Argenta Cooperative Monitoring Group
Issue Resolution Documentation – Section 13.1 of Settlement Agreement
Interpretation of Residual Stubble-Height Data
NRST Response 4.29.16

1. Brief Description of Situation

The annual report uses invalid criteria to determine success and failure to meet allowable use criteria set in the Settlement Agreement at riparian sites. Where monitoring showed that the 4-inch stubble height criteria was exceeded the report indicates that the utilization criteria was not exceeded because the confidence interval was above the 4-inch stubble height criteria. The report should reflect that the allowable use levels set in the Settlement Agreement were exceeded where the monitoring shows that the stubble height did not reach the 4-inch stubble height criteria in the Settlement Agreement.

2. Issues among parties

Corral Creek

For example, the Corral Creek DMA measurements showed that the average stubble height was 3.6 inches but the report goes on to state:

> *When the 95% confidence interval is considered the residual stubble height is within the allowable 4-inch level.*

The annual report should reflect that the 4-inch stubble height criteria was not met on Corral Creek.

Indian Creek

The Indian Creek DMA measurements showed that the average stubble height was 3.7 inches but the report goes on to state:

> *The average measurement for key species stubble height on this site was 3.7 inches with a confidence interval of +/- 0.7 inches, which is high enough to meet the allowable 4-inch limit (Table 22).*

The annual report should reflect that the 4-inch stubble height criteria was not met on Indian Creek.

3. Options for Resolution (Minimum of two)

1. The report should reflect that the use exceeded the allowable level set in the Settlement Agreement for the Corral Creek and Indian Creek DMAs.

2. The report should reflect that the use exceeded the allowable level set in the Settlement Agreement for all of the DMAs where the upper confidence interval extends below the 4-inch stubble height standard set in the Settlement Agreement.

WWP's amended dispute:

Western Watersheds Project notes that TR 1730-1[1] (page 262) says:

> *"Because the true population parameter can be anywhere inside of the confidence interval, it is quite possible that the true population parameter has, in fact, crossed the threshold."*

It goes on to say:

> *"One approach is to decide that if any part of the confidence interval crosses the threshold you will take action, based on the possibility that the true parameter has crossed the threshold. This minimizes the risk to the plant resource for which you are managing."*

The BLM should use the precautionary principle in managing these highly degraded landscapes that are experiencing severe drought conditions and take action to protect these plant and habitat resources from further degradation.

In circumstances where monitoring indicates that the stubble height criteria have been exceeded, it should also be noted that the majority of the confidence interval lies below the 4" stubble height criteria. We feel that it is incumbent on the BLM to manage these resources with the public interest in mind and to the benefit those resources over the interests of a few ranchers.

4. NRST Recommendation to Field Manager

Background -- Statistical Principles and Interpretations

The dispute over the use and interpretation of confidence intervals can be traced back to the negotiation period of the Settlement Agreement and the November CMG meeting. In both, staff and managers from the Battle Mountain office either stated or illustrated in their reports that they did not use confidence intervals to analyze and interpret utilization data. The NRST tried on these and other occasions to rectify this misconception by pointing out that confidence intervals (see CMG 2015b, pp. 5-6) are required as pointed out in the most relevant BLM technical references. For example, Elzinga et al. (1998, p. 261) stated:

> "*Because you have taken a sample (as opposed to conducting a complete census), you will not know the true population parameter (e.g., the true mean value). You will have only your estimate of the parameter (e.g., the sample mean) surrounded by a measure of precision such as a confidence interval. Interpretation then requires you to compare the parameter estimate and confidence interval to the threshold value*"

Likewise, Coulloudon et al. (1996, p. 13) stated:

> *"In rangeland monitoring, the true population total (or any other true population parameter) can never be determined. The best way to judge how well a sample estimates the true population total is by calculating a confidence interval. The confidence interval*

[1] http://www.blm.gov/style/medialib/blm/wo/blm_library/tech_refs.Par.44521.File.dat/TR_1730-01.pdf

is a range of values that is expected to include the true population size (or any other parameter of interest, often an average) a given percentage of the time (Krebs 1989). Confidence intervals are the principal means of analyzing utilization data."

In the case of upland utilization, the BLM protocol for utilization studies and residual measurements (Coulloudon et al. 1996 – Interagency Technical Reference 1734-3) calls for the measurement of at least 20 height-weight samples along a monitoring transect to estimate utilization. Those samples might represent thousands to hundreds of thousands or even millions of individual plants within the same stratum of the use area. So the average value of 20 or more samples is not some hard, indisputable, and absolute true parameter, it is a parameter estimate, or an approximation of the larger population.

Because monitoring data are collected on a very small subset of a population of interest, it is inappropriate to use the parameter estimate alone to draw conclusions. What WWP suggests is that the parameter estimate is an absolute value of certainty. It is not. One cannot conclude that a parameter estimate of 3.7" or 3.6" definitively exceed a 4" threshold when the width of the confidence interval is +/- 0.8 inches. Elzinga et al. (1998, pp. 92-93, 233-234, and 260-262) showed that a threshold (such as an annual-use limit) cannot be considered to be exceeded until both the parameter estimate and the confidence intervals all exceed the threshold. Such a situation is illustrated by example (D) in Figure 1.

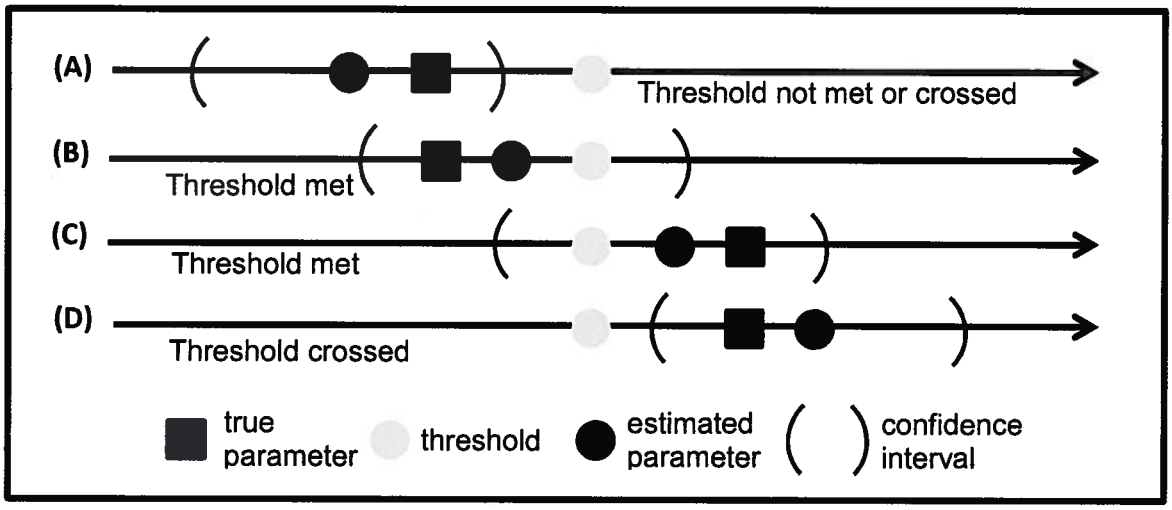

Figure 1. Illustration of how four possible monitoring results are interpreted (adopted from Elzinga et al. 1998).

Elzinga et al. (1998) illustrated how the statistical relations of four possible monitoring results are interpreted (Fig. 1). For example, in example (A), the parameter estimate along with the entire range of the confidence interval is below the threshold (in this case the end-of-season prescribed use level). In this case, the grazing use is clearly lighter than the prescribed use level, or threshold, and use at the monitoring site "meets" the prescribed use level. In example (D), the parameter estimate along with the entire range of the confidence interval is above the threshold (in this case the end-of-season prescribed use level). In this case, the grazing use is clearly greater than the prescribed use level, or threshold, and the use at the monitoring site "exceeds"

the prescribed use level. In the examples (B) and (C) (which are similar to the situations observed on Indian and Corral creeks), the confidence intervals span or "straddle" the threshold, or the prescribed use level.

In situations like (B) and (C) Elzinga et al. (1998, p. 262) pointed out: "*Now the interpretation is not nearly as clear cut.*" In these situations, there is a degree of statistical uncertainty; it is possible that the threshold has been, or has not been, exceeded. Elzinga et al. (1998, p. 262) provided some important guidance on how to interpret the situations that WWP references for Corral Creek and Indian Creek where the confidence intervals "straddle" the threshold or trigger level of use:

> *"How you will interpret situations like (B) and (C) should be **determined prior** to calculating your parameter estimate and confidence interval. In fact, you should have **decided on this prior to even initiating sampling**. **One approach** is to decide that if any part of the confidence interval crosses the threshold you will **take action**, based on the possibility that the true parameter has crossed the threshold. This minimizes the risk to the plant resource for which you are managing."* (emphasis added)

Differing Interpretations of CIs

The dispute WWP raises concerning interpretation of utilization levels is essentially the same as was made with respect to upland utilization measurements. The question in both is whether confidence intervals matter. To explore this in-depth, we will examine and report the numbers as if confidence intervals matter, confidence intervals do not matter (a practice used by BM office and requested by WWP), and by an approach that treats statistically uncertain or ambiguous measurements as "not clearly failing."

If we use confidence intervals in our analysis of riparian utilization estimates at the 10 riparian monitoring sites, we have the following results:

- 1 sites (Fire Creek) definitively[2] met the riparian use thresholds;
- 1 site (Corral Creek) is statistically uncertain or ambiguous because the confidence interval straddles the riparian stubble height threshold and the use may or may not have met the use levels, but did not definitively[2] fail; and
- 8 sites (Harry Canyon, North Fork Mill Creek, Trout Creek, Slaven Creek, Crippen Creek, Ferris Creek, The Park) clearly did not meet the riparian use thresholds (either stubble height or woody browse or both).

If we base a met/did-not-meet determination on only the definitive results, we find that 8 of 9 sites (89%) of sites exceeded riparian use thresholds and 1 of 9 sites (11%) met riparian use levels, i.e., use was less than the thresholds. One site is statistically uncertain and not figured in the calculations.

If we use the approach advocated by WWP and used by the Battle Mountain office in past monitoring reports, confidence intervals are not used. Instead, these parties treat the parameter estimate

---

[2] I.e. 'definitively' is interpreted with respect to a 95% confidence level, which means a 95% probability of being true.

as a very definitive and certain number (which it isn't according to Elzinga et al. (1998) and Coulloudon et al. (1999)) when determining if the use levels have been met or not met. Using this approach, we find the following results:

- 1 of 10 sites (10%) did not exceed the use levels, and
- 9 of 10 sites (90%) did exceed the use levels.

Finally, the approach the NRST took was the most liberal in interpreting the results. This approach was discussed during the Settlement negotiations when different parties asked how monitoring data would be interpreted in those inevitable situations where confidence intervals would straddle the use thresholds. The NRST explained that given the statistical uncertainty involved with those situations where the parameter estimate (whether it is under or over the threshold) is so close to the use threshold that the confidence interval straddles the use threshold, 'These situations would not be considered to have definitively exceeded the use levels.' What probably wasn't explicit in this discussion is whether these ambiguous cases should be treated as "passes" if they are not definitive "exceeds." In this approach the benefit of statistical doubt and uncertainty is afforded the permittees for implementing management that was so close to the use thresholds that it met the spirit and intent of the Settlement Agreement. Under this approach, we have the following results:

- 2 of 10 sites (20%; specifically Fire and Corral creeks) either met or did not definitively exceed the upland use thresholds, and
- 8 of 10 sites (80%) did not meet the use levels.

In accordance with section 6.9.1, the NRST used the methods prescribed in the BLM Technical Reference 1737-23 (Burton et al. 2011). The MIM methods specifically call for the calculation of confidence intervals. Also, the BLM Technical Reference 1730-1, Elzinga et al. (1998) pointed out that when the parameter estimate exceeds the threshold value, but the lower bound of the confidence interval does not exceed the threshold value (as is the examples cited by WWP on Corral Creek and Indian Creek), the true population parameter can be anywhere inside of the confidence interval.

The NRST has used the guidance of Elzinga et al. (1998) to interpret only those situations where the parameter estimate and the entire confidence intervals exceed the use threshold as definitive evidence (i.e. with respect to a 95% confidence level) that the use threshold was exceeded. Furthermore, the NRST has taken a view that when the monitoring data are statistically uncertain and ambiguous and the confidence intervals straddle the use thresholds, there cannot be a definitive conclusion that these data "failed" or exceeded the threshold. Some of these monitoring sites may have exceeded the use threshold, but given that the parameter estimate is so close to the threshold that a definitive conclusion could not be reached, the NRST chose a less draconian interpretation and gave the benefit of statistical doubt and uncertainty to the permittees in the "scorecard."

The purpose of the Settlement Agreement was to implement an interim management plan, not to create another enforcement tool. The real focus should be on adaptive stockmanship plans and more intensive within-season monitoring efforts to improve the overall success of parties, and more importantly to implement practices that are more likely to improve rangeland conditions. NRST required adaptive management and monitoring actions for sites that clearly exceed and for those that were unknown. The

team believes that this interpretation meets the spirit and intent of the settlement, which is less focused on keeping score and more focused on taking adaptive actions to improve resource conditions over time.

Rationale for NRST's CI Interpretation

Elzinga et al. (1998) made three important points in their guidance:

(1) How confidence intervals are interpreted in the "straddle" situations should be determined prior to sampling;
(2) One approach is described; and
(3) Decide to take action when any part of the confidence intervals crosses the threshold.

How does this guidance apply to the annual-use data from the Argenta Allotment?

(1) **Prior determination.** The NRST has been consistent in describing its approach to use confidence intervals to interpret annual-use data. These discussions began during the negotiation process of the Settlement Agreement. When the NRST was asked about those inevitable situations where the estimated parameter falls within the confidence interval of the use level (e.g., situations like Example (C) in Figure 1), Steve Smith indicated that only definitive cases (like example (D) in Figure 1) would be considered a fail, and that these statistically uncertain or ambiguous "straddle" situations would not be construed as a fail and therefore would not count against the permittees for the purposes of determining compliance with this Settlement Agreement. This approach is similar to one described in the 2004 University of Idaho Stubble-Height Report, a paper drafted by the most experienced researchers and riparian specialists in the FS, BLM, and academia and in the MIM TR (Burton et al. 2011), which drew upon the Stubble Height Review:

> "*With respect to compliance, for example, if the stubble height allowable use criterion is 4 inches and the [confidence interval] of the measurement is 0.96 inches, an observation of 3.6 inches would not imply that the criterion was exceeded*" (Burton et al. 2011, pp. 75-76 and p. 120, see also p. 129).

It is important to maintain consistency when interpreting the monitoring results. If the prior understanding was that only clear-cut fails would be counted as fails, then changing this understanding after the data have been collected, analyzed, displayed and shared with internal and external parties constitute a very poor negotiation effort. If there had been a different approach explained prior to data collection (i.e., any situation where the confidence interval straddles the threshold would be considered a failure), then it is highly probable that (i) a different threshold would have been negotiated, or (ii) no agreement would have been reached.

The CMG made written statements included in the draft of the monitoring methods paper to indicate that confidence intervals would be calculated (see section h. Data Analysis of the Key Species Method and section h. of the Height-Weight Method (CMG, 2015a)). There should have been a broad understanding that confidence intervals are required and would be used for data analysis and interpretation.

The NRST made verbal and written statements during the November CMG meeting (see CMG 2015b, pp. 5-6) in response to the failure by both consultants and BMDO staff to include confidence intervals in upland utilization data. For example,

> *"Only AG-08 in Slaven Use Area had met upland triggers by early September, though these data will need to include confidence intervals."*
>
> *"Confidence intervals should be included with all averages too."*

The NRST also included written statements in the CMG end-of-year review meeting report (see section Riparian monitoring, end-of-season, which is specific to Indian Creek and Corral Creek, and section Upland Utilization, End of Season (CMG, 2015b, pp. 5-6)):

> *Upland Utilization, End of Season – "Adam Cochran showed a PowerPoint presentation of the upland monitoring data that were collected using the height/weight (H/W) method. The presentation was incomplete because H/W curves for several key species had not yet been obtained to complete the calculation of utilization levels. Furthermore, none of the data had confidence intervals. Consequently, it was not possible to draw conclusions regarding utilization levels..."*

During the November end-of-season CMG meeting, NRST failed to realize that there was any controversy around the issue of confidence interval interpretation. At that time it was largely about the use of confidence intervals at all because BM staff were not practiced in using or reporting confidence intervals as required by Coulloudon et al. (1999, p. 13): *"Confidence intervals are the principal means of analyzing utilization data."*

As the formal disputes came in, and were amended and clarified, the NRST realized the degree of controversy over the interpretation of confidence intervals as well as their use. In order to address this, the CMG discussed the NRST's interpretation of confidence intervals during the March 9-10, 2016 meeting. The CMG agreed to use the following language in the Annual Monitoring Report to describe those sites where annual-use data "straddle" the threshold level:

> *"..., appropriate language to be included would be "monitoring data indicated with a 95% confidence interval level actual use indices and utilization level met the Settlement Agreement. Because part of this range is below the 40% allowable level, the monitoring data is interpreted as being within the allowable level."*
>
> *(DRAFT CMG meeting notes, p. 31)*

(2) **Approach**. WWP appears to contend that there is only one approach to interpreting these data. Elzinga et al. (1998, p. 262) give an example of "one approach," but they do not state or suggest that there is one and only one approach. The MIM TR (Burton et al., 2011, e.g., p. 130) describes another possible approach. Both are valid. The BLM uses different approaches at different times for different purposes.

(3) **Take action**. The Annual Monitoring Report and the notes from the CMG meeting of March 9-10, 2016, (CMG 2016) both indicate that action did occur in those sites where the

confidence intervals from annual-use data "straddle" the use threshold. For example, in the Report, the following statement is made with respect to Corral Creek:

> *"However, because part of the 95% confidence intervals extends below the 4-inch level, these sites will be prioritized to receive a higher frequency of within-season monitoring in 2016 so livestock moves can be made before prescribed utilization levels are exceeded."*

And the following statement is made in the Report with respect to Indian Creek:

> *"The residual stubble-height measurements did meet the prescribed level set in the Settlement Agreement. However, browse on the key woody species was moderate to heavy and did not meet the prescribed utilization level. Consequently, the CMG will use Appendix 1 of the Settlement Agreement to make adjustments to the 2016 grazing plan. In addition, this site will require a higher frequency of within-season monitoring so livestock moves can be made before prescribed utilization levels are exceeded."*

A general statement about increased within-season monitoring was made on p. 4 of the Annual Monitoring Report:

> *"Also, the CMG will implement more frequent within-season monitoring at the sites that did not meet the prescribed use levels [i.e. situations like example (D)] or sites where the 95% confidence interval spans the prescribed use levels [i.e. situations like example (B) or (C)] in 2015. In addition, as the within-season utilization levels reach predefined levels, the frequency of within-season monitoring will increase so the permittees can move livestock in a timely fashion before prescribed limits might be exceeded."*

And on p. 61:

> *"Within-season monitoring program at all monitoring sites with additional monitoring devoted to high priority sites (i.e. those that did not meet prescribed levels in 2015 or where the 95% confidence intervals had a range that spanned the use thresholds)."*

By raising the issue regarding the need to take action (i.e., application of the precautionary principle), WWP points out a matter which should receive more attention in the 2016 and 2017 grazing seasons. For those riparian DMAs where the annual-use measurements did not meet the prescribed use levels (Slaven, The Park, Trout, North Fork Mill Creek, Harry Canyon, Crippen, Indian Creek, Ferris DMAs) or where the measurements "spanned" the annual-use thresholds (Corral Creek DMAs) of the Settlement Agreement, the within-season monitoring attention will be intensified. When a parameter estimate is at or near the annual-use threshold, it might be a matter of only a few days or couple weeks of additional grazing for the site to exceed an annual-use threshold. The NRST and CMG believe that particular caution should be used at such sites. The sites that WWP identified in this dispute are those that were in the "cautionary" level. Therefore, to ensure continued success, these DMA sites have been prioritized for more intensive and more frequent within-season monitoring.

How frequently should these DMAs in the "cautionary" category be monitored within the grazing season? The NRST discussed this with the CMG at the 3/9-3/10 meeting. The CMG adopted a recommendation that stubble-height be measured monthly, and when the stubble-height estimate reached 5 inches or less the monitoring frequency would be reduced to 7 days, and when the stubble-height estimate reached 4.5 inches or less the permittees would prepare for an immediate move from the use area or portion of the use area represented by the DMA (CMG

2016, DRAFT CMG meeting notes, March 9-10, p. 33). For upland monitoring of utilization, the CMG agreed to monitor on a monthly (30 day) basis until a 20% utilization level is achieved at which time the monitoring frequency will be reduced to 14 days. When a 30% utilization level is reached, a move of livestock will be scheduled. (CMG 2016, DRAFT CMG meeting notes, March 9-10, p. 33).

These within-season triggers for moves coincide with roughly the lower end of the confidence interval (i.e., the most proactive end of the confidence interval) about the Settlement Agreement thresholds. This follows another recommendation in the MIM TR:

> *"It would be good for the manager to specify the range of confidence. One way to address variability is to make the lower end of the range the [trigger], and if it is reached, the operator then begins to move livestock off of the pasture. The upper end of the range could be used as a standard, which if reached, might indicate a need to adjust the grazing practice." (p. 130)*

Finally, WWP contends in the amended dispute:

> *"The BLM should use the precautionary principle in managing these highly degraded landscapes that are experiencing severe drought conditions and take action to protect these plant and habitat resources from further degradation."*

The NRST agrees with WWP regarding the need to manage highly degraded landscapes and to take action to protect resources from further degradation. However, annual-use thresholds do not constitute a management objective; and, the NRST does not believe that annual-use data are a measure of resource conditions. Instead, the NRST has advocated for the collection of long-term indicators to better determine resource conditions, and from these data establish management objectives. This work has been scheduled by the CMG for May 2016.

NRST's Recommendation Regarding CI Interpretation and 2016 Monitoring

In the explanation above, three approaches are summarized:

1. A conservative interpretation on only the definitive data (1 met, 8 did not meet) generates a scorecard of 11% pass and 89% exceed.
2. A WWP/BM approach (which ignores statistical practices and the guidance of Elzinga et al. (1998) and Coulloudon et al. (1999) to use confidence intervals) in which the parameter estimate is treated as a certain, absolute, and true estimate of the population mean (which it isn't), generates a scorecard of 10% pass and 90% exceed.
3. And a liberal interpretation in which 8 clear exceeds and 2 clear passes or ambiguous situations generates a scorecard of 80% exceed and 20% either definitively pass or did not clearly and unambiguously exceed the upland use thresholds.

NRST firmly recommends the use of confidence intervals in data interpretation and analysis, and does not support approach 2. Beyond that, the NRST does not really care how the scorecard is calculated. In year one of the interim management plan, riparian success was 10% on definitive data or 20% as determined by sites that did not definitively exceed. The NRST is far more interested in taking actions than in keeping score. Neither of these numbers (10% or 20%) changes the focus NRST placed on 2016 stockmanship plans and 2016 within-season monitoring

practices. The changes to stockmanship and monitoring are intended to improve resource conditions. The numbers used to generate the upland scorecard will not affect those plans. All three approaches would require the same adaptive actions and the same intensified monitoring plan on 9 riparian monitoring areas. There is no substantive difference in management or monitoring plans among the various approaches.

If a decision is made to interpret and portray the results in the final 2015 monitoring report differently than they were portrayed in the draft report, the NRST recommends including a comprehensive discussion of the original interpretation, as well as the alternative interpretations and results (similar to how the information is summarized within this dispute document.).

The NRST agrees with WWP that those sites, where SH spans the 4" threshold, should receive more attention in the 2016 and 2017 grazing seasons. The NRST recommends that the CMG agreed upon monitoring schedule be followed.

Finally, NRST's recommendations apply only to implementation of the Argenta settlement agreement.

Citations

Burton, T.A., S.J. Smith, and E.R. Cowley, 2011. Riparian area management: Multiple indicator monitoring (MIM) of stream channels and streamside vegetation. Technical Reference 1737-23. BLM/OC/ST-10/003+1737. U.S. Department of the Interior, Bureau of Land Management, National Operations Center, Denver, CO. 155 pp.

Cooperative Monitoring Group, 2015a. Herbaceous Removal Methods: Modifications to the Key Species Plant Method and Height Weight Method.

Cooperative Monitoring Group, 2015b. End-of-year review meeting report, November 17-18, 2015.

Cooperative Monitoring Group, 2016. DRAFT CMG meeting notes, March 9-10.

Coulloudon et al. 1996 (revised 1999). Utilization studies and residual measurements, Interagency Technical Reference 1734-3, Bureau of Land Management, National Business Center, Denver, CO.

Elzinga, C.L., D.W. Salzer, and J.W. Willoughby. 1998. Measuring and monitoring plant populations. BLM Technical Reference 1730-1. U.S. Department of Interior, Bureau of Land Management, Denver, CO.

MacDonald, L.H., A.W. Smart, and R.C. Wissmar. 1991. Monitoring guidelines to evaluate effects of forestry activities on streams in the Pacific Northwest and Alaska. EPA/910/9-91-001. U.S. Environmental Protection Agency, Region 10, Seattle, WA.

Smith, S.J., 2001. Rethinking Riparian Regrowth. Rangelands 23(3):14-16.

University of Idaho Stubble Height Review Team 2004. University of Idaho Stubble Height Study Report, University of Idaho Forest, Wildlife and Range Experiment Station Contribution No. 986, 38 p.

5. Field Manager – Acceptance with of NRST recommendation (yes X No_____).
Date Received 4/30/2016
Date of Decision 5/13/2016 **Signature of FM** _____
Comments:

After carefully considering the recommendations and rationale provided by the NRST on this dispute, and in careful coordination with the Regional Office of the Solicitor, National Upland Monitoring Lead, National Operations Center, and the Nevada State Office, I recommend to adopt the recommendations provided by the NRST pursuant to and in accordance with the MLFO responses to both NRST recommendations and the accompanying rationale provided in the attachment (see attachment "MLFO Response to Dispute Resolution: Stubble Height Confidence Intervals").

6. District Manager – Acceptance of NRST recommendation (yes_____ no____)
Date Received _____
Date of Decision _____ **Signature of DM**_____
Comments:

7. State Director – Final Determination

Acceptance of NRST recommendation (yes_____ no____)
Date Received_____
Date of Decision_____ **Signature of SD**_____
Comments:

Instructions/intent for Issue Resolution Document

Purpose of Document – This tool is intended to provide a record of how decisions are made on various issues that might arise within the CMG, including how and at what level those issues are resolved. This tool will provide transparency to options and possible consequences, including the rationales for which choices are made, and specific record of the timing and outcomes of issue resolution. It will also establish a record and help create consistency in the face of changing players over time. Close communications and an environment for resolving issues at the lowest level is encouraged. Minor issues/disagreements that are easily handled in the normal team situation will not be documented in this manner; this is intended for substantive issues that affect intended outcomes under the settlement agreement.

1. *Document the situation surrounding CMG disagreement on a given issue, including if possible, the location, essential time frames, background (including reference to Settlement Agreement section if applicable), and potential scope of consequences.*
2. *Who are the parties in disagreement; what are the differing positions and the basis for each differing side?*
3. *At least two options for resolution must be described, even if one is "no action". The CMG will be expected to problem solve and raise other possible solutions to the issue to assist the parties in disagreement. Each option will be briefly documented.*
4. *Following the problem solving activity (where needed), the NRST will select, document and provide rationale for a recommendation to the Field Manager.*
5. *If accepted, this will end the issue resolution process; if not, the NRST recommendation will be forwarded to the DM promptly, accompanied by written rationale by the FM for not agreeing to the NRST recommendation.*
6. *The DM will review the recommendation by the NRST along with the rationale statement by the FM for his/her disagreement. The DM will accept or reject the NRST recommendation; again, discussion between NRST and DM is strongly encouraged prior to determination. If the DM disagrees with the NRST, rationale should be documented and provided to the State Director along with resolution deemed more suitable by the DM.*
7. *The State Director will review the recommendation by the NRST and information in support of the differing position of the DM, and promptly render a final determination.*

Other – At each step in the process, the parties will be provided electronic copies of the various documents and determinations. The BLM Battle Mountain District Office will maintain the official record of transactions for the issue resolution process.

MLFO Response to Dispute Resolution

Stubble Height Confidence Intervals

Summary of WWP Dispute: The annual report uses invalid criteria to determine success and failure to meet allowable use criteria set in the Settlement Agreement at riparian sites. Where monitoring showed that the 4-inch stubble height criteria was exceeded the report indicates that the utilization criteria was not exceeded because the confidence interval was above the 4-inch stubble height criteria. The report should reflect that the allowable use levels set in the Settlement Agreement were exceeded where the monitoring shows that the stubble height did not reach the 4-inch stubble height criteria in the Settlement Agreement.

MLFO Response Coordination: The MLFO response was developed in coordination with the Regional Office of the Solicitor, the National Rangeland Inventory and Monitoring Lead, the National Operations Center and the Nevada State Office.

NRST Recommendations	MLFO Response to NRST Recommendations
In the explanation above, three approaches are summarized: 1. A conservative interpretation on only the definitive data (1 met, 8 did not meet) generates a scorecard of 11% pass and 89% exceed. 2. A WWP/BM approach (which ignores statistical practices and the guidance of Elzinga et al. (1998) and Coulloudon et al. (1999) to use confidence intervals) in which the parameter estimate is treated as a certain, absolute, and true estimate of the population mean (which it isn't), generates a scorecard of 10% pass and 90% exceed. 3. And a liberal interpretation in which 8 clear exceeds and 2 clear passes or ambiguous situations generates a scorecard of 80% exceed and 20% either definitively pass or did not clearly and unambiguously exceed the upland use thresholds. NRST firmly recommends the use of confidence intervals in data interpretation and analysis, and does not support approach 2. Beyond that, the NRST does not really care how the scorecard is calculated. In year one of the interim management plan, riparian success was 10% on definitive data or 20% as determined by sites that did not definitively exceed. The NRST is far more interested in taking actions than in keeping score. Neither of these numbers (10% or 20%) changes the focus	MLFO recommends to adopt approach #1, with one modification, namely that the monitoring report explain that the KMA's that have monitoring data that falls within the confidence interval be characterized as either "more likely than not" to have "met" or "not met," depending on where the data falls within the confidence interval. The report should also more clearly explain that for purposes of the adaptive management process described in the Settlement Agreement flow chart, NRST has adopted the as approach described in Elzinga (page 262) which states that "one approach is to decide that if any part of the confidence interval crosses the threshold you will take action, based on the possibility that the true parameter has crossed the threshold. This minimizes the risk to the plant resource for which you are managing." The MLFO supports this approach, which is reflected in NRST's recommendations for adaptive management, despite classifying data falling within the confidence interval as "met." NRST's recommendations describes Approach #2 (the WWP/BM approach) as "A WWP/BM approach (which ignores statistical practices and the guidance of Elzinga et al. (1998) and Coulloudon et al. (1999) to use confidence intervals) in which the parameter estimate is treated as a certain, absolute, and true estimate of the population mean." While the MLFO cannot speak for WWP, the MLFO notes that NRST's characterization of the "BM" (Battle Mountain) approach is not accurate. The MLFO is not advocating to "ignore statistical practices", or never use confidence intervals. Confidence intervals can

NRST placed on 2016 stockmanship plans and 2016 within-season monitoring practices. The changes to stockmanship and monitoring are intended to improve resource conditions. The numbers used to generate the upland scorecard will not affect those plans. All three approaches would require the same adaptive actions and the same intensified monitoring plan on 9 riparian monitoring areas. There is no substantive difference in management or monitoring plans among the various approaches.	be a useful tool, especially if enough samples can be taken so the confidence interval is small. The MLFO has never stated that it disagrees with the use of confidence intervals as a general matter. Instead, the disagreement is with characterizing all data that falls within a confidence interval as having "met" the objectives, when data within the confidence interval could actually exceed the objectives. The MLFO believes that it is more appropriate (when confidence intervals are used), to acknowledge that it cannot be definitively determined whether the data falling within the confidence interval meets the objectives, although some data will have a "more likely than not possibility" of either meeting or exceeding the threshold that was set.
If a decision is made to interpret and portray the results in the final 2015 monitoring report differently than they were portrayed in the draft report, the NRST recommends including a comprehensive discussion of the original interpretation, as well as the alternative interpretations and results (similar to how the information is summarized within this dispute document.).	The MLFO Agrees.
The NRST agrees with WWP that those sites, where SH spans the 4" threshold, should receive more attention in the 2016 and 2017 grazing seasons. The NRST recommends that the CMG agreed upon monitoring schedule be followed.	The MLFO agrees, and notes that notwithstanding the different approach taken by NRST in categorizing data within the confidence interval for purposes of making a "met" or "not met" determination under the settlement agreement, NRST has nonetheless recommended appropriate action in the way of adaptive management to address use areas where objectives were clearly not met and/or where data fell within the confidence interval. The more frequent monitoring and clearer livestock removal guidance is an appropriate response that, if implemented properly, should result in improvements this grazing year.
Finally, NRST's recommendations apply only to implementation of the Argenta settlement agreement.	The MLFO agrees.
NRST Rationale for Recommendations	**MLFO Response to NRST Rationale**
Background on use and interpretation of confidence intervals – NRST provided a general explanation regarding: (1) the	In general the MLFO agrees that confidence intervals serve a useful purpose in interpreting data that are collected and believes that it is appropriate

importance of using CIs when interpreting data, (2) the various interpretations (definitively not met, definitively met, and 'unknowns'), and (3) guidance for handling statistically uncertain situations (i.e., Elzinga et al. 1998).	to use confidence intervals in many cases. While use of confidence intervals when analyzing monitoring data is not required, it can be a useful tool, especially where a sufficient number of samples have been taken and results in a smaller confidence intervals. When confidence intervals are too large, they are of less value to those interpreting them, as noted in BLM's Technical References. Given the above, the MLFO wants to clarify that: (1) Confidence Intervals are not required or mandated to be used by BLM when interpreting monitoring data. Technical References are not policy documents and do not impose mandatory duties on the BLM. As explained in Instruction Memorandum (Washington Office IM 85-151), "Technical references are guides to completing a task and should not be construed as directives". The Technical References that apply to BLM's Rangeland Management program provide useful guidance and discussion about confidence intervals; they do not require that BLM always apply confidence intervals nor do they imply that not using confidence intervals is contrary to accepted practice norms. (2) NRST rationale states that TR 1734-3 "calls for the measurement of at least 20 height-weight samples along a monitoring transect". This statement should be clarified, in that; the Technical Reference advises to "measure at least 20 ungrazed plants to obtain a reliable cross section of ungrazed plant heights." There is no specific recommended number of samples in the Technical Reference to estimate utilization. (3) NRSTrationale states that it is "inappropriate to use the parameter estimate alone to draw conclusions." This statement is not an accurate restatement of BLM guidance. It may be useful or helpful to apply confidence intervals, when such information is available and sufficiently reliable, but it is not "inappropriate" not to do so, for example, where there is too small a sample or other factors lead to the confidence interval being too large to provide meaningful information. (4) NRST's rationalestates that per Elzinga et al. (1998), "a threshold . . . cannot be considered to be

	exceeded until both the parameter estimate and the confidence intervals all exceed the threshold." While Elzinga does say that Example (D) shows the threshold is definitively crossed, Elzinga does not state or imply that a threshold "cannot be considered to be exceeded" if it is within the confidence interval. Instead, Elzinga states that if the data falls it is within the confidence interval (e.g., Examples (B) and (C) at pages 261-262) then there is a "possibility that the true parameter has crossed the threshold," and explains that in this situation, it can be appropriate to decide that "if any part of the confidence interval crosses the threshold then action should be taken . . . [as] This minimizes the risk to the plant resource for which you are managing." Although NRST states that Examples (B) and (C) in Elzinga are characterized as "met" in the 2015 monitoring report, the fact that NRST has recommended that adaptive management actions be taken, reveals that NRST has actually treated those areas as having potentially crossed the utilization threshold. (5) The recommendations infer that only those situations where the parameter estimate and the entire confidence intervals exceed the use threshold as definitive evidence to state that thresholds are exceeded. This is true. However, Elzinga leaves the interpretation of actions needed to be taken up to the discretion of the Authorized Officer and clearly suggest that caution for the vegetative resource should be taken. John Willoughby (a co-author of Elzinga) takes the position that the prudent conclusion for Examples (B) and (C) would be that the utilization level *does not* meet the prescribed use level. Elzinga et al. (1998, page 262), state that one approach to dealing with situations B and C (in Figure 11.22 on page 261) "is to decide that if any part of the confidence interval crosses the threshold you will take action, based on the possibility that the true parameter has crossed the threshold. This minimizes the risk to the plant resource for which you are managing." Action in this case would be measures to decrease the grazing pressure on a site, such as those recommended by NRST to minimize the potential for exceeding the utilization threshold in this grazing year. The Adaptive management flow chart provided in

	the settlement agreement asked the question "were end of season use levels met?" In the answer responding "Yes" in the flow chart the recommendation is to "Continue Current Management?" In the answer responding "No" in the flow chart the recommendation is to make changes to the next years if livestock grazing was a factor. Stating that a location did not exceed the threshold, but then finding that changes were warranted is not consistent with the adaptive management flow chart in the Settlement Agreement, and reveals that while NRST categorized monitoring results within the confidence interval as "met" that it is actually recommending management consistent with thresholds not being met. The MLFO agrees that the adaptive management actions recommended for those use areas is appropriate, based on a finding that data within the confidence interval may not have met the established threshold objectives.
Comparison of different CI interpretations – NRST re-calculated % success based on varying interpretations. <u>Stubble Height</u> 1 site definitively met riparian thresholds 1 site is statistically uncertain 8 sites definitively exceeded riparian thresholds • NRST interpretation – 2 of 10 (20%) sites either met or did not definitively exceed; 8 of 10 (80%) sites definitively exceeded • WWP interpretation – (does not consider CIs) 1 of 10 (10%) sites met; 9 of 10 sites (90%) exceeded Considering unknowns – 1 of 9 (11%) sites definitively met; 8 of 9 sites (89%) definitively exceeded. One site is statistically uncertain and not figured into calculations.	(6) It is important to clarify that MLFO does not take the position that confidence intervals "do not matter" or that they should not be used. Confidence intervals can be useful tool and provide useful information. It should also be noted that using confidence intervals could result in the need for <u>more</u> grazing management changes (under the Elzinga approach and consistent with the adaptive management table), because some utilization that falls below the threshold, but within the confidence interval, could actually exceed the threshold. So values both below and above the threshold could lead to management changes, and if the confidence interval is large, this could overstate the amount of remedial management action needed, as compared to just using the utilization level to determine whether it is exceeded or not, without applying a confidence interval. (7) It should be noted that some data points within the confidence interval are more likely than not to have met or exceeded the threshold than others. It is true that rationale states it is unknown whether data within the confidence interval exceeded the use threshold; as the true value could either be above or below the threshold. However as Mike "Sherm" Karl (National Rangeland Inventory and Monitoring Lead) has stated, even when it is within the confidence interval it is possible to find that "there is a more likely than not possibility" that the

	residual stubble height was not achieved if the data is toward the lower end of the confidence interval, whereas data toward to upper end of the confidence interval has "a more likely than not possibility" that the residual stubble height was achieved. So it may not be appropriate to treat all data points within that confidence interval (especially where it is a large interval) equally unknown -- this could apply both to taking a "conservative" approach as well as to taking a more "liberal" approach of considering anything within the threshold. Each situation may require a case-by-case assessment of whether it is appropriate or useful (given sample size, or the size of the confidence interval) to use a confidence interval and how to treat data that falls within the confidence interval.
Rationale for NRST interpretation – NRST explained their rationale regarding the use of CIs and their selected interpretation in the draft year-end monitoring report. • NRST strongly supports the use of CIs in data analysis and interpretation, and has made this point since the beginning (documented in response). • For reporting purposes, NRST chose to include 'unknown' sites in the 'met' category because one cannot definitely say they exceeded. In this approach, the benefit of the doubt and uncertainty is afforded to the permittees for implementing management that was so close to the use thresholds that it met the spirit and intent of the settlement. • NRST believes the focus of the settlement is less about keeping score, and more about implementing adaptive management and monitoring to improve resource conditions over time. In the draft year-end monitoring report, the NRST required adaptive management and additional monitoring actions for sites that clearly exceeded AND for all sites that were statistically uncertain. (The agreed upon monitoring process from CMG March 2016 meeting is documented in response.)	(8) The MLFO believes that there was not a clear consensus or "prior understanding" among members of the CMG (permittees, BLM, WWP, NRST and etc…) as evidenced by WWP's formal dispute and NRST's acknowledgement that its approach differs from that the MLFO would have taken. The dispute resolution discussion is helpful in that it articulates the approach used by NRST in presenting the monitoring data that was collected and why the adopted approach was taken; while acknowledging that other approaches could have also been taken. The MLFO cautions that although NRST rationale states (italics added) that "There should have been a broad understanding that *confidence intervals are required* and would be used for data analysis and interpretation," confidence intervals are <u>not required</u> to be used. To the extent that using confidence intervals is helpful -- which BLM agrees can provide useful information for interpreting data -- the MLFO would not have categorized all data that fell within the confidence interval as having "met" the threshold, and would instead have characterized it as either "more likely than not" to have "met" or "not met," depending on where it falls within the confidence interval, and would have treated such data as "unknown" or "unable to conclude that it met the objectives." In reviewing the 2015 Monitoring Report, Michael "Sherm" Karl has pointed out that it is statistically incorrect to state "that if a portion of the 95% confidence interval is below the allowable use level, that the allowable use level is "achieved" since by definition, if any part of the confidence

	interval is above the allowable use level, then that level could also be "not achieved."
MLFO Recommendation to the Mount Lewis Field Manager for Dispute: After carefully considering the recommendations and rationale provided by the NRST on this dispute, and in careful coordination with the Regional Office of the Solicitor, National Upland Monitoring Lead, National Operations Center, and the Nevada State Office, the MLFO recommends that the field manager adopt the recommendations provided by the NRST pursuant to and in accordance with the MLFO responses to both NRST recommendations and the accompanying rationale that is specified above.	