MODERN STATISTICAL ANALYSIS OF FORAGE QUALITY ASSESSMENT WITH NIR SPECTROSCOPY

ALMANJAHIE, IBRAHIM M.[1,2] – AHMAD, I.[1,2,3] – CHIKR ELMEZOUAR, Z.[1,2,4*] – LAKSACI, A.[1,2]

[1]*Department of Mathematics, College of Science, King Khalid University, 62529 Abha, Kingdom of Saudi Arabia*

[2]*Statistical Research and Studies Support Unit, King Khalid University, Abha, Kingdom of Saudi Arabia*

[3]*Department of Mathematics and Statistics, International Islamic University, Islamabad, Pakistan*

[4]*Department of Mathematics, University Tahri Mohamed, Bechar, Algeria*

**Corresponding author*
e-mail: chikrtime@yahoo.fr

(Received 5th Jun 2019; accepted 9th Sep 2019)

Abstract. Recently, the statisticians have developed a new approach called functional Statistics to treat the data as curves or images. In parallel, the Near-Infrared Reflectance (NIR) spectroscopy approach has been used in modern Chemistry being a fast, inexpensive and accurate procedure to characterize chemical properties for an object. In this paper, we study the forage quality by analysing the spectroscopy procedure with some modern statistical models. Our contribution leads to the prediction of chemical components of Chinese ryegrass forage by analysing its spectral data using some functional models. Precisely, the functional linear quantile regression (FLQR), the functional nonparametric quantile regression (FNQR), the functional local linear quantile regression (FLLQR) and the functional local linear model regression (FLLMR) are implemented to predict the quantities of acid detergent fiber (ADF), neutral detergent fiber (NDF), and crude protein (CP) contents. The choice of these functional models is motivated by the fact that they can construct a predictive region with a given confidence level. We show that the considered models improve the prediction results significantly as compared to conventional models such as the classical partial least squares regression (PLSR) and the principal component regression (PCR). Moreover, we also show that the proposed models are more robust than their competitive models like PLSR and PCR in the sense that their efficiency is not much affected by non-homogeneity of the data.

Keywords: *Leymus chinensis, functional quantile regression, principal component regression (PCR), neutral detergent fiber, predictive regions, crude protein, neutral detergent fiber, acid detergent fiber*

Introduction

The analysis of the forage quality is crucial for livestock producers. It usually measured by its ability to provide livestock high amounts of nutrients and all kinds of livestock feeds. The forage quality depends on the composition of different chemical and biological components present in it. Precisely, the amounts of crude protein (CP), acid detergent fiber (ADF) and neutral detergent fiber (NDF) are the most critical parameters to determine the forage quality (see Mott and Moore, 1970). While the CP content increases the quantity of protein in milk and meat which is essential for dairy producers. On the other hand, NDF and ADF are also crucial for the digestibility of livestock animals (see Agnihotri et al., 2003 or Suksombat, 2004). Accurate prediction of the nutrients present in forage permits the producers to harvest, store and inventory the feed resource optimally way for both grazing and confinement fed ruminants. To get ideal results before

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346

selling, it is highly needed to apply rapid and accurate methods to evaluate the chemical properties of forage without its destruction. In literature, two methods are commonly used to analyze the quality of forage samples in a laboratory. First, one is the traditional wet chemistry analysis, and the other one is near-infrared reflectance Spectroscopy (NIRS) analysis. First, one is based upon well-established chemical principles, with the help of chemicals and drying agents to find the quality of different components present in the forage.

On the other hand, NIRS analysis is a computerized method to determine the quality of the forage using near-infrared light. However, the traditional approaches are costly, non-ecological and require a great number of samples (Yu et al., 2010; Yu et al., 2011; Albrecht et al., 2008). Alternatively, the NIRS is precise technology, rapid, inexpensive and nondestructive nature of the analysis (see for example Kamruzzaman et al., 2012; De Marchi et al., 2012; Stuth et al., 2003; Foley et al., 1998). The NIR spectroscopy assessment of the forage quality is widespread in the last few years. It was introduced by Norris et al. (1976). This approach has significant advantages over the classical chemical analyses in the form of minimal sample in the laboratory, short time for analysis, less cost in case of a single sample or large batches of samples, many constituents can be found simultaneously, non-destruction of the samples during analysis, no need of skilled operator during analysis, elimination of use of hazardous chemical reagents, and the results are usually more precise and accurate as compared with the methods usually employed. Being the responsibility of the analyst, it is utmost important not only to choose the most appropriate analytical technique but also to follow standard procedures for accurate and precise results. In the last decade this technology has received a growing interest in forage quality assessment (see Bai et al., 2004; Nie et al., 2007; Chen et al., 2007; En et al., 2009 for previous studies and Asekova et al., 2016; Monrroy et al., 2017; Yang et al., 2017 for recent advances). In all these previous studies, the spectral data are analyzed by some classical statistical methods such as the partial least regression (PLSR) or the Principal Component Regression (PCR). However, these old approaches have some drawbacks, which affect their efficiency. Indeed, in these two models the spectral data is not taken into account. The users are required to reduce the dimension of the data by transforming it over some directions. But this transformation can lead inaccurate prediction and results because the direction is chosen independently of the prediction problem. In particular, the variable of interest does not intervene in the choice of the optimal direction reduction. In order to avoid this drawback, we implemented some alternatives models recently developed in modern Statistics. Such models allow analyzing the big data without dimension reduction. These new models are based on the notion of functional Statistics that allows to model the data as curves instead of some numerical numbers which permit to explore the total information of the data. For an overview on recent developments in this topic, we cite, some special issues dedicated to this topic by various statistical journals such as Gonzalez-Manteiga and Vieu (2007), Valderrama (2007), Goia and Vieu (2016) and Aneiros et al. (2019).

More precisely, the main aim of this paper is to combine the Near Infrared Spectroscopy (NIRs) methodology (as modern Chemistry tools) with the recent development of big-data analysis in order to provide a new prediction approaches for the forage quality to circumvent the drawbacks of the previous model. For this study, we consider a sample of 151 size of Chinese ryegrass (Leymus Chinensis) forage and we examine the performance of NIR spectroscopy in estimating the CP, ADF and NDF quantities present in this sample. For this purpose, we use four statistical models the

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346

functional linear quantile regression (FLQR), the functional nonparametric quantile regression (FNQR), the functional local linear quantile regression (FLLQR) and the functional local linear model regression (FLLMR). The main feature of these models is to make two kinds of prediction namely, pointwise prediction and predictive regions. Moreover, it is well known in statistics that these predictors are more robust than the classical regression models.

Materials and methods

Sample collection and chemical analysis

The samples for this study were taken of the sites ranging from the west to east across the grassland situated in Heilongjiang Province of Northeast China. To accomplish the task 151 samples of sheepgrass hay were chosen at random from different sheepgrass fields of hay factories in 2013 with longitudes from 123.209°E to 132.944°E and latitudes from 44.475°N to 51.728°N. These sampling sites are rich in production of sheepgrass fields and also comprise of wide range of soils and climate. All of the samples were taken at the stage of blooming, recognized and collected before being cutting and packaging. Total of 151 samples from 21 different sites were distributed over 8 regions in Heilongjiang province and tested in the lab and stored at 4 °C for further process. For good quality, every sample consisted of one quarter square meter trimmed at 4 cm, oven dried (65 °C, for 48 h), ground (1 mm sieve), and finally mixed. The chemical and microbiological properties of this data set have been described in Chen et al. (2015). They state that the CP was being measured using the Kjeldahl method for nitrogen (N) determination such as (%CP = %N × 6.25) with the help of methods recommended by Association of Official Analytical Chemists-AOAC (AOAC, 1990). Similarly, the ADF and NDF were treated with the help of the methods proposed by Van Soest et al. (1991). As discussed in the introduction the core feature of the prosed models is their insensitivity to the outliers. To highlight this point, we compare their resistances to outliers to the classical models (PLS and PCR). For this aim, we have used the MAD-Median rule (Wilcox, 2005) to detect the outliers in the endogenous variables CP, ADF and ANF. Thus we observe that there is wide variability in the ADF variable compared to the others variables. This variable is ranged between 23.81% and 47.86% and MAD-test detect 11 outliers. The CP and the NDF values are more homogenous than the CP concentration. Specifically, the CP is ranged from 5.59% to 15.76%, and the MAD-test report 3 outliers in this variable. While the NDF values are between 62.96% and 81.41% and have 5 outliers from the MAD-test.

Spectroscopic analysis

NIR spectrometer Model (DA7200, Perten Corporation, Hagersten, Sweden) was used to collect the spectra of the given sample. The NIR spectra were taken at 5 nm intervals in a range from 950 to 1,650 nm. A dried sample with a weight of 50 g was scanned in a sample cell with a diameter of 7.5 cm, with a quartz window maintained at room temperature of 25 °C. Further, using DA7200 software with reflectance mode (R mode) the samples were scanned from 950 to 1,650 nm. The obtained measurements were transformed into the logarithms of the reflectance reciprocal (absorbance). To minimize or eliminate an error caused by loading the sample, each sample was repeated to scan three times. Then the average of three spectral scanning recordings was considered as the final value, which was further treated to final calculation with the help of Grams32

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346

software (Perten Corporation, Hagersten, Sweden). Because of the fineness of the grid, we can consider each subject of the spectral data as a continuous curve (see *Fig. 1*).

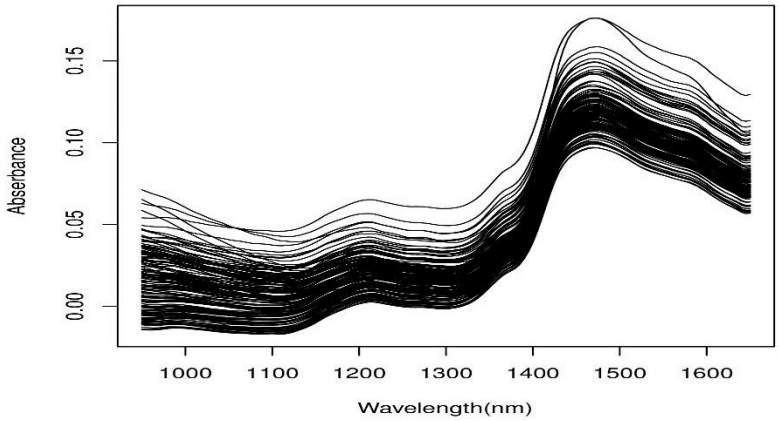

Figure 1. *The NIR spectroscopy curves of the data*

A first preliminary study with the multivariate factorial analysis shows that there is no visible structure of the data (see *Fig. 2*). This fact confirms that the spectral data cannot be summarized by a discretization on small number of grides. In sense that for this prediction issue, it would be more interesting if we incorporate the whole curve of the spectral data.

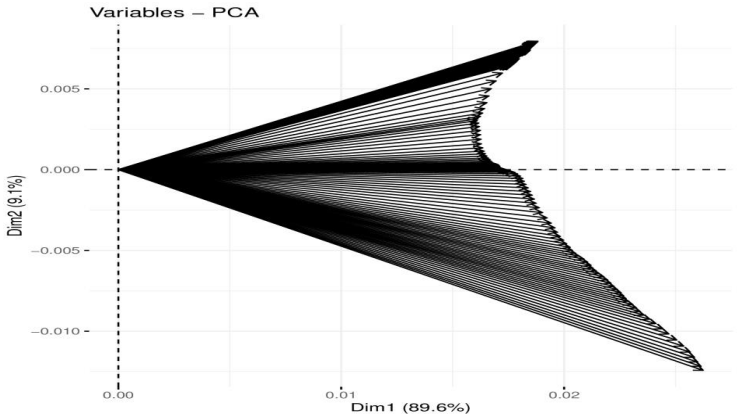

Figure 2. *Factorial analysis of the data*

Notice that the shape of the spectra contains some peaks and valleys which indicated the different chemical component characteristics of Chinese ryegrass samples.

Using this NIR sample, Chen et al. (2015) have considered three standard models (MLR, PCR and PLSR) to predict the chemical quality of this sheepgrass. However, the implementation of these models requires some pretreatment of the spectral data such as dimension reduction, linearity correlation, outliers removing, calibration etc. Such pretreatment modifies the structure of the spectral data independently to the chemical parameters CP, ADF or NDF, which ultimately influence the prediction results negatively. Alternatively, in this paper, we use some new statistical models in order to investigate the forage quality without any transformation. The main advantage of this approach is the fact

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346

that it takes into account the whole curve in its continuous path. Of course, with this consideration, we keep all the information in the sampled data and we accelerate the prediction process by avoiding the initial pretreatment and related drawbacks.

Statistical analysis

Recall that, in the previous works (see, for instance, Chen et al., 2015; Asekova et al., 2016; Monrroy et al., 2017; Yong et al., 2017) the correlation between the NIR spectroscopy and the chemical components of the forage is assumed to be linear. In this contribution, we consider the general case that covers not only the linear and nonlinear models but also the parametric and nonparametric situations. Indeed, to fix the ideas, let us present the mathematical formulation of the equation describing the relationship between the spectral data and the chemical properties of the sheepgrass:

$$Y = r(X) + error, \quad \text{(Eq.1)}$$

where Y is a real variable means either the CP, ADF or NDF. The functional variable X is the curve of the NIR spectroscopy (950-1,650 wavelength). Therefore, the prediction of Y given $X = x$ is based on the determination of the function $r(.)$. The latter is obtained by minimizing the following criterion

$$r(x) = argmin_f E[\rho(Y, f)|X = x] \quad \text{(Eq.2)}$$

Where ρ is a given loss function.

In the modern statistics, there exist various approaches which can be used to model this prediction issue. In this paper, we focus on some models related to the conditional distribution of Y given X which is constructed from the quantile regression and the modal regression. Of course, the conditional distribution of Y given X has the advantage of completely characterizing the relationship between Y and X. Thus, the two used regression models are more informative than the standard regression for analyzing the link between the spectral data and the chemical components of the hay. Furthermore, to provide to NIR spectroscopy users, we present in the rest of this section a varied collection of statistical predictors with mathematical formalism which is recently developed in functional statistics.

The functional linear quantile regression (FLQR)

The FLQR was first introduced in functional Statistics by Cardot et al. (2005). Since this pioneer work, this model had been widely studied in the context of functional Statistics (see, for instance, Qingguo and Linglong, 2017). Mathematically, the quantile regression of order α is the quantity t_α such that $P(Y \leq t_\alpha|X = x) = \alpha$. It is well known that the quantile regression can be also expressed by *Equation 2* using loss function $\rho_\alpha(z) = (2\alpha - 1)z + |z|$ (see the proof in Laksaci et al., 2009). Based on the linear formulation of the quantile regression obtained by Koenker et al. (1978) and Cardot et al. (2005) constructed an conditional quantile estimator of order α for a given sample $(X_i, Y_i)_{i=1,...,n}$ as below

$$\widehat{t_\alpha} = argmin_f \sum_{i=1}^{n}[(2\alpha - 1)(Y_i - < f, X_i >) + |Y_i - < f, X_i >| \quad \text{(Eq.3)}$$

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346

It is well established that the conditional median which corresponds to $\alpha = \frac{1}{2}$ constitutes an alternative predictor than the classical regression. Thus, for all given spectra curves x, we can predict CP, ADF and NDF contents by

$$\hat{Y} = \hat{r}(x) = argmin_f \sum_{i=1}^{n} |Y_i - < f, X_i >| \quad \text{(Eq.4)}$$

Notice that one of the main advantage of the quantile regression is to make prediction with predictive region being informative than a single-point prediction, It allows the predictions for thorough contingency planning. Indeed, for a fixed confidence level, for example $\gamma = 0.95$ the theoretical confidence interval is expressed $(t_{0.025}, t_{0.975})$. From the estimator given in *Equation 3*, we can write the interval of predictive region as

$$(\widehat{t_{0.025}}, \widehat{t_{0.975}}) \quad \text{(Eq.5)}$$

The functional nonparametric quantile regression (FNQR)

The nonparametric version of the quantile regression is defined by

$$\widetilde{t_\alpha} = argmin_f \sum_{i=1}^{n} [(2\alpha - 1)(Y_i - f) + |Y_i - f|] K\left(\frac{\|x - x_i\|}{h_n}\right) \quad \text{(Eq.6)}$$

This estimator was constructed by Laksaci et al. (2009). In fact, there exist several versions of the FNQR, the natural one is the inverse of the cumulative distribution function estimator. This last estimator was studied by Ferraty et al. (2006) and Dabo-Niang and Laksaci (2012). In this work, we have opted the estimator presented in *Equation 6* being more fast and robust. Similar to the previous case, for single-point prediction, we use the conditional median to predict the CP, ADF or NDF contents denoted by Y. So for all given spectra curves x, Y is approximated by

$$\hat{Y} = \hat{r}(x) = argmin_f \sum_{i=1}^{n} |Y_i - f|] K\left(\frac{\|x - x_i\|}{h_n}\right) \quad \text{(Eq.7)}$$

while for the predictive region we use the following nonparametric confidence interval

$$(\widetilde{t_{0.025}}, \widetilde{t_{0.975}}) \quad \text{(Eq.8)}$$

The functional local linear quantile regression (FLLQR)

The study of the local linear estimation of the quantile regression is more recent than the other models. It was introduced by Al-Awadhi et al. (2018). This version of the quantile regression inherits the advantages of the both of the previous models. In particular, from the linear property of this model, we reduce the convergence rate and from the nonparametric one we explore the local structure of the curves. Recall that as a local linear approach, the FLLQR is obtained by assuming that the function $t_\alpha(x)$ has a linear form in the neighbourhood of a given spectra curve x. Mathematically, this assumption is expressed by

$$\forall\, x_0 \; in \; neigborhood \; of \; x \;\; t_\alpha(x) \approx a + b\beta(x, x_0) \quad \text{(Eq.9)}$$

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346

Where the quantities a *and* b are obtained by minimizing

$$min_{f,g} \sum_{i=1}^{n} [(2\alpha - 1)(Y_i - f - g\beta(X_i, x)) + |Y_i - f - g\beta(X_i, x)|] K\left(\frac{\|x - X_i\|}{h_n}\right) \quad \text{(Eq.10)}$$

with $\beta(x,x)$ is known function used to measure the proximity among the curves. Thus, for fixed spectra curve x, FLLQR is estimated by $\widehat{t_\alpha}(x) = a$ (where a is solution of *Eq. 4*). Furthermore, for the pointwise prediction, we approximate the contents of CP, ADF and NDF by the conditional median $\bar{r}(x) = \bar{a}$ as solution of *Equation 11*

$$min_{f,g} \sum_{i=1}^{n} |Y_i - f - g\beta(X_i, x)| K\left(\frac{\|x - X_i\|}{h_n}\right) \quad \text{(Eq.11)}$$

Once again this estimator can be used to construct the 95% confidence interval that is

$$(\widehat{t_{0.025}}(x), \widehat{t_{0.975}}(x)) \quad \text{(Eq.12)}$$

The functional local linear modal regression (FLLMR)

The modal regression or the conditional mode is an old alternative model introduced by Collomb (1986). It is well established that this model is more pertinent than the classical regression where the white noise has a Chi-squared distribution. In functional Statistics there exist several nonparametric estimators for conditional mode. In this work we consider more recent one that is proposed by Almanjahie et al. (2018). This estimator is attractive and more robust and fast. With this model, the response variable CP, ADF and NSDF is approximated for a given spectral curve x as in *Equation 13*

$$\ddot{Y} = \ddot{r}(x) = argmin_y \hat{f}(y|X = x) \quad \text{(Eq.13)}$$

where $\hat{f}(.|X = x)$ is the local linear estimator of the conditional density of Y given $X = x$ (see Almanjahie et al., 2018 for more details on this topic). From this estimator, we construct a predictive region so-called the maximum conditional density predictive region introduced by Hyndman (1995). It is defined for a given spectral curve x as in *Equation 14*

$$R = \{y \quad such\ that \quad \hat{f}(y|X = x) > l_{0.95}\} \quad \text{(Eq.14)}$$

with

$$l_{0.95} = max\left\{l > 0\ such\ that \int_{f(y|X=x)>l} \hat{f}(y|X = x)dy > 0.95\right\}. \quad \text{(Eq.15)}$$

Results

The performance of all above mentioned models is closely linked with the use of different parameters involved in the estimation of $\hat{r}, \tilde{r}, \bar{r}$ *and* $\ddot{r}$. In fact, the distance between the curves $\|.\|$ and the smoothing parameter h_n are the most influencing parameters in this prediction issue. Concerning the norm $\|.\|$, it is observed that the most of the metrics are obtained by expanding the curve on some basis functions such as

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346

spline basis, trigonometric basis or the interpolation by wavelet basis functions see for example, Ferraty and Vieu (2006). Formally, we write

$$x(t) \approx \sum_{i=1}^{k} c_i B_i(t) \quad \text{(Eq.16)}$$

where $(B_i(.))_i$ is a given orthonormal basis function. Thereafter, the distance between the curves is computed between their interpolation functions. We refer the readers to Ferraty and Vieu (2006) for more details on the mathematical formulation for the functional metric. Meanwhile, we use the leave-one-out Cross-Validation (CV) technique to select the parameter k and the bandwidth parameter h_n. We point out that the CV procedure is implemented with respect to the Mean Square Error (MSE) defined by *Equation 17*

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \theta(X_i)\right)^2 \quad \text{(Eq.17)}$$

where $\theta(.)$ Refers to one of the previous models $\hat{r}, \tilde{r}, \bar{r}$ *and* $\ddot{r}$. On the other hand, we choose the same locative function β and the same kernel K as Rachedi et al. (2014). To evaluate the efficiency of the proposed model in this prediction issue, we divide random sample of 151 into two parts as learning sample (size = 100) and testing sample (size = 51). Further, in *Figures 3–5* we draw the box-cox plots of the absolute errors of the testing sample curves of CP, ADF and NDF respectively for four proposed regression models along with previous classical models namely PLS and PCR. The prediction results are shown in *Figures 3–5*.

It is depicted in *Figures 3–5* that there is a significant gain among the functional models as compared to the classical ones. The models FLQR, FNQR, FLLQR, and FLLMR are giving small absolute errors followed by PLS and PCR (*Table 1*).

Table 1. *Absolute error (AE) of crude protein (CP), acid detergent fiber (ADF) and neural detergent fiber (NDF) contents using proposed functional models and classical models*

Models	FLQR	FNQR	FLLQR	FLLMR	PCR	PLS
CP	0.73	1.11	0.73	0.62	2.95	1.45
ADF	0.82	1.24	0.81	1.12	7.08	2.03
NDF	0.67	1.17	0.77	0.71	5.24	1.78

The same fact is confirmed by *Table 2*, where we present the root mean squares error (RMSE) expressed by *Equation 18*

$$RMSE = \sqrt{\frac{1}{25}\sum_{j=1}^{25}\left(Y_i - \theta(X_i)\right)^2} \quad \text{(Eq.18)}$$

Table 2. *Root mean square errors (RMSE) of crude protein (CP), acid detergent fiber (ADF) and neural detergent fiber (NDF) contents using proposed functional models and classical models*

Models	FLQR	FNQR	FLLQR	FLLMR	PCR	PLS
CP	0.85	1.04	0.88	0.69	2.83	1.54
ADF	0.96	1.11	0.97	1.04	6.87	1.91
NDF	0.89	1.09	0.90	0.83	5.08	1.72

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346

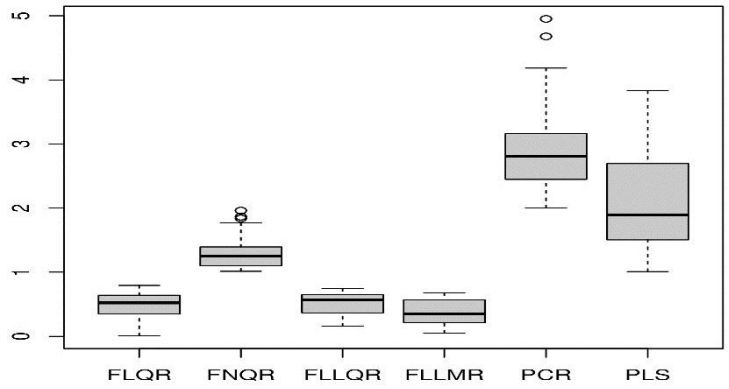

Figure 3. *Box-cox plots of absolute errors of crude protein (CP) contents using functional models and classical models*

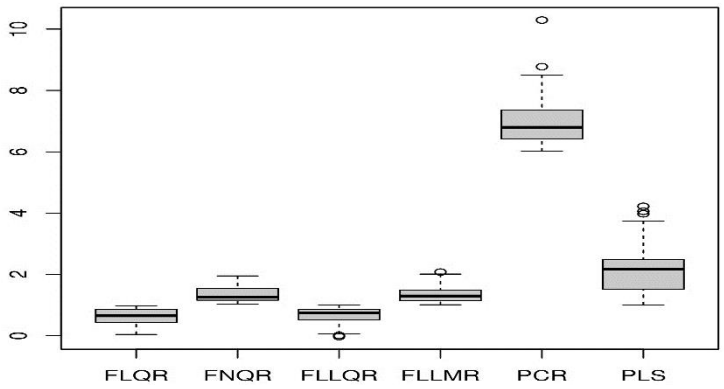

Figure 4. *Box-cox plots of absolute errors of acid detergent fiber (ADF) contents using functional and classical models*

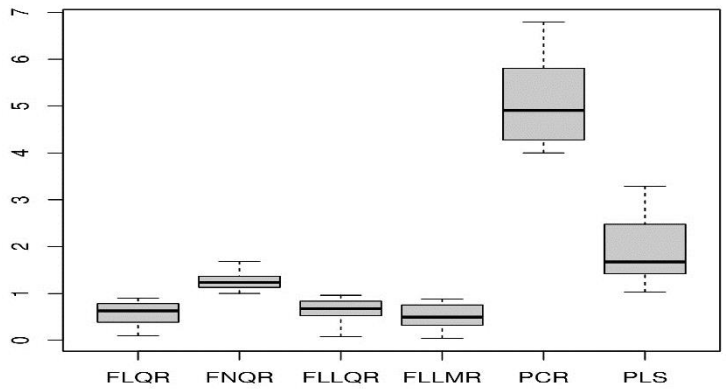

Figure 5. *Box-cox plots of absolute errors of neural detergent fiber (NDF) contents using functional models and classical models*

The values of RMSE is relatively stable and smaller for the four proposed functional models namely FLQR, FNQR, FLLQR and FLLMR as compared to the classical models namely PLS and PCR.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346

Although the performance of the studied models is varied, the variability of the RMSE is relatively stable for the four proposed models FLQR, FNQR, FLLQR and FLLMR as compared to that of the classical models such as PLS and PCR (see *Table 3*).

Table 3. *The broader range of the RMSE*

Models	FLQR	FNQR	FLLQR	FLLMR	PCR	PLS
Range for CP	(0.7, 0.97)	(0.9, 1.93)	(0.35, 0.84)	(0.37, 0.76)	(2.21, 4.95)	(1.01, 3.82)
Range for ADF	(0.4, 0.98)	(1.15, 1.81)	(0.09, 0.99)	(0.97, 1.76)	(6.01, 10.25)	(1.15, 1.81)
Range for NDF	(0.5, 0.91)	(1.03, 1.68)	(0.11, 0.96)	(0.33, 0.88)	(4.11, 6.97)	(1.02, 3.29)

In addition, we also evaluate the performance of NIRS to determine the predictive region using four functional models. The accuracy of these predictive intervals is evaluated with the help of two statistical measures namely the coverage in probability (Co-P) and the mean length of the interval (ML).

In *Table 4*, we summarize the (Co.P) and the M.L of 95% confidence intervals as mentioned in *Equations 3, 6, 10* and *13*.

Table 4. *Accuracy measures in terms of coverage in probability (Co-P) and mean length (ML) for predictive intervals using functional models*

Models	FLQR	FNQR	FLLQR	FLLMR
Co.P-CP	0.97	0.94	0.97	0.98
Co.P-ADF	0.95	0.91	0.93	0.94
Co.P-NDF	0.96	0.94	0.95	0.97
M.L-CP	0.41	0.65	0.34	0.28
M. L.-ADF	0.49	0.75	0.47	0.34
M.L-NDF	0.38	0.69	0.42	0.30

Discussion

The principal forage quality parameters were evaluated using a sample of 151 observations and the results summarized in *Tables 1–3*. The analyzed parameters are CP, ADF and NDF which are ranged between 5.59%-15.76%, 23.81%-47.86% and 62.96%-81.41%, respectively. Such a data analysis was operated using four functional models: FLQR, FNQR, FLLQR and FLLMR.

Based on the results from *Tables 1–3*, it is clear that the best model (having a small AE and RMSE) is the FLQR; it has an absolute error between 0.85-0.96 and its RMSE is valued between 0.67-0.82. Overall, it appears clearly that efficiency of the NIR Spectroscopy to predict the chemical properties differs from one parameter to another and also from one model to another. In particular, the best result is obtained by the FLLMR model in the prediction of the CP quantities. The model FLLQR is more adequate for the ADF parameter as well as for the NDF. On the other hand, both RMSE and AE emphasize the superiority of the functional model against the multivariate models (represented by the PCR and PLS regressions). Furthermore, the best multivariate model is the PCR for which RMSEs are respectively 2.83, 6.87 and 5.08 of CP, ADF and NDF. Moreover, the prediction results confirm that the robustness of the proposed functional models over the multivariate ones have lower variability in the

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346

prediction error (RMSE and AE). Undoubtedly, the performance of the functional algorithms over that for the multivariate model is justified by the adequate exploration of the information based on the NIR data. Whilst the efficiency of the multivariate model is linked to several uncontrolled arguments such as the number of regressors, number of factors, the regression equation, etc. Furthermore, the functional approach has also satisfactorily performance for the prediction by the confidence interval as shown in *Table 4*. Precisely, the results of *Table 4* show that there is no substantial difference among the four models for the Co.P parameter as well as for ML. Nevertheless, by taking into account the ratio of Co.P and ML, we can say that the most suitable model in the sense of accuracy measure for the CP is FLLMR while for ADF and NDF is the FLQR model.

Conclusion

In this work, we have developed a new approach to predict the forage quality of Chinese ryegrass in term of its chemical components. We are interested to the prediction the CP, NDF and ADF contents. The proposed approaches combine the recent development in Chemistry and the modern Statistics. Specifically, from the Chemistry, we use the NIR spectroscopy technology which is inexpensive, rapid, reliable and accurate method. Moreover, it reduces the need for conventional wet Chemistry procedures. On the other hand, form the modern statistics we use some functional models which allow to explore all the information of the spectroscopy analysis where the spectral data are viewed as curves. These models are easily implementable and improve substantially of the prediction results. Moreover, the proposed models permit to avoid the core drawback of the PLS and PCR models that is the loss of information after the predictors transformation. Indeed, it is well known that the prediction by the PLS or the PCR regressions requires some transformation of the exogenous variable to reduce the dimension, get the linearity or remove the effect of the outliers. However, all these transformations are carried out through some directions chosen independently to the prediction problem. In particular, the endogenous variable does not intervene in the choice of the optimal direction which can affect the performance of the prediction. Thus, the originality of the functional approach comes from the fact that the prediction problem is performed without any transformation of the data. In conclusion, we can say that this combination of modern Chemistry and modern Statistics is very beneficial ecologically, economically and is more flexible, robust and fast. To the end, let us point out that in addition to pointwise, the proposed models offer the possibility to precise a predictive interval with a given confidence level which being more informative than a single-point prediction in certain situations.

We mention that the NIR analysis of the forage data is a promising analysis technique. It has been regarded as a new fast and reliable method compared to traditional analytical methods. In the last decade, this kind of technique is becoming more and more popular for the screening of forage quality parameters. See, for instance, Asekova et al. (2016) for a list of references). In this study, we combine this modern chemical technique together with the recent development of mathematical statistics in order to provide performance prediction procedures permits to use the whole spectroscopy curves as regressors.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346

Acknowledgments. Authors greatly thank the Editor in chief, the Associate Editor and the anonymous referees for a careful reading of the paper and for their valuable comments and suggestions which improved the quality of this paper substantially. They also extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through research groups program under the project number R.G.P1/102/40.

REFERENCES

[1] Agnihotri, R. K., Palni, L. M. S., Singh, B., Pangtey, Y. P. S. (2003): Evaluation of fodder quality of straw of different landraces of rice (Oryza sativa L.) under cultivation in the Kumaun Region of Indian Central Himalaya. – International Journal of Sustainable Development and World Ecology 10: 391-400.

[2] Al-Awadhi, F., Kaid, Z., Laksaci, A., Ouassou, I., Mustapha, R. (2018): Functional data analysis: local linear estimation of the L1-conditional quantiles. – Statistical Methods & Applications 28(2): 217-240.

[3] Albrecht, R., Richard, J., Jean, L., Gerard, T., Claude, P. (2008): Calibration of chemical and biological changes in composting of biowastes using near-infrared Spectroscopy. – Environnemental Science & Technology 43(3): 804-811.

[4] Aneiros, G., Cao, R., Fraiman, R., Vieu, P. (2019): Special issue on functional data analysis and related topics. – J. Multivariate Anal. 170: 1-336.

[5] AOAC (1990): Protein (Crude) Determination in Animal Feed: Copper Catalyst Kjeldahl Method 984.13. 15th Ed. – Official Methods of Analysis of AOAC International, Maryland.

[6] Asekova, S., Sang-Ik, H., Hong-Jib, C., Sang-jo, P., Dong-Hyun, S., Chan-Ho, K., Grover Shannon, J., and Jeong Dong, L. (2016): Determination of forage quality by near-infrared reflectance spectroscopy in soybean. – Turkish Journal of Agriculture and Forestry 40(1): 45-52.

[7] Bai, Q. L., Chen, S. J., Dong, X. L., Meng, Q. X., Yan, Y. L., Dai, J. R. (2004): Prediction of NDF and ADF concentrations with near infrared reflectance spectroscopy (NIRS). – Guang Pu Xue Yu Guang Pu Fen Xi = Guang Pu 24(11): 1345-1347.

[8] Cardot, H., Crambes, C., Sarda, P. (2005): Quantile regression when the covariates are fonctions. – Nonparametric Statistics 17(7): 841-856.

[9] Chen, J., Zhu, R., Xu, R., Zhang, W., Shen, Y., Zhang, Y. (2015): Evaluation of Leymus chinensis quality using near-infrared reflectance spectroscopy with three different statistical analyses. – Peerj 3: e1416. DOI: 10.7717/peerj.1416.

[10] Chen, P. F., Rong, Y. P., Han, J. G., Wang, J. H., Zhang, L. D., Xu, X. Y. (2007): Evaluation of fresh sample of alfalfa silage through near infrared reflectance spectroscopy (NIRS). – Guang Pu Xue Yu Guang Pu Fen Xi = Guang Pu 27(7): 1304-1307.

[11] Chikr-Elmezouar, Z., Almanjahie, I., Laksaci, A., Rachdi, M. (2018): FDA: strong consistency of the k NN local linear estimation of the functional conditional density and mode. – Journal of Nonparametric Statistics 1-21.

[12] Collomb, G., Härdle, W., Hassani, S. (1986): A note on prediction via estimation of the conditional mode function. – Journal of Statistical Planning and Inference 15: 227-236.

[13] Dabo-Niang, S., Laksaci, A. (2012): Nonparametric quantile regression estimation for functional dependent data. – Communications in Statistics - Theory and Methods 41(7): 1254-1268.

[14] De Marchi, M., Riovanto, R., Penasa, M., Cassandro, M. (2012): At-line prediction of fatty acid profile in chicken breast using near infrared reflectance spectroscopy. – Meat Science 90(3): 653-657.

[15] En, X. Z., Guo, H. R., Jia, Y. S., Ge, G. T., Wang, K. (2009): Application and prospect of near infrared reflectance spectroscopy in forage analysis. – Spectroscopy and Spectral Analysis 29: 635-640.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346

[16] Ferraty, F., Vieu, P. (2006): Nonparametric Functional Data Analysis: Theory and Practice. – Springer Science & Business Media, New York.
[17] Ferraty, F., Laksaci, A., Vieu, P. (2006): Estimating some characteristics of the conditional distribution in nonparametric functional models. – Statistical Inference for Stochastic Processes 9(1): 47-76.
[18] Foley, W. J., McIlwee, A., Lawler, I., Aragones, L., Woolnough, A. P., Berding, N. (1998): Ecological applications of near infrared reflectance spectroscopy–a tool for rapid, cost-effective prediction of the composition of plant and animal tissues and aspects of animal performance. – Oecologia 116(3): 293-305.
[19] Goia, P., Vieu, P. (2016): An introduction to recent advances in high/infinite dimensional statistics. – J. Multivariate Anal. 146: 1-6.
[20] Hyndman, R. J. (1995): Highest-density forecast regions for nonlinear and non-normal time series models. – Journal of Forecasting 14(5): 431-441.
[21] Kamruzzaman, M., ElMasry, G., Da-Wen, S., Allen, P. (2012): Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression. – Innovative Food Science & Emerging Technologies 16: 218-226.
[22] Koenker, R., Bassett, G. Jr. (1978): Regression quantiles. – Econometrica: Journal of the Econometric Society 46(1): 33-50.
[23] Laksaci, A., Lemdani, M., Ould-Saïd, O. (2009): A generalized L1-approach for a kernel estimator of conditional quantile with functional regressors: consistency and asymptotic normality. – Statistics & Probability Letters 7998: 1065-1073.
[24] Manteiga, W. G., Vieu, P. (2007): Statistics for functional data. – Computational Statistics & Data Analysis 51(10): 4788-4792.
[25] Monrroy, M., Dehylis, G., Marissa, M., Karla, H. (2017): Determination of brachiaria spp. forage quality by near-infrared spectroscopy and partial least squares regression. – Journal of the Chilean Chemical Society 62(2): 3472-3477.
[26] Mott, G., Moore, J. (1970): Forage evaluation techniques in perspective. – Proceedings Natn Conf Forage Qual Eval Util., Univ Nebraska.
[27] Nie, Z. D., Han, J. G., Zhang, L. D., Li, J. H. (2007): Applications of near infrared reflectance spectroscopy technique (NIRS) to grassland ecology research. – Guang Pu Xue Yu Guang Pu Fen Xi Guang Pu 27(4): 691-696.
[28] Norris, K. H., Barnes, R. F., Mooret, J. E., Shenk, J. S. (1976): Predicting forage quality by infrared reflectance spectroscopy. – J. Anim. Sci. 43: 889-897.
[29] Qingguo, T., Linglong, K. (2017): Quantile regression in functional linear semiparametric model. – Statistics 51(6): 1342-1358.
[30] Stuth, J., Abdi, J., Doug, T. (2003): Direct and indirect means of predicting forage quality through near infrared reflectance spectroscopy. – Field Crops Research 84(1-2): 45-56.
[31] Suksombat, W. (2004): Comparison of different alkali treatment of bagasse and rice straw. – Asian Australasian Journal of Animal Sciences 17: 1430-1433.
[32] Valderrama, M. J. (2007): An overview to modelling functional data. – Comput Stat 22: 331-334.
[33] Van Soest, P. J., Robertson, J. B., Lewis, B. A. (1991): Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. – Journal of Dairy Science 74: 35833597.
[34] Wilcox, R. (2005): Introduction to Robust Estimation and Hypothesis Testing (Statistical Modeling and Decision Science). Vol. 2. – Academic Press, San Diego, CA.
[35] Yan, R., Chen, S., Zhang, X., Han, J., Zhang, Y., Undersander, D. (2011): Effects of replacing part of corn silage and alfalfa hay with Leymus chinensis hay on milk production and composition. – American Dairy Science Association 94(7): 3605-3608 DOI: 10.3168/jds.2010-3536.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346

[36] Yang, Z., Gang, N., Ling, P., Yan, Z., Linkai, H., Xiao, M., Xinquan, Z. (2017): Development and validation of near-infrared spectroscopy for the prediction of forage quality parameters in Lolium multiflorum. – PeerJ 5: e3867.
[37] Yu, G., Min-Jie, W., Yi-Hong, L., Xing-Ming, Y., Wei, R., Qi-Rong, S. (2011): Fluorescence excitation–emission spectroscopy with regional integration analysis for assessment of compost maturity. – Waste Management 31(8): 1729-1736.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1706_1433314346