

## Research Article

# Combining Diffusion Models and Macroeconomic Indicators with a Modified Genetic Programming Method: Implementation in Forecasting the Number of Mobile Telecommunications Subscribers in OECD Countries

**Konstantinos Salpasaranis, Vasilios Stylianakis, and Stavros Kotsopoulos**

*Department of Electrical and Computer Engineering, Polytechnic Faculty, University of Patras, Rio Campus, 26504 Patras, Greece*

Correspondence should be addressed to Konstantinos Salpasaranis; [salpk@upatras.gr](mailto:salpk@upatras.gr)

Received 21 September 2013; Revised 20 February 2014; Accepted 28 April 2014; Published 16 June 2014

Academic Editor: Imed Kacem

Copyright © 2014 Konstantinos Salpasaranis et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a modified Genetic Programming method for forecasting the mobile telecommunications subscribers' population. The method constitutes an expansion of the hybrid Genetic Programming (hGP) method improved by the introduction of diffusion models for technological forecasting purposes in the initial population, such as the Logistic, Gompertz, and Bass, as well as the Bi-Logistic and LogInLog. In addition, the aforementioned functions and models expand the function set of hGP. The application of the method in combination with macroeconomic indicators such as Gross Domestic Product per Capita (GDPpC) and Consumer Prices Index (CPI) leads to the creation of forecasting models and scenarios for medium- and long-term level of predictability. The forecasting module of the program has also been improved with the multi-levelled use of the statistical indices as fitness functions and model selection indices. The implementation of the modified-hGP in the datasets of mobile subscribers in the Organisation for Economic Cooperation and Development (OECD) countries shows very satisfactory forecasting performance.

## 1. Introduction

Forecasting is an endogenous process intertwined with the evolution of science. Forecasting methodology is divided into two categories: qualitative and quantitative. Qualitative methods employ the judgment of experts group to produce forecasts [1]. These procedures are mainly applied without using historical data. Quantitative forecasting methods are used when historical data are available as well as the assumption that some of the past patterns will be repeated in the future [2].

There is a variation of quantitative methods such as the time series forecasting which use past trend to forecast the future values of the variable and causal methods that, besides the past trend assumption, also examine the correlation of the variable with other indicators.

The adoption of innovative technologies by a society such as the mobile telecommunications adoption has been discussed and some widely used forecasting models have been

proposed. The diffusion processes as well as the produced models are described in the literature [3–8].

The most commonly used diffusion models are Gompertz, Logistic, and Bass [6] which are dynamic models and follow a sigmoid curve against time. In order to follow the overall diffusion process of the mobile wireless penetration in time, we also employ the Bi-Logistic and LogInLog models which are described in the next section of this paper. The parameters of the models have been estimated by regression analysis with the Least Squares Method [9].

In addition to time response, we investigate the relationship of the produced models with some macroeconomic indicators such as GDPpC and CPI. The core work is an expansion, modification, and implementation of the hybrid Genetic Programming (hGP) method which was presented in [8] in terms of the insertion of new diffusion models as well as the macroeconomic indicators dependence.

The term Genetic Programming (GP) method is a generalization of the Genetic Algorithm (GA) which represents a

heuristic method that employs the Darwinian principle of natural selection in finding an appropriate solution of a well-defined problem and every produced solution corresponds to a new program [10, 11].

The basic structure of the paper follows. Firstly, a brief reference to the GP method and the diffusion models are presented. The hGP technique analysis follows as well as the description of the modifications and expansion on it. The next section analyses the results of the hGP implementation. After that, we discuss the forecasting results, and, finally, the conclusion is presented.

## 2. Genetic Programming Method

GP was introduced by Koza in [11]. In his work, the solution of a problem corresponds to a chromosome-program. The main difference between GP and GA is the representation of solutions. The tree-based representation is adopted by GP method, while a string of numbers represents the solution in GA methodology. The tree-based representation consists of nodes. The nodes represent functions or leaves which correspond to the terminals of the solution, such as variables or constants [12, 13].

The steps for the GP construction are generally the following. Firstly, GP produces an initial population of random programs-solutions composed of the functions and terminals of the problem. The next step iteratively performs the following substeps until a termination criterion will be satisfied: execution of each program and assignment of fitness value according to the precision of each solution [12]. Then, GP generates a new generation of solutions by applying the operations of reproduction, crossover, and mutation. The selection of the candidate solution is performed by probability-based criteria on the fitness value. Reproduction refers to the copy of a solution to the new population. In crossover operation, the selected chromosomes are randomly combined per two and, recombining its chosen parts, generate new chromosomes (offspring) [12]. The mutation changes a function in a chromosome structure with another function. The chromosomes of the new generation have better overall fitness value. The whole process is repeated until a termination criterion is satisfied [9, 11, 12].

## 3. Diffusion Process and Forecasting with Models

Rogers [7] considers that the adoption of an innovative product by a society follows the diffusion process and it has the sigmoid curve format. In this paper, besides the well-known Logistic, Gompertz, and Bass models, we investigate the Bi-Logistic [14, 15] and the LogInLog which is inspired by the solution of the Dodd model in [16, 17].

A diffusion process is described by dynamic or nondynamic models according to whether the level of saturation is changing over time (“carrying capacity”) [14] or constant, respectively. The differential equation which describes the

fundamental diffusion model follows the following formulation:

$$\frac{dy(t)}{dt} = [S - y(t)] \cdot f(t), \quad (1)$$

where  $S$  is the estimated diffusion saturation level for time  $t$  and  $y(t)$  is the diffusion penetration and function  $f(t)$  is the diffusion coefficient.

**3.1. Logistic Diffusion Model.** The Logistic model is the solution of the differential equation (1) which describes the diffusion process. The Logistic model is described by

$$y(t) = \frac{S}{1 + e^{f(t)}}, \quad (2)$$

where  $y(t)$  is the diffusion of a new product in a society, at time  $t$ . Also,  $f(t) = a + b \cdot t$  is a time dependent function and  $a, b$  are constant parameters. The  $S$  constant is the upper limit of the function  $y(t)$ , known as the saturation level. When time  $t \rightarrow \infty$ , then  $y(t) \rightarrow S$  [9].

**3.2. Gompertz Diffusion Models.** The Gompertz model has been extensively used in forecast processes [3, 8, 9]. The Gompertz I format of the equation that this paper proposes is that of

$$y(t) = S \cdot e^{e^{f(t)}}. \quad (3)$$

Also, a variation of (3) format is the following Gompertz II format with constant, in

$$y(t) = S \cdot e^{e^{f(t)} + c}, \quad (4)$$

where, in both formats,  $f(t) = a + b \cdot t$  is a time dependent function and  $a, b, c$  are constant parameters [9].

**3.3. Bass Model.** Bass proposes that the adoption of a new product by a market consists of two major categories: innovators and imitators. The overall diffusion process starts with the innovators adoption of the new product or the innovative technology and then the imitators follow.

The cumulative adoption of the new technology  $y(t)$  for time  $t$  is presented in

$$y(t) = \frac{A - C \cdot e^{-Bt}}{1 + D \cdot e^{-Bt}}. \quad (5)$$

In (5), 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,  $B = p + q$ . Parameter  $C$  is  $C = r \cdot p$ , where  $r$  is a constant and  $D$  parameter is  $D = r \cdot (q/A)$  [6, 9].

**3.4. Bi-Logistic Model.** In some cases, the overall life of a product, like mobile telecommunications, has many phases-generations. For this purpose we employ the Bi-Logistic curve which is the sum of two Logistic curves [14]. So,

$$y(t) = \frac{S_1}{1 + e^{f_1(t)}} + \frac{S_2}{1 + e^{f_2(t)}}, \quad (6)$$

where  $f_1(t) = b_1 + c_1 \cdot (t - t_{m1})$  and  $f_2(t) = b_2 + c_2 \cdot (t - t_{m2})$ . In the first generation, saturation  $S_1$  is constant as well as  $S_2$  of the second generation. Parameters  $b_1, c_1, b_2,$  and  $c_2$  are constants and  $t_{m1}$  and  $t_{m2}$  are the introduction time of the first and second generation, respectively [15, 17].

**3.5. Logistic Growing Saturation Level in Logistic Model (LogInLog).** In this case, the saturation level is time dependent  $S(t)$  and it follows the Logistic diffusion model until the upper saturation of  $S_u$  [15–17]. The whole diffusion process follows the Logistic model [17], as

$$y(t) = \frac{S_u}{1 + c \cdot e^{f(t)} + d \cdot e^{f_u(t)}}, \quad (7)$$

where  $f(t) = b + a \cdot t$  and  $f_u(t) = b_u + a_u \cdot t$ ; parameters  $a, b, a_u, b_u, c, d$  are constants. The saturation level follows in

$$S(t) = \frac{S_u}{1 + g \cdot e^{f_u(t)}}, \quad (8)$$

where  $g$  parameter is a constant.

This model describes the diffusion process when an innovative technology has created generations which are not clearly separated [17]. It should be noted that this model is derived by the generalization of the solution of the Dodd model in [16, 17]. The parameters of the model are optimized by the least square regression.

## 4. Modified Genetic Programming Method

The hybrid Genetic Programming method in fitting and forecasting was presented in a previous work [9]. In this paper, the modified-hGP is presented extensively. The modified hGP implements a strategy which consists of three parts, the nonlinear regression analysis, the genetic algorithm part, and the final model selection. The flowchart of Figure 1 shows the parts of the modified hGP.

**4.1. Initialization of Preparation Steps Parameters.** This stage of the method contains the preparation steps for the program execution process [11, 12, 18]. The first step is the function set  $\Phi$  definition. In the modified hGP, the  $\Phi$  set has two subsets for the arithmetic and mathematical functions,  $\Phi_A$  and  $\Phi_M$ , respectively, [18]. So,  $\Phi = \{\Phi_A, \Phi_M\}$ , where  $\Phi_A = \{+, -, *, /\}$  and  $\Phi_M = \{\exp, \log, \sin, \cos, \text{Logistic}, \text{Gompertz I}, \text{Gompertz II}, \text{Bass}, \text{Bi-Logistic}, \text{LogInLog}\}$ . It should be noted that division  $\{/ \}$  is zero protected for the denominator and  $\{\log\}$  is the natural logarithm.

In the second step, the terminal set  $T = \{M, \Sigma\}$  of the variables  $M$  and constants  $\Sigma$  sets is defined. The variables set  $M = \{t, \text{GDPPC}, \text{PCIn}\}$  and  $\Sigma = \{1, \text{random}(-100, 100)\}$ , where  $t$ , GDPPC, and PCIn are the variables for time, GDP per Capita, and normalized CPI, respectively, and  $\text{random}(-100, 100)$  is the randomly generated constants with domain in  $(-100, 100) \in \mathfrak{R}$ .

The next step is to define the fitness function for each solution. Various statistical indicators can be used for the fitness function during the evaluation process. Following the

previous implementation of the hGP [9], two different fitness functions are used as follows.

In the fitting process, each chromosome is evaluated with the Sum of Squared Error (SSE), as in

$$\text{SSE} = \sum_{t=1}^T [r(t) - y(t)]^2. \quad (9)$$

In (9), the sum is over the time period  $t = 1, \dots, T$ . Also,  $r(t)$  is the real data for time  $t$  and  $y(t)$  is the model's value [9].

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

$$\text{wSSE} = \sum_{t=1}^T w_t [r(t) - y(t)]^2. \quad (10)$$

In this function, a weight  $w_t = t/T$  is used, in order to give greater weight at the time interval near the last training data [9].

Finally, the maximum number of generations is defined to end the execution of the GP.

**4.2. Initial Population.** As mentioned before, the function set of the modified hGP is extended compared to hGP. Apart from the primary arithmetic functions set  $\Phi_A = \{+, -, *, /\}$ , a mathematical functions set has been inserted;  $\Phi_M = \{\exp, \log, \sin, \cos, \text{Logistic}, \text{Gompertz I}, \text{Gompertz II}, \text{Bass}, \text{Bi-Logistic}, \text{LogInLog}\}$ . So, the modified hGP has simplified the chromosomes structure, Figures 2 and 3, while, at the same time, their mathematical efficiency has been improved.

The expressions of the randomly created solutions combine the following primary block format, whereas each part is randomly chosen.

$$\begin{aligned} \text{Block: } & \langle \text{constant} \in \Sigma \rangle \langle \text{variable} \in M \rangle \\ & \langle \text{arithmeticfunction} \in \Phi_A \rangle \langle \text{mathematicalfunction} \in \Phi_M \rangle \\ & \langle (\text{variable} \in M) \rangle. \end{aligned}$$

The solutions of the initial population are the combination of random chosen functions, variables, constants, and primary blocks. Also, the optimized Logistic, Gompertz I, Gompertz II, Bass, Bi-Logistic, and LogInLog diffusion models are being inserted in the population. The parameters of the diffusion models are optimized by nonlinear regression analysis and the Levenberg-Marquardt algorithm has been used [9].

**4.3. Solution Representation.** In modified hGP, each chromosome is a string of characters and corresponds to a program that is a possible solution to the problem [12]. The inner representation of a string of characters is considered as a parse tree using the abstract syntax trees of Python Programming Language. For example, the chromosomes  $0.5/(2 \cdot \text{Log}(1 - 0.2 \cdot t))$  and  $0.25 \cdot \text{GDPPC} \cdot \text{Gompertz I}(5 - 0.2 \cdot t)$  are presented in Figures 2 and 3 as strings and parse trees, respectively [12].

The parse tree consists of nodes. There are two types of nodes, the terminal and nonterminal nodes. The terminal nodes (leaves) of the tree contain the variables or the constants. In contrast, the nonterminal nodes of the tree consist of the modified-hGP functions [9].

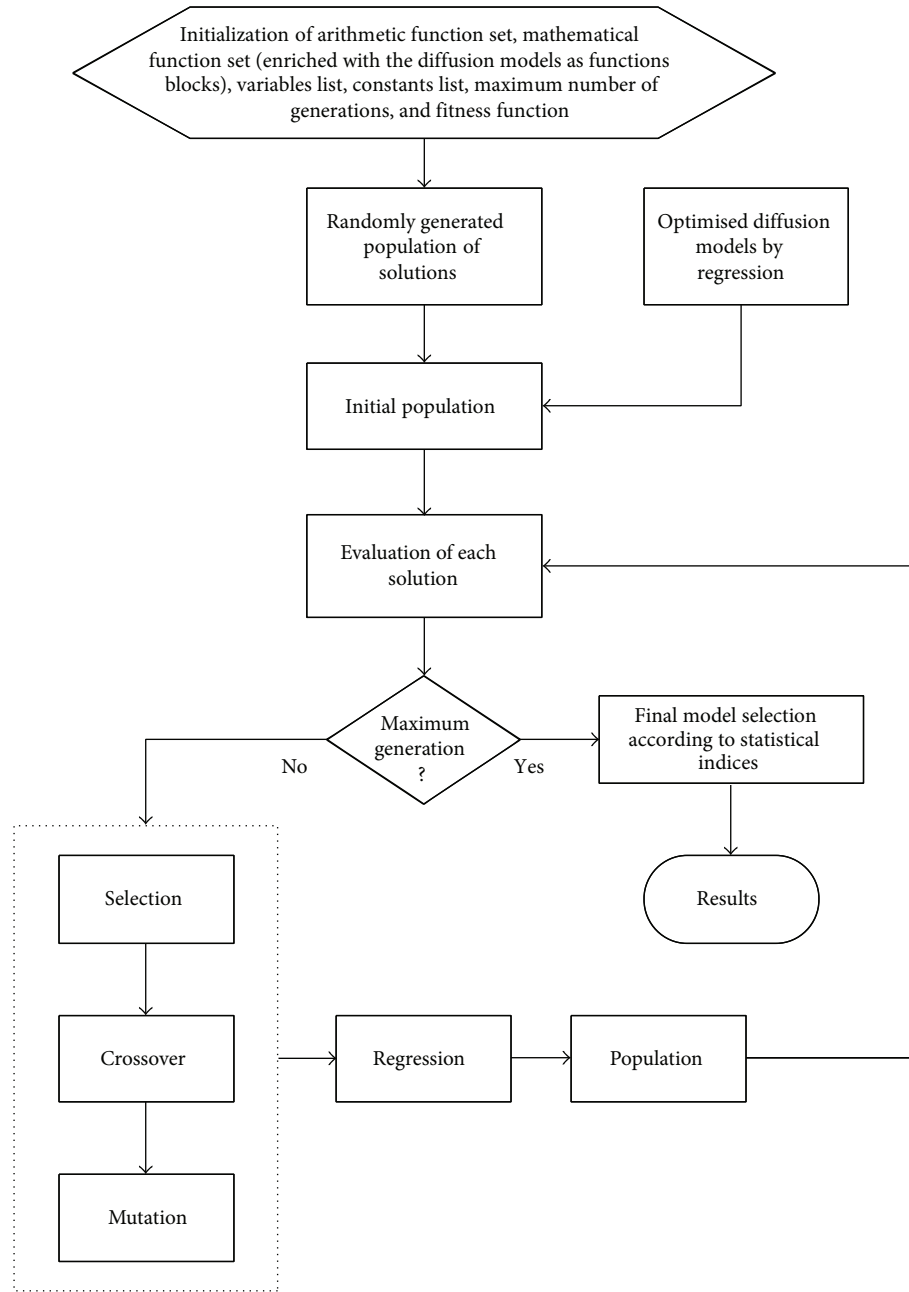


FIGURE 1: Flowchart of the modified hGP.

4.4. *Evaluation of Solutions according to Fitness Function Value.* As stated above, the best solution is selected according to (9) for fitting and (10) for forecasting purposes. The evaluated solutions are inserted into a sorted Python’s list. The solutions that are not satisfying a precision limit criterion are removed. The remaining accepted solutions of the list are sorted according to their fitness value and they are candidates to become parents for the crossover operation or to be chosen in mutation. In Figure 4, the structure of the list is depicted. It should be noted that the problem of the solutions trapping

into local optimum is solved keeping one of all the individuals having the same fitness value in the list.

In tournament selection, a number of solutions from the sorted solutions’ list are selected at random and, then, the best is chosen for the crossover or mutation operation.

4.5. *Crossover.* In the crossover operation, two parents are randomly selected, according to the tournament selection process, from the sorted by the best fitness value solutions’ list.

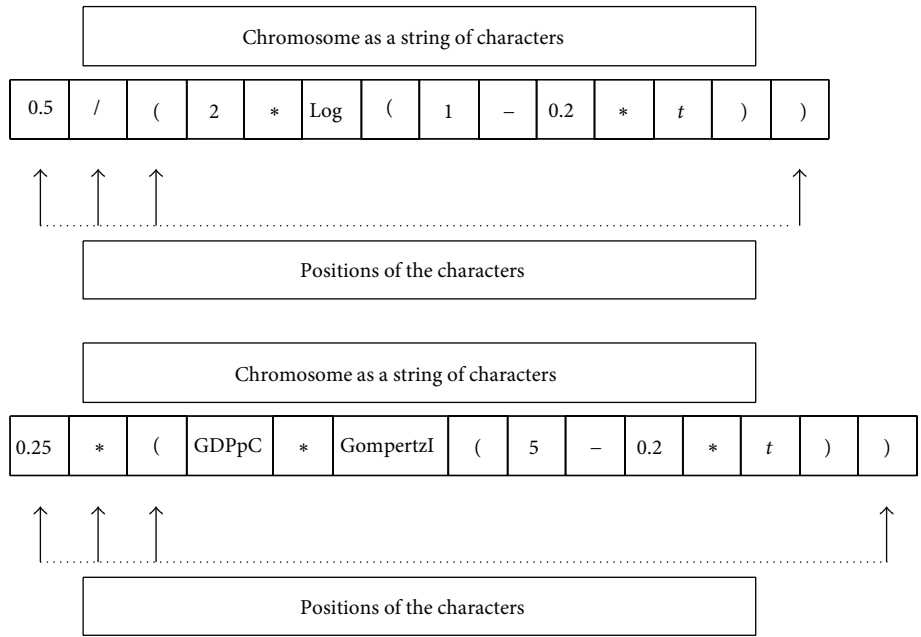


FIGURE 2: Representation of chromosomes in modified hGP as strings.

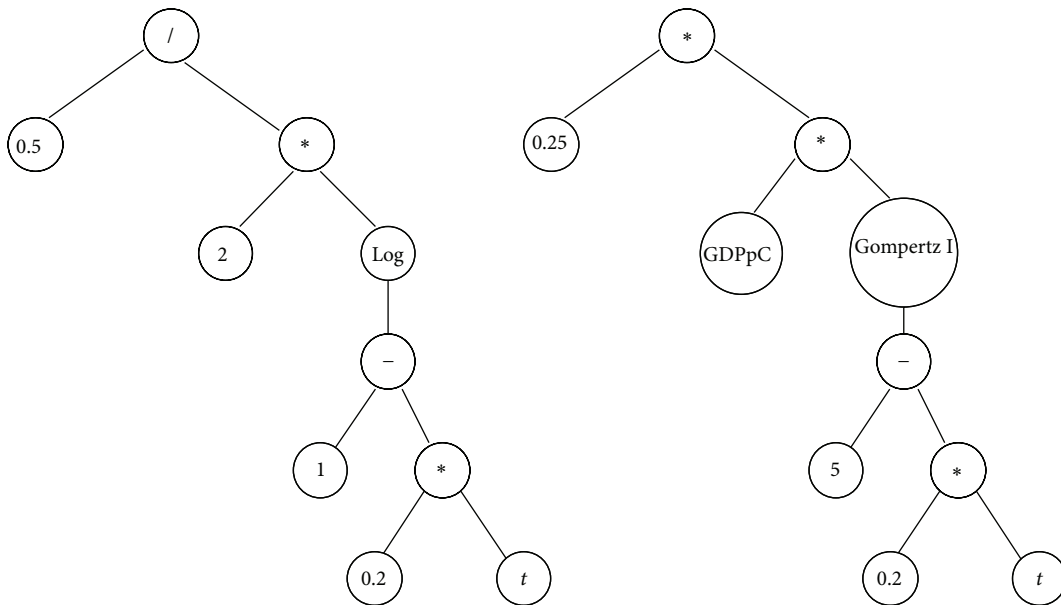


FIGURE 3: Representation of chromosomes in modified hGP as parse trees.

In each parent solution, a crossover point is randomly chosen. The substring of each parent beginning at the crossover point is interchanged between two parents' solutions and the children (offspring) are generated. The crossover operation is presented in Figures 5 and 6.

4.6. *Mutation.* In the mutation process, a solution is chosen by tournament selection from the tournament list. Once again, a string's point, which depicts a function, is randomly chosen. The mutation replaces the chosen function from the  $\Phi = \Phi_A \cup \Phi_M = \{+, -, *, /, \exp, \log, \sin, \cos, \text{Logistic},$

$\text{Gompertz I}, \text{Gompertz II}, \text{Bass}, \text{Bi-Logistic}, \text{LogInLog}\}$  set, with a new random function in the solution.

The mutation operation is presented in Figures 7 and 8.

4.7. *Fitness Function and Final Model Selection: Statistic Indicators.* The fitness function of each individual in the modified hGP method is the Sum of Squared Error (SSE) for the fitting process, as in (9), and the Weighted Sum of Squared Error (wSSE), as (10) presents.

The statistical indices in the modified hGP [9] are the Mean Absolute Percentage Error (MAPE), the Mean Square

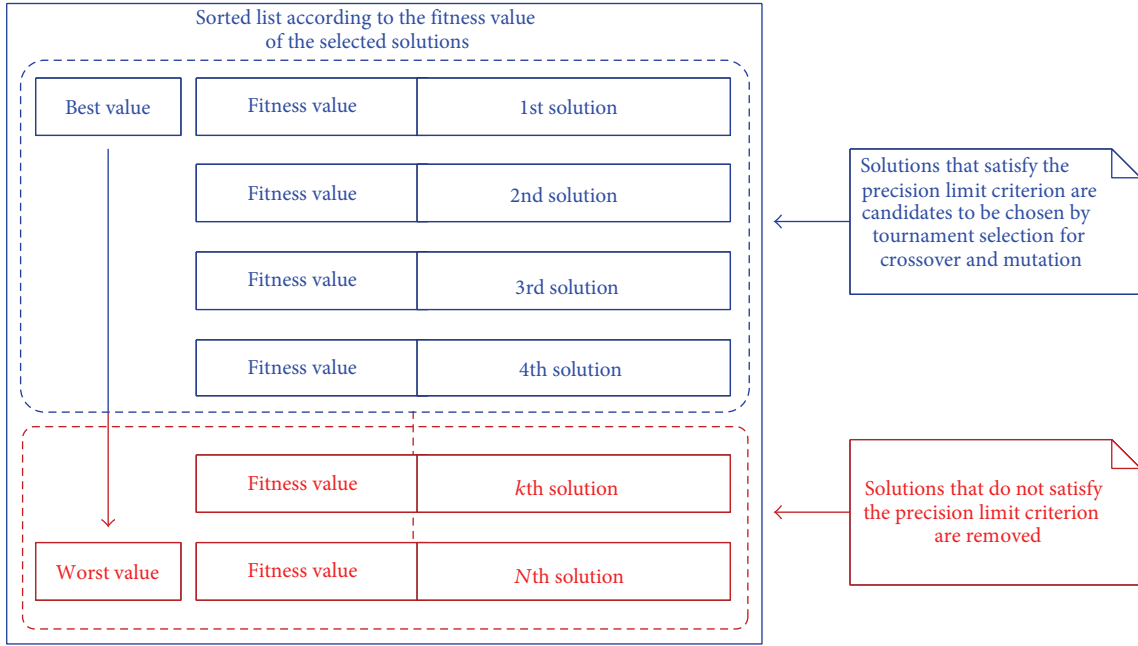


FIGURE 4: Representation of the solutions' sorted list in the selection operation of the modified-hGP.

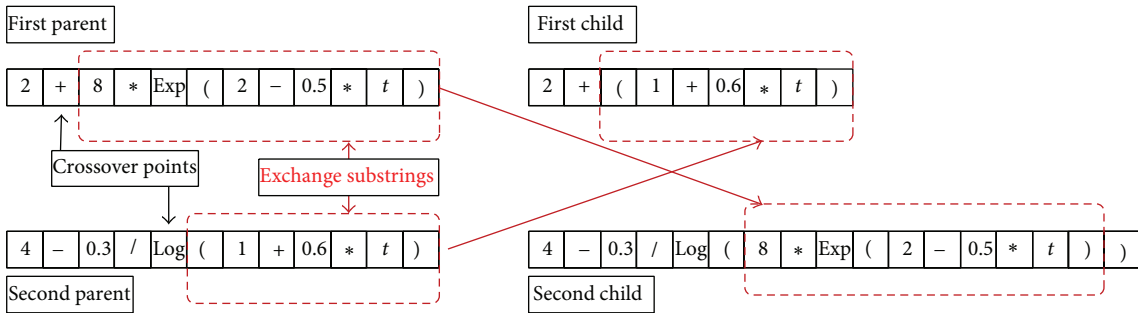


FIGURE 5: Crossover of the modified-hybrid Genetic Programming method (string representation).

Error (MSE), the Mean Absolute Error (MAE), and the Root Mean Square Error (RMSE).

MAPE is presented in (11). The  $r(t)$  is the raw data values,  $y(t)$  corresponds to the predicted value, and  $T$  is the number of the data points:

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \left| \frac{r(t) - y(t)}{r(t)} \right|. \quad (11)$$

MSE, MAE, and RMSE are presented in (12), (13), and (14), respectively:

$$\text{MSE} = \sum_{t=1}^T \frac{[r(t) - y(t)]^2}{T}, \quad (12)$$

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |r(t) - y(t)|, \quad (13)$$

$$\text{RMSE} = \sqrt{\sum_{t=1}^T \frac{[r(t) - y(t)]^2}{T}} = \sqrt{\text{MSE}}. \quad (14)$$

In addition, this study has deployed a Bayesian's Information Criterion (BIC) inspired format [19] in the final model selection for the forecasting, as the following depicts:

$$\text{BIC} = 2 \cdot \log\left(\frac{F}{N}\right) + p \cdot \log(N). \quad (15)$$

In (15),  $F$ ,  $p$ , and  $N$  correspond to Fitness Function Value (wSSE for forecasting), parameters of the model, and the number of data points, respectively.

It should be noted that in the final selection process of the appropriate forecasting model we use the half of dataset before the last observed data point.

## 5. Macroeconomic Indicators

In this section, the macroeconomic indicators of Gross Domestic Product per Capita (GDPpC) and normalized

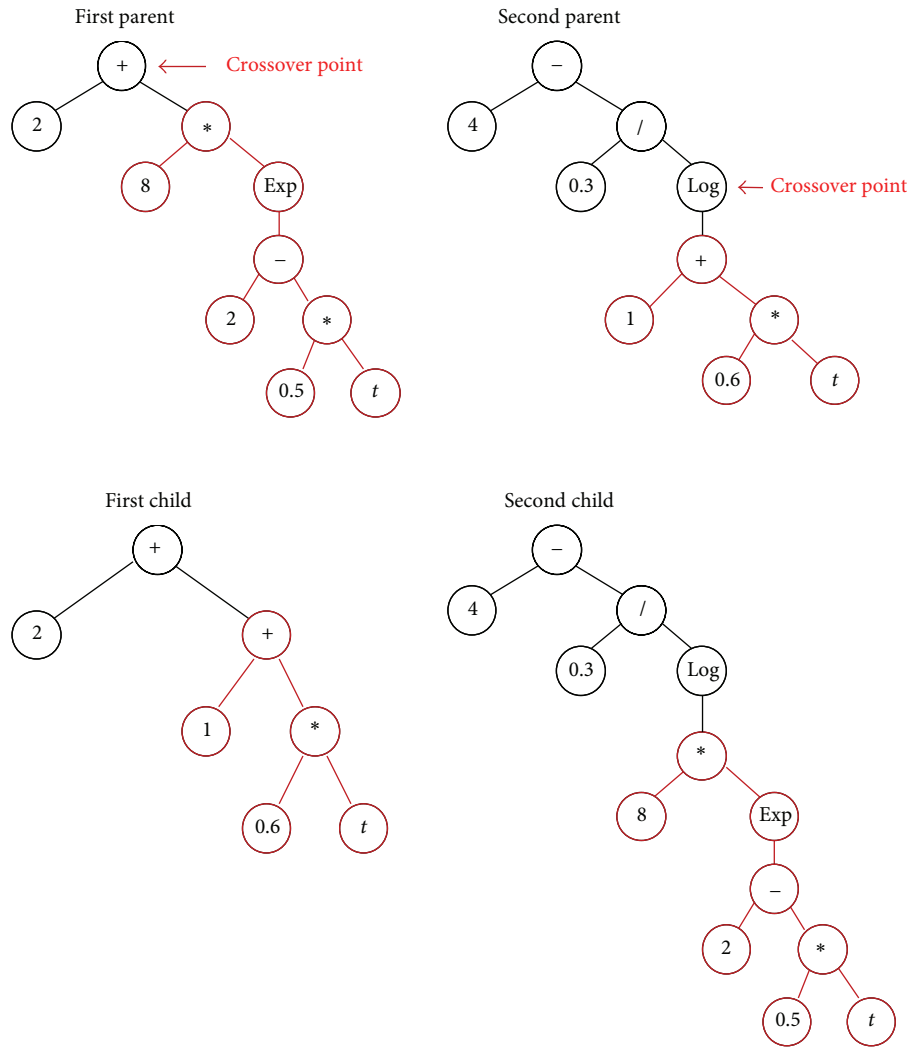


FIGURE 6: Crossover of the modified-hybrid Genetic Programming method (parse tree representation).

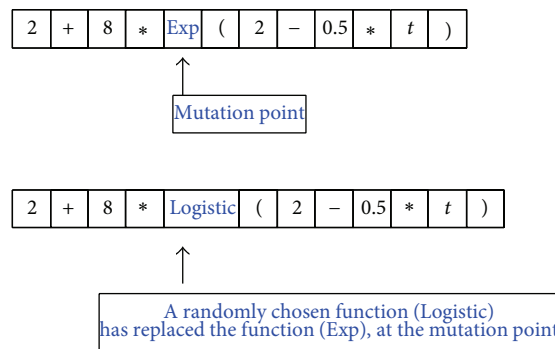


FIGURE 7: Mutation of the modified-hybrid Genetic Programming method (string representation).

Consumer Prices Index (CPI) will be presented. The GDPpC is a macroeconomic index for the productivity of a country and it could not be considered as index of personal income.

According to [20], the basic index of the value of the goods and services produced by a country is the Gross

Domestic Product (GDP). The GDPpC indicates the living standards of the economy in a country.

In general, CPI indicates a weighted average of basic consumer goods prices. Moreover, in this study, the CPI relies on the individual consumption expenditure of households, less

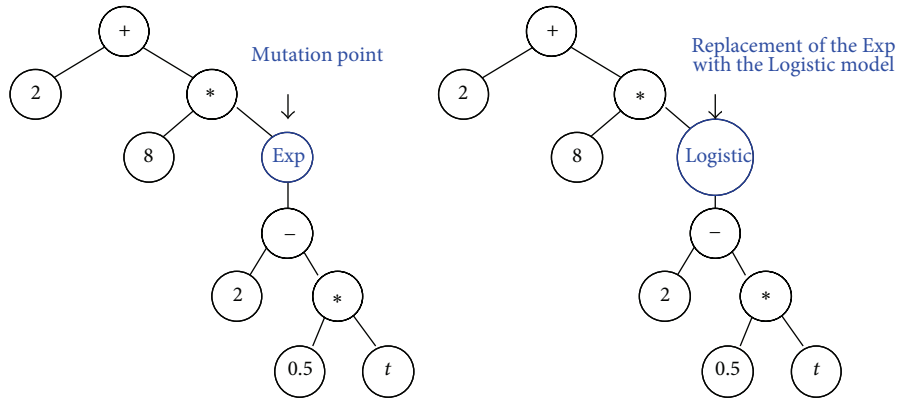


FIGURE 8: Mutation of the modified-hybrid Genetic Programming method (parse tree representation).

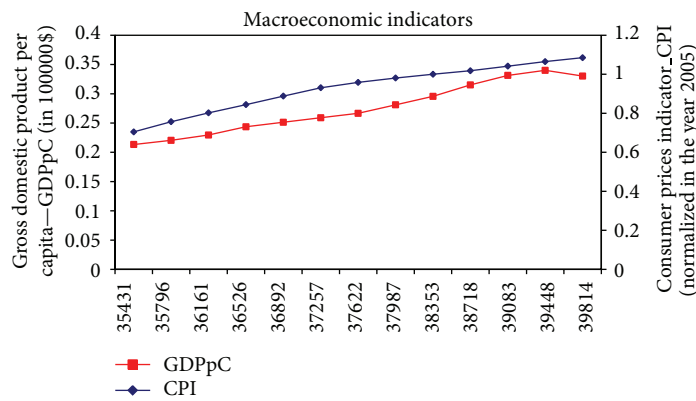


FIGURE 9: Yearly growth of the GDPpC (in 100000\$) and CPI (normalized on the year 2005).

energy, and food consumption [20]. It should be noted that CPI is normalized on the CPI of the year 2005. In Figure 9, the yearly GDPpC and CPI for the time period between the years 1997 and 2009 are presented.

It should be noted that, after the year 2008 (“economic crisis year”), the OECD’s GDPpC is decreased for 2009, but, on the other hand, the CPI has bigger tolerance.

### 6. Mobile Telecommunications Growth: A Brief History

According to [21], the first mobile telecommunications were introduced with analogue networks in the early 1980s, for voice transmission. The second generation (2G) mobile network (Global System for Mobile Communication, GSM) followed in the early 1990s and digital mobile networks were born with the first SMS service. In the late 1990s, enhanced digital generation (2.5G) was introduced for data services. The data services were changed from circuit switched transport (GSM) to packet data transport with General Packet Radio Services (GPRS) and, later, data rates grew with enhanced digital technologies such as Enhanced Data rates for GSM Evolution (EDGE).

Also, in 2003, the next generation (3G) of mobile networks, Universal Mobile Telecommunications System

(UMTS), emerged with the first video-calls and, later, (around 2006) was upgraded to High Speed Packet Access (HSPA) with data rates of 14 Mbps in the downlink and 5.76 Mbps in the uplink. Then, HSPA was upgraded to HSPA+ with theoretically 168 Mbps and 22 Mbps for downlink and uplink, respectively, and data services as videos, mobile email, and music. In 2009, Long-Term Evolution (LTE) was launched for commercial usage, while a new generation (4G) of technology is coming [21]. Figure 10 depicts the evolution of mobile technologies generations in parallel with the overall OECD mobile subscribers, contract, prepaid, and 3G subscribers, from the year 1997 to 2009 [22].

It should be noted that the number of the mobile subscribers is growing through the technology generations evolution.

### 7. Results of the Method

The results will be analysed in order to provide a satisfactory prediction for mobile subscribers which consist of mobile contract subscribers and mobile prepaid subscribers in OECD countries, as well as mobile 3G subscribers.

This study investigates the implementation of modified hGP on four different datasets. The datasets present the total yearly number of OECD mobile subscribers, the yearly



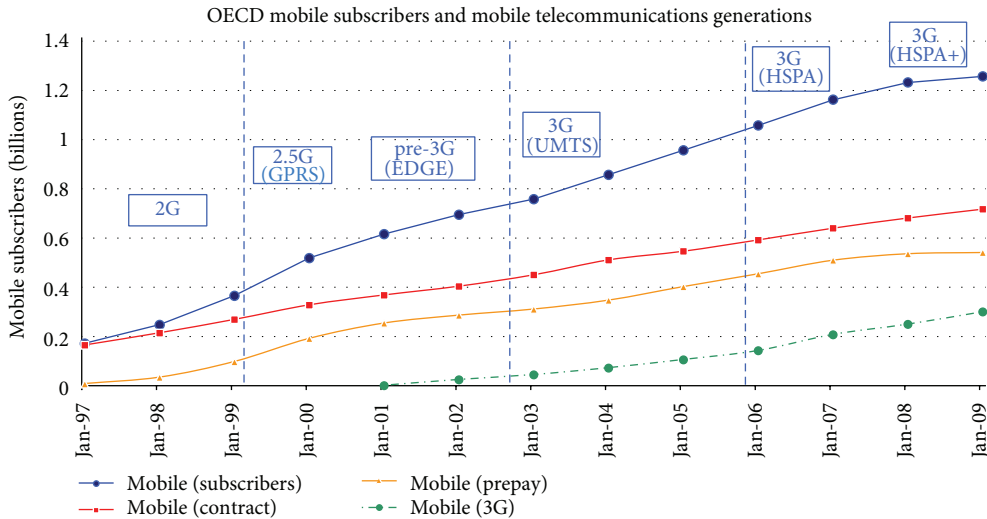


FIGURE 10: OECD mobile subscribers (in billions).

TABLE 1: Initialization parameters of the modified-hGP.

Initialization parameters of the modified-hGP	
Maximum number of generations	500
Evaluation function	SSE
Upper limit of the precision for the candidates for crossover and mutation	0.5

number of mobile contract subscribers, the yearly number of mobile prepay subscribers, and finally the yearly number of mobile 3G subscribers. The observation period begins from the year 1997 to 2009, which is comprised of 13 data points.

7.1. *Fitting Results for the OECD Mobile Telecommunications Subscribers.* Table 1 contains the initialization parameters for the execution of the modified-hGP concerning the data sets in OECD countries.

The fitting performance of the first modified-hGP model for the total number of OECD subscribers, according to its fitness value (SSE), is presented in Figure 11. Also, Figure 12 presents the errors of the models in time (residuals). The relative statistical indices SSE, MAPE, MSE, RMSE, and MAE of the modified-hGP models are presented in Table 2.

The corresponding modified-hGP model has the following format:

$$(-2.89723999239) * (((((-((-0.411428819241 * \text{Exp}(\text{GDPpC}) - 0.439389285344) * (\text{CPI}) - ((-0.296205298307 * \text{GDPpC}) * (\text{Exp}((-0.199818005671) * t) - 0.439389285344) * (\text{CPI}) - (-0.0777146764017) * (\text{CPI}) - ((-0.411428819241 * \text{Exp}(\text{GDPpC}) - 0.439389285344) * (\text{CPI}) - 0.199968576899 * \text{GDPpC}) * ((\text{CPI}) - ((((-46.45853141) * \text{Exp}(\text{GDPpC}) * ((\text{Exp}((-4.64087434687) * \text{Exp}(\text{GDPpC}) * \text{Exp}((-0.0630943612445) * t) * \text{Exp}(\text{GDPpC}) * 0.411428819241 * \text{Exp}(\text{GDPpC}) - 0.439389285344) * (\text{CPI}) - 0.199968576899 * \text{GDPpC}) * ((\text{CPI}) -$$

Fitting performance-OECD mobile subscribers

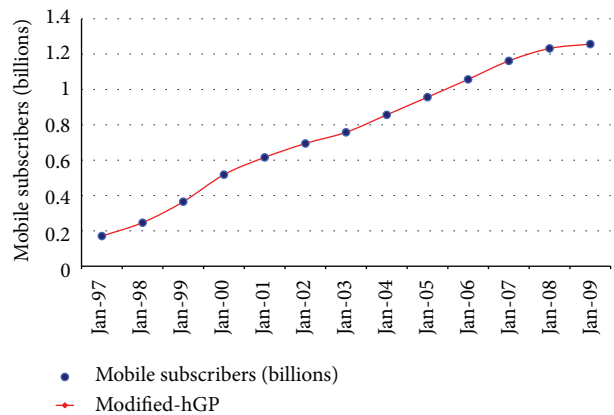


FIGURE 11: The fitting performance of the modified-hGP model for the total number of OECD subscribers.

Modified-hGP model fitting performance OECD mobile subscribers-residuals

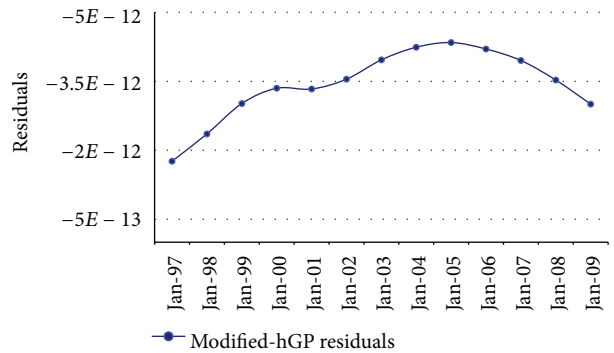


FIGURE 12: Residuals of the modified-hGP model in fitting of the total number of OECD subscribers.

TABLE 2: Statistical indices in fitting process of the modified-hGP model for the total number of OECD subscribers.

Model name	Statistical indices of the modified-hGP for OECD mobile subscribers				
	MAPE	SSE	MSE	RMSE	MAE
Modified-hGP Model	5.5719E - 12	1.59872E - 22	1.22978E - 23	3.50682E - 12	3.42953E - 12

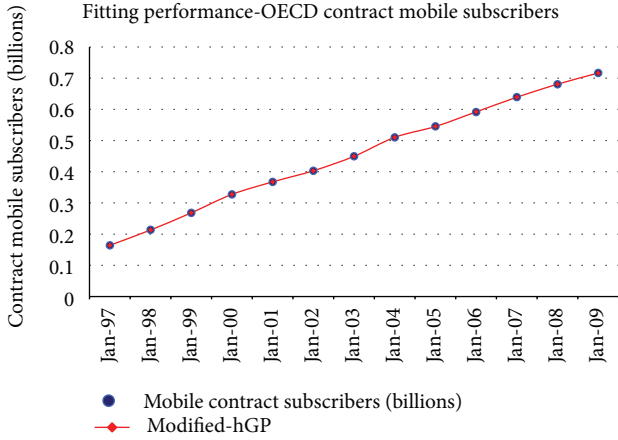


FIGURE 13: The fitting performance of the modified-hGP model for the OECD mobile contract subscribers.

$$(((((-46.45853141) * \text{Exp}(\text{GDPpC}) * ((\text{Exp}((-4.64087434687) * \text{Exp}(\text{GDPpC}) * \text{Exp}((-0.0630943612445) * t) * \text{Exp}(\text{GDPpC}) * \text{Exp}((-0.0630943612445) * t) + (-1.3303614878) * \text{Exp}(\text{CPI}) * \text{Exp}((-5.90641430423) * ((\text{Exp}((-1 * ((.86811449329) * t) * 0.08679650977 * t))))))))))))))))))$$

As one can see, this method combines different variables like GDPpC or CPI with the independent variable of time. In Table 2, the modified-hGP method achieves excellent statistical performance, showing an SSE value of 1.59872E - 22.

The fitting performance and the residuals for the remaining data sets are presented in Figures 13 and 14 for contract subscribers, Figures 14 and 15 for prepay subscribers, and Figures 16 and 17 for 3G users, respectively. The relative statistical indices SSE, MAPE, MSE, RMSE, and MAE of the produced modified-hGP models are presented in Table 3 for contract, Table 4 for prepay, and Table 5 for 3G subscribers.

The corresponding modified-hGP model for contract subscribers has the following format:

$$1.10988090241 * \text{Logistic}(t) + (-0.0870458635523 - (0.830551978552 * \text{GDPpC}) * (-0.0821332034235 * \text{GompertzI}(t) - 1.62717732509 * \text{GDPpC} * \text{Bass}(-441.621630873 * (\text{GompertzI}(t) + 1.68026476957 - 1.62717732509 * \text{GDPpC} * \text{Bass}(97.6225484069 * \text{GDPpC})) * (9.73983478741 - 14.8619093427 * \text{GDPpC} - 0.206004836329 * \text{GDPpC} * \text{GDPpC} - 0.206004836329 * \text{GDPpC} * \cos(-35.0063961093 * \text{CPI}))))))$$

It should be noted that this method combines different variables like GDPpC or CPI with the independent variable of time and a variation of diffusion models' blocks. The performance of the model corresponds to a good enough behavior

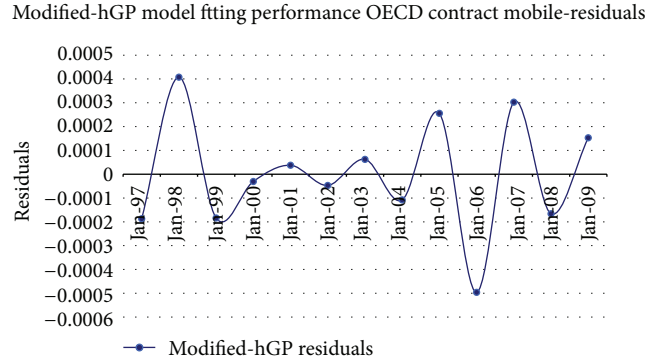


FIGURE 14: Residuals of the modified-hGP model in fitting for the OECD mobile contract subscribers.

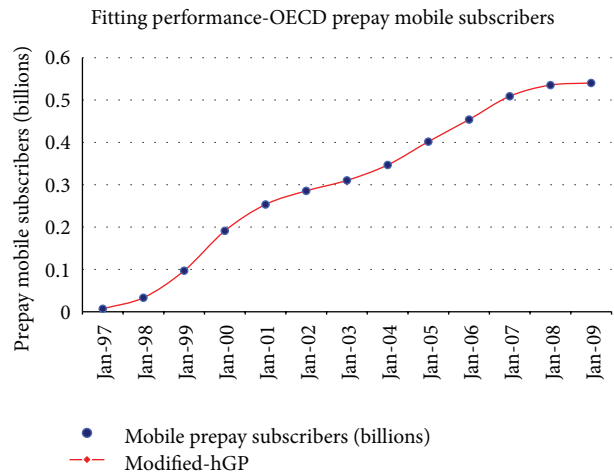


FIGURE 15: The fitting performance of the modified-hGP model for the OECD mobile prepay subscribers.

in fitting process. The error performance in fitting is depicted in Figure 14.

The corresponding modified-hGP model for prepay subscribers has the following format:

$$0.81582585124 * \text{GompertzI}(t-1) - 0.184310132872 + 0.416581273741 * \text{GDPpC} * \text{GDPpC} * \text{Bass} (+202.832908656 * \text{GDPpC}) * (+32.2627880991 * \text{GompertzI}(t-1) - 1.6566435914 * \text{GDPpC} * \text{Bass}(-61.9029092844 * \text{GDPpC}) * (-52.5926318506 - 186.401025667 * \text{GDPpC} - 416.890907258 * \text{GDPpC} + 270.47302209 * \text{GompertzI}(t-1) + 112.258418744 - 1.6566435914 * \text{GDPpC} * \text{Bass}(-61.9029092844 * \text{GDPpC}) * (-52.5926318506 - 186.401025667 * \text{GDPpC} + 112.258418744 - 1.6566435914 * \text{GDPpC} * \text{Bass}(-61.9029092844 * \text{GDPpC}) * (-52.5926318506$$

TABLE 3: Statistical indices in fitting process of the modified-hGP model for the OECD mobile contract subscribers.

Statistical indices of the modified-hGP for OECD mobile contract subscribers					
Model name	MAPE	SSE	MSE	RMSE	MAE
Modified-hGP model	0.00051034	7.08765E - 07	5.45204E - 08	0.0002335	0.000187453

TABLE 4: Statistical indices in fitting process of the modified-hGP model for the OECD mobile prepay subscribers.

Statistical indices of the modified-hGP for OECD mobile prepay subscribers					
Model name	MAPE	SSE	MSE	RMSE	MAE
Modified-hGP model	0.000405	9.00192E - 09	6.92455E - 10	2.63145E - 05	2.0048E - 05

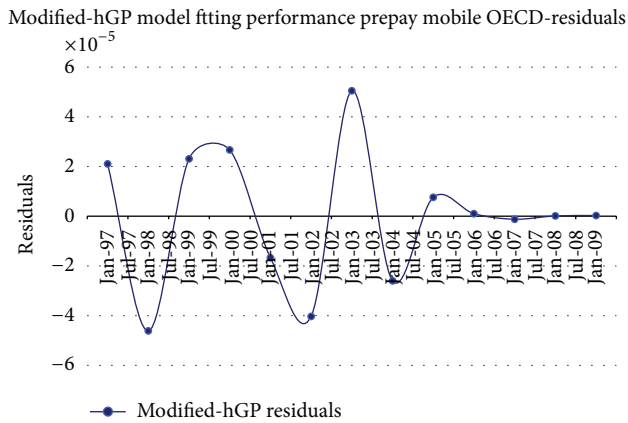


FIGURE 16: Residuals of the modified-hGP model in fitting for the OECD mobile prepay subscribers.

$$\begin{aligned}
 & -186.401025667 * \text{GDPpC} * \text{Bass}(-61.9029092844 * \\
 & \text{GDPpC}) * (-52.5926318506 - 186.401025667 * \text{GDPpC} + \\
 & 112.258418744 - 1.6566435914 * \text{GDPpC} * \text{Bass}(- \\
 & 61.9029092844 * \text{GDPpC}) * (-52.5926318506 \\
 & - 186.401025667 * \text{GDPpC} - 416.890907258 * \text{GDPpC} * \\
 & \cos(-21.0683173917 * \text{CPI}))))))
 \end{aligned}$$

Once again, model yields a good enough behavior in fitting process. The error performance in fitting is depicted in Figure 16

Finally, the corresponding modified-hGP model for 3G subscribers has the following format:

$$\begin{aligned}
 & -0.392941862339 * \text{GDPpC} - (-0.364795898833 - (- \\
 & 0.567322689601 * \text{CPI} + 0.0261425792836 * t \\
 & + 0.0241257362461 * \text{CPI} * t * \cos(-8.33823945644 * \\
 & (\cos(-111.042952295 * \text{CPI}) * \text{GompertzII}(\cos(- \\
 & 10978.8993729 / \log(28717123954.6 * \text{GDPpC}))))))
 \end{aligned}$$

Finally, the modified-hGP model yields a satisfactory fitting performance. The error performance in fitting is depicted in Figure 18.

7.2. Forecasting Results for the OECD Mobile Telecommunications Subscribers. The forecasting results of the generated models by the modified-hGP method are presented in this section, as well as the combined diffusion models with the modified-hGP models. As mentioned before, the statistic indicator wSSE has been used for the forecasting process.

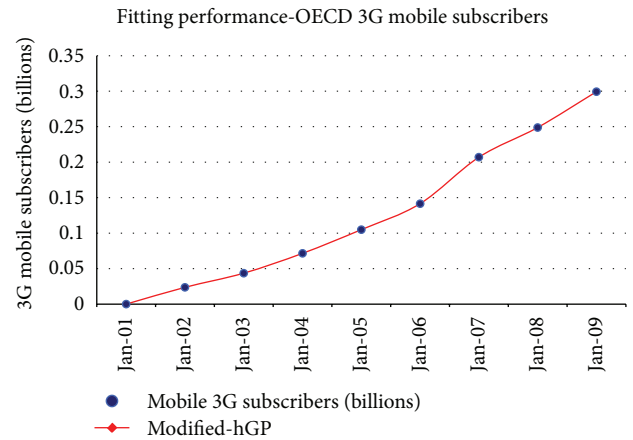


FIGURE 17: The fitting performance of the modified-hGP model for the OECD 3G mobile subscribers.

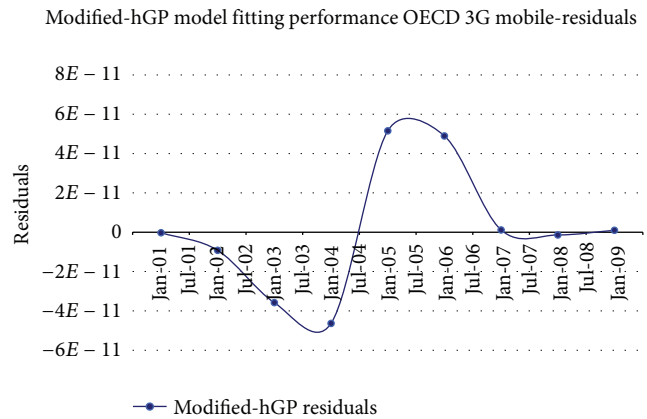


FIGURE 18: Residuals of the modified-hGP model in fitting for the OECD 3G mobile subscribers.

The initialization parameters for the execution of hGP are presented in Table 6, for the forecasting process. The forecasting method for a 2-year prediction uses 11 data points as training set of the GP method, except for the 3G training set which consists of 7 points. The forecasting performance of the modified-hGP models concerning total OECD mobile subscribers, contract, prepay, and 3G is depicted in Figures 19,

TABLE 5: Statistical indices in fitting process of the modified-hGP model for the OECD 3G mobile subscribers.

Statistical indices of the modified-hGP for OECD 3G mobile subscribers					
Model name	MAPE	SSE	MSE	RMSE	MAE
Modified-hGP model	6.05835E - 10	8.57305E - 21	9.52562E - 22	3.08636E - 11	2.17376E - 11

TABLE 6: Initialization parameters of modified-hGP.

Initialization parameters of the modified-hGP	
Maximum number of generations	500
Evaluation function	wSSE
Upper limit of the precision for the candidates for crossover and mutation	2

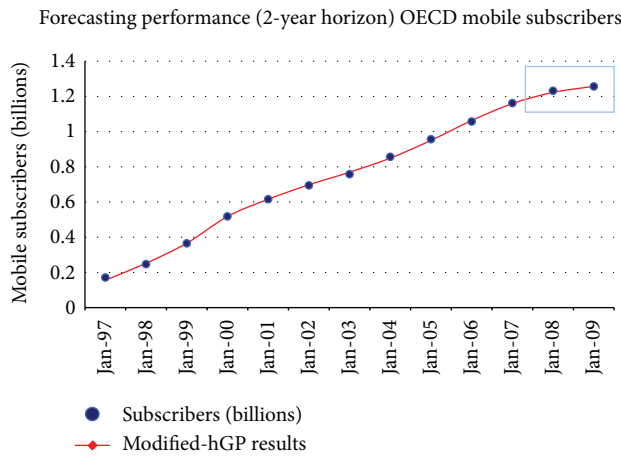


FIGURE 19: The forecasting performance of modified-hGP (forecast period—2 years ahead).

22, 24, and 25, respectively. In every graph, the forecast period window is presented in the blue rectangle.

The corresponding modified-hGP model has the following format, which has a Bi-Logistic behavior and it is time dependent:

$$\text{Bi-Logistic}(t)/(1+\text{EXP}(-9859.00409322-165.278808156 * t)+(165.278808156/(165.278808156+37407.349151/(1+(-172320995104/(-172320995104+49535561996.5 * t))))))$$

The forecasting performance of the optimized diffusion models, according to their fitness value (wSSE) for the 11 training points, is presented in Figure 20. Also, the relative statistical indices, concerning the whole dataset, of the produced forecasting modified-hGP and diffusion models are presented in Table 7.

Considering Table 7, it can be concluded that the modified-hGP method achieves good statistical indices combining some optimized diffusion models. We can see that the first hGP model achieves a wSSE value of 0.000226, while the best of diffusion models, Bi-Logistic, has a similar 0.000281. It should be noted that modified-hGP model residuals against time (data points), especially for the 2 last data points (the

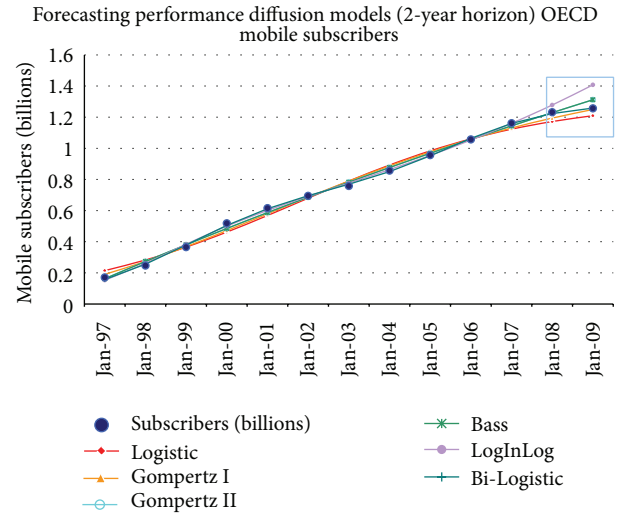


FIGURE 20: The forecasting performance of the diffusion models (forecast window—2 years ahead) for the total number of OECD subscribers.

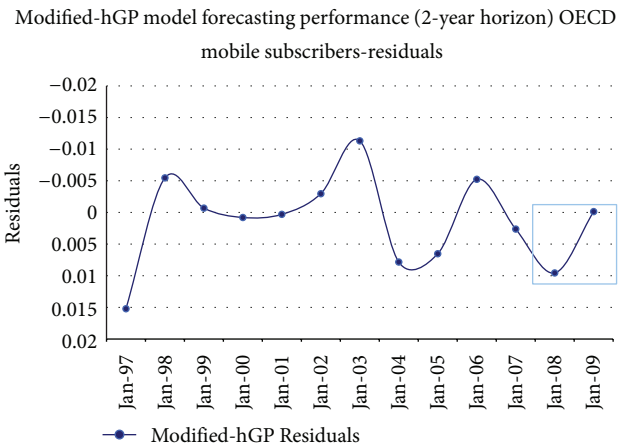


FIGURE 21: Residuals of the modified-hGP model in forecasting (2 years window period) of the total number of OECD subscribers.

forecast period), show the error response of the GP model (see Figure 21).

The modified-hGP model's performance, concerning OECD mobile contract forecasting, is depicted in Figure 22. The corresponding modified-hGP model has the following format, which is time and GDPpC dependent:

$$0.043436247449*(t-(-4.73270273612))+ (0.0230982341541/(1+(983648.677808/(983648.677808-316094351.68))^t+(1.67610337084/$$

TABLE 7: Statistical indices in forecasting process of the modified-hGP and diffusion models for the total number of OECD subscribers.

Statistical indices of the modified-hGP and diffusion models in forecasting the total number of OECD mobile subscribers					
Model name	MAPE	wSSE	MSE	RMSE	MAE
Modified-hGP	0.01271669	0.000226	$4.84E - 05$	0.006957	0.005281
Logistic	0.064808	0.009513	0.00148	0.038466	0.034057
Gompertz I	0.04259	0.004252	0.000732	0.027062	0.023146
Gompertz II	0.034539	0.004214	0.000574	0.023967	0.019402
Bass	0.034737	0.004238	0.000575	0.023983	0.019426
LogInLog	0.036436	0.023129	0.002048	0.045259	0.024329
Bi-Logistic	0.018125	0.000281	$7.68E - 05$	0.008764	0.007522

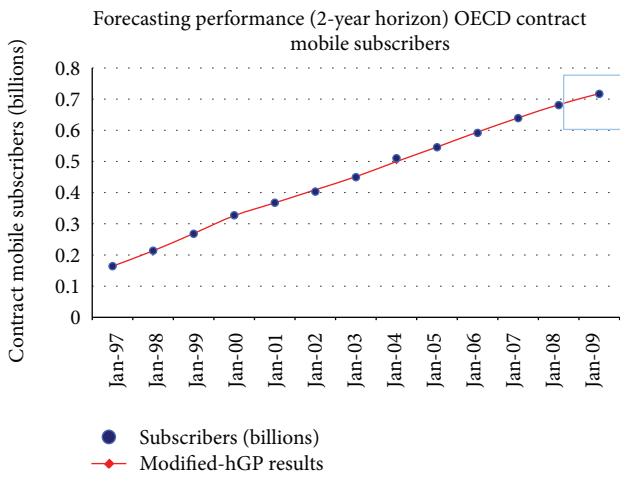


FIGURE 22: The forecasting performance of the modified-hGP (forecast window—2 years ahead) for the number of OECD mobile contract subscribers.

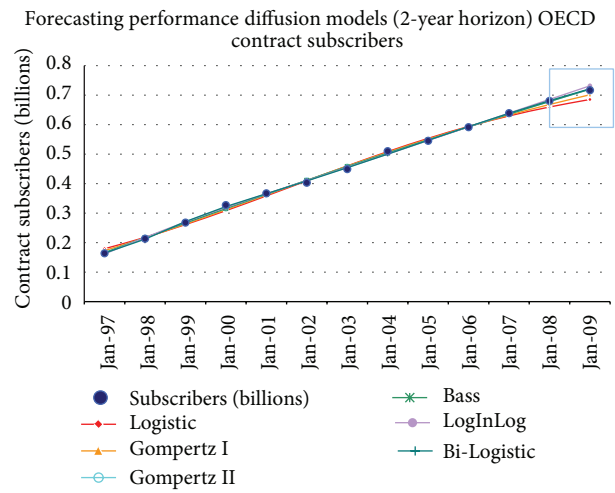


FIGURE 23: The forecasting performance of the diffusion models (forecast window—2 years ahead) for the number of OECD mobile contract subscribers.

$$(1.67610337084 - 2.03073368583) * GDPpC * (51 - 51.061701415 - 0.0249570883638 * t + 0.127734108836 * t^2)$$

The forecasting performance of the diffusion models is presented in Figure 23. Also, the statistical indices of the produced forecasting modified-hGP and diffusion models are presented in Table 8.

From Table 8, one can conclude that the modified-hGP method as well as diffusion models achieves good statistical indices. We can see that the hGP model achieves a wSSE value of  $8.58E - 05$  and the Bi-Logistic 0.000116. Once again, the modified-hGP model residuals against time (data points), especially for the 2 last data points (the forecast period), show the error response of the GP model (see Figure 24).

The modified-hGP model's performance concerning OECD mobile prepay forecasting is depicted in Figure 25. The corresponding modified-hGP model has the following format, which is time, GDPpC, and CPI dependent:

$$(0.20718880307 * ((t/CPI)/CPI) - 0.676908861729) * GDPpC/CPI + 0.052157792812$$

The forecasting performance of the diffusion models is presented in Figure 26. Also, the statistical indices of the

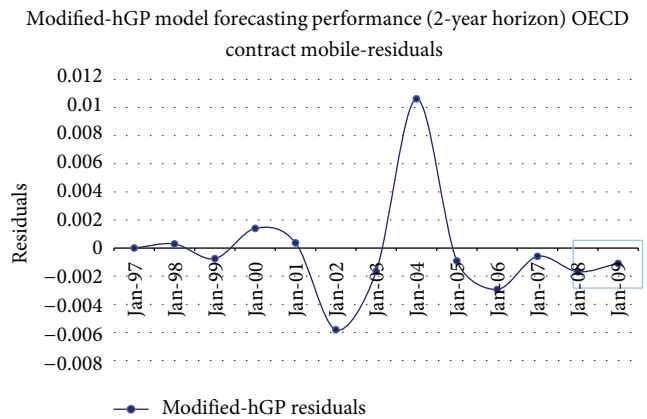


FIGURE 24: Residuals of the modified-hGP model in forecasting (2 years window period) of the number of OECD mobile contract subscribers.

produced forecasting modified-hGP and diffusion models are presented in Table 9.

Table 9 shows that the modified-hGP method achieves a satisfactory performance. The modified-hGP model achieves

TABLE 8: Statistical indices in forecasting process of the modified-hGP and diffusion models for the number of OECD mobile contract subscribers.

Statistical indices of the modified-hGP and diffusion models in forecasting the number of OECD mobile contract subscribers					
Model name	MAPE	wSSE	MSE	RMSE	MAE
Modified-hGP	0.004613	$8.58E - 05$	$1.28E - 05$	0.003572	0.002164
Logistic	0.028749	0.001562	0.00019	0.013771	0.011197
Gompertz I	0.02024	0.000547	$7.92E - 05$	0.008898	0.007826
Gompertz II	0.011585	0.000148	$2.94E - 05$	0.005422	0.004472
Bass	0.011497	0.000149	$2.93E - 05$	0.005415	0.004456
LogInLog	0.011592	0.00034	$4.09E - 05$	0.006392	0.004874
Bi-Logistic	0.008066	0.000116	$1.8E - 05$	0.004239	0.003334

TABLE 9: Statistical indices in forecasting process of the modified-hGP and diffusion models for the number of OECD mobile prepay subscribers.

Statistical indices of the modified-hGP and diffusion models in forecasting the number of OECD mobile prepay subscribers					
Model name	MAPE	wSSE	MSE	RMSE	MAE
Modified-hGP model	0.493749	0.000725	0.000363	0.019041	0.013627
Logistic	0.615593	0.006608	0.001076	0.032806	0.029652
Gompertz I	0.361444	0.003908	0.000617	0.024838	0.022462
Gompertz II	0.239243	0.001791	0.00032	0.017901	0.015845
Bass	0.24657	0.001829	0.000323	0.01798	0.015852
LogInLog	0.066263	0.010377	0.000887	0.029782	0.014094
Bi-Logistic	0.030654	0.000938	$7.98E - 05$	0.008933	0.004127

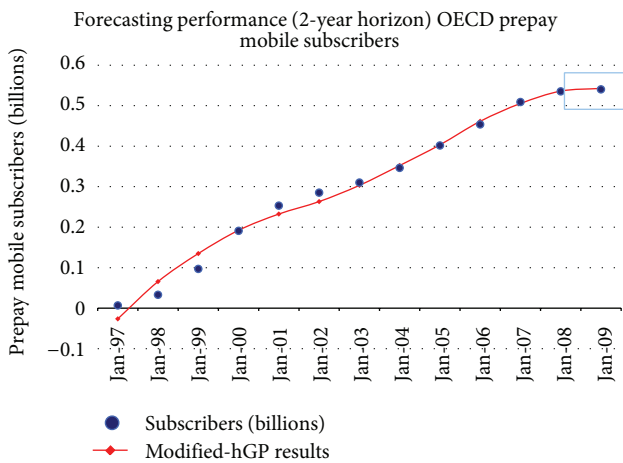


FIGURE 25: The forecasting performance of the modified-hGP (forecast window—2 years ahead) for the number of OECD mobile contract subscribers.

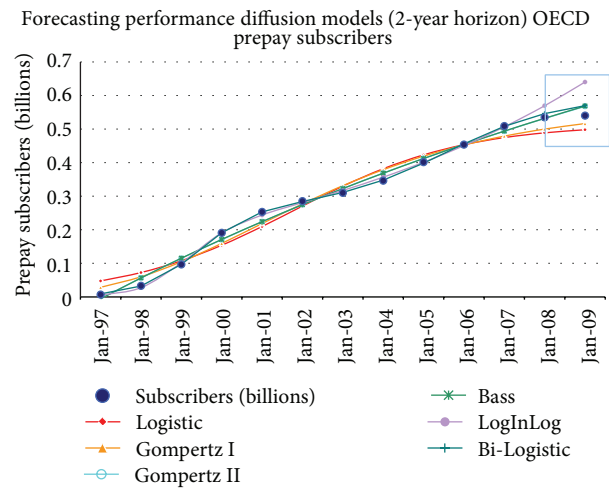


FIGURE 26: The forecasting performance of the diffusion models (forecast window—2 years ahead) for the number of OECD mobile prepay subscribers.

a wSSE value of 0.000725 and the Bi-Logistic 0.000938. Once again, the modified-hGP model residuals against time (data points), especially for the 2 last data points (the forecast period), show the error response of the modified-hGP model, as Figure 27 depicts.

Finally, the modified-hGP model’s performance, concerning OECD mobile 3G forecasting, is depicted in Figure 28. The corresponding modified-hGP model has the following format, which is time dependent:

$$0.0985374468592/(1+\text{EXP}(4.77304440434-0.569385591335*t))-0.363939773796*(1-\text{EXP}(-0.864258735151*t))-0.377831760256)$$

The forecasting performance of the diffusion models is presented in Figure 29. Also, the statistical indices of the produced forecasting modified-hGP and diffusion models are presented in Table 10.

TABLE 10: Statistical indices in forecasting process of the modified-hGP and diffusion models for the number of OECD mobile 3G subscribers.

Statistical indices of the modified-hGP and diffusion models in forecasting the number of OECD mobile 3G subscribers					
Model name	MAPE	wSSE	MSE	RMSE	MAE
Modified-hGP model	17.10841	$9.95E-05$	$4.12E-05$	0.006422	0.005037
Logistic	17.11114	$9.95E-05$	$4.12E-05$	0.006423	0.005038
Gompertz I	13.05382	0.000887	0.000132	0.011469	0.008277
Gompertz II	4.674663	0.004189	0.000538	0.023185	0.012231
Bass	4.66249	0.004268	0.000548	0.023399	0.012315
LogInLog	17.11736	$9.97E-05$	$4.13E-05$	0.006425	0.005019
Bi-Logistic	1.760939	0.011841	0.002242	0.047353	0.041684

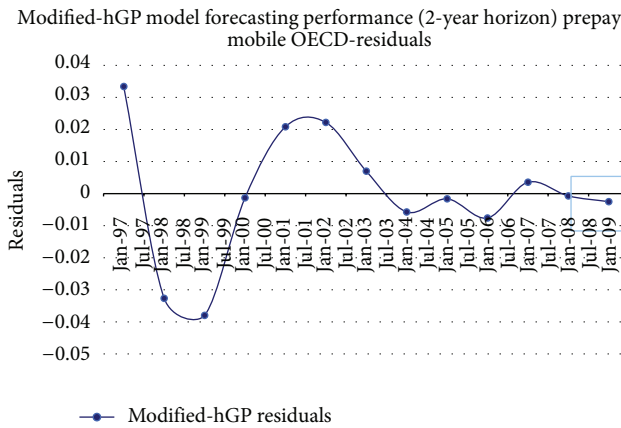


FIGURE 27: Residuals of the modified-hGP model in forecasting (2 years window period) of the number of OECD mobile prepay subscribers.

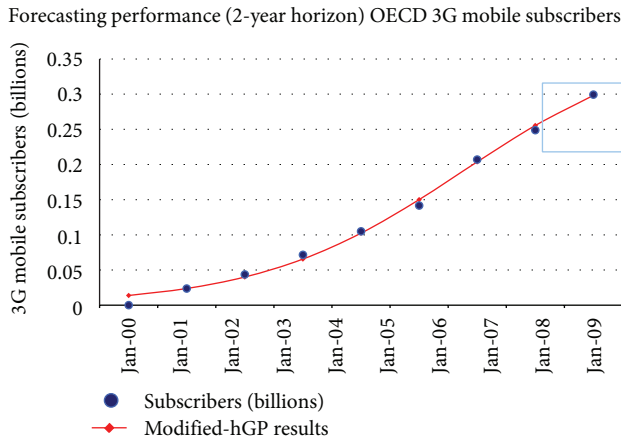


FIGURE 28: The forecasting performance of the modified-hGP (forecast window—2 years ahead) for the number of OECD mobile 3G subscribers.

Table 10 shows that the modified-hGP method achieves a good performance. The modified-hGP model achieves a wSSE value of  $9.95E-05$  similar to the Logistic and LogInLog. The MAPE indicator has the specific performance cause of the initial error at the first data point. As Figure 30 depicts, the modified-hGP model residuals against time (data points),

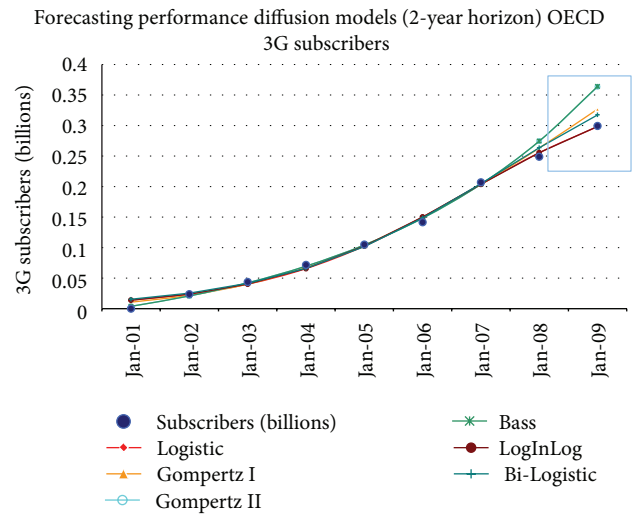


FIGURE 29: The forecasting performance of the diffusion models (forecast window—2 years ahead) for the number of OECD mobile 3G subscribers.

especially for the 2 last data points (the forecast period), show the error response of the modified-hGP model.

7.3. Comparison of the Results with ARIMA Model. The forecasting results of the generated models by the modified-hGP method are compared with those of the ARIMA method derived. As mentioned before, the statistic indicator wSSE has been used for the forecasting process.

ARIMA is an acronym for Auto-Regressive Integrated Moving Average. The ARIMA( $p, d, q$ ) model can be written as

$$(1 - \phi_1 \cdot B \cdots - \phi_p \cdot B^p) \cdot (1 - B)^d \cdot y_t = c + (1 + \theta_1 \cdot B + \cdots + \theta_q \cdot B^q) \cdot e_t. \quad (16)$$

In (16),  $B$  is the backward shift operator. The backward shift operator for  $k$ th-order difference shifts the data  $k$  periods back. In general, a  $k$ th-order difference can be written as

$$\frac{d^k y}{dt^k} = (1 - B)^k \cdot y_t. \quad (17)$$

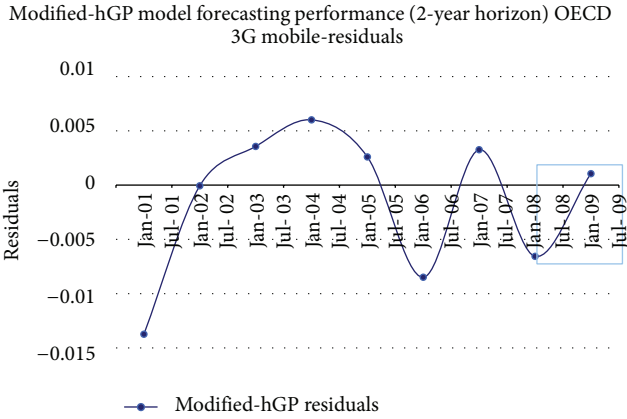


FIGURE 30: Residuals of the modified-hGP model in forecasting (2 years window period) of the number of OECD mobile 3G subscribers.

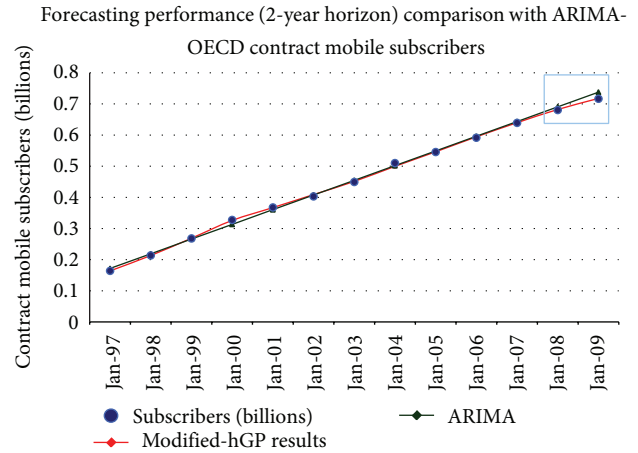


FIGURE 32: Comparison of the forecasting performances of modified-hGP with ARIMA model (forecast window—2 years ahead) for the number of OECD mobile contract subscribers.

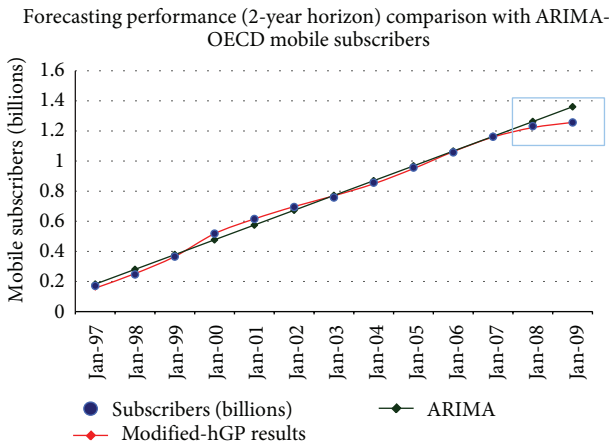


FIGURE 31: Comparison of the forecasting performances of modified-hGP with ARIMA model (forecast period—2 years ahead).

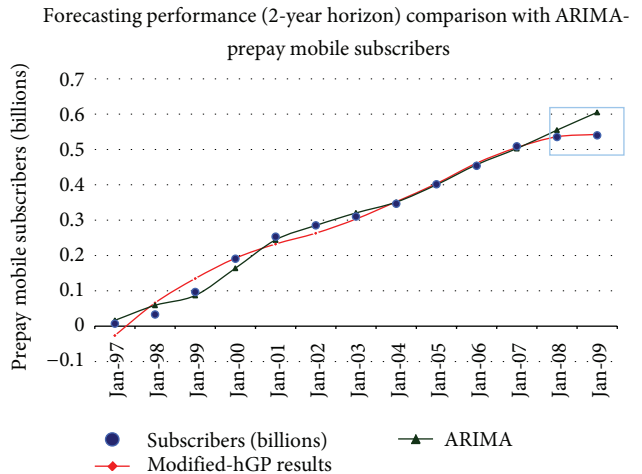


FIGURE 33: Comparison of the forecasting performances of modified-hGP with ARIMA model (forecast window—2 years ahead) for the number of OECD mobile prepay subscribers.

The  $p$  operator stands for the order of the autoregressive part  $(1 - \phi_1 \cdot B - \dots - \phi_p \cdot B^p)$  and operator  $d$  for the degree of the derivative of  $y$  part  $(1 - B)^d \cdot y_t$  and the  $q$  is order of the moving average part  $(1 + \theta_1 \cdot B + \dots + \theta_p \cdot B^p) \cdot e_t$  of (16). The  $\phi$  are the parameters of the autoregressive part of the model, the  $\theta$  are the parameters of the moving average part, and  $e$  are the errors [2].

The ARIMA models that are derived by the implementation of the “Gretl, Gnu Regression, Econometrics and Time-series Library” for the aforementioned datasets are depicted below. The forecasting performance of the same modified-hGP models and ARIMA models concerning total OECD mobile subscribers, contract, prepay, and 3G, is depicted in Figures 31, 32, 33, and 34, respectively. In every graph, the forecast period window is presented in the blue rectangle. In Tables 11, 12, 13, and 14, the comparison results of the statistical indices MAPE, MSE, RMSE, and MAE for the two predicted points are presented.

Considering Table 11, it can be concluded that the modified-hGP method achieves better forecasting performance than ARIMA model.

From Table 12, one realizes that the hGP method presents better performance than ARIMA model.

Also, in Tables 13 and 14, the modified-hGP method achieves better forecasting statistics than ARIMA model.

It could be concluded that the overall performance of the modified-hGP achieves better statistic indices than ARIMA model for the predicted data points.

**7.4. Robustness of the Proposed Modified-hGP.** The proposed method has been tested for the stability and the robustness. The program was executed 20 times in the same dataset of the mobile subscribers. The mean gap between the best and



TABLE 11: Statistical indices in forecasting process of the modified-hGP and ARIMA model for the total number of OECD subscribers.

Statistical indices of the modified-hGP and ARIMA model in forecasting the total number of OECD mobile subscribers				
Model name	MAPE	MSE	RMSE	MAE
Modified-hGP	0.003932	$7.03497E - 06$	0.002652	0.004845
ARIMA	0.053389	0.000591421	0.024319	0.066797

TABLE 12: Statistical indices in forecasting process of the modified-hGP and ARIMA model for the total number of OECD contract mobile subscribers.

Statistical indices of the modified-hGP and ARIMA model in forecasting the total number of OECD contract mobile subscribers				
Model name	MAPE	MSE	RMSE	MAE
Modified-hGP	0.001999	$3.97E - 07$	0.00063	0.001388
ARIMA	0.022104	$7.45E - 05$	0.008629	0.015574

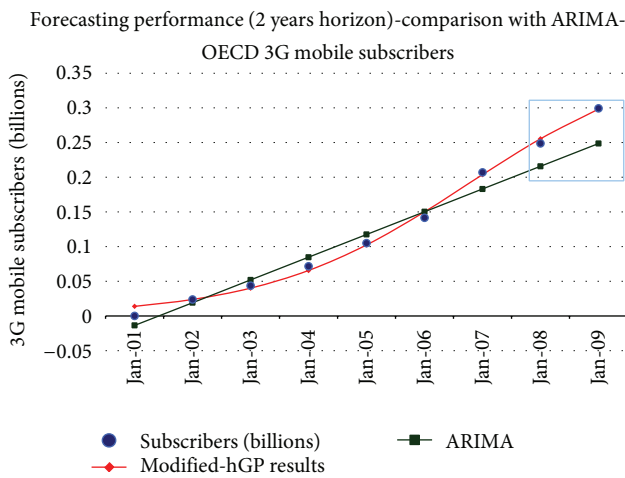


FIGURE 34: Comparison of the forecasting performances of modified-hGP with ARIMA model (forecast window—2 years ahead) for the number of OECD 3G mobile subscribers.

worst solutions was decreasing as the generation was increasing. Also, the curve of the total average of fitness value per generation was decreasing. It should be noted that in Table 15 the program parameters for the testing process are presented. In Figure 35, the mean value for the fitness value for the best and worst solutions (mean gap) per generation of the testing modified-hGP and the average fitness value per generation for the program executions are presented.

The difference between the worst and the best solutions is decreasing. In particular, after the 25th generation, the indices above are converging. The mean gap of wSSE for the worst-best solution begins from value 0.004146752 and ends up to 0.000165101.

### 8. Towards a Causal Forecasting Model: A Study of the Mobile Subscribers in OECD Countries

The introduction of GDPpC and CPI outside the time variable leads to the creation of causal forecasting models. This method provides a scenario based approach to forecasting. In order to study the future of mobile subscription in OECD

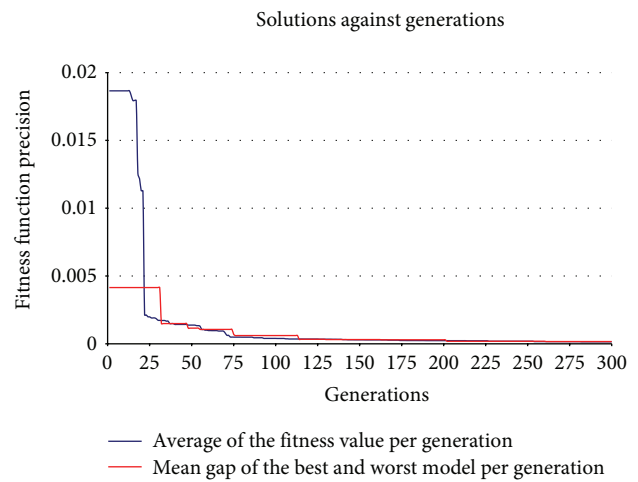


FIGURE 35: Testing the robustness of the forecasting modified-hGP.

countries, three scenarios are presented, according the GDPpC and CPI growth.

The pessimistic one concerns a continuing crisis scenario, so that the GDPpC and CPI growth rates are not increased. The second is a moderate growth scenario and the last one is the optimistic scenario, with GDPpC and CPI getting increased.

A variation of models is generated by the implementation of the modified-hGP method. According to Bayesian's criterion as well as the wSSE criterion, two models which combine all the variables, GDPpC, CPI, and time, are chosen. Figures 36, 37, and 38 depict the pessimistic, moderate, and optimistic scenario, respectively.

In Table 16, the selected models and their statistics are presented.

The BIC depends on the number of the parameters. The generated models with one variable, like time dependent models, in many cases, have better BIC performance, but not always better forecasting performance.

In contrary, multivariable models, with good enough BIC, yield a good enough forecasting performance.

In the pessimistic scenario, the first model (most likely) achieves 1.538472 billion of OECD mobile subscribers, in the

TABLE 13: Statistical indices in forecasting process of the modified-hGP and ARIMA model for the total number of OECD prepay mobile subscribers.

Statistical indices of the modified-hGP and ARIMA model in forecasting the total number of OECD prepay mobile subscribers				
Model name	MAPE	MSE	RMSE	MAE
Modified-hGP	0.00302	$1.72E - 06$	0.001312	0.001628
ARIMA	0.078342	0.00114	0.033759	0.042227

TABLE 14: Statistical indices in forecasting process of the modified-hGP and ARIMA model for the total number of OECD 3G mobile subscribers.

Statistical indices of the modified-hGP and ARIMA model in forecasting the total number of OECD 3G mobile subscribers				
Model name	MAPE	MSE	RMSE	MAE
Modified-hGP	0.014951	$4.91024E - 06$	0.002216	0.00381
ARIMA	0.151172	0.000407353	0.020183	0.041896

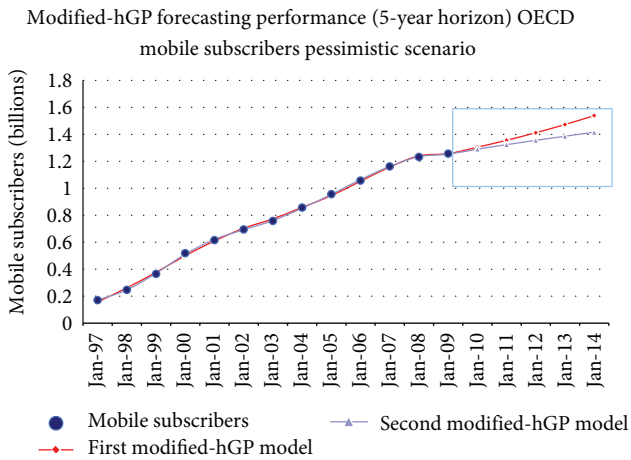


FIGURE 36: The forecasting performance (5 years ahead) of modified-hGP models—the pessimistic scenario.

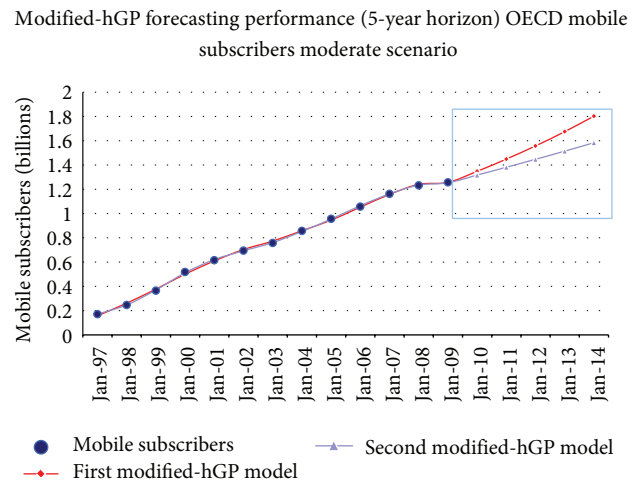


FIGURE 37: The forecasting performance (5 years ahead) of modified-hGP models—the moderate scenario.

year 2014. The second has a total number of 1.415613 billion subscribers. It should be noted that the GDPpC and CPI growths are unchanged.

In the moderate scenario (the most likely scenario), the first model achieves 1.8 billion of OECD mobile subscribers, in the year 2014. The second has a total number of 1.58 billion subscribers. It should be noted that the average GDPpC rate is 2.5% and average CPI rate growth 1.7%.

In the optimistic scenario, the first model achieves 1.948 billion of OECD mobile subscribers, in the year 2014. The second has a total number of 1.686 billion subscribers. It should be noted that the average GDPpC rate is about 4.4% and average CPI rate growth 2.2%.

### 9. Conclusions

This paper is a modification of our previous work [9] where the dataset was bigger, but in different area of interest. In this paper, an improved-hGP method was presented. The improved program achieved interesting forecasting models

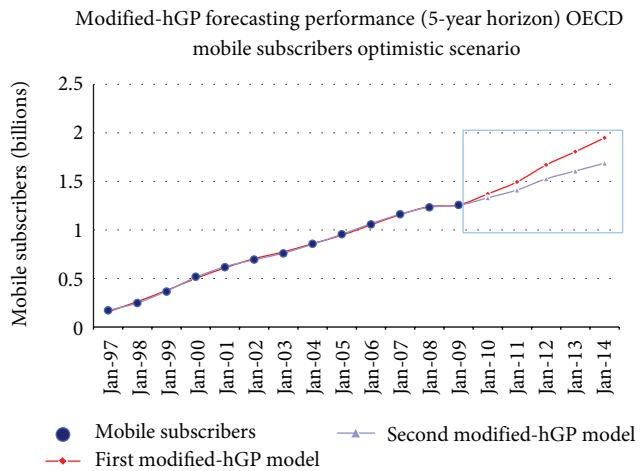


FIGURE 38: The forecasting performance (5 years ahead) of modified-hGP models—the optimistic scenario.

TABLE 15: The hGP parameters for the testing process.

Initialization parameters of the modified-hGP for the testing process	
Number of runs (executions of the program)	20
Maximum number of generations	300
Evaluation function	wSSE
Target upper limit of the precision for the solution candidates	0.1

TABLE 16: Models' expression and statistical indices of two modified-hGP models in forecasting the total number of OECD subscribers.

Modified-hGP models in forecasting the total number of OECD mobile subscribers				
Model name	Model's expression	MAPE	wSSE	BIC
First modified-hGP model	$(0.775676467651/(1 + \text{Exp}(437.848202929 - 0.578442691486 * (t - (-749.533938063)))))) + (4.22932908617/(1 + (-48176.7645458/(-48176.7645458 - 1307360612.91)))^f + (3.8401940855e - 06/(3.8401940855e - 06 - 1.24868528122e - 06))^f + (3.8401940855e - 06 + \text{Exp}(437.848202929 - 0.578442691486 * (t - (-749.533938063)))))) + (4.22932908617/(1 + (-48176.7645458/(-\text{GDPpC})/(-42.31487519/1 * \text{GDPpC})/(-102.941426203 * t - 2 - 1 * ((598.328254158 * \text{Exp}(-\text{CPI} * t/((\text{CPI}))))))))))$	0.004733	9.02E - 05	-9.31366
Second modified-hGP model	$(-10.4403884897/1 * \text{GDPpC})/(0.100668160215 * t - 2 - 2 - 1 * ((179388.4533 * \text{Exp}(-\text{CPI} * 13.8576512968))))$	0.021657	0.000433	-9.32401

with more variables than one. This GP method was implemented in dataset concerning the mobile subscribers of the OECD countries. The forecasting performance of the modified-hGP as well as the diffusion and ARIMA models was presented and the method presented satisfactory statistical indices.

The proposed method differs from the hGP in some points. Firstly, the diffusion models' set is extended with Bi-Logistic and LogInLog except for Logistic, Gompertz, and Bass so that the forecast horizon is improved, for long-term forecasting. Also, the functions' set of the method is extended by the insertion of new functions and function blocks. According to this technique, chromosomes with complicated syntax expressions can be presented with short length expression stings. The tournament selection is implemented for the crossover and mutation operations in order to maximize the algorithm's efficiency. Finally, a Bayesian inspired criterion has been implemented which, in combination with wSSE, improves the final selection of the forecasting models.

In general, the method could be considered as a forecasting tool that produces time dependent models and causal models for long-term forecasting with more variables than one. It should be noted that this method is compared with ARIMA model and achieved satisfactory performance. Also, the robustness of the proposed method has been analyzed. The implementation of the method is going to be continued on more datasets and it will be compared with other prediction methods in future work.

**Conflict of Interests**

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

**Acknowledgment**

The authors wish to express their acknowledgments to Professor Imed Kacem, University of Lorraine, France, for his constructive comments and suggestions, which helped to improve the quality of this paper.

**References**

- [1] J. S. Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer Academic Publishing, 2001.
- [2] R. J. Hyndman and G. Athanasopoulos, "Forecasting: principles and practice," 2012, <https://www.otexts.org/fpp/>.
- [3] 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.
- [4] Z. Griliches, "Hybrid corn: an exploration in the economics of technological change," *Econometrica*, vol. 25, no. 4, pp. 501–522, 1957.
- [5] E. Mansfield, "Technical change and the rate of imitation," *Econometrica*, vol. 29, pp. 741–766, 1961.
- [6] F. M. Bass, "A new product growth for model consumer durables," *Management Science*, vol. 15, no. 5, pp. 215–227, 1969.
- [7] E. M. Rogers, *Diffusion of Innovations*, The Free Press, New York, NY, USA, 5th edition, 2003.
- [8] 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.
- [9] K. Salpasaranis and V. Stylianakis, "A hybrid genetic programming method in optimization and forecasting: a case study of the broadband penetration in OECD countries," *Advances in Operations Research*, vol. 2012, Article ID 904797, 32 pages, 2012.
- [10] J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, 1975.

- [11] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, The MIT Press, 1992.
- [12] J. R. Koza, "Genetic programming for economic modeling," *Statistics and Computing*, vol. 4, no. 2, pp. 187–197, 1994.
- [13] J. Duda and S. Szydło, "Collective intelligence of genetic programming for macroeconomic forecasting," in *Proceedings of the Computational Collective Intelligence. Technologies and Applications (ICCCI '11)*, vol. 6923 of *Lecture Notes in Computer Science*, pp. 445–454, Springer, Berlin, Germany, 2011.
- [14] P. Meyer, "Bi-logistic growth," *Technological Forecasting and Social Change*, vol. 47, no. 1, pp. 89–102, 1994.
- [15] P. S. Meyer and J. H. Ausubel, "Carrying capacity: a model with logistically varying limits," *Technological Forecasting and Social Change*, vol. 61, no. 3, pp. 209–214, 1999.
- [16] M. N. Sharif and K. Ramanathan, "Binomial innovation diffusion models with dynamic potential adopter population," *Technological Forecasting and Social Change*, vol. 20, no. 1, pp. 63–87, 1981.
- [17] C. Chen and C. Watanabe, "Diffusion, substitution and competition dynamism inside the ICT market: the case of Japan," *Technological Forecasting and Social Change*, vol. 73, no. 6, pp. 731–759, 2006.
- [18] W. B. Langdon, R. Poli, N. F. McPhee, and J. R. Koza, "Genetic programming: an introduction and tutorial, with a survey of techniques and applications," *Studies in Computational Intelligence*, vol. 115, pp. 927–1028, 2008.
- [19] R. Kass and A. Raftery, "Bayes factors," *Journal of the American Statistical Association*, vol. 90, pp. 773–795, 1995.
- [20] OECD Factbook 2011, "Economic, Environmental and Social Statistics," 2013.
- [21] GSMA, "European Mobile Industry Observatory," 2011, London, UK, <http://www.gsma.com/>.
- [22] OECD iLibrary, "OECD Communications Outlook," 2011.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

