

# A GP-evolved Formulation for the Relative Permittivity of Water and Steam

S. V. Fogelson, W. D. Potter  
Artificial Intelligence Center  
The University of Georgia, USA  
Contact Email Address: sergeyf1@uga.edu

*The relative permittivity (or static dielectric constant) of water and steam has been experimentally calculated at a relatively wide range of temperatures and pressures. Two separate functions for predicting the relative permittivity of water and steam in two distinct thermodynamic regions are evolved using genetic programming. A data set comprised of all of the most accurate relative permittivity values, along with temperature, pressure, and density values from the entire experimentally calculated range of these values, found in [Fern95], is used for this task. The accuracy of these two functions is evaluated by comparing the values for the relative permittivity calculated using the evolved functions and the values calculated using the latest formulation of Fernandez et al., found in [Fern97] to the aforementioned data set. In both regions, the newly evolved functions outperform the most current formulation in terms of difference between calculated and experimentally obtained values for the dielectric constant.*

*Keywords: genetic programming, relative permittivity, thermodynamic properties*

## I. Introduction

The relative permittivity (or static dielectric constant) of water and steam,  $\epsilon_r$ , has been experimentally calculated at a relatively wide range of temperatures and pressures. The relative permittivity is an important indicator of the solvent behavior of water in a variety of biological (cell membrane electrophysiology, intracellular biochemical processes), and industrial (geochemical high temperature, high pressure processes in deep sea vents) settings [Fern97]. There have been many attempts at creating a single function that accurately predicts the relative permittivity of water and steam, the earliest of which was done by Quist and Marshall in 1965 [Quis65], but these have suffered from a lack of experimental values across the entire temperature and pressure range. Recently, Fernandez et al. compiled all of the experimentally available data for the relative permittivity of water and steam in a single database [Fern95]. Furthermore, Fernandez et al. evaluated the methods used to experimentally derive the relative permittivity and chose a subset of the total data set that was the most accurate and that should be used in data correlation. Fernandez et al. proposed a new formulation in [Fern97] that used this subset and approximated the relative permittivity very well across the entire temperature and pressure range.

Our proposal is that in order to more accurately model the behavior of the relative permittivity of water across all temperature and pressure values, two formulations should be created, so that each may be applied in a distinct thermodynamic region. In our approach, two functions are evolved that separately approximate the relative permittivity of water and steam in two thermodynamically distinct regions. These two functions

collectively approximate the relative permittivity of water across the entire range of temperature and pressure values. The accuracy of these two functions is evaluated by comparing their values for the relative permittivity with the values obtained using the latest formulation of Fernandez et al., against the subset of dielectric constant values that Fernandez et al. chose for data correlation mentioned earlier. Any differences between the functional forms of the two evolved functions are explained and ideas for future work regarding a more accurate formulation are offered.

## II. Background: The Static Dielectric Constant

The static dielectric constant (hereon relative permittivity) of a substance,  $\epsilon_r$ , is roughly defined as the ability of a substance to transmit or allow the existence of an electric field. More formally, the relative permittivity of a substance,  $\epsilon_r$ , is the ratio of the static permittivity of the substance,  $\epsilon_s$ , to the static permittivity of a vacuum,  $\epsilon_0$  [Fern95]. The relative permittivity of a substance is used for practical purposes in the design of capacitors. The behavior of the relative permittivity of water is related to its physical state (as a liquid or as steam), temperature, and pressure. This allows the entire range of temperatures and pressures to be divided into 4 regions, A,B,C, and D. Region A is the normal liquid water state between the normal freezing and boiling points (~273K to ~373K). Region B refers to water along the vapor-liquid phase boundary. Region C is the region above 373.15K. At lower pressures and temperatures within region C, water is in the normal gas (steam) phase. At higher pressures and temperatures in this region, water becomes a supercritical fluid, that is, water recondenses back into a liquid state, but exhibits the properties of both a liquid and gas. Finally, region D refers to super cooled water (water below the normal freezing point of 273.15K at the standard pressure of ~.1MPa).

The behavior of the relative permittivity exhibits discontinuities along the liquid-vapor phase boundary (region B) and in the supercritical part of the region above the normal boiling point (region C), with very small changes in the temperature and pressure causing very large changes in density and in the value of the relative permittivity [Harv06]. As a result, theoretical formulations for calculating the relative permittivity of water have mainly focused on a narrow range of temperatures (~270K to ~315K) and pressures (~.1MPa to 100MPa) below the phase boundary [Fern95]. Furthermore, data points along the phase boundary (region B) are sparse and thus have not figured in any data-driven correlations. The most current formulation for approximating the relative permittivity across the entire range of experimental temperatures and pressures may be found in [Fern97]. Fernandez et al.'s formulation uses an extensive adaptive regression algorithm to create an appropriate function taking a wide variety of domain specific thermodynamic values (including first, second, and third derivatives of the temperature and pressure inputs with respect to each other) into account. The final function uses 5 adjustable parameters and a total of 25 constants and domain specific non-adjustable parameters and approximates well across the entire range of experimentally available values.

### **III. Background: Genetic Programming**

What follows is a very brief summary of the genetic programming technique, for further explanation and clarification see [Koza92]. Genetic Programming (GP) is the evolutionary computing technique that attempts to evolve computer programs using a tree based representation scheme and GP-specific modified versions of the traditional evolutionary operators of crossover and mutation [Ghan03]. This technique attempts to evolve an executable computer program that solves a specific user-defined problem from a set of functions, which are individual processes that manipulate and convert data elements, and terminals, which are the data elements themselves. The GP approach involves determining a set of functions and terminals to be used in solving the problem, defining a fitness measure by which individual programs may be evaluated regarding the extent to which they may solve the specified problem, setting the specific parameters and operator probabilities that are involved in program tree generation (crossover and mutation probabilities, initial tree depth limit, maximum tree length, etc.), and developing a set of rules to determine when to end a specific GP run (whether after a certain number of generations have elapsed, or after an individual program with a desired fitness threshold has been found).

The genetic operators of crossover and mutation, as well as the way in which individuals are ranked according to their fitness level are modified from the GA approach (described in detail in [Holl92]) to suit the GP technique. Crossover occurs by selecting two nodes on different parent trees and then swapping all of the children of the selected nodes (as well as the selected nodes themselves) between the two individuals. Mutation, on the other hand, involves selecting a node at which mutation will occur, deleting all of the nodes that are children of the selected node, and then generating a random tree with this node as its root. The fitness evaluation and ranking method in GP is slightly different from the classic GA approach (where fitness maximization is standard) in the fact that the highest ranking individual programs in GP have the lowest fitness values (in effect, a minimization problem). Thus, GP attempts to find a program with the globally minimal fitness value in the search space of all possible programs that may be created using the function and terminal sets used in the problem, to the tree depth or program length specified in the GP setup.

### **IV. Experimental Set-Up**

In our approach, a variety of different function and terminal sets were explored in an effort to evolve two functions that could model the relative permittivity of water as a function of pressure, temperature, and density. Initially, a continuous function was evolved to approximate the relative permittivity of water across all regions, but the results of this function were highly unsatisfactory because the function could not accurately model the behavior of the relative permittivity of water in regions where discontinuities in the relative permittivity were observed (regions B and C, as described earlier). As a result, evolving 2 different functions, one specific to regions A and D, the other specific to region C, became the most logical next step in function development. Unfortunately, no empirical temperature/pressure data for region B (along the phase boundary) is

currently available [Fern95], and thus a function approximating the dielectric constant in region B was not evolved.

The functions for regions A, C, and D were evolved using data sets taken from [Fern95] and were then compared to relative permittivity values calculated with the same input values (taken from the same data sets) using the newest formulation for dielectric constant prediction, found in [Fern97]. These data sets were compiled from all previous experimentally available data, and were then corrected by Fernandez et al. to coincide with the most recent internationally accepted temperature scale, ITS-90. In most cases, values were provided for the temperature (in degrees Kelvin, or K), pressure (in megapascals, or MPa), and the corresponding dielectric constant. However, in some cases, temperature/density/dielectric constant values were given instead of temperature/pressure/dielectric constant values. In these circumstances, density values were converted into their corresponding pressures, and pressure values were converted to their corresponding densities using the IAPWS-95 formulation for the equation of state of water found in [Wagn02]. With this completed, the final data set uniformly represented the dielectric constant at every temperature, pressure, and density value that was experimentally available.

Both functions were evolved by generating a population of possible functions (represented as trees) as with standard genetic programming implementations. Each candidate function's fitness was taken to be the sum of the absolute values of the difference between the calculated and the experimentally measured value for the relative permittivity at every input value in the corresponding data set. The combination of input values for each function (that is, what combination of the three possible adjustable inputs was to be used) was determined by the GP module. The population of possible functions was then evolved with a variety of crossover/mutation probabilities and function sets. The data set of experimentally calculated relative permittivity values used to create a function for regions A and D consisted of 291 data points. The data set used to create the function for the one-phase supercritical region (region C) consisted of 353 data points. These data sets include all of the data points (644 total data points) that Fernandez et al. recommend for data correlations [Fern95]. The two evolved functions with the lowest sum of absolute errors across the data points that were found were used as the final equations for approximating the dielectric constant across the three regions.

During any given GP run, all function and terminal sets used during function evolution always included addition, subtraction, multiplication, and division as function operators, and temperature,  $T_k$ , pressure,  $p$ , and density,  $\rho$ , as terminal values. All runs also used a population of 10 random floating-point constants in the range between 0 and 1, which were generated at runtime. Other function operators ( $\sin()$ ,  $\cos()$ ,  $\ln()$ ,  $\log_{10}$ ,  $\log_2$ , and  $x^y$ ) and terminal operators (Avogadro's number,  $N_A$ , permittivity of free space,  $\epsilon_0$ , elementary charge,  $e$ , Boltzmann's constant,  $k$ , molar mass of water,  $M_w$ , mean molecular polarizability of water,  $\alpha$ , the dipole moment of water,  $\mu$ ) were also used in certain GP runs. The aforementioned terminal operators are provided in table 1. A range of crossover probabilities (between .5 and 1.0, in increments of .05) and mutation probabilities (between 0 and .5, in increments of .05) were explored for all combinations of function and terminal sets. Each combination of parameter settings was implemented in 10 GP runs, each on a population of one million individuals that were evolved for 200

generations. The function length of any individual solution (a tree representing a given candidate function) never exceeded 50 functional units (where a functional unit is taken to be a single operator from the function set or a terminal value from the terminal set), as maintaining the readability of any given evolved function was a priority.

## V. Results

Both of the two optimal functions that were evolved were found during a run that used multiplication, division, subtraction, and addition as operators in the function set and temperature, pressure, and the molar mass of water as terminal operators (with the 10 additional random ephemeral constants described earlier). In addition to the above terminals, the function evolved for region C used density,  $\rho$ , Avogadro's number,  $N_A$ , and Boltzmann's constant,  $k$ , as terminal operators. Both optimal function runs used a probability of crossover of 0.7 and a probability of mutation of 0.05. These functions (simplified with all redundancies eliminated), along with Fernandez et. al's formulation, follow:

$$\varepsilon = \varepsilon_{r,AD} \quad \text{when } T_k \leq 373.15,$$

$$\varepsilon_{r,C} \quad \text{elsewhere}$$

### Evolved Eq. 1.1, regions A, D-

$$\varepsilon_{r,AD} = \frac{2.5203 \cdot 10^4}{T_k} - \frac{M_w^3 p^2}{0.5865} - 0.08133T_k + 0.0355p + 18.08$$

### Evolved Eq. 1.2, region C-

$$\varepsilon_{r,C} = \frac{0.727539 \left( \frac{N_A}{k} \rho + M_w \rho^2 \right)}{0.97121T_k - 88.4082} - \frac{\rho p + p^2}{T_k^2} + 0.971213$$

### Fernandez' formulation:

$$\varepsilon_r = \frac{1 + 5A + 5B + \sqrt{9 + 2A + 18B + A^2 + 10AB + 9B^2}}{4 - 4B}$$

where  $A$  and  $B$  are given by

$$A = \frac{N_A \mu^2 \rho g}{\varepsilon_0 k T_k}$$

$$B = \frac{N_A \alpha}{3\varepsilon_0} \rho$$

and where  $g$  is given by

$$g = 1 + \sum_{k=1}^{11} N_k \left(\frac{\rho}{\rho_c}\right)^{i_k} \left(\frac{T_c}{T}\right)^{j_k} + N_{12} \left(\frac{\rho}{\rho_c}\right) \left(\frac{T}{228K} - 1\right)^{-q}$$

with  $\rho_c = \frac{322}{M_w}$  and  $T_c = 647.096K$  and values for  $N_k, i_k, j_k$ , and  $q$  given in table 2.

The results of applying these functions to their respective partitions of the total data set are found in table 3. The evolved functions shown above are significantly smaller than the formulation developed by Fernandez et al. and use at most three adjustable parameters (temperature, pressure, and density), three non-adjustable domain specific parameters (Avogadro's number, Boltzmann's constant, and the molar mass of water), and three of the ten possible random ephemeral constants that were available during function evolution. No domain-specific knowledge (aside from the data sets themselves) was applied to the formulation of the functions. Furthermore, the evolved functions selected different terminal values for both regions, so that the region C function uses density as an input value along with temperature and pressure, whereas the region A and D function uses temperature and pressure exclusively. This is telling because density is a much more relevant predictive parameter (varying discontinuously along with the relative permittivity while temperature and pressure monotonically increase) for the relative permittivity in the single phase and super critical region (region C) than in regions A and D. The fact that the GP approach was able to selectively choose the relevant parameters for each region is notable and significant.

As can be seen from table 3, both evolved functions outperformed Fernandez et al.'s formulation across both thermodynamic regions. For regions A and D the evolved function outperformed Fernandez et al.'s formulation strictly because of one data point value (notably, a data point that occurred immediately preceding the phase boundary around 373.15K). At this temperature, Fernandez et al.'s formulation may have rounded the temperature input parameter (at 373.147K) up, causing a very sharp discontinuous drop in the calculated relative permittivity value. In region C, the evolved function consistently outperformed Fernandez et al.'s formulation, leading to an improvement in calculation accuracy across the entire range of experimentally available relative permittivity values.

## VI. Conclusions and Future Work:

Two functions that approximate the relative permittivity of water and steam at a variety of temperatures and pressures have been proposed. These functions were evolved using the GP technique with a specific function and terminal set, and their accuracy has been compared to that achieved by Fernandez et al.'s most recent formulation. The evolved functions approximate the relative permittivity of water and steam for a wide range of temperature and pressure values quite well, improving on Fernandez et al.'s formulation across the entire experimentally available temperature and pressure range while being much simpler computationally. Further refinements to create more accurate

approximations of the relative permittivity of water and steam will include creating an evolved function that can be used across all thermodynamically distinct temperature and pressure regions. This can be done when experimental values for the temperature, pressure, and relative permittivity along the phase boundary and more values in the supercritical region are obtained. A refined fitness function that takes more than the absolute distance between expected and calculated values may also prove useful in creating a new formulation. However, significant improvements to the evolution of an appropriate function will most surely come from an increase in experimentally verifiable values for the relative permittivity, and thus any new accurate data that may be found should be used to refine the current formulation.

## VII. Tables and Figures

*Table 1. Constants used in the relative permittivity formulation*

Parameter	Value
Permittivity of free space, $\epsilon_0$	$[4 * 10^{-7} \pi (299792458)^2]^{-1} C^2 J^{-1} m^{-1}$
Elementary charge, $e$	$1.60217733 * 10^{-19} C$
Boltzmann's constant, $k$	$1.380658 * 10^{-23} JK^{-1}$
Avogadro's number, $N_A$	$6.0221367 * 10^{23} mol^{-1}$
Molar mass of water, $M_w$	$0.018015268 kg * mol^{-1}$
Mean molecular polarizability of water, $\alpha$	$1.636 * 10^{-40} C^2 J^{-1} m^{-2}$
Dipole moment of water, $\mu$	$6.138 * 10^{-30} Cm$

*Table 2. Coefficients  $N_k$ , and exponents  $i_k$ ,  $j_k$ , and  $q$  of the equation for  $g$*

$k$	$N_k$	$i_k$	$j_k$
1	0.978224486826	1	0.25
2	-0.957771379375	1	1
3	0.237511794148	1	2.5
4	0.714692244396	2	1.5
5	-0.298217036956	3	1.5
6	-0.108863472196	3	2.5
7	$0.949327488264 * 10^{-1}$	4	2
8	$-0.980469816509 * 10^{-2}$	5	2
9	$0.165167634970 * 10^{-4}$	6	5
10	$0.937359795772 * 10^{-4}$	7	0.5
11	$-0.123179218720 * 10^{-9}$	10	10
12	$0.196096504426 * 10^{-2}$		$q=1.2$

Table 3. Results and numeric comparison

	REGION C Evolved GP		REGION A,D Evolved GP	
	result	Fernandez	result	Fernandez
Sum Absolute Difference	57.35	69.15	44.85	80.58
Mean Absolute Difference	0.16	0.20	0.15	0.27
Standard Deviation				
Absolute Difference	0.25	0.25	0.26	3.18
Sum Squared Difference	30.83	35	26.9	2991.42
Mean Squared Difference	0.09	0.1	0.09	10.18
Standard Deviation				
Squared Difference	0.75	0.70	0.41	173.92
Minimum Absolute Difference	0	0.00	0.001	0
Maximum Absolute Difference	3.76	3.6	1.97	54.61
# Data Points Absolute Difference formulation < Absolute Difference Fernandez	178		142	
Percentage of Total Data Points better than Fernandez	50.20%		48.80%	
Total Data Points	353	353	291	291

## VIII. References

- [Fern95] Fernandez, D.P., Y. Mulev, A.R.H. Goodwin, and J.M.H. Levelt-Sengers. 1995. A Database for the Static Dielectric Constant of Water and Steam. *Journal of Physical and Chemical Reference Data* 24(1): 33-69.
- [Fern97] Fernandez, D.P., A.R.H. Goodwin, E.W. Lemmon, J.M.H. Levelt-Sengers, and R.C. Williams. A Formulation for the Static Permittivity of Water and Steam at Temperatures from 238K to 873K at Pressures up to 1200MPa, Including Derivatives and Debye-Huckel Coefficients. *Journal of Physical and Chemical Reference Data* 26(4): 1125-1166.
- [Ghan03] Ghanea-Hercock, R. 2003. *Applied Evolutionary Algorithms in Java*. New York, NY: Springer-Verlag.
- [Harv06] Harvey, Allan. Personal communication.
- [Holl92] Holland, J. 1992. *Adaptation in Natural and Artificial Systems: 2<sup>nd</sup> Edition*. Cambridge, MA: MIT Press.
- [Koza92] Koza, J.R. 1992. *Genetic Programming*. Cambridge, MA: MIT Press.

- [Quis65] Quist, A.S., and W.L. Marshall. 1965. Estimation of the Dielectric Constant of Water to 800° . *Journal of Physical Chemistry* 9: 3165.
- [Wagn02] Wagner, W and Pruss, A. 2002. The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use. *Journal of Physical and Chemical Reference Data* 31(2): 387-535.