



A Comprehensive Comparison of Regression Models for the Estimation of Liquidus Temperature of Aluminum Electrolytic Cell

Konstantinos Betsis, Nikolaos Karkalos, Anthimos Xenidis

National Technical University of Athens, Zografou Campus, Athens 15780, Greece

Keywords: Hall-Héroult process, aluminum electrolytic cell, liquidus temperature, regression models

Abstract

Aluminum production in industrial practice is mainly based on the Hall-Héroult process in the form of an electrolytic cell in which cryolite is employed as solvent. One fundamental parameter in these cells is the determination of the liquidus temperature of the cryolitic bath, which in turn leads to the determination of the superheat, i.e. the temperature required to overcome energy losses and achieve the appropriate electrolytic cell operation. As the liquidus temperature is dependent on the bath constituents' concentration, it is not a trivial task to determine the liquidus temperature with sufficient accuracy using a reliable approach. Although computational approaches, especially multiphysics and multilevel ones can provide some reliable estimates of the liquidus temperature, their high computational cost renders them inappropriate for everyday use in the industry, compared to regression equations. In this work, a comprehensive investigation on the appropriate regression equation for the prediction of the liquidus temperature of the cryolitic bath is carried out based on a large dataset of liquidus temperature values developed by thermodynamic software. Several types of regression equations are tested and evaluated based on different performance metrics. The final results allow for the determination of the form of the appropriate equation which can then be used for the prediction of the liquidus temperature in real experimental or industrial level aluminum electrolytic cells.

Introduction

Aluminum production is based on the Hall-Héroult process which involves the dissolution of aluminum oxide into a cryolitic bath including mainly cryolite and AlF_3 and eventually other compounds as well due to reactions with impurities. The existence of these compounds is shown to produce a noticeable effect on the cryolitic bath properties such as the liquidus temperature, the density and viscosity among others [1]. Due to the increased concern about the sustainability of the Hall-Héroult process it is rather essential to minimize the heat losses during aluminum production, something that can be achieved by more accurate estimation of the necessary preheat in respect to the bath



composition [2-5]. As the direct measurement of liquidus temperature for specific bath compositions during the electrolytic cell operation is difficult and costly to be carried out extensively, it is important to establish appropriate predictive models for the estimation of liquidus temperature and ultimately the necessary superheat.

During the past decades, a few authors have developed semi-empirical formulas for the estimation of liquidus temperature based on experimental measurements [7-9]. These formulas, usually in the form of non-linear regression equations have been used with different degrees of reliability, with the equation of Solheim et al. [7, 10] being the most widely used one, as it was shown to provide sufficient accuracy for a wide range of cryolitic bath constituents' compositions. In specific, the model proposed by Solheim et al. [10] included terms relevant to the different bath constituents but also some interaction terms and high order terms of the most important compounds such as AlF_3 , Al_2O_3 or CaF_2 . Apart from the equation of Solheim et al., Peterson and Tabereaux [11] developed an equation for the liquidus temperature of the aluminum cryolitic bath which focused on four terms, namely the cryolite, AlF_3 , CaF_2 and Al_2O_3 . Moreover, Di Yuezhong et al. [12] proposed a model including both first and second order terms of the bath constituents and achieved improved results than the Solheim model in some cases such as for a specific range of KF values. Recently some interesting works involving other machine learning models such as Multilinear Perceptron and Support Vector Regression among others have been presented in the relevant literature with very promising results in terms of accuracy [13,14], but in these cases the use of a black-box model has limited practical applicability in industry and does not easily provide an insight into the physical interpretation of the variation of the cryolitic bath properties such as the liquidus temperature.

However, regarding the semi-empirical equations, given the deviations of the composition of aluminum ores in different regions, these equations may have limited applicability in special cases or need to be revised in order to adapt to different situations [15]. Thus, in the current work, a comprehensive framework for the development of an improved equation for the prediction of the liquidus temperature of the cryolitic bath is proposed, based initially on the testing of different regression models trained by a large dataset provided by a thermodynamic software and then on the final derivation of model coefficients of the chosen equation and its validation by an experimental dataset.

Materials and methods

Procedure for derivation of regression equation

In this work, the determination of an appropriate regression equation for the prediction of aluminum cryolitic bath liquidus temperature is proposed using a specific methodology, depicted in Fig.1. At first, a large dataset including 2860 combinations of bath compositions and their respective predicted liquidus temperatures was developed by means of FactSage thermodynamic software, mentioned hereafter as the “first dataset”, and then, four different regression models were evaluated based on



their accuracy for the prediction of the bath liquidus temperature. The first model, derived from our previous work, included first order terms and interaction terms, as follows:

$$T_{liq} (^{\circ}C) = a_0 + a_1 xAlF_3 + a_2 xCaF_2 + a_3 xAl_2O_3 + a_4 xLiF + a_5 xMgF_2 + a_6 xAlF_3 * xCaF_2 + a_7 xAlF_3 * xAl_2O_3 + a_8 xAlF_3 * xKF + a_9 xCaF_2 * xAl_2O_3 + a_{10} xCaF_2 * xKF + a_{11} xAl_2O_3 * xKF + a_{12} xLiF * xKF \quad (1)$$

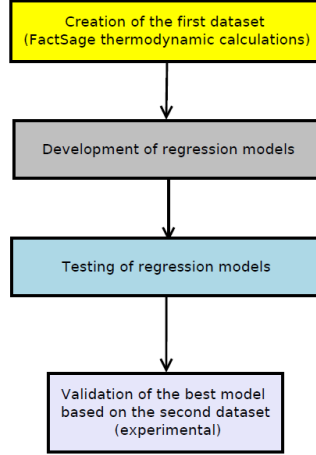


Figure 1: Flowchart of the procedure followed in this work

In the first dataset, the following range was assumed for each bath component: for AlF_3 0.0-15.0%, for CaF_2 0.0-6.0%, for Al_2O_3 0.0-5.5%, for LiF 0.0-1.1%, for MgF_2 0.0-1.1% and for KF 0.0-1.1%. Although this model can provide low percentage error values, the strict requirements for the prediction of the liquidus temperature with a deviation of at most a few degrees Celsius, rendered necessary the testing of three additional equations with non-linear terms. Specifically, one of these equations included two exponential terms in the form of $a_i x_i^m$ for AlF_3 and Al_2O_3 respectively (Eq.2), the second of these equations included two rational terms in the form of $\frac{x_i}{1+a_i x_i}$ for AlF_3 and Al_2O_3 (Eq.3) and the third equation was similar to the first one, with $m = 2$, thus equivalent to a model with second order terms (Eq.4). The selection of the tested model types, represented by the following equations, was carried out based on the relevant literature and preliminary investigations. The general form of the regression equation includes mostly linear terms for the sake of simplicity but also as the physical meaning can be retained. The use of additional non-linear terms reflects the experimentally observed trends which indicated a stronger non-linear correlation for some of the terms e.g. AlF_3 or Al_2O_3 .

$$T_{liq} (^{\circ}C) = a_0 + a_1 xAlF_3 + a_2 xCaF_2 + a_3 xAl_2O_3 + a_4 xLiF + a_5 xMgF_2 + a_6 xAlF_3 * xCaF_2 + a_7 xAlF_3 * xAl_2O_3 + a_8 xAlF_3 * xKF + a_9 xCaF_2 * xAl_2O_3 + a_{10} xCaF_2 * xKF + a_{11} xAl_2O_3 * xKF + a_{12} xLiF * xKF + a_{13} xAlF_3^m + a_{14} xAl_2O_3^n \quad (2)$$

$$T_{liq} (^{\circ}C) = a_0 + \frac{a_1 xAlF_3}{1+a_{13}xAlF_3} + a_2 xCaF_2 + \frac{a_3 xAl_2O_3}{1+a_{14}xAl_2O_3} + a_4 xLiF + a_5 xMgF_2 + a_6 xAlF_3 * xCaF_2 + a_7 xAlF_3 * xAl_2O_3 + a_8 xAlF_3 * xKF + a_9 xCaF_2 * xAl_2O_3 + a_{10} xCaF_2 * xKF + a_{11} xAl_2O_3 * xKF + a_{12} xLiF * xKF \quad (3)$$



$$T_{liq} (^{\circ}\text{C}) = a_0 + a_1 x\text{AlF}_3 + a_2 x\text{CaF}_2 + a_3 x\text{Al}_2\text{O}_3 + a_4 x\text{LiF} + a_5 x\text{MgF}_2 + a_6 x\text{AlF}_3 * x\text{CaF}_2 + a_7 x\text{AlF}_3 * x\text{Al}_2\text{O}_3 + a_8 x\text{AlF}_3 * x\text{KF} + a_9 x\text{CaF}_2 * x\text{Al}_2\text{O}_3 + a_{10} x\text{CaF}_2 * x\text{KF} + a_{11} x\text{Al}_2\text{O}_3 * x\text{KF} + a_{12} x\text{LiF} * x\text{KF} + a_{13} x\text{AlF}_3^2 + a_{14} x\text{Al}_2\text{O}_3^2 \quad (4)$$

In each case, the root mean squared error (RMSE) and mean absolute percentage error (MAPE) were evaluated for the dataset produced by the thermodynamic software and these values were also compared to the ones of the Solheim equation. After the best model was selected, its terms were determined based on the experimental dataset, named hereafter as “second dataset” and finally, it was compared to the experimental data as well as the values of the Solheim equation for the specific bath compositions.

Experimental procedure

Natural Greenland cryolite, with a melting point of $1011^{\circ}\text{C} \pm 1^{\circ}\text{C}$ and high purity ($>99.99\%$), along with AlF_3 , CaF_2 , Al_2O_3 , KF , LiF , and MgF_2 , were mixed in specific ratios to prepare synthetic cryolitic baths. These mixtures were then placed in platinum crucibles and used for liquidus temperature measurements by differential scanning calorimetry (DSC).

Pure aluminum, silver, and gold were used as calibration standards, with calibration performed at a heating rate of $1^{\circ}\text{C}/\text{min}$ under a helium atmosphere to prevent thermal hysteresis and ensure accurate measurements. The liquidus temperature measurement procedure involved progressively increasing the sample temperature at a rate of $10^{\circ}\text{C}/\text{min}$ up to a predefined temperature (approximately 30°C below the liquidus temperature estimated from literature and/or thermodynamic data), followed by equilibration for 5 minutes. The temperature was then further increased at a rate of $1^{\circ}\text{C}/\text{min}$ until it exceeded the estimated liquidus temperature by approximately 20°C .

The liquidus temperature was determined from the second derivative of the temperature change with respect to time. As the sample absorbs energy to increase its temperature at a rate of $1^{\circ}\text{C}/\text{min}$, the solids begin to melt, and the temperature of the sample no longer increases at the same rate. The determination of the liquidus temperature relies on the second derivative of the sample temperature with respect to time. Changes in the temperature-time relationship reflect the onset and completion of melting, the latter corresponding to the bath liquidus temperature. The second dataset includes 47 experimental data samples. Additionally, in this dataset, the range of bath component values was as follows: AlF_3 : 0.0–14.91%, CaF_2 : 0.0–8.01%, Al_2O_3 : 0.0–6.38%, KF : 0.0–3.75%, MgF_2 : 0.0–3.07%, and LiF : 0.0–2.92%.



Results and discussion

Determination of the best performing regression equation

At first, the different ML models were developed and evaluated based on the first dataset, derived from calculations in the FactSage software. Regarding the multiple linear regression model, presented in Eq.1 the final model coefficients using the experimental dataset is given in Eq. 5.

$$T_{liq} (^{\circ}C) = 1018.0 - 1.277 xAlF_3 - 1.440 xCaF_2 - 4.441 xAl_2O_3 - 9.256 xLiF - 6.405 xMgF_2 - 0.2562 xAlF_3 * xCaF_2 - 0.3089 xAlF_3 * xAl_2O_3 - 0.6099 xAlF_3 * xKF + 0.2541 xCaF_2 * xAl_2O_3 + 1.190 xCaF_2 * xKF + 0.798 xAl_2O_3 * xKF - 2.319 xLiF * xKF \quad (5)$$

The values of the RMSE and MAPE metrics of this model were 5.814°C and 0.574% respectively. Although the MAPE value is sufficiently low, given that an acceptable error limit for predictive models is 10%, the requirements for accurate prediction of the liquidus temperature, in order to regulate the superheat as close to the optimum value as possible, render the achieved RMSE value as marginally acceptable. This is the reason why improved models are also tested.

Regarding the model which included exponential terms, presented in Eq. 2, the derived equation is the following:

$$T_{liq} (^{\circ}C) = 1011 - 270.648 * xAlF_3 - 4.420 * xCaF_2 - 7.407 * xAl_2O_3 - 6.929 * xMgF_2 - 11.048 * xLiF - 0.028 * xAlF_3 * xCaF_2 - 0.130 * xAlF_3 * xAl_2O_3 + 0.087 * xAlF_3 * xKF + 0.473 * xCaF_2 * xAl_2O_3 + 0.290 * xCaF_2 * xKF + 0.536 * xAl_2O_3 * xKF - 0.823 * xLiF * xKF + 279.81 * xAlF_3 ^{0.986} - 12.313 * xAl_2O_3 ^{0.001} \quad (6)$$

For this model, the values of the RMSE and MAPE metrics are 3.849°C and 0.147 %, respectively, which are lower than those of the first model, especially the MAPE.

The derived model which includes rational terms, described in Eq.3 is formulated as follows:

$$T_{liq} (^{\circ}C) = 999.32 - \frac{7.933e+9 xAlF_3}{1-1.861e9 xAlF_3} - 0.452 xCaF_2 - \frac{1.827e10 xAl_2O_3}{1+3.223e+9 xAl_2O_3} - 5.312 xLiF - 5.900 xMgF_2 - 0.343 xAlF_3 * xCaF_2 - 0.722 xAlF_3 * xAl_2O_3 - 0.188 xAlF_3 * xKF + 0.258 xCaF_2 * xAl_2O_3 + 0.220 xCaF_2 * xKF + 0.114 xAl_2O_3 * xKF + 3.085 xLiF * xKF \quad (7)$$

For this model, the values of the RMSE and MAPE metrics are 2.680°C and 0.128 %, respectively, which are also lower than those of the first and second model.

Finally, for the model with second order terms, the relevant equation is the following:

$$T_{liq} (^{\circ}C) = 1009.95 + 0.615 * AlF_3 - 2.769 * CaF_2 - 5.818 * Al_2O_3 - 10.341 * LiF - 6.589 * MgF_2 + 0.041 * AlF_3 * CaF_2 + 0.024 * AlF_3 * Al_2O_3 + 0.243 * AlF_3 *$$



$$\begin{aligned} & \text{KF} - 0.285 * \text{CaF}_2 * \text{Al}_2\text{O}_3 + 0.001 * \text{CaF}_2 * \text{KF} + 0.375 * \text{Al}_2\text{O}_3 * \text{KF} - \\ & 1.847 * \text{LiF} * \text{KF} - 0.220 * \text{AlF}_3^2 + 0.033 * \text{Al}_2\text{O}_3^2 \end{aligned} \quad (8)$$

In this model, the values of the RMSE and MAPE metrics are 1.702°C and 0.056%, respectively, which are the lowest among all models.

By comparing the performance of the four different models regarding their RMSE and MAPE values when trained by the first dataset, it can be also seen from Fig.2a and Fig. 2b, it can be determined that the best performing model is the one described by Eq.8, which includes the second order terms. Moreover, also in comparison with the results of the Solheim equation this equation outperforms the Solheim model, which has an RMSE value of 2.514°C and a MAPE value of 0.176% for the first dataset.

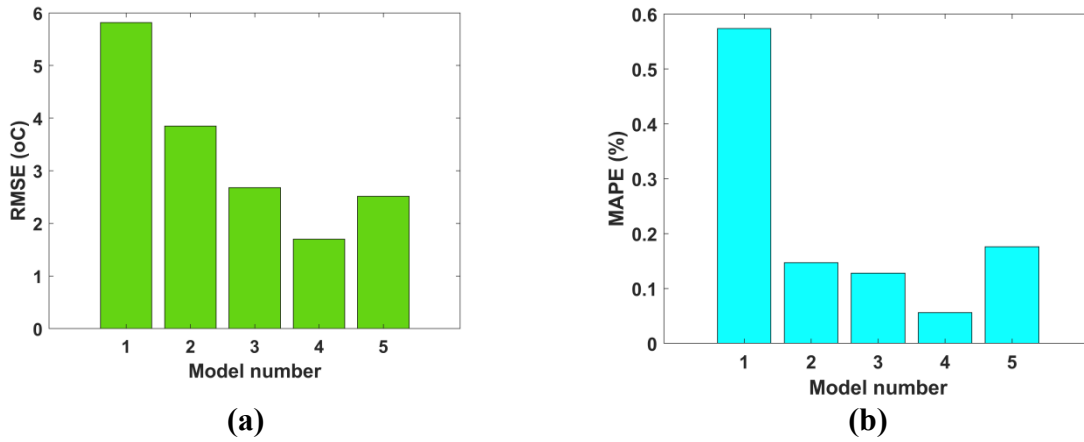


Figure 2: (a) RMSE and (b) MAPE values of the different models compared in this work (1 denotes the model described by Eq. 5, 2 the model described by Eq.6, 3 the model described by Eq. 7, 4 the model described by Eq.8 and 5 the Solheim model)

In order to check the appropriateness of the final model regarding the regression assumptions, the normality of the errors was evaluated by creating the Q-Q plot and carried out statistical tests for homoscedasticity. As can be seen in the Q-Q plot of Fig. 3, most of the points follow a straight line with some slight deviations close to the two ends. This is anticipated as much fewer data exist close to the extreme temperatures. Thus, the distribution is very close to normal as the small deviations towards the tails cannot indicate reasons for statistically significant deviations from normality.

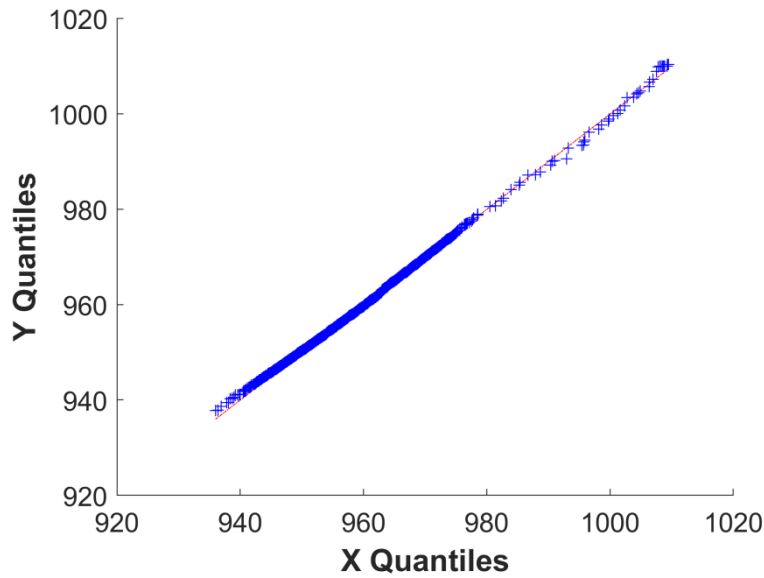


Figure 3: Q-Q plot for the selected regression model.

Regarding homoscedasticity, the Bartlett's test and Brown-Forsythe test were carried out. The results of Bartlett's test indicated a p-value of $0.37 > 0.05$ suggesting that the null hypothesis of equal variances cannot be rejected. Moreover, the results of Brown-Forsythe test also indicated a p-value of $0.5327 > 0.05$ confirming that there is statistically equal variance between the actual and predicted datasets.

Thus, based on the results of the RMSE and MAPE metrics, the Eq. 8 which has the first order, interaction and second order terms was selected as the most appropriate equation for the prediction of liquidus temperature of the aluminum cryolitic bath.

Validation of the best model using experimental data

The next step involves the determination of the final coefficients of the selected model based on the experimental data which are more trusted than the ones provided by the thermodynamic software used in the first step of this research. After the model was fitted to the general form of Eq. 4 the following equation was finally determined:

$$\begin{aligned}
 T_{liq}(^{\circ}\text{C}) = & 1008.95 + 0.054 \text{ xAlF}_3 - 2.485 \text{ xCaF}_2 - 8.94 \text{ xAl}_2\text{O}_3 - 5.26 \text{ xMgF}_2 - 7.22 \text{ LiF} - \\
 & 0.161 \text{ xAlF}_3^2 + 0.821 \text{ xAl}_2\text{O}_3^2 - 0.0036 \text{ xAlF}_3 \text{ xCaF}_2 + 0.175 \text{ xAlF}_3 \text{ xAl}_2\text{O}_3 + \\
 & 0.95 \text{ xAlF}_3 \text{ xKF} - 0.109 \text{ xCaF}_2 \text{ xAl}_2\text{O}_3 + 1.13 \text{ xCaF}_2 \text{ xKF} - 4.5 \text{ xAl}_2\text{O}_3 \text{ xKF} - \\
 & 6.44 \text{ xKF xLiF}
 \end{aligned} \tag{9}$$

The RMSE and MAPE values for the model described by Eq.9 are 2.860°C and 0.205% respectively, whereas the results of the Solheim equation show that the respective values are 7.619°C and 0.528% , which are clearly higher. Thus, the superior capabilities of the proposed model were confirmed also by the experimental data and this model will be used in future studies in order to be compared with



real industrial data. More specifically, as the regression model has an explicit form, it can be easily integrated in industrial control or process simulation systems in order to more accurately represent the liquidus temperature of the bath given its composition. Thus, the regression equation derived in this work could be used in an embedded system for real-time temperature prediction, to dynamically adjust PID parameters e.g. for the regulation of desired electrolytic cell inputs, or in a larger process simulation model in order to get more reliable results about the system state and determine its optimum operating parameters.

Conclusions

In the present work, the evaluation of different regression models for the prediction of liquidus temperature of the aluminum cryolitic bath, with respect to its constituents was carried out. A comprehensive dataset based on thermodynamic simulations was used to train the different models in order to be evaluated and then the best performing model was applied to actual experimental data in order to be validated. Several conclusions were drawn from this study:

Due to the non-linear nature of the correlation between some of the bath constituents and the liquidus temperature, the models including non-linear terms clearly outperformed the multiple linear regression model based on every metric.

Among the tested models, the model using first and second order polynomial terms was found to be the best performing, allowing for a rather low RMSE value of 1.702°C, which was also proven to be lower than the error of an established empirical equation.

Using the selected model, the prediction of the experimental data was also considerably accurate and it was further confirmed that for the range of parameters considered in this study, the proposed model was superior to already established ones, with its parameters reflecting the physical meaning of the correlation between bath constituents and liquidus temperature thus providing a robust and accurate solution to the prediction of liquidus temperature in aluminum cryolitic bath by using a time efficient approach.

Acknowledgments

This research has been implemented within the framework of the National Recovery and Resilience Plan ‘Greece 2.0’ with funding from European Union - NextGenerationEU from European Recovery and Resilience Fund (project code: TAEΔK-06198).

References

- [1] RATVIK, A.P., MOLLAABBASI, R. & ALAMDARI, H. (2022): Aluminum production process: from Hall-Heroult to modern smelters, *ChemTexts*, vol. 8, 10.



- [2] HAUPIN, W. (1991): The influence of additives on Hall-Heroult Bath Properties, *JOM*, vol. 43, pp. 28-34.
- [3] HENGWEI, Y., HUANG, L., MA, W., LIU, Z. & QIN, B. (2021): Alumina solubility and dissolution rate in NaF-KF-AlF₃-LiF-CaF₂ low-temperature electrolyte, *Metall Mater Trans B*, vol. 52, pp. 3490-3496.
- [4] KOULOUMIES, A., BECERRA, A.M., MERLIN, P., PIECHOWIAK, L. & ZAPKE, M. (2020): How to improve the environmental efficiency of the Hall-Heroult process producing and using carbon anodes, in *Light Metals 2020*, San Diego, USA, pp. 1171-1173.
- [5] RAI, V.P. & RAJU, D.D. (2024): Smelting 4.0: Digital strategy for aluminum production, in *Light Metals 2024*, Orlando, USA, pp. 655-663.
- [6] LIU, J., FALLAH-MEHRJARDI, A., SHISHIN, D., JAK, E., DORREEN, M. & TAYLOR, M. (2017): Investigation of the influence of heat balance shifts on the freeze microstructure and composition in aluminum smelting bath system: Cryolite-CaF₂-AlF₃-Al₂O₃, *Metall Mater Trans B*, vol. 48, pp. 3185-3195.
- [7] SOLHEIM, A., ROLSETH, S., SKYBAKMOEN, E., STØEN, L., STERTEN, Å. & STØRE, T. (1996): Liquidus temperatures for primary crystallization of cryolite in molten salt systems of interest for aluminum electrolysis, *Metall Mater Trans B*, vol. 27, pp. 739-744.
- [8] BINGXU, C., PENG, J., WANG, Y. & DI, Y. (2020): Study on liquidus temperature of NaF-KF-LiF-AlF₃ system with low cryolite ratio, *Metall Mater Trans B*, vol. 51, pp. 1181-1189.
- [9] SHI, D., GAO, B.L., WANG, Z.W., SHI, Z.N. & HU, X.W. (2013): Liquidus temperature of electrolytes for aluminum reduction cells, in *Light Metals 2013*, San Antonio, USA, pp. 701-704.
- [10] SOLHEIM, A., ROLSETH, S., SKYBAKMOEN, E., STØEN, L., STERTEN, Å. & STØRE, T. (1995): Liquidus temperature and alumina solubility in the system Na₃AlF₆-AlF₃-LiF-CaF₂-MgF₂, in *Light Metals 1995*, Las Vegas, USA, pp. 73-82.
- [11] PETERSON R.D. & TABEREAUX, A.T. (1987): Liquidus curves for the cryolite – AlF₃ – CaF₂ – Al₂O₃ system in aluminum cell electrolytes, in *Light Metals 1987*, Denver, USA, pp. 33-38.
- [12] YUESHONG, D., JIANPING, P., YUNBIN, B. & NAIXIANG, F. (2013): Liquidus temperatures of Na₃AlF₆ – AlF₃ – CaF₂ – KF – LiF – Al₂O₃ melts, in *Light Metals 2013*, San Antonio, USA, pp. 681-684.
- [13] LU, H., HU, X., CAO, B., CHAI, W. & YAN, F. (2019): Prediction of liquidus temperature for complex electrolyte systems Na₃AlF₆-AlF₃-CaF₂-MgF₂-Al₂O₃-KF-LiF based on the machine learning methods, *Chemometr Intell Lab Syst*, vol. 189, pp. 110-120.
- [14] LU, H., HU, X., CAO, B., MA, L., CHAI, W. & YANG, Y. (2020): Data mining assisted prediction of liquidus temperature for primary crystallization of different electrolyte systems, *Chemometr Intell Lab Syst*, vol. 196, 103885, Jan. 2020.



- [15] MASOUMEH, N., TAFTI, M.G. & NIKDEL, M. (2024): Discovery and primary study of the lithium content in the karst-type bauxite deposit, Jajarm, Iran, *Carbonate Evaporite*, vol. 39, 25.