible that not all corrections were relayed to me or integrated into my analysis, but I'm not aware of any such case. The types of corrections we were finding affected so few records that I have no reason to suspect they would influence the direction of results. I do recommend that all analyses be checked and finalized using the most correct version of data.

Question: What work have the SRC members (and/or the Monitoring Team members) done since receiving the incomplete analysis from Erickson (August 2007) to increase our understanding of how the seasonal shutdowns have been affecting raptor mortality at the APWRA?

See Smallwood 10/6/07 response. I also view this Q&A as a continuing effort by the SRC to help the Settlement parties understand the effects of seasonal shutdown on focal raptor species mortality.

The SRC views all of the work as our attempt to continue to understand and refine the analysis, and to provide the parties with our best scientific analysis and approach.

Question: The SRC has now recommended that the companies implement a 4-month shutdown from November 2007 to February 2008 in an effort to reach the 50% reduction requirement. What factors led to this recommendation from the SRC? How confident is the SRC that this winter shutdown will have a positive impact on reducing avian mortality?

See Smallwood 10/6/07 response for factors.

On confidence, I can only relay what the statistical results tell me. Document M16b estimates the impact of a 4-month winter shutdown on annual fatality rates for each of the four focal species separately and combined. Estimates #12, #24, #36, #48 are the estimated annual reductions for the AMKE, BUOW, GOEA, and RTHA respectively. Estimates #55 and #62 are the estimated annual reductions for the combined focal species (BUOW excluded and included respectively). Several members of the SRC have expressed the need to separate BUOW from the other species. Estimate #55 shows the estimated reduction on the three species (AMKE, GOEA, and RTHA) combined. The estimated reductions across all versions of models (i.e., per turbine or per MW) range from 19 to 25% for these species. The 90% confidence interval for these reductions are completely negative, meaning that we can be at least 90% confident that there is a positive impact on reducing mortality for at least

these three species regardless of the model used. By comparison, a 2-month shutdown is estimated to have about half that reduction (est. #53).

I agree with Smallwood. Our decision (at least from my perspective) was made with the intent of helping the settling parties arrive at the agree-upon 50% reduction.

Question: Are there other measures that the SRC can recommend with similar confidence that will have a positive impact on reducing avian mortality?

There is not yet enough data on any other measures. Sufficient studies have not been done to estimate mortality from other measures.

Erickson's 9/12/07 Update also points out some problems with the actual on the ground monitoring.

Question: What is needed to improve the monitoring of avian mortality during the seasonal shut down?

It's important to synchronize the timing of switching between operation and non-operation status to coincide with the monthly search, so that fatalities could be classified into the correct operating or nonoperating interval. The analyses of M16, documented in P55, worked around this problem. However, the models are complex. If there had been better synchronization, the data would have been easier to examine and interpret, and Erickson's analysis would have been more useful.

We cannot emphasize enough the important of synchronizing the shutdown with data collection. Otherwise, it is difficult to assign particular mortalities to a given condition. Immediately after any change in treatment, a search should be conducted.

Question: How does the SRC envision phasing in the seasonal shutdown to match the monitoring team's on-the-ground capabilities?

See Smallwood 10/6/07 response.

Question: How often will the turbines be monitored during a four month seasonal shutdown?

See Smallwood 10/6/07 response.

Question: With all the turbines shutdown does the existing monitoring team have the resources to gather the necessary data, so the SRC can evaluate the effectiveness of a four month shutdown? If not, what does the SRC/Monitoring Team require to ensure that the turbines are monitored per the protocol during a 2007-08 winter shutdown?

If all the turbines are shutdown for four months in winter, it compromises the ability to gain further data on the wintertime shutdown effect because none of the turbines would be left operating to make the comparison. People might try using year-to-year comparisons, but those analyses will always be potentially confounded by naturally occurring year-to-year variation.

Partly, the SRC response was based on recognizing the importance of reaching the 50 % mortality (over optimizing the study design).

Question: Can the mortality data from a four month seasonal shut down be compared to the mortality data from the ½ - ½ seasonal shutdown?

I believe certain comparisons can be done. I do not share the same opinion as Smallwood 10/6/07 that the comparison would necessarily be confounded by the experimental design. I tend to believe that, with regular monthly searches, we can always estimate fatality rates, and compare them between years regardless of the differences in mitigation that occur in those years. However, I believe it will be difficult tying any mitigation effect to a year-to-year difference. For example, if a 4-month winter shutdown occurs next year and the total annual number of fatalities is the same as it was last year (i.e. zero difference), then this does not mean that 4-month shutdown was not effective. It may be that next year turns out to be a year with a naturally higher fluctuation, and the extended shutdown prevents the number of fatalities from fluctuating as high.

I believe that the decision to implement a 4-month winter shutdown requires an acceptance that the reduction effect is real beyond reasonable doubt. If implemented, this presumes no need for further analysis. Further analyses might require year-to-year comparisons that will potentially be confounded with annual fluctuation, though this is not to say that they can't or shouldn't be done.

I also believe some comparisons can be made, although there are limitations. However, there are always limitations because no two conditions are ever exactly alike (due to seasonal differences). We just have to be aware of the limitations.

QUESTIONS FOR THE SRC

SETTLEMENT BACKGROUND:

The Settling Parties agreed, 3. (a), that the baseline for reduction in raptor mortality was 1300. This number was established from data published in Smallwood, K. S., and C. Thelander. 2004, page 74:

The Wind Power Companies are required to achieve a 50% reduction of this 1300-raptor mortality figure within 3 years of the effective date of the Settlement, the date established as Nov. 1, 2009:

The Settling Parties have agreed upon a method to determine the percentage reduction in raptor mortality:

Question 1: Has the SRC seen results of any field monitoring for any complete year where mortality for the focal raptor species has been determined for that year?

M16b includes annualized fatality rates, by separate and combined species. These rates are model-based estimates derived from the same model used to assess the shutdown effect (document P55) and data received by WEST covering the period from October 2005 through May 2007 when the 2-month winter shutdown was in effect. The estimates for AMKE, BUOW, GOEA, and RTHA are estimates #7, 19, 31, and 43 respectively. The estimate for all four species

combined, under a 2-month shutdown, is #57. The estimated rates are 0.2535/turbine/yr for the per-turbine model or 2.48/MW/yr for the per-MW model. The SRC has not yet received precise numbers for the numbers of turbines in operation or total MW capacity, but I understand there are roughly 5000 turbines in operation Altamont-wide among settling and non-settling parties, and about 450 MW. This puts the annual mortality for that period at around $0.2535 \times 5000 = 1267.5$ (based on per-turbine model) or $2.48 \times 450 = 1116$ (based on per-MW model).

Question 2: Has the SRC approved any Searcher/Scavenger (S/S) scaling factors to be applied to the data results observed in Question 1 above? The model-based rates above have not been corrected for searcher efficiency, so the actual rates will increase after correction. The rates were corrected for scavenger removal using the correction factors developed by Smallwood (P44). To my knowledge, this is the best available information we have on the scavenger removal rates, however the SRC may later approve using improved factors after the special study on AMKE and BUOW is completed.

We have not.

Question 3: Based on responses to Questions 1 & 2 above, wherever applicable, please complete the following information relating to the agreed focal raptor species mortality figures at the Altamont:

This data cannot be filled in at this time.

| Date by Species of Focal Raptors | Raw Count of Focal Raptors Approved SRC S/S Scaling Factor | Raw Count of Focal Raptors Adjusted Count |
|----------------------------------|--|---|
| | Adjusted Count of Focal Raptor by species | |
| 2004 | AMKE | AMKE |
| | BUOW | BUOW |
| | GOEA | GOEA |
| | RTHA | RTHA |
| 2005 | AMKE | AMKE |
| | BUOW | BUOW |
| | GOEA | GOEA |
| | RTHA | RTHA |
| 2006 | AMKE | AMKE |
| | BUOW | BUOW |
| | GOEA | GOEA |
| | RTHA | RTHA |
| 2007 | AMKE | AMKE |
| | BUOW | BUOW |
| | GOEA | GOEA |
| | RTHA | RTHA |

Question 4: Please identify all information used to calculate the Approved SRC S/S Scaling Factor for each year such a calculation was made and please provide such information if it is not available on the SRC Web site.

n/a.

Question 5: Assuming the Approved SRC S/S Scaling Factor is equal to or below 2.5 in all years, what is the trend of the Adjusted Count figures relative to 1300?

n/a.

Question 6: If the Approved SRC S/S Scaling Factor for any year is above 2.5 please identify all information used to calculate the scaling factor and provide such information if it is not already available on the SRC Web site?

n/a.

Question 7: Does any of the above mortality data include any mortality counts deemed to be caused by parties that were not signatories of the Settlement Agreement?

See Smallwood 10/6/07 response. I agree.

Question 8: Does the SRC's analysis of seasonal shutdown compare mortality data from one year to the next or does it compare mortality data from operating turbines vs. non-operating turbines within a particular year?

My analysis pools data from the entire time period (Oct 2005 - May 2007) and compares the operating vs. non-operating effect for turbines within that time period.

Question 9: It appears that the SRC applied at least 3 filters to the Winter Shutdown data and possibly annual monitoring. How were they weighted in the analysis? Can the data be repeated and verified? Is the information available on the SRC Web site?

I don't understand the question about weighted filters. The analysis should be repeatable and verified.

Question 10: Document M13, mentioned in the recommendation, states, after discussing effects on red-tailed hawk mortality, that "[t]he data for the other 3 raptor species is too limited to make any inferences." Why did the SRC discount this conclusion? Is documentation of this decision available on the SRC Web site?

See my response to Audubon's question #4.

Question 11: Is the SRC aware of any fluctuations in bird presence, use, or bird populations at the Altamont or regionally during the monitoring period compared with other periods? Are such fluctuations taken into account, or are factors such as bird use and population considered static for the purpose of evaluating the monitoring data?

See Smallwood 10/6/07 response.

The Wind Power Companies reserve the right to ask additional questions once we have received answers to the questions above.

The SRC appreciates the continued interest of the settling parties, and that all concerned (the settling parties and the SRC) are striving to reach the 50 % mortality. We wish to point out that the SRC was not formed at the beginning of the monitoring work, which made it difficult for us to have adequate oversight. Further, the data were not presented to us in a timely manner, which partly accounts for the lateness of our recommendation. It would have been far better for this

discussion to have occurred six months ago, and we appreciate your patience. Please, however, understand our frustration also.