Data Analysis with the Vogel−Fulcher−Tammann−Hesse Equation

João F. Mano*,†,‡ and Eduardo Pereira§

Department of Polymer Engineering, University of Minho, Campus de Azurém, 4800-058 Guimarães, Portugal, 3B's Research Group—Biodegradables, Biodegradables and Biomimetics, University of Minho, Campus de Azurém, 4710−057 Braga, Portugal, and Universidade do Minho, Escola de Ciências, Departamento de Física, Campus de Gualtar, 4710-057 Braga, Portugal

Received: April 8, 2004; In Final Form: August 2, 2004

The Vogel−Fulcher−Tammann−Hesse (VFTH) equation has been the most widespread tool for describing the temperature dependence of the characteristic times of the cooperative dynamics of glass-forming liquids, both in low-molecular-weight and macromolecular systems near and above the glass transition temperature. Previous arguments have shown that statistical correlation could exist between the three parameters. The analysis of a representative experimental data set of poly(vinyl acetate) dielectric relaxation (Richert, R. *Physica A* **2000**, *287*, 26) allowed us to conclude that correlation exists mainly between the B and T_0 parameters. It was found that the best fitting procedure should use a singular value decomposition implementation of the Levenberg−Marquardt nonlinear fitting algorithm. Simple criteria to diagnose (fitting matrix singular values) and minimize (linear combinations of fitting parameters responsible for instabilities edited to zero) ill-conditioned problems are given. The constant χ^2 confidence limit ellipsoids are presented, and it was concluded that the expected association between fitted parameters should always be made with the partial correlation instead of the more familiar correlation matrix. Some qualitative arguments are presented, stressing the need to properly weight experimental data points whenever unequal statistical weights are present in actual data because of either the merging of data sets obtained from different methods or the transformation from the measured value into a different fitted variable. It is expected that the same trends will be evident to a greater or lesser degree for all instances of VFTH data analysis.

Introduction

Glass-forming materials are able to maintain a disordered liquidlike structure below their melting temperature, if crystallization is prevented. The viscosity η or the relaxation time τ of the supercooled liquid increases strongly with decreasing temperature.[1−3] The glass transition is attained at temperatures where $\eta \approx 10^{13}$ Poise or $\tau \approx 100$ s, below which the system is no longer capable of equilibrating within the time scale of the experiment. This approach to the glass transition features a number of common properties for chemically different substances, such as those involving van der Waals forces or hydrogen, metallic, or covalent bonds. In all cases, the response function is clearly nonexponential and the associated relaxation behavior (usually labeled α-relaxation) is observed through a wide frequency range and is characterized by a broad spectrum of relaxation times. Moreover, the temperature dependence of its mean value, τ, often deviates from a simple thermally activated or Arrhenius behavior, $\tau(T) = \tau_0 \exp(E_a/RT)$.[1−3] In fact, it seems that the relaxation time increases more rapidly in the approach to the glass transition, diverging at temperatures below T_g but well above $T = 0$ K.

The three-parameter Vogel−Fulcher−Tammann−Hesse (VFTH) function[4−6] serves as a basis for the treatment of relaxation phenomena in glass-forming and viscous systems, being able to model $\tau(T)$ over a broad temperature range

$$\tau(T) = \tau_0 \exp\frac{B}{T - T_0} \quad T_0 < T_g \tag{1}$$

where τ_0 is a preexponential factor and B and T_0 are specific adjustable parameters. T_0 is a diverging temperature, close to the so-called Kauzmann temperature, and $D = B/T_0$ is the strength parameter, which can be related with fragility, in the classification proposed by Angell:[7] a D value (>30) represents a *strong* behavior, and a low D value (<30) is for a *fragile* behavior.[8]

The VFTH function is equivalent to the so-called Williams−Landel−Ferry (WLF) equation[9]

$$\log a_T = \log \frac{\tau(T)}{\tau(T_{ref})} = -\frac{C_1(T - T_{ref})}{C_2 + (T - T_{ref})} \tag{2}$$

where C_1 and C_2 depend on the material and on T_{ref} (chosen reference temperature). This equation can be obtained from the free volume theory. It is based on the Doolittle equation that relates the viscosity to the fraction of free volume ($f = v_f/v$) as $\eta \approx \exp(b/f)$, where v_f is the free volume, v is the specific volume, and b is a material parameter. Equation 2 may be derived if the free volume decreases linearly with decreasing temperature.

A theoretical rationalization of the VFTH (or WLF) equation can also be derived with the Adam−Gibbs theory.[10] This theory introduced the concept of *cooperatively rearranging regions*,

* Corresponding author. E-mail: jmano@dep.uminho.pt. Phone: +351-253510320. Fax: +351-253510339.
† Department of Polymer Engineering.
‡ 3B's Research Group—Biodegradables, Biodegradables and Biomimetics.
§ Departamento de Física.

10.1021/jp0484433 CCC: $27.50
Published on Web 11/13/2004

where the temperature dependence of their sizes allows one to relate the structural relaxation time with both the temperature and the configurational entropy, S_c

$$\tau(T) = \tau_0 \exp[C/TS_c(T)] \tag{3}$$

where τ_0 and C are constants. S_c can be calculated from the excess heat capacity, $\Delta C_p = C_p(\text{liquid}) - C_p(\text{glass})$, $S_c(T) = \int_{T_0}^{T}[\Delta C_p(T')/T']\,dT'$, where at T_0 the entropies of the liquid and solid states are equal ($S_c = 0$). Assuming that in the supercooled state ΔC_p varies with temperature as $\Delta C_p = \alpha/T$, the VFTH equation is obtained, with $B = CT_0/\alpha$.

From the models exposed, one may conclude that relevant information may be obtained from the VFTH parameters (τ_0, B, and T_0) associated to a given system; this includes insights about the fragility, free volume, temperature of diverging relaxation times, or limit frequency that may be correlated to the system's structure. Moreover, the VFTH parameters may be used to predict the system's dynamics at temperatures beyond the experimental ones.

The VFTH model and its modifications are widely used for the study of the cooperative thermal relaxation process. Most of the works dealing with the dynamics of liquids above or close to T_g use it to treat the temperature dependence on viscosity, relaxation times, or relaxation frequencies ($\omega \propto \tau^{-1}$). In most cases, a direct nonlinear fitting algorithm is used. An immediate and fundamental question arises about the reliability of fitted VFTH parameters. What are the error bounds for a given confidence level, and what are the statistical correlations between individual parameters? Are there situations in which a very different set of parameters fits equally well the experimental data? And, in these ill-conditioned situations, which are far more common than the experimentalist would like to admit, does a high correlation between parameters signal the inability of the data to discriminate between linear combinations of parameter values? Or, does it quantify a real, statistically meaningful association between corresponding physical quantities? Signs from previous work point out the possible *statistical fragility* of the VFTH equation. For example, in the structural relaxation investigation of polymers, Goméz Ribelles et al.[11,12] used a phenomenological model, based on the configurational entropy, to describe the glass transition. The experimental data were adjusted to the model using a conventional minimization procedure. It was found that convergence during the fitting would demand one to fix, for example, the B parameter. However, different fixed values of B lead to the same final $\tau(T)$ curves. This suggests that the actual data is not able to discriminate the B value, and therefore, that a statistical correlation between the VFTH parameters should exist. These are precisely the questions that this paper will try to answer.

Thermally stimulated recovery is a mechanical spectroscopy technique that allows us to investigate relaxation processes in the low-frequency region, being complementary to the dynamic mechanical spectroscopy technique.[13−15] It is usually assumed in the data treatment that the dynamics of the studied process follows Arrhenius behavior within a narrow frequency range. This allows one to obtain, from the thermal sampling procedure, the temperature dependence of the apparent activation energy of the relaxation, which is normally characterized by a distribution of characteristic times. If the Arrhenius parameters are calculated from a direct fit of the experimental data to a simple Voigt−Kelvin model, a clear dependence between τ_0 and E_a is observed. Mathematically, the Arrhenius equation is a particular case of the VFTH equation when $T_0 = 0$; thus, one should expect that the study of the VFTH fitting procedure will also reveal the same trends present in the Arrhenius case.

This study is an attempt to clarify relevant issues related to the use of the VFTH equation, when it is pretended to extract the corresponding parameters (plus realistic error bounds and possible correlations) from the fit. General quantitative criteria to diagnose (and minimize) numerical instabilities in the fit and possible artifacts in the expected (linear) correlation coefficients will be given. As a general rule of thumb, the use of singular value decomposition techniques within the well-known Levenberg−Marquardt nonlinear fitting algorithm is advocated. It will also be demonstrated that, in order to discuss the physical intrinsic association between VFTH parameters, the estimation of the partial correlation matrix is mandatory.

The experimental data explored in this work correspond to dielectric results obtained in the frequency domain on poly(vinyl acetate), compiled by Richert.[16] The data cover 16 decades in frequency, and 45 points are numerically available in that work. We will assume that the main conclusions that come from the analysis of this set of experimental data may be extended to the general features of the VFTH equation.

Linearization of the VFTH Equation

It is appealing to use linear fitting procedures, instead of the more cumbersome nonlinear counterparts, because one can rearrange experimental data in such a way as to judge the goodness-of-fit criteria by an easy graphical method based on how good the data is described by a linear relationship. The VFTH equation may be linearized by introducing the form[17]

$$\left[\frac{-\partial \ln \tau(T)}{\partial T}\right]^{-1/2} = B^{-1/2}(T - T_0) \tag{4}$$

The VFTH function is now expressed as a linear relationship between the left-hand side of eq 4 and temperature, where B and T_0 can be obtained from the corresponding slope and intercept. The major disadvantage of using eq 4 is the need to numerically derivate the $\ln \tau(T)$ experimental data with temperature, a numerically nontrivial point. Lunkenheimer et al.[3] numerically differentiate a high-order interpolating polynomial, but given the possible artifacts, cubic-spline interpolation should be used instead.[18] Apart from this detail, the use of this method (Stickel's method) has been tremendously sucessful.[3,17,19] The real value of this approach lies, however, in being especially suited to the following tasks: (i) clarification of a transition from a VFTH to an Arrhenius behavior (the derivative in the later case is a constant) and (ii) identification of the range of validity of a VFTH law.[17] Although both B and T_0 can be obtained from the slope and intercept of the Stickel plot, we do not advocate this approach, because the alternative full nonlinear fit is free from possible numerical artifacts from the numerical differentiation of experimental data and the information on confidence intervals, and the possible association between fitted parameters (derived from the correlation matrix) is much more reliable. Figure 1 shows the results of nonlinear Levenberg−Marquardt fit of Table 1 data together with the weighted residuals and the eq 4 linearized form of the VFTH model. Random residuals and a linear relationship in the inset, according to eq 4, are easy graphical judgments of how good the experimental data is following the VFTH empirical law. The best fits of Figure 1 are $\log(\tau/s) = -11.845 \pm 0.019$, $T_0/K = 250.81 \pm 0.12$, and $B/K = 1651.2060 \pm 0.0025$, the error bounds for a 95.4% confidence level.

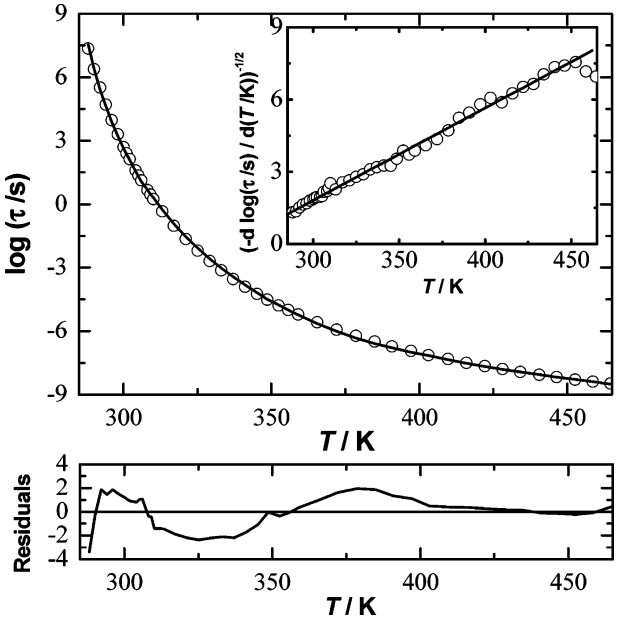

Figure 1. Singular value decomposition Levenberg−Marquardt nonlinear fit of Table 1 reference data with Vogel−Fulcher−Tammann-Hesse equation. Best fit parameters for a 95.4% confidence level: $\log(\tau/s) = -11.845 \pm 0.019$, $T_0/K = 250.81 \pm 0.12$, and $B/K = 1651.2060 \pm 0.0025$ (see text). Also shown are weighted residuals distribution (below) and Stickel's plot method showing that the VFTH equation adequately describes experimental data (inset).

TABLE 1: Experimental Data by Richert[16] on Dielectric Relaxation in Poly(vinyl acetate) Covering ca. 16 Decades in Time[a]

T/K	log τ/s	*T*/K	log τ/s	*T*/K	log τ/s	*T*/K	log τ/s
288.15	**7.53**	308.15	**0.68**	345.15	−4.19	403.12	−7.16
290.15	**6.40**	309.15	0.47	348.61	−4.51	409.67	−7.35
292.15	**5.41**	310.15	**0.31**	352.44	−4.77	415.94	−7.52
294.15	**4.63**	313.15	−0.27	355.78	−5.01	422.07	−7.67
296.15	**3.88**	317.15	−0.94	359.09	−5.24	428.16	−7.81
298.15	**3.23**	321.15	−1.54	365.49	−5.64	434.11	−7.94
300.15	**2.63**	325.15	−2.08	372.03	−6.01	440.51	−8.06
301.15	2.35	329.15	−2.58	378.60	−6.33	446.36	−8.17
302.15	**2.08**	333.15	−3.03	384.91	−6.59	452.63	−8.28
304.15	**1.56**	337.15	−3.43	390.83	−6.79	458.61	−8.39
305.15	1.30	341.15	−4.19	397.17	−7.00	464.52	−8.51
306.15	**1.06**						

[a] Bold-faced (derived from time-domain measurements) and other values (based upon frequency domain data) were obtained with different methods and probably have different experimental accuracies.

Convergence Surfaces

Nonlinear routines for fitting data to functions are based on methods for minimizing the sum of squared weighted residuals. To make it explicit that the actual experimental errors should be used in the computation of residuals, it is customary to call the appropriately weighted sum the χ^2 of the fit . Then, a goodness-of-fit criterion is the extent to which the reduced χ^2, χ^2_ν, the χ^2 divided by the number of degrees of freedom, is close to one.[20] If the actual experimental errors are not know a priori, then it is customary to assume unit errors and represent the sum of the residuals as S^2. The absolute value of this sum cannot be used to judge the quality of the fit, and the information that the experimentalist can have is reduced to use its value to estimate the a posteriori errors and to judge, on the basis of his knowledge, if the estimated errors are meaningful.[20] During the nonlinear fitting procedure, the space of the possible parameter values is explored until S^2 reaches a minimum, preferably an absolute minimum. The minimization problem is usually solved either by derivative methods, such as the Levenberg−Marquardt (LM) method, or by nonlinear search algorithms, such as the Simplex method. The quality of the result depends not only on the algorithm used but also on the fitting function. In fact, the adjustable parameters may or may not be intrinsically interdependent. Let's take, as a simple example, a function with two adjustable parameters, A_1 and A_2. Two limit cases may be addressed: If A_1 and A_2 are independent and the convergence is efficient, the 3D plot of S^2 versus (A_1, A_2) should display a deep, well-like surface, where S^2 decreases quickly in all directions when A_1 and A_2 goes throughout the estimate values. A dependency between A_1 and A_2 would be evident if S^2 would present a slow convergence, at least throughout one direction in the (A_1, A_2) space. Usually, this leads to a valley-shaped S^2 surface, where good estimates of A_1 and A_2 are found in the long, shallow bottom. In this case, there is definitely an association between A_1 and A_2 parameters, but this can reflect an intrinsic, physically meaningful association of the inability of experimental data to discriminate linear combinations of A_1 and A_2 values (numerical artifact). If the function has more than two fitting parameters, the same features could also be found, though in a higher-dimensional space.

For the case of the VFTH equation, we will first analyze the S^2 versus (log τ_0, B, T_0) surface by fixing independently the three parameters around the best values obtained from the nonlinear fitting. The corresponding 3D representation is then represented in a 2D plot as the projection of S^2 on the planes of the three parameters. The results of S^2 calculated with the analyzed experimental data set are shown in Figure 2. In Figure 2a, T_0 was fixed at different values around the best estimation. It can be seen that the S^2 surface has a single minimum, but a certain interdependency is observed between B and log τ_0. For a fixed T_0, if B increases relative to the best values, one reaches again a reasonable fitting by decreasing log τ_0.

A smaller association is expected from the data obtained if one changes log τ_0 and T_0 while keeping B fixed (Figure 2b). The shape of the surface toward the best values for a fixed B is much more round and isotropic than that observed in Figure 2a. Finally, a much higher correlation between B and T_0 for a fixed log τ_0 is found (Figure 2c). It should be noted in Figure 2 that the change of one fixed parameter leads to S^2 surfaces that changes locations in the plane of the other two parameters, but the shape tends to be maintained. This is a clear indication that a combined correlation between the three parameters must exist.

The analysis made in Figure 2 was made by looking at the VFTH parameters two by two (the third being fixed). We should extend the analysis of the S^2 surface by independently changing the three VFTH parameters. This analysis was done with the studied data. The S^2 hypersurface is projected toward the three possible planes in Figure 3. The three ellipse-like shapes observed in these graphics suggest that, in a (log τ_0, B, T_0) 3D representation, S^2 below a certain value should have a cigar-like shape. The more elongated shapes in Figure 3 correspond to the projection of S^2 toward the (B, T_0) and, to a lesser degree, the (B, log τ_0) planes. The data in Figures 2 and 3 give essentially the same qualitative information: There is always a strong association between any pair of VFTH fitted parameters, and the correlation should decrease in the order (B, T_0) > (B, log τ_0) > (log τ_0, T_0).

Levenberg−Marquardt Nonlinear Least-Squares Algorithm

The Levenberg−Marquardt (LM) algorithm has become the standard of the nonlinear least-squares fitting routines. Let us

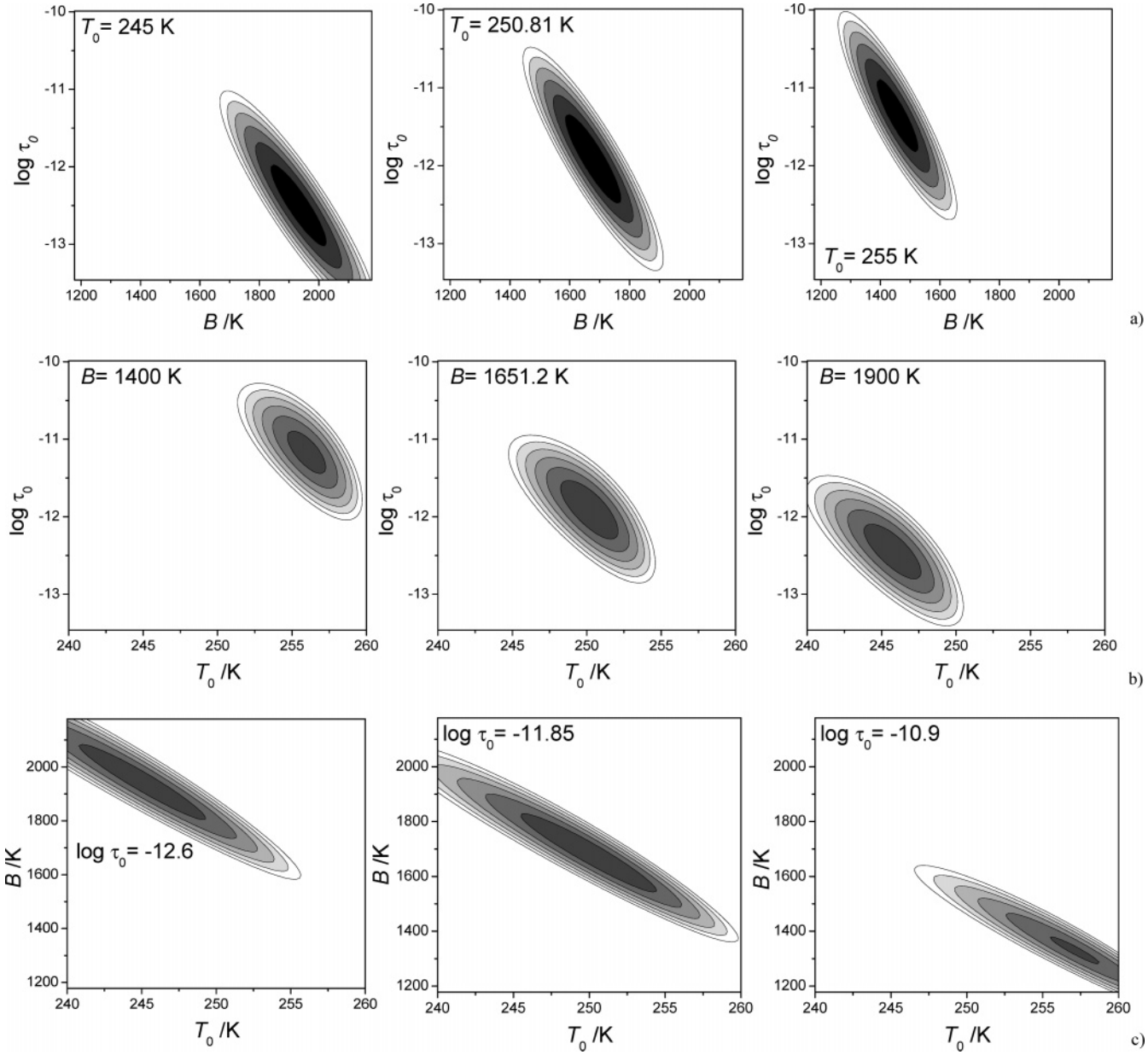

Figure 2. S^2 surface obtained by fixing two by two the three VFTH parameters and letting the third parameter vary around the best estimate (three possibilities). (part a) mapping throughout the (log τ_0, B) space. (part b) mapping throughout the (log τ_0, T_0) space. (part c) mapping throughout the (B, T_0) space. S^2 decreases with increasing darkness.

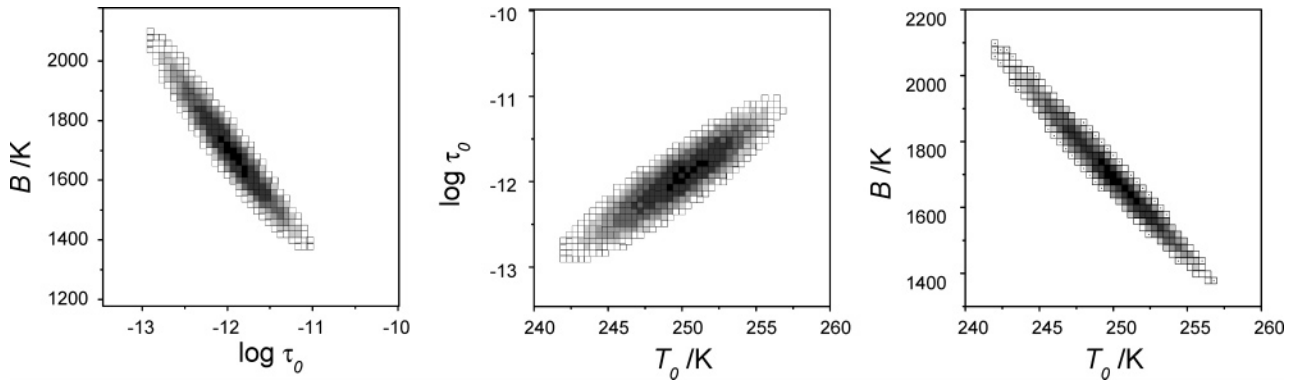

Figure 3. Projection of the S^2 surface along three possible planes, resulting from the analysis of the data set analyzed in this work letting the three VFTH parameters vary freely. S^2 decreases with increasing darkness.

define the notation by saying that one has a table of N experimental data points (index i) and wants to fit to these data to a nonlinear function of M parameters (index k). Although the dependence of the χ^2 value on the fitting parameters is nonlinear, the algorithm makes the method an iterative procedure of solving a set of linear equations in increments of the fitted

parameters $\delta\mathbf{a}$ that, added to the current approximation, give rise to a better estimate of the sought (true) parameters

$$\boldsymbol{\alpha}'\,\delta\mathbf{a} = \boldsymbol{\beta} \tag{6}$$

where $\boldsymbol{\alpha}'$ is obtained from the $\boldsymbol{\alpha}$ curvature matrix[18,21]

$$[\boldsymbol{\alpha}]_{kl} = \alpha_{kl} \equiv \frac{1}{2}\frac{\partial^2\chi^2}{\partial a_k\,\partial a_l} \approx \sum_{i=1}^{N}\frac{1}{\sigma_i^2}\left[\left(\frac{\partial y(x_i;\mathbf{a})}{\partial a_k}\right)\left(\frac{\partial y(x_i;\mathbf{a})}{\partial a_l}\right)\right] \tag{7}$$

the summation extending over all experimental data points.

The $\boldsymbol{\beta}$ vector is given by

$$[\boldsymbol{\beta}]_k = \beta_k \equiv -\frac{1}{2}\frac{\partial\chi^2}{\partial a_k} \tag{8}$$

The solution to the linear set (6) can be obtained by a matrix inversion procedure

$$\delta\mathbf{a} = [\boldsymbol{\alpha}']^{-1}\boldsymbol{\beta} \tag{9}$$

and this is the most usual implementation. Reference 18 has a popular implementation of the algorithm that uses Gauss−Jordan (GJ) elimination with full pivoting.

In the minimum of the χ^2 hypersurface, one computes the matrix $\mathbf{C} = [\alpha]^{-1}$, which is the estimated formal covariance matrix of the standard errors in the fitted **a** parameters.[18] The discussion of possible correlations between parameters is best made after normalization of the individual parameter variances, because the covariance matrix entries are scale-dependent. This is made by transforming the covariance $\mathbf{C} = [\sigma_{kl}^2]$ into the correlation matrix by

$$\mathbf{R} = \mathbf{D}^{-1}\mathbf{C}\mathbf{D}^{-1} \tag{10}$$

where **D** is a diagonal matrix given by $\mathbf{D} = \mathbf{D}^{-1} = [1/\sigma_{kk}]$.[22] The entries of the correlation matrix are the familiar (linear) correlation coefficients.

It is well-known that GJ elimination is rather susceptible to roundoff error. Moreover, whenever the fitting matrix is very close to singular (one or more very small pivot elements), one gets fitted parameters with large magnitudes but delicately and unstably balanced to cancel out almost precisely when the fitting function is evaluated.[18] This can occur if the experimental data is not very sensible to individual (or combinations of) fitting parameters. Because the LM is, in fact, an iterative algorithm for a linear set of equations, one can take advantage of the extensive knowledge of computational linear algebra available[18,23−25] and suggest an alternative to the solution of eq 9 based on the well-known singular value decomposition (SVD) of a matrix

$$\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^T \tag{11}$$

where **U** and **V** are column orthonormal matrices and $\mathbf{W} = [w_k]$ is a diagonal matrix with positive or zero elements (the singular values). Among the most desirable properties of the SVD solution of the LM method is minimization of roundoff error and numerical instabilities.[18,24] Moreover, the magnitudes of the singular values allow a clear diagnosis of possible ill-conditioned problems (reciprocal of the condition number, ratio of the largest to the smallest of the w_k's, comparable to machine precision). The inverses of the matrices in eq 11 are trivial to compute: $\mathbf{A}^{-1} = \mathbf{V}[\mathrm{diag}(1/w_k)]\mathbf{U}^T$. This should substitute the inverse in eq 9 and the SVD solution to the set (eq 6) is, finally

$$\delta\mathbf{a} = \mathbf{V}[\mathrm{diag}(1/w_k)](\mathbf{U}^T\boldsymbol{\beta}) = \left[\sum_{k=1}^{M}\left(\frac{\mathbf{U}_{(k)}\cdot\boldsymbol{\beta}}{w_k}\right)\mathbf{V}_{(k)}\right] \pm \sum_{k=1}^{M}\frac{\mathbf{V}_{(k)}}{w_k} \tag{12}$$

where the last summation is the error term. The last form shows that the columns of the matrix **V** are orthogonal unit vectors defining the principal axes of the constant χ^2 boundaries,[18] and therefore, these constant χ^2 ellipsoids have semi-axes with lengths equal to the reciprocal of the singular values w_k, smaller w_k's contributing more to the overall error than higher ones.

Whenever the curvature matrix is numerically close to singular, the solution obtained by editing to zero the values of $1/w_k$ for all the singular values smaller than a given threshold is often better than both the direct GJ elimination and the SVD solution where the small w_k's are left nonzero: When some combination of fitted parameters is irrelevant to the fit, that combination is driven down to a small value, rather than pushed up to numerically instable canceling infinities.[18,24,25]

The covariance matrix is obtained from the **V** columns as

$$\mathbf{C} = \sum_{k=1}^{M}\frac{1}{w_k^2}\mathbf{V}_k \otimes \mathbf{V}_k \tag{13}$$

where again the $1/w_k$ should be edited to zero for sufficiently small w_k's. Equation 12 also shows that the increments in the fitted parameters are linear combinations of the **V** columns with the weighted $\boldsymbol{\beta}$ vector, a point to be addressed later.

To test the ability to recover meaningful parameters from the fitting procedure, a Monte Carlo scheme of obtaining constant χ^2 confidence limit ellipsoids from synthetic data sets was adopted. The reference data of Table 1 was Levenberg−Marquardt (SVD)-fitted, and a set of fitted parameters $\mathbf{a}_0$ thus obtained. This was adopted as our best estimate of the sought (true) parameters $\mathbf{a}_{\mathrm{true}}$. The $\mathbf{a}_0$ set was then used to generate 1000 synthetic data sets adding Gaussian noise, and each data set was fitted as if it was experimental data to obtain the recovered parameter set $\mathbf{a}_i$. To the extend that the $\mathbf{a}_i - \mathbf{a}_0$ distribution resembles the shape of the $\mathbf{a}_i - \mathbf{a}_{\mathrm{true}}$ distribution (reasonable), the distribution of the $\mathbf{a}_i - \mathbf{a}_0$ parameter sets will quantify the distribution of the probable errors in the $\mathbf{a}_0$ set. For a more sound justification of the procedure, see ref 18.

Table 2 shows the pairwise correlation coefficients for both the inverse GJ matrix method as well as the SVD solution (editing to zero singular values smaller than 10^{-5} times the higher value) for a single fit to the Table 1 reference data. It

TABLE 2: Linear Correlation Coefficients for Levenberg−Marquardt Fit of Table 1 Data Obtained by the Inverse Matrix Gauss−Jordan Elimination (GJ) and by the Singular Value Decomposition (SVD, Maximum Condition Number 10^5) Methods[a]

		$\boldsymbol{B} - \log(\tau_0)$	$\log(\tau_0) - T_0$	$\boldsymbol{B} - T_0$
GJ	full correlation	−0.955	0.894	−0.981
	partial correlation	−0.894	−0.737	−0.956
SVD	full correlation	−0.993	−0.993	1.00
	sample y	−0.725	−0.748	0.987
	sample x, y	−0.65	−0.733	0.994

[a] Full and partial correlation are for single fits of the Table 1 data with errors in the dependent variable (y, log τ) only, and sample is the sample correlation obtained from fits of 1000 synthetic data sets (text and Figure 4) with errors in the dependent variable (y, log τ) only or in both dependent and independent variables (y, log τ; x, T).

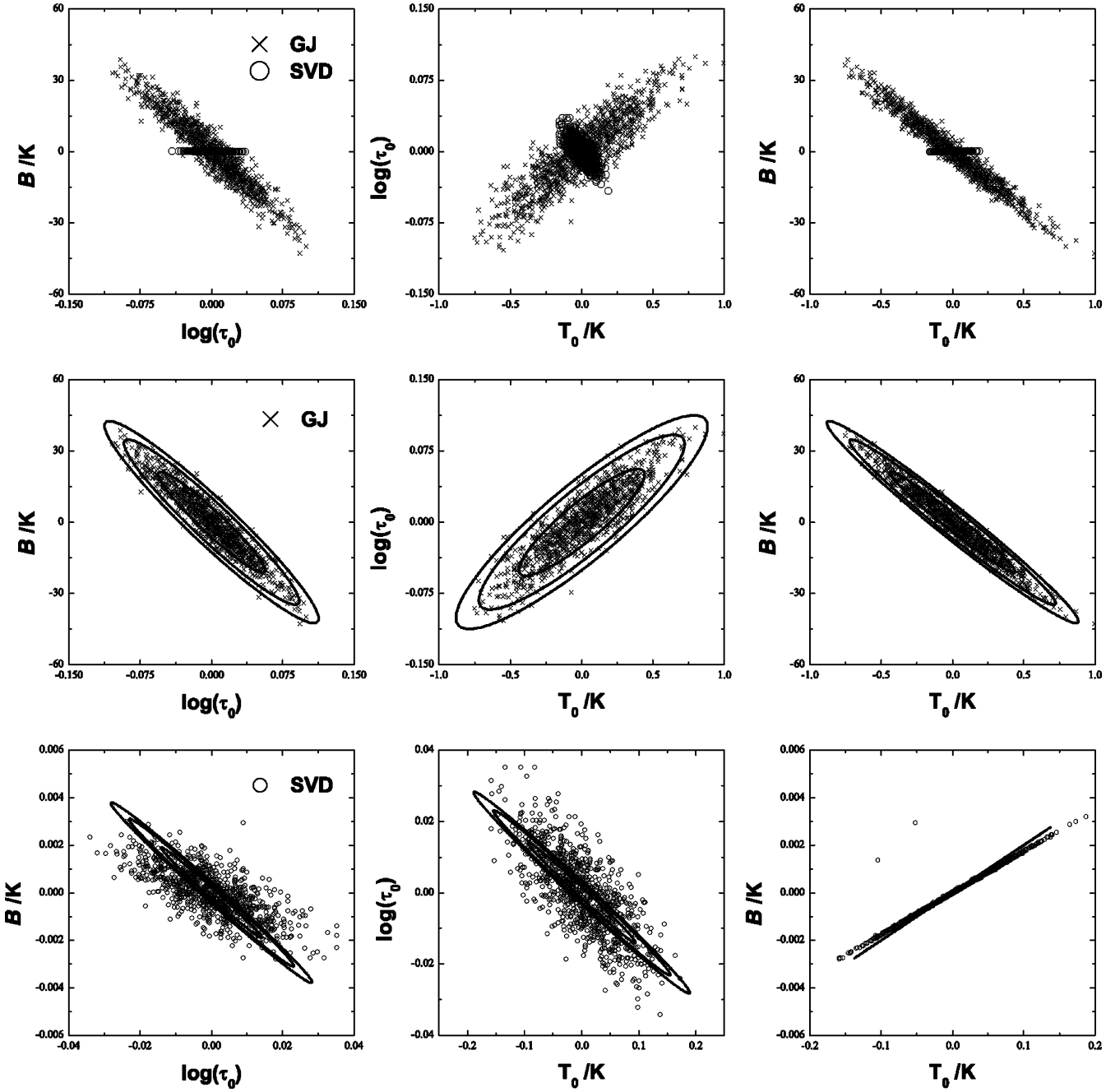

Figure 4. Distribution of the differences of the actual fitted parameters from the sample mean corresponding to the same 1000 sets of synthetic data obtained from a single fit of the reference data of Table 1 and adding $\sigma = 0.05$ Gaussian noise to the log τ values (errors in dependent variable only). Points correspond to fitted recovered parameters, while the curves are the constant χ^2 confidence regions for 68.3%, 95.4%, and 99% probability predicted from a GJ and SVD implementation of the Levenberg−Marquardt algorithm (see text). The SVD result edited to zero all combinations of singular values/**V** column vectors giving rise to condition numbers higher than 10^5. Please note the different scale of the last two rows of the figure.

shows that all of the parameters are pairwise correlated. The GJ results further show that the correlation between parameters $\log_{10} \tau$ and T_0 is slightly smaller than the mutual correlations involving parameter *B*. The discussion of SVD correlations is postponed for now. The first row of Figure 4 shows the distribution of the difference of fitted parameters from the sample mean $\mathbf{a}_i - \mathbf{a}_{\text{mean}}$ for 1000 synthetic data sets with constant $\sigma = 0.05$ Gaussian noise in the value of the fitted dependent variable (y). From the recovered distribution, sample correlation coefficients were computed in the usual way. The agreement between the correlation coefficients predicted from the GJ covariance matrix of the single fit of actual real data and the sample correlation coefficients from synthetic counterparts was, in all cases, at least five significant digits, thus validating the Monte Carlo procedure.

Consider now the data of Table 3: the singular values and corresponding orthonormal $\mathbf{V}_{(k)}$ vectors of the SVD decomposition of the curvature matrix in the minimum SVD fit of the experimental data. The condition number is 8×10^5, signaling precisely the type of numerical situation in which SVD (curvature matrix very close to singular) is most appropriate. Table 2 and Figure 4 show the results obtained with the SVD modification of the standard LM algorithm, editing to zero all of the $1/w_k$ values corresponding to condition numbers higher than 10^5 in eqs 12 and 13: The contribution from the last singular value/column of the **V** matrix was eliminated. The first

TABLE 3: Singular Values w_k and Corresponding Column Vector of the V Matrix for the Singular Value Decomposition of the Curvature Matrix in the Final Step of the Levenberg−Marquardt SVD Minimization Algorithm for the Single Fit of Table 1 Data

$\mathbf{V}\backslash w_k$	2.02×10^5	1.47×10^4	2.61×10^{-1}
$\log(\tau_0)$	−0.940	0.340	0.00577
T_0	−0.340	−0.940	0.0203
B	−0.0123	−0.0171	−1.00

row of Figure 4 immediately shows that the dispersion of the recovered parameters from synthetic data to the mean value is significantly smaller. As a side effect, the predicted correlation between parameters increases and, more important, changes signal. These data alone show that the SVD modification of the Levenberg−Marquardt model minimizes the error in the fitted NLLS parameters. In the case of not editing to zero any of the singular values, the SVD and GJ algorithms are equivalent at a point verified (not shown) with the data of Figure 4.

Figure 4 also shows constant χ^2 boundaries as confidence regions of constant significant level, estimated according to the procedure described in ref 18. If one does not want the full M-dimensional confidence region, but individual confidence regions for some smaller number ν of parameters (Figure 4), then it is necessary to make the projections of the full M-dimensional space into the ν-dimensional subspace of interest using a $\nu \times \nu$ matrix $[C_{\text{proj}}]$ from the full covariance matrix.[18] If the experimental measurement errors are normally distributed, then $\delta\mathbf{a} \equiv \mathbf{a}_k - \mathbf{a}_0$ has a multivariate normal distribution, and the equation for the elliptical boundary of the desired confidence region in the ν-dimensional subspace of interest is

$$\Delta = \delta\mathbf{a}^T[\mathbf{C}_{\text{proj}}]^{-1}\,\delta\mathbf{a} \tag{14}$$

where Δ is a function of the number of degrees of freedom of the fit. The confidence levels for 68.3%, 95.4%, and 99% probability are shown in Figure 4 for both the GJ and SVD algorithms. The agreement with the sample distributions is excellent for the GJ procedure and fairly good for the SVD case. Furthermore, the expected GJ confidence regions reproduce the correlation of the data in Figures 2 and 3, showing that realistic χ^2 hypersurfaces can be obtained from single-fit covariance matrices.

The signal change of the (linear) correlation coefficients between the GJ and SVD models deserves a more careful analysis. In a large number of cases, much effort is spent in interpreting and explaining the causes of large correlations between fitted parameters. The correlation can have several causes, but in general, these can be divided into two (possible connected) classes: (i) The fitting equation and actual experimental data are not sufficiently sensible to combinations of fitted parameters and (ii) there is a true correlation between parameters on the basis of physical arguments. In general, one wants to discuss the class ii effects, minimizing as much as possible class i artifacts. However, one must always keep in mind that picking up isolated entries from a correlation matrix can be misleading and can promote incorrect inferences, because all values of the correlation matrix are interrelated. Whenever one wants to discuss the significance of a high correlation between two parameter variables a_k and a_l, one must be aware of the possibility that the high correlation has arisen because of the mutual association of a_k and a_l with some other variable, and this is precisely the present case. For the a_k and a_l estimated association to be of intrinsic interest, it must remain high when the effect of additional variables has been removed: The *true* association between a_k and a_l must be computed with the values of additional variables held fixed in order not to influence the observed correlation. This can be done with the following procedure (actual details in ref 22): The vector of fitted parameters $\mathbf{a}$ is partitioned into two parts $\mathbf{a} = (\mathbf{a}_A, \mathbf{a}_B)$, where $\mathbf{a}_A$ has p elements (whose association one wants to investigate) and $\mathbf{a}_B$ has the remaining $q = M - p$ elements (whose influence on the $\mathbf{a}_A$ elements is to be corrected). The full correlation matrix is partitioned as

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{AA} & \mathbf{C}_{AB} \\ \mathbf{C}_{BA} & \mathbf{C}_{BB} \end{pmatrix} \tag{15}$$

The matrices in this matrix give the correlation in $\mathbf{a}_A$ ($\mathbf{C}_{AA}$), in $\mathbf{a}_B$ ($\mathbf{C}_{BB}$), and the covariances between the elements of $\mathbf{a}_A$ and $\mathbf{a}_B$ ($\mathbf{C}_{AB} = \mathbf{C}_{BA}$). Then, the conditional distribution of $\mathbf{a}_A$, given that $\mathbf{a}_B$ has a fixed value, is multivariate normal with a covariance matrix estimated as $\mathbf{C}_{A,B} = \mathbf{C}_{AA} - \mathbf{C}_{AB}\mathbf{C}_{BB}^{-1}\mathbf{C}_{BA}$, and the correlation is obtained from this covariance matrix normalizing in the manner described previously for eq 10

$$\mathbf{R}_{A,B} = \mathbf{D}_{A,B}^{-1}\mathbf{C}_{A,B}\mathbf{D}_{A,B}^{-1} \tag{16}$$

where $\mathbf{D}_{A,B}$ is a diagonal matrix containing the square roots of the diagonal elements of $\mathbf{C}_{A,B}$. $\mathbf{R}_{A,B}$ is the matrix of partial correlations and has the physical sough information: its $(k, l)^{\text{th}}$ element is an estimate of what the correlation between the k^{th} and l^{th} variables of $\mathbf{a}_A$ would be if $\mathbf{a}_B$ were held constant at any value.[22]

Table 2 has full and partial linear correlations for GJ. The most distinctive feature is the fact that the partial correlation between $\log_{10}\tau$ and T_0 is significantly different from the full value. The intrinsic correlation is masked by a stronger correlation between any one of these parameters and the remaining B: There is a considerable variation in the observed value of B and strong linear relationships between B and each $\log_{10}\tau$ and T_0 parameters, and this alone induces a strong linear relationship between $\log_{10}\tau$ and T_0. The induced association is so strong that the sign of the intrinsic correlation changes from −0.74 to an observed value of 0.89. This interpretation is confirmed by data in Table 3: The eigenvectors of the SVD algorithm show that the information about the fitted B value is obtained almost exclusively from the third eigenvector. Because the fitted parameters are given by linear combinations of the eigenvectors (eq 12), and because the SVD solution effectively discards the third singular value/eigenvector combination (previous discussion), the SVD solution only uses the first two eigenvectors, thus minimizing the correlation between parameters $\log\tau$ and T_0 on one hand and parameter B on the other. The actual sample SVD correlation is −0.75, remarkly close to the −0.74 partial GJ prediction: SVD effectively minimizes the coupling between parameters ($\log\tau$, B) and (T_0, B) to the extent that the intrinsic ($\log\tau$, T_0) correlation becomes evident without the necessity to correct for extraneous variables. The dramatic effect that editing to zero eigenvalues/eigenvectors, giving rise to high condition numbers in SVD, can have in the error distribution of the fitted parameters is best illustrated in Figure 5. The constant χ^2 confidence limit regions are considerably smaller, and the expected correlation between parameters can even change signal (middle graphic). The use of the full correlation matrix confirms the conclusions drawn from Figures 2 and 3 (the expected correlations should increase in the order (B, T_0) > (B, $\log\tau$) > ($\log\tau$, T_0)) but also illustrates the need to use partial correlation data, because the full correlation matrix will always pick up the M-1 higher mutual correlations,

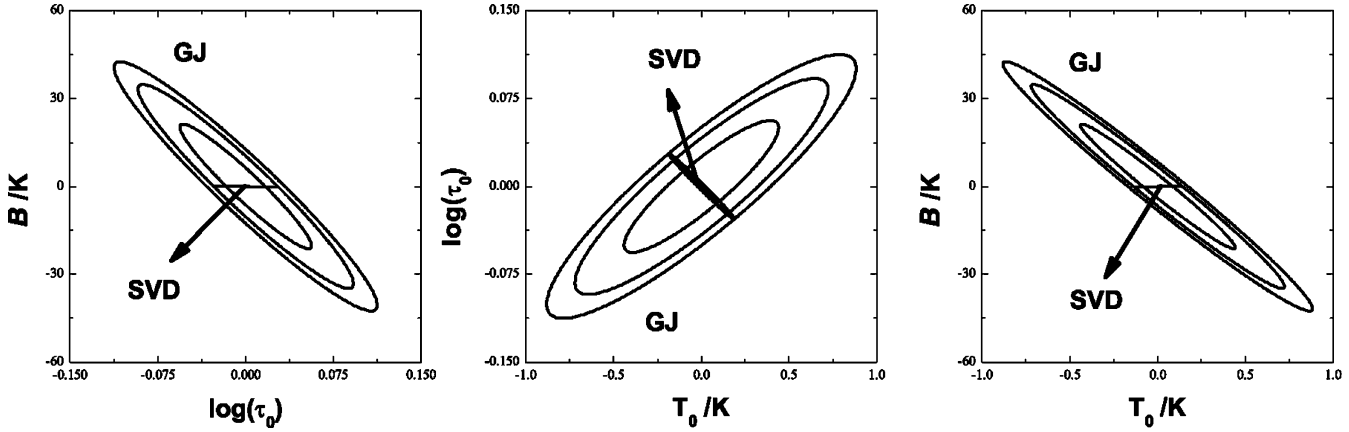

Figure 5. Comparison of the constant χ^2 confidence regions for 68.3%, 95.4%, and 99% probability predicted from a GJ and SVD Levenberg−Marquardt single fit to the reference data of Table 1. The SVD result edited to zero all combinations of singular values/**V** column vectors giving rise to condition numbers higher than 10^5.

unrevealing the smaller one. The results are also in agreement with the findings of Goméz Ribelles et al.[11,12] mentioned in the Introduction. Their experimental results are probably even more severely ill-conditioned than the data used in this work, to the point that their minimization procedure is not able to converge if *B* is left free to change.

From this discussion and from Figure 4, one can now give definitive advice on the best LM fitting procedure for VFTH data. The SVD algorithm is advocated instead of the simpler GJ inversion, because (i) it diagnoses ill-conditioned problems and (ii) minimizes their consequences. From the singular values, one can infer possible numerical instabilities: Singular values much smaller than the bigger ones show that experimental data is not sensitive to combinations of parameters which are multiples of the corresponding eigenvectors. These singular values/eigenvectors combinations should be edited to zero, thus minimizing errors in the fitted parameters. The results are better, but at the expense of a reduced curvature matrix. The discarded information can be important, however. The edited-to-zero eigenvectors should be inspected in each case to infer possible additional artifacts introduced by the fact that now the solution is obtained from a reduced set of eigenvectors bases. This is illustrated by data in Table 3. The SVD solution is obtained from the first two eigenvectors alone, and these have a very small contribution from the parameter *B*. The fitting procedure will have a reduced ability to minimize χ^2 by changing the *B* value, and this stresses the need to have a good initial estimate for *B*. If care is not exercised to test this point, the fitted parameters can become initial-approximation-dependent, a point confirmed but not illustrated in this paper. Although strictly correct from a mathematical point of view, the 95.4% confidence level $B/K = 1651.2060 \pm 0.0025$ interval should not be taken too seriously from a statistical point of view, because its value was estimated from a reduced basis set with a very small contribution from the *B* parameter. To maximize the useful physical information, as a rule of thumb, one should obtain fitted parameters from the SVD algorithm and extract meaningful correlation coefficients from an auxiliary GJ correlation matrix at the best-fit parameter values. The intrinsic correlation between pairs of parameters must always be estimated from partial correlations. This will ensure that the estimated correlation is as close as possible to the intrinsic value and free from extraneous induced associations.

Finally, Table 2 tests empirically the sensitivity of the fitting procedure to probable errors also found in independent (temperature) variables. The standard LM nonlinear fitting procedures assume errors in fitted dependent variables alone ($\log_{10}\tau$, fitting *y* values).[18] It is, however, unrealistic to expect that temperature is error-free. Table 2 then show results for SVD fits of 1000 synthetic data sets with Gaussian $\sigma = 0.2$ noise in the temperature. This was judged as a realistic error. At least for unbiased errors in the temperature, the results obtained were statistically equivalent for the standard *y*-only error algorithm and for the *x*,*y*-error case. The only difference was in the value of sample correlations (Table 2). This empirically confirms the standard procedure of neglecting statistical errors in the dependent variable for the nonlinear LM fitting algorithm.

Heteroscedastic Data

The statistical properties of nonlinear least-squares estimators are well-known: If the data is distributed independently and normally, the estimators of the parameters are unbiased, minimum variance, and maximum likelihood, and this holds true even if the data is heteroscedastic (i.e., of unequal uncertainty (σ_{yi})), but this is true only to the extent that the residuals are weighted properly

$$w_i = \frac{1}{\sigma_{yi}^2} \tag{17}$$

In general, the experimenter should have an a priori knowledge of experimental uncertainties and weight each squared residual according to the previous equation. Otherwise, nonlinear models generally yield nonnormal, biased, or inconsistent parameter estimates, and extreme care must be exercised when interpreting error bounds and correlation information extracted from fits.[20] Even if the knowledge of experimental uncertainties is only approximate, it is better to use $w_i \propto {}^1/_{\sigma_{yi}^2}$ instead of unweighted fits. And, this is mandatory for heteroscedastic data. In this case, although the closeness between the reduced χ^2 value and 1 cannot be used to access the quality of the fit, each residual will be minimized in the fit according to its intrinsic relative importance. This is standard undergraduate knowledge. Even so, this is completely overlooked in the majority of the VFTH manipulations involving relaxation times. By ignoring statistical weights one can bring even more uncertainty into computation results than operating with full, instead of partial, correlation matrices. This point is especially important for the analysis of data sets accumulated in relaxation measurements. Usually, several complementary techniques must be used to have access to a maximally broad range of relaxation times.[3] Merging the

data obtained by different methods, each with its own accuracy, causes one to inevitably face the problem of unequal statistical weights in combined data sets. This is precisely the case of the Richert reference data used in this paper.[16] Bold-faced row entries in Table 1 were derived from time-domain dielectric modulus data, which allow access to extremely long relaxation times, while light-faced entries were obtained from dielectric relaxation data as the peak frequency of the dielectric loss curve. Although this author does not provide information about experimental uncertainties, we have used a constant error of 0.05 based on our own experience. On the basis of this value, we have obtained as the best fitted parameter $\log(\tau/s) = -11.845 \pm 0.019$ and $T_0/K = 250.81 \pm 0.12$ with a 95.4% confidence level. To test the sensitivity of the data on unequal weights, we have made an alternative fit with bold entries in Table 1, given a constant error estimate of 0.10. We have thus obtained as the best fit $\log(\tau/s) = -11.908 \pm 0.015$ and $T_0/K = 250.32 \pm 0.14$, values with differences from the previous fitted parameters ($\Delta\log(\tau/s) = 0.063$ and $\Delta T_0/K = 0.49$), largely exceeding the statistical uncertainties. Even though this is a rather uninteresting test, it clearly stresses the care that must be exercised in properly weighting experimental data. This is especially important in the context of relaxation data for which data points differing by several orders of magnitude must enter the fit with different statistical weights (perhaps also differing by orders of magnitude and not just the two factors tested here).

Apart from the influence of merging data collected with different methods, there is an additional point concerning statistical weights that should be adequately addressed. This is the change of eq 17 to take into account the change in statistical weights whenever the fit is made in a transformed instead of the directly measured variable. This is again standard knowledge but usually not discussed. Whenever there is a change between measured y and actually fitted u variable, eq 17 should be substituted by

$$w_i = \frac{1}{\left(\sigma_{yi}\dfrac{\partial u}{\partial y}\right)^2} \tag{18}$$

the additional derivative term reflecting the change of variable. Only the use of eq 18 ensures that the fitting procedure minimizes actual experimental errors, and this is especially important for two types of transformation made in the mathematical handling of relaxation dynamics data: inversion and logarithmic transformations.[20] The light-faced entries of Table 1 were measured as peak frequency maxima of the broad distribution of dielectric loss curves: The Havriliak−Negami (HN) empirical curve is fitted to actual experimental data, and the peak frequency maxima f_{max} is extracted analytically from the null derivative condition for the maximum.[16,17] From the HN fit, it is then possible to extract the uncertainty in f_{max}, and this is the σ_{yi} value that should be used in the VFTH fit. From the frequency maximum raw data sets, one can do the VFTH data analysis, either in the form of eq 1 or in the more common form of

$$\log \tau = \log \tau_0 + \frac{1}{\ln 10}\frac{B}{T - T_0} \tag{19}$$

In the first case, the transformation is $u = 1/2\pi f$ and the statistical weight $\sigma_\tau = \sigma_f/2\pi f^2$. Special care must be taken in this case, because the lifetime data has, in principle, infinite variance.[20] In the more important case of eq 19, the form used for all data analysis in this paper, we have $u = \log(\tau) = \log(1/2\pi f)$ and $\sigma_{\log\tau} = \sigma_f/(f \ln 10) = 2\pi\tau\sigma_f/(\ln 10)$ showing that, in principle, one can have large differences in statistical weights. In the event of the uncertainties in frequencies being a fraction of the measured frequency maximum, $\sigma_{\log\ \tau}$ is constant, and one is allowed to use constant weights in fitting data with eq 19, but this must be checked in each case.

Conclusions

The well-known Vogel−Fulcher−Tammann−Hess (VFTH) equation has been extensively used in the description of cooperative molecular motion in glass-forming liquids. Experimental evidence has been pointed out for the statistical correlation between its three adjustable parameter, B, $\log \tau_0$, and T_0, which may bring questions on the reliability of fitted VFTH parameters. However, to the best of our knowledge this issue has not been adequately described in the literature.

It was shown in this work that the (relaxation time, frequency, or viscosity versus temperature) data may be treated in order to obtain the VFTH parameters from either linear or nonlinear fits. The well-known Stickel's linearization method can be used to obtain B and T_0. However, its main merit is to allow an easy graphical criterion to discuss the applicability of the VFTH model and we advocate the use of nonlinear algorithms due to their superior ability to minimize numerical artifacts and ability to extract reliable error bounds and correlation information.

We advocate the use of the Levenberg−Marquardt (LM) nonlinear fitting algorithm instead of several simpler but more error-prone possibilities of linearization of experimental data. The LM algorithm should use singular value decomposition (SVD) methods to solve the linear part of computation task, because these will do the following: (i) diagnose numerical instabilities (high condition numbers of fitting matrix) giving rise to ill-conditioned problems and, at the same time, (ii) minimize numerical artifacts (singular values/eigenvectors responsible for instabilities edited to zero). The distribution of singular values/eigenvectors should be checked in each case to prevent additional surprises (e.g., in the present case, the SVD fit will be unable to do an efficient minimization in the B parameter, thus making the fitted parameters slightly dependent in the initial approximation for B). The LM algorithm will accurately give error bounds and constant χ^2 confidence limit ellipsoids, but additional care must be exercised to unravel statistical association between fitted parameters, because the intrinsic correlation between any two parameters can be masked by additional stronger associations between any member of these two parameters and additional parameters. The partial correlation matrix should always by used instead of the familiar full correlation matrix. Although restricted to a single set of experimental data, it is expected that the same trends will be evident to a greater or lesser degree for all instances of Vogel−Fulcher−Tammann−Hess (VFTH) data analysis.

A brief comment is introduced to stress the need to use adequate statistical weights in the fitting procedure, which is very important for heteroscedastic (i.e., of unequal uncertainty) experimental data points. These unequal weights can be derived from either the merging in the same fit of experimental data originally obtained with different methods or the statistical weight correction necessary whenever the actual fitted data is transformed from the measured one.

Acknowledgment. Financial support for this work was provided by FCT, through the POCTI (POCTI/FIS/32901/1999) and FEDER programs. E.P. would like to acknowledge fruitful

discussions with M.N. Berberan-Santos, J.M.G. Martinho (CQFM, IST), and S. Lanceros-Mendez (UM).

References and Notes

(1) Angell, C. A. *Science* **1995**, *267*, 1924.
(2) Ediger, M. D.; Angell, C. A.; Nagel, S. R. *J. Phys. Chem.* **1996**, *100*, 13200.
(3) Lunkenheimer, P.; Schneider, U.; Brand, R.; Loidl, A. *Contemp. Phys.* **2000**, *41*, 15−36.
(4) Vogel, H. *Phys. Z.* **1921**, *22*, 645.
(5) Fulcher, G. A. *J. Am. Chem. Soc.* **1923**, *8*, 339.
(6) Tammann, G.; Hesse, W. *Z. Anorg. Allg. Chem.* **1926**, *156*, 245.
(7) Angell, C. A. *J. Non-Cryst. Solids* **1991**, *131*−*133*, 13.
(8) Böhmer, R.; Ngai, K. L.; Angell, C. A.; Plazek, D. J. *J. Chem. Phys.* **1993**, *99*, 4201.
(9) Williams, M. L.; Landel, R. F.; Ferry, J. D. *J. Am. Chem. Soc.* **1995**, *77*, 3701.
(10) Adams, G.; Gibbs, J. H. *J. Chem. Phys.* **1958**, *43*, 139.
(11) Gómez Ribelles, J. L.; Monleon Pradas, M. *Macromolecules* **1995**, *28*, 5867.
(12) Gómez Ribelles, J. L.; Monleon Pradas, M.; Vidaurre Garayo, A.; Romero Colomer, F.; Más Estelles, J.; Meseguer Dueñas, J. M. *Polymer* **1997**, *38*, 963.
(13) Cuesta Arenas, J. M.; Mano, J. F.; Gómez Ribelles, J. L. *J. Non-Cryst. Solids* **2002**, *307*−*310*, 758.
(14) Alves, N. M.; Mano, J. F.; Gómez Ribelles J. L. *Polymer* **2002**, *43*, 3627.
(15) Alves, N. M.; Gómez Ribelles, J. L.; Gómez Tejedor, J. A.; Mano, J. F. *Macromolecules* **2004**, *37*, 3735.
(16) Richert, R. *Physica A* **2000**, *287*, 26.
(17) Stickel, F.; Fischer, E. W.; Richert, R. *J. Chem. Phys.* **1995**, *102*, 6251; **1996**, *104*, 2043.
(18) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in Fortran*, 2nd ed.; Cambridge UP: Cambridge, 1992.
(19) Hansen, C.; Stickel, F.; Berger, T.; Richert, R.; Fischer, E. W. *J. Chem. Phys.* **1997**, *107*, 1086. Hansen, C.; Stickel, F.; Richert, R.; Fischer, E. W. *J. Chem. Phys.* **1998**, *108*, 6408.
(20) Tellinghuisen, J. *J. Phys. Chem. A* **2000**, *104*, 2834, 11829.
(21) Lampton, M. *Comput. Phys.* **1997**, *11*, 110.
(22) Krzanoswski, W. J. *Principles of Multivariate Analysis*, revised ed.; Oxford UP: Oxford, 2000.
(23) Anderson, E.; Bai, Z.; Bischof, C.; Blackford, S.; Demmel, J.; Dongarra, J.; Du Croz, J.; Greenbaum, A.; Hammarling, S.; McKenney, A.; Sorensen, D. *LAPACK users' guide*, 3rd ed.; SIAM: Philadelphia, 1999.
(24) Watkins, D. S. *Fundamentals of Matrix Computations*, 2nd ed.; John Wiley: New York, 2002.
(25) Lawson, C. L.; Hanson, R. J. *Solving Least-Squares Problems*; SIAM: Philadelphia, 1995. Björck, Å. *Numerical Methods for Least-Squares Problems*, SIAM: Philadelphia, 1996.