Postprint: Calibration Sample Selection for Models Predicting Available Energy Values of Corn Distillers Dried Grains with Solubles (DDGS)

Authors: Yang Xia, Zhao Feng, Li Ke, Dang Fangkun, Zhang Hu, Yin Liting, Zhang Hongfu

Date: 2017-10-23T00:00:00+00:00

Abstract

This experiment aimed to investigate the selection principles of calibration samples for models estimating the effective energy value of corn distillers dried grains with solubles (DDGS). From 23 corn DDGS samples (defined as the full sample bank), nine calibration corn DDGS samples were selected based on a gradient of approximately 0.21 MJ/kg difference in enzymatic hydrolysate gross energy (EHGE), constituting the selective sample bank; the remaining 14 corn DDGS samples were defined as the non-selective sample bank. Subsequently, differences in chemical composition content and variation between the selective and non-selective sample banks were compared, and regression models of chemical composition versus EHGE were established separately using the full sample bank and selective sample bank, followed by comparison of differences in EHGE values calculated from the regression models for the non-selective sample bank. The results showed that corn DDGS from the selective and non-selective sample banks exhibited no significant differences (P>0.05) in mean values of crude protein (CP), crude ash (Ash), ether extract (EE), crude fiber (CF), neutral detergent fiber (NDF), acid detergent fiber (ADF) contents, and EHGE, and no significant differences (P<0.05) in variance of variation for CP, Ash, EE, CF, ADF, NDF contents, and EHGE. Regarding the distribution of chemical composition content of the selective and non-selective sample banks on the first and second principal component score loadings, only one corn DDGS sample from the selective sample bank did not overlap with the distribution range of the non-selective sample bank. The EHGE prediction model established using samples from the selective sample bank was EHGE=(3 566+53.94×EE−32.68×NDF)\$×4.184/1000($R^{\{2\}= 0.7981, RSD = 0.43MJ/kg); thepredictionmodelestablishedusingsamples}$ 0, RSD=0.44 MJ/kg). The mean residuals between calculated and measured EHGE values for corn DDGS in the non-selective sample bank (n=14) obtained from the two models were 0.46 and 0.33 MJ/kg, respectively, with no significant difference ($P>0.05$). In conclusion, using EHGE as the basis for selecting calibration samples is feasible in modeling for estimating the effective energy value of corn DDGS.

Full Text

Study on the Method of Selecting Calibration Samples for Predicting the Effective Energy Value of Corn Dried Distiller' s Grains with Solubles

YANG Xia[1], ZHAO Feng[1]*, LI Ke[2], DANG Fangkun[1], ZHANG Hu[1], YIN Liting[2], ZHANG Hongfu[1]
[1]State Key Laboratory of Animal Nutrition, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing 100193, China
[2]New Hope Liuhe Co., Ltd., Beijing 100102, China

Abstract: This study investigated the principles for selecting calibration samples to establish prediction models for the effective energy value of corn dried distiller' s grains with solubles (DDGS). Nine corn DDGS samples were selected from a total of 23 samples (defined as the full sample pool) based on enzymatic hydrolyzate gross energy (EHGE) values spaced at intervals of approximately 0.21 MJ/kg; these nine samples were defined as the selected sample pool, while the remaining 14 samples constituted the non-selected sample pool. The study compared the chemical composition content and variation between the selected and non-selected sample pools, and developed regression models to predict EHGE from chemical composition using both the full sample pool and the selected sample pool. The predicted EHGE values for the non-selected sample pool from these two models were then compared. The results showed no significant differences in the mean contents of crude protein (CP), crude ash (Ash), ether extract (EE), crude fiber (CF), neutral detergent fiber (NDF), acid detergent fiber (ADF), or EHGE between the selected and non-selected sample pools ($P>0.05$). Additionally, no significant differences were observed in the variance of CP, Ash, EE, CF, NDF, ADF, or EHGE between the two pools ($P<0.05$). Principal component analysis revealed that only one corn DDGS sample from the selected pool fell outside the distribution range of the non-selected pool on the score plots of the first and second principal components. The prediction model developed from the selected sample pool was EHGE=(3566+53.94×EE−32.68×NDF)\$×4.184/1000($R^{\{2\}= 0.7981, RSD = 0.43MJ/kg), whilethemodelfromthefullsam$
RSD=0.44 MJ/kg). The mean residuals between calculated and measured EHGE values for the non-selected sample pool (n=14) were 0.46 MJ/kg and 0.33 MJ/kg for the two models, respectively, with no significant difference between them ($P>0.05$). In conclusion, using EHGE as a criterion for selecting calibration samples in predictive modeling of effective energy values for corn DDGS is feasible.

Keywords: corn DDGS; enzymatic hydrolyzate gross energy; calibration samples

The selection of calibration samples directly affects the accuracy and precision of prediction models for digestible chemical component content in feedstuffs. Therefore, establishing clear principles for calibration sample selection is crucial. Currently, when developing feed effective energy prediction models, researchers typically collect samples randomly and develop prediction models through stepwise regression after obtaining datasets on chemical composition and measured effective energy values. However, this approach may result in insufficient variation in independent or dependent variables within the calibration set, thereby compromising model precision and stability. Furthermore, inconsistent variation among calibration samples can lead to substantial differences in prediction factors and weightings among models developed by different researchers for the same feed ingredient.

According to fundamental principles of regression analysis, independent and dependent variables should be arranged at equal intervals. Considering the empirical observation that biological methods can detect differences in dietary metabolizable energy values of approximately 0.21–0.42 MJ/kg between feed treatments, this study hypothesized that selecting nine calibration corn DDGS samples from 23 corn DDGS samples based on enzymatic hydrolyzate gross energy (EHGE) intervals of about 0.21 MJ/kg would adequately represent the characteristics of all 23 samples. The selected nine samples were defined as the selected sample pool, while the remaining 14 samples constituted the non-selected sample pool. This design would test whether the selected pool could represent the full pool in terms of chemical composition variation and regression relationships between chemical composition and EHGE.

To test this hypothesis, we compared the chemical composition and variation between selected and non-selected sample pools, conducted principal component analysis on the full sample pool to examine distribution differences between the pools on the first and second principal components, and assessed the consistency of chemical composition-EHGE prediction relationships between the selected and full sample pools to evaluate whether EHGE can serve as a viable criterion for selecting calibration samples in effective energy prediction models.

1.1 Feed Ingredients

Twenty-three corn DDGS samples were collected from various sources including the United States and Chinese provinces such as Shandong, Henan, Anhui, Jilin, and Heilongjiang. Using the quartering method, each sample was divided into two portions: one portion was ground in a universal grinder and passed through a 40-mesh sieve for chemical composition analysis, while the other was ground and passed through a 60-mesh sieve for EHGE determination. All samples were sealed under vacuum with nitrogen and stored at -20°C until analysis. The origins and chemical composition of the corn DDGS samples are presented in Table 1 .

1.2 Experimental Design

The 23 corn DDGS samples (full sample pool) were ranked by EHGE values, which ranged 2.39 MJ/kg. Nine calibration samples were selected at intervals of approximately 0.21 MJ/kg to form the selected sample pool, while the remaining 14 corn DDGS samples constituted the non-selected sample pool.

First, basic statistics for chemical composition were calculated for the full, selected, and non-selected sample pools, and significant differences in chemical composition between the selected and non-selected pools were tested. Principal component analysis was then performed on the chemical composition and EHGE data of all 23 corn DDGS samples to examine whether the first and second principal component scores overlapped between the selected and non-selected pools. Finally, regression models relating chemical composition to EHGE were developed separately for the full sample pool (23 samples) and the selected sample pool (9 samples). These two models were used to calculate EHGE values for the non-selected sample pool to compare their predictive accuracy. These comprehensive statistical analyses were employed to evaluate whether the nine calibration corn DDGS samples selected based on EHGE gradients could adequately represent the variation present in all 23 corn DDGS samples.

1.3 Measurement Indicators and Methods

Moisture content was determined according to GB/T 6435-2014 and used to calculate dry matter content. Gross energy (GE) of feed samples and undigested residues after simulated digestion was measured according to ISO 9831:1998. Contents of crude protein (CP), crude ash (Ash), ether extract (EE), crude fiber (CF), neutral detergent fiber (NDF), and acid detergent fiber (ADF) were determined following GB/T 6432-1994, GB/T 6438-2007, GB/T 6433-2006, GB/T 6434-2006, GB/T 20806-2006, and NY/T 1459-2007, respectively. EHGE was measured according to the procedure described by Zhao et al. for determining the enzymative hydrolyzate gross energy of chicken feed using the Single-Phase Digestion System (SDS-2). Dialysis bag specifications and pretreatment, gastric buffer solution, small intestinal front and rear segment buffer solutions, and simulated digestive fluid preparation followed the Technical Regulations for Determination of Enzymative Hydrolyzate Gross Energy in Chicken Feed.

1.4 Data Processing and Statistical Analysis

The MEANS procedure in SAS 9.3 was used to calculate basic statistics for the chemical composition of corn DDGS in the full, selected, and non-selected sample pools. The TTEST procedure was employed to test for significant differences in chemical composition between the selected and non-selected sample pools based on the statistical principle of two-sample mean difference testing. Principal component analysis of chemical composition data from the full sample pool was conducted using the PRINCOMP procedure to obtain eigenvectors for the first and second principal components and component scores for both sample pools, with loading plots generated using Simca-P 11.5. Stepwise regression of chemical composition against EHGE for both the selected and full sample pools was performed using the REG procedure with the Stepwise option to develop prediction models (designated Model 1 and Model 2). EHGE values for the non-selected sample pool were calculated using these models, and the MEANS procedure was used to compare differences between the predicted values from the two models.

2.1 Differences and Variation in Chemical Composition Between Selected and Non-Selected Sample Pools

As shown in Table 2 , when the full sample pool of 23 corn DDGS samples was divided into selected and non-selected pools based on EHGE, no significant differences were observed in the mean contents of CP, Ash, EE, CF, NDF, ADF, or EHGE between the two pools ($P>0.05$). Similarly, no significant differences were found in the variances of CP, Ash, EE, CF, ADF, NDF, GE, or EHGE between the selected and non-selected sample pools ($P<0.05$).

2.2 Principal Component Analysis and Score Distribution of Chemical Composition

As illustrated in Figures 1 [Figure 1: see original paper] and 2 [Figure 2: see original paper], the first and second principal components derived from CP, Ash, EE, CF, NDF, ADF, GE, and EHGE of the 23 corn DDGS samples accounted for 51.97% and 19.89% of total variation, respectively, with a cumulative contribution of 71.86%. Neutral detergent fiber, ADF, and Ash were the primary weighting factors for the first principal component, while EHGE, GE, CF, and EE were the main contributors to the second principal component. On the score plots of the first and second principal components, only one sample (B23) from the selected sample pool (n=9) fell outside the 95% confidence interval of the non-selected sample pool (n=14). The scatter plot showed no clear separation between the two groups, indicating similar overall variation in chemical composition.

2.3 Differences in Prediction Models Relating Chemical Composition to EHGE Between Full and Selected Sample Pools

In the full sample pool of 23 corn DDGS samples, Ash and EE contents and GE showed significant or highly significant positive correlations with EHGE ($P<0.05$ or $P<0.01$), while CP, CF, and NDF contents exhibited significant or highly significant negative correlations ($P<0.05$ or $P<0.01$). In the selected sample pool of nine corn DDGS samples, CP content showed a significant negative correlation with EHGE ($P<0.05$), while EE content and GE showed highly significant positive correlations ($P<0.01$); other chemical components showed weak correlations with EHGE.

Since the simple positive correlation between Ash content and EHGE contradicts basic nutritional principles, Ash was excluded from stepwise regression analysis. The resulting regression model for the selected sample pool (Model 1) was EHGE=(3566+53.94×EE−32.68×NDF)$×4.184/1000($R^{\{2\}= 0.7981, RSD = 0.43MJ/kg), whilethemodelforthefulls$ RSD=0.44 MJ/kg). Both models included the same prediction factors.

To further compare the equivalence of the two EHGE prediction models, EHGE values were calculated for each sample in the non-selected pool. As shown in Table 4 , the mean predicted EHGE values from Models 1 and 2 were 13.61 and 13.69 MJ/kg, respectively, with no significant difference between them (P>0.05). The mean residual between calculated and measured EHGE values was 0.46 MJ/kg for Model 1, with individual sample deviations within 0.62 MJ/kg except for sample B24. For Model 2, the mean residual was 0.33 MJ/kg, with individual deviations within 0.69 MJ/kg except for sample B24. The standard error of prediction (SEP) was 0.56 MJ/kg for Model 1 and 0.40 MJ/kg for Model 2 (0.33 and 0.44 MJ/kg when sample B24 was excluded). These statistical results indicate that both models provide similar accuracy in predicting EHGE of corn DDGS.

3.1 Basis for Selecting Calibration Samples in Prediction Models

Appropriate selection of calibration samples not only reduces workload but also improves the precision and generalizability of prediction models. In near-infrared spectroscopy modeling for feed chemical composition (GB/T 18868-2002), calibration samples are selected either based on the maximum and minimum scores of each principal component of the spectra, or according to Mahalanobis distances greater than 0.6 after spectral cluster analysis. This approach essentially expresses the overall variation of all independent variables in the sample set through dimensionality reduction based on the intrinsic relationships among variables, thereby avoiding over-concentration of samples in certain ranges. Theoretically, this provides good representativeness.

Zhao et al., when developing a prediction model for duck metabolizable energy of corn, created 30 calibration samples based on survey data from 427 corn samples to ensure the calibration set had a normal distribution of proximate nutrients similar to the full set of 427 samples, achieving satisfactory results. However, this method considers the distribution probability of individual chemical components but not the correlations among them. Meloche et al. selected 15 calibration samples based on EE range (3.15%-13.23%) when developing a prediction model for broiler metabolizable energy of low-oil corn DDGS, yet still encountered over-concentration of samples in certain regions (six samples with EE content between 9.6%-10.8% and four samples between 11%-12%). Moreover, this selection principle did not consider whether the variation range of other chemical components was representative.

In the present study, using EHGE as the selection criterion for corn DDGS calibration samples is feasible because EHGE is significantly correlated with chemical composition, thereby allowing indirect control of chemical composition variation through controlling EHGE variation in the sample set. Additionally, the non-significant differences in chemical composition between selected and non-selected pools demonstrate that samples selected based on EHGE are similar to those not selected, thus preventing entry of overly similar samples into the calibration set.

3.2 Evaluation of Calibration Sample Representativeness

Sample representativeness refers to the similarity between selected samples and the overall sample population in terms of chemical composition content and biological effects. Statistically, representativeness is typically evaluated by comparing whether the means and variances of investigated indicators follow the same probability distribution as the total population.

Recently, principal component analysis has been applied to samples from different sources, with score plots of the first and second principal components used to visually assess whether sample sets are separated, thereby identifying samples belonging to the same category or detecting outliers. This method can also serve as a reference for evaluating calibration sample representativeness. In this study, no significant differences were observed in the means or variances of chemical composition and EHGE between selected and non-selected sample pools, and the score distributions on the first and second principal components overlapped between the two pools. This indicates that the selected sample pool can represent the non-selected pool in terms of chemical composition content and variation. Furthermore, the consistency between the regression model developed from the selected pool and that from the full pool in predicting EHGE of the non-selected pool demonstrates that the intrinsic relationship between chemical composition and EHGE in the selected pool can also represent the non-selected pool. Based on these three aspects of evidence, selecting corn DDGS calibration samples based on EHGE gradients provides satisfactory representativeness and can therefore be used to establish a calibration sample pool.

Conclusion

The calibration sample pool constructed by selecting corn DDGS samples based on EHGE intervals of approximately 0.21 MJ/kg can adequately represent the full sample pool in terms of chemical composition content and variation, as well as the relationships between chemical composition and EHGE. Therefore, using EHGE as a criterion for selecting calibration samples in corn DDGS effective energy prediction models is feasible.

References

[1] HE X W, GUO W. Chemometrics—Recent advances in sampling theory[J]. Chinese Journal of Analytical Chemistry, 1995, 23(12): 1455-1461.
[2] LI T T, CAI H Y, YAN H J, et al. Prediction model for apparent metabolizable energy of corn dried distiller' s grains with solubles for poultry[J]. Chinese

Journal of Animal Nutrition, 2014, 26(6): 1556-1562.
[3] ROCHELL S J, KERR B J, DOZIER W A. Energy determination of corn co-products fed to broiler chicks from 15 to 24 days of age, and use of composition analysis to predict nitrogen-corrected apparent metabolizable energy[J]. Poultry Science, 2011, 90(9): 1999-2007.
[4] CARRÉ B. Predicting the dietary energy value of poultry feeds[C]//WISEMAN J, COLE D J A. Feedstuff Evaluation. Guildford: Butterworths, 1990: 283-300.
[5] ZHAO F, REN L Q, MI B M, et al. Developing a computer-controlled simulated digestion system to predict the concentration of metabolizable energy of feedstuffs for rooster[J]. Journal of Animal Science, 2014, 92(4): 1537-1547.
[6] URRIOLA P E, LI M, KERR B J, et al. Evaluation of prediction equations to estimate gross, digestible, and metabolizable energy content of maize dried distillers grains with solubles (DDGS) for swine based on chemical composition[J]. Animal Feed Science and Technology, 2014, 198: 196-202.
[7] YUAN Z F, ZHANG J Y. Experimental Design and Analysis[M]. Beijing: Higher Education Press, 2000.
[8] DALE N M, FULLER H L. Repeatability of true metabolizable energy versus nitrogen corrected true metabolizable energy values[J]. Poultry Science, 1986, 65(2): 352-354.
[9] General Administration of Quality Supervision, Inspection and Quarantine of the People' s Republic of China, Standardization Administration of China. GB/T 6435—2006 Determination of moisture and other volatile matter content in feeds[S]. Beijing: Standards Press of China, 2007.
[10] State Bureau of Technical Supervision. GB/T 6432—1994 Method for determination of crude protein in feedstuffs[S]. Beijing: Standards Press of China, 1994.
[11] General Administration of Quality Supervision, Inspection and Quarantine of the People' s Republic of China, Standardization Administration of China. GB/T 6438—2007 Determination of crude ash in feeds[S]. Beijing: Standards Press of China, 2007.
[12] General Administration of Quality Supervision, Inspection and Quarantine of the People' s Republic of China, Standardization Administration of China. GB/T 6433—2006 Determination of crude fat in feeds[S]. Beijing: Standards Press of China, 2006.
[13] General Administration of Quality Supervision, Inspection and Quarantine of the People' s Republic of China, Standardization Administration of China. GB/T 6434—2006 Determination of crude fiber content in feeds—Method with intermediate filtration[S]. Beijing: Standards Press of China, 2006.
[14] General Administration of Quality Supervision, Inspection and Quarantine of the People' s Republic of China, Standardization Administration of China. GB/T 20806—2006 Determination of neutral detergent fiber (NDF) in feeds[S]. Beijing: Standards Press of China, 2007.
[15] Ministry of Agriculture of the People' s Republic of China. NY/T 1459—2007 Determination of acid detergent fiber in feeds[S]. Beijing: Agriculture Press, 2008.

[16] ZHAO F, ZHANG H F, ZHANG Z Y. Manual for Single-Phase Digestion System in Monogastric Animals[M]. 2nd ed. Beijing: Chinese Academy of Agricultural Sciences, 2011.

[17] YEGANI M, SWIFT M L, ZIJLSTRA R T, et al. Prediction of energetic value of wheat and triticale in broiler chicks: a chick bioassay and an in vitro digestibility technique[J]. Animal Feed Science and Technology, 2013, 183(1/2): 40-50.

[18] LU Y J, QU Y L, PIAO R G, et al. Algorithm for eliminating similar samples in calibration and prediction of near infrared spectroscopy analysis[J]. Spectroscopy and Spectral Analysis, 2004, 24(2): 158-161.

[19] General Administration of Quality Supervision, Inspection and Quarantine of the People' s Republic of China. GB/T 18868—2002 Method for rapid determination of moisture, crude protein, crude fiber, crude fat, lysine and methionine in feeds—Near infrared reflectance spectroscopy[S]. Beijing: Standards Press of China, 2003.

[20] ZHAO F. Methodological study on evaluating duck feed metabolizable energy using enzymative methods[D]. PhD thesis. Beijing: Chinese Academy of Agricultural Sciences, 2006.

[21] MELOCHE K J, KERR B J, SHURSON G C, et al. Apparent metabolizable energy and prediction equations for reduced-oil corn distillers dried grains with solubles in broiler chicks from 10 to 18 days of age[J]. Poultry Science, 2013, 92(12): 3176-3181.

[22] PEDERSEN M B, DALSGAARD S, KNUDSEN K E B, et al. Compositional profile and variation of distillers dried grains with solubles from various origins with focus on non-starch polysaccharides[J]. Animal Feed Science and Technology, 2014, 197: 130-141.

[23] GALLO A, MOSCHINI M, CERIOLI C, et al. Use of principal component analysis to classify forages and predict their calculated energy content[J]. Animal, 2013, 7(6): 930-939.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.