Universidad Nacional Autónoma de México
Posgrado en Ciencias de la Tierra
Instituto de Geofísica
Exploración, Aguas Subterráneas, Modelación y
Percepción Remota

Copula-based modeling for petrophysical property prediction using seismic attributes as secondary variables

T E S I S

PARA QUE OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS DE LA TIERRA

PRESENTA:

VAN HUONG LE

Director de Tesis:
Dr. Martín Alberto Díaz Viera
Instituto Mexicano del Petróleo
Miembros del Comité Tutor:
Dr. José Manuel Grajales Nishimura
Instituto de Geología
Dr. Ricardo Casar González
Posgrado en Ciencias de la Tierra

Ciudad Universitaria, CD. MX., Marzo 2021

"I declare to know the Code of Ethics of the National Autonomous University of Mexico, embodied in the University Legislation. Based on the definitions of integrity and honesty specified therein, I assure by my signature below that this work is original and entirely my authorship. All citations to, or references to, the work of other authors appear duly and adequately indicated, as well as accredited through conventional editorial resources."

Van Huong Le, University City, Mexico City, March 2021.

"Declaro conocer el Código de Ética de la Universidad Nacional Autónoma de México, plasmado en la Legislación Universitaria. Con base en las definiciones de integridad y honestidad ahí especificadas, aseguro mediante mi firma al calce que el presente trabajo es original y enteramente de mi autoría. Todas las citas de, o referencias a, la obra de otros autores aparecen debida y adecuadamente señaladas, así como acreditadas mediante los recursos editoriales convencionales".

Van Huong Le, Ciudad Universitaria, CD. MX., Marzo de 2021.

Credits for the oil information provided by the CNIH to the UNAM

The data were provided by the National Hydrocarbons Commission of Mexico, according to appendix C of the license to use the information in favor of Universidad Nacional Autónoma de México, dated December 11, 2017, under the nomenclature CNIH-C-00417. This reference information is the property of Mexico and its collection, safekeeping, use, administration and updating, as well as its publication is only authorized to the National Hydrocarbons Commission.

Créditos de la información petrolera proporcionada por el CNIH a la UNAM

Los datos fueron proporcionados por la Comisión Nacional de Hidrocarburos de México, según anexo C de la licencia de uso de la información a favor de la Universidad Nacional Autónoma de México, de fecha 11 de diciembre de 2017, bajo la nomenclatura CNIH-C-00417. Esta información de referencia es propiedad de México y su recolección, custodia, uso, administración y actualización, así como su publicación, solo está autorizada a la Comisión Nacional de Hidrocarburos.

Dedicatory

This work is dedicated to my children, Le Tra Nhat Nam (Bean) and Le Tra Anh Thu (Marina), who gives me daily happiness.

To my love, my wife, Tra Thi Doan Dung, who is always by my side through good and bad times.

To my mom, Nguyen Thi Teo, who always gives me the valuable advice and love eternally.

Acknowledgements

To the director of my thesis, Dr. Martín Alberto Díaz Viera for giving me his teaching, understanding and help at all times.

To the researchers: Dr. José Manuel Grajales Nishimura, Dr. Ricardo Casar González, Dr. Arturo Erdely, Dr. Raúl Del Valle García, Dr. Dario Grana, Dr. Elsa Leticia Flores, and Dr. Ursula Xiomara Iturrarán Viveros who reviewed and advised my research work.

To the staff of the Graduate Program in Earth Sciences, UNAM; especially to Araceli Chaman, Gloria and Erika for helping me unconditionally.

To UNAM and the Institute of Geophysics for giving me the opportunity to be a scientist.

To the National Council of Science and Technology (CONACyT) for giving me its scholarships during my study time.

To the national hydrocarbon information center (CNIH) for allowing me to use its data for my research work.

To all those who somehow made this achievement possible in my life.

Abstract

A new method for the prediction of the spatial distribution of petrophysical properties using the elastic seismic attributes as secondary variables is presented. This method basically consists of two stages: first the dependency relationship between the variables is established by means of the joint probability distribution function using copulas and later a global optimization method (such as simulated annealing or differential evolution) is applied using a variogram model and previously obtained joint probability distribution as objective functions. The proposed method is very flexible because unlike traditional methods, which assume linear dependencies, it allows modeling more general dependency relationships between variables. Depending on the approach in the construction of the dependency model, the method can be: parametric, semiparametric and nonparametric.

The proposed method is first validated in one-dimensional cases at a well log scale and then is applied in two-dimensional cases in a seismic section for a marine hydrocarbon reservoir located in the Gulf of Mexico that belongs to the province of the Mexican Cordilleras. The reservoir is siliciclastic and it is formed mainly of alternative marine sequences of sands and shales.

A bivariate case comparison between two simulation methods: traditional Sequential Gaussian CoSimulation (SGCS) and Bernstein Copula-based Spatial CoSimulation (BCSCS) is done. The results show that the BCSCS method significantly reduces the uncertainty compared to the traditional SGCS method in terms of its univariate, bivariate and spatial distributions, respectively.

The method is extended to multivariate joint distributions by applying Vine copulas or bivariate copula blocks, and is applied to bivariate and multivariate cases using simulation and estimation approaches, respectively. The results show that the simulation approach has the advantage of reproducing with high precision the univariate and bivariate behaviors, and the spatial variability of the petrophysical properties, and its main disadvantage is a higher computational cost. In contrast, the estimation approach underestimates the spatial variability of petrophysical properties but at a lower computational cost.

Resumen

Se presenta un nuevo método para la predicción de distribución espacial de propiedades petrofísicas utilizando los atributos sísmicos elásticos como variables secundarias. Este método, básicamente consta de dos etapas: primero se establece la relación de dependencia entre las variables mediante la función de distribución de probabilidad conjunta usando cópulas y posteriormente se aplica un método de optimización global (como recocido simulado o evolución diferencial) usando un modelo de variograma y la distribución de probabilidad conjunta previamente obtenida como funciones objetivo. El método propuesto es muy flexible porque a diferencia de los métodos tradicionales, que suponen dependencias lineales, éste permite modelar relaciones más generales de dependencia entre variables. Dependiendo del enfoque en la construcción del modelo de dependencia el método puede ser: paramétrico, semiparamétrico y no paramétrico.

El método propuesto se valida primero en casos unidimensionales a escala de registro de pozo y luego se aplica en casos bidimensionales en una sección sísmica para un yacimiento de hidrocarburos marino ubicado el Golfo de México que pertenece a la provincia de las Cordilleras Mexicanas. El yacimiento es siliciclástico y está formado principalmente por secuencias marinas alternativas de arenas y lutitas.

Se realiza una comparación de un caso bivariado entre dos métodos de simulación: la cosimulación secuencial Gaussiana tradicional (SGCS: por sus siglas en inglés) y la cosimulación espacial basada en cópula de Bernstein (BCSCS: por sus siglas en inglés). Los resultados muestran que el método BCSCS reduce significativamente la incertidumbre en comparación con el método tradicional SGCS en cuanto a sus distribuciones univariada, bivariada y espacial, respectivamente.

El método se extiende a distribuciones conjuntas multivariadas aplicando cópulas de Vine o bloques de cópulas bivariadas y se aplica a casos bivariados y multivariados utilizando los enfoques de simulación y estimación, respectivamente. Los resultados muestran que el enfoque de simulación tiene la ventaja de reproducir con alta precisión los comportamientos univariados y bivariados, y la variabilidad espacial de las propiedades petrofísicas, y su desventaja principal es un mayor costo computacional. Por el contrario, el enfoque de estimación subestima la variabilidad espacial de las propiedades petrofísicas pero a un menor costo computacional.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Petrophysical modeling in reservoir characterization consists of predicting petrophysical properties and their spatial configuration within the reservoir model (Cosentino, 2001). These properties cannot be measured directly in the reservoir, and they are usually inferred from other measurements (e.g. elastic seismic attributes). Petrophysical property modeling in hydrocarbon reservoirs is challenging because of the limited amount of data available and the uncertainty in the measurements. For this reason, in recent years, stochastic simulation approaches have been adopted for the spatial distribution of petrophysical properties (Caers, 2005; Coburn et al., 2006; Deutsch, 2002; Doyen, 2007; Dubrule, 2003; Yarus and Chambers, 1994).

Seismic attributes are commonly used as secondary variables for predicting rock and fluid properties. Among the estimation approach, the most common methods are regression models (Chatterjee et al., 2016; Yenwongfai et al., 2017), neural networks (Alfarraj and AlRegib, 2018; Gogoi and Chatterjee, 2019; Iturrarán-Viveros, 2012; Iturrarán-Viveros and Parra, 2014; Maurya and Singh, 2019), and cokriging (Babak and Deutsch, 2009; Doyen et al., 1996; Moon et al., 2016; Xu et al., 2016). The advantage of these estimation methods is the low computational cost. Therefore, they can work with large data sets and give a quick result. But they have several disadvantages, since they require a large amount of data, produce smoothed results underestimating the variability due to natural heterogeneities, do not reproduce statistical properties, and do not give a systematic way to quantifying the uncertainty.

To overcome these challenges, spatial simulation approach has been proposed (Caers, 2005; Chilès and Delfiner, 2012; Coburn et al., 2006; Deutsch, 2002; Doyen, 2007; Dubrule, 2003; Horta and Soares, 2010; Yarus and Chambers, 1994). The most common simulation methods are Sequential Gaussian CoSimulation (SGCS) (Afshari and Shadizadeh, 2015; Almeida and Frykman, 1994; Almeida and Journel, 1994; Cao et al., 2014; Emery and Parra, 2013; Gómez-Hernández and Journel, 1993; Parra and Emery, 2013; Verly, 1993), direct sequential cosimulation (Azevedo and Soares, 2017; Horta and Soares, 2010; Soares, 2001, 2005; Soares et al., 2017), simulated annealing cosimulation (Dafflon and Barrash, 2012; Deutsch and Cockerham, 1994a,b; Vejbæk and Kristensen, 2000), Gaussian mixture cosimulation (Figueiredo et al., 2019; Grana et al., 2017, 2012;

Grana and Rossa, 2010; Lang and Grana, 2017), and nonparametric mixture approach using kernel smoothing (Corina and Hovda, 2018; Grana, 2018). Spatial simulation methods reproduce the spatial variability and statistical properties; as they are often used for uncertainty quantification and do not depend on large dataset availability (Doyen, 2007).

The SGCS method assumes stationary random functions and it is generally applied with a linear co-regionalization model (Chilès and Delfiner, 2012). These assumptions imply a linear dependence between the primary and secondary variables. An extension of the SGCS method is the GauMix method where the model distribution is assumed to be a linear combination of multiple Gaussian pdfs. However, Gaussian distributions and the linear assumption do not correctly capture the dependency relationships between the variables in the real dataset.

For those reasons, this thesis presents copula-based spatial distribution prediction method that can correctly model and reproduce complex relationships between variables without any assumption of linearity or Gaussian distribution. Copulas have become popular in the financial sector to model complex data relationships and have been recently applied in geosciences (Bárdossy, 2006; Bevacqua et al., 2017; Bárdossy and Li, 2008; Díaz-Viera and Casar-González, 2005; Erdely and Díaz-Viera, 2010, 2015; Gräler, 2014; Gräler and Pebesma, 2011; Haslauer et al., 2010; Hernández-Maldonado et al., 2012, 2014; Kazianka and Pilz, 2010; Krupskii and Genton, 2019; Stien and Kolbjørnsen, 2008). There are two approaches to estimate the joint dependence: parametric and nonparametric copula.

In the published literature, there are examples of parametric copula approach by several authors. Díaz-Viera et al. (2006); Díaz-Viera and Casar-González (2005) applied copula using dependency measures such as Kendall and Spearman to simulate permeability using porosity as secondary variable; Bárdossy and Li (2008) applied a Gaussian copula for spatial interpolation of nitrate concentration; Kazianka and Pilz (2010) applied a Gaussian copula for spatial interpolation of benchmark geostatistical dataset; Erdely and Díaz-Viera (2015) applied a Vine trivariate parametric copula to predict permeability conditioned to porosity and P wave velocity. The nonparametric approach has been developed primarily by Erdely and Díaz-Viera (2010) and Hernández-Maldonado et al. (2012), who applied the bivariate Bernstein copula to predict permeability conditioned to porosity; Hernández-Maldonado et al. (2014) applied the trivariate Bernstein copula to predict permeability conditioned to porosity and S-wave velocity; Díaz-Viera et al. (2017) applied Bernstein copula to predict total porosity conditioned to P-impedance in 2D applications.

This research work is a natural extension of the bivariate case by Díaz-Viera et al. (2017) to the multivariate case of more than two variables base on Vine copula and shows the application of proposed method to reservoir properties modeling conditioned to elastic attributes. The method is based on realizations (sample the values) obtained using the simulated annealing or differential evolution where the joint probability distribution is estimated by copulas. The novelty of this research work is in the use of copula-based method for geophysical inverse problems. It combines the inference of joint cumulative distribution function with the optimization method (simulation annealing, differential evolution) to predict geostatistical realizations of reservoir properties.

A bivariate case comparison at well log scale was done between the simulation methods: BCSCS

and traditional SGCS. The uncertainty quantification analysis in all terms the univariate and joint distribution function and the spatial distribution function shows that the BCSCS method significantly reduces the uncertainty compared to the traditional SGCS method. The two copula-based spatial distribution estimation and simulation approaches were performed and compared in the bivariate and multivariate cases at the well log scale. Furthermore, these approaches are extended to multidimensional problems with applications to 2D sections of seismic attributes.

Brief reviews of State-of-the-Art is presented in chapter 2 where the prediction methods developed so far from two estimation and simulation approaches are shown. Then, the the antecedent, the assumptions, mathematical formulation, algorithm, computational implementation and validation of proposed method based on copula is described in chapter 3. The methodology is presented in chapter 4 where the workflow is shown within the context of the reservoirs geological-petrophysical modeling such as exploratory data analysis, variographic analysis, copula-based dependency modeling, method validation and application, and uncertainty quantification. The conceptual geological model of the study area is presented in chapter 5. The application of the methodology to the bivariate case is shown in chapter 6 and to the multivariate case in chapter 7. Finally, the conclusions and future work are described in chapter 8.

Chapter 2

Brief review of State-of-the-Art

There are two approaches in the literature to predict petrophysical properties using seismic attributes as secondary variables: estimation and simulation. The advantage of estimation methods is the fast execution time. Therefore, they can be applied with large data sets and give a fast result. But they have several disadvantages, since they produce smoothed results underestimating the spatial variability due to natural heterogeneities, do not reproduce statistical properties such as the mean, median, variance, maximum, minimum, etc., and do not give a systematic way to quantifying the uncertainty. On the other hand, the simulation methods reproduce both the statistical properties or behavior of univariate and joint, and the spatial variability of the random variable, but with a high computational cost.

2.1 Estimation approach

In the literature there are many methods of the estimation approach, the following is a brief description of the most used methods: multiple linear regression, neural networks, and cokriging.

2.1.1 Multiple linear regression

In statistics, linear regression is a linear approach to model the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression (Freedman, 2009). The following are the major assumptions made by standard linear regression models: linearity, constant variance, independence of errors (more detail in Poole and O'farrell (1971)). The general linear regression model can be expressed in the following equation

$$Y = a + \sum_{i=1}^{n} b_i.X_i + u \tag{2.1}$$

where Y is the dependent variable; $X_1, X_2, ...X_i...X_n$ are n independent variables; a and b_i are the regression coefficients, representing the parameters of the model for a specific population; and u is a stochastic disturbance-term.

The limitations of this model are in the assumptions, such as the assumption of linearity between the variables, the constant variance, etc. The advantage of this model is a simple and efficient model when assumptions are fulfilled. In the literature, there are works that applied the linear regression model (Equation 2.1) to establish the dependency relationship between petrophysical properties and seismic attributes, and then predict petrophysical properties using seismic attributes as secondary variables, e.g. Chatterjee et al. (2016); Jalalalhosseini et al. (2015); Vernik and Nur (1992); Yenwongfai et al. (2017).

2.1.2 Neural networks

Neural networks can be programmed to recognize patterns, to store and retrieve database entries, to solve optimization problems, to filter noise from experimental data and to estimate sampled functions when the analytical form of the function is unknown. An artificial neural network is basically a system of several simple processing units known as nodes or neurons associated with each other by simple connections. The strength of these connections may be changed by varying the weight attached to them. The process of adjusting the weight values is known as the training process. The training process is the fundamental procedure through which the neural network is calibrated to the particular estimation problem. Therefore, in order to apply this technique, a good data set of reference information is needed.

In contrast to any other estimation method, neural networks do not make use of a pre-defined relationship, since the estimation function is built through experience during the training phase. In this respect, neural networks are model-free estimators. Another interesting point is that the predicted distribution does not obey any statistical rule, e.g., the preservation of the mean value. In fact, one of the drawbacks of the predictions made through regressions, i.e., the smoothing effect and the loss of the extreme values, is not a concern in the case of neural networks. In fact, the technique allows the actual variability of the data to be preserved.

Neural networks have some disadvantages, too. Firstly, the training process has to be done with caution and can be a lengthy process, which requires a good calibration data set. On the other hand, failing in correctly calibrating the network may result in aberrant results. Another point to take into consideration is that the methodology not yet an off the shelf application and requires expertise by the geoscientist (Cosentino, 2001). In the literature, there are works that applied the neural networks to predict petrophysical properties such as Alfarraj and AlRegib (2018); Gogoi and Chatterjee (2019); Iturrarán-Viveros (2012); Iturrarán-Viveros and Parra (2014); Maurya and Singh (2019).

2.1.3 Cokriging

The basic principle in this case is to estimate the variable of interest by means of a correlation function defined by well data (the variogram), plus a cross-correlation function (linear corregionalization model) that quantifies the relationship with another, related variable, i.e., the seismic information. This procedure is called cokriging. From a theoretical viewpoint, the correct way to operate is to solve the full cokriging system of the 2 variables, porosity and seismic attribute. However, from a practical standpoint, cokriging is a cumbersome procedure and moreover, when applied to the dense population of seismic data, often gives problems of matrix instability in the solution. Therefore, simplified forms of cokriging, called collocated cokriging. Th advantage in this case is that tile implementation of collocated cokriging does not requier the construction of the cross-variograms, but only the knowledge of the correlation coeficient between the 2 variables (Cosentino, 2001).

The advantage of the method is that the spatial variability of the random variables is reproduced. The disadvantage is in its assumptions of the linear corregionalization model or the linear correlation coefficient, which require high linear dependence between the random variables, more detail see Díaz-Viera (2002). In the literature, there are works that applied cokriging to predict petrophysical properties such as Babak and Deutsch (2009); Doyen et al. (1996); Moon et al. (2016); Xu et al. (2016).

2.2 Simulation approach

2.2.1 Sequential Gaussian cosimulation

The Gaussian sequential Co-simulation method is a well-known method in the Geostatistical literature, so here we will only mention the details of its application. The first assumption is to assume that each random function is strictly stationary. It is also assumed that the univariate probability distribution functions and the joint probability distribution function are Gaussian (Normal) (Verly, 1992). Ussually this method is applied with a linear coregionalization model (Chilès and Delfiner, 2012) that is mostly unnatural, forced, very complicated and difficult to establish. The method assumes the existence of a very strong linear dependence between the primary and secondary variables, which is its main assumption and, at the same time, its main drawback (Díaz-Viera et al., 2017). This method was implemented by (Remy et al., 2009) in SGeMS with different variants. There are some methods related to this method such as Sequential Gaussian mixture cosimulation and Direct Sequential Cosimulation.

In the literature, there are works that applied these methods to predict petrophysical properties such as, sequential Gaussian cosimulation (SGCS) (Afshari and Shadizadeh, 2015; Almeida and Frykman, 1994; Almeida and Journel, 1994; Cao et al., 2014; Emery and Parra, 2013; Gómez-Hernández and Journel, 1993; Parra and Emery, 2013; Verly, 1993), direct sequential cosimulation (Azevedo and Soares, 2017; Horta and Soares, 2010; Soares, 2001, 2005; Soares et al., 2017), and

Gaussian mixture cosimulation (Figueiredo et al., 2019; Grana et al., 2017, 2012; Grana and Rossa, 2010; Lang and Grana, 2017).

2.2.2 Annealing cosimulation

This review is primarily based on Deutsch and Cockerham (1994a), and Deutsch and Journel (1998).

Applying the numerical technique known as "simulated annealing" to geostatistical simulation is relatively new. The name Annealing Cosimulation (ACS) is derived from: 1) through common usage simulated annealing has been shortened to "annealing", and 2) stochastic simulation of one attribute conditioned to others may be referred to as "cosimulation" (compare with cokriging). Interest in the annealing methodology is based on its ability to honor a wide variety of input data. The technique of simulated annealing is based on an analogy with the physical process of annealing and is typically applied to global optimization problems. Annealing is the process by which a material undergoes extended heating and is slowly cooled. Thermal vibrations permit a reordering of the atoms/molecules to a highly ordered lattice, i.e., a low energy state. In the context of 3-D numerical modeling, the annealing process may be simulated by the following steps:

1. Create an initial 3-D numerical model (analogous to the initial alloy in true annealing) by assigning a permeability value to each grid node at random from the population distribution.
2. Define an energy or objective function (analogous to the Gibbs free energy in true annealing) as a measure of difference between desired features and those of the realization, e.g., the objective function could include the squared difference between the variogram of the realization and a model variogram derived from core data.
3. Perturb the model (analogous to the thermal vibrations in true annealing) by visiting a random location in the 3-D numerical model and assigning a new permeability value. The new value is a random drawing from the conditional distribution of permeability given the collocated porosity value.
4. Accept the perturbation (thermal vibration) if the objective function is decreased; reject it if the energy has increased (the Boltzmann probability distribution of true annealing).
5. Continue the perturbation procedure until the low energy state is achieved. Low energy states correspond to plausible 3-D numerical models of the reservoir.

In general, the objective function is made up of the weighted sum of N_c components:

$$O = \sum_{c=1}^{N_c} w_c.O_c \tag{2.2}$$

where O is the total objective function, W_c and O_c are weights and component objective functions respectively. Each component is designed to account for a source of data. The weights are calculated such that all components of the objective function are lowered to zero at the end of the

annealing process.

In addition to the definition of an objective function, a critical aspect of simulated annealing-based simulation algorithms is a prescription for when to accept or reject a given perturbation. The acceptance probability distribution is given by the Boltzmann distribution:

$$P(accept) = \begin{cases} 1 & \text{if } O_{new} \leq O_{old} \\ e^{\frac{O_{new} - O_{old}}{t}} & \text{otherwise} \end{cases} \tag{2.3}$$

All favorable perturbations ($O_{new} \leq O_{old}$) are accepted and unfavorable perturbations are accepted with an exponential probability distribution. The parameter t of the exponential distribution is analogous to the "temperature" in annealing. The higher the temperature, the more likely an unfavorable perturbation will be accepted.

The temperature t must not be lowered too fast or else the image may get trapped in a suboptimal situation and never converge. However, if lowered too slowly, then convergence may be unnecessarily slow. The specification of how to lower the temperature t is known as the "annealing schedule". The following empirical annealing schedule is one practical alternative.

The idea is to start with an initially high temperature t_0 and lower it by some multiplicative factor λ whenever enough perturbations have been accepted (K_{accept}) or too many have been tried (K_{max}). The algorithm is stopped when efforts to lower the objective function become sufficiently discouraging. The following parameters describe this annealing schedule:

- t_0: the initial temperature.
- λ: the reduction factor $0 < \lambda < 1$.
- K_{max}: the maximum number of attempted perturbations at any one temperature (on the order of 100 times the number of nodes). The temperature is multiplied by λ whenever K_{max} is reached.
- K_{accept}: the acceptance target. After K_{accept} perturbations are accepted, the temperature is multiplied by λ (on the order of 10 times the number of nodes).
- S: the stopping number. If K_{max} is reached S times then the algorithm is stopped (usually set at 2 or 3).
- ΔO: a low objective function indicating convergence.

The advantage of this method is if it reproduces the behaviors of univariate and bivariate distribution and also the spatial variability of the random variables. The only drawback of this method is that the conditional distribution function is based on a discretized bivariate probability distribution. In the literature, there are works that applied this method to predict petrophysical properties such as Dafflon and Barrash (2012); Deutsch and Cockerham (1994a,b); Vejbæk and Kristensen (2000).

Chapter 3

Copula-based dependency modeling

3.1 Brief introduction to Copula

A copula is a multivariate cumulative distribution function for which the marginal probability distribution of each variable is uniform on the interval $[0, 1]$. Copulas are used to describe the dependence between random variables. Sklar's theorem states that any multivariate joint distribution can be written in terms of univariate marginal distribution functions and a copula which describes the dependence structure between the variables. Copulas are popular in high-dimensional statistical applications as they allow one to easily model and estimate the joint distribution of random vectors by estimating marginals and copula separately.

In order to apply the copula-based model for predicting petrophysical properties using seismic attributes as secondary variables, it is required that the petrophysical properties (total and effective porosity, permeability, water saturation, and clay volume) and elastic seismic attributes (P-velocity, S-velocity, density, P-impedance, S-impedance, etc.) are random variables and its univariate distributions are continuous. Then, the joint distribution of these random variables is constructed by modeling the univariate distributions and the copula. Finally, with the joint distribution it is possible to predict the petrophysical properties conditioned to the elastic seismic attributes.

3.1.1 Bivariate copula

According to Sklar's theorem in Sklar (1959): Let F_{12} be a bivariate joint probability distribution function and F_1 and F_2 be univariate (marginal) probability distribution functions. A copula is a function C: $[0, 1]^2 \mapsto [0, 1]$ such that for all x_1, x_2 in $\mathbb{R}$,

$$F_{12}(x_1, x_2) = C_{12}(F_1(x_1), F_2(x_2)) \tag{3.1}$$

If F_1 and F_2 are continuous, then C_{12} is unique; otherwise, C_{12} is uniquely determined on

$RanF_1 \times RanF_2$. Copula associated to a bivariate random vector (X_1, X_2) describes the relationship between X_1 and X_2, and independently from their univariate probabilistic behavior.

Let $F_1(x_1) = u, F_2(x_2) = v$. According to Nelsen (2006) and Joe (2014), the properties of copulas are:

1. $C_{12}(0, v) = 0 = C_{12}(v, 0)$.
2. $u = C_{12}(u, 1)$.
3. $v = C_{12}(1, v)$.
4. For each u_1, u_2, v_1, v_2 in $\mathbb{I} = [0, 1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$ then: $C_{12}(u_2, v_2) - C_{12}(u_1, v_2) - C_{12}(u_2, v_1) + C_{12}(u_1, v_1) \geq 0$.
5. C_{12} is uniformly continuous on its domain $[0, 1]^2$.
6. A convex linear combination of copula functions is also a copula function.
7. The horizontal, vertical, and diagonal sections of a copula C_{12} are all nondecreasing and uniformly continuous on $[0, 1]$.
8. The random variables u y v are independent if and only if $C_{12}(u, v) = \Pi(u, v) = uv$.
9. $W(u, v) \leq C_{12}(u, v) \leq M(u, v)$ where $W(u, v) = max(u + v - 1, 0)$ and $M(u, v) = min(u, v)$ are also copulas known as the lower and upper Fréchet-Hoeffding bounds.
10. If α and β are strictly increasing on $RanX_1$ and $RanX_2$, respectively, then $C_{\alpha(X_1)\beta(X_2)} = C_{12}$. Thus C_{12} is invariant under strictly increasing transformations of X_1 and X_2.
11. $0 \leq \dfrac{\partial C_{12}(u, v)}{\partial u} \leq 1$ and $0 \leq \dfrac{\partial C_{12}(u, v)}{\partial v} \leq 1$.
12. $u \longmapsto \dfrac{\partial C_{12}(u, v)}{\partial v}$ and $v \longmapsto \dfrac{\partial C(u, v)}{\partial u}$ are nondecreasing.
13. The copula density function: $c_{12}(u, v) = \dfrac{\partial^2 C_{12}(u, v)}{\partial u \partial v}$ exists when C is absolutely continuous.
14. The conditional copula function: $c_u(v) = F_{V|U}(v|u) = P[V \leq v \mid U = u] = \lim\limits_{\Delta u \to 0} \dfrac{C_{12}(u + \Delta u, v) - C_{12}(u, v)}{\Delta u} = \dfrac{\partial C_{12}(u, v)}{\partial u}$.

The graphical representation of copula functions is presented below:

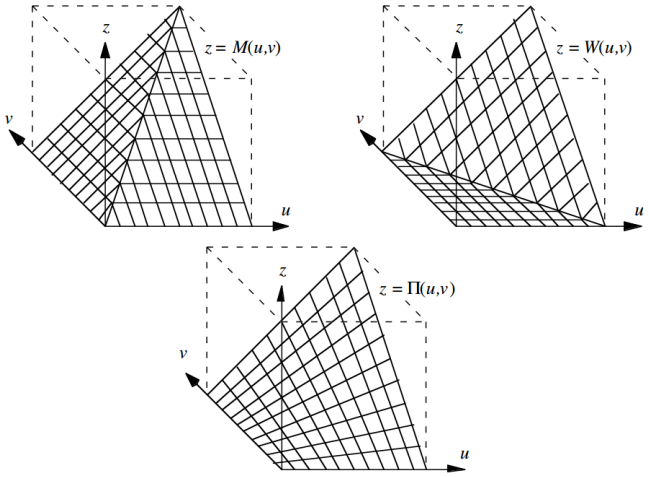


Figure 3.1: Graphs of the copulas M, Π y W (Nelsen, 2006).

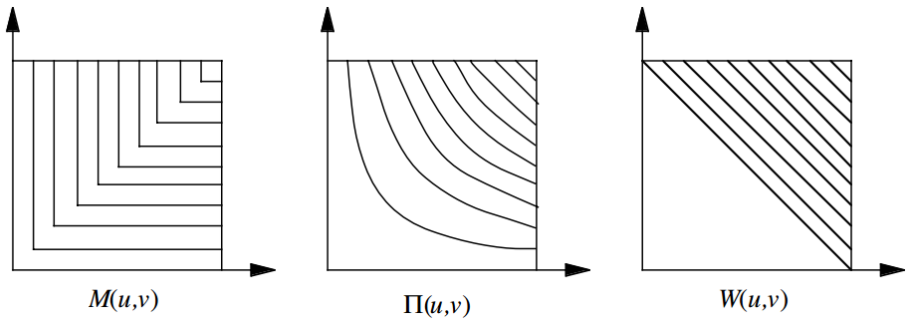


Figure 3.2: Contour diagrams of the copulas M, Π y W (Nelsen, 2006).

3.1.2 Multivariate copula: Vine copula

The trivariate probability distribution function can be written as a trivariate copula function of the univariate distributions in Equation 3.2 and the trivariate density function can be written as a function of the univariate distributions and the density copula function in Equation 3.3 by deriving the Equation 3.2.

$$F_{123}(x_1, x_2, x_3) = C_{123}\{F_1(x_1), F_2(x_2), F_3(x_3)\}. \tag{3.2}$$

$$f_{123}(x_1, x_2, x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{123}\{F_1(x_1), F_2(x_2), F_3(x_3)\} \tag{3.3}$$

According to Aas et al. (2009), the trivariate density function can be decomposed into the marginal functions and blocks of the bivariate copulas in Equation 3.4. The decomposition of

the high-dimensional copula to the bivariate copula blocks is called vine copula. This is very significant when constructing the multivariate copula function. It is very laborious to build the multivariate copula. But it is usually easier by means of the blocks of the bivariate copulas.

$$f_{123}(x_1,x_2,x_3) = f_1(x_1)\cdot f_2(x_2)\cdot f_3(x_3)\cdot c_{12}\{F_1(x_1),F_2(x_2)\} \cdot c_{23}\{F_2(x_2),F_3(x_3)\}\cdot c_{13|2}\{F_{1|2}(x_1|x_2),F_{3|2}(x_3|x_2)\} \quad (3.4)$$

Similarly, the distribution function and the density function of n variables can be written in Equation 3.5 and Equation 3.6.

$$F_{1\ldots n}(x_1,\ldots,x_n) = C_{1\ldots n}\{F_1(x_1),\ldots,F_n(x_n)\} \quad (3.5)$$

$$f_{1\ldots n}(x_1,\ldots,x_n) = f_1(x_1)\cdots f_n(x_n)\cdot c_{1\ldots n}\{F_1(x_1),\ldots,F_n(x_n)\}\cdot \quad (3.6)$$

The density function in Equation 3.6 can be rewritten by D-vine copula in Equation 3.7 or by C-vine (canonical vine) copula in Equation 3.8, see more detail in Aas et al. (2009).

$$f_{1\ldots n}(x_1,\ldots,x_n) = \prod_{k=1}^{n} f_k(x_k)\cdot \prod_{j=1}^{n-1}\prod_{i=1}^{n-j} c_{i,i+j|i+1,\ldots,i+j-1} \{F(x_i|x_{i+1},\ldots,x_{i+j-1}),F(x_{i+j}|x_{i+1},\ldots,x_{i+j-1})\} \quad (3.7)$$

$$f_{1\ldots n}(x_1,\ldots,x_n) = \prod_{k=1}^{n} f_k(x_k)\cdot \prod_{j=1}^{n-1}\prod_{i=1}^{n-j} c_{j,j+1|1,\ldots,j-1} \{F(x_j|x_1,\ldots,x_{j-1}),F(x_{j+i}|x_1,\ldots,x_{j-1})\} \quad (3.8)$$

The conditional distribution function and conditional density function can be written in Equation 3.9 and Equation 3.10.

$$F_{X|V}(x|v) = \frac{\partial C_{xv_j|v_{-j}}(F(x|v_{-j}),F(v_j|v_{-j}))}{\partial F(v_j|v_{-j})} \quad (3.9)$$

$$f_{X|V}(x|v) = c_{xv_j|v_{-j}}(F(x|v_{-j}),F(v_j|v_{-j}))\cdot f(x|v_{-j}) \quad (3.10)$$

3.2 Data-driven copula-based dependency modeling

Copula-based dependency modeling for petrophysical property prediction using elastic seismic attributes as secondary variables are presented below. Data for both petrophysical properties and elastic seismic attributes are assumed to be available at the well log scale. Then, those data are used to model the dependency between them based on copula. Note that the copula-based dependency model can be applied to any data. This section gives an example of the application for petrophysical properties and elastic seismic attributes.

3.2.1 Reference data

Given the data of n observations of the random vector (X, Y): (x_i, y_i) where $i = \{1, 2, ..., n\}$ at the well log scale. Where X is an elastic seismic attribute and Y is a petrophysical property. Note that each observation i is associated with a spatial position.

3.2.2 Univariate distribution function modeling

It is necessary to know the univariate probability distribution function of the random variables of the petrophysical property (Y) and the elastic seismic attribute (X) to fully understand the univariate behavior of these variables. Its univariate (marginal) probability distribution function can be inferred using the data for these variables using the empirical cumulative distribution functions (ecdf) of these variables in Equation 3.11. But those functions are step functions and the variables X and Y are continuous, so it is necessary to fit a continuous univariate probability distribution function. There are two approaches to fitting a univariate probability distribution function to empirical functions that are parametric and nonparametric approaches.

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{x_i \leq x\}, \quad G_n(y) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{y_i \leq y\} \tag{3.11}$$

where $\mathbb{I}$ represents an indicator function equal to 1 when its argument is true, and 0 otherwise.

Figure 3.3 show that the ecdfs are discontinuous for the data set of the P-impedance and the effective porosity: $(X, Y) = \{(11000, 0.14), (8000, 0.15), (10000, 0.18), (5000, 0.21), (6000, 0.22)\}$. In Figure 3.3 for the values $y < 0.14$ the value of the cumulative probability distribution function $v = G_n(y) = 0$, for the values of $y \geq 0.14$ or $y < 0.15$ then the value of $v = G_n(y) = 0.2$, thus for all values of and a corresponding value of v can be calculated. But for all the values $v \in [0, 1]$ the value of y cannot be known, for example for $v = 0.3$, it is not known which value of y corresponding. Therefore, as mentioned above, it is necessary to fit a univariate probability distribution function to the data of the random variables.

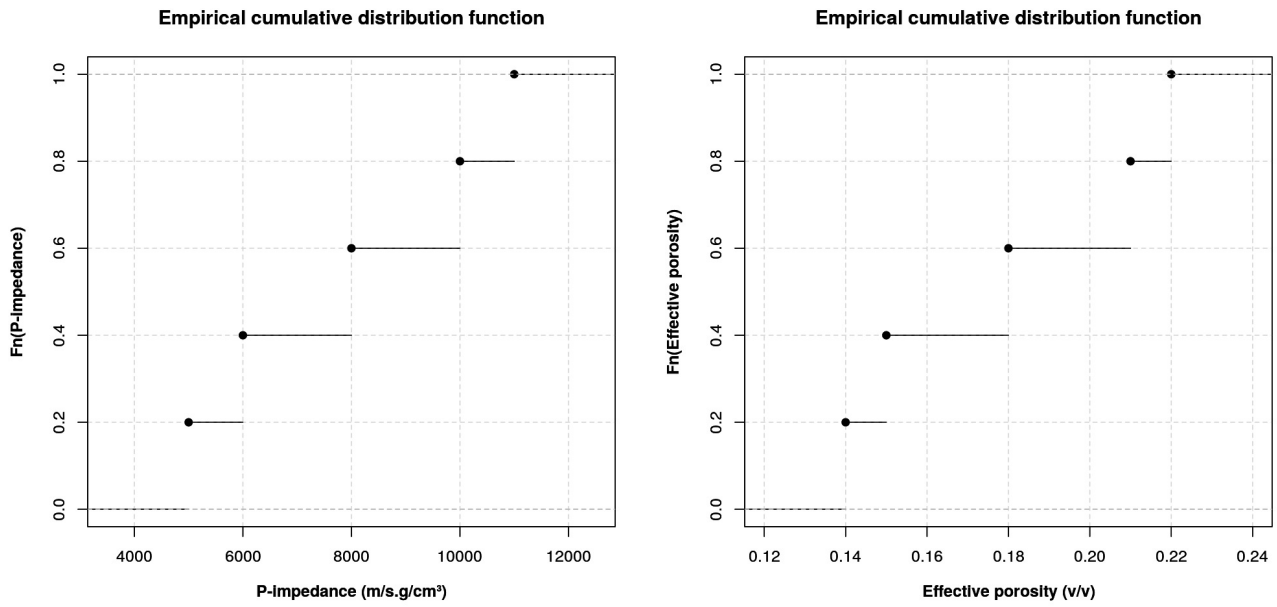


Figure 3.3: Examples of the ecdf of (X, Y).

The parametric approach is not chosen because of its limitation, the model parameters are usually two parameters that are the mean and the variance, which does not fit the data well. Therefore the non-parametric approach is chosen because of its good fit to the data. Within the nonparametric approach, several smoothing methods are available, such as B-spline (Shen et al., 2008), Kernel density (Nagler and Czado, 2016), Bernstein polynomial (Erdely and Díaz-Viera, 2010), etc. In this work, Bernstein polynomial is used because of its analytical tractability.

The univariate quantile function $Q(u)$ of the random variable X:

$$Q(u) = F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad 0 \leq u \leq 1 \tag{3.12}$$

can be approximated using Bernstein polynomials as (Perez and Palacín, 1987):

$$\tilde{Q}(u) = \sum_{k=0}^{n} \frac{1}{2}(x_{(k)} + x_{(k+1)}) \binom{n}{k} u^k (1-u)^{n-k} \tag{3.13}$$

Similarly the univariate quantile function of the random variable Y can be derived:

$$R(v) = G^{-1}(v) = \inf\{y : G(y) \geq u\}, \quad 0 \leq v \leq 1 \tag{3.14}$$

can be approximated using Bernstein polynomials:

$$\tilde{R}(v) = \sum_{j=0}^{n} \frac{1}{2}(y_{(j)} + y_{(j+1)}) \binom{n}{j} v^j (1-v)^{n-j} \tag{3.15}$$

Some important characteristics of the Bernstein polynomial are continuous, invertible and differentiable. Those characteristics are necessary for a probability distribution function. Once the cumulative probability distribution function is continuous, then for any value of the random variable both X and Y, it can always get a corresponding u and v value using the function $F(x)$ and $G(y)$ (Figure 3.4).

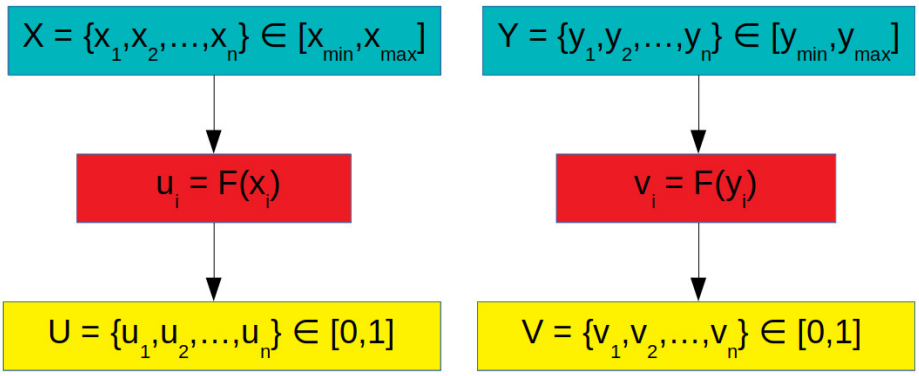


Figure 3.4: Flow diagram to obtain the values of the univariate probability distribution function of the random variables X and Y. In this case it can be an elastic seismic attribute (X) and a petrophysical property (Y).

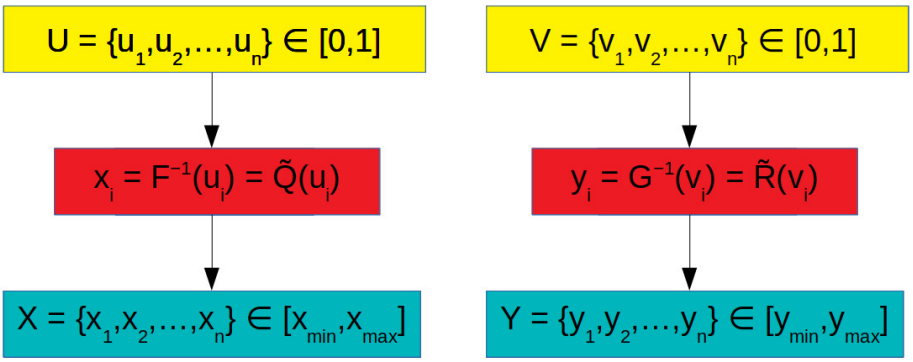


Figure 3.5: Flow diagram to obtain a non-conditional simulation of the random variables X and Y.

In the opposite direction, given any value of u and v, it can always simulate the corresponding x and y values of the random variables using the function $\tilde{Q}(u)$ and $\tilde{R}(v)$ (Figure 3.5).

Figure 3.6 shows illustrative examples of the flow diagrams of Figure 3.4 and Figure 3.5 where for a given value u_i a corresponding value of x_i can be obtained and also for a given value x_j a corresponding value of u_j can be obtained, in a similar way for v and y.

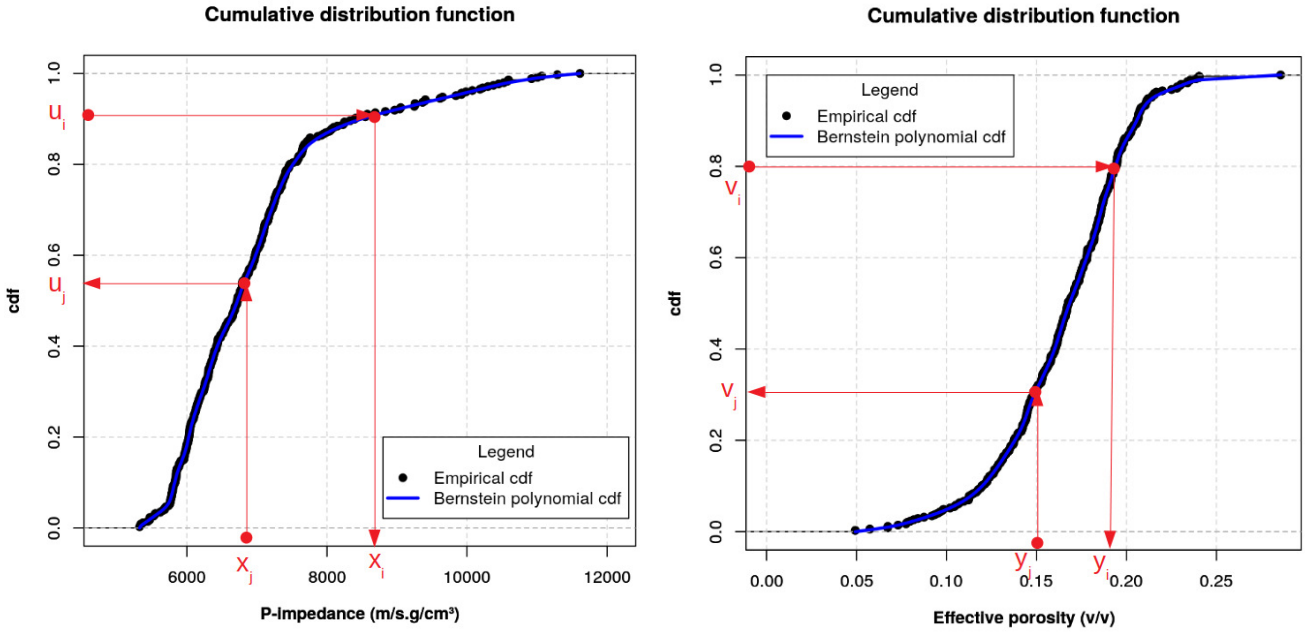


Figure 3.6: Illustrative examples of the flow diagram in Figure 3.4 and Figure 3.5.

3.2.3 Copula function modeling

Similarly, the copula function has similar behavior to the cumulative univariate probability distribution function such as, its ranges are in $[0, 1]$, they are monotone increasing or bi-increasing functions. The difference is that the copula function is two-dimensional and the univariate distribution function is one-dimensional. Therefore, as we have the empirical cumulative probability distribution function, then we also have the empirical bivariate copula function. The empirical copula is a function C_n with domain $\{\frac{i}{n}; i = 0, 1, ..., n\}^2$ defined as:

$$C_n(x, y) = C_n(F(x), G(y)) = C_n(u, v) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(u_i \leq u, v_i \leq v) \tag{3.16}$$

where $\mathbb{I}$ represents an indicator function equal to 1 when its argument is true, and 0 otherwise.

Now we look at an example to better understand the empirical copula function in equation 3.16. Same data $(X, Y) = \{(11000, 0.14), (8000, 0.15), (10000, 0.18), (5000, 0.21), (6000, 0.22)\}$ of P-impedance and effective porosity, and $\{(1.0, 0.2), (0.6, 0.4), (0.8, 0.6), (0.2, 0.8), (0.4, 1.0)$ are the corresponding values (U, V). Now to calculate the empirical copula function:

- $C_n(0.0, 0.0) = \frac{1}{5}\sum_{i=1}^{5}\mathbb{I}(u_i \leq 0.0, v_i \leq 0.0) = \frac{1}{5}(0+0+0+0+0) = \frac{0}{5} = 0$
- $C_n(0.0, 0.2) = \frac{1}{5}\sum_{i=1}^{5}\mathbb{I}(u_i \leq 0.0, v_i \leq 0.2) = \frac{1}{5}(0+0+0+0+0) = \frac{0}{5} = 0$
- $C_n(0.0, 0.4) = \frac{1}{5}\sum_{i=1}^{5}\mathbb{I}(u_i \leq 0.0, v_i \leq 0.4) = \frac{1}{5}(0+0+0+0+0) = \frac{0}{5} = 0$
- $C_n(0.0, 0.6) = \frac{1}{5}\sum_{i=1}^{5}\mathbb{I}(u_i \leq 0.0, v_i \leq 0.6) = \frac{1}{5}(0+0+0+0+0) = \frac{0}{5} = 0$
- $C_n(0.0, 0.8) = \frac{1}{5}\sum_{i=1}^{5}\mathbb{I}(u_i \leq 0.0, v_i \leq 0.8) = \frac{1}{5}(0+0+0+0+0) = \frac{0}{5} = 0$
- $C_n(0.0, 1.0) = \frac{1}{5}\sum_{i=1}^{5}\mathbb{I}(u_i \leq 0.0, v_i \leq 1.0) = \frac{1}{5}(0+0+0+0+0) = \frac{0}{5} = 0$
- $C_n(0.2, 0.0) = \frac{1}{5}\sum_{i=1}^{5}\mathbb{I}(u_i \leq 0.2, v_i \leq 0.0) = \frac{1}{5}(0+0+0+0+0) = \frac{0}{5} = 0$
- $C_n(0.4, 0.0) = \frac{1}{5}\sum_{i=1}^{5}\mathbb{I}(u_i \leq 0.4, v_i \leq 0.0) = \frac{1}{5}(0+0+0+0+0) = \frac{0}{5} = 0$
- $C_n(0.6, 0.0) = \frac{1}{5}\sum_{i=1}^{5}\mathbb{I}(u_i \leq 0.6, v_i \leq 0.0) = \frac{1}{5}(0+0+0+0+0) = \frac{0}{5} = 0$
- $C_n(0.8, 0.0) = \frac{1}{5}\sum_{i=1}^{5}\mathbb{I}(u_i \leq 0.8, v_i \leq 0.0) = \frac{1}{5}(0+0+0+0+0) = \frac{0}{5} = 0$
- $C_n(1.0, 0.0) = \frac{1}{5}\sum_{i=1}^{5}\mathbb{I}(u_i \leq 1.0, v_i \leq 0.0) = \frac{1}{5}(0+0+0+0+0) = \frac{0}{5} = 0$
- ...
- $C_n(1.0, 1.0) = \frac{1}{5}\sum_{i=1}^{5}\mathbb{I}(u_i \leq 1.0, v_i \leq 1.0) = \frac{1}{5}(1+1+1+1+1) = \frac{5}{5} = 1.0$

A matrix of the empirical values of the empirical copula function is obtained and its graphical representation can be seen in Figure 3.7. It can be seen that its value range is always within zero and one.

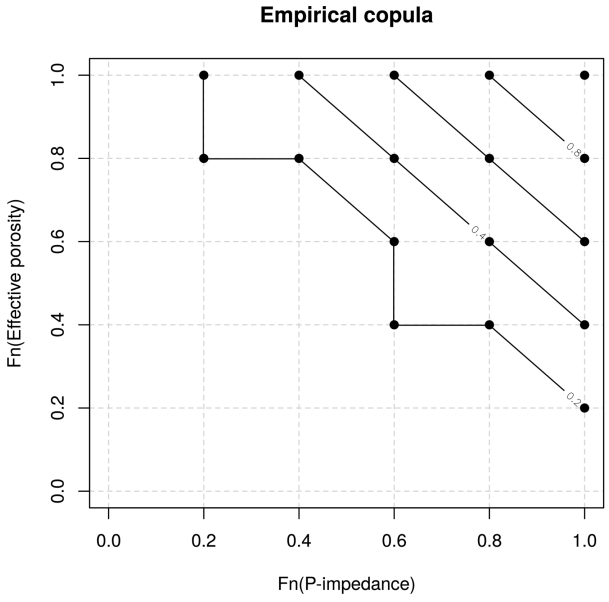


Figure 3.7: The matrix of the empirical copula and its graphic representation.

A univariate variable has two functions that are characterized as the cumulative probability distribution function and the distribution density function, which is the derivative of the cumulative distribution function. Similarly, the copula also has two functions, which are the copula function and the density copula function. Figure 3.7 shows the copula function and the copula density function is calculated as follows:

$$c_{XY}(u,v) = \frac{\partial^2 C_{XY}(u,v)}{\partial u \partial v} \tag{3.17}$$

The histogram is usually like the graphical representation of the univariate distribution density function. Similarly, there are also several ways to represent the copula density function such as, the Pseudo-observation plot, bivariate histogram. Next, it is presented how the Pseudo-observation plot is obtained. Using the same values of (X, Y) and corresponding values of (U, V). The scatterplot is used to plot the values of U vs V (see Figure 3.8).

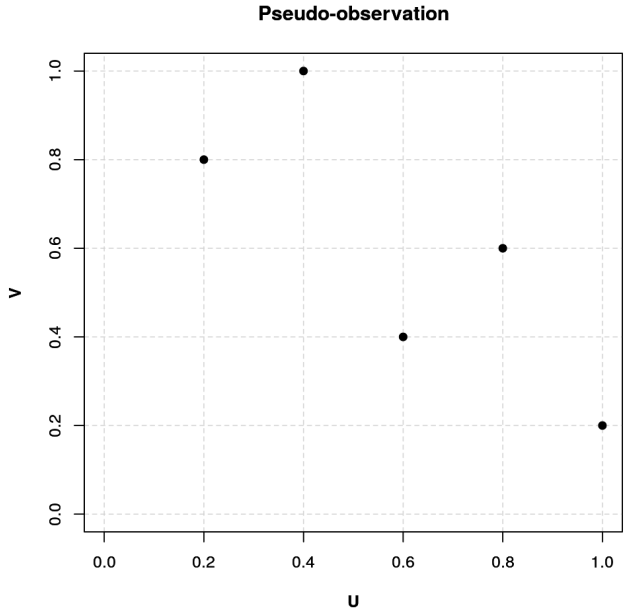


Figure 3.8: Pseudo-observation of the random variables X and Y.

Similarly, the empirical copula function is also a step function like the empirical univariate probability distribution function. But the random variables X and Y are continuous. Therefore, it is necessary to fit a continuous copula function to the empirical copula function. There are two approaches to fitting the continuous copula to the empirical copula which are parametric and nonparametric. The methods for fitting a continuous copula to the empirical copula are presented below.

3.2.3.1 The parametric approach

Parametric copula functions are usually characterized by finite parameter values: $\delta = \{\delta_1, \delta_2, ..., \delta_n\}$. There are many methods to estimate these parameters from the data such as, maximum pseudo-likelihood estimator, inversion of Kendall's tau estimator, and inversion of Spearman's rho estimator. A brief explanation of the most common method, maximum pseudo likelihood estimator, is explained below.

- **Maximum pseudo likelihood estimator**

  According to Joe (2014) the copula parameters can be estimated from the data (X, Y) and corresponding data (U,V) using maximum pseudo likelihood estimator in Equation 3.18 and Equation 3.19.

$$L_{pseudo}(\delta) = \sum_{i=1}^{N} \log c_{XY}(u_i, v_i; \delta) \tag{3.18}$$

$$\delta = argmax \sum_{i=1}^{N} \log c_{XY}(u_i, v_i; \delta) \tag{3.19}$$

3.2.3.2 The nonparametric approach

In literature, there are some methods such as, transformation estimator based on the classical bivariate kernel estimation (eg, Geenens et al. (2014)), a transformation estimator with local log-linear likelihood estimation (Geenens et al., 2014), transformation estimator with log-quadratic local likelihood estimate (Geenens et al., 2014), transformation estimator with local log-linear likelihood estimate and nearest neighbor bandwidths (Geenens et al., 2014) , transformation estimator with local log-quadratic likelihood estimation and nearest neighbor bandwidths (Geenens et al., 2014), tapered transformation estimator with plug-in bandwidths (Wen and Wu, 2020), conic transformation estimator with cross-validation bandwidths of profiles (Wen and Wu, 2020), mirror reflection estimator (Gijbels and Mielniczuk, 1990), beta nucleus estimator (Charpentier et al., 2007), Bernstein copula estimator (Díaz-Viera et al., 2017; Erdely and Díaz-Viera, 2010; Hernández-Maldonado et al., 2012, 2014; Le et al., 2020; Mendoza-Torres et al., 2017; Sancetta, 2007; Sancetta and Satchell, 2004). In the following, only the method based on the Bernstein copula estimator will be explained as an example.

- **Bernstein copula estimator**

  The empirical copula function can be approximated and smoothed by the Bernstein copula function:

$$\tilde{C}_n(u, v) = \sum_{i=0}^{n} \sum_{j=0}^{n} C_n(u, v) \binom{n}{i} u^i (1-u)^{n-i} \binom{n}{j} v^j (1-v)^{n-j} \tag{3.20}$$

for each (u, v) in the unit square $[0, 1]^2$, and where C_n is as defined in (3.16) (Sancetta and Satchell, 2004) and (Sancetta, 2007).

3.2.4 Copula-based conditional simulation algorithm

Once both the univariate probability distribution function and the copula function are modeled. The multivariate joint distribution function can now be fully modeled. Then, simulations and estimates of the variables of interest will be carried out using the following algorithms:

3.2.4.1 Copula-based bivariate conditional simulation algorithm

To simulate realizations Y using the random variables X as secondary variable, according to the measured data $(X, Y) = \{(x_1, y_1), ..., (x_n, y_n)\}$, we follow the algorithm in Nelsen (2006):

1. Generate one independent and continuous Uniform (0,1) random variates t.
2. Calculate $u = F(x)$ where $F(x)$ is the modeled cumulative distribution function of the variable X.
3. Set $v = c_u^{-1}(t)$ where

$$c_u(v) = \frac{\partial \tilde{C}_n(u, v)}{\partial u} \tag{3.21}$$

   and $\tilde{C}_n$ is computed as in Equation 3.20.
4. The couple is $(x, y) = (\tilde{Q}_n(u), \tilde{R}_n(v))$, where $\tilde{Q}_n$ y $\tilde{R}_n$ are the smoothed quantile functions of X, Y, in Equation 3.13 and Equation 3.15.

3.2.4.2 Copula-based multivariate conditional simulation algorithm

This algorithm is quite similar to the bivariate conditional simulation algorithm, only the number of univariate and copula functions are increased.

1. Generate one independent and continuous Uniform (0,1) random variates t.
2. Calculate $u_i = F_i(x_i)$, where F_i are the estimated univariate cumulative distribution functions of the secondary variables X_i, para $i = 1, ..., n$, respectively.
3. Set $v = c_{u^{(i)}}^{-1}(v) = F_{V|U_{1,...,n}}^{-1}(t)$ where

$$c_{u^{(i)}}(v) = F_{V|U_{1,...,n}}(v|u_1, ..., u_n) = \frac{\partial \tilde{C}_{vu_j|u_{-j}}(F(v|u_{-j}), F(u_j|u_{-j}))}{\partial F(u_j|u_{-j})} \tag{3.22}$$

   y $\tilde{C}$ is the bivariate copula function estimated by the random vector $(V, U_j) : j = 1, ..., n$, which can be modeled by the Equation 3.20.

4. The conditional simulated value $y = \tilde{R}_n(v)$, where $\tilde{R}_n$ is the estimated univariate quantile function of the random variable Y in Equation 3.15.

3.2.4.3 Spatial dependency modeling

The conditional simulation algorithms above are only guaranteed to reproduce the univariate distributions of the petrophysical properties and the dependency relationships between the petrophysical properties and the elastic seismic attributes. To reproduce the spatial dependence of petrophysical properties, a global optimization method is used (eg: simulated annealing, differential evolution). More details on these methods can be found in (Deutsch and Journel, 1998) and (Ardia et al., 2020). The results of the copula-based conditional simulations are used for the parameters of the global optimization methods. For each spatial position i, m simulated values of the petrophysical property are obtained. So each position i has a maximum value and a minimum value, or a range of possible values associated with the position i. These values are like the limits or boundary condition associated with each spatial position. Then from this boundary condition and the simulated values, the global optimization methods optimize the objective function, in this case, this function is the variogram function, which represents the spatial dependence of the petrophysical property. This step ensures the spatial dependency structure of the random variable is reproduced.

3.2.5 Copula-based conditional estimation algorithm

3.2.5.1 Copula-based bivariate conditional quantile regression algorithm

. For a value x of the random variable X and $0 < \alpha < 1$ let $y = \varphi_\alpha(x)$ be the solution of

$$P(Y \leq y | X = x) = \alpha. \tag{3.23}$$

Then the graph of $y = \varphi_\alpha(x)$ is the *α-quantile regression curve* of Y conditioned to $X = x$. Nelsen (2006), shows that

$$P(Y \leq y | X = x) = c_u(v)|_{u=F(x),v=G(y)} \tag{3.24}$$

The results of Equation 3.21, Equation 3.23, and Equation 3.24 leads to the algorithm for the *α-quantile regression curve* of Y conditioned to $X = x$:

1. Set $c_u(v) = \alpha$.
2. Solve for $v = g_\alpha(u)$.
3. Substitute u by $\tilde{Q}_n^{-1}(x)$ and v by $\tilde{R}_n^{-1}(y)$.
4. Solve for $y = \varphi_\alpha(x)$.

3.2.5.2 Copula-based multivariate conditional quantile regression algorithm

The algorithm to obtain the quantile regression curve α of the random variable Y using the random variables $X_i = x^{(i)}, i = 1, ..., n$ as secondary variables:

1. Set $c_{u^{(i)}}(v) = \alpha$ where $c_{u^{(i)}}(v)$ is defined in Equation 3.22.
2. Solve for $v = g_\alpha(u^{(i)})$.
3. Substitute $u^{(i)}$ by $(\tilde{Q}_n^{-1})^{(i)}(x^{(i)})$ and v by $\tilde{R}_n^{-1}(y)$ in the above equation.
4. Solve for the regression curve: $y = \varphi_\alpha(x^{(i)})$.

3.2.6 Computational implementation

The implementation of the algorithms was done in the R programming language (RCoreTeam, 2020). The empirical cumulative probability distribution function is calculated using the *ecdf* function implemented in the stats package (RCoreTeam, 2020). Furthermore, this function can be implemented using rank of each observation divided by observation number (n). It can be easily implemented in R software using the following code *rank(variable)/length(variable)*. The univariate probability distribution function estimated by the Bernstein polynomial has been implemented.

The empirical copula function and estimated by the Bernstein copula are implemented. The copula (Hofert et al., 2020), vine copula (Nagler et al., 2020), rvinecopulib (Nagler and Vatter, 2020) packages are mainly used for modeling parametric copulas. The kdecopula (Nagler and Wen, 2018), kdevine (Nagler, 2018), rvinecopulib (Nagler and Vatter, 2020) packages are principally used for modeling non-parametric copulas.

The conditional simulation algorithm based on the Bernstein copula is implemented. The CDVineCopulaConditional (Bevacqua, 2017) package is used for the multivariate conditional simulation based on the parametric vine copula. The vinereg (Nagler and Kraus, 2020) package is used for the quantile regression on both the parametric and non-parametric D-vine copulas.

Additionally, the GSLIB Geostatistical Software (Deutsch and Journel, 1998), the DEoptim (Ardia et al., 2020) and GenSA (Xiang et al., 2018) packages in R are used, which are the global optimization methods to reproduce the spatial variability of the random variables.

3.3 The simulation approach methods

3.3.1 Bernstein Copula-based Spatial CoSimulation (BCSCS)

This method basically consists of first modeling the bivariate probability distribution and second optimizing the spatial dependency model (variogram function). To carry out the first step, the univariate probability distribution functions and the copula function are modeled using Bernstein polynomials in Equation 3.15 and Equation 3.20. The algorithm for this method is found in subsubsection 3.2.4.1. To perform the second step, a global optimization method is used, which

is simulated annealing (Deutsch and Journel, 1998). Note that this method is applied for the bivariate case.

3.3.2 Vine Bernstein Copula-based Spatial CoSimulation (VBCSCS)

This method follows the same steps as the BCSCS method, but this can be applied for n variables ($n \geq 3$). The difference is that instead of modeling just one bivariate copula function, many bivariate copula functions are modeled. Univariate probability functions and bivariate copula functions are modeled by Bernstein polynomials in Equation 3.15 and Equation 3.20. The algorithm for this method is found in subsubsection 3.2.4.2.

3.3.3 Vine Parametric Copula-based Spatial CoSimulation (VPCSCS)

The steps in this method are similar to the VBCSCS method. The only difference is that instead of using the Bernstein polynomial to model the bivariate copula functions, the bivariate parametric copulas are used. The algorithm for this method is found in subsubsection 3.2.4.2.

3.4 The estimation approach methods

3.4.1 Bernstein Copula-based Quantile Regression (BCQR)

This method basically consists of modeling the bivariate probability distribution function by constructing the univariate functions and the bivariate copula function using Bernstein polynomials in Equation 3.15 and Equation 3.20. The difference with the BCSCS method is that it does not perform the optimization step of the variogram function. The algorithm can be found in subsubsection 3.2.5.1. Note that this method is applied for the bivariate case.

3.4.2 Vine Bernstein Copula-based Quantile Regression (VBCQR)

This method is very similar to the BCQR method, but this can be applied for n variables ($n \geq 3$). The difference with the BCQR method is that instead of modeling just one bivariate copula function, many bivariate copula functions are modeled. Univariate probability functions and bivariate copula functions are modeled by Bernstein polynomials in Equation 3.15 and Equation 3.20. The algorithm for this method is found in subsubsection 3.2.5.2.

3.4.3 Vine Parametric Copula-based Quantile Regression (VPCQR)

The steps in this method are similar to the VBCQR method. The only difference is that instead of using the Bernstein polynomial to model the bivariate copula functions, the bivariate parametric copulas are used. The algorithm for this method is found in subsubsection 3.2.5.2.

Chapter 4

Methodology

The static reservoirs characterization or it is also said reservoir geological-petrophysical modeling is based on the use of geostatistical techniques, which allow obtaining numerical models that represent both the geological model and the petrophysical model of a reservoir (Figure 4.1). Geostatistical techniques facilitate combining different sources of both quantitative and descriptive information in a integral way and analyzing the influence of the various parameters that intervene in the model. The objective when applying geostatistical techniques is to obtain a reliable numerical model of the reservoirs, in addition to the fact that, since they are probabilistic techniques, it is feasible to perform an analysis of the degree of uncertainty associated with the models.

In the process of characterizing a reservoir, it is assumed that known data is never sufficient or exhaustive to describe a reservoir in a complete and exact way. Under this approach, geostatistical methods represent a practical tool through which it is possible to make distributions of facies and petrophysical properties consistent with known data and with geological knowledge of the processes that formed the reservoirs.

It is important to note that the reservoir geological-petrophysical modeling starts from a conceptual geological model, which includes the environment in which the sediment deposition was carried out and a stratigraphic framework product of the sedimentation and the subsequent formation of the rocks that constitute the deposit. The work of formulating a geological model involves disciplines such as: geology, sequential stratigraphy, sedimentology, interpretation of well logs, biostratigraphy, geological studies of analogous outcrops and seismic interpretation. The depositional sedimentological model of the reservoir should provide a semi-quantitative evaluation of the geometric parameters in terms of the extension and distribution of the facies, as well as the relationship between them. This information is basic to the stochastic reservoir modeling process.

Regarding the structural model, the structural characteristics of the reservoir must be identified and defined, such as regional and local faults and fractures. This model is based mainly on the interpretation of seismic information, well log data, the study of cores extracted from wells, and field geology work on reservoir analogs that can provide useful information or corroborate information that is interpreted from seismic.

The stratigraphic model is in charge of defining the framework or internal structure of the reservoir, the process considers the identification and definition of the surfaces that delimit the main flow units of the reservoir from the correlation of geological units between the wells. Another important part is the definition and construction of the stratigraphic mesh, whose main objective is to define the internal stratigraphic geometry, in a vertical sense, which must represent the architecture of the units that compose the reservoir.

The facies model in the reservoir represents a powerful tool to guide the distribution of petrophysical properties of the reservoir, since in most cases these two aspects are closely related. Thus, the facies distribution model implies the definition of facies, which is based on the conceptual geological model and on the structural and stratigraphic model of the reservoir (Díaz-Viera et al., 2013).

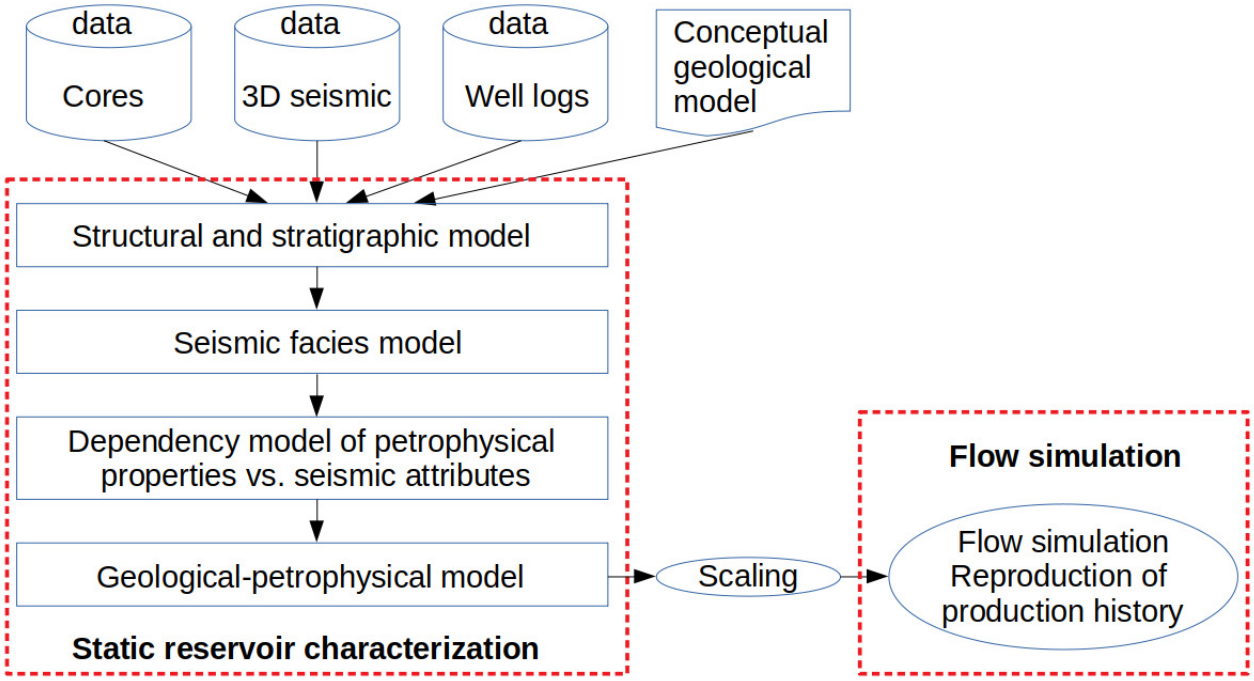


Figure 4.1: General workflow diagram for reservoir geological-petrophysical modeling (modified from Díaz-Viera et al. (2013)).

Once the facies model has been obtained, the dependency relationship between the petrophysical properties and the elastic seismic attributes is modeled for each facies using the copula function. Finally, a representative model of the reservoir is obtained that represents both geological and petrophysical models.

From the inversion point of view, Figure 4.2 shows the general scheme of the petrophysical seismic inversion. The scheme begins from a forward model where given the petrophysical properties such as porosities (e.g., effective porosity), fluid saturations (e.g., water saturation), mineral

fractions (e.g., clay volume), and between other assumptions such as the textures of rock, which is the rock construction mode, describes the relationships between the components (e.g., the coordination number), the environmental conditions such as temperature, pressure, etc. Then the facies are defined, which are characterized by a combination of a lithology and a fluid (e.g, brine sandstone). For each facies, a physical rock model is established and from this model, the elastic seismic attributes can be obtained, which are mainly P-wave velocity, S-wave velocity and density.

From the elastic seismic attributes, one can calculate the impedances that are the density multiplied by the velocities. Then, the reflection coefficient is calculated and multiplied to a wavelet, this process is called the convolution. Finally, a synthetic seismic trace is obtained that is characterized by amplitude, frequency, and phase. This process is called the forward model. Now, the inversion model is that from the actual seismic trace obtained from the seismic acquisition, then it is required to obtain the elastic seismic attributes by deconvolution which can be described by the inverse operation to a convolution. This process is called seismic inversion.

A petrophysical inversion is the process of obtaining the petrophysical properties given the elastic seismic attributes. To carry out this inversion, geostatistical methods are usually used, which model the dependency relationships between the petrophysical properties and the elastic seismic attributes in the places where both are known (e.g., well logs data). Then making use of the dependency model to predict petrophysical properties in places where only elastic seismic attributes are known (e.g., seismic data).

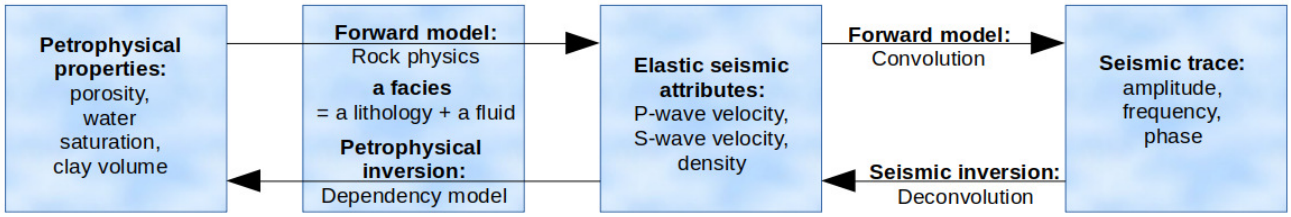


Figure 4.2: Petrophysical seismic inversion.

Figure 4.1 shows the general workflow for the reservoir geological-petrophysical modeling. Below is the specific workflow of copula-based modeling for petrophysical property prediction using the elastic seismic attributes as secondary variables applying for a given facies. The specific workflow consists of 6 steps in Figure 4.3:

1. Exploratory data analysis
2. Spatial correlation analysis
3. Copula-based dependency modeling
4. Validation
5. Application
6. Uncertainty quantification

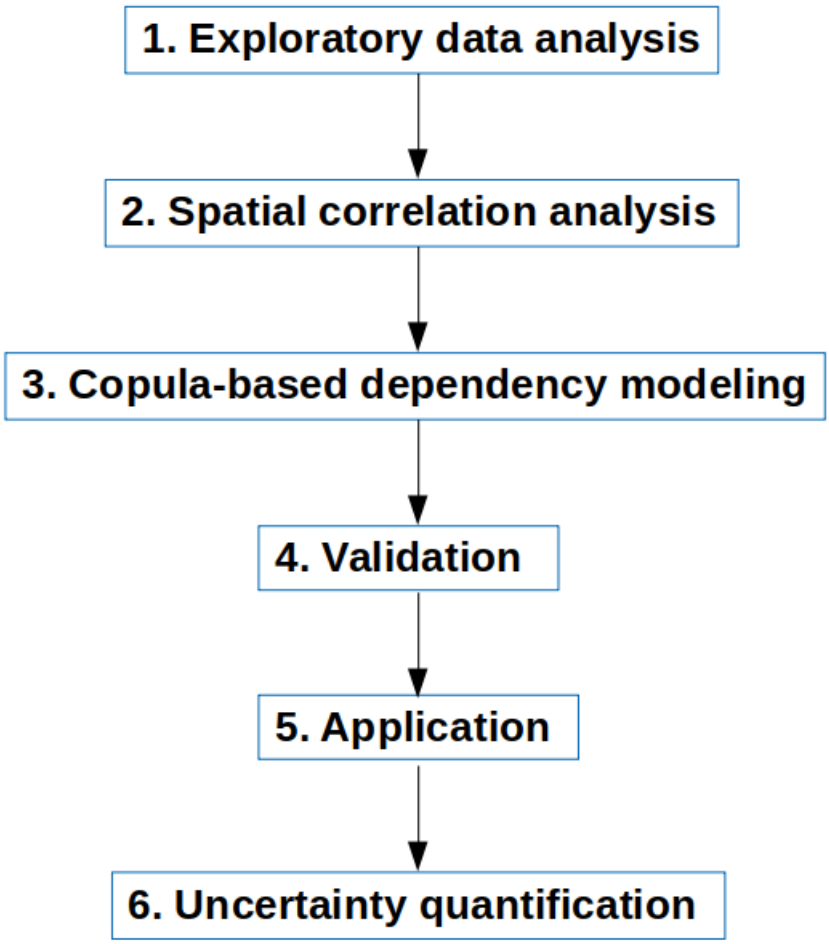


Figure 4.3: Copula-based modeling workflow for petrophysical property prediction using seismic attributes as secondary variables.

In the following, each steps will be explained in detail.

4.1 Exploratory data analysis

It is a set of statistical and graphical techniques that allow to establish a good basic understanding of the behavior of the data and of the dependency relationships existing between the variables that are studied. Exploratory data analysis (EDA) is a preliminary and indispensable step for the successful application of any statistical method. In particular, it allows the detection of flaws in the design and data collection, the treatment and/or evaluation of missing data, the identification of

outliers and the verification of the assumptions required by geostatistical techniques (Díaz-Viera, 2002).

4.1.1 Univariate analysis

The statistics of each random variable (petrophysical properties and elastic seimic attributes) such as the minimum, the maximum, the rank, the mean, the median, the first quartile, the third quartile, the interquartile rank, the variance, the standard deviation, the skewness, the kurtosis and the atypical distributional values (outliers) are described using sets of tables and figures combined to understand the univariate behavior of random variables. Note that outliers and skewness strongly affect average-type statistics such as mean, variance, and semivariogram.

A univariate probability distribution function is fitted to the random variable data using non-parametric approach (Bernstein polynomial) in subsection 3.2.2. Once this function is obtained, then, the univariate behavior of the random variable is fully known. Therefore, the realizations or simulations of the random variable can be easily performed through generating values of the uniform distribution function and the quantile function of the random variable (see Figure 3.5). If the random variable does not have the hard data but a conceptual prior distribution model for that variable is known. So, based on that conceptual model, the realizations of the random variables can also be done. This shows the versatility of this analysis that both conceptual and hard data sources can be used.

4.1.2 Dependency analysis

In earth sciences, the scatterplot of random variables (petrophysical properties and elastic seimic attributes) on their original scale and their Pearson's linear correlation coefficient are often used to analyze the dependency relationship. The univariate probability distribution of the original random variables is commonly transformed into Gaussian distributions to facilitate the dependency analysis between them. But Sklar's theorem showed that the dependency relationship between the random variables is independent of their univariate probabilistic behaviors and the copula function that contains the dependency information between the variables. Therefore, the scatterplot is not a suitable means for the analysis of dependency relationships because it is mixing the univariate probabilistic behavior of the variables and their dependencies. The correct way to analyze the dependency relationship is through the copula function and its representative graph is the Pseudo-observation. The Pseudo-observation is the cross diagram of the univariate empirical distribution functions of the variables (see more detail in subsection 3.2.3).

Pearson's linear correlation coefficient is not as robust as the rank correlation coefficient like Spearman and Kendall's because it is only used to measure the linear dependence relationship between variables. If the variables do not present linear dependence, then it is not useful. Therefore, in this work, in addition to Pearson, Spearman and Kendall are used. The definitions of these coefficients as a function of the copula can be seen in Appendix A.

4.2 Spatial correlation analysis

In this section, the spatial correlation of the random variables (petrophysical properties and elastic seismic attributes) is analyzed using the semivariogram function. The experimental semivariogram is first calculated and then a semivariogram model is fitted. To carry out this process, a stationarity analysis is needed (more detail in Díaz-Viera (2002)). If perhaps there is no data to estimate a semivariogram model but a conceptual geological model exists, then a semivariogram model can also be inferred according to the conceptual geological model. Note that the semivariogram estimator used in this work is an estimator of the average type. Therefore, outlier values and skewness can affect the semivariogram model.

4.3 Copula-based dependency modeling

Once the Pseudo-observation and empirical copula are obtained, then, a copula function is modeled, which can be parametric or non-parametric (see more detail in in subsection 3.2.3). Once the copula function is modeled, simulations of dependency relationships between petrophysical properties and elastic seismic attributes can easily be performed. Petrophysical properties can be simulated and estimated by copula-based models using seismic attributes as secondary variables by algorithms in subsection 3.2.4.

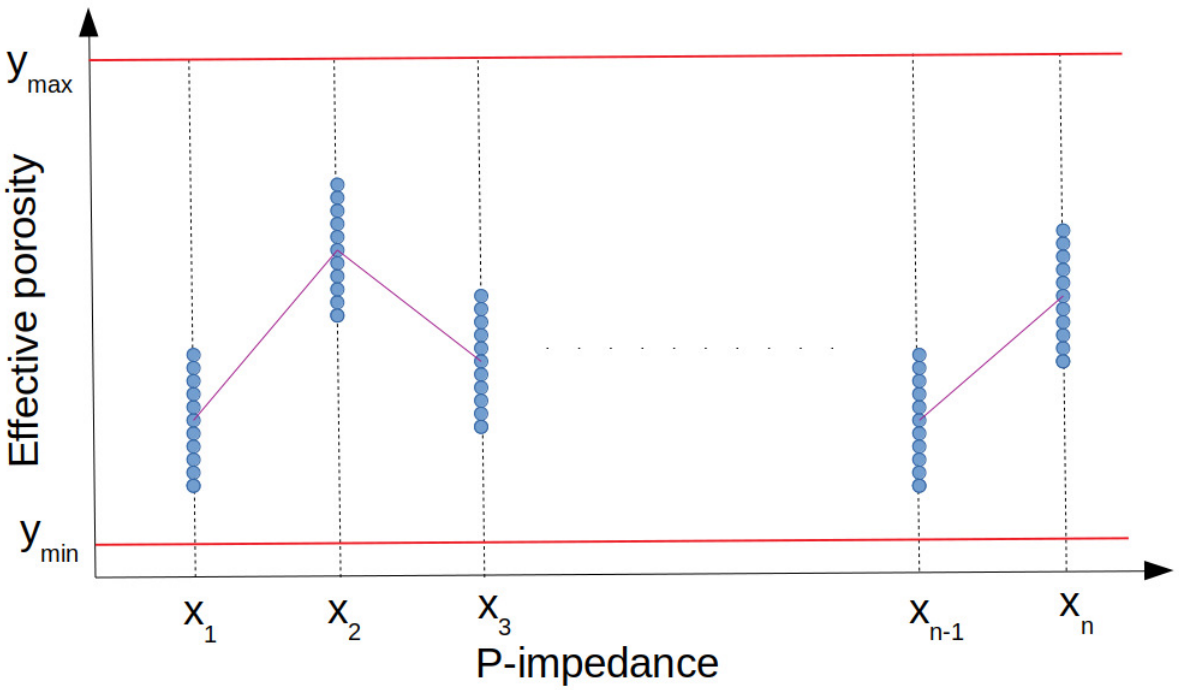


Figure 4.4: Schematic representation of copula-based conditional simulations.

Figure 4.4 shows a schematic representation of copula-based conditional simulations for effective porosity prediction using P-impedance as a secondary variable. Each P-impedance value x_i is associated with a spatial position i. Therefore, for each spatial position i has one P-impedance value and k effective porosity values (blue color). At each spatial position i, the quantiles are calculated, then a quantile regression is obtained, e.g., the median regression in purple color in Figure 4.4.

A simulation of the petrophysical properties is a realization of these variables when the univariate and multivariate behaviors are reproduced, and also the spatial correlation of those variables. With the algorithms in subsection 3.2.4, univariate and multivariate behaviors can be reproduced, but spatial variability is not reproduced. Therefore, to reproduce the spatial correlation of the petrophysical properties, the global optimization method such as simulated annealing or differential evolution is applied.

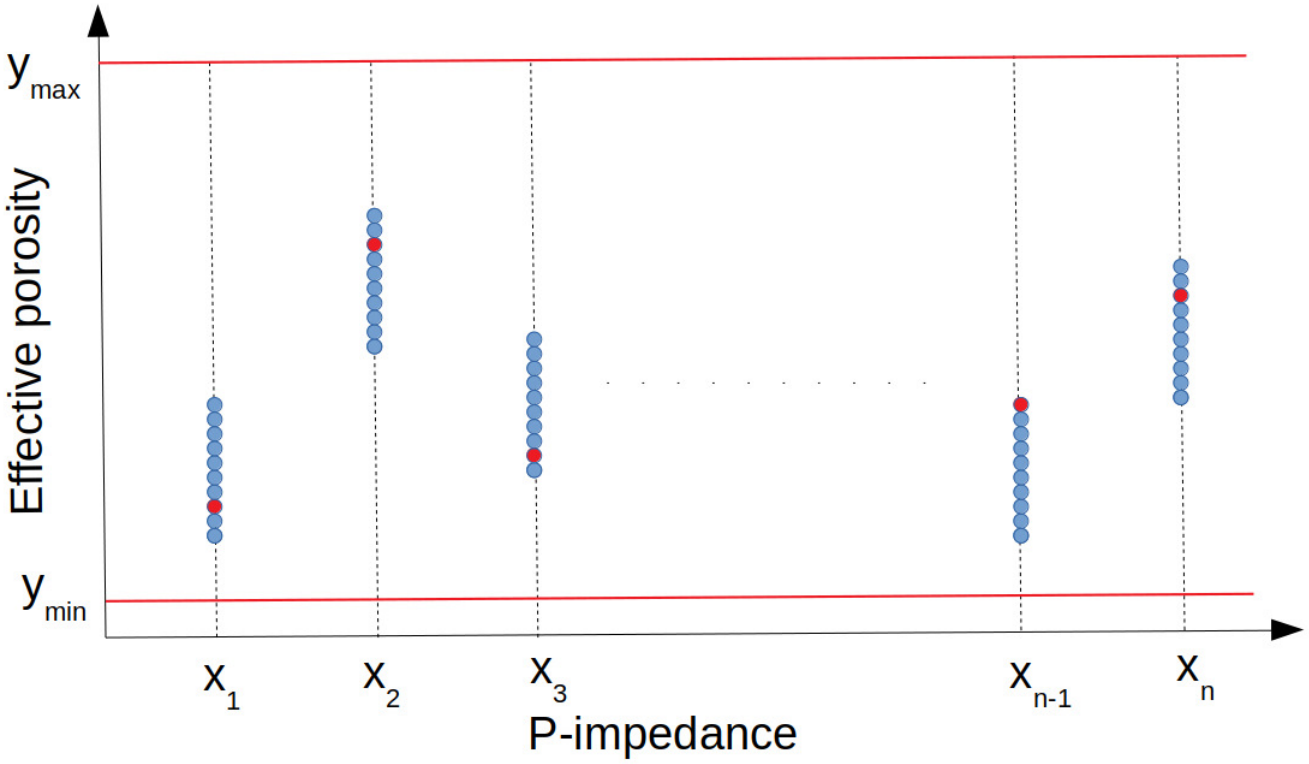


Figure 4.5: Schematic representation of a simulation (red color) after applying copula-based model and global optimization method.

Figure 4.5 shows a schematic representation of an effective porosity simulation (red color) after applying the copula-based model and the global optimization method. What the global optimization method does, is to select the values within n simulated values of the effective porosity

at all spatial positions until it reaches a spatial correlation model (semivariogram model).

4.4 Validation

This step is to validate the copula-based dependency relationship model by comparing the simulated petrophysical properties using the seismic attributes as secondary variables at the well-log scale. Furthermore, the uncertainty analysis will be performed on the results of this method with a traditional simulation method (SGCS).

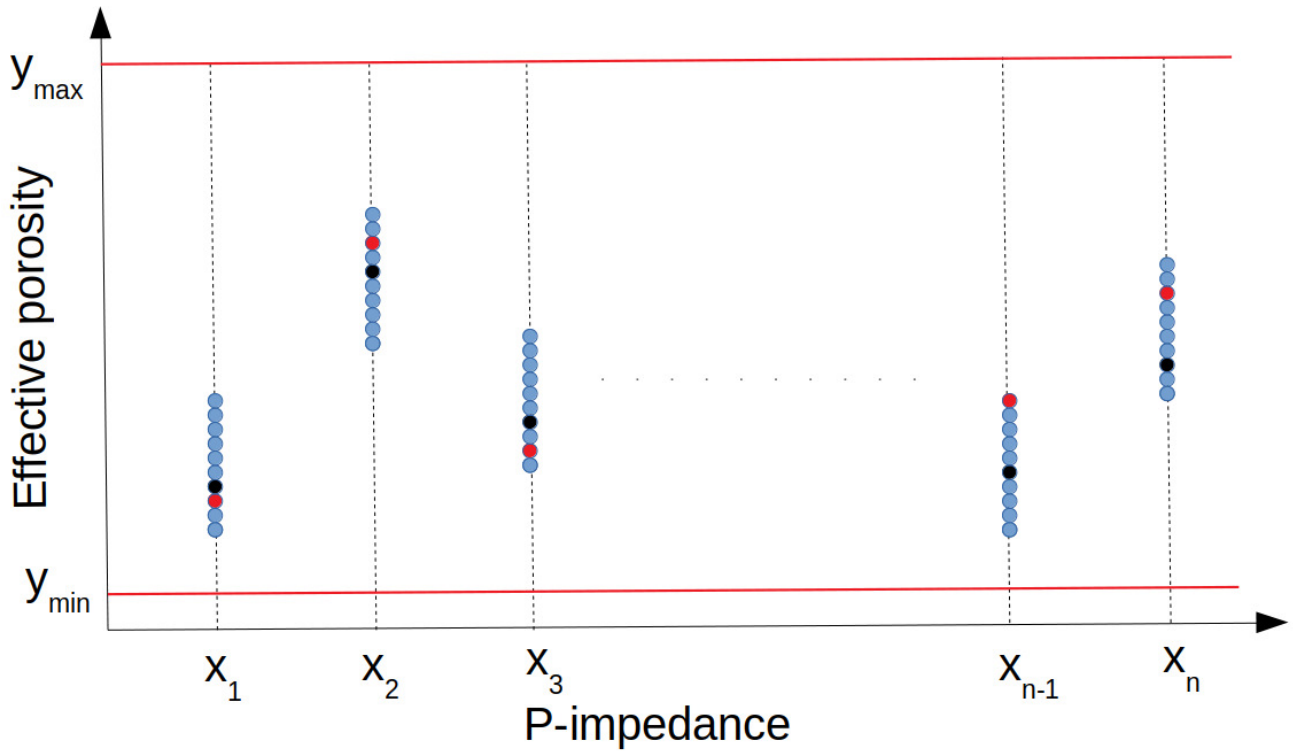


Figure 4.6: Schematic representation of k simulations (red color) after applying copula-based model and global optimization method, and reference data (black color).

Figure 4.6 shows a schematic representation of a simulation (red color) after applying the copula-based model and the global optimization method and comparing it with the reference data (black color). With the simulation results, the errors or differences between the simulated values and the original values can be analyzed. But with a simulation, you cannot perform an uncertainty analysis or the accuracy and precision of the model results relative to the reference data. Therefore, k simulations are needed to achieve this goal.

4.5 Application

Once the dependency relationship model between the petrophysical properties and the elastic seismic attributes at well log scale is validated, this model will be applied at the seismic scale with the assumption that this model is the same in both scales of the well log and seismic. The spatial dependence models of the petrophysical properties on the seismic scale can be inferred from the spatial dependence models of the elastic seismic attributes at the seismic scale.

4.6 Uncertainty quantification

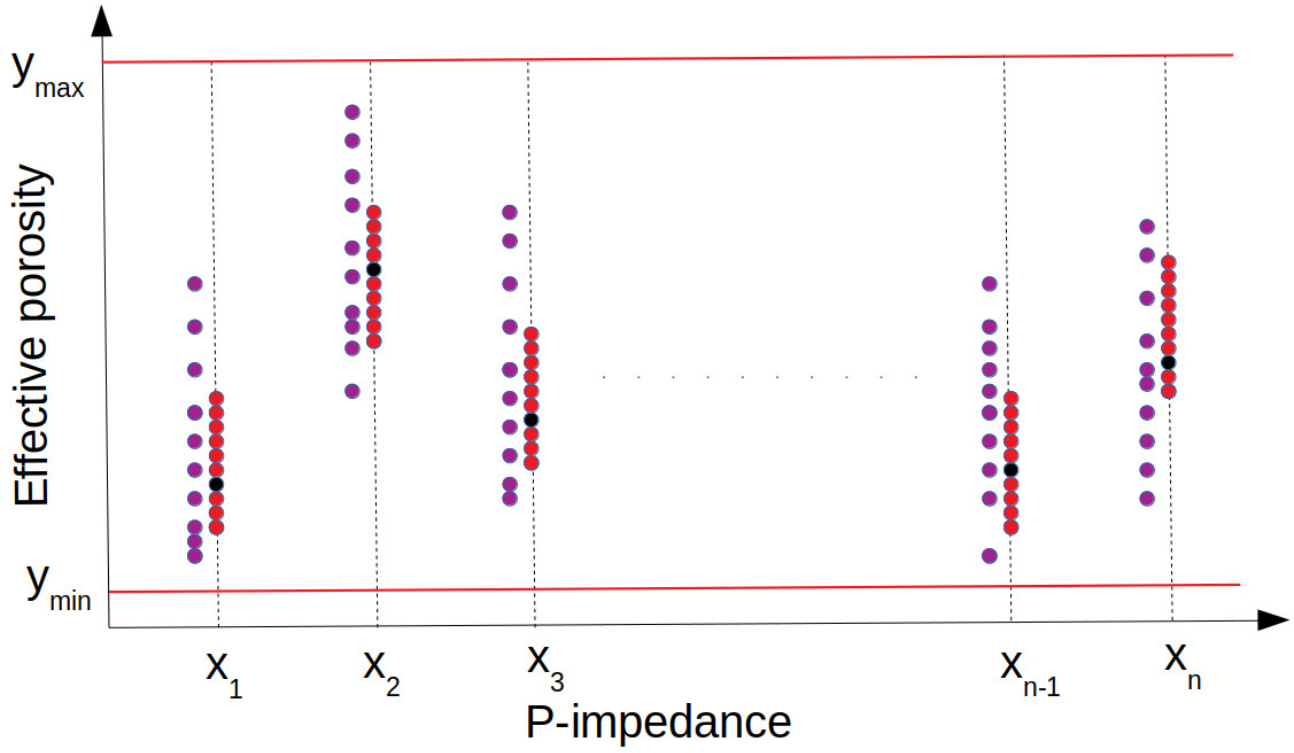


Figure 4.7: Schematic representation of k simulations (red color) after applying copula-based model and global optimization method, k simulations (purple color) of traditional simulation method (SGCS), and reference data (black color).

Given a value of the seismic attribute $X = x$, a range of possible values of the petrophysical property $Y = y \pm \Delta y$ at the spatial point of interest will be predicted. The k simulations of the primary variable Y are obtained and then validated with the reference values of Y. The simulations Y^* are conditioned to the secondary variable X applying two cosimulation methods:

traditional SGCS and proposed BCSCS. The uncertainty ranges were compared to the reference Y values. In addition, the descriptive statistics, the probability distribution functions and the spatial distribution functions of the sets of k simulations Y^* were compared to the reference solution.

Figure 4.7 shows the schematic representation showing the simulations of the copula-based method (red color), the SGCS method simulations (purple color), and the reference data (black color) at well log scale. This figure and a descriptive statistical analysis of the simulations of both methods allow a quantification analysis of the uncertainty such as the accuracy and precision of the simulations with respect to the reference data. Figure 4.7 shows that the simulations in red present better accuracy and precision than those simulated in purple.

Chapter 5

Conceptual geological model

The conceptual geological model in the study area is presented, showing the geographic location, regional geological context, stratigraphic and structural framework, petroleum system, facies distribution.The geological description in this section is mainly taken from CNH (2005). Then, the application of the methodology for well logs and seismic data.

5.1 Geographic location

The study area is the Lakach field reservations that geologically is located in the extreme south of the Cordilleras Mexicanas Province (see in Figure 5.1). The available wells are four wells: Lakach 1, 2, 11, and 2DL that are aligned in the NNW-SSE direction with the approximate distance of 5.7 km. The 3D seismic data are from the Holok Alvarado seismic cube that are located within the geological provinces: Mexican Ridges (Cordilleras Mexicanas), catemaco fold belt (Cinturón Plegado de Catemaco), and isthmus saline basin (cuenca salina istmo).

5.2 Regional geological context

Mexican Ridges: it is a contraction system of elongated folds of NNW-SSE direction, which stretches 500 km and covers nearly $70,000$ km^2 in water depths between $1,000$ and $3,000$ m. It is characterized by long and narrow asymmetric anticlines, usually with vergence to the East. The dual system extension - compression had as main detachment takeoff plane of argillaceous horizons of the upper Eocene and some other secondary detachment takeoff planes the Tertiary sequence. It has been reported many emanations of oil and gas in the seafloor, showing the functioning of an active petroleum system. These emanations are mainly concentrated in the area of the Perdido Fold Belt, Subsalt Belt and the minibasins sector, also in relation to the salt diapirism, and the distal face of the compression system front with a detachment level in the Eocene clays from Mexican Ridges (CNH, 2005).

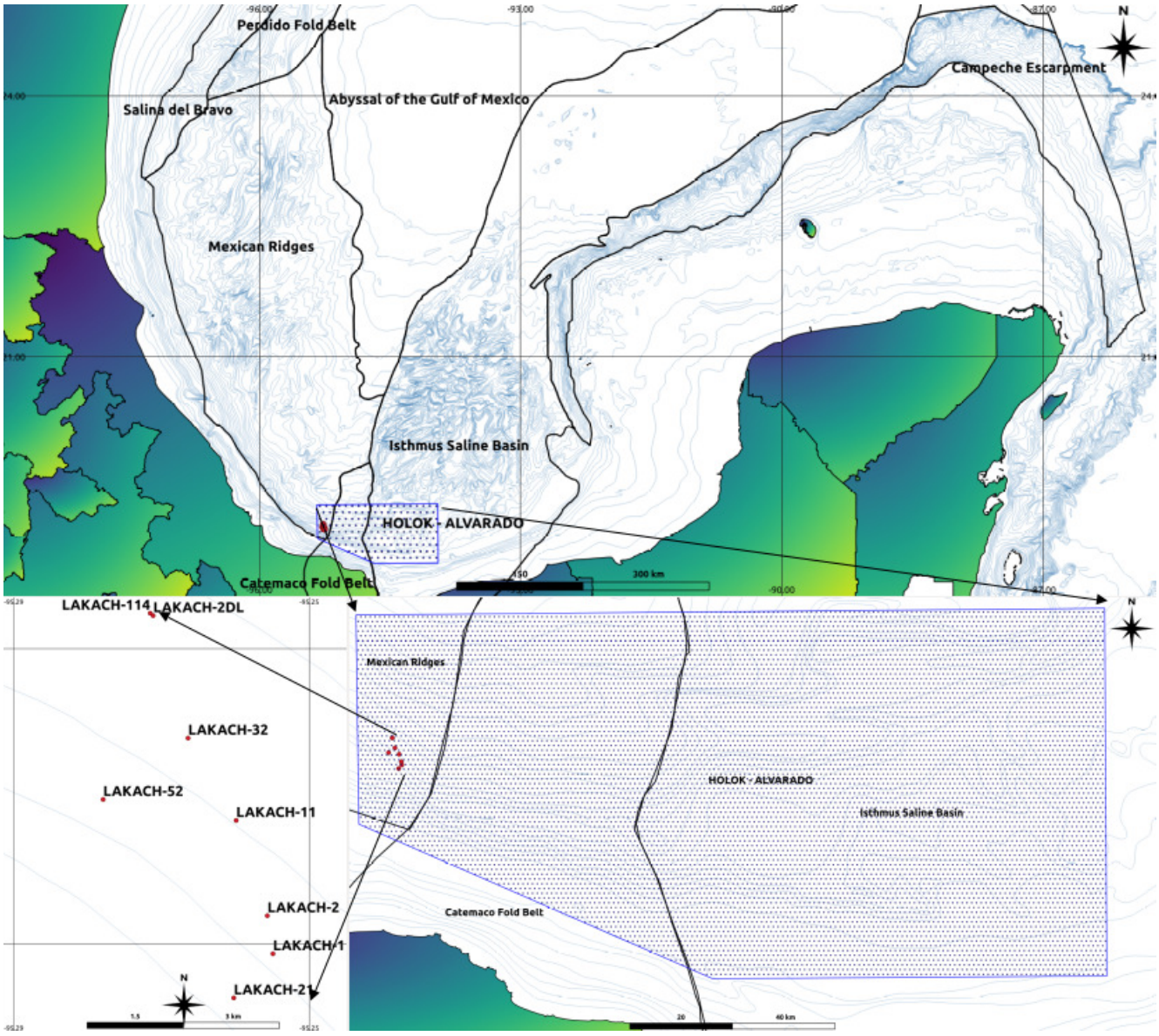


Figure 5.1: Location map of the work area: Holok Alvarado seismic cube, Lakack wells. Map processed from data from the National Hydrocarbons Information Center (CNIH) of Mexico.

5.3 Stratigraphic framework

The stratigraphic characteristics are described below according to geological age:

- **Miocene-Pliocene**: Sedimentation in bathyal and neritic environments, facies mainly clayley and silty related to turbiditic environments, sandstones packages associated to submarine

fans and channels.

- **Oligocene**: Corresponds to fine-grained sands interbedded with clays, deposited in meandering channels and distal lobes, contemporary to the first episodes of deformation caused by the allochthonous salt movements. Therefore, at this epoch the structural configuration modifies due to the plastic deformation of the salt bodies.

- **Middle-Upper Eocene**: Sedimentation was carried out in a lower bathyal environment, forming a seal of good thickness in the Middle Eocene, composed of clays with thin interbedded siltstones in hemipelagic environments. There are also thin layers of sandstones facies related to lobes and crevasse splays from channels in the Upper Eocene.

- **Lower Paleocene-Eocene**: The Paleocene corresponds to a period of low sea level in the Gulf of Mexico. Lower Paleocene is composed of basal sandstones bodies with a long lateral extension called the "Whooper" sand (equivalent to Lower Wilcox Formation) in channel facies and turbiditic layers, evolving to a sandy submarine fan system with lobes and amalgamated channels to the Lower Eocene, with a high clay content corresponding to a high sea level period. There are silt-clay intervals deposited in hemipelagic basin facies interbedded with sandy Paleocene intervals (Midway Formation and "Big Shale" member).

- **Cretaceous**: Carbonate facies identified in deep basin environments with high cyclical fluctuations of sea level. To the middle part, a of secondary source rock level is inferred of Turonian age, with high content of organic matter deposited in anoxic basin environments. In the Upper Cretaceous are identified calcareous-sandy bodies locally distributed, related to turbiditic lobes.

- **Upper Jurassic**: Mainly carbonate sediments from inner to external ramp facies, with lateral variations to dolomites and terrigenous sediments deposited in middle ramp environments. Toward the top, a maximum transgression level represented by basin carbonates rich in organic matter, particularly in the Tithonian, shows features of anoxic environments deposition.

- **Middle Jurassic**: Formed by red beds overlaying igneous-metamorphic basement, evolving to evaporitic rocks of great thickness related to the Gulf of Mexico opening.

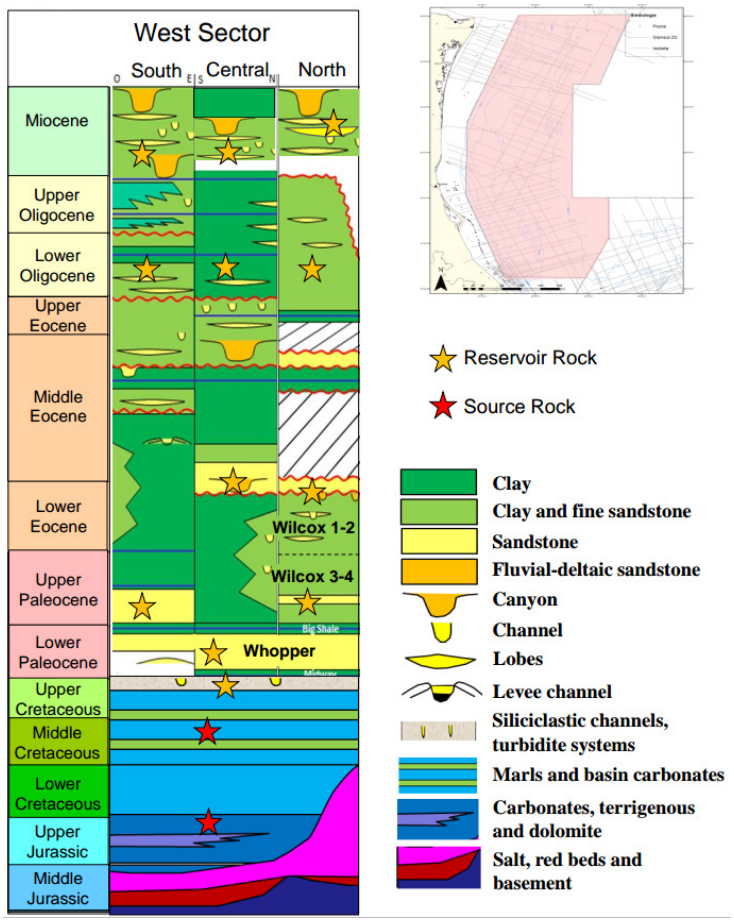


Figure 5.2: Stratigraphic column of geological provinces: Mexican Ridges, catemaco fold belt, and isthmus saline basin (CNH, 2005).

5.4 Structural framework

Mexican Ridges located East from the continental shelf of the Gulf of Mexico, offshore Veracruz and Tamaulipas states. In this area, a large and long folds were formed during the Neogene time from the southern part of Salina de Bravo Province to the Southern Gulf of Mexico.

It covers over 500 km long in a surface area nearly of 70,000 km^2, in water depths between 1,000 and 3,000 m, Mexican Ridges were generated in response to the gravitational extensional processes developed from the south of Burgos and Tampico - Misantla Basins. It is composed by long and tight eastward oriented symmetrical anticlines. The extensional-compressional combined deformation system of this Province is extended through more than one detachment level within the Paleogene interval, as well as other secondary detachment levels within the Tertiary sequence. The structures are mainly present in the Tertiary sedimentary sequence, particularly deforming

the Miocene to Recent rocks. The younger and broader folds are located towards the central part of the Mexican Ridges.

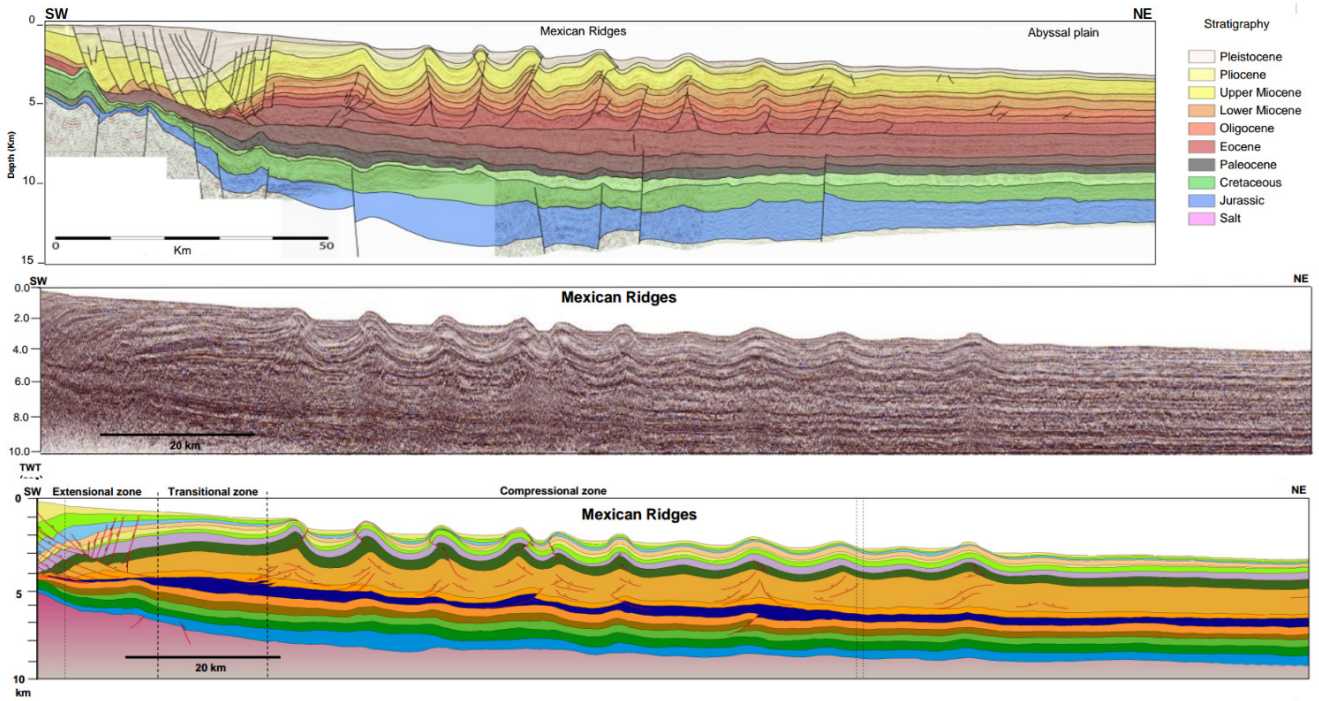


Figure 5.3: Structural section in the SW-NE direction of the Mexican ridges. The Mexican ridges contractional domain is characterized fault-related anticlines formed in the Neogene time. Showing different structural styles: dipping folds, fault propagation and detachment folds (CNH, 2005).

5.5 Petroleum systems

Figure 5.4 shows the oil system for the southern zone of the Mexican ridges, where indicate that the regional tectonics are faulting, rifting and passive margin; and traps formation are folding and wedging.

In general, the migration controlling factors for the model of hydrocarbons saturation in the area are:

- The lateral continuity of permeable facies of Paleocene / Upper Cretaceous age.
- The role of fault migration routes: the contribution of several compartments vs local migration compartment.
- The efficiency of the seals based on fracturing and thickness.
- The mobility of hydrocarbons according to the hydrocarbon phase (maturation or degradation grade).

- Tilt angle and surface of drainage areas.

Tertiary gas is charged mostly by light hydrocarbons, condensed or gas with an RGA $>$ 300 m^3/m^3 and an API grade $>$ 30. For reasons of synchronization, this load occurs most likely in the areas where the Tithonian is in condensed or dry gas generating window to present day. Similarly, the structures charge will be more efficient in a larger drainage area.

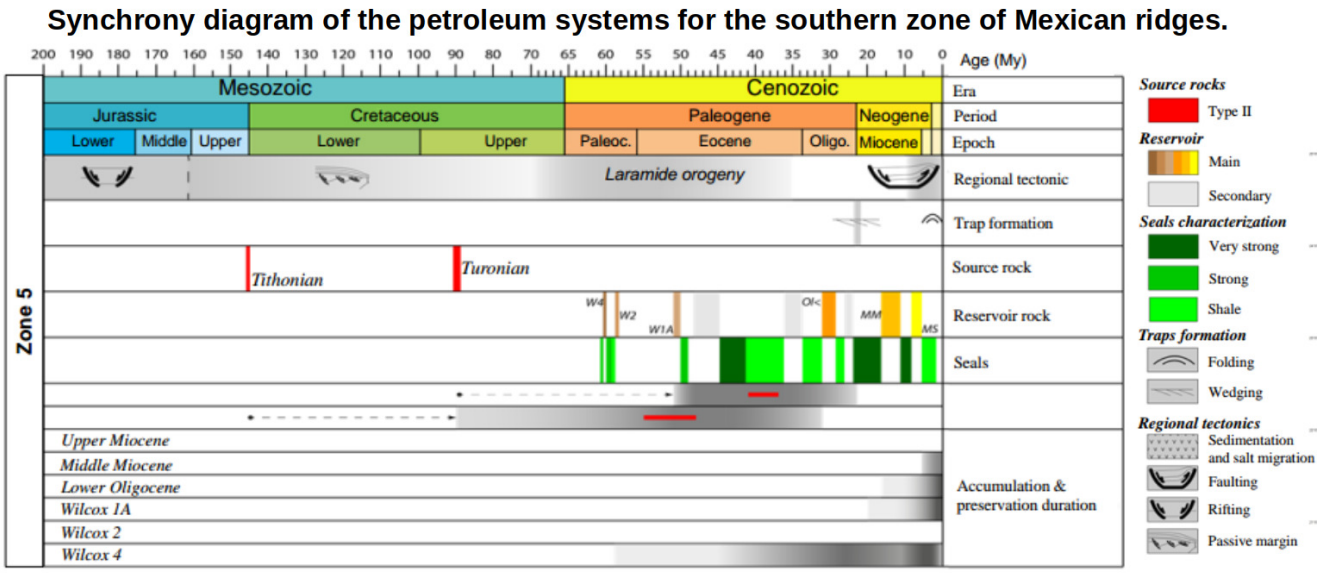


Figure 5.4: The petroleum systems for the southern zone of Mexican ridges, where the geological age, regional tectonics, trap formation, source rock, reservoir, seals, and accumulation and preservation duration are described (CNH, 2005).

5.6 Facies distribution

Facies distribution classified by the delimitation of plays will be presented below:

- **Lower Paleocene Wilcox**: The lithologies distribution map in Figure 5.5 shows an increased potential of reservoir rock to the north (Zone 2, Zone 1 north - the source of the sediments were located to the north). To the south (Zones 4 and 5) is associated with fans/smaller lobes. Abyssal plain area (south of Zone 1) unfavorable to Paleocene reservoir lithology presence. The best sandstone facies would be located in Zone 2 (with few salt intrusions) and at the north of Zone 1. The "Whopper" sandstone member, Lower Paleocene Wilcox equivalent is located at a depth range between 5,000 and 14,000 m (in a water depth $>$ 600 m). In most of Zones 1-5, burial is included between 6,000 and 9,000 m. The minimum depth ($\sim$5000 m), which also corresponds to the minimum burial range ($\sim$3500 m), is located at the south of Saline Basin and at the north of Perdido Fold Belt shows the

largest salt structures (Zone 2). In addition, there are some locally wedged or stratigraphic traps along the Southern Continental Shelf, at a depth range of 5,000-6,000 m (4,000-5,000 m burial).

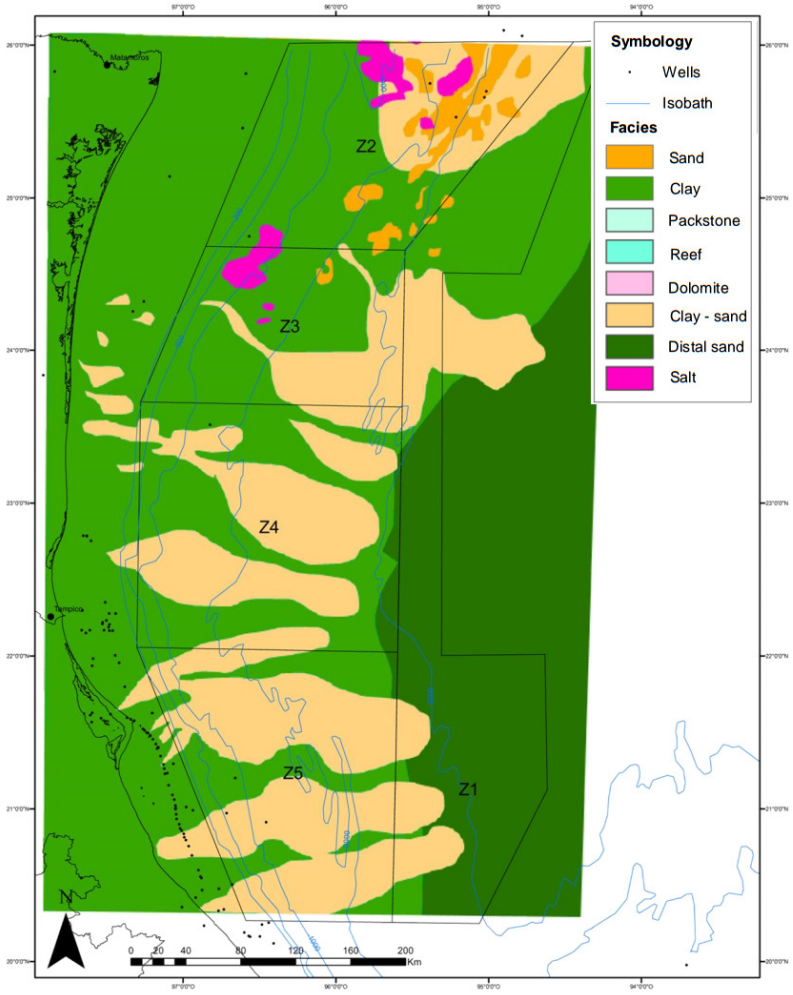


Figure 5.5: Spatial distribution map of facies in the Lower Paleocene Wilcox and the symbology Z1: zone 1, Z2: zone 2, Z3: zone 3, Z4: zone 4, Z5: zone 5 (CNH, 2005).

- **Upper Paleocene Wilcox**: The lithology distribution maps are very similar between the Upper Wilcox play and the Lower Wilcox play (increased potential of reservoir rock to north - Zone 2 and north of Zone 1). The difference in depth/burial between the Wilcox Lower Paleocene play and the Wilcox Upper Paleocene play is in general less than 500 m (Figure 5.6).

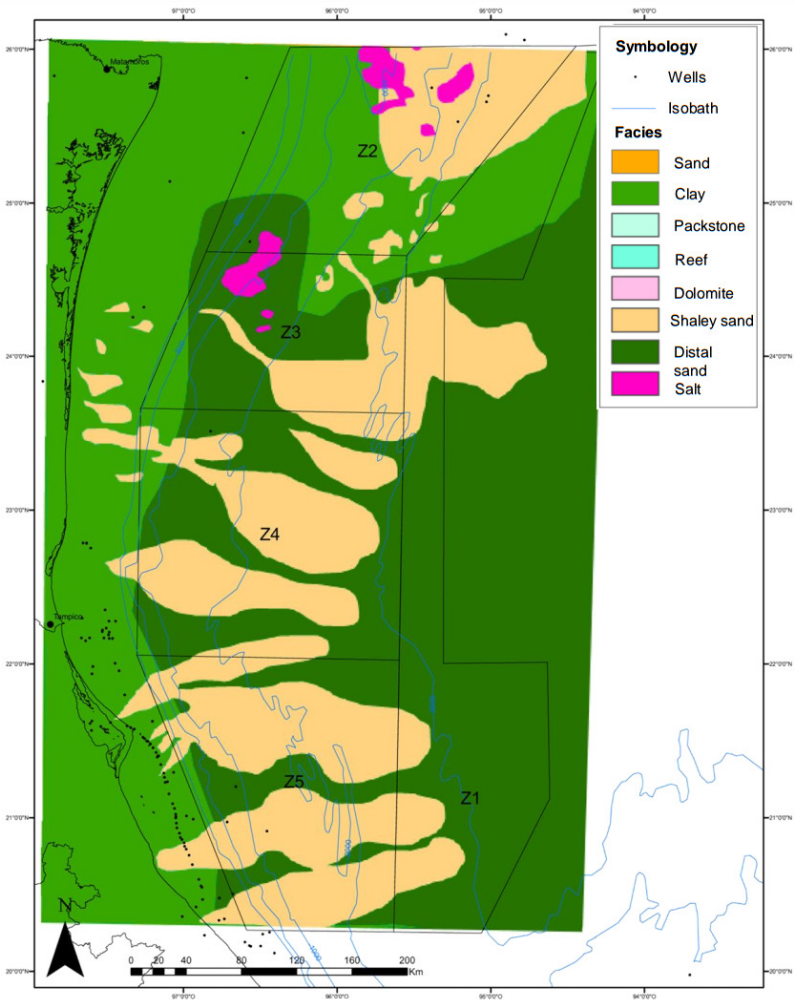


Figure 5.6: Spatial distribution map of facies in the Upper Paleocene Wilcox and the symbology Z1: zone 1, Z2: zone 2, Z3: zone 3, Z4: zone 4, Z5: zone 5 (CNH, 2005).

- **Lower Eocene Wilcox**: The distribution of sand bodies in Lower Eocene Wilcox play in Figure 5.7 shows that the best reservoirs are located in the north of Zone 2, and possibly in Zones 3 and 4. The proportion of sediments that come directly from the Mexican platform increases. In other areas, reservoirs are low quality and are deposits related to fans. The Lower Eocene Wilcox is located at a depth range between 4,000 and 12,000 m. The minimum depth, which also corresponds to the minimum burial, is located near the shoreline, outside the deepwater zone. In most of the Zones 1-5, the burial is included between 4,500 and 8,500 m, except in areas with large amplitude structures (Zone 2, 2,500m locally burial - depth of about 4,000 m minimum).

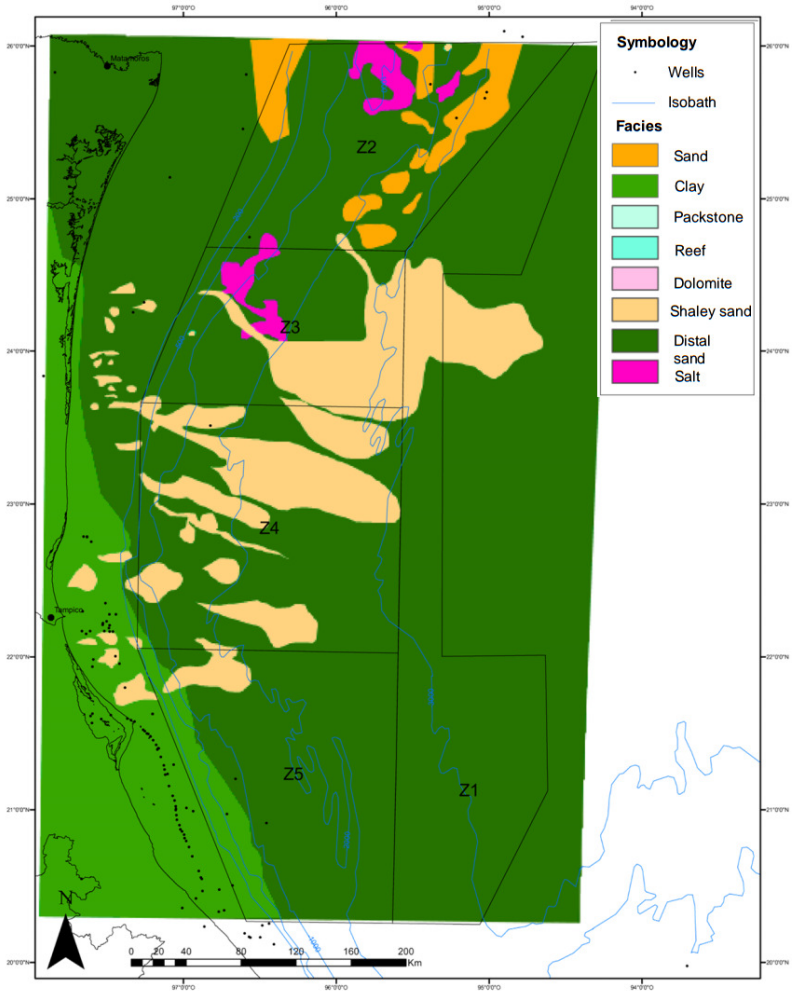


Figure 5.7: Spatial distribution map of facies in the Lower Eocene Wilcox and the symbology Z1: zone 1, Z2: zone 2, Z3: zone 3, Z4: zone 4, Z5: zone 5 (CNH, 2005).

- **Oligocene**: The lithology distribution map in deepwater Oligocene play area in Figure 5.8 shows that the sand bodies are found in almost all areas. It is mainly fans at the toe of the slope with uncertain reservoir quality. Zones 2 and 3 are largely covered with salt intrusions. The Oligocene is at a depth range between 3,000 and 7,000 m. The maximum values of depth and burial are on the shoreline in the Burgos Basin and the transitional zone between shelf and slope, where submarine slides are located. In most of the areas 1-5 burial is included between 750 and 3,750 m with minimum values to the west of the Southern Continental Platform and in the structural highs of Perdido Fold Belt (locally 1,500 m of burial to east of Zone 2, 3,000-4,000 depth under the sea level).

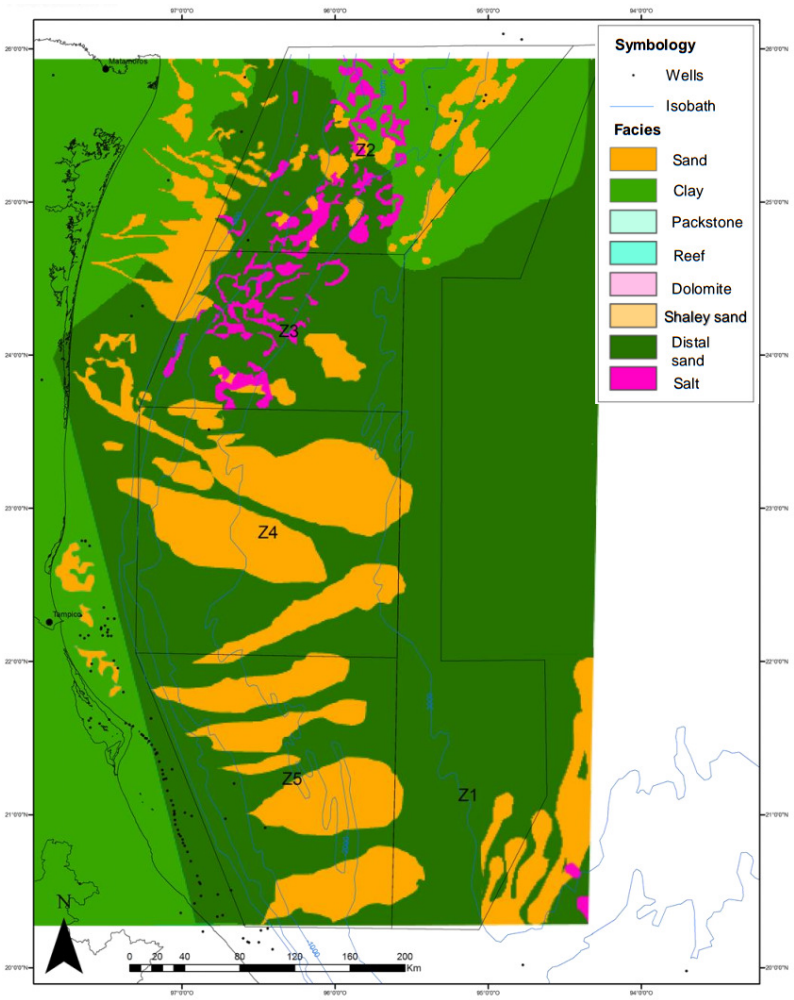


Figure 5.8: Spatial distribution map of facies in the Oligocene and the symbology Z1: zone 1, Z2: zone 2, Z3: zone 3, Z4: zone 4, Z5: zone 5 (CNH, 2005).

- **Middle Miocene**: Figure 5.9 show that the best reservoir rocks may be found in the northern part of the map (note: the reservoirs are interrupted by numerous salt intrusions). In the southern part, the reservoir rock quality is lower (in the proposed geological model), but the presence of large lobes are expected. The depth of the Middle Miocene play varies from a few hundred meters at the southern Continental Shelf and of 5,000 m in the abyssal zone. In most of the slope area (Zone 2-5), the Middle Miocene play is in between 3,000 and 4,500 m depth and compressive structures in zones 3-5 are well recognized. Throughout the entire area, burial rate is the lesser (less than 1,500 m), except where slides occurred (up to 4,500 m). In most of the Salina del Bravo and the Perdido Fold Belt (Zone 2), the burial rate is less than 750 m, which could limit the ability of preservation of hydrocarbon traps (seals may not be efficient enough to maintain an hydrocarbon column in place, particularly

if gas is expected).

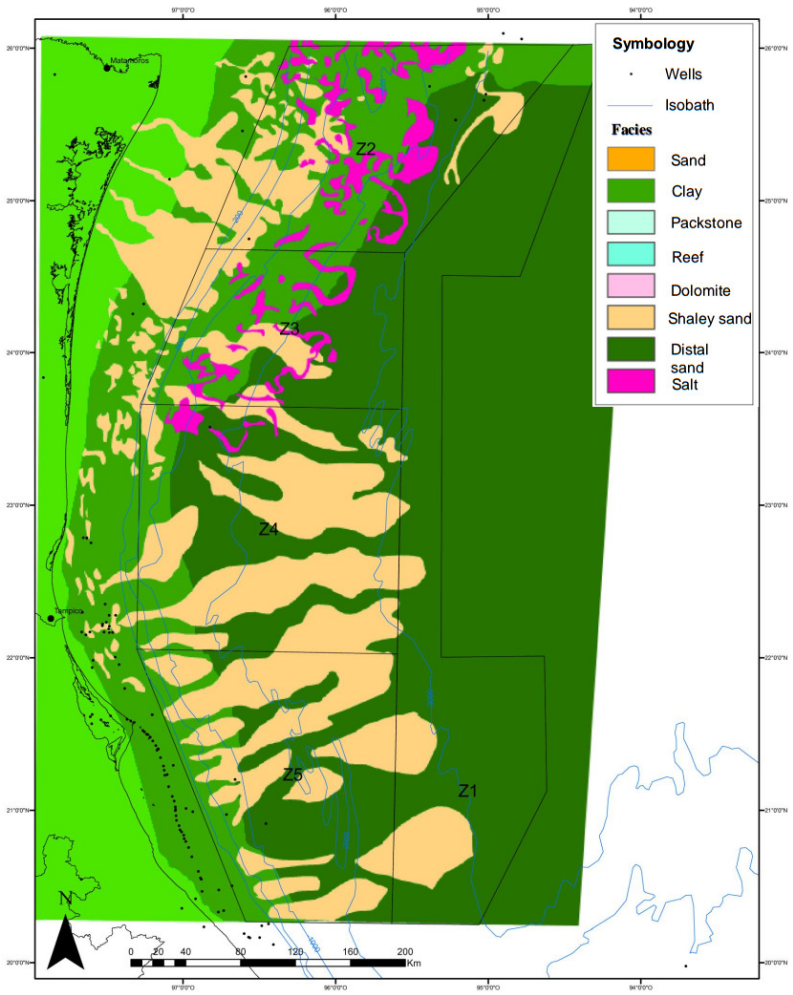


Figure 5.9: Spatial distribution map of facies in the Middle Miocene and the symbology Z1: zone 1, Z2: zone 2, Z3: zone 3, Z4: zone 4, Z5: zone 5 (CNH, 2005).

- **Upper Miocene**: Figure 5.10 show that the best reservoir lithologies are inferred at the northern part of the map, but reservoirs are small and are interrupted by the salt intrusions. The southern part has not good quality reservoir properties in the proposed geological model. The Upper Miocene play depth ranges from a few hundred meters on the southern Continental Shelf and about 5,000 m at the abyssal zone. In most of the slope area (Zone 2-5), the Middle Miocene play is between 1,500 and 4,500 m depth. Throughout the entire area, the depth is relatively shallow (less than 1,500 m), except where slides occur (up to 3,500 m). In the most part of the Salina del Bravo, in the Perdido Fold Belt (Zone 2), and in the abyssal plain, the burial rate is less than 750 m, which could limit the ability of preservation

of hydrocarbon traps (seals may not be efficient enough to maintain a hydrocarbon column in place, particularly if gas is expected).

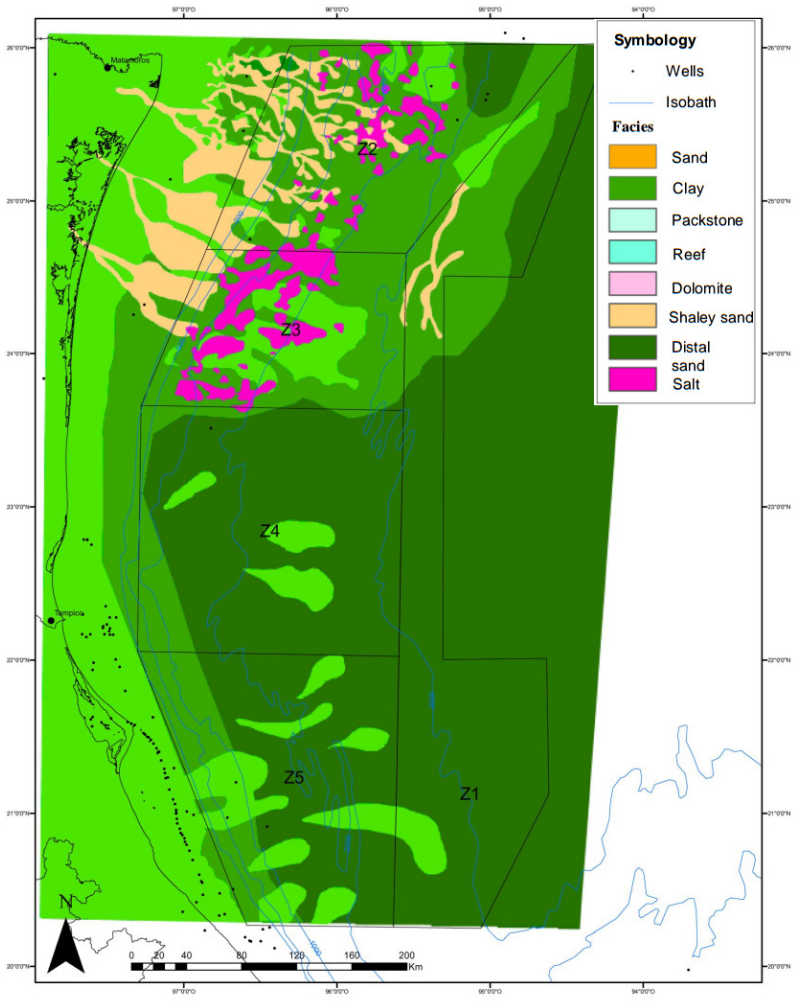


Figure 5.10: Spatial distribution map of facies in the Upper Miocene and the symbology Z1: zone 1, Z2: zone 2, Z3: zone 3, Z4: zone 4, Z5: zone 5 (CNH, 2005).

Chapter 6

Bivariate case study

In this section, the methodology is applied to bivariate data. The BCSCS method is validated and compared with the SGCS method in a 1-dimensional case using well log data. The petrophysical property, effective porosity, and the elastic attribute, P-impedance (Ip), come from the well logs of Lakach-1 with a sampling interval of 1 m and a depth from 3035 m to 3380 m. P-impedance is used as conditioning variable to simulate effective porosity. The simulated effective porosity is validated with the reference data at the well log scale.

First, the bivariate joint probability distribution function of effective porosity and acoustic impedance from the available data (well log) is inferred, then the effective porosity is simulated using the P-impedance as secondary variable and compare with the results with the reference data at the well location. Second, the prediction of the effective porosity in the inline section of seismic data conditioned by the P-impedance available in this section. The results were published in the journal of petroleum science and engineering.

6.1 Reference data at the well log scale

Figure 6.1 (a) and (b) show the spatial distribution, mean, and median of the reference data. P-impedance shows a moderate depth trend. I assume that effective porosity is stationary since the local mean and variance do not vary with the coordinates. In order to confirm the stationary behavior, it is necessary to analyze the semivariogram (Díaz-Viera, 2002).

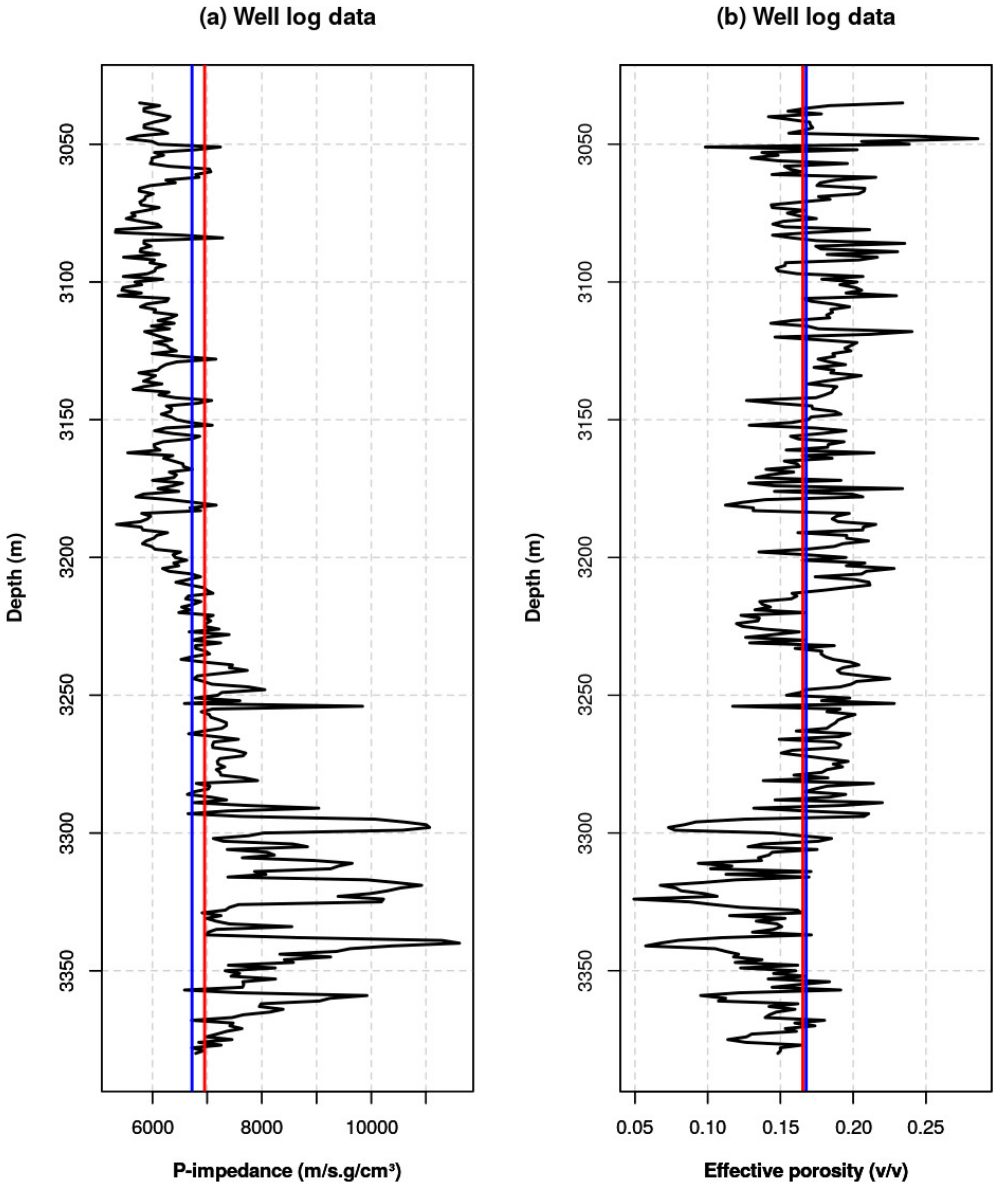


Figure 6.1: Spatial distribution of P-impedance and the effective porosity at the well log scale (black color), and its mean (red color), median (blue color).

6.2 Exploratory data analysis

6.2.1 Univariate analysis

Table 6.1 shows the statistics of P-impedance and effective porosity. P-impedance has a high positive skewness 1.5870, effective porosity has a low negative skewness -0.4323. Figure 6.2 (a), (b), (c), and (d) show that the univariate probability distribution of P-impedance is skewed whereas effective porosity is close to be unimodal and symmetric. Figure 6.2 (c) and (d) show the existence of uni-distributional outliers. P-impedance has high outlier values and effective porosity has both high and low outlier values. The further the skewness coefficient is from zero, the more skewed the distribution will be. The uni-distributional outliers strongly affect average estimators such as the mean, the semivariogram (Díaz-Viera, 2002). Therefore, when estimating the semivariogram, it is recommended to eliminate the uni-distributional outliers.

Table 6.1: The statistics of P-impedance and effective porosity.

Statistics	P-impedance	Effective porosity
Observation number	346	346
Minimum	5324.4324	0.0493
1st. Quartile	6112.2847	0.1444
Median	6723.4581	0.1677
Mean	6953.9369	0.1654
3rd. Quartile	7347.0286	0.1896
Maximum	11612.4245	0.2857
Rank	6287.9922	0.2364
Interquartile Rank	1234.7439	0.0452
Variance	1395507.9379	0.0012
Standard Deviation	1181.3162	0.0351
Variation Coeff.	0.1699	0.2123
Skewness	1.5870	-0.4323
Kurtosis	5.6322	3.5427

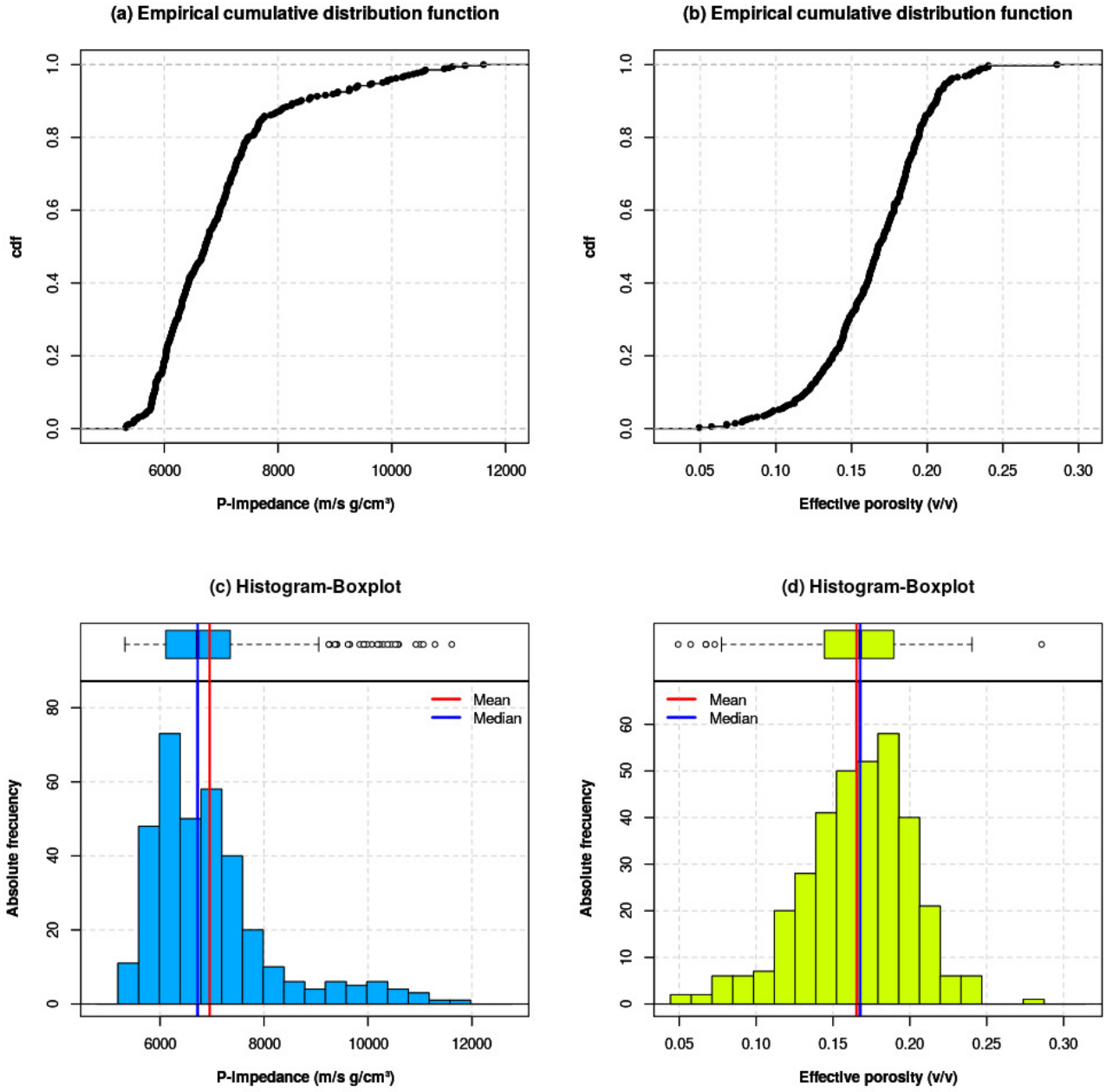


Figure 6.2: Empirical cumulative distribution function and histogram-boxplot of P-impedance and effective porosity at the well log scale, and its mean (red color) and median (blue color).

6.2.2 Dependency analysis

Figure 6.3 shows the scatterplot and pseudo-observation of P-impedance and effective porosity. It is observed that P-impedance and effective porosity show an average negative correlation, which means that, in general, as the P-impedance increases, effective porosity decreases. The bi-distributional outliers (where the uni-distributional outliers match) are also shown in the lower right of the scatterplot.

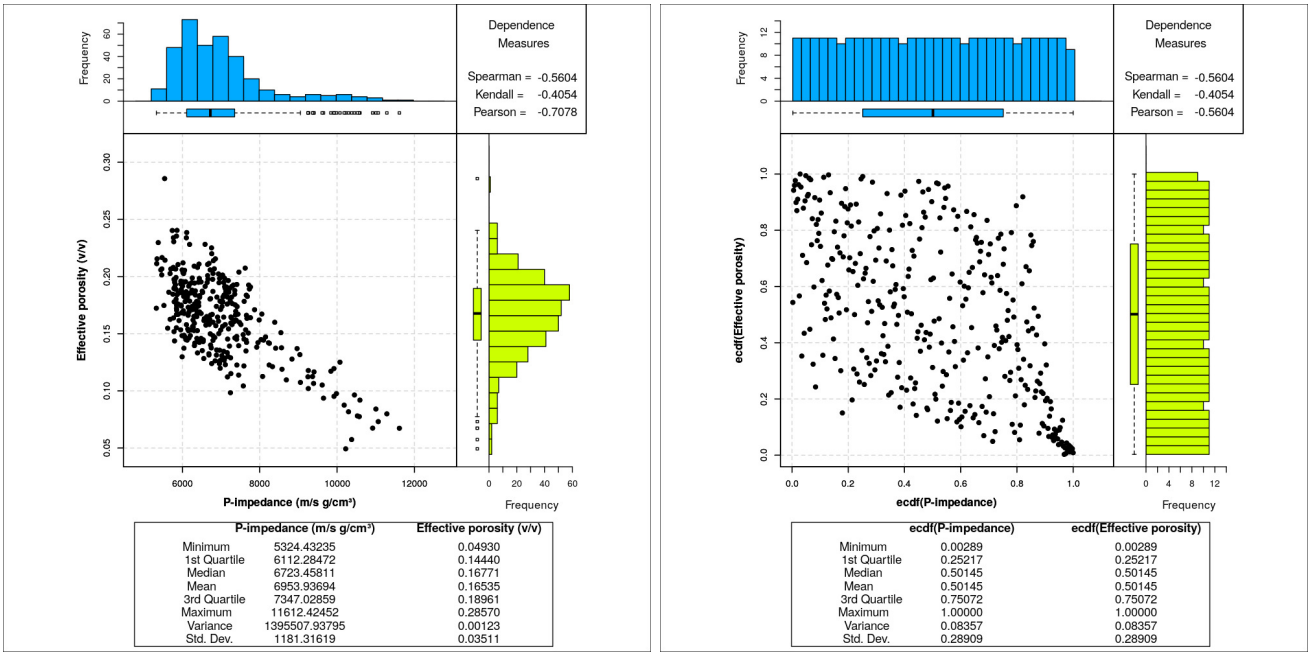


Figure 6.3: Scatterplot and Pseudo-observation of P-impedance versus effective porosity at the well log scale, and its correlation coefficients and statistics.

Figure 6.3 also showed that the Pearson correlation coefficient changed from -0.7078 to -0.5603 when scatterplot was changed to pseudo-observation, but the Spearman and Kendall correlation coefficients remained unchanged. Which indicated that Spearman and Kendall are more robust than Pearson. Pearson is useful when the dependency is linear, Spearman and Kendall are useful for any type of dependency. Pseudo-observation in Figure 6.3 showed that the marginal or univariate functions are uniformly identically distributed, which showed that the copula function is independent of the marginal behaviors, in other words, the copula does not depend on the marginal behaviors of the variables.

6.3 Spatial correlation analysis

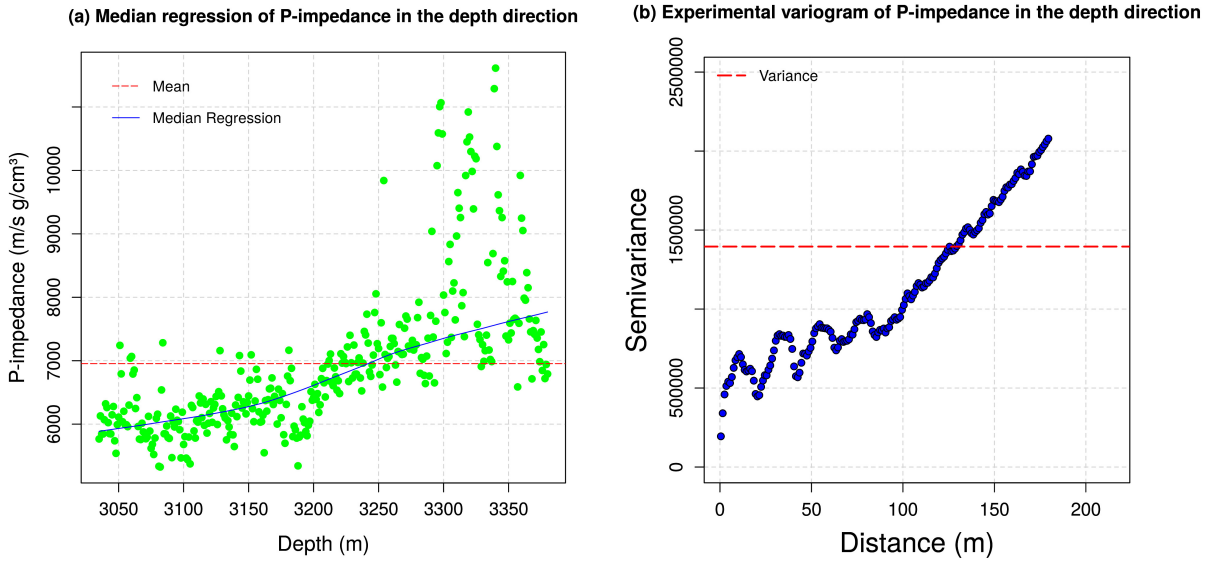


Figure 6.4: (a) Scatterplot of P-impedance versus depth and its median regression and (b) Experimental semivariogram of P-impedance in the depth direction.

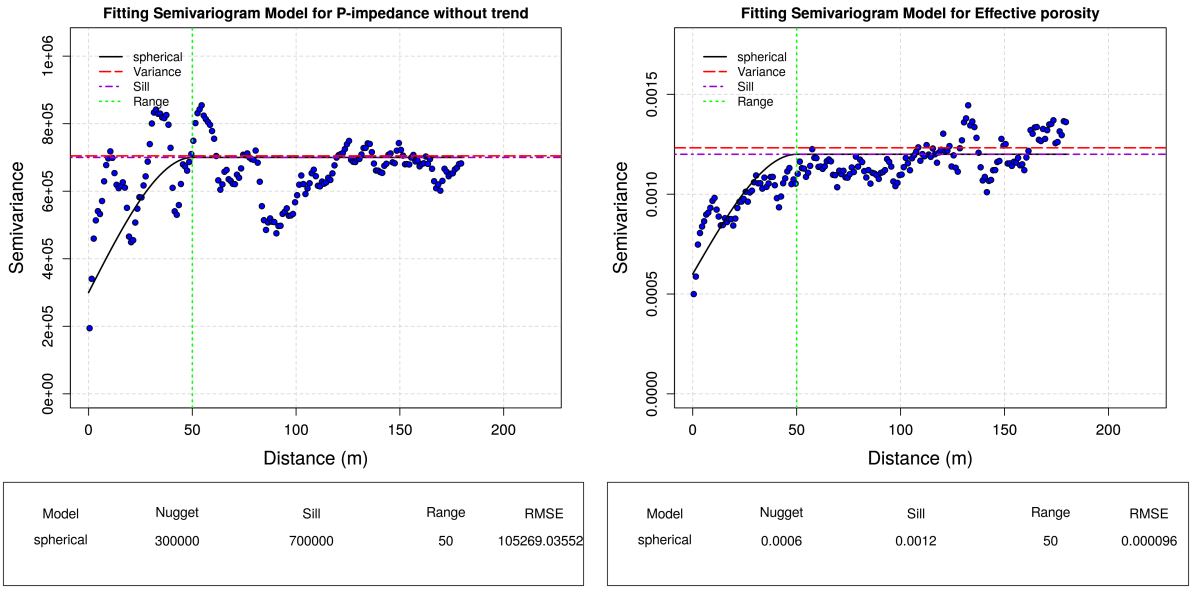


Figure 6.5: Semivariogram models and empirical semivariogram for P-impedance and effective porosity in depth direction.

When the semivariogram is bounded to the extent of the variance, then it is said that the variable has at least intrinsic hypothesis stationarity (Díaz-Viera, 2002). Figure 6.4(a) and (b) show that P-impedance has non-stationary behavior, because its semivariogram grows with the square of the depth. In order to estimate a semivariogram model for P-impedance, it is necessary to remove the trend. In Figure 6.5, the estimated semivariogram model for P-impedance and effective porosity is presented. The spatial correlation range is approximately 50 m. This assumption complies with the assumption of the linear co-regionalization model of the traditional SGCS method (Chilès and Delfiner, 2012). The semivariogram model of the effective porosity $\gamma(h)$ is a spherical model with a nugget equal to 0.0006, a sill equal to 0.0012, and a scope equal to 50.0 m.

6.4 Copula-based dependency modeling

Figure 6.6 (a) and (b) show that empirical cdf of P-impedance and effective porosity are step functions (in black), but P-impedance and the effective porosity are continuous variables. Therefore, a smoothing technique is necessary for those functions. According to the methodology, the empirical cdf of P-impedance and effective porosity are approximated by Bernstein polynomial. Figure 6.6 shows the results of the approximation and also illustrates that Bernstein polynomial fits well to the empirical cdf of P-impedance and effective porosity. For comparison, the parametric Gaussian cdf does not fit the P-impedance distribution (Figure 6.6 (a)).

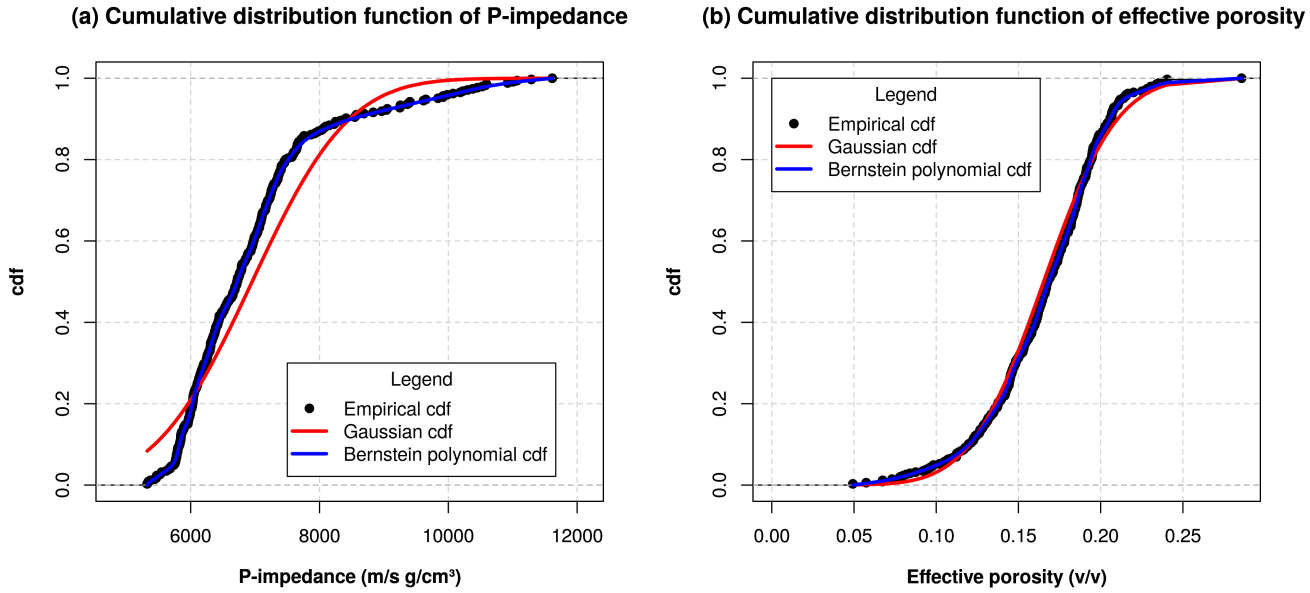


Figure 6.6: Empirical cdf of P-impedance and effective porosity approximated using Bernstein polynomial (in blue) compared to Gaussian cdf with mean and standard deviation of the variable (in red).

Figure 6.7 (a) shows the step function representing the empirical copula of P-impedance and effective porosity (in black). The empirical copula of P-impedance and effective porosity is approximated by Bernstein copula (in blue) and it is shown in Figure 6.7 (a). Pseudo-observation (empirical copula density) and Bernstein copula density are shown in Figure 6.7 (b). Figure 6.7 shows that the Bernstein copula function fitted the empirical copula function very well.

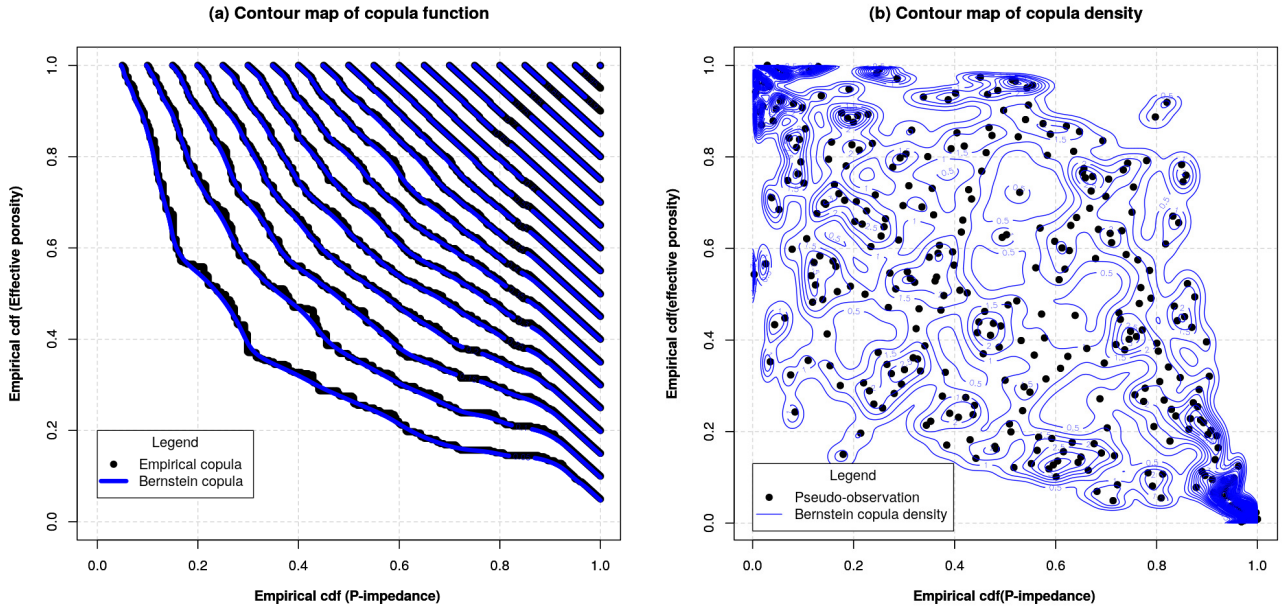


Figure 6.7: (a) Empirical copula and copula approximated by Bernstein copula of P-impedance and effective porosity, (b) Pseudo-observation and copula density approximated by Bernstein copula density.

6.5 Validation

Based on the previous results, 100 simulations of the effective porosity conditioned to P-impedance are simulated using the conditional simulation algorithm proposed in the methodology section. To reproduce the semivariogram model, we apply the simulated annealing method with the objective function of variogram.

The proposed method is compared to sequential Gaussian cosimulation (SGCS) based on a conditioning Markov model available in the open source code SGeMS (Remy et al., 2009). The variable of interest is effective porosity and secondary variable is P-impedance without trend. 100 simulations of the effective porosity conditioned to P-impedance are simulated by SGCS using the same semivariogram model (Figure 6.8).

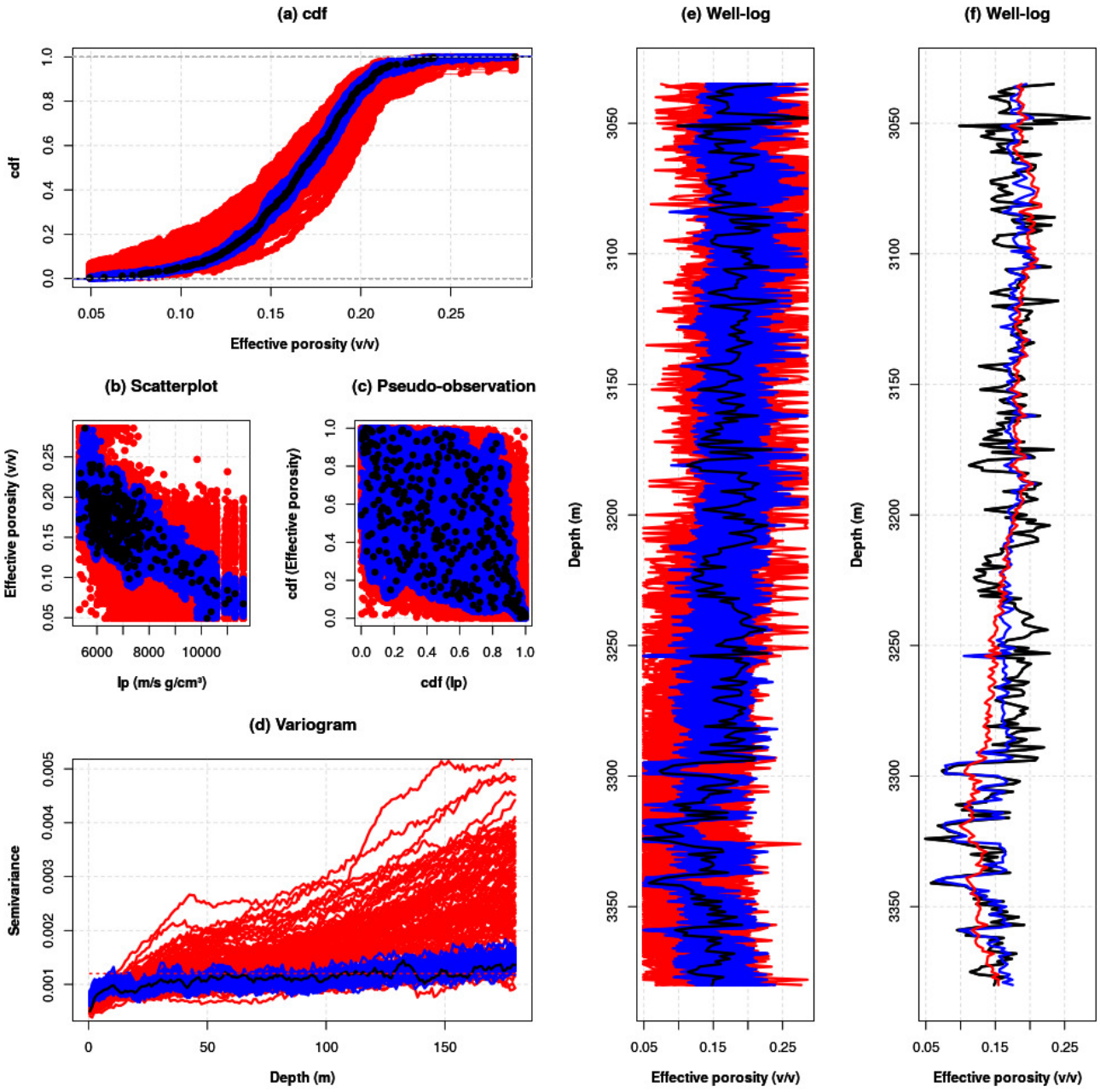


Figure 6.8: (a) Empirical cdf of effective porosity, (b) Scatterplot of P-impedance versus effective porosity; (c) Pseudo-observation of cdf(P-impedance) versus cdf(effective porosity), (d) Variograma and (e) Well-log of effective porosity (reference in black, 100 simulations by BCSCS in blue, 100 simulations by SGCS in red and variance in dashed red), and (f) Well-log of effective porosity (reference in black, mean of 100 simulations by BCSCS in blue, mean of 100 simulations by SGCS in red).

6.6 Uncertainty quantification

The simulations of the effective porosity conditioned to P-impedance obtained with the two simulation methods are compared and validated with the reference effective porosity. The univariate, bivariate, spatial distribution, as well as the modeling error the reference and simulated effective porosity are compared.

Table 6.2: Statistics of 100 simulations of effective porosity by SGCS and BCSCS, and reference effective porosity.

Statistics	BCSCS	Reference data	SGCS
Minimum	0.0493	0.0493	0.0492
Median	0.1688	0.1677	0.1673
Mean	0.1654	0.1654	0.1631
Maximum	0.2856	0.2857	0.2858
Variance	0.0012	0.0012	0.0020
Skewness	-0.4691	-0.4323	-0.3331

In Figure 6.8(a), the empirical univariate cdf results of the BCSCS method (in blue) show the smaller uncertainty compared to those of the SGCS method (in red). Table 6.2 shows the statistics of results of 100 simulations of effective porosity by the two methods and the reference effective porosity. In particular, the variance of the results of the SGCS method is almost the double the variance of the results of the BCSCS method and the reference effective porosity.

Comparing the bivariate aspect through scatterplot and pseudo-observation, Figure 6.8(b) and Figure 6.8(c) show that the results of the BCSCS method (in blue) reproduce the property dependence better than the results of the SGCS method (in red), compared to the reference data (in black). Table 6.3 shows that the dependency coefficients (Pearson, Kendall, Spearman) of the BCSCS method are closer to the references than for the SGCS method. Therefore, the results of the bivariate behavior of the SGCS method show greater uncertainty compared to those of the BCSCS method.

Table 6.3: Correlation coefficients of P-impedance and effective porosity for BCSCS, reference data and SGCS.

Coefficients	BCSCS	Reference data	SGCS
Pearson	-0.6966	-0.7078	-0.5720
Spearman	-0.5491	-0.5603	-0.6275
Kendall	-0.3937	-0.4051	-0.4409

In terms of the semivariogram model reproduction, Figure 6.8(d) and Figure 6.8(e) show that the uncertainty of 100 simulated effective porosity by the BCSCS method (in blue) is smaller

than the results of the SGCS method (in red) with respect to the reference semivariogram and well-log data (in black). Moreover, Figure 6.8(f) shows that the expected value of 100 simulations of effective porosity by BCSCS method matches the reference data. Therefore, the accuracy in Figure 6.8(f) and precision in Figure 6.8(e) improve with the BCSCS method.

Table 6.4: Error statistics of BCSCS and SGCS.

Statistics	**BCSCS**	**SGCS**
Minimum	-0.1351	-0.2183
Median	0.0002	-0.0005
Mean	0.0001	-0.0023
Maximum	0.1305	0.1771
Variance	0.0012	0.0021
Absolute Sum	908.1104	1267.168

Finally, the error between simulated values and reference data is analyzed. Table 6.4 and Figure 6.9 show the histogram-boxplot and the statistics of the errors between the reference data and 100 simulations of effective porosity and confirm that the BCSCS method provides better results.

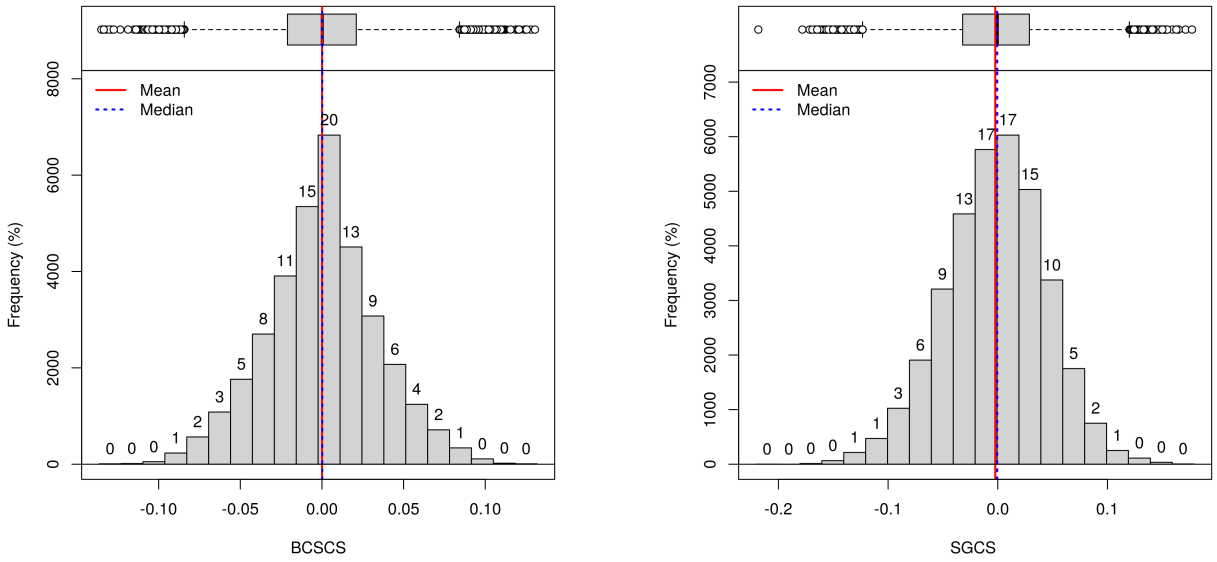


Figure 6.9: Histogram-boxplot of errors of 100 simulations of effective porosity by BCSCS and SGCS.

6.7 Application

The proposed method is finally applied to an in-line section of inverted P-impedance section (Figure 6.10(a)). The section shows a depth interval including sand and shale lithologies. P-impedance has been obtained from seismic inversion of amplitudes and travel-times, using a traditional convolutional method where the seismic response is assumed to be approximated as a convolution of the source wavelet and the reflection coefficients computed from P-wave velocity and density. The results of seismic inversion have been converted from time to depth by using the predicted seismic velocity. The low P-impedance in the upper part of the interval suggests a potential high-porosity reservoir.

The goal of the application is to predict the effective porosity distribution along the 2D section. The well log data in the previous section are used to model the rock physics relation between P-impedance and effective porosity. Seismic data and seismic properties estimated from the data have a lower resolution than well log data due to the limited bandwidth of the acquisition frequencies. Therefore, the models obtained from seismic data are typically smoother, in the vertical direction, compared to the corresponding well log data. The lateral continuity in the data is generally due to the geological continuity of the geobodies in the subsurface. The semivariograms are estimated from the P-impedance models.

Figure 6.11 shows the anisotropic behavior of the semivariograms of P-impedance in the vertical and lateral directions. It is observed that the range in the horizontal direction (200 m) is much greater than in the vertical direction (50 m), which indicates that the horizontal direction has greater spatial continuity compared to the vertical direction, which is reflected in Figure 6.10(a).

The BCSCS method is applied to predict effective porosity conditioned to inverted P-impedance at the seismic scale using their joint dependency model estimated at the well-log scale. This ensures that univariate and bivariate distributions of the effective porosity in the well-log scale will be reproduced. The vertical and lateral semivariogram models of effective porosity are assumed to be similar to the directional semivariogram models of P-impedance, due to the correlation between the two properties shown in the well logs.

Figure 6.10(b) and Figure 6.10(c) show the median value and the standard deviation of 100 simulations of effective porosity conditioned to inverted P-impedance. The predicted model shows a high-porosity region in the upper part of the interval, that corresponds to the potential reservoir layer according to geological interpretation and nearby wells. The spatial structure of effective porosity is not identically the same as the spatial structure of P-impedance because the joint dependence between them is not linear. The semivariograms of the median value of 100 simulations of effective porosity in the vertical and lateral directions are shown in Figure 6.12.

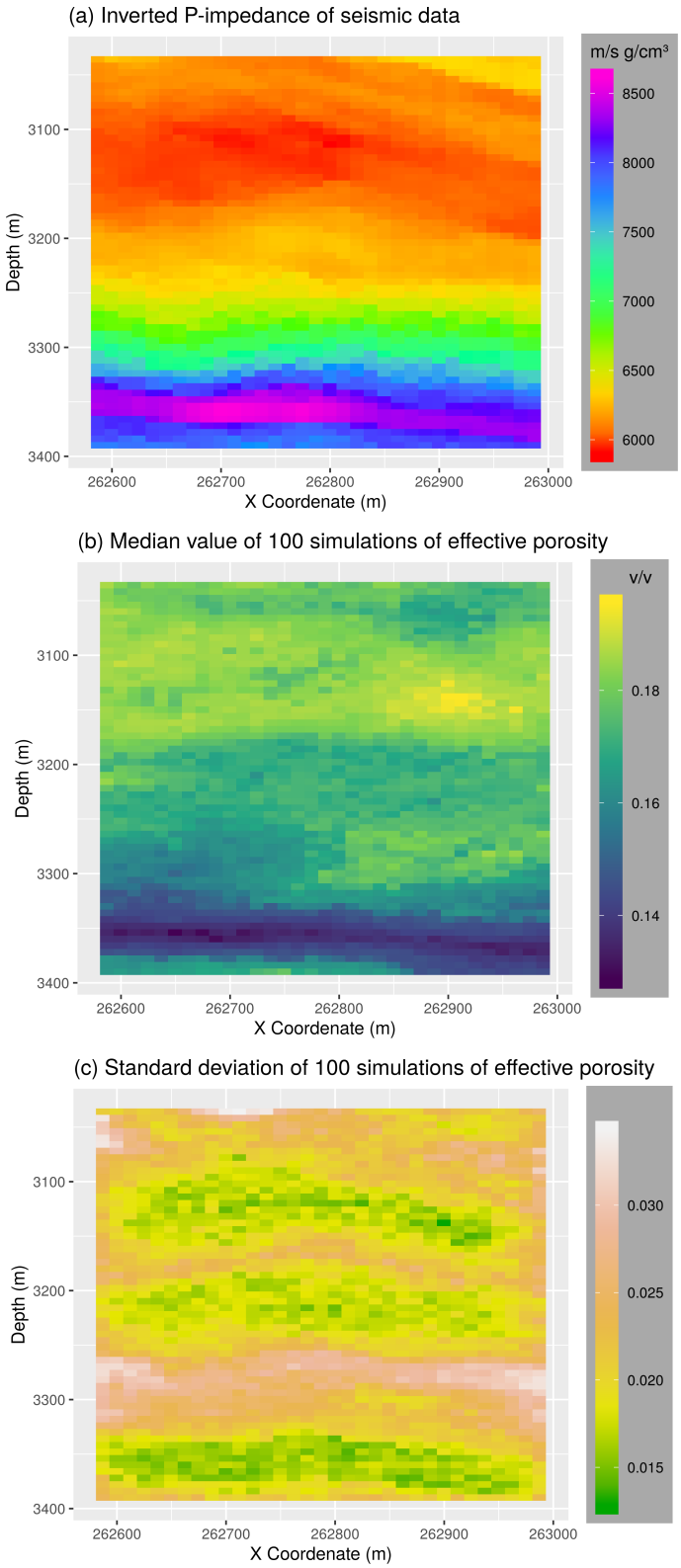
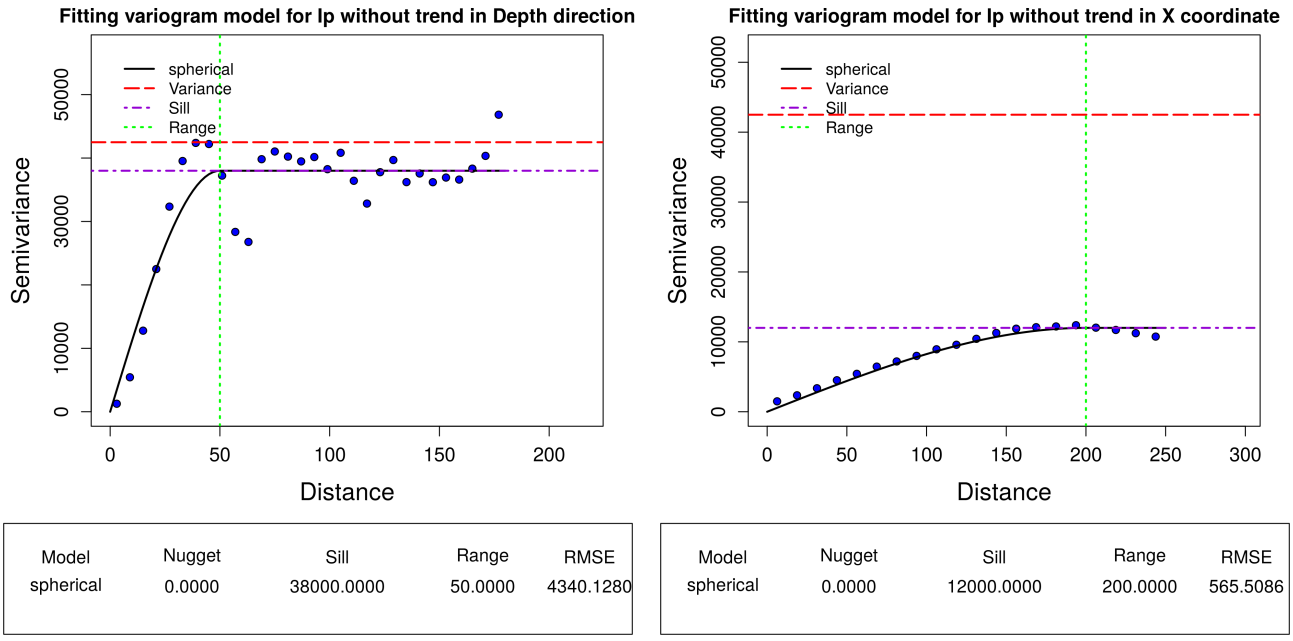


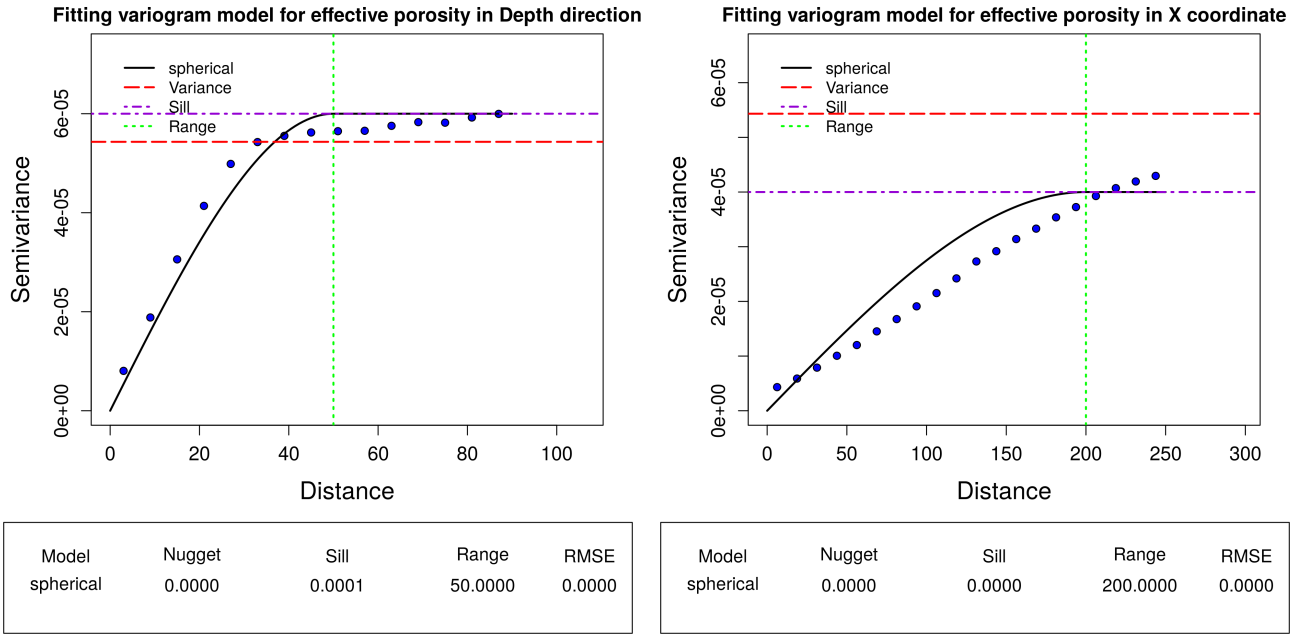
Figure 6.10: (a) In-line section of inverted P-impedance of seismic data; (b) Median value and (c) Standard deviation of 100 simulations of effective porosity.



Model	Nugget	Sill	Range	RMSE
spherical	0.0000	38000.0000	50.0000	4340.1280

Model	Nugget	Sill	Range	RMSE
spherical	0.0000	12000.0000	200.0000	565.5086

Figure 6.11: Semivariogram models and empirical semivariogram of P-impedance along the in-line section.



Model	Nugget	Sill	Range	RMSE
spherical	0.0000	0.0001	50.0000	0.0000

Model	Nugget	Sill	Range	RMSE
spherical	0.0000	0.0000	200.0000	0.0000

Figure 6.12: Semivariogram models and empirical semivariogram of the median value of 100 simulations of effective porosity.

Chapter 7

Multivariate case study

In this section, the methodology is applied to data sets of petrophysical properties (total and effective porosity, permeability, water saturation, and clay volume) and elastic seismic attributes (P-wave and S-wave velocity, density, acoustic and elastic impedance, compressibility modulus and the transverse elastic modulus, Lamé's first parameter, and the Poisson's ratio). It is not intended to use all the variables of the elastic seismic attributes to predict petrophysical properties. Therefore, before applying the methodology, a secondary variable selection analysis is required where the secondary variables that have the highest dependence on the primary variable are selected.

7.1 Reference data at the well log scale

The well being worked on in this section is the Lakach-1 well. The well log section used in this section is within the Miocene geological age. The rocks are sandstones and shales in the channel and the basin floor fans. The depth is from 3035 m to 3404 m. The data was originally sampled at 0.1m, but it was subsampled at 1.0 m due to the amount of data.

Figure 7.1 and Figure 7.2 show the reference data in the well logs which are the petrophysical properties: total and effective porosity (ϕ_t, ϕ_e), log (Permeability) (logKtimur), clay volume (Vcl), water saturation (Sw); and the elastic seismic attributes: P-wave and S-wave velocity (Vp, Vs), density (Rhob), P-wave and S-wave impedance (Ip, Is), Poisson's ratio (Poisson), compressibility and transverse elastic modulus (K, μ), and Lamé's first parameter (λ). Figure 7.1 shows a trend from high to low values of total and effective porosity and log(permeability); a trend from low to high values of water saturation; and no trend in clay volume as depth increases. Figure 7.2 shows a trend from low to high values for all elastic seismic attributes, except the Poisson's ratio, which appears to show no trend as depth increases.

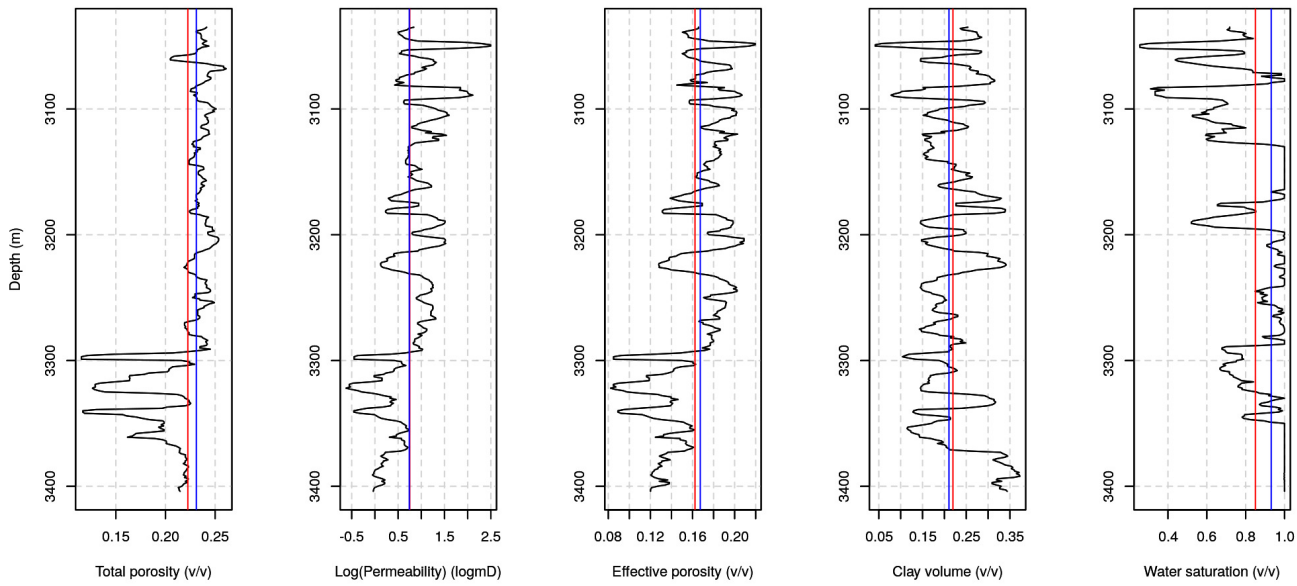


Figure 7.1: Spatial distribution of the petrophysical properties in the well log (black color), and their means (red color) and medians (blue color).

There are 9 variables of the seismic attributes, it is not intended to use all these variables as secondary variables to predict the petrophysical properties. Therefore a variable selection analysis is required below. Based on the Figure 7.3 and the Table 7.1, the pair of variables between the petrophysical properties and elastic seismic attributes that presents the highest dependence is the effective porosity and the density with the Spearman correlation coefficient -0.98. Due to the very high value of Spearman's coefficient between effective porosity and density, therefore, it is not necessary to add another secondary variable to predict effective porosity, just use density variable as secondary variable.

The next pair of variables between the petrophysical and elastic properties that presents very high dependence is the log(Permeability) and the density with the Spearman correlation coefficient -0.92. Due to the very high coefficient value, therefore, it is decided that to predict the log (Permeability) it is sufficient to use the density variable, that is, it is not necessary to add another secondary variable. Note that predicted effective porosity can also be used as the secondary variable due to its high dependence on log(Permeability). The density was chosen because it was the original data.

The pair of variables that follows is the total porosity and density with the Spearman correlation coefficient -0.81. Due to this not so very high value, therefore, it is necessary to look for another secondary variable that presents high dependence on total porosity and low dependence on density. The most suitable is the S-wave velocity that presents the Spearman correlation coefficient with the total porosity -0.41 and with the density 0.10.

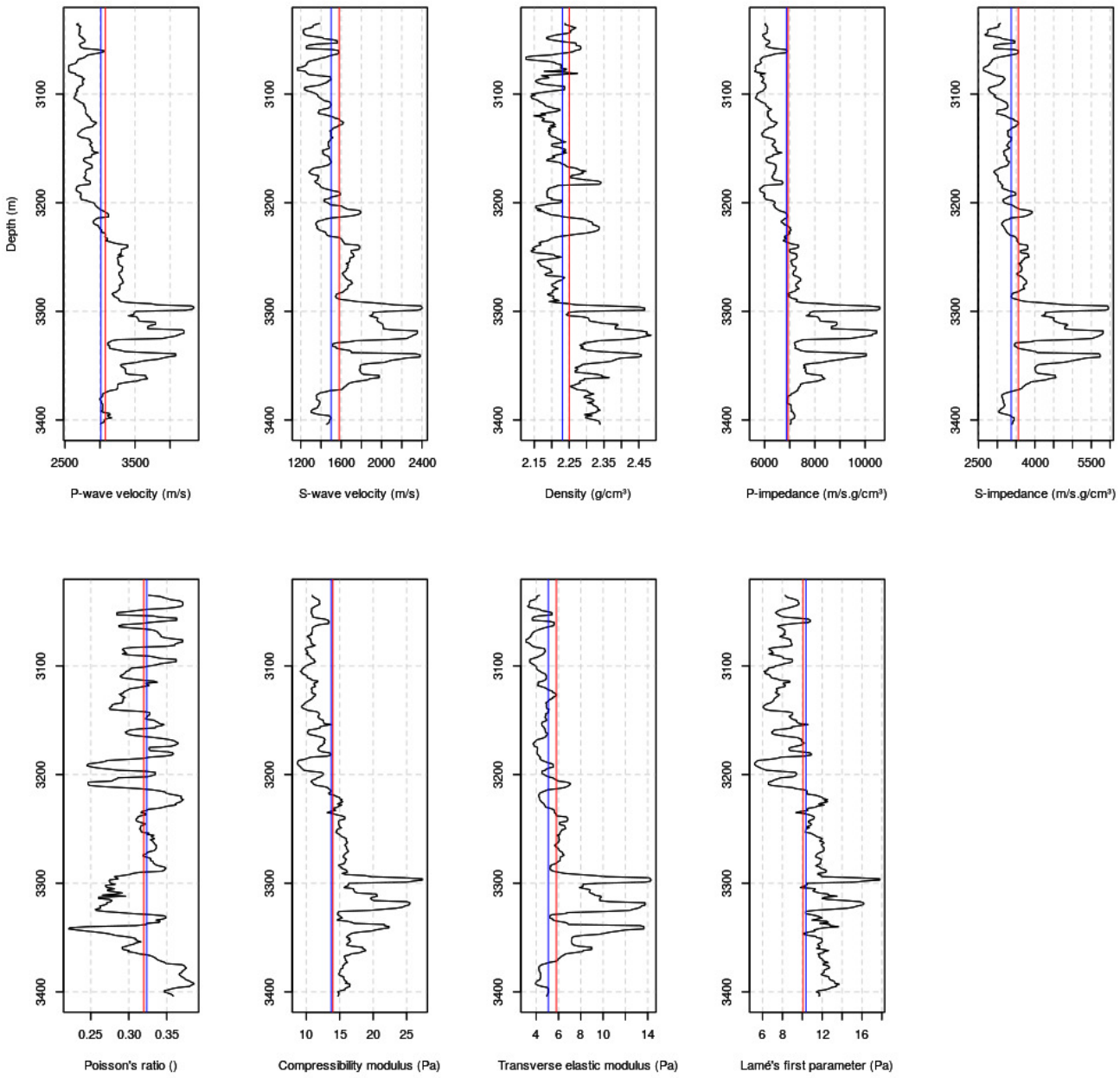


Figure 7.2: Spatial distribution of the elastic seismic attributes in the well log (black color), and their means (red color) and medians (blue color).

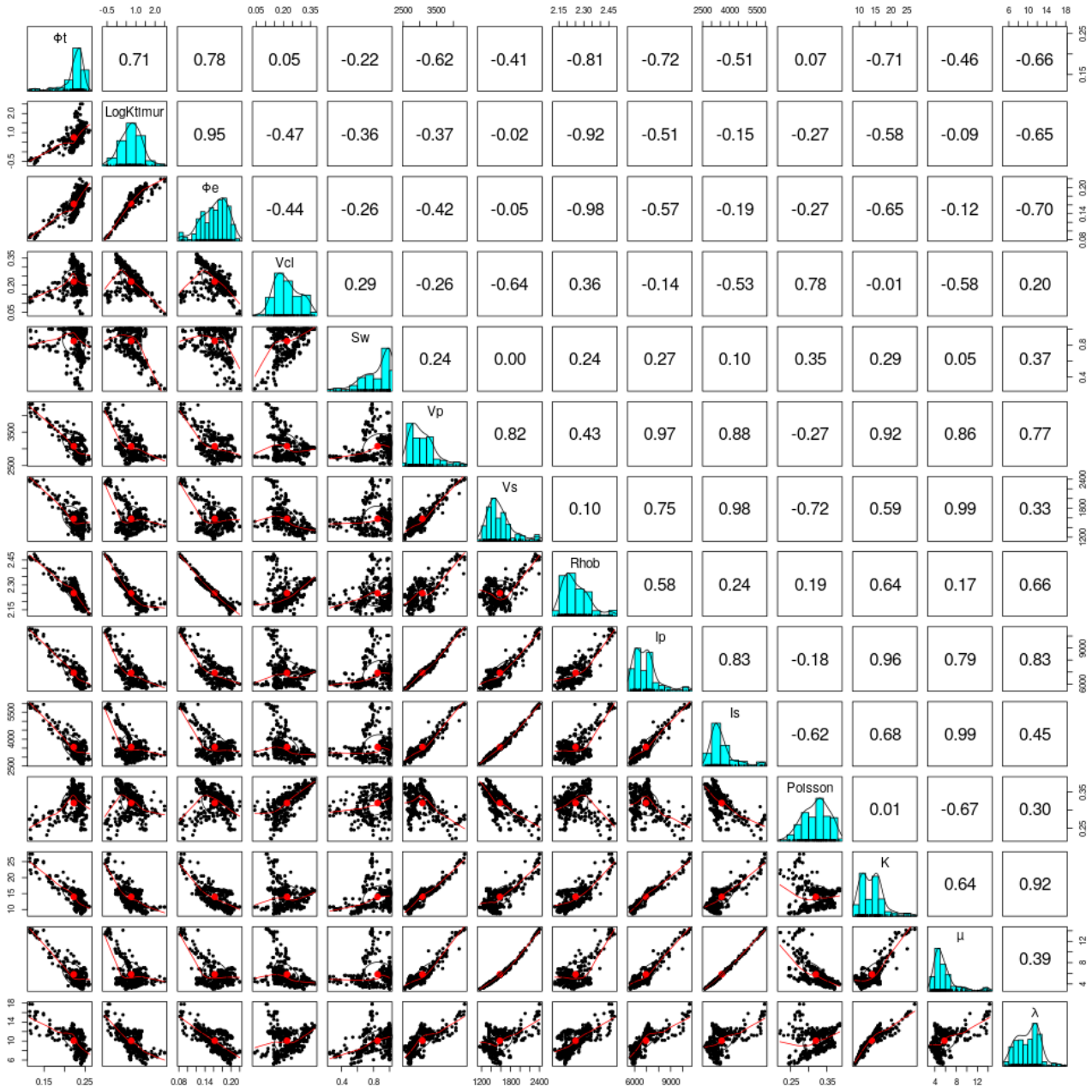


Figure 7.3: Scatterplot matrix of petrophysical properties and elastic seismic attributes, and their Spearman correlations.

The secondary variables that are selected to predict the clay volume are Poisson's ratio and density, because these two variables have a high dependence on clay volume and a low dependence on each other. With the same reasoning, the two variables are selected: the Poisson's ratio and

the compressibility modulus as secondary variables to predict the water saturation.

Table 7.1: The Spearman correlation coefficient between the petrophysical properties and the elastic seismic attributes.

	ϕ_t	LogKtimur	ϕ_e	Vcl	Sw	Vp	Vs	Rhob	Ip	Is	Poisson	K	μ	λ
ϕ_t	1.00	0.71	0.78	0.05	-0.20	-0.62	-0.41	-0.81	-0.72	-0.51	0.07	-0.71	-0.46	-0.66
LogKtimur	-	1.00	0.95	-0.47	-0.35	-0.37	-0.02	-0.92	-0.51	-0.15	-0.27	-0.58	-0.09	-0.65
ϕ_e	-	-	1.00	-0.44	-0.26	-0.42	-0.05	-0.98	-0.57	-0.19	-0.27	-0.65	-0.12	-0.70
Vcl	-	-	-	1.00	0.30	-0.26	-0.64	0.36	-0.14	-0.53	0.78	-0.01	-0.58	0.20
Sw	-	-	-	-	1.00	0.24	-0.01	0.23	0.27	0.09	0.36	0.30	0.05	0.38
Vp	-	-	-	-	-	1.00	0.82	0.43	0.97	0.88	-0.27	0.92	0.86	0.77
Vs	-	-	-	-	-	-	1.00	0.10	0.75	0.98	-0.72	0.59	0.99	0.33
Rhob	-	-	-	-	-	-	-	1.00	0.58	0.24	0.19	0.64	0.17	0.66
Ip	-	-	-	-	-	-	-	-	1.00	0.83	-0.18	0.96	0.79	0.83
Is	-	-	-	-	-	-	-	-	-	1.00	-0.62	0.68	0.99	0.45
Poisson	-	-	-	-	-	-	-	-	-	-	1.00	0.01	-0.67	0.30
K	-	-	-	-	-	-	-	-	-	-	-	1.00	0.64	0.92
μ	-	-	-	-	-	-	-	-	-	-	-	-	1.00	0.39
λ	-	-	-	-	-	-	-	-	-	-	-	-	-	1.00

7.2 Exploratory data analysis

There are 5 case studies: two bivariate cases and three trivariate cases. A bivariate case was already shown in chapter 6. Therefore, the bivariate case procedure is not presented. The procedure of a trivariate representative case is shown to avoid repetitions of trivariate cases. But the results of the five cases are showed in the end.

7.2.1 Univariate analysis

Figure 7.4 (a) to (f) show empirical cumulative distribution function and histogram-boxplot of density, S-wave velocity and total porosity. Table 7.2 shows the statistics of the variables: density, S-wave velocity, and total porosity. The density and S-wave velocity have positive skewness and the total porosity present negative skewness. The asymmetry behavior of these variables can be observed once again in Figure 7.4. It is possible due to the existence of the unidistributional outliers that are presented in the histogram-box plot of these variables.

Figure 7.4 (a) to (c) show empirical cumulative distribution functions are discontinuous and the variables are continuous, therefore it is necessary to fit a continuous function to these functions. Figure 7.5 (a) to (c) show the empirical cumulative probability distribution functions for density, S-wave velocity, and total porosity fitted by the Bernstein polynomial. This fitting approach is a non-parametric approach. Kernel smoothing is another method of this approach. The advantage of this approach is that it fits the empirical distribution function very well. The downside is that when there is a lot of data, the execution time is high, and vice versa for the parametric approach.

The specific advantage of the Bernstein polynomial is that this polynomial can be derived and its inverse function can be calculated, which is needed for a probability distribution function. The derivative of the cumulative probability distribution function is the distribution density function and its inverse function is the quantile function of the random variable.

Once the univariate probability distribution function is modeled and known. Then, the univariate behavior of the random variable can be fully known. Then, the simulations or realizations of the values of the random variable can be performed by generating the values [0,1] of the uniform distribution and then using the quantile function to obtain the values of the random variable.

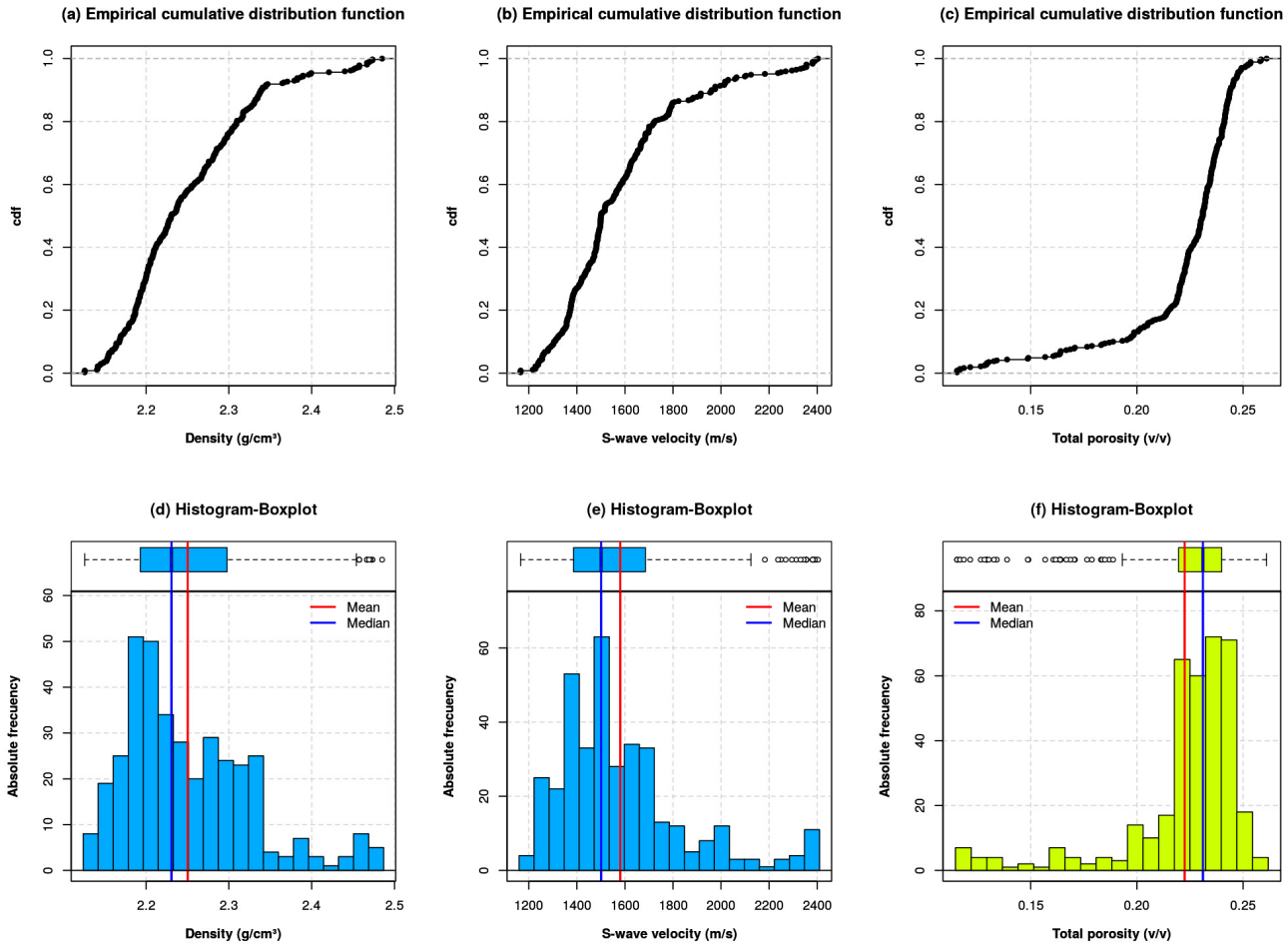


Figure 7.4: Empirical cumulative distribution function and histogram-boxplot of density, S-wave velocity and total porosity at the well log scale, and its mean (red color) and median (blue color).

Table 7.2: The statistics of density, S-wave velocity, and total porosity.

Statistics	Density	S-wave velocity	Total porosity
Observation number	370	370	370
Minimum	2.1260	1165.6275	0.1153
1st. Quartile	2.1932	1386.3900	0.2197
Median	2.2307	1500.8185	0.2311
Mean	2.2503	1580.4548	0.2225
3rd. Quartile	2.2978	1685.1024	0.2399
Maximum	2.4851	2403.9558	0.2611
Rank	0.3591	1238.3283	0.1458
Interquartile Rank	0.1047	298.7124	0.0202
Variance	0.0058	69867.6711	0.0008
Standard Deviation	0.0762	264.3249	0.0282
Variation Coeff.	0.0339	0.1672	0.1266
Skewness	0.9508	1.2642	-2.1490
Kurtosis	3.6300	4.4642	7.5972

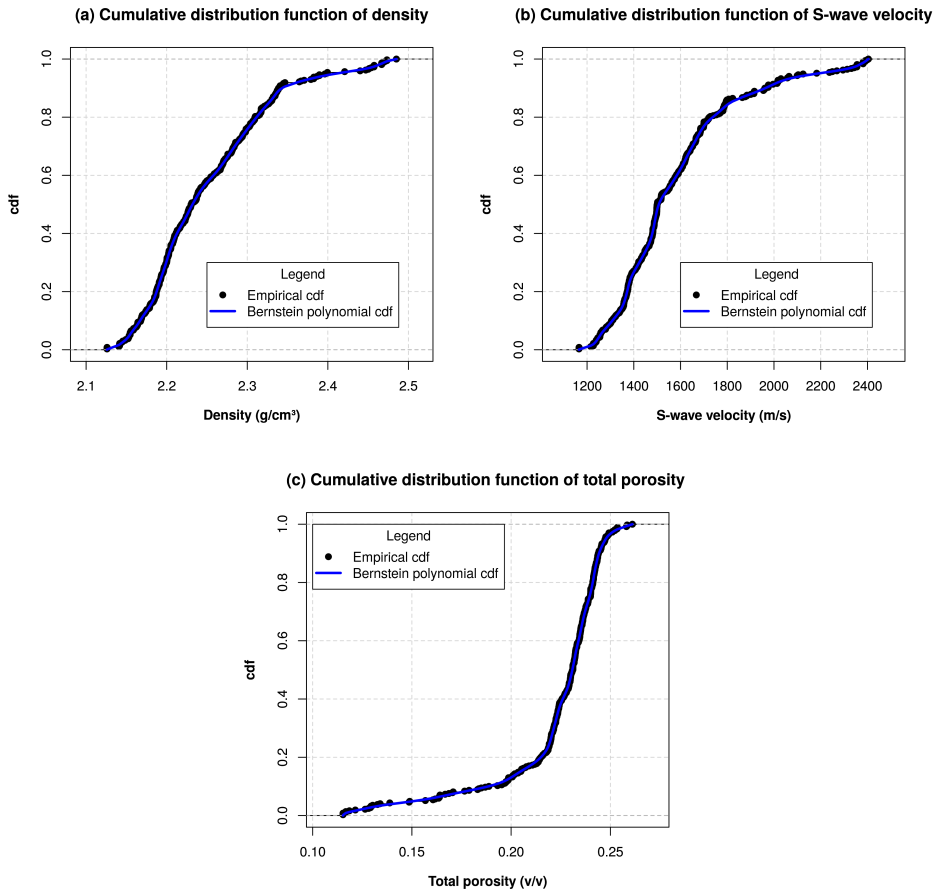


Figure 7.5: The empirical cumulative probability distribution functions of the density, S-wave velocity, and total porosity are fitted by the Bernstein polynomial.

7.2.2 Dependency analysis

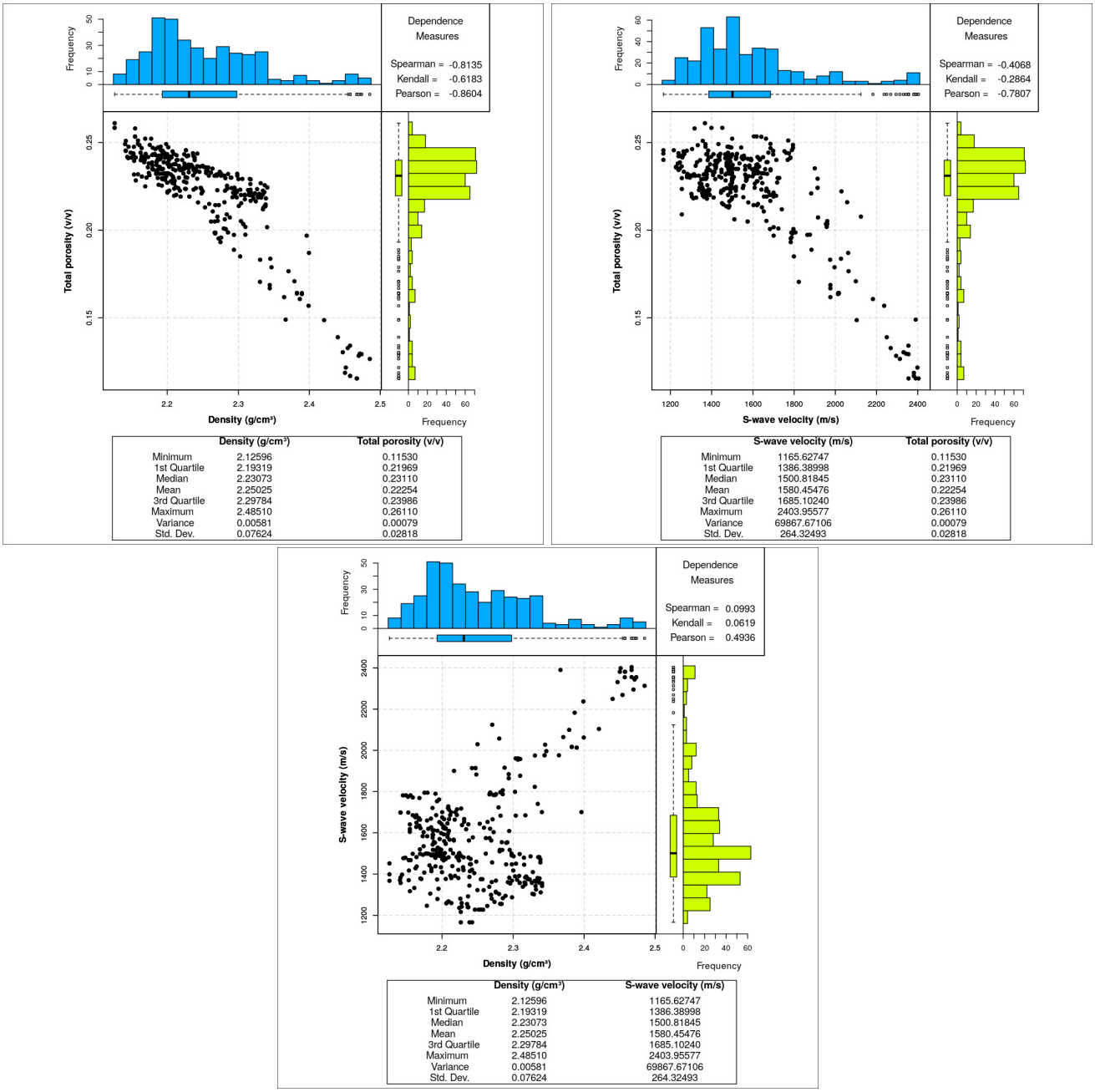


Figure 7.6: The scatterplots of the total porosity, density and S-wave velocity, and their correlation coefficients.

Figure 7.6 shows the scatterplot of the variables: total porosity, density, and the velocity of the S wave; their correlation coefficients: Spearman, Kendall, and Pearson; and its boxplot histrogram and its statistics. The advantage of the scatterplot is that the original units of the variables are shown, but the disadvantage is that the dependency pattern between the variables can be distorted by the different units, and by the existence of both uni and bi-distributional outlier values.

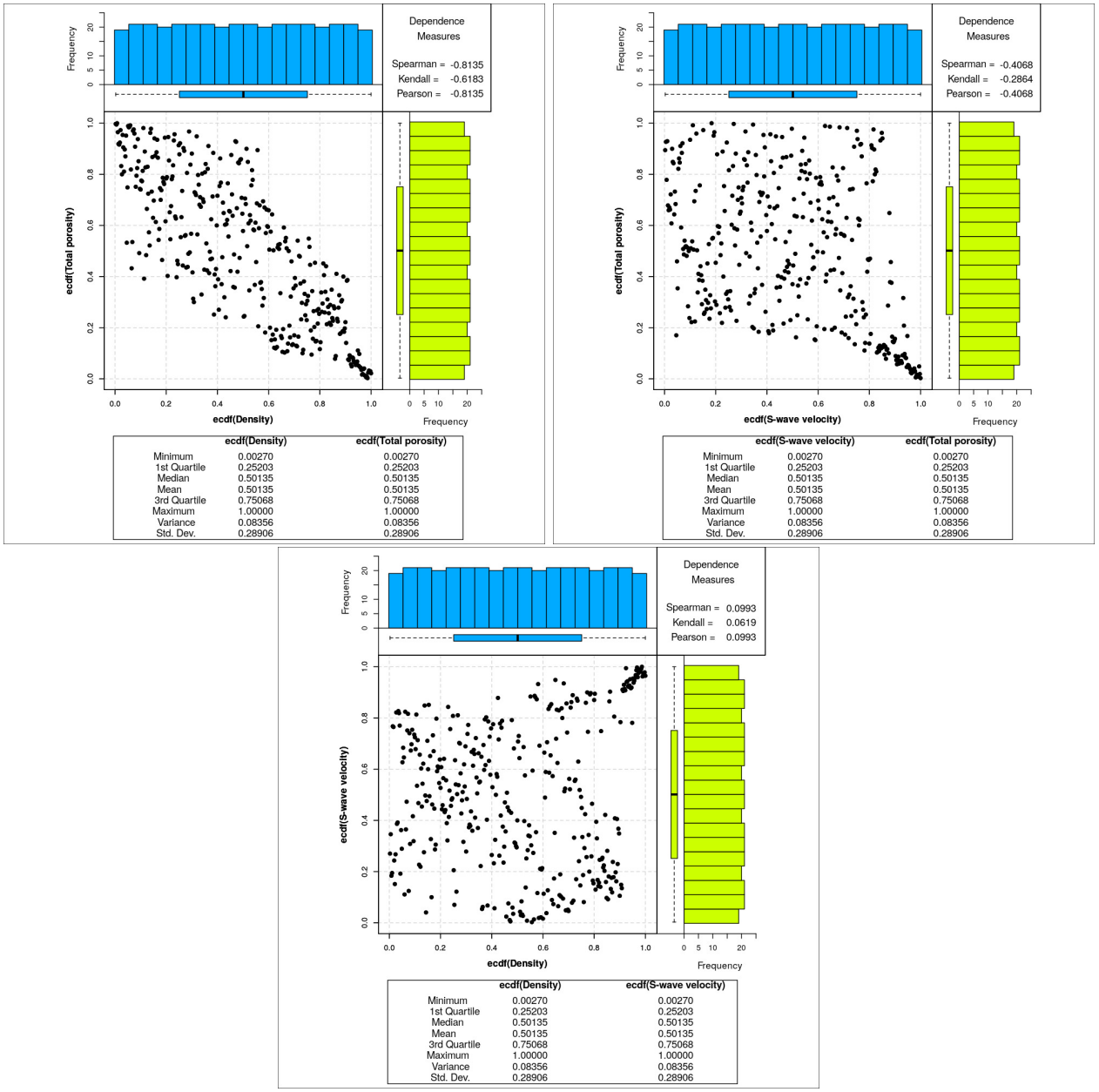


Figure 7.7: The pseudo-observations of the total porosity, density and S-wave velocity, and their correlation coefficients.

As explained in chapter 3 , the joint probability distribution function of more than two variables can be estimated using bivariate copulas and marginal functions. In this case there are three variables, therefore there are three bivariate copulas. Figure 7.7 shows the Pseudo-observations of the variables: total porosity, density, and the S-wave volocity, where it can be observed that the univariate distributions are now uniform. There are no uni and bi distributional outliers, and the dependency pattern does not depend on univariate distributions because those distributions are uniform.

Figure 7.6 and Figure 7.7 showed that the Spearman and Kendal correlation coefficients do not change when they are changed from scatterplot to Pseudo-observation, but the Pearson linear correlation coefficients changed, the abrupt change from 0.5 to 0.1 is observed in the case of density vs S-wave velocity. Which indicates that the Spearman and Kendall coefficients are more robust than Pearson. Figure 7.7 showed that the density and the velocity of the S wave have a low dependence and these two variables have a high and medium dependence on the total porosity, so the joint use of these two variables as secondary variable helps to predict the total porosity variable.

7.3 Spatial correlation analysis

Figure 7.8 shows the empirical semivariogram of the total porosity in the depth direction and its fitted semivariogram model, which is a spherical model with the nugget 0, the sill 0.00045 and the range 8 meters. That means that the shape of the spatial correlation of the total porosity in the depth direction is spherical with a spatial correlation range of 8 meters.

7.4 Copula-based dependency modeling

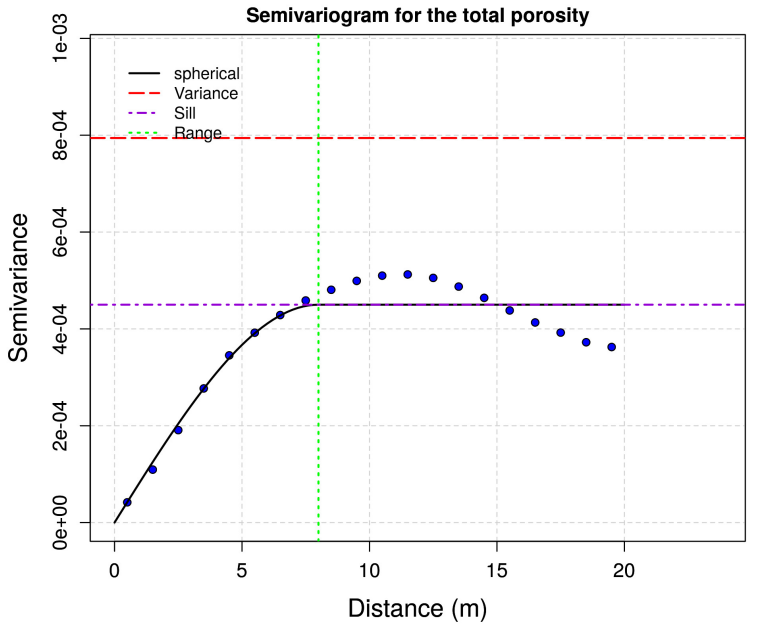
From the pseudo observations of total porosity, density, and S-wave velocity, the parametric copulas are adjusted and the copula $BB8\_270$ is obtained for the pair of variables S-wave valocity and total porosity, the copula $BB8$ for the pair of variables S-wave velocity and density, and the copula t for the pair of variable density and total porosity (Figure 7.9).

Figure 7.10 showed the 3D representation and contour map of the bivariate copula density functions of the variables total porosity, density, and S-wave velocity. Once the copula functions and the univariate functions of the random variables are known, then the joint distribution function of them can be known. Then, the simulations or realizations both conditional or unconditional can be done easily. The algorithms of the simulations and estimates can be seen in chapter 3.

7.5 Validation

Petrophysical properties are predicted based on copulas constructed using elastic seismic attributes as secondary variables. Then the predicted simulations and estimations will be compared with

the petrophysical properties of references.



Model	Nugget	Sill	Range	RMSE
spherical	0	0.00045	8	0.000042

Figure 7.8: Empirical semivariograms of the total porosity and its fitted semivariogram model.

```
tree   edge | family      cop    par    par2 |   tau   utd   ltd
-------------------------------------------------------------
   1    3,1 |     40  BB8_270  -2.15  -1.00 | -0.39     -     -
        3,2 |     10      BB8   1.48   1.00 |  0.21     -     -
   2  2,1;3 |      2        t  -0.80   8.21 | -0.59  0.00  0.00
---
type: C-vine    logLik: 332.26    AIC: -652.51    BIC: -629.03
---
1 <-> ecdf(Total porosity),   2 <-> ecdf(Density),   3 <-> ecdf(S-wave velocity)
```

Figure 7.9: Summary of the vine parametric copula model fitted to the empirical values of the cumulative distribution function of total porosity, density, and S-wave velocity.

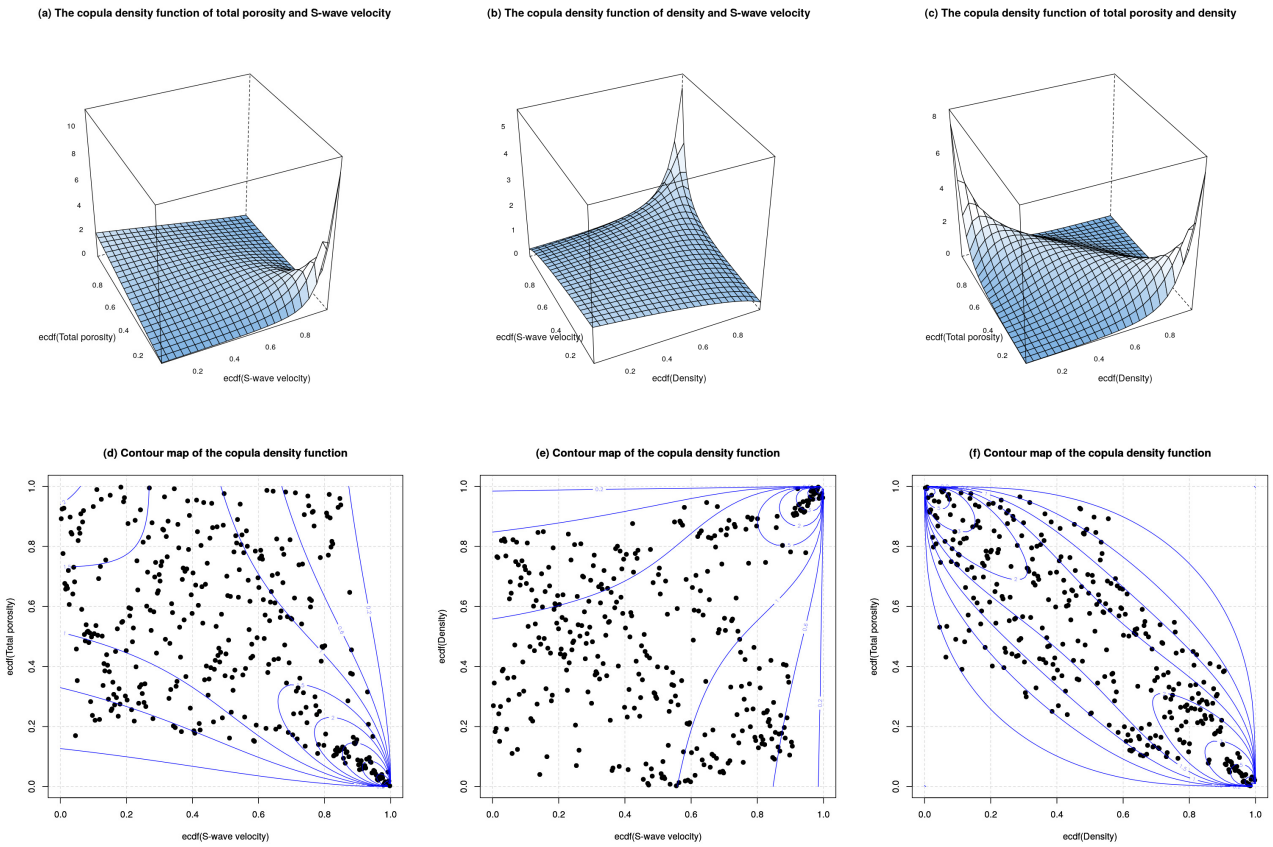


Figure 7.10: 3D representation and contour map of the bivariate copula density functions of the empirical values of the cumulative distribution function of total porosity, density, and S-wave velocity.

7.5.1 Simulations

For reproducing the joint and univariate probability distribution functions of the petrophysical properties and the elastic seismic attributes, the algorithms of chapter 3 are applied. The global optimization method (simulated annealing) is applied to reproduce the spatial correlation (the semivariogram model) of the petrophysical properties.

Left figures of Figure 7.11, Figure 7.12, Figure 7.13, Figure 7.14, and Figure 7.15 show the original data and the right figures show the simulations, where it was observed that the univariate statistics of the petophysical properties, the dependency coefficients and the dependency pattern between the properties petrophysical and elastic seismic attributes are very similar. Which indicate that the univariate and dependency models are quite good and represent the distributional behaviors of the random variables.

Figure 7.16, Figure 7.17, Figure 7.18, Figure 7.19, and Figure 7.20 show the spatial distribution of the original and simulated petrophysical properties, the errors or differences between them,

their experimental semivariograms, and their semivariogram models. Which indicate that the semivariogram models are reproduced. Table 7.3 shows the statistics of the errors.

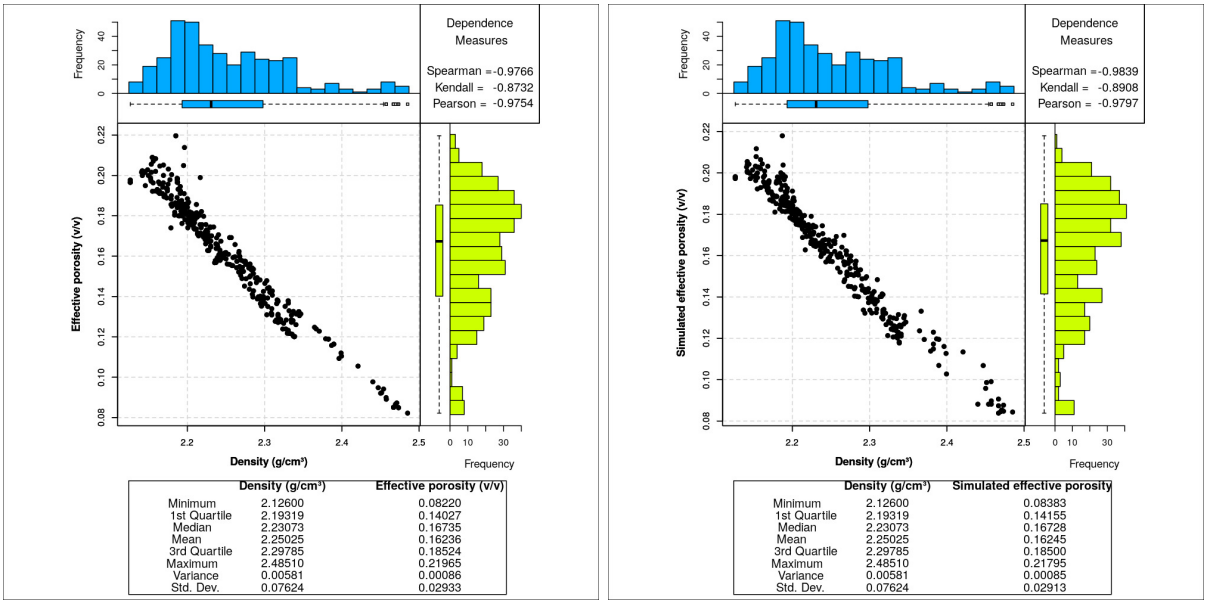


Figure 7.11: Left figure shows the scatterplot of effective porosity and density; Right figure presents scatterplot of simulated effective porosity and density, and their dependency coefficients.

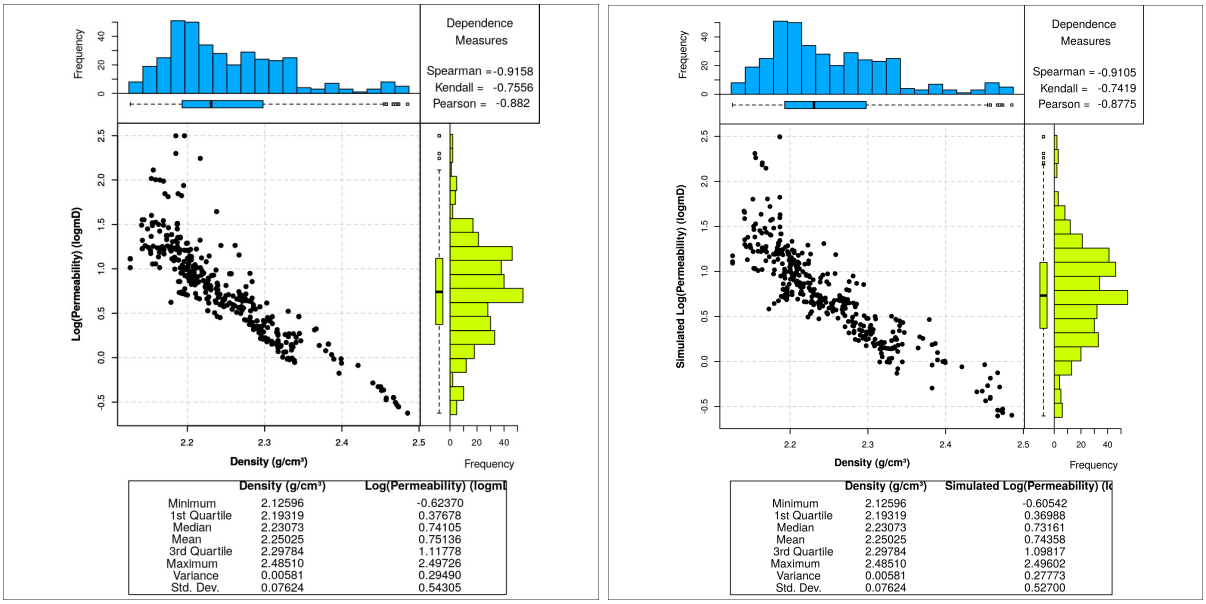


Figure 7.12: Left figures show the scatterplot of log(Permeability) and density; right figures present scatterplot of simulated log(Permeability) and density, and their dependency coefficients.

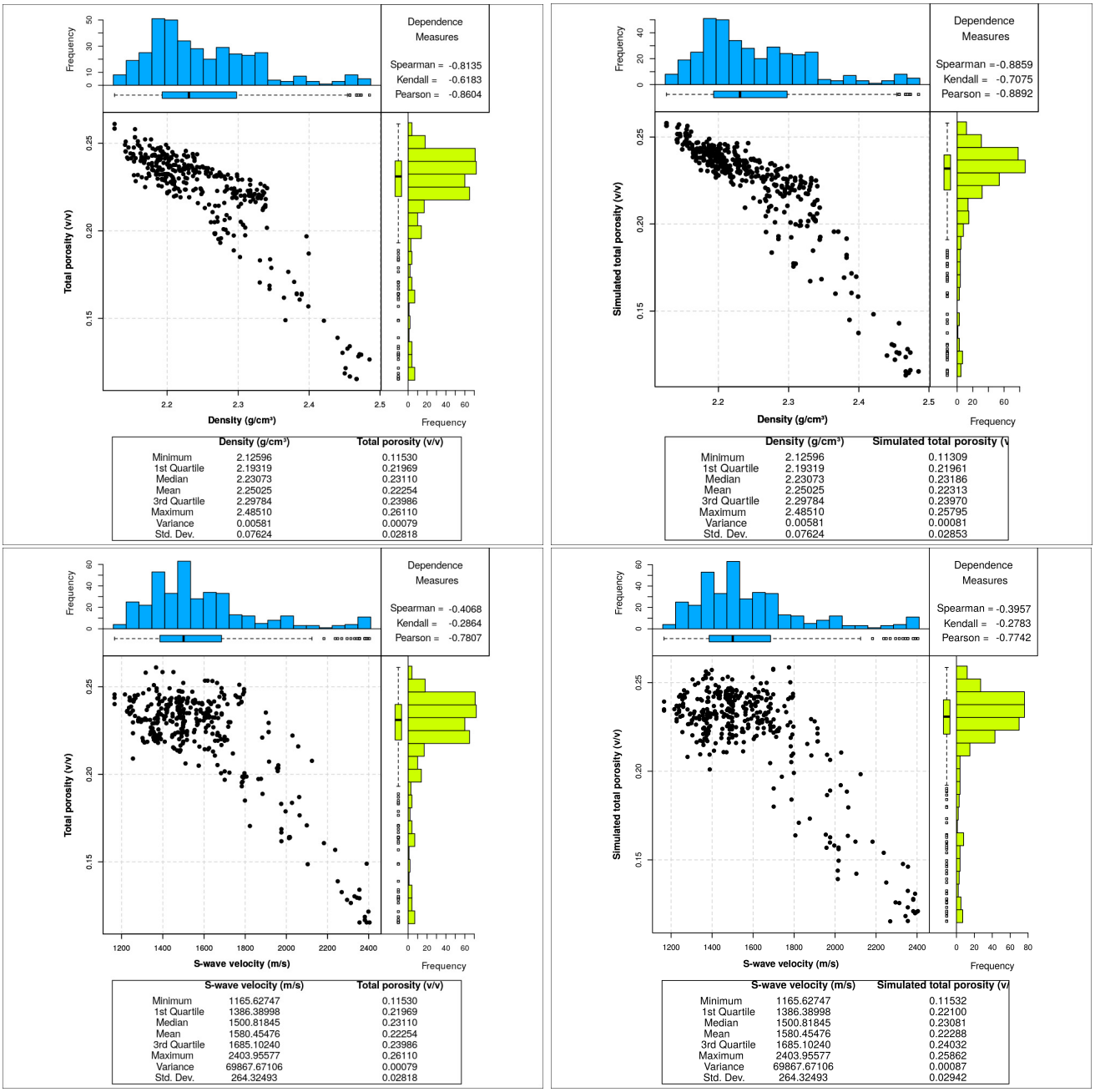


Figure 7.13: Left figures show the scatterplot of total porosity, S-wave velocity and density; right figures present scatterplot of simulated total porosity, S-wave velocity and density, and their dependency coefficients.

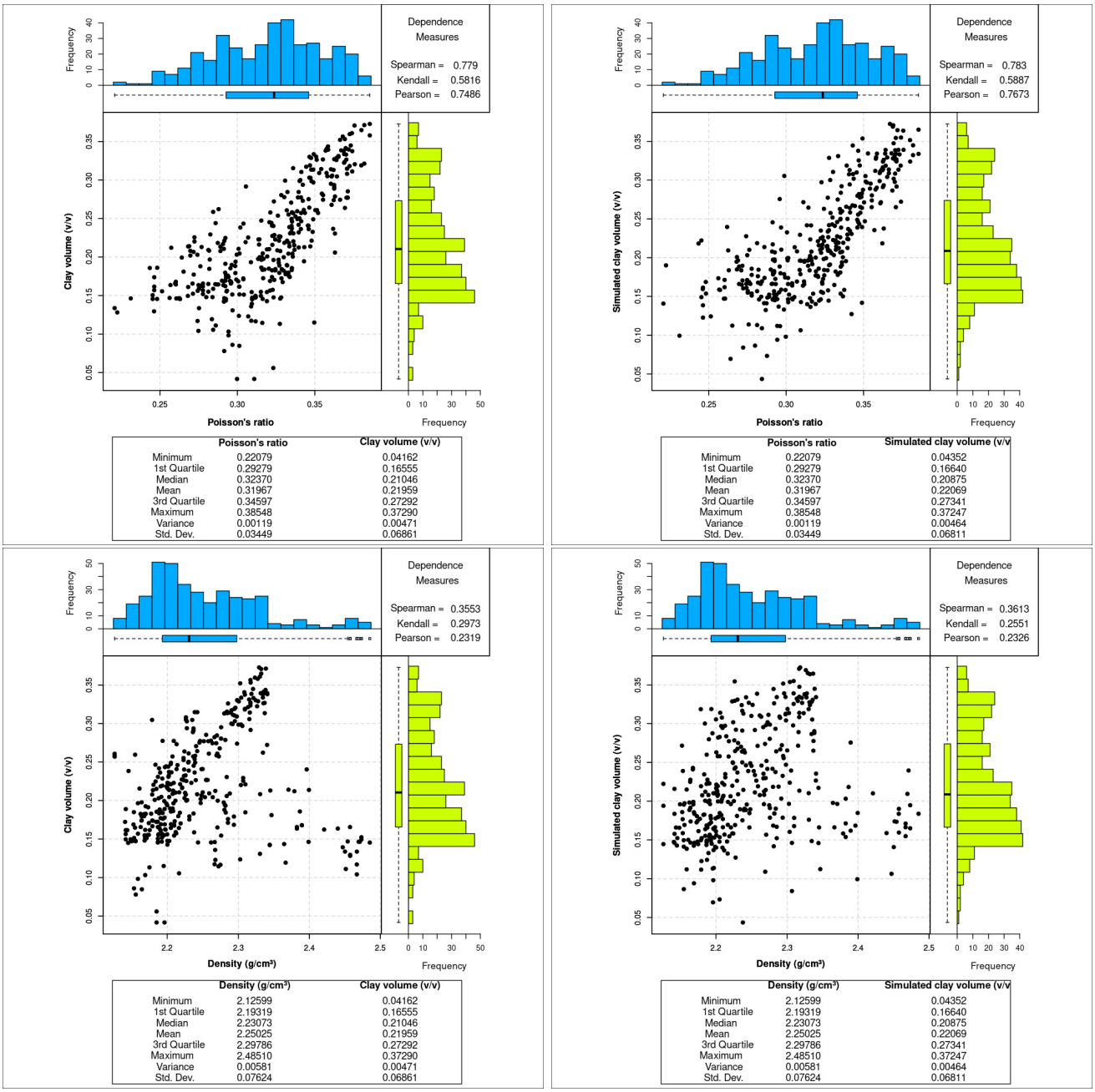


Figure 7.14: Left figures show the scatterplot of clay volume, Poisson's ratio and density; right figures present scatterplot of simulated clay volume, Poisson's ratio and density, and their Spearman correlation.

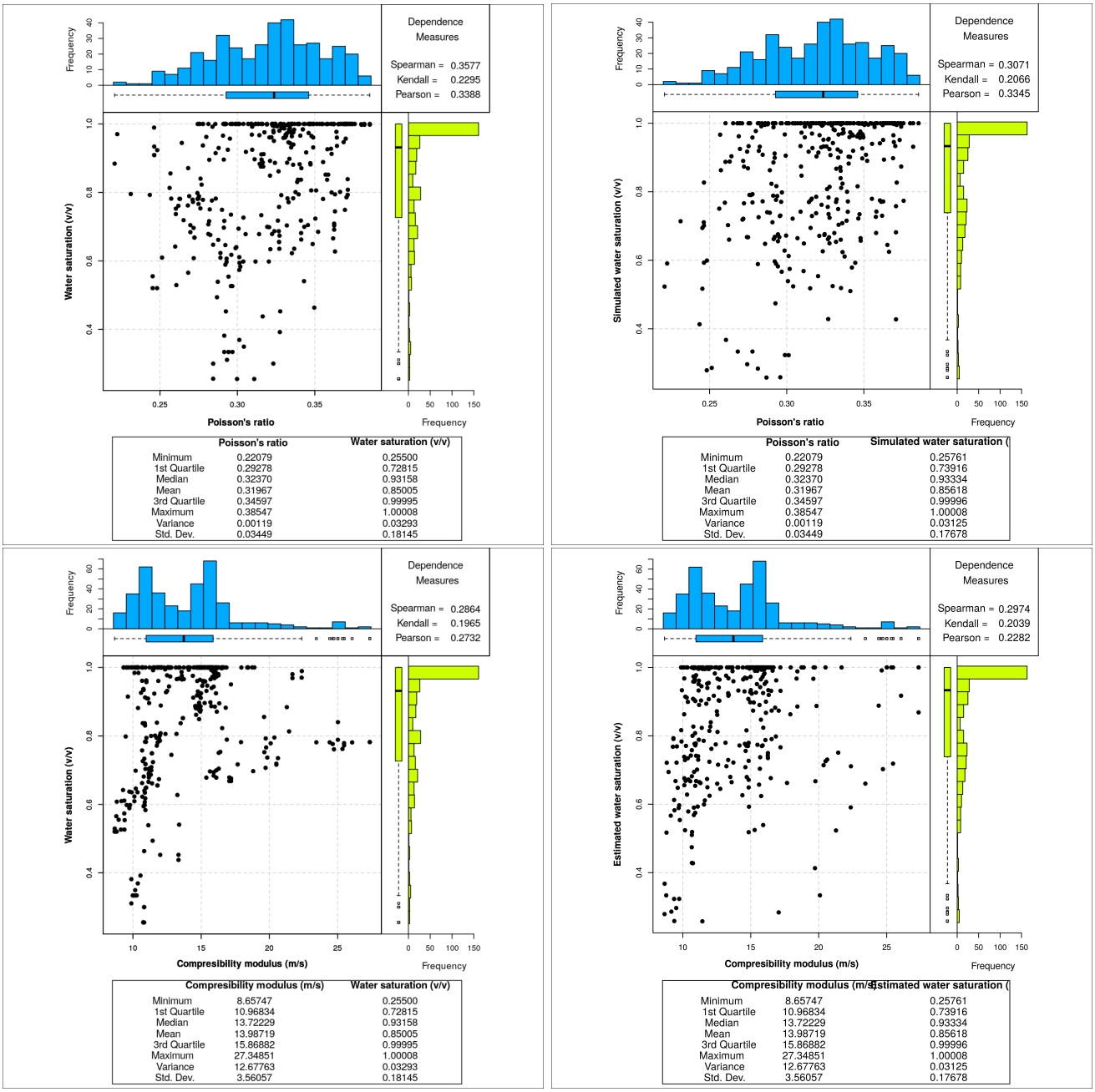


Figure 7.15: Left figures show the scatterplot of water saturation, Poisson's ratio and compressibility modulus; right figures present scatterplot of simulated water saturation, Poisson's ratio and compressibility modulus, and their Spearman correlation.

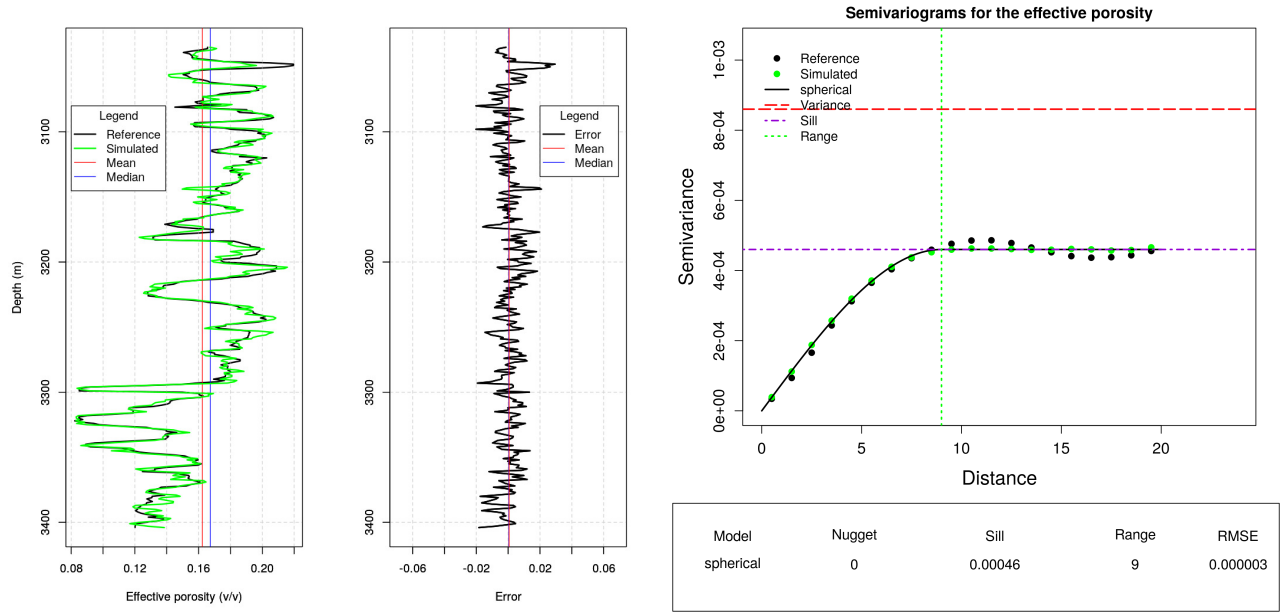


Figure 7.16: Spatial distribution, empirical semivariograms, and the fitted semivariogram model of the reference effective porosity (black color) and the corresponding simulated (green color).

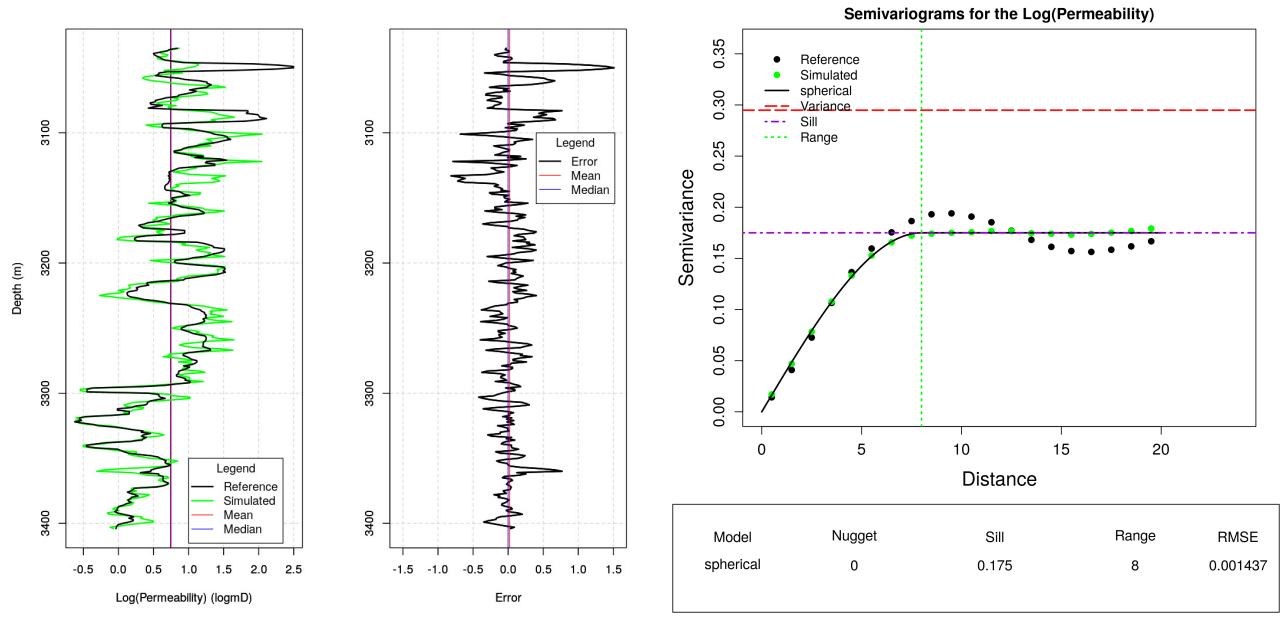


Figure 7.17: Spatial distribution, empirical semivariograms, and the fitted semivariogram model of the reference log(Permeability) (black color) and the corresponding simulated (green color).

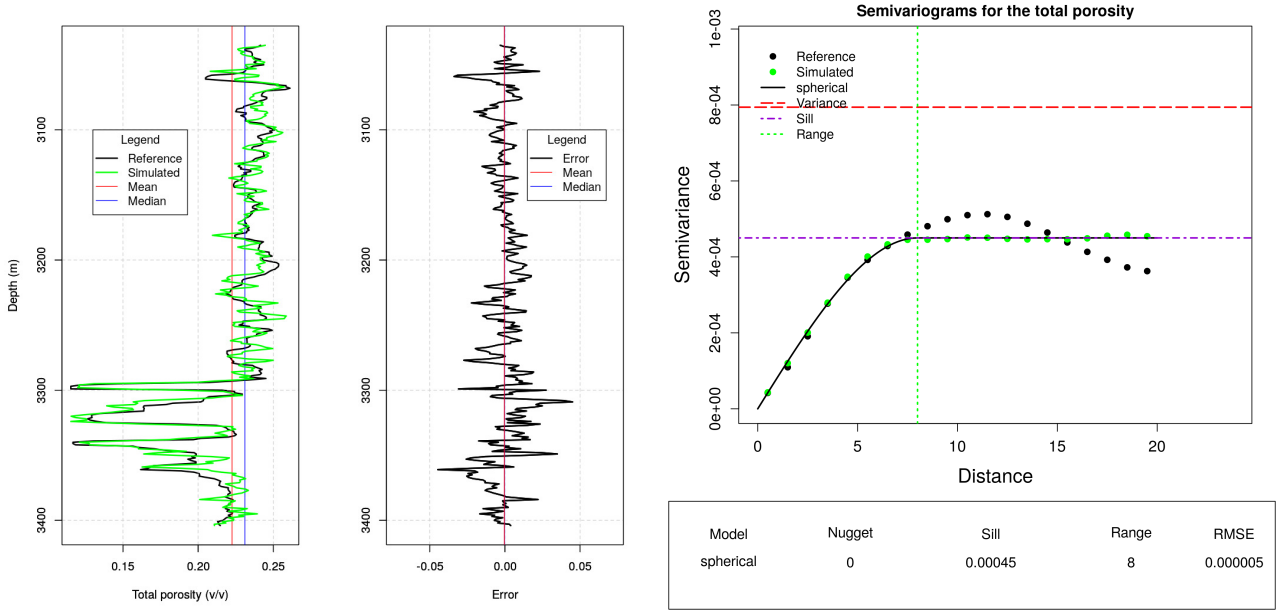


Figure 7.18: Spatial distribution, empirical semivariograms, and the fitted semivariogram model of the reference total porosity (black color) and the corresponding simulated (green color).

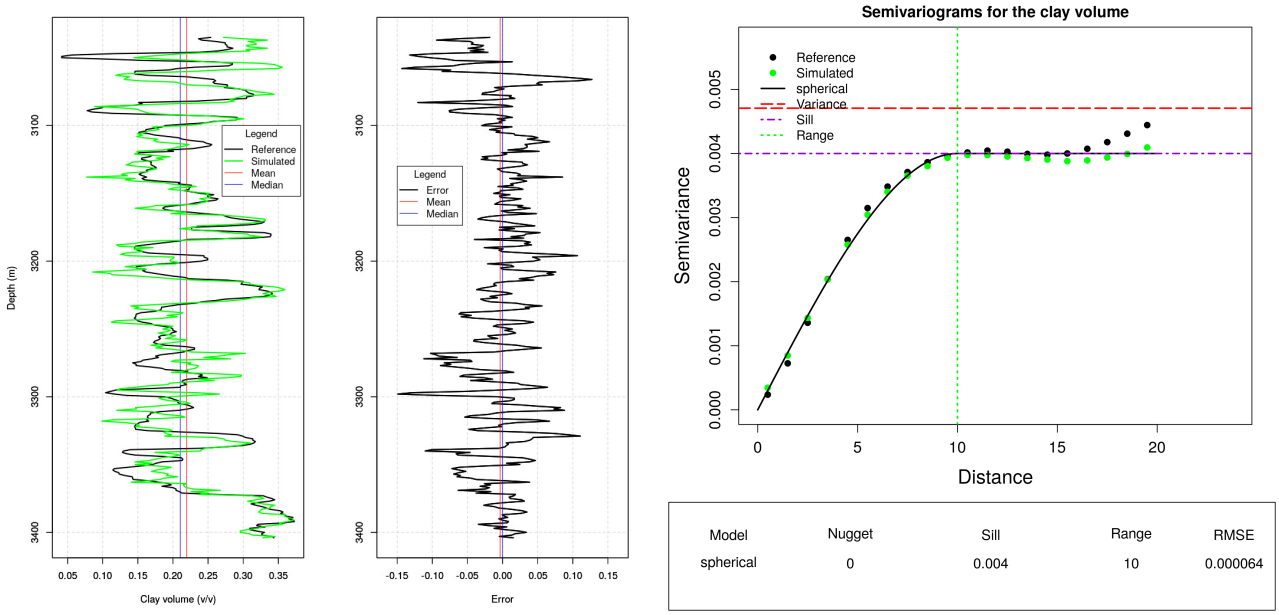


Figure 7.19: Spatial distribution, empirical semivariograms, and the fitted semivariogram model of the reference clay volume (black color) and the corresponding simulated (green color).

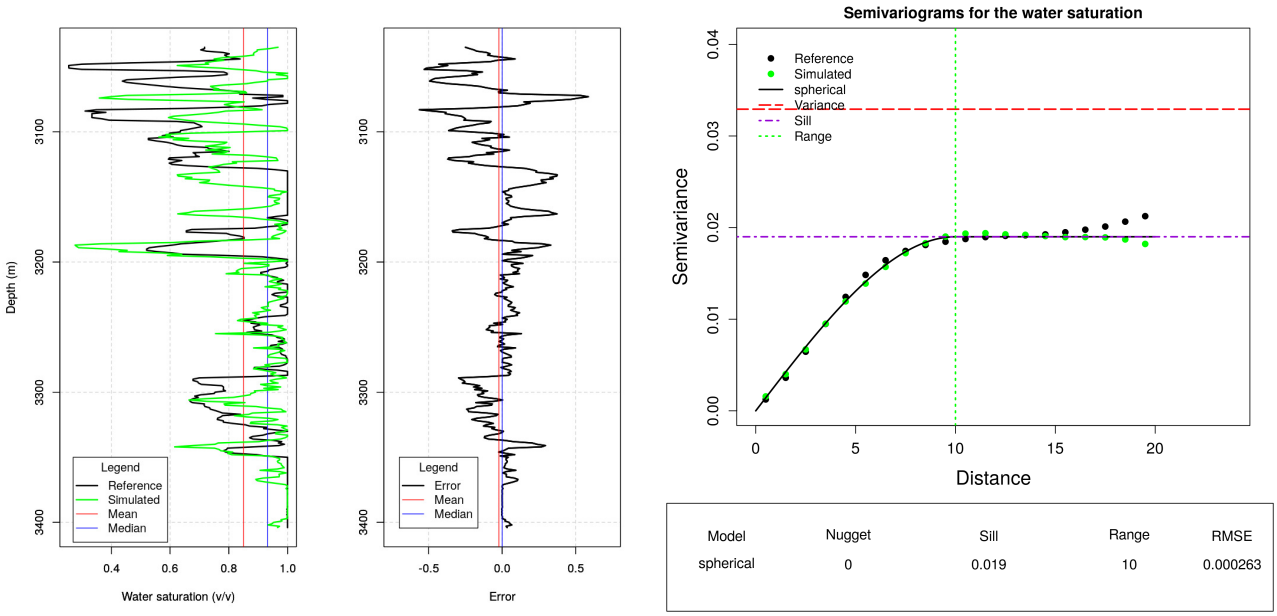


Figure 7.20: Spatial distribution, empirical semivariograms, and the fitted semivariogram model of the reference water saturation (black color) and the corresponding simulated (green color).

Table 7.3: Statistics of errors.

Statistics	Total porosity	log(Permeability)	Effective porosity	clay volume	Water saturation
Observation number	370	370	370	370	370
Minimum	-0.0447	-0.8158	-0.0202	-0.1590	-0.5639
1st. Quartile	-0.0067	-0.1218	-0.0041	-0.0267	-0.1019
Median	-0.0002	-0.0008	0.0004	-0.0002	0.0005
Mean	-0.0003	0.0215	0.0006	-0.0033	-0.0219
3rd. Quartile	0.0063	0.1307	0.0052	0.0259	0.0503
Maximum	0.0450	1.5120	0.0295	0.1272	0.5869
Rank	0.0897	2.3278	0.0497	0.2762	1.1507
Interquartile Rank	0.0130	0.2525	0.0093	0.0526	0.1522
Variance	0.0001	0.0769	0.0001	0.0020	0.0297
Standard Deviation	0.0110	0.2773	0.0075	0.0444	0.1723
Variation Coeff.	-32.2576	12.8913	12.1171	-13.2862	-7.8690
Skewness	-0.0357	1.2952	0.3115	-0.4218	-0.1462
Kurtosis	4.8064	8.9487	3.9704	3.7263	4.3762
RMSE	0.0110	0.2778	0.0075	0.0444	0.1735

7.5.2 Estimations

Estimation of petrophysical properties applying quantile regression method based on D-vine copula using elastic seismic attributes as secondary variables is presented.

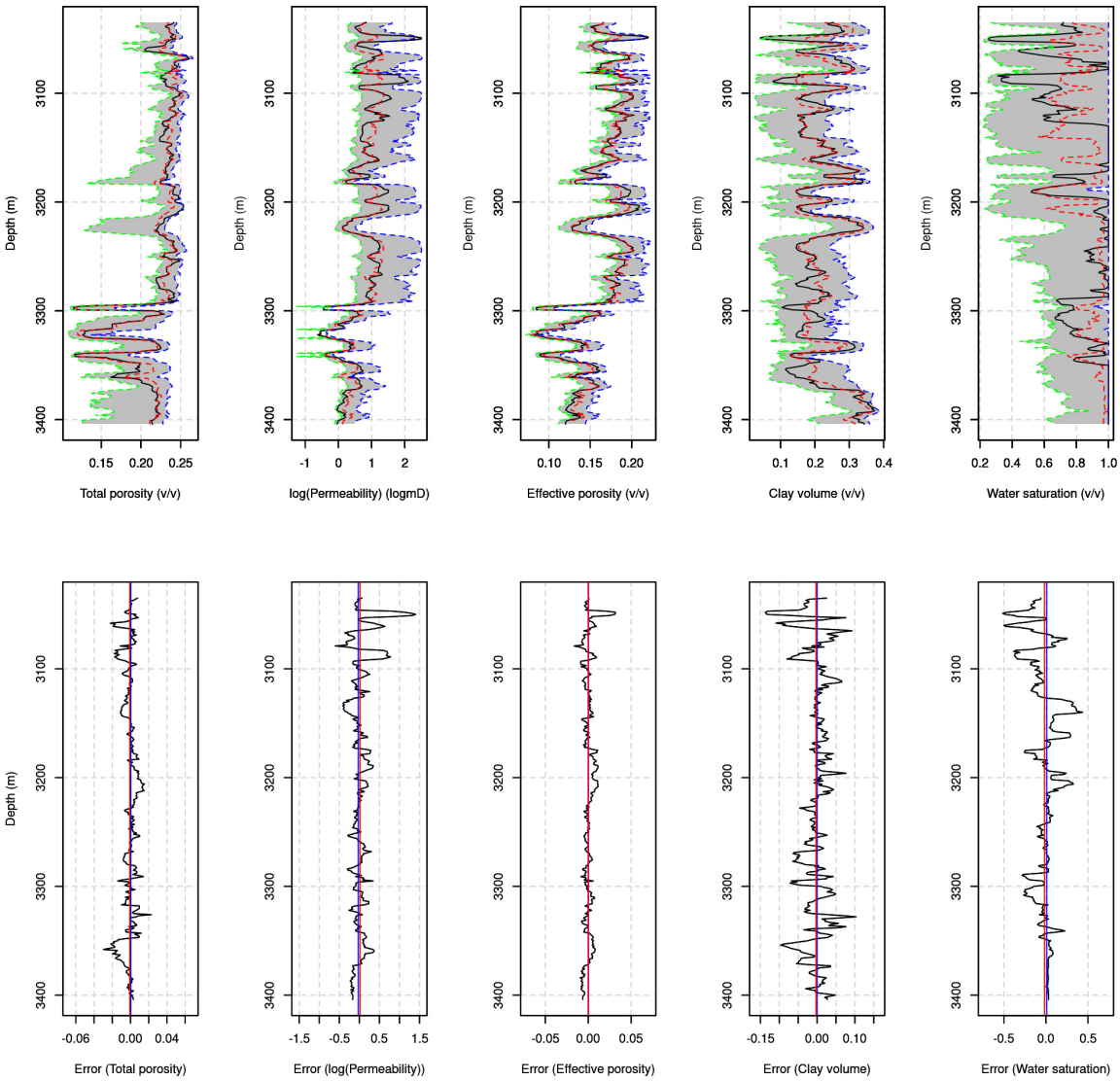


Figure 7.21: Upper figure show spatial distribution of the reference petrophysical properties (solid black line) and the median (50th percentile) (dashed red line), the 10th percentile (dashed green line), 90th percentile (dashed blue line) of the quantile regression, and grey regions represent the 90% confidence interval. Lower figure shows the difference between the reference and the median or errors (black line), the mean (red line) and the median (blue line) of the errors.

Figure 7.21 shows the quantile regression curves as the median, 10th percentile, 90th percentile, and the differences between the originals and the median curve. Table 7.4 shows the error statistics and Figure 7.22 shows the boxplot histograms of the errors. Figure 7.23 shows the semivariograms of the estimated petrophysical properties that do not follow the semivariogram model. This is logical because the estimation approach does not require reproducing the spatial variability of the random variables.

Table 7.4: Statistics of errors.

Statistics	Total porosity	log(Permeability)	Effective porosity	Clay volume	Water saturation
Observation number	370	370	370	370	370
Minimum	-0.0297	-0.6006	-0.0165	-0.1369	-0.5155
1st. Quartile	-0.0046	-0.1316	-0.0036	-0.0196	-0.0925
Median	0.0004	-0.0280	-0.0000	0.0000	0.0094
Mean	-0.0007	0.0167	0.0004	0.0029	-0.0179
3rd. Quartile	0.0037	0.1130	0.0039	0.0186	0.0444
Maximum	0.0229	1.4016	0.0317	0.1030	0.4368
Rank	0.0527	2.0022	0.0481	0.2400	0.9523
Interquartile Rank	0.0083	0.2445	0.0075	0.0383	0.1370
Variance	0.0001	0.0619	0.0000	0.0013	0.0233
Standard Deviation	0.0074	0.2488	0.0059	0.0355	0.1528
Variation Coeff.	-10.0550	14.9322	13.2447	-12.2270	-8.5217
Skewness	-0.6676	2.0391	1.4488	-0.6597	-0.2558
Kurtosis	3.9927	10.4498	8.5187	4.5934	4.2991
RMSE	0.0074	0.2490	0.0059	0.0356	0.1536

Figure 7.21 upper gives a result of the prediction through the median and also allows an analysis of uncertainty through the range of 10th and 90th quantile. That's valuable because it not only gives the most probable value, but also the range of uncertainty associated with that value.

It is observed that the spatial distribution of the medians of the prediction and the references are quite similar. As the dependencies between the petrophysical properties and the elastic seismic attributes are greater, the similarity between the median of the prediction and the reference is greater, e.g. the case of effective porosity, and vice versa, e.g. the case of water saturation in Figure 7.21.

Table 7.4 shows that the errors of log(Permeability) and effective porosity present high asymmetry, the errors of total porosity and clay volume present low asymmetry and the water saturation error presents very low asymmetry. Figure 7.22 shows that the errors for all regressions have high and low distributional outliers. Figure 7.23 shows that in general the medians of the regressions present the spatial variability less than the original spatial variability, in other words the spatial variability is underestimated. Which is once again confirming that the estimation approach does not reproduce the spatial variability of the random variable and underestimating it.

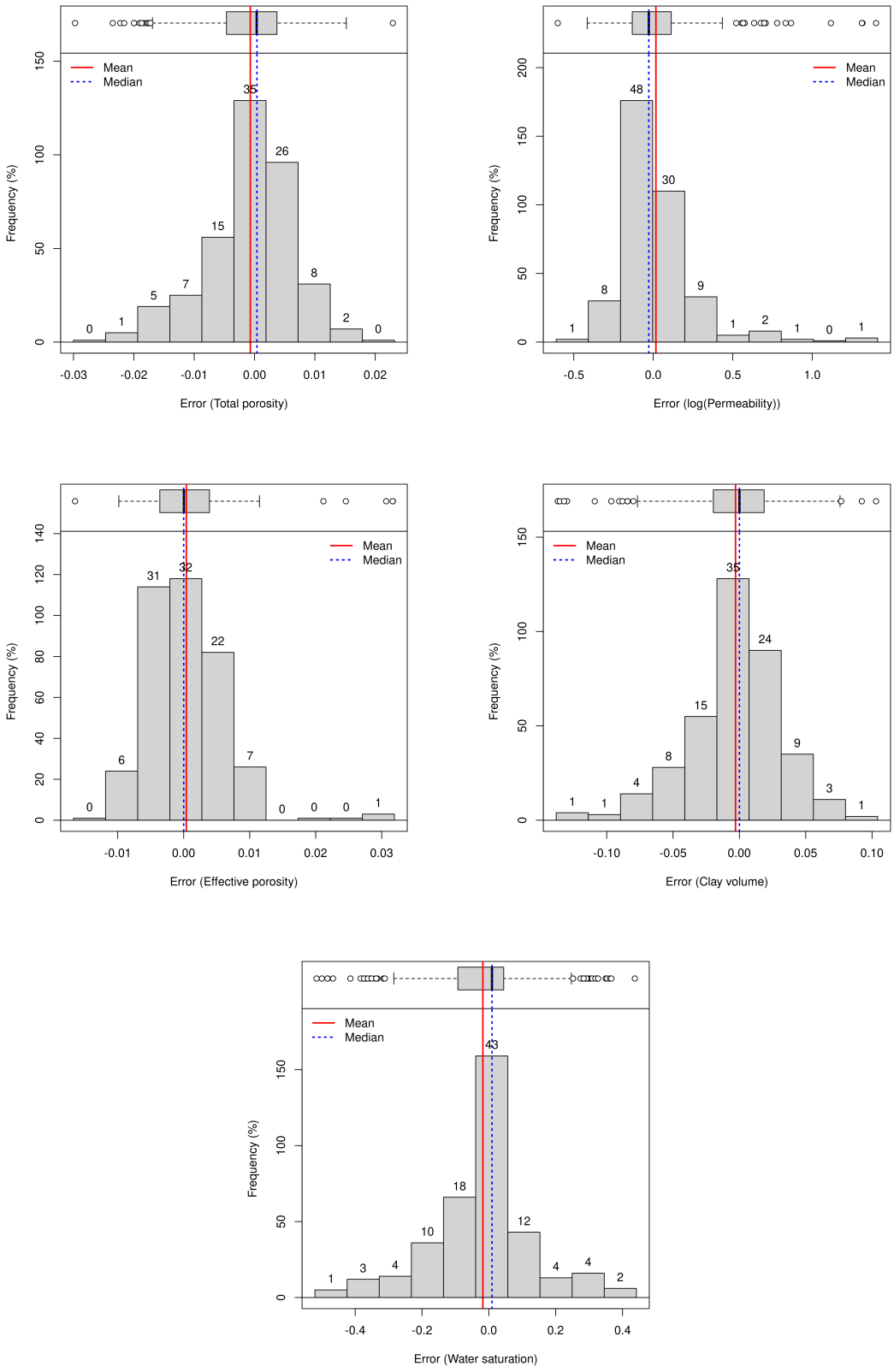


Figure 7.22: Histograms of the errors.

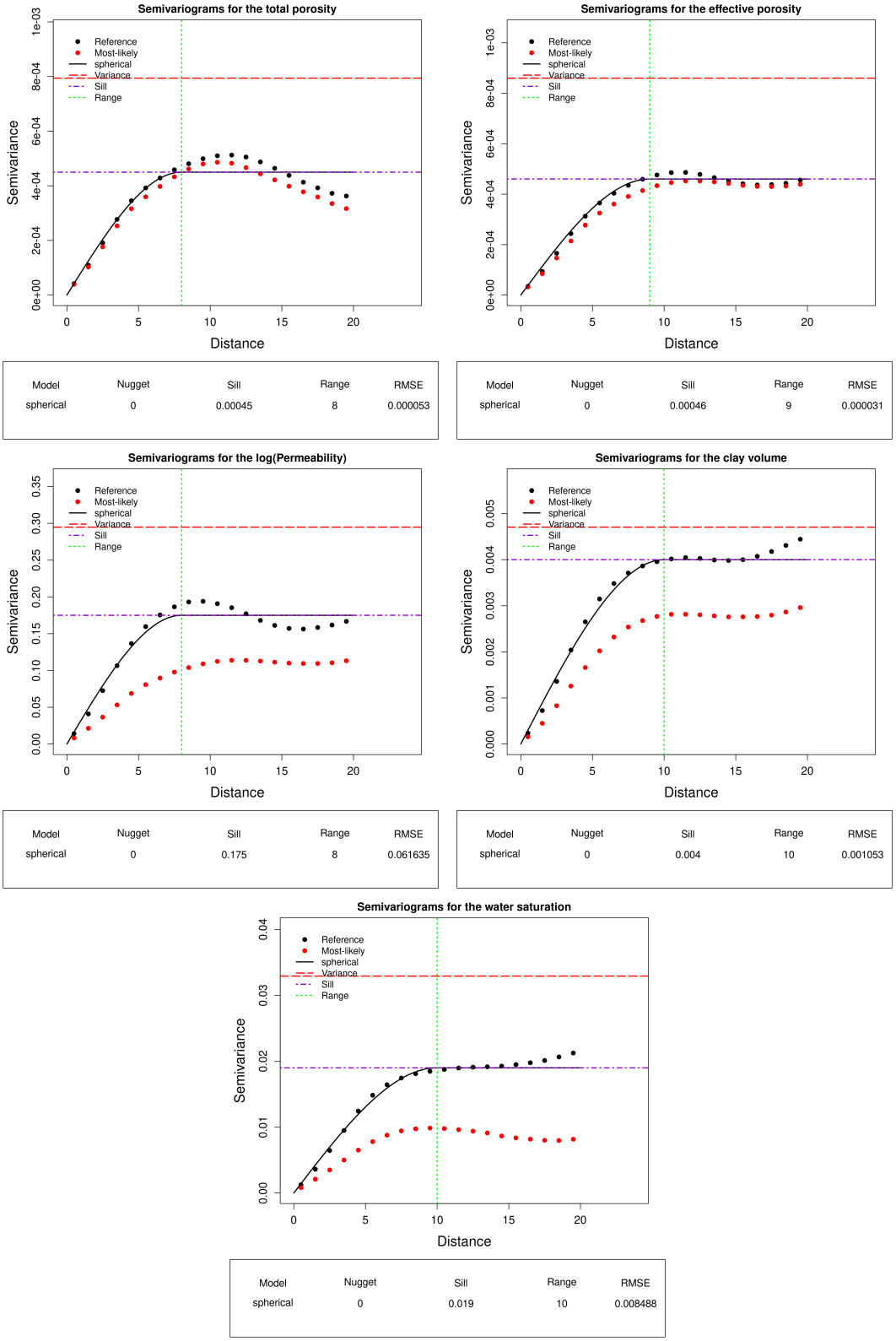


Figure 7.23: The experimental semivariogram of petrophysical properties: original in black points, estimated in red points, and its fitted semivariogram model (solid line).

7.6 Application

The dependency model built at the well log scale between the petrophysical properties and the elastic seismic attributes is applied to the seismic data in the 2D inline section that follows the west-east direction. According to the conceptual geological model, this direction presents the folds and symmetrical anticline. The elastic seismic attributes that are used as secondary variables are the same secondary variables of the built dependency model at well log scale. Figure 7.24, Figure 7.25, Figure 7.26, and Figure 7.27 show the attributes which are the S-wave velocity, the density, the Poisson's ratio and the compressibility modulus. Each section contains 2,112 cells: 64 vertical cells (384 meters) and 33 horizontal cells (412.5 meters). A cell represents approximately 6 meters vertically and 12.5 meters horizontally.

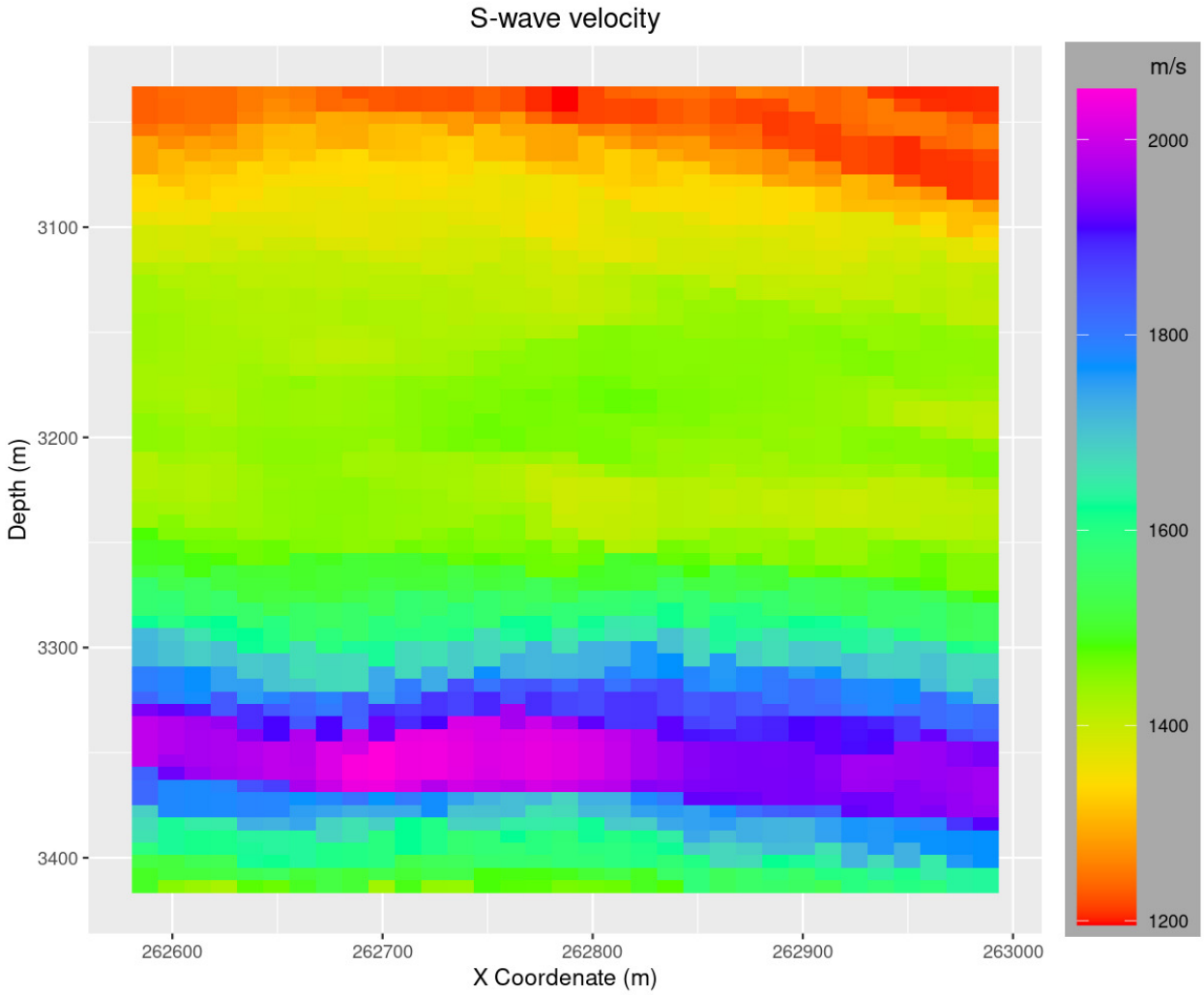


Figure 7.24: 2D seismic section of S-wave velocity.

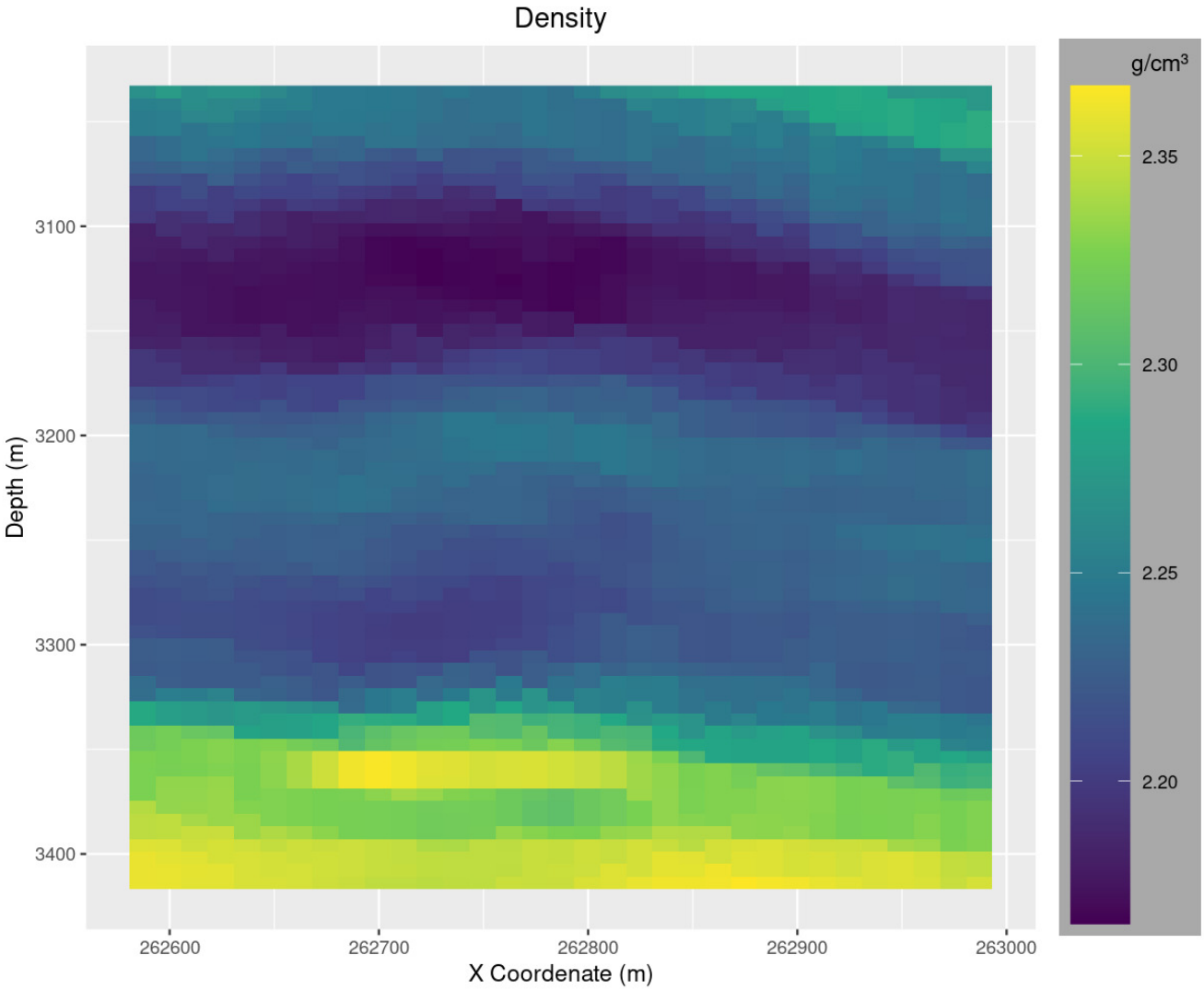


Figure 7.25: 2D seismic section of density.

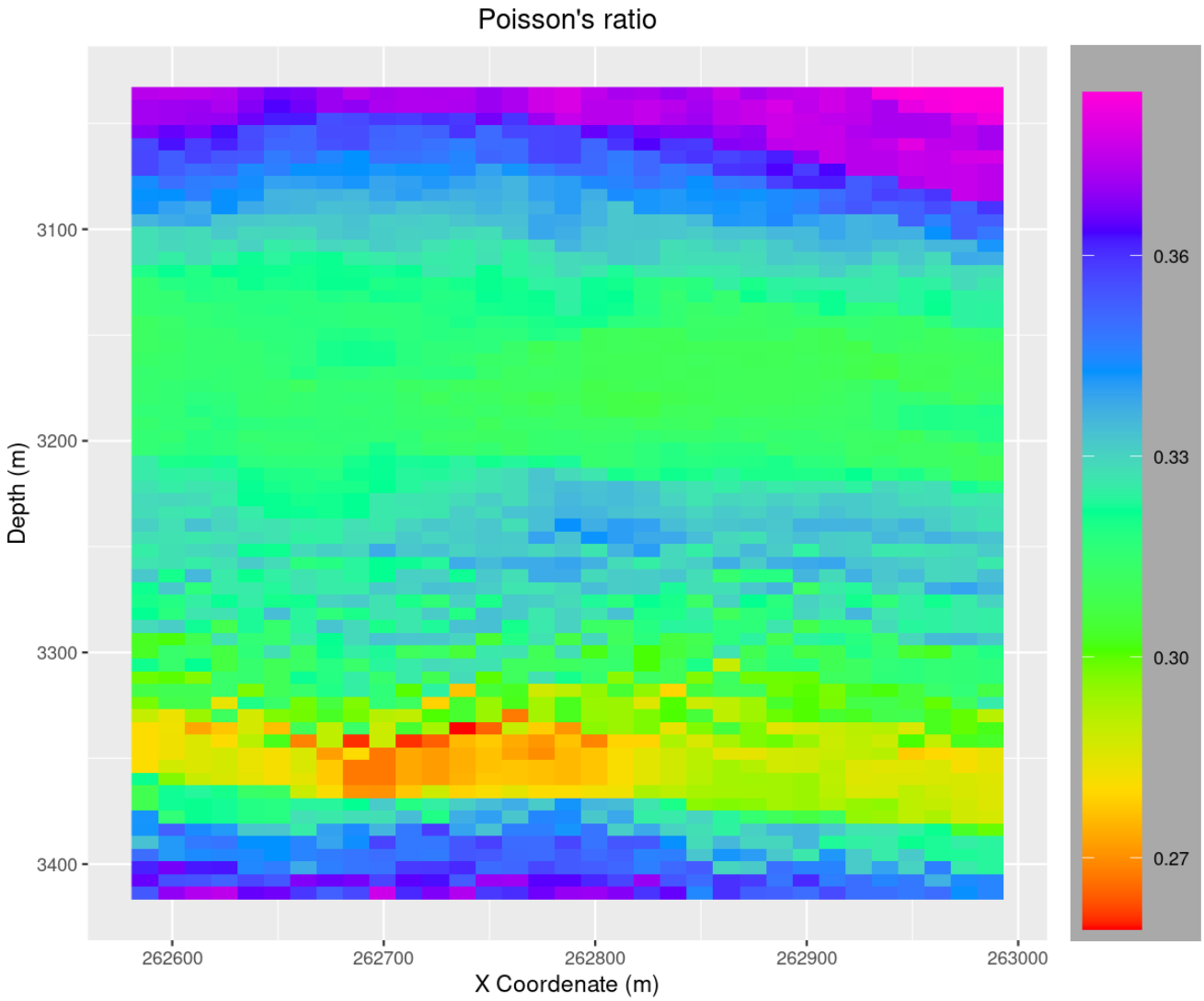


Figure 7.26: 2D seismic section of Poisson's ratio.

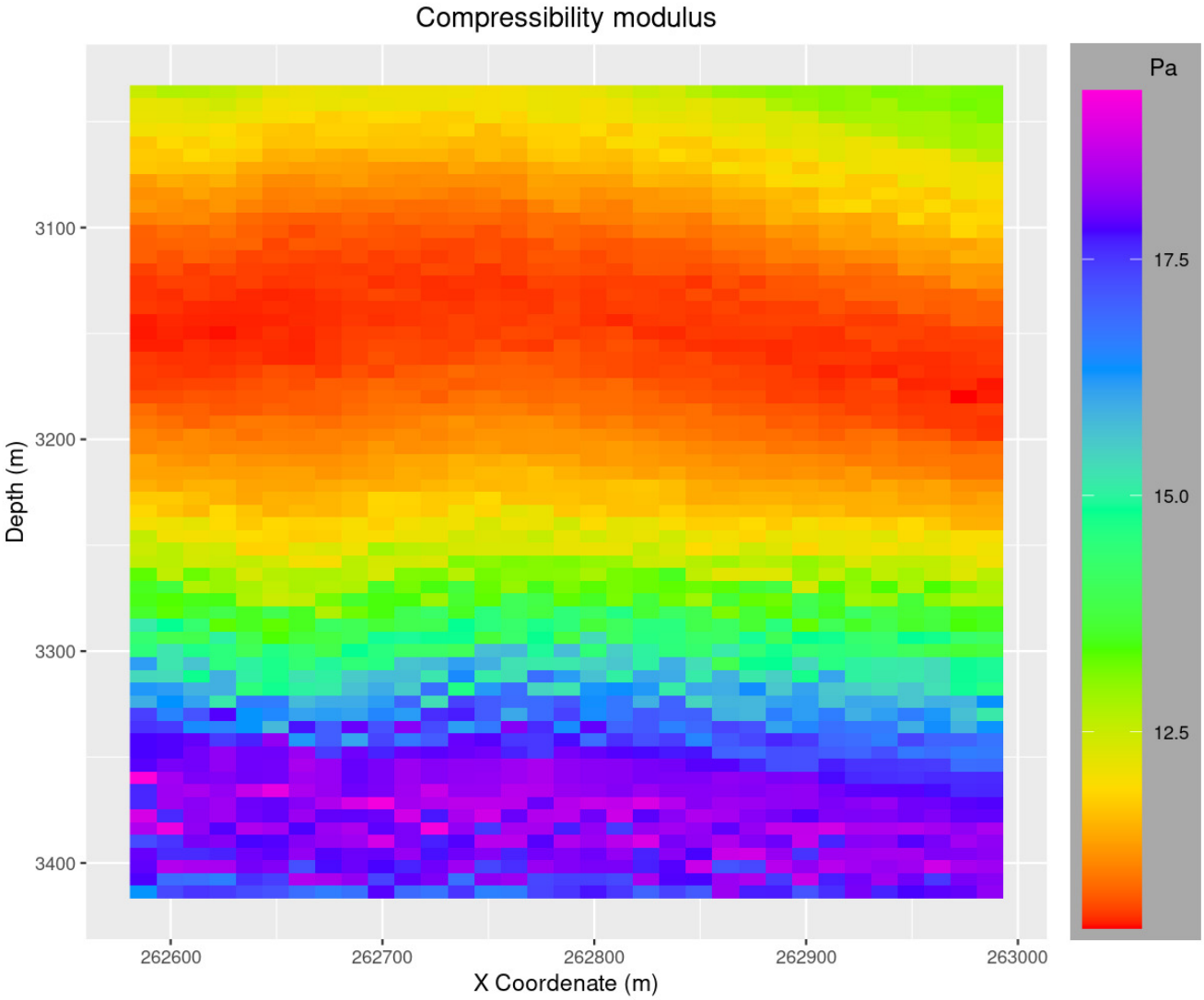


Figure 7.27: 2D seismic section of compressibility modulus.

Figure 7.28 to Figure 7.47 show the simulation, the mean (the most likely value), the median, and the uncertainty range of the predicted petrophysical properties. Which can answer questions such as spatial variability, the most likely value and the associated uncertainty in each spatial position in the section.

Figure 7.28 shows a simulation of the total porosity where the spatial variability of the variable is reproduced. The upper zone shows the high values, the lower zone shows the low values, and the middle zone shows a mixture of both. Figure 7.29 and Figure 7.30 show the most likely and median values of the total porosity where it can be seen that there are two zones: an upper middle zone with values between [0.225,0.245] and a lower zone with values between [0.165,0.225], and

there are no very high values of [0.245,0.261] and very low [0.115,0.165] as in the simulation. This confirms once again that the results of the estimation approach are smoothed out and do not reproduce spatial variability. Figure 7.31 shows the uncertainty range that is calculated by the difference between 10th and 90th percentile. It can be seen that the area that has high values of total porosity presents less uncertainty (upper middle area), and the area that presents low values of the same presents higher uncertainty (lower area). An interpretation from the four figures is the upper middle zone can be a sandstone and the lower zone can be a shale.

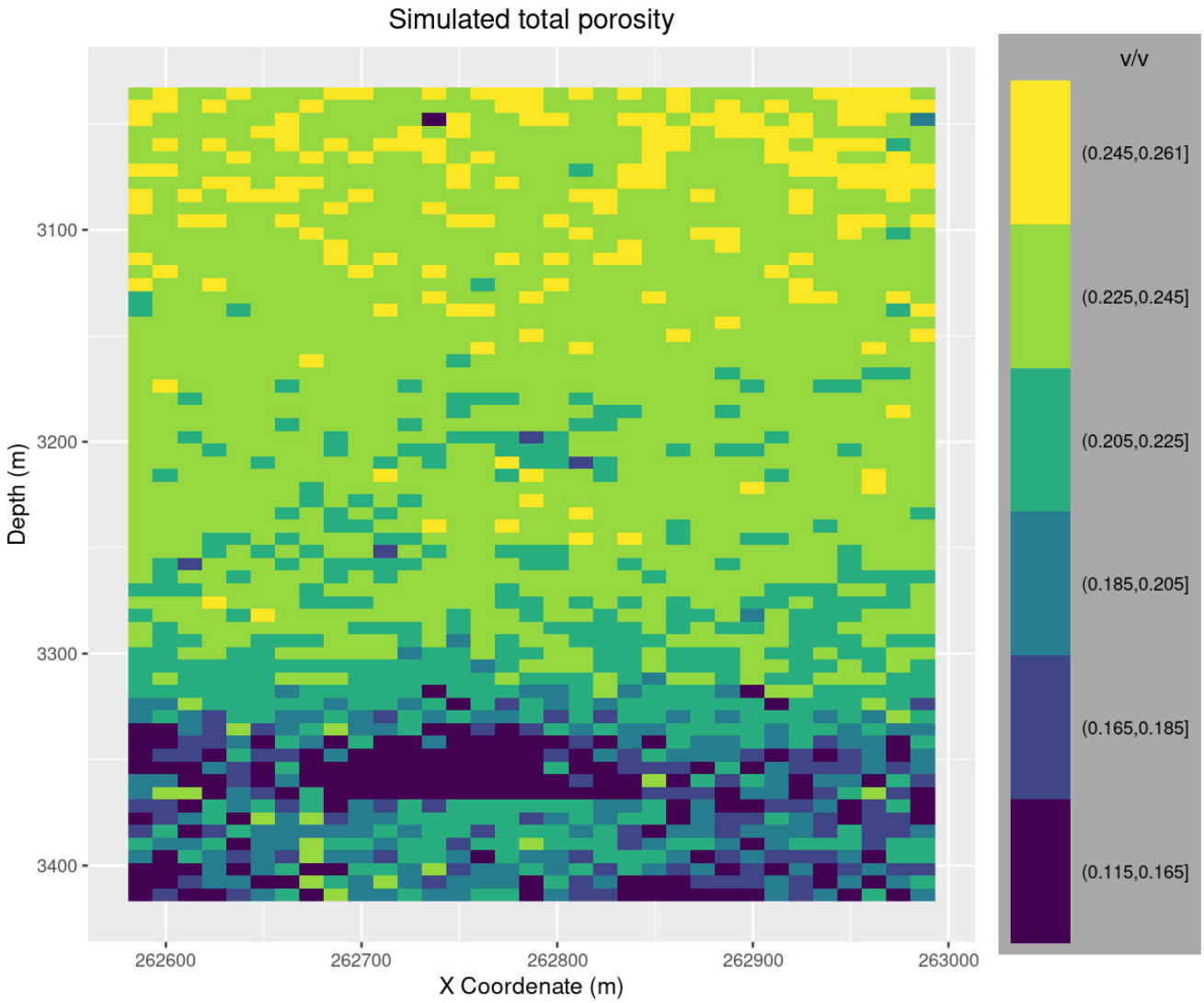


Figure 7.28: 2D seismic sections of the simulated total porosity.

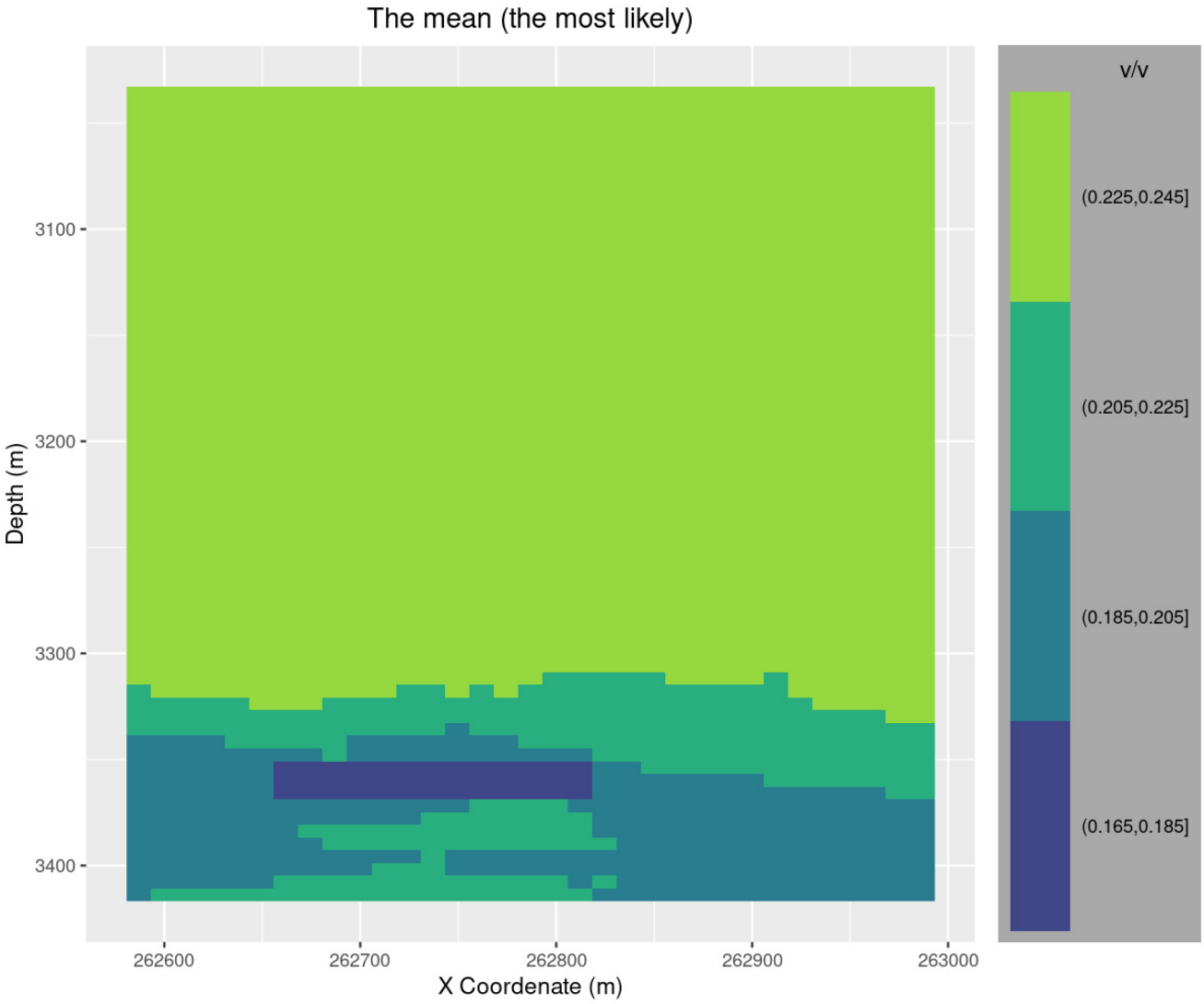


Figure 7.29: 2D seismic sections of the predicted mean total porosity.

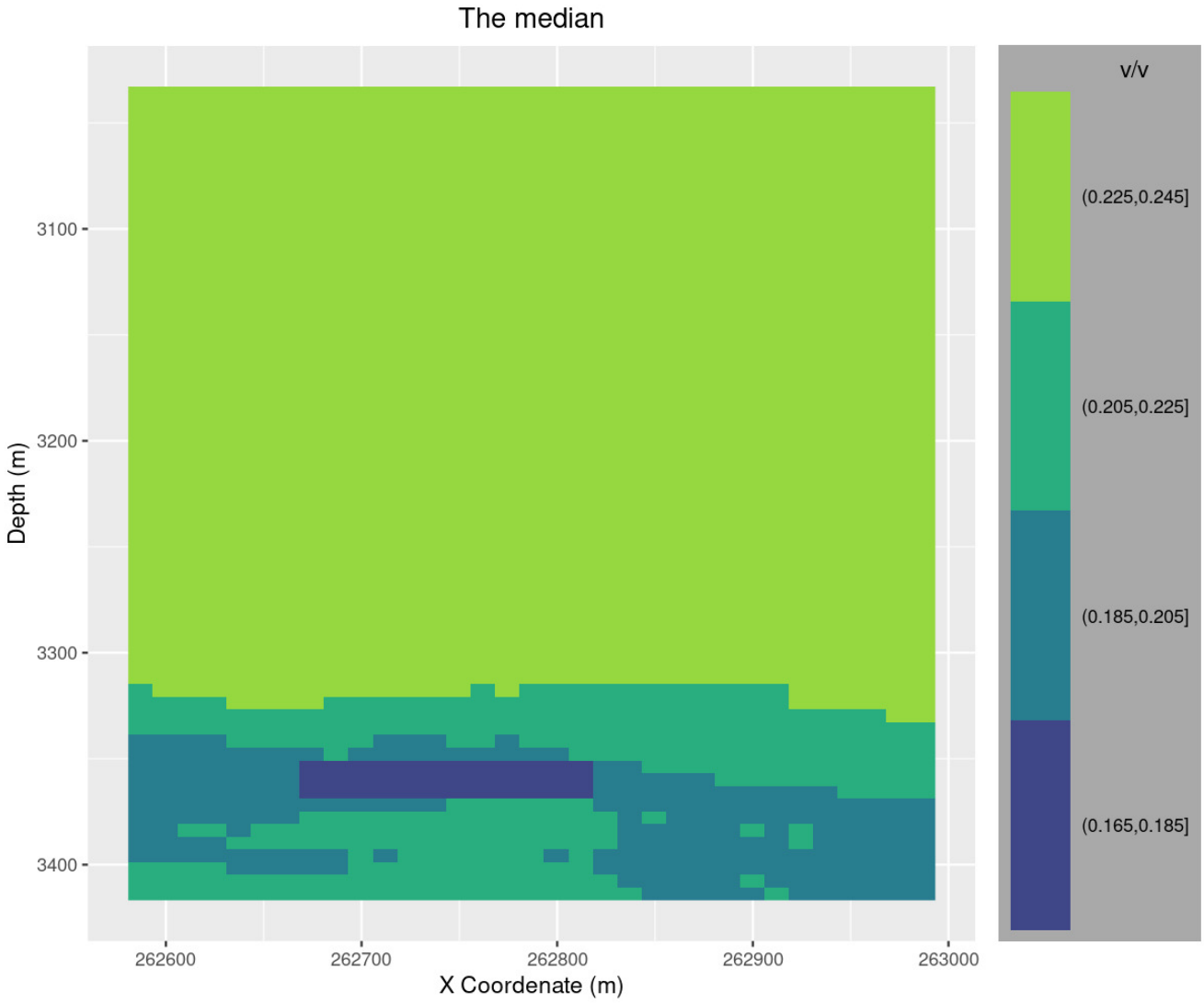


Figure 7.30: 2D seismic sections of the predicted median (50th percentile) total porosity.

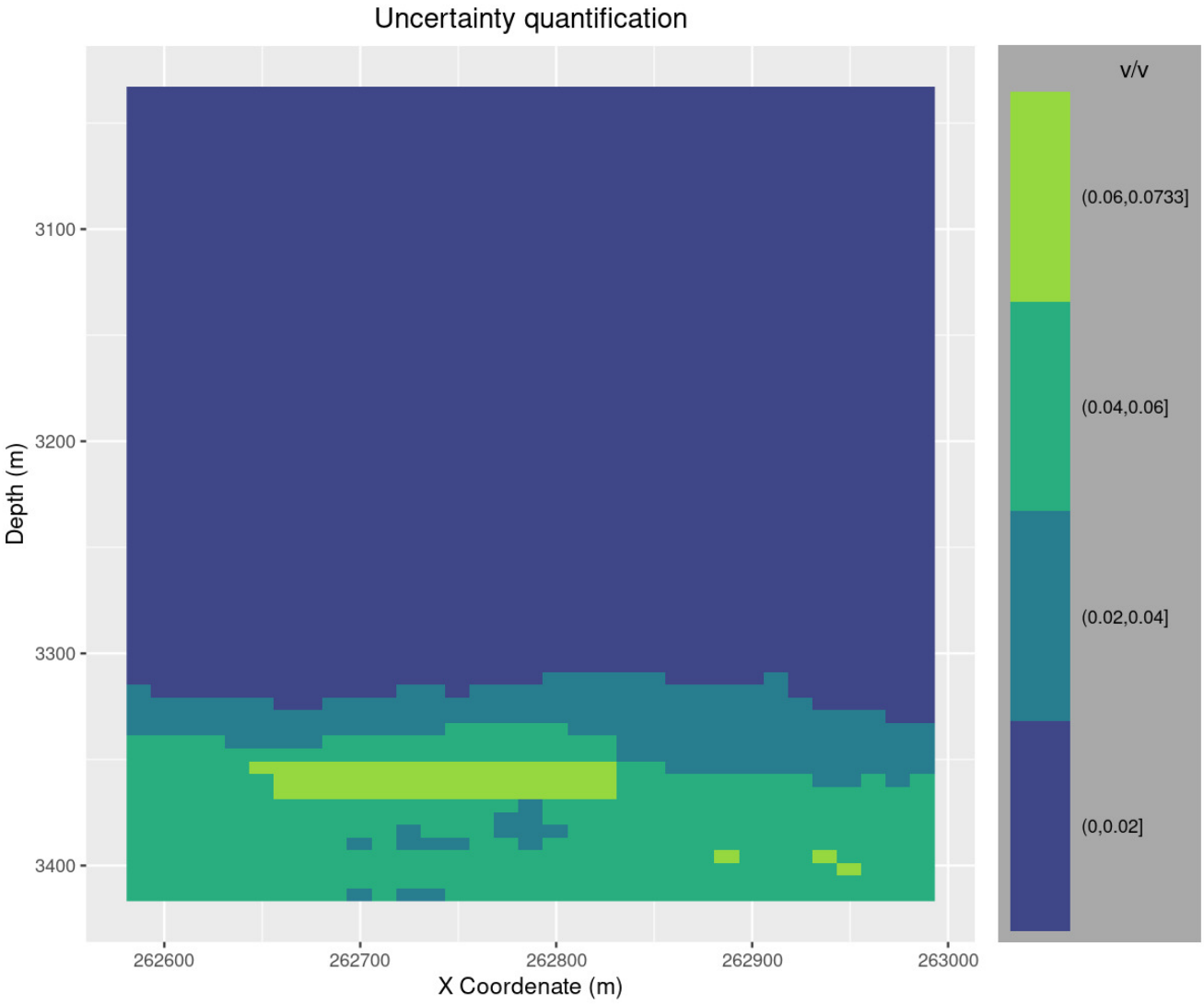


Figure 7.31: 2D seismic sections of the predicted total porosity uncertainty range.

Figure 7.32 shows a simulation of the effective porosity where the spatial variability of the variable is reproduced, and shows two areas of high values in the lower part of the intermediate values (upper middle area) and an area of low values (lower zone). Figure 7.33 and Figure 7.34 show the most likely value and the median of the effective porosity, which does not show the high values of [0.2,0.22] as in the simulation, but the areas of low and high values are agreed with simulation. Figure 7.35 shows the uncertainty range, which indicates that the areas that have high values of effective porosity present a range of uncertainty $(0.15, 0.175]$ (upper mean area), and the areas that have low values of the same present less uncertainty (lower zone). One interpretation is that the high zone with high effective porosity values can be an area of interest to produce oil

or gas, and to confirm this, the permeability and water saturation in this zone must be analyzed. An interesting observation is that this zone or body of interest follows a shape of an anticline that is consistent with the structural model in the W-E direction of the study area.

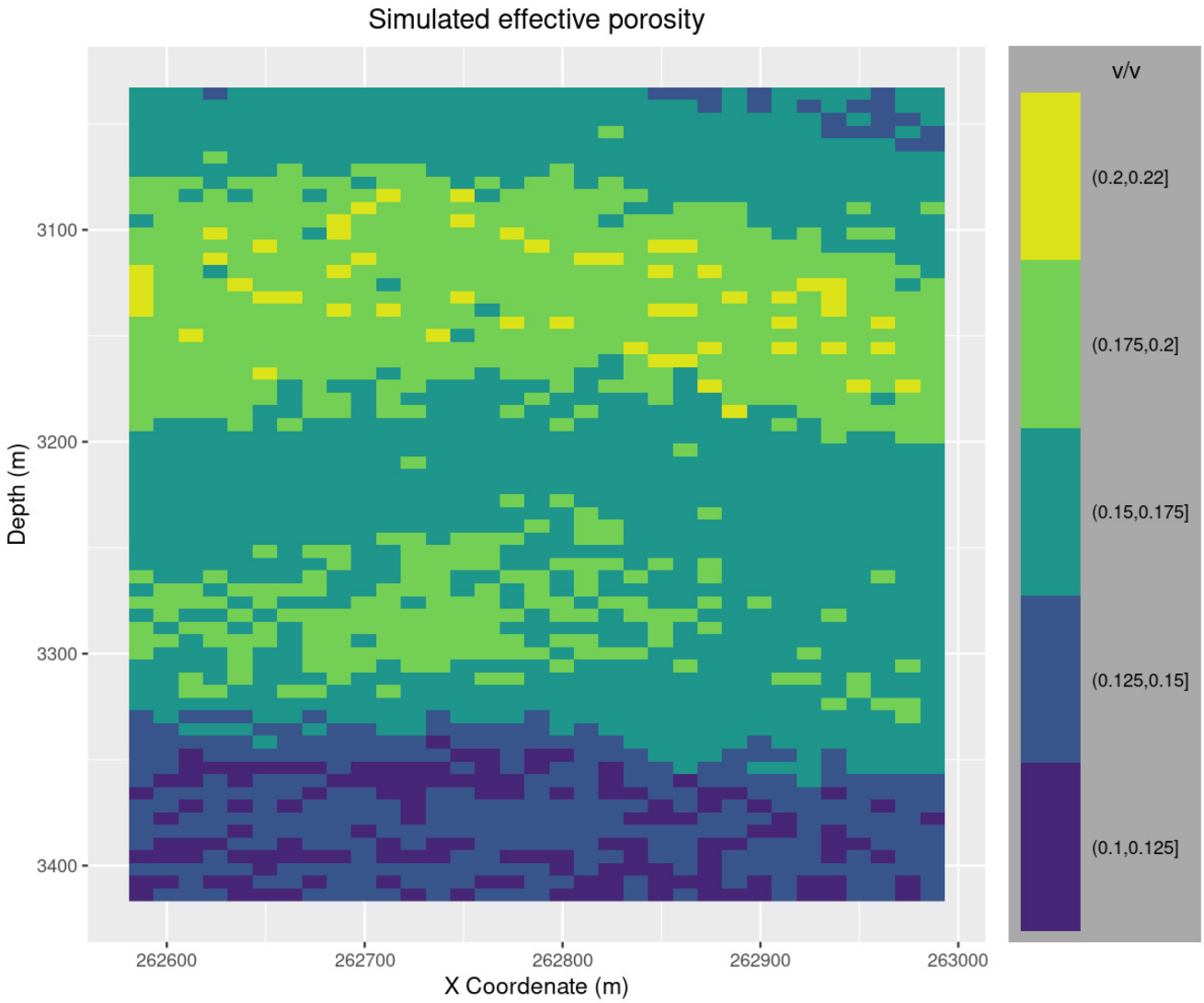


Figure 7.32: 2D seismic sections of the simulated effective porosity.

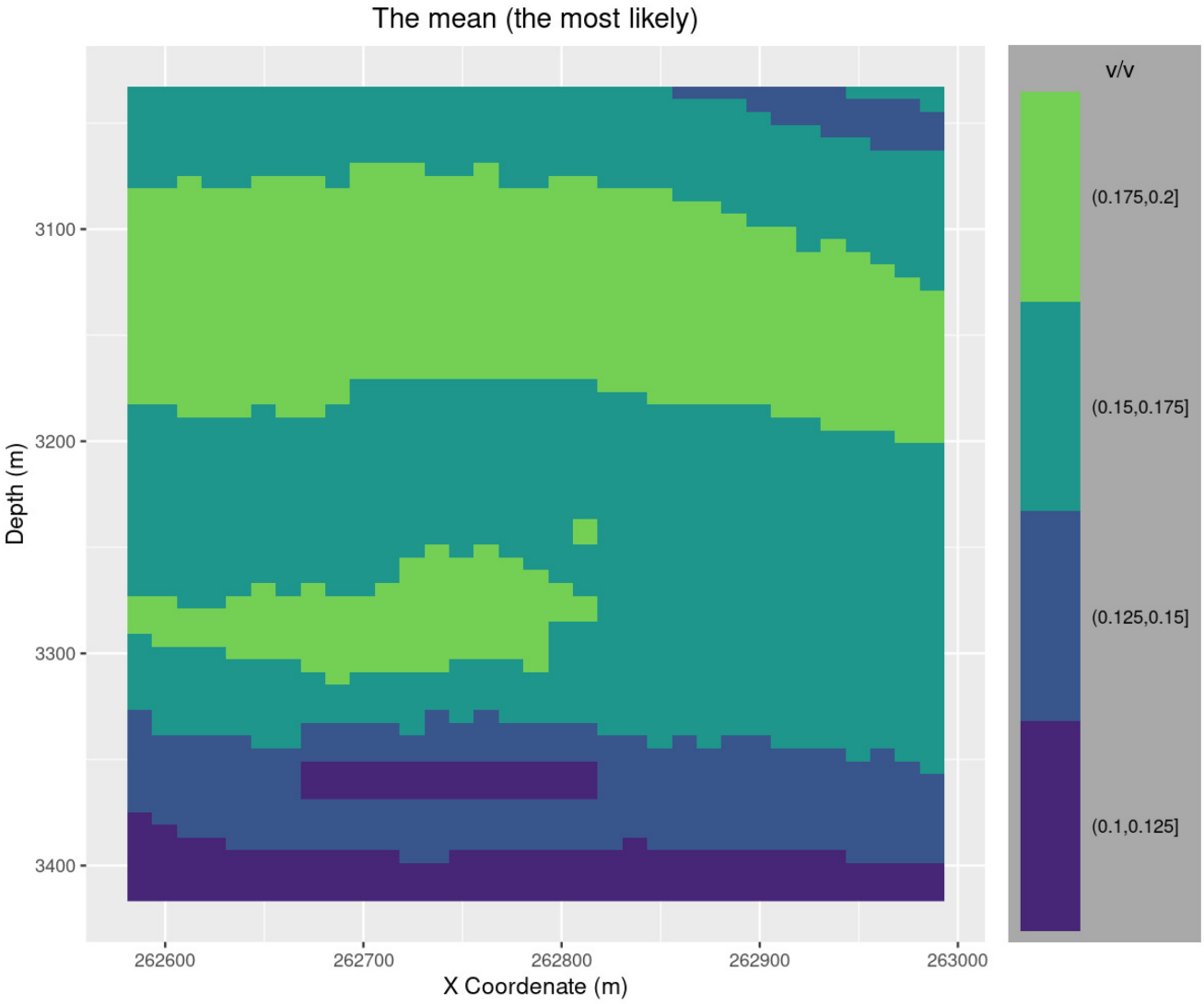


Figure 7.33: 2D seismic sections of the predicted mean effective porosity.

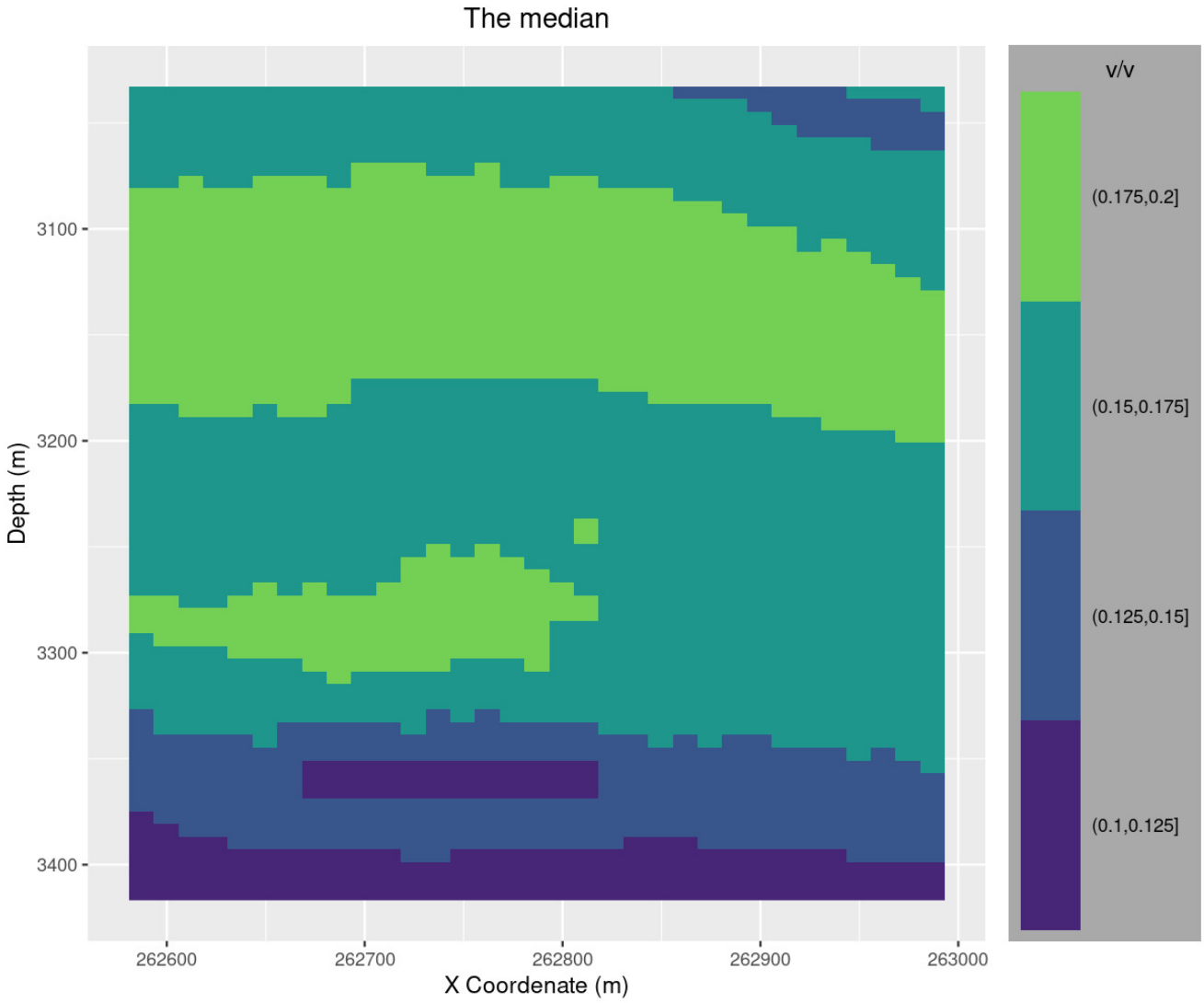


Figure 7.34: 2D seismic sections of the predicted median (50th percentile) effective porosity.

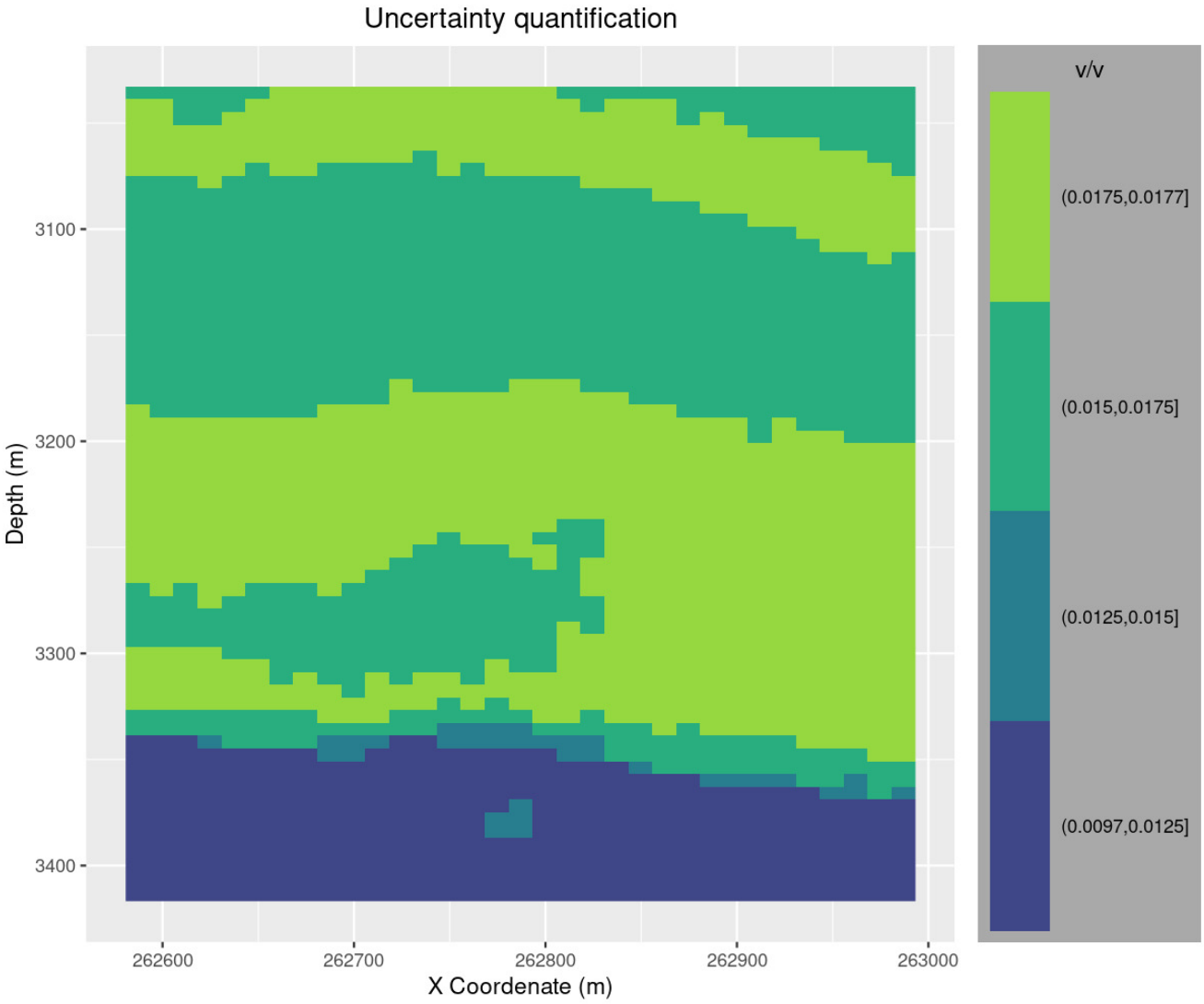


Figure 7.35: 2D seismic sections of the predicted effective porosity uncertainty range.

Figure 7.36 shows the simulation of the log(permeability) where the spatial variability of the variable is reproduced. And the areas of high and low values are quite similar with the areas of high and low values of effective porosity. This is logical because these two variables are closely related, that is, they present high dependence between them. Figure 7.37 and Figure 7.38 show the most likely value and the median of the log(permeability), which indicate that they do not present the very high [1.35,2.49] and very low [-0.302, -0.013] values as the simulation. It can be seen that the areas of high and low values are consistent between the simulation and the estimation. The upper zone of interest that presents high values of log(permeability) coincides with the zone of high values of effective porosity, and the zones of low values also coincide. Figure

7.39 shows the uncertainty range of the log(permeability) where it shows the area of high values of log(permeability) is associated with higher uncertainty and the areas with low values of the same are presenting less uncertainty. The results of the log(permeability) support the previous interpretation that the zone of interest for hydrocarbon or gas production is associated with the upper zone. But for the interpretation to be more reliable, the results for clay volume and water saturation need to be analyzed.

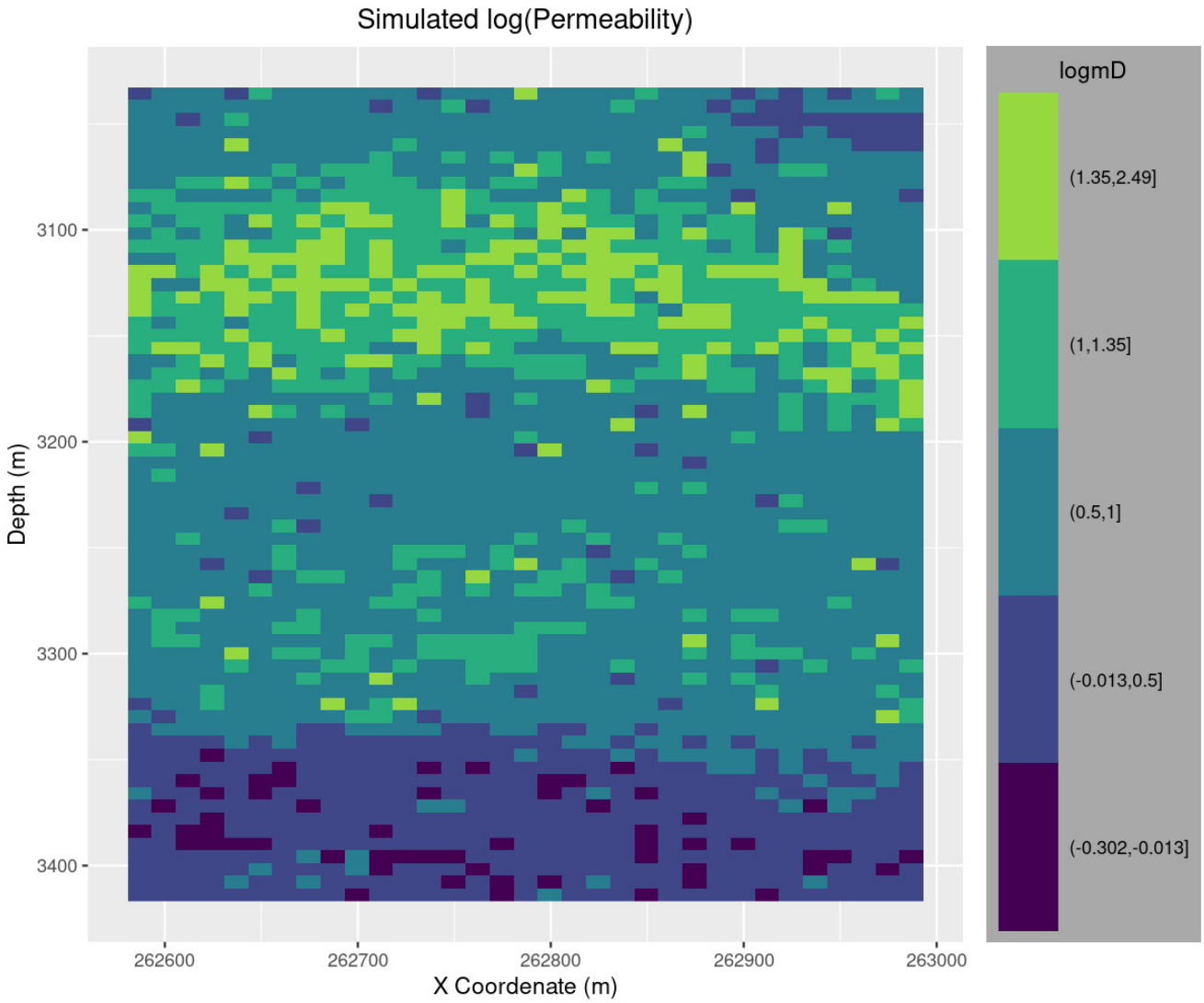


Figure 7.36: 2D seismic sections of the simulated log(Permeability).

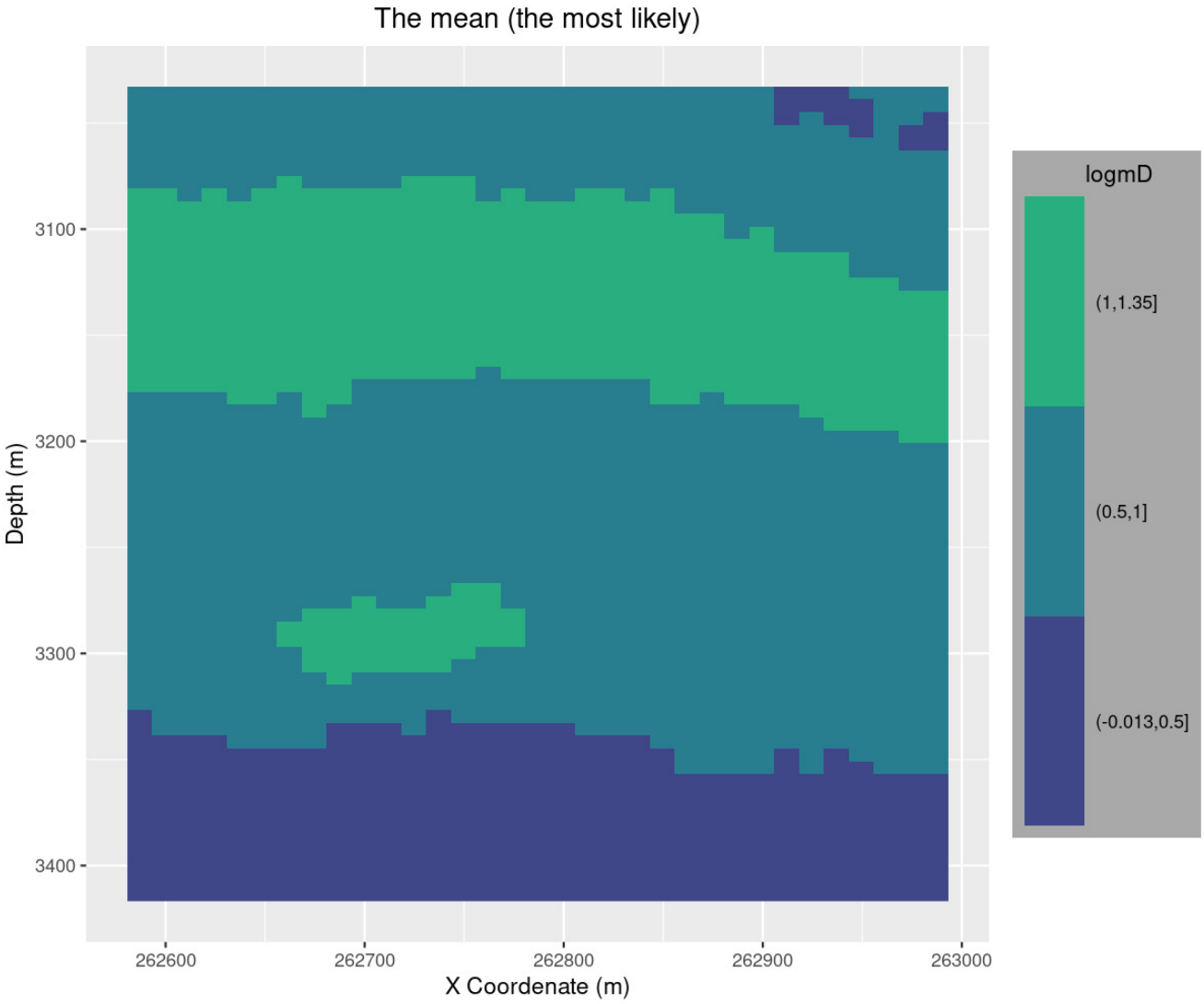


Figure 7.37: 2D seismic sections of the predicted mean log(Permeability).

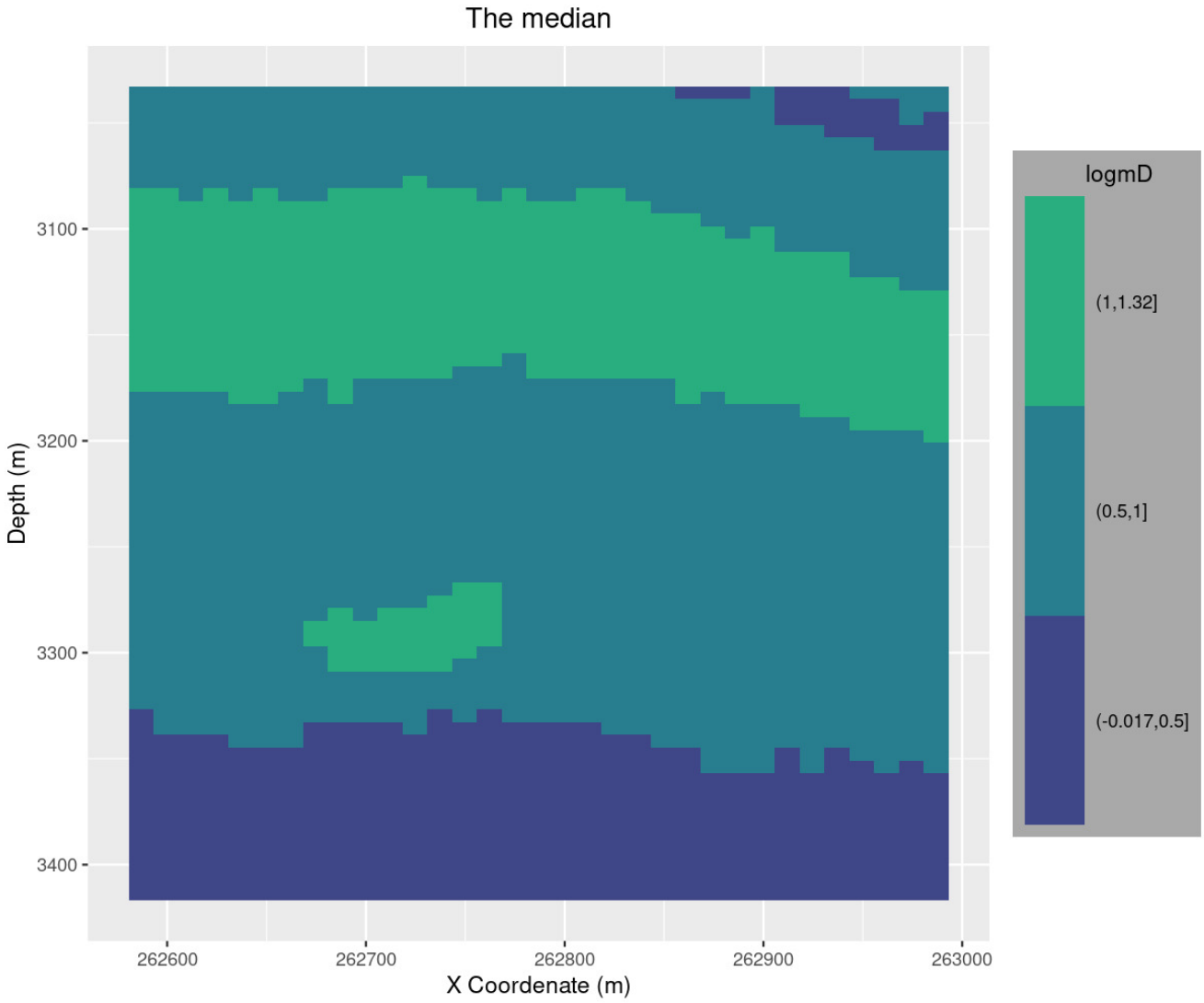


Figure 7.38: 2D seismic sections of the predicted median (50th percentile) log(Permeability).

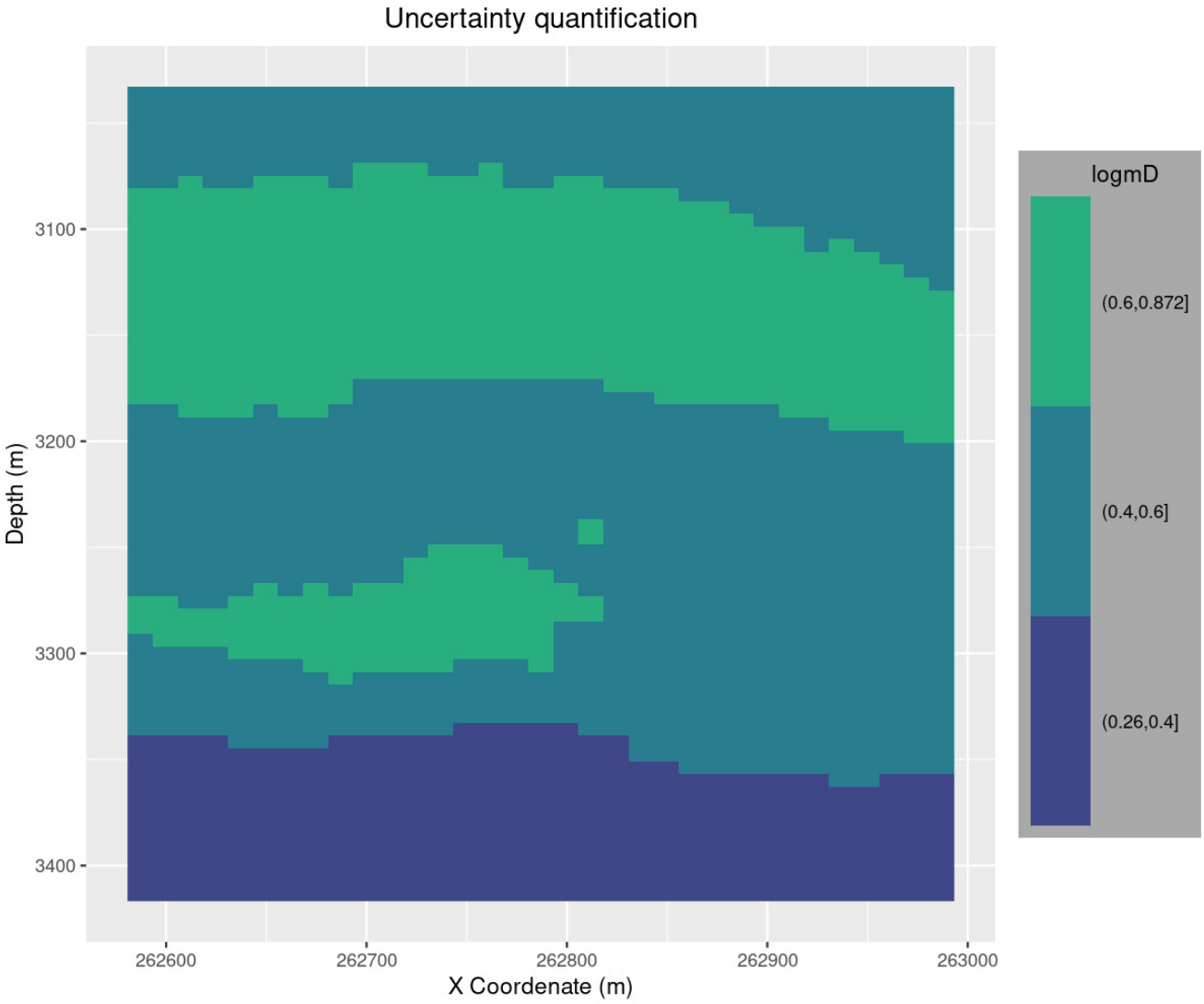


Figure 7.39: 2D seismic sections of the predicted log(Permeability) uncertainty range.

Figure 7.40 shows the simulation of the clay volume where the spatial variability of this variable occurs. Figure 7.41 and Figure 7.42 show a consistency of the high and low zones of the clay volume values, but very high [0.35,0.373] and very low [0.04,0.16] values are not observed. This confirms once again that the results of the estimation method are smoothed out and do not reproduce spatial variability. In the high zone that presents low values of the clay volume coincides with the high values of effective porosity and log (permeability). This leads to the interpretation that this zone is a sandstone zone with high effective porosity and high permeability. Figure 7.43 shows the range of uncertainty associated with the clay volume prediction where it shows that high values of the volume of clay associated with less uncertainty and low or medium values of the clay volume

associated with higher uncertainty.

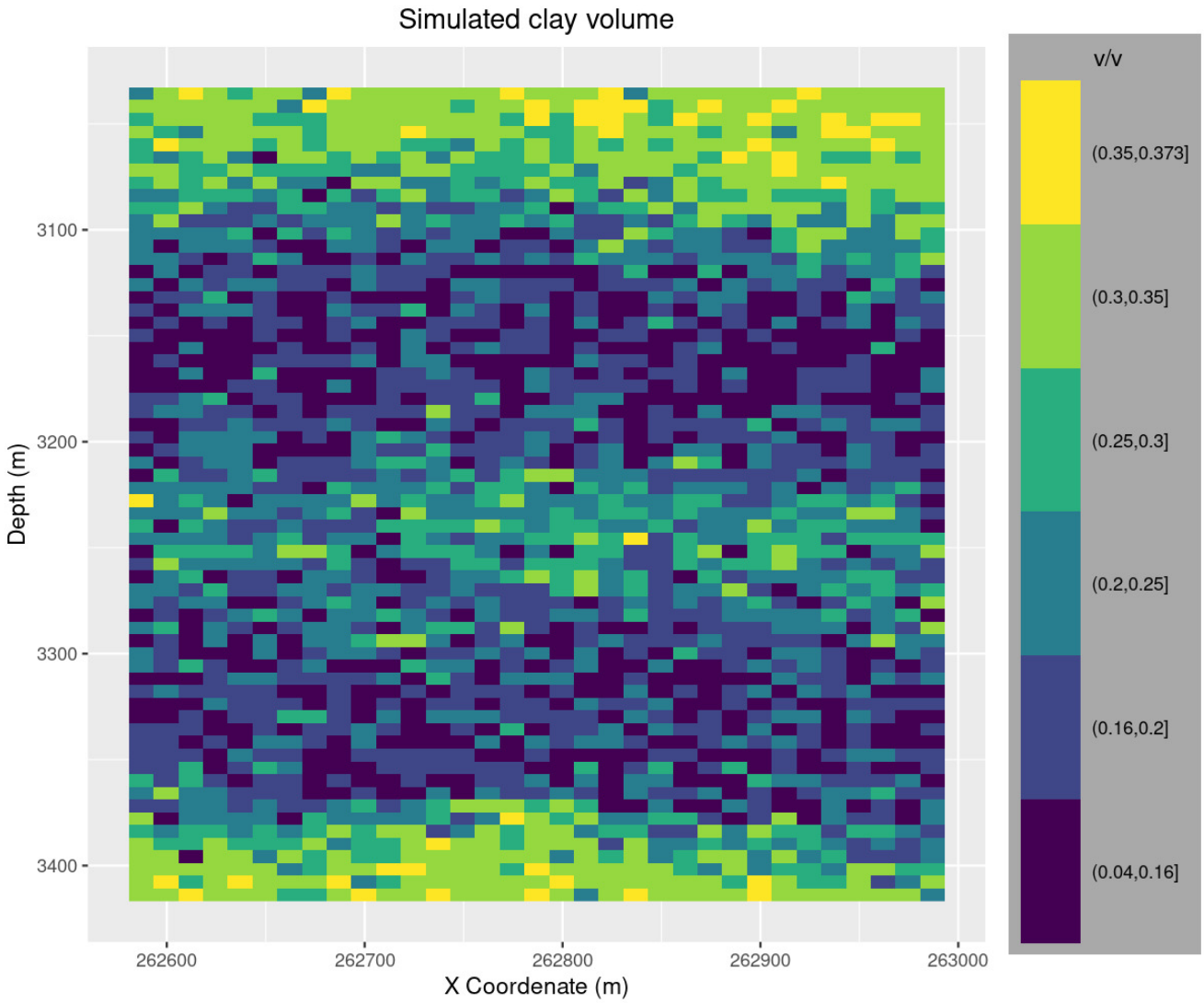


Figure 7.40: 2D seismic sections of the simulated clay volume.

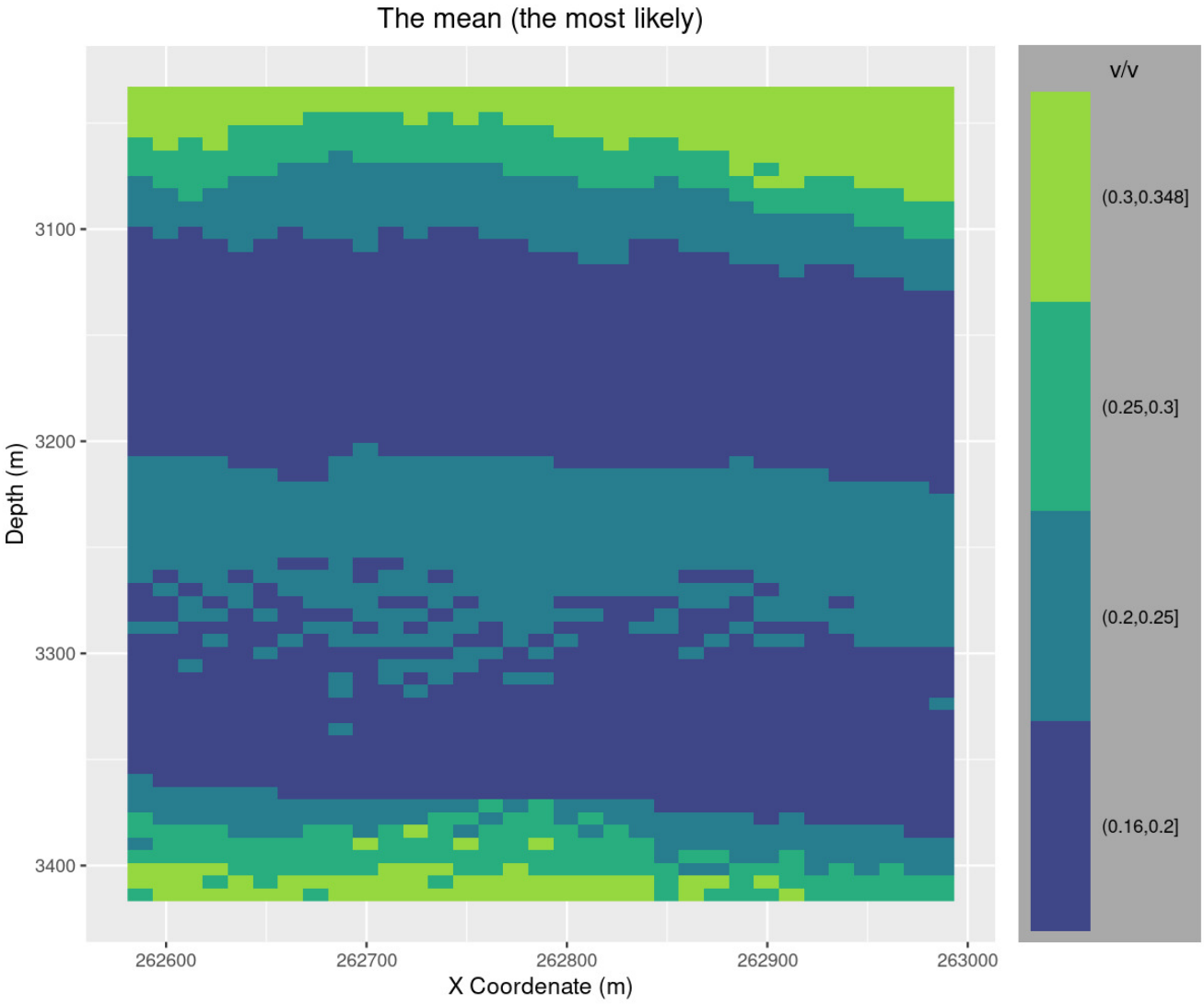


Figure 7.41: 2D seismic sections of the predicted mean clay volume.

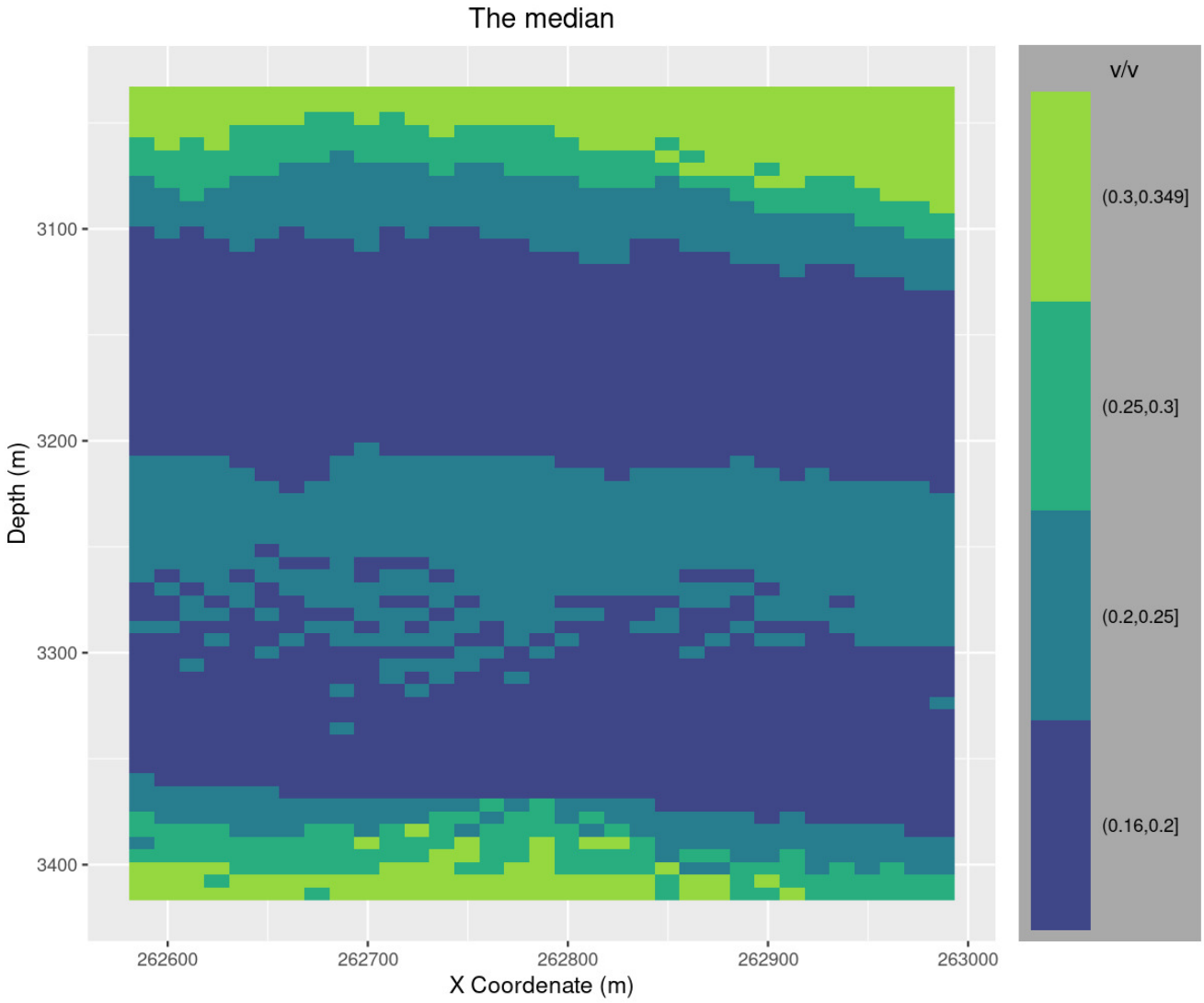


Figure 7.42: 2D seismic sections of the predicted median (50th percentile) clay volume.

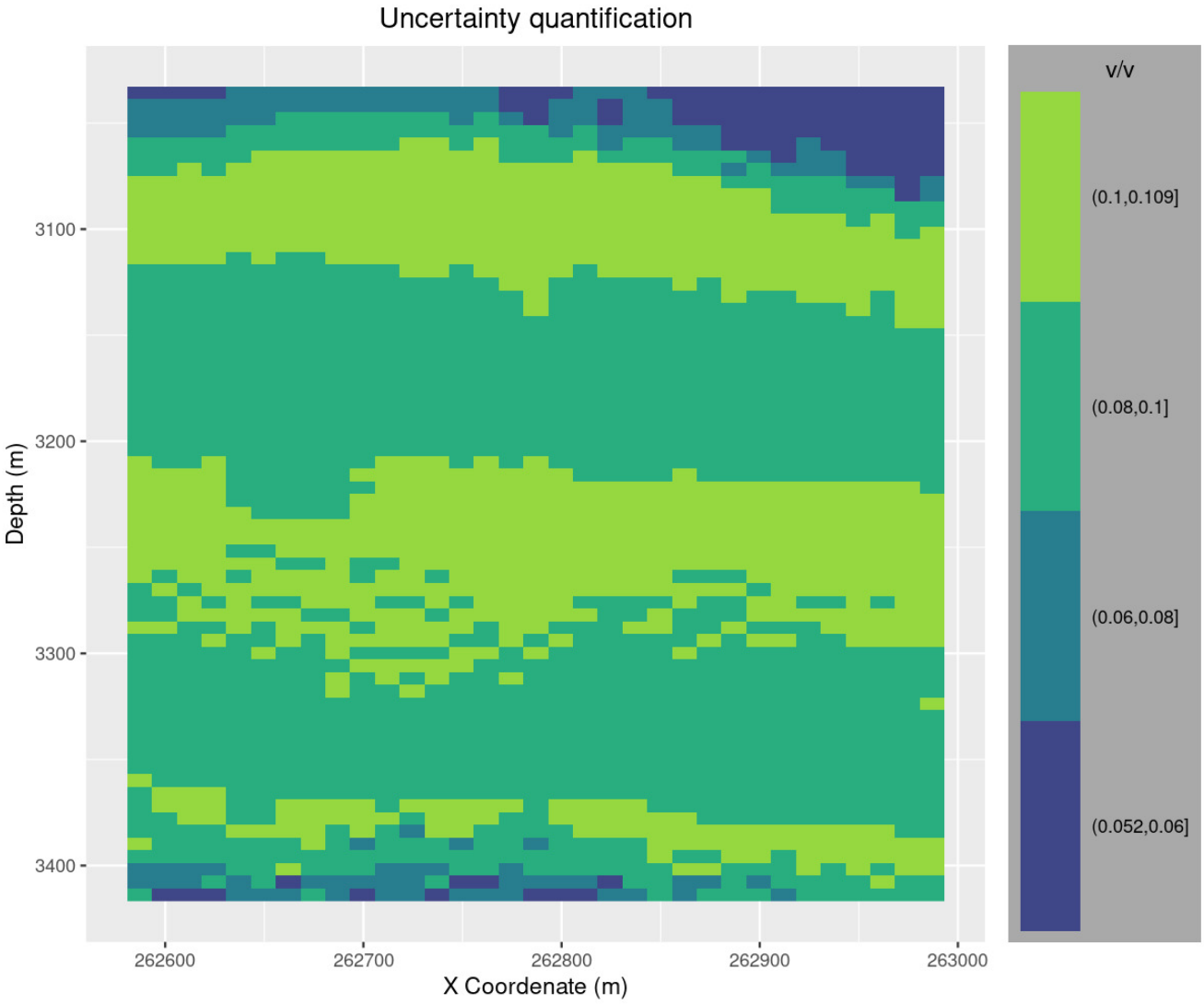


Figure 7.43: 2D seismic sections of the predicted clay volume uncertainty range.

Figure 7.44 shows the simulation of water saturation where it shows that the spatial variability of this variable is reproduced. Figure 7.45 and Figure 7.46 show the most likely value and the median of the water saturation where the upper zone is observed presents low values of the water saturation, and this zone coincides with the low values of the same in the simulation . But in this zone, the values of the most likely value are [0.75,0.8] and the simulation values are [0.25,0.75]. This case is a critical case of the estimation because the estimation results soften so much that it gives us an idea that this zone has enough water. But in reality, according to the simulation, there is not so much water in this area. This is one of the advantages of simulation. This zone that presents low values of water saturation is the same zone of low values of the volume of clay

and high values of effective porosity and high values of permeability. Therefore, once again it is confirmed that the upper zone is a zone of interest to produce hydrocarbon or gas with a sandstone lithology with high effective porosity, high permeability, and low water saturation. Figure 7.47 shows the range of uncertainty associated with the results of the prediction of water saturation, where it indicates that low values of water saturation associated with higher uncertainty and high values of the same associated with less uncertainty.

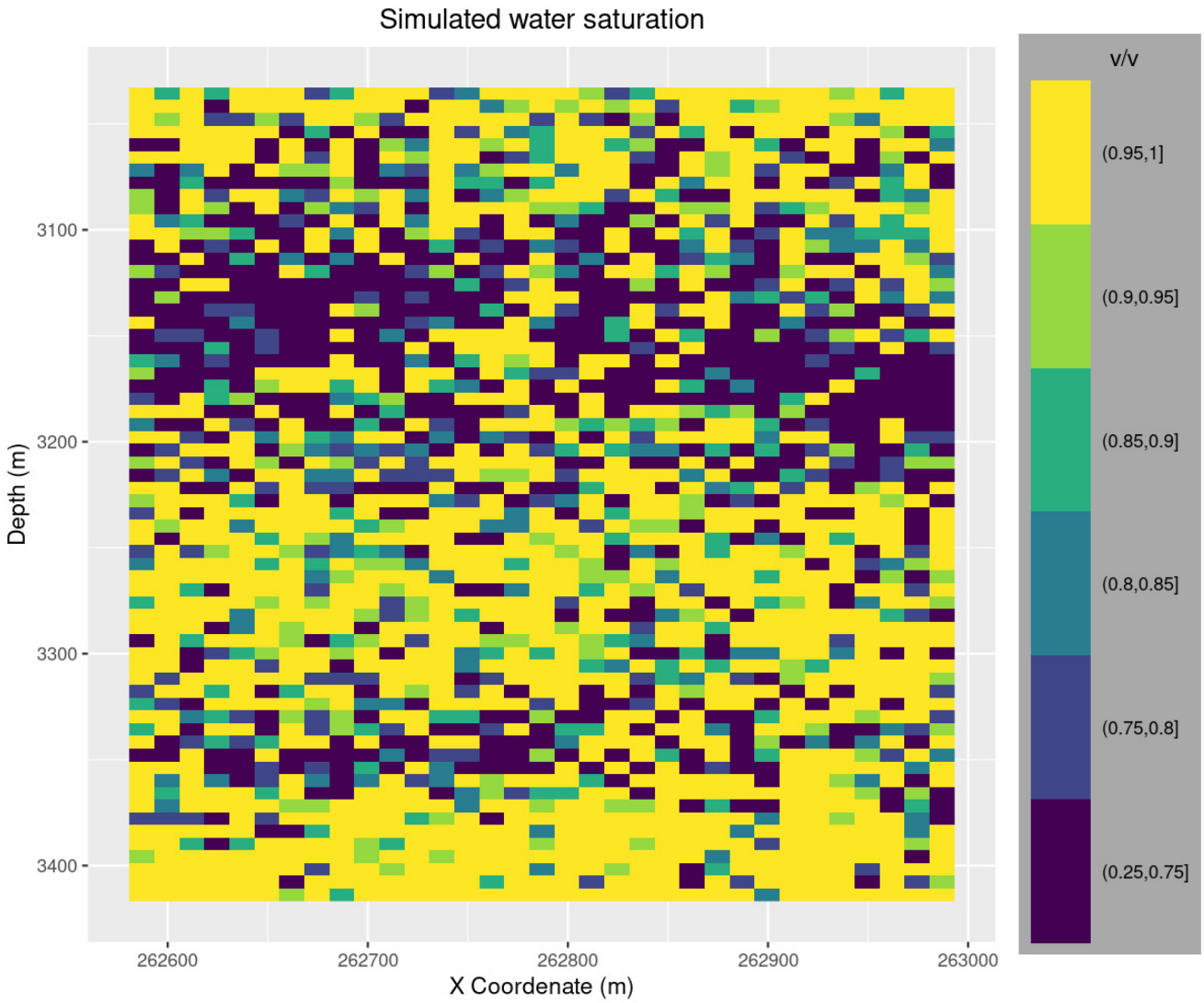


Figure 7.44: 2D seismic sections of the simulated water saturation.

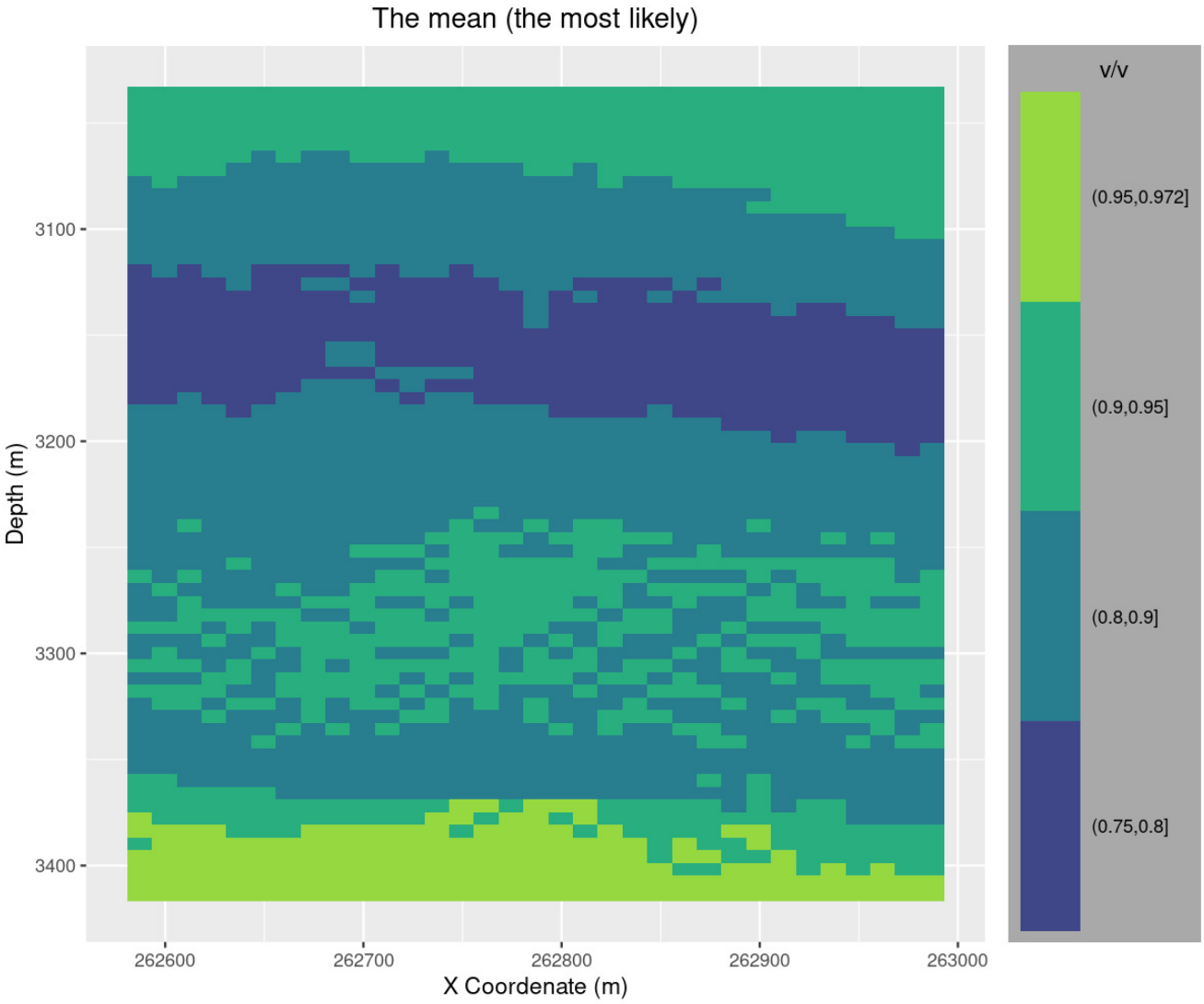


Figure 7.45: 2D seismic sections of the predicted mean water saturation.

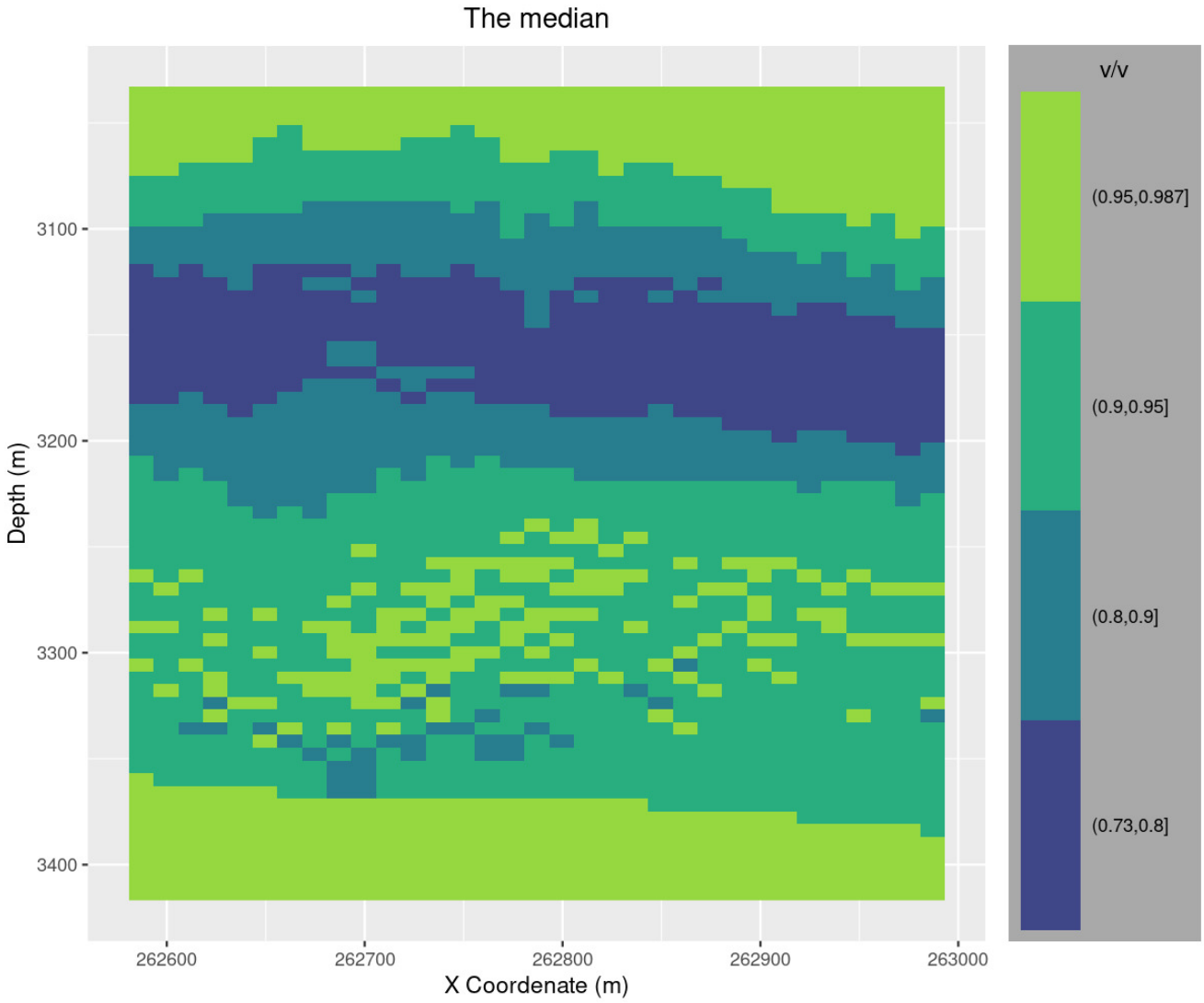


Figure 7.46: 2D seismic sections of the predicted median (50th percentile) water saturation.

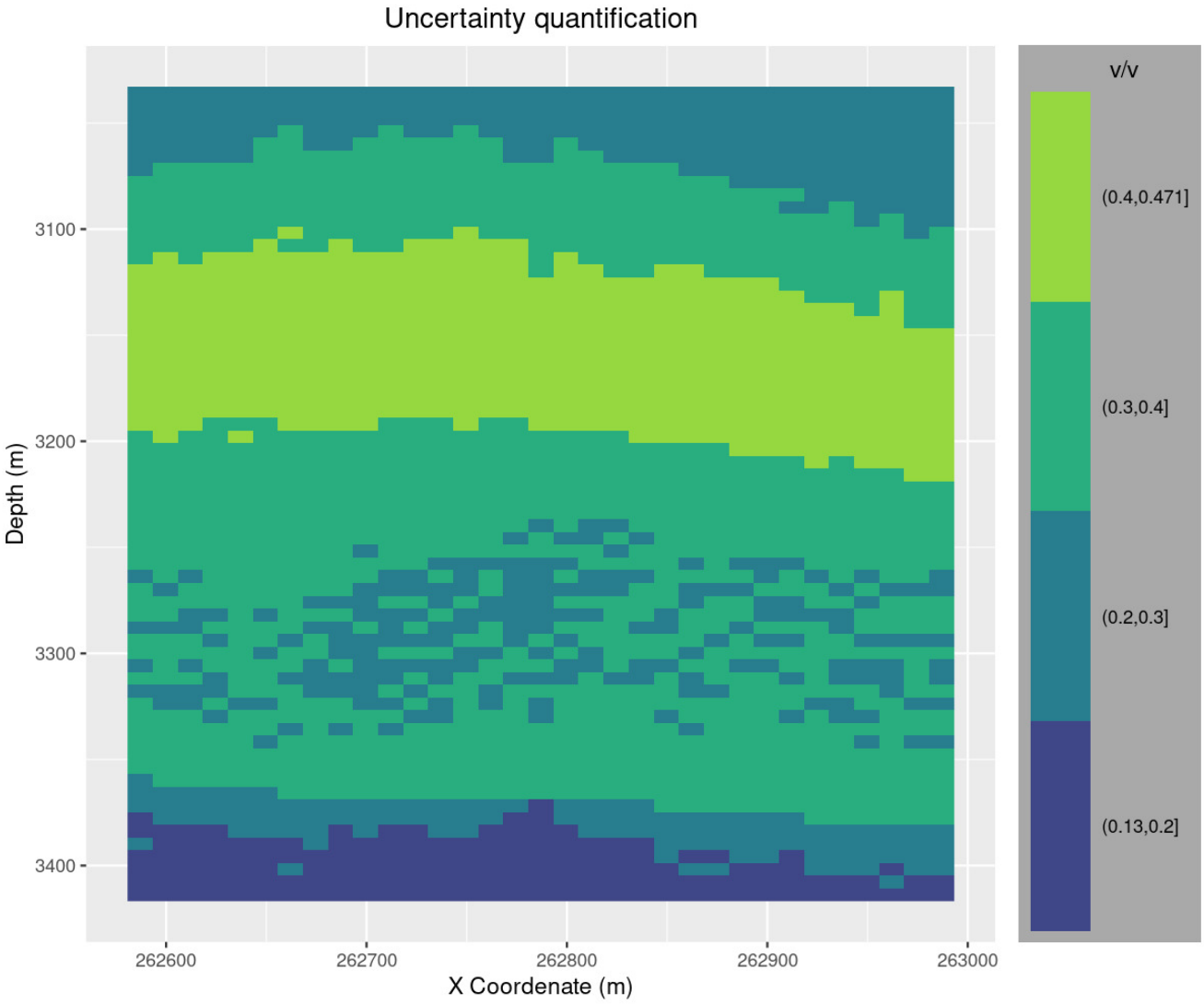


Figure 7.47: 2D seismic sections of the predicted water saturation uncertainty range.

Chapter 8

Conclusions and future work

In the bivariate case study, uncertainty quantification analysis was used to validate and compare the BCSCS and traditional SGCS method. Where it showed that the results of the BCSCS method reduce uncertainty compared to the results of the SGCS method in aspects such as univariate and bivariate distributions, and also spatial distribution. Indeed, the BCSCS method can model the dependency relationship between the random variables of interest, without assuming a linear dependence between variables nor parametric probability distribution functions. The application showed an example of how the BCSCS method can be applied by combining the data from different scales, such as well log and seismic scale.

In the multivariate case study, the proposed method was extended to multivariate problems applying vine copula with several elastic attributes to predict a set of reservoir properties of interest, including porosity, mineralogy, and fluid saturations. Copula-based estimation and simulation show the versatility of applying different approaches such as parametric, semi-parametric, and non-parametric. Copula-based estimation and simulation approaches serve not only to predict, but also allow a uncertainty quantification analysis. The simulation approach reproduces the spatial variability of the petrophysical properties of interest with the high computational cost. On the other hand, the estimation approach underestimates the spatial variability of petrophysical properties with a low computational cost.

The proposed methodology was very valuable where different sources of information such as quantitative data (well log and seismic data) and qualitative data (conceptual geological model) can be integrated. Although the geological model is not integrated into the prediction, it is used to measure the agreement between this model and the results of the predictions. The area of interest is associated with the symmetric anticline structure in the E-W direction.

Other alternatives to select the secondary variables are through the principal component and factor analysis methods. Instead of using the variables of the elastic seismic attributes, the main components or the factors are used. The advantage of these methods is that the number of secondary variables are reduced and independent, and collecting most of the variability of the data.

Future work could be the application of the Bayesian approach, which allows more systematic uncertainty quantification. Then, a comparison will be made of the results of the proposed method with the Bayesian approach.

Future work could be the application of the proposed methodology for each seismic facies classified by rock physics. The joint petrophysics and seismic inversion could be carried out through the proposed methodology using the real seismic traces.

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182 – 198. 13, 14

Afshari, A. and Shadizadeh, S. R. (2015). Integration of petrophysical and seismic data to construct areservoir permeability model. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, pages 679–686. 1, 7

Alfarraj, M. and AlRegib, G. (2018). Petrophysical property estimation from seismic data using recurrent neural networks. *SEG Technical Program Expanded Abstracts 2018. Society of Exploration Geophysicists*, pages 2141–2146. 1, 6

Almeida, A. S. and Frykman, P. (1994). Geostatistical modeling of chalk reservoir properties in the dan field, danish north sea. *in Jeffrey M. Yarus; Richard L. Chambers (eds), Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies . AAPG Computer Applications in Geology. American Association of Petroleum Geologists*. 1, 7

Almeida, A. S. and Journel, A. G. (1994). Joint simulation of multiple variables with a markov-type coregionalization model. *Mathematical Geology*, 26(5):565–588. 1, 7

Ardia, D., Mullen, K., Peterson, B., and Ulrich, J. (2020). Global optimization by differential evolution. *The R Project for Statistical Computing*. 23, 24

Azevedo, L. and Soares, A. (2017). *Geostatistical Methods for Reservoir Geophysics*. Springer International Publishing. 1, 7

Babak, O. and Deutsch, C. V. (2009). Improved spatial modeling by merging multiple secondary data for intrinsic collocated cokriging. *Journal of Petroleum Science and Engineering*, 69(1):93 – 99. 1, 7

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *WATER RESOURCES RESEARCH*. 2

Bevacqua, E. (2017). Sampling from conditional c- and d-vine copulas. *The R Project for Statistical Computing*. 24

Bevacqua, E., Maraun, D., Haff, I. H., Widmann, M., and Vrac, M. (2017). Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in ravenna (italy). *Hydrology and Earth System Sciences*, 21:2701–2723. 2

Bárdossy, A. and Li, J. (2008). Geostatistical interpolation using copulas. 2

Caers, J. (2005). *Petroleum Geostatistics*. Society of Petroleum Engineers. 1

Cao, R., Ma, Y. Z., and Gomez, E. (2014). Geostatistical applications in petroleum reservoir modelling. *Journal of the Southern African Institute of Mining and Metallurgy*. 1, 7

Charpentier, A., Fermanian, J., and Scaillet, O. (2007). The estimation of copulas : theory and practice. 21

Chatterjee, R., Singha, D. K., Ojha, M., Sen, M. K., and Sain, K. (2016). Porosity estimation from pre-stack seismic data in gas-hydrate bearing sediments, krishna-godavari basin, india. *Journal of Natural Gas Science and Engineering*, 33:562–572. 1, 6

Chilès, J.-P. and Delfiner, P. (2012). *Geostatistics: modeling spatial uncertainty*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2 edition. 1, 2, 7, 55

CNH (2005). *Perdido fold belt, subsalt belt, mexican ridges. Petroleum geological synthesis. Gulf of mexico - deep water- north sector*. xviii, 37, 40, 41, 42, 43, 44, 45, 46, 47, 48

Coburn, T. C., Yarus, J. M., and Chambers, R. L. (2006). *Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies, Volume II*, volume 5. American Association of Petroleum Geologists. 1

Corina, A. and Hovda, S. (2018). Automatic lithology prediction from well logging using kernel density estimation. *Journal of Petroleum Science and Engineering*, 170:664 – 674. 2

Cosentino, L. (2001). *Integrated Reservoir Studies*. TECHNIP, Paris, France. 1, 6, 7

Dafflon, B. and Barrash, W. (2012). Three-dimensional stochastic estimation of porosity distribution: Benefits of using ground-penetrating radar velocity tomograms in simulated-annealing-based or bayesian sequential simulation approaches. *WATER RESOURCES RESEARCH*. 1, 9

Deutsch, C. V. (2002). *Geostatistical Reservoir Modeling*. Oxford University Press; 1st edition. 1

Deutsch, C. V. and Cockerham, P. W. (1994a). Geostatistical modeling of permeability with annealing cosimulation (acs). *SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana*, page 523–532. 1, 8, 9

Deutsch, C. V. and Cockerham, P. W. (1994b). Practical considerations in the application of simulated annealing to stochastic simulation. *Mathematical Geology*, 26(1):67–82. 1, 9

Deutsch, C. V. and Journel, A. G. (1998). Gslib: Geostatistical software library and user's guide. second edition. *Oxford University Press, New York*. 8, 23, 24, 25

Díaz-Viera, M., Casar-González, R., Hernández-Maldonado, V., Méndez-Venegas, J., Mendoza-Torres, F., and Checa-Rojas, P. (2013). Nuevas metodologías y herramientas de caracterización estática y dinámica considerando las propiedades fractales de los yacimientos petroleros. *Proyecto Fondos Sener-Conacyt No. 143935 (Y-00114)*. xvii, 28

Díaz-Viera, M. A. (2002). *Geoestadística aplicada*. Instituto de Geofísica, UNAM. Instituto de Geofísica y Astronomía, CITMA, Cuba . 7, 31, 32, 49, 51, 55

Díaz-Viera, M. A., Anguiano-Rojas, P., Mousatov, A., Kazatchenko, E., and Markov, M. (2006). Stochastic modeling of permeability in double porosity carbonates applying a monte-carlo simulation method with t-copula. *In SPWLA 47th Annual Logging Symposium*. 2

Díaz-Viera, M. A. and Casar-González, R. (2005). Stochastic simulation of complex dependency pattern of petrophysical properties using t-copulas. *Proceedings of IAMG'05: GIS and Spatial Analysis*, 2:749–755. 2

Díaz-Viera, M. A., Erdely, A., Kerdan, T., del Valle-García, R., and Mendoza-Torres, F. (2017). *Bernstein Copula-Based Spatial Stochastic Simulation of Petrophysical Properties Using Seismic Attributes as Secondary Variable*, pages 487–504. Springer International Publishing, Cham. 7, 21

Díaz-Viera, M. A., Erdely, A., Kerdan, T., del Valle-García, R., and Mendoza-Torres, F. (2017). Bernstein copula-based spatial stochastic simulation of petrophysical properties using seismic attributes as secondary variable. *In: Gómez-Hernández J., Rodrigo-Ilarri J., Rodrigo-Clavero M., Cassiraga E., Vargas-Guzmán J. (eds) Geostatistics Valencia 2016. Quantitative Geology and Geostatistics. Springer, Cham*, 19:487–504. 2

Doyen, P. (2007). *Seismic Reservoir Characterization: An Earth Modelling Perspective*. European Association of Geoscientists and Engineers. 1, 2

Doyen, P. M., den Boer, L. D., and Pillet, W. R. (1996). Seismic porosity mapping in the ekofisk field using a new form of collocated cokriging. *Society of Petroleum Engineers Annual Technical Conference and Exhibition, Denver*, page 21–30. 1, 7

Dubrule, O. (2003). *Geostatistics for Seismic Data Integration in Earth Models*. Society of Exploration Geophysicists. 1

Embrechts, P., McNeil, A., and Straumann, D. (1999). Correlation: Pitfalls and alternatives. *Risk Magazine 5*. 121

Emery, X. and Parra, J. (2013). Integration of crosswell seismic data for simulating porosity in a heterogeneous carbonate aquifer. *Journal of Applied Geophysics*, 98:254 – 264. 1, 7

Erdely, A. (2011). Modelo probabilístico no paramétrico para dependencia bivariada de propiedades petrofísicas. *Reporte Técnico del Proyecto Y.00102 WAG.* 121

Erdely, A. and Díaz-Viera, M. A. (2010). Nonparametric and semiparametric bivariate modeling of petrophysical porosity-permeability dependence from well log data. *In: Jaworski P., Durante F., Härdle W., Rychlik T. (eds) Copula Theory and Its Applications. Lecture Notes in Statistics. Springer, Berlin, Heidelberg*, 198:267–278. 2, 16, 21

Erdely, A. and Díaz-Viera, M. A. (2015). A vine and gluing copula model for permeability stochastic simulation. In *Stochastic and data analysis methods and applications in statistics and demography*, pages 273–281. 2

Figueiredo, L. P. D., Grana, D., Roisenberg, M., and Rodrigues, B. B. (2019). Gaussian mixture markov chain monte carlo method for linear seismic inversion. *GEOPHYSICS*, 84(3). 1, 8

Freedman, D. A. (2009). *Statistical Models: Theory and Practice*. Cambridge University Press. 5

Geenens, G., Charpentier, A., and Paindaveine, D. (2014). Probit transformation for nonparametric kernel estimation of the copula density. 23:1848–1873. 21

Gijbels, I. and Mielniczuk, J. (1990). Estimating the density of a copula function. *Communications in Statistics - Theory and Methods*, 19(2):445–464. 21

Gogoi, T. and Chatterjee, R. (2019). Estimation of petrophysical parameters using seismic inversion and neural network modeling in upper assam basin, india. *Geoscience Frontiers*, 10(3):1113–1124. 1, 6

Gómez-Hernández, J. and Journel, A. (1993). Joint sequential simulation of multigaussian fields. *In: Soares A. (eds) Geostatistics Tróia '92. Quantitative Geology and Geostatistics. Springer, Dordrecht*, 3. 1, 7

Grana, D. (2018). Joint facies and reservoir properties inversion. *Geophysics*, 83(3):M15–M24. 2

Grana, D., Fjeldstad, T., and Omre, H. (2017). Bayesian gaussian mixture linear inversion for geophysical inverse problems. *Mathematical Geosciences*, 49. 1, 8

Grana, D., Mukerji, T., Dvorkin, J., and Mavko, G. (2012). Stochastic inversion of facies from seismic data based on sequential simulations and probability perturbation method. 77(4):M53–M72. 1, 8

Grana, D. and Rossa, E. D. (2010). Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion. *GEOPHYSICS*, 75:O21–O37. 2, 8

Gräler, B. (2014). Modelling skewed spatial random fields through the spatial vine copula. *Spatial Statistics*, 10:87 – 102. 2

Gräler, B. and Pebesma, E. (2011). The pair-copula construction for spatial data: a new approach to model spatial dependency. *Procedia Environmental Sciences*, 7:206 – 211. Spatial Statistics 2011: Mapping Global Change. 2

Haslauer, C. P., Li, J., and Bárdossy, A. (2010). *Application of Copulas in Geostatistics*, pages 395–404. Springer Netherlands, Dordrecht. 2

Hernández-Maldonado, V., Díaz-Viera, M. A., and Erdely, A. (2012). A joint stochastic simulation method using the bernstein copula as a flexible tool for modeling nonlinear dependence structures between petrophysical properties. *Journal of Petroleum Science and Engineering*, 90-91:112–123. 2, 21

Hernández-Maldonado, V., Díaz-Viera, M. A., and Erdely, A. (2014). A multivariate bernstein copula model for permeability stochastic simulation. *Geofísica Internacional*, 53:163–181. 2, 21

Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2020). Multivariate dependence with copulas. *The R Project for Statistical Computing*. 24

Horta, A. and Soares, A. (2010). Direct sequential co-simulation with joint probability distributions. *Mathematical Geosciences*, 42(3):269–292. 1, 7

Ingle-González, J. E. (2016). Modelación multivariada por medio de cópulas de dependencias no lineales para la simulación espacial conjunta de propiedades georeferenciadas. *Facultad de Estudios Superiores Acatlán, UNAM*. 121

Iturrarán-Viveros, U. (2012). Smooth regression to estimate effective porosity using seismic attributes. *Journal of Applied Geophysics*, 76:1 – 12. 1, 6

Iturrarán-Viveros, U. and Parra, J. O. (2014). Artificial neural networks applied to estimate permeability, porosity and intrinsic attenuation using seismic attributes and well-log data. *Journal of Applied Geophysics*, 107:45 – 54. 1, 6

Jalalalhosseini, S. M., Eskandari, S., and Mortezazadeh, E. (2015). The technique of seismic inversion and use of the relation between inversion results and porosity log for predicting porosity of a carbonate reservoir in a south iranian oil field. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 37(3):265–272. 6

Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman and Hall/CRC. 12, 21

Kazianka, H. and Pilz, J. (2010). *Spatial Interpolation Using Copula-Based Geostatistical Models*, pages 307–319. Springer Netherlands, Dordrecht. 2

Krupskii, P. and Genton, M. G. (2019). A copula model for non-gaussian multivariate spatial data. *Journal of Multivariate Analysis*, 169:264 – 277. 2

Lang, X. and Grana, D. (2017). Geostatistical inversion of prestack seismic data for the joint estimation of facies and impedances using stochastic sampling from gaussian mixture posterior distributions. *GEOPHYSICS*, 82(4):M55–M65. 2, 8

Le, V. H., Díaz-Viera, M. A., Vázquez-Ramírez, D., [del Valle-García], R., Erdely, A., and Grana, D. (2020). Bernstein copula-based spatial cosimulation for petrophysical property prediction conditioned to elastic attributes. *Journal of Petroleum Science and Engineering*, 193:107382. 21

Maurya, S. and Singh, K. (2019). Predicting porosity by multivariate regression and probabilistic neural network using model-based and coloured inversion as external attributes: a quantitative comparison. *Journal geological society of india*, 93:207–212. 1, 6

Mendoza-Torres, F., Díaz-Viera, M., and Erdely, A. (2017). Bernstein copula modeling for 2d discrete fracture network simulations. *Journal of Petroleum Science and Engineering*, 156:710 – 720. 21

Moon, S., Lee, G. H., Kim, H., Choi, Y., and Kim, H.-J. (2016). Collocated cokriging and neural-network multi-attribute transform in the prediction of effective porosity: A comparative case study for the second wall creek sand of the teapot dome field, wyoming, usa. *Journal of Applied Geophysics*, 131:69 – 83. 1, 7

Nagler, T. (2018). Multivariate kernel density estimation with vine copulas. *The R Project for Statistical Computing*. 24

Nagler, T. and Czado, C. (2016). Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis*, 151:69 – 89. 16

Nagler, T. and Kraus, D. (2020). D-vine quantile regression. *The R Project for Statistical Computing*. 24

Nagler, T., Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., and Erhardt, T. (2020). Statistical inference of vine copulas. *The R Project for Statistical Computing*. 24

Nagler, T. and Vatter, T. (2020). *The R Project for Statistical Computing*. 24

Nagler, T. and Wen, K. (2018). Kernel smoothing for bivariate copula densities. *The R Project for Statistical Computing*. 24

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer-Verlag New York. xvii, 12, 13, 22, 23, 122, 123

Parra, J. and Emery, X. (2013). Geostatistics applied to cross-well reflection seismic for imaging carbonate aquifers. *Journal of Applied Geophysics*, 92:68–75. 1, 7

Perez, J. M. and Palacín, A. F. (1987). Estimating the quantile function by bernstein polynomials. *Computational Statistics & Data Analysis*, 5(4):391 – 397. 16

Poole, M. A. and O'farrell, P. N. (1971). The assumptions of the linear regression model. *Transactions of the Institute of British Geographers*, page 145. 5

RCoreTeam (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing. Vienna, Austria.* 24

Remy, N., Boucher, A., and Wu, J. (2009). *Applied geostatistics with SGeMS: a user's guide*. Cambridge University Press, New York. 7, 56

Sancetta, A. (2007). Nonparametric estimation of distributions with given marginals via bernstein-kantorovic polynomials: L 1 and pointwise conver- gence theory. *J. Multivariate Anal. 98, 1376-1390.* 21, 22

Sancetta, A. and Satchell, S. (2004). The bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, 20(03):535–562. 21, 22

Shen, X., Zhu, Y., and Song, L. (2008). Linear b-spline copulas with applications to nonparametric estimation of copulas. *Computational Statistics & Data Analysis*, 52(7):3806 – 3819. 16

Sklar, A. (1959). Fonctions de répartition á n dimensions et leurs marges. *Publ. Inst. Statistique Univ. Paris*, 8:229–231. 11

Soares, A. (2001). Direct sequential simulation and cosimulation. *Mathematical Geology*, 33(8):911–926. 1, 7

Soares, A. (2005). *Classification of Mining Reserves using Direct Sequential Simulation*, pages 511–522. Springer Netherlands, Dordrecht. 1, 7

Soares, A., Nunes, R., and Azevedo, L. (2017). Integration of uncertain data in geostatistical modelling. *Mathematical Geosciences*, 49(2):253–273. 1, 7

Stien, M. and Kolbjørnsen, O. (2008). D-vine creation of non-gaussian random fields. *Norwegian Computing Center.* 2

Vejbæk, O. V. and Kristensen, L. (2000). Downflank hydrocarbon potential identified using seismic inversion and geostatistics: Upper maastrichtian reservoir unit, dan field, danish central graben. *Petroleum Geoscience.* 1, 9

Verly, G. (1992). Sequential gaussian cosimulation: A simulation method integrating several types of information. *Proceedings of 4th International Geostatistical Congress, 11 p.* 7

Verly, G. W. (1993). Sequential gaussian cosimulation: A simulation method integrating several types of information. *Soares A. (eds) Geostatistics Troia 92. Quantitative Geology and Geostatistics, Springer, Dordrecht*, 5:543–554. 1, 7

Vernik, L. and Nur, A. (1992). Petrophysical Classification of Siliciclastics for Lithology and Porosity Prediction from Seismic Velocities1. *AAPG Bulletin*, 76(9):1295–1309. 6

Wen, K. and Wu, X. (2020). Transformation-kernel estimation of copula densities. *Journal of Business & Economic Statistics*, 38(1):148–164. 21

Xiang, Y., Gubian, S., Suomela, B., and Hoeng, J. (2018). Generalized simulated annealing for efficient global optimization: the GenSA package for R. *The R Project for Statistical Computing*. 24

Xu, H., Sun, J., Russell, B., and Innanen, K. (2016). Porosity prediction using cokriging with multiple secondary datasets. *Geoconvention. Optimizing Resources*. 1, 7

Yarus, J. M. and Chambers, R. L. (1994). *Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies*. American Association of Petroleum Geologists. 1

Yenwongfai, H. D., Mondol, N. H., Faleide, J. I., Lecomte, I., and Leutscher, J. (2017). Prestack inversion and multiattribute analysis for porosity, shale volume, and sand probability in the havert formation of the goliat field, southwest barents sea. 5(3). 1, 6

Appendix A

Copula-based dependence measures.

A.1 The linear or Pearson's correlation coefficient

According to Erdely (2011) y Ingle-González (2016), the linear or Pearson's correlation coefficient between two random variables X_1 and X_2 is defined by

$$r(X_1, X_2) := \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)Var(X_2)}} \tag{A.1}$$

where $Cov(X_1, X_2) = \mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$ is the covariance of (X_1, X_2), and $Var(X_1)$ and $Var(X_2)$ are the variances of X_1 and X_2.

Which is likewise expressible in terms of its underlying copula:

$$r(X, Y) := \frac{1}{\sqrt{Var(X)Var(Y)}} \iint_{\mathbb{I}^2} [C_{12}(u, v) - uv] dF_1^{-1}(u) dF_2^{-1}(v). \tag{A.2}$$

Embrechts et al. (1999) cited by Erdely (2011) states a set of elements to consider when using linear correlation:

1. Linear correlation is simply a scalar measure of dependency. It cannot tell us everything we would like to know regarding the risk dependency structure.
2. The possible values of the linear correlation depend on the marginal distributions of the risks. All values between -1 and $+1$ are not necessarily achievable.
3. Risks with a perfect positive dependency do not necessarily have a linear correlation equal to $+1$; Risks with a perfect negative dependency do not necessarily have a linear correlation equal to -1.
4. Linear correlation equal to zero does not imply independence of risks.

5. Linear correlation is not invariant under increasing risk transformations. For example, $logX$ and $logY$ generally do not have the same linear correlation as X and Y.
6. Correlation is only defined when the risk variances are finite, so it is not a suitable dependency measure for heavy-tailed risks where the variances are infinite.

A.2 The association or concordance measure

Informally, a pair of random variables are concordant if the large values of one tend to be associated with large values of the other and the small values of one with small values of the other. To be more precise, let (x_{1i}, x_{2i}) and (x_{1j}, x_{2j}) denote two observations of a continuous random vector (X_1, X_2). We say that (x_{1i}, x_{2i}) and (x_{1j}, x_{2j}) are concordant if $x_{1i} < x_{1j}$ and $x_{2i} < x_{2j}$, or if $x_{1i} > x_{1j}$ and $x_{2i} > x_{2j}$. Similarly, we say that (x_{1i}, x_{2i}) and (x_{1j}, x_{2j}) are discordant if $x_{1i} < x_{1j}$ and $x_{2i} > x_{2j}$, or if $x_{1i} > x_{1j}$ and $x_{2i} < x_{2j}$. Note the alternative formulation: (x_{1i}, x_{2i}) and (x_{1j}, x_{2j}) are concordant if $(x_{1i} - x_{1j})(x_{2i} - x_{2j}) > 0$ and discordant if $(x_{1i} - x_{1j})(x_{2i} - x_{2j}) < 0$ (Nelsen, 2006).

- **Kendall's tau**

  Let $\{(x_{11}, x_{21}), (x_{12}, x_{22}), ..., (x_{1n}, x_{2n})\}$ denotes a random sample of n observations of a continuous random vector (X_1, X_2). There are $\binom{n}{2}$ distinct pairs (x_{1i}, x_{2i}) y (x_{1j}, x_{2j}) of observations in the sample and each pair is either concordant or discordant. Let c denote the number of concordant pairs and d the number of discordant pairs. So Kendall's tau for the sample is defined as

$$t = \frac{c-d}{c+d} = \frac{(c-d)}{\binom{n}{2}} \tag{A.3}$$

  Let (X_1, X_2) and $(\hat{X}_1, \hat{X}_2)$ be independent ramdom vectors of continuous random variables with joint distribution functions F_{12} and $\hat{F}_{12}$, respectively, with common margins F_1 (of X_1 and $\hat{X}_1$) and F_2 (of X_2 and $\hat{X}_2$). Let C_{12} and $\hat{C}_{12}$ denote the copulas of (X_1, X_2) and $(\hat{X}_1, \hat{X}_2)$, respectively, so that $F_{12}(x_1, x_2) = C_{12}(F_1(x_1), F_2(x_2))$ and $\hat{F}_{12}(x_1, x_2) = \hat{C}_{12}(F_1(x_1), F_2(x_2))$. Kendall's tau is defined as the probability of concordance minus the probability of discordance.

$$\tau = \tau_{12} = \tau_{X_1X_2} = P[(X_1 - \hat{X}_1)(X_2 - \hat{X}_2) > 0] - P[(X_1 - \hat{X}_1)(X_2 - \hat{X}_2) < 0] \tag{A.4}$$

  Let X_1 and X_2 be continuous random variables whose copula is C_{12}. Then, Kendall's tau for X_1 and X_2 (which we will indicate with τ_{12} or $\tau_{C_{12}}$) is given by:

$$\tau_{12} = \tau_{C_{12}} = 4 \iint_{\mathbb{I}^2} C_{12}(u, v) dC_{12}(u, v) - 1 \tag{A.5}$$

- **Spearman's rho**

  As with Kendall's tau, the measure of association known as Spearman's rho is based on concordance and discordance. Now let (X_1, X_2), $(\hat{X}_1, \hat{X}_2)$ and $(\check{X}_1, \check{X}_2)$ be three independent random vectors with a common joint distribution function F_{12} (whose margins are again F_1 and F_2) and the copula C_{12}. The rho_{X_1,X_2} of Spearman's rho is defined to be proportional to the probability of concordance minus the probability of discordance for the two vectors (X_1, X_2) and $(\hat{X}_1, \check{X}_2)$, that is, a pair of vectors with the same margins, but one vector has the distribution function F_{12}, while the components of the other are independent.

$$\rho_{X_1,X_2} = 3(P[(X_1 - X_2)(\hat{X}_1 - \check{X}_2) > 0] - P[(X_1 - X_2)(\hat{X}_1 - \check{X}_2) < 0]) \tag{A.6}$$

  Let X_1 and X_2 be continuous random variables whose copula is C_{12}. Then, Spearman's rho for X_1 and X_2 (denoted by ρ_{X_1,X_2} or $\rho_{C_{12}}$) is given by

$$\rho_{X_1,X_2} = \rho_{C_{12}} = 12 \iint_{\mathbb{I}^2} uv dC_{12}(u, v) - 3, \tag{A.7}$$

$$= 12 \iint_{\mathbb{I}^2} C_{12}(u, v) du dv - 3. \tag{A.8}$$

  Another interpretation of Spearman's rho can be obtained from its representation in A.8. The integral in that expression represents the volume under the graph of the copula and over the unit square, and hence ρ_C is a scaled volume under the graph of the copula (scaled to lie in the interval $[-1, 1]$). Indeed, A.8 can also be written as

$$\rho_{C_{12}} = 12 \iint_{\mathbf{I}^2} [C_{12}(u, v) - uv] du dv \tag{A.9}$$

  so that $\rho_{C_{12}}$ is proportional to the signed volume between the graphs of the copula $\rho_{C_{12}}$ and the product copula Π. Thus $\rho_{C_{12}}$ is a measure of average distance between the distribution of X_1 and X_2 (as represented by $\rho_{C_{12}}$) and independence (as represented by the copula Π) (Nelsen, 2006).