The Use of Principle Component Analysis in Type Classification of Air-dry Peat

N Chukhareva, A Zarubin and O Bulgakova
Tomsk Polytechnic University, Tomsk, Russia

E-mail: zagtpuru@gmail.com

Abstract. The use of principle component analysis (PCA) in the variant of projection on latent values with discriminant analysis (PLS-DA) in type classification of Siberian region air-dry peat by a set of properties is presented. A statistical analysis in principle component space by PCA of different physico-chemical properties of peat such as component composition, concentration of paramagnetic centers and IR-spectra is presented and shows a developed PLS-DA classification model allows estimating peat type by a set of physico-chemical properties with minimum prediction errors.

1. Introduction

The study of peat's composition and properties has always been paid great attention to because of its value in the organic part [1–3]. Peat-based materials are widely used for various tasks in the areas of agriculture, industry, as well as resource-efficient solutions for environmental problems [4–5]. But, because peat is an initially difficult natural object, including both organic and inorganic components, there is no uniform approach to choosing the direction of its use at the present time. This is caused by a different degree of influence in botanical composition, type of peat and degradation amount of the organic component on the final product's properties. Therefore, it is important to develop a classification model that contains comprehensive information based on raw material properties. On one hand, the collection of many different characteristics of peat, based on physico-chemical and technical analysis methods, does not always allow excluding part of the incorrectly obtained data considering all of the investigated properties. Not one of their series is possible to identify the specificity of the object. On the other hand, it is impossible to identify objects by baseline characteristics, such as type, if data are not initially.

2. Experiment

PCA is widely used in the analysis of the properties and composition of natural objects. So in [6], the authors used the space of the principal components to explain the distribution of peat on the ground, depending on the composition.

The aim of this work is to apply the PLS-DA method to classify the type of air-dry peat by a set of physico-chemical properties.

One of the solutions for the task is to apply PCA to classify objects by a set of different properties [7–9]. This method allows representing the observed data matrix (X) as the product of two matrices, plus the residual error, – E [10]:

$$X = SL + E, \tag{1}$$

where L is the intrinsic properties of the data matrix (loadings), and S is the matrix of individual parameters (scores).

The instrumentation of PCA allows the reduction in dimensionality of multivariate data by highlighting the principal components (PC) and interpreting the properties of investigated objects in the new space of PC's [11]. The implementation of PLS in an Excel spreadsheet editor is described by A.L. Pomerantsev [12].

In accordance with the intended aim we put forward following tasks:

- calculation of the scores matrix by physico-chemical properties of different types of air-dry peat;
- building and validation of the PLS-model.

The objects of research: twenty samples of two types (lying and lowland) of air-dry peat of eleven fields in the Tomsk region (Russia) were examined for the set of physico-chemical parameters such as botanical composition [13], elemental composition [14], group composition, functional composition by IR spectroscopy and electron paramagnetic resonance. The set of physico-chemical parameters were presented in the initial data matrix by fourteen variables (table 1).

Table 1. The set of physico-chemical parameters.

Variable	Physico-chemical parameter
V1	lg(S) – logarithm of the concentration of paramagnetic centers (electron paramagnetic resonance)
V2	bitumen content, % on daf
V3	content of humic acids, % on daf
V4	nonhydrolyzable residue content, % on daf
V5	content of water soluble and readily substances, % on daf
V6	fulvic acid content, % on daf
V7	cellulose content, % on daf (the method of determination of group composition)
V8	carbon content, % on daf
V9	hydrogen content, % on daf
V10	nitrogen content, % on daf
V11	the total content of oxygen and sulfur, % on daf (method of determining the elemental composition)
V12	$D_{3400/1620\text{-}1600}$ – ratio of optical densities corresponding to the characteristic lines of the IR spectrum
V13	$D_{2920/1620\text{-}1600}$ – ratio of optical densities corresponding to the characteristic lines of the IR spectrum
V14	$D_{125\text{-}1200/1620\text{-}1600}$ – ratio of optical densities corresponding to the characteristic lines of the IR spectrum

Data structure of two types of peat (lying – Class1 and lowland – Class2) properties investigation are presented in table 1 and it was divided for calibration and validation sets.

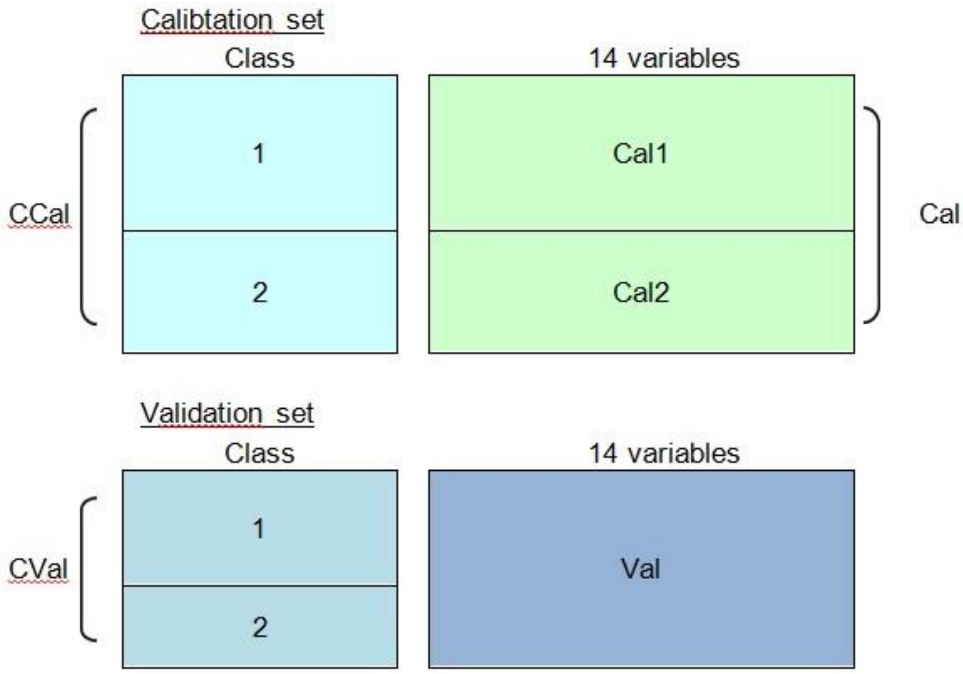

Figure 1. Data structure.

3. Results and discussion

Based on the data sets, the mathematical model of classification was constructed and validated. The results were obtained for the three latent variables and two of them are shown in figure 2.

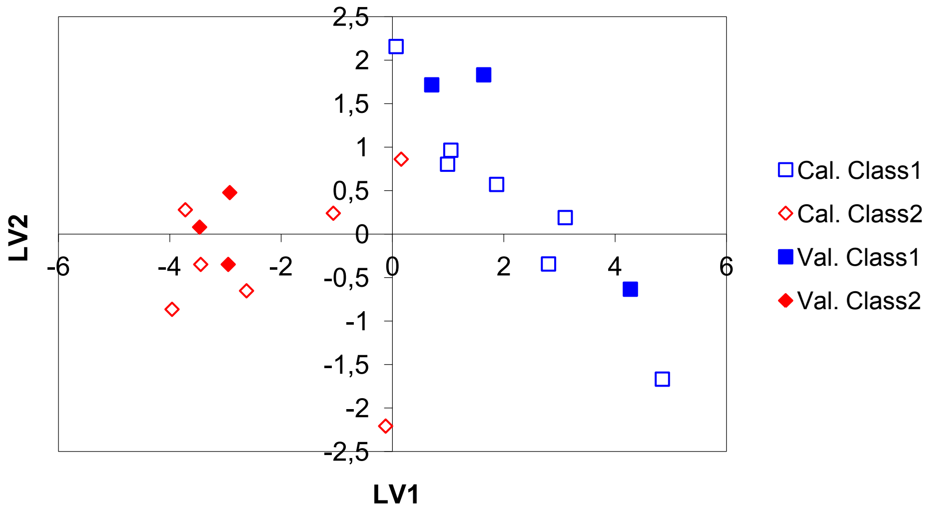

Figure 2. Scores of latent variable 1 (LV1) and latent variable 2 (LV2).

The constructed mathematical model with level of discrimination 0.5 allows classifying air-dry peat by type (prediction errors are presented in figure 3).

	Number of latent variables in model:		
	1 LV	**2 LVs**	**3 LVs**
Training set			
Type I errors:	0	0	0
Type II errors:	1	1	0
Validation set			
Type I errors:	0	0	0
Type II errors:	0	0	0

Figure 3. Prediction errors.

The plot of PLS-predictions of three latent variables is presented in figure 4.

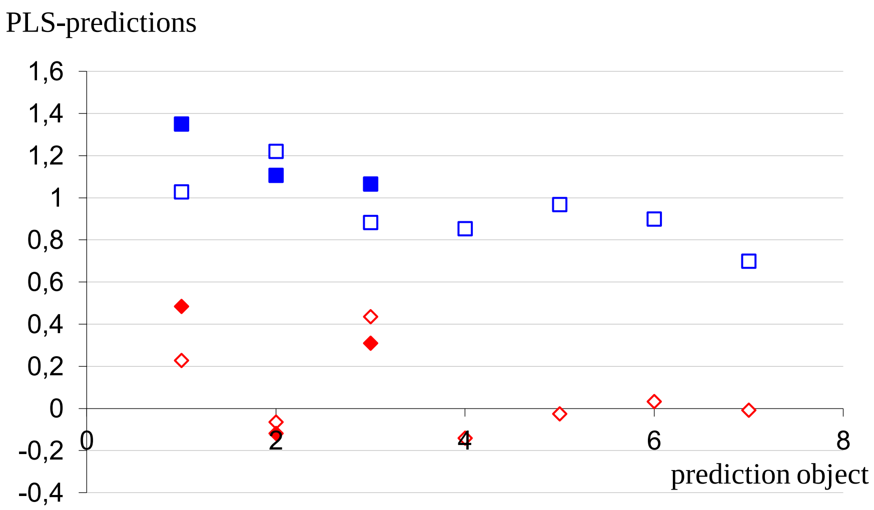

Figure 4. PLS-predictions of tree latent variables.

4. Conclusion

The principal component analysis in variant of projection on latent values with discriminant analysis was used for estimating changes in the composition and properties of air-dried peat, depending on its type. The PLS mathematical model was constructed and validated with minimum prediction errors. The analysis clearly shows that lowland Siberian peat very easy to discriminate from the lying Siberian peat. With PLS discriminating models using the whole set of 14 physico-chemical parameters, it is possible to achieve excellent discrimination almost all the time. It must be underlined that the statistical model shown in this article is built for Siberian peat only, but it can be extended for the others peat types with model correction.

References

[1] Arkhipov V S, Preis U P, Bernatonis V K, Maslov S G and Antropova N A 2008 Bituminous peat of the Tomsk region Tomsk: STT [in Russian]

[2] Lishtvan I I and Korol N T 1975 Main properties of peat and methods of their definition Minsk: Science and technology 320 12 [in Russian]

[3] Kujala K, Seppälä M, and Holappa T 2008 Physical properties of peat and palsa formation. *Cold Regions Science and Technology* **52** (3) 408-414

[4] Sokolov B N and Yampolski A L 1986 Socio-economic consequences of industrial development of peat deposits in the USSR In *Socio-economic impacts of the utilization of peatlands in industry and forestry: proceedings of the IPS Symposium, Oulu, Finland, June 9-13, 1986* p 69. International Peat Society, Finnish National Committee.

[5] Shishmina L V, Chukhareva N V and Kravtsov A V 2002 Influence of Peat Humates on Flotation of Coal *Coke and Chemistry* **(2)** 7-9

[6] Muller J, Kylander M, Martinez-Cortizas A, Wüst R A, Weiss D, Blake K, and Garcia-Sanchez R 2008 The use of principle component analyses in characterising trace and major elemental distribution in a 55 kyr peat deposit in tropical Australia: implications to paleoclimate *Geochimica et Cosmochimica Acta* ***72*****(2)** 449-463

[7] Brereton, R G 2007 *Applied chemometrics for scientists* John Wiley & Sons

[8] Wold S, Sjöström M and Eriksson L 2001 PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems* **58**(2) 109-130

[9] Forina M, Oliveri P, Lanteri S and Casale M 2008 Class-modeling techniques, classic and new, for old and new problems *Chemometrics and Intelligent Laboratory Systems* **93**(2) 132-148

[10] Kowalski B R (Ed) 1984 *Chemometrics: mathematics and statistics in chemistry* No 138 Springer

[11] Wehrens R 2011 *Chemometrics with R: multivariate data analysis in the natural sciences and life sciences* Springer

[12] Pomerantsev A L 2014 *Chemometrics in Excel* John Wiley & Sons

[13] GOST 28245-89 *Turf. Method of determination of botanical composition and degree of composition* Russia: Federal Agency on Technical Regulating and Metrology

[14] Fadeeva V P, Tikhova V D and Nikulicheva O N 2008 Elemental analysis of organic compounds with the use of automated chns analyzers Journal of Analytical Chemistry **63**(11) 1094-1106