ISSN 0036-0236, Russian Journal of Inorganic Chemistry, 2016, Vol. 61, No. 5, pp. 604–609. © Pleiades Publishing, Ltd., 2016.
Original Russian Text © N.N. Kiselyova, A.V. Stolyarenko, V.V. Ryazanov, O.V. Sen'ko, A.A. Dokukin, 2016, published in Zhurnal Neorganicheskoi Khimii, 2016, Vol. 61, No. 5, pp. 633–639.



# Prediction of New Halo-Elpasolites

**N. N. Kiselyova[a], A. V. Stolyarenko[a], V. V. Ryazanov[b], O. V. Sen'ko[b], and A. A. Dokukin[b]**

[a] *Baikov Institute of Metallurgy and Materials Science, Russian Academy of Sciences, Leninskii pr. 49, Moscow, Russia*
[b] *Dorodnitsyn Computing Center, Russian Academy of Sciences, Moscow, Russia*
*e-mail: kisi@met.ac.ru*

Received September 10, 2014

**Abstract**—We predicted elpasolites having the composition $A_2BCHal_6$ (A and C stand for different monovalent metals; B stands for trivalent metals; and Hal for F, Cl, or Br) that have not been obtained yet. Prediction was based on the data on properties of elements and simple halides only. The computations were carried out using specialized software for computer-assisted data analysis relying on the precedent-based pattern recognition method.

**DOI:** 10.1134/S0036023616050090

Searching for and studying halide compounds having the composition $A_2^+B^{3+}C^+Hal_6$ (where A and C stand for monovalent metals; B stands for trivalent metals; and Hal for F, Cl, Br, and I) with the elpasolite-like crystal structure (space group $Fm\bar{3}m$) is related to the development of new luminescent [1–4], laser [5], and magnetic [6, 7] materials. According to the database (DB) on properties of inorganic compounds [8], elpasolites in fluoride and chloride systems are the best-studied elpasolites so far. Much fewer studies have focused on bromo- and iodo-elpasolites, although bromo-elpasolites have recently been intensively studied as scintillation [1, 3] and laser materials [5]. This work aimed at designing new compounds with the elpasolite structure in A–B–C–Hal (Hal = F, Cl, Br) systems using theoretical methods.

## THEORETICAL ANALYSIS

Since the unit cell of elpasolite can be regarded as a distorted unit cell of perovskite with doubled parameters [9], attempts have been made [9–14] to use the well-known Goldschmidt's tolerance factor and its modification that includes functions of ionic radii of the components, to predict new compounds with this crystal structure. Our attempt to use the modification [10–14] of tolerance factor $t = \sqrt{2}(r_A + r_{Hal})/(r_B + r_C + 2r_{Hal})$, where $r$ are the Shannon ionic radii, demonstrated that there is significant overlapping between the stability regions of cubic elpasolites and the stability regions of compounds with crystal structures of other types ($Cs_2NaCrF_6$, cryolite, etc.). The use of other criteria [9] based on size factors did not make it possible to completely demarcate the existence regions of $A_2^+B^{3+}C^+Hal_6$ compounds with different types of crystal structure. Liu et al. [15] arrived at similar conclusions. In addition to using the size factors, they have suggested using the electronegativities of chemical elements. It should be mentioned that the authors of studies [16–18] indicated that size factors are not the only factors that need to be taken into account when predicting halide compounds. However, although ion size and/or electronegativity constitute an important, but yet incomplete, set of component properties determining the crystal structure type for $A_2^+B^{3+}C^+Hal_6$ compounds. We proposed [19] a method for searching for regularities in data relying on precedent-based pattern recognition algorithms [20–22], to allow one to find criteria including a wide range of component properties. This enables demarcating the compounds with different types of crystal structures more accurately.

## COMPUTATIONAL METHODS

A specialized information analysis system (IAS) designed by us was used for computations [22]. This system integrates the set of databases on the properties of inorganic substances and materials with the data analysis subsystem, which includes pattern recognition software. The computational procedure can be conventionally divided into several stages.

**Collection of data about compounds for computer-assisted analysis.** At this stage, we searched for data on $A_2^+B^{3+}C^+Hal_6$, compounds as well as $AHal–BHal_3–CHal$ systems where a compound with this composition is not formed under standard ambient conditions (room temperature and atmospheric pressure), over

**Table 1.** Predicted crystal structure types for $A_2BCF_6$ compounds

| C | Li | | | | Na | | | K | | | Rb | | | Cs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A / B | Na | K | Rb | Cs | K | Rb | Cs | Na | Rb | Cs | Na | K | Cs | Li | Na | K | Rb |
| Al | #3 | #2 | #2 | #3 | #1 | #1 | #2 | #1 | #1 | #1 | 1 | 1 | #1 | 4 | | #4 | |
| Sc | 4 | | | #4 | #1 | #1 | #1 | #4 | #1 | #1 | 4 | 1 | 1 | #4 | 4 | 1 | 1 |
| Ti | | #1 | #2 | | #1 | #1 | 2 | 4 | #1 | #1 | 4 | 1 | 1 | 4 | 4 | | 1 |
| V | 1 | #1 | #2 | #3 | #1 | #1 | | 4 | #1 | #1 | 4 | 1 | 1 | 4 | 4 | | 1 |
| Cr | 1 | #1 | #2 | #3 | #1 | #1 | #2 | 4 | #1 | #1 | 4 | 1 | #1 | 4 | 4 | | 1 |
| Mn | | #1 | 2 | 3 | #3 | #3 | #3 | #4 | #3 | #3 | 4 | | | 4 | 4 | | |
| Fe | | 1 | #2 | 3 | #1 | #1 | #2 | 4 | #1 | #1 | 4 | 1 | 1 | 4 | 4 | | 1 |
| Co | | #1 | #2 | 3 | #1 | #1 | #2 | 4 | #1 | #1 | 4 | 1 | #1 | 4 | 4 | | 1 |
| Ga | | #1 | #2 | #3 | #1 | #1 | 2 | | #1 | #1 | | 1 | 1 | 4 | 4 | 1 | 1 |
| Y | 4 | | | #4 | #1 | #1 | #1 | #4 | #3 | #1 | 4 | | #1 | #4 | 4 | 1 | |
| Mo | | 1 | | 3 | #1 | #1 | 1 | 4 | #1 | #1 | 4 | | 1 | 4 | 4 | | 1 |
| In | | #1 | 1 | 3 | #1 | #1 | #1 | | #1 | #1 | | 1 | 1 | 4 | | 1 | #1 |
| Sb | | | | 3 | 1 | 1 | 1 | | 1 | 1 | | 1 | 1 | 4 | 4 | 1 | 1 |
| La | #4 | #4 | #4 | #4 | #4 | #4 | #4 | #4 | #4 | #1 | #4 | #4 | #4 | #4 | #4 | | #4 |
| Ce | #4 | 4 | | | | 1 | #1 | 4 | #1 | #1 | 4 | #1 | #1 | 4 | 4 | #1 | #1 |
| Pr | 4 | 4 | | | 1 | 1 | 1 | 4 | 1 | #1 | 4 | 1 | #1 | 4 | 4 | 1 | 1 |
| Nd | 4 | 4 | | | 1 | 1 | 1 | 4 | 1 | #1 | 4 | 1 | #1 | 4 | 4 | 1 | 1 |
| Pm | 4 | | | | | 1 | 1 | 4 | | 1 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Sm | 4 | 4 | | | 1 | #1 | #1 | 4 | 1 | #1 | 4 | 1 | #1 | 4 | 4 | 1 | 1 |
| Eu | 4 | | | | 1 | #1 | #1 | 4 | | #1 | 4 | | #1 | 4 | 4 | 1 | 1 |
| Gd | 4 | | | | | #1 | #1 | 4 | | #1 | 4 | 1 | #1 | 4 | 4 | 1 | 1 |
| Tb | 4 | | | | | #1 | #1 | 4 | #3 | #1 | 4 | 1 | #1 | 4 | 4 | 1 | 1 |
| Dy | 4 | | | | #3 | #1 | #3 | 4 | #3 | #1 | 4 | | #1 | 4 | 4 | 1 | |
| Ho | 4 | | | | | #1 | #1 | 4 | #3 | #1 | 4 | | #1 | 4 | 4 | 1 | |
| Er | 4 | | | | | #1 | #1 | 4 | | #1 | 4 | | #1 | 4 | 4 | 1 | |
| Tm | 4 | 1 | 1 | 1 | #1 | #1 | #1 | 4 | #1 | #1 | 4 | 1 | #1 | 4 | 4 | 1 | 1 |
| Yb | 4 | 1 | 1 | 1 | 1 | #1 | #1 | 4 | 1 | #1 | 4 | 1 | #1 | 4 | 4 | 1 | 1 |
| Lu | 4 | 1 | | | 1 | #1 | #1 | 4 | | #1 | 4 | | #1 | 4 | 4 | 1 | 1 |
| Bi | | | 1 | 1 | 1 | #1 | #1 | | #3 | #1 | | | #1 | | | | |
| U | 4 | 4 | | | | | 1 | 4 | 1 | 1 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Pu | #4 | 4 | | | | 1 | 1 | 4 | 1 | 1 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |

the database compiling the properties of inorganic compounds [8] integrated into the IAS. Prior to being included into the sample for computations, the retrieved data were subjected to a thorough expert assessment aimed at analyzing inconsistent data. Taking into account the polymorphism of the compounds under study, comparative analysis of several thousand full-text articles, monographs, and reference books included into the database played a great role at this labor-intensive stage that cannot be completely formalized.

**Selection of properties of the components** (chemical elements and/or simple halides). The original sets of properties were formed on the basis of physicochemical views about the nature of the classes of compounds under study. The data about properties of the elements were taken from the database compiled by us (http://phases.imet-db.ru/elements), while the information about halides was taken from the database on the properties of simple halides integrated into the IAS. We used the specialized IAS software to search for the parameters of components for inclusion into the desired criterion in order to select most important

**Table 2.** Predicted crystal structure types for $A_2BCCl_6$ compounds

| C | Li | | | | | Na | | | | | K | | | | Rb | | | | Cs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A / B | Na | K | Rb | Cs | Tl | Li | K | Rb | Cs | Tl | Li | Na | Rb | Cs | Li | Na | K | Cs | Li | Na | K | Rb |
| Al | 4 | #4 | | 4 | 1 | 4 | #4 | | #4 | 1 | #4 | #4 | | 4 | 4 | 4 | | 4 | 4 | #4 | | 4 |
| Sc | 4 | #3 | #1 | #2 | #1 | 4 | | #1 | #1 | #3 | 4 | 4 | 1 | #1 | 4 | 4 | | #4 | 4 | 4 | | #4 |
| Ti | 4 | 3 | | | | 4 | 1 | 1 | #1 | 1 | 4 | 4 | | 1 | 4 | 4 | | 1 | 4 | 4 | | |
| V | 4 | | #3 | #3 | | 4 | #1 | #1 | 1 | 1 | 4 | 4 | | 1 | 4 | 4 | | 1 | 4 | 4 | | |
| Cr | 4 | 3 | | #3 | 1 | 4 | #1 | #1 | 1 | #1 | 4 | 4 | #3 | #1 | 4 | #4 | | | 4 | 4 | | |
| Fe | 4 | #4 | 3 | | | 4 | #4 | 1 | #1 | 1 | #4 | #4 | | | 4 | 4 | | #4 | 4 | 4 | | #4 |
| Y | 4 | | #1 | #1 | 1 | 4 | #4 | #3 | #1 | | 4 | #4 | 1 | #1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| In | 4 | | #1 | #3 | | 4 | 1 | #1 | #3 | | 4 | 4 | | #3 | 4 | 4 | | | 4 | 4 | | |
| La | #4 | #4 | | #1 | 1 | #4 | 4 | | #1 | | #4 | 4 | 1 | #1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Ce | 4 | #4 | | #1 | | 4 | 4 | | #1 | | #4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Pr | #4 | #4 | #4 | #1 | 3 | #4 | #4 | #4 | #1 | | #4 | #4 | 1 | #1 | #4 | #4 | 1 | 1 | 4 | #4 | 1 | 1 |
| Nd | 4 | 4 | | #1 | 3 | 4 | #4 | | #1 | | 4 | #4 | 1 | #1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Pm | 4 | 4 | | 1 | 3 | 4 | 4 | | 1 | | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Sm | 4 | | #3 | #1 | 3 | 4 | #4 | | #1 | | 4 | #4 | 1 | #1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Eu | 4 | | #3 | #1 | 3 | 4 | 4 | #3 | #1 | | 4 | 4 | 1 | #1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Gd | 4 | | #3 | #1 | 3 | 4 | 4 | #3 | #1 | | 4 | 4 | 1 | #1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Tb | 4 | | #1 | #1 | 3 | 4 | 4 | 3 | #1 | | 4 | 4 | 1 | #1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Dy | 4 | | #1 | #1 | 3 | 4 | #4 | #3 | #1 | | 4 | #4 | 1 | #1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Ho | 4 | | #1 | #1 | 3 | 4 | 4 | #3 | #1 | | 4 | 4 | 1 | #1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Er | 4 | | #1 | #1 | 3 | 4 | 4 | #3 | #1 | | 4 | 4 | 1 | #1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Tm | 4 | #3 | #1 | #1 | #3 | 4 | | #3 | #1 | #3 | 4 | 4 | 1 | #1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Yb | 4 | | #1 | #1 | #3 | 4 | | #3 | #1 | | 4 | 4 | 1 | #1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Lu | 4 | | #1 | #3 | #3 | 4 | | #3 | #1 | | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 | 1 |
| Tl | 4 | | 1 | | | 4 | 1 | 1 | #1 | | 4 | | 1 | 1 | 4 | 4 | 1 | 1 | 4 | | 1 | 1 |
| U | #4 | #4 | 1 | #1 | 3 | #4 | | #3 | #1 | 3 | #4 | 4 | | 1 | 4 | 4 | | 1 | 4 | 4 | | 1 |
| Pu | 4 | 4 | | 1 | 3 | 4 | 4 | 3 | #1 | 3 | 4 | 4 | 1 | 1 | 4 | 4 | | 1 | 4 | 4 | | 1 |

properties for classifying properties. Algebraic functions of the initial properties of components can be automatically generated at this stage and the most classifying ones can be selected, which substantially facilitates the formation of criteria.

The first two stages yield a sample for computer-assisted analysis (training sample) in the form of a matrix where each row contains a set of values of properties of the components of a compound with specific composition selected at stage 1, with mentioning one of the a priori known classes (e.g., the crystal structure type) that this compound belongs to.

**Selection of algorithms for pattern recognition.** The resulting sample is subjected to computer-assisted analysis using the software complex for precedent-based pattern recognition integrated into the IAS. This analysis aims at searching for the criteria that allow one to select both the compounds with the information about them included into the training sample and the combinations of components that have not been investigated yet into the pre-determined classes. In this study, these classes were $AHal–BHal_3–CHal$ chemical systems where compounds with the composition $A_2^+B^{3+}C^+Hal_6$ were either formed or absent, as well as the compounds of this composition with various types of crystal structures under standard ambient conditions. Fifteen pattern recognition programs that are currently integrated into the IAS formed 15 criteria. Their prediction accuracy was assessed using the cross-validation procedure that is widely used in applied computer science; the details of this procedure were given in [20, 23]. The most accurate pattern recognition algorithms and the most informative subsets of component properties were selected at this stage.

**Formation of criteria.** Due to the specific character of the algorithm, the selected programs for pattern

recognition complement each other, providing accurate results for different sets of compounds. The collective decision-making strategy using specialized software [20, 22] integrated into the IAS is used to form the generalized criterion employing the advantages of different algorithms. However, these programs form criteria with different prediction accuracy, which were assessed using examination recognition for 100 compounds the data on which were randomly selected from the training samples and were not used during computer training (at the final stage of prediction, these control examples were returned to the analysis set).

The last two stages yield the generalized criterion of formation of compounds belonging to different classes, which was obtained using a set of the most accurate pattern recognition algorithms and the collective decision-making algorithm and includes the most informative set of component properties.

**Prediction.** The resulting criterion includes only component properties and can be written in various forms (Boolean expression, set of equations and inequalities, trained neural or growing pyramidal network, etc.). Taking into account the fact that the resulting criterion often has a complex form, a specialized subsystem was integrated into the IAS to automatize the prediction procedure; this subsystem substitutes the values of component properties into the resulting criterion and predicts whether a compound belongs to a certain preset class or not.

The problem of predicting new halo-elpasolites included solving three intermediate problems. Formation of compounds with composition $A_2^+B^{3+}C^+Hal_6$ was predicted in the first of them (problem 1). The next problem included searching for criteria and predicting the formation of compounds with given composition and the most common types of crystal structures (elpasolite or $Cs_2NaCrF_6$ (space group $R\bar{3}m$)). The latter problem was divided into two smaller ones. When solving the first of them, the multiclass prediction of belonging to four classes (elpasolites, compounds with the $Cs_2NaCrF_6$ structure, compounds with the structure different from those shown above, and the systems containing no compounds with composition $A_2^+B^{3+}C^+Hal_6$ (problem 2)) was performed. Next, halide systems were consecutively divided into three classes: the target class 1—elpasolites; class 2—compounds with non-elpasolite structure; and class 3—the $AHal–BHal_3–CHal$ systems containing no compounds with composition $A_2^+B^{3+}C^+Hal_6$ (problem 3). All the results were predicted for standard ambient conditions. The final decision regarding the class that a compound being predicted belongs to, was made by comparing the predictions obtained when solving all three problems. If the results were inconsistent, the prediction was regarded to be uncertain and the prediction table cell was left empty.

This study differed from the previous ones in that the composition of the predicted compounds was more complex. Earlier, even when predicting quaternary compounds [19], the fourth component was fixed. In this study, we varied all four elements that were components of the compounds. The more complex composition made it necessary not only to design a database on the properties of quaternary compounds, which currently holds information about more than 31 thousand compounds formed by four elements [8], but also to modify the IAS programs.

## COMPUTATIONS

After the expert assessment, the sample for computer-assisted analysis included information about 289 $A_2^+B^{3+}C^+Hal_6$ ($A \neq C$) compounds having the elpasolite structure; 20 compounds with $Cs_2NaCrF_6$-type crystal structure; 57 compounds with crystal structures other than the ones given above under standard ambient conditions; and 81 $AHal–BHal_3–CHal$ systems where $A_2^+B^{3+}C^+Hal_6$ compounds are not formed.

The following properties of chemical elements A, B, C, and Hal were included in the initial set of component parameters (see values of the properties in the database http://phases.imet-db.ru/elements): pseudo-potential radius (Zunger); ionic radius (Shannon); distance from the core and outer-shell electrons (Schubert); energies of ionization of the first, second, and third electrons ($E5$, $E6$, and $E7$); the Mendeleev–Pettifor numbers ($M1–M11$); the group number in the periodic table; the quantum number; the number of outer-shell electrons; electronegativity (Pauling), Miedema's chemical potential; the melting and boiling points, thermal conductivity, molar heat capacity, enthalpy of atomization; standard entropy, etc., as well as thermal parameters of simple halides of mono- and trivalent elements (melting (decomposition) point; standard entropy ($S$), standard heat of formation, isobaric heat capacity and isobaric potential of formation), that is, 134 values altogether for each A–B–C–Hal system.

## RESULTS AND DISCUSSION

The computations showed that $E6(A)/E7(B)$ and $S(BHal_3)/S(AHal)$ are the most important algebraic functions for classifying the $AHal–BHal_3–CHal$ systems where $A_2BCHal_6$ compounds are either formed or not. However, an analysis of the position of the points corresponding to these two classes of chemical systems on the projections whose coordinates are the selected functions, demonstrates that complete demarcation of the classes was not achieved. Further simulation experiments also showed that the mean accuracy of test prediction using different pattern recognition programs in the cross-validation mode using

**Table 3.** Predicted types of crystal structures for compounds with the $A_2BCBr_6$ composition

| C | Li | | | | Na | | | | K | | Rb | | Cs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A / B | Na | K | Rb | Cs | Li | K | Rb | Cs | Rb | Cs | K | Cs | Li | Na | K | Rb |
| Sc | | 3 | | #3 | 4 | 1 | 1 | #1 | 1 | #1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Y | 4 | 1 | | #1 | 4 | 1 | 1 | #1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| La | 4 | | 1 | #1 | 4 | 3 | | #3 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Ce | 4 | | 1 | #1 | 4 | 3 | 1 | #1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Pr | 4 | 4 | 1 | #1 | 4 | 3 | 1 | #1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Nd | 4 | 4 | 1 | #1 | 4 | 3 | 1 | #1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Pm | 4 | | 1 | 1 | 4 | | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Sm | 4 | | 1 | #1 | 4 | 3 | 1 | #1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Eu | 4 | | 1 | #1 | 4 | | 1 | #1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Gd | 4 | | | #1 | 4 | | 1 | #1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Tb | 4 | | | #1 | 4 | 1 | 1 | #1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Dy | 4 | | | #1 | 4 | 1 | 1 | #1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Ho | 4 | | | #1 | 4 | 1 | 1 | #1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Er | 4 | | | #1 | 4 | 1 | | #1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Tm | 4 | | | #2 | 4 | 1 | 1 | #1 | 1 | #1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Yb | 4 | | | #2 | 4 | | | #1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |
| Lu | 4 | | | #2 | 4 | 1 | | #1 | 1 | 1 | 1 | 1 | 4 | 4 | 1 | 1 |

the criteria that included only two of the aforementioned functions was 11% lower than that when using the criteria including only the initial properties of components. In the latter, the criterion obtained using the precedent-based pattern recognition software (algorithms for constructing logical regularities, training neural networks, the *k*-nearest neighbors algorithm, and the support vector machine) that ensures the best accuracy of prediction in the cross-validation mode and the collective decision-making software based on the algorithm of generalized polynomial corrector, provided the best estimate for prediction accuracy, namely 95%.

The criteria that include only functions obtained by computation $M7(A) \cdot M11(A)$, $E6(A)/E7(Hal)$, and $E6(A) + E6(B)$, which are most important for classifying $A_2^+B^{3+}C^+Hal_6$ compounds with respect to the type of crystal structure under standard ambient conditions, failed to provide better results in the cross-validation mode of examination recognition than the results obtained using the initial parameters of the components. When solving problem 2 of multiclass prediction, the set of algorithms including the binary decision tree algorithm, the *k*-nearest neighbors algorithm, the support vector machine, training a multilayer perceptron and neural network, and the algorithm of the convex stabilizer for collective decision-making, ensured the best accuracy of examination prediction: 89% when only component properties were included in the desired criterion. When forming the criterion that allows one to demarcate elpasolites from compounds with differing crystal structures and from systems where no $A_2BCHal_6$ compounds are formed (problem 3), the best accuracy (80%) was provided by the set of algorithms that included the algorithms of constructing logical regularities, training neural networks, the *k*-nearest neighbors algorithm, the support vector machine, and the Bayesian model of collective decision-making; only component properties were included in the criterion.

The resulting criteria were used to predict the possibility of formation and crystal structure type under standard ambient conditions for compounds that have not been obtained yet. Some results of comparing the predictions found by solving all three classification problems are summarized in Tables 1–3. The following notations are used: 1 denotes prediction of compounds with the elpasolite crystal structure; 2 denotes prediction of compounds with the $Cs_2NaCrF_6$-type structure; 3 denotes prediction of compounds having crystal structure other than the aforementioned ones; and 4 denotes prediction of the absence of an $A_2^+B^{3+}C^+Hal_6$ compound in the $AHal–BHal_3–CHal$ system. Here and below, the # symbol is used to denote previously studied compounds; the information about them was used for computer training.

The analysis of the resulting prediction tables shows that the greatest number of new elpasolites is predicted in chloride systems. Computer simulations on searching for component properties that are most important for classifying the compounds with respect to the type of their crystal structure or their ability to be formed under certain conditions give grounds for claiming that the use of two or three parameters of elements or simple compounds, or several algebraic

functions of these properties does not allow one to demarcate compounds belonging to different classes with sufficient accuracy. Only a broad range of component properties can ensure a fair reliability of prediction, since it makes it possible to take into account the features of compounds with different compositions more thoroughly. Searching for these multidimensional criteria can be automatized using data analysis software relying upon precedent-based pattern recognition.

The main objection from the developers of the conventional criteria including a limited number of component parameters is associated with the complexity of the regularities obtained using computer-assisted analysis methods. Indeed, it is difficult for a chemist to understand a set of nodes in a trained growing pyramidal network or a set of equations. One of the ways to enhance the interpretability of the resulting multi-dimensional criteria is to visualize the projections with their coordinates being the key parameters of the components and the algebraic functions of these properties. This subsystem was integrated into the IAS designed by us and allows obtaining colored projections of points that correspond to compounds belonging to different classes in any preset coordinate system. It should be mentioned that the prediction process in IAS is fully automated. The user only chooses sets of elements in a special menu, while the prediction subsystem forms a sample for further prediction, substitutes component parameters into the resulting criteria, and outputs the final prediction table. The long-term experience with computations gives grounds for claiming that not only does the designed IAS provide an access to a big data on inorganic compounds for chemists, but it also provides a tool for data analysis to search for the empirical criteria of formation of different types of inorganic compounds.

An important advantage of the designed IAS is that it can be quickly adapted to new experimental data that are inconsistent with the existing classifying regularity. While writing this article, we obtained data on two compounds having the elpasolite structure that had been improperly predicted by the IAS. As opposed to the labor-intensive conventional methods for processing criteria, retraining of the computer system only involved adding several lines (the data on two elpasolites and new compounds the information about which has been recently published) to the analysis sample and took less than an hour. Hence, the data analysis system is a convenient tool for adjusting for new experimental data and analyzing it in order to search for more accurate criteria and make predictions.

## ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research (project nos. 12-07-00142, 15-07-00980, and 14-07-00819).

## REFERENCES

1. S. Kim, G. Rooh, H. J. Kim, et al., J. Cryst. Growth **317** (1), 84 (2011).
2. P. A. Tanner, C.-K. Duan, G. Jia, and B.-M. Cheng, J. Solid State Chem. **188**, 105 (2012).
3. E. C. Samulon, G. Gundiah, M. Gascon, et al., J. Lumin. **153**, 64 (2014).
4. J. Glodo, R. Hawrami, and K. S. Shah, J. Crystal Growth **379**, 73 (2013).
5. P. J. Deren, W. Strek, E. Zych, and J. Drozdzynski, Chem. Phys. Lett. **332**, 308 (2000).
6. L. P. Sosman, F. Yokaichiya, and H. N. Bordallo, J. Magn. Magn. Mater. **321**, 2210 (2009).
7. A. V. Malakhovskii, I. S. Edelman, A. L. Sukhachev, et al., Opt. Mater. **32**, 243 (2009).
8. N. N. Kiselyova, V. A. Dudarev, and V. S. Zemskov, Russ. Chem. Rev. **79**, 145 (2010).
9. K. S. Aleksandrov and B. V. Beznosikov, *Perovskites Now and in Future* (Izd–vo SO RAN, Novosibirsk, 2004) [in Russian].
10. J. Arndt, D. Babel, R. Haegele, and N. Rombach, Z. Anorg. Allg. Chem. **418** (3), 193 (1975).
11. I. N. Flerov, M. V. Gorev, J. Grannec, and A. Tressaud, J. Fluorine Chem. **116**, 9 (2002).
12. G. Friedrich, H. Fink, and H. J. Seifert, Z. Anorg. Allg. Chem. **548** (5), 141 (1987).
13. H. Faget-Guengard, J. M. Bobe, J. Senegas, et al., J. Alloys Compd. **219** (1–2), 73 (1995).
14. J. Graulich, S. Drucke, and D. Babel, Z. Anorg. Allg. Chem. **624**, 1460 (1998).
15. L. Liu, W. Lu, and N. Chen, J. Phys. Chem. Solids **65**, 855 (2004).
16. P. P. Fedorov and P. I. Fedorov, Zh. Neorg. Khim. **18**, 205 (1973).
17. A. G. Dudareva, A. K. Molodkin, and G. A. Lovetskaya, Zh. Neorg. Khim. **33**, 1613 (1988).
18. V. N. Chikanov, Russ. J. Inorg. Chem. **54**, 1099 (2009).
19. N. N. Kiselyova, Eng. Appl. Artificial Intelligence, No. **3**, 533 (2000).
20. Yu. I. Zhuravlev, V. V. Ryazanov, and O. V. Sen'ko, *"RASPOZNAVANIE:" Mathematical Methods. Program System. Applications* (FAZIS, Moscow, 2006) [in Russian].
21. V. P. Gladun, *Formation of New Knowledge* (SD "Pedagog 6, Sofia) [in Russian].
22. N. N. Kiselyova, A. V. Stolyarenko, V. V. Ryazanov, et al., Patt. Recogn. Image Anal. **21** (1), 88 (2011).
23. N. N. Kiseleva, Russ. J. Inorg. Chem. **59**, 496 (2014).

*Translated by D. Terpilovskaya*