Review of "Bird Fatality Study at APWRA October 2005 to September 2007," by Alameda County APWRA Monitoring Team

Written by Julie Yee
March 5, 2008

Short questions for the MT:

1. I'm considering how you might adjust for turbine non-operation (top paragraph on p. 11). The forthcoming operating hours information will help, but your survey also contains valuable information that was not reported in the recent draft. I'm interested in this information because you might use it to develop an crude adjustment, even in the absence of operating hours information (though it might be rough). On each survey, under what conditions did you not search that site (i.e. turbine and tower gone, turbine disassembled, blades missing,…)? How many such sites were there?
2. My early thinking was that fatalities from the first search would be tossed out (as a clearance survey) because you wouldn't have the same confidence as later searches that the fatalities occurred in a well defined interval. On the other hand, I realize you were able to sometimes backdate with confidence. How did you handle data for the first survey? And when did you backdate?
3. Speaking of first survey, is the "T" in the second column of Table 4 (p. 15) a typo meant as "Fall 2005"? Or is there another meaning to T?
4. p. 28. Under Winter Shutdown, create a subsection for Exceptions. There were other exceptions as well, correct? There is the one set of turbines that are shutdown every year the entire winter. Please mention those as well as any other exceptions.

Summary recommendations for analysis and report:

1. Keep equation 1.
2. Incorporate uncertainty about R and p.
3. Clarify R_C.
4. Account for different R's for different lengths of survey periods.
5. Reconsider the unit of analysis.
6. Make special considerations for WRRS data.
7. Refine confidence intervals and report standard errors.
8. Refine the fatality count model.
9. Decide on assumptions about mean fatality rates.
10. Sort out the reasons for estimation inconsistencies.

<u>1. Keep Equation 1</u>. I stand by equation 1. However, the derivation of this equation is not explained much anywhere. Since the public and especially settling parties have questions about the inverse relationship between ($R \times p$) and M_a, and about the validity of multiplying $R \times p$, I will offer my own explanation for this equation.

The goal is to estimate the actual number of fatalities. I will term this X, which is somewhat like M_a in equation 1, but to start this simple I want to make the distinction that X represents a single random variable, such as what random number of fatalities might occur and settle at a single turbine plot on a single day, whereas M_u is used to represent the mean of X. The reality is that X is unobserved, because a proportion of actual fatalities will disappear due to scavenger removal and a proportion of fatalities that remain on the day of the search will be missed due to detection error. Rather than observing X, we end up observing Y=the counted number of fatalities. The purpose of equation 1 is to identify the relationship between X and Y with respect to rates of scavenger removal and searcher detection.

Now, I think it would be helpful to describe the situation in terms of introductory probability. In order for a fatality to be counted, two events must occur: The carcass has to avoid getting scavenged (i.e. remain on the plot by the time of the next survey), and then it must get detected given that it remains on the plot. The probability of this can be expressed in terms of elementary probability as $\Pr\{A \cap B\}$, where A is the event that the carcass has remained, B is the event that the carcass has been detected, and $\cap$ basically means "and". In other words, both A and B must occur in order for the carcass to be counted.

It would be incorrect to state that $\Pr\{A \cap B\}=\Pr\{A\} \times \Pr\{B\}$, unless A and B are independent events which they are not. Members of the settling parties have raised questions about this and therefore the validity of multiplying R and p in the denominator of equation 1. I will explain why equation 1 is still valid. R represents the probability that a carcass is still remaining, or $\Pr\{A\}$. p represents the *conditional probability* that the carcass is detected *given that it is on the plot*, or in probability notation $\Pr\{B|A\}$. It is a fact that $\Pr\{A \cap B\}=\Pr\{A\} \times \Pr\{B|A\}$ regardless of whether or not A and B are independent. This is why $R \times p$ represents $\Pr\{A \cap B\}$, the probability that a carcass escapes scavenging and detected by surveyors, or in other words the probability that a fatality that occurred on a plot is eventually counted.

So now we have the counted fatalities (Y) and the probability that an actual fatality is counted ($\Pr\{A \cap B\}$). There is also a number for the actual fatalities (X) but we won't ever know what it is (remember X is a random unobserved variable). We do know that Y represents the portion of X fatalities that are eventually counted and that the probability of getting counted is $\Pr\{A \cap B\}$. So,

on average, $Y=X \times R \times p$, and therefore, on average, $X = \dfrac{Y}{R \times p}$ (eq. 1).

If M_u represents Y on average, and M_a represents X on average, then we can simply say $M_a = M_u/(R \times p)$. Although we will never know X, we can statistically infer M_u and therefore also M_a by applying this equation.

Example: Suppose Y=10 fatalities are counted and R and p are both 0.2. The probability that any particular carcass avoids scavenging and then detected is $0.2 \times 0.2=0.04$. So, the 10 counted fatalities probably represent roughly 4% of the fatalities that actually occurred. We can't know the number of actual fatalities, but we know it is likely on the order of 10/0.04=250.

Recommendation to MT: The equation is fine as far as I see it. I recommend elaborating, as necessary, on the equations or calculations in your analysis, or provide a reference where the information can be found. Otherwise it will not pass peer review.

2. Incorporate uncertainty about *R* and *p*. R and p are estimated from independent research and there is uncertainty surrounding these estimates, measured in terms of standard error. We cannot assume to precisely know R and p any more than M_u. The current analysis incorporates error for M_u, but I believe the errors for R and p are ignored. By ignoring our uncertainty about R and p, then the analysis will overestimate the precision of M_a estimates. Although the settling parties hinted they are not concerned with precision (they are only concerned with whether the estimated reduction falls above or below 50%), I hold that accurate precision estimates are important for any meaningful scientific assessment of data. Estimates of M_a can be quite sensitive to fluctuations in R and p. By incorporating the uncertainty over R and p, then that will allow for a more realistic assessment of uncertainty in M_a.

Recommendations to MT: Don't simply do a sensitivity analyses as it will not be adequate. Sensitivity analysis entails varying R and p up and down and examining the change in M_a (often displayed in a large table or spreadsheet). Because of the small size of R and its inverse relationship to M_a, then small changes in R can easily lead to large changes in M_a. For example, changing R from 0.05 to 0.03 will increase M_a by 67%. The issue is "how likely is R=0.03 rather than 0.05?" and therefore "how likely is it that true M_a is 67% higher than our calculated estimate?" This is awkward to answer with sensitivity analysis, because it doesn't incorporate likelihoods or probabilities. So, specifically, incorporate uncertainty of R and p by incorporating the probability distribution of potential R and p values (perhaps using normality assumptions and standard errors) to evaluate how this uncertainty impacts the resulting probability distribution (uncertainty) of M_a estimates. To do this, I'm considering recommending a hierarchical model (see below).

3. Clarify R_C. It's easy to misunderstand R_C as presented in the report, so I encourage more clarity here. First of all, R_C is defined as the average of (R_i/100). Note that R_i is in percentages and R_i/100 is a proportion. The R in my example above is a proportion, or probability, so it is like the R_i/100 in the report. The R_C is defined as the average of the R in my example, or in terms of report notation it is $R_C = \sum_{i=1}^{I} R_i \big/ (100I)$. Although, conveniently, one is just the average of the other, there is an important distinction between R_i, which corrects for scavenging on fatalities that occurred exactly i days prior to survey, and R_C which corrects for scavenging on fatalities that occurred throughout the 1 to I days prior to survey. Note that 1 day prior to survey is the end of the search period and the I^{th} day prior to survey is the beginning of it, so the range 1 to I covers the entire search period.

The end goal is to estimate the fatalities from a period, or $\sum_{i=1}^{I} X_i$ where I extend my notation from equation 1 by defining X_i as the number of fatalities on the i^{th} day prior to survey. Then $\sum_{i=1}^{I} Y_i = \sum_{i=1}^{I} X_i \times (R_i/100) \times p$, where Y_i is the number of counted fatalities that occurred on the i^{th} day prior to survey, $R_i/100$ is the proportion of fatalities from the i^{th} day prior to survey that were not scavenged, and p the detection probability is assumed constant. In this scenario, the observed datum is $\sum_{i=1}^{I} Y_i$, the number of counted fatalities occurring across days 1 to I of the survey period. The individual Y_i and X_i and $\sum_{i=1}^{I} X_i$ are not known, but we aim to estimate the mean of $\sum_{i=1}^{I} X_i$. The fatalities on a particular search plot are assumed to occur randomly at a mean rate of μ fatalities per day over the course of an I-day search period so that the mean of X_i is μ, the mean of $\sum_{i=1}^{I} X_i$ is $I \times \mu$, and the mean of $\sum_{i=1}^{I} Y_i$ is

$\sum_{i=1}^{I} \mu \times (R_i/100) \times p = \mu \times p \times \sum_{i=1}^{I} (R_i/100) = \mu \times p \times I \times \sum_{i=1}^{I} (R_i/(100 \times I))$. Then the mean of $\sum_{i=1}^{I} X_i$, or $I \times \mu$, is estimated as

$$\frac{\sum_{i=1}^{I} Y_i}{p \times \sum_{i=1}^{I} R_i/(I \times 100)}.$$

The second part of this denominator explains the use of averages for the carcass remains rates, i.e. $R_C = \sum_{i=1}^{I} R_i / (100I)$. The mean of $\sum_{i=1}^{I} X_i$ represents the adjusted mortality (M_a in report) and the data $\sum_{i=1}^{I} Y_i$ represents the unadjusted mortality (M_u in report), leading to the equation $M_a = M_u/(R_C \times p)$, also a variant of equation 1.

Note: it has been incorrectly believed by members of the settling party that the use of R based on 37 or 44 days requires or assumes that the fatalities occurred exactly 37 or 44 days prior to the survey period, which would be an obviously unrealistic assumption. Their concern would be valid if the R term in the denominator of equation was R_i, with i=37 or 44. However, the R term in the denominator is R_C. This requires that the fatalities occur uniform randomly over the survey period.

Recommendation to MT: Always clearly distinguish between R_i and R_C. Avoid using plain R which is ambiguous. There may be exceptions but be careful with this. Indicate that the R in the tables on page 6 is actually R_C. It would also help to print the entire table or cite the source. I found a complete table in the Appendix of the Smallwood 2007 publication in JWM.

4. Account for different *R*'s for different lengths of survey periods. The analysis is based on R derived for 44 and 37 day lengths, which represent the average lengths of survey periods through March 2007 and after March 2007 respectively. There is not a huge difference in estimates of R for 37 and 44. But I'm more concerned that the variation around the 37 and 44, particularly the 44, is being overlooked.

Consider a hypothetical example. The mean length of survey period during Oct 2005-Mar 2007 is 44 days with a SD=18.6. I don't know what the range is, but assuming a normal distribution then most of the observations will fall within 2 SD, or 7-51. The distribution is probably skewed due to the 30-day target tending to become extended rather than shortened. So, for purposes of illustrating the argument, I will guess and suppose a range of more like 25-80 days. The R_C's for small raptors for 25, 44 and 80 days are 0.37, and 0.22 and 0.12, which means most of the unadjusted rates are corrected for scavenging by a factor of 1/0.22 = 4.55 when they ought to be corrected for scavenging by a range of factors between 1/0.37 and 1/0.12, or 2.71 and 8.33.

The effects of overcorrection and under-correction can offset one another, but there can still be a net bias. The net bias can be positive or negative. Positive biases can occur if the over-corrections (using 44 days for <44-day periods) outweigh the under-corrections (using 44 days for >44-day periods). Heavy overcorrection could occur when the majority of search periods are below the mean of 44 days (very likely when the search periods are skewed). Large under-corrections can occur for search periods that extend far beyond the 30-day target interval. I suspect there to be more over-corrections than there are under-corrections and this to be a potential contributor to upward biases in the small raptor estimates. I can't be certain without a closer examination of the data.

<u>Recommendation to MT</u>: Report the range of search interval durations, as well as the mean and SD which you already have. Assign separate R_C to separate lengths of survey period for greater accuracy and reduced chance of bias. Do not sum total unadjusted fatalities prior to correcting with $R_{C.}$ It's ok to sum the total unadjusted fatalities for data description purposes, but not for purposes of analysis. Note for the settling parties: because of the variation in R_C, then the Total Adjusted Fatalities would not simply be a straight product between Total Unadjusted Fatalities and its correction factors.

<u>5. Reconsider the unit of analysis</u>. In Feb 2008, the SRC recommended that the unit of analysis be fatalities/string/season. The reason is that fatalities cannot always be confidently assigned to smaller units such as to turbine and to survey period. I'm backing off on this recommendation. The analysis I ran in September showed virtually no difference in estimates when analyzing data on a per string basis versus on a per turbine basis. Because of the confidence issue, I support the recommendation that the data be analyzed per string (i.e. aggregating the fatalities at each turbine in a string to arrive at a count of fatalities per string). But I also believe it is not critical and that an analysis on data per turbine would perform just as well. As for analyzing data per season, however, I am not comfortable aggregating the unadjusted fatalities at each survey in a season to arrive at an unadjusted subtotal of fatalities for that season. Due to several search intervals having potentially different lengths each season, then some data units per season may have no single suitable R_C correction factor. This is like the issue I expressed immediately above: do not sum total unadjusted fatalities per season prior to correcting with $R_{C.}$

There will always be the potential to misclassify a fatality into the wrong survey period. Although one can argue that misclassification of season is less likely than a misclassification of month, I would counter that a misclassification of month is less severe than a misclassification of season. Overall, I think the reasons for analyzing data by survey period are more compelling than data by season. The unit of inference can still be by season.

<u>Recommendation to MT</u>: Use fatalities/string/survey as the unit of analysis. The resulting unit of inference should be standardized to per MW, turbine, season, and year (i.e. fatalities/MW/season, fatalities/MW/year, fatalities/turbine/season, and fatalities/turbine/year). Some weighting is necessary since different strings have

different numbers of turbines and different amounts of MW, and survey periods occur in different lengths.

6. Make special considerations for WRRS data. Technically, fatalities that were removed by WRRS and retroactively included back into the MT database should not be subjected to the same R_C correction factor as the rest of the fatalities that were detected at the end of the search interval. I'm having trouble thinking of any R_C that would be appropriate, and I'm not comfortable with assigning a *p* because the probability of WRRS discovery is not known. We know the number of fatalities found, but not the number missed by WRRS. I have a couple of ideas.

The first idea is to separate the WRRS from the rest of the data. Calculate an estimate of total adjusted fatalities on surveyed plots sans WRRS. Then adjust the total by adding in the WRRS fatalities on surveyed plots. Then extrapolate these rates Altamont-wide. You can imagine the fatalities at Altamont as falling into one of two groups: those that are "destined" for collection under WRRS and those that are not. We don't know the relationship between the two and it doesn't really matter, because we know what the first group is, and we can calculate a statistical estimate for the second group using the current monitoring program.

The second idea is to include WRRS, but pre-adjust it for having been found early. For example, suppose a fatality was discovered and removed by WRRS ten days into a 30-day search interval. If it had remained in place, then it would have been subjected to an additional 20 days of possible scavenging followed by a possibility of non-detection by the survey team. What is the probability, *q*, that a carcass known to be present on the tenth day will be detected on the 30^{th} day? A reasonable equation for *q* can be derived. Either the carcass would have been counted or not, which would add either a 1 or 0 to the unadjusted count respectively, but statistically the mean or expected increment is *q*. So, the idea is to calculate *q* for every WRRS discovery, add that to the unadjusted count. Then adjust that by the usual R_C and *p*.

Recommendation to MT: I would recommend either of these ideas over the current method of assuming WRRS fatalities would have been found on the next survey, even if WRRS fatalities were few. I can't think of a flaw with either method, but I haven't thought about it long either.

7. Refine confidence intervals and report standard errors: Provide standard errors on Tables 2 and 3.

Any method that leads to negative confidence intervals is suspect. I think the problem is more likely with the choice of method rather than the implementation of the method. Although the report does not describe the method used, it's a fair bet that a normal approximation was assumed. The estimates are always positive. So, the fact that as much as ¼ of the range of some of these intervals is negative is a sure sign that the normal distribution is not doing the estimator enough justice. The negative lower confidence bounds are too low (realistically, estimated fatalities should be no lower than zero). You can simply truncate the interval at zero. However, when the lower bound is too low, it's usually the case that the upper bound of the confidence interval is too low too. What is believed to be a 95% confidence interval could actually have a much lower reliability than 95%.

Recommendation to MT: Something should be changed. If there's no error in the implementation of the current CI method, then I recommend that the method be changed because I don't think it's producing CI at the accurate confidence level for this data. Bootstrapping is a reliable alternative that will produce nonnegative confidence intervals. Another idea is to use a Poisson model similar to my September 2007 analysis.

8. Refine the fatality count model. This section is meant to formally address the issues mentioned above and to provide a framework for addressing statistical issues that can later arise. The current analysis can be expressed in terms of a statistical model which relies on a number of assumptions. I don't fully understand the monitoring team's model. If I could see how they performed their calculations, then I suppose I could translate the corresponding model and then sort out what assumptions are at the crux of their analyses.

In the meantime, I will outline a proposal of a fatality count model and discuss the assumptions. Most of these assumptions are standard or commonly invoked. Other assumptions are made at the discretion of the analyst, and there can be differences in opinion among scientists about their reliability. Some examples of assumptions include 1) R_C is known and constant, 2) observer detection (p) is known, 3) fatality rates are independent of turbine size, or conversely 4) fatality rates are dependent on turbine size. I'm going to try to make this presentation only as mathematical as it needs to be, which is not very mathematical, but there is a lot of notation.

I extend my earlier notation (comment #1), where X = actual fatalities and Y = counted fatalities. You can think of X and Y as being generated by a series of random events driven by a systematic process defined by underlying rates of avian-turbine collisions (μ, introduced in #3), scavenger removals (R_C), and observer detectability (p)[1]. In statistical terms, X is randomly generated by a stochastic process based on μ, and then Y is randomly generated by a stochastic process based on X, R_C and p. This hierarchy of processes is the motivation for constructing a hierarchical statistical model. I have discussed the relationship between X and Y for only one turbine (comments #1 and #3) and only one survey period (comment #3). Now I formalize the model by extending the discussion to a sample of turbines and accounting for uncertain R and p (but still limiting it to one survey period and one type of species). After that, I briefly discuss extension to multiple survey periods and multiple species. I also hope this paves the way to discussions about extrapolation. Consider the following steps of reasoning:

1. Define X_{td} = # of fatalities that dropped at turbine t on day d.
2. Assume X_{td} follows a Poisson distribution with mean μ_{td}. The value μ_{td} represents the rate of expected fatalities at turbine t on day d.[2]
3. Define Y_{td} = # of fatalities counted at turbine at end of survey period for fatalities that occurred on day d. Also define $X_{t.}$= # of fatalities that dropped at turbine t over the course of the survey period, and $Y_{t.}$= # of fatalities counted at turbine at

[1] There are other factors including, but not limited to, loss of information due to crippling, accuracy in identifying cause of death, geographic variation, seasonal variation, and importantly bird abundance. I hold off on these for now.

[2] For more information about Poisson distributions, look up "Poisson process" in Wikipedia.

end of survey period. Note $X_{t.}= \sum_{d=1}^{D} X_{td}$ and $Y_{t.}= \sum_{d=1}^{D} Y_{td}$, where D = # of days in survey period. Also note that $Y_{t.}$ are observable, but Y_{td} , X_{td} , and $X_{t.}$ are not.

4. Notice that Y_{td} is the portion of X_{td} that is eventually counted, and fatalities are counted with probability q_{td} defined as $q_{td} = R_{i(td)} \times p_{td}$ where $i(t,d)$ is the number of days between the fatality date, d, and the survey date for turbine t (here, $R_{i(t,d)}$ is like R_i from the Monitoring Team report, which is not the same as R_C). Therefore,…
5. …assume that Y_{td} follows a binomial distribution with binomial parameters X_{td} and binomial probability $R_{i(td)} \times p_{td}$.[3]
6. At this point, we can consider the possibility of observer detection probability p_{td} varying between turbines or search dates. Fatalities at turbines on steep hill slopes could have lower rates of observer detectability, while fatalities in open areas following grazing have higher rates of observer detectability. Variations in detectability associated with turbine location and survey timing can potentially be tied to p_{td}, but this would require extensive observer detection study. With the limited studies conducted thus far, we might be restricted to relying on the assumption that p_{td} is a constant (denote p).
7. So, assume p_{td} is constant and equal to p which is not known with certainty. We can't pretend to know p with absolute certainty, because that could lead to overstated precision and unintended biases. We can't pretend to know nothing about p because the analysis would fall apart without a correction factor. Previous study provides an estimate of p (denote $\hat{p}$) and there should also be a standard error (square root of Var($\hat{p}$)). This prior information can be fed into the model, allowing us to state exactly what we know about p and how certain we are. We could say that our prior for p is a normal distribution centered around $\hat{p}$ with a variance equal to Var($\hat{p}$), or perhaps use a beta distribution[4].
8. As for $R_{i(t,d)}$, there is also the possibility that scavenging rates vary geographically and seasonally. As with observer detection, I don't think we have enough information yet to consider that level of variability. Independent research by Shawn Smallwood revealed consistent and predictable patterns in scavenging rates as they relate to number of days carcasses were exposed. So, I believe it is reasonable to structure $R_{i(t,d)}$ as a function of days exposed, i, which in turn is a function of fatality date and survey date, which is dependent on turbine (different turbines surveyed on different dates).
9. So, assume $R_{i(t,d)}$ can be assigned by the regression equation $R_i = a + b\ln(i+1)$, where $i(t,d)$ is defined in step 4 above. The regression equation is not known with certainty. Shawn Smallwood's research estimates the vector (a, b) (i.e. ($\hat{a}$, $\hat{b}$)) and that research should also yield a variance-covariance matrix, Cov($\hat{a}$, $\hat{b}$). We could use this information to specify a prior distribution for (a, b) incorporating our uncertainty about $R_{i(t,d)}$.

[3] See Wikipedia for more information about binomial distributions.
[4] Normal and beta distributions can also be found in Wikipedia.

These lines of reasoning nearly complete the statistical model for one survey period. Since some unknown variables follow distributions depending on other unknown variables with their own distributions, then I call this a hierarchical model. This model can be extended to multiple survey periods – only the notation gets tricky. The model can also be performed separately for separate species – only the priors for $\hat{p}$ and $(\hat{a}, \hat{b})$ change. There is one last component to the model which has to do with the assumption(s) on μ_{td}, but first I want to summarize what the model is so far:
In summary, the hierarchical model can be specified in statistical software as:

$$X_{td} \sim \text{Poisson}(\mu_{td})$$
$$Y_{td} \sim \text{Binomial}(X_{td}, q_{td})$$
$$q_{td} = R_{i(t,d)} \times p$$
$$i(t,d) = \text{\# days between fatality date } (d) \text{ and survey date for turbine } t$$
$$R_i = a + b \times \ln(i+1)$$
$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{cases} \begin{pmatrix} a_{sm} \\ b_{sm} \end{pmatrix} & \text{, for small raptors} \\ \begin{pmatrix} a_{lg} \\ b_{lg} \end{pmatrix} & \text{, for large raptors} \end{cases}$$
$$\begin{pmatrix} a_{sm} \\ b_{sm} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} 121.86 \\ -34.54 \end{pmatrix}, \text{ Cov}\begin{pmatrix} \hat{a}_{sm} \\ \hat{b}_{sm} \end{pmatrix} \right)$$
$$\begin{pmatrix} a_{lg} \\ b_{lg} \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} 106.43 \\ -5.16 \end{pmatrix}, \text{ Cov}\begin{pmatrix} \hat{a}_{lg} \\ \hat{b}_{lg} \end{pmatrix} \right)$$
$$p = \begin{cases} p_{sm} & \text{, for small raptors} \\ p_{lg} & \text{, for large raptors} \end{cases}$$
$$p_{sm} \sim \text{Normal}(0.74,\ \text{var}(\hat{p}_{sm})) \text{ or Beta}(\text{mean} = 0.74 \text{ and variance} = \text{var}(\hat{p}_{lg}))$$
$$p_{lg} \sim \text{Beta}(\text{mean near 1.0 and variance} = \text{var}(\hat{p}_{lg}))$$

When we make certain regularity assumptions about μ_{td} (for example the most basic is that μ_{td} is a constant (μ) across all dates and turbines), then this model can be used to estimate μ_{td} based on input from the data $Y_{t\cdot} = \sum_{d=1}^{D} Y_{td}$. We can also estimate what X_{td} had been along with total actual fatalities $\sum_{t=1}^{T} \sum_{d=1}^{D} X_{td}$ and estimation errors. Depending on which particular camp of statistical analysis the analyst favors (frequentist or Bayesian), model solutions and estimates can be calculated by method of Maximum Likelihood (for frequentists) or Gibbs sampling using Markov Chain Monte Carlo (for Bayesians). I think the Bayesian approach would be the easiest to implement.

Recommendation to MT: Detail either your model or your calculation method at a level of detail similar to what I just demonstrated above. This information will help me and the rest of the SRC evaluate the analysis. Or, you can consider adopting the above model which I just outlined.

9. Decide on assumptions on mean fatality rates. I discuss this element of the model separately because of its relevance to the method of extrapolation across the Altamont. It is easy to believe that mean fatality rates μ_{td} (expected fatalities at turbine *t* on day *d*) vary by turbine and by day. However, it is impossible to independently estimate a different μ_{td} for every turbine and day with the current level of sampling, exhaustive as it is. The model relies on constraints on μ_{td} that ensures some constancy about μ_{td} over time and space. I've noticed several competing assumptions about μ_{td}:

1. The simplest is to assume $\mu_{td} = \mu$. In other words, the mean daily fatality rate is similar across all turbines and days. Actually, we often relax this assumption by allowing different fatality constants for different seasons or between geographic portions of Altamont (i.e. NW, SW, NE, SE). Based on this assumption, turbines have similar fatality rates even when they are sized differently. Under this assumption, it would be appropriate to calculate an expected rate of fatalities per turbine, and then extrapolate a total expected rate of fatalities by multiplying with the number of turbines in the Altamont (or totals in NW, SW, NE, and SE geographic regions if those geographic distinctions are assumed).
2. The next simplest is to assume $\mu_{td} = k\mu$, where *k* represents size of turbine in kilowatt capacity and μ is fatality rate per kW. This assumption is based on the notion that rates of fatalities increase (roughly) linearly with the turbine capacity size. Under this assumption, it would be appropriate to calculate an expected rate of fatalities per kW (i.e. μ), and then extrapolate a total expected rate of fatalities by multiplying with the number of total kW in the Altamont.
3. Assume μ_{td} varies by strata. The turbines in the sample were selected by stratified random sampling from the following strata: very small (40-65kW), small (100-150kW), and medium (250-400kW). Within strata, one might assume the fatality per turbine is constant. Under this assumption, it would be appropriate to calculate expected rates of fatalities and extrapolate separately by strata.
4. Assume a spatial and/or temporal model for μ_{td}. Here, the mean fatality rates would vary across turbines and dates in a predictable deterministic pattern. The pattern could be a function of many possible inputs including, but not limited to, a) geospatial information that key on hazardous factors associated with configurations to neighboring turbines or topography, b) turbine size or type, and c) temporally varying factors such as weather, bird abundance, and turbine operating hours. Under this assumption, the extrapolation would entail first developing a model for estimating μ_{td} for all of Altamont's turbines and at every time interval over the three-year period, and then totaling these estimates.

The first two assumptions and methods of extrapolation are consistent with the 2004 analysis by Smallwood and Thelander. To some degree, we should continue using these approaches, since the agreement defines a specific baseline value extracted from the 2004 analysis. Although there might be better methods for estimating Altamont-wide rates (possibly assumptions/methods 3 or 4), I would still place higher importance on preserving comparability with the baseline. Supposing we operate under assumptions 1 or 2, which assume no differences between strata, then I find that the fact that the current monitoring sample was selected by stratified random sampling does not preclude the validity of extrapolation methods 1 and 2. Similarly, supposing that we operate under assumption 3, then we should use method 3 for extrapolation.

Recommendation to MT: I think careful consideration should be made to assumptions on μ_{td}. I recommend that the SRC be involved. Assumptions/methods #1-3 are not difficult to implement. With #4, the possibilities are wide open. The model for μ_{td} can be as simple or as complex as needed. For example, a simple model for μ_{td} might be just a step up from assumption 1 where, instead of assuming constant rates within seasons (causing discontinuities where the season switches over), the μ_{td} is modeled to smoothly fluctuate up and down across time with a seasonal periodicity.

10. Sort out the reasons for estimation inconsistencies. I took fatalities/turbine/yr and fatalities/MW/yr from my M16b document and compared them to MT's Tables 2 and 3 (see Tables 1 and 2 below). My estimates are already adjusted for scavenging removal, but not searcher detection. So, I applied the same searcher detection corrections and performed the same extrapolation that the MT did (to 4489 turbines or to 580 MW). My numbers are considerably lower. Some differences were expected (different algorithm, and my estimates excluded summer 2007 data) but these are substantial differences that I wouldn't have expected. If anything, I expected my estimates to be biased high because they included all fatalities whereas MT only included turbine-related fatalities. BTW, all of my lowers and uppers are for 90% CI.

Recommendation to MT: Check your calculations. I'm willing to investigate by trying to reproduce your estimates using your own methods, but I would need to be detailed with much more information than the current report. To anyone who wants to investigate my calculation for errors, I can send over details and code.

Table 1. Estimated annual fatalities in the Altamont (using extrapolation based on assumption/method 1 from comment #9 above) using fatality rates on a per turbine basis from document M16b and searcher detection factors consistent with MT report.

	fatalites/turbine/year			adjusted for searcher detection			extrapolated to 4489 turbin		
	estimate	lower	upper	estimate	lower	upper	estimate	lower	upper
.AMKE	0.06721	0.0496	0.08481	0.090824	0.067027	0.114608	408	301	514
BUOW	0.1193	0.09682	0.1417	0.161216	0.130838	0.191486	724	587	860
GOEA	0.009273	0.005989	0.01256	0.009273	0.005989	0.01256	42	27	56
RTHA	0.05777	0.05011	0.06543	0.05777	0.05011	0.06543	259	225	294

Table 2. Estimated annual fatalities in the Altamont (using extrapolation based on assumption/method 1 from comment #9 above) using fatality rates on a per MW basis from document M16b and searcher detection factors consistent with MT report.

	fatalites/MW/year			adjusted for searcher detection			extrapolated to 580 MW		
	estimate	lower	upper	estimate	lower	upper	estimate	lower	upper
.AMKE	0.65700	0.48430	0.82970	0.88784	0.65446	1.12122	515	380	650
BUOW	1.17760	0.95730	1.39790	1.59135	1.29365	1.88905	923	750	1096
GOEA	0.09034	0.05860	0.12210	0.09034	0.05860	0.12210	52	34	71
RTHA	0.56160	0.48780	0.63540	0.56160	0.48780	0.63540	326	283	369