COMPARISON OF LINEAR REGRESSION MODELS OF THERMOPHYSICAL PROPERTIES WITH MODELS BASED ON MACHINE LEARNING

[1]Mario MACHŮ, [1] Ľubomíra DROZDOVÁ, [1]Bedřich SMETANA

[1]*VSB - Technical University of Ostrava, Ostrava, Czech Republic, EU, mario.machu@vsb.cz*

https://doi.org/10.37904/metal.2022.4451

Abstract

The paper compares classical models for determining the thermophysical properties of steels based primarily on empirical equations derived using linear regression methods with models created using machine learning methods. The selected investigated quantities include phase transformation temperatures, specific heat capacity, coefficient of thermal expansion. The results of both approaches are verified on the measured data by methods of thermal analysis such as differential scanning calorimetry, differential thermal analysis and dilatometry. The methods are evaluated both in terms of the accuracy of predictions and in terms of the adequacy of use for a specific purpose, or in terms of the complexity of creating and using the model.

Keywords: Metallurgy, steel, properties, applications, testing methods

1. INTRODUCTION

Steel is still remaing one of the most important materials in world because of its properties and it is not going to change any soon [1,2]. Thermophysical properties of steel are very important for both manufacturing processes (control of melting, casting, heat treatment processes) both research purposes (modelling of technology operations) [3]. Results in work were obtained by using of a linear regression (liquidus and solidus temperature, specific heat capacity), and two methods of machine learning. Multi-layer perceptron (artificial neural network) was used for predicting of liquidus and solidus temperature and method of decision tree for predicting of specific heat capacity.

1.1. Linear regression

In a number of technical areas, the dependence of the output quantity y on the values of the input variables x is monitored [4]. The sought-after quantity y can also be the thermophysical properties of materials, such as e.g. liquid temperature, solidus or heat capacity. The input variable x is then usually the composition of the material (proportion of individual elements or components), and if the required quantity is temperature dependent, then also the temperature. Determining the specific relationship between composition, temperature and the thermophysical quantity sought is then the task of regression analysis. Regression analysis is one of statistical methods by which we estimate the value of a certain random variable (so-called dependent variables) based on knowledge of other variables (independent variables). The most common type of regression is linear regression, where a dependence in the form of a linear function is assumed, although the arguments of the function may not be linear [5]. The linear regression can be multiple and the searched function then has the form:

$$y=\beta_0+\beta_1 \cdot x_1+\beta_2 \cdot x_2+\beta_3 \cdot x_3+\ldots+\beta_n \cdot x_n \quad (1)$$

where:

y - dependent variable (1)

β_i - coefficient (1)

x_i - independent variable (1).

If a given phenomenon has a nonlinear dependence on one of the quantities, the problem can be linearized and the sought dependence can take on a form combining linear and nonlinear terms:

$$y=\beta_0+\beta_1\cdot(\ln(x_1))+\beta_2\cdot x_2^2+\beta_3\cdot x_3+\ldots+\beta_n\cdot x_n \quad (2)$$

The actual search for βi coefficients is then routinely performed by the least squares method, which minimizes the sum of the squares of the deviations between the measured value and the estimated value [6]. Coefficient estimates can be obtained using regression tools e.g. in Excel or using software such as Matlab, Octave, R or some of the Python libraries (Scikit, Sklearn), which also offer tools for determining the suitability of a given linear model. Of these, the most frequently used coefficient of determination is R^2, the mean squared error MSE (mean squared error), or the square root of the mean squared error *RMSE* (root mean squared error). Furthermore, these software tools allow you to identify variables that do not have a significant impact and can be neglected when looking for dependencies. A description of all regression diagnostics tools that can be applied using the above SW is in [7].

1.2. Machine learning

Machine learning techniques expanded in last decades into almost all fields of science, material engineering is not an exception. Rapid development of algorithms and increasing computational power of computers allows to model even highly non-linearly dependent properties [8-10]

An ANN consists of non-linear basic processing units called neurons. The neuron model and architecture of a neural network describe how a network transforms its inputs into outputs. The neural network architecture consists of multiple layers of neurons which have a summing up junction and a transfer function. A single neuron transmits an input p through the connection that multiplies its strength by the weight w to form a product wp. A bias b is then applied - it is much like a weight with constant value of 1 but can be omitted. The transfer function then produces the neuron output Y using the product wp and bias. There are various transfer functions, most commonly used are sigmoid and linear. The central idea of an ANN is to adjust weights and biases or the network itself adjusts these parameters to achieve accurate results - desired output values [11-13].

Decision tree learning or induction of decision trees is one of the predictive modelling approaches used in statistics, data mining and machine learning. It uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees. Decision trees are among the most popular machine learning algorithms given their intelligibility and simplicity [14-16].

2. METHODOLOGY

Over 150 alloyed steel grades were used for obtaining linear regression equations. The same data set was used for training, validating and testing of ANN. These element contents represent inputs to the model. Output of model is a value of T_L and T_S. The chemical composition of the steels used in modelling is summarized in **Table 1**. Source of these data is [17]. **Table 2** shows range of 37 steel grades used for validation of obtained models. Also linear regression equations for specific heat capacity were obtained based on this dataset which properties were described in [18].

Table 1 The composition of steels used for the creation of the ANN

Content (wt%)	C	Mn	Si	P	S	Cu	Ni	Cr	N	Mo	V	Fe
minimum	0.01	0.02	0.12	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	bal.
maximum	1.20	1.63	2.07	0.04	0.04	0.2	3.30	5.00	0.02	0.99	0.14	bal.
mean	0.46	0.72	0.43	0.01	0.01	0.04	0.54	0.92	0.01	0.16	0.01	bal.

Table 2 The composition of steels used for evaluating of the ANN models and obtaining equations for c_p

Content (wt%)	C	Mn	Si	P	S	Cu	Ni	Cr	N	Mo	V	Fe
minimum	0.07	0.32	0.17	0.01	0.00	0.03	0.01	0.03	0.00	0.00	0.00	bal.
maximum	1.03	1.43	0.94	0.02	0.01	0.15	2.34	5.00	0.01	1.22	0.92	bal.
mean	0.35	0.85	0.32	0.01	0.00	0.08	0.22	0.93	0.01	0.18	0.09	bal.

3. RESULTS

3.1. Temperature of liquidus

Using linear least squares regression, described previously, the liquid temperature equation for Fe-C-O-Cr, Fe-C-O-Ni and Fe-C-O-Cr-Ni alloys was derived using SW Matlab. The experimentally obtained liquidus temperatures of 36 alloys were measured by direct thermal analysis. The obtained equation has the form:

$$T_L = 1543.5 - 75.068 \cdot wtC - 3{,}456 \cdot wtCr - 2.904 \cdot wtNi + 16.580 \cdot wtO - 67.876 \cdot wtMn + 1671.400 \cdot wtSi + 2120.900 \cdot wtP - 150.610 \cdot wtS - 421.450 \cdot wtAl - 1261.700 \cdot wtCu + 202.740 \cdot wtCo - 22.839 \cdot wtN \quad (3)$$

where:

wt - mass fraction (%)

Figure 1 shows a comparison of the experimentally obtained liquidus temperatures and the temperatures obtained using the new equation. The coefficient of determination is 0.99, which means a very good agreement. The standard deviation of the measured and calculated values is 3.36 ° C.

3.2. Temperature of solidus

Using same procedure as for liquidus temperature, the temperature of solidus was derived. The experimentally obtained solidus temperatures of 36 alloys were measured by direct themal analysis. The obtained equation has the form:

$$T_L = 1539.8 - 183.3 \cdot wtC - 4.6 \cdot wtCr - 4.8 \cdot wtNi - 9.7 \cdot wtMn - 48.1 \cdot wtSi - 99.4 \cdot wtP - 934.1 \cdot wtS - -109.9 \cdot wtCu - 16.9 \cdot wtV + 7.4 \cdot wtMo \quad (4)$$

Figure 2 shows a comparison of the experimentally obtained solidus temperatures and the temperatures obtained using the new equation. The coefficient of determination is 0.945, which means a very good agreement. The standard deviation of the measured and calculated values is 11.1 ° C.

3.3. Specific heat capacity

Two equations for the heat capacity calculation were derived using least squares linear regression. They were divided into an area from 30 to 650 ° C and from 850 to 1450 ° C. For simplicity, phase transformations were not included in the calculation. The use of neural networks would be more appropriate to describe the course of heat capacity in the whole range, ie from 30 to 1580 ° C.

Equation for temperature interval of 30 - 650 °C:

$$c_p = 0.366 + 0.001 \cdot T + 0.021 \cdot wtC + 0.001 \cdot wtCr - 0.002 \cdot wtNi + 0.535 \cdot wtO + 0.499 \cdot wtMn - 0.855 \cdot wtP - 0.064 \cdot wtS + 1.482 \cdot wtAl + 0.890 \cdot wtCu \quad (J \cdot K^{-1} \cdot g^{-1}) \quad (5)$$

Equation for temperature interval of 840 - 1 450 °C:

$$c_p = 0.829 + 0.063 \cdot wtC - 0.009 \cdot wtCr - 0.005 \cdot wtNi - 0.923 \cdot wtO - 1.202 \cdot wtMn + 5.740 \cdot wtP - 0.373 \cdot wtS - 1.550 \cdot wtAl - 1.042 \cdot wtCu \ (J \cdot K^{-1} \cdot g^{-1}) \quad (6)$$

Figure 3 shows a comparison of experimentally obtained heat capacities and values obtained using the new equation. The coefficient of determination is 0.99, which means a very good agreement. The standard deviation of the measured and calculated values is 0.01 $J \cdot K^{-1} \cdot g^{-1}$.

Figure 4 shows that there is less agreement between the measured and calculated values than in the case of up to 650 °C. The coefficient of determination is 0.88. The standard deviation is 0.05 $J \cdot K^{-1} \cdot g^{-1}$. Less agreement may be due to the fact that up to 900 °C heat capacities could still be affected by the course of phase transformations they could also be affected by the dissolution of carbides in the range of 900 - 1000 ° C. The fact that the heat capacities were measured up to 700 °C on the Setaram Sensys Evo TG / DSC and from 700 to 1580 °C on the Setaram MHTC 96 Line can also play a role.

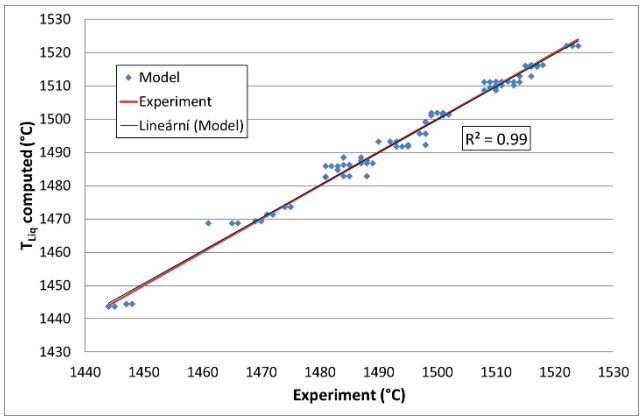

Figure 1 Liquidus temperature - model vs. measurement

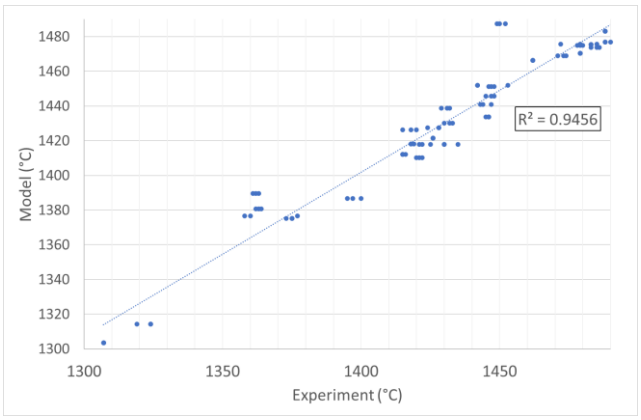

Figure 2 Solidus temperature - model vs. measurement

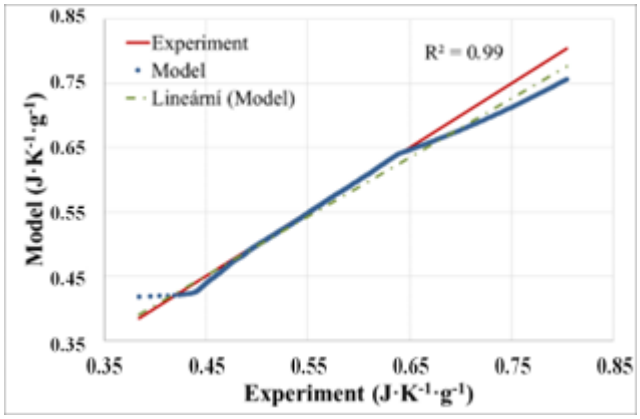

Figure 3 c_p - model vs. measurement (30 - 650 °C)

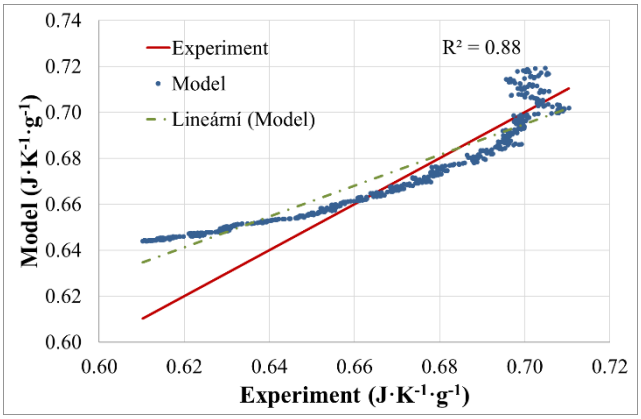

Figure 4 c_p - model vs. measurement (840 - 1 450 °C)

Table 3 shows results obtained by linear regression compared with results obtained by machine learning methods in previous work [19-21]. As for liquidus and solidus temperature, both methods seem to be suitable but linear regression slightly overperforms ANN. Usability of derived equations is strictly limited by the range of chemical composition of steel grades used. As for specific heat capacity, linear regression fails in giving reasonable values near the temperatures of phase changes (both in low- and high-temperature area) and using of decision tree modelling gives more accurate results in whole range of temperatures.

Table 3 Comparison of results obtained by different methods

	R^2 (1)	
	Linear regression	ANN
Liquidus temperature	0.99	0.92
Solidus temperature	0.95	0.94
Specific heat capacity	0.99/0.88	0.98

4. CONCLUSION

Paper shows possibility of using classic approach of modelling thermo-physical properties, on example of temperatures of liquidus and solidus and specific heat capacity. The method of linear regression is suitable for phase change temperature predictions but fails in describing continuous property dependent not only on composition but also on temperature. Reason is non-linearity in course of the specific heat capacity dependency on temperature during phase changes. For predicting properties dependent on temperature, the method of decision tree from broad family of machine learning methods seems to be much more suitable. As for the multi-layer perceptron artificial neural network method, it seems to be good for predicting temperatures of phase changes, but it is slightly overperformed by linear regression model both for liquidus and solidus temperature. One must be careful when using obtained equations and avoid using them outside of the validity range - outside of the chemical composition range mentioned in the paper.

ACKNOWLEDGEMENTS

This paper was created within the frame of the project No. CZ.02.1.01/0.0/0.0/17_049/0008399 from the EU and CR financial funds provided by the Operational Programme Research, Development and Education, Call 02_17_049 Long-Term Intersectoral Cooperation for ITI, Managing Authority: Czech Republic - Ministry of Education, Youth and Sports, and student project SP2022/39, SP2022/13.

REFERENCES

[1] MITRA A. Use of Steel as a Sustainable Concept. In: Hashmi S, Choudhury IA, editors. *Encyclopedia of Renewable and Sustainable Materials*, vol. 4. Amsterdam: Elsevier; 2020.

[2] SUN, W., WANG, Q., ZHOU, Y., WU, Y. Material and energy flows of the iron and steel industry: Status quo, challenges and perspectives. *Applied Energy*. 2020, vol. 268, 114946.

[3] GUR, C.H., PAN, J., editors. *Handbook of Thermal Process Modeling Steels*. 1st ed. Boca Raton: CRC Press; 2008.

[4] MELOUN, M., MILITKÝ, J. *Statistická analýza experimentálních dat*. Praha: Academia, 2004.

[5] De FAOITE, D., BROWNE, D.J., STANTON, K.T. Regression analysis of temperature-dependent mechanical and thermal properties of dielectric technical ceramics. *Journal of Materials Science*. 2013, vol. 48, no. 1, pp. 451-461.

[6] DORESWAMY, HEMANTH, K.S., MANOHAR, M.G. Linear Regression Model for Knowledge Discovery in Engineering Materials. **In:** *First International Conference on Artificial Intelligence, Soft Computing and Applications 2011*.

[7] GÉRON, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems*. Second edition. Sebastopol: O'Reilly, 2019. ISBN 978-1-492-03264-9.

[8] SCHMIDT, J., MARQUES, M.R.G., BOTTI, S. *et al.* Recent advances and applications of machine learning in solid-state materials science. *npj Computation Materials.* 2019, vol.5, no. 1, p. 83.

[9] NISBET, R., ELDER, J.E., MINER, G. *Handbook of statistical analysis and data mining applications*. Burlington: Elsevier/Academic Press, 2009.

[10] LÓPEZ-MARTÍNEZ, E.r et al. Artificial Neural Networks to Estimate the Thermal Properties of an Experimental Micro-Alloyed Steel and Their Application to the Welding Thermal Analysis. *Strojniški vestnik - Journal of Mechanical Engineering*. 2015, vol. 61, no. 12, pp. 741-750.

[11] PEET, M.J., HASAN, H.S., BHADESHIA, H.K.D.H. Prediction of thermal conductivity of steel. *International Journal of Heat and Mass Transfer*. 2011, vol. 54, no. 11-12, pp. 2602-2608.

[12] ŽMAK, I., FILETIN, T. Predicting thermal conductivity of steels using artificial neural networks. = Predicting thermal conductivity of steels using artificial neural networks. *Transactions of FAMENA*. 2010, vol. 34, no. 3, pp. 11-20.

[13] SUN, Y., TU Y., SUN, J., NIU, D. Effect of temperature and composition on thermal properties of carbon steel. *In: Chinese Control and Decision Conference*. [online]. 2009, pp. 3756-3760.

[14] BHOWMIK, R., SIHN, S., PACHTER, R., VERNON, J.P. Prediction of the specific heat of polymers from experimental data and machine learning methods. *Polymer*. 2021, vol. 220:123558.

[15] WU, X., KUMAR, V., ROSS QUINLAN, J. *et al.* Top 10 algorithms in data mining. *Knowledge and Information Systems.* 2008, vol. 14, pp**.** 1-37.

[16] WAGNER, N., RONDINELLI, J.M. Theory-Guided Machine Learning in Materials Science. *Frontiers in Materials*. 2016, vol. 3:28.

[17] MIETTINEN, J. HOWE, A. Estimation of liquidus temperatures for steels using a thermodynamic approach. *Ironmaking and Steelmaking*. 2000, vol. 27, no. 3, pp. 212-227.

[18] DROZDOVÁ, Ľ., SMETANA, B., PRESOLY, P., NOVÁK, V., MACHŮ, M., BERNHARD, M., et al. Investigation of Fe-C-Cr and Fe-C-Cr-Ni-based systems with the use of DTA and HT-LSCM methods. *Journal of Thermal Analysis and Calorimetry*. 2020, vol. 142, no. 2, pp. 535-546.

[19] MACHŮ, M., DROZDOVÁ, L., SMETANA, B., ZIMNÝ, O., VLČEK, J. Estimation of liquidus temperatures of steel using artificial neural network approach In: *METAL 2018 - 27th International Conference on Metallurgy and Materials, Conference Proceedings*. Brno: Tanger, 2018, pp. 56-62.

[20] MACHŮ, M., DROZDOVÁ, Ľ., SMETANA, B., ZLÁ, S., KAWULOKOVÁ, M. Artifical neural network usage for determining solidus temperature of steels. In: *METAL 2019: conference proceedings : peer reviewed : 28th International Conference on Metallurgy and Materials : May 22nd-24th 2019, Hotel Voronez I, Brno, Czech Republic, EU*. Ostrava: Tanger, 2019. pp. 48-53.

[21] MACHŮ, M., DROZDOVÁ, Ľ., SMETANA, B. On methods for determing the specific heat capacity of steels. In: METAL 2021: Proceedings 30th Anniversary International Conference on Metallurgy and Materials: May 26th-28nd 2021, Brno, Czech Republic. Ostrava: Tanger, 2021. p. 77-82.