Bayesian Analysis to Determine Relative Significance of Inputs of a Rock-Physics Model

Kyle T. Spikes[1]* and Mrinal K. Sen[1,2]

[1]Department of Geological Sciences, Jackson School of Geosciences, The University of Texas at Austin, Austin, TX, United States, [2]Insitute for Geophysics, Jackson School of Geosciences, The University of Texas at Austin, Austin, TX, United States

Rock-physics models relate rock properties to elastic properties through non-unique relationships and often in the presence of seismic data that contain significant noise. A set of inputs define the rock-physics model, and any errors in that model map directly into uncertainty in target seismic-scale amplitudes, velocities, or inverted impedances. An important aspect of using rock-physics models in this manner is to determine and understand the significance of the inputs into a rock-physics model under consideration. Such analysis enables the design of prior distributions that are informative within a reservoir-characterization formulation. We use the framework of Bayesian analysis to find internal dependencies and correlations among the inputs. This process requires the assignments of prior distributions, and calculation of the likelihood function, whose product is the posterior distribution. The data are well-log data that come from a hydrocarbon-bearing set of sands from the Gulf of Mexico. The rock-physics model selected is the soft-sand model, which is applicable to the data from the reservoir sands. Results from the Bayesian algorithm are multivariate histograms that demonstrate the most frequent values of the inputs given the data. Four analyses are applied to different subsets of the reservoir sands, and each reveals some correlations among certain model inputs. This quantitative approach points out the significance of a singular or joint set of rock-physics model parameters.

Keywords: rock physics, bayesian analyses, model sensitivity, elastic properties, fluid effect

INTRODUCTION

The application of a rock-physics model to a relevant data set has a number of uses. An appropriate model based on geologic context can provide an aid in the interpretation of depositional and diagenetic history of a formation or sequence of formations of interest. Another use is to understand what seismic velocities to expect from in scenarios both represented and not represented in a relevant data set. In either case, inputs into rock-physics models can be treated as qualitative nearly arbitrary values that satisfy a fit to data. That fit could be deemed successful from a visual standpoint, or a quantitative comparison with the data could be used to determine a successful fit. This process, however, potentially could exclude any unknown correlations among inputs, where some physical connection could be used to determine more confidently the values of the inputs. Our work here identifies the most significant sets of model inputs as well as significant correlations among them. We do this by quantitatively fitting an established rock-physics model to appropriate well-log data using a Bayesian analysis process. In the

FIGURE 1 | Log data from the Gulf of Mexico data set. Included are **(A)** gamma ray, **(B)** water saturation, **(C)** porosity, **(D)** density, **(E)** P-wave velocity, and **(F)** P-impedance. In gray are the M10-M60 sands, also labeled in **(A)**. The sands have lower impedances than the shales and tend to separate cleanly in that domain.
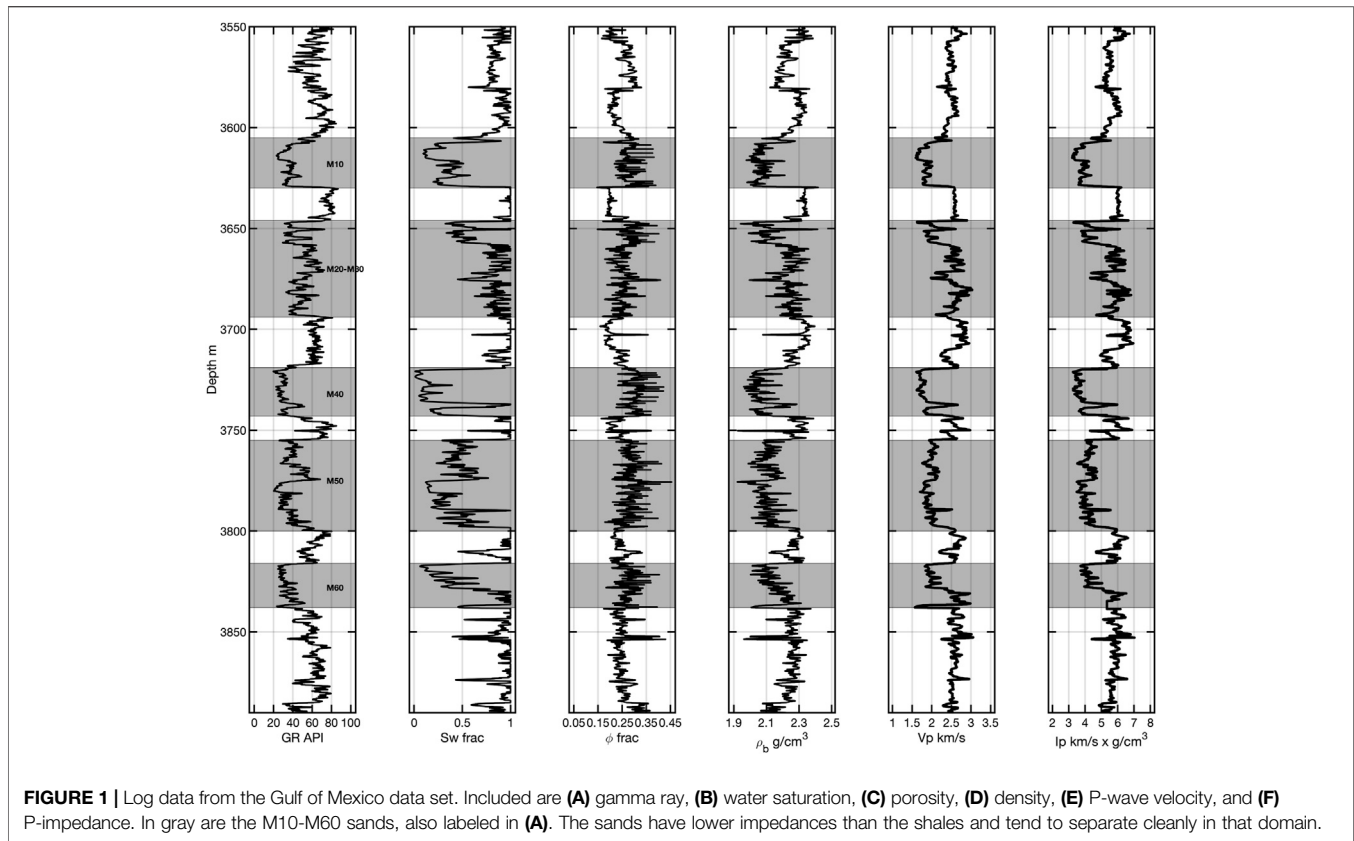
context of seismic reservoir characterization, this type of statistical analysis is important because it provides information to define realistic prior distributions in the rock-physics part of a reservoir-characterization workflow.

Bayesian approaches are commonly used to infer rock or elastic properties from geophysical observations such as Bosch et al. (2010), Ulvmoen and Omre (2010), Ulvmoen et al. (2010), Nawaz and Curtis (2017), Grana (2018). The goal of these works is to obtain subsurface distributions of rock properties of interest. The work here determines the sensitivity of model parameters for a model fit to a data set. The work in this paper uses one model and well-log data from the Gulf of Mexico. Our work is unique and important because it directly interrogates the rock-physics model to determine the significance and any internal dependencies and/or correlations of the inputs. Many other models and relevant data could be treated in this way, so the results are not necessarily useful for application to other models. However, the results for an individual model are applicable to other data sets where the inputs can be considered from most to least significant.

MATERIALS AND METHODS

Data and Rock-Physics Model

We use a data set from a clastic reservoir in the Gulf of Mexico with water depth about 1300 m. Selected well logs (**Figure 1**) from one well included gamma ray (GR), water

saturation (S_w), porosity (ϕ), density (ρ_b), P-wave velocity (V_P), and the calculated P-impedance (I_P). The stratigraphic sequence consists of alternating sands and shales with P-impedances being higher in shales than in the unconsolidated sands. The hydrocarbon-bearing sands are named M10–M60, labeled in **Figure 1A**. These sands are Tertiary age, and they resulted from deep water sand deposition. Hydrocarbon traps resulted from salt movement and growth faults (Contreras, 2006).

The model selected for this analysis work is the soft-sand model (Dvorkin and Nur, 1996). It has been used with this data set in other publications (e.g., Xie and Spikes, 2021). The model is a combination of Hertz-Mindlin (1949) contact theory and modified Hashin-Strikman forms. The Hertz-Mindlin (1949) theory expresses the effective bulk (K_{HM}) and shear moduli (μ_{HM}) for an identical set of spherical grains in the dry condition for a specific hydrostatic pressure (P),

$$K_{HM} = \left[\frac{C_n^2 \left(1 - \phi_c\right)^2 \mu^2}{18\pi^2 \left(1 - \nu\right)^2} P \right]^{\frac{1}{3}} \tag{1}$$

and

$$\mu_{HM} = \frac{5 - 4\nu}{5\left(2 - \nu\right)} \left[\frac{3C_n^2 \left(1 - \phi_c\right)^2 \mu^2}{2\pi^2 \left(1 - \nu\right)^2} P \right]^{\frac{1}{3}}. \tag{2}$$

The terms in **Eqs. 1,2** are the Lamé parameters (μ and ν) of the effective homogeneous mineral, the coordination number (C_n),
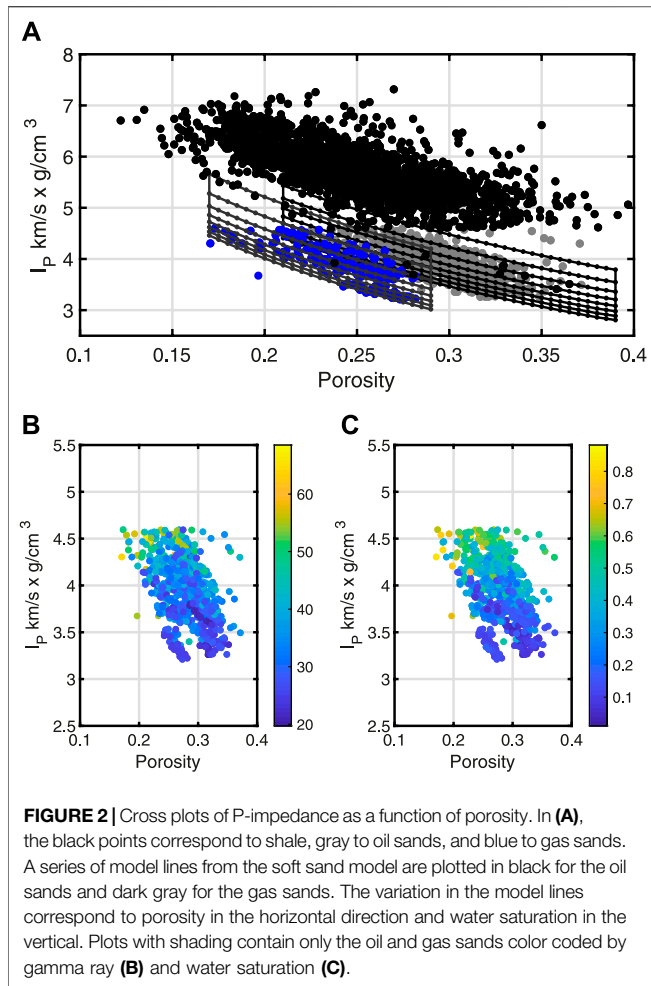
FIGURE 2 | Cross plots of P-impedance as a function of porosity. In **(A)**, the black points correspond to shale, gray to oil sands, and blue to gas sands. A series of model lines from the soft sand model are plotted in black for the oil sands and dark gray for the gas sands. The variation in the model lines correspond to porosity in the horizontal direction and water saturation in the vertical. Plots with shading contain only the oil and gas sands color coded by gamma ray **(B)** and water saturation **(C)**.

TABLE 1 | Rock physics model input values.

	ϕ	S_w	P (MPa)	ϕ_c	C_n	SSR
Gas sand	0.17–0.29	0.2–1.00	20	0.36	5	3
Oil sand	0.21–0.39	0.2–1.00	20	0.38	5	4

TABLE 2 | Elastic moduli and densities used in the rock-physics models [modified from Mavko et al. (2009)].

	Gas	**Oil**	**Brine**	**Clay**	**Quartz**
K (GPa)	0.37	0.8	3	16.8	36
μ (GPa)	–	–	–	5.60	45
ρ (g/cc)	0.15	0.60	1.05	2.58	2.65

Figure 2B is *GR*, and in **Figure 2C** it is S_w, both of which show variation in this domain.

The models plotted in **Figure 2A** are repeated in **Figure 3A** without the shale data for clarity. Plots in **Figures 2B–D** are the same data, but the models have been perturbed relative to those in **Figure 3A**. In **Figure 3B**, C changed from 20 to 30%. In **Figure 3C**, C was changed back to the original, P was reduced from 20 to 15 MPa. Last, the pressure change was removed, and ϕ_c was changed from 0.36 to 0.35. For simplicity these perturbations were the same for both the gas and oil-sand models. In all four cases, the models qualitatively fit the data, but these fits do not give an indication of which inputs are more significant to change. The Bayesian approach provides a way to assess the significance of the model parameters.

Bayesian Analysis

The Bayesian approach includes a quantitative match of model to data. Within it we must compute a prior, $p(\mathbf{m})$, and likelihood function, $l(\mathbf{d}|\mathbf{m})$ (i.e., Ulrych et al., 2001; Tarantola, 2005; Sen 2006). The prior (**Eq. 3**),

$$p(\mathbf{m}) \propto exp\left[\left(\mathbf{m} - \mathbf{m}_{prior}\right)^T \mathbf{C}_M^{-1}\left(\mathbf{m} - \mathbf{m}_{prior}\right)\right], \qquad (3)$$

is proportional to the exponential that contains differences between the model $\mathbf{m}$ and $\mathbf{m}_{prior}$, which is a second and more informed model. Inputs to the rock physics model populate vectors of $\mathbf{m}$ and $\mathbf{m}_{prior}$. The differences are scaled by the inverse of the covariance matrix of the prior model, $\mathbf{C}_M$.

Next, we define the objective function [$E(\mathbf{m})$],

$$E(\mathbf{m}) = \left(\mathbf{d} - g(\mathbf{m})\right)^T \mathbf{C}_D^{-1}\left(\mathbf{d} - g(\mathbf{m})\right), \qquad (4)$$

where $g(\mathbf{m})$ is the rock-physics model that calculates simulated values based on the model values in $\mathbf{m}$, $\mathbf{d}$ is the real data vector, and the data covariance matrix is represented by $\mathbf{C}_D$. The likelihood function (**Eq. 5**) is then proportional to the negative exponent of the objective function

$$l(\mathbf{d}|\mathbf{m}) \propto \exp\left(-E(\mathbf{m})\right). \qquad (5)$$

Last the posterior distribution ($\sigma(\mathbf{m}|\mathbf{d})$) is proportional to the product of the prior and the likelihood (**Eq. 6**)

and the critical porosity (ϕ_c). The soft sand model represents a modified Hashin and Strikman (1963) lower bound for porosities from zero to ϕ_c. Additional equations are necessary to determine the elastic moduli at ϕ_c. Gassmann (1951) fluid substitution translates the moduli to the saturated condition. In this work, we deal with four fitting parameters: S_w, mineralogy, P, and ϕ_c. Variations in S_w control the effective fluid bulk modulus (computed using the Reuss average) and the fluid density. Mineralogy is limited to quartz and clay (C) content, where $C = (1$–the fraction of quartz). Preliminary work demonstrated that the contributions of C_n and shear stiffness reduction (SSR) were not significant when fitting this model to laboratory data, so they are held constant here.

Figure 2A contains cross plots of I_P as a function of ϕ for the shales (black points), oil sands (gray points), and gas sands (blue points). Overlain on the sand points are rock-physics model lines from the soft sand model. Those models for both the sand types vary as a function of ϕ and S_w while the other inputs are constant. **Tables 1,2** contains the values for the inputs and respective moduli and densities. **Figures 2B,C** are also plots of I_P as a function of ϕ but only for the two sands. The color code in
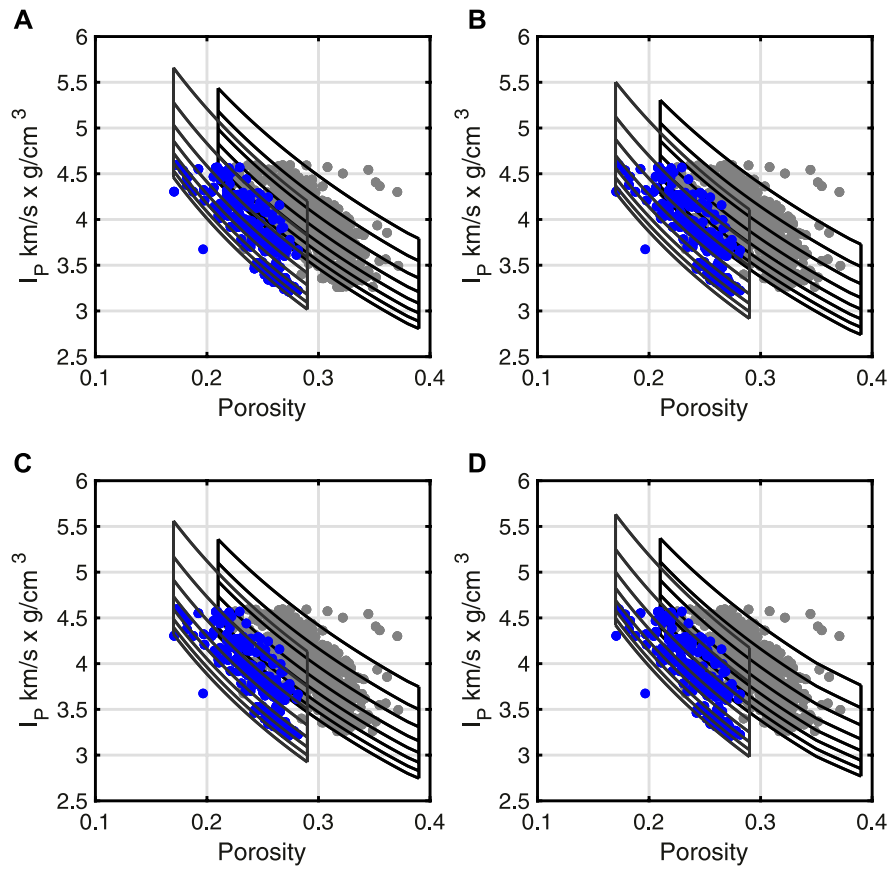
FIGURE 3 | The four plots contain perturbed versions of the rock physics models in **Figure 2**. The oil sands are in gray points and gas sands in blue. The models in **(A)** are the same as those in **Figure 2A**. The composition was altered slightly in **(B)** relative to **(A)**. In **(C, D)** the pressure and critical porosity, respectively, were changed relative to the models in **(A)**. All four sets of models match the data from a qualitative perspective.

$$\sigma\left(\mathbf{m}|\mathbf{d}\right) \propto l\left(\mathbf{d}|\mathbf{m}\right)p\left(\mathbf{m}\right). \qquad (6)$$

The number of model parameters is four, which is small enough to allow for a tractable analytical solution. No sampling methods are required. We analyze posterior multivariate histograms, so a normalizing factor to obtain probabilities is not needed.
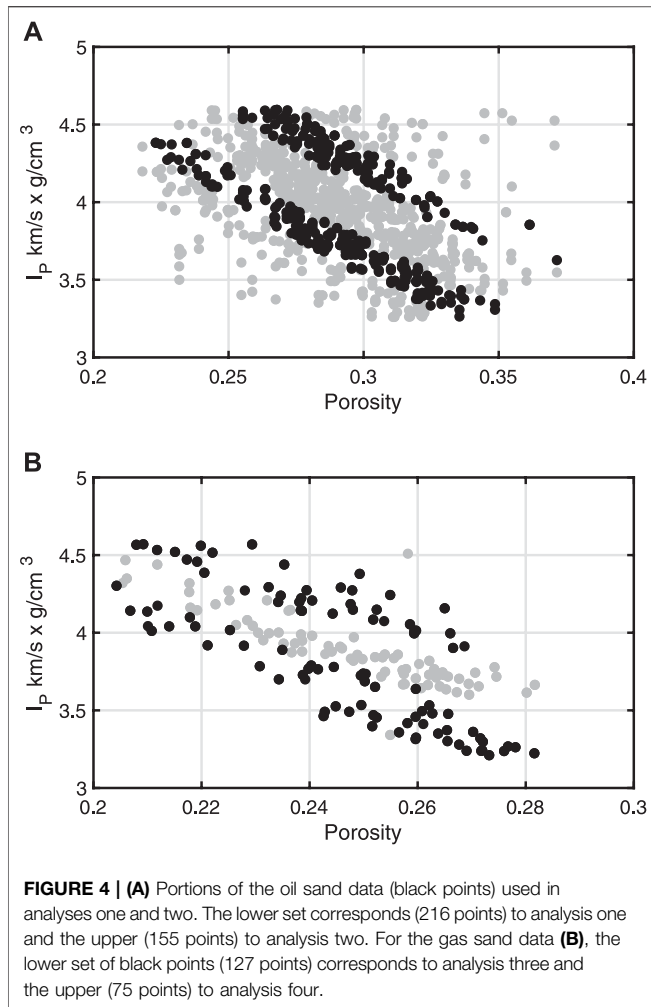
Data Selection

Data input into the analysis is a subset of the oil or gas sands. The selection of the subset is a two-step process. First, one of the rock-physics model lines shown in **Figure 2A** or **Figure 3A** is used as a reference. The reference is required because the computed models must be relatively close to the data used in the Bayesian framework. If the data and model are significantly far apart, then the objective function in **Eq. 4** has large values, and so the likelihood function (**Eq. 5**) will have exceedingly small values. Second, data are selected that occur within a certain percent of both I_P and ϕ from that reference model. Some scatter in the data is necessary as to not overfit the models to the data. **Figure 4A** shows two subsets (black points) on which we perform the analysis separately.

The reference models had S_w values of 0.4 for the lower set and 0.8 for the upper set. For both, the percent away from the reference was 2%, which resulted in 216 data points for the lower set and 155 for the upper set. Similarly, for the gas sands, two reference models were used with S_w values of 0.4 and 0.8 (**Figure 4B**). The percent away from the reference was 1%. The numbers of points were, respectively, 127 and 75. We conduct the analyses on these four subsets of data and call them analysis 1, 2, 3, and 4. Analyses 1 and 2 correspond to the lower and upper subsets, respectively, for the oil sands. Likewise, analyses 3 and 4 refer to the lower and upper subsets, respectively, for the gas sands.

RESULTS

The first set of results comes from analysis 1, for one subset of the oil sands. The model priors are user provided based on some knowledge of the data. The models **m** and $\mathbf{m}_{prior}$ were normal distributions of size 200 × 4, so 200 values of each of the four inputs. Mean values for $\mathbf{m}_{prior}$ were $[S_w, C, P, \phi_c]$ = [0.4, 0.2, 20, 0.38] and covariance matrix

FIGURE 4 | (A) Portions of the oil sand data (black points) used in analyses one and two. The lower set corresponds (216 points) to analysis one and the upper (155 points) to analysis two. For the gas sand data **(B)**, the lower set of black points (127 points) corresponds to analysis three and the upper (75 points) to analysis four.

$$
\begin{bmatrix}
0.0020 & 0 & 0 & 0 \\
0 & 0.0010 & 0 & 0 \\
0 & 0 & 0.1000 & 0 \\
0 & 0 & 0 & 0.019
\end{bmatrix}
$$

The mean of **m** was slightly different, and the covariance matrix for it had slightly larger values on the diagonal. All combinations of the priors were computed, so 200^4 models with no sampling. The data to model comparison (**Eq. 4**) used I_P. The posterior is four dimensional with size $200 \times 200 \times 200 \times 200 \times$, which is not possible to display nor visualize. Instead, we sum over two given dimensions to compute bivariate histograms, both priors and posterior, which are easy visualize. Each bivariate histogram is a joint histogram of the two variables. Thus, a combination of two values of two variables has a frequency for that pair of values together. Prior bivariate histograms for pairs of $C - S_w$, $P - S_w$, and $\phi_c - S_w$ are in **Figures 5A–C**. **Figures 5D–F** contain their accompanying bivariate posterior histograms. The priors have smooth appearances with relatively small frequencies. The posteriors show much different appearances among themselves. In **Figure 5D**, many pairs of C and S_w have relatively large frequencies, whereas in **Figures 5E, F**, a wide

range S_w values are present for narrow ranges of P and ϕ_c, respectively. Priors for pairs of P–C, $\phi_c - C$, and $\phi_c - P$ are in **Figures 5G–I**, respectively, with their corresponding posteriors in **Figures 5J–L**. The priors are relatively smooth and have wide ranges of possible pairs. Posteriors in **Figures 5J, K** display a wide range of C values for narrow ranges of P and ϕ_c. The plot in **Figure 5L**, however, shows isolated pairs of ϕ_c and P.

Figure 6 contains the bivariate priors and posterior histograms for analysis 2, the second oil sand analysis. The procedure for analysis 2 was the same as analysis 1. The only differences were the mean of $\mathbf{m}_{prior}$, which was $[S_w, C, P, \phi_c] = [0.6, 0.2, 20, 0.38]$, and the covariance matrix, which was

$$
\begin{bmatrix}
0.0030 & 0 & 0 & 0 \\
0 & 0.0010 & 0 & 0 \\
0 & 0 & 0.1000 & 0 \\
0 & 0 & 0 & 0.019
\end{bmatrix}
$$

Similar updates were made for **m**. The layout of **Figure 6** is the same as **Figure 5**. All six priors are relatively smooth. Four of the six posteriors (**Figures 6E,F,J,K**) have wide ranges of one parameter with a narrow range of the other. **Figure 6D** indicates many possible pairs, but the posterior in **Figure 6L**, like **Figure 5L**, shows isolated, joint values of ϕ_c and P.

Results from analyses 3 and 4 are plotted in **Figures 7**, **8**, respectively. The same analysis procedure was repeated again. The means and covariance matrices of the models in analysis 3 were the same as in analysis 1. The difference is that the effective fluid bulk modulus and density is a function of the gas-brine mixture rather the oil-brine mixture. In **Figure 7**, the posterior histograms resemble the counterparts in **Figures 5**, **6** although with different peak frequency values. Most notably, the histogram in **Figure 7F** again shows an isolated joint pair of ϕ_c and P. Last, in analysis 4, the mean for $\mathbf{m}_{prior}$ was $[S_w, C, P, \phi_c] = [0.72, 0.2, 20, 0.38]$ and covariance matrix

$$
\begin{bmatrix}
0.0036 & 0 & 0 & 0 \\
0 & 0.0010 & 0 & 0 \\
0 & 0 & 0.1000 & 0 \\
0 & 0 & 0 & 0.019
\end{bmatrix}.
$$

The bivariate histograms in **Figure 8** also resemble their counterparts in **Figures 5–7**. **Figure 8L** displays an isolated pair of ϕ_c and P.

DISCUSSION

The Bayesian analysis method allowed us to determine quantitatively the most and least significant model inputs. Importantly, it revealed correlations between parameters that are not obvious. In all four analyses, the posterior of pressure and critical porosity indicated isolated pairs. Each analysis had a different set of values. If another prior is used, the result would change in terms of the value of pressure and critical porosity, but an isolated set occurs. This correlation is not intuitive, but it indicates that these two parameters are the most significant. Posterior histograms that contain saturation or clay content is more intuitive. More specifically, it is
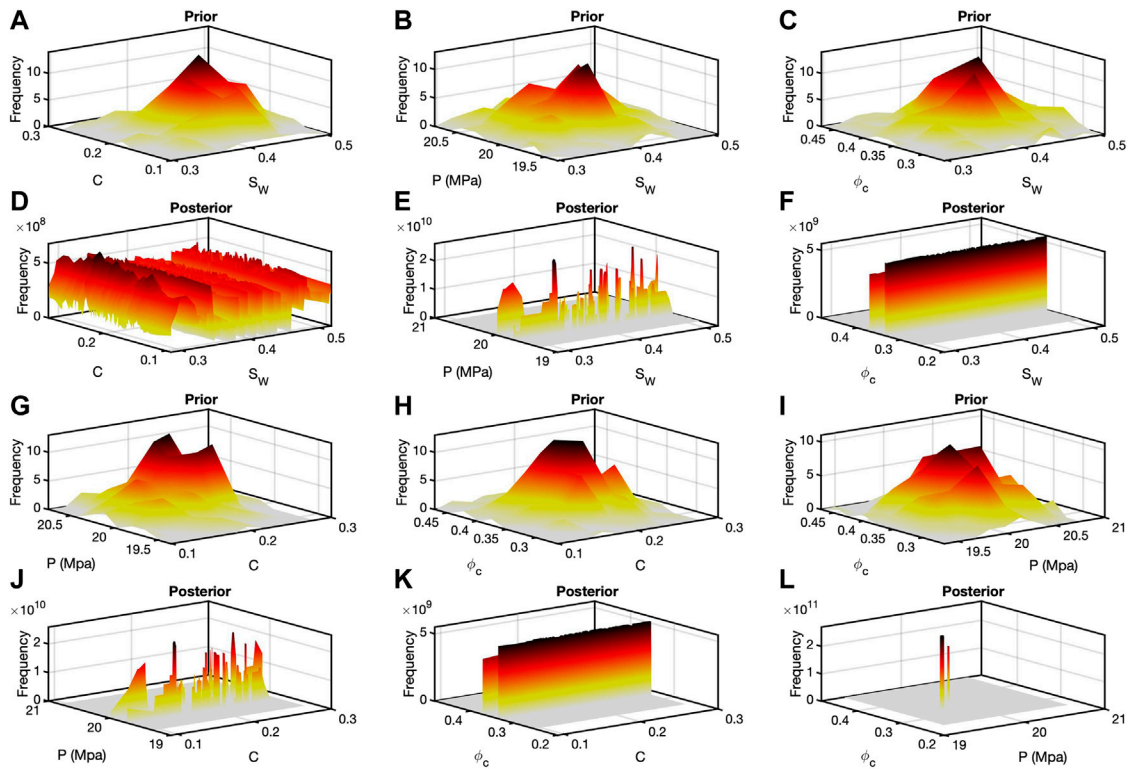
FIGURE 5 | Bivariate histograms for analysis one. Plots **(A–C)** are the bivariate prior histograms for $C - S_w$, P-S_w, and $\phi_c - S_w$, respectively. Plots **(D–F)** contain the posterior histograms corresponding to the priors in **(A–C)**. Priors for P-C, $\phi_c - C$, and $\phi_c - P$ are in **(G–I)** with their corresponding posteriors in **(J–L)**. When many pairs in a posterior occur together, such as in **(D)** $(C - S_w)$ that indicates relative insensitivity to those two variables. Counter to that is **(L)** where isolated pairs of ϕ_c and P occur together in even larger numbers.
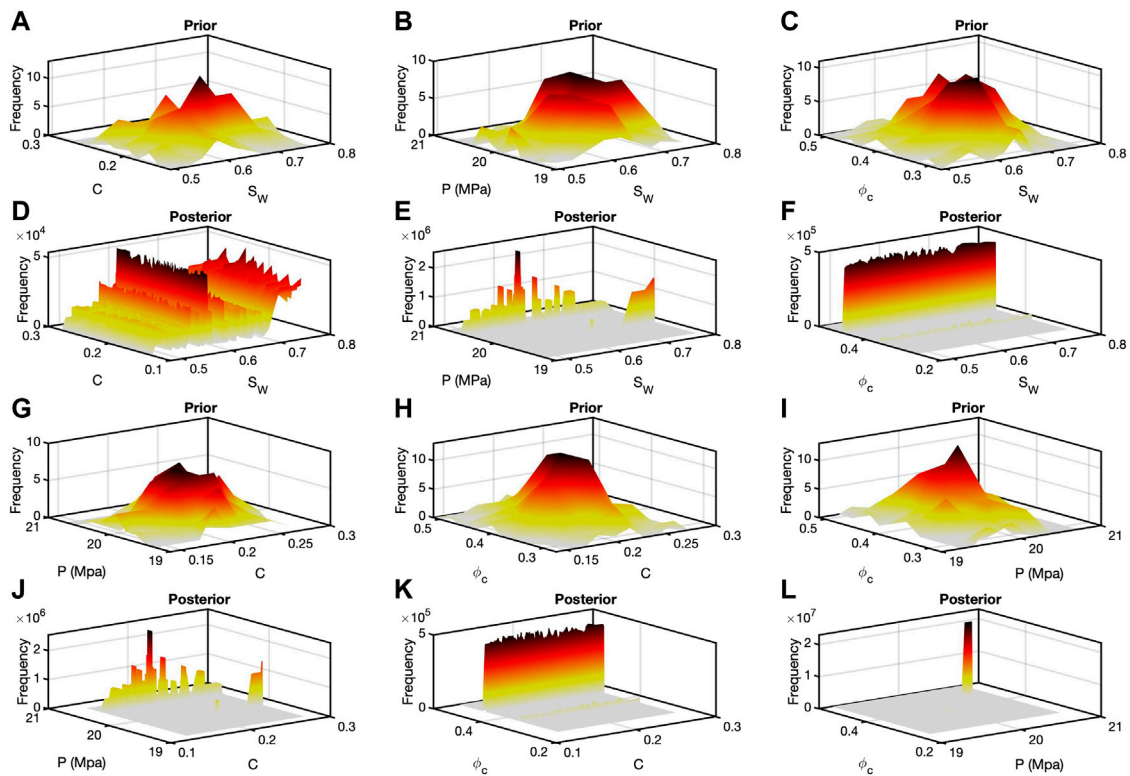
FIGURE 6 | Bivariate histograms for analysis two. The juxtaposition is the same as in **Figure 5**. The prior model differed between analysis one and two, but the bivariate posterior histograms show patterns similar to those in **Figure 5**.
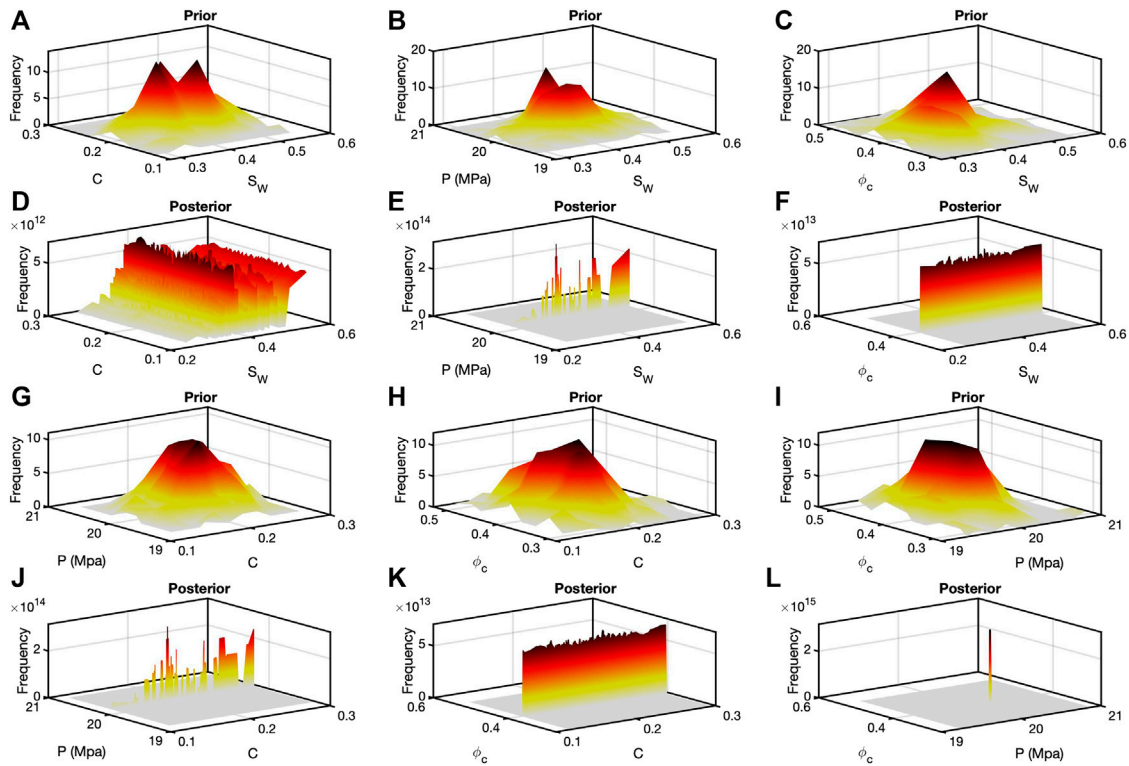
FIGURE 7 | Results from anallysis three for gas sands displayed again as bivariate prior and posterior histograms. The layout is identical to **Figure 5**. The prior was different from inversions one and two including different fluid properties of gas relative to oil. However, the posterior histograms, in particular **(D)** and **(L)**, resemble their counterparts in **Figures 5, 6**.
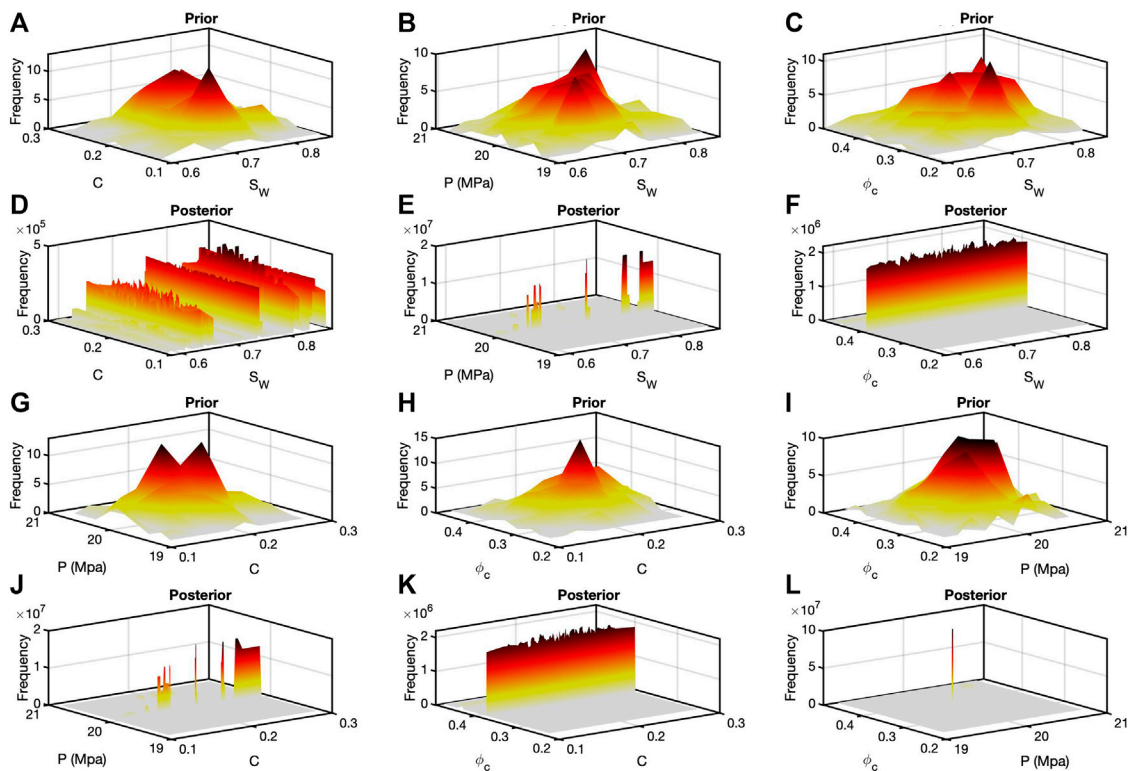

FIGURE 8 | Analysis four, for gas sands, resulted in similar bivariate histograms compared with those in **Figure 7**. However, the frequencies are considerably lower. This is the result of the prior model not being as close to the data as they were in the other inversions. Nonetheless, the patterns repeated themselves among the different posteriors.

understandable that a small change in either might not change the velocity very much. When one or the other is paired with pressure or critical porosity, the result is a relatively broad range of with a narrow selection of pressure or critical porosity. When applying this method to well data, we considered, fluid types and saturations, and mineralogy along with pressure and critical porosity. Preliminary work indicated similar correlations between pressure and critical porosity on dry, clean sands. The extension to include fluids and composition indicates similar patterns and correlations that were not expected.

This study demonstrates a way to determine the significance of inputs into one rock-physics model. The significance becomes evident of any singular or joint set of rock-physics model parameters. In an application to reservoir characterization, the most significant terms should be set first, and the others can be more loosely defined. If this model was deemed appropriate on a different data set, then in a deterministic application, the most significant inputs should be selected first and jointly. The other inputs would then provide more subtle fine tuning. This approach would limit the number of variables to consider within the requisite sensitivity study for that data set. If this model was used in a statistical seismic inversion, a user could set narrow limits on the priors for the most significant parameters and wider ranges for the lesser ones. The effect of this would be to reduce the size of the model parameter space to explore. This was work done on one model one data set, and the results are relevant specific to this model and applicable to other data sets. Knowledge about this one particular model might be useful for similar types of models, given an appropriate data set, but not likely for other types of models.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: Data release contract does not allow of public dissemination of the data used. Requests to access these datasets should be directed to KS, kyle.spikes@jsg.utexas.edu.

AUTHOR CONTRIBUTIONS

KS wrote the codes, the manuscript, and generated the figures MS provided initial motivation for the work, edited the manuscripts, and critiqued the results.

REFERENCES

Bosch, M., Mukerji, T., and Gonzalez, E. (2010). Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: a review. *Geophysics* 75 (5), 75A165–75A176. doi:10.1190/1.3478209

Contreras, A. J. (2006). Spatial delineation, fluid-lithology characterization, and petrophysical odeling of deepwater Gulf of Mexico reservoirs through joint AVA deterministic and stochastic inversion of 3D partially-stacked seismic amplitude data and well logs. *Doctoral dissertation*. Austin, (TX): The University of Texas at Austin.

Dvorkin, J., and Nur, A. (1996). Elasticity of high-porosity sandstones: theory for two North Sea data sets. *Geophysics* 61, 1363–1370. doi:10.1190/1.1444059

Gassmann, F. (1951). Uber die elastizitat poroser medien. *Vier. der Natur.* 96, 1–23.

Grana, D. (2018). Joint facies and reservoir properties inversion. *Geophysics* 83 (3), M15–M24. doi:10.1190/geo2017-0670.1

Hashin, Z., and Shtrikman, S. (1963). A variational approach to the theory of the elastic behavior of multiphase materials. *J. Mech. Phys. Solids* 11, 127–140. doi:10.1016/0022-5096(3)90060-7

Mavko, G., Mukerji, T., and Dvorkin, J. (2009). *The rock physics handbook*. Cambridge, United Kingdom: Cambridge University Press, 254.

Mindlin, R. D. (1949). Compliance of elastic bodies in contact. *ASME J. Appl. Mech.* 16, 259–268.

Nawaz, M. A., and Curtis, A. (2017). Bayesian inversion of seismic attributes for geological facies using a hidden Markov model. *Geophys. J. Int.* 208, 1184–1200. doi:10.1093/gji/ggw411

Sen, M. K. (2006). *Seismic inversion*. Richardson, TX: Society of Petroleum Engineers, 120.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Paris, France: Society for Industrial and Applied Mathematics, 352.

Ulrych, T. J., Sacchi, M. D., and Woodbury, A. (2001). A Bayes tour of inversion: a tutorial. *Geophysics* 66 (1), 55–69. doi:10.1190/1.1444923

Ulvmoen, M., Omre, H., and Buland, A. (2010). Improved resolution in bayesian lithology/fluid inversion from prestack seismic data and well observations: part 2—real case study. *Geophysics* 75 (2), B73–B82. doi:10.1190/1.3335332

Ulvmoen, M., and Omre, H. (2010). Improved resolution in bayesian lithology/fluid inversion from prestack seismic data and well observations: part1—methodology. *Geophys.* 75 (2), R21–R35. doi:10.1190/1.3294570

Xie, W., and Spikes, K. T. (2021). Reservoir facies design and modeling using probabilistic rock-physics templates. *Geophysics* 86 (1), 1–12. doi:10.1190/geo2020-0044.1