Scientific paper

Crystal Structure Prediction in Orthorhombic ABO_3 Perovskites by Multiple Linear Regression and Artificial Neural Networks

Slobotka Aleksovska*, Sandra Dimitrovska and Igor Kuzmanovski

Institute of Chemistry, Faculty of Natural Sciences and Mathematics, University "Sts. Cyril and Methodius" Skopje, Republic of Macedonia

Corresponding author: E-mail: bote@iunona.pmf.ukim.edu.mk; tel: ++389 02 3117055 ext. 910; fax: ++389 02 3226865

Received: 17-07-2006

Abstract

The unit cell parameters and the fractional atomic coordinates of the orthorhombic perovskites of ABO_3 type are expressed as a function of the effective ionic radii of the constituents using two approaches: multiple linear regression and artificial neural networks. For this purpose, 46 orthorhombic perovskites of $GdFeO_3$ type (spa ce group *Pnma*) with accurately refined structures are included in the analysis: 41 in calibration set, and 5 in test set. The predictive strength of the proposed model is very high. This is shown by the values of the coefficients of correlation $(R_{adj})^2$ which are higher than 0.9 for all dependent variables and by the agreement between the actual and predicted values for the dependent variables, obtained by both methods. This simple mathematical model can be used: to predict the crystal structure of members in this series; as starting model for crystal structure refinement; to test the actual crystallographic data of ABO_3 perovskites.

Keywords: Perovskites, crystal structures, artificial neural networks, multiple linear regression.

1. Introduction

The study of perovskites is of academic and technical interest due to their wide variety of interesting physical properties[1–3], which could be modified by composition-driven structural variances.

The basic perovskite composition is ABO_3 where A is large ion suitable for twelve-coordinated cube-octahedral site and B is a smaller ion suitable for the six coordinated octahedral site, formed by the anions. Firstly, it was thought that perovskites have an ideal cubic structure, but it was later found that they could be divided in several isostructural and isomorphous subgroups with orthorhombic, tetragonal, rhombohedral etc. structures. In each group of compounds ions in different oxidation states ($A^{2+}B^{4+}O_3$, $A^{3+}B^{3+}O_3$, $AB'_xB''_{1-x}O_3$, $A'_xA''_{1-x}BO_3$ etc.) can exist.[1,2]

Many of ABO_3 compounds adopt the $GdFeO_3$ structure[1,2] even the mineral perovskite – $CaTiO_3$[1–4], which was formerly assumed as cubic. The $GdFeO_3$ structure is derived from the ideal cubic structure (space group $Pm\bar{3}m$) by tilting of the FeO_6 octahedra about the [110] and [001] directions of the cubic subcell which results in a reduction of symmetry to orthorhombic, *Pbnm*, space group.[2,5] This tilting maintains the B–O distances, while changing the A–O distances. Thus, generally, four A–O distances are elongated and eight are shortened. Therefore, the coordination number of A cations is reduced to 8, instead of 12, as in the ideal perovskite structure. Thus, in the most cases, in orthorhombic perovskites the coordination number of A cation is 8, although there are some cases when it is 9 or 10.

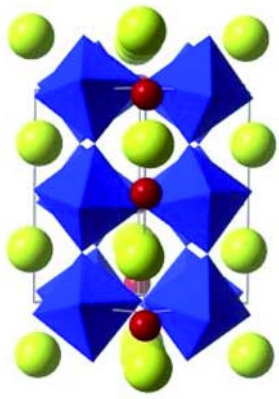

Fig. 1. Tilting of BO_6 octahedra in $GdFeO_3$ type perovskites.

It is well known that physical and chemical properties of the compounds (the perovskites as well) depend on the structure, which is, on the other hand, closely related to the radii of the constituents and some other relevant physical variables. Therefore, numerous attempts have been made to correlate the structural parameters in perovskites with physical variables of the constituent elements.[6–23]

In our previous work, the unit cell parameters and crystal structures in some isomorphous series of compounds were predicted using multiple linear regression (MLR).[16–20] Recently, we have reported on the prediction of unit cell parameters in orthorhombic ABO_3 perovskites[21] and on cubic[22] and monoclinic[23] $A_2BB´O_6$ perovskites using MLR and artificial neural networks (ANN).

Continuing our research on structural correlations in perovskites, in this work the complete crystal structure in isomorphous group of perovskites is predicted. For this purpose, the orthorhombic perovskites, which crystallize in $GdFeO_3$ structural type, were chosen. Thus, the unit cell parameters and atomic coordinates of this isomorphous group of compounds were expressed as a function of the effective ionic radii of the cations using MLR and ANN.

2. Data Analysis

2. 1. Selection of the Samples and Independent Variables

The lattice parameters and the atomic coordinates of isomorphous ABO_3 ($A^{2+}B^{4+}O_3$, $A^{3+}B^{3+}O_3$ and $A^{1+}B^{5+}O_3$) orthorhombic perovskites were used as dependent variables in the analysis. The unit cell edges, as well as, the atomic coordinates (refined at room temperature) for 46 orthorhombic perovskites, which crystallize in $GdFeO_3$ structural type, were taken from the literature.[4, 24–48] Five of these compounds were chosen for prediction of the crystal structures (test set). The rest of the compounds were used for construction of the calibration set.

Since the crystal structures of the compounds were refined by various groups of authors, there are differences in the choice of the cell, orientation of axis and atom designation. The majority of structures, probably of historical reasons, were refined in nonstandard *Pbnm* space group in accordance with the structure of $GdFeO_3$. In order to avoid possible confusion, all data were transformed to match the standard *Pnma* space group ($Z = 4$). In this structure the four B-atoms are located at the center of symmetry (4b) and the four A-atoms are on mirror planes (4c). There are two crystallographically different types of O-atoms: four (O1) lying on mirror plane (4c) and eight (O2) in general positions. Thus, ten dependent variables (seven for atomic coordinates and three for lattice parameters) must be included in the analysis.

Only the accurately refined structures by diffraction methods were taken into consideration (small standard deviation for lattice parameters and fraction atomic coordinates). However, although in the literature there are reported refined structures for series of rare earth manganites[49], they were not included in the analysis because of the significant distortion of the MnO_6 octahedra due to strong Jahn-Teller effect of the Mn^{3+} ion at room temperature. Also, the rare earth orthoscandates[50] were not included in the analysis, as well, due to highly distorted coordination polyhedron of Ln (Ln = rare earth) and Sc.

The independent variables are the effective ionic radii of the constituents introduced by Shannon.[51] These values refer to corresponding oxidation state and coordination number of the ions. As mentioned previously, the cations in A-position were treated as eight-coordinated and the cations in B-position as six-coordinated and in high spin state for *d*-ions.

The collected data (dependent and independent variables) for the calibration set are given in Table 1.*

2. 2. Modeling

Two methods with powerful predictive abilities were used in this work: multiple linear regression (MLR) and artificial neural networks (ANN).

Multiple linear regression. The MLR was performed using the program package *Statgraphic Plus Ver. 3.0*.[52] Each dependent variable *d* (numerical value of a unit cell parameter or fractional atomic coordinate) was presented as a function of the type:

$$d = e + f \cdot r_A + g \cdot r_B \quad (1)$$

where r_A and r_B are the effective ionic radii of A and B cations in Å, *f* and *g* are regression coefficients and *e* is the intercept.

Artificial neural networks. In the last decade the ANN – computation systems (implemented most often in terms of software) designed on the basis of the biological neurons capable for parallel signal processing – have been proven as valuable and efficient method for handling of noisy, nonlinear and incomplete multivariate data[53] in different aspects of chemistry[54], and today are valuable tool for chemometricians. ANN and their application in chemistry in details are described in the literature.[54,55]

In this work three layered cascade-forward (Fig. 2.) ANNs were used with one input, one output and one hidden layer. This type of ANNs was chosen instead of feed-

* It should be mentioned that there is a misprint in one of the fractional coordinates of $YbVO_3$.[44] Thus, the value for O2*z* coordinate (O2*y* in *Pnma* setting) in the paper is 0.58, whereas it should be 0.058.

Table 1. Input data (independent and dependent variables) in the analysis[1].

Formula	r_A(Å)	r_B(Å)	a(Å)	b(Å)	c(Å)	Ax	Az	O1x	O1z	O2x	O2y	O2z	Ref.
$GdAlO_3$	1.053	0.535	5.3049	7.4485	5.2537	0.53778	0.50808	0.4862	0.0722	0.2851	0.03831	0.7153	24
$HoAlO_3$	1.015	0.535	5.3234	7.3764	5.182	0.55185	0.51157	0.4789	0.0829	0.2936	0.04352	0.7056	24
$ScAlO_3$	0.87	0.535	5.2322	7.2042	4.9371	0.5695	0.5204	0.4561	0.1211	0.3066	0.0609	0.6897	25
$YAlO_3$	1.019	0.535	5.331	7.37	5.179	0.55308	0.51197	0.4775	0.0845	0.295	0.0442	0.7045	24
$BaCeO_3$	1.42	0.87	6.235	8.781	6.212	0.523	0.501	0.487	0.071	0.278	0.041	0.726	26
$SrCeO_3$	1.26	0.87	6.145	8.575	6.000	0.5445	0.5117	0.4578	0.1044	0.2997	0.055	0.6998	27
$DyFeO_3$	1.027	0.645	5.598	7.623	5.302	0.56648	0.51707	0.4624	0.106	0.3049	0.0549	0.693	28
$ErFeO_3$	1.004	0.645	5.582	7.591	5.263	0.56913	0.51845	0.4594	0.1137	0.3059	0.0573	0.691	28
$EuFeO_3$	1.066	0.645	5.606	7.685	5.372	0.56012	0.51445	0.468	0.0978	0.3006	0.0506	0.6977	28
$GdFeO_3$	1.053	0.645	5.611	7.669	5.349	0.56284	0.51556	0.4672	0.1005	0.3016	0.0506	0.6957	28
$HoFeO_3$	1.015	0.645	5.591	7.602	5.278	0.56801	0.51781	0.4605	0.1091	0.3052	0.056	0.6924	28
$LuFeO_3$	0.977	0.645	5.547	7.565	5.213	0.57149	0.51997	0.4539	0.1199	0.3071	0.0621	0.6893	28
$NdFeO_3$	1.109	0.645	5.584	7.768	5.453	0.54881	0.51069	0.4759	0.0876	0.2936	0.0462	0.7052	28
$PrFeO_3$	1.126	0.645	5.578	7.786	5.482	0.54367	0.50903	0.4788	0.0817	0.2919	0.0437	0.7075	28
$TbFeO_3$	1.04	0.645	5.602	7.635	5.326	0.56408	0.51597	0.464	0.1035	0.3026	0.0538	0.695	28
$TmFeO_3$	0.994	0.645	5.576	7.584	5.251	0.56913	0.51896	0.4559	0.1148	0.3057	0.0587	0.6907	28
$YFeO_3$	1.019	0.645	5.5877	7.5951	5.2743	0.56852	0.51787	0.4604	0.1103	0.3045	0.0567	0.6924	29
$YbFeO_3$	0.985	0.645	5.557	7.57	5.233	0.57076	0.51936	0.4537	0.1169	0.3077	0.0599	0.6886	28
$NdGaO_3$	1.109	0.62	5.4979	7.7078	5.4276	0.54142	0.50908	0.4826	0.08	0.2903	0.0422	0.7107	30
$PrGaO_3$	1.126	0.62	5.4901	7.7275	5.4557	0.53526	0.50743	0.4848	0.076	0.2874	0.0405	0.7132	31
$SrHfO_3$	1.26	0.71	5.7646	8.1344	5.7516	0.516	0.504	0.486	0.063	0.2789	0.0335	0.7189	32
$DyNiO_3$	1.027	0.6	5.5056	7.4455	5.2063	0.5697	0.5178	0.4729	0.0983	0.301	0.0489	0.6964	33
$EuNiO_3$	1.066	0.6	5.45857	7.5371	5.29413	0.5574	0.5128	0.4767	0.089	0.2947	0.0444	0.7058	34
$GdNiO_3$	1.053	0.6	5.48544	7.51116	5.26063	0.56307	0.515	0.4765	0.0885	0.2974	0.0471	0.7038	33
$SmNiO_3$	1.079	0.6	5.43283	7.56483	5.32693	0.5514	0.5101	0.4865	0.0825	0.2932	0.0457	0.7086	34
$BaPrO_3$	1.42	0.85	6.1787	8.7261	6.2137	0.5135	0.5016	0.4933	0.0703	0.2709	0.0379	0.7294	35
$BaPuO_3$	1.42	0.86	6.193	8.744	6.219	0.5134	0.503	0.4884	0.0703	0.2719	0.0368	0.7275	36
$SrRuO_3$	1.26	0.62	5.5302	7.8441	5.5639	0.5201	0.5016	0.5	0.0541	0.2777	0.0288	0.7225	37
$CaSnO_3$	1.12	0.69	5.662	7.8814	5.5142	0.5506	0.5141	0.4644	0.0997	0.2982	0.0517	0.6988	38
$CaTiO_3$	1.12	0.605	5.447	7.654	5.388	0.5341	0.50626	0.4842	0.0704	0.2884	0.0369	0.7109	4
$CdTiO_3$	1.1	0.605	5.4215	7.6176	5.3053	0.53873	0.50847	0.4722	0.0902	0.2969	0.0472	0.7008	39
$LaTiO_3$	1.16	0.67	5.6156	7.9145	5.6336	0.5457	0.5084	0.4913	0.0799	0.2941	0.0417	0.7096	40
$YTiO_3$	1.019	0.67	5.6901	7.613	5.3381	0.57339	0.52105	0.45736	0.1209	0.30942	0.05824	0.69031	41
$DyVO_3$	1.027	0.64	5.598	7.586	5.292	0.567	0.5186	0.4618	0.1074	0.3031	0.0549	0.6923	42
$GdVO_3$	1.053	0.64	5.62	7.643	5.35	0.5635	0.5169	0.4685	0.101	0.2987	0.0509	0.6954	42
$NdVO_3$	1.109	0.64	5.582	7.738	5.451	0.552	0.513	0.479	0.085	0.296	0.048	0.702	43,44
$TbVO_3$	1.04	0.64	5.621	7.605	5.319	0.568	0.518	0.464	0.103	0.301	0.052	0.692	43,44
$TmVO_3$	0.994	0.64	5.582	7.548	5.244	0.572	0.523	0.455	0.121	0.302	0.058	0.69	43,44
$YbVO_3$	0.985	0.64	5.578	7.54	5.23	0.572	0.521	0.454	0.114	0.306	0.058	0.684	43,44
$CaZrO_3$	1.12	0.72	5.7616	8.0171	5.5912	0.5496	0.5121	0.4619	0.1032	0.3007	0.0548	0.6974	45
$SrZrO_3$	1.26	0.72	5.817	8.171	5.796	0.524	0.504	0.487	0.073	0.285	0.035	0.716	46

[1] The unit cell parameters and fractional atomic coordinates match for *Pnma* space group. *y*-coordinates for A and O1 atoms are fixed by symmetry to ¼. Fractional atomic coordinates for B-atom are 0, 0, ½.

forward ANNs (which are the most often used), because the cascade-forward networks are capable of solving the same problem with smaller number of neurons in the hidden layer, due to direct connections between the input and output neurons. However, since the number of weights in cascade-forward ANNs is bigger, their optimization is slower compared to feed-forward networks.

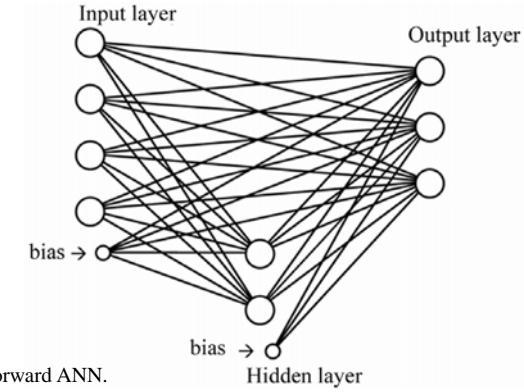

Fig. 2. Graphical presentation of the cascade-forward ANN.

3. Results and Discussion

The predictive strength of the regression equations for each dependent variable is very high. This is shown by the value of the adjusted coefficient of determination $(R_{adj})^2$ which is higher than 0.9 for all dependent variables. The coefficient of determination was calculated using the following formula:

$$(R_{adj.})^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \Big/ [n-(k+1)]}{\left[\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2\right] \Big/ (n-1)} \quad (2)$$

where: $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is sum of squares of error; $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ is sum of squares of regression; n is the number of analyzed compounds, and k is the number of independent variables.

The regression coefficients and $(R_{adj})^2$ are listed in Table 2 and the predicted values for lattice parameters and atomic coordinates for perovskites in calibration set are given in Table 3. As can be seen, there is excellent agreement between the actual and predicted values for the dependent variables of calibration set.

The performance of the ANN architecture is determined by the number of input, output and hidden neurons. However, the number of input, as well as of output neurons, is defined by the data set. Thus, the number of input neurons was 2, while the number of output neurons was 10. The generalization performances of the networks during the training were controlled by early stopping procedure. Optimal network architecture was located by changing the number of hidden neurons from one to ten. The performances of the trained ANNs with different number of hidden neurons were compared using the values of the root mean squared error of prediction (*RMSEP*):

$$RMSEP = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}\left(d_{i,j} - \bar{d}_{i,j}\right)^2 \Big/ (m \cdot n) \right)^{1/2} \quad (3)$$

In this equation $d_{i,j}$ represents the dependent variables for the samples in the test set, $\bar{d}_{i,j}$ represents the predicted values for $d_{i,j}$ obtained by the optimized neural network, m is number of the samples in the test set, while n is

Table 2. The coefficients from the regression analysis for each variable and their standard deviations (σ) and the coefficients of determination $(R_{adj})^2$.

Variable	*e*	*f*	*g*	$(R_{adj})^2$/%
a	3.806	0.066	2.652	98.14
σ	0.044	0.065	0.098	
b	4.492	1.462	2.534	99.19
σ	0.047	0.070	0.106	
c	2.828	1.565	1.358	99.00
σ	0.042	0.061	0.093	
A*x*	0.684	–0.202	0.136	91.70
σ	0.008	0.011	0.017	
A*z*	0.555	–0.072	0.056	93.82
σ	0.002	0.003	0.005	
O1*x*	0.407	0.167	–0.180	93.49
σ	0.005	0.007	0.011	
O1*z*	0.190	–0.233	0.243	96.64
σ	0.005	0.007	0.011	
O2*x*	0.365	–0.116	0.089	92.92
σ	0.004	0.006	0.008	
O2*y*	0.092	–0.110	0.118	96.03
σ	0.002	0.004	0.005	
O2*z*	0.628	0.139	–0.110	93.98
σ	0.004	0.006	0.009	

Table 3. Predicted values for unit cell parameters and fractional atomic coordinates for randomly selected compounds of calibration set obtained by MLR.

Formula	***a*** (Å)	***b*** (Å)	***c*** (Å)	A***x***	A***z***	O1***x***	O1***z***	O2***x***	O2***y***	O2***z***
$GdAlO_3$	5.2939	7.3873	5.2018	0.54473	0.50932	0.48616	0.07491	0.29042	0.03929	0.70927
$BaCeO_3$	6.207	8.773	6.231	0.516	0.502	0.487	0.071	0.278	0.038	0.724
$PrFeO_3$	5.590	7.773	5.465	0.54498	0.51027	0.47847	0.08465	0.29174	0.04420	0.70736
$YFeO_3$	5.5834	7.6164	5.2979	0.56656	0.51793	0.46066	0.10960	0.30414	0.05594	0.69250
$NdGaO_3$	5.5230	7.6846	5.4048	0.54501	0.51009	0.48015	0.08253	0.29149	0.04313	0.70774
$SrHfO_3$	5.7717	8.1334	5.7634	0.5268	0.5043	0.4891	0.0692	0.2820	0.0371	0.7188
$EuNiO_3$	5.46715	7.57102	5.31039	0.5510	0.5120	0.4766	0.0877	0.2947	0.0455	0.7040
$BaPrO_3$	6.1535	8.7221	6.2038	0.5136	0.5008	0.4904	0.0660	0.2759	0.0361	0.7257
$BaPuO_3$	6.180	8.747	6.217	0.5150	0.5013	0.4886	0.0684	0.2768	0.0372	0.7246
$SrRuO_3$	5.5330	7.9053	5.6412	0.5146	0.4993	0.5053	0.0473	0.2740	0.0266	0.7287
$CaSnO_3$	5.7094	7.8781	5.5171	0.5523	0.5132	0.4694	0.0970	0.2964	0.0501	0.7016
$CaTiO_3$	5.484	7.663	5.402	0.54075	0.50846	0.48469	0.07631	0.28888	0.04016	0.71091
$LaTiO_3$	5.6590	7.8858	5.5525	0.5415	0.5092	0.4796	0.0828	0.2900	0.0434	0.7093
$TbVO_3$	5.571	7.634	5.324	0.5616	0.5161	0.4651	0.1035	0.3013	0.0530	0.6960
$SrZrO_3$	5.798	8.159	5.777	0.528	0.505	0.487	0.072	0.283	0.038	0.718

Table 4. Actual and predicted values for lattice parameters and fractional atomic coordinates for the compounds in test set.

Formyula	*a*(Å)	*b*(Å)	*c*(Å)	A*x*	A*z*	O1*x*	O1*z*	O2*x*	O2*y*	O2*z*
$SmFeO_3$										
Act. (I)	5.6001	7.7060	5.3995	0.55728	0.51335	0.47070	0.09530	0.29910	0.04980	0.69930
MLR (II)	5.5873	7.7041	5.3918	0.55446	0.51364	0.47064	0.09561	0.29719	0.04936	0.70083
Δ(I–II)	0.0128	0.0019	0.0077	0.0028	–0.0003	0.0001	–0.0003	0.0019	0.0004	–0.0015
ANN(III)	5.5690	7.6944	5.3784	0.5547	0.5137	0.4693	0.0970	0.2980	0.0500	0.7002
Δ(I–III)	0.0311	0.0116	0.0211	0.0026	–0.0004	0.0014	–0.0017	0.0011	–0.0002	–0.0009
$ErVO_3$										
Act. (I)	5.589	7.554	5.254	0.5693	0.5199	0.46070	0.1140	0.3043	0.0570	0.6910
MLR (II)	5.569	7.582	5.268	0.5689	0.5187	0.4591	0.1119	0.3054	0.0570	0.6910
Δ(I–II)	0.02	–0.028	–0.014	0.0004	0.0012	0.0016	0.0021	–0.0011	0	0
ANN(III)	5.549	7.579	5.256	0.5691	0.5186	0.4581	0.1135	0.3056	0.0582	0.6912
Δ(I–III)	0.04	–0.025	–0.002	0.0002	0.0013	0.0026	0.0005	–0.0013	–0.0012	–0.0002
$LaGaO_3$										
Act. (I)	5.49139	7.77250	5.52298	0.5170	0.5036	0.4950	0.0570	0.2850	0.0290	0.7200
MLR (II)	5.52640	7.75912	5.48466	0.5347	0.5064	0.4886	0.0706	0.2856	0.0375	0.7148
Δ(I–II)	–0.03501	0.01338	0.03832	–0.01770	–0.00280	0.00640	–0.01360	–0.00060	–0.00850	0.00520
ANN(III)	5.51100	7.7641	5.49880	0.5305	0.5053	0.4906	0.0668	0.2840	0.0350	0.7166
Δ(I–III)	–0.01961	0.00840	0.02418	–0.01350	–0.00170	0.00440	–0.00980	0.00100	–0.00600	0.00340
$SrSnO_3$										
Act. (I)	5.7035	8.0645	5.7079	0.5193	0.5062	0.4921	0.0601	0.2802	0.0325	0.7184
MLR (II)	5.7186	8.0827	5.7362	0.5241	0.5032	0.4927	0.0643	0.2802	0.0348	0.7210
Δ(I–II)	–0.0151	–0.0182	–0.0283	–0.0048	0.003	–0.0006	–0.0042	0	–0.0023	–0.0026
ANN(III)	5.7122	8.0882	5.7485	0.5215	0.5029	0.4924	0.0623	0.2808	0.0328	0.7202
Δ(I–III)	–0.0087	–0.0237	–0.0406	–0.0022	0.0033	–0.0003	–0.0022	–0.0006	–0.0003	–0.0018
$NaUO_3$										
Act. (I)	5.9051	8.2784	5.7739	0.5306	0.5075	0.4671	0.0959	0.2984	0.0502	0.6982
MLR (II)	5.8989	8.1432	5.7060	0.5497	0.5129	0.4667	0.1000	0.2957	0.0518	0.7022
Δ(I–II)	0.0062	0.1352	0.0679	–0.0191	–0.0054	0.0004	–0.0041	0.0027	–0.0016	–0.004
ANN(III)	5.8996	8.1795	5.7319	0.5482	0.5123	0.4654	0.0990	0.2982	0.0512	0.7001
Δ(I–III)	0.0055	0.0989	0.042	–0.0176	–0.0048	0.0017	–0.0031	0.0002	–0.001	–0.0019

number of the dependent variables in the data set. The comparison of the *RMSEP* values showed that the most suitable of ANN architecture was the one with 3 hidden neurons. These ANN showed the best prediction abilities.

As previously stated, in order to check the performances of the proposed models, the complete crystal structures for five compounds (test set) were predicted and compared with the actual ones. Three compounds in test set were randomly chosen: two from the largest series ($SmFeO_3$ from the series of orthoferites[47] and $ErVO_3$ from the orthovanadites[42]) and $SrSnO_3$[38] from $A^{2+}B^{4+}O_3$ type of perovskites. Recently, the crystal structure of $NaUO_3$ was reported to be of $GdFeO_3$ type.[48] This is probably the only one example of orthorhombic perovskite of $A^{1+}B^{5+}O_3$ type with refined structure. Therefore, this compound was also chosen for construction of the test set. In our opinion, another interesting compound was $LaGaO_3$, as an example with $c > a$.[30] It seemed interesting to predict the crystal structure of all these compounds, using the developed models. The predicted values, obtained by both approaches, for lattice parameters and fractional atomic coordinates of the compounds in test set are given in Table 4.

With the aim to obtain even better picture for the predictive strength of the models, the distances and angles in the structures of the compounds from test set were calculated using the actual and predicted coordinates. Both, the actual and the predicted distances and angles were calculated in the same programme, in order to avoid some small discrepancies due to application of different programes for solving crystal structure by different authors. For this purpose the programme package *Crystals* 3.0 was used.[56] Selected distances and angles for the compounds of the test set are given in Table 5.

It seemed also interesting to check how much the differences between actual and predicted values for the distances and angles affected the crystal chemistry of the compounds in test set. Thus, for each compound in the test set the theoretical and observed tolerance factors, tilt angles, the distortion of AO_8 and BO_6 polyhedra, bond valences and global instability indices were calculated

Table 5. Comparison between selected distances (Å) and angles (°) for the compounds in test set calculated using the actual and predicted (by MLR and ANN) coordinates.

$SmFeO_3$	act.	MLR	ANN
Sm-O1	2.388	2.399	2.398
Sm-O1	2.308	2.302	2.291
Sm-O2 2x	2.702	2.709	2.706
Sm-O2 2x	2.340	2.340	2.328
Sm-O2 2õ	2.573	2.569	2.564
Fe-O1 2x	2.000	2.001	2.000
Fe-O2 2x	2.012	2.008	2.003
Fe-O2 2x	2.028	2.018	2.015
O1-Fe-O2 2x	88.74	88.61	88.50
O1-Fe-O2 2x	88.74	88.76	88.70
O2-Fe-O2 2x	90.14	90.14	90.25
$LaGaO_3$	act.	MLR	ANN
La-O1	2.646	2.544	2.566
La-O1	2.495	2.404	2.421
La-O2 2x	2.723	2.727	2.726
La-O2 2x	2.450	2.433	2.446
La-O2 2x	2.722	2.642	2.660
Ga-O1 2x	1.969	1.979	1.976
Ga-O2 2x	1.957	1.984	1.978
Ga-O2 2x	1.994	1.991	1.986
O1-Ga-O2 2x	89.76	89.87	89.76
O1-Ga-O2 2x	89.82	90.49	90.97
O2-Ga-O2 2x	91.20	90.83	90.55
$NaUO_3$	act.	MLR	ANN
Na-O1	2.406	2.406	2.419
Na-O1	2.646	2.543	2.543
Na-O2 2x	2.934	2.895	2.894
Na-O2 2x	2.414	2.453	2.445
Na-O2 2x	2.850	2.712	2.745
U-O1 2x	2.151	2.123	2.132
U-O2 2x	2.142	2.125	2.133
U-O2 2x	2.151	2.133	2.142
O1-U-O2 2x	88.43	88.20	88.07
O1-U-O2 2x	88.56	88.40	88.44
O2-U-O2 2x	90.86	90.46	90.66
$ErVO_3$	act.	MLR	ANN
Er-O1	2.298	2.279	2.268
Er-O1	2.217	2.228	2.216
Er-O2 2x	2.665	2.669	2.675
Er-O2 2x	2.264	2.262	2.253
Er-O2 2x	2.481	2.493	2.482
V-O1 2x	1.993	1.998	2.000
V-O2 2x	2.004	2.003	1.998
V-O2 2x	2.021	2.023	2.020
O1-V-O2 2x	87.70	87.70	87.76
O1-V-O2 2x	88.84	88.48	88.30
O2-V-O2 2x	89.35	89.72	89.94
$SrSnO_3$	act.	MLR	ANN
Sr-O1	2.723	2.708	2.716
Sr-O1	2.551	2.524	2.538
Sr-O2 2x	2.853	2.865	2.840
Sr-O2 2x	2.530	2.556	2.557
Sr-O2 2x	2.786	2.770	2.795
Sn-O1 2x	2.046	2.055	2.054
Sn-O2 2x	2.044	2.054	2.055
Sn-O2 2x	2.055	2.062	2.060
O1-Sn-O2 2x	89.61	89.59	89.82
O1-Sn-O2 2x	90.46	90.20	90.33
O2-Sn-O2 2x	90.94	91.27	91.33

using the actual (experimental) and predicted values for structural data using both modelling techniques. The results are presented in Table 6.

The observed tolerance factors (t_o) were calculated using the mean interatomic distances for A (eight coordinated) and B (six coordinated) coordination polyhedra. The actual t_o was calculated using the experimental values for mean interatomic distances extracted from the literature, while MLR and ANN values for t_o were obtained by the mean interatomic distances calculated using the predicted values for lattice parameters and atomic coordinates. As can be seen from Table 6 the values for observed tolerance factors obtained using the predicted values by MLR and ANN are close to each other, and also to the value obtained by the experimental data. The values for t_o are close, as well as, to the values for $t_{cal.}$ calculated using the radii by Shannon.[51]

The angles of tilting were calculated by fractional atomic displacement of anions from the special positions of the cubic unit cell to new positions.[58] This approach was used in order to check the differences between the predicted and the actual values for fractional atomic coordinates, and their influence to the calculated angles of tilting. It might be seen (Table 6) that in the most cases there is very good agreement between the calculated values by the experimental data and by the predicted ones.

It might be noticed that the values for tolerance factors (both actual and predicted) are in good agreement with the tilting angles. Namely, as the tolerance factors decrease, the tilting angles increase. Thus, the tilting is the most pronounced at $ErVO_3$ having the smallest value for tolerance factor and the less pronounced at $SrSnO_3$ having the highest value for tolerance factor.

Another important crystal chemistry data are the bond length distortions of the coordination polyhedra calculated using the equation by Shannon[51]. It should be noticed that for $SmFeO_3$, $ErVO_3$, $SrSnO_3$ and $NaUO_3$ the obtained values for Δ_8 and Δ_6, calculated with the predicted distances, are in very good agreement with the actual ones. However, the actual values for Δ_6 for $LaGaO_3$ are evidently higher than the predicted. As it was mentioned in the paper[30] the GaO_6-octahedron is almost regular with $\Delta_6 = 0.006$. But, it must be emphasised that, in the paper, the value of r (the average Ga–O distance) is miscalculated and consequently the Δ_6 value is also not correct. Therefore, taking into account the lattice parameters and fractional atomic coordinates given in this paper, we have calculated the distances and angles using the programme package Crystals.[56] The obtained value for Δ_6 using these distances is 0.057, which indicates that distortion of GaO_6-octahedra in actual structure exists. This value for Δ_6 is higher than the predicted values, also due to the differences in the coordination number of La in actual structure. Namely, contrary to the other perovskites with this structure that have eight coordinated A-cations, the coordination number of La in this

Table 6. Crystal chemistry of the perovskites in test set*

	$^{VIII}t_o$	$^{VIII}t_{cal.}$	θ(°)	φ(°)	Φ(°)	Δ_8	Δ_6	δ(°)	*V(A)*	*V(B)*	*GII*
$SmFeO_3$											
act.	0.875	0.857	15.66	11.28	19.22	3.863	0.029	1.162	3.072	3.017	0.074
MLR	0.877		15.69	10.90	19.03	3.981	0.013	1.269	3.078	3.054	0.080
ANN	0.875		15.93	11.06	19.31	4.207	0.010	1.455	3.157	3.075	0.121
$ErVO_3$											
act.	0.852	0.833	18.67	12.75	22.47	4.974	0.032	2.567	2.852	2.947	0.103
MLR	0.852		18.44	12.87	22.36	5.158	0.028	2.772	2.843	2.934	0.109
ANN	0.851		18.67	12.87	22.55	5.543	0.024	2.877	2.911	2.949	0.080
$LaGaO_3$											
act.	0.937	0.896	9.24	7.43	11.83	2.295	0.057	0.556	2.945	3.107	0.061
MLR	0.915		11.43	8.06	13.96	2.371	0.006	0.344	3.175	3.015	0.103
ANN	0.922		10.82	7.68	13.24	2.192	0.004	0.369	3.095	3.049	0.067
$SrSnO_3$											
act.	0.933	0.900	9.74	7.05	12.00	2.431	0.006	0.454	1.993	4.075	0.041
MLR	0.929		10.41	6.75	12.39	2.340	0.003	0.725	1.967	3.978	0.024
ANN	0.931		10.12	6.91	12.23	2.269	0.002	0.670	1.940	3.983	0.036
$NaUO_3$											
act.	0.883	0.845	15.83	11.33	19.38	7.031	0.004	1.919	0.935	4.926	0.059
MLR	0.875		16.51	10.59	19.53	4.931	0.004	2.186	0.990	5.207	0.119
ANN	0.875		16.43	11.10	19.74	5.112	0.004	2.398	0.976	5.098	0.060

* t_o is the observed tolerance factor[57] for eight coordinated A-cation (t_o = <A–O>/$\sqrt{2}$·<B–O>), while ^{VIII}t are the tolerance factors calculated using the crystal radii by Shannon.[51] θ, φ and Φ are the octahedral tilt angles calculated by the fractional atomic coordinates.[58] Δ_8 and Δ_6 are the distortion indices of the AO_8 and AO_6 polyhedra, respectively[51] ($\Delta = \frac{\sum[(r_i - r)/r]^2 \cdot 10^3}{n}$, r_i – individual and r – average bond length). V(A) and V(B) are the valences determined by bond valence model.[59] *GII* are the global instability indices.[60]

structure is 10.[30] It should be pointed out that in the performed analysis, the value for the effective ionic radii corresponds to eight-coordinated La^{3+} ion, in accordance to the coordination number for A cations for all compounds in calibration set. Probably, the discrepancies in the actual and predicted values for lattice parameters and fractional atomic coordinates, and consequently for calculated crystallographic parameters, are due to the differences in the values of ionic radii for eight and ten coordinated La^{3+} ion.

The bond valences for A and B cations for each compound in the test set were calculated using the equation proposed by Brown *et al.*:[59]

$$V_i = \sum_i e^{[(r_o - r_{i,j})/B]} \tag{5}$$

As can be seen from Table 6 the values for bond valences of A and B cations for the compounds in the test set are almost equal to the theoretical values. Further, the bond valence results were used for calculations of the values of global instability indices using the equation proposed by Salinas-Sanchez *et al.*:[60]

$$GII = \left\{ \left[\sum_{i=1}^{N} \left(d_i^2 \right) \right] / N \right\}^{1/2} \tag{6}$$

where

$$d_i = V_{i(\mathrm{ox})} - V_{i(\mathrm{cal.})}. \tag{7}$$

As can be seen, the values for *GII* (both actual and predicted) are close to each other and are less than 0.2 which means that all structures of the compounds in test set are stable with small lattice strain in some of the compounds.

4. Conclusions

Two different approaches (ANN and MLR) were used for prediction of the complete crystal structure of orthorhombic ABO_3 perovskites of $GdFeO_3$ type (space group *Pnma*). The obtained values of the coefficients of correlation $(R_{adj})^2$, higher than 0.9 for all ten dependent variables, as well as, the agreement between the actual and predicted values for the dependent variables, indicates that both approaches could be successfully used for prediction of the crystal structures of new members of this series.

The models developed here were tested on five selected compounds (test set) for which the complete structure (lattice parameters and fractional atomic parameters) and consequently distances, angles and several

crystallochemical parameters were calculated. The obtained results for the compounds in the test set, show that there is very good agreement between actual and predicted values for all parameters. Some small disagreements in the case of $LaGaO_3$ are thought that are caused by the differences in the coordination number of La^{3+} ion. Namely, the value for effective ionic radii for La^{3+} used in the analysis corresponds to the eight-coordinated ion instead of actual structure where it is ten-coordinated.

It should be pointed out that this methodology might be extended to similar or some other perovskite structures, such as complex perovskites. An extended model (including some other fundamental variables) for prediction of complete crystal structure of perovskites with Jahn-Teller cations is in progress. Thus, simple models for crystal structure predictions might be constructed for almost each serie of perovskites. However, the choice of relevant variables for describing some special structural peculiarities connected with special properties (such as superconductivity, colossal magnetoresistance etc.) might be difficult.

5. References

1. F. S. Galasso, Perovskites and High T_c Superconductors, Gordon and Breach Science Publishers, New York, USA, **1990**, pp. 3–10.
2. R. H. Mitchell, Perovskites: Modern and Ancient, Almaz Press, Thunder Bay, Ontario, Canada, **2002**, pp. 1–29.
3. E. J. Baran, *Catalysis Today* **1990**, *8*, 133–151.
4. R. H. Butner, E. N. Maslen, *Acta Cryst.* **1992**, *B48*, 644–649.
5. N. W. Thomas, *Acta Cryst.* **1996**, *B52*, 16–31.
6. N. W. Thomas, *Acta Cryst.* **1996**, *B52*, 954–960.
7. O. Fukunaga, T. Fujita, *J. Solid State Chem.* **1973**, *8*, 331–338.
8. N. W. Thomas, *Acta Cryst.* **1991**, *B47*, 180–191.
9. N. W. Thomas, *Acta Cryst.* **1991**, *B47*, 597–608.
10. D. M. Giaquinta, H. Conrad zur Loye, *Chem. Mater.* **1994**, *6*, 365–372.
11. N. W. Thomas, A. Beitollahi, *Acta Cryst.* **1994**, *B50*, 549 –560.
12. N. W. Thomas, *Acta Cryst.* **1998**, *B54*, 585–594.
13. A. A. Bokov, N. P. Protsenko, Z.-G. Ye, *J. Phys. Chem. Solids* **2000**, *61*, 1519–1527.
14. M. W. Lufaso, P. M. Woodward, *Acta Cryst.* **2001**, *B57*, 725–738.
15. S. Aleksovska, V. M. Petru'evski, B. Šoptrajanov, *Acta Cryst.* **1998**, *B54*, 564–567.
16. V. Petruševski, S. Aleksovska, *Croat. Chem. Acta* **1991**, *64*, 577–583.
17. V. Petruševski, S. Aleksovska, *Croat. Chem. Acta* **1994**, *67*, 221–230.
18. S. Aleksovska, V. Petruševski, Lj. Pejov, *Croat. Chem. Acta* **1997**, *70*, 1009–1019.
19. S. Aleksovska, S. C. Nyburg, Lj. Pejov, V. M. Petruševski, *Acta Cryst.* **1998**, *B54*, 115–120.
20. V. M. Petruševski, S. Aleksovska, *Croat. Chem. Acta* **1999**, *72*, 71–76.
21. I. Kuzmanovski, S. Aleksovska, *Chemometr. Intell. Lab. Syst.* **2003**, *67*, 167–174.
22. S. Dimitrovska, S. Aleksovska, I. Kuzmanovski, *Cent. Eur. J. Chem.* **2005**, *3*, 1–18.
23. S. Dimitrovska, I. Kuzmanovski, S. Aleksovska, Proceedings of the International Scientific Conference, South-West University "Neofit Rilsky", Blagoevgrad, **2005**, *2*, pp. 66–74.
24. D. du Boulay, PhD Theses, The University of Western Australia, 1996.
25. N. L. Ross, *Phys. Chem. Minerals* **1998**, *25*, 597–602.
26. A. J. Jacobson, B. C. Tofield, B. E. F. Fender, *Acta Cryst.* **1972**, *B28*, 956–961.
27. J. Ranløv, K. Nielsen, *J. Mater. Chem.* **1994**, *4*, 867–868.
28. M. Marezio, J. P. Remeika, P. D. Dernier, *Acta Cryst.* **1970**, *B26*, 2008–2022.
29. D. Du Boulay, E. N. Maslen, V. M. Streltsov, N. Ishizawa, *Acta Cryst.* **1995**, *B51*, 921–929.
30. L. Vasylechko, A. Matkovskii, D. Savytskii, A. Suchocki, F. Wallrafen, *J. Alloys Comp.* **1999**, *291*, 57–65.
31. L. Vasylechko, Ye. Pivak, A. Senyshyn, D. Savytskii, M. Berkowski, H. Borrmann, M. Knapp, C. Paulmann, *J. Solid State Chem.* **2005**, *178*, 270–278.
32. B. J. Kennedy, C. J. Howard, B. C. Chakoumakos, *Phys. Rev. B* **1999**, *60*, 2972–2975.
33. J. A. Alonso, M. J. Martínez-Lope, M. T. Casais, J. L. Martínez, G. Demazeau, A. Largeteau, J. L. García-Muñoz, A. Muñoz, M. T. Fernández-Díaz, *Chem. Mater.* **1999**, *11*, 2463–2469.
34. J. A. Alonso, M. J. Martínez-Lope, M. T. Casais, M. A. G. Aranda, M. T. Fernández-Díaz, *J. Am. Chem. Soc.* **1999**, *121*, 4754–4762.
35. N. Rosov, J. W. Lynn, Q. Lin, G. Cao, J. W. OžReilly, P. Pernambuco-Wise, J. E. Crow, *Phys. Rev. B* **1992**, *45*, 982–986.
36. G. G. Christoph, A. C. Larson, P. G. Eller, J. D. Purson, J.D. Zahrt, R. A. Penneman, G. H. Rinehart, *Acta Cryst.* **1988**, *B44*, 575–580.
37. J. S. Gardner, G. Balakrishnan, D. McK. Paul, *Physica C* **1995**, *252*, 303–307.
38. E. H. Mountstevens, J. P. Attfield, S. A. T. Redfern, *J. Phys.: Condens. Matter* **2003**, *15*, 8315–8326.
39. S. Sasaki, C. T. Prewitt, J. T. Bass, W. A. Schulze, *Acta Cryst.* **1987**, *C43*, 1668–1674.
40. M. Cwik, T. Lorenz, J. Baier, R. Müller, G. André, F. Bourée, F. Lichtenberg, A. Freimuth, E. Müller-Hartmann, M. Braden, *Phys. Rev. B* **2003**, *68*, 060401(R).
41. J. R. Hester, K. Tomimoto, H. Noma, F. P. Okamura, J. Akimitsu, *Acta Cryst.* **1997**, *B53*, 739–744.
42. J. Pickardt, Th. Schendle, M. Kolm, *Z. Anorg. Allg. Chem.* **1988**, *560*, 153–157.
43. G. V. Bazuev, G. P. Schveickin, *Neorg. Mat.* **1975**, *11*, 1333–1334.
44. V. G. Zubkov, I. F. Berger, Z. M. Pesina, G. V. Bazuev, G. P. Schveickin, *Kristallografia* **1985**, 881–883.

45. H. J. A. Koopmans, G. M. H. Van De Velde, P. J. Gellings, *Acta Cryst.* **1983**, *C39*, 1323–1325.
46. B. J. Kennedy, C. J. Howard, B. C. Chakoumakos, *Phys. Rev. B* **1999**, *59*, 4023–4027.
47. E. N. Maslen, V. A. Strelov, N. Ishizava, *Acta Cryst.* **1996**, *B52*, 406–413.
48. S. Van den Berghe, A. Leenaers, C. Ritter, *J. Solid State Chem.* **2004**, *177*, 2231–2236.
49. J. A. Alonso, M. J. Martínez-Lope, M. T. Casais, M. T. Fernández-Díaz, *Inorg. Chem.* **2000**, *39*, 917–923.
50. R. P. Liferovich, R. H. Michell, *J. Solid State Chem.* **2004**, *177*, 2188–2197.
51. R. D. Shannon, *Acta Cryst.* **1976**, *A32*, 751–756.
52. STATGRAPHICS PLUS, Ver. 3.0, *Statistical Graphics Package,* Educational Institution Edition, Statistical Graphics Corporation, 1994–1997.
53. L. Hadjiiski, P. Geladi, P. Hopke, *Chemometr. Intell. Lab. Syst.* **1999,** *49,* 91–103.
54. J. Zupan, J. Gasteiger, Neural Networks in Chemistry and Drug Design, WCH, Weinheim, **1999**, pp. 125–141.
55. A. Bos, M. Bos, W. E. van der Linden, *Anal. Chim. Acta* **1992**, *256*, 133–144.
56. P. W. Betteridge, J. R. Carruthers, R. I. Cooper, K. Prout, D. J. Watkin, *J. Appl. Cryst.* **2003**, *36*, 1487.
57. S. Sasaki, C. T. Prewitt, R. C. Liebermann, *Am. Mineral.* **1983**, *68*, 1189–1198.
58. Y. Zhao, D. J. Weidner, J. B. Parise, D. E. Cox, *Phys. Earth Planet. Interiors* **1993**, *76*, 1–16.
59. I. D. Brown, D. Altermatt, *Acta Cryst.* **1985**, *B41*, 244–247.
60. A. Salinas-Sanchez, J. L. García-Mu!noz, J. Rodriguez-Carvajal, R. Saez-Puche, J. L. Martínez, *J. Solid State Chem.* **1992**, *100*, 201–211.

Povzetek

Parametre osnovnih celic in atomske koordinate ortorombskih perovskitov ABO_3 smo izrazili kot funkcijo efektivnih ionskih radijev gradnikov z metodama večkratne linearne regresije in nevronskih mrež. V analizo smo vključili 46 ortorombskih perovskitov tipa $GdFeO_3$ (prostorska skupina *Pnma*) z zanesljivo določenimi kristalnimi strukturami: 41 v kalibracijski skupini in 5 za testiranje metode.
Napovedna sposobnost modela je velika, kar potrjujejo vrednosti korelacijskih koeficientov $(R_{adj})^2$, ki presegajo 0.9 za vse odvisne spremenljivke, in ujemanje dejanskih in napovedanih vrednosti odvisnih spremenljivk, določenih z obema metodama. Ta preprosti matematični model lahko uporabimo za napovedovanje kristalnih struktur posameznih članov skupine, kot začetni model za prilagajanje kristalne strukture, in za preverjanje kristalografskih podatkov za perovskite ABO_3.