ROG.e 2024
ISSN 2525-7579
biblioteca.ibp.org.br

Technical Paper

Promoted and Organized by:

Data-aware PVT solver: a machine learning approach for multiphase CO2 rich mixtures modelling

Reza Arefidamghani , Gustavo Eduardo Oviedo Celis , Jorge Trujillo , Raiza Gomes de Souza , Heloisa Althoff , Daniel Onofre de Almeida Cruz, Hamidreza Anbarlooei.

Abstract:
Carbon Capture, Utilization and Storage (CCUS) holds significant promise the transition to zero-emission energy systems. To effectively transfer carbon dioxide (CO2) from its sources to geological sinks, it is imperative to gain understanding of the thermodynamics and hydrodynamics, particularly in multi-phase flow mixtures, of CO2 with impurities. Equations of state and flash calculation play a vital role in this domain. Binary interaction parameters play a critical role in flash calculations as they quantify the molecular interactions within mixtures. However, accurately estimating these parameters poses a significant challenge due to the complex nature of molecular interactions. Fine-tuning, adjusting, and optimization techniques are commonly employed to determine these parameters. However, traditional approaches often require intuitive comprehension and are limited by the availability of data. This study aims to employ Machine Learning (ML) techniques, specifically Multi-Fidelity simulation (MF) and Gaussian process-based calibration (GPC), to tackle this problem. These techniques have been integrated into an in-house PVT solver. The implementation leverages GPU acceleration, enabling the solver to process faster flash calculations for different practical mixtures. ML-based methods offer a distinct advantage over traditional fitting approaches by autonomously pinpointing the most sensitive parameters, optimizing them, and quantifying the prediction uncertainties. It renders them a data-aware solution, capable of seamlessly adapting to new experimental observations.
Keywords: flash calculation; machine learning; multi-fidelity simulation; Gaussian process calibration; CCUS
Palavras-chave:

Received: September 8th, 2024 | **Accepted:** July 7th, 2024 | **Available online:** September 23th, 2024
Article nº: 2978
DOI: https://doi.org/10.48072/2525-7579.roge.2024.2978

1. universidade Federal do Rio de janeiro. Institute of Mathematics - UFRJ. BRASIL. E-mail: reza.arefi@matematica.ufrj.br. (https://orcid.org/0000-0001-9885-2532). 2. Universidade Federal do Rio de Janeiro. Engenharia mecânica. BRASIL. E-mail: g.oviedo@mecanica.coppe.ufrj.br. (https://orcid.org/0000-0002-4042-9093). 3. GALP. Upstream. PORTUGAL. E-mail: jorge.trujillo@galp.com. (https://orcid.org/0000-0003-4692-5386). 4. GALP. Upstream. PORTUGAL. E-mail: raiza.souza@galp.com. (https://orcid.org/0000-0001-7019-6377). 5. petrogral Brasil. Inovação. BRASIL. E-mail: heloisa.althoff@petrogalbrasil.com. (https://orcid.org/0009-0005-0534-1407). 6. Universidade Federal do Rio de Janeiro. BRASIL. E-mail: doac@mecanica.coppe.ufrj.br. 7. universidade Federal do Rio de janeiro. matematica aplicada. BRASIL. E-mail: hamid@matematica.ufrj.br. (https://orcid.org/0000-0002-3220-7255).

1. Introduction

Carbon Capture, Utilization and Storage (CCUS) promises to play a big role in the transition to zero emission energy systems, involving the transport from the source to the sink (geological formations such as saline aquifers or depleted oil and gas reservoirs). A more precise thermodynamic understanding of the CO_2 mixture and impurities is crucial for the deployment of those projects, and performance of the multi-phase flow modelling of such fluids. In this context, a PVT table obtained from flash-calculations is frequently required for flow assurance simulations. Several equations of state and numerical methods have been developed to perform Vapor-Liquid Equilibrium (VLE) and Vapor-Liquid-Liquid Equilibrium (VLLE) calculations. Cubic Equations of State (EoS) accompanied by van-der Waals quadratic mixing rules form the foundation of many commercial and research softwares. Binary interaction parameters are crucial in this regard, since they quantify the nature of molecular interaction that occurs between various chemical species in a mixture. Since these parameters typically aggregate the effects of several different phenomena into a single number, a formal theory has not been put out to estimate them precisely (if possible). While various methods exist for computing these parameters [Coutinho, Kontogeorgis and Stenby (1994), Jaubert and Mutelet (2004), Kordas, Magoulas, Stamataki and Tassios (1995)], yet evidence suggests that these calculations may be inaccurate and necessitate adjustments. Fine-tuning, adjusting, and optimization techniques are commonly employed to determine the values of these parameters [Ghafri, Maitland and Trusler (2014), Celis, Arefidamghani, Anbarlooei and Cruz (2023)]. Nevertheless, conventional approaches often demand intuitive understanding and are limited by the availability of data, with optimization techniques becoming expensive when dealing with numerous parameters. For instance, in a multi-component mixture comprising 20 species, there are roughly 200 binary interaction parameters, highlighting their susceptibility to requiring adjustments ($n^2/2$ parameters, where n is the number of species). Unfortunately, there is typically insufficient experimental data available to precisely and rigorously tune these parameters. While sensitivity analysis techniques can be utilized to identify candidate parameters for optimization, their implementation typically requires significant computational calculations, which may not always be feasible. In practice, physical intuition or previous experience are required to identify one or few more sensitive parameters to tweak. The goal of the present work is to use Machine Learning (ML) techniques to fine tune binary interaction parameters. For this purpose, two candidate ML techniques have been explored:

1. Multi-Fidelity simulation (MF) Paleyes, Mahsereci and Lawrence (2023), which leverages multiple levels of data fidelity to predict experimental outcomes, rather than directly tuning parameters. This approach effectively captures uncertainty within the system.
2. Gaussian Process-based Calibration (GPC) Higdon, Kennedy, Cavendish, Cafeo and Ryne (2004), Kennedy and O´Hagan (2001) where not only calibrates the parameters to match experimental observations (same as traditional optimization techniques) but also captures the posterior distribution around these parameters, along with the assessing the confidence or uncertainty surrounding the experimental data. Additionally, this method employs sensitivity analysis at a considerably lower cost than traditional approaches and provides insights into the most crucial parameters for optimization.

Unlike traditional fitting approaches, these ML-based methods excel at seamlessly and automatically fine-tuning binary interaction parameters and eliminate the need for a priori physical insights into parameter tuning. Furthermore, these methods enable the quantification of uncertainties in predictions, offering insights into regions of the domain where experimental observations are lacking. This information can signal the necessity of conducting specific experimental campaigns. These uncertainties can also propagate downstream into flow assurance simulations, offering a means to evaluate uncertainties across various transportation scenarios. An essential characteristic of these methods is their ability to continuously adjust and fine-tune parameters in response to new experimental observations. This permits to transform traditional PVT solvers into a data-aware solution.

In the current study, both MF and GPC methods have been implemented into an in-house PVT solver. This solver has the ability to calculates VLE and VLLE configurations, using both PR Jaubert and Mutelet (2004) and SRK Soave (1972) cubic EoSs. Flash calculations can be considered time-consuming, especially when utilized in conjunction with optimization processes. Consequently, there have been endeavors to accelerate these calculations in the past [Michelsen (1986), Hendriks and Bergen (1992), Rasmussen, Krejbjerg, Michelsen, Bjurstøm (2006), Gaganis and Varotsis (2014)]. However, in this study, the developed PVT solver has been implemented on a GPU framework to accelerate the calculations and bypass the complexities associated with the previously mentioned methods. This solver is capable of performing over one million flash calculations in fractions of a second for various practical mixtures.

2. Governing equations and problem setting

In the context of analyzing multi-component mixtures, the role of flash-calculations, provide crucial insights into the phase equilibria and composition distribution (for more information, see Montel (1993). To conduct such calculations, a thermodynamic model (EoS) must be specified. The present study focuses on two commonly used equation of states: the Peng-Robinson (PR) Jaubert and Mutelet (2004) and the Soave-Redlich-Kwong (SRK) Soave (1972). These equations of states can be expressed as follows:

$$P = \frac{RT}{v - b_i} - \frac{a_i(T)}{(v+\delta_1 b_i)(v+\delta_2 b_i)}, \quad (1)$$

where *P*, *T* and *v* stand for the thermodynamics properties of pressure, temperature, and molar volume respectively, while *R* is the ideal gas constant. By employing two parameters, δ_1 and δ_2, one can make a seamless transition between SRK and PR equations ($\delta_1 = 1$ and $\delta_2 = 0$ for SRK and $\delta_1 = 1 + \sqrt{2}$ and $\delta_2 = 1 - \sqrt{2}$ for PR). The coefficients $a_i(T)$ and $b_i(T)$ are temperature dependent parameters linked to the energy and co-volume of the pure component, see Jaubert and Mutelet (2004) and Soave (1972). In order to extend the use of Equation 1 for multi-component mixtures, additional equations, known as mixing rules, for *a* and *b* need to be introduced. Equation 2 shown the van-der Waals mixing rules approach used in the present study

$$a = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_i \sqrt{a_i a_j}\left(1 - k_{ij}(T)\right), \qquad b = \sum_{i=1}^{N} x_i b_i, \quad (2)$$

where *x*, represent the molar fraction of every component in a mixture, *N* is the number of components in the mixture, and $k_{ij}(T)$ stand for the binary interaction parameter (BIP) between the components *i* and *j*. In the present work, the semi-analytical model developed by Jaubert and Mutelet (2004), based on a group contribution method, has been implemented for BIPs. This model predicts BIPs as:

$$k_{ij}(T) = \frac{-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_{ik}-\alpha_{jk})(\alpha_{il}-\alpha_{jl})\, A_{kl}\left(\frac{298.15}{T}\right)^{\left(\frac{B_{kl}}{A_{kl}}-1\right)} - \left(\frac{\sqrt{a_i(T)}}{b_i} - \frac{\sqrt{a_j(T)}}{b_j}\right)^2}{2\frac{\sqrt{a_i(T)a_j(T)}}{b_i b_j}} \quad (3)$$

where a_i and b_i depending on the choice of the EoS (PR, or SRK), and α_{ij}, A_{kl} and B_{kl} are constant parameters (for more information regarding this model, please refer to [Jaubert and Mutelet (2004) and Jaubert, Privat and Mutelet (2010)]).

Although this model provides initial values for BIPs, more accurate and finely-tuned values are still necessary for practical applications. In some cases, such as mixtures with a high

percentage of CO_2, the values provided by this model or similar models may not be accurate enough. As a result, the predictions from the flash calculation may deviate from experimental observations significantly. As mentioned before, the typical solution to this problem involves fitting BIPs using experimental observations. However, it is important to note that the number of parameters to calibrate grows rapidly with the square of the number of components in the mixture, as BIPs scale accordingly. On the other hand, the number of experimental observations is typically limited, and it may not be feasible to use them to fine-tune all parameters. Additionally, BIPs are temperature-dependent, further complicating the situation. In practice, physical intuition or previous experience are required to identify one or few more sensitive parameters to tweak.

In the present research, the behavior of a hydrocarbon mixture comprising more than 20 components in the presence of a high percentage of CO_2 has been investigated (see Ghafri et al. (2014) for more detail). Specifically, the vapor-liquid phase envelope of the mixture at certain fixed temperatures for various concentrations of CO_2 has been obtained and compared with a few experimental observations. Comparing the results (presenting on the next section) of flash calculations (using Equation 3 without tuning BIPs) with available experimental data reveals discrepancies, especially when CO_2 is present in the mixture. In this case, there are 210 parameters to be tuned, while the available observations consist of only around 20 points. It is evident that fine-tuning all parameters here is impractical. Based on some physical intuition, it was previously observed that adjusting certain BIPs (such as the BIP corresponding to CO_2 and CH_4) significantly improves the results Ghafri et al. (2014). Our ultimate goal is to automate the selection of parameters for tuning or eliminate the need for it entirely using machine learning techniques. The two mentioned techniques (multi-fidelity simulation and calibration using a Gaussian process) have been selected based on this objective.

In order to define the criteria for comparing numerical and experimental data the Mean Squared Error (MSE) and the Integral Error of the area between the curves (Int-Err) emerge as two primary metrics for this purpose.

3. Multi-fidelity

In many cases in science and engineering, several models are available to define a system of interest. These diverse models have varying evaluation costs and fidelity levels. In general, a high-fidelity model despite requiring significant computational resources, are capable of accurately representing the complexities of the system, thus meeting the precision requirements

specified by the application. Lower-fidelity models, on the other hand, provide lower accuracy while being computationally less costly than their high-fidelity counterparts. In addition, there are experimental data classified as the highest-fidelity data. Multi-fidelity modeling is a powerful technique that combines information from different fidelity levels to enhance the accuracy of predictions. The fundamental concept behind the multi-fidelity method involves using low-fidelity models to accelerate processes, while also integrating the high-fidelity model or data to maintain accuracy. This approach enhances the efficiency and effectiveness of predictive modeling. In our specific context (flash calculation), two levels of data fidelity can be identified:

- Low-fidelity model predictions, obtained by flash calculation using PR or SRK equations. Figure 1 shows the experimental observations and also the results of flash calculation without any fine-tuning for both models. The discrepancies between the models and the experimental data are more evident for higher concentrations of CO_2. Specifically, PR exhibits a substantially higher MSE of 6.6830 and an Int-Err of 1.8642, whereas SRK demonstrates notably lower values with an MSE of 2.0988 and an Int-Err of 0.9929. One may train a surrogate model based on these predictions. The results of such a surrogate model will fall into this category again.
- High-fidelity experimental data (Ground Truth), which refers to limited experimental observations about VLE phase envelope.

Comparison of experimental observations with phase envelopes derived from PR and SRK equations suggests the presence of a potentially linear relationship between low and high-fidelity functions. Therefore, in the following we just discuss the linear multi-fidelity model Kennedy and O´Hagan (2001). In this model, the high-fidelity function is expressed as a combination of the low-fidelity function, scaled appropriately, and an associated error term as:

$$f_{high}(x) = \rho\, f_{low}(x) + f_{err}(x) \quad (4)$$

In this equation, we assume that f_{low} represents a Gaussian process modeling the low-fidelity function, while ρ serves as a scaling factor indicating the magnitude of correlation with the high-fidelity data. f_{err} indicates another Gaussian process, independent of f_{low}, which models the bias term. The multi-fidelity model then reads as

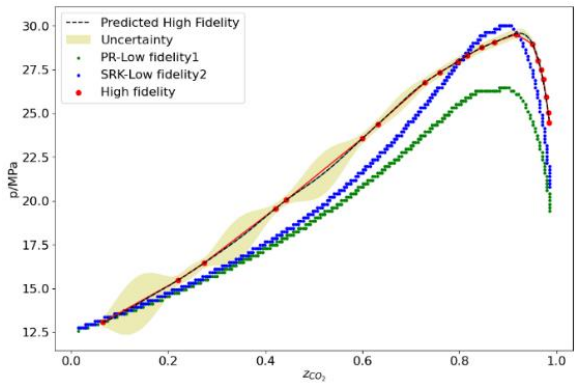
$$\begin{bmatrix} f_{low}(h) \\ f_{high}(h) \end{bmatrix} \sim GP\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k_{low} & \rho\, k_{low} \\ \rho\, k_{low} & \rho\, k_{low}^2 + k_{err} \end{bmatrix}\right) \quad (5)$$

where k is the covariance (assuming zero mean for the Gaussian process).

Given the smoothness nature of the phase envelope curves (see Figure 1), a radial basis function (RBF) kernel for both the bias and correlation components of the model. We also assume that our numerical predictions, using both PR and SRK have same fidelity. In absence of enough information about experimental observations, same has been assumed for the noise of the high-fidelity observations. This model has been implemented using **Emukit** (A Python-based toolbox of various methods in decision making, uncertainty quantification and statistical emulation) Paleyes et al (2019), Paleyes et al. (2023).

The linear multi-fidelity model provides a systematic framework to integrate and leverage information from these different sources, ultimately aiming to improve the accuracy of predictions. Figure 1 depicts the outcomes of applying multi-fidelity techniques to this dataset. In this figure, the red points represent the high-fidelity data, while the green and blue points correspond to the low-fidelity predictions generated by the PR and SRK models, respectively. The red solid line indicates the interpolation of experimental data, while the black dashed line signifies the posterior mean, generated by the linear multi-fidelity model. As evident, the posterior closely approximates the ground truth (high-fidelity experimental observations). The yellow shadow represents the uncertainty predicted by the multi-fidelity model. As observed, regions with insufficient experimental data exhibit higher uncertainty in the predictions. This insight can guide the planning of additional experiments in these areas. Alternatively, this uncertainty in thermodynamic modeling can be incorporated into downstream flow assurance solvers to account for the variability in flow behavior.

Figure 1 - Linear MF model fit to low and high-fidelity data for VLE hydrocarbons mixture at T=373.15 K.



Source: Made by the authors.

4. Calibration strategy

As demonstrated in the previous section, the multi-fidelity method operates akin to a black box, enhancing prediction accuracy while also providing insights into uncertainties; nevertheless, its inability to furnish precise values for the BIPs restricts its applicability across diverse problem domains. In this section, we investigate a Gaussian process-based calibration method to address the aforementioned limitation. Specifically, we seek a technique capable of calibrating the BIPs while also accommodating inaccuracies in the physics underlying our flash calculations. Our primary objective is to determine uncertainties in predictions akin to those provided by the MF method. For this purpose, we utilize the method outlined in Nguye, Francom, Luscher, and Wilkerson (2021), which is rooted in the framework devised by Kennedy and O'Hagan (2001). Below, we provide a brief overview of the method.

The Kennedy and O'Hagan (2001) model utilizes a Gaussian process (GP) model to represent the simulator output at unseen inputs, extending it into a comprehensive framework that facilitates parameter calibration, output prediction, and uncertainty quantification. This method is suitable for univariate outputs from the simulator (PVT solver), while our response is highly multivariate. Just to reiterate, our simulator output represents phase envelope curves, depicted as lists of more than 100 points in the $P - z_{CO_2}$ plane. The model developed in Nguye et al. (2021) extends the previous method to the multivariate scenario. Additionally, recognizing the time-consuming nature of function calls, it incorporates training of a surrogate model based on a subset of numerical simulations to expedite calculations.

Let us consider a simulator $\eta(x, \theta)$, where x represents the controllable experimental conditions (in our case, temperature, and composition of the mixture), and θ represents the parameters to be calibrated, i.e. binary interaction parameters (BIPs). For simplicity, let us assume we are calibrating three parameters $k_1 = k_{CO_2,CH_4}, k_2 = k_{CO_2,Squ}$ and $k_3 = k_{CH_4,Squ}$. The flash calculations will be performed m (in our case $m < 100$) times for different settings to obtain training data for the surrogate model. We using a p_η-dimensional basis representation to model the simulator output as

$$\eta(x, \theta) = \sum_{i=1}^{p_\eta} k_i w_i(x, \theta) + \varepsilon \quad (6)$$

where k_i's are orthogonal basis vectors, $w_i(x, \theta)$'s are GPs over the input space and ε is the error term. The basis vectors can be obtained using Principle Component Analysis (PCA) of the output matrix (see Nguye et al. (2021) for more detail). This forms our surrogate model.

The experimental observations y can be modeled as

$$y(x) = \eta(x,\theta) + \delta(x) + \epsilon \quad (7)$$

where $\eta(x,\theta)$ is the output obtained by using the true parameter θ, $\delta(x)$ models the discrepancy between the simulator and physical reality, and ϵ represents the observational error, Nguye et al. (2021). Now, it is possible to use experimental observations to calibrate the parameter θ. An optimization algorithm is utilized to iteratively adjust model parameters to minimize the disparity between model predictions and observed outcomes. For a comprehensive explanation of the method, please refer to Nguye et al. (2021).

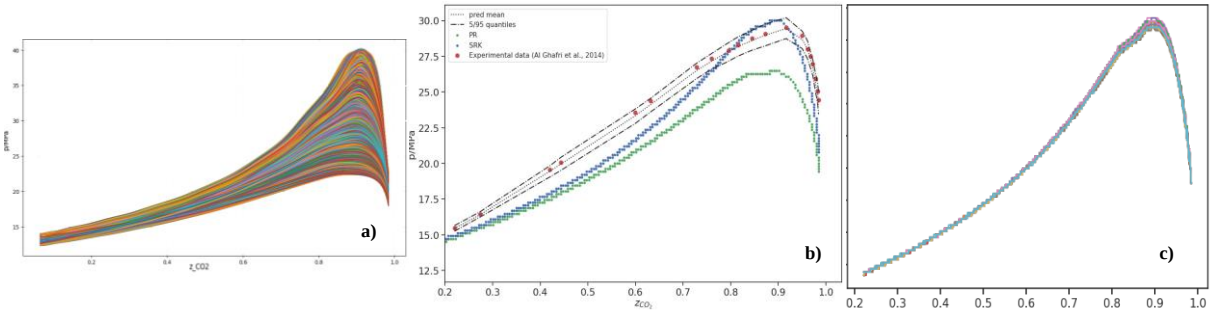
Importantly, this process also entails the estimation of parameter distributions, enabling a probabilistic characterization of model uncertainty. By obtaining parameter distributions, we gain insights into the variability and uncertainty associated with model parameters, thereby enhancing our ability to make informed decisions based on predictive outcomes. Parameter inference within the model is executed via Markov Chain Monte Carlo (MCMC) sampling to approximate the posterior distribution of model parameters.

The efficacy of this GPC model in parameter calibration and uncertainty analysis is demonstrated in Figure 2. In Figure 2-a, the data used to train the surrogate model (simulations) are shown. This data is obtained by placing a normal distribution around the values obtained from group theory for each selected BIP, with a standard deviation of 0.1. Figure 2-b shows the results of calibration process and related uncertainties. This figure clearly demonstrates the effectiveness of the GPC model in adjusting the parameters to closely match the observed in data. Finally, Figure 2-c displays the results of simulations (using flash calculations) from 100 different samples drawn from the posterior distribution of the parameters. As clearly visible, this figure completely corroborates the predictions in Figure 2-b and the results are predominantly fell within the 90 percent confidence interval. Please note that the results depicted in the middle figure (2-b) are obtained using the surrogate model and not the flash calculations.

Our results indicate that the mean of the posterior distribution for each parameter obtained by GPC, closely aligns with the values obtained from expensive optimization methods like grid search and AdaGrad (Adaptive Gradient Method); see Celis et al. (2023). Furthermore, the small standard deviation indicates a concentrated distribution around the mean, underscoring the effectiveness of our calibration approach (see Figure 3). Additionally, the relatively narrow uncertainty range (5/95) further highlights the superior and effective performance of the GPC model. The narrow uncertainty range signifies a high level of confidence in the calibrated

parameters (see Figure 3). These outcomes underscore the robustness and precision of our calibration technique, showcasing its ability to adapt to varying levels of complexity within the data and generate reliable parameter estimates.

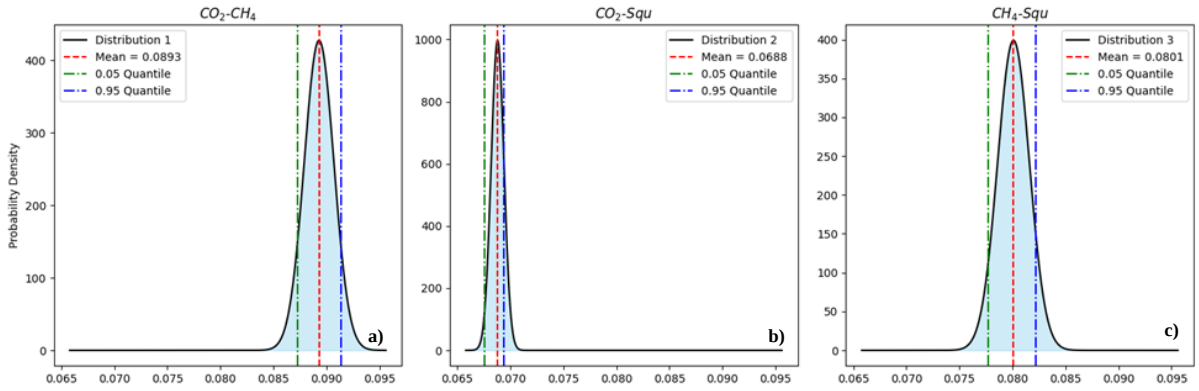
Figure 2 - Results of GPC applied to the flash calculations. **a)** Initial data used to train the surrogate model, **b)** The mean and uncertainty in predictions after calibration, **c)** 100 phase envelopes were generated using flash calculations with parameters drawn from the posterior distribution obtained in the calibration process.



Source: Made by the authors.

This successful validation not only confirms our ability to efficiently calibrate binary interaction parameters at a significantly lower cost compared to traditional methods but also highlights the utility of our surrogate model for future analyses, particularly in sensitivity analysis. Leveraging the GPC model for sensitivity analysis offers a substantially more cost-effective alternative to running expensive flash calculations conventionally required for such investigations. This promising avenue of future work holds immense potential for advancing parameter fine-tuning and optimization processes across various scientific and engineering disciplines.

Figure 3 - Posterior distribution of three selected parameters after calibration.



Source: Made by the authors.

5. Concluding remarks and future works

In the current study, both multi-fidelity and Gaussian process-based calibration models have been implemented into an in-house PVT solver. This solver has the ability to calculates vapor-liquid and vapor-liquid-liquid equilibrium cases, using both PR and SRK cubic EOSs. The models were applied to a test case involving a hydrocarbon mixture with a high percentage of CO_2. The multi-fidelity model facilitates the integration of real-time data by leveraging various levels of fidelity data to predict experimental outcomes. By incorporating information from multiple sources, such as high-fidelity experimental observations and low-fidelity calculations, this model adaptively adjust the predictions to reflect the latest data trends.

The results of the current study demonstrate the significant potential of both methods in integrating experimental observations into numerical simulations to enhance the quality of predictions. In particular, if surrogate models have been pre-trained, the GPC method can be effectively employed to dynamically adjust their predictions based on new experimental or field observations. The remarkable adaptability and responsiveness to evolving data environments are evident in both MF and GPC methods, enhancing their data-awareness and robustness. Importantly, this adaptability facilitates the transition of conventional PVT solvers into data-aware solutions. Although the current study utilized three pre-selected parameters in the analysis, it is feasible to leverage the surrogate model for efficient sensitivity analysis and selection of parameters for fine-tuning. This area constitutes the focus of our ongoing research. Furthermore, in many practical applications, especially in the presence of CO_2, a second liquid phase may form in the equilibrium. Our next objective is to extend the methods implemented here to address such cases of VLLE.

6. Acknowledgments

The authors acknowledge the technical and financial support awarded by GALP/Petrogal Brazil and ANP–Agência Nacional de Petróleo, Gás Natural e Biocombustível.

Referências

Celis, G. E. O., Arefidamghani, R., Anbarlooei, H., & Cruz. D. O. A. (2023). An optimization approach to study the phase changing behavior of multi-component mixtures. ArXiv Preprint, 12. https://doi.org/10.48550/arXiv.2306.16327

Coutinho, J. A. P., Kontogeorgis, G. M., & Stenby, E. H. (1994). Binary interaction parameters for nonpolar systems with cubic equations of state: a theoretical approach 1. co2/hydrocarbons using SRK equation of state. Fluid Phase Equilibria, 102, 31–60. https://doi.org/10.1016/0378-3812(94)87090-X

Gaganis, V., & Varotsis, N. (2014). An integrated approach for rapid phase behavior calculations in compositional modeling. Journal of Petroleum Science and Engineering, 118, 74–87. https://doi.org/10.1016/j.petrol.2014.03.011

Ghafri S. Z. A., Maitland, G. C., & Trusler, J. P. M. (2014). Experimental and modeling study of the phase behavior of synthetic crude oil + CO2. Fluid Phase Equilibria, 365, 20–40. https://doi.org/10.1016/j.fluid.2013.12.018

Hendriks, E. M., & Bergen, A. R. D. V. (1992). Application of a reduction method to phase equilibria calculations. Fluid Phase Equilibria, 74, 17–34. https://doi.org/10.1016/0378-3812(92)85050-I

Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., & Ryne, R. D. (2004). Combining field data and computer simulations for calibration and prediction. SIAM Journal on Scientific Computing, 26(2), 448–466. https://doi.org/10.1137/S1064827503426693

Jaubert, J. N., & Mutelet, F. (2004). VLE predictions with the Peng–Robinson equation of state and temperature dependent kij calculated through a group contribution method. Fluid Phase Equilibria, 224(2), 285–304. https://doi.org/10.1016/j.fluid.2004.06.059

Jaubert, J. N., Privat, R., & Mutelet, F. (2010). Predicting the phase equilibria of synthetic petroleum fluids with the PPR78 approach. AIChE Journal, 56(12), 3225–3235. https://doi.org/10.1002/aic.12232

Kennedy, M. C., & O'Hagan, A. (2002). Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(3), 425–464. https://doi.org/10.1111/1467-9868.00294

Kordas, A., Magoulas, K., Stamataki, S., & Tassios, D. (1995). Methane hydrocarbon interaction parameters correlation for the Peng-Robinson and the t-mpr equation of state. Fluid Phase Equilibria, 112(2), 33–44. https://doi.org/10.1016/0378-3812(95)02787-F

Michelsen, M. L. (1986). Simplified flash calculations for cubic equations of state. Industrial & Engineering Chemistry Process Design and Development, 25(1), 184–188. https://doi.org/10.1021/i200032a029

Montel F. (1993). Phase equilibria needs for petroleum exploration and production industry. Fluid Phase Equilibria, 84, 343–367. https://doi.org/10.1016/0378-3812(93)85132-6

Nguyen, T., Francom, D. C., Luscher, D. J., & Wilkerson, J. W. (2021). Bayesian calibration of a physics-based crystal plasticity and damage model. Journal of the Mechanics and Physics of Solids, 149, 104284. https://doi.org/10.1016/j.jmps.2020.104284

Paleyes, A., Mahsereci, M., & Lawrence, N. D. (2023). Emukit: A python toolkit for decision making under uncertainty. In Proceedings of the Python in Science Conference, 8. https://doi.org/10.25080/gerudo-f2bc6f59-009

Paleyes, A., Pullin, M., Mahsereci, M., McCollum, C., Lawrence, N., & González, J. (2019). Emulation of physical processes with Emukit. In Second Workshop on Machine Learning and the Physical Sciences, NeurIPS, 8. https://doi.org/arxiv.org/abs/2110.13293v1

Rasmussen, C. P., Krejbjerg, K., Michelsen, M. L., & Bjurstrøm, K. E. (2006). Increasing the computational speed of flash calculations with applications for compositional, transient simulations. SPE Reservoir Evaluation & Engineering, 9(1), 32–38. https://doi.org/10.2118/84181-PA

Soave, G. (1972). Equilibrium constants from a modified redlich-kwong equation of state. Chemical Engineering Science, 27(6), 1197–1203. https://doi.org/10.1016/0009-2509(72)80096-4