

A Simulation Study for the Distribution Law of Relative Moments of Evolution

LORENTZ JÄNTSCHI,^{1,2} SORANA D. BOLBOACĂ,³ AND RADU E. SESTRĂȘ¹

¹Department of Genetics and Plant Breeding, University of Agricultural Science and Veterinary Medicine Cluj-Napoca, Cluj-Napoca 400372, Romania; ²Department of Physics and Chemistry, Technical University of Cluj-Napoca, Cluj-Napoca 400641, Romania; and ³Department of Medical Informatics and Biostatistics, "Iuliu Hatieganu" University of Medicine and Pharmacy Cluj-Napoca, Cluj-Napoca 400349, Romania

Received January 28, 2012; accepted April 26, 2012

Nine selection-survival strategies were implemented in a genetic algorithm experiment, and differences in terms of evolution were assessed. The moments of evolution (expressed as generation numbers) were recorded in a contingency of three strategies (i.e., proportional, tournament, and deterministic) for two moments (i.e., selection for crossover and mutation and survival for replacement). The experiment was conducted for the first 20,000 generations in 46 independent runs. The relative moments of evolution (where evolution was defined as a significant increase in the determination coefficient relative to the previous generation) when any selection-survival strategy was used fit a Log-Pearson type III distribution. Moreover, when distributions were compared to one another, functional relationships were identified between the population parameters, revealing a degeneration of the Log-Pearson type III distribution in a one-parametrical distribution that can be assigned to the chosen variable—evolution strategy. The obtained theoretical population distribution allowed comparison of the selection-survival strategies that were used. © 2012 Wiley Periodicals, Inc. *Complexity* 17: 52–63, 2012

Key Words: genetic algorithm; evolution; molecular descriptors family; quantitative structure-activity relationship; multivariate linear regression

INTRODUCTION

The issue of optimizing quantitative structure–activity relationships (QSARs) belongs to an interdisciplinary field that includes chemistry, informatics, and biology. The continuous accumulation of information and its orga-

nization into vast databases (e.g., PubMed, PubChem, Genome, etc., which have been developed by the National Institute of Health) has led to the need for efficient technologies capable of processing this huge amount of data.

Genetic algorithms (GAs) have evolved since their introduction and have since become strong informatics tools for solving difficult problems of decision, classification, optimization, and simulation in different research areas [1–6], including drug design [7–9] and especially QSAR analyses [10–12]. Studies of the main operators associated with GAs usually report only the algorithm effectiveness

Corresponding author: Sorana D. Bolboaca, Department of Medical Informatics and Biostatistics, "Iuliu Hatieganu" University of Medicine and Pharmacy Cluj-Napoca, Cluj-Napoca 400349, Romania (e-mail: sbolboaca@umfcluj.ro)

(expressed as the speed required to achieve the objective and the closeness to the global maximum for optimization) [13]. Therefore, different crossing operators [14], mutation and cross operators [15], and dynamic parameters [16] have been studied.

Researchers have often focused on identifying the optimum solution to a difficult problem using a GA [17, 18], although the effectiveness of the GA (expressed as execution time and required memory resources) is also of interest [19, 20]. New selection methods, such as combined data-splitting feature selection [21], mirrored sampling and/or sequential selection [22], three-dimensional feature vectors that integrate the value of the objective function, the degree and number of constraints violations [23], and the keep-best reproduction strategy [24], have been introduced and used to identify solutions to different problems.

However, although it is recognized that selection plays a central role in finding the optimal solution [23], few studies have compared different selection strategies [25]. In the studies that have, the efficiency of the various selection strategies is different in different contexts. For example, Roulette wheel selection worked more efficiently for the production of feasible course timetables [25]. On the other hand, rank selection proved more effective in hill-climbers experiments [26], whereas the keep-best reproduction strategy performed better at solving constrained ordering problems [24]. To date, we have found no studies that have compared the different selection-survival strategies using GAs and conducted comparisons for evolution on QSARs analyses.

Our research aimed to identify and assess the distribution law for the moments of evolution in multiple linear regressions that was applied to the structure–activity relationship for the octan-1-ol/H₂O partition coefficient of polychlorinated biphenyls (PCBs). This article presents the first comparison of different selection-survival strategies of the distribution law for the moments of evolution using a GA on a quantitative structure–activity problem.

MATERIALS AND METHODS

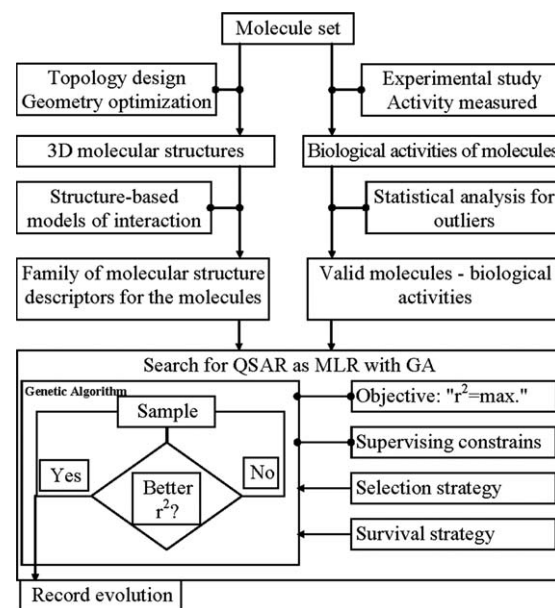
PCBs

The relationship between the octan-1-ol/H₂O partition coefficient (expressed as log K_{ow}) and the structure of PCBs was previously studied using two strategies to identify the number of molecular descriptor family (MDF) descriptors that explain the link between the structure and activity of the compounds [27]. The applied search strategies were systematic (four descriptors [28]) and heuristic (random search in i variables [29]).

QSAR and GA

We conducted an evolution experiment using a GA on nine selection-survival strategies (TT, TD, TP, DT, DD, DP,

FIGURE 1



Workflow of the applied method.

PT, PD, and PP, where T = tournament, D = deterministic, and P = proportional to find the structure–activity relationship of the octan-1-ol/H₂O partition coefficient on a series of PCBs (Figure 1). The supervised evolution was applied to a sample of molecular descriptors drawn from the MDF. MDF individuals were obtained by applying a sequence of operators (the family genome) to the molecular structures.

The molecular structures were topologically designed and geometrically optimized using the OPLS (Orthogonal Partial Least-Squares) molecular mechanics [30] and PM3 (Parameterized Model number 3) semiempirical methods [31, 32] until convergence was assured less than a 1% gradient. The values for the octanol/water partition coefficient of the PCBs were taken from the literature, whenever valid measurements recorded in the same environmental conditions were available. The supervised evolution analysis was conducted using different selection-survival strategies; therefore, we supplied our GA [33] with the same input data for 46 repeated runs. The following criteria were instituted for each selection-survival strategy: (1) eight chromosomes were in the cultivar; (2) 20,000 was the maximum number of generations; (3) the probability of mutation was set to 0.05; (4) two genes were implied in the mutation; (5) multilinear models had two variables; (6) the determination coefficient was the parameter to be optimized; and (7) maximum value of the determination coefficient was the optimization objective. The moments

TABLE 1

Example Demonstrating the Transformation of the Evolution Value into a Relative Moment of Evolution

Generation number	0	15	136	188	246	528	5423	11887
Evolution moment	1	16	137	189	247	529	5424	11888
Duration until the next evolution	15	121	52	58	282	4895	6464	
Relative duration according to evolution	15.0000 ^a	7.5625 ^b	0.3796	0.3069	1.1417	9.2533	1.1917	1.7×10^{-4}

^a(16 - 1)/1 = 15.

^b(137 - 16)/16 = 7.5625.

of evolution were recorded and further analyzed. The results obtained after the 46 independent runs of the GA were used to identify the distribution law of the relative moments of evolution.

Assessment of the Distribution Law

The moments when the evolution occurred (defined as a significant increase in the determination coefficient) were transformed to relative data by dividing the difference between the two consecutive generations when evolution occurred by the time of evolution (the first of these two generations; see Table 1 for example). The relative durations were obtained for each run (1..46) and for each of the applied selection-survival strategies. A total of 11,347 moments of evolution were the input data for analyzing the distribution law. The EasyFitXL (version 5.1) was used to test the distribution law of the relative moments of evolution as an intermediary step for the initial selection of the probable distribution function. A total of 65 distribution laws for quantitative variables were applied in the intermediary step. Specifically, the Kolmogorov–Smirnov [34, 35], Anderson–Darling [36], and Chi-squared [37] tests were used, and their associated probabilities were set to a significance level of 5% to identify the most suitable distribution law.

The identified probability distribution function was validated by analyzing each pair of survival-selection strategies. Furthermore, the parameters of the probability distribution function were investigated using the maximum likelihood estimation (MLE) approach [38, 39]. This approach was applied to reduce the inference of the parameters in the observed probabilities that were calculated using the Kolmogorov–Smirnov [34, 35], Anderson–Darling [36], and Chi-Squared [37] statistics. Dedicated software that controls for identified relationships among the parameters of a probability distribution function using a maximum likelihood approach was implemented in software that was designed by the authors. The Pearson correlation coefficient (r) [40] and the semiquantitative correlation coefficient (λ) [41] were used to test the association between shape and location, which were two parameters of the identified distribution law. The nonlinear associa-

tion between scale and location, as parameters of the identified distribution law, was tested using the SlideWrite software (Advanced Graphics Software, CA). Finally, a principal components analysis (PCA) was used to isolate the factors that can identify similarities between the statistical parameters that are associated with the identified distribution law (Statistica software version 8, StatSoft, OK).

RESULTS AND DISCUSSION

The top three distribution laws that were identified as being the most suitable for the relative moments of evolution are presented in Table 2.

The analysis of the results presented in Table 2 revealed the following:

- One distribution law (Log-Pearson type III) out of 65 proved to be suitable for defining the relative moments of evolution. The Log-Pearson type III distribution had a risk of error greater than 1%, and in two out of three cases, the errors exceeded 5% (see p_{KS} and p_{AD} , Table 2). Therefore, the above-mentioned observation holds true for the Log-Pearson type III distribution at a significance level of 5.6% for the Anderson–Darling test, 7.7% for the Kolmogorov–Smirnov test, and $\sim 0.08\%$ for the Chi-Squared test.
- In accordance with the calculus method used, the Kolmogorov–Smirnov statistic measures the agreement between the observation ranks, whereas the Chi-Squared statistic measures the agreement between the observed values.
- The agreement between the ranks of the observed relative moments of evolution and the ranks of the Log-Pearson type III distribution is more probable than the agreement between the observed values of the evolution moments and the values of the Log-Pearson type III distribution. This result is expected, because the observations of all the investigated selection-survival strategies were included in the analysis.

Therefore, the results presented in Table 2 reveal that the Log-Pearson type III is the distribution law for the relative moments of evolution.

TABLE 2

Distribution Laws for Relative Moments of Evolution: Results

Dist	Stat								
	KS	p _{KS}	Rank	AD	p _{AD}	Rank	CS ^a	p _{CS}	Rank
Log-Pearson type III	0.01197	0.07683	1	2.4264	0.05617	1	41.731	7.3×10^{-5}	1
Burr	0.01635	4.57×10^{-3}	3	6.7901	3.23×10^{-4}	3	46.345	1.25×10^{-5}	2
Burr-4P	0.01592	6.27×10^{-3}	2	6.0813	7.48×10^{-4}	2	51.408	1.71×10^{-6}	3

Dist = distribution law; Stat = statistics; Rank = the rank of statistics over all 65 alternatives; KS = Kolmogorov–Smirnov statistics; AD = Anderson–Darling statistics; CS = Chi-squared statistics; p = p-value associated with statistics.

^aDegrees of freedom = 13.

The Pearson type III distribution is a particular case of a family of distribution laws that were developed and classified in 1895 by Pearson [40]. The Pearson type III distribution is a normal distribution when the parameters governing the skewness tend to infinity, making it possible to analyze a family of processes that deviate from a normal distribution. The Log-Pearson type III distribution is obtained, when a log transformation is applied to populations of events that are positively skewed. Like the Fisher–Tippett distribution, the Log-Pearson type III distribution is defined by three continuous parameters [42]: form ($\alpha > 0$), scale ($\beta \neq 0$), and location ($\gamma \in \mathcal{R}$). The Log-Pearson type III distribution law not only found its usefulness in environmental studies [43–50] but also has applications in other research fields [51–55].

Our analysis further investigated all the applied selection-survival strategies. These results are included in Table 3.

The results presented in Table 3 supported the following hypothesis: “Log-Pearson type III is the distribution law of the relative moments of evolution regardless of selection-survival strategies.” Accordingly, results presented in Table 3 reveal that:

- The above hypothesis was not rejected at a significance level $\geq 10\%$.
- The PD (Proportional-Deterministic) strategy and the TP (Tournament-Proportional) strategy were rejected by the Kolmogorov–Smirnov and Chi-squared tests at a significance level $\leq 20\%$. This was expected, because 7.4% error (2/27) was observed at a significance level of 20%.
- The Chi-squared statistic rejected the hypothesis with a less than 0.08‰ risk of error in the overall assessment. The risk of rejecting the null hypothesis increased dramatically when each selection-survival strategy was analyzed (minimum value 16%; mean value 53%).

Because a disagreement is observed in Table 2, and an agreement is noted in Table 3, it could be concluded that the parameters of a Log-Pearson type III distribution

depend on the selection-survival strategy. As a result, the parameters of the Log-Pearson type III distribution were further estimated.

The values for shape, scale, and location of the Log-Pearson type III distribution that were applied to the relative moments of evolution are presented in Table 4.

The statistics presented in Table 5 were obtained based on the values presented in Table 4.

An association between shape and location was identified, when all the selection-survival strategies were investigated ($r = 0.994$, r = Pearson correlation coefficient). This relationship could have been attenuated by the MLE approach, as the strength of the relationship decreased from 1 to 0.994. Starting with this hypothesis, the relationship between shape and location was plotted for each selection-survival strategy (Figure 2).

TABLE 3

The Agreement between Observations (Relative Moments of Evolution) and the Theoretical Log-Pearson Type III Distribution

Stra	No. obs	Stat					
		KS	p _{KS}	AD	p _{AD}	CS (df)	p _{CS}
TT	1379	0.02284	0.46	0.63251	0.47	12.3 (10)	0.27
TD	1429	0.01224	0.98	0.23477	0.75	3.3 (10)	0.97
TP	1318	0.02691	0.29	1.2118	0.24	14.7 (10)	0.16
DT	996	0.02845	0.39	0.73496	0.41	10.6 (9)	0.30
DD	1084	0.01919	0.81	0.34184	0.66	8.1 (10)	0.62
DP	851	0.02416	0.69	0.6234	0.47	6.9 (9)	0.65
PT	1463	0.02030	0.58	0.70531	0.43	12.5 (10)	0.25
PD	1474	0.03055	0.13	0.93998	0.33	8.7 (10)	0.56
PP	1353	0.01212	0.99	0.23201	0.75	3.6 (10)	0.97

Stra = survival-selection strategy; Stat = statistics; p = p-value associated with statistics; df = degrees of freedom.

TABLE 4

Shape, Scale, and Location of the Log-Pearson Type III for the Relative Moments of Evolution According to the MLE

SS	Shape (α)	Scale (β)	Location (γ)
TT	27.666	-0.41137	9.0948
TD	140.850	-0.18338	23.409
TP	37.404	-0.37736	11.725
DT	81.525	-0.26053	19.298
DD	211.900	-0.16693	33.136
DP	294.760	-0.14430	40.665
PT	92.711	-0.22803	18.666
PD	48.108	-0.31208	12.615
PP	73.500	-0.25278	16.289

SS = selection-survival strategy.

The observed distributions of the relative moments of evolution in the nine investigated selection-survival strategies were compared using a series of alternative theoretical distributions. The hypothesis for the Log-Pearson type III theoretical distribution cannot be rejected for any of the nine selection-survival strategies. Figure 2 shows the observed relationship established between two (shape and location) of the three parameters (shape, location, and scale), when the MLE method was applied to identify population parameters. Figure 2 illustrates the possibility of a slight attenuation of linear dependence between the parameters in the process of maximization of the likelihood; therefore, the points can be observed close to the line. Two strategies, the Tournament-Tournament and the Deterministic-Tournament, are over- and under-estimated, respectively.

The semiquantitative correlation coefficient was also calculated, and the results were as follows:

- For the {TP, PD, PP, PT, TD, DD, DP} selection-survival strategies: $\lambda = 0.998$, $t(7,2, \lambda) = 51$, $p_t(t,5) = 5 \times 10^{-8}$ (where t = statistics associated with the correlation coefficient and p_t = p -value).
- For the {TP, PD, PP, PT, TD, DD, DP, TT, DT} selection-survival strategies: $\lambda = 0.988$, $t(9,2, \lambda) = 17$, $p_t(t,7) = 5 \times 10^{-7}$.

The mathematical model of the linear regression analysis for the data associated with the {TP, PD, PP, PT, TD, DD, DP} selection-survival strategies (Table 4) [56] was:

$$b_1 \cdot \alpha + b_2 \cdot \gamma = 1 \text{ from which } b_1 = \frac{S(\alpha) \cdot S(\gamma^2) - S(\gamma) \cdot S(\alpha\gamma)}{S(\alpha^2) \cdot S(\gamma^2) - S^2(\alpha\gamma)}$$

$$b_2 = \frac{S(\gamma) \cdot S(\alpha^2) - S(\alpha) \cdot S(\alpha\gamma)}{S(\alpha^2) \cdot S(\gamma^2) - S^2(\alpha\gamma)} \quad (1)$$

TABLE 5

Statistics of the Log-Pearson Type III for the Relative Moments of Evolution

SS	μ	$\hat{\mu}$	$\tilde{\mu}$	σ	γ_1	γ_2
TT	0.645	7.18×10^{-5}	0.117	2.11	15.6	$6.31 \times 10^{+2}$
TD	0.736	3.37×10^{-4}	0.095	4.01	74.6	$3.98 \times 10^{+4}$
TP	0.778	3.24×10^{-5}	0.104	3.26	26.8	$2.32 \times 10^{+3}$
DT	1.526	1.15×10^{-4}	0.156	8.91	69.6	$2.78 \times 10^{+4}$
DD	1.528	1.09×10^{-4}	0.113	13.57	258.1	$9.22 \times 10^{+5}$
DP	2.543	1.40×10^{-4}	0.162	26.88	443.6	$3.94 \times 10^{+6}$
PT	0.685	2.19×10^{-4}	0.091	3.42	53.3	$1.55 \times 10^{+4}$
PD	0.636	1.58×10^{-4}	0.101	2.50	26.5	$2.44 \times 10^{+3}$
PP	0.759	2.65×10^{-4}	0.110	3.41	39.8	$7.22 \times 10^{+3}$

Mean
$$\mu = \int_0^{e^\gamma} LP3_{PDF}(z)zdz$$

Mode
$$\frac{\partial LP3_{PDF}(\hat{\mu})}{\partial \hat{\mu}} = 0$$

Median
$$LP3_{CDF}(\tilde{\mu}) = 0.5$$

$$\tilde{\mu} = LP3_{InvCDF}(0.5)$$

Standard deviation
$$\sigma = \int_0^{e^\gamma} (z - \mu)^2 LP3_{PDF}(z)dz$$

Skewness
$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} \text{ where}$$

$$\mu_k = \int_0^{e^\gamma} (z - \mu)^k LP3_{PDF}(z)dz,$$

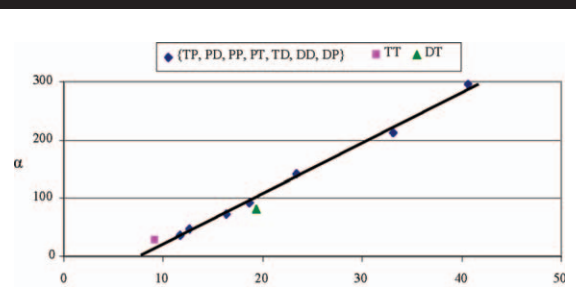
$$k > 1$$

Kurtosis excess
$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3$$

PDF = probability density function; CDF = cumulative distribution function.

where α = shape of the distribution; γ = location of the distribution; b_1 and b_2 = coefficients of the model; S = sum of values.

FIGURE 2



Association between shape (α) and location (γ) in the family of Log-Pearson type III curves.

TABLE 6

Shape, Scale, and Location of the Log-Pearson Type III for the Relative Moments of Evolution under the Assumption of a Linear Association between the Shape and Location and the Results of Agreement

SS (n)	LP3(x; $\alpha_j, \beta_j, 0.011386\alpha_j + 7.7824$)			MLE	KS (p)	AD (p)	CS/df (p)
	Shape (α)	Scale (β)	Location (γ)				
TT (n = 1379)	58.463	-0.28599	14.439	147.2	0.02991 (0.17)	1.14190 (0.26)	14.3/10 (0.16)
TD (n = 1429)	135.98	-0.18892	23.265	323.7	0.01108 (0.99)	0.19340 (0.79)	1.5/10 (1.00)
TP (n = 1318)	41.507	-0.35880	12.508	192.5	0.02707 (0.28)	1.24890 (0.23)	14.2/10 (0.16)
DT (n = 996)	33.879	-0.40104	11.640	-335.4	0.02688 (0.46)	0.52884 (0.53)	7.6/9 (0.58)
DD (n = 1084)	249.43	-0.15403	36.183	-72.80	0.01865 (0.84)	0.34461 (0.66)	7.9/10 (0.64)
DP (n = 851)	277.43	-0.14866	39.370	-387.4	0.02449 (0.68)	0.61738 (0.48)	7.2/9 (0.62)
PT (n = 1463)	83.812	-0.23611	17.325	400.9	0.02255 (0.44)	0.94039 (0.33)	16.3/10 (0.09)
PD (n = 1474)	62.578	-0.27655	14.907	316.9	0.02962 (0.15)	0.98832 (0.31)	11.5/10 (0.32)
PP (n = 1353)	73.132	-0.25152	16.109	140.3	0.01343 (0.96)	0.27687 (0.71)	3.7/10 (0.96)

n = sample size; p = p-value associated with statistics.

The statistical significance of the b_1 and b_2 coefficients was calculated using Fisher's formula [57]:

$$\Delta = \begin{vmatrix} S(\alpha^2) & S(\alpha\gamma) \\ S(\alpha\gamma) & S(\gamma^2) \end{vmatrix}; \quad t(b_1) = \frac{b_1}{\sqrt{S(\varepsilon)}} \frac{\sqrt{\Delta}}{\sqrt{\Delta_{11}}};$$

$$t(b_2) = \frac{b_2}{\sqrt{S(\varepsilon)}} \frac{\sqrt{\Delta}}{\sqrt{\Delta_{22}}}; \quad \Delta_{11} = S(\gamma^2); \quad \Delta_{22} = S(\alpha^2) \quad (2)$$

where $\varepsilon = 1 - b_1 \cdot \alpha - b_2 \cdot \gamma$; t = Student's statistic.

The following regression model was identified for the shape and location of the Log-Pearson type III distribution:

Equation : $-1.436 \times 10^{-2} \alpha + 1.285 \times 10^{-1} \cdot \gamma = 1$

Statistics : $t(-1.436 \times 10^{-2}) = -4.98$; $t(1.285 \times 10^{-1}) = 6.89$

p - values : $p_t(4.98, 5) = 4\%$; $p_t(6.89, 5) = 1\%$

(3)

The location parameter could be extracted from Eq. (3) and was used to align the parameters of the Log-Pearson type III distribution according to the MLE for each range of observations, $SS_j \in \{TP, PD, PP, PT, TD, DD, DP, TT, DT\}$ [Eq. (4)]:

$$\sum_{i=1}^{n_{SS_j}} \log(LP3_{PDF}(x; \alpha_j, \beta_j, 0.11386\alpha_j + 7.7824)) \rightarrow \max. \quad (4)$$

The obtained results and the associated statistics are presented in Table 6.

The results presented in Table 6 revealed that there was no reason to reject the hypothesis of a linear association

between the shape and location (the parameters of the Log-Pearson type III distribution) for the relative moments of evolution in the series of investigated evolutions (DD, DP, DT, PD, PP, PT, TD, TP, and TT). Moreover, the location (γ) and scale (β) proved to be associated with a power function (see Figure 3). Consequently, the relationship presented in Figure 3 was embedded into the Log-Pearson type III theoretical distribution laws, and the distribution became degenerated with one independent statistical parameter.

The equations illustrating the association between shape and location and between scale and location of the Log-Pearson type III distribution parameters for the relative moments of evolutions were obtained based on a nonlinear regression analysis (Figure 3), as well as a linear association (Figure 2). The obtained results are presented in Eq. (5):

$$\alpha = 8.77\gamma - 68.3 \quad \beta = -0.14 - 144\gamma^{-2.57} \quad (5)$$

FIGURE 3

Location (γ)	Scale ($-\beta$)	$\gamma = \text{est}(-\beta)$	Non-linear regression			
14.439	0.28599	0.290678	Power $y=a0+a1*\text{pow}(x,a2)$			
23.265	0.18892	0.184222	r2 Coef Det DF Adj r2 Fit Std Err F-Statistic			
12.508	0.35880	0.357915	0.9990	0.9986	0.003194	2926
11.640	0.40104	0.402157	Value Std Error t-Value 95% Confidence Limits			
36.183	0.15403	0.154214	0.140	0.0032	43.3	0.132 0.148
39.370	0.14866	0.151442	144	31.51	4.57	66.8 221
17.325	0.23611	0.234335	-2.57	0.0907	-28.3	-2.79 -2.35
14.907	0.27655	0.278819				
16.109	0.25152	0.253735				
r (correlation coefficient) = 0.999488			r ² (determination coefficient) = 0.998976			

The association between location and scale in the Log-Pearson type III curves of evolution.

TABLE 7

Shape, Scale, and Location of the Log-Pearson Type III for the Relative Moments of Evolution under the Assumption of a Linear Association between Shape and Location and a Nonlinear Association between Scale and Location, Including the Corresponding Significances

SS (n)	LP3 ($x; 8.77 \cdot \gamma_j - 68.3, -0.14 - 144 \cdot \gamma_j^{-2.57}, \gamma_j \rightarrow \max.$)				Significance					
	Shape (α)	Scale (β)	Location (γ)	MLE	p_{KS}	p_{AD}	p_{CS}	p_{KS}	p_{AD}	p_{CS}
TT (n = 1379)	82.293	-0.23659	17.171	146.3	0.46 ^a	0.09 ^b	0.47 ^a	0.17 ^b	0.27	0.12 ^b
TD (n = 1429)	72.113	-0.25562	16.011	323.9	0.98	0.98	0.75	0.74	0.97	0.77
TP (n = 1318)	43.589	-0.34725	12.758	192.4	0.29	0.30	0.24	0.19	0.16	0.10
DT (n = 996)	33.783	-0.40234	11.640	-335.3	0.39	0.47	0.41	0.52	0.3	0.55
DD (n = 1084)	250.61	-0.15404	36.364	-72.80	0.81	0.88	0.66	0.66	0.62	0.47
DP (n = 851)	222.51	-0.15780	33.160	-390.5	0.69	0.14	0.47	0.15	0.65	0.21
PT (n = 1463)	66.295	-0.26890	15.347	401.3	0.58	0.68	0.43	0.46	0.25	0.36
PD (n = 1474)	73.919	-0.25189	16.216	316.8	0.13	0.08	0.33	0.24	0.56	0.44
PP (n = 1353)	82.373	-0.23645	17.180	140.2	0.99	0.90	0.75	0.64	0.97	0.80

p = probabilities of a random observation for KS, AD, and CS tests.

^aFrom MLE independent parameters.

^bFrom MLE with an independent parameter - γ .

where α = shape parameter; β = scale parameter; and γ = location parameter.

The above-referenced equations were included in the MLE procedure. The parameters associated with the Log-Pearson type III distribution and their corresponding significances are presented in Table 7.

The results presented in Table 7 revealed no justification for rejecting the hypothesis of a linear association between shape and location or of a nonlinear association between scale and location. Because the hypotheses for the above-mentioned associations were accepted, the values presented in Table 5 were subsequently modified, and the new values are presented in Table 8.

The results presented in Table 8 revealed that different selection-survival strategies had similar statistical values. To identify these similarities, a PCA was conducted, and the results associated with the two main components are presented in Figure 4. For the degenerated Log-Pearson type III theoretical distribution (one independent parameter), a MLE approach was applied. The locations were then identified, and a series of population statistics were computed (i.e., mean, mode, median, standard deviation, skewness, and kurtosis excess). A PCA was then conducted on these population statistics, and Figure 4 depicts the analysis results. The first two factors split the evolution strategies into four or five groups. Based on comparisons of these population statistics, some of the more closely related strategies are the PP (Proportional-Proportional) and PT (Proportional-Tournament) strategies, which have statistics that almost overlap, as well as

the TD (Tournament-Deterministic) and PD (Proportional-Deterministic) strategies, which have statistics that are quite similar. Paired values for location, mean, standard deviation, skewness, and kurtosis excess were identified for the DP (Deterministic-Proportional) and DD (Deterministic-Deterministic) strategies. Likewise, paired values for location and mode were observed for the DT (Deterministic-Tournament) and TP (Tournament-Proportional) strategies.

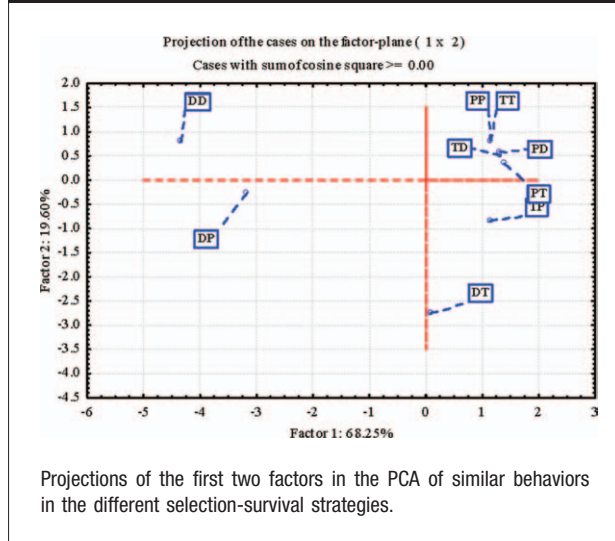
TABLE 8

Statistics Associated with the Log-Pearson Type III for the Relative Moments of Evolution under the Assumption of Two Associations between the Distribution Parameters

SS	γ	μ	$\hat{\mu}$	$\tilde{\mu}$	σ	γ_1	γ_2
TT	17.171	0.737	3.28×10^{-4}	0.109	3.34	42.1	$8.53 \times 10^{+3}$
TD	16.011	0.668	2.23×10^{-4}	0.097	3.00	39.4	$7.01 \times 10^{+3}$
TP	12.758	0.791	5.03×10^{-5}	0.104	3.45	30.6	$3.32 \times 10^{+3}$
DT	11.640	1.242	2.96×10^{-5}	0.162	5.15	25.3	$1.97 \times 10^{+3}$
DD	36.364	1.583	1.13×10^{-4}	0.112	14.98	319.9	$1.66 \times 10^{+6}$
DP	33.160	1.746	2.38×10^{-4}	0.150	13.95	206.9	$5.51 \times 10^{+5}$
PT	15.347	0.643	1.72×10^{-4}	0.092	2.88	38.0	$6.25 \times 10^{+3}$
PD	16.216	0.677	2.40×10^{-4}	0.098	3.04	39.9	$7.27 \times 10^{+3}$
PP	17.180	0.738	3.29×10^{-4}	0.109	3.34	42.1	$8.54 \times 10^{+3}$

γ = location; μ = mean; $\hat{\mu}$ = mode; $\tilde{\mu}$ = median; σ = standard deviation; γ_1 = skewness; and γ_2 = kurtosis excess.

FIGURE 4



Analyses from Figure 4 reveal the following:

- The DD (Deterministic-Deterministic) and DP (Deterministic-Proportional) strategies were projected along one of the principal factors (one accounted for 68.25% of the variance), because the values of all the statistics were high (see Table 8).
- The DT (Deterministic-Tournament) strategy was projected along one of the principal factors that accounted for 19.6% of the variance; the location, mean, and standard deviation had relatively high values compared to the other investigated strategies, whereas the values for skewness and kurtosis excess were low.
- A group of strategies—PP (Proportional-Proportional), TT (Tournament-Tournament), TD (Tournament-Deterministic), PD (Proportional-Deterministic), and PT (Proportional-Tournament)—had similar values with only small differences among them (see Figure 4).
- The TP (Tournament-Proportional) strategy was located with the second principal factor at a relatively equal distance from the compact group (PP = Proportional-Proportional, TT = Tournament-Tournament, TD = Tournament-Deterministic, PD = Proportional-Deterministic, and PT = Proportional-Tournament) and the DT (Deterministic-Tournament) strategy, but the TP strategy was in the same quadrant as the DT strategy.

The observed probabilities were used to measure the differences among the Log-Pearson type III curves of the relative moments of evolution for the different selection-survival strategies using Eq. (6):

$$LP3_i(x) = LP3(x; 8.77 \cdot \gamma_i - 68.3, -0.14 - 144 \cdot \gamma_i^{-2.57}, \gamma_i) \quad (6)$$

where TT = Tournament-Tournament ($i = 1$), TD = Tournament-Deterministic ($i = 2$), TP = Tournament-Proportional ($i = 3$), DT = Deterministic-Tournament ($i = 4$), DD = Deterministic-Deterministic ($i = 5$), DP = Deterministic-Proportional ($i = 6$), PT = Proportional-Tournament ($i = 7$), PD = Proportional-Deterministic ($i = 8$), and PP = Proportional-Proportional ($i = 9$).

The difference in probabilities is given by Eq. (7):

$$\text{diff}_{i,j}(x) = \max(0, LP3_i(x) - LP3_j(x));$$

$$\text{Diff}_{i,j} = \int_0^{\min(e^{\gamma_i}, e^{\gamma_j})} \text{diff}_{i,j}(x) dx \quad (7)$$

where $\text{diff}_{i,j}(x)$ indicates that an evolution was observed at the “ x ” relative moment of evolution for the “ i ” selection-survival strategy, but the evolution was not observed for the “ j ” selection-survival strategy. In other words, $\text{diff}_{i,j}(x)$ is the probability of observing (during an undefined period of time, from 0 to 8) two evolutions: one for the “ i th” selection-survival strategy, and the other for the “ j th” selection-survival strategy.

The calculated values for the $\text{diff}_{i,j}(x)$ function are presented in Table 9.

The results presented in Table 9 illustrate the following:

- An evolution following the Tournament-Tournament strategy was less likely to occur if an evolution process following the Proportional-Proportional strategy did not occur in the same relative moment, and vice versa. The probability of this event was less than 1‰ ($\text{Diff}_{TT,PP} =$

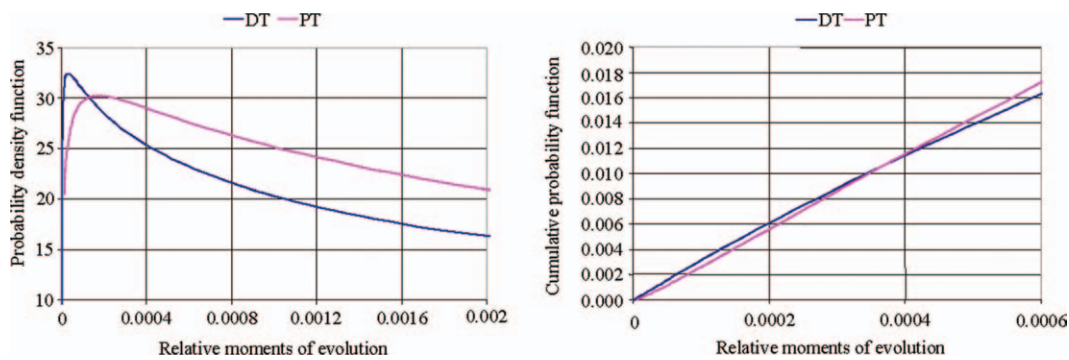
TABLE 9

Differences in the Observed Probabilities of Evolutions Using the Different Selection-Survival Strategies

$\text{Diff}_{i,j}(x)$	TT	TD	TP	DT	DD	DP	PT	PD	PP
TT		0.021	0.030	0.083	0.063	0.070	0.031	0.018	0.000
TD	0.020		0.030	0.099	0.061	0.084	0.010	0.003	0.020
TP	0.021	0.023		0.075	0.034	0.058	0.027	0.022	0.021
DT	0.070	0.087	0.070		0.068	0.024	0.095	0.084	0.069
DD	0.042	0.042	0.021	0.060		0.043	0.042	0.041	0.042
DP	0.052	0.068	0.049	0.020	0.047		0.075	0.065	0.052
PT	0.028	0.008	0.033	0.106	0.061	0.090		0.011	0.028
PD	0.017	0.003	0.030	0.097	0.062	0.082	0.013		0.017
PP	0.000	0.021	0.030	0.083	0.063	0.070	0.031	0.018	

$\text{Diff}_{i,j}(x)$ = function defined in Eq. (7).

FIGURE 5



Probabilities of the relative moments of evolution: Deterministic-Tournament (DT) and Proportional-Tournament (PT) strategies expressed as a Probability Density Function (left-hand graph) and a Cumulative Density Function (right-hand graph).

$\text{Diff}_{\text{PT,DT}} = 0.000$) as shown in Figure 4 (the events almost overlap in the compact group of the selection-survival strategies).

- An evolution process following the Proportional-Tournament strategy was more likely to occur, if an evolution process following the Deterministic-Tournament strategy did not generate evolutions in the same relative moments. The probability of this event was 10.6% ($\text{Diff}_{\text{PT,DT}} = 0.106$, Table 9). As shown in Figure 4, the Proportional-Tournament had the highest value for the first principal component. The Deterministic-Tournament had the lowest value for the first principal component, and the opposite event had a 9.5% probability ($\text{Diff}_{\text{PT,DT}} = 0.095$, Table 9).

Graphical representations of the probability distribution function (left-hand image) and of the cumulative distribution function (right-hand image) of the relative moments of evolution for the Deterministic-Tournament and Proportional-Tournament strategies are presented in Figure 5. The Deterministic-Tournament strategy is characterized by the lowest value for the principal component of the population statistics, whereas the Proportional-Tournament strategy yielded the highest positive value for the first component. The figure reveals a shift in the probability to induce evolution near to the relative moment of evolution of 0.00036.

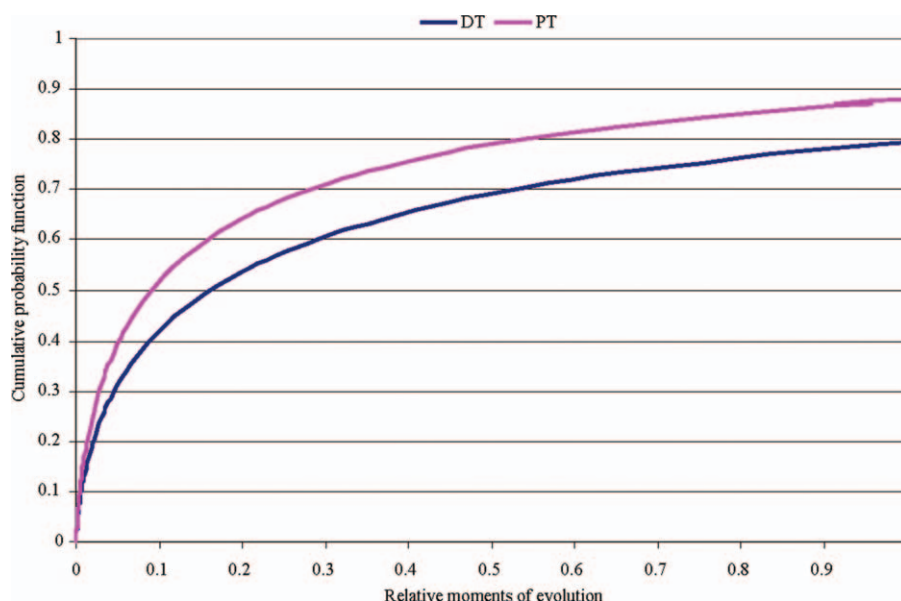
The representation of the cumulative distribution function for the Deterministic-Tournament and Proportional-Tournament strategies on a larger relative time scale (from 0 to 1) is presented in Figure 6. As shown in Figure 6, the Deterministic-Tournament strategy is characterized by the lowest values (in both absolute and positive values) for the principal component of the population statistics, whereas

the Proportional-Tournament strategy is characterized by the highest positive value of the first component. The figure reveals a significant difference between the strategies (approximately a 10% greater probability of the Proportional-Tournament strategy relative to the Deterministic-Tournament strategy) for lower values of the relative moments of evolution (<1).

The difference in the occurrence of the relative moments of evolution between the Deterministic-Tournament and Proportional-Tournament strategies based on the results presented in Table 9 could be explained by analyzing the graphical representations of the distribution laws (Figures 5 and 6). A 10.6% difference ($\text{Diff}_{\text{PT,DT}} = 0.106$, Table 9) in favor of the Proportional-Tournament strategy over the Deterministic-Tournament strategy was observed for the following event: ‘an evolution process by Proportional-Tournament strategy occurred while an evolution process by Deterministic-Tournament strategy failed to occur in the same relative moment of time.’ This difference could be explained by the values for the locations ($\gamma_{\text{PT}} = 15.347$ and $\gamma_{\text{DT}} = 11.640$; Table 8).

The opposite event had a probability of 9.5% ($\text{Diff}_{\text{PT,DT}} = 0.095$, Table 9) that could be explained by the graphical representation of the probability density function (PDF; the Deterministic-Tournament strategy is more productive at lower relative moments). In addition, the Proportional-Tournament and Deterministic-Tournament strategies generated an equal number of evolutions up to the relative moment of evolution equal to 0.00038 (Figure 5, cumulative probability function). From that moment on, the Proportional-Tournament strategy was likely to generate more evolutions than the Deterministic-Tournament strategy, thus producing evolutions more slowly than the Deterministic-Tournament strategy.

FIGURE 6



Cumulative probability function of the relative moments of evolution: the Deterministic-Tournament (DT) and Proportional-Tournament (PT) strategies.

In summary, we achieved our primary research aim by identifying the distribution law of the relative moments of evolution. These moments fit degeneration with one parameter of the Log-Pearson type III distribution for which selection and survival strategies of a GA are influential factors. There are important ramifications of this finding. First, knowing the probability distribution function makes it possible to calculate the probability associated with a random observation. Moreover, given the statistical parameter of a probability distribution function, it can then be determined with high confidence whether an observation is random. The results of this study show that the relative moments of evolution follow a natural process; therefore, their probability distribution function can be defined by estimating one unknown parameter. Another question arises here: “Could the investigated QSAR problem be representative of QSAR-building in general?” Both selection and survival are measures of overall survival and are calculated using functions that influence the entire genetic code. The median quality depended not only on the total number of experimental observations but also on the number of statistical replications that were performed (45 runs were performed for this study); therefore, our “narrow and unique” problem (e.g., investigation of the probability distribution function for just one set of compounds and one activity) should not influence the obtained results. It is thus expected that the moments of evolution in the QSAR analysis will hold, even if we were to change the

descriptors. This study showed that one factor that can change is the value of the unknown parameter of the degenerated Log-Pearson type III probability distribution law that characterized the evolution. Ongoing studies in our laboratory aim to demonstrate whether the results of this study reflect the probability distribution function of the relative moments of evolution in other QSARs.

CONCLUSIONS

The relative moments of evolution followed the Log-Pearson type III law when all pairs of the selection-survival strategies were investigated. The Log-Pearson type III distribution could not be rejected when each selection-survival strategy was analyzed separately. Furthermore, changing the evolution strategy (i.e., selection-survival strategy) has only one degree of freedom, as indicated by the two dependencies given in the three degrees of freedom of the Log-Pearson type III distribution for the relative moments of evolution. These findings suggest that permitting one degree of freedom (e.g., one change in the evolution strategy) under the same environmental constraints will parameterize the changes to the evolution outcome.

Acknowledgments

The study was supported by POSDRU/89/1.5/S/62371 through a postdoctoral fellowship for L. Jäntschi.

REFERENCES

1. Deep, K.; Singh, K.P.; Kansal, M.L.; Mohan, C. An interactive method using genetic algorithm for multi-objective optimization problems modeled in fuzzy environment. *Expert Syst Appl* 2011, 38, 1659–1667.
2. Pasandideh, S.H.R.; Niaki, S.T.A.; Nia, A.R. A genetic algorithm for vendor managed inventory control system of multi-product multi-constraint economic order quantity model. *Expert Syst Appl* 2011, 38, 2708–2716.
3. Wu, Y.T.; Chan, K.Y. Optimal design and impact analysis of urban traffic regulations under ambient uncertainty. *Stoch Environ Res Risk Assess* 2011, 25, 271–286.
4. Chen, S.C.; Lin, S.W.; Chou, S.Y. Enhancing the classification accuracy by scatter-search-based ensemble approach. *Appl Soft Comput* 2011, 11, 1021–1028.
5. Furuta, H.; Frangopol, D.M.; Nakatsu, K. Life-cycle cost of civil infrastructure with emphasis on balancing structural performance and seismic risk of road network. *Struct Infrastruct Eng* 2011, 7, 65–74.
6. Brylinski, M.; Skolnick, J. Comprehensive structural and functional characterization of the human kinome by protein structure modeling and ligand virtual screening. *J Chem Inf Model* 2010, 50, 1839–1854.
7. Ho, W.H.; Chan, A.L.F. Hybrid Taguchi-differential evolution algorithm for parameter estimation of differential equation models with application to HIV dynamics. *Math Probl Eng* 2011, 514756 (article number). Available at: <http://www.hindawi.com/journals/mpe/2011/514756/>. Accessed on December 10, 2011.
8. Reymond, J.L.; van Deursen, R.; Blum, L.C.; Ruddigkeit, L. Chemical space as a source for new drugs. *Med Chem Commun* 2010, 1, 30–38.
9. Marques, J.M.C.; Llanio-Trujillo, J.L.; Abreu, P.E.; Pereira, F.B. How different are two chemical structures? *J Chem Inf Model* 2010, 50, 2129–2140.
10. Xu, J.; Zhang, H.; Wang, L.X.; Xu, W.L.; Yi, C.H.; Liang, H.; Li, Z.C. Quantitative structure–property relationship analysis for optical limiting of organic compounds based on genetic algorithm multivariate linear regression. *Asian J. Chem.* 2011, 23, 92–96.
11. Golmohammadi, H.; Dashtbozorgi, Z. Quantitative structure–property relationship studies of gas-to-wet butyl acetate partition coefficient of some organic compounds using genetic algorithm and artificial neural network. *Struct Chem* 2010, 21, 1241–1252.
12. Mallakpour, S.; Hatami, M.; Golmohammadi, H. Prediction of inherent viscosity for polymers containing natural amino acids from the theoretical derived molecular descriptors. *Polymer* 2010, 51, 3568–3574.
13. Martin, W.N.; Spears, W.M. *Foundations of Genetic Algorithms 6*; San Francisco, CA: Morgan Kaufmann, 2001.
14. Prügel-Bennett, A. The mixing rate of different crossover operators. In: *Foundations of Genetic Algorithms*, Morgan Kaufmann: San Francisco, USA, Vol. 6, 2001; pp 261–274.
15. Spears, W.M. The equilibrium and transient behavior of mutation and recombination. In: *Foundations of Genetic Algorithms*, Morgan Kaufmann; San Francisco, USA, Vol. 6, 2000; pp 241–260.
16. Droste, S.; Jansen, T.; Wegener, I. Dynamic parameter control in simple evolutionary algorithms. In: *Foundations of Genetic Algorithms*, Morgan Kaufmann: San Francisco, USA, Vol. 6, 2001; pp 275–294.
17. Solimanpur, M.; Foroughi, A. A new approach to the cell formation problem with alternative processing routes and operation sequence. *Int J Prod Res* 2011, 49, 5833–5849.
18. Liu, P.; Cai, X.; Guo, S. Deriving multiple near-optimal solutions to deterministic reservoir operation problems. *Water Resour Res* 2011, 47, art. no. 7208.
19. Rahmani, H.; Bonyadi, M.R.; Momeni, A.; Moghaddam, M.E.; Abbaspour, M. Hardware design of a new genetic based disk scheduling method. *Real-Time Syst* 2011, 47, 41–71.
20. Vale, Z.A.; Ramos, C.; Faria, P.; Soares, J.P.; Canizes, B.; Teixeira, J.; Khodr, H.M. Comparison between deterministic and meta-heuristic methods applied to ancillary services dispatch. *Lect Notes Comput Sci* 2010, 6096, 731–741.
21. Hemmateenejad, B.; Javidnia, K.; Miri, R.; Elyasi, M. Quantitative structure–retention relationship study of analgesic drugs by application of combined data splitting-feature selection strategy and genetic algorithm-partial least square. *J Iran Chem Soc* 2012, 9, 53–60.
22. Auger, A.; Brockhoff, D.; Hansen, N. Analyzing the impact of mirrored sampling and sequential selection in elitist evolution strategies. In: *FOGA'11—Proceedings of the 2011 ACM/SIGEVO Foundations of Genetic Algorithms XI*, 22. Association for Computing Machinery: New York, USA, 2011; pp. 127–138.
23. Tang, K.-Z.; Sun, T.-K.; Yang, J.-Y. An improved genetic algorithm based on a novel selection strategy for nonlinear programming problems. *Comput Chem Eng* 2011, 35, 615–621.
24. Wiese, K.C.; Goodwin, S.D. Keep-best reproduction: A local family competition selection strategy and the environment it flourishes in. *Constraints* 2001, 6, 399–422.
25. Chinnasri, W.; Sureerattan N. Comparison of performance between different selection strategies on genetic algorithm with course timetabling problem. In: *ICAMS 2010—Proceedings of 2010 IEEE International Conference on Advanced Management Science*. Institute of Electrical and Electronics Engineers: Chengdu, China, Vol. 2, 2010; pp 105–108.
26. MacFarlane, A.; Secker, A.; May, P.; Timmis, J. An experimental comparison of a genetic algorithm and a hill-climber for term selection. *J. Documentation* 2010, 66, 513–531.
27. Jäntschi, L. MDF—A new QSAR/QSPR molecular descriptors family. *Leonardo J Sci* 2004, 3, 68–85.
28. Jäntschi, L.; Bolboaca, S.D. Molecular descriptors family on structure activity relationships. 6. Octanol–water partition coefficient of polychlorinated biphenyls. *Leonardo El J Pract Technol* 2006, 8, 71–86.
29. Jäntschi, L.; Bolboaca, S.D.; Sestraş, R.E. A study of genetic algorithm evolution on the lipophilicity of polychlorinated biphenyls. *Chem Biodivers* 2010, 7, 1612–1872.

30. Jorgensen, W.L.; Tirado-Rives, J. The OPLS potential functions for proteins. Energy minimization for crystals of cyclic peptides and crambin. *J Am Chem Soc* 1988, 110, 1657–1666.
31. Stewart, J.J.P. Optimization of parameters for semi-empirical methods I. Method. *J Comput Chem* 1989, 10, 209–220.
32. Stewart, J.J.P. Optimization of parameters for semi-empirical methods. II. Applications. *J Comput Chem* 1989, 10, 221–264.
33. Jäntschi, L.; Bolboaca, S.D. A genetic algorithm for structure–property relationships: Software implementation. Available at: <http://arxiv.org/ftp/arxiv/papers/0906/0906.4846.pdf>. Accessed on December 4, 2011.
34. Kolmogorov, A. Confidence limits for an unknown distribution function. *Ann Math Stat* 1941, 12, 461–463.
35. Smirnov, N.V. Tables for estimating the goodness of fit of empirical distributions. *Ann Math Stat* 1948, 19, 279–281.
36. Anderson, T.W.; Darling D.A. Asymptotic theory of certain ‘goodness-of-fit’ criteria based on stochastic processes. *Ann Math Stat* 1952, 23, 193–212.
37. Pearson, K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos Mag* 1900, 50, 157–175.
38. Fisher, R.A. On an absolute criterion for fitting frequency curves. *Messenger Math* 1912, 41, 155–160.
39. Jäntschi, L. Distribution fitting. 1. Parameters estimation under assumption of agreement between observation and model. *Bull UASVM Hort* 2009, 66, 684–690.
40. Pearson, K. Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. *Philos Trans R Soc Lond Ser A* 1896, 187, 253–318.
41. Bolboaca, S.D.; Jäntschi, L. Pearson versus Spearman, Kendall’s Tau correlation analysis on structure–activity relationships of biologic active compounds. *Leonardo J Sci* 2006, 9, 179–200.
42. Phinea, H.N.; Hira, M.A. Log Pearson type-3 distribution: Parameter estimation. *J Hydrol* 1983, 64, 25–37.
43. Foster, H.A. Theoretical frequency curves. *Trans ASCE* 1924, 87, 142–203.
44. Nemichandruppa, M.; Balakrishnan, P.; Senthilvel, S. Probability and confidence limit analysis of rainfall in Raichur region. *Karnataka J Agric Sci* 2010, 23, 737–741.
45. Cotaescu, I.I.; Paul, G.; Paulescu, M. Applying the Dirac equation to derive the transfer matrix for piecewise constant potentials. *Phys Lett A* 2007, 366, 363–366.
46. Gubareva, T.S.; Gartsman, B.I. Estimating distribution parameters of extreme hydrometeorological characteristics by L-moments method. *Water Resour* 2010, 37, 437–445.
47. Spencer, C.S.; McCuen, R.H. Detection of outliers in Pearson type III data. *J Hydrol Eng* 1996, 1, 2–10.
48. Cotaescu, I.I.; Gravila, P.; Paulescu, M. Pseudo-Gaussian oscillators. *Int J Mod Phys C* 2008, 19, 1607–1615.
49. Naghavi, B.; Yu, F.X. Selection of parameter-estimation method for LP3 distribution. *J Irrig Drain Eng* 1996, 122, 24–30.
50. Paulescu, M.; Dughir, C.; Tulcan-Paulescu, E.; Lascu, M.; Gravila, P.; Jurca, T. Solar radiation modeling and measurements in Timisoara, Romania: Data and model quality. *Environ Eng Manage J* 2010, 9, 1089–1095.
51. Malekinezhad, H.; Nachtnebel, H.P.; Klik, A. Regionalization approach for extreme flood analysis using L-moments. *J Agric Sci Technol (Iran)* 2011, 13, 1183–1196.
52. Chen, L.-H.; Hong, Y.-T.; Hsu, C.-W. A study on regional drought frequency analysis using self-organizing map and L-moments. *J Taiwan Agric Eng* 2011, 57, 57–77.
53. Jou, P.H.; Akhoond-Ali, A.M.; Behnia, A.; Chinipardaz, R. Parametric and nonparametric frequency analysis of monthly precipitation in Iran. *J Appl Sci* 2008, 8, 3242–3248.
54. Josse, J.; Pagès, J.; Husson, F. Testing the significance of the RV coefficient. *Comput Stat Data Anal* 2008, 53, 82–91.
55. Grover, N.B.; Eward, K.L.; Helmstetter, C.E. Synchronous cultures from the baby machine. A model for animal cells. *Math Biosci* 2004, 190, 87–96.
56. Jäntschi, L.; Bolboaca, S.D. The jungle of linear regression revisited. *Leonardo El J Pract Technol* 2007, 10, 169–187.
57. Fisher, R.A. The goodness of fit of regression formulae and the distribution of regression coefficients. *J R Stat Soc* 1922, 85, 597–612.