

Research Article

Multiple Linear Regressions by Maximizing the Likelihood under Assumption of Generalized Gauss-Laplace Distribution of the Error

Lorentz Jäntschi,^{1,2,3} Donatella Bálint,² and Sorana D. Bolboacă^{4,5}

¹Department of Physics and Chemistry, Faculty of Materials and Environmental Engineering, Technical University of Cluj-Napoca, Muncii Boulevard No. 103-105, 400641 Cluj-Napoca, Romania

²Doctoral School of Chemistry, Institute for Doctoral Studies, Babeş-Bolyai University, Kogălniceanu Street No. 1, 400084 Cluj-Napoca, Romania

³Department of Chemistry, Faculty of Science, University of Oradea, Universităţii Street No. 1, 410087 Oradea, Romania

⁴Department of Medical Informatics and Biostatistics, Faculty of Medicine, Iuliu Haţieganu University of Medicine and Pharmacy, Louis Pasteur Street No. 6, 400349 Cluj-Napoca, Romania

⁵Doctoral School, University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, Calea Mănăştur No. 3-5, 400372 Cluj-Napoca, Romania

Correspondence should be addressed to Sorana D. Bolboacă; sbolboaca@umfcluj.ro

Received 22 August 2016; Accepted 23 October 2016

Academic Editor: Irini Doytchinova

Copyright © 2016 Lorentz Jäntschi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multiple linear regression analysis is widely used to link an outcome with predictors for better understanding of the behaviour of the outcome of interest. Usually, under the assumption that the errors follow a normal distribution, the coefficients of the model are estimated by minimizing the sum of squared deviations. A new approach based on maximum likelihood estimation is proposed for finding the coefficients on linear models with two predictors without any constrictive assumptions on the distribution of the errors. The algorithm was developed, implemented, and tested as proof-of-concept using fourteen sets of compounds by investigating the link between activity/property (as outcome) and structural feature information incorporated by molecular descriptors (as predictors). The results on real data demonstrated that in all investigated cases the power of the error is significantly different by the convenient value of two when the Gauss-Laplace distribution was used to relax the constrictive assumption of the normal distribution of the error. Therefore, the Gauss-Laplace distribution of the error could not be rejected while the hypothesis that the power of the error from Gauss-Laplace distribution is normal distributed also failed to be rejected.

1. Introduction

The first report on multiple linear regression appears on 1885 [1] and was detailed in 1886 [2]. The classical treatments of the multiple regressions were built on the product-moment method implemented in 1846 [3] and later connected with the optimal correlation [4].

In his first published paper, Fisher introduces the method of likelihood maximization [5], later used in conjunction with Pearson's correlation [6]—a paper which started a contradictory debate between the method of central moments and

the method of likelihood estimation [7] replied to in [8] and finally linked with the partial correlation coefficients [9].

A multiple linear regression model involves more than two variables, one (y) being assumed dependent and the others (x_1, x_2, \dots, x_m) being assumed to be independent, and is considered here as a continuation of a previous study [10]. The most important assumption is that the data are paired; for example, a natural association between the values of the variables exists. This kind of association is accomplished when for instance a multiple linear regression is constructed involving a measured property/activity (y) for a series of compounds

for which other compounds measured properties/activities or structure-based descriptors are available (x_1, x_2, \dots, x_m), the natural association being in this case the (chemical) compound responsible for that property/activity/descriptor value.

The least squares method is the standard approach for regression analysis, the method being credited to Legendre [11] (for a debate about the inventor, please see [12]), which also (implicitly) assumes that the error is normally distributed.

Iteratively applying local quadratic approximation to the likelihood (through the Fisher information [13]), the least squares method was used to fit a generalized linear model as a way of unifying classical, logistic, and Poisson (linear) regression in [14] by iteratively reweighing the least squares method in the way to the maximum likelihood estimation of the model parameters.

Generalized Gauss-Laplace distribution is the natural extension [15] from Gauss's [16] and Laplace's [17] symmetric distributions. It is a triparametric distribution (location, scale, and shape) and parameter estimation via maximum likelihood and the method of moments have been reported in [18], concluding that the estimates do not have a closed form and must be obtained numerically.

A more general result regarding the maximum likelihood estimation can be found in [19] but unfortunately provides only conditions in which maximum likelihood estimate exists and is unique without providing the reciprocal (namely, there exist also other conditions than the one given, in which maximum likelihood estimate exists and is unique). Even more, for numerical estimates, it is hardly to discuss unicity.

The problem of estimating the parameters of a multiple linear regression under assumption of generalized Gauss-Laplace distribution of the error is a hard problem which can be solved only numerically and it involves an optimization problem with $m + 3$ constraints, where m is the number of unknown (to be determined) coefficients of the multiple linear regression. In this paper a mathematical and a numerical treatment of the problem is proposed.

In order to provide a proof of the facts for the proposed method of relaxing the distribution of the error when linear regression is used to link between chemical information and biological measurements, ten previously reported datasets were considered, all with significant role in human medicine or ecology.

2. Mathematical Treatment

One may define the generalized Gauss-Laplace (GL) distribution as

$$\begin{aligned} \text{GL}(x; \mu, \sigma, q) \\ = \frac{q}{2\sigma} \frac{\Gamma^{1/2}(3/q)}{\Gamma^{3/2}(1/q)} \exp\left(-\frac{|(x - \mu)/\sigma|^q}{(\Gamma(1/q)/\Gamma(3/q))^{q/2}}\right), \end{aligned} \quad (1)$$

where $\Gamma(\cdot)$ is the Gamma function and μ (location), σ (scale), and q (shape) are the parameters of the distribution.

This definition will be used here for the Gauss-Laplace distribution to relax the normal distributed constraint for the distribution of the error (ε).

2.1. Statement of the Problem. Multiple linear regressions under assumption of GL distribution (see (1), for the error ε ; $(a_j)_{1 \leq j \leq m}$, σ and q are to be determined from sampled data) are stated in the following equation:

$$\text{MLRGL}(\varepsilon; \sigma, q) = \text{GL}\left(\varepsilon; (a_j)_{1 \leq j \leq m}, \sigma, q\right), \quad (2)$$

where $\varepsilon = y - \sum_{1 \leq j \leq m} a_j x_j$ (and $\hat{\varepsilon} \approx 0$ and $\bar{\varepsilon} = 0$).

The case with intercept ($y \approx \hat{y} = a_0 + \sum_{1 \leq j \leq m} a_j x_j$) is reduced to the case without intercept by increasing with one the number of the independent variables ($a_{m+1} \leftarrow a_0$; $x_{m+1} \leftarrow 1$; and $m \leftarrow m + 1$, when $y \approx \hat{y} = \sum_{1 \leq j \leq m+1} a_j x_j$) and therefore will not be mentioned further. The substitution given as (2) transforms the distribution from univariate to a multivariate one and can be mathematically characterized by a series of properties, such as is given in [29] (results applicable resizing x from 0 and $x_0 \leftarrow y$).

Let us take a sample of n paired measurements (e.g., $(y_i, (x_{j,i})_{1 \leq j \leq m})_{1 \leq i \leq n}$, where n is the number of paired measurements and m is the number of independent measures). The likelihood for the sample is

$$\text{LMLRGL}(\cdot) = \sum_{i=1}^n \ln(\text{MLRGL}(\cdot)). \quad (3)$$

Doing the substitution given in (2) and expressing in full its parameters, the expression of the likelihood function from (3) becomes

$$\begin{aligned} \text{LMLRGL}\left((a_j)_{1 \leq j \leq m}, \sigma, q\right) \\ = n \cdot \ln\left(\frac{q}{2\sigma} \frac{\Gamma^{1/2}(3/q)}{\Gamma^{3/2}(1/q)}\right) \\ - \sigma^q \left(\frac{\Gamma(1/q)}{\Gamma(3/q)}\right)^{q/2} \sum_{i=1}^n \left|y_i - \sum_{1 \leq j \leq m} a_j x_{j,i}\right|^q. \end{aligned} \quad (4)$$

The likelihood is at maximum when all its partial derivatives are zero:

$$\begin{aligned} 0 = \frac{\partial}{\partial a_1} \text{LMLRGL}\left((a_j)_{1 \leq j \leq m}, \sigma, q\right) = \dots \\ = \frac{\partial}{\partial q} \text{LMLRGL}\left((a_j)_{1 \leq j \leq m}, \sigma, q\right). \end{aligned} \quad (5)$$

2.2. Simplification of the Problem. The problem of finding the maximum of the likelihood is a typical problem of finding the extreme points, but not easy to be solved because it depends on a large number of variables. The easiest way is to eliminate

one variable, namely, σ . The derivative of LMLRGL by σ provides the value of σ :

$$n = q \cdot \frac{S}{\sigma^q \cdot (\Gamma(1/q) / \Gamma(3/q))^{q/2}}, \quad (6)$$

$$S(q, (a_j)_{1 \leq j \leq m}) = \sum_{i=1}^n \left| y_i - \sum_{1 \leq j \leq m} a_j x_{j,i} \right|^q.$$

Please note that $S = S(q, (a_j)_{1 \leq j \leq m})$ does not depend on σ . Therefore, let $\text{LMLRGLS}(S, n, q)$ be the function having this constraint. After some calculations, the expression of $\text{LMLRGLS}(S, n, q)$ is

$$\text{LMLRGL}(S, n, q) = \frac{n}{q} \left(q \cdot \ln \left(\frac{n^{1/q} q^{1-1/q}}{2S^{1/q} \cdot \Gamma(1/q)} \right) - 1 \right) \quad (7)$$

$$\begin{aligned} S &= S(q, (a_j)_{1 \leq j \leq m}) \\ &= \sum_{i=1}^n \left| y_i - \sum_{j=1}^m a_j x_{j,i} \right|^q \\ &= \sum_{i=1}^n T(i, (a_j)_{1 \leq j \leq m})^{q/2}, \end{aligned} \quad (8)$$

where $T(i, (a_j)_{1 \leq j \leq m}) = (y_i - \sum_{1 \leq j \leq m} a_j x_{j,i})^2$.

On the other hand, only S depends on $(a_j)_{1 \leq j \leq m}$, and therefore when the derivative of S relative to a_1, \dots, a_m is zero, then also the derivative of the maximum likelihood estimation function (either any of LMLRGL and LMLRGLS) is zero. Doing the partial derivatives of S , with the above given substitution (function T), the following equation is the result:

$$\sum_{i=1}^n y_i x_{u,i} W(i, q) = \sum_{j=1}^m a_j \cdot \sum_{i=1}^n x_{u,i} x_{j,i} W(i, q); \quad (9)$$

for $1 \leq u \leq m$,

where $W(i, q) = (y_i - \sum_{1 \leq j \leq m} a_j x_{j,i})^{q-2}$.

At this point only the expression of the likelihood function (see (7) and (8)) must be included in the new statement of the problem (see (9)) in order to keep in full the derivatives constraints of the initial problem (see (4) and (5)). There is no obvious further reduction of the problem. However, revising the results obtained till this point, the cancelling of the (likelihood function) derivative relative to σ was included at the beginning of the simplification (see (6)) while the cancelling of the (likelihood function) derivatives relative to the regression coefficients a_1, \dots, a_m is equivalent to the previous given equation for the regression coefficients (see (9)). On the other hand, (7)–(9) facilitate an iterative solution of the problem.

3. Fixed-Point Theory for Iterating to the Optimal Solution

A convenient notation was used in (9) to suggest the further treatment of the problem. Actually, Fisher and Mackenzie

proposed for the first time to use such numerical treatment in statistics [30]. This is based on the assumption that, near to the optimal solution, an iterative evaluation of the coefficients (here q and $(a_j)_{1 \leq j \leq m}$) conducted using their previous values (hidden in (9) inside of the function $W(i, q)$) leads to the optimum. The optimum is obtained when no significant change from one step to another is on their values, and, at that time, the $W(i, q)$ function acts as the argument of a contraction mapping [31].

There are some inconveniences for a smooth application of the fixed-point theory. One of them is that the obtaining of the maximum of the LMLRGLS function (see (7), being obtained for known values of S and n (where $(\partial/\partial_q)\text{LMLRGLS}(S, n, q)$ vanishes)) is not a very simple problem; it is expected from its explicit expression to have more than one local maximum. Fortunately, some clues exist, such as the domain of q (ranging from 0) and the expectance from the power of the error (here let us say that it is expected to have q from 0.1 to 10 and is very unlikely but possible to have q from 0.01 to 100, but outside of this range also precision of computations often fails). But the biggest inconvenience is that (9) is not an equation, but a system of equations, and here we may only provide different strategies of iteration, hoping that at least one of them provides the contraction mapping. Namely, we may

- (i) start from some initial values of the regression coefficients $((a_{j,0})_{1 \leq j \leq m})$ and for the power of the error q_0 ;
- (ii) use initial values to obtain the likelihood function LMLRGLS (from (7)) as a function depending only on q ; it requires only the evaluation of S (see (8));
- (iii) find the maximum (let this be ϑ) of LMLRGLS function from (7) (where its derivative is 0 and the point is a global extreme point);
- (iv) prepare starting of a loop on k , by setting it to 0 ($k \leftarrow 0$);
- (v) it is possible, especially at the beginning of the iteration (when $k = 0$), that ϑ and q_k be largely different one to each other; a major change in q will accelerate the convergence but will also increase the likelihood of divergence; therefore, use the new (ϑ) and old (q_k) value to indicate the gradient of the change in q , such as $q_{k+1} \leftarrow \delta \cdot q_k + (1 - \delta) \cdot \vartheta$, with small $0 < \delta < 1$ to be determined;
- (vi) do a loop ($k \leftarrow k+1$) using the new value of q (namely, q_{k+1}) to calculate the new values of the coefficients $((a_{j,k+1})_{1 \leq j \leq m})$ with (9) and using the new values of the coefficients calculate the new value of q (turn back to find the maximum from (7));
- (vii) repeat until $((a_{j,k+1})_{1 \leq j \leq m})$ and (q_{k+1}) have no changes during iteration.

At arriving in the stationary point, all criteria for the maximization of the likelihood are accomplished; namely, the equations corresponding to all derivatives cancellation are assured. The great advantage of this proposed method is that it reduces the problem of finding the maximum of a

Require: $1 \leq m \leq n$ {Regression multiplicity & Sample size}
 Require: $(y_i, (x_{k,i})_{1 \leq k \leq m})_{1 \leq i \leq n}$ {Data}
 Require: $q_j, (a_{k,j})_{1 \leq k \leq m}$ {Estimates previously obtained}

$$S \leftarrow \sum_{i=1}^n \left(y_i - \sum_{k=1}^m a_{k,j} x_{k,i} \right)^{q_j}$$

ALGORITHM 1: Calculate “S” at some step “j” from (8).

Require: $1 < n$ {Sample size}
 Require: S {Calculated with Algorithm 1}
 $q \leftarrow \vartheta$ from: {finding of the maxima of a univariate function}

$$0 = \frac{\partial}{\partial \vartheta} \left[\frac{n}{\vartheta} \cdot \left(\vartheta \cdot \ln \left(\frac{n^{1/\vartheta} \cdot \vartheta^{1-1/\vartheta}}{2 \cdot S^{1/\vartheta} \cdot \Gamma(1/\vartheta)} \right) - 1 \right) \right]$$

ALGORITHM 2: Calculate “q” at some step “j” from (7).

function with $m + 3$ variables to the finding of the maximum of a function with one variable (q), in a repeated process, of course.

The disadvantage is that the evolution is through a contracting functional of which contraction cannot be assured all the time. This is the reason why there are different strategies of finding such kind of contracting functional (see example 6.1 in [32] for construction of a contracting functional from resampling).

Some calculations are the same regardless the strategy used and are given in the next as Algorithms 1–3.

One strategy is to use the equations from cancellation of regression coefficients derivatives (see (9)) to iterate the values of the coefficients, while another one is to treat (9) as a system of equations and to solve it as a whole (Algorithm 4).

Another strategy that is required to be specified is that if (7)–(9) is used to simply iterate for new values or if (9) should be used in a loop to converge to new estimates for the coefficients associated with the new q (Algorithm 5). The expected assumption is that the errors are normally distributed ($q = 2$) and the optimal solution of (5) is near to this.

The contingency of 2×2 strategies given above was tested on sampled data (see Section 4), and the pair (Algorithms 4 and 5) turned out to be the only one providing a contraction functional. Thus, for convenience, the working algorithm is given in full (see Algorithm 6) and was used to obtain the results given in the next section.

In order to assure the numerical stability of the calculations, Algorithm 6 was used with fixed and reasonable value $\varepsilon = 10^{-5}$, and in order to assure a smooth convergence, the value of the new error’s power estimate (q_j) was replaced by an exponential smoothing (a technique commonly applied to time series [33]), $q_j \leftarrow 0.1 \cdot q_j + 0.9 \cdot q_{j-1}$.

Therefore, in all scenarios, the initial (starting) values of the estimates to be determined will be the one given by the

classical multiple linear regression models as presented in the following:

$$q_0 \leftarrow 2;$$

$$(a_{k,0})_{1 \leq k \leq m} \leftarrow \text{MLR}(y, x_1, \dots, x_k), \quad (10)$$

$$\text{MLR}(y, x_1, \dots, x_k) = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y},$$

where $\text{MLR}(y, x_1, \dots, x_k)$ uses the classical strategy of ordinary least squares $((\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y})$ to find the parameters.

4. Case Study

Ten datasets of chemical compounds with different sample size (Table 1) along with their measured outcome activity were considered to illustrate Algorithms 1–6.

For all datasets, the experimental values of the dependent variable (y) and for the selected previously reported independent variables (under the assumption of the normal distribution of the error) on multiple linear regressions with two ($m = 2$) independent variables are given in Table 2.

Different descriptors (independent variables) were used to explain the activity/property of interest on models presented in Table 2. The names of these descriptors are

- (i) TIE: state topological parameter [20];
- (ii) TICl: total information content index (neighborhood symmetry of 1) [20];
- (iii) IHDMkMg and IHDDFMg: MDF descriptors [21];
- (iv) SAG: molecular surface area grid; $f(0)_n$: Fukui index [22];
- (v) TPSA(NO): topological polar surface area expressed by nitrogen and oxygen contributions; Aegm: Dragon descriptor [23];
- (vi) RDF035m: radial distribution function on a spherical volume of a 3.5 Å radius weighted by atomic mass;

Require: $1 \leq m \leq n$ {Regression multiplicity & Sample size}
 Require: $(y_i, (x_{k,i})_{1 \leq k \leq m})_{1 \leq i \leq n}$ {Data}
 Require: $q_j, (a_{k,j})_{1 \leq k \leq m}$ {Estimates previously obtained}
 for $i = 1$ to n do

$$W_i \leftarrow \left(y_i - \sum_{k=1}^m a_{k,j} x_{k,i} \right)^{q_j-2}$$

 end for

ALGORITHM 3: Calculate “ $(W_i)_{1 \leq i \leq n}$ ” at some step “ j ” from (9).

Require: $1 \leq m \leq n$ {Regression multiplicity & Sample size}
 Require: $(y_i, (x_{k,i})_{1 \leq k \leq m})_{1 \leq i \leq n}$ {Data}
 Require: $(W_i)_{1 \leq i \leq n}$ { $(W_i)_{1 \leq i \leq n}$ iterated at step j by Algorithm 3}
 $(a_{u,j+1})_{1 \leq u \leq m} \leftarrow (a_u)_{1 \leq u \leq m}$ {Solution of linear & homogeneous system Eq. (9)} from

$$\left\{ \sum_{i=1}^n y_i x_{u,i} W_{i,j} = \sum_{k=1}^m \alpha_k \sum_{i=1}^n x_{k,i} x_{u,i} W_{i,j} \right\}_{1 \leq u \leq m}$$

ALGORITHM 4: Block solves providing “ $(a_{u,j+1})_{1 \leq u \leq m}$ ” at some step “ j ” with (9).

Require: $1 \leq m \leq n$ {Regression multiplicity & Sample size}
 Require: $(y_i, (x_{k,i})_{1 \leq k \leq m})_{1 \leq i \leq n}$ {Data}
 Require: $q_0, (a_{k,0})_{1 \leq k \leq m}$ {Coefficients iterated at step 0}
 $j \leftarrow 0$
 repeat
 $j \leftarrow 0 + 1$
 $S_j \leftarrow$ Algorithm 1
 $q_j \leftarrow$ Algorithm 2
 repeat
 $(W_i)_{1 \leq i \leq n} \leftarrow$ Algorithm 3
 $(a_{u,j+1})_{1 \leq u \leq m} \leftarrow$ Algorithm 4
 until $\sum_{1 \leq k \leq m} |a_{k,j+1}/a_{k,j-1}| < \varepsilon$
 until $|q_{j+1}/q_{j-1}| + \sum_{1 \leq k \leq m} |a_{k,j+1}/a_{k,j-1} - 1| < \varepsilon$

ALGORITHM 5: Double loop with (9) for (7) and (8).

Require: $1 \leq m \leq n$ {Regression multiplicity & Sample size}
 Require: $(y_i, (x_{k,i})_{1 \leq k \leq m})_{1 \leq i \leq n}$ {Data}
 $q_0 \leftarrow 2; (a_{k,0})_{1 \leq k \leq m} \leftarrow (X^T X)^{-1} X^T y, j \leftarrow 0$
 repeat

$$S \leftarrow \sum_{i=1}^n \left(y_i - \sum_{k=1}^m a_{k,j} x_{k,i} \right)^{q_j}$$

$$q_j \leftarrow \vartheta \text{ s.t. } 0 = \frac{\partial}{\partial \vartheta} \left[n \cdot \left(\ln \left(\frac{n^{1/\vartheta} \cdot \vartheta^{1-1/\vartheta}}{2 \cdot S^{1/\vartheta} \cdot \Gamma(1/\vartheta)} \right) - \vartheta \right) \right] \& \left[n \cdot \left(\ln \left(\frac{n^{1/\vartheta} \cdot \vartheta^{1-1/\vartheta}}{2 \cdot S^{1/\vartheta} \cdot \Gamma(1/\vartheta)} \right) - \vartheta \right) \right] = \max.$$

 repeat

$$(W_i)_{1 \leq i \leq n} \leftarrow \left(y_i - \sum_{k=1}^m a_{k,j-1} x_{k,i} \right)^{q_j-2}$$

$$(a_{u,j+1})_{1 \leq u \leq m} \leftarrow \left\{ \sum_{i=1}^n y_i x_{u,i} W_{i,j} = \sum_{k=1}^m a_{k,j} \sum_{i=1}^n x_{k,i} x_{u,i} W_{i,j} \right\}_{1 \leq u \leq m}$$

 until $\sum_{1 \leq k \leq m} |a_{k,j+1}/a_{k,j-1}| < \varepsilon$
 until $|q_{j+1}/q_{j-1}| + \sum_{1 \leq k \leq m} |a_{k,j+1}/a_{k,j-1} - 1| < \varepsilon$

ALGORITHM 6: Contraction functional for MLR-MLE-GL.

TABLE 1: Datasets characteristics.

Set	Sample size (n)	Class	Property/activity	Reference
1	132	Estrogens	Estrogen binding affinity—log RBA	[20]
2	37	Carboquinone derivatives	Minimum effective dose (MED)—log(1/MED)	[21]
3	33	Organic pollutants	Oxidative degradation—log(k')	[22]
4	97	Benzotriazoles	Fish toxicity—pEC50	[23]
5	136	Thiophene and imidazopyridine derivatives	Inhibition of polo-like kinase 1—pIC ₅₀	[24]
6	14	Substituted phenylaminoethanones	Average antimicrobial activity—pMICam	[25]
7	110	Acetylcholinesterase inhibitors	Inhibition activity—pIC ₅₀	[26]
8	107	Polychlorinated biphenyl ethers	298 K supercooled liquid vapor pressures—log(p_L)	[27]
9	107	Polychlorinated biphenyl ethers	Aqueous solubility—log(S_w, L)	[27]
10	47	Para-substituted aromatic sulphonamides	Carbonic anhydrase II inhibitors—log(K_i)	[28]

TABLE 2: Reported bivariate models.

Set	Model under assumption of normal errors	Determination coefficient (r^2)
1	$-4.284 - 0.0263 \cdot \text{TIE} + 0.0368 \cdot \text{TIC1}$	0.3976
2	$7.780 - 579 \cdot \text{IHDMkMg} + 0.049 \cdot \text{IHDDFMg}$	0.7700
3	$-2.703 + 0.00515 \cdot \text{SAG} + 9.703 \cdot f(0)_n$	0.6859
4	$4.110 - 0.0172 \cdot \text{TPSA(NO)} + 0.0097 \cdot \text{Aeigm}$	0.7161
5	$2.5651 + 0.1899 \cdot \text{RDF035m} + 2.9825 \cdot \text{Small-RSI-mol}$	0.5101
6	$0.780 + 0.0339 \cdot {}^0\chi^v + 0.004 \cdot \mu$	0.8357
7	$5.446 + 0.716 \cdot \text{nR10} + 1.113 \cdot \text{N-070}$	0.6838
8	$1.476 - 0.588 \cdot \text{N}_{\text{Cl}} - 5.029 \cdot 10^{-2} \cdot V_s^+$	0.9880
9	$-4.080 - 0.880 \cdot \text{N}_{\text{Cl}} + 5.996 \cdot \sigma_{\text{tot}}^2$	0.9619
10	$4.055 - 0.154 \cdot {}^0\chi^v - 1.284 \cdot \text{FNSA1}$	0.7058

TABLE 3: Differences between values of coefficients obtained by classical linear regression approach compared to the proposed approach.

Set	diff(q)	diff(a_0)	diff(a_1)	diff(a_2)	diff(LMLRGL)	diff(σ)
1	0.3400	-0.00073	-0.00315	0.24400	-0.30000	-0.00200
2	-0.4150	-0.00034	-16.30000	0.17400	-0.10100	-0.00020
3	-0.3830	-0.28700	0.00009	-0.04000	-0.06000	-0.00030
4	-0.1680	0.00006	0.00007	-0.01400	-0.05000	0.00000
5	0.9420	0.34500	-0.00880	-0.62400	-6.10000	-0.00850
6	0.5000	0.00027	0.00078	-0.02140	-0.09000	-0.00006
7	0.5210	-0.10300	0.03490	-0.01800	-1.10000	0.00030
8	-0.5690	0.00090	-0.00330	-0.01100	-0.42000	-0.00010
9	-0.4370	-0.27700	-0.00020	0.04000	-0.30000	-0.00020
10	-0.9370	0.01400	-0.00700	0.06000	-0.70600	0.49310

diff: difference between value obtained by classical approach and value obtained by the proposed approach.

a_0 , a_1 , and a_2 : coefficients of the independent variables; q : power of the error (Algorithm 6 for the proposed approach).

σ : population standard deviation; LMLRGL: likelihood for multiple linear regressions under assumption of GL distribution.

small-RSI-mol: the smallest value of atomic steric influence in a molecule [24];

(vii) ${}^0\chi^v$: Kier's molecular connectivity index; μ : dipole moment [25];

(viii) nR10: number of 10-membered rings; N-070: number of Ar-NH-Al fragments [26];

(ix) N_{Cl} : the number of the chlorine atoms on the two phenyl rings; V_s^+ : the surface maxima values of the electrostatic potential; and σ_{tot}^2 : total variance of the electrostatic potential at a point r_i [27];

(x) FNSA1: fractional partial positive surface area 1 $\text{PPS}_{\text{Al}}/\text{TMA}_S$; where PPS_A = Partial Positive Surface Area and TMS_A = Total Molecular Surface Area.

All sets subjected to analysis converged maximizing the likelihood and Table 3 provides the differences between values obtained by classical MLR approach and values obtained by the proposed approach (MLR-MLE-GL).

The results presented in Table 3 reveal different estimates for the coefficients in the assumption of the more general generalized Gauss-Laplace distribution of the error. In 6 out

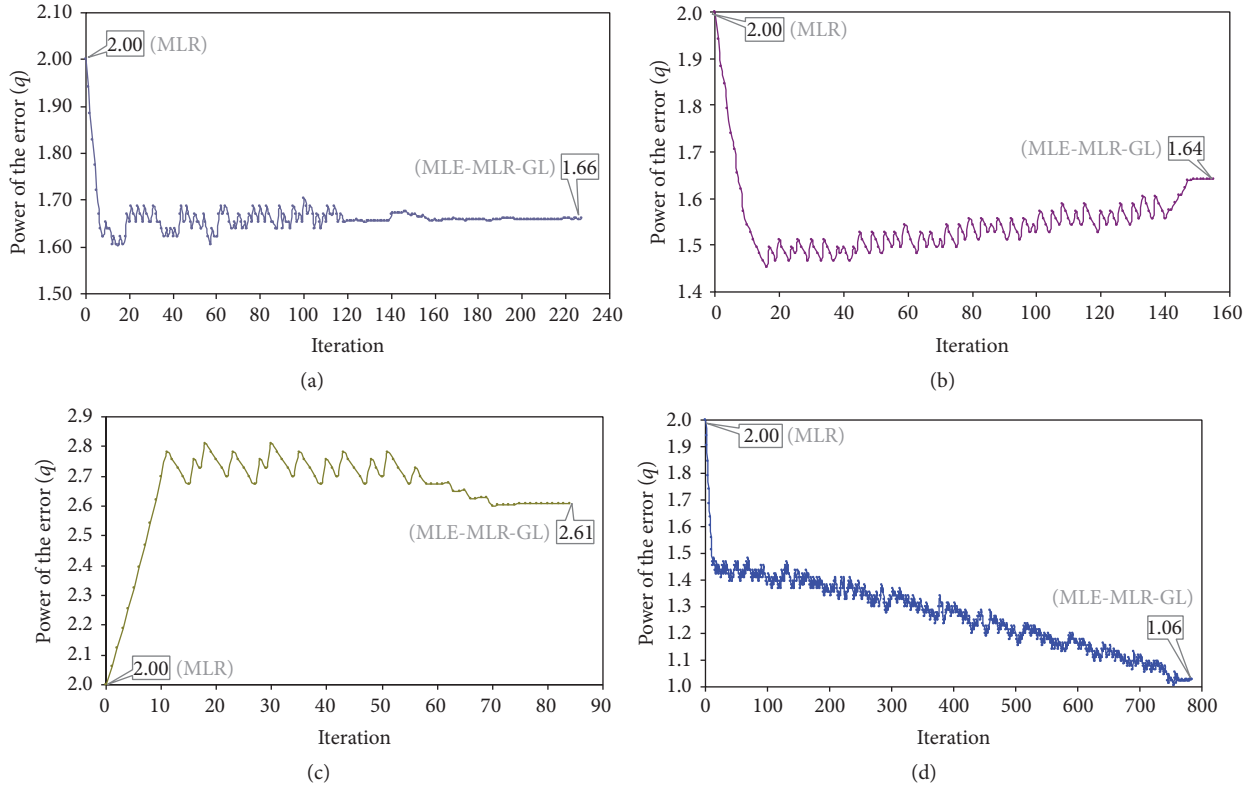


FIGURE 1: Evolution of the power of the errors (q) by optimization iteration: (a) set 1 (converged at 226); (b) set 6 (converged at 154); (c) set 8 (converged at 83); and (d) set 5 (converged at 784).

of 10 cases, the power of the error proved to be higher compared with convenient value of 2, the highest values being observed for set 10 ($q = 2.937$). Opposite, the power of the error proved to be almost half of the expected value for set 5 ($q = 1.058$). The values of coefficients obtained by applying the MLE and the proposed approach were close to each other in two cases (set 4 and set 8). With one exception, represented by set 2, the sum of the absolute differences of a_0 , a_1 , and a_2 was less than 1. The values obtained for the population standard deviation by the two investigated methods proved to be closest to each other, with highest difference of 0.49310 observed on set 10.

The power of the error follows different patterns according to the model, decrease-fluctuation-plateau (set 1, Figure 1(a)), decrease-increase in steps-plateau (set 6, Figure 1(b)), increase-fluctuation-decrease in steps-plateau (set 8, Figure 1(c)), and decrease-fluctuation (set 5, Figure 1(d)).

A question (hypothesis) can be raised about the power of the error: *if its distribution can be assumed normal*. This hypothesis (the distribution of the power of the error can be assumed to be normal) can be tested on the results even if the sample is small (10 cases) to provide an answer. However, the tendency to have a mean of two in convergence is clear ($\bar{q} = 2.06$ from the 10 cases) and the hypothesis of its normality cannot be rejected (Anderson-Darling statistic measures that only 14.72% ($p_{\text{to-reject}} = 0.8528 > 0.05$) of the random samples are in better agreement with the normal distribution

while Kolmogorov-Smirnov statistic measures only 28.7% ($p_{\text{to-reject}} = 0.713 > 0.05$)).

5. Conclusions

The proposed algorithm (Algorithm 6 in this paper) was found to provide an appropriate contraction mapping to be used for maximum likelihood estimation of the multiple linear regression parameters in the generalized Gauss-Laplace distribution assumption of the measurement's errors. The analysis conducted on 10 samples demonstrated that, in general, it is not appropriate to assume that the measurement error is normally distributed, and when it is possible a deeper treatment of the distribution of the error need to be conducted. From a sample of 10 cases, the analysis of the distribution of the error showed that the normal distribution of the power of the error could not be rejected, being very likely to have a mean equal to two.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] F. Galton and H. Section, "The british association reports," *Nature*, vol. 32, pp. 502–510, 1885.

- [2] F. Galton, "Regression towards mediocrity in hereditary stature," *Journal of the Anthropological Institute of Great Britain and Ireland*, vol. 15, pp. 246–263, 1886.
- [3] A. Bravais, "Analyse mathématique sur les probabilités des erreurs de situation d'un point," *Mémoires par Divers Savants*, vol. 9, pp. 255–332, 1846.
- [4] K. Pearson, "Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia," *Philosophical Transactions of the Royal Society of London*, vol. 187, pp. 253–318, 1896.
- [5] R. A. Fisher, "On an absolute criterion for fitting frequency curves," *Messenger of Mathematics*, vol. 41, pp. 155–160, 1912.
- [6] R. A. Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, vol. 10, no. 4, pp. 507–521, 1915.
- [7] H. E. Soper, A. W. Young, B. M. Cave, A. Lee, and K. Pearson, "On the distribution of the correlation coefficient in small samples. appendix II to the papers of 'student' and R. A. Fisher," *Biometrika*, vol. 11, no. 4, pp. 328–413, 1917.
- [8] R. A. Fisher, "On the 'probable error' of a coefficient of correlation deduced from a small sample," *Metron*, vol. 1, pp. 3–32, 1921.
- [9] R. A. Fisher, "The distribution of the partial correlation coefficient," *Metron*, vol. 3, pp. 329–332, 1924.
- [10] L. Jäntschi, L. L. Pruteanu, A. C. Cozma, and S. D. Bolboacă, "Inside of the linear relation between dependent and independent variables," *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 360752, 11 pages, 2015.
- [11] A. M. Legendre, *Nouvelles Méthodes pour la Détermination des Orbites des Comètes*, F. Didot, Paris, France, 1805.
- [12] S. M. Stigler, "Gauss and the invention of least squares," *The Annals of Statistics*, vol. 9, no. 3, pp. 465–474, 1981.
- [13] R. A. Fisher, "Theory of statistical estimation," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 22, no. 5, pp. 700–725, 1925.
- [14] J. A. Nelder and R. W. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society. Series A*, vol. 135, no. 3, pp. 370–384, 1972.
- [15] L. Jäntschi and S. D. Bolboacă, "Observation vs. observable: maximum likelihood estimations according to the assumption of generalized gauss and laplace distributions," *Leonardo Electronic Journal of Practices and Technologies*, vol. 8, no. 15, pp. 81–104, 2009.
- [16] C. F. Gauss, *Theoria Motus Corporum Coelestium*, Perthes et Besser, Hamburg, Germany, 1809, (translated in 1857 as: *Theory of Motion of the Heavenly Bodies Moving about the Sun in Conic Sections* (trans. C. H. Davis), Little Brown, Boston, Mass, USA and reprinted in 1963, Dover, New York, NY, USA).
- [17] P. S. Laplace, *Théorie Analytique des Probabilités*, Courcier, Paris, France, 1812.
- [18] M. K. Varanasi and B. Aazhang, "Parametric generalized Gaussian density estimation," *The Journal of the Acoustical Society of America*, vol. 86, no. 4, pp. 1404–1415, 1989.
- [19] T. Makelainen, K. Schmidt, and G. Styan, "On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples," *The Annals of Statistics*, vol. 9, no. 4, pp. 758–767, 1981.
- [20] J. Li and P. Gramatica, "The importance of molecular structures, endpoints' values, and predictivity parameters in QSAR research: QSAR analysis of a series of estrogen receptor binders," *Molecular Diversity*, vol. 14, no. 4, pp. 687–696, 2010.
- [21] S. D. Bolboacă and L. Jäntschi, "Comparison of quantitative structure-activity relationship model performances on carboquinone derivatives," *TheScientificWorldJournal*, vol. 9, pp. 1148–1166, 2009.
- [22] L. Jia, Z. Shen, W. Guo et al., "QSAR models for oxidative degradation of organic pollutants in the Fenton process," *Journal of the Taiwan Institute of Chemical Engineers*, vol. 46, pp. 140–147, 2015.
- [23] S. Cassani, S. Kovarich, E. Papa, P. P. Roy, L. van der Wal, and P. Gramatica, "Daphnia and fish toxicity of (benzo)triazoles: validated QSAR models, and interspecies quantitative activity-activity modelling," *Journal of Hazardous Materials*, vol. 258–259, pp. 50–60, 2013.
- [24] N. C. Comelli, P. R. Duchowicz, and E. A. Castro, "QSAR models for thiophene and imidazopyridine derivatives inhibitors of the Polo-Like Kinase 1," *European Journal of Pharmaceutical Sciences*, vol. 62, pp. 171–179, 2014.
- [25] D. Verma, P. Kumar, B. Narasimhan et al., "Synthesis, antimicrobial, anticancer and QSAR studies of 1-[4-(substituted phenyl)-2-(substituted phenyl azomethyl)-benzo[b]-[1,4]diazepin-1-yl]-2-substituted phenylaminoethanones," *Arabian Journal of Chemistry*, 2015.
- [26] M. D. Vitorović-Todorović, I. N. Cvijetić, I. O. Juranić, and B. J. Drakulić, "The 3D-QSAR study of 110 diverse, dual binding, acetylcholinesterase inhibitors based on alignment independent descriptors (GRIND-2). the effects of conformation on predictive power and interpretability of the models," *Journal of Molecular Graphics and Modelling*, vol. 38, pp. 194–210, 2012.
- [27] X. Hui-Ying, Z. Jian-Wei, H. Gui-Xiang, and W. Wei, "QSPR/QSAR models for prediction of the physico-chemical properties and biological activity of polychlorinated diphenyl ethers (PCDEs)," *Chemosphere*, vol. 80, no. 6, pp. 665–670, 2010.
- [28] J. Singh, B. Shaik, S. Singh et al., "Comparative QSAR study on para-substituted aromatic sulphonamides as CAII inhibitors: information versus topological (distance-based and connectivity) indices," *Chemical Biology & Drug Design*, vol. 71, no. 3, pp. 244–259, 2008.
- [29] F. Sinz, S. Gerwinn, and M. Bethge, "Characterization of the p -generalized normal distribution," *Journal of Multivariate Analysis*, vol. 100, no. 5, pp. 817–820, 2009.
- [30] R. A. Fisher and W. A. Mackenzie, "Studies in crop variation. II. The manurial response of different potato varieties," *The Journal of Agricultural Science*, vol. 13, no. 3, pp. 311–320, 1923.
- [31] I. A. Rus, A. Petruşel, and M. A. Şerban, "Weakly Picard operators: equivalent definitions, applications and open problems," *Fixed Point Theory*, vol. 7, no. 1, pp. 3–22, 2006.
- [32] C. A. Klaassen and H. Putter, "Efficient estimation of Banach parameters in semiparametric models," *The Annals of Statistics*, vol. 33, no. 1, pp. 307–346, 2005.
- [33] L. Jäntschi, *Genetic algorithms and their applications [Ph.D. thesis in Horticulture] [Ph.D. thesis]*, University of Agricultural Sciences and Veterinary Medicine, Cluj-Napoca, Romania, 2010, PhD Advisor R. E. Sestraş (Romanian).

