Microporous and Mesoporous Materials 130 (2010) 309–313

Contents lists available at ScienceDirect

Microporous and Mesoporous Materials

journal homepage: www.elsevier.com/locate/micromeso

Machine learning study of the heulandite family of zeolites

Shujiang Yang [a], Mohammed Lach-hab [a], Estela Blaisten-Barojas [a,b,*], Xiang Li [a,c], Vicky L. Karen [c]

[a] Computational Materials Science Center, George Mason University, MSN 6A2, Fairfax, VA 22030, USA
[b] Department of Computational and Data Sciences, George Mason University, MSN 6A2, Fairfax, VA 22030, USA
[c] Ceramics Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

ARTICLE INFO

Article history:
Received 20 August 2009
Received in revised form 22 November 2009
Accepted 24 November 2009
Available online 27 November 2009

Keywords:
Zeolite
Heulandite
Clinoptilolite
Data mining clustering
Machine learning

ABSTRACT

Heulandite and clinoptilolite form the most abundant family of natural zeolite crystals. The topology of both of them is characterized by the framework type HEU. Despite many studies on these crystals, the mineral assignment to a zeolite as heulandite or clinoptilolite is still controversial and unresolved today. Based on a machine learning clustering analysis of crystallographic data of zeolite crystals, we show that zeolites belonging to the HEU framework type are divided into three groups of minerals instead of two. Two of the groups, HEU-h and HEU-c, contain crystals with names heulandite and clinoptilolite, respectively. The third newly proposed group HEU-m is composed of mixed zeolites named under both traditional names. The grouping is based on the EM algorithm and a set of descriptors built from data collected in the NIST/FIZ Inorganic Crystal Structure Database. Verification of the division of the HEU family into three groups is provided based on a battery of machine learning tests.

1. Introduction

Clinoptilolite and heulandite are two closely related natural zeolites and commonly coexist in sedimentary rocks of volcanic origin. The topological arrangement of the TO_4 (T-atom is Si or Al) building blocks in these two minerals form a network recognized by the International Zeolite Association as HEU [1]. There are five distinct positions of T-atoms in the framework with different coordination sequences and vertex symbols. The highest symmetry space group in this framework is C2/m. The underlying structure contains oblate channels confined by alternating 10- and 8-member rings parallel to the *c*-axis. These channels are crosslinked by additional 8-member ring channels along the [1 0 0] and [1 0 2] directions giving rise to a two-dimensional channel system parallel to [0 1 0] [2].

Although a variety of zeolite samples are described as either heulandite or clinoptilolite on the basis of chemical parameters and/or thermal behavior, in several instances the mineral name is assigned based only on the site where the sample originates [3]. The name heulandite is used for samples in vugs of igneous rocks, while samples in diagenetically altered vitroclastic sediments are assigned the name clinoptilolite. Therefore, the naming of heulandite or clinoptilolite in the literature is ambiguous. Some publications refer to the type of zeolite only based upon the sample origin and without further analysis. Other authors argue that two mineral names are not necessary to describe these zeolites and suggest that silica-rich heulandite could replace the name clinoptilolite [4]. The Structure Commission of the International Zeolite Association provides the unique HEU framework type code for the heulandite family of zeolites, without commenting on differences between heulandite and clinoptilolite.

However, despite having identical framework topology, there are apparent differences between natural clinoptilolite and heulandite species. For example, the majority of clinoptilolites have slightly higher Si/A1 ratio in the framework than heulandites and are richer in monovalent alkali ions (e.g. Na^+, K^+) than in divalent alkaline earth ions (e.g. Ca^{2+}). As a result, some properties of clinoptilolites such as unit cell dimension, thermal stability, optical refraction, NMR spectroscopy, incoherent inelastic neutron scattering, differ from those of heulandites [5–8]. Empirical methods have been proposed to distinguish clinoptilolites from heulandites based on one or two of the following properties: Si/Al ratio [6,9–12], cation composition [10–13], unit-cell parameters [6], thermal stability [9,10], optical property [6], or spectroscopic properties [7]. It is indisputable that thermal stability and spectroscopic properties are useful tools for compound identification, but these properties are based on the structure and composition and therefore are not fundamental to define a mineral category [5,14]. Unit-cell parameters from X-ray diffraction are not appropriate for distinguishing heulandite from clinoptilolite because they are not uniquely defined in crystallography and are sensitive to changes in both water content and extra cation composition [5]. A distinction method based on cation composition is disputable since cations in

---

[*] Corresponding author. Address: Computational Materials Science Center, George Mason University, MSN 6A2, Fairfax, VA 22030, USA. Tel.: +1 703 9931988; fax: +1 703 9939300.

E-mail address: blaisten@gmu.edu (E. Blaisten-Barojas).

1387-1811/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.micromeso.2009.11.027

zeolites are highly exchangeable. Definitely, cation-exchange cannot convert clinoptilolite to heulandite or the other way around.

The Si/Al ratio reflects the chemistry of the rigid framework but has no topological content. Zeolite mineral species are not normally distinguished solely on Si/Al ratio. Indeed, many zeolite species may possess variable Si/Al ratio and display the same framework structure. Despite this fact, the subcommittee on zeolites of the International Mineralogical Association, Commission on New Minerals and Mineral Names, has supported the identification of heulandite and clinoptilolite based on the Si/Al ratio [14]. This is an exception based on the established usage of the heulandite and clinoptilolite names and on the convenience for recognizing a chemical feature. Therefore, the official definition of *heulandite* is a zeolite mineral with framework topology HEU and the ratio $Si/Al < 4$ and the definition of *clinoptilolite* is a zeolite with the same framework topology and $Si/A1 \geqslant 4.0$. The threshold value of 4 is arbitrary and adopted following reference [6].

Adding to the ambiguity of the distinction between clinoptilolite and heulandites within the HEU family, intermediate species were reported as a third group of the family. The so-called Type-II heulandite is an intermediate type of the isomorphous series heulandite–clinoptilolite end members based on the thermal resistance behavior [10,15]. Perraki and Orfanoudaki [16] analyzed zeolite samples from Thrace, Greece, and based on a series of characterization techniques on the thermal and cation-exchange properties, they concluded that their samples are characteristic of the intermediate class heulandite Type-II.

Hawkins was one of the first to recognize the existence of multiple groups in the heulandite family based entirely on chemical grounds [17]. Using cluster analysis on data of 78 zeolite crystals he showed that at least five compositional subgroups exist within the heulandite family of zeolites, with the traditional heulandites and clinoptilolites being the most dissimilar subgroups. Multiple polymorphs in the heulandite family of zeolites with identical chemical composition may exist due to different distributions and ordering of Si and Al atoms occupying the tetrahedral sites of the framework [18]. In analogy to alkali feldspars it can be postulated that each clinoptilolite or heulandite may be structurally different, even if a constant Si/Al ratio and cation composition are maintained. The problem of how many groups do really exist within the heulandite family of zeolites is not only of academic interest but has also strong influence on cation diffusion, cation exchange, gas sorption, and catalytic properties [18].

In this work, we present a machine learning clustering methodology for identifying groups of zeolites within the HEU framework type. Unlike Hawkins [17] who used compositions of six exchangeable cations in addition to those of Si and Al as the descriptors for clustering, our methodology aims at using compositional and structural information solely from the rigid framework. Our goal is to predict *a priori* sub-grouping information of the heulandite family of zeolites without knowledge of their thermal stability and spectroscopic properties. This paper is organized as follows. Section 2 is a description of the methodology, Section 3 is a compendium of our results and the paper is concluded in Section 4.

2. Machine learning clustering

The crystallographic information was obtained from the FIZ/NIST Inorganic Crystal Structural Database (ICSD) [19]. The ICSD is the most complete library of zeolite crystal structures assembled from publications of X-ray diffraction experiments. There are 78 zeolite crystals in the ICSD queried as either heulandite or clinoptilolite, which should display the HEU framework. A cross-search in the American Mineralogist Crystal Structure Database [20] yielded 40 HEU type zeolites, out of which 39 are in the ICSD. After examining thoroughly their crystal structure records, eight of the 78 crystals were removed from our consideration due to incomplete or erroneous information in the ICSD records. Six of these problematic records have been reported in Tables 1 and 2 of Ref. [21] and records of the remaining two do not contain the Si and Al concentration. This exclusion leaves 70 zeolites with HEU topology in the dataset, which are all confirmed to have the HEU framework type by calculating coordination sequences and vertex symbols using the zeoTsites program [22] plus tiling sequences using the TOPOS program [23]. The mineral names assigned to them in the original publications are either “clinoptilolite” or “heulandite” (except for one named “heulandite B” [24], indicating that it is the heat-collapsed phase of heulandite). Since naming from different authors is not consistent, we first provide a generic name to each crystal according to the nomenclature of the International Mineralogical Association based on the Si/Al ratio [14]. With this criterion, our data set has 50 heulandites and 20 clinoptilolites, which are listed in Table 1. In the rest of this paper we refer to entries in the data set of 70 zeolites with these names as listed in Table 1.

Within the realm of machine learning, a “feature” is the specification of a property that has a value for each zeolite crystal under study. A machine learning model is based on a feature vector composed of all pertinent features. In previous works, we explored tens of topological, chemical and physical features for the machine learning classification of zeolite framework types [25–28]. Due to the small number of available zeolites within the HEU framework, a reasonable feature vector should include as few independent features as possible. After examining the data, two features are chosen to be the most characteristic of the zeolite backbone. One feature is the silicon-to-aluminum molar ratio Si/Al that reflects the chemical composition of the framework. The second feature is the average tetrahedrality $\langle T \rangle$, which is chosen to characterize the framework topology [25]. Tetrahedrality T is the degree of distortion from a perfect tetrahedron for which $T = 0$. The feature $\langle T \rangle$ is calculated over tens of thousands of tetrahedra obtained through Delaunay tessellation of points associated with all the T-sites in the zeolite framework [26–28]. Values of these two features for the studied zeolites are given in Table 1. Although framework density and normalized reduced cell parameters are better features than unit-cell parameters in characterizing the framework dimension, neither one is used in the clustering analysis done din this paper due to their sensitivity to the adsorbent phase and extra-framework cations [3].

Cluster analysis is a data segmentation procedure for grouping a collection of samples into subsets identified as “clusters,” such that samples within a cluster are more similar to each other than samples assigned to different clusters. Each sample is described by a set of features. The degree of similarity between the individual samples being clustered is a central ingredient of any cluster analysis because any method will attempt to group the samples based on the type of similarity fed to the model. In our case there are 70 zeolites available and each of them is characterized by two features as given in Table 1. Our unsupervised learning model is a cluster analysis based on the Expectation Maximization (EM) algorithm [29] as implemented in WEKA [30,31]. EM is an iterative top-down clustering procedure, closely related to the K-means method [31], and proceeds by performing an expectation step assigning weights to each data point based on relative local densities and a maximization step by re-computing densities based on current values of the weights. In this clustering algorithm dissimilarities are based on the matrix of squared Euclidian distances between samples:

$$d(x_i, x_j) = \sum_{\alpha=1}^{2} (x_{i\alpha} - x_{j\alpha})^2 \tag{1}$$

Table 1
The 70 HEU type zeolites used in the study, their characteristic features, given mineral names and clustering results.

ICSD code	chemical formula	*Si/Al*	*<T>*	ICSD name	IMA [14] name	this work
22050	$C_{0.6} (Al_{2.4} Si_{6.6} O_{18}) (H_2 O)_{4.77}$	2.750	0.185	*Heulandite*	Heulandite	**HEU-h**
27526	$Ca_{1.5} (Al_{2.394} Si_{6.597} O_{18}) (H_2 O)_6$	2.756	0.181	*Heulandite*	Heulandite	**HEU-h**
31278	$(Na_{.26} K_{0.89} Ca_{3.37} Sr_{0.24} Ba_{0.03}) Al_{.48} Si_{26.61} O_{72} (H_2 O)_{24.84} H_{1.03}$	2.798	0.183	*Heulandite*	Heulandite	**HEU-h**
37061	$K_{8.48} (Al_9 Si_{27}) O_{72} (H_2 O)_{18}$	3.000	0.184	*Heulandite*	Heulandite	**HEU-h**
37062	$K_{6.92} (Al_9 Si_{27}) O_{72} (H_2 O)_{9.7}$	3.000	0.184	*Heulandite*	Heulandite	**HEU-h**
96825	$Sr_{4.23} (Al_{8.96} Si_{27.04} O_{72}) (H_2 O)_{25.76}$	3.016	0.183	*Heulandite*	Heulandite	**HEU-h**
96826	$Sr_{4.19} (A_{l8.96} Si_{27.04} O_{72}) (H_2O)_{24.98}$	3.016	0.183	*Heulandite*	Heulandite	**HEU-h**
96827	$Sr_{4.16} (Al_{8.96} Si_{27.04} O_{72}) (H_2 O)_{18.58}$	3.016	0.181	*Heulandite*	Heulandite	**HEU-h**
96828	$Sr_{4.56} (Al_{8.96} Si_{27.04} O_{72}) (H_2 O)_{17.16}$	3.016	0.180	*Heulandite*	Heulandite	**HEU-h**
96829	$Sr_{3.46} (Al_{8.96} Si_{27.04} O_{72}) (H_2 O)_{16.6}$	3.016	0.180	*Heulandite*	Heulandite	**HEU-h**
38399	$Ca_{4.48} Al_{8.96} Si_{7.04} O_{72} (H_2 O)_{24.5}$	3.018	0.183	*Heulandite*	Heulandite	**HEU-h**
92924	$Na_{1.72} K_{0.4} Ca_{2.65} Ba_{0.03} Sr_{0.87} (Al_{8.92} Si_{27.08}) O_{72} (H_2 O)_{26}$	3.036	0.183	*Heulandite*	Heulandite	**HEU-h**
92925	$Na_{1.72} K_{0.4} Ca_{2.65} Ba_{0.03} Sr_{0.87} (Al_{8.92} Si_{27.08}) O_{72} (H_2 O)_{26}$	3.036	0.181	*Heulandite*	Heulandite	**HEU-h**
92927	$Na_{1.72} K_{0.4} Ca_{2.65} Ba_{0.03} Sr_{0.87} (Al_{8.92} Si_{27.08}) O_{72} (H_2 O)_{26}$	3.036	0.183	*Heulandite*	Heulandite	**HEU-h**
75295	$Ca_{1.94} (Na_{0.91} Ca_{1.76}) (Na_{0.39} K_{0.13}) (Al_{8.9} Si_{27.1} O_{72}) (H_2 O)_{24.76}$	3.045	0.183	*Heulandite*	Heulandite	**HEU-h**
75296	$Na_{7.2} (Al_{8.9} Si_{27.1} O_{72}) (H_2 O)_{25.92}$	3.045	0.184	*Heulandite*	Heulandite	**HEU-h**
75297	$Pb_4 (Al_{8.9} Si_{2\ 7.1} O_{72}) (H_2 O)_{16.44}$	3.045	0.183	*Heulandite*	Heulandite	**HEU-h**
85696	$(NH_4)_{8.96} Al_{8.79} Si_{27.21} O_{72} (H_2 O)_{19.52}$	3.096	0.187	*Heulandite*	Heulandite	**HEU-h**
151180	$((CH_3) NH_3)_{8.16} Na_{0.52} ((Al_{8.7} Si_{27.3}) O_{72}) (H_2 O)_{10.77}$	3.132	0.189	*Heulandite*	Heulandite	**HEU-h**
151182	$((C_3 H_7) NH_3)_{0.56} Na_{8.68} ((Al_{8.7} Si_{27.3}) O_{72}) (H_2 O)_{19.68}$	3.132	0.182	*Heulandite*	Heulandite	**HEU-h**
151183	$((CH_3)_2 NH_2)_{6.92} Na_{0.36} ((Al_{8.7} Si_{27.3}) O_{72}) (H_2 O)_{7.80}$	3.132	0.184	*Heulandite*	Heulandite	**HEU-h**
97912	$Cd_{4.36} (Al_{8.7} Si_{27.3} O_{72}) (H_2 O)_{29.08}$	3.137	0.186	*Heulandite*	Heulandite	**HEU-h**
97913	$Cd_{4.15} (Al_{8.7} Si_{27.3} O_{72}) (H_2 O)_{26.92}$	3.137	0.185	*Heulandite*	Heulandite	**HEU-h**
97914	$Cd_{4.02} (Al_{8.7} Si_{27.3} O_{72}) (H_2 O)_{25.21}$	3.137	0.183	*Heulandite*	Heulandite	**HEU-h**
97915	$Cd_{4.02} (Al_{8.7} Si_{27.3} O_{72}) (H_2 O)_{25.58}$	3.137	0.183	*Heulandite*	Heulandite	**HEU-h**
40143	$K_{8.4} H_{.2} (Al_{8.6} Si_{27.4} O_{72}) (H_2 O)_{19.28}$	3.186	0.184	*Heulandite*	Heulandite	**HEU-h**
82119	$Na_{5.68} Ca_{1.52} (Al_{8.6} Si_{27.4} O_{72}) (H_2 O)_{21.4}$	3.186	0.183	*Heulandite*	Heulandite	**HEU-h**
82120	$K_{8.4} (Al_{8.6} Si_{27.4} O_{72}) (H_2 O)_{19.28}$	3.186	0.184	*Heulandite*	Heulandite	**HEU-h**
82121	$Rb_{8.44} (Al_{8.6} Si_{27.4} O_{72}) (H_2 O)_{17.36}$	3.186	0.184	*Heulandite*	Heulandite	**HEU-h**
68259	$Ca_{3.45} Rb_{1.5} (Al_{8.4} Si_{27.6} O_{72}) (H_2 O)_{23.5}$	3.286	0.182	*Heulandite*	Heulandite	**HEU-h**
91669	$Cd_{4.11} (Al_{8.22} Si_{27.78} O_{72}) (H_2 O)_{29.6}$	3.377	0.185	*Heulandite*	Heulandite	**HEU-h**
25029	$Ca_{1.16} (Al_2 Si_{6.95} O_{18}) (H_2 O)_6$	3.390	0.183	*Heulandite*	Heulandite	**HEU-h**
64767	$Ca_{1.23} (Al_2 Si_7 O_{18}) (H_2 O)_6$	3.390	0.183	*Heulandite*	Heulandite	**HEU-h**
66457	$(Ca_{1.88} Mg_{.08} Na_4 K_{0.28}) (Al_{8.16} Si_{27.84} O_{72}) (H_2 O)_{25.52}$	3.412	0.182	*Clinoptilolite*	Heulandite	**HEU-h**
66458	$(Ca_{1.8} Mg_{.16} Na_{4.24} K_{0.28}) (Al_{8.16} Si_{27.84} O_{72}) (H_2 O)_{24.88}$	3.412	0.182	*Clinoptilolite*	Heulandite	**HEU-h**
66459	$(Ca_{1.54} Mg_{0.1} Na_{3.28} K_{0.2}) (Al_{8.16} Si_{27.84} O_{72}) (H_2 O)_{17.64}$	3.412	0.182	*Clinoptilolite*	Heulandite	**HEU-h**
66461	$(Ca_{1.76} Na_{2.4} K_{0.52}) (Al_{8.16} Si_{27.84} O_{72}) (H_2 O)_{3.72}$	3.412	0.183	*Clinoptilolite*	Heulandite	**HEU-h**
97899	$Cu_{3.62} (H_3O)_{1.36} (Al_8 Si_{28} O_{72}) (H_2 O)_{28.88}$	3.500	0.188	*Heulandite*	Heulandite	**HEU-h**
97900	$(Cu_{2.44} (NH_3)_{5.44}) (NH_4)_{3.72} (Al_8 Si_{28} O_{72}) (H_2 O)_{22.08}$	3.500	0.184	*Heulandite*	Heulandite	**HEU-h**
87650	$Na_{1.56} H_{2.34} Al_{1.32} (Al_{7.86} Si_{28.14} O_{72}) (H_2 O)_{28.56}$	3.580	0.188	*Heulandite*	Heulandite	**HEU-h**
87651	$Na_{1.52} H_{2.71} Al_{1.21} (Al_{7.86} Si_{28.14} O_{72}) (H_2 O)_{25.84}$	3.580	0.187	*Heulandite*	Heulandite	**HEU-h**
87652	$Na_{0.76} H_{3.71} Al_{1.13} (Al_{7.86} Si_{28.14} O_{72}) (H_2 O)_{8.48}$	3.580	0.184	*Heulandite*	Heulandite	**HEU-h**
87653	$Na_{0.52} H_{4.88} Al_{0.82} (Al_{7.86} Si_{28.14} O_{72}) (H_2 O)_{6.20}$	3.580	0.184	*Heulandite*	Heulandite	**HEU-h**
37063	$K_{6.22} (A_{l9} Si_{27)} O_{72}$	3.000	0.173	*Heulandite*	Heulandite	**HEU-m**
37064	$K_{7.06} (A_9 Si_{27}) O_{72}$	3.000	0.175	*Heulandite*	Heulandite	**HEU-m**
92926	$Na_{1.72} K_{0.4} Ca_{2.65} Ba_{0.03} Sr_{0.87} (Al_{8.92} Si_{27.08}) O_{72} (H_2 O)_{26}$	3.036	0.175	*Heulandite*	Heulandite	**HEU-m**
97916	$Cd_{4.05} (A_{l8.7} Si_{27.3} O_{72}) (H_2 O)_{13.06}$	3.137	0.177	*Heulandite*	Heulandite	**HEU-m**
100745	$Ca_{3.17} Na_2 Al_{8.3} Si_{27.7} O_{72}$	3.329	0.176	*Heulandite*	Heulandite	**HEU-m**
66460	$(Ca_{1.32} Na_{3.12} K_{72}) (Al_{8.16} Si_{27.84} O_{72}) (H_2 O)_{15.92}$	3.412	0.178	*Clinoptilolite*	Heulandite	**HEU-m**
9262	$Ca_{.8} Na_{.4} Al_2 Si_7 O_{18} (H_2 O)_2$	3.500	0.170	*Heulandite B*	Heulandite	**HEU-m**
73413	$(Na_{0.28} Ca_{0.222})_4 (Ba_{0.08} K_{0.42})_4 (Al_{6.96} Si_{29.04} O_{72}) (H_2 O)_{6.72}$	4.172	0.178	*Clinoptilolite*	Clinoptilolite	**HEU-m**
73414	$(Na_{0.21} Ca_{0.11})_4 (Ba_{0.08} K_{0.53})_4 (Al_{6.96} Si_{29.04} O_{72}) (H_2 O)_{5.16}$	4.172	0.179	*Clinoptilolite*	Clinoptilolite	**HEU-m**
73415	$(Na_{0.19} Ca_{0.08})_4 (Ba_{0.08\ K\ 0.54})_4 (Al_{6.96} Si_{29.04} O_{72}) (H_2 O)_{4.84}$	4.172	0.178	*Clinoptilolite*	Clinoptilolite	**HEU-m**
87847	$(Na_{0.52} K_{2.44} Ca_{1.48}) (Al_{6.59} Si_{29.41} O_{72}) (H_2 O)_{28.64}$	4.464	0.181	*Clinoptilolite*	Clinoptilolite	**HEU-m**
97838	$Cs_{7.1} (Al_6 Si_{30} O_{72})$	5.000	0.181	*Clinoptilolite*	Clinoptilolite	**HEU-m**
34179	$Ca_{3.0} Ag_{1.3} Al_{7.2} Si_{28.8} O_{72} (H_2 O)_{17.5}$	4.000	0.184	*Heulandite*	Clinoptilolite	**HEU-c**
34180	$Ca_{2.9} Na_{1.1} Al_{7.2} Si_{28.8} O_{72} (H_2 O)_{20.5}$	4.000	0.184	*Heulandite*	Clinoptilolite	**HEU-c**
68258	$Na_{2.88} K_{0.37} Mg_{0.80} C_{0.84} Ba_{0.15} (Al_{6.84} Si_{29.16} O_{72}) (H_2 O)_{20.48}$	4.263	0.184	*Clinoptilolite*	Clinoptilolite	**HEU-c**
87846	$(Na_{1.32} K_{1.28} Ca_{1.72} Mg_{0.52}) (Al_{6.77} Si_{29.23} O_{72}) (H_2 O)_{26.84}$	4.319	0.184	*Clinoptilolite*	Clinoptilolite	**HEU-c**
72712	$(Cs_{5.62} K_{0.44} Mg_{0.26}) (Al_{6.58} Si_{29.42}) O_{72} (H_2 O)_{10.92}$	4.464	0.184	*Clinoptilolite*	Clinoptilolite	**HEU-c**
100095	$Ca_2 Na_{2.24} K_{1.48} Mg_{0.08} Al_6 Si_{30} O_{72} (H_2 O)_{22.76}$	4.999	0.183	*Clinoptilolite*	Clinoptilolite	**HEU-c**
100096	$Ca_{1.24} Na_{1.84} K_{1.76} Mg_2 Al_6 Si_{30} O_{72} (H_2 O)_{21.32}$	4.999	0.184	*Clinoptilolite*	Clinoptilolite	**HEU-c**
97837	$Cs_{7.39} (Al_6 Si_{30} O_{72}) (H_2 O)_{7.39}$	5.000	0.185	*Clinoptilolite*	Clinoptilolite	**HEU-c**
97839	$Cs_6 (Al_6 Si_{30} O_{72})$	5.000	0.183	*Clinoptilolite*	Clinoptilolite	**HEU-c**
97840	$Na_8 (A_{l6} Si_{30} O_{72}) (H_2 O)_{9.04}$	5.000	0.184	*Clinoptilolite*	Clinoptilolite	**HEU-c**
97841	$Na_6 (Al_6 Si_{30} O_{72})$	5.000	0.185	*Clinoptilolite*	Clinoptilolite	**HEU-c**
97842	$Na_6 (Al_6 Si_{30} O_{72})$	5.000	0.185	*Clinoptilolite*	Clinoptilolite	**HEU-c**
69390	$Na_{1.66} K_{2.56} Ca_{1.9} (Al_{5.48} Si_{30.52} O_{72}) (H_2 O)_{19.16}$	5.569	0.184	*Clinoptilolite*	Clinoptilolite	**HEU-c**
69391	$Cs_{3.98} Ca_{1.2} (Al_{4.76} Si_{31.24} O_{72}) (H_2 O)_{14.56}$	6.563	0.185	*Clinoptilolite*	Clinoptilolite	**HEU-c**
201219	$Ag_{3.88} Si_{32.12} Al_{3.88} O_{72} (H_2 O)_{15.72}$	8.091	0.186	*Heulandite*	Clinoptilolite	**HEU-c**

where $x_{i\alpha}$ are values of the αth feature for the ith sample given in Table 1. Summarizing, the EM method makes probabilistic assignments of points to cluster centroids, while K-means makes deterministic assignments. The advantage of the EM algorithm over other clustering methods [31] is its ability of determining the optimal number of clusters without human intervention.

3. Results and discussion

Multiple random ordering of the input data and multiple random seeds to the EM algorithm consistently clustered the 70 HEU type zeolites into three groups: 43 heulandites are clustered together in HEU-h; 15 clinoptilolites are clustered together in HEU-c; the remaining seven heulandites and five clinoptilolites mix together to form a cluster, HEU-m. The clustering results are shown in Fig. 1, and are also included in Table 1. The ratio of the between-cluster point scatter B over the within-cluster point scatter W is $B/W = 7.1$, indicating a good split into distinctive clusters. This point scatter ratio is quite high despite the fact that the 70 zeolites are HEU, meaning that they have equal topology features.

To test the effectiveness of the feature set and facilitate clustering result analysis, we did a backward machine learning classification test. The three clusters obtained from the unsupervised EM clustering method were treated as classes in a supervised classification. This classification was performed using the C4.5 decision tree [32], Random Forest [33], Naïve Bayesian [34], Support Vector Machine [35], and Multilayer Perceptron [31,36] algorithms. Leave-one-out cross validation was used to estimate the classification performance. For each model built, 69 instances were used in the training with the remaining one used for testing; this process was repeated 70 times to cover all the zeolites in the dataset. The first four above-mentioned classification methods predicted correctly to which class each zeolite instance belongs. Thus, the classification accuracy is 100%. The fifth above-mentioned method has strong dependency on the random seed that decides the starting neural network. The leave-one-out experiment was then repeated 10 times with different seeds. As expected, one case yielded 100% perfect prediction and the other nine cases misclassified one or two zeolites belonging to the HEU-c cluster. The C4.5 decision tree classification is illustrated in Fig. 2. It is clear that both features determine the clustering instead of solely the Si/Al feature. Indeed, the ratio Si/Al $\leqslant$ 3.580 is for all HEU-h zeolites; Si/Al > 3.580 is for all HEU-c zeolites; but the mixed group HEU-m covers the full range of Si/Al. Feature tetrahedrality <*T*> is needed to establish further distinction among the three groups. The HEU-m group has low <*T*> relative to the other two traditional groups, which indicates that these instances give rise to Delaunay simplices that are closer to perfect tetrahedra. Worthy of note, the sole zeolite from the American Mineralogist database not included in the ICSD is predicted to belong to the HEU-h subgroup by our model, while it was named "heulandite-B" [37] in the original publication.

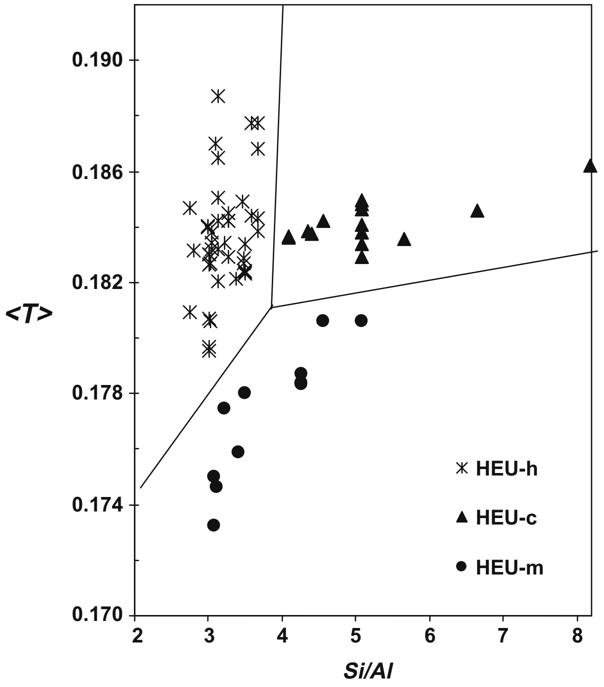

Fig. 1. Distribution of the 70 heulandite type zeolite crystals in the three clusters HEU-h, HEU-c, and HEU-m. Lines are drawn to guide the eye.

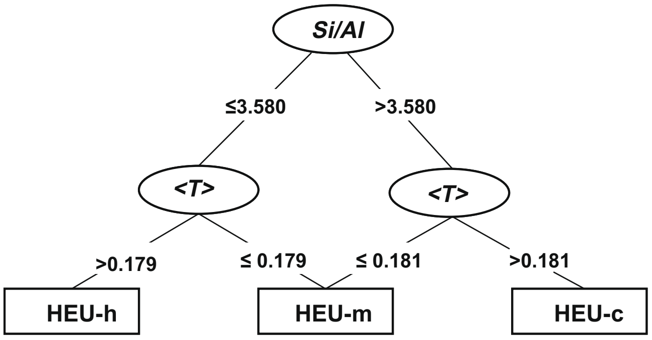

Fig. 2. Decision tree visualization of the three clusters HEU-h, HEU-c and HEU-m of the HEU type zeolites.

Our clustering result is consistent with previous experimental prediction of the existence of intermediate species between heulandite and clinoptilolite [10,15,16]. Eleven out of twelve instances in the HEU-m group are crystals experimentally dehydrated with heat treatment, or combined with cation-exchange. The other instance in the HEU-m group is a crystal treated under high pressure. Although some zeolites are found naturally to be intermediate species [16], the above facts imply that intermediate species could be formed through thermal or high pressure treatment of traditional heulandites or clinoptilolites. However, there is no proof that any thermal or high pressure treatment would necessarily transform native heulandites or clinoptilolites into intermediate species. Six clinoptilolites (ICSD-97837, 97838, 97839, 97840, 97841, 97842) reported in Ref. [38] were treated with cation exchange, dehydration and calcination. Among them, only ICSD-97838 is clustered into HEU-m, while the other five remain in the native clinoptilolite group HEU-c. It is worth noting that the "heulandite B" instance (ICSD-9262) is a heat-collapsed phase of heulandite and is clustered into the intermediate group HEU-m. The high-pressure sample in the intermediate group HEU-m (ICSD-92926) was treated at 3400 MPa [39]. However, ICSD-92925 treated under 1450 MPa, together with ICSD-92924 and 92927 reported in the same publication are clustered into the native heulandite group HEU-h.

4. Conclusions

With up to date crystal structural data from the Inorganic Crystal Structure Database, we developed a machine learning clustering model to resolve the long lasting puzzle about the nomenclature of the heulandite family of zeolites. The model utilizes modern machine learning algorithms and has as foundation characteristic properties of the zeolite framework. It is shown that there are three groups of zeolites with the HEU topology (HEU-h, HEU-m, HEU-c). Our prediction is consistent with previous experimental evidence of the existence of intermediate species between the heulandite and clinoptilolite native groups. Our machine learning results calls for a new look at the classification of the heulandite family and is important for the thorough understanding of the functionality of the framework structure.

Acknowledgments

This work was supported under the National Science Foundation grant CHE-0626111 and the National Institute of Standards and Technology Cooperative Agreement 70NANB5H1110. Additionally, TERAGRID computing time grant PHY050026 is acknowledged.

References

[1] Database of Zeolite Structures from the International Zeolite Association. <http://www.iza-structure.org/databases/>.
[2] G. Gottardi, E. Galli, Natural Zeolites, Springer-Verlag, Berlin, 1985.
[3] E. Passaglia, R.A. Sheppard, in: D.L. Bish, D.W. Ming (Eds.), Natural Zeolites: Occurrence, Properties, Applications, Mineral. Soc. Am., Washington, DC, 2001, pp. 69–116.
[4] R.W. Tschernich, Zeolites of the World, Geoscience Press Inc., Phoenix, 1992.
[5] D.L. Bish, J.M. Boak, in: D.L. Bish, D.W. Ming (Eds.), Natural Zeolites: Occurrence, Properties, Applications, Mineral. Soc. Am., Washington, DC, 2001, pp. 207–216.
[6] J.R. Boles, Am. Mineral. 57 (1972) 1463–1493.
[7] R.L. Ward, H.L. McKague, J. Phys. Chem. 98 (1994) 1232–1237.
[8] N.W. Ockwig, R.T. Cygan, M.A. Hartl, L.L. Daemen, T.M. Nenoff, J. Phys. Chem. C 112 (2008) 13629–13634.
[9] F. Mumpton, Am. Mineral. 45 (1960) 351–369.
[10] A. Alietti, Am. Mineral. 57 (1972) 1448–1462.
[11] G.P. Valueva, Russian Geol. Geophys. 35 (1994) 1–4.
[12] J.M. Boak, P. Cloke, D. Broxton, Geol. Soc. Am. Program with Abstracts 23:A186, 1991.
[13] B. Mason, L.B. Sand, Am. Mineral. 45 (1960) 341–350.
[14] D.S. Coombs et al., Min. Mag. 62 (1998) 533–571.
[15] W.J. Mortier, J.R. Pearce, Am. Mineral. 66 (1981) 309–314.
[16] T. Perraki, A. Orfanoudaki, Appl. Clay Sci. 25 (2004) 9–16.
[17] D.B. Hawkins, Contr. Mineral. Petrol. 45 (1974) 27–36.
[18] T. Armbruster, Stud. Surf. Sci. Catal. 135 (2001) 13–27.
[19] NIST/FIZ Inorganic Crystal Structure Database (ICSD), <http://www.nist.gov/srd/nist84.htm>.
[20] American Mineralogist Crystal Structure Database, <http://rruff.geo.arizona.edu/AMS/amcsd.php>.
[21] S. Yang, M. Lach-hab, I.I. Vaisman, E. Blaisten-Barojas, arXiv_0908.4115v2, <http://arxiv.org/abs/0908.4115v2>.
[22] G. Sastre, J.D. Gale, Micorpor. Mesopor. Mater. 43 (2001) 27–40.
[23] V.A. Blatov, IUCr Comp. Commun. Newsl. 7 (2006) 4–38.
[24] A. Alberti, T.M.P.M. Tschermaks, Min. Petr. Mitt. 19 (1973) 173–184.
[25] D.A. Carr, M. Lach-hab, S. Yang, I.I. Vaisman, E. Blaisten-Barojas, Micropor. Mesopor. Mater. 117 (2009) 339–349.
[26] S. Yang, M. Lach-hab, I.I. Vaisman, E. Blaisten-Barojas, in: R. Stahlbock, S.F. Crone, S. Lessmann (Eds.), Proc. 2008 Inter. Conf. Data Mining, CSREA Press, Las Vegas, 2008, pp. 702–706.
[27] S. Yang, M. Lach-hab, I.I. Vaisman, E. Blaisten-Barojas, LNCS 5545 (2009) 160–168.
[28] S. Yang, M. Lach-hab, I.I. Vaisman, E. Blaisten-Barojas, in: H.R. Arabnia, D. de la Fuente, J.A. Olivas (Eds.), Proc. 2009 Inter. Conf. Artificial Intelligence, CSREA Press, Las Vegas, 2009, pp. 340–344.
[29] A.P. Dempster, N.M. Laird, D.B. Rubin, J. Royal, Stat. Soc. Series B 39 (1977) 1–38.
[30] Weka 3: Data Mining Software in Java. <http://www.cs.waikato.ac.nz/ml/weka>.
[31] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, second ed., Springer, New York, 2009.
[32] R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA, 1993.
[33] L. Breiman, Mach. Learn. 45 (2001) 5–32.
[34] G.H. John, P. Langley, in: Proceedings of the 11th Conference on Uncertainty Artificial Intelligence, San Mateo, 1995, pp. 338–345.
[35] J. Platt, in: B. Schoelkopf, C. Burges, A. Smola (Eds.), Advances in Kernel Methods – Support Vector Learning, 1999, pp. 185–208.
[36] S. Haykin. Neural Networks – A Comprehensive Foundation, second ed., Prentice-Hall, Englewood Cliffs, 1998.
[37] A.O. Larsen, F.S. Nordrum, N. Döbelin, T. Armbruster, O.V. Petersen, M. Erambert, Eur. J. Mineral. 17 (2005) 143–153.
[38] M. Johnson, D. O'Connor, P. Barnes, C.R.A. Catlow, S.L. Owens, G. Sankar, R. Bell, S.J. Teat, R. Stephenson, J. Phys. Chem. B 107 (2003) 942–951.
[39] P. Comodi, G.D. Gatta, P.F. Zanazzi, Eur. J. Mineral. 13 (2001) 497–505.