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

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

Abstract

The relative permittivity (or static dielectric constant) of water and steam has been experimentally calculated at a relatively wide range of temperatures and pressures. A single function for predicting the relative permittivity of water and steam in three distinct thermodynamic regions is 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 this function is evaluated by comparing the values for the relative permittivity calculated using the evolved function and the values calculated using the latest formulation of Fernandez et al., found in [Fern97] to the aforementioned data set. In all regions, the newly evolved function outperforms 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

1. Introduction

The relative permittivity (or static dielectric constant) of water and steam, ε_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 geophysical/industrial (geochemical high temperature, high pressure processes in deep sea vents and in industrial processing plants) settings [Fern97]. Many prior researchers have attempted to create a single function that accurately predicts the relative permittivity of water and steam, the earliest of which being Quist and Marshall's 1965 formulation [Quis65]. However, these attempts have suffered from a lack of experimental values across the entire temperature and pressure range, and thus have only been able to approximate the relative permittivity of water with minimal uncertainty over a small range of temperatures and pressures. 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 was recommended for use in data correlation. Fernandez et al. proposed a new formulation in [Fern97] that used this subset to generate a statistical regression function that approximated the relative permittivity fairly well across the entire experimentally available temperature and pressure range.

In an earlier paper [Foge07], we proposed two individual functions evolved using genetic programming that divided the entire data set recommended for data correlation by Fernandez et al. into two distinct thermodynamic regions, with each equation applied to the temperature and pressure range specific to the given thermodynamic region. Although that proposed formulation

outperformed Fernandez et al.'s formulation across the entire range of data values in both thermodynamic regions, a formulation that utilizes a single equation to approximate the relative permittivity across the entire range of experimental values would seem both more natural and appropriate, and is an important goal for researchers in this area. It was hoped that an increase in the size of the evolving population of programs coupled with an increase in the maximum size any individual program could be would allow for the discovery of just such an equation. In the current approach, such an equation has been evolved and closely approximates the relative permittivity of water across the entire range of experimentally verified temperature and pressure values. The accuracy of this function is evaluated by comparing its output value for the relative permittivity of water at a given temperature and pressure with the output relative permittivity value 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.

2. The Static Dielectric Constant

The static dielectric constant (hereon relative permittivity) of a substance, ε_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, ε_r, is the ratio of the static permittivity of the substance, ε_s, to the static permittivity of a vacuum, ε_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 or phase (as a liquid or as steam), temperature, and pressure. Experimentally verified relative permittivity values for water in its solid state (as ice) at temperatures as low as 190K ($-83°C$) exist [Mats96], however this data did not include corresponding pressure values for any of the measurements, and as a result, could not be used. Water, in its liquid or gas (steam) state can exist within a large range of temperatures and pressures, and this range has been traditionally 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) at pressures up to 1000MPa. Region B refers to water along the liquid-vapor phase boundary. Region C is the region with a temperature above 373.15K. At lower pressures and temperatures within region C, water is in the normal vapor (steam) state. At higher pressures and temperatures in this region, water becomes a supercritical fluid, that is, water ceases to behave as if it were in either the liquid or gas state, but rather exhibits a combination of the thermodynamic properties attributable to both liquids and gases. Finally, region D refers to super cooled water (water that exists in the liquid state 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), although numerous, have not had their pressure values recorded, and thus have not figured in any data-driven correlations that correct for pressure differences. The most current formulation for approximating the relative permittivity across the entire range of experimental temperatures and pressures may be found in [Fern97] and is also reproduced in the results section. 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.

3. Evolution and Genetic Programming

Genetic Programming (GP) may be seen as an abstract algorithmic implementation broadly inspired by the main principles of Darwin's theory of evolution by means of natural selection. Roughly, Darwinian evolutionary theory involves populations of interbreeding organisms (species) competing for environmental resources over time. Species share genetic material by interbreeding, and random mutations occur to members of the species that may either hinder or further their reproductive success. As the members of a given species breed with each other over time, characteristics beneficial for the species' survival propagate throughout the population, while those characteristics that are detrimental to the survival of the species do not get expressed in the population. That is, individuals with characteristics that favor their survival within the given environment tend to propagate, whereas individuals not possessing those characteristics in the environment (or those that exhibit detrimental characteristics) tend to die out.

GP applies the broad tenets of Darwinian evolutionary theory within a heuristic framework that attempts to create automatically generated programs that evolve to optimally solve user-defined problems [Koza92]. GP is an extension of the evolutionary computational approach known as genetic algorithms (GA) first pioneered by John Holland [Holl92]. Within the GP framework, a population of candidate solutions, each represented as an executable computer program of some finite length (an individual of a given population), evolves in response to some problem to be solved (the environmental conditions) [Koza92]. Each GP individual/candidate program within the population is given a fitness value that is the output of a function (the fitness function) that determines the appropriateness or optimality of the program output (individual behavior) when given the user-defined problem (the environmental conditions). This allows each individual within the GP population to be measured against every other individual, whether the individual solves the problem (optimally responds to the environment) or not. Once all of the individuals within a population have been assigned a fitness value, certain individuals are probabilistically chosen to recombine and create offspring based on their fitness values, so that individuals with higher fitness values tend to be chosen more frequently for recombination. During recombination two unique individuals are chosen to represent the parents, and may stochastically recombine to generate two offspring. Occasionally, however, (because recombination is probabilistic and does not always occur) they do not recombine and remain unchanged as offspring. After every recombination event, an offspring individual may be mutated with some small probability. The series of steps from initial population generation, parent selection, recombination, and mutation of offspring constitutes a generation of the GP run. At the start of every generation, newly created individuals in the population are evaluated by the fitness function and assigned a fitness value. The GP run continues in this manner (after the generation of the initial population, only fitness value assignment, parent selection, recombination, and mutation of offspring occur) until some stopping criteria (such as the creation of an individual with either some given minimum or maximum fitness value, or one that adequately solves the problem at hand) has been reached.

Each GP individual uses a tree-based representation scheme, where the tree completely represents a given program. Nodes for the GP program tree either come from the terminal set or the function set (both predefined by the individual implementing the GP search).

The terminal set completely defines the kinds of inputs the given program can use to solve the problem. The members of the terminal set can only occur as leaf nodes within the program tree (that is, nodes that have no children). The function set defines the kinds of transformations that are permissible given any of the elements in the terminal set or any of the other elements within the function set as arguments to each of the elements within the function set. Thus, the members of the function set may only occur as the internal nodes of a GP-generated program tree (nodes with at least one child node). These restrictions amount to the fact that the union of the function and terminal sets of a GP implementation must possess the property of closure (where closure is defined as the ability to have any composition of functions and terminals produce an executable computer program) [Ghan03]. The program trees generated using GP do not have to be standard binary trees (trees where every node is either a leaf node, or has a maximum of two child nodes), as the experimenter may define a function operator within the function set that takes more than two arguments. Initially, GP individuals are randomly generated through a stochastic tree-building process where each node in the tree is chosen to be a random member of either the function or terminal sets. Traditionally, GP candidate programs are initially generated either strictly to some maximum initial tree depth limit (where all nodes up to the maximum initial tree depth are chosen stochastically exclusively from the function set and all nodes at the maximum initial depth limit are chosen exclusively from the terminal set), or until all of the branches of the tree have either gone to the maximum initial depth or have ended in terminal nodes before the maximum initial tree depth has been reached.

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 are 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.

Ultimately, 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 and assigned a fitness value, 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 or stopping criteria 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).

4. Experimental Set-Up

In our approach, a variety of different function and terminal sets were explored in an effort to evolve a single function that could model the relative permittivity of water as a function of pressure, temperature, and density in thermodynamic regions A, C, and D of the temperature-pressure phase space. Recall, no empirical temperature *and 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 function for regions A, C, and D was evolved using data sets taken from [Fern95] and was then compared to relative permittivity values calculated with the same input temperature/pressure/density 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 (as of December 2006).

The function was 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 errors 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 the function consisted of 644 data points, which represent the complete dataset that Fernandez et al. recommend for data correlations [Fern95]. The function with the lowest sum of absolute errors across the data points that was found after all runs had been completed was chosen as the final formulation.

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, ρ, as terminal values. In cases where a generated function divided a value by zero, the zero-generating term was replaced by 0.00001. All runs used a population of 100 random floating-point constants in the range between 0 and 1, which were generated at runtime. These constants would function as additional terminal values for the genetic program to use during function creation. 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, ε_0, elementary charge, e, Boltzmann's constant, k, molar mass of water, M_w, mean molecular polarizability of water, α, the dipole moment of water, μ) were also used in certain GP runs. The aforementioned terminal operators are provided in table 1. The function length of any individual solution (a tree representing a given candidate function) never exceeded 100 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). The large size of the function and terminal sets causes the size of the search space (representing all of the possible unique programs of length 100 or less that can be generated from the function and terminal sets) to be enormous (easily more than a googol). As a result, each GP run was done on a population of one and a half million individuals that were evolved for 200 generations. This was done in hopes that the GP

implementation would uniformly sample as much of the search space as possible in its effort to find a suitable function within a reasonable time. 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 unique parameter settings was implemented in 10 GP runs, after which the function with the lowest total absolute error was chosen.

5. Results

The optimal function that was evolved was found during a run that used multiplication, division, subtraction, and addition as operators in the function set and temperature, pressure, and density as terminal operators (with the 100 additional random ephemeral constants described earlier). The optimal function run used a probability of crossover of 0.9 and a probability of mutation of 0.05. The final evolved function, along with Fernandez et. al's formulation, are listed below (next page).

The results of applying the GP-evolved function and Fernandez et al.'s formulation to the total data set are found in table 3. Our evolved (non-simplified) function is significantly smaller (31 terms versus 112 terms) than the formulation developed by Fernandez et al. and uses only three adjustable parameters (temperature, pressure, and density), zero non-adjustable domain specific parameters, and only three of the one hundred possible random ephemeral constants that were available during function evolution. No domain-specific knowledge (aside from the data sets themselves) was applied to the formation of the function.

As can be seen from table 3, the evolved function outperformed Fernandez et al.'s formulation in all collected statistical categories except the minimum absolute difference, where both functions had at least one data point where very marginal absolute error (<0.01) existed.

6. Conclusions and Future Work:

A new function that approximates the relative permittivity of water and steam at a variety of temperatures and pressures has been developed. This function was evolved using the GP technique with a specific function and terminal set, and its accuracy has been compared to that achieved by Fernandez et al.'s most recent formulation. The evolved function approximates the relative permittivity of water and steam for a wide range of temperature and pressure values extremely well, improving on Fernandez et al.'s formulation across the entire experimentally available temperature and pressure range while being 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, including regions where water is in the solid phase, or where a phase boundary exists. This can be done when experimental values for the temperature, pressure, and relative permittivity in these regions are obtained. A refined fitness function that takes more than the absolute difference between expected and calculated values may also prove useful in creating a new, more accurate formulation. Introducing a penalty for very large and difficult to read formulations may also help in finding a function that is both compact and generalizes well across the entire thermodynamic space. 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.

GP-Evolved formulation

$$\varepsilon_r = \frac{\dfrac{\rho^2+\rho^3}{-.02036T_k\rho+.0864\rho^2+.1194T_k p} + \dfrac{(-6.75862p^2+.313T_k p-{T_k}^2)(1+\rho)}{T_k}}{55.474T_k p+55.55p^2-.076T_k p\rho+.016p+.016p\rho+\rho+\rho^2} + \frac{\rho^2}{.03264T_k(T_k+\rho)} - \frac{\dfrac{{T_k}^2}{\rho^2}+2.617p-1.617\rho+1.617T_k}{T_k+.0486-p+\rho}$$

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

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

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

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

7. Tables

Table 1. Constants used in the relative permittivity formulation

Parameter	Value
Permittivity of free space, ε_0	$[4*10^{-7}\pi(299792458)^2]^{-1}C^2J^{-1}m^{-1}$
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	$0.018015268kg*mol^{-1}$
Mean molecular polarizability of water, α	$1.636*10^{-40}C^2J^{-1}m^{-2}$
Dipole moment of water, μ	$6.138*10^{-30}Cm$

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

	Evolved GP result	Fernandez
Sum Absolute Difference	103.12	149.73
Mean Absolute Difference	0.16	0.23
Standard Deviation Absolute Difference	0.25	2.15
Sum Squared Difference	55.8	3026.43
Mean Squared Difference	0.09	4.68
Standard Deviation Squared Difference	0.54	117.24
Minimum Absolute Difference	0	0
Maximum Absolute Difference	3.55	54.61
# Data Points Absolute Difference formulation < Absolute Difference Fernandez	331	
Percentage of Total Data Points better than Fernandez	51.08%	
Total Data Points	644	

8. 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.

[Foge07] Fogelson, S., & Potter, W. 2007. A GP-Evolved Formulation for the Relative Permittivity of Water and Steam. To appear in *The Proceedings of the International Conference of Industrial and Engineering Application of Artificial Intelligence and Expert Systems, IEA-AIE'07.*

[Ghan03] Ghanea-Hercock, R. 2003. *Applied Evolutionary Algorithms in Java.* New York, NY: Springer-Verlag.

[Harv06] Harvey, Allan. NIST, Personal communication.

[Holl92] Holland, J. 1992. *Adaptation in Natural and Artificial Systems: 2nd Edition.* Cambridge, MA: MIT Press.

[Koza92] Koza, J.R. 1992. *Genetic Programming.* Cambridge, MA: MIT Press.

[Mats96] Matsuoka,T., Fujita, S., Mae, S. 1996. Effect of temperature on dielectric properties of ice in the range 5-39 GHz. *Journal of Applied Physics* 80(10): 5884-5890.

[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.