energies

MDPI

Article

An Efficient Method to Predict Compressibility Factor of Natural Gas Streams

Vassilis Gaganis [1,†], Dirar Homouz [2,†], Maher Maalouf [3,*], Naji Khoury [4] and Kyriaki Polychronopoulou [5,*,‡]

[1] Mining and Metallurgical Engineering, National Technical University of Athens, Athens 157 73, Greece
[2] Applied Mathematics and Sciences, Khalifa University, Abu Dhabi, P.O. Box 127788, United Arab Emirates
[3] Industrial and Systems Engineering, Khalifa University, Abu Dhabi, P.O. Box 127788, United Arab Emirates
[4] Civil and Environmental Engineering, Notre Dame University Louaize, Beirut, P.O. Box 72, Zouk Mikael, Lebanon
[5] Mechanical Engineering, Khalifa University, Abu Dhabi, P.O. Box 127788, United Arab Emirates
* Correspondence: maher.maalouf@ku.ac.ae (M.M.); kyriaki.polychrono@ku.ac.ae (K.P.)
† These authors contributed equally to this work.
‡ Current address: Center for Catalysis and Separations, Khalifa University, Abu Dhabi, P.O.Box 127788, United Arab Emirates.

Received: 17 April 2019; Accepted: 24 May 2019; Published: 4 July 2019

Abstract: The gas compressibility factor, also known as the deviation or Z-factor, is one of the most important parameters in the petroleum and chemical industries involving natural gas, as it is directly related to the density of a gas stream, hence its flow rate and isothermal compressibility. Obtaining accurate values of the Z-factor for gas mixtures of hydrocarbons is challenging due to the fact that natural gas is a multicomponent, non-ideal system. Traditionally, the process of estimating the Z-factor involved simple empirical correlations, which often yielded weak results either due to their limited accuracy or due to calculation convergence difficulties. The purpose of this study is to apply a hybrid modeling technique that combines the kernel ridge regression method, in the form of the recently developed Truncated Regularized Kernel Ridge Regression (TR-KRR) algorithm, in conjunction with a simple linear-quadratic interpolation scheme to estimate the Z-factor. The model is developed using a dataset consisting of 5616 data points taken directly from the Standing–Katz chart and validated using the ten-fold cross-validation technique. Results demonstrate an average absolute relative prediction error of 0.04%, whereas the maximum absolute and relative error at near critical conditions are less than 0.01 and 2%, respectively. Most importantly, the obtained results indicate smooth, physically sound predictions of gas compressibility. The developed model can be utilized for the direct calculation of the Z-factor of any hydrocarbon mixture, even in the presence of impurities, such as N_2, CO_2, and H_2S, at a pressure and temperature range that fully covers all upstream operations and most of the downstream ones. The model accuracy combined with the guaranteed continuity of the Z-factor derivatives with respect to pressure and temperature renders it as the perfect tool to predict gas density in all petroleum engineering applications. Such applications include, but are not limited to, hydrocarbon reserves estimation, oil and gas reservoir modeling, fluid flow in the wellbore, the pipeline system, and the surface processing equipment.

Keywords: natural gas stream; compressibility factor; kernel ridge regression; truncated Newton method

1. Introduction

Fluid properties are directly involved in all flow and volumetric calculations in the upstream and downstream zones of the petroleum industry. Density and its change over pressure, vaporization, or condensation ratios, as well as kinetic properties such as viscosity all affect the obtained results [1–4].

For the case of natural gas, various calculations need to be run including the estimation of hydrocarbon reserves in a reservoir, the study of the gas flow in the reservoir and the wellbore, and its thermodynamic behavior through the pipelining system until its arrival at the sales point and further transportation to the end user, whereas the latter could be a home user, a plant, or an electric power generation unit. The accurate determination of gas density is of utmost importance, directly related to the conversion of flow rates from line conditions to standard ones where rates are commonly reported. Moreover, the rate of change of density with pressure under constant temperature, known as isothermal compressibility, is related to the accumulation or withdrawal of mass in a control volume such as a reservoir when that volume is depleted and pressure undergoes reduction.

Although natural gas is a complex multi-component mixture with methane being the major compound and other compounds such as nitrogen, carbon dioxide, ethane, propane, and heavier hydrocarbons at lower concentrations [5,6], for the special case of near surface conditions, it can be treated as an ideal one. Therefore, its thermodynamic behavior is governed by the ideal gas Equation of State (EoS) $V_m = {}^{RT}/_{p}$, where V_m denotes molar volume; hence, density ρ and compressibility c at pressure p and at temperature T are given respectively by:

$$\rho = \frac{pM}{RT} \tag{1}$$

$$c = -\frac{1}{V}\frac{dV}{dp} = \frac{1}{p} \tag{2}$$

where M stands for the fluid's molar mass. To handle conditions far from the atmospheric one, the above definition needs to be extended by introducing the deviation factor Z, also known as the compressibility factor, not to be confused with isothermal compressibility, such that $V_m = {}^{ZRT}/_{p}$. In that case, the density and compressibility of the gas are given by:

$$\rho = \frac{pM}{ZRT} \tag{3}$$

and:

$$c = -\frac{1}{V}\frac{dV}{dp} = \frac{1}{p} - \frac{1}{Z}\frac{dZ}{dp} \tag{4}$$

respectively. From a thermodynamic perspective, the Z-factor basically describes the deviation between real gas and ideal gas behavior. There are a number of accurate equations of state providing estimates of the Z-factor; among them are the Soave–Redlich–Kwong (SRK) [7], the Peng–Robinson (PR) [8], the Lee–Kesler EoS [9], and the Zudkevitch–Joffe–Redlich–Kwong (ZJRK) ones [10]. Research has been published also for the calculation of the virial coefficients of the EoS [11]. Clearly, all EoS-based methods to predict the Z-factor need an accurate description of the gas composition and characterization of all its components. This last step can be very tedious, as it involves estimation of properties' components such as critical values, acentric factors, and binary interaction coefficients.

Another group of methods takes advantage of the corresponding states principle according to which fluids at the same reduced pressure and temperature exhibit very close compressibility factor values. Reduced conditions are defined by $p_r = {}^{p}/_{p_c}$ and $T_r = {}^{T}/_{T_c}$, respectively, where p_c and T_c denote the mixture's critical pressure and temperature, respectively. The most pronounced method to compute the Z-factor by this approach is the use of the industry-standard Standing–Katz (S-K) chart [5], which provides the Z-factor of natural gas as a function of its reduced properties. This chart is known to perform surprisingly accurately for hydrocarbon mixtures, although it has been generated by using experimental data from a very limited dataset. This chart, originally generated in the early 1940s, is valid for T_r values in $[1.05, 3.0]$ and p_r values up to 15. Later, it was extended [12] up to p_r equal to 30 only for a narrower range of reduced temperatures, i.e., $T_r \in [1.4, 2.8]$. When impurities are present, the correction method of Wichert and Aziz [13] needs to be utilized. Various mixing rules are available to estimate mixture's T_c and p_c when the composition and the critical properties of the gas components are known. The methods of Kay [14], Stewart–Burkhardt–Voo (SBV) [15] and Sutton [16] are the most widely used. When either the composition or the gas components' critical properties are unknown, the S-K chart is still applicable by utilizing the pseudo-critical properties p_{pc} and T_{pc}, respectively. The latter can be computed from correlations of the specific gravity [17], thus leading to the estimation of the Z-factor by means of the pseudo-reduced values defined by $p_{pr} = p/p_{pc}$ and $T_{pr} = T/T_{pc}$. For the sake of simplicity, we will use p_r and T_r to denote both reduced and pseudo-reduced pressure and temperature in the following sections.

To facilitate the use of computers, the S-K chart has been fitted by various correlations, which can be distinguished into two categories. Firstly, the iterative ones require repeated calculations to solve a non-linear algebraic equation, the root of which is the Z-factor or a direct function of that. The methods of Hall and Yarborough (H-Y) [18], as well as that of Dranchuk and Abu Kassem (DAK) [19] are the most pronounced ones in this category. Secondly, explicit correlations are available for directly predicting the Z-factor given the reduced or pseudo-reduced properties. The methods of Beggs and Brill [20], Azizi et al. [21], Kumar [22] (also known as the Shell oil one), and Heidaryan [23] are some of them. Alternative approaches have also been presented such as that of Kareem et al. [24], who aimed at fitting the H-Y approximation of the S-K chart. To achieve that, they used a transformation to simplify the complexity of the H-Y method, so as to fit it with an explicit formula. In order to improve the prediction capabilities of the Standing–Katz chart, Elsharkawy [25] presented a new simple mixing rule to calculate the pseudo-critical properties of the gas when the composition is known.

Recent studies have been focusing on the use of machine learning techniques to predict the Z-factor. Among them, Moghadassi et al. [26] applied Artificial Neural Networks (ANN) [27] to calculate the PVTproperties for pure gases, whereas Kamyab et al. [28] designed an ANN for the prediction of the Z-factor of natural gas by using the Standing–Katz chart as their library. The authors reported that their ANN method was more accurate than the iterative DAK [19], and it is applicable for the whole pressure range of the extended Standing–Katz chart, i.e., up to $p_r = 30$. Sanjari et al. [29] developed an ANN to calculate the compressibility factor, this time trained against the experimental Z-factor values rather than the ones obtained from the S-K chart, and they compared their values with empirical methods and EoS. Furthermore, Fayazi et al. [30] and Kamari [31] developed Support Vector Machines (SVMs) [32] by training them against experimental data that involved a variety of gas compositions, ranging from sour to sweet natural gas. Their approaches were shown to be far more accurate than EoS empirical correlations. Mohamadi et al. [33] conducted Z-factor calculations as they were derived by adopting empirical correlations and EoS coupled with intelligent methods. In particular, they reported on the improvement of van der Waals and Redlich–Kwong EoS by using Genetic Algorithms (GA), as well as Fuzzy Inference Systems (FIS), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and ANNs to predict the Z-factor.

From the discussion above, it becomes clear that the development of stable and accurate calculation methods to compute the Z-factor by means of the S-K chart is still a hot topic in the natural gas industry. Indeed, the S-K chart is a straightforward approach, and its H-Y computer implementation has been considered as the industry standard for decades [8]. However, despite recent improvements, none of the available methods can be freely used for any arbitrary pressure and temperature conditions, mostly due to two major drawbacks of them. Firstly, most of the available methods are applicable only to a limited pressure range, usually $p_r < 15$, thus disregarding the S-K chart extension up to $p_r = 30$. This restriction does not allow for the description of the gas thermodynamic behavior at high pressures such as those prevailing in HP/HT reservoirs or at high pressure compressors. Additionally, some of the regression models and most of the machine learning-based approaches when applied to the prediction of the Z-factor exhibit an oscillating behavior. This is attributed to the fact that such models are data driven, and they have been trained by only focusing on the accurate approximation of the Z-factor values at each of the available training data points. As a result, no physical evidence, such as a known trend of the Z-factor and its derivative, can be directly inherited by the model. In fact, the model is asked to "discover" underlying trends by itself through training. As an example, consider the relationship of the Z-factor with pressure at pressures above p_r = 15 and up to p_r = 30, which is perfectly linear. Data-driven models, when simply trained against data points, are not guaranteed to exhibit that natural straight line behavior and constant value of the Z-factor derivative. Instead, slight Z-factor deviations may lead to significant "hiccups" in the predicted compressibility value, hence to instability in the solution of the natural gas flow problem either in a reservoir or in a pipeline manifold. On the other hand, although the H-Y correlation exhibits remarkable accuracy against the Standing–Katz chart, it may also exhibit various convergence problems, leading to possible failure of the computation. As an example, consider the calculation of the Z-factor at $T_r = 1.05$, $p_r = 3.1$ by solving $f(y) = 0$, where $y = {}^{ap_r}/_Z$ and a is a function of T_r, using the recommended initial value of $y = 0.001$ [8]. Figure 1 illustrates the plot of $f(y)$, as well as the step values followed by a typical Newton–Raphson method. It can be readily seen that the flat part of $f(y)$ at $y = 0.18$ causes an overshooting, thus leading the estimate at the third iteration to an extremely high value, definitely outside the valid y parameter bounds, from which the Newton–Raphson method cannot recover. Note that such behavior is quite common at various p_r values above 20 at all reduced temperatures.

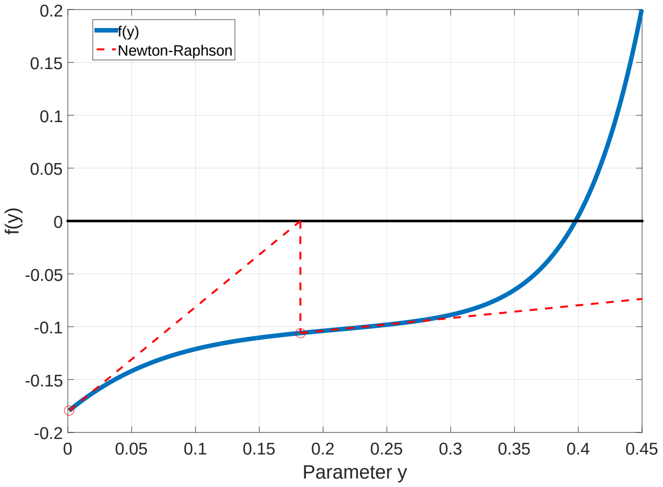

Figure 1. Example of failed solution of the H-Y method.

The objective of this study is to develop a numerical model to predict the Z-factor of hydrocarbon gases by fitting the Standing–Katz chart so as to overcome the weaknesses of the existing methods. For this task, the reduced pressure range is split into three regions, each described with its own submodel. For low reduced pressures, a Kernel Ridge Regression (KRR) model is used, whereas a linear and a quadratic one are developed for medium and high pressures. Figure 2 provides a general sketch of the problem.

The combined model acts as a rapid and inexpensive method to estimate the compressibility factor of a gas stream accurately. Compared to most existing methods, it can be safely used to provide Z-factor values in an extended operating range of reduced pressure values up to $p_r = 30$, which fully covers all upstream operations and most of the downstream ones. Unlike other numerical models, it takes advantage of the simplicity of the industry standard Standing–Katz diagram, thus ensuring a simple form and guaranteeing continuity of the Z-factor value and its derivative, hence smooth and physically-sound values. Additionally, the developed model can be utilized for the direct calculation of the Z-factor of any hydrocarbons mixture, even in the presence of impurities such as N_2, CO_2, and H_2S. The model's accuracy combined with the guaranteed continuity of the compressibility factor derivatives with respect to pressure and temperature renders it as the perfect tool to predict gas density in all petroleum engineering applications. Such applications include, but are not limited to, oil and gas reserves' estimation and reservoir modeling, as well as fluid flow in the wellbore, the pipeline system, and the surface processing equipment.

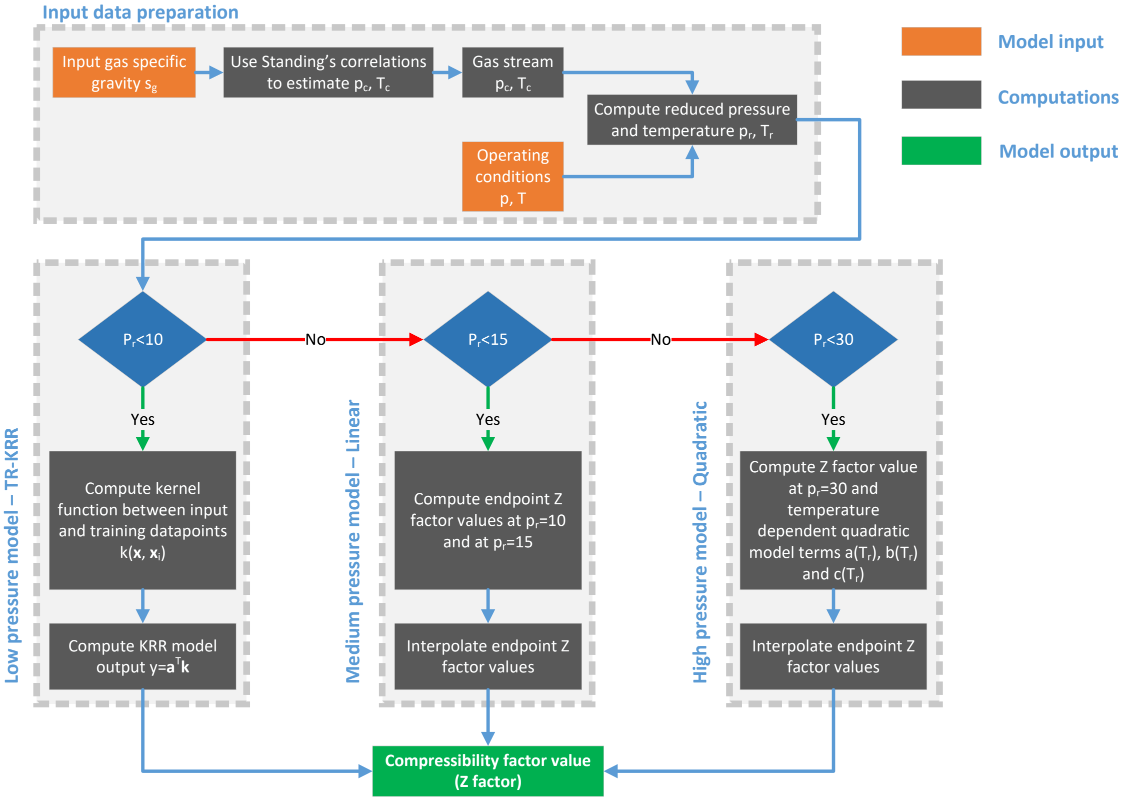
Input data preparation
Input gas specific gravity s_g
Use Standing's correlations to estimate p_c, T_c
Gas stream p_c, T_c
Compute reduced pressure and temperature p_r, T_r
Operating conditions p, T
Model input
Computations
Model output
Low pressure model – TR-KRR
Medium pressure model – Linear
High pressure model – Quadratic
P_r<10
No
P_r<15
No
P_r<30
Yes
Yes
Yes
Compute kernel function between input and training datapoints k(x, x_i)
Compute endpoint Z factor values at p_r=10 and at p_r=15
Compute Z factor value at p_r=30 and temperature dependent quadratic model terms a(T_r), b(T_r) and c(T_r)
Compute KRR model output y=a^T k
Interpolate endpoint Z factor values
Interpolate endpoint Z factor values
Compressibility factor value (Z factor)

Figure 2. General sketch of the proposed approach.

Nevertheless, it should be pointed out that the proposed method has been designed to reproduce the Standing–Katz chart; hence, it can only be used for hydrocarbon mixtures such as those against which the standard diagram has been developed. Therefore, it might not be applicable to mixtures encountered in chemical plants, especially those comprised of non-hydrocarbon and/or polar compounds. Additionally, the non-rigorous background of the method does not allow for safe extrapolation of the operating conditions beyond those of the original chart, unlike a rigorous, although computationally-demanding, tuned multi-parameter equation of state model.

The rest of the paper is organized as follows: In Section 2, the proposed methodology is described in detail, including discussions on the data generation, the Truncated Regularized Kernel Ridge Regression (TR-KRR) method, and the numerical approach followed. The method's performance and related computational issues are discussed in Section 3. Three example applications demonstrating the efficiency of the proposed method are presented in Section 4. The conclusions are stated in Section 5.

2. Methodology

Thorough examination of the standard Standing–Katz chart and its extension to high reduced pressures (Figure 3) indicates that it can be split into two regions along the reduced pressure axis based on the complexity of the isotherms shape. Indeed, at p_r values below 10, the change of the Z-factor versus reduced pressure exhibits complex behavior with specific minima and varying curvature at most of the isotherms, thus indicating proximity to critical fluid behavior. On the other hand, for $p_r \geq 10$, the isotherms exhibit a perfectly straight line shape up to the maximum p_r value of 15. By further examining the supplementary Z-factor chart [12], it is clear that the Z-factor also varies linearly with reduced pressure. Moreover, the average slope at each isotherm in the $p_r \in [15, 30]$ range exhibits a very slight change (of less than 0.5%) compared to the slope in the $[10, 15]$ range. As a result, three distinct regions can be identified in the S-K chart, and each one of them needs to be treated separately by its own modeling technique. This way, the use of "blind", data-driven machine learning methods can be limited to the complex behavior of the low pressure region, whereas simpler models directly adopting the inherent linearity of the Z-factor with respect to p_r can be generated for p_r values above 10. For the sake of notation simplicity, the low, medium, and high pressure range will be labeled by $L = [0, 10]$, $M = [10, 15]$, $H = [15, 30]$ respectively.

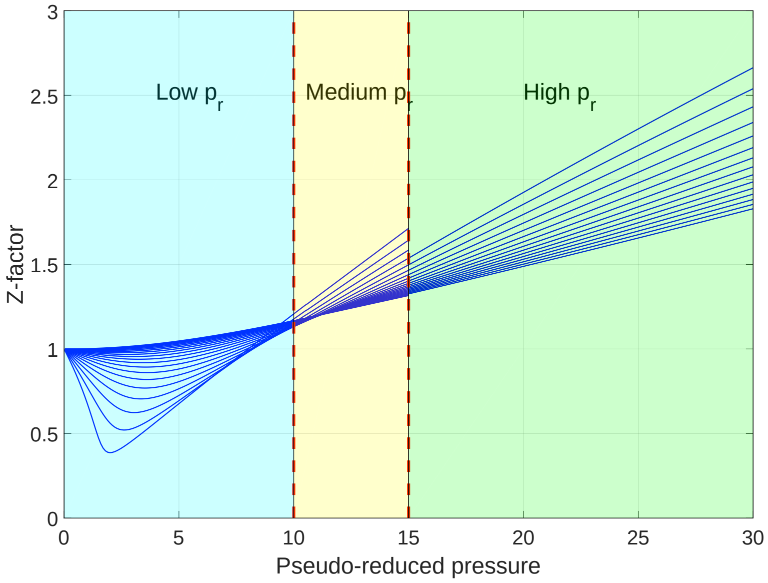

Figure 3. The standard S-K chart and its extension to high reduced pressures.

2.1. The Low Pressure Range

In this work, the utilization of the Kernel Ridge Regression method (KRR), to be discussed in detail in Appendix A, is proposed as the appropriate machine learning tool to deal with the prediction of the Z-factor at low reduced pressures. This method aims at generating a model $\hat{y} = f(\mathbf{x})$ that relates optimally a set of N recorded input and output pairs, $\mathbf{x}_i$ and y_i, $i \in [1, N]$, respectively, usually observed through an experimental process. In the present case, the input and output correspond to $\mathbf{x} = [T_r, p_r], \mathbf{x} \in \mathbb{R}^2$ and $y = Z(T_r, p_r)$, and they are obtained by digitizing the S-K chart. Therefore, for the low pressure range, the Z-factor is given by an expression of the form:

$$\hat{y} = \boldsymbol{\alpha}^T \mathbf{k}(\mathbf{x}) \tag{5}$$

where:

$$\mathbf{k} = [k(\mathbf{x}, \mathbf{x}_1), \ldots, k(\mathbf{x}, \mathbf{x}_N)]^T \tag{6}$$

and vector $\mathbf{k}$ contains kernel functions $k(\mathbf{x}, \mathbf{x}_i)$ that involve the conditions $\mathbf{x}$ at which the Z-factor needs to be computed, as well as the pressure and temperature conditions of the training points $\mathbf{x}_i$. The common choices for the kernel function $k(\mathbf{x}, \mathbf{y})$ are the polynomial one $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T\mathbf{y} + 1)^c$ with $c \in \mathbb{N}^*$ as the polynomial degree and the Radial Basis Function (RBF) kernel $k(\mathbf{x}, \mathbf{y}) = \exp(-1/\sigma||\mathbf{x} - \mathbf{y}||^2)$ where $\sigma > 0$ is the width of the kernel [34].

The coefficients vector $\boldsymbol{\alpha}$ is computed as a function of the training data $\mathbf{x}_i$ and y_i, so as to minimize the deviation between the model estimates on all training data points $\hat{\mathbf{y}} = \mathbf{K}\boldsymbol{\alpha}$ and the digitized Z-factor values of all training data points $\mathbf{y}$. The kernel matrix is defined by $\mathbf{K} = \{k_{ij}\} = \{k(\mathbf{x}_i \mathbf{x}_j)\}$, where $\mathbf{x}_i$

and $\mathbf{x}_j$ correspond to the inputs of any pair of data points. To satisfy the deviation minimization requirement, $\boldsymbol{\alpha}$ is given by:

$$\boldsymbol{\alpha} = (\mathbf{K} + \lambda \mathbf{I}_N)^{-1}\mathbf{y} \tag{7}$$

where λ corresponds to a positive regularization parameter value that is introduced to avoid overfitting [35] and $\mathbf{I}_N$ is the identity matrix of size NxN.

Although $\boldsymbol{\alpha}$ is obtained from the solution of a linear system, the fact that matrix $\mathbf{K} + \lambda \mathbf{I}_N$ can be very large and dense renders the solution for $\boldsymbol{\alpha}$ in Equation (7) as a very slow process with time complexity of $O(n^3)$ [35]. To treat that issue, one should revert to iterative linear systems solving methods such as the *Conjugate Gradient* (CG), and placing a threshold on the number of iterations leads to what is called the truncated Newton. Maalouf and Homouz [35] combined KRR with the truncated Newton method and developed a TR-KRR algorithm that is very fast to train. Interested readers should refer to Maalouf and Homouz [35] for a detailed description of TR-KRR.

The data used for the development of the proposed low reduced pressure TR-KRR model have been obtained by digitizing the original Standing–Katz chart for natural hydrocarbon gases and its extension at higher pressures [12]. Firstly, the isotherms of the standard chart, that is $T_r \in [1.05, 1.45]$ in steps of 0.05, $T_r \in [1.5, 2.0]$ in steps of 0.1, and $T_r \in [2.2, 3.0]$ in steps of 0.2, were digitized for the full pressure range of $p_r \in [0.1, 10.5]$ in steps of 0.1, thus leading to a total of 105 points per isotherm.

Subsequently, the dataset was densified by computing the estimated Z-factor values for the missing isotherms, so as to get a fixed T_r density of 0.05. As such isotherms are not plotted on the S-K chart, they were inferred by invoking the H-Y method. More specifically, to estimate $Z_{SK}^{T_r}(p_r)$ at some reduced temperature T_r and reduced pressure p_r, the digitized values at the neighboring isotherms and at p_r need first to be picked from the chart, i.e., $Z_{SK}^{T_r-}(p_r)$ and $Z_{SK}^{T_r+}(p_r)$, where subscript SK denotes the S-K chart and superscripts T_r- and T_r+ denote the closest neighboring isotherms. As an example, consider the case of $T_r = 1.75$ (which is not shown in the original S-K chart) for which $T_r^- = 1.7$ and $T_r^+ = 1.8$. Subsequently, the corresponding Z-factor values at all three reduced temperatures T_r-, T_r, and T_r+ and at p_r are also computed by means of the H-Y method to obtain $Z_{HY}^{T_r-}(p_r)$, $Z_{HY}^{T_r}(p_r)$ and $Z_{HY}^{T_r+}(p_r)$. Finally, the Standing–Katz Z-factor estimate $\hat{Z}_{SK}^{T_r}$ at T_r is computed by aligning the digitized values to the computed H-Y ones:

$$\hat{Z}_{SK}^{T_r} = Z_{SK}^{T_r-} + (Z_{SK}^{T_r+} - Z_{SK}^{T_r-}) \frac{Z_{HY}^{T_r} - Z_{HY}^{T_r-}}{Z_{HY}^{T_r+} - Z_{HY}^{T_r-}} \tag{8}$$

This way, the neighboring digitized values of the SK chart at T_r^- and at T_r^+ are fully adopted and further combined to the curve shape provided by the H-Y method.

To ensure smooth behavior of the KRR model at values close to the pressure and temperature range boundary, additional densification was run by introducing isotherms at $T_r \in [1.06, 1.09]$ and at $T_r \in [2.96, 2.99]$, both in steps of 0.01. Similarly, the pressure range was densified at $p_r \in [0.012, 0.018]$ and at $p_r \in [10.42, 10.48]$ in steps of 0.002. Eventually, the total number of data points of the form $\{p_r, T_r, Z\}$ collected by this procedure was 5616.

2.2. The Medium Pressure Range

For reduced pressure values in $10 \leq p_r \leq 15$, a linear model is proposed that interpolates temperature-dependent endpoint Z-factor values at $p_r = 10$ and at $p_r = 15$ with pressure in the following form:

$$Z_M(p_r, T_r) = Z(T_r)|_{p_r=10} + \left(Z(T_r)|_{p_r=15} - Z(T_r)|_{p_r=10}\right) \frac{p_r - 10}{5} \tag{9}$$

The endpoint Z-factor values at $p_r = 10$ and at $p_r = 15$ have been acquired from the S-K chart and further interpolated with temperature by means of polynomials of T_r, which are given in Appendix B.

The plot of the digitized points and their interpolants for $Z(T_r)|_{p_r=10}$ and $Z(T_r)|_{p_r=15}$ are shown in Figures 4 and 5, respectively.

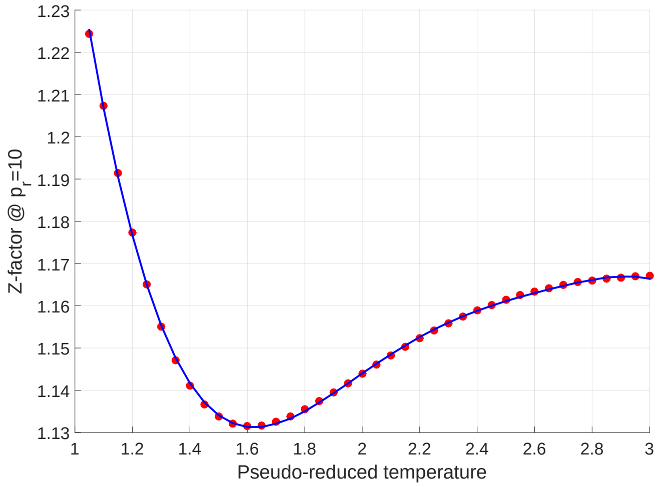

Figure 4. Z factor values at $p_r = 10$ for various T_r values.

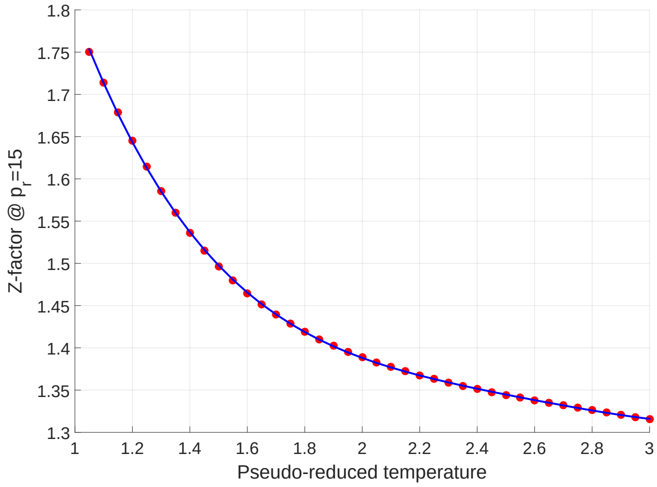

Figure 5. Z factor values at $p_r = 15$ for various T_r values.

2.3. The High Pressure Range

Although the Z-factor values in the high pressure range $15 \le p_r \le 30$ exhibit a straight line shape (as was the case with $p_r \in [10, 15]$), we propose the use of a quadratic model of reduced pressure to interpolate the temperature-dependent endpoint Z-factor values at $p_r = 15$ and at $p_r = 30$ with pressure. This additional degree of freedom allows the interpolating model not only to satisfy the endpoint values, but also to ensure continuity of the Z-factor derivative value at $p_r = 15$. This way, the continuity of the gas compressibility when switching from the medium pressure model $Z_M(p_r, T_r)$ to the high pressure one $Z_H(p_r, T_r)$ is guaranteed. Based on those requirements, the high pressure model is defined by:

$$Z_H(p_r, T_r) = a(T_r)p_r^2 + b(T_r)p_r + c(T_r) \tag{10}$$

where the temperature-dependent terms are given in Appendix C. The plot of $Z(T_r)|_{p_r=30}$ is shown in Figure 6.

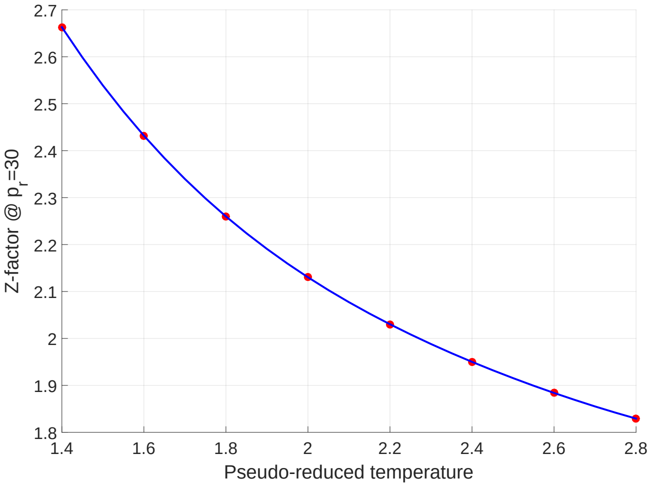

Figure 6. Z factor values at $p_r = 30$ for various T_r values.

2.4. The Combined Model

To further ensure the value and derivative continuity between $Z_L(p_r, T_r)$ and $Z_M(p_r, T_r)$, we recommend a weighted average scheme that switches mildly between the two models in the $p_r \in [10, 10.5]$ region by using:

$$Z_{(10-10.5)}(p_r, T_r) = wZ_L(p_r, T_r) + (1-w)Z_M(p_r, T_r) \tag{11}$$

where:

$$w = 1 - \frac{p_r - 10}{0.5} \tag{12}$$

Therefore, the combined Z-factor prediction algorithm is given by the following scheme:

$$Z(p_r, T_r) = \begin{cases} Z_L(p_r, T_r), & 0 \le p_r \le 10, 1.05 \le T_r \le 3.0, \\ wZ_L(p_r, T_r) + (1-w)Z_M(p_r, T_r), & 10 \le p_r \le 10.5, 1.05 \le T_r \le 3.0, \\ Z_M(p_r, T_r), & 10.5 \le p_r \le 15, 1.05 \le T_r \le 3.0, \\ Z_H(p_r, T_r), & 15 \le p_r \le 30, 1.4 \le T_r \le 2.8. \end{cases} \tag{13}$$

It is straightforward to show that derivative continuity at $p_r = 15$ is guaranteed, that is:

$$\frac{\partial}{\partial p_r} Z_M(p_r, T_r)\bigg|_{p_r=15} = \frac{Z(T_r)|_{p_r=15} - Z(T_r)|_{p_r=10}}{15-10} = \frac{\partial}{\partial p_r} Z_H(p_r, T_r)\bigg|_{p_r=15} \tag{14}$$

3. Results and Discussion

To obtain a reliable KRR model for the prediction of the Z-factor at $p_r \le 10$ parameters σ, λ, and N, corresponding to the RBF kernel width, the strength of the penalty term and the number of training data points, respectively, need to be determined. For that task, the ten-fold cross-validation technique was used to ensure generalization [36]. In the n-fold cross-validation and for each combination of possible parameters values, the data were split into n-folds, and $(n-1)$ folds were used for training, while the remaining fold was reserved for testing. This process was iterative until all folds were tested. Finally, the optimal model was selected so as to exhibit balanced error and minimum absolute relative error against the training and the validation datasets, as well as nearly equal training and prediction error. The examined range of the parameters was $\sigma \in [0.001, 0.1]$, $\lambda \in [0.0001, 0.1]$ and $N \in \{1000, 2000, 3000, 4000\}$. After repeatedly examining all possible combinations, it was found that optimal performance was obtained by selecting $\sigma = 0.01$, $\lambda = 0.001$ and $N = 3000$. Clearly, model accuracy is expected to be further improved when increasing the number of training data points. However, the selected population size is the minimum that led to a ratio of the training and validation errors very close to unity, thus minimizing the risk of overfitting.

The performance of the optimal TR-KRR model is shown in Table 1. Judging from the mean errors, the model was perfectly justified, exhibiting no bias. Moreover, it exhibited excellent agreement with the S-K chart as the mean absolute relative error was only 0.04% for both the training and the validation dataset, and the worst case error that might be observed was less than 2%, which is significantly better than all conventional methods. The scatter plots of the model performance for the training and prediction dataset are shown in Figures 7 and 8, respectively. A plot of the Z-factor surface that has been produced by interpolating the training data points together with the validation data points is shown in Figure 9.

In terms of computational time required for training, TR-KRR was very efficient, and this agrees with the findings of Maalouf and Homouz [35] that the biggest strength of TR-KRR method lies in its superior efficiency over other state-of-the-art methods, such as SVM [32]. In addition to predicting accurately the original temperature curves contained in the dataset, TR-KRR was capable of predicting new temperature curves. Figure 10 provides four examples of two additional temperature curves predicted by KRR at $T_r = 1.15$, $T_r = 1.45$, $T_r = 1.75$ and at $T_r = 2.65$. In all cases, the predicted curves fell smoothly between neighboring Standing–Katz curves.

Table 1. Performance of the Truncated Regularized Kernel Ridge Regression (TR-KRR) model.

	Training Dataset	Validation Dataset
Mean error	0.00	0.00
Mean relative error (%)	0.00	0.00
Mean absolute relative error (%)	0.04	0.04
Max error	0.01	0.01
Max relative error (%)	1.70	1.98
R^2	0.99997	0.99996

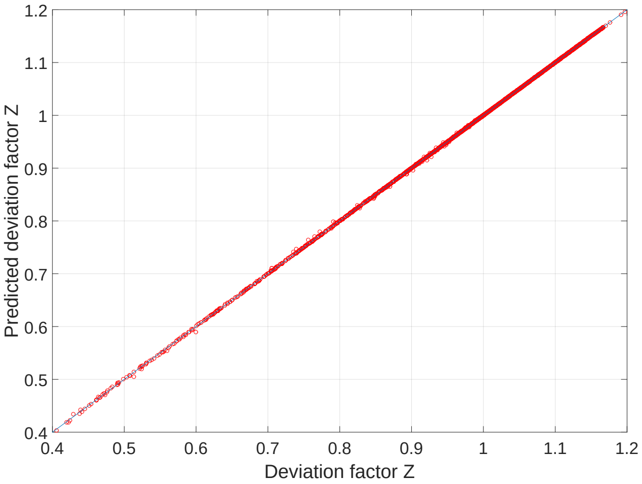

Figure 7. Performance of the TR-KRR model on the training dataset.

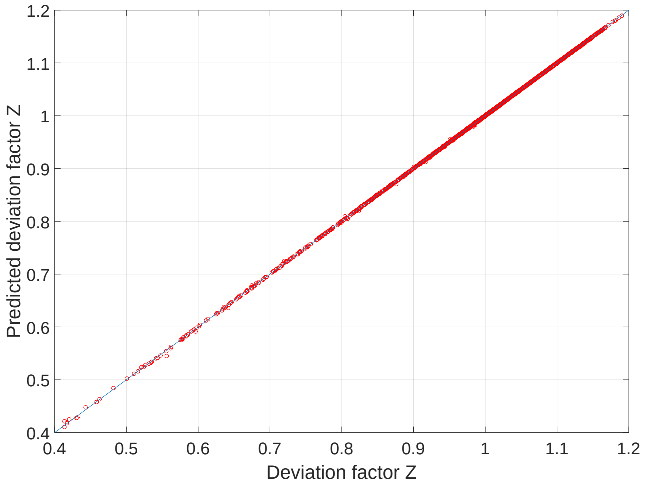

Figure 8. Performance of the TR-KKR model on the prediction dataset.

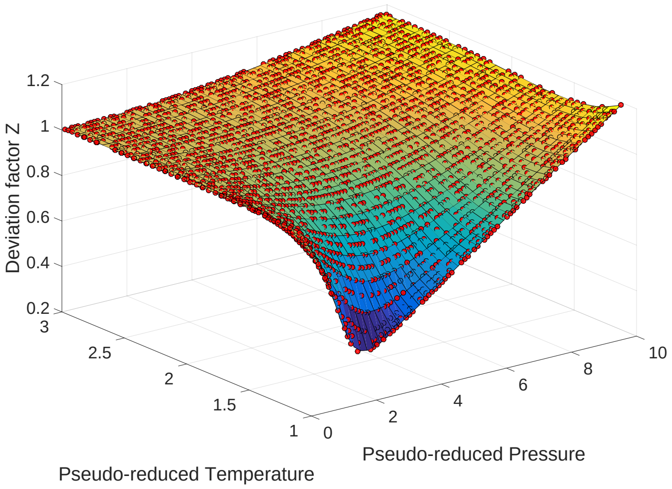

Figure 9. The Z factor surface generated by the TR-KRR model.

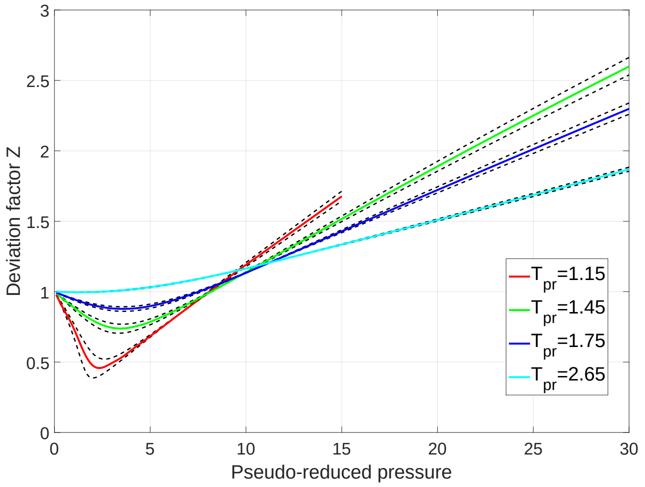

Figure 10. Examples of predicted isothermal Z factor curves vs p_r.

Implementing the TR-KRR prediction of the Z-factor for low p_r and any given T_r can be done conveniently according to the following simple formula:

$$Z = \sum_{i}^{N} \alpha_i d_i(p_r, T_r) \tag{15}$$

where

$$d_i = \exp\left(-\frac{1}{\sigma}\left[(\bar{T}_{r_i} - \bar{T}_r)^2 + (\bar{p}_{r_i} - \bar{p}_r)^2\right]\right) \tag{16}$$

p_{r_i}, T_{r_i} are the training population pseudo-critical temperature and pressure data, respectively, and the bar character denotes normalized inputs given by:

$$\bar{p}_r = \frac{p_r - \min(p_{r_i})}{\max(p_{r_i}) - \min(p_{r_i})} - \frac{1}{2} \quad \text{and} \quad \bar{T}_r = \frac{T_r - \min(T_{r_i})}{\max(T_{r_i}) - \min(T_{r_i})} - \frac{1}{2} \tag{17}$$

The derivative of the Z-factor with respect to pressure, needed to calculate the fluid's isothermal compressibility by means of Equation (4), is given by:

$$\frac{\theta Z}{\theta p} = \sum_{i=1}^{N} \frac{\theta Z}{\theta d_i} \frac{\theta d_i}{\theta \bar{p}_r} \frac{\theta \bar{p}_r}{\theta p_r} \frac{\theta p_r}{\theta p} = -\frac{2}{\sigma p_c\big(\max(p_{r_i}) - \min(p_{r_i})\big)} \sum_{i=1}^{N} \alpha_i d_i(\bar{p}_{r_i} - \bar{p}_r) \tag{18}$$

4. Case Studies

Three case studies are presented to further demonstrate the efficiency of the derived model.

4.1. Case Study 1

In this case study, we considered the effect of the Z-factor accuracy to that of the gas compressibility, which is a derived property according to Equation (4). The composition of the gas mixture studied was $z_{C_1} = 82\%, z_{C_2} = 10\%, z_{C_3} = 5\%, z_{C_4} = 3\%$, and its critical pressure and temperature were estimated at 670 psi and 215 F, respectively, by means of Standing's gas specific gravity-based correlations. The isothermal compressibility of the gas mixture has been computed at a temperature of 236 F and for a pressure range from nearly atmospheric to 5000 psi. The methods utilized were the most commonly-used ones, that is the iterative algorithm of Hall and Yarborough, the correlation of Beggs and Brill, and the one proposed in this work. Note that the test temperature has been deliberately selected so as to demonstrate the performance of the examined methods at near critical conditions.

The performance of all methods and their comparison against the exact S-K chart is shown in Figure 11.

The compressibility of the ideal gas (i.e., $c = p^{-1}$), obtained by totally neglecting the deviating effect of the Z-factor, is also shown. Judging from the difference between the ideal gas compressibility and that obtained by any other method, it can be readily seen that the effect of the Z-factor on the gas compressibility was quite significant. Indeed, at high pressures close to 5000 psi, introducing the Z-factor effect reduced the compressibility by almost one order of magnitude, thus verifying the need for accurate Z-factor determination algorithms.

As expected, at very low pressures, where the gas behaves almost ideally, the effect of the Z-factor was minimal; hence, all methods ended up with very close isothermal compressibility estimates. However, this was not the case at higher pressures where the Beggs and Brill correlation performed quite poorly and the computed compressibility exhibited deviations as high as 47% at the near critical region. For example, at 1500 psi, the compressibility value obtained by the original S-K chart was 4.3×10^{-4} psi^{-1}, whereas the Beggs and Brill one was only 2.3×10^{-4} psi^{-1}. On the other hand, the H-Y method, despite its convergence issues, performed better by predicting 5.4×10^{-4} psi^{-1}, leading to an error of 25%. The method proposed in this work was significantly more accurate, as it led to a value of 4.5×10^{-4} psi^{-1}, which corresponds to a relative error of only 5%.

A similar situation was observed at high pressures where the S-K value was 3.1×10^{-5} psi^{-1}. The values obtained by Beggs and Brill, Hall and Yarborough, and the new method were 2.3×10^{-5} psi^{-1}, 2.8×10^{-5} psi^{-1} and 3.0×10^{-5} psi^{-1}, leading to relative errors of 25%, 10%, and 3%, respectively, thus demonstrating the superiority of the new method.

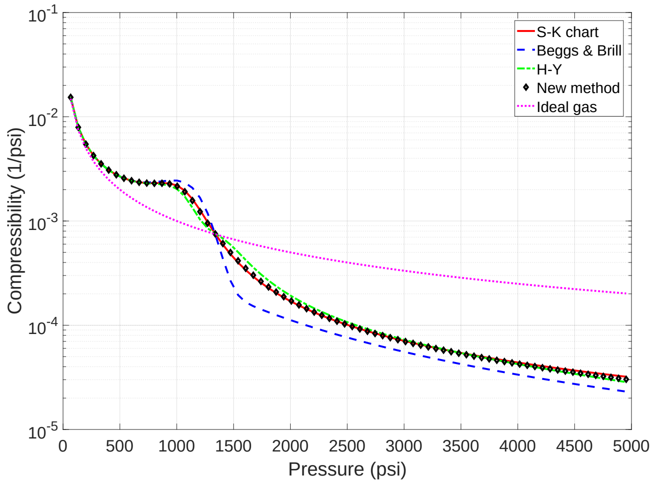

Figure 11. Performance of four methods for the prediction of natural gas compressibility.

4.2. Case Study 2

We now consider the dependency of the pressure drop of a gas stream flowing in a pipeline on the accuracy of the utilized Z-factor values. The gas was assumed to be a dry one, the exact composition of which was $z_{CH_4} = 0.78$, $z_{C_2H_6} = 0.04$, $z_{C_3H_8} = 0.02$, $z_{C_4H_{10}} = 0.01$, $z_{CO_2} = 0.15$ with a molar mass value of 21.74 g/mole and γ_g = 0.7504. The gas was assumed to be compressed at an initial pressure of 1200 psi and introduced to a 30 mile-long pipeline of internal diameter of 6″ at ambient temperature (60 F) and at a flow rate of 30 MMscf/d. The flow was assumed to be isothermal, and the heating effect of gas compression was ignored. Moreover, to simplify calculations, it was also assumed that the pipeline network exhibited no elevation changes. The outlet pressure for a given pipeline length can be estimated by:

$$p_{out} = \sqrt{p_{in}^2 - \frac{s_g f Z T L}{D^5} \left(\frac{q_{SC}}{77.54} \frac{p_b}{T_b} \right)^2} \quad (19)$$

where f is the friction factor, p_{in} and p_{out} is the inlet and outlet pressure (psi), T is the pipeline temperature (R), L is the pipeline length (miles), D is the diameter (in), p_b, T_b are the standard conditions (psi and R, respectively), and q_{SC} is the flow rate reported at standard conditions (scf/d).

From the expression above, it is clear that the outlet pressure for given pipeline geometry and gas pump efficiency was directly related to the Z-factor. Therefore, various outlet pressures will be obtained depending on the method utilized to estimate the deviation factor. In this example, the methods of Hall and Yarborough, Beggs and Brill, the proposed method in this work, and the direct use of the Standing–Katz chart were utilized. The pseudo-critical properties were computed as functions of gas s_g using Standing's recommended expressions, thus leading to $p_{pc} = 667.1$ psi and $T_{pc} = 404.9$ R. Correspondingly, the reduced conditions in this test case were equal to $p_{pr} = 1.8$ and $T_{pr} = 1.283$. The resulting computed outlet pressures are shown in Table 2.

It can be readily seen that although the prevailing conditions were well within the valid range of the tried methods, the Beggs and Brill correlation exhibited significant deviation from the original S-K chart, which, eventually, led to a significant difference of 4 bar, whereas the H-Y methods performed better, but still exhibited an error of 2.5 bar. On the other hand, the proposed approach provided an accurate estimate of the Z-factor, which, in turn, led to an outlet pressure deviation of 10 psi, i.e., less than 1 bar.

Table 2. Performance of the various methods on the prediction of the pipeline outlet pressure (Case Study 2).

Method	Z-Factor	p_{out} (psi)	Deviation (psi)
Standing–Katz	0.6950	337	-
Beggs and Brill	0.7137	278	59
Hall and Yarborough	0.7071	300	37
This method	0.6912	347	10

4.3. Case Study 3

The interpretation of production data from a gas reservoir so as to estimate the reservoir reserves by means of the gas material balance method is demonstrated in this case study. When applied to a dry gas reservoir between production start and current time, the material balance equation states that:

$$\frac{p}{Z(p)} = \frac{p_i}{Z(p_i)} - \frac{p_i}{Z(p_i)}\frac{G_p}{G} \tag{20}$$

where p and G_p refer to the pressure and recorded cumulative gas production at the current time instance, while p_i denotes the initial reservoir pressure, and G corresponds to the reserves. Evidently, by recording a single pressure drop and the observed gas production, the equation can be directly solved for G. However, as such measurements (especially the pressure one) are prone to errors, most engineers prefer to keep recording p and G_p for a longer period and then compute the statistically-optimal G value in the least squares sense. By rewriting the above equation in the following form:

$$\frac{p}{Z(p)} = \frac{p_i}{Z(p_i)} - \left(\frac{1}{G}\frac{p_i}{Z(p_i)}\right)G_p \tag{21}$$

it becomes clear that the observed p/Z ratio at any pressure is a linear function of the cumulative production G_p. Therefore, by fitting the p/Z ratios with a linear function of the form $p/Z = aG_p + b$ and considering that $G_p = G$ when $p = 0$, i.e., that the reserves will be fully produced only when the reservoir pressure becomes equal to zero, we obtain:

$$\frac{p}{Z} = a + bG_p \Rightarrow G = \lim_{p\to 0} G_p = -\frac{a}{b} \tag{22}$$

In this example, the pressure and production data, shown in Table 3, from a large gas field have been utilized. The fluid composition and properties were the same to those in Case study 1 and so were the computational methods used to estimate the p/Z ratios. The estimated gas reserves G as obtained by each method are shown in Table 4.

Table 3. Production data and the predicted Z-factor values at various pressure steps (Case Study 3).

P(psi)	T(F)	$G_p(Bscf)$	p_{pr}	T_{pr}	Z (B-B)	Z (H-Y)	Z (S-K Chart)	Z (This Work)
3600	150	0.00	5.48	1.51	0.826	0.842	0.833	0.832
3450	150	4.78	5.25	1.51	0.814	0.830	0.822	0.820
3300	150	12.65	5.02	1.51	0.804	0.820	0.811	0.809
3150	150	20.48	4.79	1.51	0.795	0.810	0.800	0.798
2850	150	38.25	4.34	1.51	0.781	0.794	0.785	0.783
2685	150	44.01	4.09	1.51	0.776	0.788	0.780	0.778

Table 4. Reserves' prediction by various methods (Case Study 3).

Method	Reserves Estimate (Bscf)	Deviation (Bscf)
Beggs and Brill	224.4	4.36
Hall and Yarborough	227.7	1.05
Standing–Katz chart	228.7	-
This work	229.3	0.61

As expected, the results exhibited significant differences depending on the method utilized. Once again, the Beggs and Brill method exhibited worst performance, leading to a missed gas volume of 4.36 billion scf of gas, whereas the Hall and Yarborough method performed better by limiting the missed reserves to 1.05 Bscf. On the other hand, the method proposed in this work underestimated reserves only by 0.61 Bscf, which is the most accurate estimate compared to all other approaches. It should be noted that, unlike density calculations, when it comes to estimating hydrocarbon reserves, the exact value of the Z-factor did not really affect the results. To show that, we rewrite Equation (20) by replacing the Z-factor with a random multiple of that, that is:

$$\frac{p}{aZ(p)} = \frac{p_i}{aZ(p_i)} - \frac{p_i}{aZ(p_i)}\frac{G_p}{G} \tag{23}$$

The reserves estimate from a single measurement (p, G_p) is given by:

$$G = \frac{p_i/(aZ(p_i))}{p_i/(aZ(p_i)) - p/(aZ(p))} G_p \tag{24}$$

where, clearly, parameter a cancels, thus showing that any multiple of the Z-factor will lead to the same reserves estimate. As was the case with the prediction of gas compressibility, this result demonstrates the need for the Z-factor prediction models to preserve the shape of the original S-K curves, thus exhibiting physically-sound derivatives and isothermal compressibility values.

5. Conclusions

In this work, a new, efficient, consistent, and physically-sound method is presented for the prediction of the Z-factor of natural gas streams. Our conclusions are summarized as follows:

- At low reduced pressures where critical behavior might be observed and dependency on pressure is quite complex, the proposed method utilized the non-linear regression TR-KRR modeling technique to predict the gas compressibility factor.
- Our results indicated that despite the abruptly changing slope of the Z-factor isotherms, the TR-KRR model was highly accurate and efficient at predicting Z-factor.
- The simplicity of the original S-K chart and its extension at medium and high pressures was directly inherited by the corresponding linear and quadratic submodels developed for the prediction of the Z-factor.
- Special attention has been paid to ensure a natural model derivative behavior, so as to end up with reliable isothermal compressibility values.

Author Contributions: Conceptualization, M.M. and N.K.; methodology, M.M. and D.H.; software, V.G.; validation, V.G. and D.H.; formal analysis, V.G., D.H., M.M., and K.P.; investigation, K.P., M.M., V.G., N.K., and D.H.; resources, K.P., M.M., D.H., and V.G.; data curation, V.G., M.M., and D.H.; writing, original draft preparation, M.M., K.P., and V.G.; visualization, V.G. and D.H.; funding acquisition, K.P.

Funding: This publication is based on work supported by the Khalifa University of Science and Technology under Award No. RC2-2018-024.

Acknowledgments: The authors would like to acknowledge the support by Khalifa University of Science and Technology under Award No. RC2-2018-024.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Details of the Z_L Model

Regression [37] aims at generating a model $y = f(\mathbf{x})$ that relates optimally a set of N recorded input-output pairs $\mathbf{x}_i$ and y_i, $i \in [1, N]$, respectively, usually observed through an experimental process. The input vector may contain $d-1$ features, also known as parameters or attributes, whereas, for convenience, unity is further added to the input vector, so that $\mathbf{x} \in \mathbb{R}^d$. Linear regression assumes a vector $\boldsymbol{\beta}$ such that:

$$y = f(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta} + \epsilon = \hat{y} + \epsilon \tag{A1}$$

By collecting all inputs in matrix $\mathbf{X} \in \mathbb{R}^{N\times d}$ and all outputs (responses) in $\mathbf{y} \in \mathbb{R}^N$, linear regression for the full dataset generalizes to:

$$\mathbf{y} = f(\mathbf{X}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \hat{\mathbf{y}} + \boldsymbol{\epsilon} \tag{A2}$$

The weights vector $\boldsymbol{\beta}$ needs to be optimized so that error $\boldsymbol{\epsilon}$ exhibits its minimum value over all available training pairs, and the optimization problem is defined by:

$$\boldsymbol{\beta} = \operatorname{argmin}\big(J(\boldsymbol{\beta})\big) \tag{A3}$$

where the total error to be minimized is given by:

$$J(\boldsymbol{\beta}) = \frac{1}{2}\sum_{i=1}^{i=N}\epsilon_i^2 = \boldsymbol{\epsilon}^T\boldsymbol{\epsilon} = \frac{1}{2}(\mathbf{y}-\hat{\mathbf{y}})^T(\mathbf{y}-\hat{\mathbf{y}}) = \frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \tag{A4}$$

The closed-form solution of this optimization problem, that is $\boldsymbol{\beta}$ for which $\nabla_{\boldsymbol{\beta}}J = \mathbf{0}$, is given by the Moore–Penrose inverse [38] of the data matrix $\mathbf{X}$, that is:

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{A5}$$

Regularization [39] needs to be used to avoid poor estimation of the regression coefficients due to the instability of the solution of Equation (A5). This idea lies in "penalizing" high coefficients values, which, essentially, improves the rank of the covariance matrix $\mathbf{X}^T\mathbf{X}$. Therefore, ridge regression aims at minimizing the modified objective function:

$$J(\boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{\epsilon}^T\boldsymbol{\epsilon} = \frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \frac{\lambda}{2}\boldsymbol{\beta}^T\boldsymbol{\beta} \tag{A6}$$

which is optimized at:

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}_d)^{-1}\mathbf{X}^T\mathbf{y} \tag{A7}$$

where $\mathbf{I}_d$ is a $d \times d$ identity matrix and λ is a positive constant that adjusts the penalty term.

In order to extend to non-linear relationships, the dual form [40] needs firstly to be introduced. Let $\boldsymbol{\beta}$ be expressed as a linear combination of the data points, i.e.,

$$\boldsymbol{\beta} = \mathbf{X}^T\boldsymbol{\alpha} \tag{A8}$$

By replacing back in the regression model, we obtain:

$$\mathbf{y} = \mathbf{X}\mathbf{X}^T\boldsymbol{\alpha} + \boldsymbol{\epsilon} = \mathbf{G}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \tag{A9}$$

where the Gramian matrix $\mathbf{G} = \mathbf{X}\mathbf{X}^T$. This way, the objective function to be minimized is now expressed in terms of the new coefficients vector $\boldsymbol{\alpha}$ and turns into:

$$f(\boldsymbol{\alpha}) = \frac{1}{2}(\mathbf{y}-\mathbf{G}\boldsymbol{\alpha})^T(\mathbf{y}-\mathbf{G}\boldsymbol{\alpha}) + \frac{\lambda}{2}\boldsymbol{\alpha}^T\boldsymbol{\alpha} \tag{A10}$$

and it is optimized at:

$$\boldsymbol{\alpha} = (\mathbf{G} + \lambda \mathbf{I}_N)^{-1}\mathbf{y} \tag{A11}$$

In this case, the benefit of the dual form is that solution $\boldsymbol{\alpha}$ depends on $\mathbf{G}$, that is only on dot products of the input vectors $\mathbf{x}_i$. The price to be paid is that the size of $\boldsymbol{\alpha}$ is equal to the number of data points N, whereas the size of $\boldsymbol{\beta}$ is equal to the number of features d.

Kernel Ridge Regression (KRR) [34] can be thought of as an extension of the linear regression dual form, so as to handle non-linear relationships between the input and output. This is done simply by including a nonlinear map, $\boldsymbol{\phi}(.)$, which projects the original input data $\mathbf{x}$ into a high-dimensional, even infinite-dimensional, feature space $\mathbb{F}$ such that:

$$\boldsymbol{\phi} : \mathbf{x} \in \mathbb{R}^d \to \boldsymbol{\phi}(\mathbf{x}) \in \mathbb{F}. \tag{A12}$$

thus leading to the generalized Gramian $\mathbf{G} = \boldsymbol{\Phi}(\mathbf{X})\boldsymbol{\Phi}(\mathbf{X})^T$ and the regression model:

$$\mathbf{y} = \boldsymbol{\Phi}(\mathbf{X})\boldsymbol{\Phi}(\mathbf{X})^T\boldsymbol{\alpha} + \boldsymbol{\epsilon} = \mathbf{G}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \tag{A13}$$

This way output, $\mathbf{y}$ is non-linear in the input $\mathbf{x}$, but it is still linear in the coefficients $\boldsymbol{\alpha}$. In general, the non-linear mapping $\boldsymbol{\phi}(.)$ is unknown; however, the solution is based mainly on linear dot products of the images $\boldsymbol{\phi}(\mathbf{x})$. To overcome the difficulty in working with dot products of high dimensionality in the feature space $\mathbb{F}$, kernel functions are utilized, which compute a dot product in $\mathbb{F}$ as a function of the dot product of the original inputs, that is $\boldsymbol{\phi}(\mathbf{x})^T\boldsymbol{\phi}(\mathbf{y}) = k(\mathbf{x}^T\mathbf{y})$ for some suitable scalar function $k(.)$. The most commonly-used kernels are the polynomial kernel $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T\mathbf{y} + 1)^c$ with $c \in \mathbb{N}^*$ as the polynomial degree and the Radial Basis Function (RBF) kernel $k(\mathbf{x}, \mathbf{y}) = \exp(-1/\sigma||\mathbf{x} - \mathbf{y}||^2)$ where $\sigma > 0$ is the width of the kernel [34]. Based on that, the Kernel Ridge Regression model (KRR) [35] is given by:

$$\mathbf{y} = \mathbf{K}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \tag{A14}$$

where $\mathbf{K} = \{k_{ij}\} = \{k(\mathbf{x}_i\mathbf{x}_j)\}$, and the coefficients vector $\boldsymbol{\alpha}$ is computed by minimizing:

$$f(\boldsymbol{\alpha}) = \frac{1}{2}(\mathbf{y} - \mathbf{K}\boldsymbol{\alpha})^{\mathrm{T}}(\mathbf{y} - \mathbf{K}\boldsymbol{\alpha}) + \frac{\lambda}{2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha} \tag{A15}$$

The solution now is given by:

$$\boldsymbol{\alpha} = (\mathbf{K} + \lambda \mathbf{I}_N)^{-1}\mathbf{y} \tag{A16}$$

Although $\boldsymbol{\alpha}$ is obtained from the solution of a linear system, the fact that matrix $\mathbf{K} + \lambda\mathbf{I}_N$ can be very large and dense renders the solution for $\boldsymbol{\alpha}$ in Equation (A16) as a very slow process with time complexity of $O(n^3)$ [35]. To treat that issue, one should revert to iterative linear systems, solving methods such as the CG, and placing a threshold on the number of iterations leads to what is called the truncated Newton. Maalouf and Homouz [35] combined KRR with the truncated Newton method and developed a TR-KRR algorithm that is very fast to train. Interested readers should refer to Maalouf and Homouz [35] for a detailed description of TR-KRR.

Once the coefficients values $\boldsymbol{\alpha}$ have been optimized, the predictive model for any new input $\mathbf{x}$ is given by the following simple, linear-in-the-weights model:

$$y = \boldsymbol{\alpha}^T\mathbf{k}(\mathbf{x}) \tag{A17}$$

where:

$$\mathbf{k} = [k(\mathbf{x}, \mathbf{x}_1), \ldots, k(\mathbf{x}, \mathbf{x}_N)]^T \tag{A18}$$

In this work, TR-KRR is terminated when the CG residual is less than a threshold $\varepsilon = 0.5$ or the maximum number of 200 CG iterations are reached.

Algorithm 1: Linear CG for computing $\hat{\boldsymbol{\alpha}}$. $\mathbf{A} = \mathbf{K} + \lambda\mathbf{I}_N$, $\mathbf{b} = \mathbf{y}$

```
Data: $\mathbf{A}, \mathbf{b}, \hat{\boldsymbol{\alpha}}^{(0)}$
Result: $\hat{\boldsymbol{\alpha}}$ such that $\mathbf{A}\hat{\boldsymbol{\alpha}} = \mathbf{b}$
begin
    $\mathbf{r}^{(0)} = \mathbf{b} - \mathbf{A}\hat{\boldsymbol{\alpha}}^{(0)}$                /* Initialize the residual */
    $c = 0$
    while $||\mathbf{r}^{(c+1)}||^2 > \varepsilon_2$ and $c \leq$ Max CG Iterations do
        if $c = 0$ then
            $\zeta^{(c)} = 0$
        else
            $\zeta^{(c)} = \frac{\mathbf{r}^{\mathrm{T}(c+1)}\mathbf{r}^{(c+1)}}{\mathbf{r}^{\mathrm{T}(c+1)}\mathbf{r}^{(c)}}$      /* Update A-Conjugacy enforcer */
        $\mathbf{d}^{(c+1)} = \mathbf{r}^{(c+1)} + \zeta^{(c)}\mathbf{d}^{(c)}$          /* Update the search direction */
        $s^{(c)} = -\frac{\mathbf{r}^{\mathrm{T}(c)}\mathbf{r}^{(c)}}{\mathbf{d}^{\mathrm{T}(c)}\mathbf{A}\mathbf{d}^{c}}$          /* Compute the optimal step length */
        $\hat{\boldsymbol{\alpha}}^{(c+1)} = \hat{\boldsymbol{\alpha}}^{(c)} - s^{(c)}\mathbf{d}^{(c)}$          /* Obtain an approximate solution */
        $\mathbf{r}^{(c+1)} = \mathbf{r}^{(c)} - \mathbf{A}\boldsymbol{\alpha}^{(c)}$                  /* Update the residual */
        $c = c + 1$
```

Appendix B. Details of the Z_M Model

The linear model providing the Z-factor in $10 \leq p_r \leq 15$ interpolates the endpoint Z-factor values at $p_r = 10$ and at $p_r = 15$ in the following form:

$$Z_M(p_r, T_r) = Z(T_r)|_{p_r=10} + \left(Z(T_r)|_{p_r=15} - Z(T_r)|_{p_r=10}\right)\frac{p_r - 10}{5} \tag{A19}$$

The temperature-dependent endpoint Z-factor values have been acquired from the S-K chart and further interpolated by means of polynomials of T_r, and they are defined by the following expressions:

$$\begin{aligned} Z(T_r)|_{p_r=10} = {} & -0.024466T_r^6 + 0.284414T_r^5 - 1.281582T_r^4 + 2.712782T_r^3 \\ & - 2.407661T_r^2 + 0.079238T_r + 1.883774 \end{aligned} \tag{A20}$$

and:

$$Z(T_r)|_{p_r=15} = 0.048200T_r^4 - 0.502345T_r^3 + 1.977248T_r^2 - 3.546957T_r + 3.820608 \tag{A21}$$

Appendix C. Details of the Z_H Model

To obtain Z-factor values in the high pressure range $15 \leq p_r \leq 30$, the digitized endpoint Z-factor values at $p_r = 15$ and at $p_r = 30$ are interpolated by the following quadratic model:

$$Z_H(p_r, T_r) = a(T_r)p_r^2 + b(T_r)p_r + c(T_r) \tag{A22}$$

where:

$$a(T_r) = \frac{Z(T_r)|_{p_r=30} - 4Z(T_r)|_{p_r=15} + 3Z(T_r)|_{p_r=10}}{225} \tag{A23}$$

$$b(T_r) = \frac{Z(T_r)|_{p_r=15} - Z(T_r)|_{p_r=10}}{5} - 30a(T_r) \tag{A24}$$

$$c(T_r) = Z(T_r)|_{p_r=15} - 225a(T_r) - 15b(T_r); \tag{A25}$$

and:

$$Z(T_r)|_{p_r=30} = 0.090371T_r^4 - 0.957066T_r^3 + 3.938661T_r^2 - 7.726749T_r + 8.039752 \tag{A26}$$

From Equation (A23), it can be seen that the denominator is very large compared to the nominator, thus verifying that the introduced curvature due to parameter a is very small and that $Z_H(p_r, T_r)$ is very close to a straight line.

References

1. Tan, S.P.; Qiu, X.; Dejam, M.; Adidharma, H. Critical point of fluid confined in nanopores: Experimental detection and measurement. *J. Phys. Chem. C* **2019**, *123*, 9824–9830. [CrossRef]
2. Qiu, X.; Tan, S.P.; Dejam, M.; Adidharma, H. Simple and accurate isochoric differential scanning calorimetry measurements: Phase transitions for pure fluids and mixtures in nanopores. *Phys. Chem. Chem. Phys.* **2019**, *21*, 224–231. [CrossRef] [PubMed]
3. Qiu, X.; Tan, S.P.; Dejam, M.; Adidharma, H. Novel isochoric measurement of the onset of vapor-liquid phase transition using differential scanning calorimetry. *Phys. Chem. Chem. Phys.* **2018**, *20*, 26241–26248. [CrossRef] [PubMed]
4. Nikpoor, M.H.; Dejam, M.; Chen, Z.; Clarke, M. Chemical-gravity-thermal diffusion equilibrium in two-phase non-isothermal petroleum reservoirs. *Energy Fuels* **2016**, *30*, 2021–2034. [CrossRef]
5. Standing, M. *Volumetric and Phase Behavior of Oil Field Hydrocarbon Systems: PVT for Engineers*; California Research Corporation: Vancouver, BC, Canada, 1951.
6. Amyx, J.W.; Bass, D.M.; Whiting, R.L. *Petroleum Reservoir Engineering: Physical Properties*; McGraw-Hill: New York, NY, USA, 1960.
7. Redlich, O.; Kwong, J.N. On the thermodynamics of solutions. v. an equation of state. Fugacities of gaseous solutions. *Chem. Rev.* **1949**, *44*, 233–244. [CrossRef] [PubMed]
8. Whitson, C.; Brule, M. *Phase Behavior*; SPE: Richardson, TX, USA, 2000.
9. Reid, R.C.; Prausnitz, J.M.; Poling, B.E. *The Properties of Gases and Liquids*, 4th ed.; McGraw-Hill: New York, NY, USA, 1987.
10. Zudkevitch, D.; Joffe, J. Correlation and prediction of Vapor-Liquid Equilibrium with the Redlich-Kwong Equation of State. *AIChE J.* **1970**, *16*, 112. [CrossRef]
11. Hayden, J.G.; O'Connell, J.P. A generalized method for predicting second virial coefficients. *Ind. Eng. Chem. Proc. Des. Dev.* **1975**, *14*, 209–216. [CrossRef]
12. Katz, D.L.; Cornell, D.; Kobayashi, R.; Poettmann, F.H.; Vary, J.A.; Elenbaas, J.R.; Weinaug, C.F. *Handbook of Natural Gas Engineering*; McGraw-Hill: New York, NY, USA, 1959.
13. Wichert, E.; Aziz, K. Compressibility Factor of Sour Natural Gases. *Cdn. J. Chem. Eng.* **1975**, *49*, 267. [CrossRef]
14. Kay, W. Gases and vapors at high temperature and pressure-density of hydrocarbon. *Ind. Eng. Chem.* **1936**, *28*, 1014–1019. [CrossRef]
15. Stewart, W.; Burkhardt, S.; Voo, D. Prediction of pseudo-critical parameters for mixtures. In Proceedings of the AIChE Meeting, Kansas City, MO, USA, 18 May 1959; Volume 18.
16. Sutton, R. Compressibility factors for high-molecular-weight reservoir gases. In Proceedings of the SPE Annual Technical Conference and Exhibition, Las Vegas, NV, USA, 22–26 September 1985; Society of Petroleum Engineers: London, UK, 1985.
17. Danesh, A. *Pvt and Phase Behaviour of Petroleum Reservoir Fluids*; Developments in Petroleum Science, Elsevier: Amsterdam, The Netherlands, 1998.
18. Hall, K.R.; Yarborough, L. A New EOS for Z-factor Calculations. *Oil Gas J.* **1973**, *71*, 82–92.
19. Dranchuk, P.; Abou-Kassem, H. Calculation of z factors for natural gases using equations of state. *J. Can. Pet. Technol.* **1975**, *14*. [CrossRef]
20. Brill, J.P.; Beggs, H.D. *Two-Phase Flow in Pipes*; University of Tulsa INTERCOMP Course: The Hague, The Netherlands, 1974.
21. Azizi, N.; Behbahani, R.; Isazadeh, M. An efficient correlation for calculating compressibility factor of natural gases. *J. Nat. Gas Chem.* **2010**, *19*, 642–645. [CrossRef]
22. Kumar, N. Compressibility Factor for Natural and Sour Reservoir Gases by Correlations and Cubic Equations of State. Master's Thesis, Texas Tech University, Lubbock, TX, USA, 2004.
23. Heidaryan, E.; Moghadasi, J.; Rahimi, M. New correlations to predict natural gas viscosity and compressibility factor. *Fluid Phase Equilibria* **2009**, *218*, 1–13. [CrossRef]

24. Kareem, L.; Iwalewa, T.; Al-Marhoun, M. New explicit correlation for the compressibility factor of natural gas: linearized z-factor isotherms. *J. Petrol Explor. Prod. Technol.* **2016**, *6*, 481–492. [CrossRef]
25. Elsharkawy, A.M. Efficient methods for calculations of compressibility, density and viscosity of natural gases. *Fluid Phase Equilibria* **2004**, *218*, 1–13. [CrossRef]
26. Moghadassi, A.; Parvizian, F.; Hosseini, S.; Fazlali, A. A new approach for estimation of pvt properties of pure gases based on artificial neural network model. *Braz. J. Chem. Eng.* **2009**, *26*, 199–206. [CrossRef]
27. Irene, A.I.; Sunday, I.S.; Orodu, O.D. Forecasting Gas Compressibility Factor Using Artificial Neural Network Tool for Niger-Delta Gas Reservoir. *Soc. Pet. Eng.* **2016**. [CrossRef]
28. Kamyab, M.; Sampaio, J.H.; Qanbari, F.; Eustes, A.W. Using artificial neural networks to estimate the z-factor for natural hydrocarbon gases. *J. Pet. Sci. Eng.* **2010**, *73*, 248–257. [CrossRef]
29. Sanjari, E.; Lay, E.N. Estimation of natural gas compressibility factors using artificial neural network approach. *J. Nat. Gas Sci. Eng.* **2012**, *9*, 220–226. [CrossRef]
30. Fayazi, A.; Arabloo, M.; Mohammadi, A.H. Efficient estimation of natural gas compressibility factor using a rigorous method. *J. Nat. Gas Sci. Eng.* **2014**, *16*, 8–17. [CrossRef]
31. Kamari, A.; Hemmati-Sarapardeh, A.; Mirabbasi, S.-M.; Nikookar, M.; Mohammadi, A.H. Prediction of sour gas compressibility factor using an intelligent approach. *Fuel Process. Technol.* **2013**, *116*, 209–216. [CrossRef]
32. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, UK, 2000.
33. Mohamadi-Baghmolaei, M.; Azin, R.; Osfouri, S.; Mohamadi-Baghmolaei, R.; Zarei, Z. Prediction of gas compressibility factor using intelligent models. *Nat. Gas Ind. B* **2015**, *2*, 283–294. [CrossRef]
34. Shawe-Taylor, J.; Cristianini, N. *Kernel Methods for Pattern Analysis*; Cambridge University Press: Cambridge, UK, 2004.
35. Maalouf, M.; Homouz, D. Kernel ridge regression using truncated newton method. *Knowl.-Based Syst.* **2014**, *71*, 339–344. [CrossRef]
36. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman & Hall/CRC: Vancouver, BC, Canada, 1994.
37. Rencher, A. *Methods of Multivariate Analysis*; Wiley Interscience: Hoboken, NJ, USA, 2002.
38. Shores, T. *Applied Linear Algebra and Matrix Analysis*; Springer: Berlin/Heidelberg, Germany, 2007.
39. Bishop, C.M. *Pattern Recognition and Machine Learning*; Oxford University Press: Oxford, UK, 2006.
40. Nocedal, J.; Wright, S. *Numerical Optimization*; Springer: Berlin/Heidelberg, Germany, 2006.