Research Article

A Hybrid Genetic Programming Method in Optimization and Forecasting: A Case Study of the Broadband Penetration in OECD Countries

Konstantinos Salpasaranis and Vasilios Stylianakis

Electrical and Computer Engineering Department, University of Patras, 26500 Rio Patra, Greece

Correspondence should be addressed to Konstantinos Salpasaranis, salpk@upatras.gr

The introduction of a hybrid genetic programming method (hGP) in fitting and forecasting of the broadband penetration data is proposed. The hGP uses some well-known diffusion models, such as those of Gompertz, Logistic, and Bass, in the initial population of the solutions in order to accelerate the algorithm. The produced solutions models of the hGP are used in fitting and forecasting the adoption of broadband penetration. We investigate the fitting performance of the hGP, and we use the hGP to forecast the broadband penetration in OECD (Organisation for Economic Co-operation and Development) countries. The results of the optimized diffusion models are compared to those of the hGP-generated models. The comparison indicates that the hGP manages to generate solutions with high-performance statistical indicators. The hGP cooperates with the existing diffusion models, thus allowing multiple approaches to forecasting. The modified algorithm is implemented in the Python programming language, which is fast in execution time, compact, and user friendly.

1. Introduction

Many methods have been proposed for predicting the penetration of new technology in a community. The subject has been described and analyzed by worldwide literature, extensively [1–6].

Examples of the above methods are the diffusion models for the adoption of new technologies. The diffusion models are mathematical functions that follow an S-shaped curve in time. The diffusion models used in this study are the Gompertz, Logistic, and Bass [4]. The parameters of the models have been estimated by regression analysis [5].

Genetic algorithm (GA) is a probabilistic search method which uses the Darwinian principle of natural selection in finding an appropriate solution of a specific problem [7].

GP is more general than GA, because the produced solution corresponds to a new program [8]. The implementation of genetic programming (GP) in optimization problems has produced some important forecasting tools [7, 8].

Generally, a GP begins with a set of initial randomly chosen functions (solutions) and this set is called population. A chromosome is a program solution of GP. Each solution has a fitness value, and this chromosome's fitness is evaluated. The next generation is the resultant of the Darwinian selection process. In this process, the best chromosomes, according to their fitness values, are selected for the next generation. Some of the selected chromosomes are randomly combined (crossover) and generate new chromosomes (offspring). The mutation process also occurs, according to which a part of a randomly selected chromosome is changing. Finally, the chromosomes with better fitness values have better probability of being selected. The chromosomes of the new generation have better overall fitness value than those of the past generations. The whole process is repeated until an end condition has occurred [8, 9].

This paper is structured as follows. At first, the diffusion models are shortly presented. The basic structure of the new modified GP (hGP) and its analysis follow. The next section contains the results analysis, and, finally, the conclusion is presented. In the Appendices A, B, and C, the syntax of some produced models and the statistic indicators as well as the estimation formulas of the modified GA are provided.

2. Diffusion Models

Typically the diffusion process of innovative technologies in a society, such as broadband access, follows the sigmoid curve [6]. In this paper, some well-known diffusion models have been used as follows.

2.1. Logistic Model

The logistic diffusion model is the most known model among the sigmoid curves. The model is described by (2.1):

$$y(t) = \frac{S}{1 + a \cdot e^{f\{t\}}},\tag{2.1}$$

where the diffusion of a new product in a society at time t is presented as $y(t)$. Also, $f(t) = a + b \cdot t$ is a time-dependant function, and a, b are constant parameters. The S parameter is the limit of the function $y(t)$. When time $t \rightarrow \infty$, then $y(t) \rightarrow S$ [5].

2.2. Gompertz Model

The Gompertz model was introduced as a law of mortality [10]. The equation that we use is the known format of (2.2), as follows:

$$y(t) = S \cdot e^{e^{f(t)}}.\tag{2.2}$$

We can also use the alternative format, Gompertz with constant, which is described in (2.3):

$$y(t) = S \cdot e^{e^{f(t)}} + c,$$ (2.3)

where $f(t) = a + b \cdot t$ is a time-dependent function and a, b, c are constant parameters [5].

2.3. Bass Model

The Bass model proposes that the market for a new product consists of two major categories: innovators and imitators.

Firstly, the innovators purchase the new product, and the imitators follow afterwards.

This model's function is an extended logistic curve, where the cumulative adoption of the new technology $y(t)$ for time t is presented in (2.4):

$$y(t) = \frac{A - C \cdot e^{-B \cdot t}}{1 + D \cdot e^{-B \cdot t}}.$$ (2.4)

In (2.4), parameter A corresponds to initial purchasers of the new technology product. Parameter B is the sum of the innovators and imitators coefficients, p and q, respectively. This is presented as $B = p + q$.

C parameter is $C = r \cdot p$, where r is a constant and D parameter is $D = r \cdot q / A$ [4, 11].

3. Genetic Programming Method

The specific hGP implements a hybrid strategy, which consists of two parts, the nonlinear regression analysis and the modified genetic algorithm part. The flowchart of Figure 1 shows the parts of the hGP.

Firstly, a random set of solutions is created. Simultaneously, the candidate diffusion models are optimized by regression analysis and the produced models are inserted into the initial set of solutions. Now, the initial population of the hGP is prepared. Then, each solution is evaluated by its fitness function and the best solutions are inserted into a sorted list. The first, initial generation is ready. The termination criterion of the program is the maximum number of generations. After that, the selection of the best solutions takes place. It consists of the random combination of the best solutions, according to the crossover law and the random change of another randomly chosen solution. These solutions are going to be optimized by regression, and the next generation is ready. Finally, the whole process is repeated.

3.1. Solution Representation

In GP, each chromosome is not a fixed-length character string but a program that is a possible solution to the problem [12]. A chromosome in hGP, specifically, is presented as a string of characters in the Python programming language or as a parse tree. For example, the chromosome $a + b^* \cdot e^{(c - d^* t)}$ is presented in Figures 2 and 3 as string and parse tree, respectively [12].
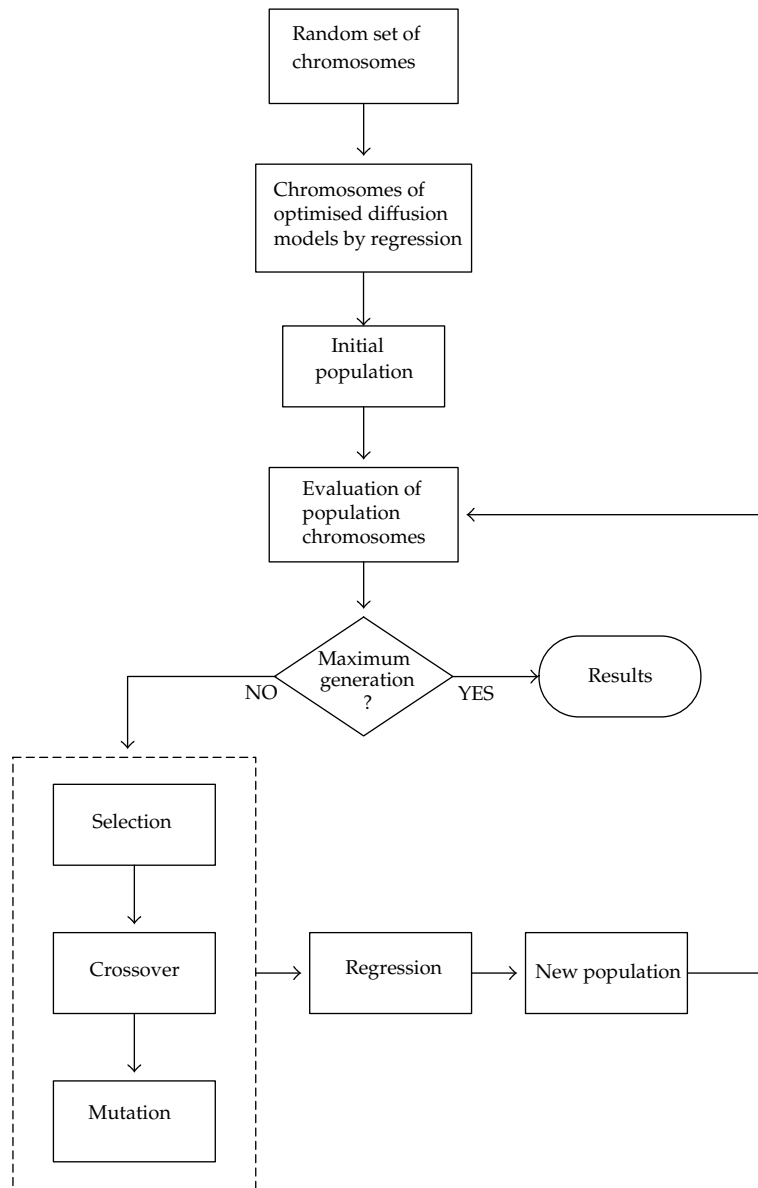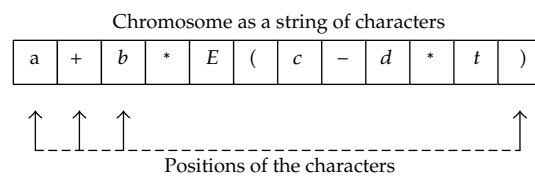
Figure 1: Flowchart of the hGP.

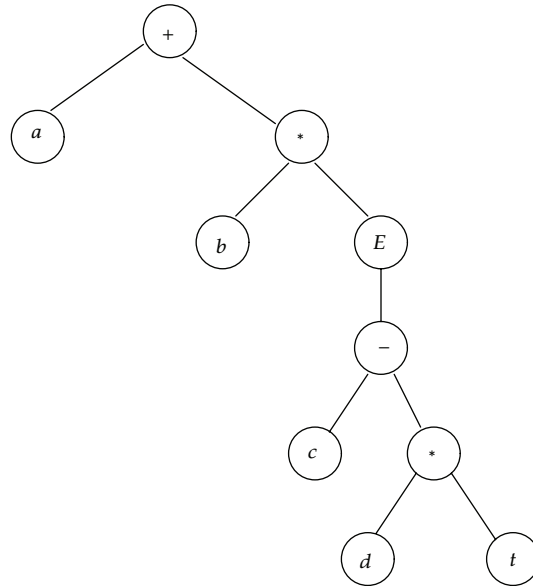**Figure 2:** Representation of a chromosome in hGP as string.

Figure 3: Representation of a chromosome in hGP as a parse tree.

The functions that are used in hGP are the addition (+), subtraction (−), protected division (/), multiplication (*), and exponential (E), because these are the same with the diffusion models' functions.

The parse tree consists of nodes. The terminal nodes are referred to as leaves. The leaves of the tree contain the variables or the constants of the program. As we can see the nonterminal nodes of the tree correspond to the function set of the hGP.

3.2. Initial Population

The initial population is generated by the fusion of a randomly produced number of chromosomes and the diffusion models which are optimized by regression analysis. The initial diffusion models' chromosomes are the optimized Logistic, Gompertz family, and Bass models. It should be noted that the diffusion models may differ according to the problem. The parameters of the initial diffusion models are optimized by nonlinear regression analysis, and the Levenberg-Marquardt algorithm has been used.

From a programmer's point of view, each chromosome is a string data type which is presented as a parse tree in the internal structure of the Python programming language. The population appears as a list of strings. This list corresponds to the first generation for the program.

3.3. Evaluation

During the evaluation process, we can use many functions as fitness indicators (with real data). In the specific implementation, we use two different fitness functions as follows.
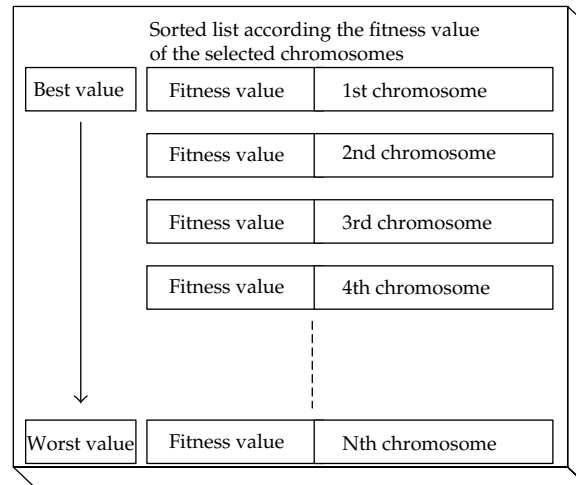
Sorted list according the fitness value of the selected chromosomes		
Best value	Fitness value	1st chromosome
	Fitness value	2nd chromosome
	Fitness value	3rd chromosome
	Fitness value	4th chromosome
Worst value	Fitness value	Nth chromosome

Figure 4: Representation of the sorted list of the selected chromosomes in hGP.

In the fitting process, each chromosome is evaluated with the sum of squared error (SSE), as in Appendix C in (C.1). In forecasting, the evaluation function corresponds to the weighted sum of squared error (wSSE) function, as in (C.2).

3.4. Selection

In hGP, the chromosomes with the best values of the fitness function (smaller than a precision limit) are inserted into a Python's list. The members of the list are sorted according to their fitness value. In Figure 4, we can see the structure of the sorted list. Then some randomly chosen chromosomes will be selected in order to produce the next generation using crossover and mutation processes.

3.5. Crossover

As it has been aforementioned, in the implementation of hGP, each chromosome is a string data type or a parse tree in the internal structure of the programming language. In the crossover process, two parents are randomly selected from the chromosomes' list with the best fitness value.

A crossover point is randomly chosen in the first parent string. Also another crossover point is randomly chosen in the second parent string. The first child (offspring) is generated when the substring of the first parent, which begins at the crossover point, is replaced by the substring from the second parent, which begins at the second crossover point. The second child is generated by the crossing over of the other parts (substrings) of the parents' strings. In Figures 5 and 6, the crossover operation is presented.
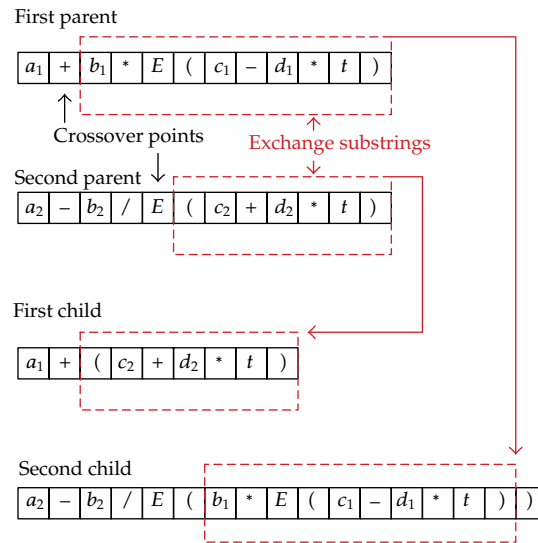
First parent

Figure 5: Crossover of the hybrid genetic programming method (string representation).

3.6. Mutation

In the mutation process, a chromosome is randomly chosen from the chromosomes list, with the best fitness value. The substring point, in which the mutation will take place, is also randomly chosen. The mutation replaces the function $(+, -, /, *, E)$ which is presented at the mutation point with a new random function in the substring.

The mutation operation is presented in Figures 7 and 8.

4. Results

In this section, the dataset, the statistic indicators, and the results of the study will be presented. The results will also be analysed in order to provide a satisfactory prediction for the broadband penetration in OECD countries.

4.1. Dataset

In this study, the proposed method has been implemented on two different datasets. The first dataset presents the overall OECD broadband penetration. The data came from the OECD portal [13], which presents the total fixed (wired) broadband penetration in OECD countries. According to OECD reports, the overall broadband penetration is rapidly growing. The dataset concerns the time period from the second quarter (Q2) of the year 2002 until the forth quarter (Q4) of the year 2010. The dataset is comprised of 18 data points.

The method is also implemented on the second dataset which concerns three innovators countries in fixed (wired) broadband technology adoption [14], namely, Sweden, The Netherlands, and Denmark. This dataset concerns the time period from the second quarter (Q2) of the year 2001 until the fourth quarter (Q4) of 2010 that is 20 data points.
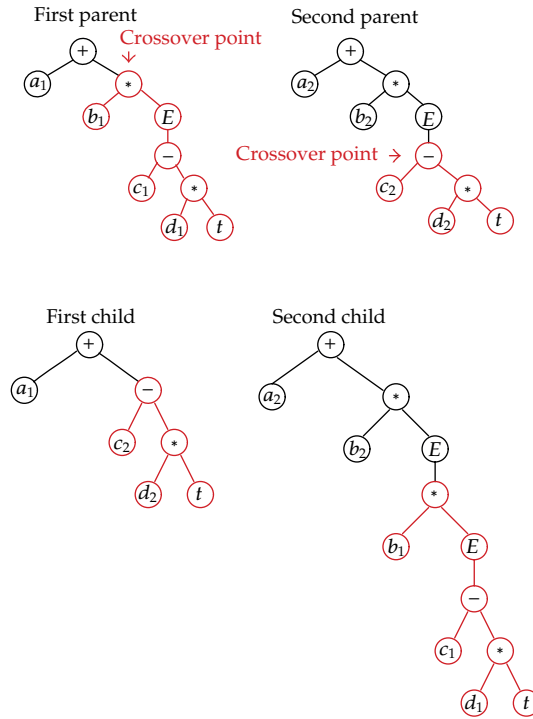
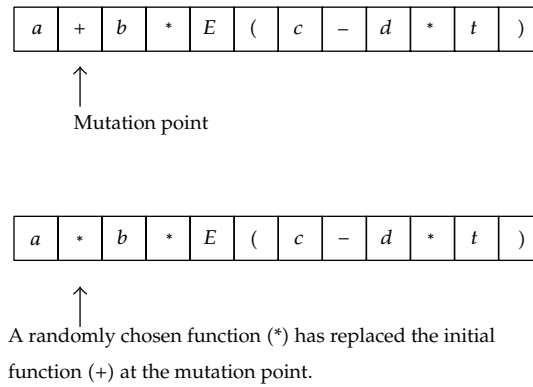**Figure 6:** Crossover of the hybrid genetic programming method (parse tree representation).

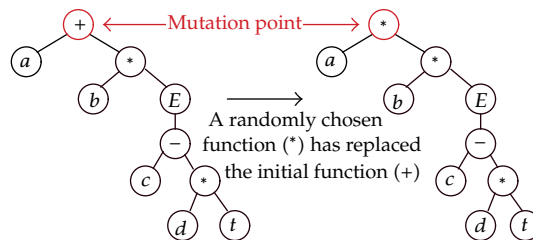**Figure 7:** Mutation of the hybrid genetic programming method (string representation).

Figure 8: Mutation of the hybrid genetic programming method (parse tree representation).

Table 1: Parameters initialization of hGP.

Parameters of the hGP	
Maximum number of generations	500
Evaluation function	SSE
Upper limit of the precision for the candidates for crossover and mutation	0.5

Table 2: Statistical indices of the first five produced hGP models (OECD overall).

Statistical indices of the hGP for OECD overall dataset				
Model name	MAPE	SSE	MSE	RMSE	MAE
hGP Model 1	0.007863	$4.53047E-05$	$2.51693E-06$	0.001586	0.001301
hGP Model 2	0.009163	$4.91669E-05$	$2.73149E-06$	0.001653	0.001271
hGP Model 3	0.010628	$4.95931E-05$	$2.75517E-06$	0.00166	0.001363
hGP Model 4	0.010628	$4.95931E-05$	$2.75517E-06$	0.00166	0.001363
hGP Model 5	0.010628	$4.95932E-05$	$2.75518E-06$	0.00166	0.001363

Thus, the fitting and forecasting ability of hGP are both tested in markets that have reached or are going to reach the saturation point of the penetration curve.

4.2. Statistic Indicators

As mentioned above, the fitness function of each individual is the sum of squared error (SSE) for the fitting process. The results are analyzed by the estimation of some widely used statistical indices. The main indices of our analysis are the mean absolute percentage error (MAPE), the mean square error (MSE), the mean absolute error (MAE), and the root mean square error (RMSE).

The statistic indicators of MAPE, MSE, MAE and RMSE are presented in Appendix B.

4.3. Fitting Results

As aforementioned, the statistic indicator used for the fitting process is the SSE. In this section, the fitting performance of the generated models by the hGP as well as the comparison of the optimized diffusion models with the hGP's models is presented. Both hGP and optimized diffusion models concerning the fitting performance are presented in Appendix A.1.

4.3.1. Fitting Results for the Overall OECD Broadband Penetration

Table 1 contains the initialization parameters for the execution of hGP concerning the whole dataset of the OECD countries.

An example of the fitting performance for the first five hGP models, according to their fitness value (SSE), is presented in Figure 9. The graphs represent the broadband penetration percentage (y-axis) and time (x-axis) as date. Each time point (x-axis) corresponds to a six-month period. The relative statistical indices SSE, MAPE, MSE, RMSE, and MAE of the produced hGP models are presented in Table 2.

Also, the fitting performance of the optimized diffusion models, according to their fitness value (SSE), is presented in Figure 10 and their statistical indices in Table 3.
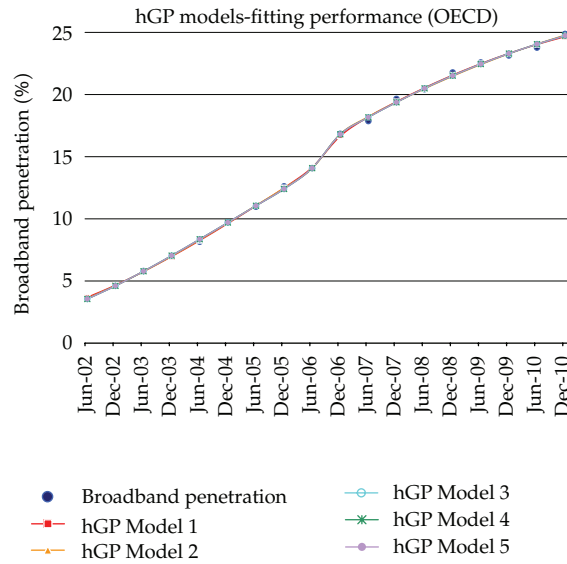
Figure 9: The fitting performance of the first five hGP models (OECD overall).
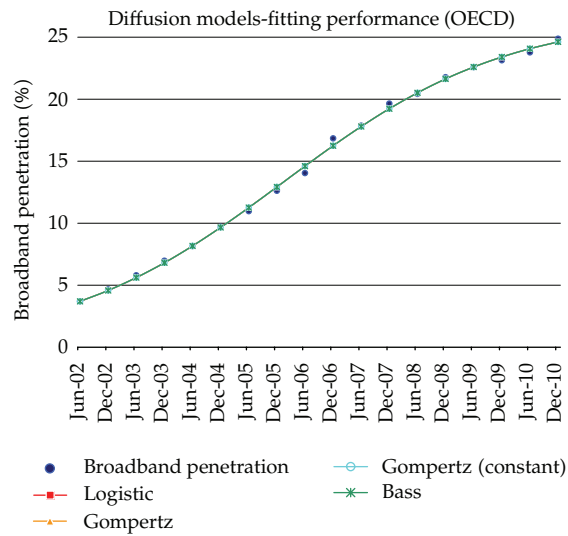
Figure 10: The fitting performance of the optimized diffusion models for OECD overall dataset.

Table 3: Representation of the statistics for the optimized diffusion models sorted by SSE.

Statistical indices of the optimized diffusion models for OECD overall dataset					
Model name	MAPE	SSE	MSE	RMSE	MAE
Logistic	0.015544	0.000134615	$7.47864E-06$	0.00273471	0.00211296
Bass	0.015544	0.000134616	$7.47867E-06$	0.00273472	0.00211297
Gompertz with constant	0.022755	0.000166627	$9.25707E-06$	0.00304254	0.00249006
Gompertz	0.033951	0.000316133	$1.7563E-05$	0.00419082	0.0037315

Table 4: Parameters initialization of the hGP.

Parameters of the hGP	
Maximum number of generations	500
Evaluation function	SSE
Upper limit of the precision for the candidates for crossover and mutation	0.5

Table 5: Statistical indices of the first five produced hGP models (Sweden).

Statistical indices of the hGP for Sweden dataset				
Model name	MAPE	SSE	MSE	RMSE	MAE
hGP Model 1	0.008675744	$7.46756E-05$	$4.63861E-05$	$3.73378E-06$	0.001932299
hGP Model 2	0.009181524	$8.90721E-05$	$5.39797E-05$	$4.45361E-06$	0.002110357
hGP Model 3	0.033258076	0.000264358	$6.12299E-05$	$1.32179E-05$	0.003635641
hGP Model 4	0.048504881	0.000515968	0.000103702	$2.57984E-05$	0.005079212
hGP Model 5	0.04856113	0.000517456	0.000104259	$2.58728E-05$	0.005086532

Table 6: Representation of the statistics for the optimized diffusion models sorted by SSE (Sweden).

Statistical indices of the optimized diffusion models for Sweden dataset				
Model name	MAPE	SSE	MSE	RMSE	MAE
Logistic	0.055827	0.001689	$8.45E-05$	0.009191	0.007991
Bass	0.055827	0.001689	$8.45E-05$	0.009191	0.007991
Gompertz with constant	0.083396	0.002071	0.000104	0.010175	0.009002
Gompertz	0.086425	0.003551	0.000178	0.013326	0.011753

When considering Tables 2 and 3, the hGP method achieves better statistical indices than those of the optimized diffusion models. For example, the first hGP model shows a SSE value of 4.53047E−05 and the Logistic model 0.000134615. It should be noted that hGP was executed for 500 generations, which is a rather medium number of generations. Thus, better results could be achieved for larger number of generations.

4.3.2. Fitting Results for Sweden, The Netherlands, and Denmark Broadband Penetration

Table 4 contains the initialization parameters for the execution of hGP concerning the dataset for The Netherlands-Sweden-Denmark countries.

The fitting performance for the first five hGP models, according to their fitness value (SSE) as well as the performance of the optimized diffusion models is presented in Figures 11 and 12 for Sweden, Figures 13 and 14 for The Netherlands, and Figures 15 and 16 for Denmark, respectively. The relative statistical indices SSE, MAPE, MSE, RMSE, and MAE of the produced hGP models are presented in Tables 5 and 6 for Sweden, Tables 7 and 8 for The Netherlands, and Tables 9 and 10 for Denmark.

hGP models-fitting performance (Sweden)

Figure 11: The fitting performance of the first five hGP models for Sweden.

Diffusion models-fitting performance (Sweden)
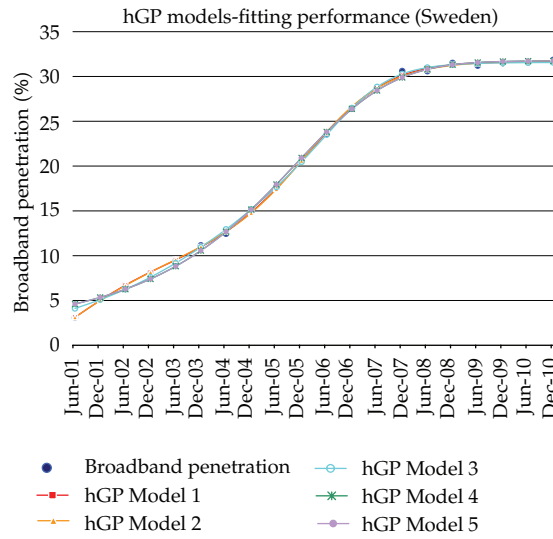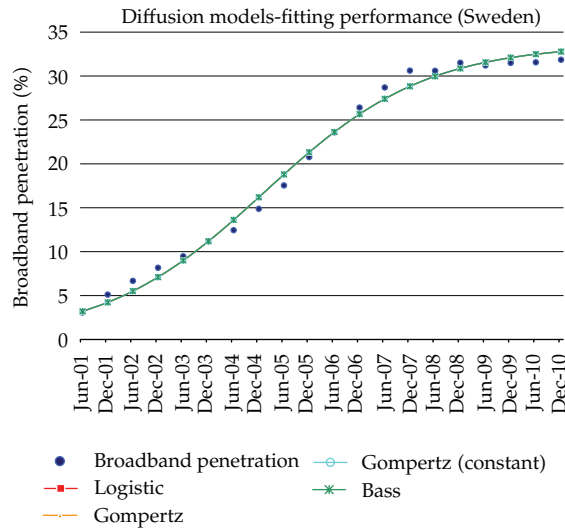
Figure 12: The fitting performance of the optimized diffusion models for Sweden.

Table 7: Statistical indices of the first five produced hGP models (The Netherlands).

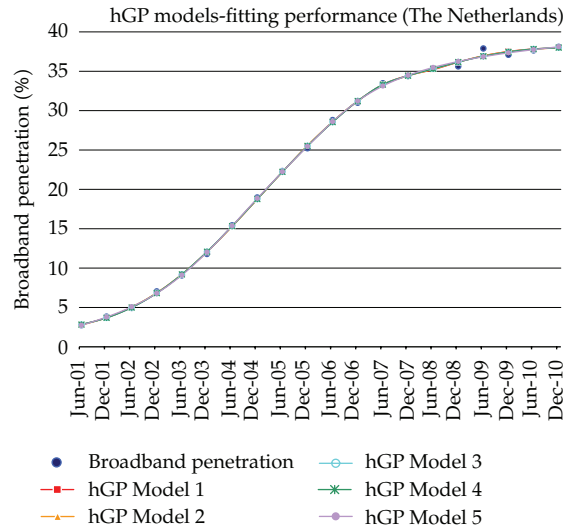| Statistical indices of hGP for The Netherlands dataset				
Model name	MAPE	SSE	MSE	RMSE	MAE
hGP Model 1	0.013729	0.00018787	$9.39E - 06$	0.003065	0.002343
hGP Model 2	0.012842	0.00018848	$9.42E - 06$	0.00307	0.002304
hGP Model 3	0.01377	0.00019367	$9.68E - 06$	0.003112	0.002264
hGP Model 4	0.01377	0.00019367	$9.68E - 06$	0.003112	0.002264
hGP Model 5	0.010985	0.00020294	$1.01E - 05$	0.003185	0.002155

Figure 13: The fitting performance of the first five hGP models for The Netherlands.
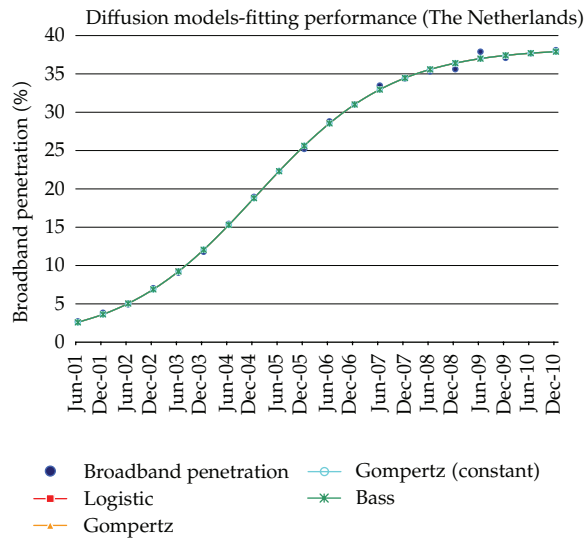
Figure 14: The fitting performance of the optimized diffusion models for The Netherlands.

Table 8: Statistics for the optimized diffusion models sorted by SSE (The Netherlands).

Statistical indices of the optimized diffusion models for The Netherlands dataset					
Model name	MAPE	SSE	MSE	RMSE	MAE
Logistic	0.015446	0.00023692	$1.18E-05$	0.003442	0.002548
Bass	0.015447	0.00023692	$1.18E-05$	0.003442	0.002548
Gompertz with constant	0.031535	0.000412654	$2.06E-05$	0.004542	0.003828
Gompertz	0.07946	0.001367898	$6.84E-05$	0.00827	0.007501

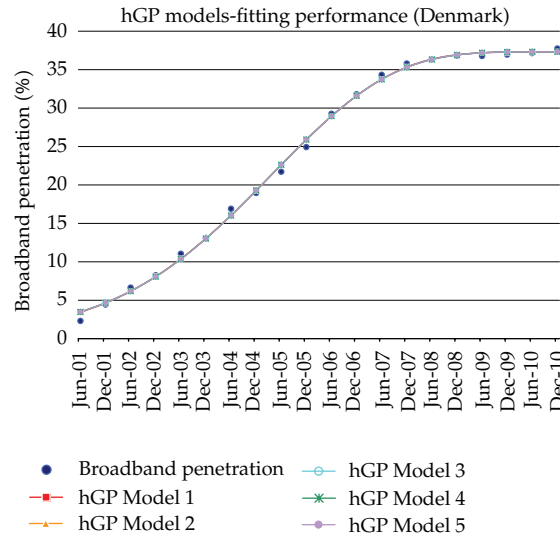**Figure 15:** The fitting performance of the first five hGP models for Denmark.
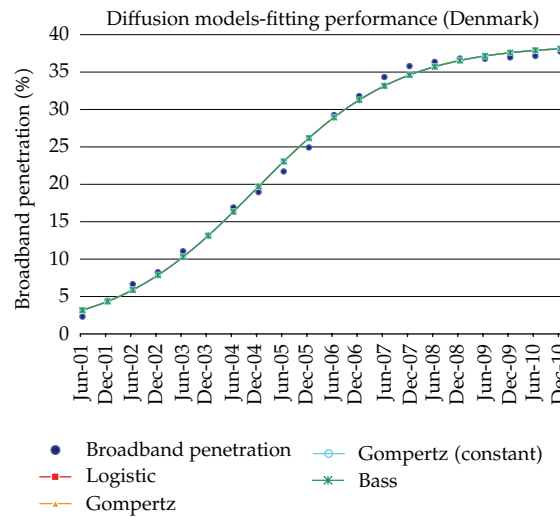
Figure 16: The fitting performance of the optimized diffusion models for Denmark.

According to Tables 5 and 6, the hGP method achieves better statistical indices than those of the optimized diffusion models. For example, the first hGP model achieves an SSE value $7.46756E−05$ while the Logistic model an SSE value of 0.001689.

Once again, according to Tables 7 and 8, the hGP method achieves better statistical indices than those of the optimized diffusion models.

Finally, according to Tables 9 and 10, the hGP method achieves better statistical indices than those of the optimized diffusion models. The first hGP model achieves an SSE value of 0.000566649 while the Logistic model 0.001110701.

Table 9: Statistical indices of the first five produced hGP models (Denmark).

		Statistical indices of hGP for Denmark dataset			
Model name	MAPE	SSE	MSE	RMSE	MAE
hGP Model 1	0.045512	0.000566649	$2.83324E-05$	0.005323	0.004332
hGP Model 2	0.047466	0.000599483	$2.99742E-05$	0.005475	0.004452
hGP Model 3	0.047487	0.000599823	$2.99912E-05$	0.005476	0.004453
hGP Model 4	0.047489	0.000599875	$2.99938E-05$	0.005477	0.004454
hGP Model 5	0.047531	0.000600572	$3.00286E-05$	0.00548	0.004456

Table 10: Statistics for the optimized diffusion models sorted by SSE (Denmark).

	Statistical indices of the optimized diffusion models for Denmark dataset				
Model name	MAPE	SSE	MSE	RMSE	MAE
Logistic	0.049134	0.001110701	$5.55E-05$	0.007452	0.006494
Bass	0.049134	0.001110704	$5.55E-05$	0.007452	0.006494
Gompertz with constant	0.049134	0.001110701	$5.55E-05$	0.007452	0.006494
Gompertz	0.049134	0.001110704	$5.55E-05$	0.007452	0.006494

Table 11: Initialization parameters of hGP.

| Parameters of the hGP	
Maximum number of generations	500
Evaluation function	wSSE
Upper limit of the precision for the candidates for crossover and mutation	0.09

4.4. Forecasting Results

The forecasting results of the generated models by hGP are presented in this section as well as the comparison of the optimized diffusion models with the hGP's models. The statistic indicator used for the forecasting process is the wSSE. Both hGP and the optimized diffusion models concerning the forecasting performance are presented in Appendix A.2.

4.4.1. Forecasting Results for the Overall OECD Broadband Penetration

Table 11 contains the initialization parameters for the execution of hGP for the forecasting process. The whole dataset of the OECD countries contains 18 data points (2 points per year since 2002 until 2010). The forecasting method for a 2-year prediction uses 14 points (data for hGP training) of the dataset and implements the statistical indices to this 14-point subset. The graph of the forecasting performance is shown in Figure 17. In every graph, the forecast period window is presented into the blue rectangle.

Table 12: Statistical indices of the first five produced forecasting hGP models for OECD overall.

	Statistical indices of the hGP for OECD forecasting (14 points training)				
Model name	MAPE	wSSE	MSE	RMSE	MAE
hGP Model 1	0.009744	$3.67091E - 05$	$4.02E - 06$	0.002005	0.001412
hGP Model 2	0.014944	$4.00502E - 05$	$4.82E - 06$	0.002195	0.001608
hGP Model 3	0.014527	$4.55281E - 05$	$5.38E - 06$	0.002319	0.001683
hGP Model 4	0.019881	$5.2309E - 05$	$6.65E - 06$	0.002579	0.002041
hGP Model 5	0.020006	$5.40818E - 05$	$6.86E - 06$	0.00262	0.002063
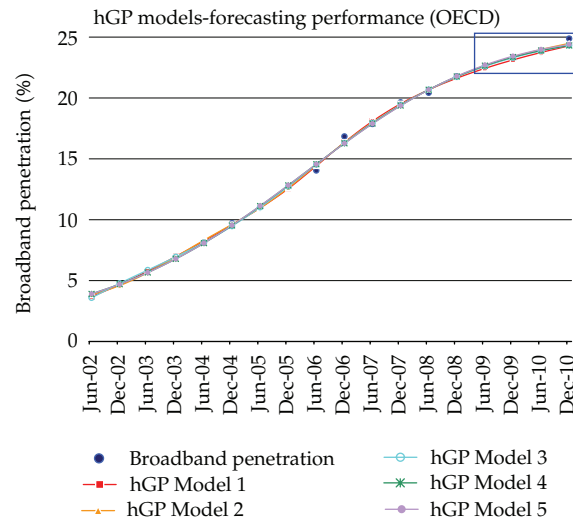
Figure 17: The forecasting performance of hGP (forecast period—2 years ahead).

The relative statistical indices, for the training points, of the produced forecasting hGP models are presented in Table 12.

The forecasting performance of the optimized diffusion models, according to their fitness value (wSSE) for the 14 training points, is presented in Figure 18 and their statistical indices in Table 13.

Considering Tables 12 and 13, the conclusion is that the hGP method achieves better statistical indices than those of the optimized diffusion models. We can see that the first hGP model achieves a wSSE value $3.67091E-05$ while the Logistic model $6.13566E-05$. The comparison of the models residuals against time (data points), especially for the 4 last data points (the forecast period), shows the predominance of the hGP models (see Figure 19).

4.4.2. Forecasting Results for the Sweden-The Netherlands-Denmark Broadband Penetration

Table 14 contains the initialization parameters for the execution of hGP in the forecasting process. The datasets of Sweden, The Netherlands, and Denmark contain 20 data points (2 points per year since 2001 until 2010). The forecasting method for a 2-year prediction uses 16 data points for hGP training of the dataset and implements the statistical indices to the subset
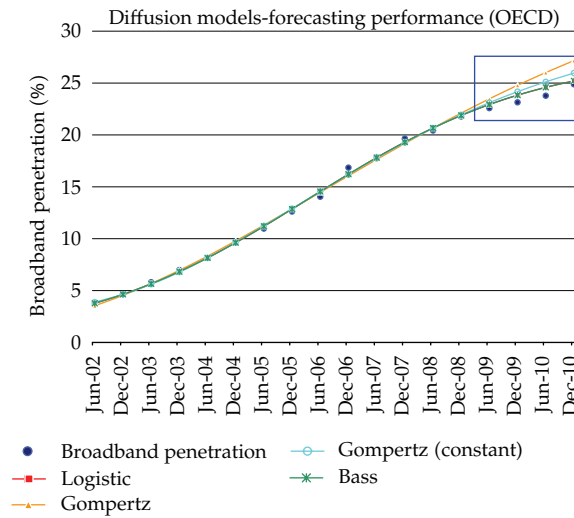
Figure 18: The forecasting performance of the optimized diffusion models (forecast period—2 years ahead).

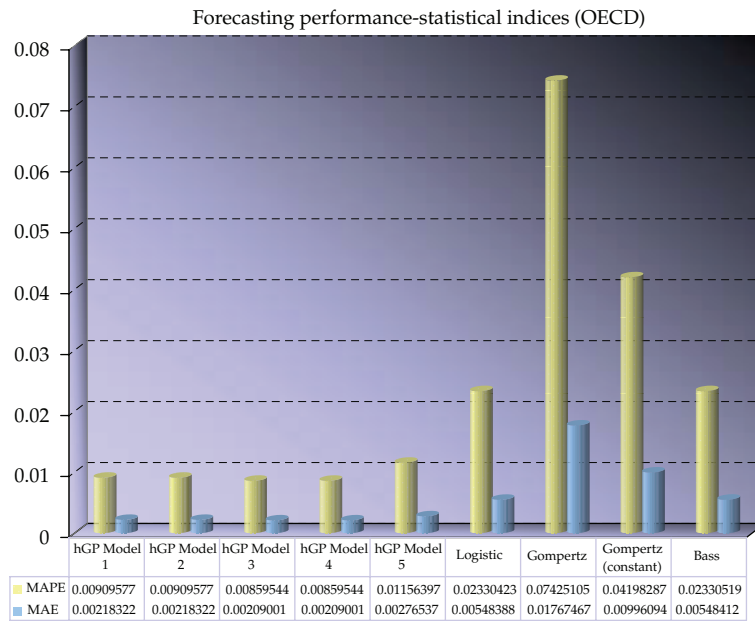|	hGP Model 1	hGP Model 2	hGP Model 3	hGP Model 4	hGP Model 5	Logistic	Gompertz	Gompertz (constant)	Bass
MAPE	0.00909577	0.00909577	0.00859544	0.00859544	0.01156397	0.02330423	0.07425105	0.04198287	0.02330519
MAE	0.00218322	0.00218322	0.00209001	0.00209001	0.00276537	0.00548388	0.01767467	0.00996094	0.00548412

Figure 19: The forecasting performance of the models (forecast period—2 years ahead).

of the training points. The graphs of the forecasting performance of the hGP and optimized diffusion models are shown in Figures 20 and 21 for Sweden, Figures 23 and 24 for The Netherlands, and Figures 26 and 27 for Denmark, respectively. The relative statistical indices for the training data are presented in Tables 15 and 16 for Sweden, Tables 17 and 18 for The Netherlands, and Tables 19 and 20 for Denmark. Finally, the statistical indices, MAPE and MAE, that describe the forecasting performance of the models in a forecasting horizon of 2-years are presented in Figures 22, 25, and 28 for the three countries.

Table 13: Representation of the statistics for the optimized diffusion models sorted by wSSE (forecast period—2 years ahead).

Statistical indices of the optimized diffusion models for OECD forecasting (14-point training)					
Model name	MAPE	wSSE	MSE	RMSE	MAE
Logistic	0.016716	$6.13566E-05$	$7.18E-06$	0.002679	0.002025
Bass	0.016716	$6.13573E-05$	$7.18E-06$	0.002679	0.002025
Gompertz with constant	0.021155	$6.98404E-05$	$8.51E-06$	0.002917	0.002349
Gompertz	0.022379	$9.5999E-05$	$1.08E-05$	0.003284	0.002678

Table 14: Parameters initialization of hGP.

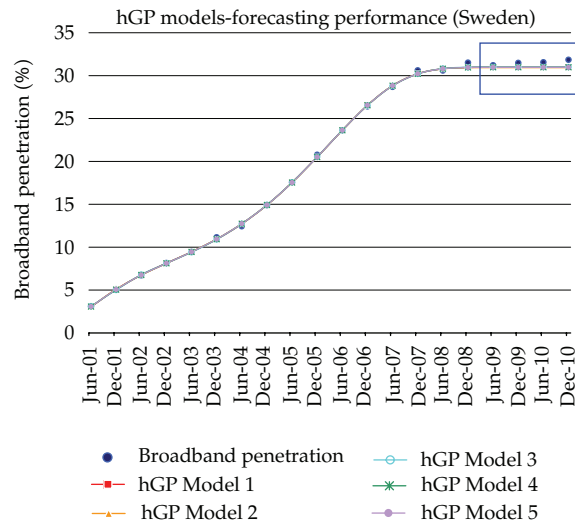| Parameters of the hGP	
Maximum number of generations	500
Evaluation function	wSSE
Upper limit of the precision for the candidates for crossover and mutation	0.9

Figure 20: The forecasting performance of the first five hGP models for Sweden (forecast period—2 years ahead).

(1) Forecasting Results for Sweden Broadband Penetration

Considering Tables 15 and 16, it is concluded that the hGP method again achieves better statistical indices than those of the optimized diffusion models for the training subset. The first hGP model has a wSSE value of $5.53E-05$ while the Logistic model 0.000433. It should be mentioned that the hGP achieves a satisfactory performance after the 16th data point, with minimized residuals (errors), MAPE and MAE for the last 4 data points (2-year forecast horizon, see Figure 22).

(2) Forecasting Results for The Netherlands Broadband Penetration

Further commenting on the forecasting results, according to Tables 17 and 18, it is concluded that the hGP method again achieves better statistical indices than those of the optimized
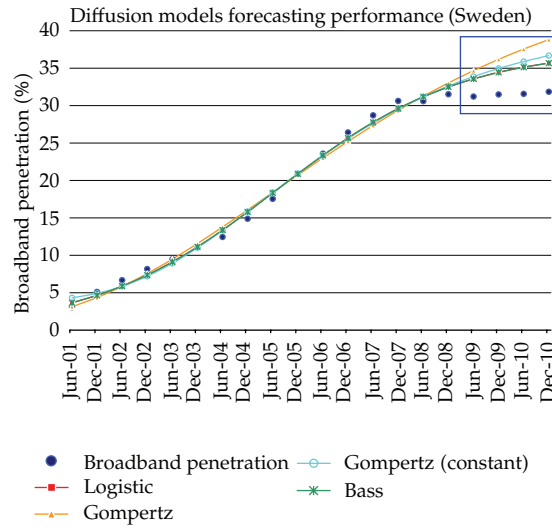
Figure 21: The forecasting performance of the optimized diffusion models for Sweden (forecast period—2 years ahead).

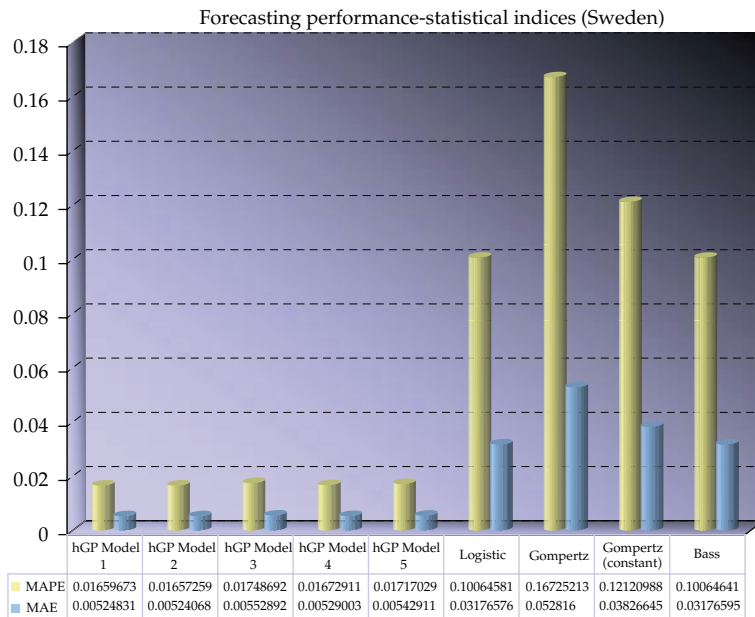|	hGP Model 1	hGP Model 2	hGP Model 3	hGP Model 4	hGP Model 5	Logistic	Gompertz	Gompertz (constant)	Bass
MAPE	0.01659673	0.01657259	0.01748692	0.01672911	0.01717029	0.10064581	0.16725213	0.12120988	0.10064641
MAE	0.00524831	0.00524068	0.00552892	0.00529003	0.00542911	0.03176576	0.052816	0.03826645	0.03176595

Figure 22: The forecasting performance of the models for Sweden (forecast period—2 years ahead).

diffusion models for the 16-point training subset. The first hGP model shows a wSSE value of 0.00010 while the Logistic and Bass models 7.31E-05. Both parts, the hGP and diffusion models, specially Logistic and Bass models, achieve well enough performance after the 16th data point, with minimized residuals (errors) for the last 4 data points (2-year forecast horizon, see Figure 25).
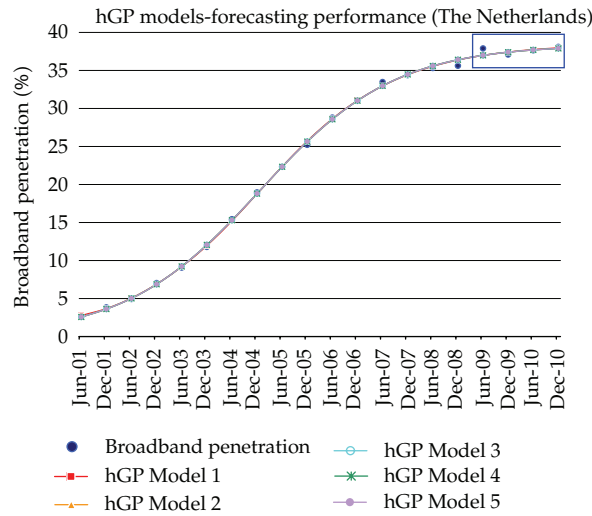
Figure 23: The forecasting performance of the first five hGP models and the optimized diffusion models for The Netherlands (forecast period—2 years ahead).
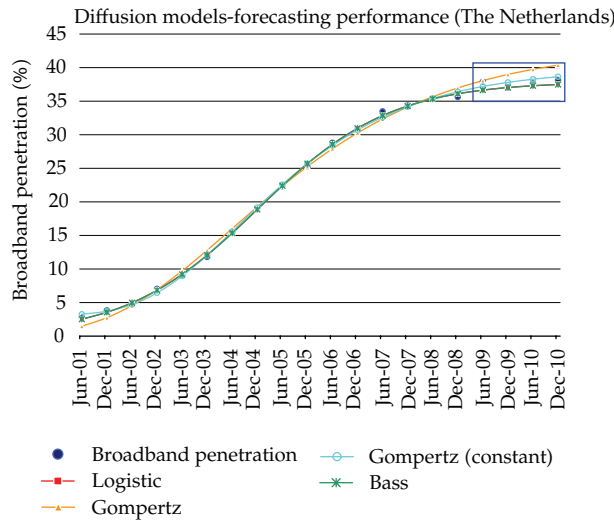

Figure 24: The forecasting performance of the optimized diffusion models for The Netherlands (forecast period—2 years ahead).

Table 15: Statistical indices of the first five produced hGP models (Sweden) forecast period—2 years.

Statistical indices of hGP for Sweden dataset (16-point training)				
Model name	MAPE	wSSE	MSE	RMSE	MAE
hGP Model 1	0.011145	$5.53E-05$	$4.81E-06$	0.002192	0.001692
hGP Model 2	0.008093	$5.66E-05$	$4.71E-06$	0.002171	0.001546
hGP Model 3	0.008504	$5.67E-05$	$4.69E-06$	0.002166	0.001538
hGP Model 4	0.008105	$5.69E-05$	$4.74E-06$	0.002176	0.001549
hGP Model 5	0.008151	$5.79E-05$	$4.81E-06$	0.002192	0.001557

Forecasting performance-statistical indices (The Netherlands)

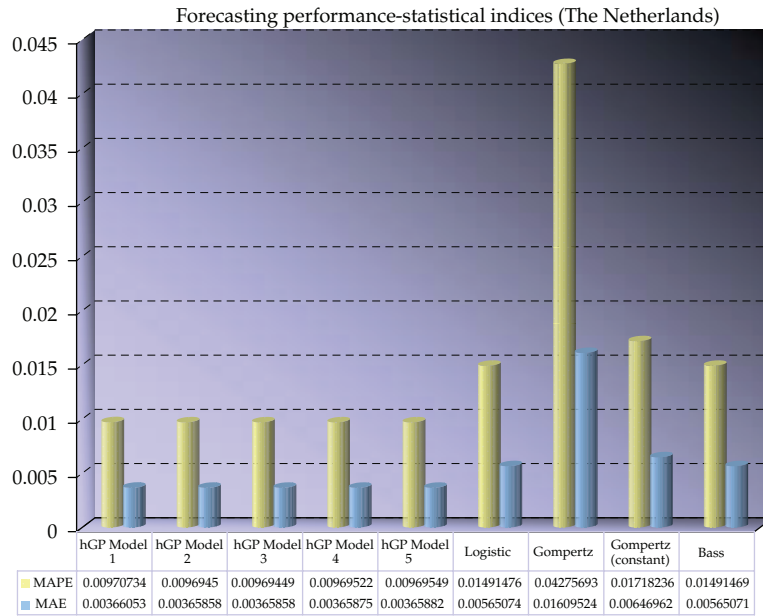|	hGP Model 1	hGP Model 2	hGP Model 3	hGP Model 4	hGP Model 5	Logistic	Gompertz	Gompertz (constant)	Bass
MAPE	0.00970734	0.0096945	0.00969449	0.00969522	0.00969549	0.01491476	0.04275693	0.01718236	0.01491469
MAE	0.00366053	0.00365858	0.00365858	0.00365875	0.00365882	0.00565074	0.01609524	0.00646962	0.00565071

Figure 25: The forecasting performance of the models for The Netherlands (forecast period—2 years ahead).
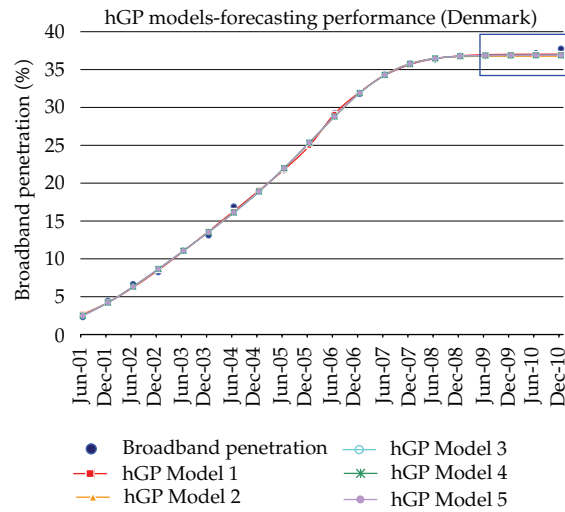
hGP models-forecasting performance (Denmark)

Figure 26: The forecasting performance of the first five hGP models for Denmark (forecast period—2 years ahead).

(3) Forecasting Results for Denmark Broadband Penetration

Finally, according to Tables 19 and 20, once again the hGP method achieves better statistical indices than those of the optimized diffusion models for the training subset. The first hGP
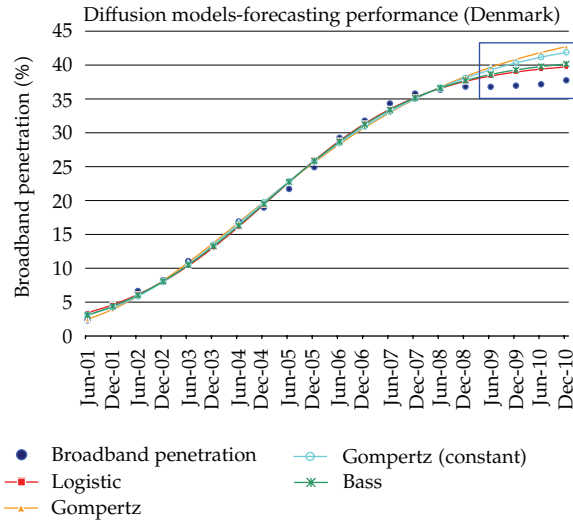
Figure 27: The forecasting performance of the optimized diffusion models for Denmark (forecast period—2 years ahead).

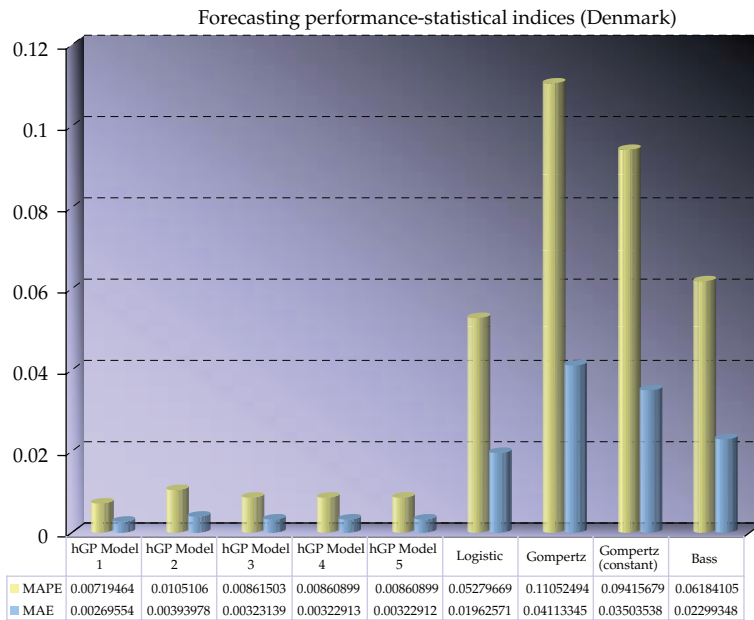|	hGP Model 1	hGP Model 2	hGP Model 3	hGP Model 4	hGP Model 5	Logistic	Gompertz	Gompertz (constant)	Bass
MAPE	0.00719464	0.0105106	0.00861503	0.00860899	0.00860899	0.05279669	0.11052494	0.09415679	0.06184105
MAE	0.00269554	0.00393978	0.00323139	0.00322913	0.00322912	0.01962571	0.04113345	0.03503538	0.02299348

Figure 28: The forecasting performance of the optimized diffusion models for Denmark (forecast period—2 years ahead).

model achieves a wSSE value of $3.39655E-05$ while the Logistic model 0.000319962. The hGP achieves well enough performance after the 16th data point, with minimized residuals (errors), MAPE and MAE for the last 4 data points (2-year forecast horizon, see Figure 28).

Table 16: Statistics for the optimized diffusion models sorted by wSSE (Sweden) forecast period—2 years.

Statistical indices of the optimized diffusion models for Sweden dataset (16-point training)					
Model name	MAPE	wSSE	MSE	RMSE	MAE
Logistic	0.054539	0.000433	$5.03E-05$	0.007089	0.006376
Bass	0.054539	0.000433	$5.03E-05$	0.007089	0.006376
Gompertz with constant	0.06964	0.000513	$6.58E-05$	0.008114	0.007267
Gompertz	0.054088	0.000825	$8.47E-05$	0.009202	0.007889

Table 17: Statistical indices of the first five produced hGP models for The Netherlands (forecast period—2 years).

Statistical indices of hGP for The Netherlands dataset (16-point training)					
Model name	MAPE	wSSE	MSE	RMSE	MAE
hGP Model 1	0.014673	0.000100	$8.58E-06$	0.002929	0.002194
hGP Model 2	0.017209	0.000103	$8.84E-06$	0.002973	0.00227
hGP Model 3	0.017209	0.000103	$8.84E-06$	0.002973	0.00227
hGP Model 4	0.017223	0.000103	$8.85E-06$	0.002975	0.002271
hGP Model 5	0.017253	0.000103	$8.86E-06$	0.002976	0.002272

Table 18: Statistics for the optimized diffusion models sorted by wSSE for The Netherlands (forecast period—2 years).

Statistical indices of the optimized diffusion models for The Netherlands dataset (16-point training)					
Model name	MAPE	wSSE	MSE	RMSE	MAE
Logistic	0.018228	$7.31E-05$	$7.36E-06$	0.002712	0.002069
Bass	0.018228	$7.31E-05$	$7.36E-06$	0.002712	0.002069
Gompertz with constant	0.033352	0.000162	$1.76E-05$	0.004199	0.003502
Gompertz	0.074251	0.000437	$5.84E-05$	0.007639	0.006336

5. Conclusions

This paper introduces a new GP method that produces fitting and forecasting solutions models with well enough performance. The whole process is assisted by the insertion of some widely used diffusion models like Logistic, Gompertz family, and Bass, in the initial population of the chromosomes. The hGP method was implemented with a dataset concerning the overall broadband penetration of the OECD countries, as well as the datasets of Sweden, The Netherlands, and Denmark which are pioneers in broadband technologies. Both the fitting and the forecasting performance of the method presented satisfactory statistical indices.

The proposed method differs from the classic GP method in several points. First, some well-known diffusion models, Logistic, Gompertz family, and Bass, which have been optimized by regression analysis, have been inserted in the randomly generated initial population. Second, the regression process has been implemented for each chromosome and after each crossover and mutation operation in order to maximize the algorithm's efficiency. Also, the crossover and mutation processes are implemented by a random

Table 19: Statistical indices of the first five produced hGP models for Denmark (forecast period—2 years).

	Statistical indices of the hGP for Denmark dataset (16-point training)				
Model name	MAPE	wSSE	MSE	RMSE	MAE
hGP Model 1	0.024844	$3.39655E-05$	$7.25E-06$	0.002692	0.002018
hGP Model 2	0.024808	$6.20484E-05$	$1.02E-05$	0.003201	0.002468
hGP Model 3	0.022951	$6.42515E-05$	$1.05E-05$	0.003233	0.002483
hGP Model 4	0.022944	$6.42636E-05$	$1.05E-05$	0.003233	0.002483
hGP Model 5	0.022944	$6.42636E-05$	$1.05E-05$	0.003233	0.002483

Table 20: Statistics for the optimized diffusion models sorted by wSSE for Denmark (forecast period—2 years).

	Statistical indices of the optimized diffusion models for Denmark dataset (16-point training)				
Model name	MAPE	wSSE	MSE	RMSE	MAE
Logistic	0.058957	0.000319962	$4.33E-05$	0.006582	0.005759
Bass	0.048066	0.000362021	$4.14E-05$	0.006434	0.005734
Gompertz with constant	0.05776	0.000562689	$6.02E-05$	0.007756	0.007007
Gompertz	0.042059	0.000683336	$6.6E-05$	0.008125	0.006943

selection of individuals from a sorted list which contains the chromosomes with the best performance.

In general, the produced hGP could be considered as a tool which generates solutions with well enough performance in fitting and forecasting of the broadband penetration.

In this paper, the forecasting horizon of hGP was two years ahead. A further investigation of the hGP performance for a longer forecast horizon would be desirable. It should be noted that the hGP method performance could be further improved with the insertion of more functions into the function set. A future study with an enrichment of the function set of the hGP could thus be considered.

Appendices

A.

In this section, some of the produced hGP models as well as the optimized diffusion models concerning the fitting and the forecasting performance that were analysed before are presented. Each produced model corresponds to a Python's string data type of the program.

A.1. Fitting Performance Models

For more details, see Tables 21, 22, 23, 24, 25, 26, 27, and 28.

A.2. Forecasting Performance Models

For more details, see Tables 29, 30, 31, 32, 33, 34, 35, and 36.

Table 21: Representation of the first five produced hGP models in fitting (OECD overall).

Model name	Model
	Fitting performance—hGP models for OECD overall
hGP Model 1	0.371142995572*E(−3.18695928215*E(−0.0179031441739*t))*E(−E(−4.31337138346e+11 *E(−0.39450322724*t))*(3.81808835539*t−0.290266301465+3.81536111536*t))
hGP Model 2	0.310471455673*E(−3.11349554655*E(−0.0204726299602*t))**E(−0.168943728131/(1+E (−(−50.6231851732+0.733563132765*t))))
hGP Model 3	0.297974863522*E(−3.04030207778*E(−0.0232771010194*t))*E(−0.12721433893*E (−109810.431808/(1+E(−(−40.4802143374+0.418592190558*t)))))
hGP Model 4	0.297974843236*E(−3.0403025557*E(−0.0232771051305*t))*E(−0.12721423417*E (−70692.9894579/(1+E(−(−40.0400529746+0.418595561014*t)))))
hGP Model 5	0.297974680462*E(−1.8377143802*E(−0.0232772109071*t))*E(−1.20258571767*E (−0.0232769748624*t))*E(−0.1272152801*E(−22946.6443056/(1+E(−(−38.9144349592 +0.418589930059*t)))))

Table 22: Representation of the optimized diffusion models in fitting (OECD overall).

Model name	Model
	Fitting performance—optimized diffusion models for OECD overall
Logistic	0.267462866067/(1+E(−(−2.58338763745+0.0419392545351*t)))
Bass	0.26746298117*(1−E(−(4.30693372098e−08+0.0419391296103)*(t−(−267.18543419))))/ (1+(0.0419391296103/4.30693372098e−08)*E(−(4.30693372098e−08+0.0419391296103) *(t−(−267.18543419))))
Gompertz with constant	0.24779967624*E(−6.09503667714*E(−0.0312988127799*t))+0.0323712682778
Gompertz	0.311174465387*E(−3.43140971411*E(−0.0230181532619*t))

Table 23: Representation of the first five produced hGP models in fitting (Sweden).

Model name	Model
	Fitting performance—hGP models for Sweden
hGP Model 1	0.316195914503/(1+E(−((−24.3613307631+((0.248179188862*t))−((+2856.86270612 −0.73365067168*(−7632.23570515−264.609582528*t))*((−188.226187141*(1/(1+E(− (−11.2194594526−0.0269181697779*t))))))))))))
hGP Model 2	0.635230289913/(1+E(−((−859.98155469−13.6682736089*t))/(1+E(−(−0.246865293706 −0.112785107324*t))+(−645.208699231864.791617138+15.8723835928*t))))
hGP Model 3	0.322283774013/(1+E(−(0.0238625468755+0.297883008597*(12.8929141638 −(−134.989713256+2.79343973525*(98.6539018027−0.344395312981*t))/(1+E (−(5.312543151582.46101246618−0.118615909978*t))+5.93885038216)))))
hGP Model 4	0.317301851063/(1+(E(−(−4.86018930519−1.40562027222+2.17969243374*E (−(−225.362926623/(1+E(5.79151748628−0.00873094773031*t))))))))
hGP Model 5	0.317329645302*(1/(1+E((−(−6.28781400917+0.520317506477*(E(−((−22.2379905754 −21.8755007663−(−42.6520973363/(1+E(−(3.89062429403+0.258885588424 −0.00909501063577*t))))))))))))))

Table 24: Representation of the optimized diffusion models in fitting (Sweden).

Fitting performance—optimized diffusion models for Sweden	
Model name	Model
Logistic	0.336282902699/(1+E(−(−2.56246214+0.0518475273171*t)))
Bass	0.336282965867*(1−E(−(2.4303762338e−08+0.0518474673078)*(t−(−231.654837703)))) /(1+(0.0518474673078/2.4303762338e−08)*E(−(2.4303762338e−08+0.0518474673078)* (t−(−231.654837703))))
Gompertz with constant	0.288901526679*E(−8.20983958007*E(−0.0446528448837*t))+0.0500790650845
Gompertz	0.366385197367*E(−3.31412845356*E(−0.0301554112794*t))

Table 25: Representation of the first five produced hGP models in fitting (The Netherlands).

Fitting performance—hGP models for The Netherlands	
Model name	Model
hGP Model 1	0.00328297044203*E(−((−1.62805354054−1.51126326886*E(−0.308998348091 *E(0.00681524096293*t))*E(−32.6517776745*E(−2.63952059485−0.0444857287095*t))*E (E(−0.00300820120766*t))*E(E(−6.16053667129*E(−7.75605043494*E(−1.862862714 −195.916190767*E(−250.192549588*E(−0.0440511725397*t)) −1.01338834582)))))))
hGP Model 2	0.00272274693312*E(−((−1.4875958793−4.39774477765*E(+1.48779357596e−06*E (+0.0712548718888*t))*E(−25.1750484592*E(−2.51499099805−0.0357061452836*t))*E (E(−0.00188093694928*t))*E(E(−59.713187786*E(−697.16935125*120.774228698*E (−0.127948050362*t))*E(−0.0885251390425*t))−2.0188443226))))
hGP Model 3	0.00279956582383*E(−((−1.87785817855−0.691671611195*E(−0.937233488237*E (−0.0154741062254*t))*E(−29.8530474333*E(−2.62003795345−0.0517069348527*t))*E(E (−0.0036068106375*t))*E(E(−20.7938314693*E(−104.667728363*E(30.4510810431 −64.4593053792*E(−3388.8924389*E(−0.174118745631*t))+31.2957260778)))))))
hGP Model 4	0.00279956582383*E(−((−1.87785817855−0.691671611195*E(−0.937233488237*E (−0.0154741062254*t))*E(−29.8530474333*E(−2.62003795345−0.0517069348527*t))*E (E(−0.0036068106375*t))*E(E(−20.7938314693*E(−104.667728363*E(+30.4510810431 −64.4593053792*E(−3388.8924389*E(−0.174118745631*t))+31.2957260778)))))))
hGP Model 5	0.0056229772664*(((E(E(−567628.398566*t))*E(E(−5.98069338075*E(−0.246097777946*t))*E(0.00111530764774*t))+E(E(0.000864502471227*t))*E(E(−18544.326574*E (−0.147581997545*t))*E(−0.0398044163448*t))*E(E(−2.07368114239*E(t))+E ((−3.39589351006*E(−0.0540701741002*t))+1.12064956451)))))

B.

The MAPE is presented in (B.1). The sum is over time period $t = 1, \ldots, T$, $r(t)$ is the raw data actual value for time t, and $y(t)$ is the estimated model's value:

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{r(t) - y(t)}{r(t)} \right|. \tag{B.1}$$

Table 26: Representation of the optimized diffusion models in fitting (The Netherlands).

Fitting performance—optimized diffusion models for The Netherlands

Model name	Model
Logistic	0.383817447797/(1+E(−(−2.99687030519+0.0615659713049*t)))
Bass	0.383817510204*(1−E(−(4.51096518494e−08+0.0615658365913)*(t−(−180.776347814)))) /(1+(0.0615658365913/4.51096518494e−08)*E(−(4.51096518494e−08+0.0615658365913)* (t−(−180.776347814))))
Gompertz with constant	0.362060867532*E(−7.28281633264*E(−0.0455738781315*t))+0.0321086407123
Gompertz	0.408631571889*E(−4.42176509808*E(−0.0373156465688*t))

Table 27: Representation of the first five produced hGP models in fitting (Denmark).

Fitting performance—hGP models for Denmark

Model name	Model
hGP Model 1	5.63426994728*E(−1618550.49488*E(−15.2920199213*E(−3.43879264146*E(−1.28765774838 *E(E(−39343593.2921*E(−19.8648947917*E(((−0.155432332908*(E(−0.740392625374*E(E (−0.43418452549−0.184927854534+0.0142460870633*t)))))))))))))))
hGP Model 2	1.54576267797*E(−65887807.219*E(−17.6521068857*E(((−0.328946437325*(E (−0.937705801568*E(E(−0.300061035521−0.249644860355+0.0126941999487*t)))))))))))
hGP Model 3	1.55049461946*E(−25211043.951*E(−16.6892960686*E(−0.348864738858*E (−0.937562881193*E(E(−0.702455330352*0.778787005609+0.012669874873*t))))))
hGP Model 4	1.55178035993*E(−21975484.4721*E(−16.5513605349*((E(−0.352162700827*(E (−0.938395470029*E(E(−0.295038781436−0.252306036288+0.0126692295308*t)))))))))
hGP Model 5	1.56240433165*E(−4303572.92965*E(−14.9161171585*E(−0.395041661739*E (−0.944193764022*E(E(−0.257102951003−0.289694506506−0.000548574587926 +0.0126461044806*t))))))

Table 28: Representation of the optimized diffusion models in fitting (Denmark).

Fitting performance—optimized diffusion models for Denmark

Model name	Model
Logistic	0.386701712691/(1+E(−(−2.76922996844+0.058481412812*t)))
Bass	0.386701824675*(1−E(−(7.48033939618e−08+0.0584811962598)*(t−(−184.676718196)))) /(1+(0.0584811962598/7.48033939618e−08)*E(−(7.48033939618e−08+0.0584811962598) *(t−(−184.676718196))))
Gompertz with constant	0.364158594712*E(−6.08132991138*E(−0.042851861025*t))+0.0339245270247
Gompertz	0.411115617425*E(−3.86441490696*E(−0.0356437099203*t))

Table 29: Representation of the first five produced hGP models in forecasting (OECD overall).

Forecasting performance—hGP models for OECD overall	
Model name	Model
hGP Model 1	0.266691660828*(1*E(−4.5558998293*E(+0.395188436078*E(+6.76421029076e−07*t))*E(−0.0364443088831*t))+E(−32.929440994*E(−14.9394803737*E(−0.0140586124973*t))*((t))−2.16676076882))
hGP Model 2	0.266691660828*(1*E(−4.5558998293*E(+0.395188436078*E(+6.76421029076e−07*t))*E(−0.0364443088831*t))+E(−32.929440994*E(−14.9394803737*E(−0.0140586124973*t))*((t))−2.16676076882))
hGP Model 3	0.260458088163*(1*E(+20.7470896309*E(−1.45363179622*E(+0.0230415722335*t))−((E(−0.0405561643755*t))*E(+0.68313768074*E(+0.549850481378*E(−6.35620276975e−06*t))+E(+0.68313768074*E(+0.549850481378*((−0.120362874712*t))−20243.8982316))))))
hGP Model 4	0.260458088163*(1*E(+20.7470896309*E(−1.45363179622*E(+0.0230415722335*t))−((E(−0.0405561643755*t))*E(+0.68313768074*E(+0.549850481378*E(−6.35620276975e−06*t))+E(+0.68313768074*E(+0.549850481378*((−0.120362874712*t))−20243.8982316))))))
hGP Model 5	0.503856681046/(1+E(−(−4.93152601976*E(−297363.42521*E(−13.3313258294*E(−0.00150407130185*t))+E(−5.0614088398*E((1−((−16.026189607*t))−284.509095804)))))))

Table 30: Representation of the optimized diffusion models in forecasting (OECD overall).

Forecasting performance—optimized diffusion models for OECD overall	
Model name	Model
Logistic	0.277910816466/(1+E(−(−2.58069789463+0.040512555172*t)))
Bass	0.277911999052*(1−E(−(1.4531815579e−07+0.0405120087548)*(t−(−245.790644296))))/(1+(0.0405120087548/1.4531815579e−07)*E(−(1.4531815579e−07+0.0405120087548)*(t−(−245.790644296))))
Gompertz with constant	0.291419722502*E(−4.90258116522*E(−0.0259253838506*t))+0.0253819182253
Gompertz	0.395094721399*E(−3.35529164695*E(−0.0182708343745*t))

MSE, MAE, and RMSE are presented in (B.2), (B.3), and (B.4), respectively:

$$\text{MSE} = \sum_{t=1}^{T} \frac{[r(t) - y(t)]^2}{T}, \tag{B.2}$$

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^{T} |r(t) - y(t)|, \tag{B.3}$$

$$\text{RMSE} = \sqrt{\sum_{t=1}^{T} \frac{[r(t) - y(t)]^2}{T}} = \sqrt{\text{MSE}}. \tag{B.4}$$

Table 31: Representation of the first five produced hGP models in forecasting (Sweden).

Forecasting performance—hGP models for Sweden

Model name	Model
hGP Model 1	7.20927380763/(1+E(−(−43.9996049107*E(+2.70293611052−(−(+2.803498099 −8.15837562235*E(−1.70915730227*E(−1.74720440508−(−(−1.20189076128 −0.464224975147*E(−1.02239739844−(−(E(+0.0141394755816*t))*E(−5.74593580027*E (−0.254782583675*t)))))))))))))
hGP Model 2	0.310010281135*E(−8.27016729973*E(−1.60830108009*E(−0.00327207897458−(−(E (−0.119609366795*t))*−0.614180508706−10774.1877418*E(−4.50941741619*E((E (−0.00432500059282*t))))))))
hGP Model 3	0.3097209777447*E(−3.32613006656*E(−1828787.41004*E(−21.0309802655*E (−0.00529622245859*t))−(−(−0.380077801407*0.0270286551485*t))*(E(E(E (−251239.306145*E(−13.3031814351*E(−0.00387849535444*t)))))))))
hGP Model 4	0.309960585568*E(−9.58732389121*E(−1.6194568498*E(+0.0756117709535−(−(E(−0.118396422714*t))*−0.553628776298−22454.3177439*E(−4.79688961086*E((E(−0.00396787850031*t))−3.14178858275e−05))))))
hGP Model 5	0.309820654157*E(−14.6146178732*E(−1.72422423849*E(+0.218999757571−(−(E (−0.116856333856*t))*−0.430780829643−509200.484595*E(−5.9633868508*E((E (−0.00290035446778*t))))))))

Table 32: Representation of the optimized diffusion models in forecasting (Sweden).

Forecasting performance—optimized diffusion models for Sweden

Model name	Model
Logistic	0.375347548406/(1+E(−(−2.50124733322+0.0454566554106*t)))
Bass	0.375348107159*(1−E(−(4.05055161391e−08+0.0454564711241)*(t−(−251.439978251)))) /(1+(0.0454564711241/4.05055161391e−08)*E(−(4.05055161391e−08+0.0454564711241) *(t−(−251.439978251))))
Gompertz with constant	0.370148682751*E(−5.03987923004*E(−0.0313473706267*t))+0.0375384813167
Gompertz	0.493581707395*E(−3.13558666165*E(−0.0214071552547*t))

C.

In the fitting process, each chromosome is evaluated with the sum of squared error (SSE), as in

$$\text{SSE} = \sum_{t=1}^{T} \left[r(t) - y(t) \right]^2. \tag{C.1}$$

Table 33: Representation of the first five produced hGP models in forecasting (The Netherlands).

Forecasting performance—hGP models for The Netherlands

Model name	Model
hGP Model 1	0.253326242314*E(−25.4808891372*E(−0.0589928095313*t))*(E(+23.1207979455*E (−0.0648090808052*t))+0.521707030392)
hGP Model 2	0.383595542499*(1−E(−(1.73831772618e−08+0.0617339056621)*(t−(−195.681455628)))) /(1+(0.0617339056621/1.73831772618e−08)*E(−(1.73831772618e−08+0.0617339056621) *(t−(−195.681455628))))
hGP Model 3	0.38359548724/(1+E(−(((−3.00265784135++0.0617339638782*t)))))
hGP Model 4	0.383619520777/(1+E(−(+30.303671841−30.3139537846*E(−6.97443602173e−06*t))*(E(+5.67696636631))))
hGP Model 5	−6.99832857574e−05/(1−1.00018242854*E(((+0.00367050894083*E(−0.0617182268173*t)) −((0.00104546915744*t))*(E(−(−(−642.723570678−0.745098043066*t)))))))

Table 34: Representation of the optimized diffusion models in forecasting (The Netherlands).

Forecasting performance—optimized diffusion models for The Netherlands

Model name	Model
Logistic	0.37898025831/(1+E(−(−3.01882999246+0.062714536225*t)))
Bass	0.378980333001*(1−E(−(2.34739032545e−08+0.0627144548701)*(t−(−187.825594969)))) /(1+(0.0627144548701/2.34739032545e−08)*E(−(2.34739032545e−08+0.0627144548701) *(t−(−187.825594969))))
Gompertz with constant	0.368021712377*E(−6.9150399399*E(−0.0442055310107*t))+0.0308188707983
Gompertz	0.433015517173*E(−4.14875467224*E(−0.0340271367285*t))

In (C.1), the sum is over the time period $t = 1, \ldots, T$. Also, $r(t)$ is the real data for time t, and $y(t)$ is the model's value [15].

In forecasting, the evaluation function corresponds to the weighted sum of squared error (wSSE) function, as in

$$\text{wSSE} = \sum_{t=1}^{T} w_t \left[r(t) - y(t) \right]^2. \tag{C.2}$$

In this function, a weight $w_t = t/T$ is used, in order to focus on the time interval near the last observed or training data.

Table 35: Representation of the first five produced hGP models in forecasting (Denmark).

Forecasting performance—hGP models for Denmark

Model name	Model
hGP Model 1	0.370274249473/(1+E(−(−1.4932887728−1.6616878321−(0.0657816856919 −0.0657829197159)*(t*(t−(−54744.3143422)−(+26.3542094658)*(t−46.1208987279)* (t+25.8109328343−(+47.3153899385−46.1208987279)*(t/(1+(−2.94897421702/ +14.4559028944)*E(−(E(−(E(−(E(−(E(−(−69.3535656648+1.1508868672*t)))))))))))))))))
hGP Model 2	0.367675661705/(1+E(−(−1.54302066592−1.68154541948−(0.0657658786124 −0.0657664779573)*(t*(t−(−56772.6811399)−(−57863.1392956−57863.1735902 −(2053.24155389*(t++9.11367676744−(0.0657743518948−0.06575839253)*(t*(t− (−95622.1907084)−(+3.71296243976)*(t++153.381465941)*(t−69.4331806343))))))))))))
hGP Model 3	0.368792284508/(1+E(−(−1.59162101334−1.73786706106−(0.0657680461973 −0.0657550491287)*(t*(t−9961.43677147−(0.0655352138407−0.0654902811262)*(t*(t− (−(−(−0.417805111612)*(t−87094.1655681+161159.989637)*(t−115.359792771))))))))))
hGP Model 4	0.368796568856/(1+E(−(−1.55630251856−1.77353320735−(0.0654733227916 −0.0655131749375)*(t*(t−(−(−(−0.0997652183839)*(t+10381.0881184+11105.3767878+ +11089.8380866−(0.0661623277964−0.0667280200974)*(t*(t−(−(+4067.51578195+ +3972.08191893)*(t−118.562082811)))))))))))
hGP Model 5	0.368796569031/(1+E(−(−1.59261997915−1.73721570272−(0.0654378132468 −0.0655261015789)*(t*(t−(−(−(−726.874412652)+740.122972564−(0.0656174439567 −0.0654501910722)*(t*(t−(−(148193.4147−(+601.176257868++624.503460915)*(t/(1+ (−4.14155341813/5.62371847802)*E(−(E(−(−26.4655872285 +0.152122244894*t)))))))))))))))

Table 36: Representation of the optimized diffusion models in forecasting (Denmark).

Forecasting performance—optimized diffusion models for Denmark

Model name	Model
Logistic	0.405722295441/(1+E(−(−2.70897871256+0.0546356858504*t)))
Bass	0.412729366859*(1−E(−(0.00143820912913+0.0495453230285)*(t−(−20.4381091217)))) /(1+(0.0495453230285/0.00143820912913)*E(−(0.00143820912913+0.0495453230285) *(t−(−20.4381091217))))
Gompertz with constant	0.433911203761*E(−4.34698049161*E(−0.033071306649*t))+0.0190975994121
Gompertz	0.476280140231*E(−3.53102816452*E(−0.0290174259708*t))

References

[1] N. Meade and T. Islam, "Modelling and forecasting the diffusion of innovation—a 25-year review," *International Journal of Forecasting*, vol. 22, no. 3, pp. 519–545, 2006.

[2] Z. Griliches, "Hybrid corn: an exploration in the economics of technological change," *Econometrica*, vol. 25, no. 4, pp. 501–522, 1957.

[3] E. Mansfield, "Technical change and the rate of imitation," *Econometrica*, vol. 29, pp. 741–766, 1961.

[4] F. M. Bass, "A new product growth for model consumer durables," *Management Science*, vol. 50, no. 12, pp. 1825–1832, 2004.

[5] S. Konstantinos and S. Vasilios, "A new empirical model for short-term forecasting of the broadband penetration: a short research in Greece," *Modelling and Simulation in Engineering*, vol. 2011, Article ID 798960, 10 pages, 2011.

[6] E. M. Rogers, *Diffusion of Innovations*, The Free Press, New York, NY, USA, 5th edition, 2003.

[7] J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, Mich, USA, 1975.

[8] J. R. Koza, "Genetic programming as a means for programming computers by natural selection," *Statistics and Computing*, vol. 4, no. 2, pp. 87–112, 1994.

[9] W. Lee and H. Y. Kim, "Genetic algorithm implementation in Python," in *Proceedings of the 4th Annual ACIS International Conference on Computer and Information Science (ICIS '05)*, pp. 8–12, July 2006.

[10] C. Christodoulos, C. Michalakelis, and D. Varoutas, "On the combination of exponential smoothing and diffusion forecasts: an application to broadband diffusion in the OECD area," *Technological Forecasting and Social Change*, vol. 78, no. 1, pp. 163–170, 2011.

[11] N. Meade and T. Islam, "Forecasting with growth curves: an empirical comparison," *International Journal of Forecasting*, vol. 11, no. 2, pp. 199–215, 1995.

[12] K. Li, Z. Chen, Y. Li, and A. Zhou, "An application of genetic programming to economic forecasting," in *Current Trends in High Performance Computing and Its Applications, Proceedings of the International Conference on High Performance Computing and Applications, Part I*, pp. 71–80, Springer, 2005.

[13] OECD Broadband portal, 2011, http://www.oecd.org/dataoecd/22/12/39574779.xls.

[14] OECD Broadband portal, 2011, http://www.oecd.org/dataoecd/22/13/39574788.xls.

[15] M. A. Kaboudan, "Forecasting with computer-evolved model specifications: a genetic programming application," *Computers and Operations Research*, vol. 30, no. 11, pp. 1661–1681, 2003.