ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Non-Crystalline Solids 342 (2004) 166–169

JOURNAL OF NON-CRYSTALLINE SOLIDS

www.elsevier.com/locate/jnoncrysol

Letter to the Editor

About testing the reliability of glass property data in binary systems

O.V. Mazurin [a,*], Yu. Gankin [b]

[a] *Joint-Stock Company 'Thermex', 86 Sredniy Prospect, St. Petersburg, Russia*
[b] *MDL Information Systems, Inc.*

Received 24 April 2002; received in revised form 7 April 2004
Available online 12 August 2004

Abstract

Reliability of the experimental data given in the publications concerning properties of glasses and melts is an important characteristic of such publications. Unfortunately, in the last decades there is an obvious tendency of increasing the percentage of publications containing gross errors in experimental data. In the present paper the ways to oppose this tendency by using glass property databases are described.

Analysis of the dynamics in the studies of properties of glasses and melts made during the last few decades [1] reveals a rather disturbing tendency, namely, a steady increase of obviously erroneous data in publications. Especially dangerous in our opinion is the fact that such erroneous results can be found in an appreciable number of cases in the highly ranked journals publishing papers on glass science. Due to the excellent refereeing procedure, the general quality of papers in these journals is high and accordingly all information that is published in such journals is considered to be highly reliable. Thus it is important to take steps to decrease the number of erroneous data published in these journals. The simplest and most obvious way to do this is to compare the published results with the results of previous studies. If these earlier obtained results are numerous enough and have been obtained by a number of various groups of scientists, they form a good base for determining the reliability of newly found property values. Until recently it has been practically impossible for referees to check the quality of the experimental data presented in the papers submitted to a journal by comparing them with the whole set of the existing data. Now the situation has changed due to the existence of two universal glass property databases, namely, INTERGLAD [2] and MDL®SciGlass [3]. It takes only a few minutes to find whether given properties of given glass systems have been studied and, if they have, to find how the data in a refereed paper correlate with the already published ones. Accordingly, we propose that this is a good time to undertake steps that should appreciably decrease the publication of incorrect experimental data in the pages of the leading journals on glass science.

There is also another aspect of this problem. Among the papers with erroneous data there is a considerable number of papers where much attention has been given to ensuring a high quality of measurements. However, it was obvious from such papers that the authors were unable to compare their results with most of the results already published by other scientists. The reasons for it seem obvious. Individual attempts to find most of the

* Corresponding author. Tel.: +7 812 320 54 22.
E-mail addresses: mazurin@itcwin.com (O.V. Mazurin), y.gankin@mdli.com (Yu. Gankin).

0022-3093/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.jnoncrysol.2004.06.012

existing data published during the last century all over the world requires too much effort. At the same time, careful comparison of new results with previous ones is one of the main conditions of a really reliable and therefore really useful study. In principle, it is possible that some high quality new data could differ from many previous ones for certain objective reasons, for example, due to some specific features of an experimental procedure. However, in this case an author should emphasize this difference and try to find the reasons of such variation. Certainly, in this case he/she should pay special attention to all details of his/her experiment and give convincing evidences of high precision of the selected study procedures. Accordingly, we are sure that it is a good time for authors of scientific papers to begin to consider a glass property database as a necessary instrument in their scientific work.

To make this letter reasonably short we will illustrate our points by only one example, though we can cite dozens of similar examples taken from papers published in this journal during the last decade, as well as in practically all high-ranking journals where papers with new data on glass properties can be found. One should state that it is a general misfortune nowadays that authors do not conduct a reliability analysis based on a thorough enough inspection of the existing data, before submitting a paper for publication.

The example presented below concerns the values of properties of binary glasses. It is the easiest and most clearly demonstrated way to determine outliers in experimental data. Quite often in various studies a binary glass is used as an initial member of a series of ternary and even four-component glasses. It is clear that, if the property value of the first member of a large series of glasses is incorrect, the correctness of all other data for these glasses is also questionable. In principle, it is also possible to compare values of new and already published property data for ternary and more complicated compositions.

To illustrate the general approach to the discussed problem we have selected the property and binary system for which the greatest amount of data can be found in the literature. This is the density of sodium-silicate glasses. The graph where these data are compiled is shown in Fig. 1. The graph was prepared by using the MDL SciGlass Information System because it proved to be somewhat more convenient for our purpose. The figure presents the results of 567 measurements of density taken from 164 publications. By a special option of the program outliers were found and excluded supposing a normal distribution of residuals and fiducial probability 0.95. The number of outliers was 29, i.e. ~5% of the whole sampling of points. After that the most probable composition dependence of density was determined by using the method of least squares. This dependence was described by Eq. (1):

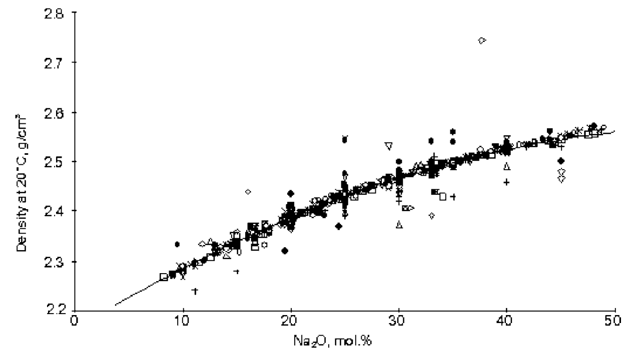

Fig. 1. The values of density of binary sodium-silicate glasses with concentration of Na_2O in the range from 8 to 49 mol%. The line is the approximation of 567 points taken from 164 different publications by a quadratic polynomial.

$$d = -1.075\text{E–}04 * ([Na_2O] - 26.65)^2 + 7.677\text{E–}03 * ([Na_2O] - 26.65) + 2.443, \quad (1)$$

where d is the density at room temperature and $[Na_2O]$ is the content of Na_2O in mol%.

By using this approach for any graph similar to that presented in Fig. 1 it is easy to find erroneous data in published papers. Table 1 shows some statistics concerning the number of outliers found in publications belonging to various periods of studies of glass properties in comparison with the total amount of experimental data on densities of sodium silicate glasses published during the corresponding period. In our opinion, the tendency to a rapid increase in percentage of erroneous data can be called a frightening one. It is reasonable to add that during the period of 1991–2000 fifteen (of the total sixteen) points describing outliers were taken from papers published in highly ranked American, English, and German journals.

As was mentioned above, outliers were selected on the basis of an assumption that the distribution of residuals is a normal (Gaussian) one. The corresponding diagram is shown in Fig. 2. The line shows the approximation of the obtained dependence by the Gaussian equation. We can see that distribution of experimental data around mean (i.e. the most probable) values is close to the normal one. It would be not surprising if all these data were obtained in one study by one experimental

Table 1
Comparison of total data on density of sodium silicate glasses published before the end of the 20th century and data containing gross errors (outliers)

Years	Total amount of data		Outliers	
	Number of papers	Number of results	Number of papers	Number of results
1889–1980	107	396	3	3
1981–1990	33	73	4	9
1991–2000	29	81	8	16

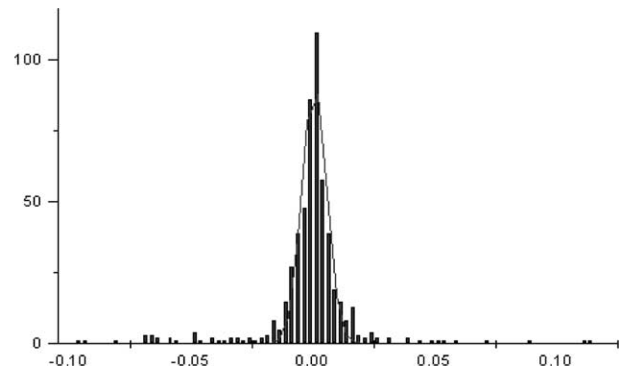

Fig. 2. Distribution of the differences between experimental and calculated (by using Eq. (1)) values of density for the selection of data presented in Fig. 1. The line shows the Gaussian fit to the distribution.

technique. However, in our case we have compiled data taken from more than two hundred various publications. It is obvious that the experimental technique used by every author differed from all the techniques used by other authors either by some details or even fundamentally. Here actually hundreds of factors, including psychological ones, influenced the values of the presented sampling. Nevertheless, we have the normal distribution of errors. The only essential deviation from the approximating curve is the central part of this distribution. It results from the fact that in several excellent studies (such as the work by Morey and Merwin [4] for example) densities for a great number of glasses were measured. This led to large number of points that practically coincide with the approximation curve. It is also seen that outside the Gaussian distribution there is a comparatively small number of points whose positions are not related to any kind of order. They are outliers. It is self-evident that, if we want to find the most probable property dependence, we should remove these outliers before determining the most probable composition dependence by an approximating polynomial (as was mentioned above).

It should be noted that the normal distribution was found for most of the data compiled in MDL SciGlass for various properties of binary and ternary glasses (a number of examples is presented in Ref. [5]), the main condition being a considerable number of points taken from publications by more than 4 or 5 different authors. In principle, most probable composition dependence of a property as well as obvious outliers could be found even in cases when a distribution of differences between calculated and experimental values appreciably deviates from the envelope. However, in any case the correct use of the described procedure can be possible only if the distribution of the main part of points around the approximation curve is symmetric enough.

We do hope that the glass community will accept the described method of determination of the most probable composition dependencies of glass properties (and hence the determination of outliers). However, in view of the importance of this problem we decided to present another evidence of the accuracy of such an approach. We selected the values of density of sodium-silicate glasses taken only from the papers where results of chemical analysis were presented. In this case the error of determination of a glass composition should be considerably smaller than in the cases of compositions given by batch. However, probably even more important is the fact that scientists who perform chemical analysis are usually especially careful in minimizing errors of property measurements. 179 points taken from 29 various papers were compiled. We used the procedure of data processing described above. The initial graph with all data is shown in Fig. 3 and the corresponding distribution of residuals is presented in Fig 4. Comparison of this graph with the graph in Fig. 1 shows that indeed the general quality of the selected data is much higher than that of all existing data. After exclusion of several outliers the following approximating equation was obtained:

$$d = -1.043\text{E–}04 * ([Na_2O] - 28.65)^2 + 7.253\text{E–}0.3 * ([Na_2O] - 28.65) + 2.458. \quad (2)$$

The comparison of Eqs. (1) and (2) shows that the difference between calculation of density for the central part of the dependence (from about 15 to 40 mol% of Na_2O) does not exceed 0.001 g/cm^3. It means that even when we collect a great number of results, most of which can not be considered as very precise (due to results of many studies of glasses with compositions given by batch) statistical evaluation of these results gives us the possibility to obtain reasonable characteristics of property-composition dependencies. Accordingly, these characteristics can be efficiently used for checking the quality of any new measurements in the same area of studies.

It is to be noted that the main features of statistical evaluation of data described above are the same for practically all properties and any kind of binary systems,

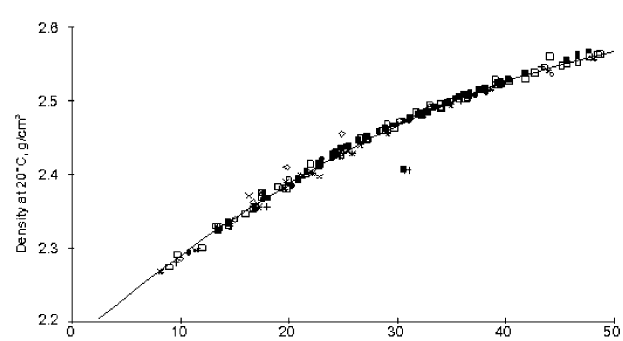

Fig. 3. The same as in Fig. 1 but only for the experimental values of density taken from the publications containing the analyzed compositions of the studied glasses. 179 points taken from 29 publications were processed.

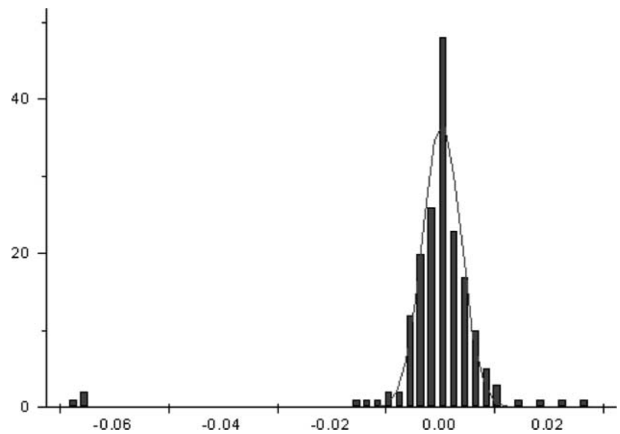

Fig. 4. The same as in Fig. 2 but only for the experimental values of density taken from the publications containing the analyzed compositions of the studied glasses (cf. Fig. 3).

if a considerable number of results of various authors could be compiled for the selected dependence.

In Fig. 3, three points from Refs. [6–9] ($d = 2.403$ for content of Na_2O equal to 30.6–31 mol%) clearly deviate from the most probable composition dependence much more than any other data. The authors of the mentioned papers have analyzed glass compositions. Thus it is evident that they paid a considerable attention to the experimental technique. In Ref. [7] they claimed that the error of density measurements was equal to ±0.005 g/cm^3. At the same time the actual residual of their measurements is near to 0.07 g/cm^3. This example shows that in principle even careful experimenters can obtain highly erroneous data if they have no means to compare them with data of other scientists.

We are sure that, if the authors were able to use one of the universal glass property databases, they would reveal this great discrepancy in their results in the first stages of their study and would improve the experimental technique. However, in any case, testing of the presented data quality in the submitted papers will help the authors to find and exclude gross errors in their publications and thus provide readers of the journal with reliable information on glass and melt properties.

References

[1] O.V. Mazurin, M.V. Streltsina, T.P. Shvaiko-Shvaikovskaya, A.O. Mazurina, Glass Phys. Chem. 28 (2002) 272.
[2] INTERGLAD International Glass Database, Version 5, New Glass Forum, 2001.
[3] MDL®SciGlass-6, MDL Information Systems, San Leandro, CA, 2003.
[4] G.W. Morey, H.E. Merwin, J. Opt. Soc. Am. 22 (1932) 632.
[5] O.V. Mazurin, M.V. Streltsina, T.P. Shvaiko-Shvaikovskaya, A.O. Mazurina, Glass Phys. Chem. 29 (2003) 555.
[6] A. Mekki, D. Holland, Kh. Ziq, C.F. McConville, J. Non-Cryst. Solids 220 (1997) 267.
[7] A. Mekki, D. Holland, Kh. Ziq, C.F. McConville, J. Non-Cryst. Solids 220 (1997) 271.
[8] J.A. Johnson, C.E. Johnson, D. Holland, A. Mekki, P. Appleyard, M.F. Thomas, J. Non-Cryst. Solids 246 (1999) 104.
[9] A. Mekki, K.A. Ziq, D. Holland, C.F. McConville, Phys. Chem. Glasses 43 (2002) 41.