Application of Evolutionary Rietveld Method Based XRD Phase Analysis and a Self-Configuring Genetic Algorithm to the Inspection of Electrolyte Composition in Aluminum Electrolysis Baths

Igor Yakimov [1], Aleksandr Zaloga [1,*], Petr Dubinin [1], Oksana Bezrukova [1], Aleksandr Samoilo [1], Sergey Burakov [2], Eugene Semenkin [2], Maria Semenkina [2] and Eugene Andruschenko [1]

[1] Department of Research, Siberian Federal University, 660041 Krasnoyarsk, Russia; i-s-yakimov@yandex.ru (I.Y.); dubinin-2005@yandex.ru (P.D.); opiksina@gmail.com (O.B.); x_lab@rambler.ru (A.S.); andryushchenkoeugene@gmail.com (E.A.)

[2] Department of Systems Analysis and Operations Research, Siberian State University of Science and Technology, 660037 Krasnoyarsk, Russia; burakov_krasu@mail.ru (S.B.); eugenesemenkin@yandex.ru (E.S.); semenkina88@mail.ru (M.S.)

* Correspondence: zaloga@yandex.ru; Tel.: +7-913-559-4935

check for updates

Abstract: The technological inspection of the electrolyte composition in aluminum production is performed using calibration X-ray quantitative phase analysis (QPA). For this purpose, the use of QPA by the Rietveld method, which does not require the creation of multiphase reference samples and is able to take into account the actual structure of the phases in the samples, could be promising. However, its limitations are in its low automation and in the problem of setting the correct initial values of profile and structural parameters. A possible solution to this problem is the application of the genetic algorithm we proposed earlier for finding suitable initial parameter values individually for each sample. However, the genetic algorithm also needs tuning. A self-configuring genetic algorithm that does not require tuning and provides a fully automatic analysis of the electrolyte composition by the Rietveld method was proposed, and successful testing results were presented.

1. Introduction

Aluminum is normally produced by the electrolysis of alumina in molten fluorides at a temperature of around 950 °C. The main component of the molten electrolyte is cryolite (Na_3AlF_6), whilst aluminum fluoride, calcium fluoride, and sometimes magnesium fluoride and potassium fluoride are added to improve the cryolite's technological properties. During the electrolysis, the composition of the electrolyte in the baths continuously changes and shifts from the optimum. The maintenance of an optimal bath composition is a vital element in electrolysis technology. An integral characteristic of the bath composition is the cryolite ratio (CR)—the ratio of molar concentrations of sodium fluoride and aluminum fluoride (1):

$$CR = \frac{C(NaF, \text{ mol. \%})}{C(AlF_3, \text{ mol. \%})} = 2 \cdot \frac{C(NaF, \text{ mass. \%})}{C(AlF_3, \text{ mass. \%})} \tag{1}$$

The express process control of the electrolyte composition is generally performed by X-ray diffraction quantitative phase analysis (QPA), which uses calibration curves. The cryolite ratio is calculated according to the Equation (1). The concentrations of NaF and AlF_3 are calculated using the results of the QPA of crystallized bath samples. The phase concentrations, in turn, are calculated from the measured intensities of their diffraction peaks. The optimal frequency of measuring the CR is once every two days, the accuracy of the analysis is $\Delta(p = 0.95)\sim0.04$, and the optimal measurement time per sample is several minutes.

X-ray diffractometers need periodic calibration using electrolyte reference materials with well-established phase composition [1,2]. However, it is difficult to create such reference materials, because, firstly, they must contain all the phases from Table 1. Secondly, the crystal structures of the phases in reference materials must match those in the electrolyte samples. Therefore, QPA by the Rietveld method is more suitable for the process control of bath composition because this method works without using reference materials (as is shown in References [3,4]), yet the low automation of this method in working with such complicated samples limits its applicability. The issue is to set such appropriate initial approximations of both the profile and structural parameters of phases that can be quickly refined automatically by the least squares method (LSM). To address this problem, we suggest applying a genetic algorithm which we have developed to set the initial values of the parameters for each sample automatically [5]. This approach provides high accuracy of measuring the cryolite ratio, but it is not yet fully applicable to the process control. This is because the algorithm must be configured as well. In this paper, we provide a self-configuring genetic algorithm, which works without a preliminary adjustment and performs the fully automated analysis of aluminum electrolyte composition by the Rietveld method.

Table 1. The phase composition of typical industrial bath samples at Russian aluminum smelters.

#	Phase	Chemical Formula	Concentration Range, wt. %	CR Range
1	Cryolite	Na_3AlF_6	0–90	>1.67
2	Chiolite	$Na_5Al_3F_{14}$	0–85	<3.0
3	Sodium fluoride	NaF	0–5	>3.0
4	Ca-cryolite 1	$NaCaAlF_6$	0–15	<3.0
5	Ca-cryolite 2	$Na_2Ca_3Al_2F_{14}$	0–20	<2.95
6	Fluorite	CaF_2	0–9	>2.45
7	Weberite	Na_2MgAlF_7	0–15	<2.85
8	Neiborite	$NaMgF_3$	0–6	>2.5
9	α-, β-, γ-alumina	Al_2O_3	2–5	

References [3,4] also describe an automated analysis of aluminum electrolyte samples using the Rietveld method. However, in these articles, a maximum of 5-phase samples of a calcium-containing electrolyte with an insignificant content (about 0.8%) of the semi-amorphous $NaCaAlF_6$, which difficult to simulate by the Rietveld method, are investigated. As shown below, the proposed approach allows analyses of 8-phase calcium- and magnesium-containing electrolytes (where a feature of Russian aluminum production is adding magnesium fluoride up to 4–5%), with a noticeable $NaCaAlF_6$ content (up to 8% rel.) in an automatic mode for a comparable time. Moreover, in calcium-containing electrolytes, magnesium can also accumulate over time from alumina.

2. Materials and Methods

2.1. The Method of Genetic Algorithm Self-Configuring

Full-profile QPA based on the Rietveld method is widely used for quantitative XRD analysis in laboratories. However, its use for technological inspection in the industry is not yet developed as the Rietveld method is based on a non-linear LSM convergence, which requires a very good approximation of the initial estimations of parameters to be tuned for each sample. In the case of

laboratory investigations, the requirements for the initial approximation are not as strict as there exists the possibility for an interactive refinement. However, for a technological inspection in the industry, a high level of automation and the ability to unify the analysis of a large number of samples are strongly required. In this case, the unified initial estimations of a large amount of profile and structure parameters do not fit well to all of them. This results in a divergence in the LSM. One of the possible approaches to tackle this problem is the application of genetic algorithms (GAs) for the choice of an initial approximation of sample parameters, for the evolutionary selection of perspective parameters, and for their automated refinement with Rietveld's LSM.

The application effectiveness of evolutionary algorithms (GAs in particular) depends on the choice of genetic operators: Selection, recombination (crossover), mutation, and substitution. However, the settings for an effective algorithm which ensure that acceptable results are obtained within the shortest possible time can be different for different problems, i.e., they cannot be determined in advance for all cases. Therefore, procedures of dynamic self-adapting and self-configuring of algorithm settings (e.g., References [6,7]) are used here. Self-configuring is an automated choice of effective genetic operators from a given set during an algorithmic run while solving the problem in hand. The configuration of operators is determined stochastically based on the probability of an operator to be used for a generating new solution. These probabilities are calculated according to their success in previous stages. The deployment probability of the most successful operator, the one that gave the best solutions on the previous generation, is increased, whereas the probabilities of other operators are decreased. It makes possible the automated choice of the best configuration of operators for increasing algorithm productivity.

The main stages of a self-configuring genetic algorithm (SGA) for unconditional optimization can be described as follows:

1. Initially, the choice of any particular variant for each kind of operator (selection, crossover, mutation) is equiprobable. More specifically, the probability of choosing a variant of an operator is equal to $p = 1/z$, where z is the number of operator variants. It means that all variants of all operators are used equiprobably before statistics of their effectiveness are collected.

2. On each generation, an effectiveness estimation is performed for each variant of each operator. It is based on the mean fitness of solutions obtained with the use of this variant of this operator: $averagefitness_i = f_i/n_i$, $i = 1, 2, \ldots, z$, where $averagefitness_i$ is the mean fitness of solutions obtained with the i-th variant of the operator; f_i is the fitness sum of all solutions obtained with the i-th variant of the operator; n_i is the number of solutions obtained with the i-th variant of the operator; $i = 1, \ldots, z$, where z is the number of operator variants.

3. For the next generation, the probability of using the most effective variant is increased by $((z-1) \cdot K)/(z \cdot N)$ and probabilities of all other variants are decreased by $K/(z \cdot N)$, where N is the number of established generations of an algorithm run, K is a constant (usually equal to 2 for the considered problems). However, the probability of all variants cannot be lower than a given threshold, whereas the sum of all variant probabilities must be equal to 1. When an operator variant reaches this threshold, it will stop giving out part of its probability, and the best variant will no longer receive it. It is organized in this way because of the possibility that a variant could be unsuccessful on the first stages but could be very useful later on, and this could not happen if its probability decreased to zero.

4. Operators used for the generation of a new solution are chosen stochastically according to obtained probability distributions.

Such self-configuring frees the end user who is not an expert in evolutionary optimization from choosing the settings of the genetic algorithm, whilst the efficiency of solving the problem remains acceptable (with the best choice of the genetic algorithm parameters, the efficiency of solving the problems is somewhat higher, but the selection of the GA parameters requires time and a highly-qualified user).

The effectiveness of an SGA can be improved by using it within the framework of the island evolutionary (cooperative-competing) model, when several populations exist separately from each other, only at times exchanging genetic material. This ensures a more uniform distribution of possible solutions to the problem within the search space. Therefore, in order to solve the problem of quantitative X-ray phase analysis, the following realization of self-configuration of the multi-population parallel genetic algorithm was created. It generates n different populations from the models of the substance being determined, and on each of the n computational nodes of the multi-core personal computers (PC), an individual single-population SGA is run. At the beginning of the process, random individuals are generated, i.e., sets of numbers consisting of the values of the refined parameters of the Rietveld method for the generated models, which are distributed over the search space. At each of the computational nodes, with the help of the recombination and selection operators, evolutionarily occurs the formation of descendants with smaller objective function values. The mutation operators randomly "scatter" them around the search space, sometimes with an increase in the value of the objective function. A proportion of the models with a smaller objective function value are refined using Rietveld's LSM method. Then, as a result of the general selection, a new population of test models is formed, i.e., descendants, on average, with better suitability. A certain number of the best test models from the populations at work nodes are sent to the control computer node of the SGA. All these decisions accumulated on the controlling node are sorted in decreasing order of the value of the objective function. Periodically, some of the best solutions accumulated at a given generation of evolution on the control node are randomly selected and randomly returned to the population at work nodes. Such a moderate migration ensures the spread of successful solutions to populations and improves overall convergence.

Self-configuring is realized for individual GA processes on work units in the way described above. The standard set of genetic operators is given for each process. They are one-point, two-point, and uniform crossover, rank-based and tournament with different size of tournament selection, and low, average, and high selection. Probability redistribution of all operators is performed locally for each work unit irrespective of their effectiveness on other units. This last point could improve the general effectiveness of the algorithm but requires a separate careful study.

2.2. Full-Profile QPA by Parallel Self-Configuring Genetic Algorithms

The essence of the QPA by the Rietveld method is an iterative minimizing of the difference between an experimental powder pattern and the calculated one by the LSM:

$$\Phi(\overline{P}_{k+1}) = \sum_i w_i (Yo(2\Theta_i) - Yc(\overline{P}_k + \Delta\overline{P}_k, 2\Theta_i))^2 \to 0, \tag{2}$$

where Yo, Yc is an experimental and a calculated intensity at a position $2\theta i$, respectively; w_i is a weight coefficient; $\overline{P}_k$ is the vector of profile, microstructural, and structural parameter values at an iteration k; $\Delta\overline{P}_k$ is the parameter increments calculated by the LSM; the initial approximations are set at $k = 0$.

A refinable part of $\overline{P}$ composes parametrical strings that play the role of individuals that are evolutionarily optimized by a GA. The full set of parameters $\overline{P}$, which includes both refinable and fixed parameters, describes a trial model of multiphase sample characteristics. In the case of the evolutionary QPA, a range must be defined within which possible values of refinable parameters fall. The best values found within the range by the GA are then refined by the Rietveld method.

A QPA feature is that it allows the GA to conduct the search for appropriate initial values of parameters within wide ranges. For example, the crystal structures of phases having been found in a sample may be used to set initial values of refinable structural parameters. Crystal structures are normally taken from crystal structure databases. In this case, the atomic coordinates of general crystallographic positions may be chosen as refinable parameters. In addition, the occupation of the positions may be refined for solid solutions. Thus, the range limits the variation of both atomic coordinates and occupation coefficients.

After the completion of the GA process, the final solution is refined using the Rietveld method, and the phase concentrations are calculated using the found parameter values. If the sample does not contain an amorphous phase, the concentrations are calculated according to the following equation:

$$C_\alpha = S_\alpha Z_\alpha M_\alpha V_\alpha / \sum_{j=1}^{N} S_j Z_j M_j V_j, \tag{3}$$

where S_a is the scale factor of a phase a, which is obtained from the calculated powder pattern Yc_a, V_a is the cell volume, Z_a is the number of structural units per cell, M_a is the molecular weight of a phase a, and N is the number of crystalline phases in the sample.

If an amorphous phase is present in the sample, the QPA uses an internal standard [8].

Narrow search ranges pose a problem for the evolutionary full-profile QPA. In such cases, the values of the R-factor vary insufficiently. Therefore, it becomes an unreliable selection criterion. To improve the sensitivity of the criterion, we suggest adding bias between the measured sample's elemental composition and the composition calculated from the phase concentrations.

$$R_{wp} = 100 \cdot \left\{ (1 - w_{Ch}) \cdot \sqrt{\frac{\sum_i w_i \cdot [Yo_i - Yc_i(\overline{P})]^2}{\sum_i w_i \cdot (Yo_i)^2}} + w_{Ch} \cdot \frac{\sum_t w_t \cdot [\sum_\alpha p_{t\alpha} \cdot \frac{S_\alpha Z_\alpha M_\alpha V_\alpha}{\sum_\alpha S_\alpha Z_\alpha M_\alpha V_\alpha} - C_t^{Ch}]^2}{\sum_t w_t \cdot (C_t^{Ch})^2} \right\}, \tag{4}$$

where C_t^{Ch} is the concentration of an element t measured by chemical analysis; P_{ta} is the mass fraction of an element t in a phase a; w_{Ch} is the weight contribution of the chemical data in R_{wp} (normally 0.5).

The combination of the suggested variant of the full-profile Rietveld method with the parallel SGA provides an automated QPA. The SGA uses the profile R-factor as a figure of merit to ensure a proper selection of trial models. To perform optimization by the SGA, the special software was written in C++ language. The ObjCryst++ library [9] was used for crystallographic calculations and Rietveld method refinements.

2.3. Objects of Investigations

We applied the SGA to the QPA of the aluminum bath electrolyte. As the model objects, we had chosen 24 branch reference materials used at five Russian aluminum smelters.

The objects had been chosen for the following reasons. Firstly, they were made from real industrial baths taken at different aluminum smelters. Therefore, they were entirely consistent with all the features of real crystallized bath samples, such as composition, impurities, and microstructure. Secondly, the balance between the chemical and phase composition of the reference samples is fully guaranteed by the correspondence among the results obtained by the different analytical methods used for the certification. The mean uncertainty of the certified CR values was 0.008. Thirdly, the quantitative phase composition significantly varies from sample to sample and covers the range of cryolite ratios from 1.9 to 3.

The samples were ground manually in an agate mortar and then pressed in cuvettes from the front side. The powder patterns were obtained with a Shimadzu-7000 powder diffractometer with scintillator detector using *CuKα* radiation in the range $10° \leq 2\Theta \leq 90°$; the exposition step was $0.01°$. The structural models were taken from the Inorganic crystal structure database (ICSD) [10].

Since the SGA QPA is fully automated, we preset the SGA for the analysis of each sample identically. We also compiled a standard excessive list of contained phases, which is provided in Table 1, for each sample. The nineteen parameters listed in the Table 2 were refined for the chosen phases. In cases when a phase was absent in a sample, its scale factor and the concentration were set to zero. The SGA was run three times for each sample. As the QPA result, we accepted the arithmetic mean of the phase concentrations that were established over three runs. The analysis of each sample lasted about five minutes.

Table 2. Parameters that were refined by the self-configuring genetic algorithm (SGA) for the bath reference materials.

Phase	List of Refinable Parameters
Na_3AlF_6	*S, a, b, c, β, U, W, Eta0, Asym1*, atomic coordinates (12 atomic positions), the texture *hkl* [220] by March-Dollase model
$Na_5Al_3F_{14}$	*S, a, c, U, W, Eta0, Asym1*, atomic coordinates (9 atomic positions)
$NaCaAlF_6$	*S, a, b, c, β, U, W*, Eta0, *Asym1*
$Na_2Ca_3Al_2F_{14}$	*S, a, U, W*, Eta0, *Asym1*
CaF_2	*S, a, U, W*, Eta0, *Asym1*
Na_2MgAlF_7	*S, a, b, c, U, W*, Eta0, *Asym1*
$NaMgF_3$	*S, a, b, c, U, W*, Eta0, *Asym1*
$α-Al_2O_3$	*S, a, b, U, W*, Eta0, *Asym1*

where *S* is the scale factor; a, *b, c, β* are unit cell parameters; *U, W* is a peak FWHM by Pseudo-Voigt; *Eta0* is the peak shape parameter; *Asym1* is the peak asymmetry parameter.

The research laboratories at aluminum smelters are equipped with combined XRD-XRF analyzers. Normally, the analyzers combine an X-ray diffractometer with a fixed x-ray fluorescence channel that provides quantification of calcium and magnesium. For this reason, we used concentrations of these two elements to calculate the R-factor according to Reference [4].

We preset the following genetic operators for the SGA:

1. Tournament selection among 3, 5, 7, 9 trial models, range selection;
2. Two-point crossover, three-point crossover, uniform crossover;
3. Low-level mutation, average-level mutation, high-level mutation, with three standard deviations each.

The probability of operators varied adaptively for each sample during the SGA performance. In addition, the SGA used a local optimization by Lamarck.

3. Results

Figure 1 shows how the parallel SGA typically converges during the search for the profile and structural parameters. The graph was plotted during the running of the full-profile QPA of a bath reference material. The *X*-axis shows the number of the generation, whilst the *Y* axis provides the best corresponding R-factor value that was found among the population at the managing unit. At the zero generation, the SGA randomly generates the initial populations of the trial models.
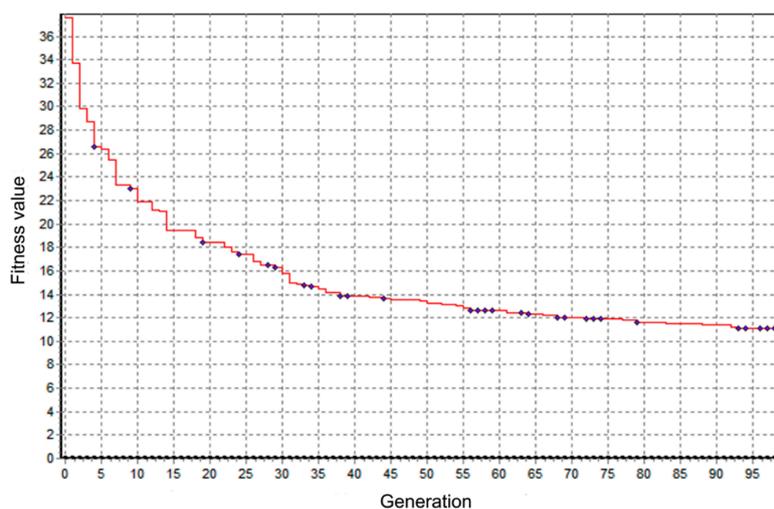
Figure 1. A typical graph of how the SGA converges when analyzing an electrolyte sample.

At the final stage of the QPA, the approximate parameter values, which have been found by the SGA, are exposed to the Rietveld refinement. Figure 2 depicts the experimental powder pattern of the analyzed bath reference sample and the profile that was calculated after the Rietveld refinement. The value of the profile R-factor, which characterizes the difference between the profiles, is 8.6%.
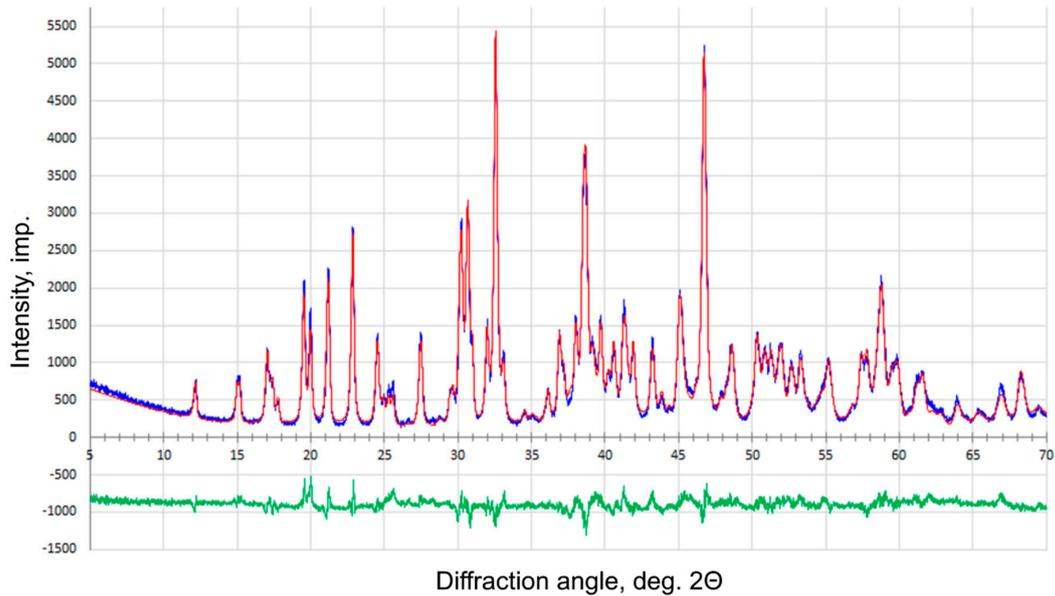
Figure 2. The model powder diffraction pattern calculated by the SGA (in red) and the experimental powder diffraction pattern (in blue) of a bath sample. The green line shows the difference between the profiles.

We propose the correspondence between the certified and calculated values of the cryolite ratio as the quality criterion for results of the evolutionary full-profile QPA. The cryolite ratios were calculated according to the Equation (1). We used the phase concentrations that were computed according to Equation (3) to find the shares of sodium fluorite and aluminum fluorite. Figure 3 shows the correspondence between the certified and calculated values of the cryolite ratio.

Figure 3 also provides the linear regression equation ($y = a + bx$) and the standard deviation, which numerically characterize the correspondence. Therefore, the bias of b from 1 characterizes the systematic error of the results, while the standard deviation describes the random error.
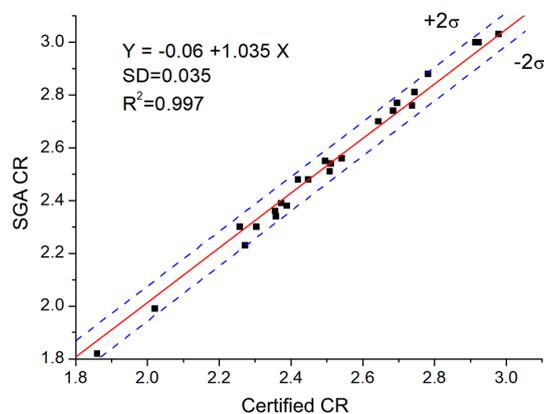
Figure 3. The correspondence between the calculated and certified CR values for the branch reference materials. Certified CR is the certified values; SGA CR is the calculated values; SD is the standard deviation.

4. Discussion

The calculated values match the certified values with an accuracy of SD = 0.035, and all the results fall within the 95% confidence interval. The linear regression equation is close to a $y = x$ form because the a coefficient is statistically insignificant. However, the b coefficient is 3.5 relative per cent higher than 1. This indicates that the results of the evolutionary full-profile QPA are slightly overestimated.

An analysis shows that this systematic error is caused by the overestimation of the Na_3AlF_6 concentration. This fact proves that the automated Rietveld method that uses SGA data refines the structure of this phase ineffectively. It appears that the structure distorts due to the incomplete transition of the Na_3AlF_6 high-temperature modification to the low-temperature modification. Such an effect is a result of the nonequilibrium crystallization of bath samples, which is caused by a specific sampling procedure being used at aluminum smelters. In addition, the structural distortion inflates the standard deviation of the results.

Overall, the results meet the technological requirements that are set for the accuracy of CR analysis at the smelters. Therefore, we recommend the automated evolutionary method of QPA for the express control of bath composition. However, prior to implementing the method in industry, we must improve its performance by eliminating the causes of the systematic error.

Author Contributions: Conceptualization, I.Y.; Data curation, S.B.; Formal analysis, E.S.; Funding acquisition, I.Y. and A.Z.; Investigation, A.S. and E.A.; Methodology, S.B. and M.S.; Project administration, I.Y.; Resources, O.B. and A.S.; Software, A.Z.; Supervision, P.D. and E.S.; Validation, S.B.; Writing—original draft, I.Y.; Writing—review & editing, A.Z. and P.D.

References

1. *Electrolytic Bath Standards*; Alcan International Ltd.: Montreal, QC, Canada, 2005.
2. Yakimov, I.S.; Dubinin, P.S.; Zaloga, A.N.; Piksina, O.E.; Kirik, S.D. Development of industry standard electrolyte samples of aluminum electrolyzers. *Stand. Samples* **2008**, *4*, 34–42. (In Russian)
3. Feret, F.R. Breakthrough in analysis of electrolytic bath using Rietveld-XRD method. In Proceedings of the TMS2008: 137th Annual Meeting & Exhibition of the Minerals, Metals & Materials Society, New Orleans, LA, USA, 9–13 March 2008; pp. 343–346.
4. Knorr, K.; Kelaar, C. Automated analysis of aluminium bath electrolytes by the Rietveld method. *Miner. Eng.* **2009**, *22*, 434–439. [CrossRef]
5. Zaloga, A.; Akhmedova, A.; Yakimov, I.; Burakov, S.; Semenkin, E.; Dubinin, P.; Piksina, O.; Andryushchenko, E. Genetic Algorithm for Automated X-ray Diffraction Full-Profile Analysis of Electrolyte Composition on Aluminium Smelters. In Proceedings of the Informatics in Control, Automation and Robotics 12th International Conference, Colmar, Alsace, France, 21–23 July 2015; pp. 79–93. [CrossRef]
6. Semenkin, E.; Semenkina, M. The Choice of Spacecrafts' Control Systems Effective Variants with Self-Configuring Genetic Algorithm. In Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics, ICINCO'2012, Rome, Italy, 28–31 July 2012; Ferrier, J.-L., Bernard, A., Gusikhin, O., Madani, K., Eds.; pp. 84–93.
7. Semenkin, E.S.; Semenkina, M.E. Self-configuring Genetic Algorithm with Modified Uniform Crossover Operator. *Adv. Swarm Intell. Lect. Notes Comput. Sci.* **2012**, *1*, 414–421. [CrossRef]
8. Bish, D.L.; Howard, S.A. Quantitative phase analysis using the Rietveld method. *J. Appl. Cryst.* **1988**, *21*, 86–91. [CrossRef]

9.   Favre-Nicolin, V.; Cerny, R. FOX, 'free objects for crystallography': A modular approach to ab initio structure determination from powder diffraction. *J. Appl. Cryst.* **2002**, *35*, 734–743. [CrossRef]
10.   Inorganic Crystal Structure Database. Available online: http://www2.fiz-karlsruhe.de/icsd_publications.html (accessed on 28 August 2018).