

## Article

# The Effect of Multi-Generational Selection in Geometric Semantic Genetic Programming

Mauro Castelli <sup>1,\*</sup>, Luca Manzoni <sup>2</sup>, Luca Mariot <sup>3</sup>, Giuliamaria Menara <sup>2</sup> and Gloria Pietropolli <sup>2</sup>

- <sup>1</sup> NOVA Information Management School (NOVA IMS), Universidade NOVA de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal
- <sup>2</sup> Dipartimento di Matematica e Geoscienze, Università degli Studi di Trieste, H2bis Building, Via Alfonso Valerio 12/1, 34127 Trieste, Italy; lmanzoni@units.it (L.M.); giuliamaria.menara@phd.units.it (G.M.); gloria.pietropolli@phd.units.it (G.P.)
- <sup>3</sup> Digital Security Group, Radboud University, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands; luca.mariot@ru.nl
- \* Correspondence: mcastelli@novaims.unl.pt

**Abstract:** Among the evolutionary methods, one that is quite prominent is genetic programming. In recent years, a variant called geometric semantic genetic programming (GSGP) was successfully applied to many real-world problems. Due to a peculiarity in its implementation, GSGP needs to store all its evolutionary history, i.e., all populations from the first one. We exploit this stored information to define a multi-generational selection scheme that is able to use individuals from older populations. We show that a limited ability to use “old” generations is actually useful for the search process, thus showing a zero-cost way of improving the performances of GSGP.

**Keywords:** evolutionary computation; genetic programming; geometric operators; geometric semantic genetic programming



**Citation:** Castelli, M.; Manzoni, L.; Mariot, L.; Menara, G.; Pietropolli, G. The Effect of Multi-Generational Selection in Geometric Semantic Genetic Programming. *Appl. Sci.* **2022**, *12*, 4836. <https://doi.org/10.3390/app12104836>

Academic Editor: Valentino Santucci

Received: 22 April 2022

Accepted: 9 May 2022

Published: 10 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Genetic programming (GP) [1] is a method for evolving programs, usually represented as trees, through operations that mimic the Darwinian process of natural selection. Among the most successful methods of GP, there is the *geometric semantic GP* (GSGP) [2]. In its inception, GSGP was only an object of theoretical interest because of the geometric properties it induced on the fitness landscape. Subsequently, an efficient implementation of the geometric semantic genetic operators [3,4] allowed for the application of GSGP in many different fields (e.g., [5]).

To implement and execute GSGP in an efficient manner, it is necessary to store *all* the information regarding the entire evolutionary process. In particular, all the populations need to be efficiently stored. Thus, it is natural to ask *how* this information can be used to improve the search process.

In this paper, we propose a method to use this additional “free” information by allowing the selection process to select individuals from the “old” populations. In particular, we present two ways for performing this multi-generational selection:

- By selecting uniformly among the last  $k$  generations;
- By selecting among all the generations with a decreasing probability (i.e., with a geometric distribution).

We compare the performance of two methods over six real-world datasets, showing that the ability to look back at the evolutionary history of GSGP is actually beneficial in the evolutionary process.

The paper is structured as follows: Section 2 explores the existing works concerning the use of old generations (or *memory*) to improve the performance of evolutionary computation algorithms. Section 3.2 recalls the basic notions of GSGP and introduces multi-generational

selection. In Section 4, the experimental settings are presented, while Section 5 discusses and analyzes the results. Section 6 provides a summary of this work and directions for future research.

## 2. Related Works

When considering an evolutionary algorithm (EA) as a dynamic system, the idea of selecting individuals from older populations may be regarded as the ability of the system to directly exploit its *memory*, or equivalently the sequence of its past states. Under this perspective, approaches that enhanced genetic algorithms (GA) with memory started to appear in the literature in the mid-90s.

As far as we know, Louis and Li [6] were the first to propose the use of a *long-term memory* to store the best solutions found so far by a GA, eventually reintegrating them into the population in a later stage. Their experimental investigation of the traveling salesman problem (TSP) showed that GA obtained a better performance when its population was seeded with sub-optimal solutions found in previous instances rather than initializing it at random. Wiering [7] experimented with a combination of GA and local search, where memory plays an analogous role in Tabu Search: if a local optimum has been found before, then it gets the lowest possible fitness to maximize the probability that the corresponding individual is replaced in the next generation. Later, Yang [8] compared two variants of the random immigrants scheme based on memory and elitism to enhance the performance of GA over dynamic fitness landscapes. In these hybrid schemes, the best individual retained either by memory or elitism from old populations was used as a basis to evolve new immigrants through mutation, thus increasing the diversity of the population and its adaptability against a dynamic environment. Still, concerning dynamic optimization problems, Cao and Luo [9] considered two retrieving strategies which selected the two best individuals from the associative memory of a GA. In particular, the environmental information associated with these two individuals was evaluated by either a survivability or diversity criterion. Similar to the methods proposed in this paper, Castelli et al. [10,11] proposed the reinsertion of old genetic material in GAs. In particular, the authors proposed a method to boost the GA optimization ability by replacing a fraction of the worst individuals with the best ones from an older population.

From the point of view of GP, the reinsertion of genetic material from old populations usually occurs in the related literature under the name of *concept and knowledge reuse*. This is indeed a proper term since GP evolves programs that can be used, in turn, for learning concepts and functions, as in symbolic regression [12]. Séront [13] set forth a method to retrieve concepts evolved by GP based on a library that saved the trees of the best individuals. This method stood on the reasonable assumption that highly fit individuals embed useful concepts in their syntactic trees for solving a particular optimization problem. The results showed that the use of a concept library to create the initial population is beneficial for GP, as compared to the random initialization. Jaskowski et al. [14] explored a different direction where a method for reusing knowledge embedded by GP was used among a group of learners that worked in parallel on a visual learning task. Therefore, in this case, the reuse of GP subprograms does not come from old populations but is rather shared among different current populations at the same time. Pei et al. [15] investigated the issue of class imbalance in GP-based classifiers and proposed a method to mitigate it by using previously evolved GP trees to initialize the population in later runs. The experimental results indicated that such a mechanism allows the training time to be reduced and increases the accuracy of multi-classifier systems based on GP. More recently, Bi et al. [16] proposed a new method to improve GP learning performance over image classification problems. Such a method is based on knowledge transfer among multiple populations, similarly to the aforementioned approach of [14].

One of the main issues of the methods proposed in the above papers is that (a part of) the evolutionary history is needed to properly exploit older populations, thus increasing the space necessary for those methods to work. It is also interesting to notice that an

increase in the population's size (under a certain limit) is also useful for GSGP [17]. Thus, as proposed in this paper, it is fundamental to explore the trade-off between population size and performance and whether this trade-off can be removed or mitigated by using part of the existing evolutionary history.

The principle of exploiting memory in evolutionary algorithms is also considered under a different guise in the area of machine learning. This is the case, for instance, of the *conservation machine learning* approach proposed by Sipper and Moore [18,19]. There, the authors explore the idea of reusing ML models learned in different ways (e.g., multiple runs, ensemble methods, etc.) and apply it to the case of random forests. The results showed that their method improves the performance over some classification tasks through ensemble cultivation.

Finally, to the best of our knowledge, in the specific area of GSGP, there is no work directly addressing the reuse of old genetic material or individuals from past populations. Nevertheless, there have been several attempts aimed at improving the performance of the basic GSGP algorithm over regression problems. One of the most relevant approaches considers the integration of *local search* in the evolutionary process of GSGP [20,21].

### 3. Geometric Semantic GP with Multi-Generational Selection

In this section, we recall the construction of GSGP with semantic operators and its current efficient implementation through dynamic programming. We then introduce a procedure to use this implementation to obtain a zero-cost way of sampling, during the selection process, not only from the current population but also for *any* previous population obtained during the evolution process.

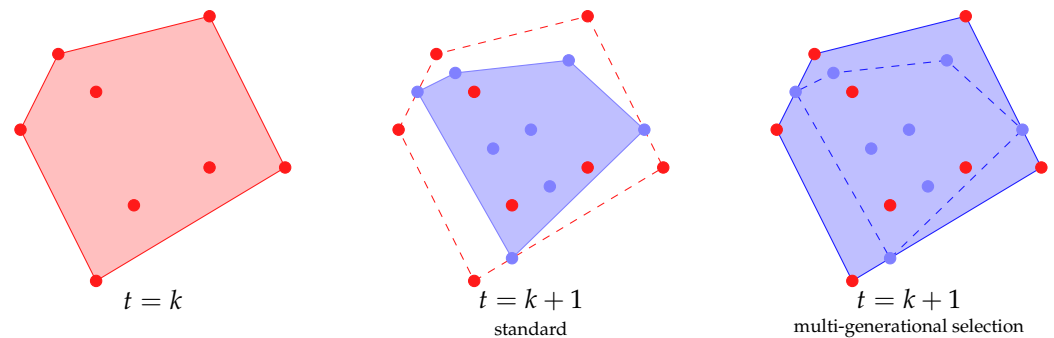
#### 3.1. Geometric Semantic Genetic Programming

In classical GP, the crossover (recombination) and mutation operators act on the genotype of the individual involved, usually modifying or changing parts of the subtrees in the case of a tree-based representation. However, the effects on the phenotype of the individuals are difficult to formalize, and even small modifications to the genotype can create a significant change in the phenotype. Therefore, in addition to the standard "syntactic" operators of GP, there have been multiple studies on *semantic* operators [22]. By relying on these operators, it is possible to predict the effect of crossover and mutation on the phenotype of the solutions.

In 2012, Moraglio and coauthors [2] defined a way of performing semantic crossover and mutation where the effect was obtaining a *geometric* crossover and *geometric* mutation in the semantic space, two concepts defined and studied in [23]. Given a metric space  $(X, d)$  where  $X$  is a set and  $d: X \times X \rightarrow \mathbb{R}_+$  is a distance, a crossover is said to be geometric if, for each  $x, y \in X$ , each element resulting from the crossover of  $x$  and  $y$  belongs to the set  $\{z \in A \mid d(x, z) + d(z, y) = d(x, y)\}$ , i.e., the result of the crossover is inside the segment connecting  $x$  and  $y$ . A mutation operator is said to be geometrical and, in particular, a geometric  $\varepsilon$ -mutation, if, for each  $x \in X$ , the individuals resulting from the mutation are all inside the ball  $B_\varepsilon(x)$ , i.e., the ball of radius  $\varepsilon$  centered in  $x$ . One of the advantages of working with geometric operators is that it is possible to reason on the geometrical effect of the different operators; e.g., a geometric crossover always generates offspring in the convex hull given by the current population (see Figure 1).

In particular, one interesting space where it is possible to define genetic operators is the *semantic space*. That is, given a set of samples  $X = \{\bar{x}^1, \dots, \bar{x}^m\} \subset \mathbb{R}^n$  and a function  $T: \mathbb{R}^n \rightarrow \mathbb{R}$ , representing a GP individual, we can define the *semantics* of  $T$ , denoted by  $s(T)$  as the vector  $(T(\bar{x}^1), T(\bar{x}^2), \dots, T(\bar{x}^m)) \in \mathbb{R}^m$ . If we perform a *geometric* crossover between the semantics of two trees  $T_1$  and  $T_2$ , the semantics of *any* individual resulting from the crossover will have outputs (on the elements of the set  $X$ ) intermediate between the ones of  $T_1$  and  $T_2$ . Similarly, a *geometric* mutation on the semantic space will generate a perturbation of the outputs of the parent individual. There is, however, the question of how to define operators working on the semantic space by manipulating the syntactic space.

The conundrum was solved by Moraglio et al. [2] in 2012 with the definition of semantic mutation and crossover, as detailed below.



**Figure 1.** A representation of the effect of multi-generational selection on the convex hull where geometric semantic crossover can generate new individuals.

**Semantic Crossover.** Let  $T_1$  and  $T_2$  be two functions from  $\mathbb{R}^n$  to  $\mathbb{R}$  representing two GP trees and let  $R: \mathbb{R}^n \rightarrow [0, 1]$  be a randomly generated tree. Then the *semantic crossover* between  $T_1$  and  $T_2$  using the random tree  $R$  is defined as:

$$\text{Cross}(T_1, T_2, R)(\vec{x}) = R(\vec{x})T_1(\vec{x}) + (1 - R(\vec{x}))T_2(\vec{x}) \quad \text{for all } \vec{x} \in \mathbb{R}^n.$$

Notice that the outputs of  $\text{Cross}(T_1, T_2, R)(\vec{x})$  will be intermediate with respect to the outputs of  $T_1$  and  $T_2$ . Thus, the crossover is geometric in the semantic space. To constrain the output of  $R$  inside the interval  $[0, 1]$ , a simple technique is to generate a random tree and pass its output to a sigmoid function.

**Semantic Mutation.** Let  $T: \mathbb{R}^n \rightarrow \mathbb{R}$  be the function defined by a GP tree,  $R: \mathbb{R}^n \rightarrow \mathbb{R}$  be a randomly generated tree, and  $m \in \mathbb{R}_+$  be a positive real number, called the *mutation step*. Then, the *semantic mutation* of  $T$  using the random tree  $R$  is defined as:

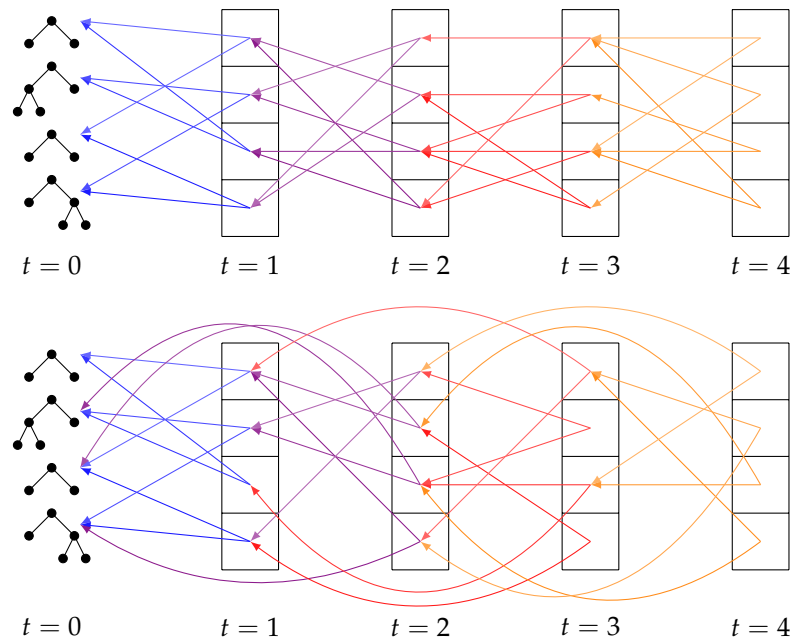
$$\text{Mut}(T, R)(\vec{x}) = T(\vec{x}) + mR(\vec{x}) \quad \text{for all } \vec{x} \in \mathbb{R}^n.$$

Notice that there is an additional parameter, the mutation step  $m$ , that allows how “big” the jumps/perturbations produced by mutation to be tuned.

One disadvantage of geometric operators and crossover in particular, at least in their naïve implementation, is that they produce trees that are exponential with respect to the number of generations: each crossover can be more than twice the size of the trees, thus generating an exponential blowup of the trees. This problem can be solved by using a dynamic programming approach [3]:

- The initial population is composed of standard GP trees;
- Each successive generation is not composed of trees; rather, each individual is a structure containing the random trees used in crossover and mutations and *pointers* or *references* to the individuals in the previous populations. This solves the problem of an exponential space blowup;
- Evaluation can be performed bottom-up, saving the intermediate results from the initial population and combining them following the application of crossover and mutations.

The resulting structure is represented schematically in Figure 2. The resulting complexity for computing the output of an individual is  $O(gp)$ , where  $g$  is the number of generations and  $p$  is the population size. Thus, the resulting complexity is polynomial rather than exponential.



**Figure 2.** A visual representation of how GSGP can be implemented in an efficient way, sharing subtrees between individuals. At the top, the standard implementation where the parents can be selected only from the previous generation is shown. At the bottom, parents can be selected uniformly at random from the previous two populations. Notice how no additional storage is required.

### 3.2. Multi-Generational Selection for GSGP

One interesting aspect to notice in the “fast” GSGP implementation is that we are able to access all the intermediate populations at zero cost: they are already stored in the data structures that we use, and their output is already computed to obtain the outputs of the current population. Hence, it is natural to ask if we can use this additional “free” information to improve the performances in GSGP. In particular, we propose allowing “older” individuals to be selected in the tournament selection.

Concerning the implementation details, as one can observe in Figure 2 at the bottom, the pointers can go back for multiple generations (two in the example figure) instead of a single one, with no additional overhead. The pseudocode for the multi-generational selection is presented in Algorithm 1. As it is possible to notice, since all the information is already available, there is no significant additional computational cost compared to standard GSGP.

---

**Algorithm 1** The pseudocode of the multi-generational (tournament) selection algorithm, where  $P$  is a two-dimensional array of individuals of  $n$  rows (generations), where  $P[i][j]$  is the  $j$ -th individual in the  $i$ -th generation,  $f$  is the fitness function,  $t \in \mathbb{N}$  is the tournament size, and  $D$  is a distribution.

---

```

function MULTI-GENERATIONAL-SELECTION( $P, n, f, t, D$ )
    Tournament  $\leftarrow \emptyset$  ▷ Individuals selected for the tournament
    for  $1 \leq i \leq t$  do ▷ Repeat for the tournament size  $t$ 
         $j \leftarrow n - \text{extract from } D$  ▷ Select the generation
         $k \leftarrow \text{uniform random integer between 1 and } |P[j]|$  ▷ Select the individual
        Tournament  $\leftarrow$  Tournament  $\cup \{P[j][k]\}$  ▷ Add the individual to the tournament
    end for
    best  $\leftarrow \arg \max_{x \in \text{Tournament}} f(x)$  ▷ Find the best individual in the tournament
    return best
end function

```

---

In what follows, we describe how this selection can be performed since multiple possibilities are available. In particular, we propose a uniform and a geometric selection strategy.

### 3.2.1. Uniform Multi-Generational Selection

The simplest case is when individuals can be selected *uniformly at random* from the populations of the last  $k \in \mathbb{N}$  generations (or all of them if less than  $k$  generations have been performed). If  $k = 1$ , then we obtain standard GSGP. However, as shown in Figure 1, one of the effects of higher values of  $k$  is to “delay” the shrinking of the convex hull given by crossover. In this way, we expect to equip GSGP with a better exploration ability.

### 3.2.2. Geometric Multi-Generational Selection

In the uniform selection method, there is a hard cut-off for participating in the selection process. Another idea is to gradually decrease the probability of an old population contributing to the selection process. In particular, it is possible to employ a *geometric distribution*, so that the probability of selecting the  $k$ -th generation before the current one is  $p(1-p)^k$  with  $p \in (0,1)$  a parameter. Thus, for example, if  $k = 0.5$ , the probability of selecting from the previous population is 0.5, two generations behind  $(0.5)^2 = 0.25$ , and so on (if we go back more than the number of existing generations we select from the initial population). Thus, we can regard the geometric method as a way of “fading out” the convex hull given by the previous populations.

Here we have defined two different multi-generation selection methods: *uniform* with a parameter  $k \in \mathbb{N}$ , which we will denote by  $U_k$  (e.g.,  $U_1$ ,  $U_2$ , etc.), and *geometric* with a parameter  $p \in (0,1)$ , which we will denote by  $G_p$  (e.g.,  $G_{0.5}$ ,  $G_{0.75}$ , etc.).

## 4. Experimental Setting

The following section introduces the experimental environment adopted to test the validity of the proposed methods: Section 4.1 describes the benchmark dataset employed, while Section 4.2 provides the experimental settings needed to render experiments fully reproducible.

### 4.1. Dataset

The datasets exploited in this paper have been purposely chosen as they consist of real-world, complex datasets ranging from different areas that have been extensively leveraged as benchmarks for GP. The reader can find [24] a comprehensive description of the reasons why these datasets represent suitable reference points to assert the validity of a genetic programming method. Table 1 outlines the key features of the considered problems, such as the number of instances and variables, the area to which they belong, and, finally, the kind of task required. It is worth pointing out that these datasets are significantly different from each other (considering, for example, the number of instances and variables). Thus, they represent an optimal choice to test the validity of our methods when applied to problems that have rather diverse characteristics.

**Table 1.** Principal characteristics of the considered datasets: the number of variables, the number of instances, and the domain.

Dataset	Variables	Instances	Area
airfoil	6	1503	Physics
bioav	242	359	Pharmacokinetic
concrete	9	1030	Physics
ppb	627	131	Pharmacokinetic
toxicity	627	234	Pharmacokinetic
yacht	7	308	Physics

The first group of datasets deals with predicting pharmacokinetic parameters of potential new drugs.

*Human oral bioavailability (%F)* measures the percentage of the initial drug dose that effectively reaches the systemic blood circulation: this problem constitutes an essential pharmacokinetic task as the oral assumption is usually the preferred way of supplying



drugs to patients, and also because it is a representative measure of the quantity of the active principle that can effectively actuate its biological effect [25].

*Protein-plasma binding level (%PPB)* characterizes the distribution into the human body of a drug. Specifically, it corresponds to the percentage of the initial drug dose which binds plasma proteins: this measure is fundamental, as blood circulation is the major vehicle of drug distribution into the human body [26].

*Median oral lethal dose (LD50)* concerns the harmful effect produced by the distribution of a drug into the human body, as it measures the lethal dose required to kill half the members of a tested population after a specified time. It is expressed as the number of milligrams of drug-related to one kilogram of cavity mass [26].

The second group of datasets originates from physical problems.

*Airfoil self-noise (air)* measures the hydrodynamic performance of sailing yachts, taking into account their dimension and velocity [27].

*Concrete compressive strength (conc)* [28] characterizes the value of the slump flow of the concrete when given as inputs concrete components such as cement, fly ash, slag, water, coarse aggregate and fine aggregate.

Finally, *Yacht hydrodynamics (yac)* measures the hydrodynamic performance of sailing yachts starting from their dimension and velocity.

#### 4.2. Experimental Study

This section describes the experimental settings summarized in Table 2 to make the method fully reproducible. To get statistically valuable results, 100 runs of the method have been performed for each benchmark. Each run consisted of 100 generations to allow the algorithm to stabilize, and, in each run, the dataset has been randomly split into training and test sets with a percentage of 70–30%. All the parameters of GSGP described in Table 2 are the standard values used in the literature (e.g., [29]).

**Table 2.** Experimental settings.

Parameter	Value
Population size	100
Number of generations	100
Number of runs	100
Max. initial depth	4
Crossover rate	0.9
Mutation rate	0.3
Mutation step	0.1
Selection method	Tournament of size 4
Elitism	Best individuals survive

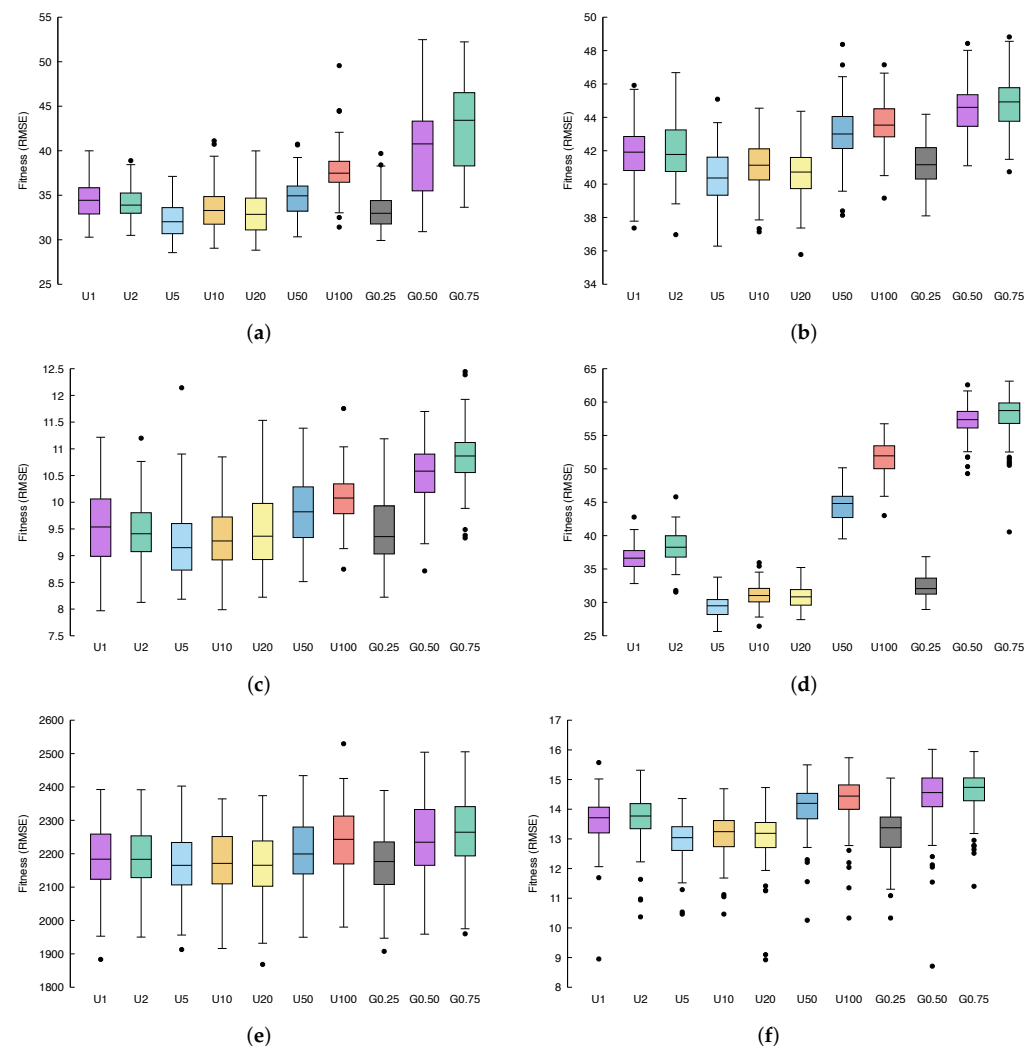
To assess the validity of both uniform and geometric distributions for performing the multi-generational selection, the fitness values achieved within these techniques were compared with classical GSGP. It is worth emphasizing that the results obtained by standard GP [1] are not listed in this paper since standard GP is consistently outperformed by GSGP.

Regarding the uniform distribution variation of the proposed method, experiments have been performed considering the following values as the number of previous generations from where parents have been selected: 2, 5, 10, 20, and 100. The geometric variation of the proposed technique has been tested by selecting 0.25, 0.50, and 0.75 as values for the parameter  $p$  that defines the geometric probability distributions. The selection of those parameters allows covering most cases (from “looking back” only a few generations to the entire evolution), thus giving a general idea of the behavior of the proposed method.

The population size for all the considered cases of study is set to 100, which is a usual trade-off between computational costs and quality of the search process [17], and the trees of the first generation are initialized with the ramped half and half technique. The fitness function selected to measure and compare the quality of different methods proposed is the root-mean-square error (RMSE).

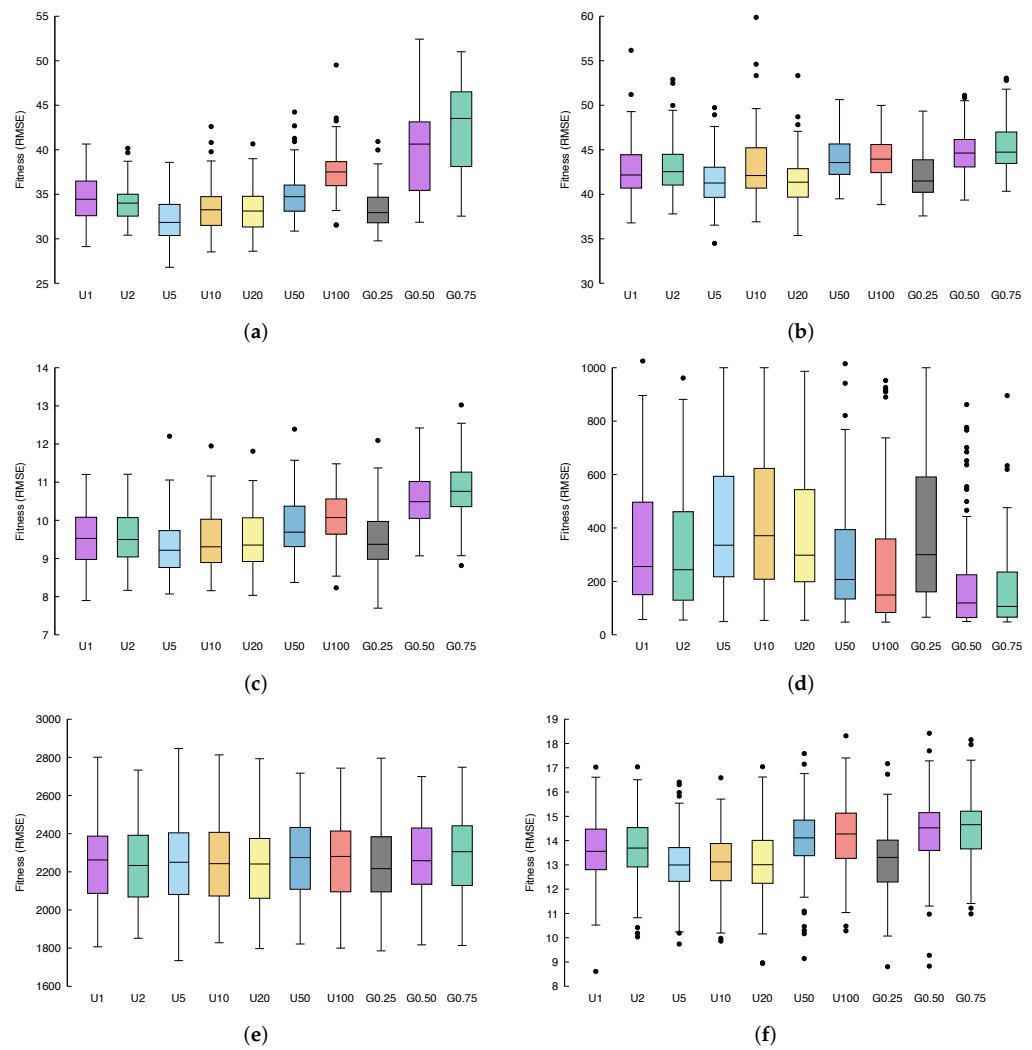
### 5. Results and Discussion

Figures 3 and 4 show, via box-plots, the distribution of the fitness, calculated over 100 independent runs, achieved by the different configurations, discussed in Section 4.2, compared to classical GSGP, which is denoted here as U1 whereas, as explained before, it considers only the previous generation from which to select parents, exactly as in GSGP. Table 3 stores the fitness values obtained by selecting the ancestors with uniform multi-generational selection for both the training and testing set among all the considered benchmark problems. Similarly, Table 4 stores the geometric multi-generational selection method’s fitness values. Table 5 reports a statistical significance assessment of the achieved result, displaying the *p*-values obtained under the hypothesis that the median fitness resulting from the considered technique is equal to the one obtained with standard GSGP. This means that, if the resulting *p*-values are zero, the fitness obtained is statistically significantly better or significantly worse with respect to the one achieved by GSGP, and to understand which case we are dealing with, the results contained in Tables 3 and 4 must be compared. Finally, Tables 3 and 4 display the fitness values obtained by selecting parents using a uniform distribution and geometric distribution, respectively.



**Figure 3.** Box-plots of the RMSE on the training set over 100 independent runs of the considered benchmark dataset for all the proposed methods. (a) Air; (b) %F; (c) conc; (d) %PPB; (e) LD50; (f) yac.





**Figure 4.** Box-plots of the RMSE on the test set over 100 independent runs of the considered benchmark dataset for all the proposed methods. (a) Air; (b) %F; (c) conc; (d) %PPB; (e) LD50; (f) yac.

**Table 3.** Fitness values obtained by selecting the ancestors with uniform multi-generational selection. The values in bold are the best results obtained.

		GSGP	U2	U5	U10	U20	U50	U100
air	train	34.43	33.89	<b>32.02</b>	33.28	32.85	34.93	34.93
	test	34.44	34.01	<b>31.83</b>	33.26	33.12	34.72	37.51
%F	train	41.92	41.78	<b>40.37</b>	41.13	40.72	43.00	43.53
	test	42.17	42.54	<b>41.27</b>	42.10	41.36	43.56	43.94
conc	train	9.54	9.41	<b>9.15</b>	9.27	9.36	9.82	10.08
	test	9.52	9.49	<b>9.21</b>	9.31	9.35	9.69	10.07
%PPB	train	36.62	38.26	<b>29.48</b>	31.02	30.82	44.80	51.94
	test	255.51	243.46	335.42	371.03	298.03	206.88	<b>148.83</b>
LD50	train	2183.65	2183.17	<b>2165.20</b>	2171.09	2165.36	2199.38	2243.06
	test	2262.15	<b>2233.41</b>	2250.19	2242.84	2240.93	2274.87	2280.51
yac	train	13.71	13.77	<b>13.04</b>	13.24	13.18	14.19	14.44
	test	13.55	13.69	<b>12.99</b>	13.12	13.00	14.11	14.27

**Table 4.** Fitness values obtained by selecting the ancestors with geometric multi-generational selection. The values in bold are the best results obtained.

		GSGP	G0.25	G0.50	G0.75
air	train	34.43	37.48	<b>32.97</b>	40.76
	test	34.44	<b>32.93</b>	40.63	43.51
%F	train	41.92	<b>41.16</b>	44.60	44.92
	test	42.17	<b>41.49</b>	44.63	44.73
conc	train	9.54	<b>9.35</b>	10.58	10.86
	test	9.52	<b>9.37</b>	10.48	10.76
%PPB	train	36.62	<b>32.06</b>	57.37	58.73
	test	255.51	300.03	119.20	<b>106.33</b>
LD50	train	2183.65	<b>2176.51</b>	2234.56	2264.43
	test	2262.15	<b>2216.47</b>	2258.10	2305.45
yac	train	13.71	<b>13.37</b>	14.55	14.73
	test	13.55	<b>13.30</b>	14.52	14.65

**Table 5.** *p*-values returned by the Wilcoxon rank-sum test under the alternative hypothesis that the median errors on the test set obtained from classical GSGP are equal with respect to the errors obtained with the methods introduced in this paper. Highlighted in bold, the *p*-values below 0.05 where the direction of the difference shows an improvement with respect to standard GSGP.

	U2	U5	U10	U20	U50	U100	G0.25	G0.50	G0.75
airfoil	0.158	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	0.280	0.000	<b>0.000</b>	0.000	0.000
bioav	0.741	<b>0.000</b>	<b>0.001</b>	<b>0.000</b>	0.000	0.000	<b>0.007</b>	0.000	0.000
concrete	0.763	<b>0.001</b>	<b>0.042</b>	0.445	0.001	0.000	0.557	0.000	0.000
ppb	0.000	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	0.000	0.000	<b>0.000</b>	0.000	0.000
toxicity	0.783	<b>0.049</b>	0.365	0.128	0.281	0.001	0.275	0.001	0.000
yacht	0.135	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	0.000	0.000	<b>0.001</b>	0.000	0.000

Concerning the *air* dataset (Figures 3a and 4a), it is possible to see that, if we consider a uniform distribution for the multi-generational selection, the best performance is obtained by selecting ancestors from the 5 previous generations (Table 3); on the other hand, with a geometric distribution, better results are achieved with  $p = 0.25$  (Table 4). Both of these methods meaningfully outperform standard GSGP, as confirmed by the statistical results of Table 5. For what concerns the other variations of the proposed methods, combining information provided by fitness distribution shown in the box plots with the corresponding *p*-values, it is possible to conclude that: U2, U10 and U20 also outperform GSGP with statistical significance, U50 leads to results similar to GSGP, while U100, G0.50 and G0.75 results are substantially less performing with respect to GSGP.

Considering the %F dataset (Figures 3b and 4b), a similar behavior as the one observed for the previous benchmark problem appears. Again, U5 and G0.25 represent the two best candidates and outperform GSGP together with U2, U10, and U20. Here, U50, U100, G0.50, and G0.75 lead to worse fitness values compared to classical GSGP.

Moving to the *conc* dataset, (Figures 3c and 4c provide us with more evidence that U5 and G0.25 are a significant improvement of the standard GSGP. Moreover, it is still the case that U2, U10 and U20 slightly outperform GSGP, while U50, U100, G0.50 and G0.75 bring poor fitness values.

Taking into account the %PPB dataset (Figures 3d and 4d), it is clear that all models are affected by overfitting, and a lower error in the training set entails a bigger error in the test set.

Applying our methods to the LD50 dataset (Figures 3e and 4e), we can once more recognize that U5 and G0.25 represent the best improvement in GSGP. Further, it is important to highlight that for this dataset, while U2, U10, U20, U50, U100, G0.50, and G0.75 are indeed less well performing than the standard GSGP, none of these methods result in actually poor fitness values; instead, each of them reaches performance very similar to that obtained by GSGP.

Finally, the *yac* dataset (Figures 3f and 4f) is the last confirmation of the behavior we have been observing so far. Once more, U5 and G0.25 are a significant improvement of the standard GSGP, U2, U10 and U20 outperform GSGP, whereas U50, U100, G0.50 and G0.75 result in poor fitness values.

All in all, experiments performed revealed that U5 and G0.25 provide us with a meaningful improvement of the standard GSGP.

Regarding results obtained with uniform multi-generational selection, a number  $k$  of previous generations equal to 10 and 20 (and 5, as aforementioned) leads to statistically significant improvement in the fitness. This confirms our intuition: selecting ancestors also from previous generations (that, anyway, are not too far away) led to better results as a more wide set of genotypes is considered for recombination, and good characteristics of an individual that may have been lost during generation can be retrieved, thus decreasing the likelihood of being stuck in local minima. On the other hand, considering only 2 previous generations results in fitness values comparable with GSGP. This is reasonable considering that individuals of two subsequent generations do not differ too much. Thus, selecting ancestors from a generation or from the directly previous one does not remarkably affect the quality of the offspring generated. On the other hand, U50 and U100 lead to significantly worse performance in terms of fitness. This is because ancestors are selected from generations too far back, where individuals were not yet improved by the genetic process.

Considering the results achieved by geometric multi-generational selection, while, as stated above, setting  $p = 0.25$  led to a significant upgrade of the fitness value; for the other choice of parameters, i.e.,  $p = 0.5$  and  $p = 0.75$ , the obtained results were worse with regards to standard GSGP.

These results are interesting for a particular reason: the expected value of the geometric distributions with  $p = 0.25$ ,  $p = 0.5$ , and  $p = 0.75$  are 3, 1, and  $\frac{1}{3}$ , respectively (i.e.,  $\frac{1-p}{p}$ ). Thus, we would expect G0.25, G0.50, and G0.75 to behave similarly to U4, U2, and between U1 and U2, respectively. In the first case, it appears to be correct, while in the second one, this happens only in some of the datasets. However, the behaviour of G0.75 is quite different from what was expected. Since the motivation cannot be traced back to the expected value of the distribution, it could be due to the fact that while it is expected that most of the individuals will be from the previous generation, only a limited number of them can be from older ones, damaging the search process. However, this is only a conjecture, and we expect to investigate this unexpected behaviour in later works together with the effect of using other distributions in the multi-generational selection.

## 6. Conclusions

In this paper, we have presented and studied a way to use the information of the previous generations, which GSGP stores anyway, to improve the performance of GSGP. This resulted in a multi-generational selection, and we presented two methods implementing this idea: a uniform selection probability and a geometric selection. The main idea consists of selecting individuals not only from the last population but also from previous ones (how many and with which probability depending on the underlying method). We have tested the proposed uniform and geometric multi-generational selections with multiple parameters on a selection of six datasets, showing that a *limited* ability to select from previous populations is beneficial to the search process. In the future, we plan to expand this research and provide even more powerful ways of exploiting the additional information that GSGP, in its fast implementation, is already storing.

**Author Contributions:** All authors contributed equally to this manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by national funds through the FCT (Fundação para a Ciência e a Tecnologia) by the projects GADgET (DSAIPA/DS/0022/2018).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

%F	Human oral bioavailability
%PPB	Protein-plasma binding level
air	Airfoil self-noise
conc	Concrete compressive strength
Gp	Geometric multi-generational selection with parameter $p$
EA	Evolutionary algorithms
GA	Genetic algorithms
GP	Genetic programming
GSGP	Geometric semantic genetic programming
LD50	Median oral lethal dose
RMSE	Root-mean-square errors
Uk	Uniform multi-generational selection with parameter $k$
yac	Yacht hydrodynamics

## References

1. Koza, J.R. Genetic programming as a means for programming computers by natural selection. *Stat. Comput.* **1994**, *4*, 87–112. [\[CrossRef\]](#)
2. Moraglio, A.; Krawiec, K.; Johnson, C.G. Geometric semantic genetic programming. In Proceedings of the International Conference on Parallel Problem Solving from Nature, Taormina, Italy, 1–5 September 2012; pp. 21–31.
3. Vanneschi, L.; Castelli, M.; Manzoni, L.; Silva, S. A new implementation of geometric semantic GP and its application to problems in pharmacokinetics. In Proceedings of the European Conference on Genetic Programming, Vienna, Austria, 3–5 April 2013; pp. 205–216.
4. Castelli, M.; Manzoni, L. GSGP-C++ 2.0: A geometric semantic genetic programming framework. *SoftwareX* **2019**, *10*, 100313. [\[CrossRef\]](#)
5. Vanneschi, L.; Silva, S.; Castelli, M.; Manzoni, L. Geometric semantic genetic programming for real life applications. In *Genetic Programming Theory and Practice XI*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 191–209.
6. Louis, S.; Li, G. Augmenting genetic algorithms with memory to solve traveling salesman problems. In Proceedings of the Joint Conference on Information Sciences, Nagoya, Japan 23–29 August 1997; pp. 108–111.
7. Wiering, M. Memory-based memetic algorithms. In Proceedings of the Benelearn'04—Thirteenth Belgian-Dutch Conference on Machine Learning, Brussels, Belgium, 8–9 January 2004; pp. 191–198.
8. Yang, S. Genetic Algorithms with Memory- and Elitism-Based Immigrants in Dynamic Environments. *Evol. Comput.* **2008**, *16*, 385–416. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Cao, Y.; Luo, W. Novel Associative Memory Retrieving Strategies for Evolutionary Algorithms in Dynamic Environments. In *Lecture Notes in Computer Science, Proceedings of the Advances in Computation and Intelligence—4th International Symposium, ISICA 2009, Huangshi, China, 23–25 October 2009*; Cai, Z., Li, Z., Kang, Z., Liu, Y., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5821, pp. 258–268.
10. Castelli, M.; Manzoni, L.; Vanneschi, L. The effect of selection from old populations in genetic algorithms. In *Companion Material Proceedings, Proceedings of the 13th Annual Genetic and Evolutionary Computation Conference, GECCO 2011, Dublin, Ireland, 12–16 July 2011*; Krasnogor, N., Lanzi, P.L., Eds.; ACM: New York, NY, USA, 2011; pp. 161–162.
11. Castelli, M.; Manzoni, L.; Vanneschi, L. A Method to Reuse Old Populations in Genetic Algorithms. In *Progress in Artificial Intelligence, Proceedings of the 15th Portuguese Conference on Artificial Intelligence, EPIA 2011, Lisbon, Portugal, 10–13 October 2011*; Lecture Notes in Computer Science; Antunes, L., Pinto, H.S., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; Volume 7026, pp. 138–152.

12. Augusto, D.A.; Barbosa, H.J.C. Symbolic Regression via Genetic Programming. In Proceedings of the 6th Brazilian Symposium on Neural Networks (SBRN 2000), Rio de Janeiro, Brazil, 22–25 November 2000; pp. 173–178.
13. Seront, G. External concepts reuse in genetic programming. In Proceedings of the AAAI Symposium on Genetic programming, MIT/AAAI, Cambridge, MA, USA, 10–12 November 1995; pp. 94–98.
14. Jaskowski, W.; Krawiec, K.; Wieloch, B. Knowledge reuse in genetic programming applied to visual learning. In Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2007, London, UK, 7–11 July 2007; pp. 1790–1797.
15. Pei, W.; Xue, B.; Shang, L.; Zhang, M. Reuse of program trees in genetic programming with a new fitness function in high-dimensional unbalanced classification. In Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO 2019), Prague, Czech Republic, 13–17 July 2019; pp. 187–188.
16. Bi, Y.; Xue, B.; Zhang, M. A Divide-and-Conquer Genetic Programming Algorithm With Ensembles for Image Classification. *IEEE Trans. Evol. Comput.* **2021**, *25*, 1148–1162. [[CrossRef](#)]
17. Castelli, M.; Manzoni, L.; Silva, S.; Vanneschi, L.; Popovič, A. The influence of population size in geometric semantic GP. *Swarm Evol. Comput.* **2017**, *32*, 110–120. [[CrossRef](#)]
18. Sipper, M.; Moore, J.H. Conservation machine learning. *BioData Min.* **2020**, *13*, 9. [[CrossRef](#)]
19. Sipper, M.; Moore, J.H. Conservation machine learning: A case study of random forests. *Sci. Rep.* **2021**, *11*, 3629. [[CrossRef](#)] [[PubMed](#)]
20. Castelli, M.; Trujillo, L.; Vanneschi, L.; Silva, S.; Z.-Flores, E.; Legrand, P. Geometric Semantic Genetic Programming with Local Search. In Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2015, Madrid, Spain, 11–15 July 2015; pp. 999–1006.
21. Castelli, M.; Manzoni, L.; Mariot, L.; Saletta, M. Extending Local Search in Geometric Semantic Genetic Programming. In *Lecture Notes in Computer Science, Proceedings of the Progress in Artificial Intelligence—19th EPIA Conference on Artificial Intelligence, EPIA 2019, Vila Real, Portugal, 3–6 September 2019*; Oliveira, P.M., Novais, P., Reis, L.P., Eds.; Springer: Berlin/Heidelberg, Germany, 2019; Volume 11804, pp. 775–787.
22. Vanneschi, L.; Castelli, M.; Silva, S. A survey of semantic methods in genetic programming. *Genet. Program. Evolvable Mach.* **2014**, *15*, 195–214. [[CrossRef](#)]
23. Moraglio, A.; Poli, R. Topological interpretation of crossover. In Proceedings of the Genetic and Evolutionary Computation Conference, Seattle, WA, USA, 26–30 June 2004; pp. 1377–1388.
24. McDermott, J.; White, D.R.; Luke, S.; Manzoni, L.; Castelli, M.; Vanneschi, L.; Jaskowski, W.; Krawiec, K.; Harper, R.; De Jong, K.; et al. Genetic programming needs better benchmarks. In Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation, Philadelphia, PA, USA, 7–11 July 2012; pp. 791–798.
25. Archetti, F.; Lanzeni, S.; Messina, E.; Vanneschi, L. Genetic programming and other machine learning approaches to predict median oral lethal dose (LD 50) and plasma protein binding levels (% PPB) of drugs. In Proceedings of the European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, Valencia, Spain, 11–13 April 2007; pp. 11–23.
26. Archetti, F.; Lanzeni, S.; Messina, E.; Vanneschi, L. Genetic programming for human oral bioavailability of drugs. In Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, Seattle, WA, USA, 8–12 July 2006; pp. 255–262.
27. Brooks, T.F.; Pope, D.S.; Marcolini, M.A. *Airfoil Self-Noise and Prediction*; Technical Report; NASA: Washington, DC, USA, 1989.
28. Castelli, M.; Vanneschi, L.; Silva, S. Prediction of high performance concrete strength using genetic programming with geometric semantic genetic operators. *Expert Syst. Appl.* **2013**, *40*, 6856–6862. [[CrossRef](#)]
29. Pietropoli, G.; Manzoni, L.; Paoletti, A.; Castelli, M. Combining Geometric Semantic GP with Gradient-Descent Optimization. In Proceedings of the European Conference on Genetic Programming (Part of EvoStar), Madrid, Spain, 20–22 April 2022; pp. 19–33.