

# Equations for Describing Behavior Tables in Thermodynamics Using Genetic Programming: Synthesizing the Saturated Water and Steam Table

Martín Montes Rivera<sup>1</sup>, Marving Omar Aguilar Justo<sup>1</sup>, Alberto Ochoa Zezzatti<sup>2</sup>

<sup>1</sup> Universidad Politécnica de Aguascalientes, Aguascalientes,  
Mexico

<sup>2</sup> Universidad Autónoma de Ciudad Juárez, Chihuahua,  
Mexico

martin.montes@upa.edu.mx, marving.aguilar@upa.edu.mx, alberto.ochoa@uacj.mx

**Abstract.** There are several tables with important data used in the calculus of different processes like machining tables, friction tables and thermodynamics processes tables, or as it is explored in this paper, the description of saturated water and steam table. We propose the generation of equations for describing the entire behavior of numerical values in a table using Genetic Programming (GP), when table data describes the variable behavior of a dependent function. This obtained equations simplify the calculus process without requiring several tables and allowing to work when tables are not available for a desired value of an independent variable, a common situation in thermodynamics. In this case it is tested the proposed algorithm for synthesizing the saturated water and steam table.

**Keywords:** Genetic Programming, Synthesizing Tables, Saturated Water and Steam Tables.

## 1 Introduction

Tables are arrays of rows and columns used as tools for presenting information, they are commonly used to show values of responses, relationships, scores, percentages and statistical results, among others [1].

The results of measurements obtained from input-output models and the evaluation of functions are commonly expressed using tables, which describe the behavior of dependent variables in function of independent variables. Examples of function applications are shown in [2] and tables of input-output models with their applications used in thermodynamics are shown in [3].

Tables are common tools in different sciences and processes; thermodynamics is not an exception and students must learn to use them [4].

Thermodynamics like other industrial sciences applied has several mathematical models, which try to describe the behavior of specific variables, sometimes the obtained models based on physical and mathematical principles are complex and tables with the data of the models are used instead [5].

Tables increase accessing information speed, but if the values on tables are used for describing the behavior of a variable, sometimes is better to plot the data or obtain an equation to describe their values for a better comprehension [6].

Nowadays several information is uploaded on internet, this include great amounts of numerical data distributed in several tables which like other information are sometimes difficult to find among other documents [7]. Equations instead tables require less space on internet and are easier to find, in this sense, is better to use equations instead tables when is possible synthesizing them.

Tables in thermodynamics are commonly used to indicate the substitution parameters in process equations, but in this case, would not be better to get an equation that represents the table? And then, substitute it at the process equation which already imply to work with an equation.

There are different ways for obtaining equations from tables related to the kind of equation that represents the numerical data, for example, the processes for obtaining the equation of a table are different if the data chart has straight line shape or parabola shape, because different math considerations are required [2].

Sometimes numerical data in a table could have behaviors that can only be represented with no linear equations, which require complicated mathematical methodologies for obtaining them, in this cases computer optimizing algorithms are an alternative [8].

The numerical data in a table has an equation with unknown structure and an easy form to realize what kind of equation would work for synthesizing it, would imply to plot the dependent variable against the independent and then, human judgment would be required to identify the data chart shape and approximate it to a general equation (supposing that there are mathematical methodologies for approximating the equation).

Optimizing algorithms are tools commonly used for synthesizing equations using known information for improving results. Evolutionary algorithms are optimizing metaheuristic programs that improve equations or parameters by using known information and objective functions combined with natural selection principles [9, 10].

An optimizing evolutionary algorithm which works well with unknown structures is Genetic Programming (GP), even if there are no mathematical techniques to approximate the required equation [8, 11].

In this paper is proposed an algorithm to use GP for obtaining the unknown equations that represent relationships among variables in a table and continue with the process using equations instead of tables. We test the proposed algorithm working with the saturated water and steam table used in thermodynamics for different processes like generation of Pressure, Volume and Temperature (PVT) diagrams.

## **2 Background**

Manufacturing industries are continually improving their models for conceptual design and engineering processes. The work in [5] shows that Thermodynamics demands better models for simulation, which has been a focus of research that will continue looking for nonexistent and more accurate models.

Tables are used in several applications and sciences concentrating important data used in health, economics, research, among others, this has motivated the emerging of new techniques for working with great amount of data in tables and their characteristics extraction like is shown in [12,13].

The work in [4], shows how Thermodynamics has been taught to students and that several tables are required in the generation of PVT diagrams and calculus of different processes, they propose to use tables for the solution of thermodynamic problems and their synthesis, which help students in their understanding.

Teaching thermodynamics has different approaches nowadays, the work in [14], compare the efficiency of use traditional and cooperative methodologies, together with the use of Technologies of Information and Communication (TIC).

The work in [15] exhibit the application of height and weight tables for obtaining equations to predict mid-arm circumference and the correct blood pressure cuff size.

Singapore like other countries concentrates its statistical information using tables, the work in [16] describes Input Output Tables (IOTs), obtained from Supply and Use of Tables (SUTs), used for a variety of proposes like economics, process modeling and impact studies, their paper is focused on the transformation of data from SUTs to IOTs format.

The work in [6] shows the use of tables in education and demonstrates that interpretation of tables by teachers and students is better understood when charts of the columns in tables are generated, improving their comprehension.

Reporting evidence tables in research have better interpretation when they are used for generating charts, like is shown in [17].

Transformation of time series tables to equations, it is a common action when modeling physical processes and GP is commonly used for their synthesis.

The work in [18] shows synthesizing monotonic stochastic processes using time series signals (which could be expressed like tables) and GP.

Similarly, the work in [19] describes the process for synthesizing functions that represent several recollected data of relevant heterogeneous catalysts of different materials, using GP.

Another application using GP for synthesizing equations is shown in [20], where GP is used for equation synthesis with simulated data of main variables and the average of their relations, this work is tested for obtaining equations of Hydraulic processes.

The work in [21], shows a comparison between Artificial Neural Networks (ANN) and GP, applied in the obtaining of new formulations for electrical resistivity based on numerical data, showing the performance of GP in data regression, like the proposed in this work.

### 3 Water Steam Tables in Thermodynamics

Calculus of different processes in Thermodynamics require tables for describing behavior of gases or liquids, which depend of temperature, pressure or volume [3].

Water and water steam tables describe the behavior of water at different conditions of temperature and pressure, the variables used in water steam table in [3] are: Temperature ( $T$ ), Pressure ( $P$ ), Specific volume for saturated liquid ( $v_f$ ) and saturated gas ( $v_g$ ), Enthalpy of saturated liquid ( $h_f$ ) and saturated gas ( $h_g$ ), Internal energy of saturated liquid ( $u_f$ ) and saturated gas ( $u_g$ ) and Entropy of saturated liquid ( $s_f$ ) and saturated gas ( $s_g$ ).

The problems that engineers solve with this table are the calculus of the main variables (mentioned before) using  $T$  or  $P$  like independent variables, they also obtain other variables that require substitution of main variables in equations, for example, determining water intrinsic expansion, behavior of refrigerant cooling with water, among others [3].

Exercise example 3-1 in [3], shows a simple application of what students must learn to do with water and water steam tables, this example asks the pressure and volume of a recipient containing 50 kg of saturated liquid water at 90 °C. In this case, the value of  $T$  is in the table, but if it is not, then, interpolation with equation (1) must be performed:

$$y = y_0 + (y_1 - y_0) \frac{x - x_0}{x_1 - x_0} \quad (1)$$

Where  $y$  is the value of the variable when  $x$  takes an absent value in the table, and  $y_0, y_1, x_0$  and  $x_1$  are the nearest known conditions in the table for  $y$  and  $x$ .

Assume that in the previous exercise  $T = 88$  °C for understand what happen if there are required absent values in the table and linear interpolation in equation (1) is applied for obtaining the absent values of  $P, v_f$  and  $V$  (volume of the recipient).

The near section to the required value of  $T$  is shown in table (1), obtained from the values in [3].

**Table 1.** Section of water and water steam table in [3] with near values for  $T = 88$  °C.

$T(^{\circ}C)$	$P(\text{kPa})$	$v_f(\frac{\text{m}^3}{\text{kg}})$
80	47.416	0.001029
85	57.868	0.001032
90	70.183	0.001036
95	84.609	0.001040

Considering table (1), the value of  $x$  represents  $T$ , which means that  $x = 88$  and the nearest known temperatures are  $x_0 = 85$  and  $x_1 = 90$ . For obtaining  $P$  when  $T = 88$  °C,  $y$  represents the pressure and its nearest known values in the table are  $y_0 = 57.868$  and  $y_1 = 70.183$ . Then,  $P = 65.257$  kPa at 88 °C like is shown in equation (2) obtained substituting the nearest known conditions in equation (1):

$$y = 57.868 + (70.183 - 57.868) \frac{88 - 85}{90 - 85} = 65.257 \quad (2)$$

Similarly,  $v_f$  (in this case  $y$ ) is calculated with nearest known conditions  $y_0 = 0.001032$ ,  $y_1 = 0.001036$  and temperature conditions represented with  $x$  are the same as before, then,  $v_f = 0.0010344 \frac{\text{m}^3}{\text{kg}}$  substituting these values in equation (1).

The calculus of  $V$  is obtained with  $v_f$  times  $m$  (equation (3)), since specific volume is the volume per mass unit, in this case  $m = 50$  kg, then,  $V = 0.05172 \text{ m}^3$ :

$$V = \left( 0.0010344 \frac{\text{m}^3}{\text{kg}} \right) (50 \text{ kg}) = 0.05172 \text{ m}^3 \quad (3)$$

There are other applications with water and water steam tables, but all of them follow the same process in previous exercise, find the required property in the table, find the desired value of independent variable ( $T$  or  $P$ ), if this is not in the table, then, interpolate using equation (1) and substitute the obtained values in the specific process.

In this work equation synthesis for testing the proposed algorithm takes  $T$  like independent variable and synthesizes dependent variables  $P$ ,  $v_f$ ,  $v_g$ ,  $h_f$  and  $h_g$ , allowing to get any value without require tables or interpolation, but only  $T$  substitution.

## 4 Genetic Programming

GP is an evolutionary algorithm proposed by John R. Koza in 1992, this algorithm like other evolutionary algorithms follows natural selection principles in [23], i.e. its main operations are population initializing, fitness value calculation, fitness-based selection, crossover and mutation (Fig. 1) [11, 22].

Most used code representations of elements in GP are tree base-individuals (Fig. 2), which have disadvantages like conversions from tree representation to equations, and uses several pointers that sometimes make very slow its execution. Nevertheless, there are other many representation with faster manipulation, in this case we need to generate equations so a variation of linear GP with native equation structure is used like described in [24, 25].

Equation structure is a linear GP representation that stores individuals in arrays with operators and numbers and parenthesis that separate operations, similarly to depth in tree-based individuals [24, 25].

Fitness evaluation comes after population initializing using fitness function, which is different depending of the desired objective.

After fitness evaluation fitness based selection must define which chromosomes will be selected for crossover operation. Tournament selection is a commonly used selection technique, because allows to control selection pressure with tournament size ( $t_s$ ) and population size ( $p_s$ ). Tournament is made by selecting  $t_s$  random individuals for match and winners according to its fitness value are taken into mating pool.

Crossover operation takes individuals from mating pool for offspring generation. Equation structure crossover is similar to haploid crossover in Genetic Algorithms (GA), which exchange genes between two parents in a random crossover point for generating offspring, like is shown in Fig. 3 with crossover point at 5th bit [24, 25].

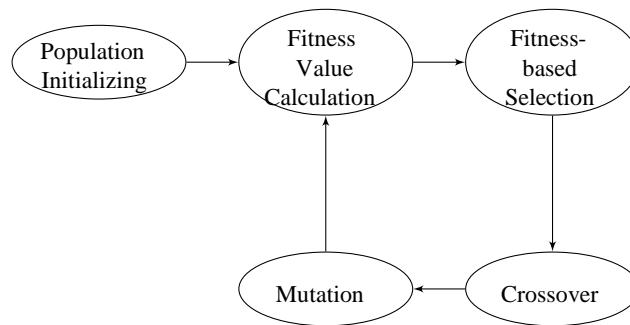


Fig. 1. Evolutionary algorithm diagram.

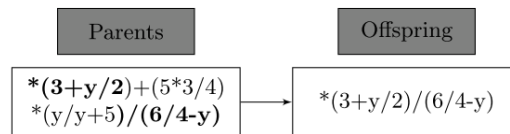


Fig. 2. GP tree-based individual that represents  $(4 - 2) \times (8 + y)$ .

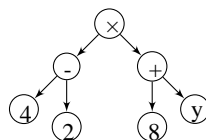


Fig. 3. GP structure equation individual and its crossover operation.

Mutation in linear GP with equation structure is similar to mutation in GA, which could be Single Gene Mutation (SGM) or Multiple Gene Mutation (MGM), where SGM changes aleatory single selected bits and MGM selects two points for changing all alleles or bits between them [8,24]. MGM has shown to be better in GP since there are more possible values than 1 or 0 for its alleles (like occurs in GA), therefore diversity plays a crucial place for obtaining good solutions in GP and MGM supply more genetic material than SGM for a single individual . SGM and MGM are shown in Fig. 4 [26].



Fig. 4. Single (left figure) and multiple gene mutation (right figure) [8].

## 5 Applying Equation Structure for Synthesizing Tables

Our proposed algorithm starts by generating two arrays, first one ( $X_i$ ) contains all values of  $T$  in water and water steam table and second one ( $Y_i$ ) with all values of required dependent variable ( $P, v_f, v_g, h_f$  or  $h_g$ ).

Since  $T$  is the independent variable, the individuals in population are expected to be functions of  $T$ , therefore, operands are  $T$  and integer numbers  $[0,9]$ ; operators are plus (+), minus(-), times, ( $\times$ ), divide (/) and power (^).

The control values generating population are number of operators ( $n_o$ ), depth or number of operations separated with parenthesis ( $d$ ), a dividing factor ( $div$ ) for increase or decrease values in equations (required when  $Y_i$  has small values) and population size ( $p_s$ ).

After generating population, fitness function ( $F(p_j)$ ) evaluates fitness in the  $j_{th}$  individual of the population by calculating accumulated absolute error produced between the dependent variable  $Y_i$  and the evaluation of substituting  $T$  in equation of  $p_j$  with all values of  $X_i$ , for every  $i$  of the  $n$  available values in the table, like is shown in equation (4), therefore, best adapted individuals has  $F(p_j) = 0$  and it worst  $F(p_j) = \infty$  :

$$F(p_j) = \sum_{i=1}^n abs[Y_i - p_j(X_i)] \quad (4)$$

Selection operation used is tournament with selected tournament size ( $t_s$ ) control and crossover operation is performed like described in section 4, mutation is MGM with control value in number of mutations per generation ( $m_{gen}$ ). Overall control in the algorithm is the number of generations ( $n_{gen}$ ).

## 6 Results

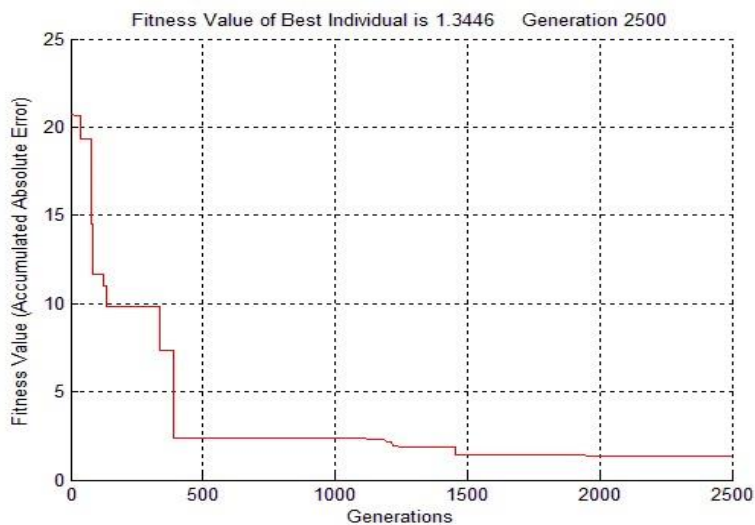
The resulting equations for dependent variables  $P, v_f, v_g, h_f$  and  $h_g$ , were obtained with seed value of 1 for MATLAB<sup>TM</sup> random numbers (platform used for implementing the proposed algorithm) and the control parameters listed in table 2, obtained heuristically since they

depend of the specific application, therefore, they are increased until find acceptable equations for the processed model.

**Table 2.** GP parameters for obtaining synthesized equations.

Parameter: \ Model:	$P$	$v_f$	$v_g$	$h_f$	$h_g$
$n_{gen}$	2500	2500	3000	2500	3000
$p_s$	200	200	200	200	200
$n_o$	5	5	6	6	6
$d$	5	5	6	6	6
$div$	1	1000	1	1	1
$t_s$	3	3	3	3	3
$m_{gen}$	20	20	20	20	20

Charts of fitness behavior (in this case, accumulated absolute error) against generations and its final accumulated error in best individuals are shown per model from Fig. 5 to 9.



**Fig. 5.** Fitness against generations synthesizing  $P$  equation.

In every chart must be consider that the fitness value or accumulated error in the best chromosome is the total absolute error obtained between the generated equation and the data in the table, substituting the 100 independent values available in the saturated water and steam table.



The relative error is calculated as the quotient of the average error in the obtained equation and the average value of data in the table.

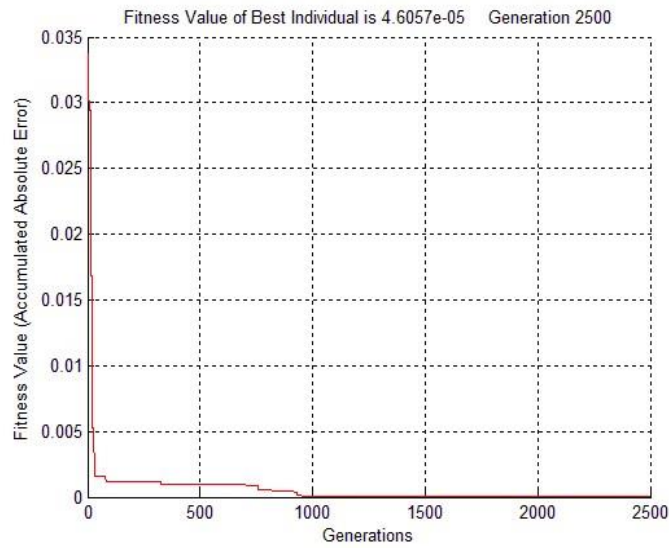


Fig. 6. Fitness against generations synthesizing  $v_r$  equation.

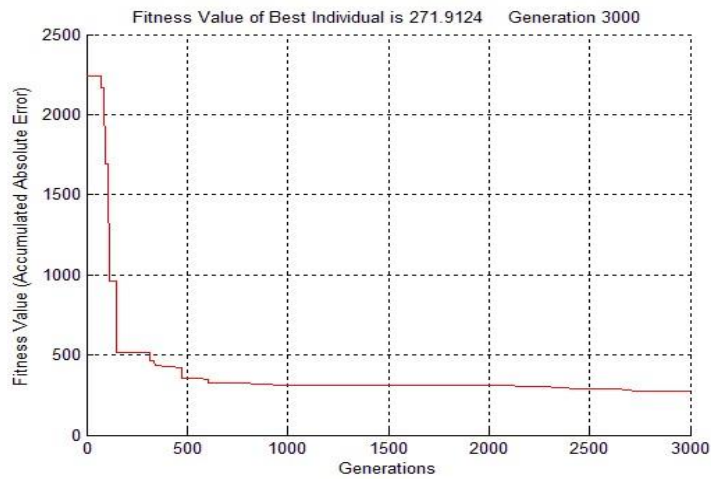


Fig. 7. Fitness against generations synthesizing  $v_g$  equation.

P synthesized equation obtains and average error of 0.013446 units, which is 5.38% relative error.

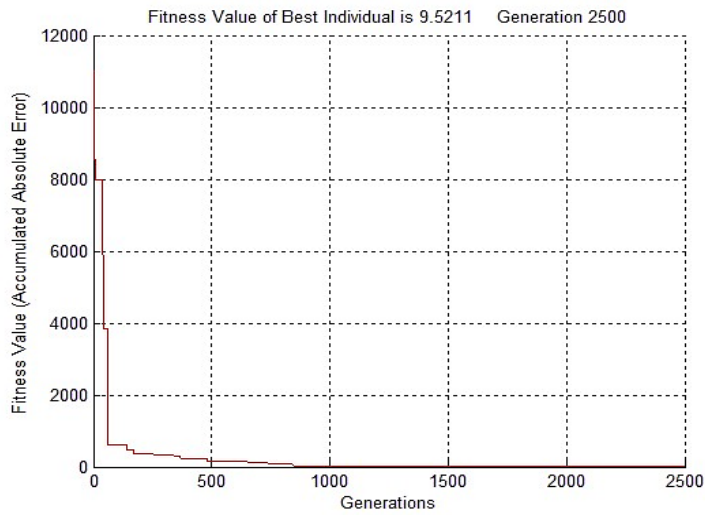


Fig. 8. Fitness against generations synthesizing  $h_r$  equation.

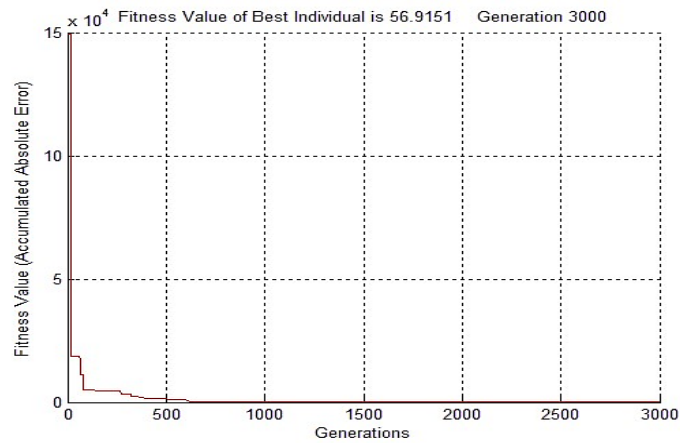


Fig. 9. Fitness against generations synthesizing  $h_g$  equation.

$v_f$  synthesized in equation obtains and average error of  $4.6057 \times 10^{-7}$  units, which is 0.0461% relative error.

$v_g$  synthesized equation obtains and average error of 2.719124 units, which is 7.77% relative error.

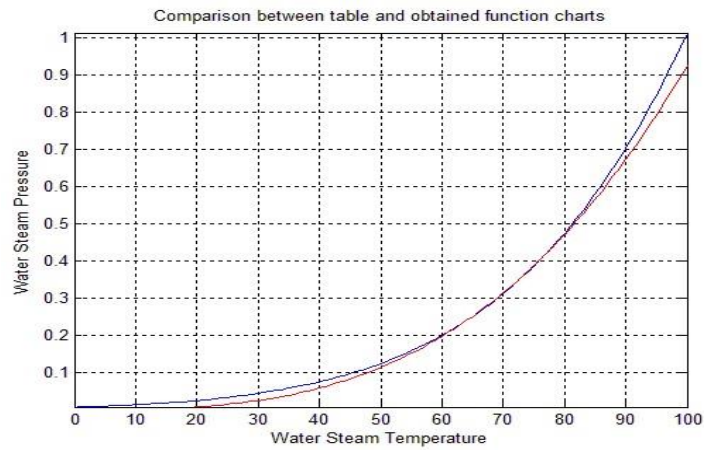


Fig. 10. Comparison between table data and obtained equation for  $P$  model.

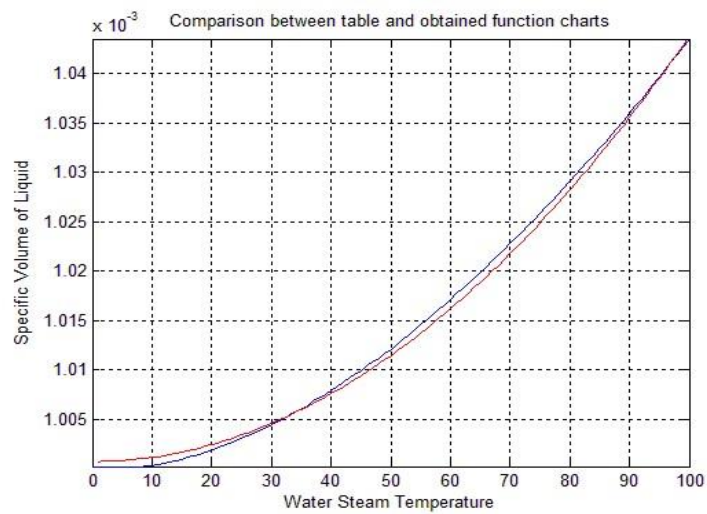


Fig. 11. Comparison between table data and obtained equation for  $v_f$  model.

$h_f$  synthesized equation obtains and average error of 0.095211 units, which is 0.045% relative error.

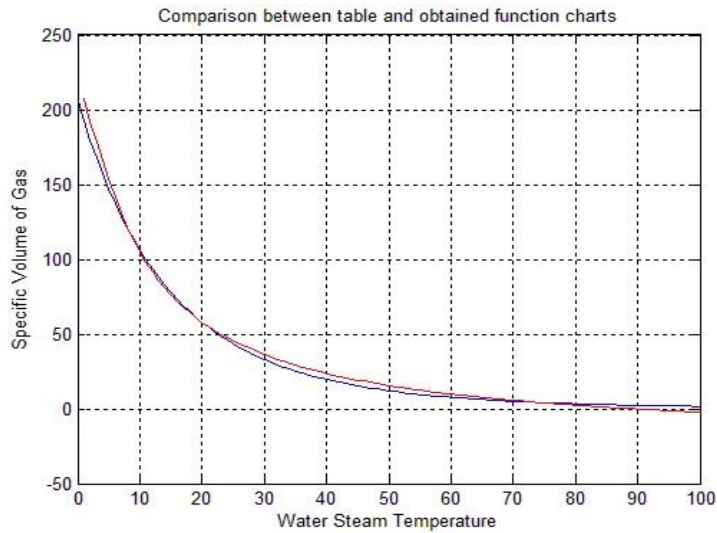


Fig. 12. Comparison between table data and obtained equation for  $v_g$  model.

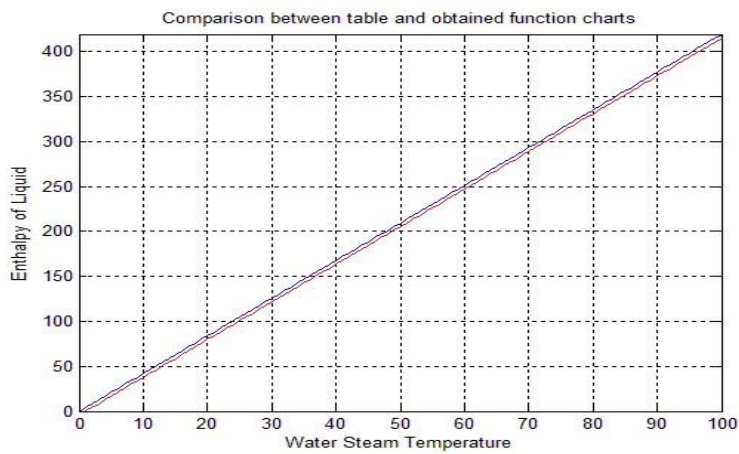


Fig. 13. Comparison between table data and obtained equation for  $h_f$  model.

$h_g$  synthesized equation obtains and average error of 0.569151 units, which is 0.0218% relative error.

Charts of comparison between data tables and obtained equations are shown from Fig. 10 to 14, this results show that error levels are almost imperceptible when plotting data

information in a table. Curves obtained from tables are shown in blue and curves of obtained equations in red.

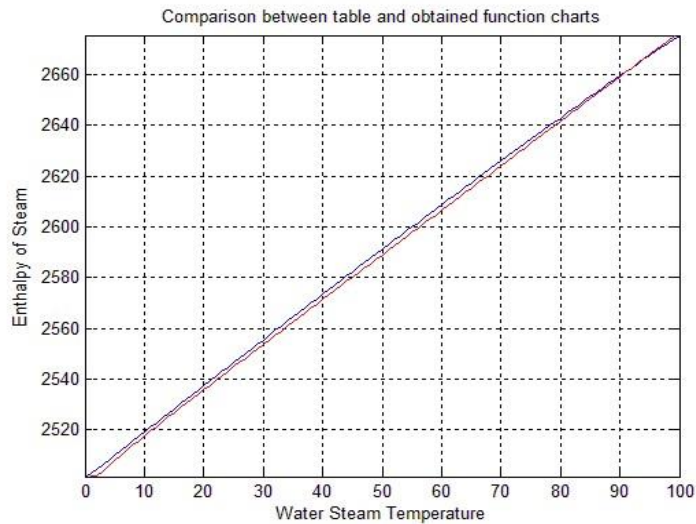


Fig. 14. Comparison between table data and obtained equation for  $h_g$  model.

Obtained equations using GP for water and water steam tables are shown from equation (5) to (9).

$$P = \frac{3.16T(3T - 7)(T + 3)}{10^7} \quad (5)$$

$$v_f = 0.0084T \left( \frac{5.1202T}{10^7} - \frac{1.1947}{10^6} + \frac{1.024}{10^6} \right) + 0.001001 \quad (6)$$

$$v_g = \frac{6T - 1938}{24(2)^T + 23} - \frac{0.02083(1081T - 95511)}{T + 8} \quad (7)$$

$$h_f = 4.1845T + 0.18371 \quad (8)$$

$$h_g = 1.75T + \frac{1.75T - 17.4}{T + \frac{2}{T} + 5} + 2502 \quad (9)$$

## 7 Conclusions

The proposed algorithm takes advantage of GP for synthesizing equations and applying it successfully for obtaining equations that model variables in water and water steam table. This algorithm also shows that linear GP with equation structure produce good results for synthesizing equations and therefore structures of equations of water and water steam tables are available for optimizing with GA in future work, which could improve its precision.

Using the obtained equations it is possible to solve easily and faster the Thermodynamics exercises with a small error of precision (which could be improved using optimizing techniques with the obtained equation structures) compared with traditional methodology, since tables are not required any more and interpolation it is not necessary with this equations.

## References

1. Arlene, F.: *How to Conduct Surveys: A Step by Step Guide*. SAGE Publications. 4th Edition (2009)
2. Leithold, L.: *The Calculus*. Oxford University Press, 7th edition (1998)
3. Çegel, Y. A., Boles, M. A.: *Termodinámica*. Mc-Graw Hill, 7th edition (2012)
4. Maldonado, L. C., Chapou, J. L., Arcos, J. M., Granados, J.: *Termodinámica: metodología para un mejor aprendizaje*. *Rev. Cub. Fís*, Vol. 27, No. 2A, pp. 125–129 (2010)
5. Chau-Chyun, C., Mathias, P. M.: *Applied Thermodynamics for Process Modeling*. *AIChE Journal*, Vol. 48, No. 2, pp. 194–200 (2002)
6. Chengping, Z.: *Using Excels Data Table and Chart Tools Effectively in Finance Courses*. *Journal of Accounting and Finance*, Vol. 15, No. 7 (2015)
7. Gelbukh, A.: *Enriquecimiento parcial de la consulta para la recuperación de información de alta calidad y flexibilidad*. *Research in Computing Science*, Vol. 73, pp. 17–44 (2014)
8. Weise, T.: *Global Optimization Algorithms: Theory and Application*. e-book obtained from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.64.8184&rep=rep1&type=pdf> (2009)
9. Luis, A., Prospero, C., Mejía, M., Ruiz, J., Vela, V.: *Aprendizaje incremental basado en población como buena alternativa al uso de algoritmos genéticos*. *Research in Computing Science*, Vol. 116, pp. 51–64 (2016)
10. Pérez-Gaspar, L. A., Caballero-Morales, S. O., Trujillo-Romero, F.: *Integración de optimización evolutiva para el reconocimiento de emociones en voz*. *Research in Computing Science*, Vol. 93, pp. 9–21 (2015)
11. Poli, R., Langdon, W. B., McPhee, N. F.: *A Field Guide to Genetic Programming*. e-book obtained from: [http://www0.cs.ucl.ac.uk/staff/wlangdon/ftp/papers/poli08\\_fieldguide.pdf](http://www0.cs.ucl.ac.uk/staff/wlangdon/ftp/papers/poli08_fieldguide.pdf) (2008)

12. Rodríguez-Flores, F., Flores-Pulido, L., Dávila de la Rosa, E.: Inteligencia de negocios y minería de datos aplicado a la industria refresquera. *Research in Computing Science*, Vol. 126, pp. 63–71 (2016)
13. León-Barranco, A., Saucedo-Lozada, S. N., Avendaño-Jimenez, I. Y., Martínez-Leyva, R., Carcaño-Rivera, L. A.: Business Intelligence in Educational Institutions. *Research in Computing Science*, Vol. 96, pp. 43–53 (2015)
14. Méndez, D.: The influence of Teaching Methodologies in the Learning of Thermodynamics in Secondary Education. *Journal of Baltic Science Education*, Vol. 12, No. 1, pp. 59–72 (2013)
15. Ostchega, Y., Dillon, C., Prineas, R. J., McDowell, M., Carroll, M.: Tables for the selection of correct blood pressure cuff size based on self-reported height and weight and estimating equations for mid-arm circumference: data from the US National Health and Nutrition Examination Survey. *Journal of Human Hypertension*, Vol. 20, pp. 15–22 (2006)
16. Department of Statistics, Ministry of Trade & Industry, Republic of Singapore: Singapore Supply and use, and Input-Output Tables. Department of Statics Singapore (2014)
17. Mignini, L., Champaneria, R., Mishanina, E., Khan, K. S., EBM-CONNECT: Graphical displays for effective reporting of evidence quality tables in research syntheses. *Reproductive Health*, Vol. 13–21 (2016)
18. Ross, B. J.: Using Genetic Programming to Synthesize Monotonic Stochastic Processes. *CI '07 Proceedings of the Third IASTED International Conference on Computational Intelligence*, pp. 71–78 (2007)
19. Baumes, L. A., Blansch, A., Serna, P., Tchougang, A., Lachiche, N., Collet, P., Corma, A.: Using Genetic Programming for an Advanced Performance Assessment of Industrially Relevant Heterogeneous Catalysts. *Materials and Manufacturing Processes*, Vol. 24, pp. 282–292 (2009)
20. Moreno, J. A., Cárdenas, E. G., Obregón, N.: A Novel Genetic Programming Heuristic Supported on Genetic Algorithms and Average Mutual Information as Fitness Evaluation Metric Some Applications from Hydraulics. *XVIII International Conference on Water Resources* (2010)
21. Karahan, I. H., Ozdemir, R., ErKayman, B.: A comparison of genetic programming and neural networks; new formulations forelectrical resistivity of ZnFe alloys. *Appl Phys A Materials Science and Processing*, Vol. 113, pp. 459–476 (2013)
22. Oltean, M., Grosan, C.: Using traceless genetic programming for solving multi-objective optimization problems. *Journal of Experimental & Theoretical Artificial Intelligence* 3, Vol. 19, pp. 227–248 (2007)
23. Darwin, C.: *The Origin of Species*. John Murray, penguin classics, 1985 edition (1859)
24. Montes-Rivera, M., Paz-Ramos, M. A., Orozco-Mora, J. L., Ochoa-Zezzatti, C. A.: An Alternative Structure to Replace the Tree-Based Individuals in a Genetic Programming Algorithm for Decoupling MIMO Systems. *6th MICAI Workshop on Hybrid Intelligent Systems* (2013)
25. Montes-Rivera, M., Paz-Ramos, M. A., Orozco-Mora, J. L.: *Generador de bloques desacoplantes empleando programación genética: Aplicado al desacoplamiento de un sistema de riego y un proceso de transporte de papel*. Editorial Académica Española (2014)
26. Bäck, T.: *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford University Press (1996)