Evaluating Alfalfa – A Blind Study of 39 NFTA Certified Laboratories[1]

Chemists: **Dr. Marcus K. Meilahn*, Seth D. Willis, and Adam B. Crooks**
Weld Laboratories, Inc., Greeley, Co 80631
Statisticians: **Dr. Robert L. Heiny (deceased)** **University of Northern Colorado**
Dr. Eric L. Heiny **Utah Valley University**

Summary

A method to prepare unground alfalfa samples with uniform composition indistinguishable from routine, cored alfalfa samples is reported in Appendix 1. These samples were sent to 40 NFTA[2] certified laboratories to determine the Relative Feed Value (RFV) by NIR and/or wet chemistry. The variation among protein and ADF results was very low (<5.6% RSD). Variation in aNDF[3] results was much higher (>7.3% RSD). The RFVs ranged from 138 (fair) to 217 (supreme); overall the RFV results were: fair (1), good (7), premium (21), and supreme (10).

A linear relationship between ADF and aNDF **(Eq. 1)** based on 16 years of NFTA alfalfa samples was calculated with an R^2 of 0.962. This relationship allows an accurate prediction of aNDF from ADF. Using the calculated aNDF, RFVs were determined within 4 RFV points from the experimentally determined aNDFs on NFTA samples. **Equation 1. aNDF = 1.12*ADF + 3.37**

Since this approach accurately predicts RFV from ADF on NFTA samples, it should work well on routine alfalfa samples. Application of this approach to the 39 laboratories in this study afforded 25 laboratories with calculated RFV within ten RFV points of the average and of these 20 were within five points.

A process to determine accurately (within 10 RFV points) the relative feed value based on ADF (RFV_{ADF}) with a probability of 96.8% based on the analysis of five samples by five different NFTA certified laboratories is reported.

Two recommendations to improve the evaluation of alfalfa are proposed: 1) use laboratories that can determine ADF at the A/B level; 2) use the calculated aNDF to determine the RFV_{ADF}.

Previous Alfalfa Articles and Studies. There have been several recent articles discussing the variability of feed analysis and laboratories.[4,5,6] These articles prompted us to report our study's results. There are four sets of data that provide information on these topics. **Study 1** (Appendix 2) is a blind study done by the National Forage Testing Association, the National Hay Association, and the University of Wisconsin in 2008. It involved 16 NIR laboratories analyzing three samples. They rated seven labs as excellent, five as good, and four as poor. Individual lab results were not published. **Study 2** (Appendix 2) is similar and was done by the University of Nebraska in 2008. They tested ten NIR labs and

reported six as excellent, two as good, and two as poor. Individual lab results were not published. Dry matter, protein, acid detergent fiber (ADF), and neutral detergent fiber (aNDF) were the major analytes in these studies. Fiber determination was a major source of error in labs rated as poor, and in Study 1 the major source of variability was aNDF. In Study 1, the % RSD for ADF was 6.48; for aNDF it was 10.74.

Study 3[5] (Appendix 2) examines the variation in rations and ration components for protein and aNDF from 14 laboratories via chemical analyses. Performed on dried, ground material, this ring test was not blind. For alfalfa, three of the 14 labs reported aNDF results that were greater than three Horwitz standard deviations (HSD) from the average of 38.31. On NFTA samples these would be failing grades. For alfalfa they stated: "a range of 34.2 to 41.3% aNDF for alfalfa hay…is not acceptable for feed evaluation or ration formulation." For the passing laboratories there were six A's, three B's, and two C's. Hristov, et al., recommended "…for feed analysis laboratories to follow the official aNDF method exactly." This study is a follow-up to these three studies.

The biggest obstacle in blind studies is obtaining samples with uniform composition. For effective blind studies, the samples must be indistinguishable from routine alfalfa samples and have uniform composition. Uniform composition refers to samples with the same percentages for stems, leaves, and fines. For a detailed discussion of these samples and their preparation see Appendix 1.

Evaluating Lab Performance by Protein, ADF, and aNDF Using Horwitz Standard Deviations. The NFTA certifies laboratories that meet requirements for four analytes: moisture/dry matter, protein, ADF and aNDF on eight samples annually (including five alfalfa samples). Results exceeding 3 HSD for an analyte are considered failing and all results are the average of triplicate analyses. The HSDs and grading scale for samples in this study are in Appendix 3**.**

Moisture was not used as an evaluating criterion. Laboratories were placed into three categories in **Table 1**: those with no failing grades, those with one failing grade, and those with two failing grades.

Table 1

Lab Results Within 3 HSD for Protein, ADF, and NDF.

Lab No.	Type	%Moist.	%Protein	% ADF	%NDF	RFV
2	NIR	6.7	23.0	29.9	35.8	171
4	NIR	6.6	22.5	28.3	33.6	185
5	NIR	7.0	23.6	27.9	34.3	182
8	NIR	8.2	25.1	27.7	33.0	190
12	NIR	5.9	23.2	28.0	33.3	187
15	NIR	11.6	25.4	29.4	35.8	171
16	NIR	6.5	22.6	28.3	35.4	176

17	NIR	8.2	24.4	27.5	35.8	175
18	NIR	7.9	25.2	28.5	35.0	177
20	Chemistry	8.1	25.3	27.0	32.1	197
23	NIR	8.2	25.4	27.6	34.2	183
24	NIR	7.1	23.4	27.3	34.6	182
25	NIR	9.2	24.8	27.7	33.4	187
28	Chemistry	8.1	24.1	28.1	33.2	188
29	NIR	7.7	23.1	27.1	33.3	189
31	NIR	6.4	24.0	28.1	33.1	188
32	NIR	7.7	23.4	29.5	35.3	174
34	NIR	7.1	23.4	29.5	34.5	178
36	Chemistry	7.1	24.1	27.8	35.6	176
37	NIR	8.7	23.5	29.4	35.0	176
38	NIR	8.4	24.6	28.0	33.5	186
39	NIR	7.6	23.7	28.9	33.2	186
40	Chemistry	7.3	24.7	28.2	33.2	188
	Average	7.7	24.0	28.2	34.2	182.3
	Stdev.	1.2	0.9	0.8	1.1	6.9

Laboratory Results Exceeding 3 HSD for 1 Analyte						
1	NIR	8.6	23.7	27.5	**36.8**	171
6	NIR	5.9	24.4	**25.6**	32	200
7	NIR	7.6	24.8	28	**31.4**	199
9	Chemistry	9.4	23.2	26.3	**31.6**	202
10	NIR	7.7	23.5	27	**30.5**	207
11	NIR	9.5	23.6	28.3	**37.2**	167
13	NIR	8.5	22.5	26.3	**30.6**	208
14	NIR	9.9	23	28.4	**36.8**	169
19	NIR	12.6	23.3	**25.5**	32.4	198
21	NIR	7.3	25	27.7	**31.3**	200
30	NIR	6.4	25	26.9	**30**	210
33	NIR	10.2	**22**	29.9	35.9	170

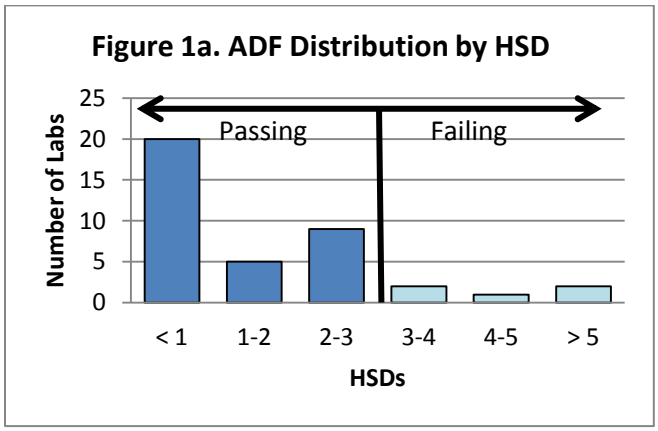
Laboratory Results Exceeding 3 HSD for 2 Analytes						
Lab No.	Type	%Moist.	%Protein	% ADF	%NDF	RFV
3	NIR	10.3	23.4	**31.6**	**36.7**	163
22	Chemistry	8.8	25.6	**33.1**	**42.6**	138
26	Chemistry	8.3	**22.0**	26.0	**29.4**	217
35	NIR	8.9	24.7	**25.2**	**37.6**	171

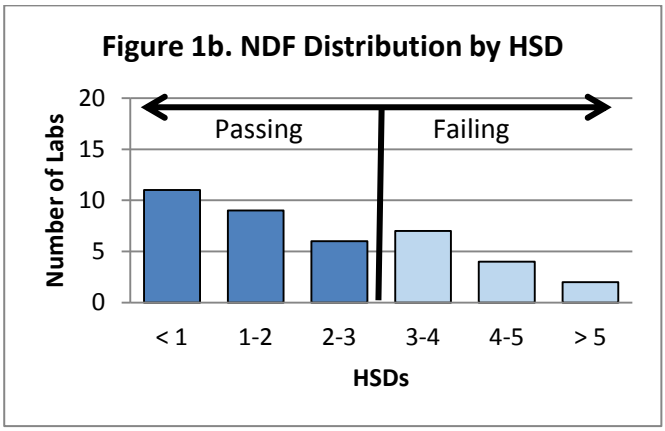
The distribution of grades in this study is presented in **Table 2.**

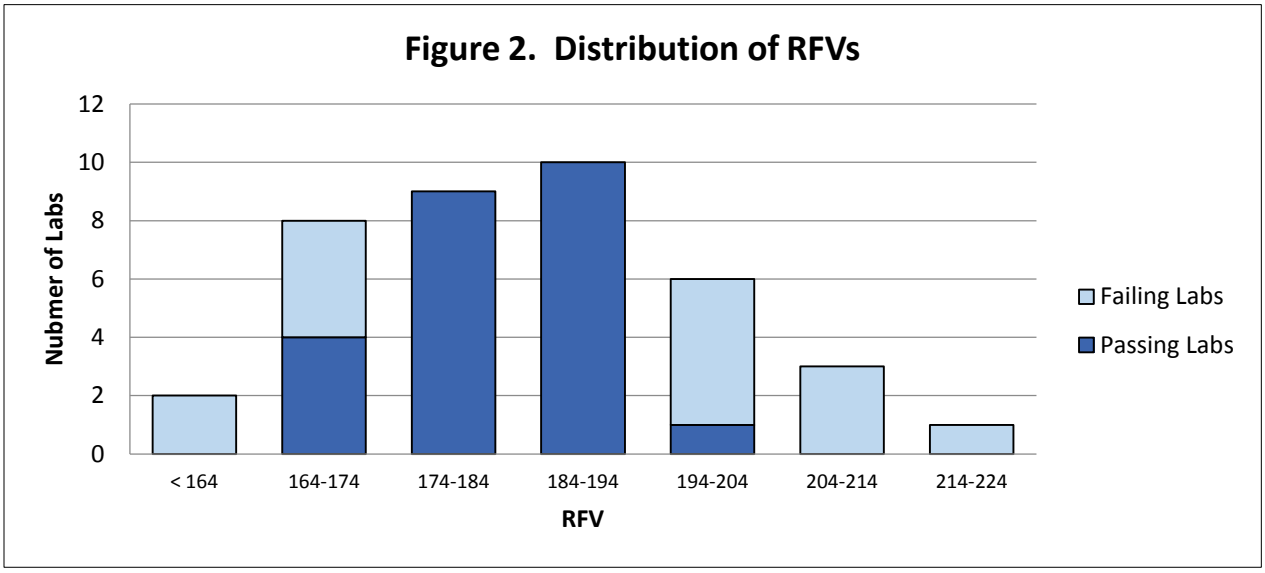
Table 2.			
Grades Based on Horwitz Standard Deviations for 39 Laboratories*			
Grade	Protein	ADF	aNDF
A	15	20	11
B	14	5	9
C	8	9	6
Laboratories Passing	37 (95%)	34 (87%)	26 (67%)
Laboratories Failing	2	5	13

*Grades based on NFTA-style data processing.

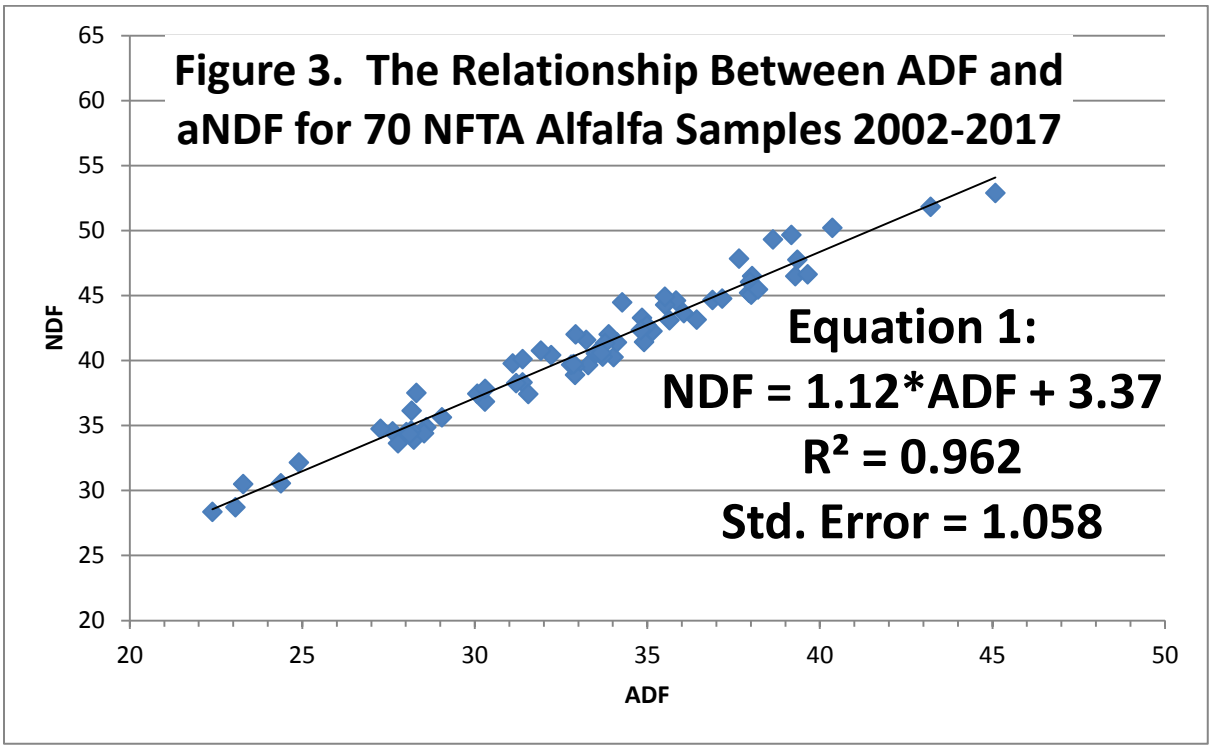
The distributions of passing and failing fiber grades are presented in **Figure 1** and the RFV distributions in **Figure 2.**







A new approach to obtaining relative feed values. Prior studies have shown that aNDF is the major source of RFV variation. In the United States, grading alfalfa by ADF is common in a number of areas while RFV/RFQ is prevalent in other areas. It is well known that there is a general correlation between ADF and aNDF. Can ADF be used to accurately predict aNDF? To examine the relationship between ADF and aNDF, we took the reference method averages for ADF and aNDF for 16 years of NFTA alfalfa samples and determined a precise relationship between ADF and aNDF. The database contains 70+ samples and represents at least 6000 individual chemical determinations for both ADF and aNDF. The standard error for the predicted aNDF is 1.06% and for RFV, 3.9 RFV points. This was established with a full cross-validation study. This means one would expect the calculated $aNDF_{ADF}$ value, on average, to be 1.06% percentage points away from the RMA aNDF value and the RFV_{ADF} to be within four points of the RFV.



For grading alfalfa, the most important RFV range in **Figure 3** is from 150 to 200+ RFV points (<32% ADF). There are 25 NFTA samples in this range. To demonstrate the effectiveness of this relationship in this range, the differences in RFV_{ADF} and the actual RFV were determined. Twenty-two (88%) of the results were within ± 3 RFV points and two (8%) differed by ± 4. Only one result exceeded the eight point range. **Equation 1** thus provides an attractive method for determining aNDF and affording the RFV_{ADF}. In **Table 3**, the RFV results are compared to the RFV_{ADF} results for laboratories obtaining a letter grade of A or B for ADF (25 labs), **Table 4** for those labs receiving letter grades of C, and the distribution by grades in **Table 5.**

Table 3.

RFV and RFV_{ADF} Data for ADF Letter Grades A or B

Lab ID	ADF	RFV_{ADF}	RFV		Lab ID	ADF	RFV_{ADF}	RFV
1N	27.5	183	171		23N	27.6	182	183
4N	28.3	177	185		24N	27.3	184	182
5N	27.9	180	182		25N	27.6	182	187
7N	28	179	199		28C	28.1	178	188
8N	27.7	181	190		29N	27.1	186	189
10N	27	187	207		30N	26.9	188	210
11N	28.3	176	167		31N	28.1	178	188
12N	28	179	187		36C	27.8	180	176
14N	28.4	176	169		38N	28.0	179	186
16N	28.3	177	176		39N	28.9	173	186
17N	27.5	183	175		40C	28.2	177	188
18N	28.5	175	177					
20C	27	187	197		**Average**	**27.8**	**180**	**186**
21N	27.7	181	200		**SD**	**0.51**	**4.00**	**10.9**

Table 4.

RFV and RFV_{ADF} Data for ADF Letter Grades of C.

Lab ID	ADF	RFV_{ADF}	RFV		Lab ID	ADF	RFV_{ADF}	RFV
2N	29.9	165	171		9C	26.3	193	202
15N	29.4	168	171		13N	26.3	193	208
32N	29.5	168	174		26C	26	196	217
33N	29.9	165	170					
34N	29.5	168	178					
37N	29.4	169	176					
Average	**29.6**	**167**	**173**			**26.2**	**194**	**209**
SD	**0.24**	**1.7**	**2.7**			**0.17**	**1.7**	**7.5**

Table 5.

Distribution of Letter Grades for ADF and aNDF

Lab Number		ADF Range		aNDF A	aNDF B	aNDF C	aNDF F
2	F	>30.0		-	-	-	2
6	C	29.3-30.0		1	2	3	-
1	B	28.6-29.3		1	-	-	-
20	A	27.4-28.6		8	6	1	5
4	B	26.7-27.4		1	-	2	1
3	C	26.0-26.7		-	-	-	3
3	F	< 26.0		-	1	1	1
39		**Totals**		**11**	**9**	**6**	**13**

Only eight NFTA labs (21%) achieved A's for ADF and aNDF. At the other end of the spectrum, ten laboratories (26%) received letter grades of C or F for both ADF and aNDF. For aNDF, 19 laboratories (49%) received letter grades of C or F.

Assigning USDA Categories. The USDA recommendations for grading alfalfa are summarized in **Table 6.** The shaded data is generated using **Equation 1** to calculate $aNDF_{ADF}$ and RFV_{ADF} based on 6000+ determinations. The USDA gives the broad ranges and, with the application of **Equation 1**, these ranges are more precisely displayed in **Table 6**. Three examples using **Table 6** to help define alfalfa categories are provided.

Table 6.

Correlation of ADF, aNDF, $aNDF_{ADF}$, RFV, and RFV_{ADF} with USDA Ranges. (shaded results from Equation 1.)

Supreme				Premium				Good		
ADF	aNDF	RFV		ADF	aNDF	RFV		ADF	aNDF	RFV
<27	<34	>185		27-29	34-36	170-185		29-32	36-40	150-170
ADF	$aNDF_{ADF}$	RFV_{ADF}		ADF	$aNDF_{ADF}$	RFV_{ADF}		ADF	$aNDF_{ADF}$	RFV_{ADF}
25.0	31.5	205		27.5	34.3	183		29.5	36.5	168
25.5	32.1	200		28.0	34.9	179		30.0	37.1	164
26.0	32.6	196		28.5	35.4	175		30.5	37.7	161
26.5	33.2	191		29.0	36.0	171		31.0	38.2	158
27.0	33.7	187						31.5	38.8	154

Example 1. If you have an alfalfa sample with an ADF of 25.0, the USDA would rate this as Supreme with an RFV greater than 185. Based on **Table 6**, the alfalfa would have an aNDF ~ 31.5 and an RFV_{ADF} of 205, consistent with a Supreme rating.

Example 2. If you select the data for sample number 36 from **Table 3** you have an ADF of 27.8, an aNDF of 34.6 and an RFV of 176. It would be rated Premium. The RFV_{ADF} is 181 in good agreement with the RFV.

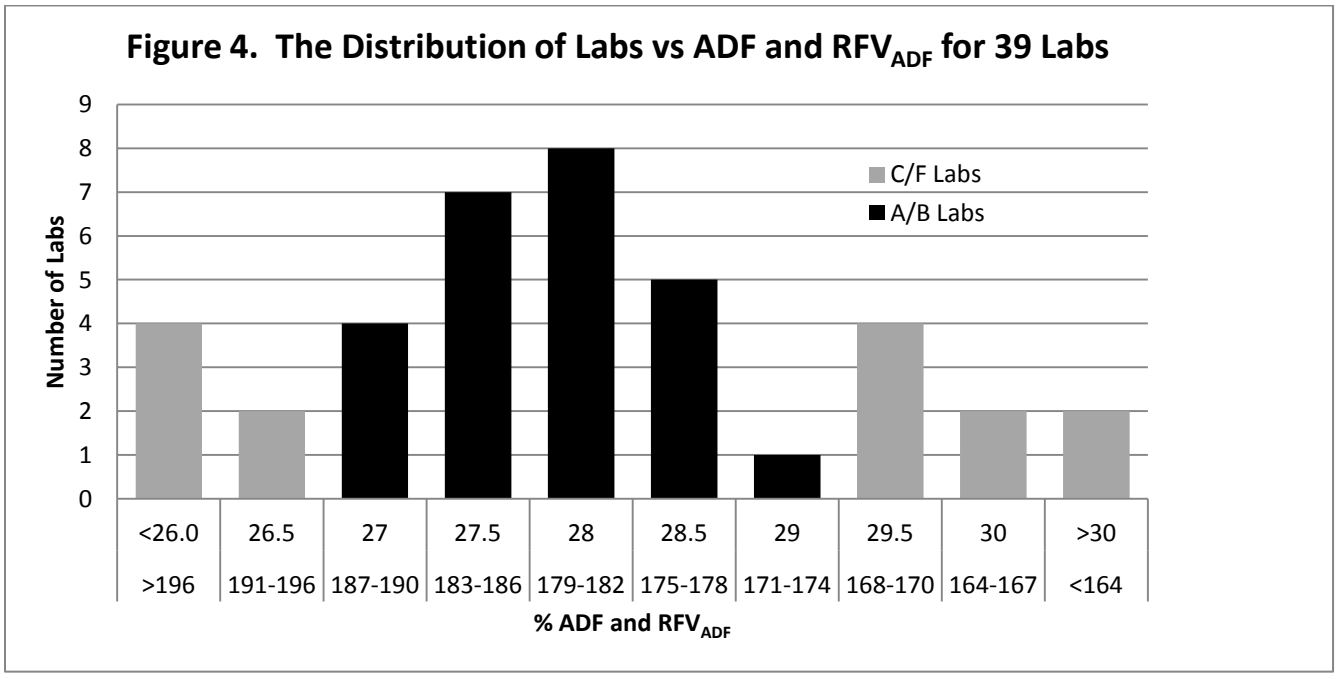
Example 3. If you selected sample 10 from **Table 3**, the ADF is 27.0 and the RFV is 207; however, **Table 6** predicts an ADF of 27.0 would have an RFV_{ADF} of 187. Which is correct: 207, 187, or neither? This is one reason to utilize RFV_{ADF} as a check for RFV. The database of 6000+ measurements suggests the aNDF from sample 10 is probably incorrect. We also know 49% of the labs in this study received letter grades C or F on aNDF. Which is the more accurate answer? **Table 6** can assist in making more informed decisions.

In **Table 7** the ADF results ranges, letter grades, the RFV_{ADF} ranges, and the USDA categories are summarized. The RFV_{ADF} distributions are presented in **Figure 4**.

Table 7.

Summary of Laboratory Performance Using ADF/RFV_{ADF}/USDA Category

No. of Labs	ADF Ranges	RFV_{ADF} Range	ADF Letter Grade	USDA Category
4	26.6-27.2	186-189	B	Supreme
8	27.3-28.0	181-185	A	Premium
4	28.0	180	A	Premium
8	28.0-28.6	176-179	A	Premium
1	28.7-29.3	173	B	Premium
At an ADF of "B" or better standard, the laboratories below would not pass.				
3	25.9-26.5	193-196	3 C	Supreme
6	29.4-30.0	165-169	6 C	Good
	Failing ADFs			
3	< 25.9	199-203	3 F	Supreme
2	> 30.0	145-154	2 F	Good/Fair



ADF Laboratories with Letter Grades of A or B. Laboratories should be able to accurately determine ADF within 2 HSDs (B or better). In **Table 3**, 25 laboratories dried, ground, and analyzed samples with

uniform composition by NIR and/or chemical analyses. These were single determinations. The ADF average of these 25 labs was 27.8 with a standard deviation of 0.51. The ADF average for all 39 labs was 28.0 with a standard deviation of 1.55. Applying the NFTA data treatment of removing the top and bottom 15 percent of results, the remaining ADF results in this study have an average of 27.95 and a standard deviation of 0.65.

Selected data from the NFTA's RMA Criteria Report for 2015 is presented in **Table 8** (NDF data for this report is included in Appendix 4). Excluding sample AH_04 which has a much higher ADF than the other five samples, the four similar alfalfa samples (AH_01, 02, 03, and 05) reported standard deviations between 0.69 and 0.73 after outliers have been deleted. These results are from laboratories running the reference method in triplicate.

Table 8.

Data from RMA Criteria Report for 2015

Sample ID	ADF (AR)	RMASD*	RMASD ADJ**	HSD	Number of Labs
AH_01	31.14	0.73	1.26	0.74	22
AH_02	31.03	0.69	1.20	0.74	23
AH_03	30.49	0.72	1.25	0.73	28
AH_04	37.38	0.76	1.32	0.87	31
AH_05	29.18	0.69	1.20	0.70	25
This Study Overall	28.0		1.55	0.68	39
This Study middle 70%	27.95		0.65	0.68	27

* Reference Method Average Standard Deviation

** Reference Method Standard Deviation adjusted to a single analytical result for NFTA data

The standard deviation of 0.65 on the middle 70% of samples in this study is lower than the standard deviation for the 22+ laboratories running the ADF reference method in triplicate, and much lower than the NFTA RMASD when adjusted to single analysis. Based on the presented data, expecting ADFs at the B or better level is a very reasonable expectation. If one disregards the five laboratories that failed ADF, 34 laboratories remain. Of these, 25 (74%) already meet this goal.

Comparing Two Results Using RFV. It is common practice for individuals to split samples and send the splits to two different laboratories. Articles have been written discussing how to properly do this.[4] Is this process going to provide useful information? Yes. There are only eight laboratories out of 39 in this study that received A's for ADF and aNDF. Their RFVs should be the most accurate. When you submit your first split sample to the laboratory of your choice you have a one in five chance of getting an A-level laboratory analyzing your sample. The odds of a second identical sample also being run by one of the remaining seven A-level laboratories is likewise approximately one in five. The chances of getting two of these laboratories to analyze your samples are only ~4%. When you receive the results on your split samples, there are only two possibilities: 1. there will be a significant difference in the

RFVs; or 2. the RFVs will be very close. In case 1, which RFV is correct? In fact, there is a one in ten chance that both results are not close to the correct result. In case 2, the results are in good agreement; therefore, we assume the RFV to be correct. Unfortunately, there are six laboratories in **Table 4** which have low RFVs within ±3 RFV points; there are 25 laboratories in **Table 3** centered on 180 RFV points that are close; three others are above 200 RFV points; and two are below 155 RFV points.

If you want to split samples and send them to two different laboratories, we recommend you make comparisons based on ADF or RFV_{ADF} to minimize the aNDF variation. We do not recommend comparing laboratories solely based on RFV. Your chances of getting an A-level laboratory to run your first sample more than double when relying only on ADF. The chances are one out of four that both samples will be run by an A-level laboratory (25% for RFV_{ADF} vs. 4% for RFV).

Selecting a laboratory or a set of laboratories. Using the results of this study and statistical analysis, we are proposing a process to determine RFV_{ADF} within 10 RFV_{ADF} points using a small number of laboratories. The key is **Figure 4**. There are 25 A/B labs centered on an ADF of 28.0 and an RFV_{ADF} of 180. There are eight labs that have significantly lower RFV_{ADF} (<170) and six laboratories that have significantly higher RFV_{ADF} (> 190). These are the laboratories in gray. How can you avoid having your hay valued by the eight laboratories that have low results or the six laboratories that have high results?

Proposed Process. Submit five samples as described in Appendix 5 to five different laboratories on a random basis. Take the RFV_{ADF} or ADF results (NOT RFV results) and delete the highest and lowest results (ADF or RFV_{ADF}). Average the remaining results. (**Table 9** shows two examples.) Applying statistics derived from this blind study, the confidence of being within five or ten points of the mean are shown in **Table 10**.

Table 9. Two Examples to Determine Alfalfa Quality

Using RFV_{ADF}				**Using ADF**		
Lab No.	RFV_{ADF}	RFV_{ADF} Dropped		Lab No.	ADF	ADF Dropped
8	182	182		3	31.6	Removed
16	176	176		10	27.0	Removed
24	185	Removed		17	27.5	27.5
32	168	Removed		24	27.3	27.3
40	178	182		31	28.2	28.2
Average	178	**179**			28.6	**27.7**
		ADF = 28.0				**RFV = 182**

The results of both examples are very similar. There are four laboratories (8, 16, 24, 40) with acceptable RFV_{ADF} average of 179 (average ADF value is 28.0). In the other example using ADF there are also four laboratories with acceptable ADF average of 28.05 (10, 17, 24, and 31). The RFV_{ADF} for this case is also 179. In either example, using actual laboratory data from **Table 1**, you would have

found four acceptable laboratories to help meet your analysis needs. There are over a half a million combinations for randomly selecting five results out of 39. Using the mean of 39 labs as the correct value, the probabilities of having results within 5 or 10 RFV_{ADF} points of the mean based on the number of samples sent to laboratories are shown below.

Table 10. Probabilities of RFV_{ADF} Based on Number of Samples Sent

Number of Samples		Probability within 5 RFV_{ADF} Points		Probability within 10 RFV_{ADF} Points
3		**70.5%**		**85.3%**
4		**71.1%**		**92.6%**
5		**76.2%**		**96.8%**
6		**80.1%**		**98.4%**
8		**86.3%**		**99.5%**
10		**90.5%**		**99.9%**

We suggest using five samples in the proposed process with a 96.8% probability the results will be within 10 RFV_{ADF} points of the mean. It is crucial to note that a single sample is not a way to evaluate the performance of any individual lab. Performance on a single sample does not predict performance on a subsequent sample.

This method requires five "identical" samples (Appendix 5) and analyses from five different laboratories. You should end up with three, four, and perhaps five laboratories that obtain RFV_{ADF} results whose average is within ten RFV points or better. This process is most appropriate for high quality alfalfa (especially samples on the border between premium and supreme) and for transactions involving high volumes of alfalfa.

An additional option is blind samples with uniform composition that are indistinguishable from routine alfalfa samples. To this end, we are working on developing the next generation of blind samples.

Other Issues of Note

There are two cautions that must be mentioned. When a laboratory receives a ground sample, such as an NFTA check sample, the laboratory knows it is being evaluated. A substantial number of NFTA-certified laboratories have reported failing results on blind samples. This is an inconsistency that needs to be addressed by the NFTA. The second caution is that a number of NIR laboratories that share the same name and perhaps NIR equations. There are other laboratories with different names sharing equations.

Addressing the issue of aNDF variability among laboratories is crucial because of the impact aNDF has on the calculation of both RFV and RFQ. RFV and RFQ are calculated using ADF and aNDF on

a dry matter basis. RFQ requires additional variables. However, the largest variation in both RFV and RFQ is caused by the variation in aNDF.

NIR and Chemical Analyses. This is the first blind study to include chemical analyses. A *t*-test showed that wet chemistry and NIR analyses were not independent, though wet chemistry methods are presumed to be more accurate since all NIR determinations are ultimately based on wet chemistry analyses.

Additional Data. This study provides additional data that supports the results from the first three studies concerning the accuracy of aNDF analyses. The NFTA RMA Criteria Report for 2015 also supports these four studies. In **Table 11,** the number of laboratories reporting fiber results from the reference method (ADF and aNDF) varies from 22 to 31 for ADF and from 22 to 28 for aNDF.

Table 11. Number of Labs Used to Determine the Alfalfa RMAs for 2015

Sample	Dry Matter	Protein	# of ADF Labs	# of aNDF Labs
AH_01	64	52	22	25
AH_02	65	50	23	22
AH_03	66	54	28	25
AH_04	68	57	31	28
AH_05	67	51	25	25

This means a significant percentage of laboratory results from laboratories running the reference methods are not used to determine the RMA. Another fiber issue that needs to be mentioned is the majority of chemical analysis laboratories do not run the reference method.

There have been 149 individual samples that have been analyzed in three blind studies. Of these, 142 were NIR analyses, four were chemical analyses, and three were analyzed by NIR and chemical methods. The major problem revealed in these three blind studies and the ring test is the substantial variation in determining aNDF by NIR and/or chemical methods.

Alfalfa growers and consumers are keenly aware of the substantial variation in alfalfa results by laboratory. In an article by Young and Severe[5] they discussed concerns about aNDF variation in alfalfa and in rations. They reported: **"**…in our survey 50 percent of the respondents reported losing money because of a business deal involving the analysis from a laboratory. Seven out of the 55 who responded to this question stated they had lost hundreds of thousands of dollars."

This study makes two recommendations to improve the evaluation of alfalfa:

1. Use ADF or RFV_{ADF} (Table 6, examples 1, 2, and 3);
2. Use laboratories that can determine ADF at the A or B level.

After removing five failing ADF labs in this study, 34 out of 34 remaining laboratories reported RFV_{ADF} results at the C or better level affording results from 165 to 196 (180 ± 15 RFV_{ADF} points). At the

strongly recommended B or better level, 25 laboratories out of 34 laboratories (74%) afforded RFV_{ADF} results from 173 to 188 (180 ± 8 RFV_{ADF} points). It should be noted that 20 laboratories (59%) reported results in the 180 ± 5 RFV_{ADF} points range.

References

1 Presented in part at the California Alfalfa Symposia –November 27-29, Reno, Nevada

2 National Forage Testing Association. www.foragetesting.org

3 The small "a" indicates the use of α-amylase (according to the NFTA reference method) for determining neutral detergent fiber.

4 Mike Rankin, Putting testing labs to the test. Hay and Forage Grower, November 2016 5

5 Jerry Severe and Alan Young, Feed analysis – A look at variability. Hay & Forage Grower, November 2017

6 John Goeser, The Forage Lab Consistency Conundrum. Hay and Forage Grower, March 2017

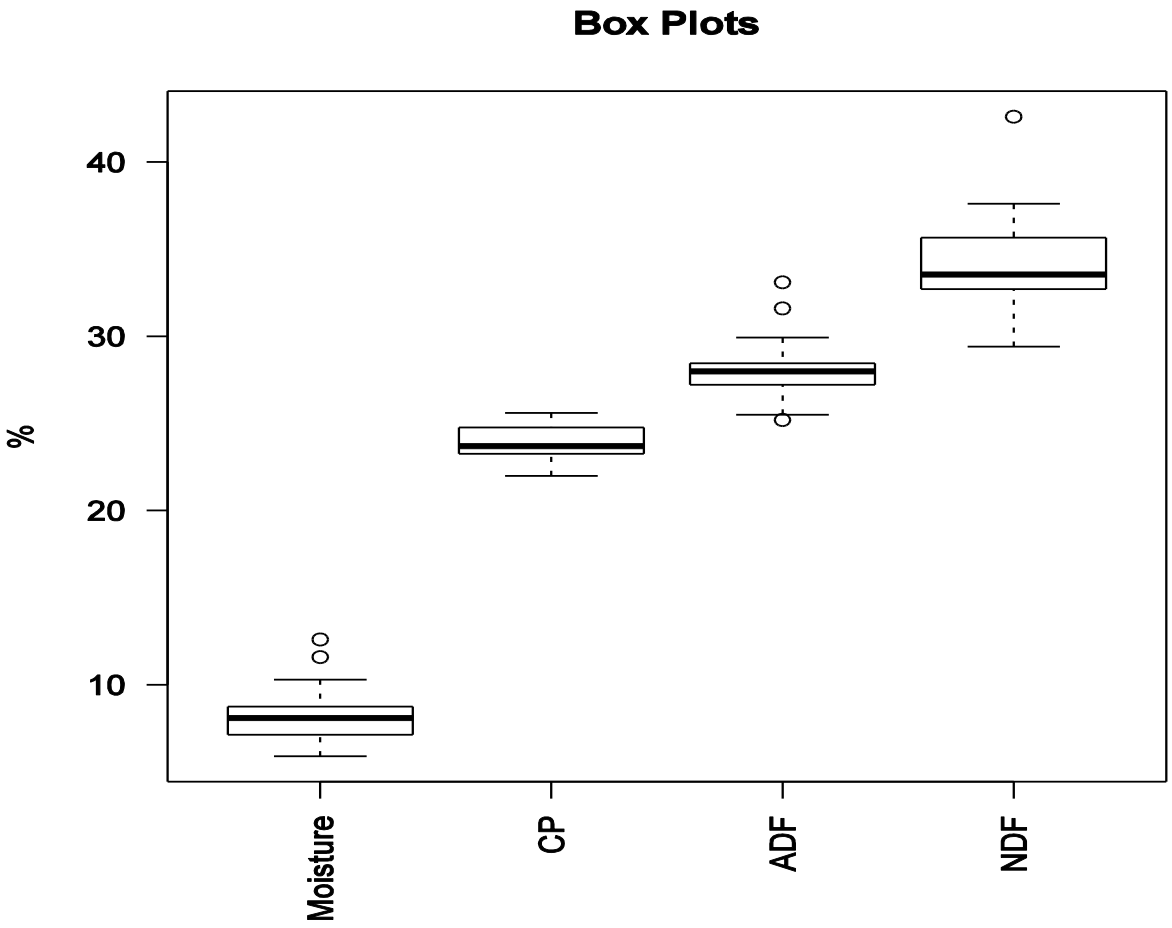
A. A.B. Hristov, D. Mertens, S. Aaman, M. Vander Pol, and W.J. Price, J. Dairy Sci. 93:348-5362 (2010).

Appendix 1.

The Study's Samples. Weld Laboratories was in the first two blind studies (one led by the NFTA, NHA, and the University of Wisconsin and the other led by the University of Nebraska), and we were rated as excellent in both. However, based on analysis of the data from those blind studies, we suspect there may have been issues with the samples' "homogeneity." This started a research project with the goal of preparing samples with uniform composition.

A large set of alfalfa cores from the face of a single alfalfa bale was obtained. The stems, leaves, and fines were separated using Tyler screens and a homemade air separator. The separator uses a column of air to assist in separating the stems from the leaves and fines. Fines were defined as the material which passed a 2 mm screen. The leaf fraction was relatively small and consisted of intact leaves, leaf particles, and non-leaf material that did not pass the 2 mm screen. The fines had slightly lower fiber levels than the leaves. The bulk stems, leaves**,** and fines were split into halves, quarters, and eighths by conventional methods to afford 24 subsets. Eight samples were assembled from the 24 subsets using an analytical balance weighed to ~20 g ± 0.02 grams so each sample had the same percentage of stems, leaves, and fines. The assembled samples were dried at 60^{o}C in a forced-air oven, ground, and analyzed by NIR. The results for moisture, protein, ADF and NDF had relative standard deviations of 3.6, 3.0, 3.1, and 2.6% respectively.

Preparation Procedure for This Study. Approximately 1,500 grams (3.3 pounds) of alfalfa cores were obtained from the face of single alfalfa bale. Five sets of eight samples were prepared and each individual sample contained the following: 37.25% stems, 7.59% leaves, and 55.16% fines. The stems were weighed to 0.001 grams and leaves and fines to 0.005 grams. This afforded sample variation of <0.03%. These samples were sent to 40 NFTA-certified laboratories in 2013 for NIR and/or chemical analysis. Thirty-nine reports were obtained; laboratory 27 did not report results. In order to determine if the results from these five sets of eight could be combined into a single data set, a basic exploratory data analysis (EDA) method and a multivariate analysis of variance was used. Based on the EDA results, the five sets were combined into one set.



A total of five sets of eight samples were prepared and sent to the laboratories. Each of the five sets was originally analyzed separately. The results by set were very similar (as expected). Since four different measurements were made on each sample, the data is fundamentally multivariate. Therefore, to determine if the five sets could be combined into a single population, a MANOVA was used. The MANOVA will determine if set had any effect on the four measurements: moisture, protein, ADF, and NDF. The results are listed below.

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall Set Effect

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.649791	0.90	16	95.344	0.5678
Pillai's Trace	0.393316	0.93	16	136	0.5403
Hotelling-Lawley Trace	0.47498	0.89	16	56.258	0.5806
Roy's Greatest Root	0.291244	2.48	4	34	0.0627

note: Roy's Greatest Root is an upper bound, and it will sometimes behave differently than the other test statistics.

None of the test statistics are significant at the 5% level, and Wilk's Lambda, Pillai's Trace, and the Hotelling-Lawley Trace all have p-values over 0.50. Therefore, there is no evidence to conclude that the set any sample comes from, has an effect on any of the four measurements.

A second issue is the potential differences between NIR and chemical analyses. To determine if the method of analysis has an effect on any of the four measurements, a MANOVA was again used. The p-values for all four test statistics were over 0.90, and it is concluded that no evidence exists that the type of method used, NIR vs. chemical analyses, has any effect on the four measurements. Results are listed below.

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of **No Overall Method Effect**

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.980548	0.17	4	34	0.9529
Pillai's Trace	0.019452	0.17	4	34	0.9529
Hotelling-Lawley Trace	0.019837	0.17	4	34	0.9529
Roy's Greatest Root	0.019837	0.17	4	34	0.9529

Data Analysis. The averages of four analytes (dry matter, protein, ADF, and aNDF) were determined using the published NFTA procedure to determine the reference method averages (RMA). According to their 2015 website, outliers are rejected and then the highest 15% of the results and the lowest 15% are removed, and the average was determined. Since outliers are not defined by the NFTA, we did not delete any samples as outliers in our study for the purposes of determining the RMAs. After dropping the top and bottom 15%, the following averages and standard deviations (SD) were determined: moisture 8.0% (0.70); protein 23.9% (0.64); ADF 28.0% (0.65); and NDF 34.0% (1.30). The NFTA only uses chemical analysis results from laboratories using the reference method average to determine the RMA. Results from one year of NFTA alfalfa results are compared to this study's results in **Table 1**.

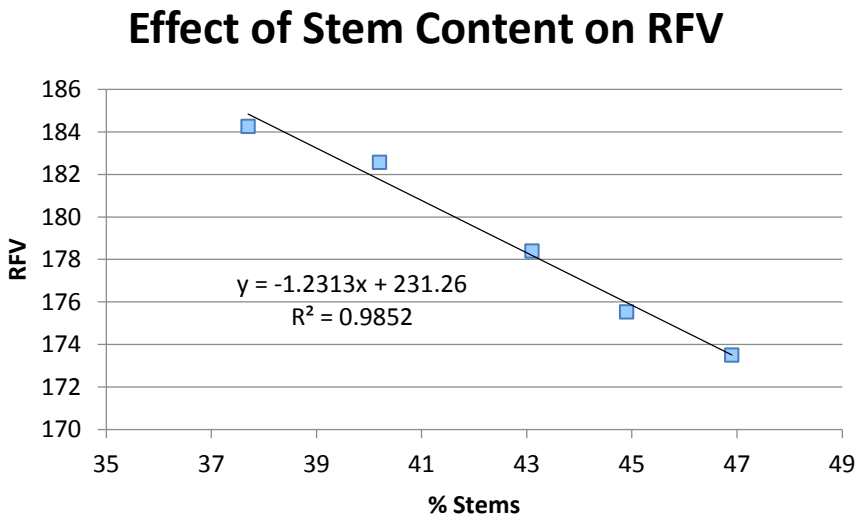
Table 12. Standard Deviations and RSD% for 2015 NFTA Alfalfa Samples and This Study*

ID	Protein		ADF		NDF	
	SD	RSD%	SD	RSD%	SD	RSD%
AH_01	0.46	2.28	0.73	2.34	1.01	2.68
AH_02	0.37	2.07	0.69	2.22	0.97	2.59
AH_03	0.41	2.35	0.72	2.36	0.80	2.06
AH_04	0.38	2.37	0.76	2.03	0.89	1.91
AH_05	0.35	2.16	0.69	2.36	0.72	1.93
This Study	0.64	2.68	0.65	2.31	1.30	3.81

*Outliers have been deleted for AH_01-05. No outliers are deleted from "This study."

The similarity of these standard deviations and RSD% strongly suggests these samples have similar composition in comparison to NFTA samples.

Additional Evidence of Uniform Composition. Could sample variation cause the variation in RFV? To examine this possibility, a sample was prepared with 37.77% stems and its RFV determined by NIR. Additional ground stems were added to the initial sample to provide a new sample with known concentrations (to the nearest 0.01%) and the NIR analysis was run. The results are plotted in **Figure 1.** There is a linear relationship over the range investigated and to lower the RFV approximately 10 RFV points would require increasing the percent stems by about 10 percentage points. Since fines, leaves, and stems are accurate to better than 0.03% in the study, sample variation is not the cause of the large variation in RFV.



The Role of Twins. In preparing the 40 samples, each sample has a "twin". Twins are eighths from the same quarter, and as such, the twins are the most similar samples in the study. In the study there are 19 sets of twins because Lab 27 did not report results. The twin of 1 is 2; the twin of 3 is 4, etc. We assigned letter grades to all the samples and these are summarized in **Figure 7**.

Figure 5. ADF Distributions of 39 Letter Grades by Twins

1	2	3	4	5	6	7	8
A	C	F	A	A	F	A	A
C	B	A	A	C	A	C	A
A	A	F	B	A	F	A	A
A	C	X*	A	B	B	A	C
C	C	F	A	C	A	B	A

There are five samples that received letter grades of F for ADF. Their twins received letter grades of 4 A's and 1 B. This is evidence the samples have uniform composition and those labs receiving F grades were the result of sample preparation, handling, and/or analysis.

Appendix 2. Additional information on Studies 1, 2, and 3.

Study 1. (NFTA, NHA, and UW).[8] This is the benchmark study which first reported the issues with NIR laboratories. The study consisted of 16 laboratories analyzing three alfalfa samples. These samples were partially ground and did not resemble cored alfalfa samples. Our analysis of the ADF and aNDF data determined a failure rate from 38% to 65% for aNDF and from 12 to 25% for ADF. This indicates that determining the RFV from the ADF would have provided more accurate results than the traditional method.

Study 2. (University of Nebraska).[8] This study consisted of 10 NIR labs and 15 different samples. Each sample was analyzed a minimum of four times and a maximum of seven. Two labs were rated as poor based on their RFVs. We do not recommend analyzing data by RFV since a laboratory could report a failing ADF and a failing aNDF could result with an accurate RFV. The University of Nebraska graciously provided us the laboratory data but not the laboratory IDs. Our analysis was based on ADF and NDF and we found, as the University found, two laboratories with significant low biases for ADF and NDF (high RFVs). These biases ranged from 2.2 to 3.2 percentage points. After removing the two labs from the data set, we agree that the remaining eight labs had minimal errors. Of the remaining eight labs, one lab had a positive ADF bias greater than 1.0 and one had a negative aNDF bias greater than 1.0. The remaining laboratory errors for ADF and aNDF were within ± 1 percentage point. Our analysis also suggested some samples were not "homogenous" and we rejected one sample. One caution is in order: doing a study between laboratories with as few as four laboratories analyzing a single sample is a concern.

Study 3. Hristov, Mertens, et.al. (Ring Study, not a blind study). This study involved fourteen chemical analysis laboratories analyzing variability in feed and total mixed rations for protein and aNDF. There were ten commercial laboratories in the study and four non-commercial labs. Four labs reported a single analysis, six labs reported duplicate results, and four reported results in triplicate.

Appendix 3. Ranges for Passing Letter Grades Using the HSD

Analyte	Average	Letter Grade A	Letter Grade B	Letter Grade C
Moisture**	8.0%	±0.508	±1.02	±1.52
Protein	23.9%	±0.593	±1.19	±1.78
ADF	28.0%	±0.677	±1.32	±2.03
aNDF	34.0%	±0.800	±1.60	±2.40

**Modified HSD used by the NFTA.

Appendix 4. Standard Deviations and RSD% for 2015 NFTA Alfalfa Samples and This Study

ID	Protein		ADF		NDF	
	SD	RSD%	SD	RSD%	SD	RSD%
AH_01	0.46	2.28	0.73	2.34	1.01	2.68
AH_02	0.37	2.07	0.69	2.22	0.97	2.59
AH_03	0.41	2.35	0.72	2.36	0.80	2.06
AH_04	0.38	2.37	0.76	2.03	0.89	1.91
AH_05	0.35	2.16	0.69	2.36	0.72	1.93
This Study	0.64	2.68	0.65	2.31	1.30	3.81

Appendix 5. Preparing Ground "Identical" Alfalfa. Alfalfa samples are mixtures of solids and, as such, you cannot have homogeneous or "identical." The best that can be achieved is samples that are very similar and relative feed values that are very close. Our method of preparing ground identical samples is to take several hundred grams of bulk alfalfa cores and grind them through a Wiley mill with a 5 mm screen. The ground material is ground using a cyclone mill with a 1 mm screen. The material must be mixed to achieve samples of "uniform composition." The initial bulk mixing is done with a cake spatula with a 1.5"X 6" offset blade. The entire sample is passed through a riffle splitter a number of times and the splits are crossed mixed to aid in the mixing process. After thoroughly mixing the several hundred grams of material, it is split into halves, quarters, and eighths using the riffle splitter and placed sealable plastic bags. The ADF concentrations are determined on each eighth (NIR) to determine if the samples have suitable uniformity.