AIP Advances

Data-driven design of glasses with desirable optical properties using statistical regression

Cite as: AIP Advances **10**, 105110 (2020); https://doi.org/10.1063/5.0022451
Submitted: 22 July 2020 • Accepted: 13 September 2020 • Published Online: 05 October 2020

Yomei Tokuda, Misa Fujisawa, Daniel M. Packwood, et al.

Data-driven design of glasses with desirable optical properties using statistical regression

Cite as: AIP Advances 10, 105110 (2020); doi: 10.1063/5.0022451
Submitted: 22 July 2020 • Accepted: 13 September 2020 •
Published Online: 5 October 2020

Yomei Tokuda,[1,a] Misa Fujisawa,[1,b] Daniel M. Packwood,[2,c] Mei Kambayashi,[1,d] and Yoshikatsu Ueda[3,e]

AFFILIATIONS

[1] Faculty of Education, Shiga University, 2-5-1, Hiratsu, Otsu City, Shiga Prefecture, Japan
[2] Institute for Integrated Cell-Material Sciences (iCeMS), Kyoto University, Sakyo-ku, Kyoto Prefecture, Japan
[3] Research Institute for Sustainable Humanosphere, Kyoto University, Uji City, Kyoto Prefecture, Japan

[a] **Author to whom correspondence should be addressed:** tokuda@edu.shiga-u.ac.jp
[b] s1019724@st.shiga-u.ac.jp
[c] packwood.danielmiles.6z@kyoto-u.ac.jp
[d] bsk.sole.12@ezweb.ne.jp
[e] ueda.yoshikatsu.4e@kyoto-u.ac.jp

ABSTRACT

In this study, we used a data-driven approach to build models for assisting the design of new glasses with high refractive index and low dispersion. Our models, which are based on multiple linear regression and kernel ridge regression, achieved high accuracy in predicting optical properties of glasses based on their composition alone. Using the predictions of these models as a guide, we fabricated new glasses in our laboratory. In agreement with model predictions, these glasses had promising optical properties. This work therefore demonstrates a successful example of data-driven materials design and can be used as a template for designing glasses or other materials with other desirable properties.

I. INTRODUCTION

Most commercial glasses consist of a multi-component mixture of oxides. The composition of this mixture can have dramatic effects on physical properties,[1] and trial-and-error attempts at optimizing composition to obtain desirable properties are costly and have a large environmental load because of experimental waste. In order to accelerate the discovery of new glasses, a method that can quickly optimize composition by elucidating the complicated relationship between composition and properties is desirable.

Among the many properties of glasses, refractive index and dispersion are particularly important because both must be tuned to create optical lenses.[2] The dispersion is defined as the Abbe number

$$\nu_D = \frac{n_d - 1}{n_F - n_C}, \tag{1}$$

where n_F, n_d, and n_C are the refractive indices at wavelengths of 486.1 nm, 587.6 nm, and 656.3 nm, respectively.[3] A large Abbe number corresponds to small dispersion. High refractive index and a large Abbe number are therefore a desirable property for optical lenses with less chromatic aberration. A plot of refractive index as a function of the Abbe number is referred to as an Abbe diagram (Fig. 1). However, according to the Abbe diagram, refractive index tends to decrease non-linearly as the Abbe number is increased. This trend is not favorable for making optical lenses, which require high refractive index and low dispersion. While several reports have proposed novel glasses with a large refractive index and large Abbe number,[4,5] the discovery of these glasses via a purely experimental approach can be time-consuming.

A recent approach to materials design involves a type of machine learning, namely, statistical regression of big data.[6–13] Other methods such as neural networks and random forests have also been used for this purpose.[14–17] A recent work in this direction used deep learning (a highly nonlinear statistical regression) to predict refractive index.[18] Statistical regression involves building models, which capture the relationship between material property and

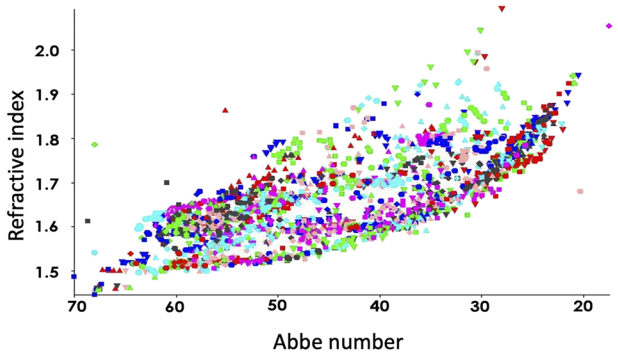

FIG. 1. Abbe number ν_D as a function of the refractive index n_d for the glass containing SiO_2 derived from the glass database "INTERGLAD." Each point corresponds to one kind of glass.

composition.[9,18,19] Arguably, the simplest such model that can be learned from data is the multiple linear regression (MLR) model,[20,21]

$$y_i = w_0 + w_1 x_i + w_2 x_i^2 + \cdots = w_0 + \sum_k w_k x_i^k, \tag{2}$$

where i denotes the ith glass in the dataset and y_i and x_i correspond to the property of interest and the amount of a specific component oxide in the glass, respectively. This model, which is called "linear" because it is linear in the coefficients w, is appropriate for glasses due to the continuous variation of property with composition. It can be easily generalized to the case of multiple components. In some cases, however, the property of interest cannot be easily fit to equations of the type in (2). In such cases, a more flexible model such as the kernel ridge regression (KRR) model is desirable.[22] In a KRR model, a potentially infinite number of basis functions are used to define the relationship between x and y. In principle, this model is able to describe any property–composition relationship of any complexity. While such a model may seem impractical at first, in practice, the basis functions can be handled indirectly using the so-called kernel trick.[23]

In this paper, we prepare new glasses with large Abbe numbers using guidelines deduced from statistical regression of glass data. Concretely, we use the glass dataset to build MLR and KRR models for predicting the Abbe number from glass composition, both of which display strong predictive power during cross-validation (CV). Using these models as a guide, we then experimentally dope a glass with oxides that are predicted to increase and decrease the Abbe number. Experimental characterization of the doped glasses confirms the predictions of our models, demonstrating how statistical regression can be used to guide the design of new materials with targeted properties.

II. METHODS

A. Multiple linear regression (MLR) model

The composition of glass 1 from the database can be expressed as

$$X_1 = x_{1,SiO_2} + x_{1,B_2O_3} + \cdots = \sum_j^m x_{1,j} = 100, \tag{3}$$

where x_{1,SiO_2} and so on indicate the fraction of SiO_2 in glass 1 and m is the number of possible components. The "100" appearing on the right side of Eq. (3) means that the fraction is expressed as a mole percentage. Assuming a MLR model, the Abbe number for glass 1 can be written as

$$y_1 = w_0 + w_{SiO_2} x_{1,SiO_2} + w_{1,B_2O_3} x_{1,B_2O_3} + \cdots = w_0 + \sum_j^m w_j x_{1,j}. \tag{4}$$

The linear relationship between the Abbe number and glass composition is clearly problematic. We therefore generalize this expression by introducing polynomial functions as follows:

$$y_1 = w_0 + \sum_j^m \left(w_{j,SiO_2} x_{1,SiO_2} + w'_{j,SiO_2} x_{1,SiO_2}^2 + w''_{j,SiO_2} x_{1,SiO_2}^3 + \cdots \right), \tag{5}$$

where w, w', and w'' are the coefficients for first, second, and third order terms, respectively. Using the matrix notations

$$\mathbf{\Phi} = \begin{pmatrix} 1 & x_{1,SiO_2} & x_{1,SiO_2}^2 & \cdots \\ 1 & x_{2,SiO_2} & x_{2,SiO_2}^2 & \cdots \\ \vdots & \vdots & \vdots & \cdots \\ 1 & x_{N,SiO_2} & x_{N,SiO_2}^2 & \cdots \end{pmatrix}, \tag{6}$$

$$\boldsymbol{w} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_m \end{pmatrix}, \tag{7}$$

where N is the number of glasses in the dataset and x_{ij} refers to the amount of component j in glass i, Eq. (5) for N glasses can be rewritten as

$$\boldsymbol{y} = \mathbf{\Phi} \boldsymbol{w}. \tag{8}$$

This model is also known as a polynomial regression.

A simple choice of error function is given by the sum of squares of errors between the predicted value **y** for the training data value **t** and the regularization term σ to avoid overfitting,

$$E = (\boldsymbol{t} - \mathbf{\Phi} \boldsymbol{w})^2 + \sigma \boldsymbol{w}^T \cdot \boldsymbol{w}. \tag{9}$$

The second term is known as a "ridge." To minimize E, Eq. (9) is differentiated with respect to **w**, yielding

$$\boldsymbol{w} = \left(\mathbf{\Phi}^T \mathbf{\Phi} + \sigma \mathbf{I} \right)^{-1} \mathbf{\Phi}^T \mathbf{y}, \tag{10}$$

where **I** is the identity matrix. Once **w** is calculated from (10), Eq. (8) can be used to make predictions.

In this MLR model, the selection of the order of Eq. (6) is important. To determine how predictive performance depends upon polynomial order, we use cross-validation (CV). In CV, the data are divided into S sets. First, $(S - 1)$ of the sets were used as training data for Eq. (10), and the coefficient of determination, R^2, was calculated for the remaining set (test data). This operation was repeated S times, with a different set used for the test data each time. Afterward, the R^2 values obtained from each iteration were averaged. CV was repeated for linear, quadratic, and cubic orders. In this study,

we select $S = 5$ for our training set of N glasses, where N = 879 was selected as described in Sec. III A. All codes were written in the R programming language[24] and executed within RStudio.[25]

B. Kernel ridge regression model (KRR model)

For the case of a nonlinear relationship between x and y, a linear model as in Eq. (8) is not suitable. To overcome this situation, several more flexible approaches such as regression trees, spline smoothing, and kernel ridge regression (KRR) have been proposed.[23] Among them, KRR is particularly versatile as it can be adopted to many kinds of problems provided a suitable kernel function, and hence, it is selected in the present study.

Suppose that y is represented by a potentially infinite number of hidden basis functions,

$$y_i = w_0 + w_1\psi_1(x_i) + w_2\psi_2(x_i) + \cdots = \sum_k w_k\psi_k(x_i) = \boldsymbol{w}^T\boldsymbol{\Psi}(x_i), \quad (11)$$

where $\psi_k(x_i)$, $\boldsymbol{\Psi}(x_i)$, and $\mathbf{w}^{\mathrm{T}}$ are the kth basis function, a column vector of the basis functions, and a row vector of weights w_i, respectively. In KRR, the basis functions are not calculated directly. Rather, the covariance kernel $\boldsymbol{K}$,[21] defined as

$$\boldsymbol{K}(x_i, x_j) = \boldsymbol{\Psi}(x_i)\boldsymbol{\Psi}(x_i)^{\mathrm{T}}, \quad (12)$$

is assumed to be known or can be approximated. Given a covariance kernel, we can make predictions for other glasses outside of our dataset. In this work, we adopt the kernel function[26] as

$$k(x_1, x_2) = \exp\left(-\beta|x_1 - x_2|^2\right), \quad (13)$$

where

$$|x_1 - x_2|^2 = \left(x_{1,\mathrm{SiO_2}} - x_{2,\mathrm{SiO_2}}\right)^2 + \left(x_{1,\mathrm{B_2O_3}} - x_{2,\mathrm{B_2O_3}}\right)^2 + \cdots + \left(x_{1,m} - x_{2,m}\right)^2, \quad (14)$$

comparing glasses 1 and 2. Here, β is a constant (called a hyperparameter) and $|x_1 - x_2|$ is the dissimilarity between the two glass compositions. A small dissimilarity indicates that the two glasses have similar composition. Equation (13) is a popular choice of kernel for KRR models; however, other types of kernels can be used as well.

In KRR, the error for glass k can be defined as

$$E_k = \left(y_k - \sum_{i=1}^{N} \alpha_i k(x_i, x_k)\right)^2, \quad (15)$$

where the values of α_i can be obtained by minimizing the right-hand side of Eq. (15).[27] To avoid overfitting, a regularization term is included as follows:

$$\boldsymbol{\alpha} = (\boldsymbol{K} + \lambda\boldsymbol{I})^{-1}\boldsymbol{y}, \quad (16)$$

where $\boldsymbol{I}$ and λ are an identity matrix and another hyperparameter, respectively.[27] $\mathbf{K}$ is the so-called kernel matrix,

$$\boldsymbol{K} = \begin{pmatrix} k(x_1, x_1) & k(x_1, x_2) & & k(x_1, x_N) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_N) \\ \vdots & \vdots & & \vdots \\ k(x_N, x_1) & k(x_N, x_2) & & k(x_N, x_N) \end{pmatrix}. \quad (17)$$

As in the case of MLR, $\lambda\boldsymbol{I}$ is known as a "ridge." In the context of statistical regression, λ represents the noise on the observed value.[21] The predicted value $\mathbf{y}^*$ for new $\mathbf{x}^*$ can then be calculated as

$$\boldsymbol{y}^* = \boldsymbol{k}^T\boldsymbol{\alpha}\boldsymbol{t}, \quad (18)$$

where $\boldsymbol{k}$ and $\boldsymbol{t}$ denote a vector with the elements $k(x_i, x^*)$ and that with the elements t_i.[27] The representation of products of basis functions using the kernel function [as in Eq. (12)] is known as a kernel trick.[23]

As with the case of the MLR model, CV was used to determine the performance of the KRR model and for selecting the values of the hyperparameters β and λ. The hyperparameters β and λ were then selected as those that provide the largest value of R^2. We again select S = 5 for 879 glasses. The Gaussian kernel was generated using the R package kernlab.[28]

C. Databases

Statistical regression requires databases in order to build models. Decades of studies on the optical properties of glasses have resulted in considerable optical data scattered throughout the glass literature. A particularly comprehensive database for glass properties is called INTERGLAD and is commercially available from New Glass Forum, Japan.[29] This database contains all the data from the glass literature and patents. As in 2020, it contains ~350 000 glasses with their reported properties. The database is updated each year. In the following, we consider the subset of glasses from INTERGLAD, which contain SiO_2 as one of the components, because practical glasses in use contain SiO_2 in most cases.

D. Experimental

A non-doped (ND) glass was selected as the composition in mol. % of La_2O_3:Nb_2O_5:Ta_2O_5:GeO_2:B_2O_3:ZrO_2:Al_2O_3:TiO_2:SiO_2:Sb_2O_3 = 21.18:16.94:4.23:3.29:19.36:9.65:5.6:4.83:4.86:0.06. The refractive index and Abbe number were reported as 1.9995 and 28.6, respectively.[30] This glass is a good starting-point for materials design because it has both high refractive index and low Abbe number.

Several kinds of oxides, namely, CaO, Ga_2O_3, La_2O_3, Nb_2O_5, Sc_2O_3, SrO, Ta_2O_5, and Yb_2O_3, were added at the level of 3 mol. % or 5 mol. % to the ND glass above. All the glass samples were prepared by the melting method, using reagent grade chemicals (FUJIFILM Wako Pure Chemical Corporation). Predetermined quantities of the starting materials were mixed and melted in a Pt crucible for 1 h at a temperature of 1400 °C. The melt was poured onto stainless steel at a temperature of 300 °C. The glass obtained was annealed below the glass transition temperature for 10 min. Both sides of the glass were polished before measuring optical properties.

For the purposes of optical characterization, refractive indices of the glass plates were measured using a Metricon 2010 prism coupler at wavelengths L of 463 nm, 633 nm, 1313 nm, and 1517 nm. The refractive index can be approximated using Cauchy's equation

$$n = A + \frac{B}{L^2} + \frac{C}{L^4}. \quad (19)$$

First, the refractive index was plotted against 1/(wavelength)^2. Thereafter, a quadratic regression was performed to determine the

constants A, B, and C. Refractive indices at wavelengths of 486 nm, 587 nm, and 656 nm were then determined, and the Abbe number was calculated using Eq. (1).

III. RESULTS AND DISCUSSION

A. Glasses

From the INTERGLAD database, more than 10 000 SiO_2-including glasses were found in which both the refractive indices at 587 nm and Abbe numbers were reported. It was observed that these glasses contained up to $m = 56$ different kinds of oxides. Figure 1 shows the refractive index n_d plotted as a function of the Abbe number ν_D for these glasses including SiO_2. As pointed out earlier, glasses with a large refractive index tend to show small Abbe numbers. In this study, we focused on the 879 glasses with high refractive index between 1.9 and 2.0 because high refractive index is desirable for optical glass.

B. Predictions from the MLR model

The MLR model was fit using these 879 glasses. As shown in Eq. (3), the inputs were the molar fraction of each component and outputs were the Abbe numbers for each glass. For the case of the third order polynomial, the averaged R^2 value resulting from cross-validation was quite low at 0.64. Figure S1 shows the value of the averaged R^2 for the first and second order polynomials plotted against σ in Eq. (9). This result indicates better fitting performance for the second order polynomial compared to the first order polynomial. For the second order polynomial, we obtain $\sigma = 0.2$ and an averaged R^2 value of 0.944. We therefore decided to use a second order polynomial with this value of α for subsequent prediction of glass properties. Figure 2 shows the best fitting for polynomial functions for the 879 glasses; (a) first, (b) second, and (c) third orders.

As pointed out above, the glasses we consider contain up to 56 kinds of oxides. To determine which oxide tends to increase the Abbe number, we proceed as follows. Suppose we wish to determine the effect of La_2O_3. We first add 5 mol. % La_2O_3 to the ND glass, making the sum of the composition percentages become 105 mol. %. The percentage of every component is then divided by 100/105 to ensure that we obtain a new glass with compositions summing to 100 mol. %. After this, the Abbe number was estimated by Eq. (8). This calculation allows us to predict how the Abbe number is affected by the addition of the new component of 5 mol. % La_2O_3. Predictions for various added components are tabulated in Table I. The data for all the added oxides are also tabulated in Table S1. As demonstrated, the ten components that are predicted to most strongly increase the Abbe number are Sc_2O_3 followed by As_2O_3, SnO_2, Al_2O_3, Yb_2O_3, Ga_2O_3, SiO_2, Y_2O_3, Gd_2O_3, and La_2O_3. These predictions can be used to guide the development of new glasses, as demonstrated later.

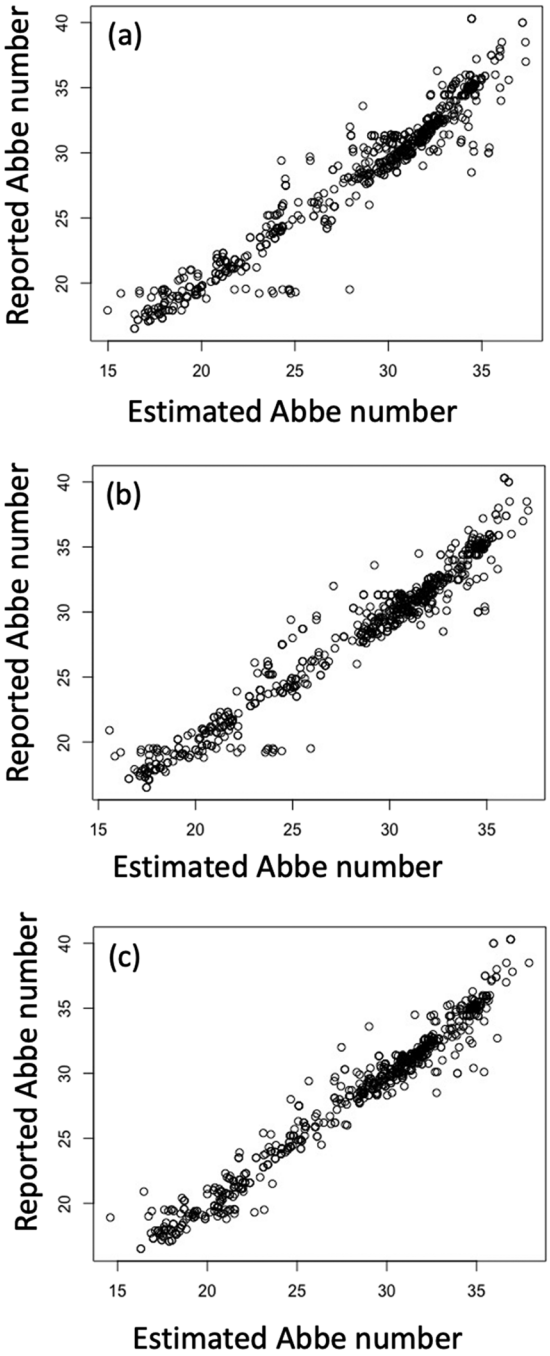

FIG. 2. Estimated Abbe number ν_D from the MLR models compared to reported Abbe numbers for 879 glasses. MLR models using (a) first order, (b) second order, and (c) third order polynomials.

C. Predictions from the KRR model

We also fitted the KRR model using the 879 glasses in our dataset. Cross-validation showed that the hyperparameter values that achieved the highest R^2 value were 0.0001 and 0.0004 for β and λ, respectively (Fig. S2). Figure 3 shows the reported Abbe numbers plotted against the ones predicted by the KRR model using these hyperparameters. The result clearly shows that the KRR model predictions have good generalization performance. With an averaged

TABLE I. Abbe number ν_D estimated by the MLR and KRR models for selected oxides. The second and fourth columns are the estimated Abbe number by using the MLR and KRR models, respectively. The third and fifth columns are the rank in descending order. The data for all the glasses are tabulated in Table S1.

Additive oxide	Ridge regression 5%	Order by ridge regression	GP regression 5%	Order by GP
Non-doped	28.8	18	28.8	39
Al_2O_3	31.5	4	29.2	11
B_2O_3	29.2	13	28.9	13
CaO	28.9	14	30.0	2
Ga_2O_3	30.1	6	29.7	4
K_2O	27.9	25	28.9	15
La_2O_3	29.3	10	29.4	9
Li_2O	29.2	12	29.4	7
Na_2O	27.7	28	30.0	3
Nb_2O_5	27.6	29	26.7	56
Sc_2O_3	51.5	1	28.0	49
SiO_2	29.5	7	28.8	17
SrO	27.6	30	28.9	14
Ta_2O_5	28.3	23	26.3	57
TiO_2	27.7	27	27.0	54
Y_2O_3	29.5	8	29.1	12
Yb_2O_3	31.0	5	30.4	1

R^2 value of 0.996, the KRR model out-performed ridge regression, as shown in Fig. 3.

As with the case of the MLR model, we identified the oxides that tend to increase the Abbe number in the KRR model. The predicted values are tabulated in Tables I and S1. The KRR model predicts the top ten components for increasing the Abbe number to be Yb_2O_3, followed by CaO, Na_2O, Ga_2O_3, GeO_2, BaO, Li_2O, Gd_2O_3, La_2O_3, and ZrO_2. In both the MLR and KRR models, Yb_2O_3, Ga_2O_3, Gd_2O_3, and La_2O_3 are predicted to enhance the Abbe number. However, the MLR model predicts that Sc_2O_3, As_2O_3, and SnO_2 are more effective for enhancing the Abbe number compared to KRR. The two sets of predictions, from the respective MLR and KRR models, should therefore be treated as complementary for the purpose of designing new glasses.

D. Experimental synthesis of new glasses based on machine-learning predictions

Using the predictions described above, we synthesized non-doped and new glasses and measured their optical properties. New glasses were created by doping CaO, Ga_2O_3, La_2O_3, Nb_2O_5, Sc_2O_3, SrO, Ta_2O_5, or Yb_2O_3 to the ND glass. The materials vitrified at a doping level of 5 mol. % for all additives, except for the cases of Sc_2O_3 and Yb_2O_3. Sc_2O_3 and Yb_2O_3 were instead doped at a level of 3 mol. % because of less vitrification ability.

The refractive index of the ND glass was 1.998, which was reported as 1.9995.[30] The discrepancy between the reported and experimental values is small. Figure 4 shows the refractive index plotted against $1/\lambda^2$ for the ND glass. The curve was fit by Eq. (19) to obtain the constants A, B, and C. The refractive indices n_F, n_d, and n_c were calculated to estimate the Abbe number. The refractive index at 587.6 nm and Abbe number for these glasses are tabulated in Table II. Because the refractive indices were almost identical for the new glasses, we ignore them in the following discussion. The experimental Abbe number for the ND glass was 26.2, which was smaller than the reported value of 28.6. The discrepancy between

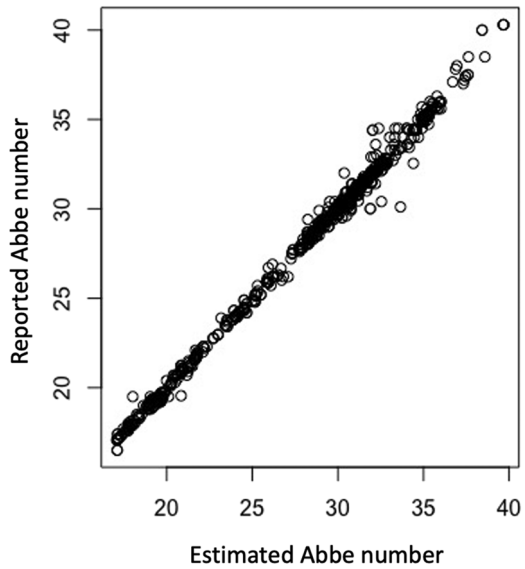

FIG. 3. Estimated Abbe number ν_D from the KRR model compared to reported Abbe numbers for 879 glasses with β of 0.0001 and λ of 0.0004. R^2 was 0.998. The hyperparameters β and λ were selected as those that provide the largest value of R^2 in CV.

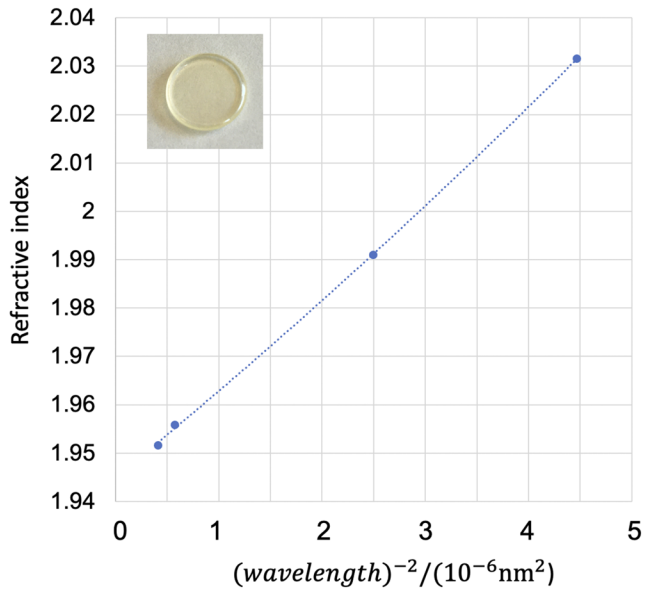

FIG. 4. Refractive index plotted as a function of (*wavelength*)$^{-2}$ for the non-doped glass. The dotted line is a quadratic approximation.

TABLE II. Measured values of the refractive index at 588 nm and Abbe number v_D for the estimated glasses.

	Doped level (%)	Refractive index	Abbe number	Variation of Abbe number[a]
Non-doped	0	1.998	26.2	...
CaO	5	1.991	26.9	+0.7
Ga_2O_3	5	1.987	27.4	+1.2
La_2O_3	5	1.998	29.4	+3.2
Nb_2O_5	5	2.017	24.3	−1.9
Sc_2O_3	3	2.004	29.6	+3.4
SrO	5	1.988	27.0	+0.8
Ta_2O_5	5	2.024	25.7	−0.5
Yb_2O_3	3	1.995	27.1	+0.9

[a]This means the difference in the Abbe number between the non-doped and doped glasses.

the reported and experimental values is not insignificant. We believe that this discrepancy originated from the use of Eq. (19) instead of Sellmeier's dispersion formula. Sellmeier's dispersion formula usually requires more parameters than considered in this study because of absorption. Therefore, we focused on the difference in the Abbe number between the ND and doped glasses, which were prepared in this study. The Abbe numbers for the glasses doped with Yb_2O_3 (3%), Ga_2O_3, and La_2O_3 were 27.1, 27.4, and 29.4, respectively, which were larger than that of ND glass experimentally observed in the present study. In the case of these dopants, both the MLR and KRR models predicted an increase in the Abbe number (Tables I and S1). Moreover, we observe that the dopants Nb_2O_5 and Ta_2O_5 cause a decrease in the experimental Abbe number, which also agrees with the predictions of both regression models.

These results demonstrate how statistical regression can provide directions for designing new materials. It is particularly important to note that there are no guidelines or clear intuitive reasons why these particular dopants would lead to increases or decreases in the Abbe number, which further justifies the use of data-driven methods for this particular case. Finally, these results also highlight that complementary predictions for different regression models should be considered when designing new materials. For example, the dopant Sc_2O_3 was observed to increase the Abbe number in the experiment; however, this increase was predicted only by the MLR model and not by the KRR model.

IV. CONCLUSION

The design of new glasses with specifically targeted properties is an outstanding challenge in materials science. In order to address this challenge, we adopted a data science approach based upon statistical regression. The advantage of statistical regression is that it can elucidate the complicated relationship between physical property and glass composition with arguably higher accuracy than any physical theory. Here, we adopted two complementary regression approaches—multiple linear regression and kernel ridge regression—and succeeded to learn the relationship between an optical property (the Abbe number) and glass composition with high accuracy. Using these relationships, we predicted how the optical properties of a specific glass could be tuned by doping. These predictions were tested experimentally, and agreement between the direction of change (increases and decreases in the Abbe number) was obtained. This work therefore demonstrates a successful example of data-driven materials design and can be used as a template for designing glasses or other materials with other desirable properties.

SUPPLEMENTARY MATERIAL

See the supplementary material for the results of cross-validation for MLR and KRR shown in Figs. S1 and S2, respectively. The Abbe numbers estimated by MLR and KRR for all the oxides are tabulated in Table S1.

AUTHORS' CONTRIBUTIONS

Y.T. conceptualized the idea, ran the software, wrote the original draft, and supervised the work. M.F. performed investigation. D.M.P. performed the methodology, validated the study, and wrote, reviewed, and edited the paper. M.K. performed investigation. Y.U. performed validation and investigation.

ACKNOWLEDGMENTS

This study was partially supported by Asahi Glass Foundation, Research Institute for Sustainable Humanosphere, Kyoto University and Shiga University. The optical experiment was carried out on the Metricon 2010 prism coupler at Osaka Prefecture University (Professor M. Takahashi).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]F. Gan, J. Non-Cryst. Solids **184**, 9 (1995).
[2]P. Hartmann, R. Jedamzik, S. Reichel, and B. Schreder, Appl. Opt. **49**, D157 (2010).
[3]J. E. Shelby, *Introduction to Glass Science and Technology* (Royal Society of Chemistry, 2005).
[4]J. Chung, Y. Watanabe, Y. Yananba, Y. Nakatsuka, and H. Inoue, J. Am. Ceram. Soc. **103**, 167 (2020).
[5]S. Fujino, H. Takebe, and K. Morinaga, J. Am. Ceram. Soc. **78**, 1179 (1995).
[6]K. Ishii, T. Tsuneoka, S. Sakida, Y. Benino, and T. Nanba, J. Ceram. Soc. Jpn. **120**, 98 (2012).
[7]L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, npj Comput. Mater. **2**, 16028 (2016).
[8]M. M. Echezarreta-López and M. Landin, Int. J. Pharm. **453**, 641 (2013).
[9]H. Liu, Z. Fu, K. Yang, X. Xu, and M. Bauchy, J. Non-Cryst. Solids: X **4**, 100036 (2019).
[10]D. S. Brauer, C. Rüssel, and J. Kraft, J. Non-Cryst. Solids **353**, 263 (2007).
[11]A. Afantitis, G. Melagraki, K. Makridima, A. Alexandridis, H. Sarimveis, and O. Iglessi-Markopoulou, J. Mol. Struct.: THEOCHEM **716**, 193 (2005).

[12]A. I. Priven, Glass Technol. **45**, 244 (2004).
[13]Y. T. Sun, H. Y. Bai, M. Z. Li, and W. H. Wang, J. Phys. Chem. Lett. **8**, 3434 (2017).
[14]B. Deng and Y. Zhang, Chem. Phys. **538**, 110898 (2020).
[15]Y. Zhang, A. Li, B. Deng, and K. K. Hughes, npj Mater. Degrad. **4**, 14 (2020).
[16]B. Deng, J. Non-Cryst. Solids **529**, 119768 (2020).
[17]E. Alcobaça, S. M. Mastelini, T. Botari, B. A. Pimentel, D. R. Cassar, A. C. P. d. L. F. de Carvalho, and E. D. Zanotto, Acta Mater. **188**, 92 (2020).
[18]R. Ravinder, K. H. Sridhara, S. Bishnoi, H. S. Grover, M. Bauchy, Jayadeva, H. Kodamana, and N. M. A. Krishnan, Mater. Horiz. **7**, 1819 (2020).
[19]C. Dreyfus and G. Dreyfus, J. Non-Cryst. Solids **318**, 63 (2003).
[20]H. F. F. Mahmoud, arXiv:1906.10221 (2019).
[21]C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).
[22]S. An, W. Liu, and S. Venkatesh, Pattern Recognit. **40**, 2154 (2007).
[23]K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (The MIT Press, 2012).
[24]RCore Team, R: A language and environment for statistical computing, https://www.r-project.org/, 2019.
[25]RStudio Team, Integrated Development for R, http://www.rstudio.com/, 2019.
[26]D. M. Packwood, *Bayesian Optimization for Materials Science*, SpringerBriefs in the Mathematics of Materials (Springer, 2017).
[27]M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Fundamentals of Machine Learning*, 2nd ed. (The MIT Press, 2018).
[28]A. Alexandros, A. Smola, K. Hornik, and A. Zeileis, J. Stat. Software **11**, 1 (2004).
[29]See https://www.newglass.jp/interglad_n/gaiyo/info_j.html for New Glass Forum.
[30]Ohara, Inc., Japanese Unexamined Patent Application Publication No. 2009-203155 (2009).