Local Basis Function Representations of Thermodynamic Surfaces: Water at High Pressure and Temperature as an Example

J. Michael Brown
Department of Earth and Space Sciences
Box 35-1310
University of Washington
Seattle, WA 98195
brown@ess.washington.edu
206-616-6058

Abstract

A flexible numerical framework using local basis function is developed in order to better represent thermodynamic properties of fluids as a function of pressure temperature and composition. A new equation of state for water to 100 GPa and 10,000 K is presented. Conventional equations of state are typically based on complex sets of global basis functions that are unique to each application. The use of series expansions in positive and negative powers of the independent variables (with additional exponential/log/trig. factors in some cases) is common and recent accurate equations of state are assembled from collections of such arbitrary terms. These individually crafted representations do not always fit data within uncertainties, can be difficult to implement, and are not easily modified to account for newer data. Multivariate (tensor) b splines overcome these shortcomings. The underlying basis functions are local, orthonormal and complete. Data can be represented to arbitrary precision and the relationships between model parameters and observational constraints are uncomplicated. The fittings of new data in more extended regimes of the thermodynamic space do not require modifying parameters in previously constrained regions. Models are transportable and easily implemented since robust and efficient standardized routines for evaluation of b splines are available in modern numerical environments. Because a local basis function representation can flexibly fit any surface, the articulation of *a priori* constraints (associated with expected physical and chemical behavior) is appropriately necessary and explicit during model construction. Basic tools are demonstrated in applications to thermodynamic properties of water where the modified equation of state matches a prior formulation below 1 GPa while providing an excellent fit to measurements at much higher pressures.

Keywords

b spline; local basis functions; equation of state; fluid thermodynamics; regularized parameter estimation; water

Highlights

- Fluid thermodynamics are represented using local basis functions (b splines)
- Representations are easily shared using compact files and standardized algorithms
- No material-specific coding is needed
- Regularization allows sparse and scattered data to be represented
- A modified equation of state for water is determined

1. Introduction

Comprehensive thermodynamic representations of fluids under extreme conditions of pressure and temperature are commonly parameterized with individually crafted combinations of arbitrary basis functions (see the overview given by Span [1]). Although able to adequately fit extant data, such representations are difficult to modify in the face of new results in extended regimes of pressure-temperature-composition. For example, no single available model for water is capable of successfully predicting its properties over the full span of pressures and temperatures. The sound speed measurements in water to high pressures [2] provided evidence that the internationally recognized equation of state for water, IAPWS-95 [3], while accurately predicting properties in a low-pressure regime, is systematically in error beyond about 1 GPa. The complexity of the representation underlying IAPWS-95 precludes simple modifications to account for the new measurements. The focus here is on pure water. However, the methods described are extendable to multi-component systems. The accurate descriptions of thermodynamic properties of fluid mixtures (whether electrolytic solutions or water with non-polar fluids) also remain problematic.

The goal is to provide a path by which high-pressure measurements can be more readily converted into accurate and comprehensive thermodynamic representations in a standardized (universal) electronic format. A framework is described for accurately representing either individual datasets of a particular thermodynamic property (*e.g.* specific heat, density, sound speed) or the comprehensive representation of the underlying energy potential (including Helmholtz, Gibbs, or internal). This framework allows arbitrary precision in fitting data and ongoing modifications to accommodate new information in more extended regimes of pressure, temperature, and composition without changing the representation elsewhere. Furthermore, physical constraints can be used to guide the behavior of the representations in regions of sparse or non-existent experimental constraints.

The methods are based on algorithms that are widely available in numerical libraries. Additional functions needed for thermodynamic applications are provided in the supplemental materials. This simplicity is in contrast to the complex and custom representations that are dominate in the literature.

The benefits of the recommended approach include: (1) the process to create a representation is straightforward and requires no extensive search for an appropriate custom numerical form for each fluid's equation of state. (2) Since the model format is universal and independent of the material being represented, the resultant thermodynamic representations can be easily communicated (shared electronically as a set of model parameters). (3) The accuracy of the fit is by choice during construction rather than being dictated by the form of a particular arbitrary equation. (4) These representations can easily be modified to account for revised information in new regimes of thermodynamic states. (5) The parameters of the representation are in units of the quantity to be represented and are approximately equal to that quantity at each specified state. (6) Numerical evaluations based on this formalism are robust and efficient (*ie.* computationally fast).

The utility of the current approach is illustrated through a re-evaluation of the thermodynamic properties of water in light of high-pressure experimental constraints. IAPWS-95 is the reference equation of state against which modifications can be compared. The methods described are equally applicable for use in multicomponent systems; the numerical framework for that has been developed and will be discussed elsewhere.

1.1 Basis functions; From global to local

Numerical terms, the "basis functions", in equations of state range from simple linear relationships, including the perennially favored series expansion with positive and/or negative powers of independent variables, to non-linear representations containing complex, and sometimes nested, combinations of model parameters and functions, including logs, exponentials and trig functions. Basis functions are "global" when each model parameter contributes everywhere.

Global basis functions may not be orthogonal; properties at any point can result from correlated interactions among the model parameters. Functions may also not be normalized; model parameters can represent complex combinations of units that make unit conversions or code development more difficult. Furthermore, it is not guaranteed that a chosen combination of global basis function is complete. The representation may not be able to fit data to arbitrary precision (*ie.* within experimental uncertainties). Because each basis functions contribute everywhere, adding new observations requires a refitting of all model parameters and frequently a re-selection of appropriate basis functions.

The varieties, differences, and complexities found in implementations of global basis functions are apparent in the literature. For example, Phutela and Pitzer [4] and separately Archer and Rard [5] investigated the behavior of aqueous $MgSO_4$ at 1 bar pressure. Data, some of it in common between the studies, were analyzed within the context of the same electrolyte chemistry theory. Different sets of nested arbitrary functions were chosen in each investigation. The numerical implementation of these models is time intensive, and applicable for only one solution chemistry. The extension to high pressure is not straightforward. In an effort to account for the high-pressure behavior, Marion *et al.* [6] linearized the thermodynamic derivatives with respect to pressure and created additional arbitrary power law representations as a function of pressure and temperature. This could only account for high-pressure behavior over a relatively small regime of pressure [7]. In another study, the high-pressure equation of state for aqueous NaCl [8] was formulated using complicated non-linear functions (including logs and exponentials) of temperature and pressure. IAPWS-95, the widely used equation of state for water, includes 58 terms with over 300 parameters. A critical evaluation or the implementation of any of these models remains challenging. The numerical coding is unique in each case. Sensitivity testing or error analysis is difficult. Extension or revision to account for new results is rare.

The utility of “local” basis functions (numerical terms that have separate regions of influence) has been recognized [1] and IAPWS-95 includes partially localized basis functions, created through multiplication of polynomial terms by a decaying exponential factor. Through a trial and error algorithm to find a basis function set that adequately fits the data, terms were chosen from a large library of possible functions, Wagner and Pruss [3] were able to fit selected data within uncertainties extending from steam to water and from temperatures below freezing to well above the critical point. Extensive data are available to 100 MPa and sparser data to 1 GPa. No effort was made to accurately fit higher-pressure measurements; their representation exhibits systematic misfits in the regime of diamond anvil and shock wave experiments [2]. Within the framework used to construct the IAPWS-95 parameterization, it is not possible to better represent the high-pressure data without undertaking a new search for appropriate basis functions and to refit all low pressure and high-pressure data.

A switch to truly local basis functions allows accurate representation of any thermodynamic surface. A fit of data to arbitrary precision is possible; parameters in one thermodynamic region need not change if new measurements in another regime require modification of the surface. Through use of standardized algorithms, a model can be communicated as an electronic list of parameters that are correctly interpreted in any numerical environment without specialized coding. The arbitrary shape of a surface represented by local basis functions is entirely a result of the fitting process (rather than being an implicit outcome of a chosen collection of global basis functions). However, strong *a priori* constraints, based on known physics or chemistry, can and should guide creation of the representation.

1.2 The Use of b Splines for Thermodynamic Representations

Local basis functions (*LBF*) in the form of “basis function splines” or ‘b splines’, as described by de Boor [9], are polynomials on intervals with matching derivatives at interval boundaries. These representations are characterized by their order (*ie.,* polynomial degree), interval (knot) spacings, and coefficients (model parameters associated with the intervals). The coefficients have units of the quantity being represented and as intervals become small, the coefficients approach the values of the quantity being represented within each interval.

A b spline of order 1 (polynomial degree 0) provides a constant value between gridded entries. Modeled values change discontinuously at interval boundaries. A b spline of order 2 (polynomial degree 1) is equivalent to linear interpolation. Values are continuous at boundaries but higher order derivatives are discontinuous. With further increase of spline order (polynomial degree), the number of continuous derivatives at interval boundaries increases. Further details of b spline basis functions are provided in appendix 1.

Through their construction, b splines basis functions are orthogonal, normalized, and can be complete [9]. Robust and efficient algorithms in standard numerical libraries allow their evaluation. Derivatives and integrals of b splines are themselves b splines. Thus, a b spline provides an analytic functional representation that can be easily integrated or differentiated (to the limit of the chosen order). Importantly, the relationship between b spline coefficients and the represented quantity is linear. Thus, the creation of a b spline representation is accomplished through the numerical solution of a set of linear equations, the “normal equations”.

The communication of an *LBF* model (its transport and implementation) requires only a small electronic file containing the spline orders, knot sequences, and associated coefficients. The requisite subroutines to evaluate splines are found in all modern numerical environments (*e.g.,* FORTRAN, Mathematica, Python, MATLAB). Example MATLAB scripts and functions are provided in the supplemental materials to illustrate the ideas. The conversion of these examples to other languages and environments is straightforward.

2. Determination of thermodynamic surfaces using local basis functions

Here an *LBF* representation is found as solution to the regularized linear problem:

$$\begin{pmatrix} data\ basis\ functions \\ \lambda \quad \begin{pmatrix} regularization \\ basis\ functions \end{pmatrix} \end{pmatrix} (model\ parameters) = \begin{pmatrix} data \\ zeros \end{pmatrix} \tag{1}$$

where the data basis functions in Equation 1 are values of the b spline basis functions evaluated at data sites, model parameters are the b spline coefficients, and "data" on the right side are quantities to be represented. In conventional least squares fitting, more data than model parameters are required. Since, when fitting thermodynamic data, model parameters are likely more numerous than data, additional constraints are required. If the surface is expected to be "smooth", second derivatives of the surface should be small everywhere and regularization, minimally evaluated at every grid point, is composed of second derivative basis functions. The "damping factor", λ, weights how strongly regularization needs to be enforced. The general framework of regularized parameter estimation is described in [10]. Data used in construction of the representation through Equation 1 can be scattered and sparse if the regularization provides an adequate constraint on the form of the representation in regimes of interpolation or extrapolation.

The representation of energy as a function of the independent variables provides a complete equation of state description. Use of the Helmholtz energy (function of temperature and volume) underlies the development of theoretical and empirical efforts including the formulation of IAPWS-95. An equivalent representation using Gibbs energy is a more direct link to experiments and applications since its independent variables, pressure and temperature, are those normally measured. For this reason, the present work focuses on Gibbs energy representations. However, the underlying numerical framework is as easily implemented for any choice of thermodynamic potential (*e.g.,* Helmholtz energy, or internal energy) with their respective independent variables.

The "industrial version", IAPWS-97, [11] is also parameterized as Gibbs energy. However, a complex geometry with three pressure and temperature domains was required with separate representations. The result is less accurate than IAPWS-95 and covers only a limited regime of pressures (<100 MPa) and temperatures (<2273 K). As shown below, an *LBF* representation for Gibbs energy can reproduce IAPWS-95 to parts-per-million accuracy over a far wider range of pressure and temperature.

Table 1 contains a summary of relationships between Gibbs free energy (G) and the standard laboratory observables. Since both specific volumes and specific heats (divided by T) are linear in model parameters for G, Equation 1, cast in the form

$$\begin{pmatrix} basis\ functions\ for\ G \\ \lambda_C \quad basis\ functions\ for\ G_{tt} \\ \lambda_V \quad basis\ functions\ for\ G_p \\ \lambda_R \quad basis\ functions \\ for\ regularization \end{pmatrix} (model\ parameters) = \begin{pmatrix} G\ data \\ \lambda_C \quad -C_p/T\ data \\ \lambda_V \quad V\ data \\ zeros \end{pmatrix}, \qquad 2$$

allows direct solution for the Gibbs energy model parameters on the basis of measured values of volumes and specific heats. The values of the various λ_x in Equation 2 determine the relative weight of each contribution to the overall fit. To the extent that measurements adequately span the modeled range of pressures and temperatures, density and specific heat measurements “regularize” the behavior of the Gibbs energy surface by forcing least squares agreement of both G and derivatives of G. With Gibbs energy specified only at a reference pressure (typically 1 bar), Equation 2 provides a solution by “collocation” [9] to the differential equations relating G to values of specific heat and volume. The bottom entries in Equation 2 can be any choice for derivative basis functions on the left and corresponding zeros for each on the right side. The belief that specific heat and volume are “smooth surfaces” suggests that their second derivatives (G_{4t} and G_{3p}) should be small and λ_R enforces the degree of smoothness. Limiting behavior for thermodynamic properties at extremes of pressure and temperature can also provide regularization of the solution. Because sound speeds involve a non-linear combination of derivatives, their use in determining Gibbs energy requires an iterative approach based on an initial guess for $G(P,T)$ (as shown in Equation A8). Implementation of high-pressure limiting behavior for the pressure derivative of the bulk modulus [12] would also require a non-linear approach.

Since sound speed measurements at high pressure and temperature are more numerous and of higher accuracy than determinations of specific heat or density, an alternate path to determining Gibbs energy is advantageous. Integration from a base pressure of the coupled differential equations of density, sound speed, thermal expansivity, and specific heat [2, 7, 13, 14, 15] provides a path to determine all thermodynamic properties at high pressure.,

$$\frac{\partial \rho}{\partial P}_T = \frac{1}{u^2} + \frac{\alpha^2 T}{C_p} \tag{3}$$

$$\frac{\partial C_p}{\partial P}_T = -\mathrm{T}\frac{\partial^2 V}{\partial T^2}_P \tag{4}$$

With knowledge of *G* at the base of isotherms (pressure P_o for temperatures T), *G(P,T)* is determined through the additional integration:

$$G(P,T) = G(P_o,T) + \int_{P_o}^{P} V(P,T)dP \tag{5}$$

The first term on the right side of Equation 3, the adiabatic derivative of density with respect to pressure, is determined entirely by sound speed measurements. The second term, the adiabatic to isothermal correction, can be comparable in size to the adiabatic contribution at 1 bar but decreases with increasing pressure. In the predictor-corrector approach to solution of Equations 3 and 4, both α and C_p are estimated at an incremented pressure based on the derivatives of volume as a function of temperature initially obtained at a prior pressure point. Newly obtained volumes are then differentiated to improve and correct estimates of the quantities in the adiabatic to isothermal term.

Gibbs energy provided by solution of Equation 5 can then be represented by local basis functions through the solution of Equation 2. By construction, all thermodynamic properties are internally consistent with measured sound speeds.

The use of local basis functions requires care in the conceptualization of a model. An *a priori* analysis of scope and objectives of the mathematical representation is essential. This includes decisions on limits for independent variables in pressures, temperatures, and chemical concentration and the degree of acceptable data misfit. The physical or chemical constraints associated with an acceptable fit must be explicitly considered. For example, the degree of surface "smoothness" and expected asymptotic behavior can, in many cases, be articulated and incorporated into the model. Standard statistical methods to evaluate the accuracy and sensitivity of the model (*eg.* jackknife or bootstrap analysis [10]) can be readily undertaken. The steps creating an *LBF* representation are illustrated in the following sections.

3. Numerical implementation of *LBF* representations

Although all numerical environments have tools suitable for this numerical analysis, here MATLAB® provides a convenient platform to articulate concepts associated with *LBF* representations and is used for the following examples. In particular, the "live script" environment of MATLAB allows a mixture of formatted text, executable code, and embedded figures. These scripts enable users to duplicate all calculations and figures given in this work. The live scripts are also converted to formatted documents to enable "off-line" study of the methods, underlying implementation, and graphical results.

Custom functions that implement *LBF* representations of thermodynamic properties are located in the supplemental materials. The basic MATLAB functions and the specialized functions are described in more detail in Appendix 2. Several key functions are noted here.

The custom function *spdft* solves Equation 1 to find an analytic (*LBF*) representation of surfaces with one, two, or three independent variables. Either scattered data (typical of measurements) or gridded data (typical if a previously determined representation is available) can be used. No units are implicit in its use. The remaining functions require consistent use of standard MKS units with pressures given in MPa. The function *spgft* solves Equation 2 in order to create an *LBF* representation for Gibbs energy as a function of pressure and temperature. The function *eqst* implements a predictor-corrector solution of Equations 3, 4 and 5. Given an *LBF* representation for Gibbs energy, the function *spGval* returns thermodynamic properties, determined from analytically determined derivatives of the energy surface.

3.1. Basic tutorial

Steps in the workflow for local basis function representations include: (1) Decide what minimum order spline is consistent with data (*ie*. what level of differentiation of the representation is required). (2) Create a provisional sequence of "control points" appropriate for the spline order and distribution of data (a set of spline knots associated with the specified control points is automatically constructed within the functions). (3) Undertake several trial fits using damping values differing initially by orders of magnitude to find the appropriate range of damping. (4) To achieve the desired misfit of data, the choice of damping is then further refined. (5) If misfit remains large for even small values of damping, control point spacing can be reduced or spline order can be increased until an adequate fit is found. (6) The "size" of the representation (number of knots and coefficients) can be adjusted (while checking that the representation still satisfies misfit expectations) by exploring what minimum number of control points are required for a given spline order.

A live script, *FitIndependentVariablesExamples* in the supplemental materials, provides a tutorial on the use of local basis functions to fit thermodynamic properties as a function of one, two, or three independent thermodynamic variables (using function *spdft*). The user is encouraged to explore results by changing parameters associated with creation of the representations or to view the (non-executable) document version of the script.

In the first section, synthetic data of one independent variable are considered. This provides an opportunity to gain insight into the use of local basis function fits with the fewest required adjustments. A smooth representation requires a spline order of at least four to provide continuous second derivatives of the associated third degree (cubic) polynomials. The nature of the variation of the data in this example makes typical polynomial fitting difficult (large second derivatives of y for small values x and smaller curvature for large x). When the script is executed, results using different values of the "damping factor", λ, are illustrated (here as Figure 1). A large value of λ forces the fit to be smooth (low second derivatives); this results in a linear least-square representation. With too little damping, the data are over-fit; the representation is free to wiggle as allowed by the knot spacing and spline order. An intermediate value for the damping provides a fit that has an acceptable reduced chi-square (χ^2) misfit near 1. The user can verify that increasing the spline order in this example has little impact on the fit; sufficient regularization insures that second and higher order derivatives are adequately small. The control point intervals can be made arbitrarily small. However, if intervals are too large, the basis functions cannot fit data (the basis is "incomplete") to the required precision.

In the second section, sound speeds for water [15] as a function of pressure and temperature are represented with a surface. An adequate fit is produced with the parameters provided. The script demonstrates how data are input and how vectors of control points are represented. The user can explore variations in damping for each independent variable and what control point spacing is required in order to achieve a fit within experimental uncertainty. A modest extension of the spline surface beyond the data is illustrated. Since, in this example, regularization is based on the second derivatives of the surface, the representation becomes increasingly linear in regions of extrapolation.

In the third section of the script, sound speeds as a function of pressure, temperature, and composition are represented by a surface of three independent variables. The fitting process remains similar to those in the previous examples with fewer independent variables. However, the larger number of possible adjustments (choices for damping, spline order, control point sequences in each of three dimensions) increases the complexity of the analysis.

3.2. Reanalysis of sound speed measurements

The analysis below is undertaken by sequentially executing sections (creating Figures 2, 3, and 4) within the script *WaterSoundSpeedFitting*. The function *spdft* is used to fit sound speed measurements, *eqst* integrates sound speeds to give the full equation of state, and *spgft* creates the local basis function representation.

In this script, a previously published analysis [15] is re-evaluated using local basis functions. The intent is to show that the current methods provide a robust analysis without the need for a specialized determination of custom basis functions. The precise sound speed measurements [15] provide a comprehensive data set for water from 252 K to 474 K and to a maximum pressure of 400 MPa. The 95% confidence intervals (accounting for propagated errors in pressure, temperature, and sound speed determination) range from less than 0.02% at low pressures to as much as 0.04% at extremes of pressure and temperatures.

In Figure 2 (reproducible by executing the first section of *WaterSoundSpeedFitting*) sound speed measurements and the sound speed surface predicted by IAPWS-95 are shown. The distribution of data, the behavior of the surface, and fluid limits (melting phase lines for water ices) are apparent. As a result of freezing, measurements are not possible on a full rectangular grid and the sound speed measurements at high pressure are constrained within a shrinking range of temperatures below 273 K. Notable is the variable curvature of the surface in different regimes of pressure and temperature, which increases the difficulty in creating a numerical representation based on global basis functions.

Following past efforts to accurately represent thermodynamic surfaces [13], in [15] a set of custom polynomial basis functions were found with pressure as a function of sound speed and temperature. This tactic reduces the number of required global polynomial terms needed to achieve an arbitrary level of misfit. Determining the requisite combination of terms, with positive and negative powers, for an acceptable fit is part of that research agenda. The use of local basis functions simplifies this process and, as shown below, allows the data to be more appropriately fit within experimental uncertainty.

In the three panels of Figure 3 (created by executing the second section of *WaterSoundSpeedFitting)* fractional misfits of the measured sound speeds from three representations are shown. On the left side, the systematic deviations of measured sound speeds from IAPWS-95 predictions indicate either that measurements have errors larger than estimated or that IAPWS-95 is not sufficiently accurate to match sound speeds within measurement uncertainty. As expected, in the middle panel, deviations of measured sound speeds from the polynomial parameterization reported in [15] are smaller than those shown on the left side. However, these deviations remain larger than reported experimental uncertainty and exhibit systematic trends. The inability to achieve a better fit is an indication that the chosen numerical form is insufficiently complete. In the panel on the right side of Figure 3 deviations of data from a local basis function representation (using function *spdft*) are shown. Here the damping term λ for second derivatives of the surface with respect to pressure and temperature was adjusted to find an adequately smooth surface able to fit data within reported uncertainty.

The equation of state for water (Gibbs energy representation) is explored in the third section of *WaterSoundSpeedFitting* using the three representations of sound speeds versus pressure and temperature. These are (1) IAPWS-95 sound speeds (here used as a control to validate the accuracy of the numerical methods), (2) sound speeds based on the parameterization given in [15] (to confirm that their analysis is replicated), and (3) sound speeds based on the local basis function representation. Equations 3, 4 and 5 are numerically solved (function *eqst*) and resulting values for G, V and C_p were converted to an *LBF* representation for Gibbs energy (function *spgft*). All thermodynamic properties (densities, specific heats, sound speeds) are then calculated as appropriate derivatives of the Gibbs energy surface (using function *fnGval*).

Deviations of densities and specific heats from IAPWS-95 are shown in Figure 4. As shown in the panels on the left side, the test of the numerical method is successful. Using IAPWS-95 sound speeds as the only high-pressure constraint on the equation of state, IAPWS-95 densities and specific heats are recovered to better than parts per million deviations with slightly larger deviations at the lowest temperatures. In the middle panels, integration of sound speeds based on the previously reported global polynomial representation gives results that agree with that analysis as illustrated in figures 11 and 12 of [15]. Negative deviations of density (associated with positive deviations of sound speeds relative to IAPWS-95) increase with pressure for temperatures above 300 K. At low temperatures, where their polynomial sound speed fit exhibits significant systematic deviations from data and from predictions of IAPWS-95, density and specific heat deviations are larger. In panels on the right side, deviations based on the *LBF* representation of sound speeds are shown. Deviations from IAPWS-95 above 350 K are greater than those in the middle panels as is expected based on the systematic sound speed misfits shown in the middle panel of Figure 3. Below 300 K, the local basis function fit reverses the sign of deviations from IAPWS-95 for both densities and specific heats in comparison with the representation in [15] that did not adequately fit the data in this regime.

In the lower panels of Figure 4, specific heats above 273 K show small deviations from IAPWS-95 despite substantial changes in C_p with pressure. Since the pressure dependence of specific heat is determined as a second derivative of volume with respect to temperature (Equation 4), it is notable that the numerical analysis is fully capable of extracting the correct pressure behavior for specific heat. Although deviations from IAPWS-95 at the lowest temperature are larger than the plotted limits shown in Figure 4, the absolute deviations are not so large compared to experimental uncertainties for volumes and specific heats at low temperature and high pressure.

Given the ease of equation of state construction demonstrated here, a collection of models can be readily created based on differing *a priori* assumptions including: (1)

ensembles of solutions using different damping factors or regularization choices, (2) ensembles based on analysis of many synthetic data sets with specified statistical distributions of errors (a method of bootstrap statistical analysis), (3) ensembles of results using sub-sets of data (jackknife statistics), (4) or through exploration of the impact of systematic errors. Considerations of the spread in predictions under these differing assumptions can serve to better illuminate uncertainties inherent in the thermodynamic analyses.

To summarize the principal ideas articulated in this section, an easily executed protocol allows laboratory measurements of sound speeds to be converted into an equation of state description. The resulting Gibbs energy *LBF* representation is a single numerical entity that provides an analytic description of all thermodynamic properties. This differs from typically reported results where separate parameterizations for sound speeds, densities, and specific heats are frequently given. Implementation of this numerical framework is straightforward and requires no specialized skill in finding appropriate basis functions. When IAPWS-95 derived sound speeds are treated as the only high-pressure "data", the resulting equation of state for water reproduces IAPWS-95 to an accuracy measured in parts per million.

3.3. The representation and modification of IAPWS using local basis functions

Analyses in this section are reproduced by executing the sections of the script *IAPWS_fit_modify.* Since IAPWS-95 is parameterized as the Helmholtz energy (a function of volume and temperature), the determination of properties at a specified pressure requires a computationally slow numerical search. It is advantageous to cast IAPWS-95 as a function of pressure and temperature, enabling faster evaluations, and to modify IAPWS-95 at higher pressures without impacting the accuracy of its predictions in the low-pressure regime where presumably accurate measurements constrain the equation of state. In general, a process is needed to allow assimilation of new observations in revised and comprehensive formulations.

Using function *IAPWS95*, which is a full implementation of the standard IAPWS-95 formulation, Gibbs free energies, specific volumes, and specific heats are determined on a 450 by 400 grid of “control points”. These have a linear spacing of control points in pressure between 1 bar (0.1 MPa) and 100 MPa and are logarithmically distributed to 300 GPa. The temperature control points are logarithmic from 240 K to 20,000 K. Points lying in a low temperature and high-pressure regime (the regime of solid stability where IAPWS-95 does not provide reasonable estimates) are identified and excluded from the analysis and from plots of results. A region around the critical point is also excluded, as are high temperature (>1000 K) properties below 30 MPa. While these are important areas for fluids, the focus here is on the higher-pressure behavior. A separate denser *LBF* representation plus near critical point numerical formulations would allow a representation in that regime to specified accuracy.

Evaluations using *IAPWS95* are computationally slow when pressure is provided as an independent variable. To reduce the time required to execute this script, pre-computed control points are provided in a separate file that is loaded by default. The full calculation can be accomplished by uncommenting a line in the script.

An *LBF* representation for Gibbs energy was determined by applying the function *spgft* to the grid of the IAPWS-95 control points. Since gridded points in the regime of ice are excluded, regularization of the solution is necessary in order to have a numerically stable inversion. The *LBF* representation based on this inversion is included with the supplemental materials and is available through call to the function *fnIAPWSval.* The first section of the script follows steps used in creation of this representation.

In the second section of the script, the accuracy and speed in execution of the local basis function representation for IAPWS-95 is evaluated. Thermodynamic properties at user-selected pressure and temperature points are determined using the exact formulation and the local basis function representation of IAPWS-95. The maximum deviations in density (less than 1 ppm), specific heat (~ppm), and sound speed (~10’s ppm) are adequately small and occur near the high-pressure low-temperature boundary where the IAPWS-95 formalism fails. Thus, the *LBF* representation can replace the full formulation with no functional loss of accuracy. As documented in the script, the *LBF* version allows evaluation of points at specified pressures and temperatures that can be three orders of magnitude faster than execution of the full implementation since no numerical search is required. The size of this local basis function representation (approximately 1 MB) is relatively small. Even smaller models (using fewer control points) could be constructed that reproduce IAPWS-95 to within specified tolerances to suit particular applications that do not need the full accuracy or full range of pressures and temperatures.

In the next sections of the script, modifications of the local basis function representation in one regime without changing the representation in another is illustrated. Abramson and Brown [2] demonstrated that measured sound speeds deviate in an essentially temperature-independent fashion from predictions of IAPWS-95 in a pressure regime beyond 1 GPa. The sound speeds measured by Decremps *et al.* [16] and Ashara *et al.* [17] confirmed a need for high pressure modifications to IAPWS-95. Furthermore, as previously noted in [2], changes relative to IAPWS-95 necessary to better fit sound speeds lead to better agreement with the Hugoniot measurements reported in [18,19 20, 21].

Beginning in section three of the script, sound speeds predicted by IAPWS-95 are integrated (using function *eqst*) and the resulting equation of state is shown to be in parts-per-million agreement with the original. This establishes that the numerical methods remain robust over a larger range in pressure (1 bar to 100 GPa) and temperature (250 K to 10,000 K),

In section four, a model for smooth perturbations from IAPWS-95 of sound speeds above 1 GPa serves as a convenient proxy for changes in the equation of state that allow better matches to high pressure sound speed measurements and determinations of densities and temperatures on the shock-wave Hugoniot.

Sound speed deviations from IAPWS-95 are decreased linearly to -4% between 1 and 10 GPa (measured sound speeds are less than IAPWS-95 predicted sound speeds). Between 10 GPa and 40 GPa the deviations are further decreased to -5%. Above 40 GPa the derivative with pressure of deviations changes sign and the total deviation is about +7% at 100 GPa. The perturbed sound speed surface was integrated from 1 GPa to 100 GPa using function *eqst*. A new *LBF* representation for Gibbs energy was then created with IAPWS-95 values used below 1 GPa and the modified values used at higher pressure. All thermodynamic properties are then determined as appropriate derivatives of this modified *LBF* representation.

Deviations of the modified *LBF* representation from IAPWS-95 are illustrated in Figure 5 (low pressure and temperature regime) and in Figure 6 (full range of pressure and temperature). By construction, the modified representation remains faithful to IAPWS-95 in the regime below 1 GPa and begins to deviate only at higher pressures. The maximum deviation in density is about 6% near 50 GPa. Specific heats are altered from IAPWS-95 by a few percent. Gibbs energy changes by up to 3 MJ/kg at the highest pressures.

In Figure 7 sound speeds measured in diamond anvil high pressure cells are compared to predictions of the modified *LBF* representation. Data and calculated isotherms are illustrated in the upper panel. The deviations of sound speed measurements from the modified representation (lower panel) are generally consistent with experimental scatter. The measurements of Ashara *et al.* [17] extend to higher pressures but relied on an indirect method to determine pressure. They exhibit greater scatter and have systematic deviations relative to the Abramson and Brown [2] and the Decremps *et al.* [16] measurements.

In Figure 8, Hugoniot measurements and calculations based on IAPWS-95 and the modified *LBF* representation are shown. The IAPWS-95 predicted Hugoniot systematically deviates from data by up to 6% for pressures beyond 10 GPa. The modified equation of state fits both Hugoniot densities and Hugoniot temperatures within experimental uncertainties.

In summary, an *LBF* representation was created that duplicates the behavior of IAPWS-95, the reference equation of state. This representation was then modified to better match a selected set of measurements at higher pressures while maintaining the accuracy of the predictions in a low-pressure regime. The modified *LBF* representation for water is constructed using scripts provided in the supplemental materials. Further modification of this equation of state is easily accomplished on the basis of new or revised data or through use of different assumptions concerning the weighting of experimental data.

4. Conclusions

The use of thermodynamic representations based on standardized local basis functions was demonstrated; b spline formalism is a natural choice for such representations. A flexible numerical framework was presented that allows accurate evaluation of thermodynamic properties as a function of pressure, temperature, and composition. In order to adequately interpolate and extrapolate sparse constraints provided by scattered measurements, numerical regularization and consideration of limiting behavior is necessary.

The communication and utilization of a local basis function (*LBF*) representation of thermodynamic properties is straightforward. Having a common format for the representation with evaluations based on standardized algorithms means that any equation of state can be reported to arbitrary precision and that the representation is usable without obtaining or writing specialized computer functions that are applicable to only a single material.

Evaluations of *LBF* representations, parameterized with pressure and temperature as the independent variables, are numerically fast compared to the search required when evaluating formulations based on volume and temperature. The conversion between representations in volume-temperature to pressure-temperature is straightforward and was demonstrated.

When provided with exact properties to fit (*i.e.* creating a representation of an existing equation of state), the local basis function formulation can match the original source to arbitrary precision. More importantly, *LBF* representations can be easily modified in order to better represent newer data in extended regimes of pressure and temperature. The methods were applied to the representation of sound speeds of water as a function of pressure and temperature. From this representation, an accurate equation of state predicting all thermodynamic properties was constructed. This *LBF* representation reproduces predictions of IAPWS-95 in a regime of pressure under 1 GPa and matches extant higher-pressure sound speed and Hugoniot measurements extending almost to 100 GPa.

Acknowledgements

This work was partially funded by NASA Outer Planets Research grant NNX13AL23G and by the Icy Worlds node of NASA's Astrobiology Institute (08-NAI5-0021). Longstanding consultations with Ken Creager are acknowledged and appreciated. S. Vance and O. Bollengier helped improving the focus of this effort. E. Abramson is thanked for his comments and suggestions on the manuscript.

References Cited

[1] R. Span, Multiparameter Equations of State: An Accurate Source of Thermodynamic Property Data, Springer, Berlin, 2000.

[2] E.H. Abramson, J.M. Brown, Equation of state of water based on speeds of sound measured in the diamond-anvil cell. Geochim. Cosmochim. Acta 68 (2004) 1827–1835.

[3] W. Wagner, and A. Pruss, The IAPWS Formulation (1995) for the thermodynamic properties of ordinary water substance for general and scientific use. J. Phys. Chem. Ref. Data 31 (2002) 387–535.

[4] R.C. Phutela, K.S. Pitzer, Densities and apparent molar volumes of aqueous magnesium sulfate and sodium sulfate to 473 K and 100 bar, J. Chem. Eng. Data 31 (1986) 320–327.

[5] D.G. Archer, J.A. Rard, Isopiestic Investigation of the Osmotic and Activity Coefficients of Aqueous $MgSO_4$ and the Solubility of $MgSO_4$ -7H_2O(cr) at 298.15 K: Thermodynamic Properties of the $MgSO_4$ + H_2O System to 440 K, J. Chem. Eng. Data 43 (1998) 791-806.

[6] G.M. Marion, J.S. Kargel, D.C. Catling, S.D. Jakubowski, Effects of pressure on aqueous chemical equilibria at subzero temperatures with applications to Europa. Geochim. Cosmochim. Acta 69 (2005) 259–274.

[7] S. Vance, J.M. Brown, Thermodynamic properties of aqueous $MgSO_4$ to 800 MPa at temperatures for -20 to 100°C and concentrations to 2.5 mol/kg from sound speeds with applications to icy world oceans, Geochem. Cosmo. Acta 110 (2013) 176–189.

[8] T. Driesner, The system H_2O–NaCl. Part II: Correlations for molar volume, enthalpy, and isobaric heat capacity from 0 to 1000°C, 1 to 5000 bar, and 0 to 1 X_{NaCl} Geochim. Cosmochim. Acta 71 (2007) 4902–4919.

[9] C. de Boor, A Practical Guide to Splines, revised ed., Springer, Berlin, 2001.

[10] R.C. Aster, B. Borchers, and C.H. Thurber, Parameter Estimation and Inverse Problems, second ed., Elsevier, New York, 2012.

[11] W. Wagner, J. R. Cooper, A. Dittmann, J. Kijima, H.J. Kretzschmar, A. Kruse, R. Mares, K. Oguchi, H. Sato, I. Stocker, O. Sifner, Y. Takaishi, I. Tanishita, J. Trubenbach, T. Willkommen, The IAPWS industrial formulation 1997 for the thermodynamic properties of water and steam. J. Eng. Gas Turbines Power 122 (2000) 150-182.

[12] F.D. Stacey, and P.M. Davis, High pressure equations of state with applications to the lower mantle and core, Phys. Earth Planet. Inter. 142 (2004) 137–184.

[13] T. Sun, S.N. Biswas, N.J. Trappeniers, C.A. Ten Seldam, Acoustic and Thermodynamic Properties of Methanol from 273 to 333 K and at Pressures to 280 MPa, J. Chem. Eng. Data 33 (1988) 395-398.

[14] J.M. Zaug, L.J. Slutsky, J.M. Brown, Equilibrium properties and structural relaxation in methanol to 30.4 GPa, J. Phys. Chem. 98 (1994) 6008-6016.

[15] C.-W. Lin, J.P.M. Trusler, The speed of sound and derived thermodynamic properties of pure water at temperatures between 253 and 473 K and at pressures up to 400 MPa, J. Chem. Phys. 136 (2012) 094511; doi: 10.1063/1.3688054

[16] F. Decremps, F. Datchi, A. Polian, Hypersonic velocity measurement using Brillouin scattering technique. Application to water under high pressure and temperature, Ultrasonics 44, (2006) 1495–1498.

[17] Y. Asahara, M. Murakami, Y. Ohishi, N. Hirao, K. Hirose, Sound velocity measurement in liquid water up to 25 GPa and 900 K: Implications for densities of water at lower mantle conditions, Earth Planet. Sci. Lett. 289 (2010) 479–485.

[18] S. Marsh, LASL Shock Hugoniot Data, University of California Press, Berkeley, 1980.

[19] A.C. Mitchell, W.J. Nellis, Equation of state and electrical conductivity of water and ammonia shocked to the 100 GPa (1 Mbar) pressure range, J. Chem. Phys. 76 (1982) 6273–6281.

[20] S.B. Kormer, Optical study of the characteristics of shock compressed condensed dielectrics, Sov. Phys. Uspekhi. 11 (1968) 229–254.

[21] G.A. Lyzenga, T.J. Ahrens, W.J. Nellis, A.C. Mitchell, The temperature of shock-compressed water, J. Chem. Phys. 76 (1982) 6282–6286.

Appendix 1 b-spline details and methods to create thermodynamic representations

The following is an overview of necessary b splines fundamentals (based on de Boor [9]) that are applicable to the thermodynamic analysis. For n control points and a knot sequence ($t_1, t_2, \ldots, t_{n+k}$), where k is the order of the b-spline and (k-1) is the "degree" of the associated polynomials, a recursion relationship (equation A1) defines the basis function, $B_{i,k}$(x), where x (the independent variable) lies between the first and last knot and the zeroth order basis function $B_{j,0}$ is equal to 1.

$$B_{j,k}(x) = \frac{x - t_j}{t_{j+k-1} - t_j} B_{j,k-1}(x) + \frac{t_{j+k} - x}{t_{j+k} - t_{j+1}} B_{j+1,k-1}(x) \qquad \text{A1}$$

Each basis function, $B_{j,k}(x)$, is non-zero only for x within the interval between t_j and t_{j+k}. The basis function $B_{j,k}$ evaluated at collection of locations x forms a matrix of values that depend on the knot sequence (chosen intervals for polynomials), the chosen order of the fit (the degree of the polynomial) and where the spline is to be evaluated.

The i^{th} model value, z_i, evaluated at point x_i, is then calculated as a linear combination of basis functions with coefficients a_j.

$$z_i = \sum_{j=1}^{n} B_{j,k}(x_i) a_j \qquad \text{A2}$$

Since the basis functions defined in equation A1, are polynomials, derivative and integrals of $B_{j,k}$ are analytic and straightforward in construction. Derivatives and integrals of $z(x)$ are determined using the coefficients a_j and the integrated or differentiated values of the basis functions, $B_{j,k}(x)$.

The spline order determines how many derivatives are continuous at interval boundaries. By repeating a knot, the number of continuous derivatives is reduced by one. Thus, "knot multiplicity" allows specified control of local discontinuities in interior of a spline representation. Multiplicity of knots equal to the spline order (*k-fold* multiplicity) at both ends of a knot sequence allows full freedom in fitting a spline to arbitrary data.

A surface of two (or more) independent variables, *z(x,y)* (in the case of thermodynamic representations, *x* and *y* are usually either pressure and temperature or volume and temperature) can be constructed from a tensor b spline representation:

$$z_{ij} = \sum_{k=1}^{n}\sum_{l=1}^{m} B_{k,q}(x)\, C_{l,p}(y)\, a_{k,l} \tag{A3}$$

where $C_{l,p}$ are basis functions associated with the order and knot sequence for the second independent variable. In this case, the model coefficients, $a_{k,l}$, form a two dimensional array. Additional independent variables (for example, a compositional variable) can be incorporated following the same steps – add an additional summation including the product of basis functions associated with order and knots sequence for each new independent variable and increase the dimensionality of the array of coefficients.

The matrixes containing basis function representations, *B* and *C,* can be re-mapped into a single matrix *D* so that the tensor spline representation given in equation A3 can be written as a tensor multiplication:

$$z_{ij}=D_{ijkl}a_{kl} \tag{A4}$$

with the implied summations based on the repeated index convention. This can be further re-mapped into a single matrix multiplication:

$$z_{i}=D_{ij}a_{j} \tag{A5}$$

This formalism can be extended to represent thermodynamic surfaces of any dimensionality. The primary “cost” of additional independent variables is in the increasing size of the basis function matrix *D*. However, since this matrix is sparse (because basis functions are non-zero for only a small number of knot intervals around each data point) the problem remains numerically tractable.

Having chosen the spline order and knot sequence and after determining basis function values (*D*) at data locations, *x, y*, the spline coefficients, a_j, can be determined for an over-determined problem through solution of the normal equations:

$$a=D^{-1}z \qquad \text{A6}$$

However, since data are not likely to fully span the parameter space (*ie.,* no data may exist within some knot intervals), the problem defined by equation A6 may not be well conditioned and is likely rank deficient.

The determination of spline coefficients for poorly conditioned systems follows from standard parameter estimation methods [10] where "regularization" introduces additional model constraints. A simple and standard regularization is to assert that models be "smooth" by some physically relevant measure. In addition, constraints on asymptotic or limiting behavior can regularize the representation in a region of extrapolation. The general approach for a linear system with regularization is:

$$a = \begin{bmatrix} D \\ \lambda L \end{bmatrix}^{-1} \begin{bmatrix} z \\ 0 \end{bmatrix} \qquad \text{A7}$$

where D contains basis functions that relate observations z to model parameters a, and the matrix L contains basis functions for the regularization and λ is the weighting factor that balances the trade-off between fitting data and meeting the criteria of the regularization. For regularization based on model smoothness, L contains second derivative basis functions. Increasing λ leads to more misfit of the data but provides a smoother representation.

If the relationships between model parameters and data are non-linear, the problem is recast as an iterative solution to find a better set of model parameters:

$$\delta a = \begin{bmatrix} C \\ \lambda Q \end{bmatrix}^{-1} \begin{bmatrix} z - z_o \\ -\lambda Q a_o \end{bmatrix} \qquad \text{A8}$$

where matrixes C and Q contain elements that are derivatives with respect to the model coefficients a of the elements in matrixes D and L from equation A7. Subscript

$_o$ refers to values associated with a previously estimated model.

Appendix 2: Numerical Implementation of Local Basis Function Representations

All numerical analysis is implemented within the MATLAB and familiarity with this environment is assumed. The “Curve Fitting Toolbox” (an optional toolbox) is required; it contains necessary utility functions for splines. The “Parallel Toolbox” is optional. However, if a multicore computer is used, execution times for some algorithms are significantly reduced. A computer with at least 16 GB of RAM is required in order to execute the example scripts.

The b spline routines are based on published algorithms [9]. Additional functions specific for thermodynamic representations are provided. The conversion of these functions to other common numerical environments (ie, FORTRAN, Python, or Mathematica) is relatively straightforward. Within the MATLAB environment, the command “help functionname” provides documentation of what each function does and how it is invoked. MKS units are used with pressures given in MPa.

Functions Provided in the MATLAB Curve Fitting Toolbox

The following functions provided in the MATLAB Curve Fitting Toolbox are required. Additional spline utility functions are called from within these functions. A short explanation is provided here.

fnval.m (**fun**ction e**val**uation) evaluates splines (tensor or univariant) at specified points – the input can be in the form of a cell (data on a grid) or as a matrix (scattered data points). Data at gridded locations are evaluated using the b spline representation. Since it is computationally efficient to use “piece-wise polynomial” (pp) splines for evaluations at scattered data sites, *fnval* converts b-form to pp-form if scattered data points are to be evaluated.

fnder.m (**fun**ction **der**ivative) returns a spline representing the specified derivative (-1 for the integral) of an input spline.

spcol.m (**sp**line **col**location) – returns values and derivatives of b spline basis functions evaluated at specified data locations for a given knot sequence and spline order.

apknts.m (**ap**propriate **kn**ot**s**) and *opknts.m* (**op**timal **kn**ot**s**) are functions to determine appropriate knot locations given a spline order and control point locations.

fnplt.m (**fu**n**c**tion **pl**ot) is an easy-to-use plotting function for splines.

Functions Provided for Thermodynamic Analysis

An overview of the thermodynamic functions is provided here. Within MATLAB, typing “help functionname” where functionname is one of the following, provides more information about required input variables and output variables. Separate live scripts (*.mlx* files) show how to use these functions.

IAPWS.m : a direct implementation of the IAPWS-95 formula for the equation of state of water. Input can be cells (for results on a specified grid of pressures and temperatures) or matrixes (for results at scattered data sites) of pressure and temperature or densities and temperature. The input thermodynamic state can be either pressure and temperature or density and temperature. Values calculated at specified pressures require a numerically costly search while calculations at specified densities are much faster.

fnIAPWSval.m returns thermodynamic properties based on an accurate local basis function representation of IAPWS. Input can be cells (for results on a specified grid of pressures and temperatures) or matrixes (for results at scattered data sites) of pressure and temperature. This function loads a spline representation contained in the file *IAPWS_sp_strct.mat* .

fnGval.m returns thermodynamic properties based on a generic local basis function representation of Gibbs free energy. An input of either a cell (gridded data) or a matrix (scattered data sites) of pressures, temperatures, and (optional) compositions is used in the evaluation. MKS units required.

spdft.m (**sp**line **d**ata **fit**ter) is a versatile fitting function that provides a solution to equation A7. It returns a local basis function representation of a surface as a function of one, two or three independent variables. Data to be represented can be scattered or gridded. The spacing of control points, the damping, the distribution of regularization points, the order of the spline, and the derivatives used in regularization are adjustable.

spgft.m (**sp**line **G**ibbs energy **fit**ter) returns a local basis function representation of Gibbs energy as an implementation of equation 2. The input includes gridded Gibbs energies, densities and specific heats with regularization based on specified derivatives of G. Gridded values can be based on separate *LBF* fits to scattered measurements. MKS units are required with pressures in MPa.

eqst.m (**eq**uation of **st**ate integration) numerically integrates sound speeds as a function of pressure and temperature to return densities, specific heats and Gibbs energy. It is a "predictor-corrector" implementation for solution of equations 3, 4, and 5. The standard inputs are sound speeds on a grid of pressures and temperatures with densities, specific heats, and Gibbs energies specified at the lowest pressure. Alternatively, specific heats given on the full grid of pressures and temperatures forces the output to be consistent with these values. Several methods are implemented within this function to reduce numerical instabilities. MKS units are required with pressure in MPa.

mk_mask4Gspline.m returns a matrix based on an input of gridded pressures and temperatures. Matrix elements are 1 if a P-T point is in a physically plausible range for water (*ie.* where IAPWS-95 returns finite and reasonable values). Matrix elements are set to NaN (not a number) otherwise and at locations near the critical point for water.

CalcWaterHug.m uses a local basis function representation for the equation of state in order to solve the Rankine-Hugoniot equations. Densities and temperatures are predicted for specified Hugoniot pressures.

therm_surf.m provides a convenient plotting tool for the given examples. It plots 3D perspectives of thermodynamic properties as a function of temperature and pressure.

Nested within *spdft.m* and *spgft.m* are two key sub-functions.

makeCmn.m remaps the tensor product (shown in equations A3 and A4) into the matrix given in equation A5. The result is a sparse matrix representing basis function for either two or three independent variables.

area_wt.m provides relative weights for regularization points proportional to fractional volume within the grid that each point represents. In the case of a non-uniform distribution of knots and regularization points this volume weighting enhances the numerical stability of the inverse problem (equation A7).

A feature of modern numerical environments is the availability of “structures” – a single variable name that contains heterogeneous information. The convention used here is that a variable name beginning with “sp” denotes a structure containing a b spline representation. For example, “sp_rho” is an appropriate name for a structure containing the b spline representation for density. The standard MATLAB format for a spline structure contains the following information.

```
sp_rho =
        form: 'B-'
        knots: {[1x36 double]  [1x26 double]}
        number: [30 20]
        order: [6 6]
        dim: 1
        coefs: [30x20 double]
```

where, in this case, the b spline has 30 pressure control points and 20 temperature control points using a 6th order (in both P and T) representation. The length of each knot sequence is equal to the sum of the spline order and the number of control points.

The function *spdft* appends a list of data and fit statistics to the spline structure. In the example above, the additional structure within sp_rho titled "Data" has contents:

```
sp_rho.Data=
PT: [263x2 double]
rho: [263x1 double]
del_rho: [263x1 double]
TPc: {[1x30 double]  [1x20 double]}
lam: [1 1]
RegFac: [4 4]
rms: 2.3317e-05
devs: [263x1 double]
chisqr: 1.446
```

which are a list of the input pressures and temperatures in a 263 by 2 matrix, the data values (263 points), a list of data uncertainties, "Control points" used to construct the spline, the damping parameter used in constructing the fit, the number of regularization points between each control point, the rms misfit in units of the data, a list of deviations between the fit and the data, and the reduced χ^2 misfit. Thus, a single entity, containing all relevant information, is packaged and easily saved and transported.

Example Scripts

"Live" scripts are provided that offer tutorial assistance in using the basic thermodynamic tools and give demonstrations of the methods to represent thermodynamic surfaces. Sections of each script can be executed sequentially by clicking on the side-bar. However, resulting text and plotting output from previous sessions are saved in the file. The pdf versions of the live scripts also contain the full record of calculations and figures.

The script *FitIndependentVariablesExamples* provides the basic tutorial on the use of the function *spdft* to fit surfaces to arbitrary data with one, two, or three independent variables. The features of the function are demonstrated in application to several different data sets.

The script *WaterSoundSpeedFitting* has sections that reproduce the analysis discussed in: “Reanalysis of Sound Speed Measurements”.

The script *IAPWS_fit_modify* has sections that reproduce the analysis discussed in “The Representation and Modification of IAPWS using Local Basis Functions**”.**

Table 1. Relationship of thermodynamic quantities to derivatives of Gibbs energy. Subscripts on G denote the appropriate partial derivative with *p* for pressure, *t* for temperature, and *m* for the chemical composition variable.

Thermodynamic Quantity	**Symbol**	**Derivatives w.r.t G**
Specific volume (inverse of density)	V	G_p
Specific Heat	C_p	$-G_{2t}\,T$
Isothermal Bulk Modulus	$\mathrm{K_t}$	$-G_p\,G_{2p}^{-1}$
Thermal expansivity	α	$G_{pt}\,G_p^{-1}$
Adiabatic Bulk Modulus	K_s	$(K_t^{-1} - \alpha^2 VTC_p^{-1})^{-1}$
Sound Speed (squared)	u^2	$\frac{G_p^2 G_{2t}}{G_{pt}^2 - G_{2t}G_{2p}}$
Chemical Potential	μ	G_m

Figures

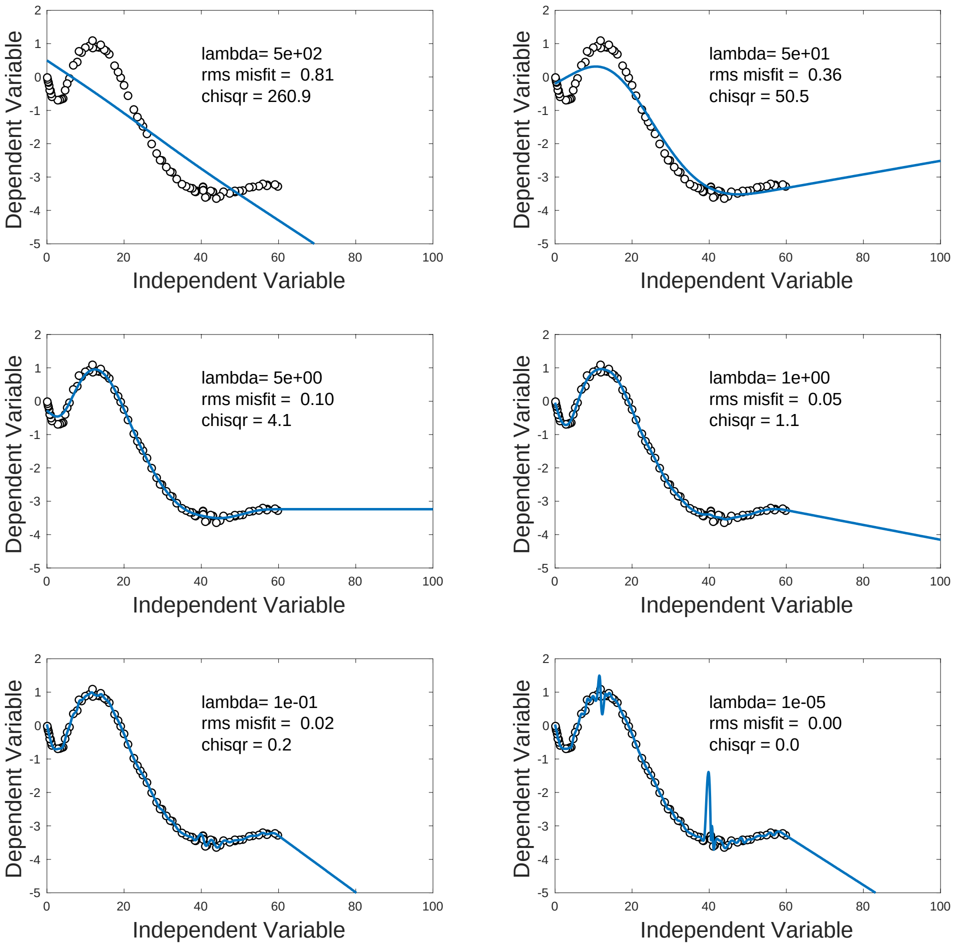

Figure 1**.** Local basis function fits (solid lines) to data (circles) using function *spdft*. The "data" are arbitrary and dimensionless with an assigned uncertainty of 0.05 and extend over an independent variable range from 0 to 60. The representations extend to 100 using a fourth-order spline that is regularized by minimizing the second derivative. The six panels illustrate the effect of different values for the damping parameter λ, The *rms* error and reduced χ^2 are given in each panel. A large value of damping forces a linear fit to data. $\lambda=1$ gives an *rms* misfit and χ^2 consistent with an acceptable fit to the data. Values of damping that are too small lead to overfitting of the data. The representations become linear in the regime of extrapolation.

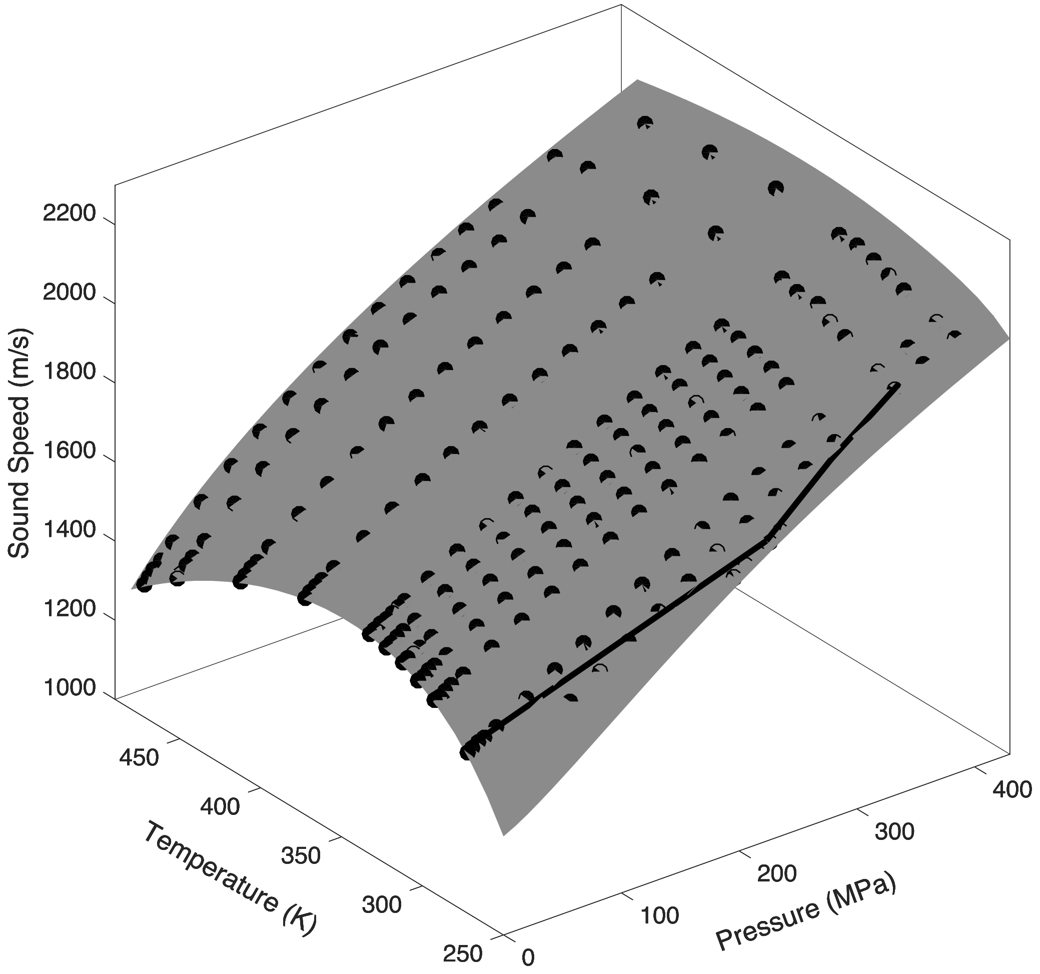

Figure 2. Sound speeds for water as a function of temperature and pressure. The surface is based on predictions using IAPWS-95 [3]. Filled circles are the measurements of Lin and Trusler [15]. Solid line gives the phase boundary between water and ice.

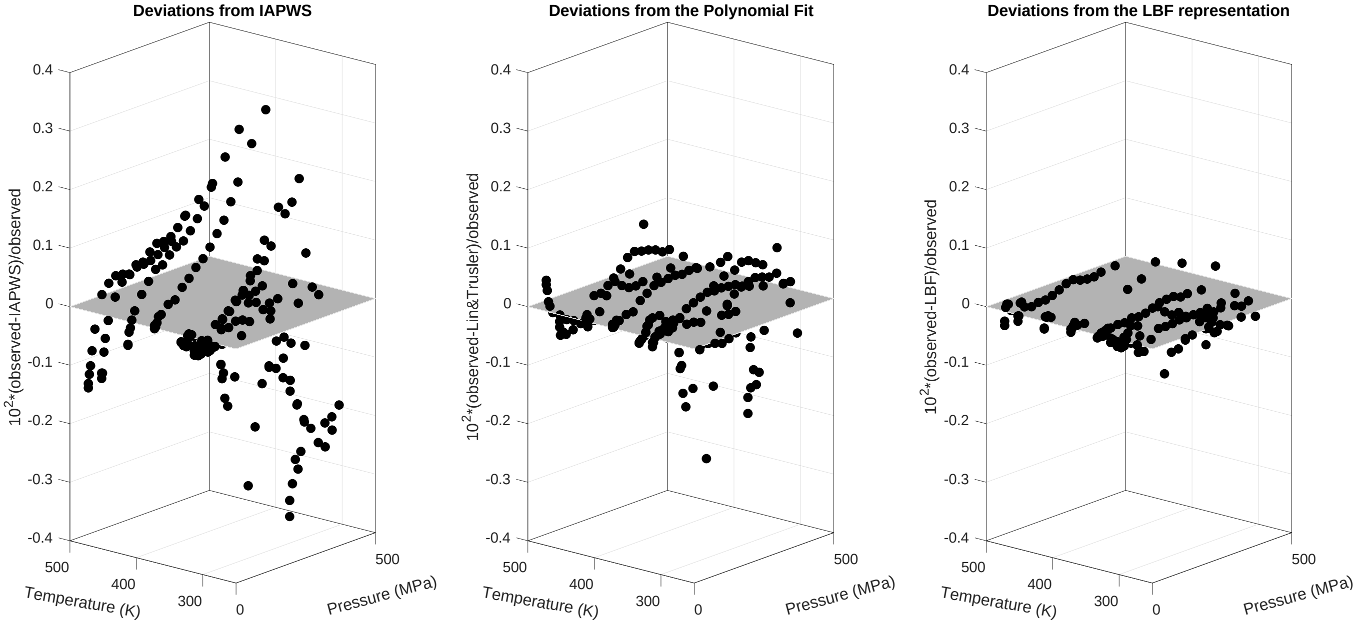

Figure 3. Deviations of sound speed measurements reported in [15] from three representations for sound speeds. The panel on the left gives deviations from IAPWS-95. The middle panel gives deviations from the polynomial fit reported in [15]. The panel on the right gives deviations from the local basis function representation that was required to fit measurements within reported experimental uncertainties.

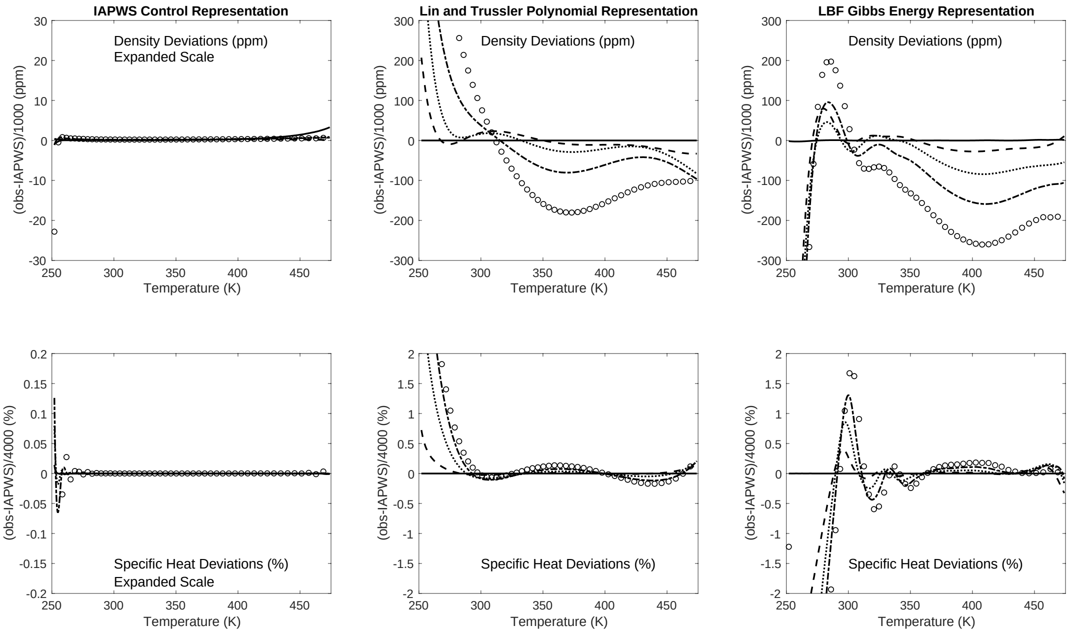

Figure 4. Equation of state deviations from IAPWS-95 based on integration of three representations for sound speeds. Density deviations are shown in the upper panels and specific heat deviations in the lower panels. Isobars are plotted as a function of temperature with solid lines for 0.1 MPa, dashed lines for 100 MPa, short dash lines for 200 MPa, dot-dash lines for 300 MPa, and circles for 400 MPa. The deviation panels on the left side test the numerical framework through integration of IAPWS sound speeds. Note that scales for these panels are expanded by factor of ten relative to the remaining panels. The middle panels give deviations from IAPWS based on integration of the polynomial representation of Lin and Trusler [15]. Panels on the rights are deviations from IAPWS based on integration of the *LBF* fit to the Lin and Trusler [15] sound speed measurements.

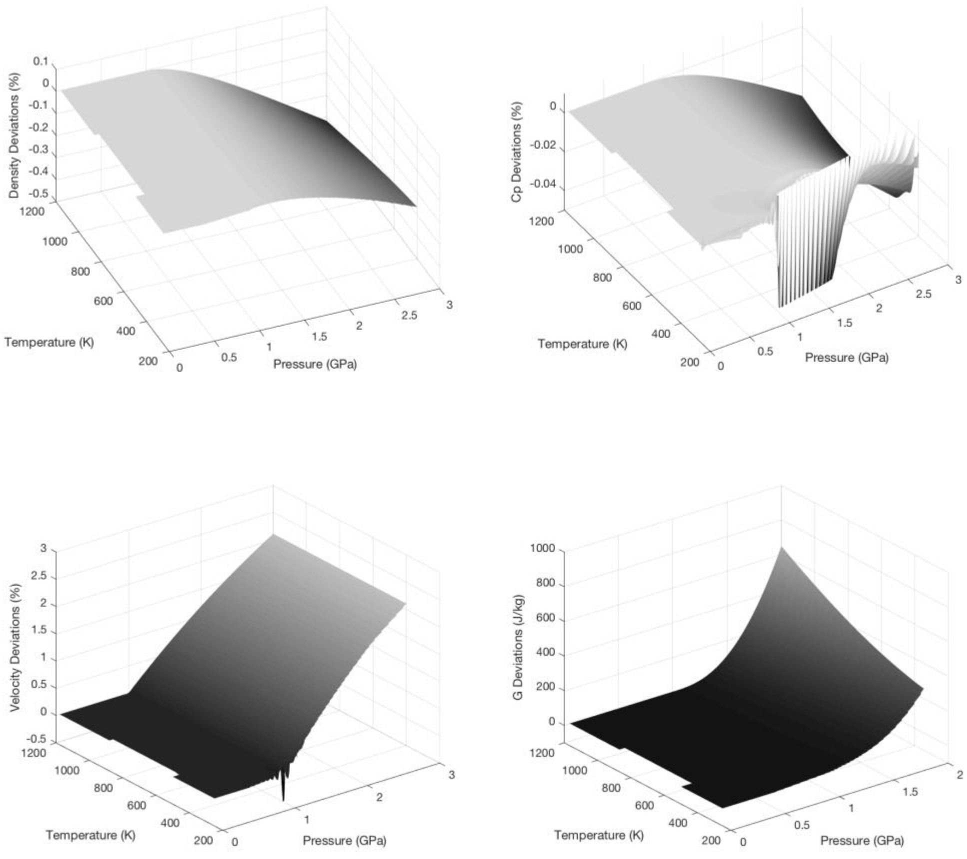

Figure 5. Deviations of selected thermodynamic properties from IAPWS-95 using the modified *LBF* representation for Gibbs energy over a limited range of pressures and temperatures. Density deviations are shown in the upper left panel, specific heat deviations in the upper right panel, and Gibbs energy in the lower right panel. Sound speed deviations are in the lower left panel. By construction, the modified representation matches IAPWS-95 below 1 GPa and begins to deviate at higher pressures. Slight edge instabilities for specific heats at low temperatures of ∼0.1% extend beyond plotting limits of 0.04%. The small excluded regime near T_c is visible as a notch in the surface.

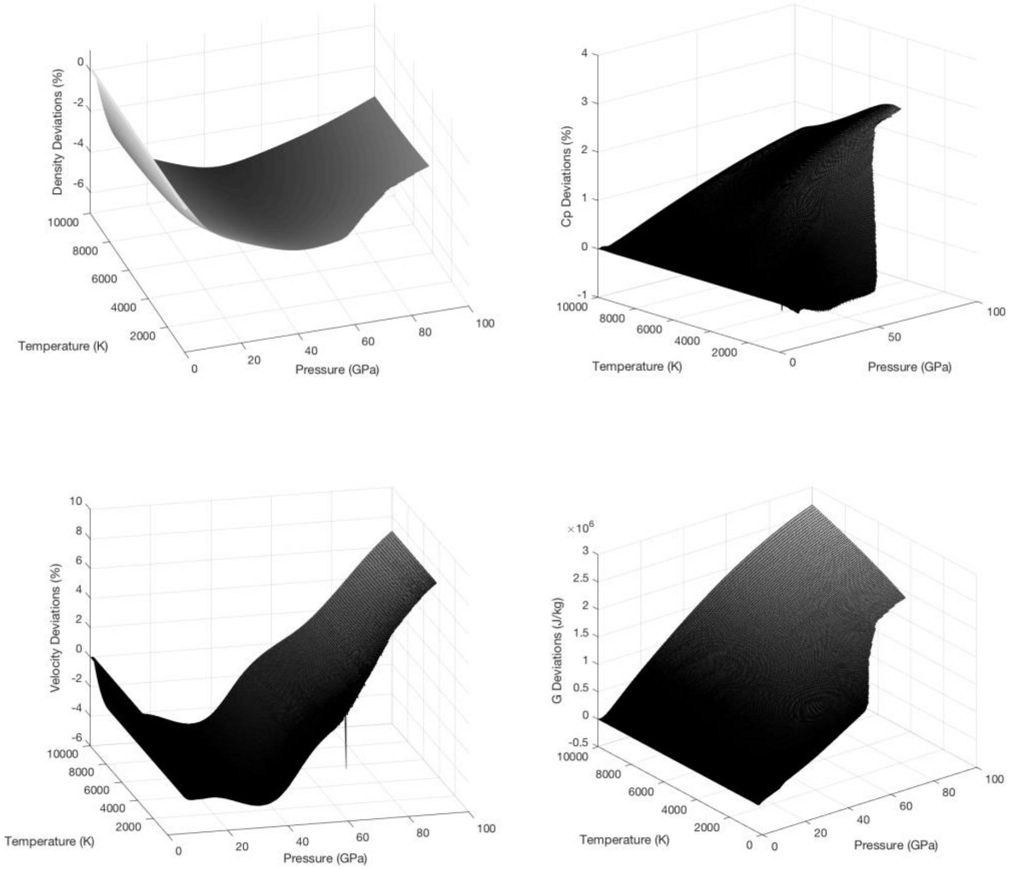

Figure 6. Deviations of selected thermodynamic properties from IAPWS-95 using the modified *LBF* representation for Gibbs energy over the entire range of pressures and temperatures. Density deviations are shown in the upper left panel, specific heat deviations in the upper right panel, and Gibbs energy in the lower right panel. The temperature-independent correction to IAPWS-95 sound speeds that was used to construct the modified equation of state is shown in the lower left panel.

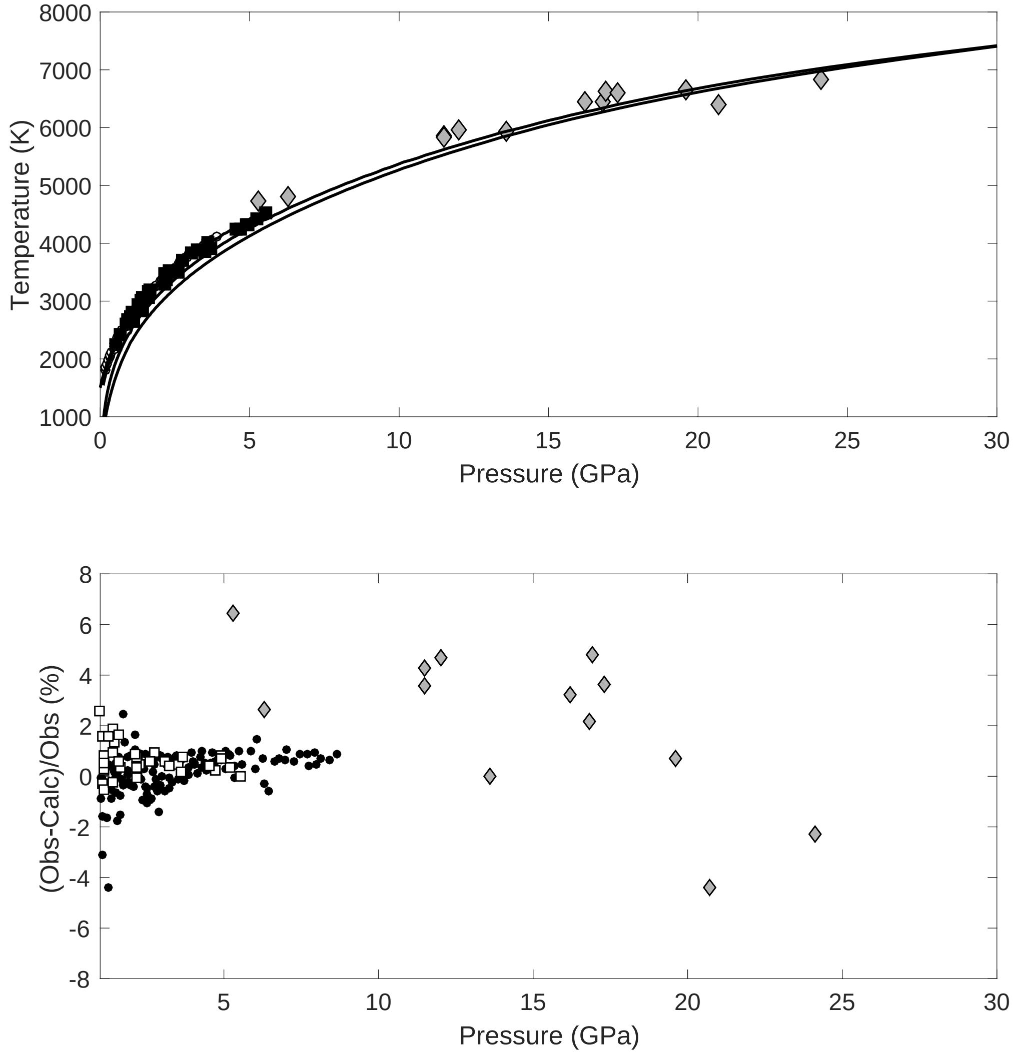

Figure 7. Sound speeds as a function of pressure and temperature. Upper panel: sound speeds measured in the diamond anvil cell and isotherms (400, 500, 750 and 1000 K) predicted by the local basis function representation for Gibbs energy. Lower panel: deviations of measured sound speeds from the local basis function representation. Open squares are from Abramson and Brown [2] filled circles are Decremps *et al.* [16], diamonds are Ashara *et al.* [17]. The sound speed measurements extend from 400 K to nearly 1000 K.

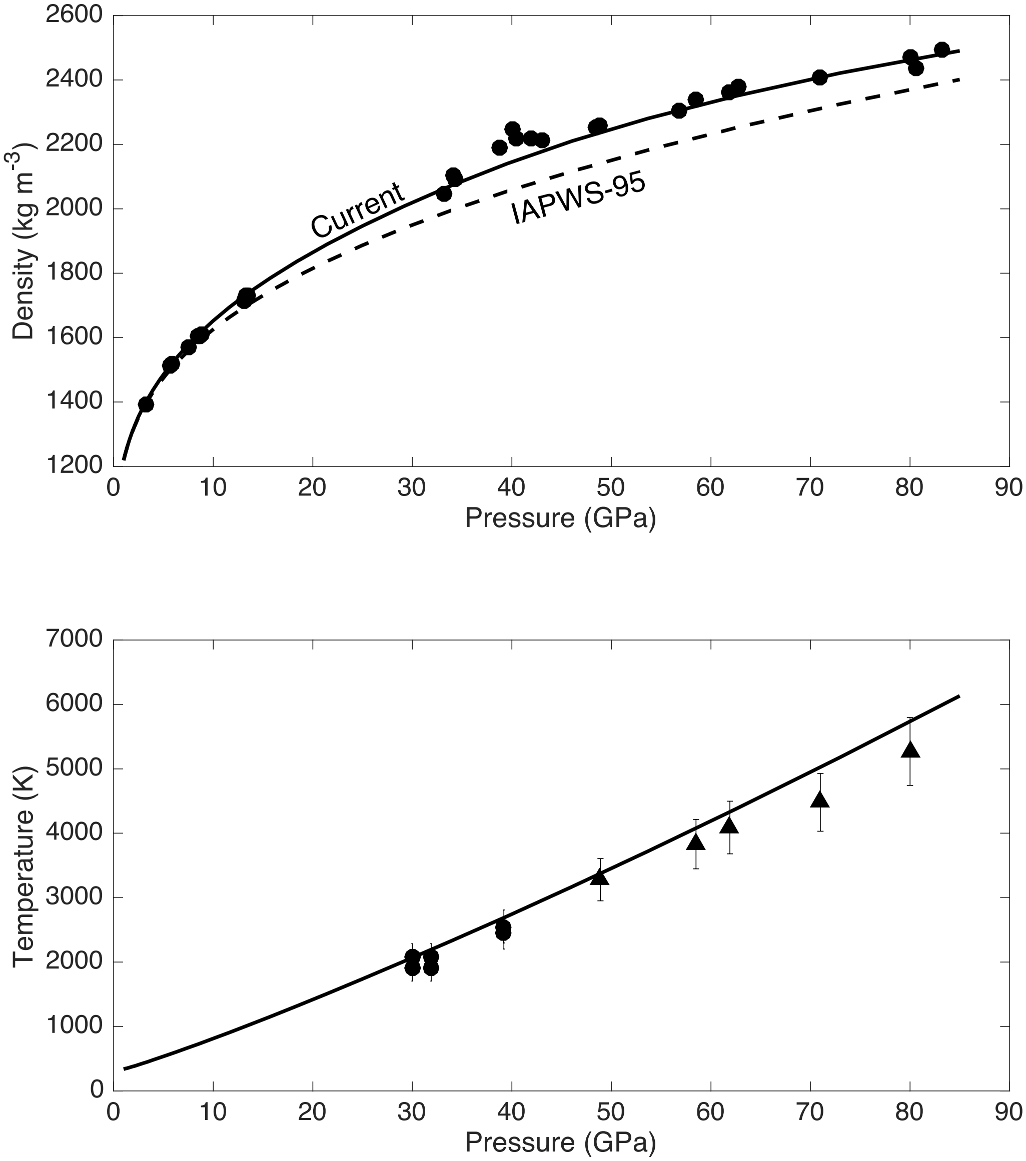

Figure 8. Upper panel: Hugoniot densities (circles) are from Marsh [18] (below 45 GPa) and from Mitchell and Nellis [19] (above 45 GPa). The solid line is based on the *LBF* representation. The dashed line for Hugoniot densities is based on the equation of state given by IAPWS-95. Lower panel: Measured Hugoniot temperatures and uncertainties by Kormer [20] (circles) and Lyzenga *et al.* [21] (triangles) are shown along with temperatures predicted using the local basis function representation for water.