THE UNIVERSITY *of* EDINBURGH

Calibrating Rock Physics Models Using Physics Guided Neural Networks

Mohammed Alkharji

Doctor of Philosophy

The University of Edinburgh

2024

Declaration

I, Mohammed Alkharji, certify that this thesis, and the work presented within, is my own original composition. Where it draws on the work of others, such contributions are explicitly acknowledged at the appropriate points within the text. This work has not been previously submitted for any other degree or professional qualification.

Mohammed Alkharji,
May, 2024

Abstract

Rock physics models characterize reservoirs by linking reservoir properties to their elastic response. The accuracy of these quantified properties depends on selecting and calibrating a model that accurately represents this underlying relationship. Complex reservoirs, such as anisotropic ones, require models with calibration parameters that are often difficult or costly to obtain. This challenge has created a barrier to applying certain rock physics models to characterize complex real-world reservoirs.

The variation in these calibration parameters can significantly influence the modeled elastic properties, which, in turn, affect seismic responses. For instance, variations in fluid distribution within the reservoir can alter the measured velocity, while anisotropy in shale layers can impact amplitudes versus offset (AVO) analysis. Therefore, it is reasonable to assume that the effects of these unknown calibration parameters are implicitly captured in the seismic response. However, due to the complex interplay of factors such as porosity, lithology, and fluid content, these parameters cannot be directly inferred from seismic data.

Neural networks are powerful tools for addressing geophysical challenges, particularly in identifying complex patterns and relationships between seismic attributes and reservoir properties. However, they require large training datasets that consist of measured reservoir properties and their seismic responses to adjust their weights accurately and capture the underlying intricate relationships. Such datasets are rarely available in real field data.

Physics-guided neural networks (PGNNs) have emerged as a promising approach, utilizing customized loss functions to penalize unrealistic predictions based on physics equations and models. This thesis proposes a novel PGNN workflow that facilitates the application of uncalibrated rock physics models, using a limited set

of basic field measurements to constrain the training process. The workflow trains the network on two weighted datasets. The first dataset is generated from uncalibrated rock physics models, allowing the network to learn the underlying relationship between known and unknown model parameters and their impact on seismic responses. The second dataset comprises field data points—such as well log measurements of porosity and saturation, along with their seismic responses—which inherently contain information about unknown model parameters, that the network can extract to help constrain the training. A hyperparameter (λ) controls the influence of each dataset in the training process.

This workflow introduces a groundbreaking approach that enables the characterization of complex reservoirs without costly calibration measurements, thereby reducing exploration risks. By analyzing variations in seismic responses, it captures information about unknown calibration parameters and optimizes the prediction of reservoir properties, even for complex reservoir conditions.

Additionally, the workflow addresses the data scarcity issue that often hinders the application of machine learning in reservoir characterization during the early stages of exploration. It provides an alternative by using rock physics models as a data augmentation tool, allowing the network to be properly trained on the underlying relationships between reservoir properties and seismic responses, even with limited field data.

Furthermore, the workflow integrates and capitalizes on field measurements as they become available, enhancing prediction accuracy over time. By incorporating these measurements into the training process, the workflow ensures continuous calibration and improved predictive performance. This workflow has shown enhanced prediction accuracy for synthetic models of dispersive and vertical transverse isotropy (VTI) reservoirs compared to networks trained on limited measurements or uncalibrated rock physics models.

The workflow relies on three key assumptions. First, variations in unknown parameters influence seismic responses. This is evident with the effects of anisotropy on AVO analysis, where changes in these parameters can significantly alter seismic data. Second, seismic modeling of rock physics data can replicate field gathers, enabling the transfer of learned relationships between reservoir properties and seismic signatures from modeled data to actual field data. Many rock physics applications use synthetic modeling to quantify field observations based on modeled elastic properties. Third, I assume that unknown parameters remain consistent across field measurements. This consistency enables the network to accurately isolate their impact and calibrate accordingly. This assumption is valid for geographically constrained reservoirs formed by similar geological processes, where these parameters are likely uniform and can be reliably isolated.

Lay Summary

Rock physics models characterize reservoirs by transforming elastic properties into reservoir properties. The accuracy of this transformation relies on selecting and calibrating an appropriate rock physics model that represents the underlying relationship between reservoir and elastic properties accurately. However, some models require calibration parameters that are difficult or costly to obtain, limiting their practical application. Physics-guided neural networks (PGNNs) are a special type of neural network that leverage physics constraints to guide training and improve predictions. This thesis introduces a novel variation of PGNN that facilitates the use of uncalibrated rock physics models with a limited set of basic field measurements to enhance reservoir property estimation. This approach trains the network to understand the underlying physics relationship between model parameters and their seismic impact, using data from uncalibrated rock physics models. The limited field measurements and their corresponding seismic amplitudes further constrain this training, employing the implicit physics relationship that links reservoir measurements to seismic amplitude. The workflow was tested on two types of reservoirs—dispersive and anisotropic—that typically require calibration and demonstrated enhanced prediction accuracy of reservoir properties.

"I dedicate this work to my parents, who have been my guiding light, to my brothers and sisters for their unwavering encouragement, and to my wife and children, whose love and understanding have been my greatest source of strength and inspiration throughout this journey."

Acknowledgment

I extend my deepest gratitude to my supervisors, Dr. Mark Chapman and Dr. Giorgos Papageorgiou, for their unwavering guidance, support, and patience throughout my PhD journey. Their invaluable mentorship and corrections of my countless learning mistakes have been instrumental in reaching this stage.

I am also grateful to Dr. Andrew Curtis for his insightful feedback on my research and to my advisor, Dr. Anton Ziolkowski, for his ongoing support.

In addition, I wish to express my heartfelt appreciation to the late Dr. Sven Treitel and Dr. John Queen for their invaluable insights shared through numerous discussions and conversations.

Lastly, I would like to acknowledge Saudi Aramco for providing me with the opportunity and sponsoring my PhD thesis at the University of Edinburgh. Their support has been crucial in enabling me to pursue my research goals and academic aspirations.

Contents

List of Figures

List of Tables

Parameters	Description	Units
ρ^o	Mineral density	gm/cm^3
V_p^*	Reference p-wave velocity	km/sec
V_s^*	Reference s-wave velocity	km/sec
ϕ^*	Reference porosity	-
σ^*	Reference crack density	-
ρ^*	Reference bulk density	gm/cm^3
S_w^*	Reference saturation	-
σ	Crack density	-
ρ_g	Gas density	gm/cm^3
ρ_w	Water density	gm/cm^3
K_g	Gas bulk density	GPa
K_w	Water bulk density	GPa
η_g	Gas viscosity	cP
η_w	Water viscosity	cP

Table 1: Papageorgiou and Chapman (2017) model parameter's description.

Property	Shale	Sandstone	units
ρ^o	2.60	2.65	g/cm^3
V_p^*	3.00	3.80	km/s
V_s^*	1.50	2.1	km/s
ϕ^*	0.15	0.22	-
ϵ^*	1×10^{-4}	1×10^{-4}	-
ω_0	5	5	Hz
ρ^*	2.37	2.30	gm/cm^3
S_w^*	1.0	1.0	-
ρ_g	0.11	0.11	gm/cm^3
ρ_w	1.040	1.040	gm/cm^3
K_g	0.20	0.20	GPa
K_w	2.0	2.0	GPa
η_g	0.02	0.02	Pa S
η_w	0.47	0.47	Pa S

Table 2: Reference rock and fluid parameters of shale and sandstone layers. Parameter descriptions can be found in Table 1.

Acronym	Definition
PGNN	Physics-Guided Neural Network
AVO	Amplitude Versus Offset
VTI	Vertical Transverse Isotropy
RMSE	Root Mean Squared Error
SD	Standard Deviation
SNR	Signal to Noise Ratio
VSP	Vertical Seismic Profile
DHI	Direct Hydrocarbon Indicators

Table 3: List of Acronyms

1 Chapter 1: Introduction

Rock physics models are essential tools for predicting reservoir properties such as lithologies, fluid content, and porosity from seismic attributes (Avseth et al., 2005). Using a set of key reservoir parameters, models are based on equations and assumptions about the underlying physic relationships between these parameters and their impact on elastic properties (Mavko et al., 1998). The accuracy of these predictions hinges on selecting the appropriate model that accurately represents the reservoir's physics and calibrating its parameters accordingly (Avseth et al., 2012). This requires understanding whether the reservoir is isotropic or anisotropic and whether it exhibits dispersive or elastic behavior, as these characteristics significantly impact model selection and calibration.

Calibration of rock physics models involves comparing ground truth data from logs and core analysis with elastic responses predicted by models (Dvorkin et al., 1995; Avseth et al., 2000; Avseth et al., 2005). However, in complex reservoirs, the relationship between parameters and modeled elastic responses is often intricate and costly to obtain (Avseth et al., 2012; Bosch et al., 2010). Anisotropic models, for instance, require elastic property measurements from multiple directions, but most logging occurs vertically. Dispersive models need a characteristic frequency (f_c) or time-scale constant (τ) data (Mavko and Mukerji, 1998), which require broadband measurements of rock samples, posing additional challenges (Payne et al., 2007).

Advanced core analysis can help obtain reservoir property measurements, but replicating in-situ conditions and scaling data across core, log, and seismic datasets is challenging (Dewhurst et al., 2011; Bredesen et al., 2015). Advanced log measurements and vertical seismic profiles (VSP) can provide information about calibration parameters, but they are often costly and unavailable during early exploration. Furthermore, the number of calibration parameters needed to represent elastic responses increases with reservoir complexity (Tsvankin et al., 2010). For

example, isotropic rocks are described by bulk and shear moduli, while the presence of horizontal or vertical fractures increases the number of elastic constants to five, and two orthogonal fracture sets may require nine (Tsvankin, 2012). These parameters often necessitate advanced tools and result in additional costs, which may not be feasible at early exploration stages.

Thomsen (1993) formulated an equation to calculate offset-dependent reflection coefficients in a vertical transverse isotropy (VTI) medium, demonstrating that anisotropic parameters significantly impact seismic amplitudes at near and far offsets. Similarly, Chapman et al. (2006) numerically showed that dispersion within the seismic frequency range leads to frequency-dependent reflection coefficients, which can be used to quantify reservoir properties. These findings suggest that the impacts of anisotropy and dispersion on calculated or measured reflectivity could potentially be leveraged to estimate anisotropic and dispersive model parameters directly from seismic data.

However, due to the complex interplay between reservoir parameters such as porosity, lithology, and fluid content on the seismic signal, estimating these calibration parameters is often difficult (Sen, 1995). For instance, Payne et al. (2007) attempted to quantify permeability from measured attenuation but encountered challenges due to uncertainties associated with multiple input model parameters. The interactions between these parameters led to multiple combinations producing similar attenuation observations. Additionally, Xiao (2006) discussed various methods to estimate anisotropic parameters from seismic data but concluded that achieving accurate estimates requires larger acquisition offsets and measured sonic velocities, which may not be feasible in all scenarios.

As a result of these challenges, calibration difficulties have led most rock physics applications to rely on simplified models with minimal calibration parameters that do not accurately characterize reservoirs (Han and Batzle, 2004; Avseth et al.,

2010). Consequently, these difficulties have hindered the widespread application of advanced rock physics models.

Neural networks have demonstrated their effectiveness in extracting hidden patterns between input features and target labels (Bishop, 1994). They have been used to solve various geophysical challenges (Röth and Tarantola, 1994; Maleki et al., 2014; Lu et al., 2018; Russell, 2019). However, training a network to extract the underlying relationships between extracted seismic attributes and key reservoirs parameters requires large training dataset, which might be unavailable. This data scarcity problem has led researchers to explore alternative approaches to effectively apply neural networks.

Recently, physics-guided neural networks (PGNNs) have emerged that integrate physics equations into the training process to enhance the accuracy of predictions (Karpatne et al., 2017b). By incorporating physics models into their loss functions, these networks penalize unrealistic predictions, ensuring that their outcomes align with known physical principles. They traditionally address issues like data scarcity by incorporating seismic forward modeling operators into the loss function (Alfarraj and AlRegib, 2019) or leveraging rock physics models with known calibration parameters (Das et al., 2019). They can also be used to train on large dataset with noisy features or incorrect labels (Karpatne et al., 2017a).

If calibration parameters influence the response of elastic properties, which then affect the extracted seismic response, a complex relationship likely exists between unknown calibration parameters and seismic responses. This research hypothesizes that a neural network could learn this relationship by creating a training dataset where reservoir properties like porosity and saturation serve as target labels and seismic attributes extracted from modeled gathers with varying combinations of model parameters act as features.

In this thesis, I develop a novel workflow using physics-guided neural networks to address the complex calibration challenges associated with anisotropic and dispersive rock physics models. Gas saturation estimation using dispersive seismic attributes has been shown to be sensitive to a 'characteristic frequency' (Wu et al., 2014). My approach improves neural network training by simulating calibrated models, enabling accurate estimates of porosity and saturation from dispersive gathers, despite the unknown value of this 'characteristic frequency.' This method is also applied to amplitude versus offset (AVO) analysis in the presence of VTI anisotropy, where inversion results strongly depend on unknown Thomsen parameters. The workflow learns these implicit Thomsen parameters from a small number of field measurements, enhancing predictive accuracy for porosity and saturation.

This workflow introduces an innovative approach to characterizing complex reservoirs without the need for costly calibration measurements, thereby reducing exploration risks. By analyzing variations in seismic responses, it captures information about unknown calibration parameters and optimizes the prediction of reservoir properties, even under complex reservoir conditions. The workflow also addresses data scarcity in early exploration stages by using rock physics models as a data augmentation tool, allowing machine learning networks to be effectively trained on the underlying relationships between reservoir properties and seismic responses, even with limited field data. As new field measurements become available, the workflow continuously incorporates them to enhance prediction accuracy over time through ongoing calibration.

The structure of the thesis is as follows: Chapter 2 outlines the challenges in predicting reservoir properties for fizz-gas and VTI reservoirs. It begins by examining the limitations of Gassmann's equations in distinguishing between low and high gas saturations in unconsolidated sandstones. The chapter then transitions to a discussion on dispersive rock physics models, highlighting the calibration dif-

ficulties associated with them. The second section addresses the challenges of assuming isotropic behavior in anisotropic VTI reservoirs by introducing Thomsen parameters, which simplify the understanding of these reservoirs, and discusses methods to measure these parameters.

Chapter 3 shifts focus to neural networks, detailing their design, training process, and key hyper-parameters. It further introduces physics-guided neural networks, highlighting their applications in geophysics and their relevance to this thesis.

Chapter 4 details the design and testing of the standard neural network for predicting porosity and saturation, assessing its accuracy across various scenarios. It underscores the challenge of predicting reservoir properties from dispersive reservoirs using uncalibrated models, particularly without explicit knowledge of the characteristic frequency.

Chapter 5 presents the theoretical derivation of the Physics-Guided Neural Network workflow, applying it to an uncalibrated dispersive rock physics model and analyzing factors that may affect the prediction accuracy.

Chapter 6 explores the workflow's adaptability in addressing calibration challenges in a VTI medium without explicit knowledge of Thomsen parameters. It also analyze the impact of varying unknown model parameter(s) within field data.

Chapter 7 provides a discussion of the research outcomes and their implications. Additionally, it outlines potential avenues for future research, identifying areas for further investigation and advancement.

In Chapter 8, I list the main findings of the research, providing a comprehensive overview of the key outcomes and contributions made.

2 Challenges in Rock Physics: Dealing with Unknown Calibration Parameters

2.1 Chapter Overview

This chapter describes two primary challenges in reservoir characterization: the fizz-gas inversion challenge and the inversion for reservoir properties using amplitude versus offset (AVO) analysis in vertical transverse isotropy (VTI) reservoirs. It demonstrates that while modeling solutions for such reservoirs exist, a common challenge is the lack of advanced measurements required to calibrate these models effectively. The discussion includes dispersive rock physics models used to predict gas saturation in dispersive reservoirs and Thomsen parameters that simplify AVO analysis in VTI reservoirs. The chapter highlights the difficulties in acquiring accurate measurements for these models and why they are often rarely applied in real-world scenarios. The chapter also outlines the Papageorgiou and Chapman (2017) model that will be used to generate the elastic properties of geological layers for this thesis.

2.2 Introduction

Rock physics has advanced seismic reservoir characterization in the last two decades, demonstrated by its diverse applications in solving exploration and reservoir characterization challenges (Avseth et al., 2005; Dvorkin et al., 2014). For example, Smith et al. (2003) utilized Gassmann (1951)'s equations for fluid content characterization distant from well locations. Avseth et al. (2000) applied the contact-cement model to map reservoir porosity from velocity predictions. Grana et al. (2017) employed various rock physics models for facies characterization in carbon dioxide (CO_2) storage, considering porosity and permeability. Li and Li (2018) implemented multiple rock physics models for inverting the reservoir brittleness index in China's Fuling shale gas field, Sichuan Basin.

While rock physics inversion workflows have been successful, several challenges

in reservoir characterization remain. Examples include distinguishing high from low gas saturation in unconsolidated sandstone, often referred to as fizz-gas (Han and Batzle, 2002), analyzing AVO curves in an anisotropic reservoir (Kim et al., 1993), characterizing fractured reservoirs from seismic amplitudes (Liu et al., 2010), and assessing total organic content and reservoir brittleness in shale plays (Liu and Sun, 2015).

Various rock physics models have been developed to address some of these challenges. However, their practical application often encounters limitations due to the difficulty of their model calibration (Tsvankin, 1996). For instance, laboratory studies show that attenuation, which is the decay of seismic amplitude with frequency (Aki and Richards, 1980), and dispersion, the variation of elastic properties with frequency (Yilmaz, 2001), change more noticeably with gas saturation (Domenico, 1975; Winkler and Nur, 1979). This led to the creation of various dispersive rock physics models (O'Connell and Budiansky, 1977; Chapman et al., 2002; Papageorgiou and Chapman, 2017). A major challenge with these models is the calibration of a parameter called τ, which is the inverse of a characteristic frequency where maximum attenuation occurs due to fluid movement induced by a passing wave (Dvorkin et al., 1995). Measuring τ is difficult with standard logging and core analysis methods (Chapman et al., 2002). For VTI reservoirs, Thomsen (1986) introduced three parameters for simpler tensor value relations in VTI anisotropic reservoirs. But accurately estimating these parameters through core analysis can be challenging, especially in replicating in-situ reservoir conditions (Dewhurst et al., 2011).

One approach to the calibration problem is to select a simpler rock physics model with calibration parameters that can be obtained from basic well log measurements (Avseth et al., 2010). However, this carries the risk that the model would not be simulating the correct physics (Gurevich et al., 2010). An alternative approach is to calibrate models based on unrepresentative measurements (Maultzsch

et al., 2003). These approaches often carry significant uncertainties because predicted elastic properties vary greatly between rocks due to differences in depth, temperature, cementation, porosity, fluid content, and mineralogy (Schön, 2015). Therefore, selecting the appropriate rock physics model that accurately represents the reservoir's physics, while using calibration measurements from the same or similar reservoir, is crucial in quantifying reservoir properties (Avseth et al., 2005).

Next, I discuss the challenge of inverting for gas saturation in weakly consolidated dispersive sandstone when using simplified rock physics models.

2.3 Fizz Gas Inversion Challenge

Gassmann's (1951) equations are a widely utilized rock physics model that establish a connection between fluid saturation and elastic moduli. Using Mavko et al.'s (2020) formulation, Gassmann's equations can be written as:

$$\frac{K_{sat}}{K_o - K_{sat}} = \frac{K_{dry}}{K_o - K_{dry}} + \frac{K_{fl}}{\phi(K_o - K_{fl})} \tag{2.1}$$

$$\mu_{dry} = \mu_{sat} \tag{2.2}$$

where K_{sat} is the saturated bulk modulus, K_{dry} is bulk modulus of the dry framework, K_0 is bulk modulus of the mineral, K_{fl} is bulk modulus of the fluid, ϕ is porosity, μ_{dry} is the dry shear modulus, and μ_{sat} is the saturated shear modulus. For a mixture of gas and water, the fluid bulk modulus K_{fl} is calculated using Wood's averages (Wood and Lindsay, 1956), which assume a uniform distribution of fluids within the pores, as shown below:

$$K_{fl} = \frac{1}{\frac{S_g}{K_g} + \frac{S_w}{K_w}} \tag{2.3}$$

where S_w is the water saturation, S_g is gas saturation, K_g is the bulk modulus of the gas, and K_w is the bulk modulus of the water.

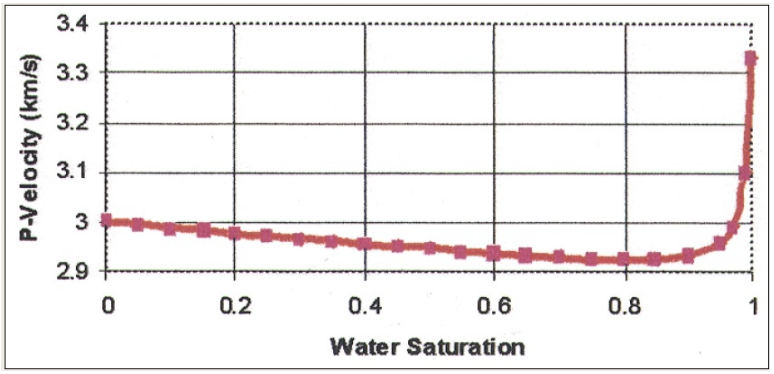


Figure 2.1: Velocity-saturation curve illustrating the fizz-gas challenge, with saturation <90% showing similar velocity responses. Adapted from Han and Batzle, 2002.

Using Gassmann's equations, introducing a small percentage of gas saturation into an unconsolidated sandstone wet reservoir leads to a significant reduction in the rock's velocity, as illustrated in Figure 2.1 (Domenico, 1975; Han and Batzle, 2002). This reduction complicates distinguishing between small and commercial gas saturations, particularly in shallow gas reservoirs (Han and Batzle, 2005). The challenge stems from the similar velocities observed across a wide saturation range (less than 90%). Consequently, standard interpretation tools like AVO and direct hydrocarbon indicators (DHIs) often yield similar responses for both low and high gas saturation levels, making accurate estimation of fluid saturation difficult (Avseth et al., 2005). Efforts to address this include employing multi-component data (Zhu et al., 2000) and analyzing extended elastic impedance of P to S converted waves (González et al., 2003), although these methods are costly and not routinely conducted.

However, if the reservoir is dispersive, dispersion could be utilized to enhance predictions of gas saturation using dispersive rock physics models. Many lab experiments showed that fluid flow within the reservoir can cause velocity dispersion within the seismic bandwidth (<100 Hz). Batzle et al. (2006b) demonstrated that velocity dispersion is highly correlated to fluid mobility, defined as the ratio of permeability to viscosity. Heavy oils, being highly viscous, can cause sig-

nificant attenuation at seismic frequencies. Behura et al. (2007) observed peak dispersion/attenuation of signal at seismic frequencies in heavy oil due to high viscosity. Similar effects were noted by Batzle et al. (2006a), Das and Batzle (2008) and Boitnott et al. (2011).

Patchy saturation, where fluids distribution within reservoirs are non-uniform (Cadoret et al., 1998), can lead to attenuation within the seismic frequency range due to squirt flow between different fluid phases (Müller et al., 2010). This phenomenon is expected in CO_2 storage projects where gas is injected into water-bearing aquifers, causing non-uniform fluid distribution (Kim et al., 2013). Caspari et al. (2011) used time-lapse sonic log measurements in a CO_2 sequestration project to estimate patch size and relate velocity to CO_2 saturation at ultrasonic frequencies, suggesting that similar relationships could be observed at seismic frequencies with larger patches. Sengupta et al. (2003) demonstrated that patchy saturation modeling in a gas-injected field provides a better match with field-observed seismic responses than models with uniformly distributed water saturation.

Numerically, White (1975) predicted dispersion at seismic frequencies ($<$70 Hz) using patchy saturation models. Additionally, Maultzsch et al. (2003) modeled dispersion within the seismic frequency range when large fractures ($>$1 m) were present.

These observations underscore the importance of advanced dispersive rock physics models for accurately quantifying reservoir properties. By incorporating complex reservoir settings such as velocity dispersion from fluid mobility, heavy oil viscosity, patchy saturation, and fractures, these models enable more precise predictions of reservoir characteristics.

2.4 Squirt Flow Models

Early lab experiments revealed that adding fluids into dry rock samples introduced velocity dispersion (Born, 1941). Biot (1956) attempted to model this elastic behavior across frequencies. Yet, subsequent lab studies indicated that the high-frequency limit of Biot's model often underestimated velocity dispersion, suggesting alternative attenuation mechanisms (Winkler, 1985; Winkler, 1986; Wang and Nur, 1988).

Squirt flow models offer an alternative explanation for attenuation, which occurs due to fluid movement between compliant cracks and stiffer pores as waves pass through (Murphy et al., 1984). Murphy (1982) demonstrated that varying gas saturation levels produce different attenuation measurements, suggesting that attenuation could be used to estimate gas saturation. Similarly, Klimentos and McCann (1990) found that P-wave attenuation in sandstones increases with clay content and developed an empirical relationship to quantify attenuation based on porosity and clay content within the rock. Additionally, Domenico (1975) and Batzle et al. (2006a) showed that squirt flow between adjacent fluids can occur due to a non-uniform fluid distribution, known as patchy saturation, within reservoirs. This finding suggests the potential for using squirt flow models to quantify fluid saturation and distribution.

The development of squirt flow models has advanced along two distinct paths, each characterized by different approaches to representing rock pore space (Wu et al., 2014). One approach (Hsu et al., 1991; Dvorkin et al., 1995; Gurevich et al., 2010) avoids explicit mathematical descriptions of pore structures, instead emphasizing averaged bulk properties of the rock. The advantage of this method is its simplified mathematical representation of rock response. However, it may lack the accuracy needed to capture the complexity of pore systems essential for quantifying reservoir properties. In contrast, the second approach (Hudson et al., 1996; Chapman et al., 2002; Chapman, 2009) uses inclusion models to represent

the squirt effect, relying on precise mathematical descriptions of pore stiffness. The limitation of this method is that it often assumes idealized geometries rarely encountered in natural formations (Avseth et al., 2010). For a detailed analysis of these models' advantages and limitations, refer to Chapman et al., 2002, Wu et al., 2014, and Mavko et al., 2020.

White (1975) modeled the impact of patchy saturation using two concentric spheres representing regions with different fluid saturations to predict dispersive behavior in elastic moduli at seismic frequencies. The model assumes fluid interactions at a mesoscale level, where the fluid heterogeneities are larger than the grain scale but smaller than the seismic wavelength (Payne et al., 2007). However, in White's model, the impact of pore structure on fluid flow is ignored (Mavko et al., 2020).

Papageorgiou and Chapman (2017) presented a significant advancement in squirt flow models by developing a model that integrates wave-induced fluid movement between compliant cracks and pores while also considering patchy saturation. This model offers a more comprehensive approach to capturing the complex interactions resulting from squirt flow between compliant cracks and stiff pores, in addition to squirt flows caused by patchy saturation. Although the model still assumes idealized pore geometries—which may not fully represent natural formations—it nonetheless marks an important step forward in accurately modeling these intricate mechanisms. The model will be described in detail later in this section.

Mavko and Mukerji (1998) investigated various dispersion mechanisms, categorizing the behavior into three frequency regimes. In the low-frequency regime, minimal attenuation and low apparent velocity occur due to long wavelengths that allow pore pressure to equalize, resulting in a state known as the relaxed-state. In contrast, the high-frequency regime exhibits negligible attenuation but high

velocity, as short wavelengths prevent pore pressure from equalizing, leading to increased apparent incompressibility, referred to as the unrelaxed-state. Between these two regimes lies a transition zone, where both dispersion and attenuation become significant. The characteristic frequency, denoted as f_c or sometimes its inverse τ, marks the frequency at which maximum attenuation occurs.

Squirt flow model parameters can be derived from basic log measurements such as porosity, saturation, and density, with additional parameters like permeability obtained from core analysis (Ahmed et al., 1991). However, accurately calibrating the characteristic frequency (f_c), or its inverse (τ), is challenging due to its dependence on multiple factors such as fluid viscosity, rock permeability, grain size, and pore geometry (Dvorkin et al., 1994; Chapman et al., 2002). Precise determination of τ is crucial for modeling elastic behavior in dispersive reservoirs and for effectively applying rock physics in reservoir properties inversion (Maultzsch et al., 2003; Payne et al., 2007).

Current methods for calibrating τ involve fitting laboratory broadband measurements of elastic properties from a reference reservoir sample to a dispersive rock physics model (Dvorkin et al., 1995). These models can yield varying τ estimates, reflecting their different assumptions about reservoir conditions and pore geometry (Payne et al., 2007; Moyano et al., 2011; Wu et al., 2014). An added layer of uncertainty arises when the lab samples used for calibrating τ do not represent the actual reservoir rocks. For instance, due to the unavailability of appropriate broadband measurements on reservoir rock samples, Maultzsch et al. (2003) used synthetic sandstone measurements to calibrate τ for fracture characterization. This reliance on synthetic samples may introduce discrepancies between model predictions and real reservoir behavior. Similarly, Payne et al. (2007) highlight the challenge of accurately measuring τ because it depends on difficult-to-measure parameters like crack density. The difficulty in obtaining representative samples and measuring critical parameters adds significant uncertainty to τ estimation,

consequently hindering the application of squirt flow models to accurately characterize real field scenarios. As a result, estimating τ remains an active area of research (Wu et al., 2022).

Despite these challenges, employing squirt flow models to invert for gas saturation has been numerically demonstrated. Wu et al. (2014) utilized Chapman et al. (2002)'s model in a cross-plot technique involving Gassmann-equivalent reflectivity and reflectivity gradient with frequency for Bayesian inversion of porosity and saturation on synthetic data with added noise, yielding accurate predictions. Jin et al. (2017) developed a convolutional modeling workflow that enables seismic modeling of dispersive elastic moduli. They used Chapman et al. (2002)'s model to calculate moduli and generate synthetic gathers. Bayesian inversion, based on modeled and observed gathers, provided precise estimations of porosity and saturation for various gas saturation levels. Both inversion workflows assumed knowledge of τ.

In this thesis, I aim to address the challenge of needing a known τ value to accurately predict reservoir properties using squirt flow models. Specifically, in Chapter 5, I will apply the Papageorgiou and Chapman (2017) squirt flow model along with my physics-guided neural network (PGNN) training workflow to predict porosity and saturation from synthetic gathers without prior knowledge of the key model parameter τ. This approach removes a primary barrier to applying squirt flow models for characterizing complex real-world reservoirs, enabling more accurate reservoir characterization.

Chapman et al. (2002) derived a formulation to calculate the effective bulk and shear moduli, $K_{eff}(\omega)$ and $\mu_{eff}(\omega)$, respectively at different frequencies (ω) assuming full saturation, random connected cracks with uniform radius size and spherical shape pores, see Appendix A. Papageorgiou and Chapman (2017) modified the model to allow for multiple fluids with non-uniform patchy saturation

by introducing a hyperparameter q such that:

$$K_f = \frac{q}{\frac{S_w}{K_w} + \frac{qS_g}{K_g}}, \text{ where } \quad \frac{K_g}{K_w} < q \leq 1, \tag{2.4}$$

$$\frac{1}{\overline{\eta}} = \frac{k_w}{\eta_w} + \frac{k_g}{\eta_g}, \tag{2.5}$$

where S_w and S_g are the water and gas saturations, respectively; K_w and K_g are the bulk moduli of water and gas, respectively; η_w and η_g are the viscosities of water and gas; k_w and k_g are the relative permeabilities of water and gas; and $\overline{\eta}$ is the average fluid viscosity. Setting $q = 1$ results in uniform mixing (Wood and Lindsay, 1956).

The dispersive seismic velocities are determined as follows:

$$V_p(\omega) = \sqrt{\frac{\text{Re}(K(\omega)) + \frac{4}{3}\text{Re}(\mu(\omega))}{\rho}}, \tag{2.6}$$

$$V_s(\omega) = \sqrt{\frac{\text{Re}(\mu(\omega))}{\rho}}, \tag{2.7}$$

where V_p and V_s represent P-wave and S-wave velocities, respectively, and ρ is the bulk density, calculated as:

$$\rho = (1 - \phi)\rho_m + \phi(\rho_w S_w + \rho_g S_g). \tag{2.8}$$

Here, ρ_w and ρ_g denote the densities of water and gas, respectively, while ρ_m is the matrix density.

2.5 VTI Anisotropy

Rock elastic moduli, dictating stiffness and deformation response to stress, are defined by Hooke's Law (Hooke, 1678) as follows:

$$\sigma_{ij} = C_{ijkl}\epsilon_{kl}, \tag{2.9}$$

Here, σ_{ij} is the stress applied to the rock, ϵ_{kl} is the resultant strain, and C_{ijkl} is the elastic tensor linking stress and strain.

In isotropic rocks, where stress-strain responses are uniform regardless of wave direction, the elastic tensor can be represented by two moduli: bulk and shear (Mavko and Mukerji, 1998). However, in seismic anisotropy, where the elastic response varies by wave direction (Grechka, 2009), more elastic tensor values are needed. Mineral alignment, such as in shale (Aleksandrov, 1961; Johnston and Christensen, 1995), or vertical fracturing result in vertical transverse isotropy (VTI), requiring five tensor values. Horizontal fractures cause horizontal transverse isotropy (HTI), also requiring five values (Bakulin et al., 2000). Multiple fracture sets result in orthorhombic anisotropy, which requires nine tensor values (Tsvankin et al., 2010). My research will focus on VTI anisotropy caused by shale layering.

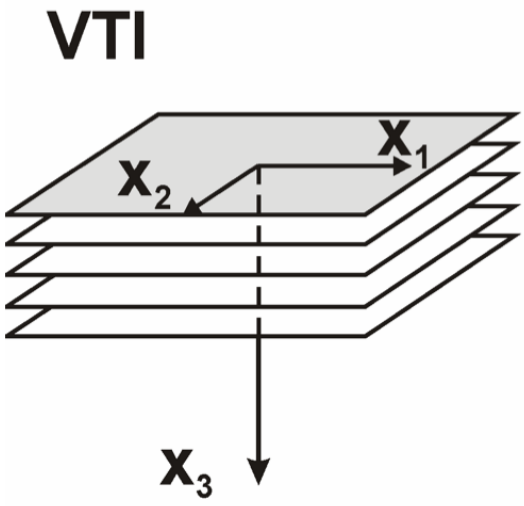


Figure 2.2: VTI model with x_3 as symmetry axis. Figure taken from Danek et al., 2010

Shales are the primary clastic component in sedimentary basins and serve as seals and barriers for most reservoirs (Dewhurst and Siggins, 2006). Neglecting their impact can lead to persistent challenges, such as depth conversion issues in seismic

exploration (Banik, 1984), imaging difficulties (Alkhalifah and Larner, 1994), and complications in amplitude variation with offset analysis (Wright, 1987).

The VTI anisotropy of shale is characterized by the elastic medium having a symmetry axis along $\mathbf{x_3}$, as illustrated in Figure 2.2. This anisotropy can be described using five independent elastic tensor components: C_{11}, C_{33}, C_{44}, C_{66}, and C_{13} (Mavko and Bandyopadhyay, 2009). However, acquiring independent measurements for each of these components is challenging. Experimental studies on shales are notably limited due to several reasons. The extremely friable nature of shales often complicates efforts to preserve samples (Jones and Wang, 1981). In addition, the testing process is time consuming due to the very low permeability of shales, which slows the equilibration of pore pressure throughout the sample (Wang, 2002).

Thomsen (1986) provided three dimensionless parameters that describe the relationship between tensor values in VTI medium as follow:

$$\epsilon = \frac{C_{11} - C_{33}}{2C_{33}}, \tag{2.10}$$

$$\gamma = \frac{C_{66} - C_{55}}{2C_{55}}, \tag{2.11}$$

$$\delta = \frac{(C_{13} + C_{55})^2 - (C_{33} - C_{55})^2}{2C_{33}(C_{33} - C_{55})}, \tag{2.12}$$

ϵ and γ describe the P-wave and SH-wave anisotropy, respectively (Mavko et al., 2020). On the other hand, δ does not have an exact interpretation, but it plays a crucial role in normal moveout (NMO) corrections and reflection amplitudes (Alkhalifah and Tsvankin, 1995; Grechka, 2009). Thomsen assumed weak anisotropy, where $|\epsilon| << 1$ and $|\delta| << 1$ (Tsvankin, 2012), and approximated three equations to calculate P, SV, and SH phase velocities at an incident angle

θ, as follows:

$$V_P(\theta) \approx V_{p_0}(1 + \delta \sin^2\theta(\cos^2\theta + \epsilon \sin^2\theta),) \tag{2.13}$$

$$V_{SV}(\theta) \approx V_{s_0}\left(1 + \left(\frac{V_{p_0}}{V_{s_0}}\right)^2 (\epsilon - \delta)\sin^2\theta\cos^2\theta\right), \tag{2.14}$$

$$V_{SH}(\theta) \approx V_{s_0}(1 + \gamma \sin^2\theta), \tag{2.15}$$

where V_{p_0} and V_{s_0} are vertical P-wave and SV-wave velocities, calculated as:

$$V_{p_0} = \sqrt{\frac{C_{33}}{\rho}}, \tag{2.16}$$

$$V_{s_0} = \sqrt{\frac{C_{55}}{\rho}}. \tag{2.17}$$

Thomsen parameters provide a method for modeling seismic responses in VTI reservoirs. By utilizing only ϵ and δ, along with log measurements of vertical P-wave (V_{p_0}) and S-wave velocities (V_{s_0}), it is possible to accurately model P-P and P-SV reflections in VTI anisotropic media by deriving essential elastic tensors C_{11} and C_{13} (Tsvankin, 1996). These derivations are facilitated by Equations 2.10 and 2.12. SH-waves are excluded because they polarize perpendicular to the incident plane and do not interact with P and SV waves, remaining completely decoupled (Rüger, 1996).

Accurately measuring Thomsen parameters is challenging due to the difficulty of replicating in-situ reservoir conditions (Dewhurst et al., 2011). Laboratory experiments on core plugs can estimate ϵ and δ by measuring velocities at various orientations (Jones and Wang, 1981; Vernik and Nur, 1992; Sarout et al., 2015). While effective, these measurements are costly, and results may vary as Thomsen parameter values change with stress, highlighting the importance of accounting for reservoir conditions (Sarout and Guéguen, 2008).

Advanced dipole sonic logs have been suggested for anisotropy characterization in

deviated wells (Walsh et al., 2007), which provide accurate data but are expensive to implement and rarely feasible in early exploration due to high costs. Walkaway vertical seismic profile (VSP) surveys can also quantify anisotropy by measuring vertical and horizontal phase slowness (Hsu et al., 1991; Grechka and Mateeva, 2007); however, their expense makes them uncommon in initial exploration stages.

For surface seismic data, several inversion methods exist to estimate ϵ and δ in VTI media (Thomsen, 1986; Tsvankin and Thomsen, 1994; Alkhalifah and Tsvankin, 1995). Although less costly, these methods require large offsets, which may not always be available in the seismic survey (Xiao, 2006).

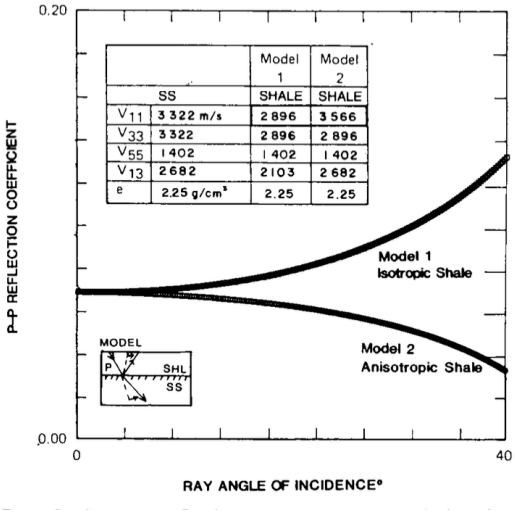


Figure 2.3: Wright (1987) AVO modeling in VTI media showed reversal in gradient behavior compared to isotropic case.

Accounting for anisotropy is critical in AVO analysis because it significantly impacts the amplitude response (Dvorkin et al., 2014). Wright (1987) explored the effect of an anisotropic shale layer over an isotropic sandstone layer, finding that an ϵ value of 0.20 could reverse the AVO curve gradient compared to isotropic scenarios. This reversal complicates the interpretation of AVO anomalies, as shown in Figure 2.3. Similarly, Kim et al. (1993) investigated anisotropy effects on AVO

at shale-sand interfaces, discovering that $\Delta\delta$ (the difference in δ values between layers) significantly affects AVO responses at incidence angles between 20° and 30°, while $\Delta\epsilon$ (the difference in ϵ values) becomes influential at higher angles. He noted that positive $\Delta\delta$ values might falsely suggest significant hydrocarbon accumulations due to increased amplitude. In parallel, Blangy (1994) observed similar AVO effects in VTI media for class II and III sandstone reservoirs, noting that viscoelastic modeling could lead to even greater amplitudes than in isotropic cases. This is due to the fact that at higher angles, there is less attenuation due to less heterogeneity, which leads to more transmitted energy and fewer reflected waves.

Thomsen (1993) approximated the offset-dependent reflection coefficients for VTI media using the Thomsen parameters ϵ and δ, demonstrating that these parameters significantly impact the seismic signal and are captured within the measured seismic reflectivity at interfaces. However, accurately measuring Thomsen parameters is challenging due to the difficulty of replicating in-situ reservoir conditions and limitations in measurement techniques. This challenge introduces uncertainties in standard AVO analysis when predicting reservoir properties.

In Chapter 6, I address this issue using a geological model representing a simplified real-field scenario: a VTI shale layer overlying an elastic sandstone reservoir. Assuming that the Thomsen parameters for the shale layer are unknown reflects a common challenge in real-world reservoir characterization. To overcome this, I apply my novel physics-guided neural network (PGNN) training workflow, which assumes that changes in unknown parameters impact the seismic response (discussed further in Chapter 5, Section 5.4). This assumption is valid in VTI media, where ϵ and δ significantly affect AVO analysis. Using this workflow and basic log measurements along with their corresponding seismic responses, I demonstrate that it improves predictions of reservoir properties within the sandstone layer, even without explicit knowledge of the Thomsen pa-

rameters in the overlying shale. By effectively incorporating anisotropic models without direct parameter measurements, the workflow enhances prediction accuracy and facilitates advanced reservoir characterization techniques when dealing with anisotropic reservoirs.

3 Feed Forward Neural Networks

3.1 Chapter Overview

This chapter introduces feed forward neural networks and covers their components, training cycle and hyperparameters. The chapter then transitions to examining physics-guided neural networks (PGNN), focusing on their current roles in tackling geophysical issues and laying the foundation of my PGNN training workflow.

3.2 Neural Network Model

Suppose a dataset D defined as:

$$D = \{(\mathbf{x_1}, \mathbf{y_1}), (\mathbf{x_2}, \mathbf{y_2}), ..., (\mathbf{x_k}, \mathbf{y_k})\}. \tag{3.1}$$

In D, $\mathbf{x_k} \in R^{(1\times n)}$ denotes a vector of seismic features at the reservoir interface, and $\mathbf{y_k} \in R^{(1\times m)}$ represents corresponding reservoir properties. A neural network transforms these input features, $\mathbf{x_k}$, into output predictions, $\hat{\mathbf{y}}_\mathbf{k}$, that approximate the actual reservoir properties, $\mathbf{y_k}$ (Svozil et al., 1997). This transformation can be described as follows:

$$\mathcal{F}(\mathbf{x_k}) = \hat{\mathbf{y}}_\mathbf{k}, \tag{3.2}$$

where $\mathcal{F}$ is called a forward propagation operator that propagates input features $\mathbf{x_k}$ to output the prediction vector $\hat{\mathbf{y}}_\mathbf{k}$.

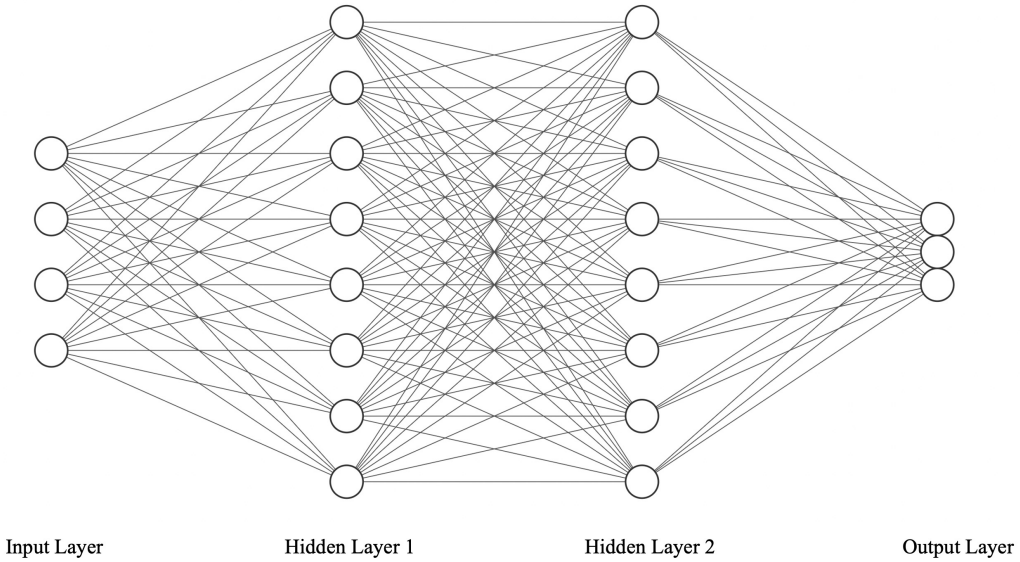


Figure 3.1: Neural Network architecture showing two hidden layers, an input layer with four features , and an output layer with three predictions

A feed forward neural network, often referred to simply as a neural network, consists of a series of layers; see Figure 3.1. The first and last layers are called the input and output layers, respectively, and are sized by the feature and label vectors, respectively. Each layer consists of a number of nodes that combine the outputs from preceding layers using the following equation:

$$s_i^h = \sum_{j=1}^{n} w_{ji} a_j{}^{(h-1)} + b_i, \tag{3.3}$$

Here, a_j^{h-1} represents the output of the j^{th} node from the previous layer, n is the number of nodes in the previous layer, w_{ji} is the weight between the j^{th} node of the previous layer and the i^{th} node of the current layer, and b_i is called the node bias. The total sum s_i^h is then passed through an activation function σ, which is a function that converts the total input of the node to an output based on a predefined mathematical function (discussed in Section 3.4.2), producing a value a_i^h:

$$a_i^h = \sigma(s_i^h). \tag{3.4}$$

The output serves as input for the next layer, progressing until the final layer outputs the prediction vector $\hat{\mathbf{y}}_\mathbf{k}$. The number of hidden layers, nodes, and the type of activation function are determined by the complexity of the prediction task. More complex tasks, like image classification, require deeper networks to analyze the intricate non-linear relationships between various objects within the input features and target labels (Bianchini and Scarselli, 2014).

3.3 Neural Network Training

Neural networks should yield predictions that closely match actual labels on training and unseen datasets (Svozil et al., 1997). To achieve this, the dataset (D) is first segmented into three subsets: D_T for training, D_V for validation, to monitor training convergence, and D_B acts as a blind dataset, for post-training evaluation of prediction accuracy.

Neural networks undergo a training process to refine their predictions by minimizing a function, called a loss function, that measures the difference between the network's predictions and the actual target labels (Goodfellow et al., 2016). Training starts by initializing the network's weights and biases to small random values. Feature vectors from dataset D_T are then propagated through the network, resulting in prediction vectors $\hat{\mathbf{y}}$, using equation 3.2. To evaluate the network's performance, the loss (C) is calculated as the average of the absolute differences between these predictions and the actual reservoir property vectors as follows:

$$C = \frac{1}{z} \sum_{i=1}^{z} |\hat{\mathbf{y}}_\mathbf{i} - \mathbf{y}_\mathbf{i}| \,, \tag{3.5}$$

where z represents the total instances. The loss is minimized through gradient descent (Cauchy, 1847), an iterative process that optimizes the network's weights and biases by reducing the loss function (Géron, 2017). As formulated by Goodfellow et al. (2016), the gradient descent equations used to update the network's

weights and biases are as follows:

$$w_{ji}^{L*} = w_{ji} - \gamma \frac{dC}{dw_{ji}^{L}}, \tag{3.6}$$

$$b_{i}^{L*} = b_{i} - \gamma \frac{dC}{db_{i}^{L}}, \tag{3.7}$$

w_{ji} and b_i represent the current weights and biases, respectively. The terms w_{ji}^{L*} and b_i^{L*} denote the updated weights and biases at layer L after applying the gradient descent updates. The learning rate, γ, is a crucial hyperparameter that determines the magnitude of each update step and will be elaborated on in Section 3.4.1.

In my research, I utilize Kingma and Ba (2015) optimization method for updating the weights and biases, which introduces the adaptive moment optimization technique known as ADAM. The weight update equation is as follows:

$$w_{ji}^{*} = w_{ji} - \gamma \frac{m_{ji}^{*}}{\sqrt{v_{ji}^{*}} + \epsilon}, \tag{3.8}$$

Here, ϵ is a small constant added to prevent division by zero. The adaptive moment estimates, m_{ji}^{*} and v_{ji}^{*}, are calculated using the formulas:

$$m_{ji}^{*} = \beta_1 m_{ji} + (1 - \beta_1) \frac{dC}{dw_{ji}}, \tag{3.9}$$

$$v_{ji}^{*} = \beta_2 v_{ji} + (1 - \beta_2) \left(\frac{dC}{dw_{ji}} \right)^2, \tag{3.10}$$

m_{ji} and v_{ji} represent the estimates from the previous update step. β_1 and β_2 are hyperparameters that I have kept at their default values of 0.9 and 0.999, respectively, to minimize the number of parameters I have to tune. The learning rate, γ, is the only parameter adjusted through fine-tuning. The process for updating biases is similar, with derivatives taken with respect to the biases.

In neural network training, a pass through the entire dataset is termed an "epoch".

However, it is often preferable to update the network's weights using smaller subsets of the dataset, known as "batches" (Masters and Luschi, 2018). Each batch contains "m" training instances on which the network is trained before updating its weights (Brownlee, 2018). The training process continues until the loss, C, is minimized across all batches. The impact of batch size on network training will be analyzed in Section 4.4.

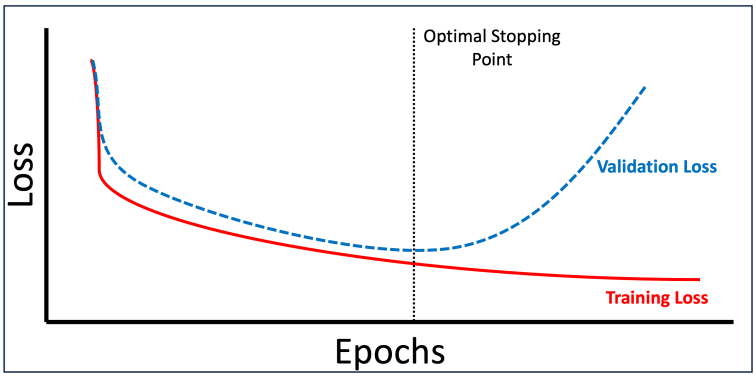


Figure 3.2: Losses on the training and validation datasets are monitored throughout training. Training halts when the validation loss stabilizes or begins to increase.

Assessing training convergence extends beyond monitoring the loss on the training dataset D_T due to the risk of overfitting. Overfitting occurs when a network performs well on the training data but poorly performs on unseen data (Lawrence et al., 1997). It often arises from misinterpreting noise as real data, small datasets that lack sufficient instances to capture feature-label relationships or flawed data assumptions and network design (Hawkins, 2004). To mitigate this, a separate validation dataset D_V is used, that contains instances not used in training, to evaluate convergence. Training continues until the losses on both D_T and D_V are minimal, with D_V loss stabilizing or increasing, signaling the optimal stopping point to avoid tuning to noise or irrelevant features, as illustrated in Figure 3.2. Finally, network accuracy is confirmed by evaluating the prediction error between the actual labels and the network's predictions on a blind test dataset D_B.

3.4 Neural Network Hyperparameters

3.4.1 Learning rate

The learning rate is a dimensionless hyperparameter that dictates the step size the network takes during optimization (Zeiler, 2012). A too-small learning rate can lead to overfitting (Smith, 2018). On the other hand, a large learning rate risks causing the network to diverge, failing to converge to a solution (Yang and Ma, 2019). I implement and evaluate different learning rates through a grid search method to select optimal learning rate in section 4.4.

3.4.2 Activation Functions

An activation function is a mathematical operation that is applied to a node's input to produce the node's output (Karlik and Olgac, 2010). The selection of an appropriate activation function is dependent upon the specific objectives of the network (Singh and Singh, 2020). For instance, activation functions like the sigmoid and softmax are commonly used in classification networks due to their ability to model probability distributions (Bridle, 1990; Karsoliya, 2012). In contrast, activation functions such as the rectified linear unit (ReLU) are often employed in regression analysis to facilitate efficient training and mitigate vanishing gradients—a phenomenon where gradients diminish as they backpropagate through deep layers, leading to slow or stalled learning (Schmidt-Hieber, 2020). In designing my neural network in Section 4.4, I will test common activation functions and evaluate their performance. A summary of the activation functions being considered is provided in Figure 3.3.

In designing my neural network in Section 4.4, I will test common activation functions and evaluate their performance. A summary of the activation functions being considered is provided in Figure 3.3.

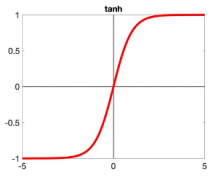
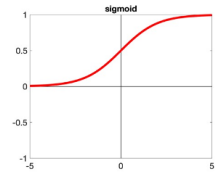
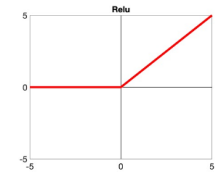
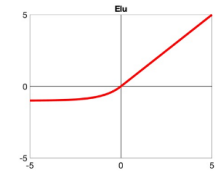
Activation Function Graph	Activation Function Equation
	$y(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
	$y(x) = \frac{1}{1 + e^{-x}}$
	$y(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$
	$y(x) = \begin{cases} (e^x - 1), & x < 0 \\ x & x \geq 0 \end{cases}$

Figure 3.3: Key activation functions to be experimented with for the design of my neural network.

3.4.3 Hidden layers and Nodes

Complex tasks like image classification and speech recognition require deep networks with multiple layers and nodes to map intricate non-linear relationships within images or speeches and accurately predict target labels (Bianchini and Scarselli, 2014). However, having too many parameters within the network, such as an excessive number of hidden layers and nodes, could cause the network to memorize the training dataset and poorly generalize to unseen datasets due to overfitting (Hunter et al., 2012; Karsoliya, 2012). Extensive discussions have taken place among researchers regarding the most effective approaches to determine the optimal number of layers and neurons per layer (Bebis and Georgiopou-

los, 1994; Benardos and Vosniakos, 2007; Uzair and Jamil, 2020). In my thesis, I tackle this challenge by employing a grid search-based analysis to explore various combinations of neurons and hidden layers networks in Section 4.4.

3.4.4 Regularization

Regularization refers to any modification applied to the network design or loss function to improve its generalization on unseen data and avoid overfitting (Goodfellow et al., 2016). Various regularization techniques have been developed. Common approaches include the utilization of a separate validation dataset (Caruana et al., 2001), L1 regularization (Tibshirani, 1996), and L2 regularization (Hoerl and Kennard, 1970). More recently, dropout regularization has gained prominence (Hinton et al., 2012).

Dropout regularization effectively trains multiple networks by stochastically turning off different nodes in each epoch and then averaging their predictions (Hinton et al., 2012; Ba and Frey, 2013). A key benefit of dropout is its ease of implementation, governed by a single hyperparameter: the dropout rate, which determines the fraction of nodes deactivated per epoch. In my research, I use a validation dataset alongside dropout regularization to mitigate overfitting. I experiment with various dropout rates to optimize this parameter in Section 4.4.

3.5 Physics-Guided Neural Networks

Physics-guided neural networks (PGNN), integrate physics constraints into their loss function during training (Karpatne et al., 2017a). A generic PGNN loss function can be expressed as follows:

$$C_T = C_d + \lambda_{phy} \cdot C_{phy}. \tag{3.11}$$

Here, C_d measures the loss between the network's predictions and the actual labels from the training dataset. C_{phy} calculates the loss based on how well the network predictions align with a specified scientific model. The total loss C_T incorporates

both the data-driven loss and the physics-informed loss. The coefficient λ_{phy} fine-tunes the impact of physics loss on the training process. When λ_{phy} is set to zero, the PGNN's loss function becomes equivalent to a standard neural network loss function.

PGNNs have been primarily utilized in geophysics for three purposes. First, they ensure network predictions adhere to physics-based constraints (Karpatne et al., 2017b). Second, they assist in creating training datasets based on rock physics models (Das et al., 2019; Das and Mukerji, 2020). Finally, PGNNs employ seismic modeling equations to alleviate the problem of data scarcity by ensuring that network predictions closely match observed seismic data (Alfarraj and AlRegib, 2019; Biswas et al., 2019; Sun et al., 2020; Jin et al., 2022).

Data mislabeling and noise can affect neural network predictions, leading to outcomes misaligned with actual physical principles (Frenay and Verleysen, 2014; Lazer et al., 2014). To address this, Karpatne et al. (2017a) proposed integrating physics constraints into neural network training. They demonstrated this in a study on lake temperature prediction, employing a physics-guided loss function to ensure that temperature predictions were physically plausible, particularly in terms of density variations with depth. The study found that neural networks without physics-based adjustments had higher error rates and produced unrealistic predictions. Conversely, employing PGNN not only enhanced the accuracy of predictions but also ensured their consistency with physical laws. Furthermore, they proved to be more data-efficient, requiring less training data to achieve accurate results compared to traditional neural network training, highlighting their suitability in scenarios with limited data availability.

Another application of PGNN involves using rock physics models to to generate training dataset for the network from limited field measurements. Das et al. (2019) used Avseth et al. (2000) constant-cement rock physics model to generate

elastic properties from reservoir parameters. Their approach involved simulating pseudo-wells with varying facies, porosity, and clay volume, and converting these into seismic data. The objective was to train the network on predicting impedance log from input seismic trace. While the network showed high predictive accuracy under optimal conditions, discrepancies in source wavelet characteristics between training and testing datasets led to significant errors. Additionally, uncalibrated key model parameters like the grain contact parameter further increased prediction inaccuracies. This underscores two crucial aspects. First, the need for accurate seismic modeling workflow to closely mirror actual data. Second, the importance of precise calibration of model parameters for effective training and reliable network predictions.

In a modified PGNN approach, the network trains on observed seismic traces to overcome limited data challenges. It predicts elastic properties from these traces, which are then used to generate synthetic seismic data through a seismic modeling workflow. This training process employs a loss function that calculates the misfit between the observed and modeled seismic data, thus eliminating the need for dedicated label vectors. Biswas et al. (2019) applied this method to predict P and S wave velocities along with density directly from seismic data, using a known low-frequency model to generate synthetic traces and calculate training loss. This approach achieved accuracy comparable to traditional neural networks trained with complete feature-label dataset. Its primary advantage is using seismic traces for both input and outcome validation, reducing the reliance on costly labeled datasets. However, its effectiveness depends on the accurate execution of the seismic modeling workflow and the presence of a reliable low-frequency model.

Addressing the challenge of limited field data measurements is a key obstacle in applying neural networks effectively, particularly in geophysical problems. Das et al. (2019) highlighted the potential of rock physics models to mitigate this

limitation, provided known model parameters are derived from field data. However, as explored in Chapter 2, acquiring complex reservoir parameters, such as time constant τ and Thomsen parameters, proves challenging during the early stages of exploration. A possible solution is the utilization of a physics-guided neural network loss function based on forward modeling. This approach, however, faces its own challenges when modeling anisotropic VTI reservoirs, which require knowledge of Thomsen parameters to accurately model the elastic response of reservoir properties.

In Chapter 5, I propose a novel physics guided neural network approach that combines some of these concepts. This workflow is particularly tailored for scenarios where key model parameters remain unknown and field measurements are limited.

4 Porosity and Saturation Estimation From Dispersive Seismic Data Using Neural Network and Squirt-Flow Rock Physics Model

4.1 Chapter Objectives

Chapter 2 discussed the successful use of dispersive rock physics models for gas saturation inversion, as demonstrated by Wu et al. (2014) and Jin et al. (2017), where Bayesian inversion algorithms were used to estimate porosity and saturation from seismic gathers generated by a squirt flow model with a known key model parameter, τ (see Section 2.4). This chapter introduces an alternative approach, utilizing neural networks to predict porosity and saturation from seismic gathers modeled by the squirt flow rock physics model, assuming τ is known.

The chapter begins by outlining the geological model and the seismic modeling workflow used to generate seismic gathers. It then details the design of the neural network employed throughout the research. The chapter concludes by addressing a common challenge in real-world neural network applications: the unknown value of a key model parameter, τ, which affects the application of squirt flow models for characterizing complex reservoirs. This issue is further complicated by limited field measurements, which are often insufficient to properly train a neural network.

4.2 Seismic Modeling

I selected a geological model with a shale (non-reservoir) layer overlying a sandstone (reservoir) layer, reflecting a common exploration scenario where non-reservoir units enclose reservoir formations. In this chapter and Chapter 5, I assume isotropic behavior for the shale layer to simplify the analysis, while allowing the sandstone layer to exhibit dispersive behavior. In Chapter 6, I introduce anisotropy to the shale layer to examine the effects of anisotropic behavior on

AVO analysis, while keeping the sandstone layer isotropic and elastic. This model captures varying complexities of real-field conditions, providing a robust test of my thesis's physics-guided neural network methodology for both dispersive and anisotropic cases.

The shale layer is tight, elastic, and has water saturation above 90%, but it can accommodate up to 10% gas saturation. The underlying reservoir layer consists of porous sandstone, capable of exhibiting dispersion and attenuation. The fluid within the sandstone can be gas, brine, or a mixture of both. Each layer has a temporal thickness of 200 ms. The shale layer is not assumed to be fully brine saturated to introduce complexity and test the network's ability to predict reservoir properties as conditions change across the interface.

When modeling gathers, I will adopt the convolutional modeling workflow outlined in Jin et al. (2017). The workflow takes the dispersive moduli $K(\omega)$ and $\mu(\omega)$ outputted from Papageorgiou and Chapman (2017) model, discussed in Section 2.4 to calculate a frequency-dependent reflectivity $R(\omega, \theta)$ using the formulations outlined by Schoenberg and Protazio (1990).

The reflection coefficient R can vary with frequency (ω) due to dispersion. Therefore, a special manipulation is required to convolve a dispersive reflectivity that varies at wide range of frequencies with a wavelet. Jin et al. (2017) provided the following equation to resolve this issue:

$$x_t(\theta) = \mathcal{F}^{-1}(W(\omega)\ .\ R(\theta, \omega)), \tag{4.1}$$

here x_t is the seismic trace, θ is the incident angle, R is the dispersive reflectivity , $\mathcal{F}^{-1}$ is the inverse Fourier transform, $W(\omega)$ is the inverse Fourier transform of the source wavelet at frequency ω. The source wavelet used is a Ricker wavelet,

which, according to Wang (2015), can be expressed as:

$$w(t) = (1 - \frac{1}{2}\omega_p^2 t^2)e^{-\frac{1}{4}\omega_p^2 t^2}, \tag{4.2}$$

where t is the time sample and ω_p is the dominant frequency. The Fourier transform of $w(t)$, using Wang (2015) formulation, is written as:

$$W(\omega) = \frac{2\omega^2}{\sqrt{\pi}\omega_p^3} e^{\frac{-\omega^2}{\omega_p^2}}. \tag{4.3}$$

Finally, the inverse Fourier transform is calculated by using the following equation:

$$x_t(\theta) = \int_{-\infty}^{+\infty} [W(\omega) \cdot R(\theta, \omega)]\, e^{-i\omega t}\, d\omega. \tag{4.4}$$

The noise level within the gather is controlled by the signal to noise ratio (SNR), in dB units, by applying the following equation:

$$SNR = 10log\left(\frac{P_{x_t}}{P_{n_t}}\right), \tag{4.5}$$

where P_{x_t} and P_{n_t} represent the power spectra for vector $x_t(\theta)$ and noise vector n_t respectively. The power spectrum for x_t is calculated as follows:

$$P_{x_t} = \sum_{t=0}^{N-1} |x_t(\theta)|^2, \tag{4.6}$$

where N is the number of samples within the trace. The power spectrum for P_{n_t} is calculated as follows:

$$P_{n_t} = \sum_{t=0}^{N-1} |n_t|^2\ . \tag{4.7}$$

Now, I rewrite equation 4.5 with equations 4.6 and 4.7 and rearrange to solve for n_t:

$$\sum_{t=0}^{N-1} |n_t|^2 \cdot dt = \frac{\sum_{t=0}^{N-1} |x_t(\theta)|^2}{10^{\frac{\text{SNR}}{10}}}, \tag{4.8}$$

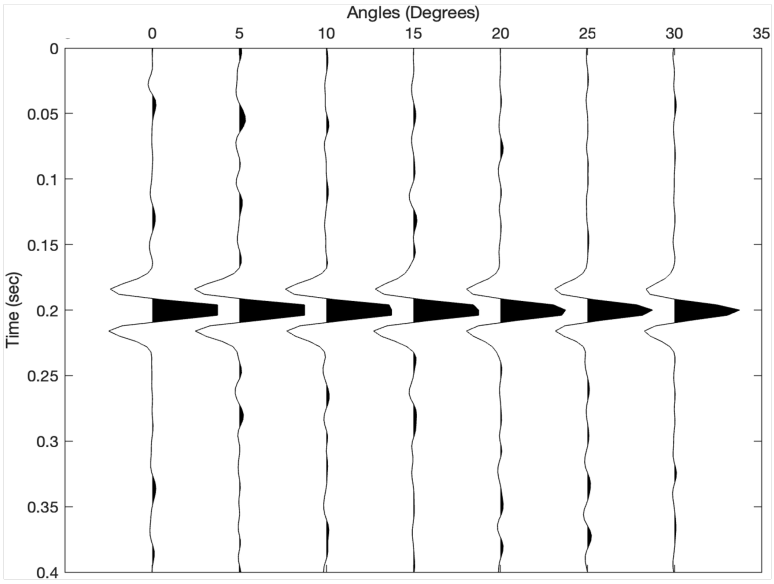


Figure 4.1: Synthetic gather generated using the modeling workflow from Jin et al. (2017) with a 25 Hz Ricker wavelet. White Gaussian noise is added with a 25 dB SNR.

where the noise vector n_t is sampled from a Gaussian distribution with mean of 0 such that the power spectrum of the noise vector in equation 4.8 is satisfied. Next, I convolve n_t with the source wavelet $w(t)$, see equation 4.2, and add to the trace $x_t(\theta)$ as follows:

$$x_t^*(\theta) = x_t(\theta) + w(t) * n_t, \tag{4.9}$$

where $x_t^*(\theta)$ is the seismic trace at incident angle θ with added noise. Figure 4.1 illustrates a gather with a 25 dB SNR.

4.3 Features and Labels Preparation

Neural networks rely on datasets comprising labels and features to train their weights. The labels of the training dataset are sandstone porosity and saturation, which were used in Papageorgiou and Chapman's (2017) model to calculate the elastic properties of the sandstone layer.

I extract frequency-dependent amplitudes from the seismic gather as features.

Calculating these frequency-dependent amplitudes requires spectral analysis, which can be affected by spectral leakage, a phenomenon where energy from one frequency spreads into neighboring frequencies when analyzing a finite signal segment (Wu et al., 2012). To reduce spectral leakage, I extract a 100 ms window above and below the interface for each seismic trace $x_t^*(\theta)$ and apply a taper $T(t)$ to the window, as follows:

$$x_t^W(\theta) = |x_t^*(\theta)|_{t=100}^{t=300} \cdot T(t). \tag{4.10}$$

Here, $|x_t^*(\theta)|$ represents the selected noise trace vector, and $x_t^W(\theta)$ denotes the tapered trace. The Tukey window tapering was applied to extract frequency-dependent amplitudes for its tuning flexibility, spectral leakage reduction, and frequency content preservation (Stoffer and Bloomfield, 2000). However, it can be less effective at reducing sidelobes of the windowed signal, which may impact spectral analysis (Oppenheim and Schafer, 1998). The Tukey cosine taper, $T(t)$, is defined as follows:

$$\begin{cases} T(t) = \frac{1}{2}[1 + \cos(\frac{2\pi}{r}\frac{t-1}{N-1} - \pi)] & t < \frac{r}{2}(N-1) + 1 \\ T(t) = 1 & \frac{r}{2}(N-1) + 1 \leq t \leq N - \frac{r}{2}(N-1) \\ T(t) = \frac{1}{2}[1 + \cos(\frac{2\pi}{r} - \frac{2\pi(t-1)}{r(N-1)} - \pi)] & N - \frac{r}{2}(N-1) < t, \end{cases}$$

where r is the ratio of tapered to constant signal and N is the total number of samples. I set $r = 0.3$ and I calculate the fast Fourier transform (FFT) of the tapered trace as follows:

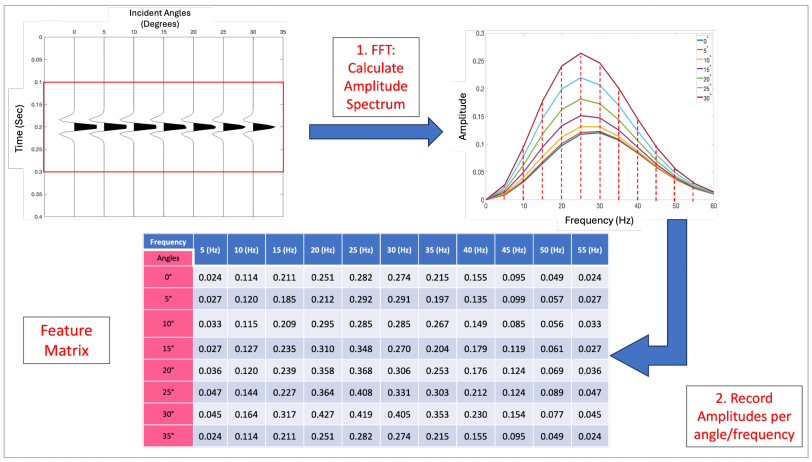
$$X(\omega) = \sum_{t=0}^{N-1} x_t^W(\theta) e^{i\omega t}, \tag{4.11}$$

here, $X(\omega)$ is the FFT output at frequency ω, N is the number of samples per trace. Next, I calculate the amplitude spectrum of the tapered trace $|X(\omega)|$ as follows:

$$|X(\omega)| = \sqrt{Re(X(\omega))^2 + Im(X(\omega))^2}. \tag{4.12}$$

I select m values of frequency ω, and calculate the amplitude spectra at those

values for all incident angles z. Finally, I arrange the calculated amplitudes in an $(m \times z)$ feature matrix F, where m is the number of frequency bins selected and z is the number of incident angles. Figure 4.2 summarizes the steps for generating feature matrix F.



Frequency / Angles	5 (Hz)	10 (Hz)	15 (Hz)	20 (Hz)	25 (Hz)	30 (Hz)	35 (Hz)	40 (Hz)	45 (Hz)	50 (Hz)	55 (Hz)
0°	0.024	0.114	0.211	0.251	0.282	0.274	0.215	0.155	0.095	0.049	0.024
5°	0.027	0.120	0.185	0.212	0.292	0.291	0.197	0.135	0.099	0.057	0.027
10°	0.033	0.115	0.209	0.295	0.285	0.285	0.267	0.149	0.085	0.056	0.033
15°	0.027	0.127	0.235	0.310	0.348	0.270	0.204	0.179	0.119	0.061	0.027
20°	0.036	0.120	0.239	0.358	0.368	0.306	0.253	0.176	0.124	0.069	0.036
25°	0.047	0.144	0.227	0.364	0.408	0.331	0.303	0.212	0.124	0.089	0.047
30°	0.045	0.164	0.317	0.427	0.419	0.405	0.353	0.230	0.154	0.077	0.045
35°	0.024	0.114	0.211	0.251	0.282	0.274	0.215	0.155	0.095	0.049	0.024

Figure 4.2: Workflow to generate feature matrix F.

4.4 Neural Network architecture

Designing a neural network involves key decisions like choosing the number of layers and nodes, selecting activation functions, and setting the learning rate. To make this process more efficient, I have adopted two main guidelines. First, I prioritize improving prediction accuracy without notably extending training time. Second, I assess each parameter's impact individually, and focus my final grid search on the most influential ones. I use the root mean square error (RMSE) to measure the accuracy of predictions, which is calculated as follows:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(\hat{y}_i - y_i)^2}{N}}, \tag{4.13}$$

where $\hat{y}$ is the network prediction for porosity and saturation, y is the true label value for porosity and saturation and N is the number of instances in the test

dataset.

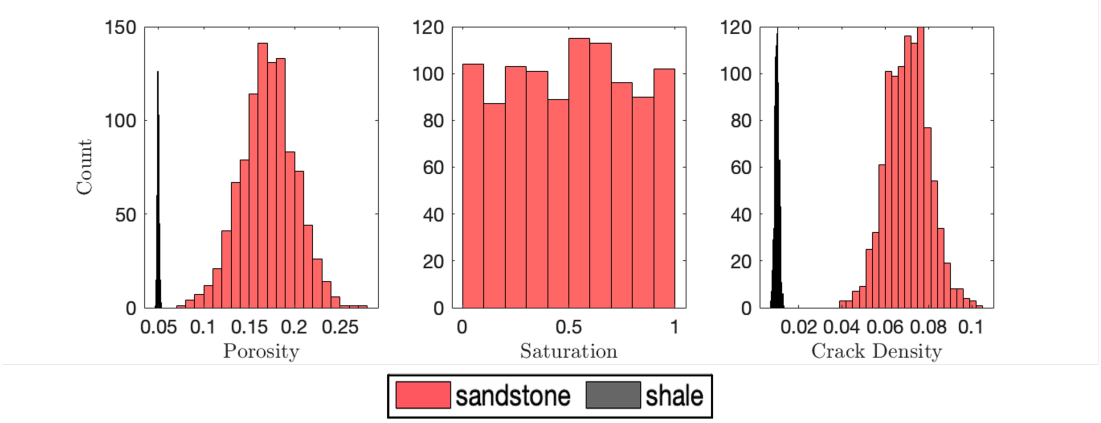


Figure 4.3: Papageorgiou and Chapman's (2017) Input parameter distribution for generating training and test instances. Shale saturation is assumed to be 90% or more brine-saturated.

I assembled a training dataset of 10,000 instances, each with sandstone porosity and saturation as labels and features extracted from the corresponding gather. Shale is assumed to be tight, >90 % brine-saturated, elastic, and with properties selected from the distributions in Figure 4.3. For sandstone, saturation is uniformly distributed, allowing for brine, gas, or a mixture of both fluids. Porosity is sampled from a Gaussian distribution with a mean of 0.17 to represent typical sandstone conditions at an effective pressure of 30 MPa (Mavko et al., 2020). Crack density is similarly sampled from a Gaussian distribution with a mean of 0.07, based on the empirical relationship between effective pressure (30 MPa) and crack density provided by Chapman et al. (2002). Fluid distribution in both layers is assumed to be uniform ($q = 1$) when calculating fluid properties according to Equation 2.4. The Papageorgiou and Chapman's (2017) model requires a reference rock to adjust and calibrate the model-calculated values. To fulfill this requirement, I utilize a shale-sandstone sequence with elastic properties measured from field data provided in the appendices of Mavko et al. (2020), ensuring realistic properties for both shale and sandstone layers. The specific parameters used are detailed in Table 2.

The training was conducted on 8,000 instances and the remaining 2,000 reserved for validation. These validation instances serve the purpose of monitoring convergence and avoid overfitting (see Section 3.3). The training will stop if the validation loss remains constant for 10 epochs or starts to increase. The network's performance evaluation is carried out by making predictions on a test dataset consisting of 100 instances that exhibit attenuation within seismic frequencies similar to the random instances shown in Figure 4.4.

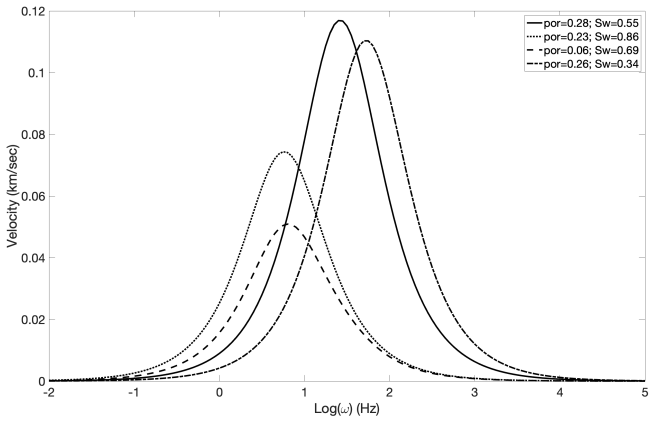


Figure 4.4: Attenuation curves at sandstone reservoir derived from random samples of porosity and saturation with $\tau = 1.0 \times 10^{-1}$ sec. High attenuation values are observed within the seismic frequency range.

Within these datasets, the target labels pertain to sandstone porosity and saturation. The features are represented by an amplitude matrix denoted as $\mathbf{X}$, which is a 13×7 matrix comprising real numbers. In this matrix, the rows correspond to frequency bins at 5 Hz intervals from 0 to 60 Hz. The columns, on the other hand, represent the number of traces within the gather. The traces were modeled to have incident angles spanning from 0° to 30°, incremented every 5 degrees.

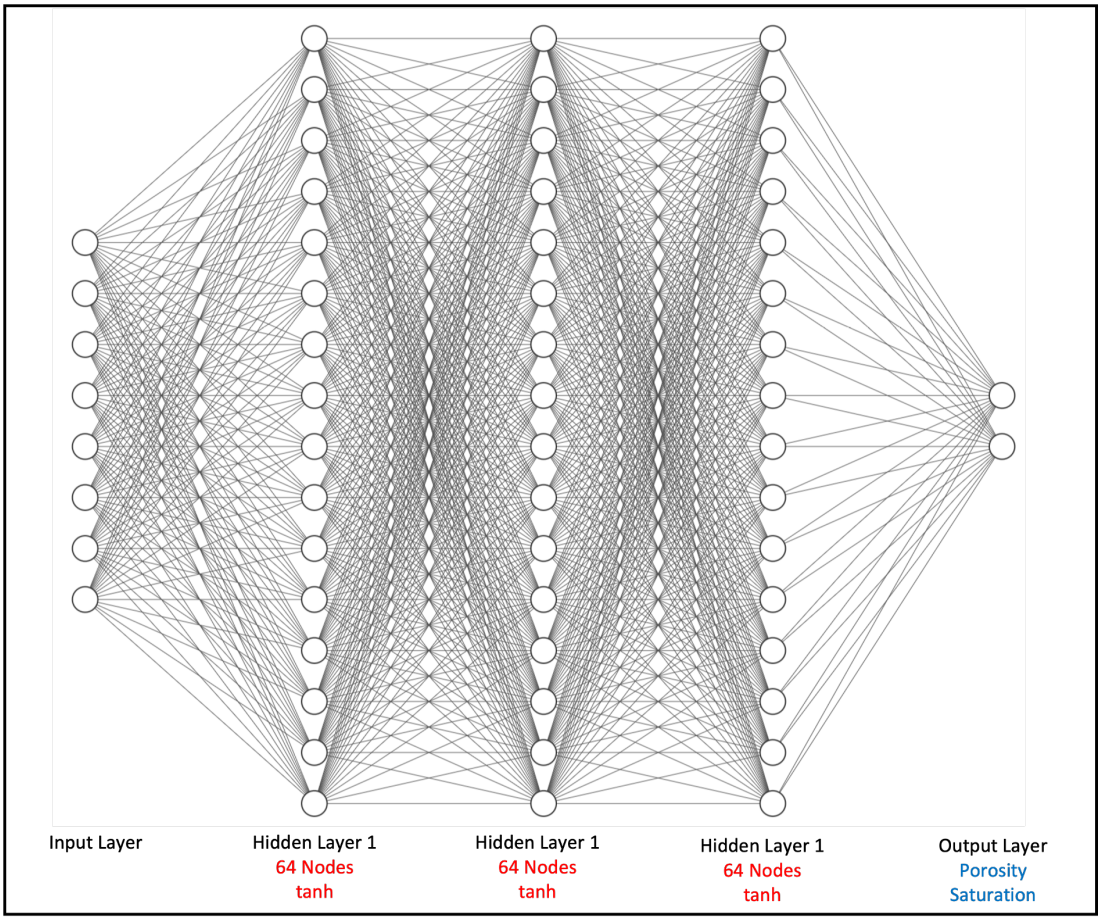


Figure 4.5: The basic neural network architecture employed as a framework to evaluate the impact of network parameters on prediction accuracy and training time.

Research indicates that a neural network with three hidden layers can handle complex regression tasks with satisfactory accuracy (Uzair and Jamil, 2020). However, as Karsoliya (2012) notes, there is no one-size-fits-all approach to network design, making it common to use a trial-and-error method for optimization. Based on these insights, I established a baseline network with an input layer, three hidden layers, and an output layer, as illustrated in Figure 4.5. I then incrementally investigate the impact of individual parameters while keeping other neural network parameters constant to achieve optimal prediction accuracy.

The input layer has nodes corresponding to the number of input features. Each hidden layer contains 64 nodes, a moderate choice given the lack of consensus

on optimal node numbers per layer in the literature (Bebis and Georgiopoulos, 1994). The output layer includes two nodes, representing porosity and saturation estimates. I selected the *tanh* activation function for the hidden layers based on its effectiveness and accuracy in regression analysis using neural networks (Karlik and Olgac, 2010), while the output layer uses a linear activation function to validate prediction accuracy and prevent any distortion of data that could occur with a non-linear activation function.

The learning rate is initially set to 1.0×10^{-4}. Loss is calculated as per Equation 3.5. Weights and biases are updated in batches of 32 instances, a commonly balanced batch size in neural network training (Masters and Luschi, 2018). Weights are initialized following the Glorot and Bengio, 2010 method, drawn from a zero-mean distribution with variance dependent on the input and output units for each layer, enhancing weight updates. Biases are initialized at zero. A detailed description of all network parameters can be found in Section 3.4.

I begin my network design by evaluating the impact of the number of dense hidden layers and their nodes. I adopt an incremental approach, adding one layer at a time and assessing the network's prediction accuracy and training time at each stage. I keep other neural network parameters constant, gradually increasing the hidden layers from one to a maximum of seven.

Figure 4.6 shows an increase in RMSE for saturation as the number of layers increases, leveling off at four layers. This increase may be due to the network's tendency to overfit with additional parameters, making it easier to memorize training data, particularly for complex relationships (Goodfellow et al., 2016). In contrast, porosity is less sensitive to layer count, showing consistent RMSE across different numbers of layers. This likely indicates a simpler relationship between seismic features and porosity, as increasing the layer count does not lead to overfitting or reduced generalization error, and simpler networks yield

similar prediction RMSE. Training time, however, decreases exponentially with additional layers, inversely related to prediction RMSE for saturation. To balance training time efficiency and saturation prediction accuracy, a grid search will be conducted on networks with two to five hidden layers to determine the optimal layer count for porosity and saturation predictions.

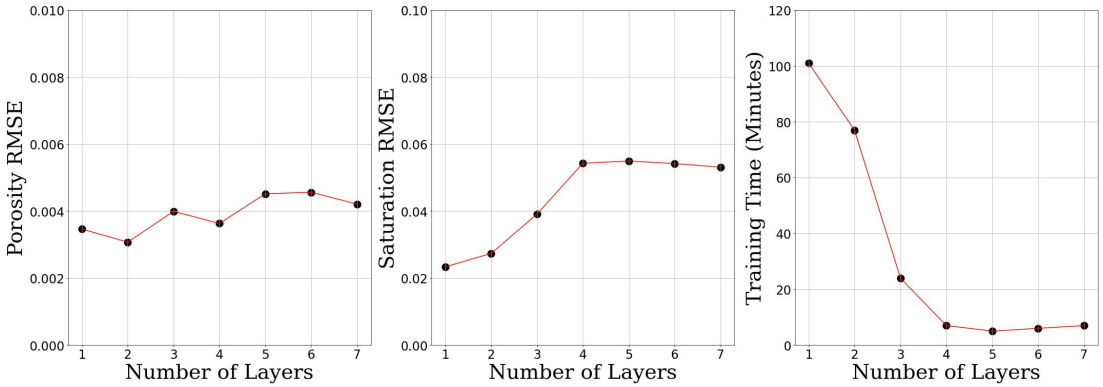


Figure 4.6: Impact of the number of hidden dense layers on neural network prediction RMSE and training time.

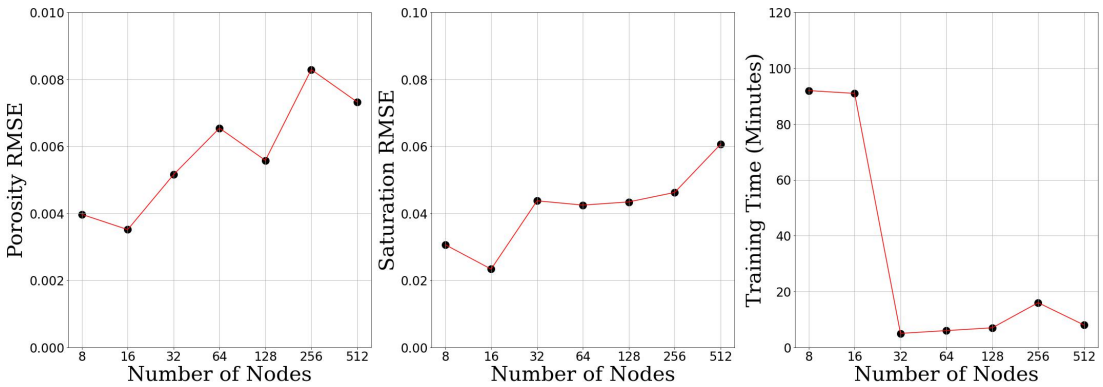


Figure 4.7: Impact of the number of nodes per dense layers on neural network prediction RMSE and training time.

With the number of hidden layers fixed at three, I systematically explore the impact of number of nodes per layer on porosity and saturation predictions. Figure 4.7 shows that increasing node count generally raises RMSE for porosity and saturation, likely due to overfitting. Conversely, using fewer than 32 nodes significantly prolongs training time, suggesting the network struggles to map complex relationships with limited nodes. A grid search between 16 and 128 nodes will

be conducted to achieve a balance between training efficiency and prediction accuracy. Research suggests that pruning nodes with depth can improve training efficiency and enhance accuracy by reducing redundant nodes (Frankle and Carbin, 2019). In the final grid search, I test this approach by testing the impact of nodes reduction with depth on network prediction accuracy and training time.

The choice of learning rate (γ) is crucial in controlling updates to a neural network's weights and biases, as discussed in Section 3.4.1. To assess its impact, I hold all other network parameters constant and test various learning rate values for their effects on prediction accuracy and training time. Figure 4.8 shows that extremely low learning rates significantly prolong training and result in good porosity and saturation predictions. In contrast, very high learning rates speed up training but often degrade prediction quality for both porosity and saturation. Therefore, I have confined the learning rate within the range $\gamma \in \text{Log}[-3, -2]$ to achieve accurate predictions with efficient training time.

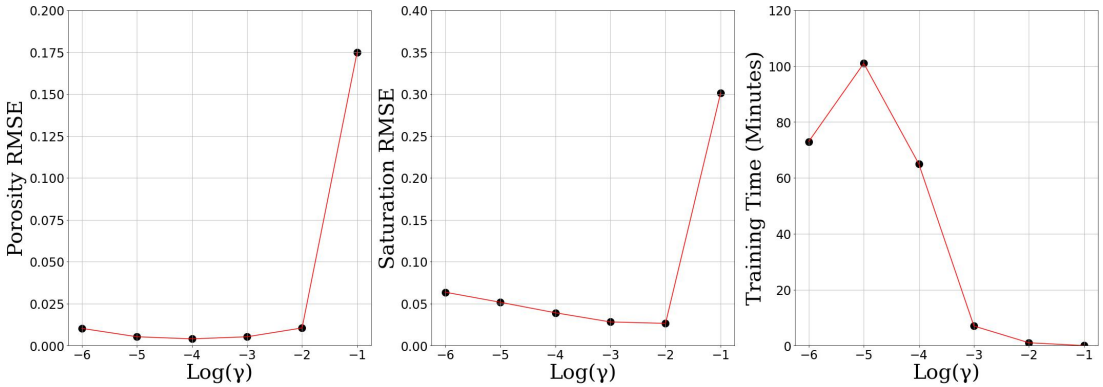


Figure 4.8: Impact of learning rate on neural network prediction RMSE and training time.

I now examine the impact of activation functions on neural network training and prediction. Four commonly used functions in regression analysis are tested: *tanh*, *sigmoid*, *elu*, and *relu* (see Section 3.4.2 for descriptions). As shown in Figure 4.9, the *tanh* function results in higher RMSE for porosity and saturation predictions, while *sigmoid* significantly increases training time due to its complex

gradient calculations and slow updates at extreme values—a phenomenon known as the vanishing gradient problem (Aggarwal, 2018). The *relu* function provides the lowest RMSE and shortest training time, balancing accuracy and efficiency. Therefore, I select *relu* as the activation function for all hidden layers. The output layer initially uses a linear activation function during testing but will later switch to a sigmoid function to constrain porosity and saturation predictions between 0 and 1.

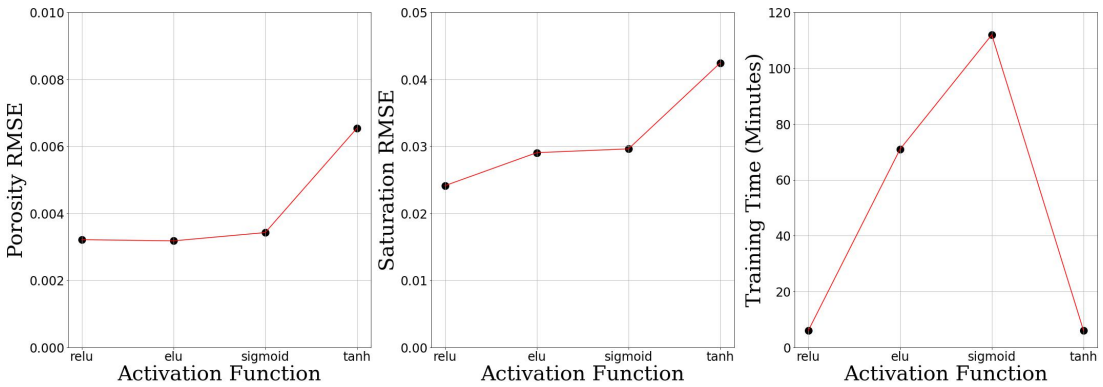


Figure 4.9: Impact of activation functions on neural network prediction RMSE and training time.

Network updates can be performed using batch processing, with batch size as a controllable hyperparameter, as outlined in Section 3.3. Keeping all other parameters constant, I tested batch sizes from 1 to 1024. As shown in Figure 4.10, a batch size of 1 significantly prolongs training due to frequent updates, while larger batch sizes improve training efficiency by reducing update frequency. In terms of prediction accuracy, porosity and saturation errors tend to be lower with smaller batch sizes, which also increases training time. Consequently, I limited the final grid search to batch sizes between 8 and 256 to balance training efficiency with prediction RMSE.

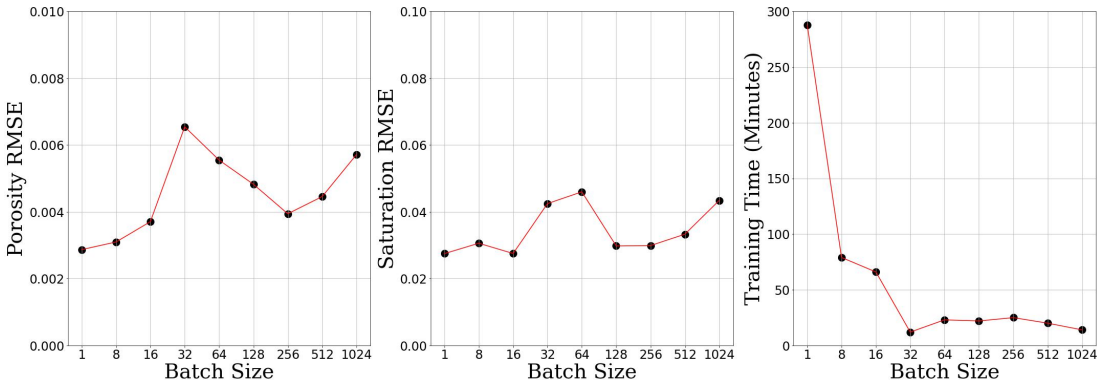


Figure 4.10: Impact of batch size on neural network prediction RMSE and training time.

In my thesis, training, validation, and testing datasets are generated using the same rock physics model. Although random noise is added to each gather to introduce variability, relying solely on data from this model risks overfitting, as the network may memorize specific feature-label pairs rather than learning underlying relationships, potentially reducing its generalization on test datasets. To address this, I incorporate dropout layers as a regularization technique (see Section 3.4.4), placing them after the first and second hidden layers with a set deactivation percentage during training. Figure 4.11 shows that porosity prediction error increases gradually as dropout rates rise from 0 to 0.4, with a sharp increase beyond this. Saturation predictions are less affected by changes in dropout rate, though no clear explanation for this reduced impact could be identified. The increase in RMSE may result from excessive neuron deactivation during training, which limits the network's ability to learn complex patterns. Therefore, I limit my dropout range to 0.1–0.4 in the final grid search.

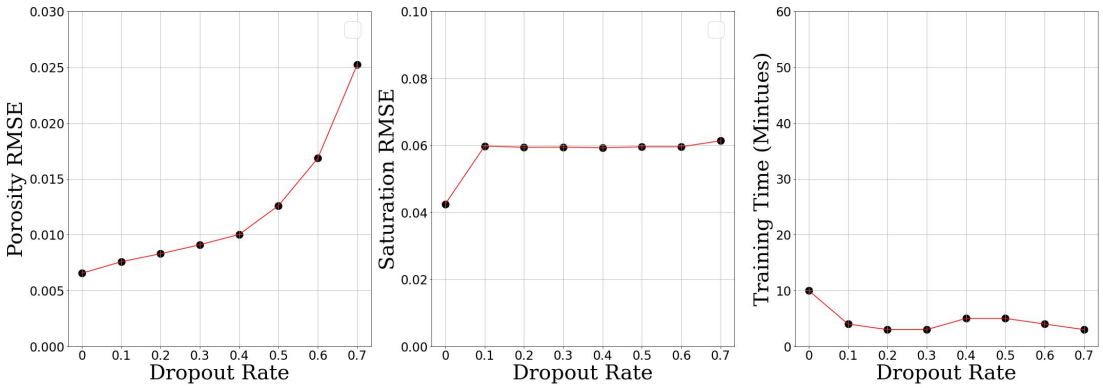


Figure 4.11: Impact of dropout rate on network prediction RMSE and training time.

Figure 4.12 provides a comprehensive overview of the previous parameter tuning analysis regarding prediction RMSE and network training time. Several observations emerge from this analysis. Porosity prediction consistently exhibits low error rates in most variations of network parameters, suggesting a robust correlation between porosity and the extracted features that are easier to learn, independent of the specific network design. Saturation prediction appears more sensitive to certain network parameters, including learning rate, the number of dense layers, nodes, dropout and batch size. The finding suggests that achieving accurate saturation predictions necessitates deeper layers with a higher number of nodes, effectively capturing the complex relationship between saturation and the extracted features. In the context of training time, most parameters appear to maintain similar level of training time efficiency, with the exception of using extreme low learning rate and a batch size of 1 due to reasons mentioned earlier in the Section.

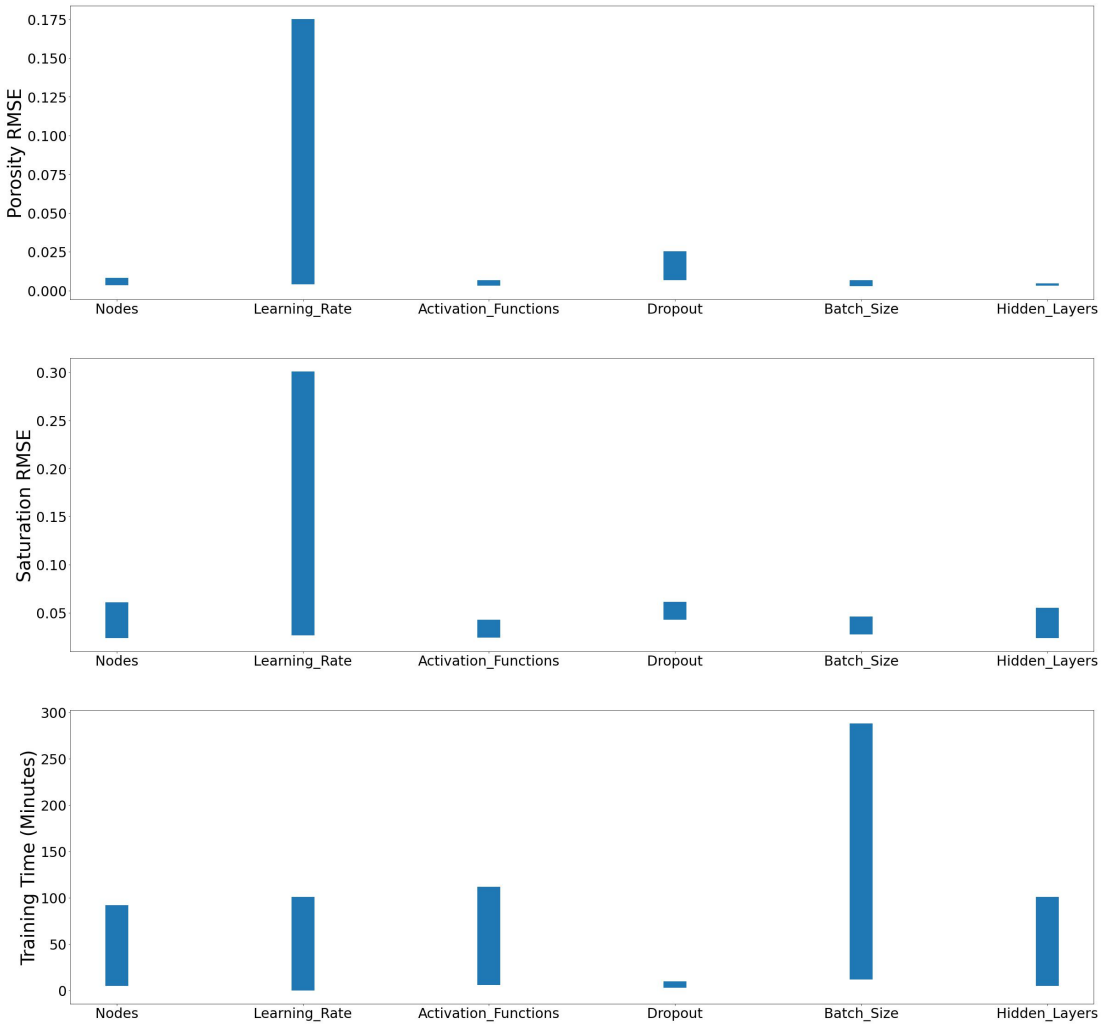


Figure 4.12: Impact of various neural network hyperparameters on prediction RMSE for porosity, saturation and training time.

I now finalize the grid search strategy to determine the optimal neural network architecture. Given that porosity prediction is more robust and less sensitive to network design, the grid search prioritizes optimizing saturation prediction while maintaining efficient training time. As shown in Figure 4.12, the learning rate has the most significant impact on prediction RMSE, directly influencing network convergence (Géron, 2017). Based on prior analysis, I apply a learning rate grid range of $log(\gamma) = [-3, -2]$, which demonstrated the best balance of RMSE and efficiency (Figure 4.8).

The next key parameters are the number of hidden layers and nodes. Based on Figures 4.6 and 4.7, I test configurations with two to five dense hidden layers, each containing between 2^4 and 2^7 nodes. I explore both constant node counts per layer and configurations where nodes decrease with each subsequent layer. For batch size, a grid search between 2^3 and 2^8 is conducted to balance accuracy and training efficiency. The activation function for all hidden layers is set to *relu*. With over 480 configurations tested, I present only the final optimal network, which achieved the best prediction RMSE and training time.

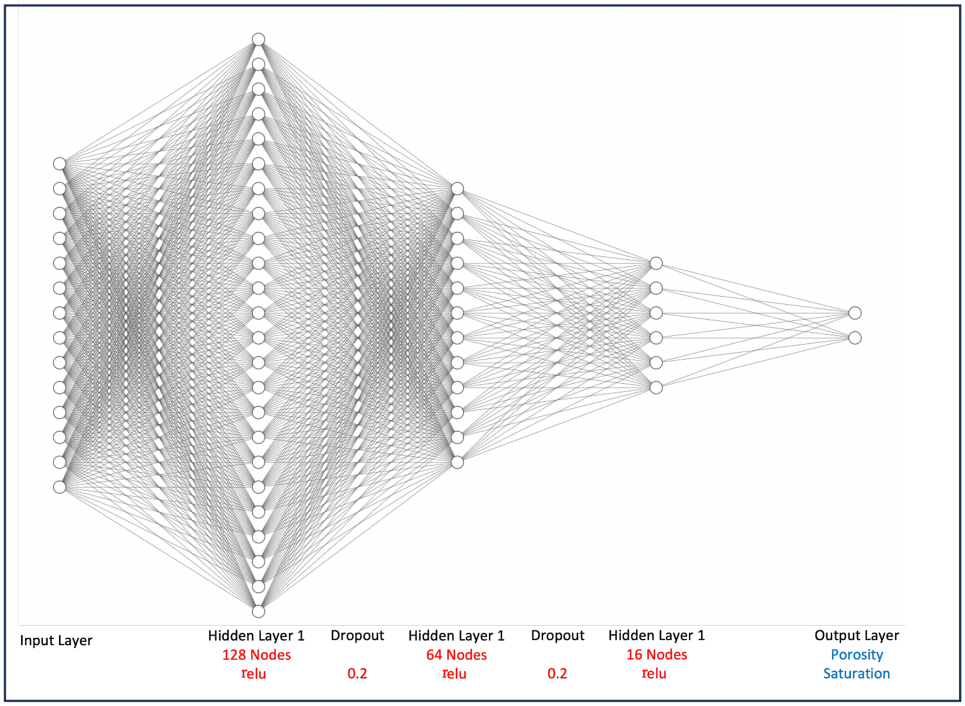


Figure 4.13: Optimal network design for porosity and saturation prediction.

Figure 4.13 illustrates the final neural network design for predicting porosity and saturation. The network starts with an input layer that processes a feature matrix of dimensions 13×7, which is then flattened for dense layer processing. Three dense hidden layers with 128, 64, and 16 nodes further process these features. The output layer has two nodes for porosity and saturation predictions and uses a sigmoid activation function to constrain outputs between 0 and 1, matching the expected range of these properties, see Section 3.4.2, Figure 3.3 for sigmoid func-

tion limits. All hidden layers employ the *relu* activation function, as previously determined. The learning rate is set to 1.0×10^{-3}, and the batch size is 32.

Now, I assess the performance of my neural network through a numerical example.

4.5 Numerical example

Using the geological model discussed in section 4.2, I proceed with predicting porosity and saturation within the sandstone layer by utilizing the neural network detailed in Section 4.4. This example assumes that the dataset comprises a sufficient number of instances to ensure effective network training and explores three distinct scenarios. In the first scenario, the generated seismic gathers for the training and test datasets exhibit no dispersion or attenuation within the seismic frequency range. In the second scenario, I adjust τ to introduce attenuation effects with a fixed τ value applied consistently across both the training and test datasets. In the third scenario, the network's capacity to predict porosity and saturation is tested with a fixed but unknown τ value in the test data, while the network is trained on a dataset with a wide range of τ, porosity, and saturation values. In all cases, the fluid is assumed to be uniformly distributed within both layers ($q = 1$) when calculating fluid properties using Equation 2.4.

In each case, the network was trained on a dataset that consist of $10,000$ training instances. Each instance consist of sandstone porosity and saturation as labels, and the features are extracted from the corresponding gather using the workflow described in Section 4.3.

Training is conducted on 8000 instances, while the remaining 2000 instances are used for validation. Training will stop, if no prediction enhancement is observed on validation data for 100 epochs. The network accuracy is evaluated by calculating the RMSE, equation 4.13, on a separate 100 test instances that were not used in training or validation.

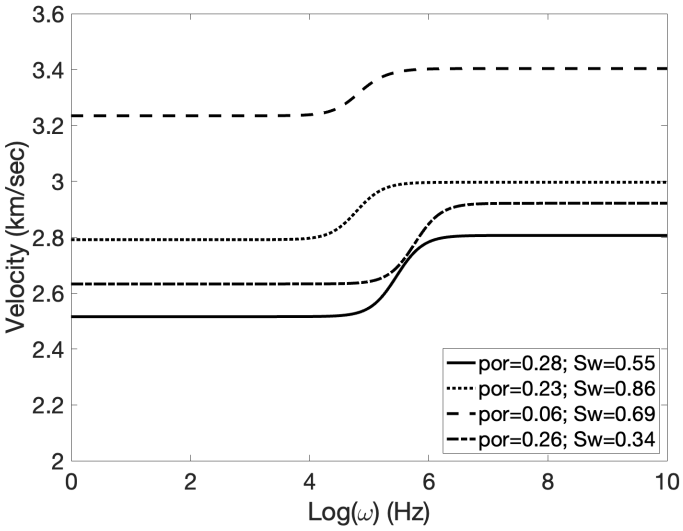


Figure 4.14: The modeled velocity shows consistent elastic properties across seismic frequencies ranging from 10 to 100 Hz when using τ value of 1.0×10^{-5}.

4.5.1 Case 1: Predicting Porosity and Saturation from Non-Dispersive Gathers

The attenuation/dispersion frequency range in Papageorgiou and Chapman's (2017) model is determined by the parameter τ. In this case, I have set τ to a value of 1.0×10^{-5}, effectively shifting the dispersion behavior to ultrasonic frequencies. Figure 4.14 provides a visualization of the modeled velocities, which were computed using equation 2.6, for various combinations of porosity and saturation under the condition $\tau = 1 \times 10^{-5}$. It is worth noting that these velocities exhibit elastic behavior within the seismic frequency range of 10 to 100 Hz.

Now, I generate the elastic training and test datasets based on the histogram distributions presented in Figure 4.3.

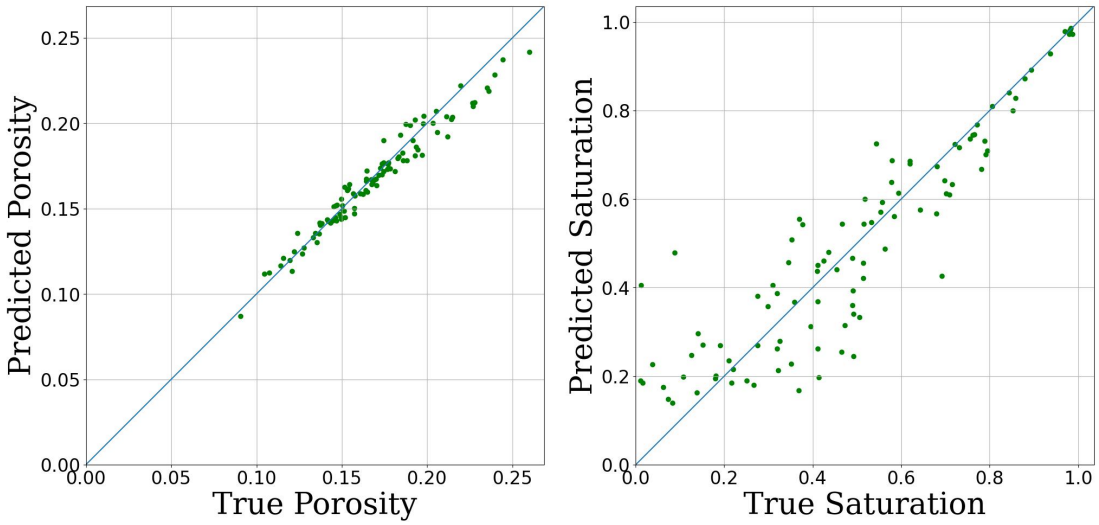


Figure 4.15: True vs. predicted porosity and saturation on an unseen elastic test dataset, with the network trained on an elastic dataset of 10,000 instances.

Figure 4.15 presents the network's prediction results on a test dataset exhibiting similar elastic behavior to the training dataset. Porosity predictions demonstrate low RMSE of 7.43×10^{-3}. However, saturation predictions perform well only for water saturations above $S_w = 80\%$ but deteriorate rapidly at lower levels, yielding an RMSE prediction error of 11.20×10^{-2}. The network accuracy is biased by the velocity-saturation relationship, detailed in Section 2.3 and will be discussed in Section 4.6, for saturation below 80% in addition to the Gaussian noise within the gathers.

To improve the training and accuracy of neural networks, one effective method is to expand the training dataset (Simard et al., 2003). I have applied the strategy by generating six training datasets, each with an increasing number of instances. These datasets will be employed for training individual networks, where their prediction accuracy will be evaluated on the same test dataset.

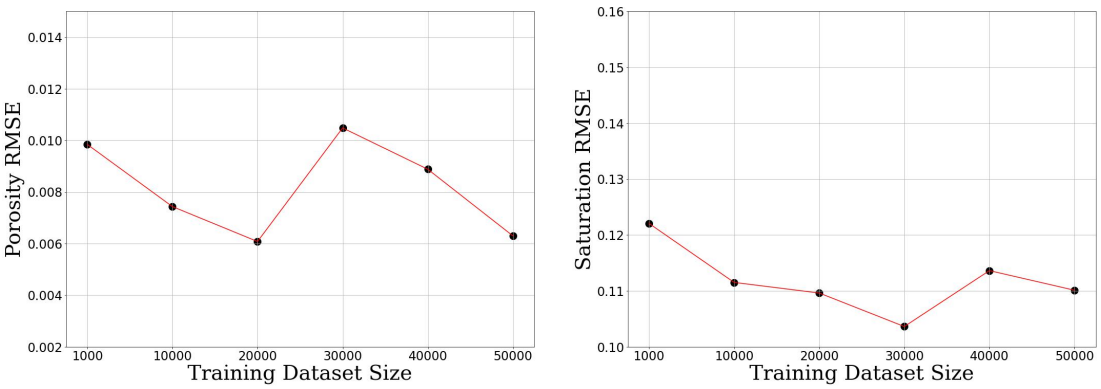


Figure 4.16: Network prediction RMSE of porosity and saturation with increasing training dataset size.

Figure 4.16 shows the network prediction across different training dataset sizes for the network. On average, the porosity prediction RMSE oscillate around 8.00×10^{-3} which is 7% increase in RMSE compared to the network trained on 10,000 elastic instances. In terms of saturation, the average RMSE across different network dataset sizes is 11×10^{-2}, which shows a slight decrease of 2 % in RMSE compared to the network trained on 10,000 instances.

Expanding the training dataset to improve prediction accuracy may seem beneficial, but two major concerns have prompted caution. First, the lack of a theoretical model to effectively use amplitude information for gas saturation limits the potential benefits of dataset augmentation. Second, the exponential increase in data augmentation time, as shown in Figure 4.17, suggests that the time and computational costs might outweigh the potential gains. Consequently, I have decided to cap the dataset at 50,000 instances, where the error in saturation prediction seems to have stabilized. The next step involves introducing attenuation into the gather modeling process and evaluating the network's prediction accuracy.

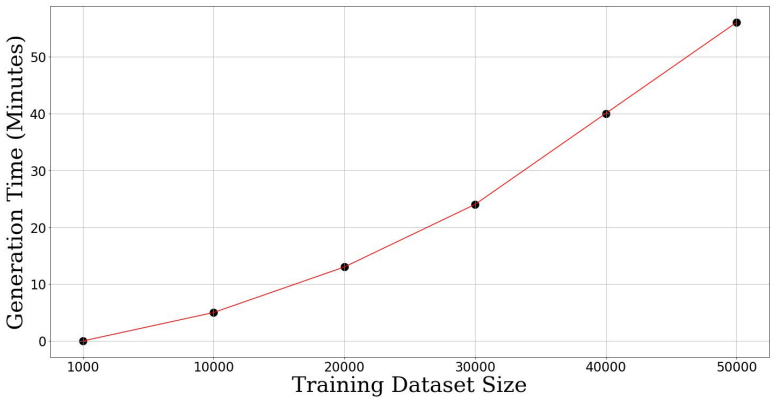


Figure 4.17: Training dataset size versus generation time of those datasets shows a non-linear increase of time with increasing size of dataset.

4.5.2 Case 2: Predicting Porosity and Saturation from Dispersive Gathers

Now I generate dispersive seismic gathers by setting $\tau = 1.0 \times 10^{-1}$, which let attenuation/dispersion occur within the seismic frequency range, as seen in Figure 4.18. I generate training and test datasets by sampling from histogram distributions shown in Figure 4.3.

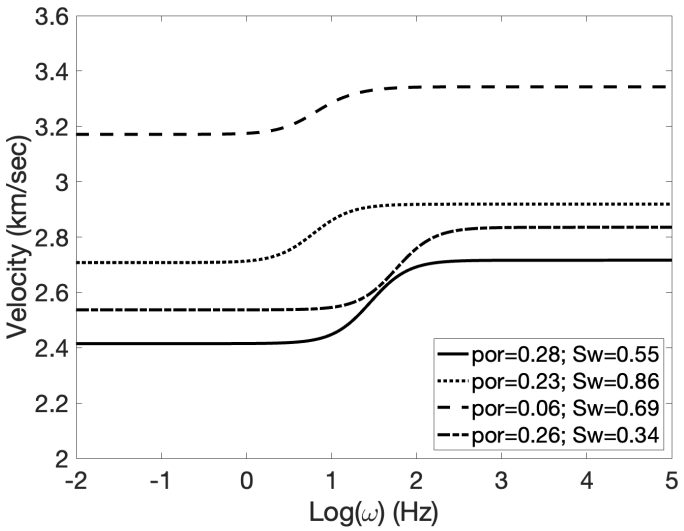


Figure 4.18: Modeled velocity with different porosity and saturation combinations shows dispersion within the seismic frequency range when $\tau = 1.0 \times 10^{-1}$.

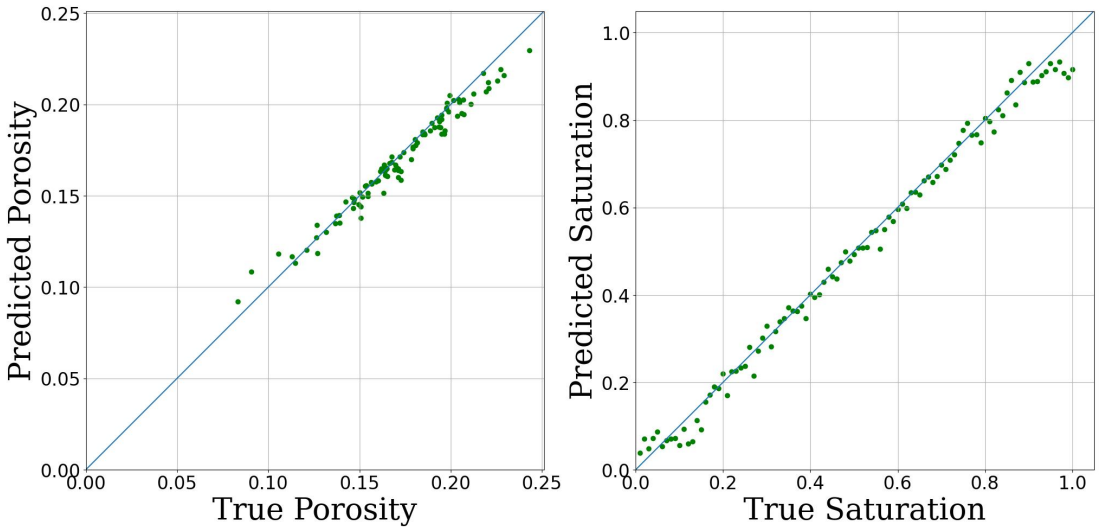


Figure 4.19: True vs. predicted porosity and saturation on an unseen dispersive test dataset, with the network trained on a dispersive dataset of 10,000 instances, assuming a fixed τ for both datasets.

Figure 4.19 presents the network's prediction results, showing a marked improvement in saturation RMSE with the incorporation of dispersion in neural network training. The prediction RMSE for saturation is 2.81×10^{-2}, representing a 75% reduction compared to predictions based on an elastic dataset of similar size. This accuracy gain is influenced by the dispersive behavior of velocity, which maintains a semi-linear relationship with saturation as gas saturation increases, enabling easier prediction of gas saturation (discussed in Section 4.6). This contrasts with the non-linear velocity behavior under elastic assumptions, known as the fizz-gas effect, which exhibits a sharp drop in velocity at low gas saturation but remains constant thereafter (see Section 2.3). This finding demonstrates the potential of neural networks for predicting saturation from dispersive gathers and underscores the importance of including attenuation as an additional input for inverting saturation from seismic data.

Porosity predictions show an RMSE of 5.94×10^{-3}, reflecting a 20% improvement in prediction accuracy over the elastic dataset. However, this improvement is less pronounced due to the weaker influence of attenuation on the porosity-velocity

relationship (further discussed in Sections 4.5.3 and 4.6). The scatter in predicted values may be attributed to Gaussian noise within the gathers

I aim to determine how the size of the training dataset affects the accuracy of porosity and saturation predictions. I prepare nine datasets with incrementally increasing numbers of instances and analyze the resulting prediction RMSE for porosity and saturation. As illustrated in Figure 4.20, datasets below 400 instances shows high RMSE for porosity and saturation prediction, most likely due to the lack of sufficient instances that they network can use to extract the underlying relationship between features and labels. After that, the prediction stabilizes although minor reduction in RMSE can still be achieved with increasing dataset size. These findings help in estimating the expected RMSE range for porosity and saturation based on the available training data.

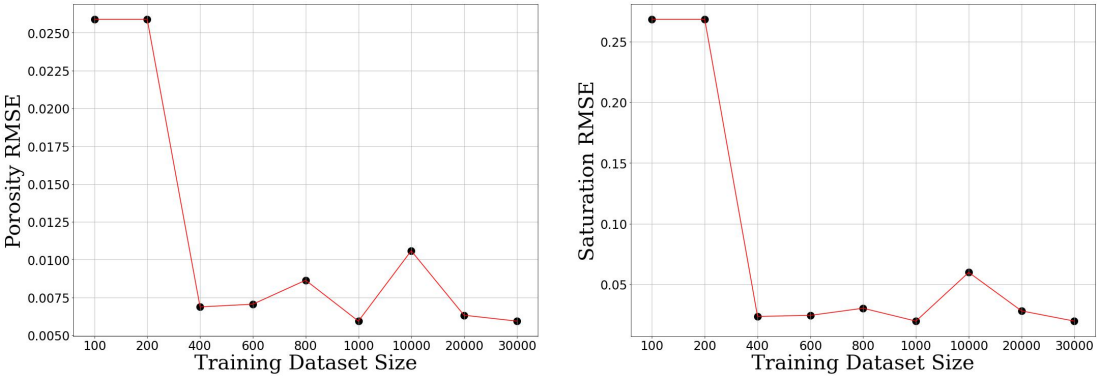


Figure 4.20: Porosity and saturation prediction RMSE from dispersive gathers as training dataset size increases.

Noise present in the training dataset inevitably affects the network's prediction accuracy. Field settings introduce noise from multiple sources, including inherent acquisition noise in the gathers. My objective is to establish a correlation between noise and network prediction errors. Using this correlation, I aim to select a threshold for acceptable noise levels in the data.

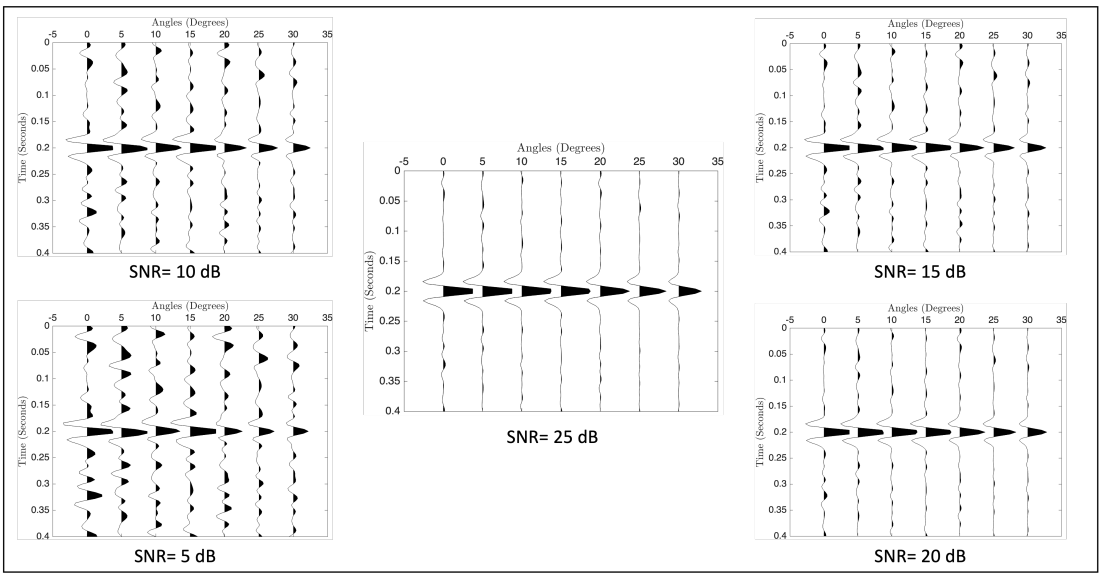


Figure 4.21: Modeled gathers with different SNR.

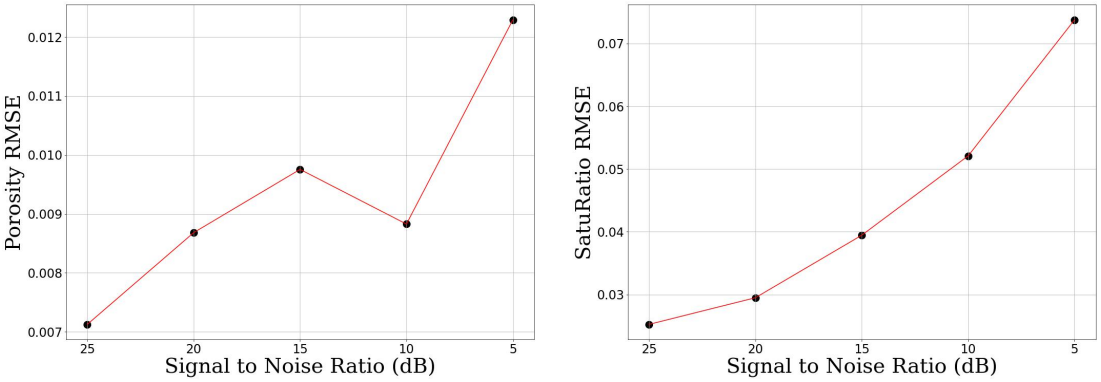


Figure 4.22: Porosity and saturation prediction RMSE error as a function of signal to noise ratio within modeled gathers.

I investigate the impact of Gaussian noise by creating five distinct training datasets, each with 10,000 instances, modeled with SNR of 25, 20, 15, 10, and 5 dB using the equations discussed in Section 4.2. Each dataset is accompanied by corresponding test datasets featuring similar noise levels, see Figure 4.21. Figure 4.22 illustrates that there is a general increase in RMSE for both porosity and saturation as the SNR decreases. In particular, there is a substantial increase in RMSE for saturation once the SNR drops below 15 dB in the training dataset. Therefore, I establish 15 dB as the minimum acceptable SNR in the training dataset based

on this observation. Although, increased noise level within features can act as a regularization tool to enhance network prediction accuracy (Bishop, 1995), it appears here that increased noise levels within the training gathers led to masking of underlying relationships between features and labels, resulting in increasing values of prediction RMSE.

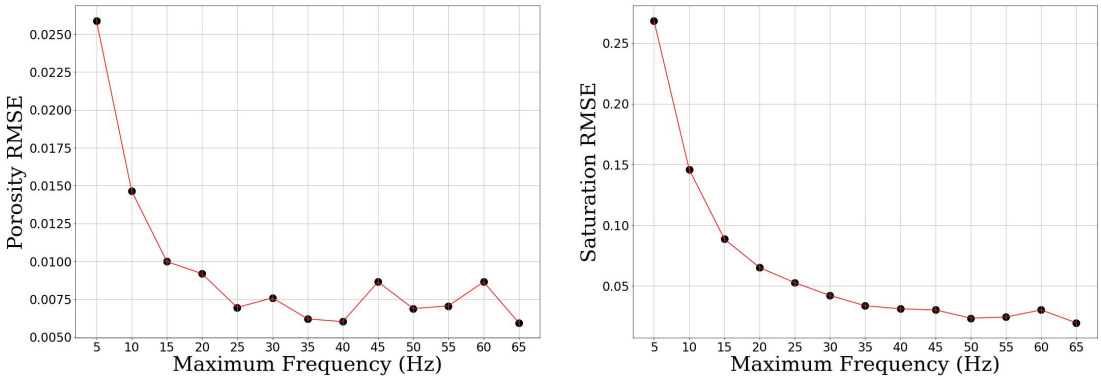


Figure 4.23: Impact of reducing frequency information on network prediction RMSE.

Seismic data processing play a key role in standard AVO analysis of the data. The main goal of data processing is to improve the signal-to-noise ratio without compromising the useful signal bandwidth inherent in the recorded data throughout the analysis stages (Yilmaz, 2001). I have discussed before the impact of noise level on seismic gather. Here, I would like to discuss the impact of frequency content of the data and maximum offset on the prediction accuracy of the network. First, I work on reducing the frequency information given to the network by incrementally removing frequency bins from the input feature matrix.

Figure 4.23 shows that the network's RMSE for porosity and saturation gradually increases when the frequency falls below 40 Hz. This implies that for an effective utilization of dispersion and attenuation to predict saturation, the features extracted must encompass frequency content up to 40 Hz. Next, I explore the effects of eliminating offset information from the feature matrix by systematically reducing maximum incident angle traces from the feature matrix. Figure 4.24 illustrates that there is a notable improvement in saturation predictions between

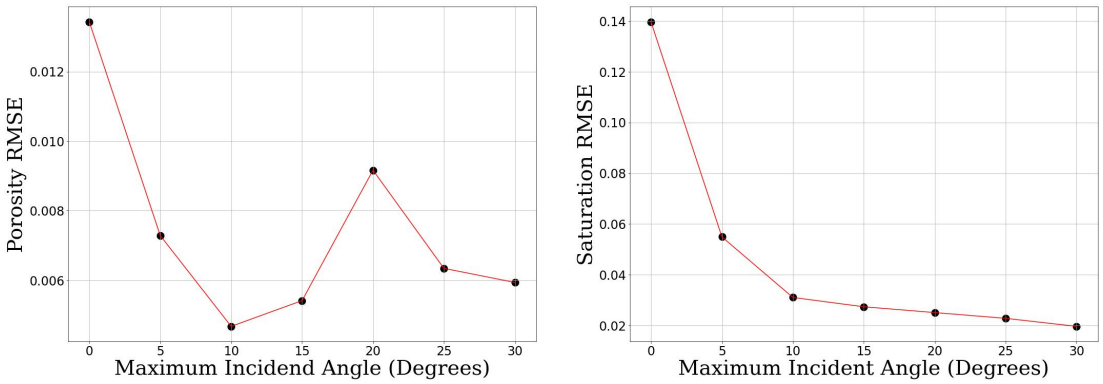


Figure 4.24: Impact of reducing maximum offset on network prediction RMSE for porosity and saturation.

0° and 10°. However, traces beyond a 10° incident angle contribute less to improvements in porosity and saturation predictions. This may be attributed to the smaller amplitude spectrum magnitudes at far offsets, which are more susceptible to noise contamination and therefore have a reduced impact on network training, see Figure 4.2.

4.5.3 Case 3: Predicting Porosity and Saturation from Gathers with Unknown τ Values

Until now, I have assumed knowledge of τ, allowing the generation of sufficient training instances for the network by varying porosity and saturation. However, in real field settings, limited well data often restricts proper network training, and obtaining τ from the reservoir is both challenging and costly (see Section 2.4). To better reflect real conditions, I analyze the network's performance when trained on a range of τ values, porosity, and saturation but tested on a dataset with varying porosity and saturation while τ remains fixed and unknown, set at $log(\tau) = -1.0$ to allow attenuation within the seismic frequency range. I generated a training dataset of 10,000 instances using input parameter distributions shown in Figure 4.3. For τ, values were sampled uniformly from $\log(\tau) \in [-4, 2]$ (Figure 4.25).

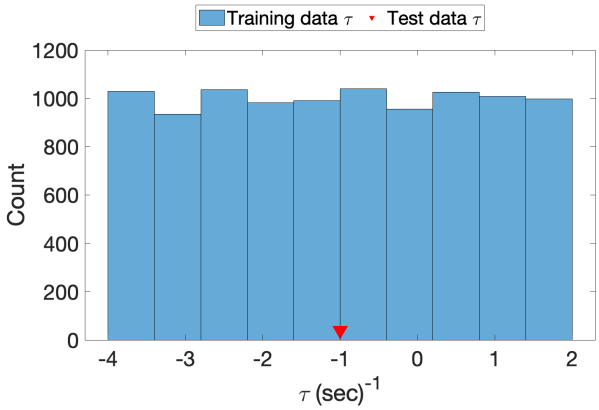


Figure 4.25: Distribution of training dataset τ value (blue) and test dataset τ (red).

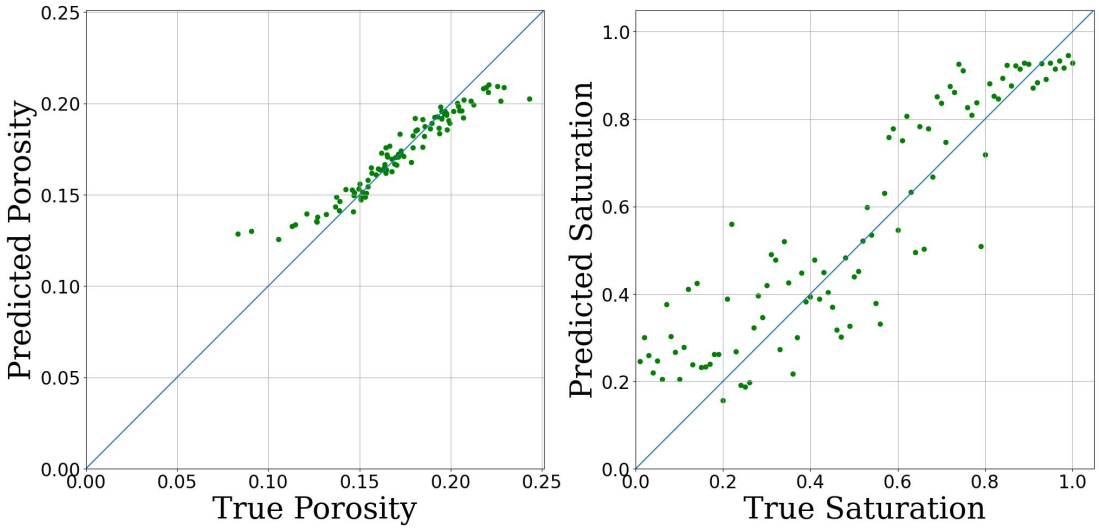


Figure 4.26: True vs. predicted porosity and saturation on a dataset with a fixed but unknown τ value, using a network trained on 10,000 instances with gathers modeled across various τ, porosity, and saturation values.

Figure 4.26 displays the network's prediction results, obtained after training it on a wide range of τ values and subsequently applying it to a test dataset with varying porosity and saturation but a fixed τ value. The network exhibits poor accuracy in predicting saturation, particularly when saturation falls below 85%, resulting in an RMSE of 13.00×10^{-2}. In contrast, porosity prediction yields an RMSE of 1.10×10^{-2}, with higher errors observed at extreme high and low porosity values. The RMSE for porosity increased by 85%, while the RMSE for

saturation surged by 330%, compared to the network trained on a fixed, known τ with a similar dataset size (see Section 4.5.2),

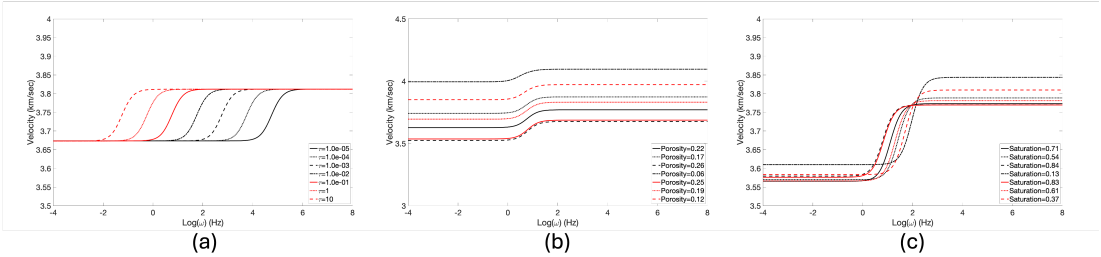


Figure 4.27: Modeled velocity curves under the following conditions: (a) varying τ values with fixed porosity and saturation; (b) varying porosity with fixed τ and saturation; (c) varying saturation with fixed τ and porosity.

Figure 4.27 illustrates the effects of varying τ, porosity, and saturation on modeled velocity. Variations in τ shift the transition zone along the frequency axis (see Section 2.4). Similarly, changes in saturation alter the position of the transition zone due to fluid viscosity effects (Equation 2.4), which also shift the transition zone and create non-uniqueness, as both parameters similarly influence velocity across the frequency spectrum. In contrast, porosity primarily affects velocity vertically by impacting the bulk modulus (Equation 2.8), though minor shifts in the transition zone may still occur due to fluid effects in more porous rocks compared to tighter formations.

Therefore, it is expected that saturation predictions would be more adversely affected by training the network on multiple τ and saturation values than porosity predictions. This expectation is supported by the significant increase in RMSE for saturation (330%) compared to porosity (85%).

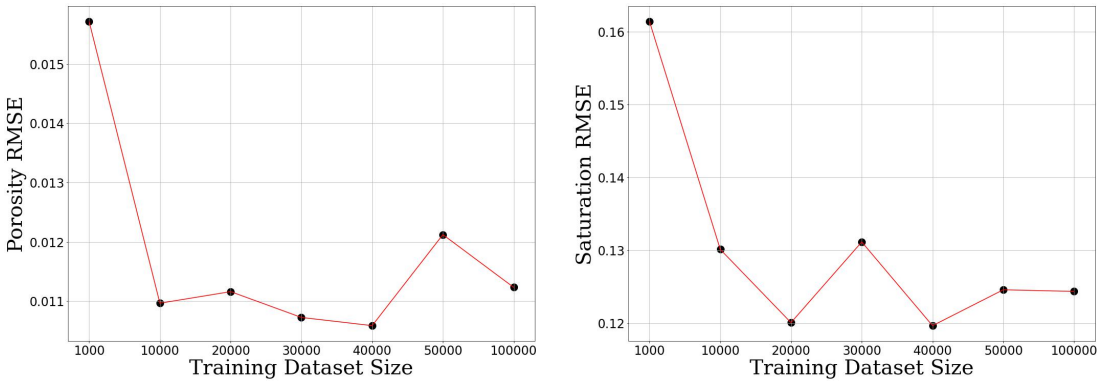


Figure 4.28: Network prediction RMSE of porosity and saturation using different training dataset sizes with multiple τ values.

I explored methods to improve network predictions trained on an uncalibrated rock physics model. One strategy was to enlarge the training dataset. I created seven different datasets, each with increasing numbers of instances, and trained seven networks with these. The networks' porosity and saturation prediction accuracies were assessed using a consistent test dataset. Figure 4.28 reveals that saturation RMSE reached about 12.00×10^{-2} with a dataset of 20,000 instances, a 7% improvement over the 10,000-instance dataset. Further increases in dataset size did not significantly change saturation RMSE and the network prediction RMSE seem to oscillate around 12.5×10^{-2}. As for porosity, the prediction RMSE oscillates around 1.10×10^{-2} and reaching a minimum of 1.05×10^{-2} when trained on dataset size of 40,000 instances, showing a minor improvment of 5% in RMSE compared to dataset size of 10,000. The best results for saturation and porosity were observed with the 40,000-instance dataset, as shown in Figure 4.29.

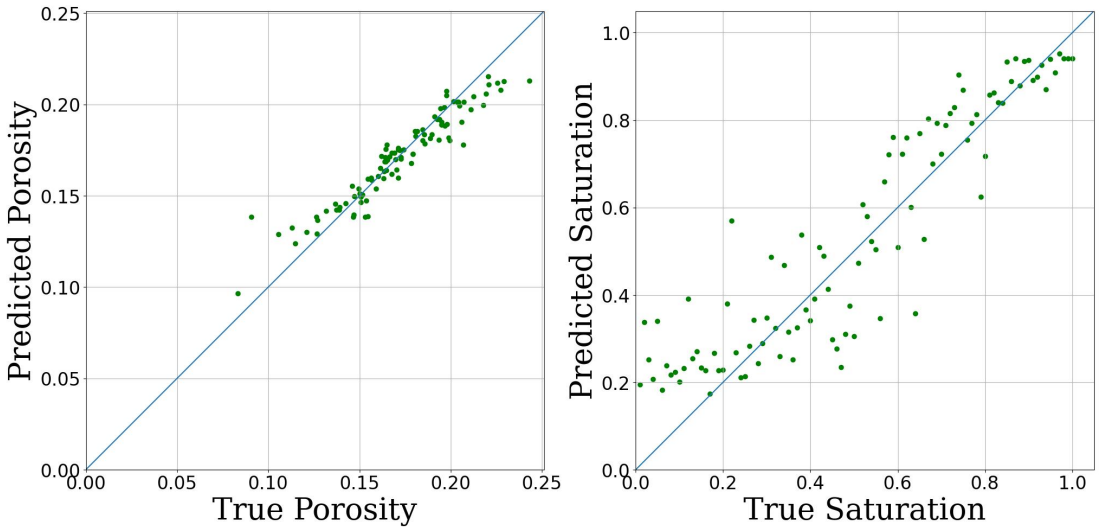


Figure 4.29: True vs. predicted porosity and saturation on a dataset with a fixed but unknown τ value, using a network trained on 40,000 instances with gathers modeled across various τ, porosity, and saturation values.

Another way to enhance the prediction accuracy of the network is by modifying the neural network architecture. From the analysis I have conducted on Section 4.4, I found that three factors strongly impact the saturation prediction accuracy. Those are the learning rate, batch size and number of hidden layers. I have excluded the dropout since it tend to deteriorate prediction accuracy, see Figure 4.11. I use a dataset size of 10,000 as a benchmark for my analysis to reduce network training time and since it gave representative RMSE for porosity and saturation as observed in Figure 4.28.

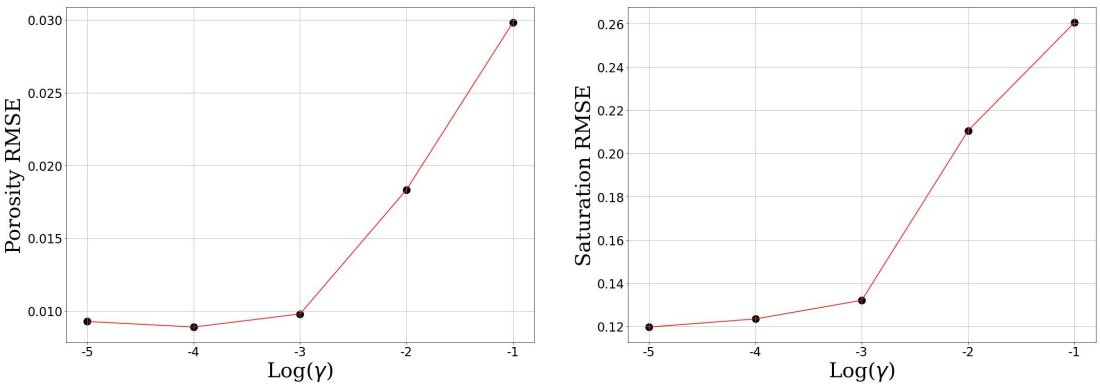


Figure 4.30: Network prediction RMSE for porosity and saturation on a fixed τ test dataset, with the network trained on multiple τ, porosity, and saturation values using different learning rates.

Starting with the learning rate, I experiment with five different learning rate values and plot the network prediction RMSE. Figure 4.30 shows no significant prediction improvement in RMSE over the original learning rate of 1.0×10^{-3} for porosity. As for saturation, there is a minor improvement of 7.5% between the learning rate of 1.0×10^{-3} and 1.0×10^{-5}. However, choosing this learning rate tend to prolong the training time by 900 % compared to 1.0×10^{-3} as observed in Figure 4.8, while giving only a minor improvement of 7.5 %. Therefore, no lower learning rates were assessed below 1.0×10^{-5}.

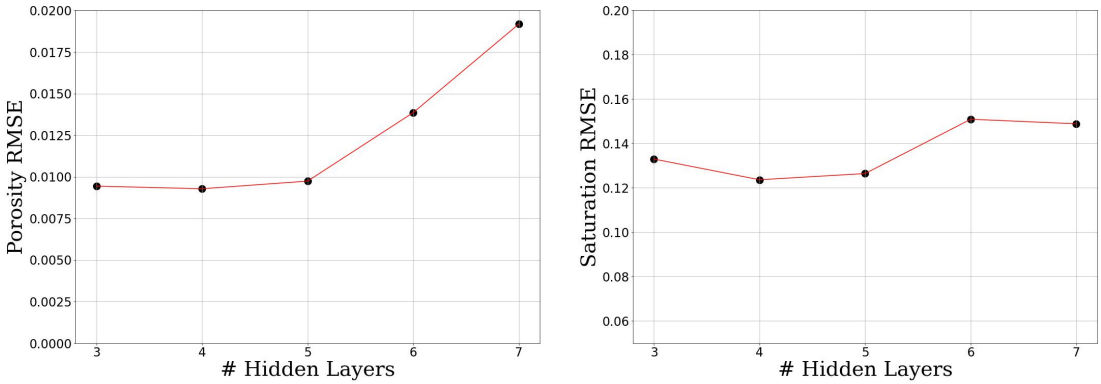


Figure 4.31: Network prediction RMSE for porosity and saturation on a fixed τ test dataset, with the network trained on multiple τ, porosity, and saturation values using different number of hidden layers.

Next, I experiment with the number of dense layers, by changing the number of hidden layers . Figure 4.31 illustrates that increasing the number of dense layers between three and 5 hidden layers results in a 5% reduction in saturation prediction RMSE while maintaining similar RMSE for porosity. However, going beyond that there is an overall increase in RMSE, probably due to data overfitting with increase model parameters. Therefore, increasing network complexity by adding more hidden layers does not appear to be an effective approach for enhancing predictions when τ is unknown.

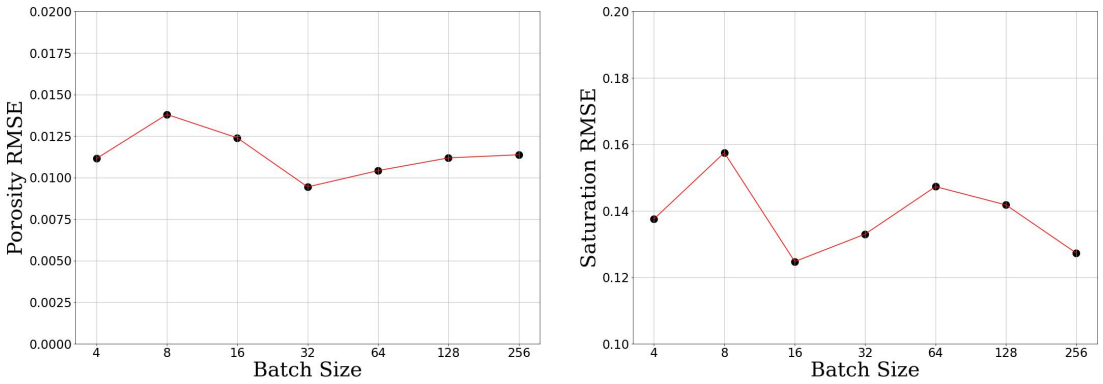


Figure 4.32: Network prediction RMSE for porosity and saturation on a fixed τ test dataset, with the network trained on multiple τ, porosity, and saturation values using different batch sizes.

Finally, I examine the impact of varying batch sizes on the network prediction accuracy when trained on multiple τ values. Figure 4.32 demonstrate the impact of batch size on network prediction accuracy. The minimum RMSE for porosity and saturation occurs at batch sizes 32 and 16, respectively. The RMSE for porosity is 9.7×10^{-3} and for saturation is 12.5×10^{-2} at those batch sizes, which are only 8% and 4% lower in RMSE than the analysis in Figure 4.26. This minimal improvement suggests that batch size parametrization has a negligible impact on the overall prediction accuracy.

4.6 Discussion

In Case 1, where the network was trained on elastic gathers (see Section 4.5.1), the results clearly demonstrate the challenge of distinguishing between high and low gas saturation when saturation drops below 85% in a scenario representing the purely elastic, low-frequency relaxed limit. This difficulty arises from a significant drop in the velocity when a small amount of gas initially displaces water. After this initial change, the velocity remains relatively constant as more gas enters the pores (see Section 2.3).

The velocity-saturation relationship at $\log(\tau) = -5$ is shown in Figure 4.33.

Using reference rock parameters listed in Table 2, velocity was modeled with Equation 2.6 at $\omega = 25$ Hz, assuming constant porosity. From this figure, it is evident that relying solely on the velocity profile makes it difficult to differentiate between high and low gas saturation, as multiple saturation values have similar velocity measurements. Conversely, the relationship between porosity and velocity (assuming constant saturation) displays a more linear trend, making it easier for the neural network to learn and predict, which explains the lower prediction RMSE for porosity in Figure 4.15.

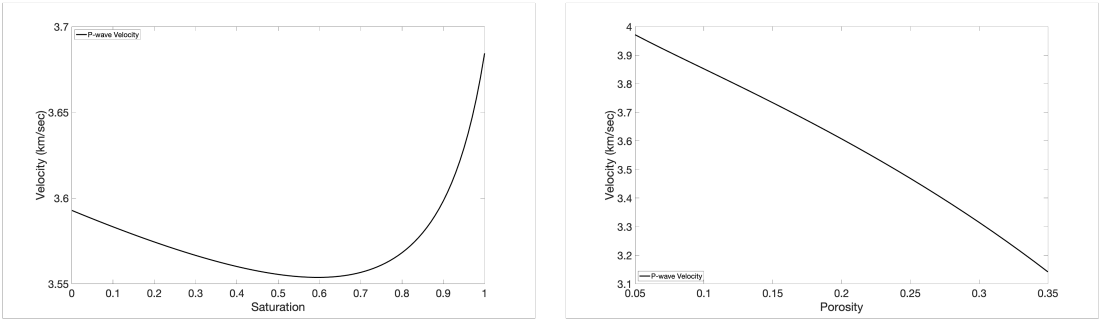


Figure 4.33: Left: Velocity-Saturation relationship at constant porosity shows small changes in velocity below Sw=85% over wide range of saturation. Right: Velocity-Porosity exhibits a more linear trend at constant saturation.

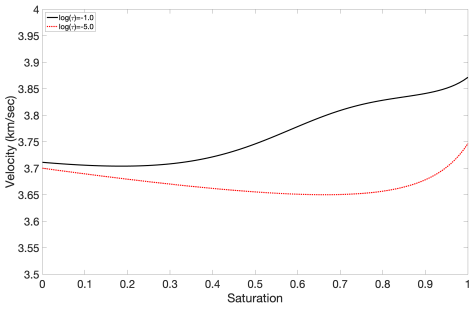


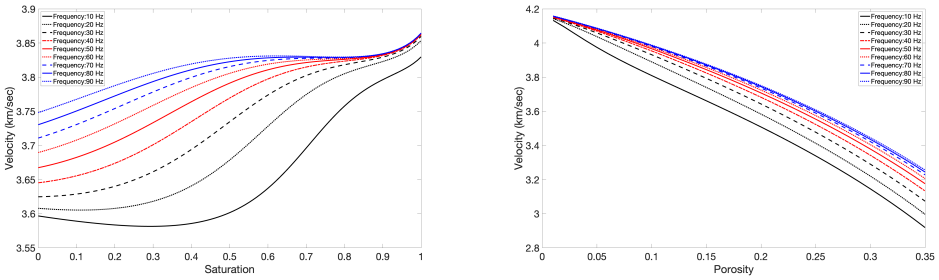
Figure 4.34: Velocity-Saturation relationship at dispersive ($log(\tau) = -1.0$) and elastic ($log(\tau) = -5.0$) values, assuming fixed porosity.

In Case 2, where dispersive gathers were modeled with a τ value that allowed attenuation within the seismic frequency range (see Section 4.5.2), introducing attenuation into the network training had a notable impact on predictions of both saturation and porosity compared to the purely elastic Case 1. For porosity,

RMSE decreased by 20%, while for saturation, RMSE decreased by 75% with the addition of attenuation.

Figure 4.34 provides a comparative view of dispersive and elastic velocity behaviors with respect to porosity and saturation. Both velocity types were modeled using the same parameters according to Equation 2.6 at a frequency of $\omega = 25$ Hz, with the only difference being the value of τ. The figure shows that the dispersive velocity curve does not exhibit a significant drop in P-wave velocity as gas saturation increases. This is primarily due to the rock being in an unrelaxed state, resulting in an apparent velocity increase (see Section 2.4).

Figure 4.35 illustrates the relationship between velocity and both saturation and porosity at different frequencies, using $\log(\tau) = -1.0$ as a reference. At lower frequencies, distinguishing saturation between 0.01 to 0.2 is challenging, as multiple saturation values yield the same velocity. However, as frequency increases, velocity changes become more pronounced, particularly at lower saturations. In contrast, porosity shows minimal change with frequency compared to saturation, as dispersion is primarily influenced by fluid movement rather than porosity (see Section 4.5.3). The main impact of dispersion will becomes only prominent at high porosities, since fluids will have larger impact on bulk density as porosity increases (see equation 2.8).



(i) **(ii)**

Figure 4.35: (**i**) Velocity-Saturation curves at various frequencies when $log(\tau) = -1.0$ and assuming fixed porosity. (**ii**) Porosity-Saturation curves at various frequencies when $log(\tau) = -1.0$ and assuming fixed saturation. The black and red colors represent velocity curves within the seismic frequency range.

In Case 3, I explored a scenario where the key calibration parameter, τ, was unknown along with porosity and saturation. The network was trained on a dataset encompassing a wide range of τ values, resulting in a 330% increase in RMSE for saturation and a 85% increase for porosity compared to a network trained with a fixed, known τ. To improve prediction accuracy, various strategies were tested, including expanding the dataset and adjusting network parameters and architecture. While these adjustments led to some improvements, no significant gains in accuracy were observed, underscoring the importance of calibrating τ for accurate saturation and porosity predictions.

In scenarios such as heavy oil or CO_2 sequestration projects, dispersion and attenuation within seismic frequencies can play a crucial role in enhancing saturation prediction accuracy, as discussed in Chapter 2 and shown in Section 4.5.2. However, obtaining sufficient measurements to build a robust neural network training dataset can be challenging.

This highlights the need for advanced rock physics models capable of simulating reservoir elastic behaviors to improve neural network training datasets for more accurate property predictions. During early exploration stages, key parameters like τ may be unknown. Case 3 demonstrated that attempting to train a network across various τ values, due to its unknown value from field measurements, significantly reduces accuracy, as illustrated in Figure 4.36.

In Chapter 3, I explored the varied applications of Physics-Guided Neural Networks (PGNNs) in geophysical studies. In the following chapter, I introduce a novel PGNN workflow designed to facilitate the application of uncalibrated rock physics models using only basic field measurements. This workflow incorporates a physics-guided loss function that leverages available field data for implicit calibration within the network, aiming to improve the network's accuracy in predicting reservoir properties even with limited field data and some unknown model pa-

rameters.

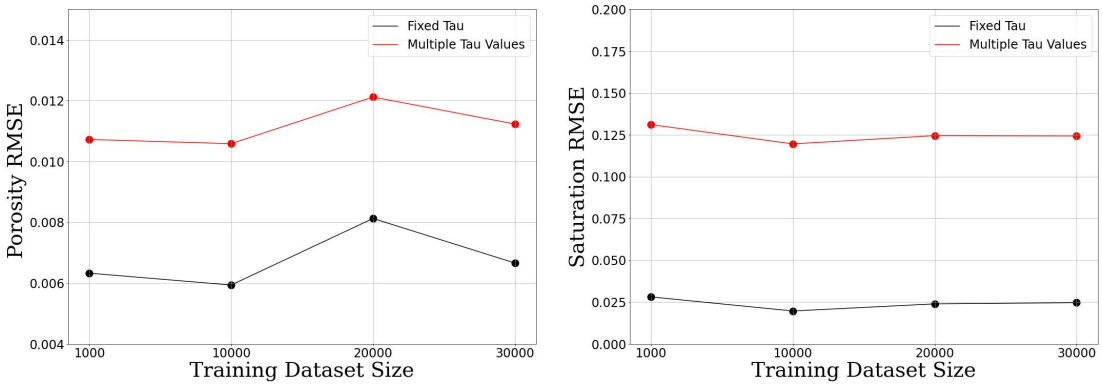


Figure 4.36: Network predictions of porosity and saturation when trained on a dataset with a fixed τ value versus a wide range of τ values.

5 Chapter 5: Physics Guided Neural Network Training Theory and Application to a Dispersive Rock Physics Model with Unknown τ

5.1 Chapter Objectives

The previous chapter concluded by addressing the challenge of training a neural network with a dataset generated from uncalibrated rock physics models, due to the scarcity of advanced field measurements. Specifically, I focused on the difficulty of calibrating τ in the dispersive rock physics model by Papageorgiou and Chapman (2017), and how using a wide range of τ values affects the accuracy of neural network predictions for reservoir saturation and porosity.

In this chapter, I present a new variant of physics-guided neural networks, tailored to implicitly calibrate advanced rock physics models using limited basic field measurements that are often obtained in the early stages of field exploration. Specifically, I demonstrate this approach for calibrating the τ parameter when its actual value is unknown.

5.2 Introduction

At the early stages of exploration, we typically have extensive seismic data paired with limited basic core and well measurements. This includes logs such as sonic, shear, porosity, density, saturation, and clay volume, along with core analyses like porosity and facies description. These basic measurements are sufficient for straightforward rock physics modeling in isotropic and homogeneous reservoirs using models like Gassmann's equations (Connolly, 1999; Sbar, 2000). However, for more complex reservoirs, such as those with seismic frequency dispersion or anisotropic reservoirs , advanced rock physics modeling is necessary, which these fundamental measurements alone cannot be used to calibrate.

In Chapter 4, I presented a specific example concerning dispersive seismic data, highlighting the importance of knowing the time scale constant (τ). I demonstrated that training a neural network with an uncalibrated rock physics model in such complex scenarios results in poor predictions of reservoir properties, as detailed in Section 4.5.3.

In this chapter, I present a unique PGNN workflow that facilitates the application of uncalibrated rock physics models using only limited basic field measurements. I assume that certain key model parameters significantly affect seismic responses in predicting and inverting reservoir properties from seismic data but cannot be predicted or measured from seismic or core data—a challenge highlighted in Section 4.5.3. Despite this, I propose that a neural network can be trained to capture the general relationships between reservoir properties, unknown parameters, and seismic attributes using diverse combination of model parameters generated by an uncalibrated rock physics modeled. By applying the physics-guided training principle, the network training can be constrained by the implicit relationship between reservoir properties and extracted seismic attributes within the field data, which indirectly contain the relationship between unknown model parameters and reservoir properties. The level of this constraint is controlled with the hyperparameter λ. A crucial assumption of this workflow is that the unknown parameter(s) remain consistent across different field measurements.

5.3 Theoretical Derivation of PGNN Loss Function

I assume a small number N of field measurements from drilled wells and core analysis, complemented by seismic responses extracted from nearby seismic gathers. These measurements are limited and not sufficient for comprehensive neural network training. Together, they form my field dataset D_F, represented as follows:

$$D_F = \{(\mathbf{x_i}, \mathbf{y_i})\}; i \in (1, N), \tag{5.1}$$

where $\mathbf{y_i}$ is a label vector of reservoir properties at well i, $\mathbf{x_i}$ is a feature vector of extracted seismic amplitudes from gathers around the same well.

Next, I define a rock physics model $\mathcal{M}$ that takes a vector of reservoir parameters $(\mathbf{y}; \mathbf{p})$, where $\mathbf{p}$ is a vector of unknown model parameters that requires calibration, and perform a seismic modeling exercise to output seismic features $\mathbf{x}$ as follows:

$$\mathbf{x_i} = \mathcal{M}\{\mathbf{y_i}; \mathbf{p_i}\}; i \in (1, M), \tag{5.2}$$

where M is arbitrary large. From equation 5.2, I form a rock physics modeled neural network training dataset D_M such that:

$$D_M = \{(\mathbf{x_i}, \mathbf{y_i})\}; i \in (1, M). \tag{5.3}$$

Now I define a neural network $\mathcal{F}$ that takes input features $\mathbf{x}$ and output predicted labels $\hat{\mathbf{y}}$ as follows:

$$\mathcal{F}(\mathbf{x}) = \hat{\mathbf{y}}. \tag{5.4}$$

I seek to train $\mathcal{F}$ on a combination of instances from D_F and D_M. To that end, I define a batch B as:

$$B = B_F \cup B_M, \tag{5.5}$$

where B_F and B_M are Z and Q instances from D_F and D_M respectively that the network train on before it updates its weights as the weighted sum of loss functions over field and model data batches:

$$B_F = \{(\mathbf{x_i}, \mathbf{y_i}\}; i \in (1, Z), \tag{5.6}$$

$$B_M = \{(\mathbf{x_i}, \mathbf{y_i}\}; i \in (1, Q), \tag{5.7}$$

Finally, I write my custom loss function C over the batch B as:

$$C(B) = (1 - \lambda) \sum_{B_M} |\mathbf{y} - \mathcal{F}(\mathbf{x})| + (\lambda) \sum_{B_F} |\mathbf{y} - \mathcal{F}(\mathbf{x})|, \tag{5.8}$$

where λ is a hyperparameter between $[0, 1]$ that controls the loss input between B_M and B_F.

5.4 Assumptions of PGNN Loss Function

To effectively apply Equation 5.8, it is essential that three key assumptions are satisfied. First, any variations in calibration's unknown parameters should impact the extracted seismic features, allowing the network to understand their effect on the relationship between reservoir and elastic properties. Second, the seismic modeling of elastic properties by the rock physics model must accurately mirror the field seismic gathers in terms of seismic features, ensuring that the relationships learned from modeled and field data are interchangeable. Lastly, the unknown parameters must remain consistent within a specific geographical area without erratic variations, enabling the network to isolate and respond to their influence on seismic attributes.

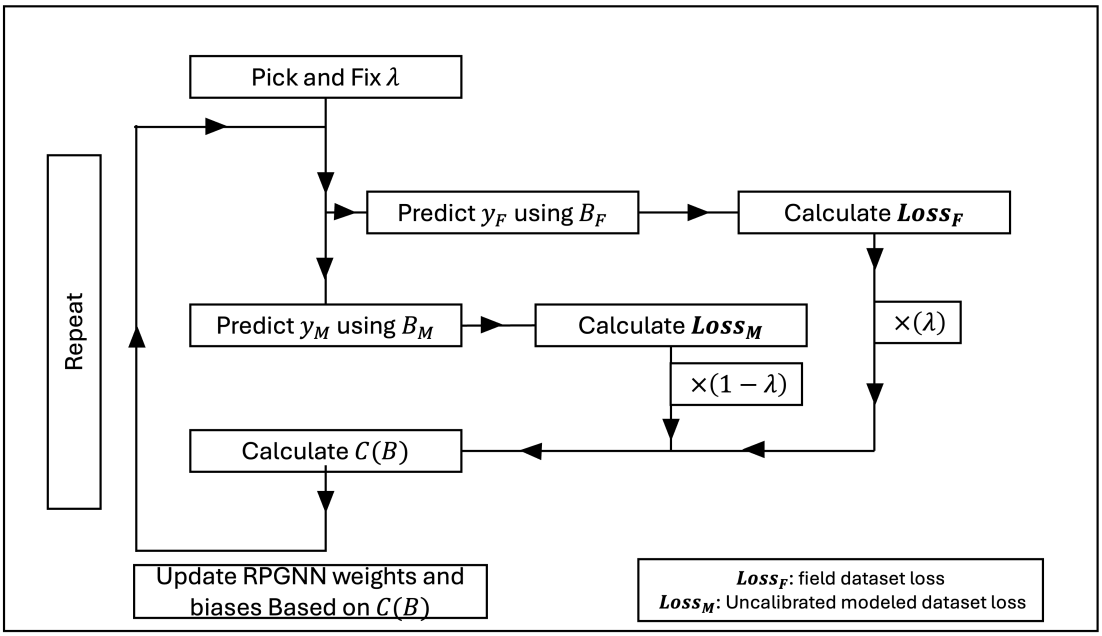


Figure 5.1: Physics-guided neural network training cycle.

5.5 Implementing the PGNN Loss Function: Methodology and Considerations

Figure 5.1 shows the training cycle of the network. Initially, I set a constant value for λ for the duration of the training. The network weights are initialized using the method described by Glorot and Bengio (2010), see Section 4.4. Next, the network processes feature vectors from batches B_F and B_M, producing respective predictions $\hat{\mathbf{y}}_\mathbf{F}$ and $\hat{\mathbf{y}}_\mathbf{M}$. The loss C is calculated as per equation 5.8. Based on this loss, the network weights are updated, using the training method outlined in Section 3.3, and this cycle repeats until the network reaches convergence. In principle, a network that minimizes C based on batches from B_F and B_M should effectively learn the underlying relationships between reservoir properties as labels and seismic amplitudes as features, allowing it to minimize B_M while also being implicitly calibrated to the field data to minimize B_F.

Implementing the physics-guided neural network loss function presents two major challenges. The first is the absence of a designated validation dataset, which is crucial for monitoring training convergence and preventing overfitting, see Section 3.3. The second challenge involves selecting an appropriate value for the hyperparameter λ, essential for balancing the losses from the field dataset (D_F) and the modeled dataset (D_M) in the PGNN workflow.

To tackle the convergence challenge, I split the field dataset D_F into two balanced subsets based on the distribution of field measurement labels. One subset is used for training with Equation 5.8 and the other for validation. The validation loss is checked after each training epoch, and training halts if the loss stabilizes or increases for a set number of epochs. I record the epoch count, reset the network, and switch the roles of the subsets for a second training cycle, recording the epoch count again. Finally, I reset the network and train on the entire D_F dataset using the average epoch count from both cycles as the stopping criterion. Although this method does not guarantee absolute convergence, it provides a

practical way to assess convergence with the limited data in D_F. The impact of this cross-validation method compared to using a dedicated validation dataset on the PGNN workflow is discussed in Section 5.7.3.

For choosing the optimal λ value, I assume having a separate test dataset. Various λ values are selected and tested on this dataset, which is similar to D_F but not used in training. The optimal λ value is determined by the one that results in the lowest loss on this test dataset. Although having such a test dataset is often unlikely in real-world scenarios, this method helps set guidelines for adjusting λ in such cases.

5.6 Numerical Example

I apply my PGNN loss function on the Case 3 synthetic example discussed in Section 4.5.3. I use the same neural network architecture, shown in Figure 4.4. My rock physics modeling exercise $\mathcal{M}$ uses Papageorgiou and Chapman's (2017), discussed in Section 4.3. Random noise is added to each gather to generate gathers with a SNR of 25 dB. I assume knowledge of some reservoir property distributions based on basic log measurements and core analysis, depicted in Figure 4.3. These are used as inputs for my rock physics model. I also assume that fluid is uniformly distributed ($q = 1$) within both layers when calculating fluid properties using equation 2.4.

For my modeling dataset D_M, I selected a dataset size of 10,000 training instances, with τ as the unknown parameter sampled from the uniform distribution shown in Figure 4.25. Training on this dataset resulted in RMSE values of 1.1×10^{-2} for porosity and 13.0×10^{-2} for saturation when using neural networks (see Section 4.5.3, Figure 4.28), which closely represents the average RMSE for porosity and saturation across different modeled training dataset sizes discussed in Section 4.5.3.

The dataset includes porosity and saturation as target labels ($\mathbf{y_m}$) alongside the feature matrix $\mathbf{x_m}$, where $\mathbf{x_m}$ is a (13×7) matrix created using the workflow described in Section 4.3. The rows of the feature matrix represent amplitude spectrum values recorded at 13 frequency bins from 0 to 60 Hz, sampled at 5 Hz intervals. The columns of $\mathbf{x_m}$ represent the amplitude spectrum values extracted at each gather trace, with traces sampled at 5° intervals between 0° and 30° incident angles. I define D_M as:

$$D_M = \{\mathbf{x_m}, [\phi_m, S_{w_m}]\}; m \in (1, 10,000), \tag{5.9}$$

where ϕ_m and S_{w_m} represent the input rock physics model values for porosity and saturation, respectively, utilized to generate the seismic gather.

I constructed a synthetic field dataset by assuming 40 measurements of porosity (ϕ_f) and saturation (S_{w_f}), which is significantly fewer than the 400 instances required for accurate porosity and saturation predictions (see Section 4.5.2, Figure 4.20). I set τ for the sandstone layer at a value that allows for dispersion within the seismic frequency range as shown previously in Figure 4.18. I follow the modeling exercise workflow that uses Papageorgiou and Chapman's (2017) rock physics model discussed in Section 4.2 to generate 40 dispersive seismic gathers using randomly sampled porosity and saturation measurements from the distributions in Figure 4.3. Random noise is added to each gather to maintain a signal to noise ratio of 25 dB. Now I construct my neural network field dataset D_F. The target labels ($\mathbf{y_f}$) are the 40 porosity and saturation measurements. The feature matrix is $\mathbf{x_f}$, where $\mathbf{x_f}$ is a (13×7) matrix that is extracted from each dispersive gathers at the same frequency bins and incident angle traces that were used to extract features matrix $\mathbf{x_m}$. Now, I write D_F as:

$$D_F = \{\mathbf{x_f}, [\phi_f, S_{w_f}]\}; f \in (1, 40), \tag{5.10}$$

Note that τ is set at a fixed value, but that value is unknown to the network.

I generate a linearly spaced range of λ values ranging from 0 to 1. For each λ value within this range, I train the network using Equation 5.8. Subsequently, I employ the trained network to predict the loss on a test dataset labeled as D_T. This dataset consists of 100 training instances, generated using the same workflow and reservoir distributions as D_F. The labels for D_T are porosity (ϕ_t) and saturation (S_{w_t}), and the feature matrix $\mathbf{x_t}$ is extracted at the same frequency bins and incident angles as those used for D_F. I write D_T as:

$$D_T = \{\mathbf{x_t}, [\phi_t, S_{w_t}]\}; t \in (1, 100), \tag{5.11}$$

I select a λ value from the linearly spaced range between 0 and 1, then initialize the network weights. The batch size for D_M is set to 32, as this batch size yielded the lowest RMSE for porosity and saturation predictions during network design (see Section 4.4). This setup means that the network loss is calculated using Equation 5.8 for 32 instances of D_M alongside all instances of D_F before updating the network weights. Convergence during training is assessed using the cross-validation method described in Section 5.5.

Once training is complete, the features of D_T are passed through the trained network, using Equation 5.4, to obtain a prediction vector $\hat{\mathbf{y}}_\mathbf{t}$ for porosity and saturation, which is then compared to the true labels in D_T. The RMSE for porosity and saturation predictions is plotted against the training λ value, and the cycle repeats for the next λ value. The RMSE is calculated using Equation 4.13.

To Understand the impact of label distribution in D_F on training and prediction accuracy, particularly in terms of convergence criteria, I generated 25 unique seeds of D_F, each with distinct porosity and saturation combinations. These datasets will be trained with D_M to assess how property distribution in D_F influences

prediction precision. To quantify the variability in predictions, the standard deviation (SD) for porosity and saturation will be computed for these 25 seeds using the equation:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}. \tag{5.12}$$

In this formula, σ represents the RMSE standard deviation across the 25 seeds, x_i denotes the RMSE of a single seed over D_T, and μ is the average RMSE for all seeds.

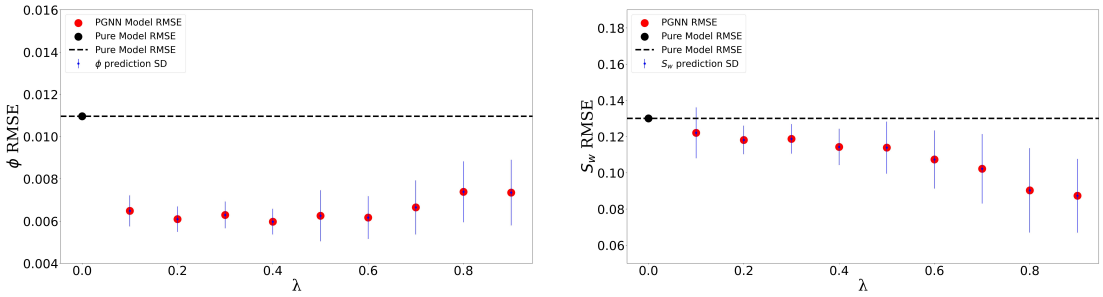


Figure 5.2: Average network predictions for porosity and saturation on the test dataset D_T using PGNN training with 25 unique seeds of D_F at regularly sampled λ values.

5.6.1 Results

Figure 5.2 displays the RMSE for porosity and saturation predictions on the test dataset D_T using 25 seeds of D_F, along with their standard deviations. On average, porosity predictions show a decrease in RMSE up to $\lambda = 0.6$, after which it increases. The minimum RMSE for porosity represents a 40% reduction compared to training solely with model data at $\lambda = 0.4$. Saturation predictions using PGNN show a continuous decrease in RMSE as λ increases, with a minimum RMSE representing a 33% reduction at $\lambda = 0.9$. Regarding standard deviation, both predictions exhibit relatively low variability at smaller λ values, increasing gradually. This trend lends greater confidence to network predictions within a smaller λ range and suggests increasing uncertainty as the network weights are more influenced by the field dataset (D_F).

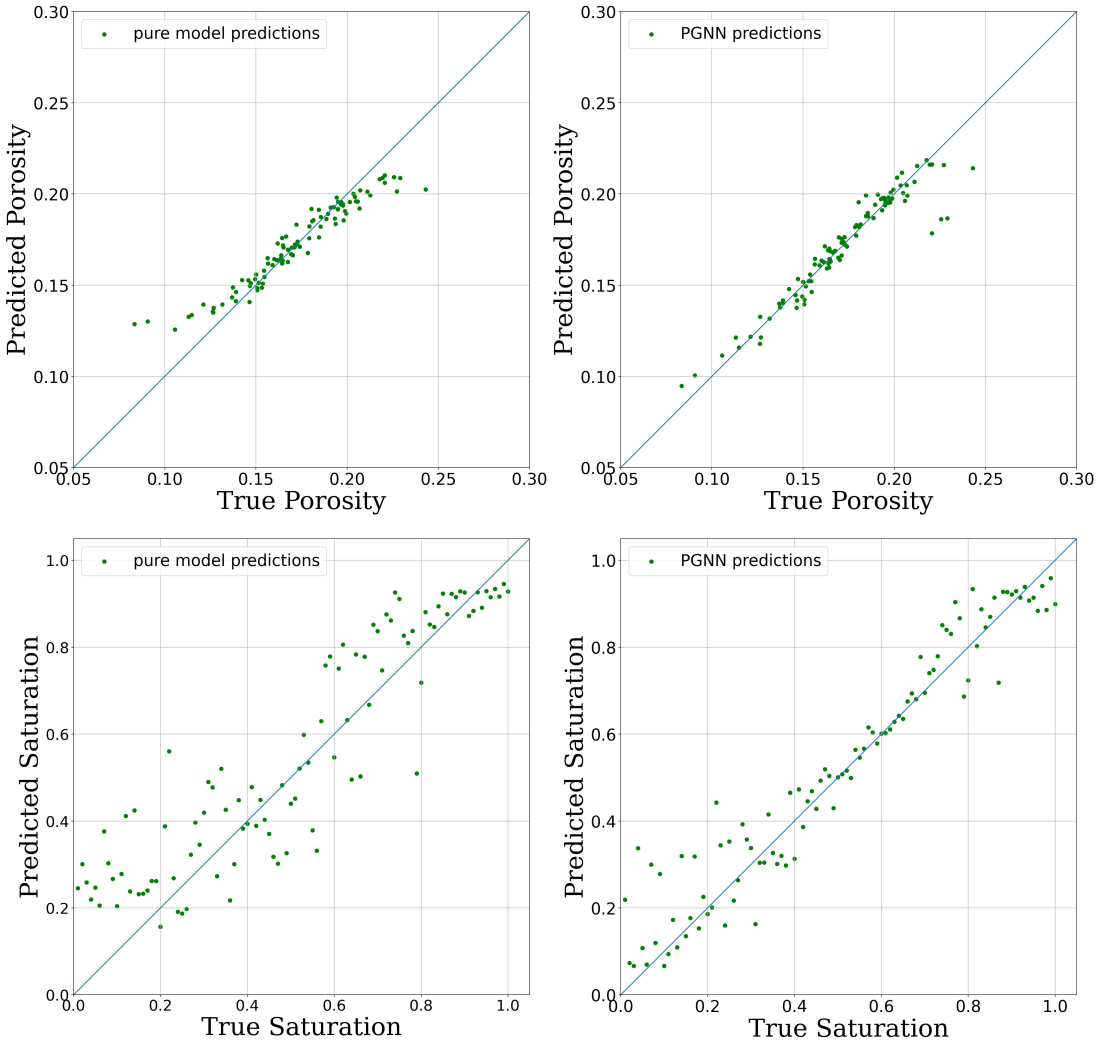


Figure 5.3: True vs. predicted porosity and saturation using the PGNN workflow on the test dataset D_T at $\lambda = 0.9$, compared to training on an uncalibrated modeled dataset.

Figure 5.3 shows the network predictions for porosity and saturation at $\lambda = 0.9$. Porosity predictions exhibit noticeable improvement with reduced variance when using PGNN training, compared to standard neural network training with only the model dataset D_M. For saturation, the accuracy enhancement is particularly significant within the range of $S_w = 0.3$ to 0.8.

In general, PGNN training results in improved porosity and saturation predictions. Saturation predictions show a decrease in RMSE with increasing λ values,

while porosity reaches a minimum RMSE at $\lambda = 0.6$, which then increases as λ continues to rise. This difference could be attributed to the distinct dynamic ranges of the two properties, with porosity ranging from 0.01 to 0.40 and saturation ranging from 0.01 to 1.0. This topic will be discussed further in Section 5.8.

Some seeds of D_F yield better predictions than others, as indicated by the standard deviation. For example, Figures 5.4 and 5.5 show network predictions on D_T using two different seeds of D_F. Both networks were initialized with the same weights and trained alongside D_M using Equation 5.8 with $\lambda = 0.9$. The results demonstrate improved porosity and saturation predictions in Figure 5.4, whereas Figure 5.5 shows prediction deterioration, highlighting the impact of D_F on overall prediction accuracy.

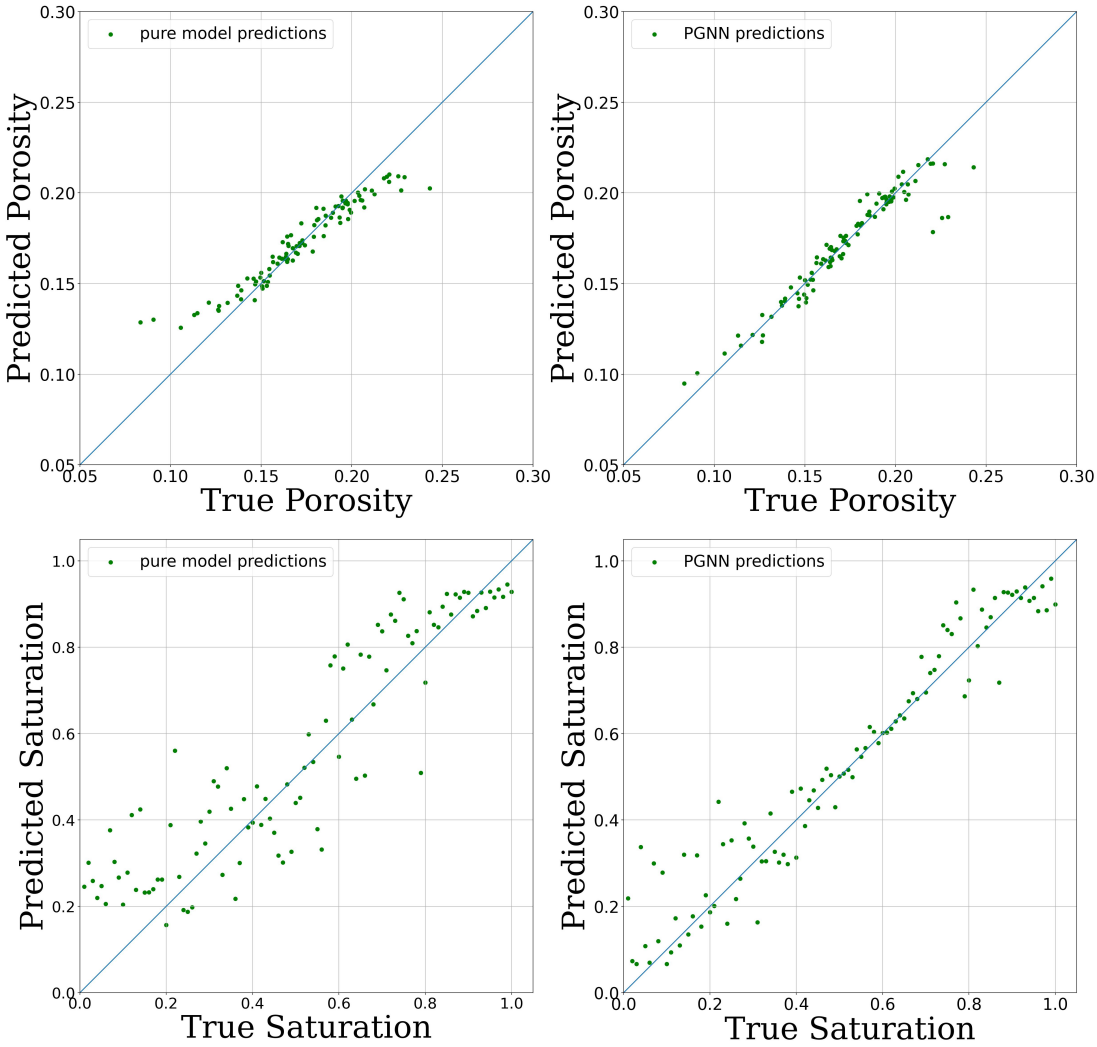


Figure 5.4: True vs. predicted porosity and saturation using the PGNN workflow with a random seed of D_F on the test dataset D_T at $\lambda = 0.9$, showing overall improvement in porosity and saturation predictions compared to pure model training.

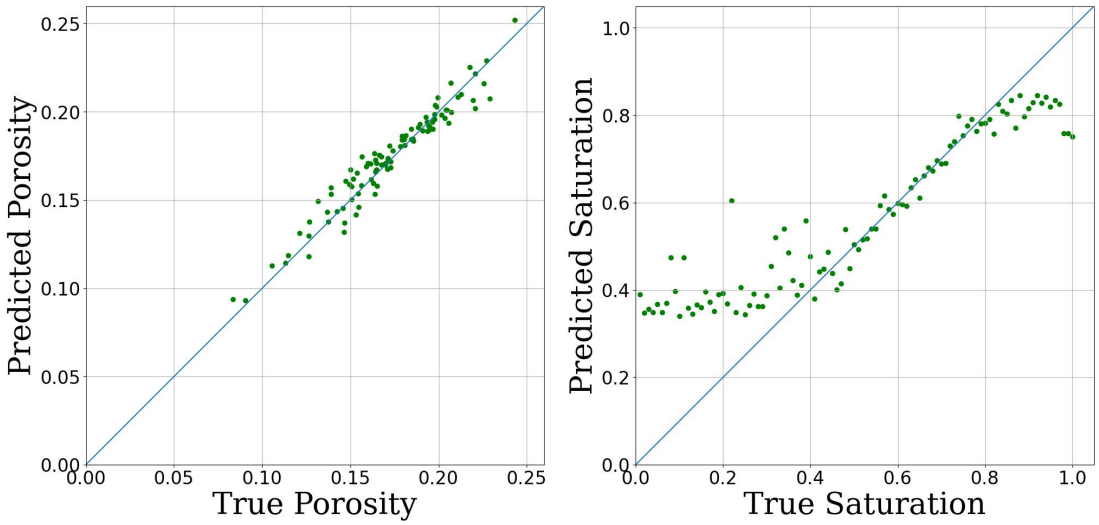


Figure 5.5: True vs. predicted porosity and saturation using the PGNN workflow with a random seed of D_F on the test dataset D_T at $\lambda = 0.9$, showing higher RMSE for saturation at lower saturation levels.

5.6.2 Numerical Example Discussion

The general accuracy of predictions improves when using PGNN training, especially at higher λ values in prediction both porosity and saturation compared to training purely on uncalibrated rock physics modeled data (see Section 4.5.3). Notably, there is a greater enhancement in predicting porosity compared to saturation. This difference can be attributed to the observation in Section 4.6, where it was shown that porosity exhibits a linear relationship with velocity, making it easier for the neural network to learn.

As λ increases, there is a noticeable variability in predictive performance as demonstrated by the high standard deviation. During PGNN training, the weights are initialized to the same values, all the networks were trained alongside the same model dataset D_M and the convergence criteria was kept the same for all seeds of D_F. Therefore, there could be two main sources of this variability. First, are the sampled instances of D_F. Second is the noise level within the feature extraction window.

To evaluate the impact of the field dataset D_F on neural network training, I conducted two separate training sessions using different D_F datasets, labeled 'a' and 'b', each characterized by unique porosity and saturation combinations. Both networks were initialized in the same way and trained using the PGNN workflow, then evaluated on a test dataset D_T. Figure 5.6 displays the PGNN prediction results for porosity and saturation from these distinct D_F datasets. The results highlighted notable differences in predictive accuracy. Specifically, with λ set at 0.9, the network trained with dataset 'a' achieved a 16% decrease in porosity prediction RMSE and a 39% decrease in saturation prediction RMSE. Meanwhile, the network trained with dataset 'b' showed a 29% decrease in porosity RMSE and 37% increase in saturation RMSE predictions compared to training the network on pure modeled uncalibrated dataset.

Figure 5.7 showcases a side-by-side comparison of the porosity and saturation distributions within these datasets. The spread of porosity values was consistent in both, but notable variations appeared in the saturation distributions. Seed 'b' tended towards higher saturations, whereas seed 'a' presented a more balanced range, including a significant portion of lower saturation values (Sw<0.3). These lower values are more challenging to predict due to a lower dispersion gradient at lower frequencies, as seen in Figure 4.35. This uneven distribution in seed 'b', I believe, contributed to its less accurate saturation predictions, particularly for Sw<=0.5, a trend clearly depicted in Figure 5.5.

These findings highlight that effective PGNN training depends not just on the quantity of data instances in D_F, but also on having a representative distribution of reservoir properties.

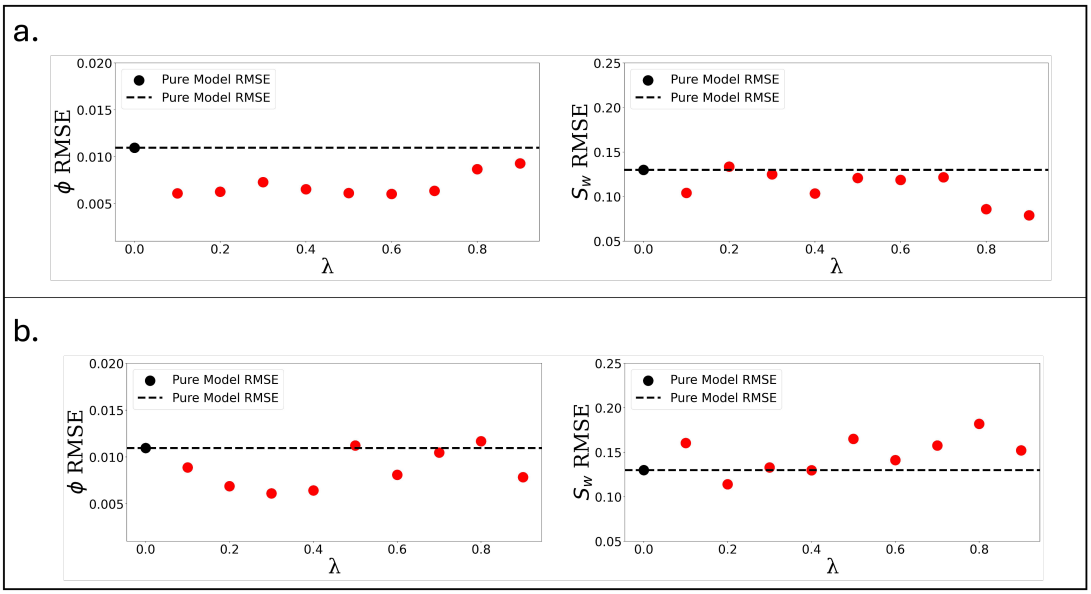


Figure 5.6: PGNN predictions on test dataset D_T with two different datasets D_F. **a**. shows an improvement of porosity and saturation RMSE on D_T across different λ values. **b**. shows less enhancement of porosity and saturation prediction RMSE on D_T across different λ values.

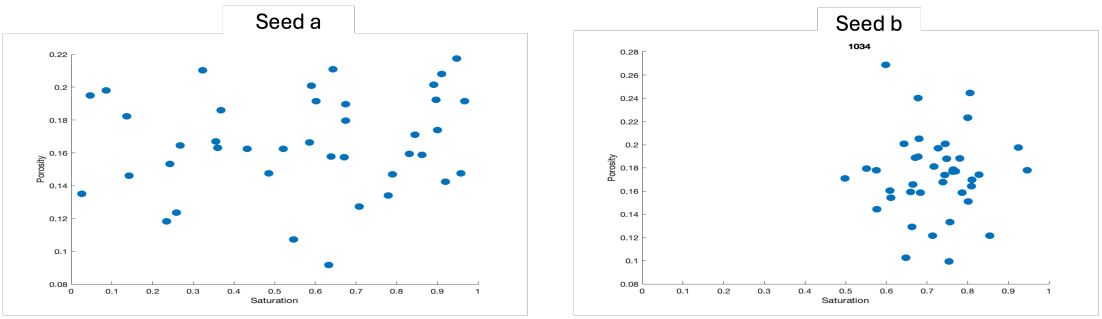


Figure 5.7: Comparative scatter of porosity and saturation in two field seeds. Seed (**a**) depicts a uniformly distributed porosity and saturation. In contrast, seed (**b**) exhibits a skewed distribution, predominantly featuring higher saturation values.

The noise level within all seismic gathers is determined by the signal-to-noise ratio, as discussed in Section 4.2. However, the features were extracted from a window that does not cover the entire trace. Given the uniformity of D_M across all PGNN training, the primary source of noise impact arises from features within D_F. Consequently, certain features may experience higher levels of noise than others.

To evaluate the noise impact, I generated 10 seeds of D_F with identical modeling parameters but different random noise added to the gathers. Subsequently, I applied the same PGNN training process to all seeds and plotted their average RMSE predictions for porosity and saturation across various λ values. The results, depicted in Figure 5.8, show the average prediction for the network on D_T.

Even with identical model parameters, varying noise levels introduce noticeable differences in the PGNN training and prediction outcomes. For λ values up to 0.6, the standard deviation in both porosity and saturation predictions remains relatively stable. However, a significant increase in standard deviation is observed when λ exceeds 0.6. This indicates that the training process is sensitive to noise in the extracted features, leading to less accurate predictions at higher λ values. Despite this, networks trained with the PGNN workflow still show improved accuracy in predicting porosity and saturation compared to those trained exclusively on a purely modeled dataset. The following section will explore additional factors contributing to this increased standard deviation, especially when the training emphasizes the field dataset D_F.

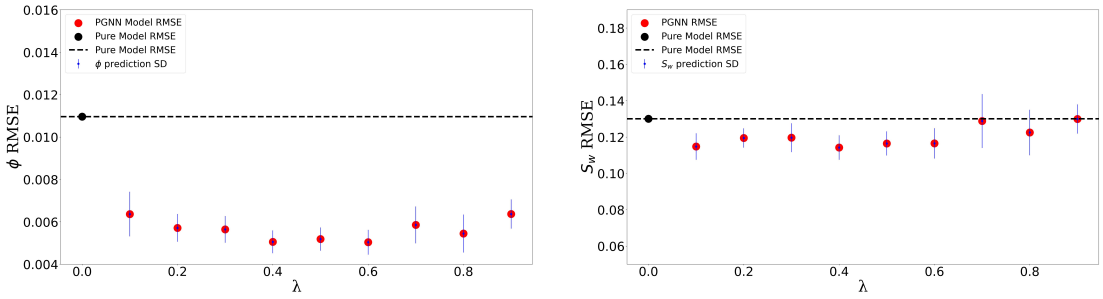


Figure 5.8: Averaged RMSE in network prediction of porosity and saturation over D_T when trained with 10 seeds of D_F that share the same model parameters, but differ in random noise within the gather.

An additional factor potentially amplifying prediction uncertainty at higher λ values is insufficient training of the network weights. This phenomenon is primarily attributable to the convergence criteria discussed in Section 5.3. The convergence

process is driven by the cross-validation of D_F instances. Consequently, assigning a high weight to D_F accelerates the adjustment of network weights towards minimizing the loss on those instances and those in the validation dataset, ultimately leading to premature termination of training. This results in an undertrained network that exhibits poor generalization on the unseen dataset D_T.

Identifying the ideal λ value is a significant challenge in implementing PGNN training. Figure 5.2 shows that both porosity and saturation predictions generally improve across various λ values when using PGNN training. This analysis depends on having a test dataset D_T with the same τ values as D_F, but not used in training. However, such a dataset is often not available in real-world scenarios. Additionally, the figure indicates that at very high λ values, certain seeds produce porosity and saturation predictions that are as accurate or even superior to those obtained with lower-range λ values. On the other hand, other seeds result in predictions that are comparable in RMSE values to the network trained solely on uncalibrated rock physics model data that were discussed earlier in Section 4.5.3.

Two key factors influence the optimal selection of λ in PGNN training. First, the composition and diversity of dataset D_F significantly affect training. As illustrated in Figure 5.6, a diverse range of reservoir measurements allows for greater emphasis on D_F during training. Seed 'a' shows improved prediction accuracy across a broader λ range compared to seed 'b', which yields poorer predictions than training solely with modeled across different values of λ. This discrepancy is likely due to the limited variety of theoretical relationships in D_F for predicting porosity or saturation, leading to overfitting and failure to generalize to unseen data in D_T, as shown by the high error rates for low saturation values in Figure 5.5.

The second factor is the noise level within D_F, which can significantly impact the optimal λ range. When predictions rely on patterns from a limited number

of instances, noise interference can cause the network to misinterpret noise as critical information, leading to overfitting and poor generalization, especially at higher λ values, as depicted in Figure 5.8. This demonstrates the importance of carefully balancing λ to avoid overfitting while still capturing meaningful patterns from the data.

5.7 Further Analysis

The numerical example, discussed in Section 5.6, demonstrated the potential predictive power of PGNN training and discussed some issues that could enhance or deteriorate that prediction accuracy. In this section, I would like to address other issues that were not discussed in the numerical example, but could impact the PGNN training.

5.7.1 Constraining the unknown modeling parameter τ range within D_M

During the numerical example in Section 5.6, the range of the unknown parameter τ was sampled from a uniform distribution that spans $log(\tau) \in (-4, 2)$. However, it is expected that as more information becomes available, this range will be constrained. In this section, I will test the impact of such a constrain on the predictive power of PGNN.

I create a modeled dataset D_{M_c} with constrained τ range that is sampled from a uniform distribution that closely resemble the seismic frequency range $\log(\tau) \in (-2, 0)$. All other model parameters will be sampled from the same distributions used for D_M. I use the exact same D_F seeds and network initialization that were used in the numerical example to train the constrained model dataset D_{M_c} and average the results.

Figure 5.9 shows the prediction results averaged over 25 seeds of D_F along with the standard deviation. In general, the prediction RMSE of porosity remains

consistent up to $\lambda = 0.6$ and gradually increases thereafter. Saturation, however, shows a lower RMSE with increasing λ values. As for standard deviations, lower values are observed when training on D_{M_c} compared to training on the dataset D_M, which has a wider range of τ values (see Section 5.6, Figure 5.2).

Figure 5.10 presents the results of an analysis comparing training with a wide range of τ values in D_M to training with a narrower range in D_{M_c}. Comparing the predictions from the two purely model-based datasets, D_M and D_{M_c}, the latter shows significantly lower RMSE for porosity and less improvement in RMSE for saturation. When comparing PGNN predictions between the two datasets, D_M and D_{M_c}, there is minimal improvement in saturation predictions, while porosity predictions remain almost identical. This suggests that the information extracted from D_F is consistent, regardless of the range of the unknown model parameter in the uncalibrated modeled dataset. The main impact of narrowing the range of the unknown parameter—τ, in this case—is a reduction in prediction RMSE for porosity and saturation when training on the pure, uncalibrated modeled dataset compared to training on a wider range of unknown τ values with a standard neural network workflow (see Section 4.5.3)

In Section 4.5.3 (Figure 4.27), porosity predictions showed less sensitivity to dispersion, with only minor shifts in frequency as porosity increases or decreases. In contrast, saturation curves are more frequency-sensitive. This suggests that the significant reduction in RMSE for porosity with constrained τ is due to the reduced importance of precise τ accuracy; being in the general range is likely sufficient to calibrate the network effectively. For saturation, however, only a slight RMSE reduction is observed with constrained τ, indicating that achieving significantly lower RMSE for saturation may require further narrowing of the τ range.

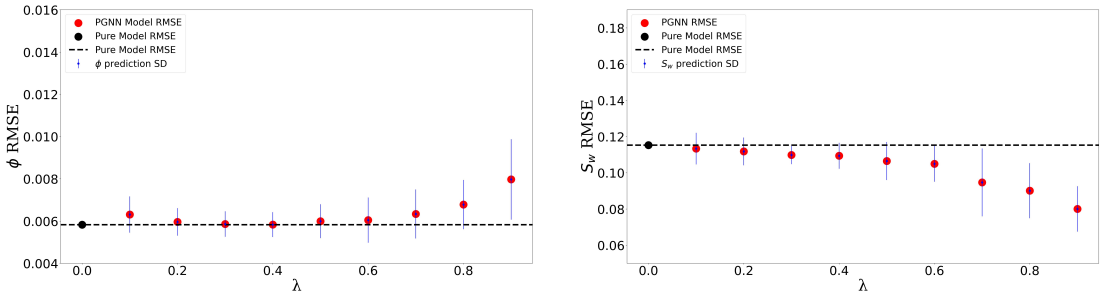


Figure 5.9: Average RMSE in network predictions for porosity and saturation over D_T, when trained using a modeled dataset with a constrained range of τ values D_{M_C}.

In Figure 5.10, training with the PGNN workflow shows a semi-consistent RMSE for porosity predictions up to $\lambda = 0.6$. Since the pure model was likely calibrated for porosity prediction by constraining τ, the calibration influence of the field dataset D_F was minimal within this range of λ values. However, as λ increases, it appears the network begins overfitting to field instances in D_F, possibly due to premature termination of the training process, noise within the D_F dataset or difference in dynamic range between porosity and saturation (see Sections 5.6.2 and 5.8). Saturation predictions, on the other hand, show a continuous decrease in RMSE with increasing λ values, likely due to the importance of precisely calibrating τ from the field dataset using the PGNN workflow. Additionally, Figure 5.9 shows higher standard deviations for $\lambda > 0.6$, potentially due to noise and premature data training (see Section 5.6.2).

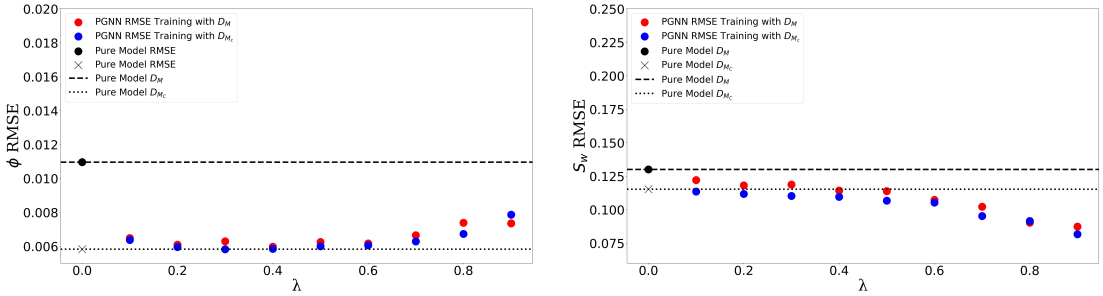


Figure 5.10: Comparison of average RMSE in network predictions for porosity and saturation over D_T, when trained on a modeled dataset with a wide range of τ values (D_M) versus a modeled dataset with a constrained range of τ values (D_{M_C}).

5.7.2 Field dataset size D_F

In this Section, I replicate the numerical example but vary the number of instances within D_F from 40 to 60. Additionally, I vary the number of instances from 40 to 20 to assess the impact of lowering the number of instances on PGNN training.

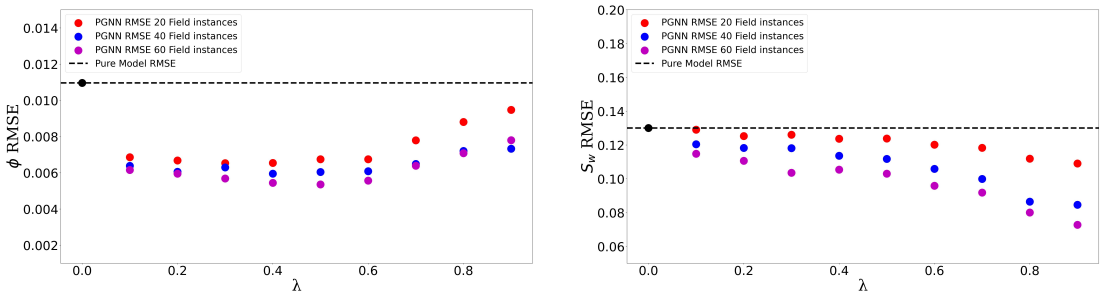


Figure 5.11: Comparison of Averaged RMSE in network predictions for porosity and saturation over D_T between networks trained with D_F comprising 20, 40 and 60 instances.

Figure 5.11 shows that increasing the size of the field dataset, D_F, consistently improves PGNN training results across various λ settings. Specifically, when the number of field data instances increases from 40 to 60, the RMSE for saturation prediction at $\lambda = 0.9$ decreases from 33% to 44%, compared to the RMSE when training on the purely uncalibrated τ-modeled dataset (see Section 4.5.3). For porosity predictions, increasing D_F to 40 results in a similar RMSE at $\lambda = 0.9$. However, the minimal RMSE for porosity is observed at $\lambda = 0.5$, with a 51% reduction compared to the pure modeled dataset.

Reducing D_F from 40 to 20 instances still outperforms training solely with modeled data but with diminished benefits. However, the PGNN's decrease in RMSE prediction for saturation and porosity drops to 15% and 14%, respectively, compared to the 35% and 33% improvements with 40 field points at $\lambda = 0.9$.

This analysis yields two principal insights. Firstly, increasing the size of the field dataset D_F generally results in an enhancement in saturation prediction accuracy.

Secondly, the optimal λ value for minimizing the RMSE in porosity predictions gradually shifts towards giving more weight to D_F as its size grows. Specifically, for D_F sizes of 20, 40, and 60, the ideal λ values are 0.3, 0.4, and 0.5, respectively. This is demonstrated in Figure 5.11. Figure 5.12 illustrates saturation predictions for a single seed, matching the average RMSE for 25 seeds, at $\lambda = 0.9$ for each corresponding D_F size.

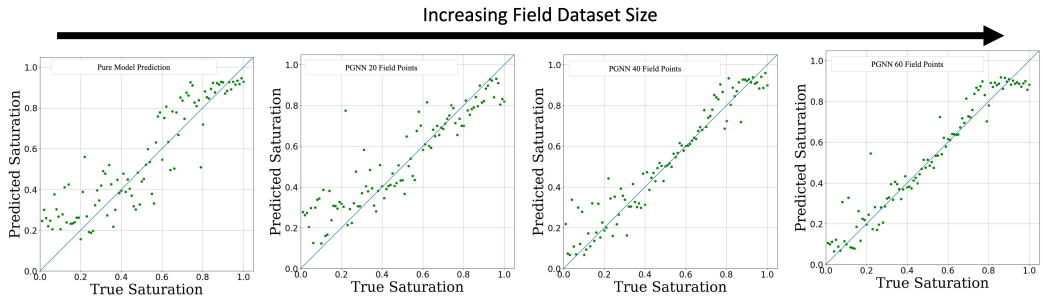


Figure 5.12: Comparative analysis of saturation predictions on a pure uncalibrated modeled dataset with those trained using PGNN incorporating field datasets of 20, 40, and 60 instances.

Furthermore, I analyze the standard deviation in PGNN training when predicting with D_F that contains 60 instances across 25 seeds. For porosity and saturation predictions, there is a semi-consisten decrease in standard deviation across different λ values as oppose to training with D_F that contained only 40 instances in D_F. This observation suggests that the high standard deviation observed in Figure 5.2 is mainly driven by the lack of training instances to adequately train the network weights when relying heavily on D_F, leading to an under-trained network.

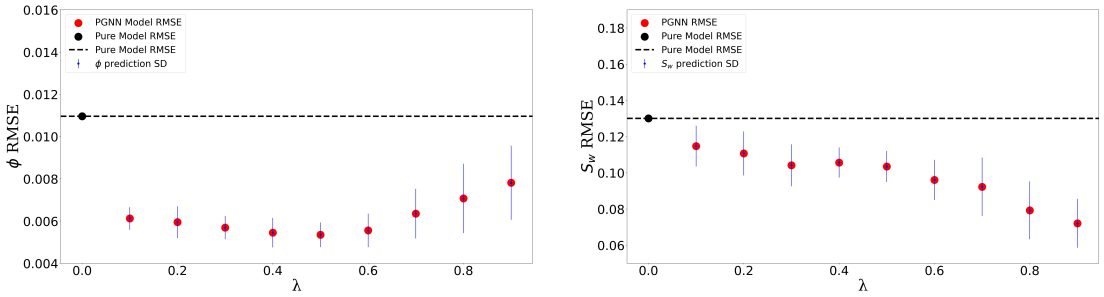


Figure 5.13: Average RMSE of network predictions for porosity and saturation over D_T, when trained with a field dataset of 60 instances.

5.7.3 Impact of PGNN training Using Cross-Validation of D_F

In this section, I explore the effectiveness of the PGNN cross-validation method for ensuring convergence. Following the setup described in Section 5.6, I employ the test dataset D_T as a validation dataset. Training is halted if the loss, calculated using equation 3.5, on D_T remains consistent for 50 epochs or begins to increase.

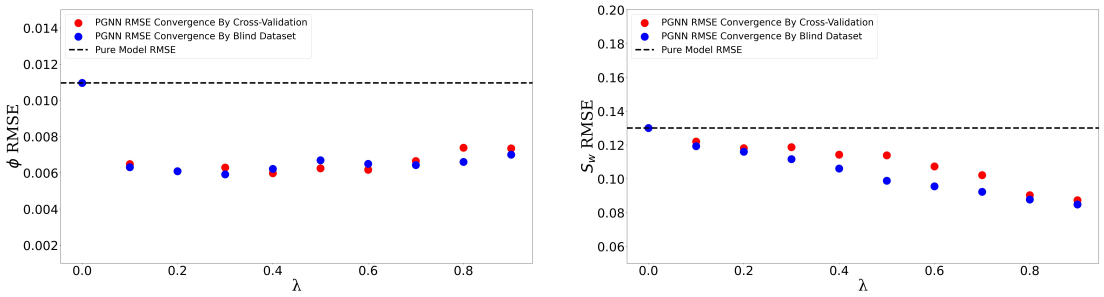


Figure 5.14: A comparative analysis of PGNN training shows lower RMSE predictions for saturation when convergence is assessed using a validation dataset rather than the cross-validation method.

Figure 5.14 presents averaged RMSE predictions for porosity and saturation using two different convergence methods. At $\lambda <= 0.2$ values, the difference between the methods is minimal, reflecting the limited influence of D_F on the network training loss. However, as λ increases, a divergence emerges between the outcomes of using cross-validation with D_F and employing a separate validation dataset in saturation RMSE. Specifically, the saturation RMSE is lower when using a

dedicated validation dataset compared to cross-validation. Porosity, on the other hand, shows similar RMSE for both convergence methods, likely due to its linear relationship, which facilitates convergence, or the reduced need for precise τ values for network calibration (see Sections 4.5 and 5.7.1).

This analysis supports that while cross-validation can provide reasonable results, employing a dedicated validation dataset tends to yield more accurate predictions, especially at higher λ values.

5.7.4 Impact of Transferred learning training on PGNN

Transfer learning is a technique in machine learning, involving the adaptation or fine-tuning of a model originally trained for one task to make it suitable for another related task (Torrey and Shavlik, 2010). In a study conducted by Wu et al. (2020), the authors employed transfer learning to predict reservoir properties using seismic data. In their research, they initially trained a neural network on synthetic SEG Marmousi2 seismic data, which served as the input features, and the target labels were the impedance log that represents each trace. This initial training aimed to teach the model to accurately predict acoustic impedance based on seismic traces. After successful training, the model's weights were saved.

The next step involved obtaining a limited set of instances from an over thrust model. These instances were used to fine-tune and calibrate the pre-trained model to make accurate impedance predictions in an over-thrust geological setting. The key to this process was initializing the model's weights using the saved weights from the previous training on synthetic data. The results demonstrated that this transfer learning approach achieved similar prediction accuracy to a neural network trained using seismic data and impedance logs from the over-thrust fault model.

Applying a similar methodology, I employ transfer learning in the context of my

PGNN training. Firstly, I train the neural network model on the pure modeled dataset D_M. Next, I save the final weights once the training is complete. Subsequently, I initialize the weights of my neural network that I will be using for PGNN training from those saved weights. After that, I continue the training cycle of PGNN as showen in Figure 5.1 and compare the prediction of the network to that that initialize the weights using Glorot and Bengio (2010) method.

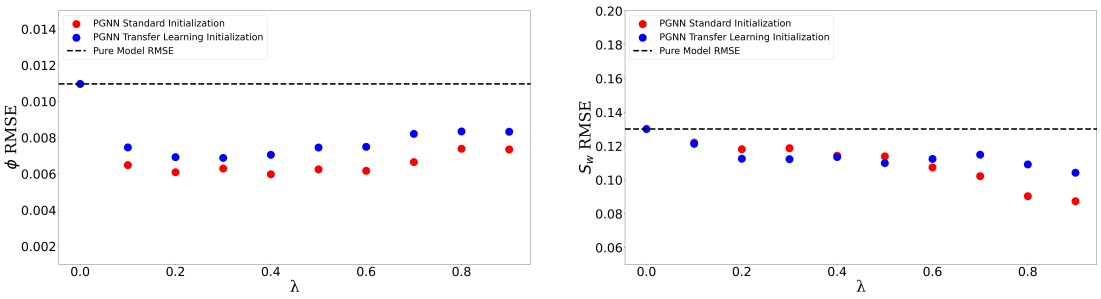


Figure 5.15: Comparison of the averaged RMSE in network predictions for porosity and saturation over dataset D_T between networks trained with transfer learning and standard initialization.

Figure 5.15 presents the prediction results for porosity and saturation on dataset D_T when the transfer learning technique was applied. The outcomes revealed that standard initializing of the network using the method proposed by Glorot and Bengio (2010) led to slightly better predictions for porosity and saturation in general.

Figure 5.16 presents the standard deviation in RMSE from PGNN training using transfer learning. The data reveals a consistent standard deviation across both lower and higher λ values for porosity and saturation predictions. This consistency suggests that using weights from a pre-trained model is beneficial for training the network, particularly in establishing relationships between seismic features and reservoir properties. Additionally, this observation supports the idea that the high standard deviation at elevated λ values, as noted in Figure 5.2, is likely due to under-training of the network weights.

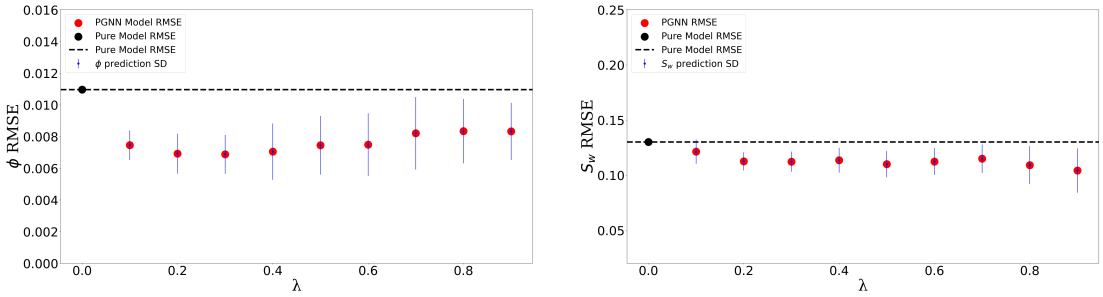


Figure 5.16: Average RMSE of network predictions for porosity and saturation over D_T, when trained using transfer learning initialization.

5.7.5 Batch size of D_M

The PGNN loss function undergoes updates based on a batch comprising Q instances from D_M and Z instances from D_F, as detailed in Section 5.3. Given the limited size of Z for D_F, all instances are utilized during the final training cycle of PGNN. However, the size of Q can vary from 1 to 64, as explained in Section 4.4, Figure 4.10. In this Section, I explore the impact of the model data batch size D_M on the predictive efficacy of PGNN training. I replicate a numerical example with a batch size of $Q = 32$ and conduct additional trials with batch sizes of 8 and 64, comparing their predictions to those obtained with the original batch size of 32.

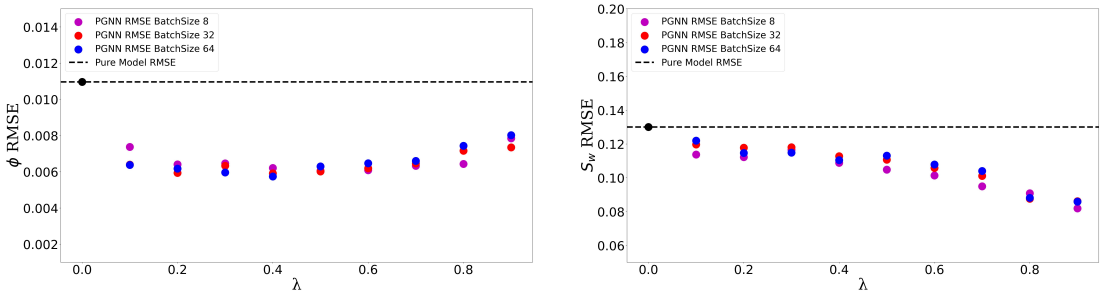


Figure 5.17: Comparative analysis of PGNN predictions for porosity and saturation on the test dataset D_T with different modeled dataset batch size during training.

Figure 5.17 illustrates the impact of varying batch sizes on PGNN training. For porosity and saturation, there is a minimum impact of varying batch size on the network prediction RMSE.

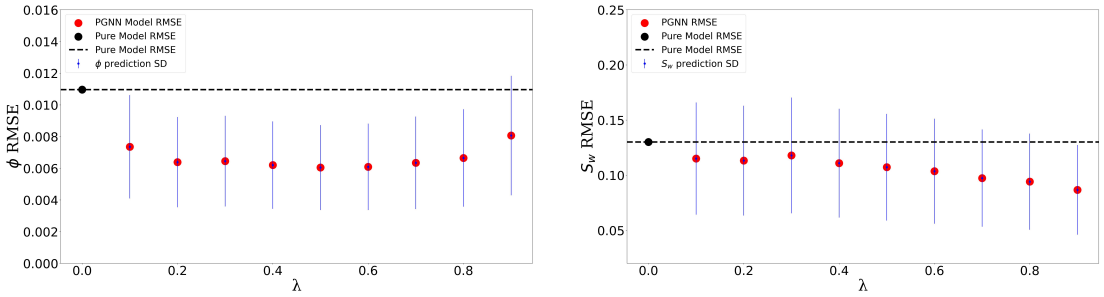


Figure 5.18: Average RMSE of network predictions for porosity and saturation over D_T, when trained with D_M using a batch size of 8.

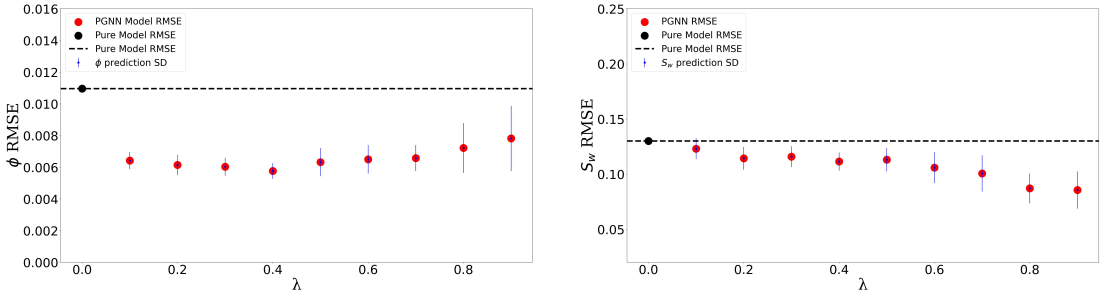


Figure 5.19: Average RMSE of network predictions for porosity and saturation over D_T, when trained with D_M using a batch size of 64.

Figures 5.18 and 5.19 illustrate the standard deviation in RMSE from PGNN training over 25 seeds of D_F, with batch sizes of 8 and 64, respectively. A key observation is that reducing the batch size significantly increased the standard deviation across all λ values. While it is often reported in machine learning literature that larger batch sizes can degrade network performance (Araya-Polo et al., 2018), my study found that smaller batch sizes introduced greater uncertainty during training, and that using larger batch size resulted in more stable predictions within the PGNN workflow.

5.7.6 Processing of D_F

In real-world field scenarios, the processing of seismic data significantly affects the quality of extracted features from processed gathers. This impact can include the removal of valid data signals, changes in frequency content, alterations in AVO signatures, or insufficient removal of random noise. Fully evaluating the effects of

diverse seismic processing methods on the efficacy of PGNN is a complex challenge beyond the scope of this thesis. However, I provide an example of how seismic processing, specifically changes in SNR, can influence PGNN's performance using dataset D_F.

Previous analysis involved creating all seismic gathers with a 25 dB SNR. In this Section, I duplicated the setup, maintaining identical modeling parameters for both D_F and D_M, but with a key variation: the SNR of D_F was intentionally lowered to 20 dB, while keeping D_M at a constant 25 dB SNR. This change was implemented to specifically explore how variations in SNR during seismic processing affect the accuracy of the PGNN training.

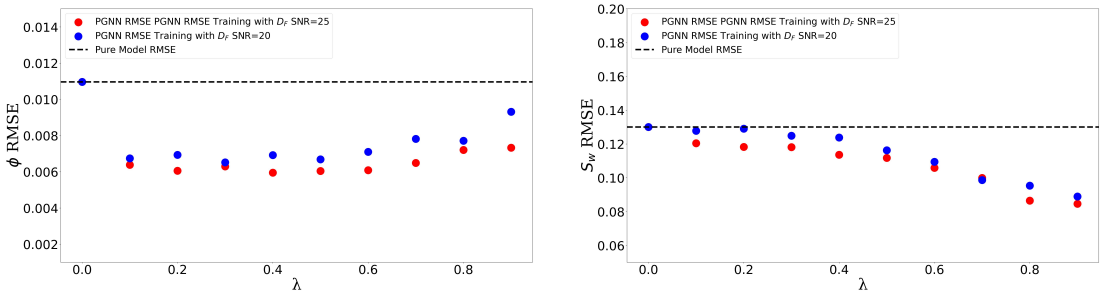


Figure 5.20: PGNN average RMSE in saturation and porosity predictions on D_T with a 25 and 20 SNR. Increased level of noise in D_F slightly deteriorated PGNN accuracy.

Figure 5.20 displays the average RMSE in porosity and saturation predictions on D_T following PGNN training with altered SNR. The results show a decline in prediction accuracy as the noise level in the field dataset D_F increases. This finding underscores the significant impact that even minor changes in SNR during seismic data processing can have on PGNN's effectiveness.

Furthermore, Figure 5.21 illustrates that the standard deviation slightly decreased in higher λ values compared to the 25 dB SNR scenario shown in Figure 5.2. This could be attributed to the fact that noise addition could be used as a regularization method to minimize overfit because it forces the network to learn underlying

patterns rather than memorizing specific details (Bishop, 1995).

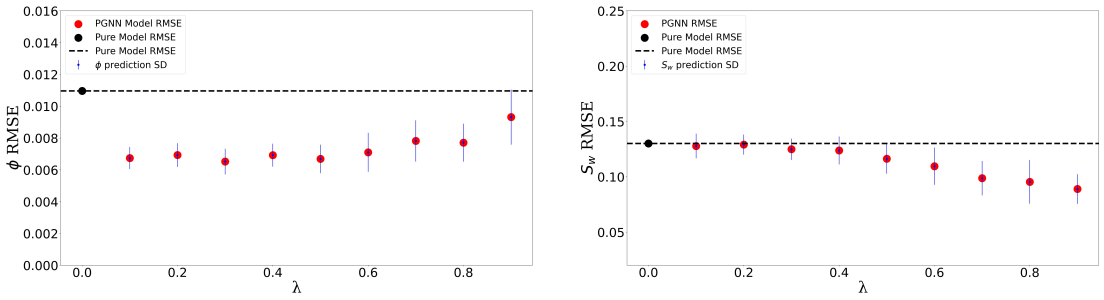


Figure 5.21: Average RMSE of network predictions for porosity and saturation over D_T, when trained with D_F at 20 dB SNR.

5.7.7 Wrong Model Assumption of D_M

Ideally, the selection of a rock physics model should align closely with the characteristics of the field reservoir. However, due to the limited information available in the early stages of exploration, there is a risk of adopting inaccurate rock physics assumptions for the targeted reservoir. For instance, one might assume an elastic behavior of the reservoir that does not align with the dispersive behavior that occurs in reality.

In this section, I investigate the potential implications of such assumptions by generating an elastic model dataset, denoted as D_{M_E}. This dataset is created by setting $\tau = 1.0 \times 10^{-5}$, resulting in an elastic response as illustrated in Section 4.5.1, Figure 4.15. All other model parameters are sampled from the same distribution as D_M. Utilizing the same seeds of D_F employed in the numerical example, which exhibit dispersion within the seismic frequency range, I train the model along D_{M_E} using PGNN training and analyze the results.

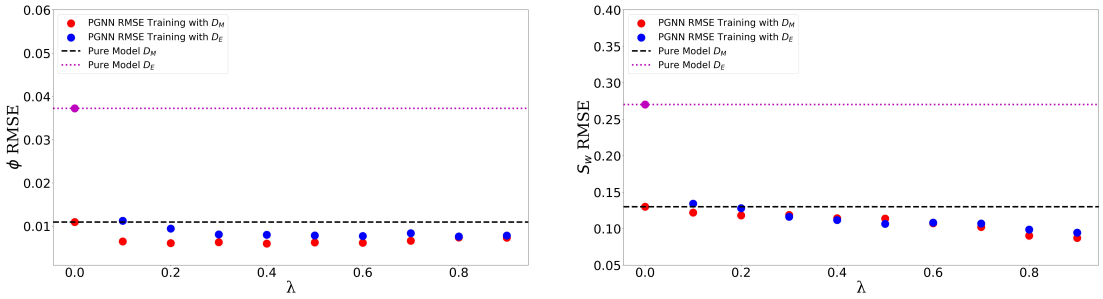


Figure 5.22: Comparative analysis of PGNN predictions for porosity and saturation on the test dataset D_T using two independent networks—one for porosity and one for saturation—versus a single network predicting both properties.

Figure 5.22 presents a comparative analysis of the PGNN workflow applied to two distinct datasets: D_{M_E}, a new rock physics model, and the existing D_M. Upon introducing D_{M_E}, I established a baseline at $\lambda = 0.0$, focusing solely on modeled data. The subsequent analysis revealed notable improvements in both porosity and saturation predictions when incorporating dispersive field data D_F into the PGNN framework.

From Figure 5.22, the RMSE for porosity and saturation significantly drops when using the PGNN workflow, even when training on purely elastic modeled dataset D_E. Figure 5.23 provides a visual representation of this comparison at $\lambda = 0.9$, illustrating the difference between standard network training and the enhanced approach offered by PGNN training when predicting on the test dataset D_T.

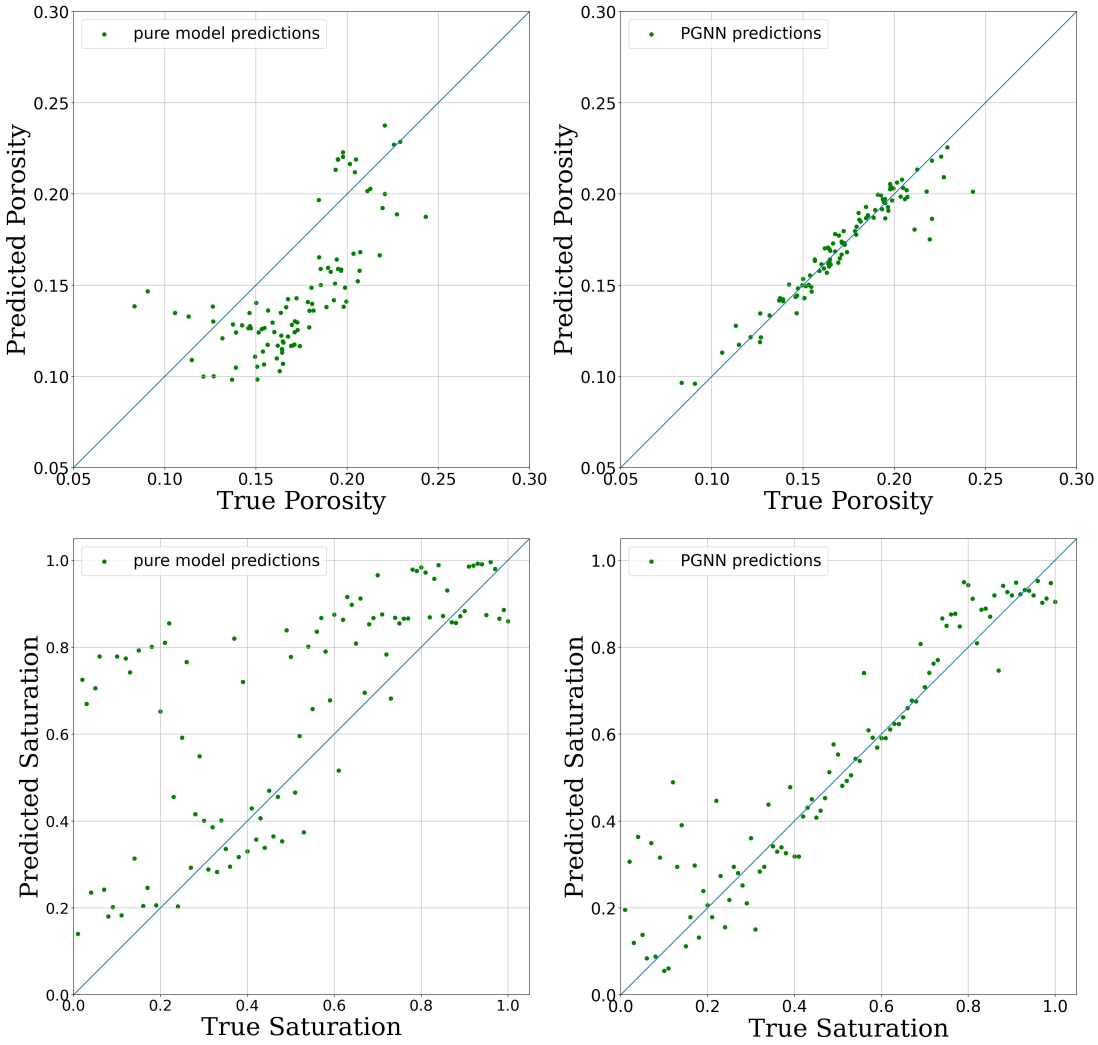


Figure 5.23: True vs. predicted porosity and saturation using the PGNN workflow trained on a purely elastic, uncalibrated rock physics model alongside dispersive field data D_F, evaluated on the test dataset D_T at λ=0.9.

Figure 5.24 highlights a higher standard deviation in porosity and saturation predictions when using the less suitable D_{M_E} model, indicating that incorrect model assumptions are likely to yield greater prediction uncertainty. This is likely due to fewer common factors between the uncalibrated rock physics model dataset D_{M_c} and the field dataset D_F during training.

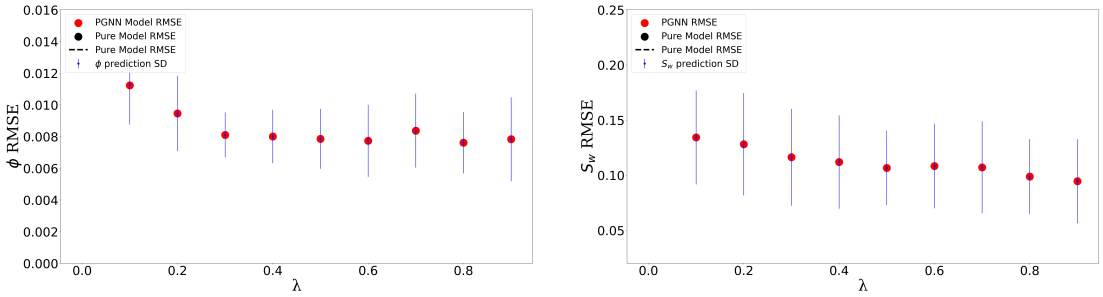


Figure 5.24: Average RMSE of network predictions for porosity and saturation over D_T, when trained with purely elastic modeled dataset D_E.

A crucial insight from this analysis is the ability of the field dataset D_F to compensate for inaccuracies in the modeled dataset D_{M_E}. By training the network with field data features and labels, significant reductions in prediction error are achieved compared to training solely with an inaccurate rock physics model.

5.8 Discussion

In the numerical example discussed in Section 5.6, and in much of the subsequent analysis (Section 5.7), there is a difference in the optimal λ values for predicting porosity versus saturation. For instance, Figure 5.2 shows that the optimal λ value yielding the lowest RMSE for porosity is 0.4, whereas for saturation, the optimal value is 0.9. Additionally, the RMSE for porosity gradually increases beyond $\lambda = 0.6$, while for saturation, it decreases over the same range. This discrepancy can create confusion when selecting an optimal λ value, as no single value minimizes RMSE for both saturation and porosity. Given that the dynamic range for porosity is 0.01 to 0.40 and for saturation is 0.01 to 1.0 (see Figure 4.3), it is likely that this difference in dynamic range impacts prediction accuracy, with saturation errors exerting a greater influence on weight updates than porosity errors due to saturation's larger dynamic range.

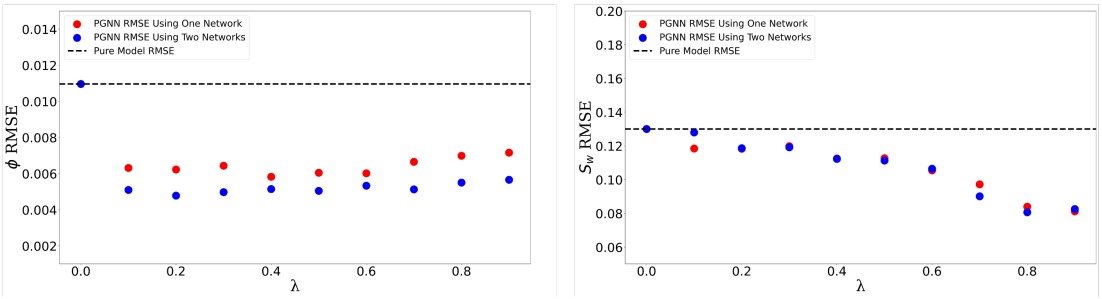


Figure 5.25: Comparative analysis of PGNN predictions for porosity and saturation on the test dataset D_T using two independent networks—one for porosity and one for saturation—versus a single network predicting both properties.

To test this hypothesis, I repeated the numerical example (Section 5.6) using separate networks dedicated to each reservoir property, predicting them individually to eliminate any potential training interference. Figure 5.25 shows the prediction results for porosity and saturation using two independent networks on the same test dataset D_T and field dataset D_F. Porosity predictions exhibit a consistently lower RMSE across all λ values and do not increase significantly at higher λ values compared to using a single network. Saturation, on the other hand, shows no major difference in prediction accuracy between the two approaches. This supports the hypothesis that using a single network to predict both properties may have compromised porosity predictions, especially at higher λ values, to improve saturation predictions. Therefore, I recommend normalizing the labels or using separate networks for different reservoir properties when their dynamic ranges differ.

The accuracy of the PGNN workflow is influenced by several factors, including the number of field measurements, their distribution, and data noise. The workflow achieved its lowest saturation RMSE of 8.0×10^{-2} using a PGNN with 60 field instances at $\lambda = 0.9$, as shown in Figure 5.11. In contrast, a neural network trained on instances from a calibrated rock physics model exhibited a lower RMSE of about 0.27×10^{-1}, as detailed in Figure 4.19. The discrepancy in accuracy stems from the loss function employed in the PGNN, referenced in 5.8. Modeled dataset

included instances that do are misaligned with the actual reservoir properties of field data. As the network aims to minimize loss across all instances, including those misaligned with actual field data, the network weights are not exclusively optimized for minimizing loss from field measurements. This results in some inherent errors when predicting field reservoir properties. To reduce this error, it is essential to increase the amount of field data for more precise network training and to narrow the range of unknown parameters to decrease the loss from modeled data, thereby enhancing overall accuracy.

When large field data are available for full rock physics model calibration, using a standard neural network training to train neural network, as shown in Section 4.5.2, is recommended for even more precise predictions.

This chapter, while addressing modeling aspects like unknown parameters and rock physics model assumptions, did not explore the impact of seismic modeling workflows on replicating field data gathers. These workflows, crucial in extracting features from D_F and D_M, vary greatly in complexity. Additionally, the potential effect of varying dynamic ranges in D_F gathers, due to diverse seismic data acquisition and processing, was not investigated. A practical approach would involve normalizing extracted features before PGNN application, crucial for unbiased network training.

6 Chapter 6: Physics-Guided Neural Networks Application on VTI Layer with Unknown Thomsen Parameters

6.1 Chapter Objective

My thesis focuses on using rock physics models to predict reservoir properties in complex formations where key calibration parameters are unknown. Building on the previous chapter's discussion of a dispersive rock physics model and the critical parameter τ, this chapter shifts focus to applying this method to reservoirs with vertical transverse isotropy (VTI) resulting from shale layering. This shift highlights challenges in calibrating models using Thomsen parameters, often unavailable in the early stages of exploration, necessary for the accurate inversion of reservoir properties.

6.2 Seismic Modeling of of VTI Anisotropy

In Section 2.5, I discussed the importance of considering shale anisotropy in amplitude versus offset (AVO) analysis to prevent the misinterpretation of seismic anomalies attributed to this anisotropy. I now create a synthetic geological model to quantify the impact of an anisotropic VTI shale layer on the predicted properties of a sandstone layer situated beneath it using neural networks.

I apply the same geological model discussed in Section 4.2, but assume that the top shale layer has a preferred alignment of clay minerals, inducing VTI anisotropy above a sandstone reservoir. I then calculate the elastic properties $K(\omega)$ and $\mu(\omega)$, using Equations A.1 and A.2 from Papageorgiou and Chapman (2017) model for each layer. Setting $\tau = 1.0 \times 10^{-5}$ simulates an elastic Gassmann response at seismic frequencies, as outlined in Section 4.5.1. All elastic properties are calculated at 25 Hz, which is the dominant frequency of the source Ricker wavelet. Reservoir properties are derived from distributions in Figure 4.3 to define

each layer's elastic properties.

I first compute the bulk density (ρ) of each layer using equation 2.8. Subsequently, P-wave and S-wave velocities are determined using equations 2.6 and 2.7, respectively. The isotropic elastic tensor components C_{33} and C_{55} are then calculated by soloving equation 2.16 and 2.17 respectively.

In Chapter 4, I demonstrated in Section 4.5.1 that utilizing elastic gathers to predict gas saturation is challenging due to the complex non-linear relationship between saturation and velocity, known as the fizz-gas challenge (see also the discussion in Section 4.6). I also showed that this challenge can be mitigated if the network is trained on dispersive gathers with a fixed and known τ value (see Section 4.5.2). However, my objective in this chapter is to incorporate anisotropy involving Thomsen parameters without the added complexity of attenuation or τ. Therefore, I intend to simplify the velocity-saturation relationship to distinguish between high and low gas saturation based solely on elastic seismic amplitudes by employing the patchy saturation mixing law. In the model by Papageorgiou and Chapman (2017), patchy saturation is regulated by the hyperparameter q, as discussed in Section 2.4 . Setting $q = 1$ results in uniform mixing (Wood and Lindsay, 1956), which is complex and extremely non-linear. However, changing the value of q simplifies the relationship between velocity and saturation due to patchy saturation, as shown in Figure 6.1. I set $q = 1$ for shale, since it is highly saturated ($> 90\%$), which should be distinguishable even with uniform fluid distribution (see Section 4.5.1), and set $q = 0.2$ for the sandstone layer to facilitate a more straightforward relationship between velocity and saturation.

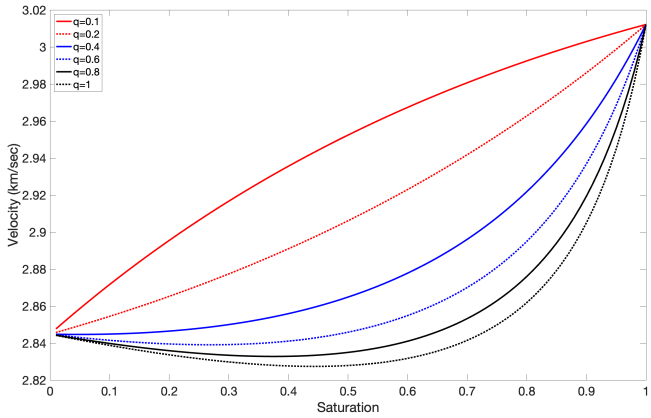


Figure 6.1: Comparison of Velocity-Saturation relationships: From uniform distribution (with $q = 1$) to patchy saturation (q<1). Sandstone reservoir is set at $q = 0.2$.

I calculate the VTI anisotropic tensor for the shale layer. P-SV propagation in VTI anisotropy is characterized by four elastic tensors: C_{11}, C_{33}, C_{44}, and C_{13} (Tsvankin, 1996). I randomly sampled values of ϵ and δ and solve for C_{11} and C_{13} in equations 2.10 and 2.12 respectively.

With these tensors, I use Jin et al. (2017)'s formulations of Schoenberg and Protazio (1990) paper to calculate the reflection coefficient between the two layers $R(\theta)$. I then convolve a Ricker wavelet (equation 4.2) with $R(\theta)$ to generate seismic gathers. Noise is set at 25 dB SNR and added to the gather as described in 4.2.

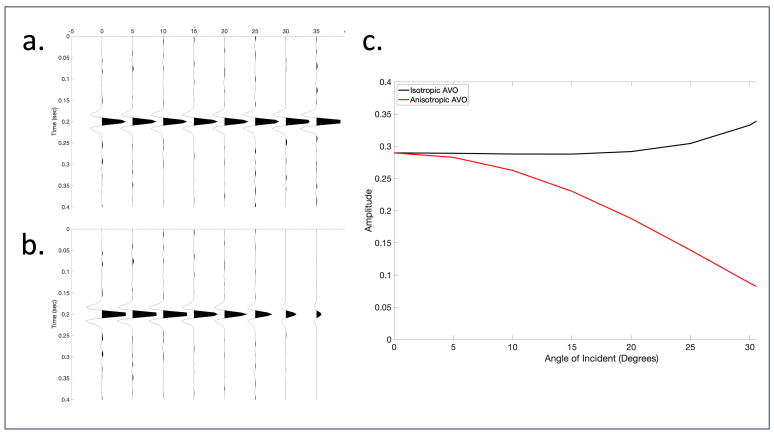


Figure 6.2: **a**. seismic gather with 25 db noise showing isotropic shale over sandstone reservoir. **b**. seismic gather with 25 db noise depicting VTI anisotropic shale over sandstone reservoir. **c** : AVO curves comparing isotropic and anisotropic shale scenarios.

Figures 6.2 presents seismic gather examples with identical reservoir properties, illustrating scenarios of isotropic shale and VTI anisotropic shale along with their respective AVO curves.

6.3 Labels and Features Extraction

In my neural network training, the dataset comprises labels and features. The labels include porosity and saturation of the sandstone reservoir. For features, I isolate a 100 ms window above and apply the same Tukey tapering technique used in Section 4.3, focusing on the seismic trace above and below the reservoir reflection. The windowed trace is represented as follows:

$$x_t^W(\theta) = |x_t^*(\theta)|_{t=100}^{t=300} \cdot T(t) \tag{6.1}$$

Next, I compute the mean of the windowed trace x_t^W at each incident angle as follows:

$$\overline{x}_t = \frac{1}{N} \sum_{i=1}^{N} x_i^W \tag{6.2}$$

where $\overline{x}_t$ is the mean amplitude of the windowed trace. These amplitude values

are organized into a feature vector $F \in \mathbb{R}^{z\times 1}$, where z is the count of incident angles. This vector, also known as an amplitude versus offset (AVO) curve since it tracks amplitude variations with offset or angle, serves as an input feature for my neural network's training dataset.

6.4 Numerical Example

My objective is to predict the porosity and saturation of a sandstone reservoir, assuming a VTI shale layer on top. The analysis is divided into four cases.

In the first case, I focus on demonstrating the neural network's ability to predict porosity and saturation of sandstone when all model parameters for both shale and sandstone are known. This serves as a baseline to assess the network's predictive capability under ideal conditions.

In the second case, I investigate a scenario in which the network is trained using isotropic rock physics model data and then applied to predict properties in a VTI reservoir. This setup tests the network's effectiveness at predicting reservoir properties under simplified modeling assumptions within a complex anisotropic geological context.

For the third case, I train the network across a wide range of Thomsen parameters ϵ and δ. This approach allows me to explore the network's behavior in a more complex setting, where a broader spectrum of anisotropic conditions is considered. The results from this case offer an understanding of the neural network's adaptability to varied anisotropic parameters.

Lastly, I apply my physics-guided neural network (PGNN) training workflow, incorporating a limited number of field measurements. This case is particularly relevant to real-world applications where comprehensive field data is not always

available, and it demonstrates how the network manages predictions with constrained data availability.

I utilize the neural network architecture and parametrization previously detailed in Section 4.4 and depicted in Figure 4.5. In each scenario, the shale in my geological model is considered VTI anisotropic with Thomsen parameters $\epsilon = 0.1$ and $\delta = 0.2$. It is also assumed to have a brine saturation of 90% or higher, with all other model parameters sampled from the distributions shown in Figure 4.3. The primary reservoir, sandstone, may contain varying mixtures of gas and water saturation, with its parameters also drawn from the distributions in Figure 4.3.

To evaluate the network's proficiency in predicting sandstone porosity and saturation, I use a test dataset D_T comprising 100 instances. This dataset reflects the shale layer's Thomsen parameters and sandstone properties selected from the aforementioned distributions. The prediction accuracy is assessed by calculating the room mean square error (RMSE) for the difference between the network prediction and the actual label within D_T, see equation 4.13.

6.4.1 Case 1: Training Neural Network with Fully Calibrated Rock Physics Modeled Data

I create a training dataset of 10,000 training instances where I sample from the sandstone and shale parameter distributions. I also assume knowledge of Thomsen parameters of the shale layer which allowed me to calculate VTI anisotropic modeled gathers. I train the network on 80% of the data and keep the remaining 20% as a validation dataset.

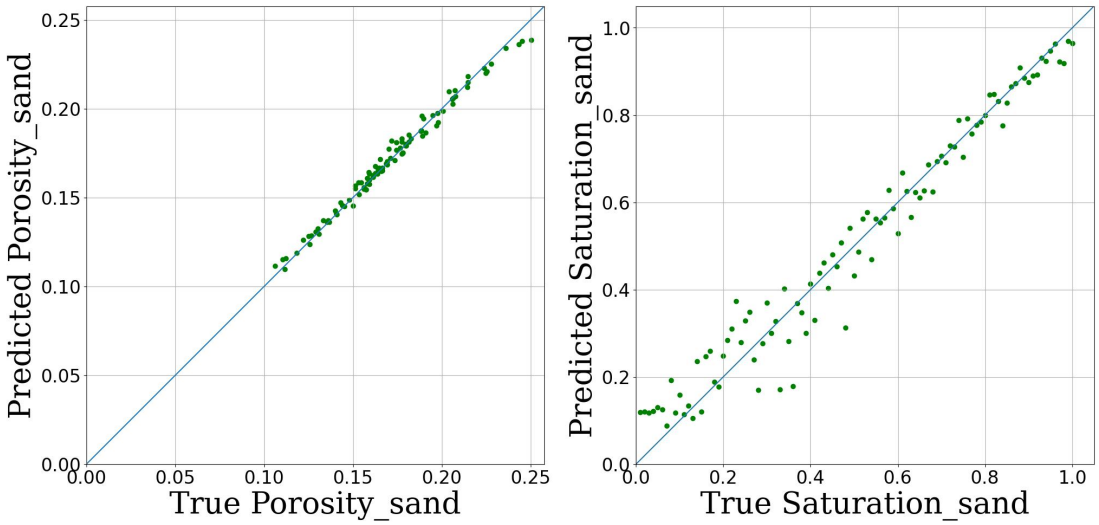


Figure 6.3: True vs. predicted porosity and saturation on the VTI test dataset D_T, using a network trained on VTI-calibrated anisotropic modeled dataset.

Figure 6.3 presents the prediction results on the VTI anisotropic test dataset D_T. The RMSE for porosity prediction is 3.57×10^{-2}, and for saturation prediction, it is 5.81×10^{-2}. These low RMSE values indicate that the network accurately predicts both porosity and saturation within VTI settings when trained on a modeled dataset with known calibration Thomsen parameters, ϵ and δ. Despite utilizing a patchy saturation fluid distribution, the network exhibits higher variance at lower saturation levels compared to higher values. This increased variance may be attributed to less linear velocity-saturation relationships at low saturation levels (see Figure 6.1). This analysis demonstrates that neural networks can effectively characterize porosity and saturation from seismic gathers in VTI settings, provided that a sufficiently representative training dataset exists.

6.4.2 Case 2: Training Neural Network Assuming Isotropic Rock Physics Modeled Data

I now assume a simplified rock physics model with isotropic shale over my sandstone reservoir. This simplification involves setting both ϵ and δ to 0 in the shale layer. Following the established seismic modeling workflow, I generate training instances and conduct predictions on the test dataset D_T, which is created under

the assumption of shale VTI anisotropy. The network is trained on 80% of this data, while the remaining 20% of the isotropic dataset is used for validation.

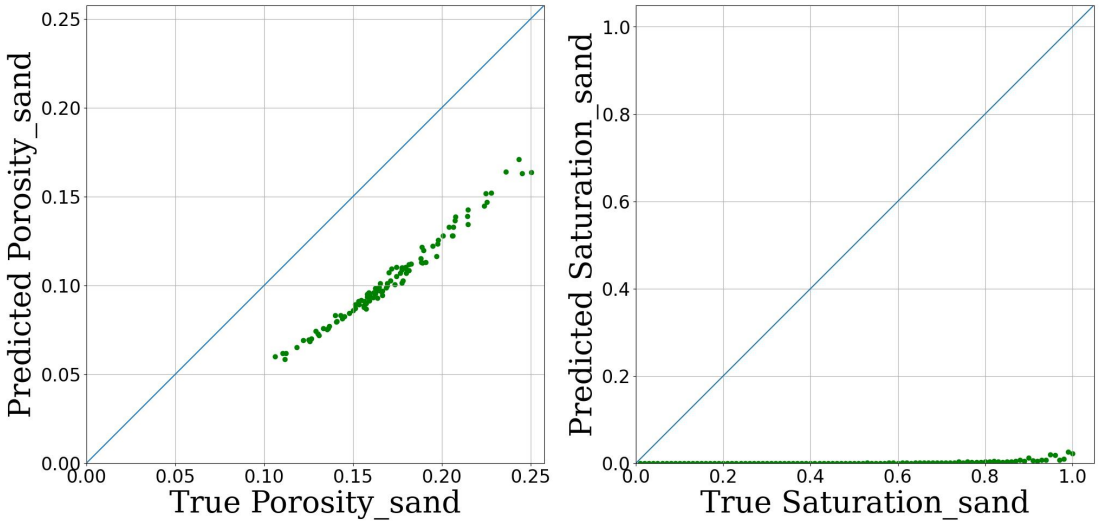


Figure 6.4: True vs. predicted porosity and saturation on the VTI test dataset D_T, using a network trained on an isotropic modeled dataset.

Figure 6.4 presents the network's prediction results on the VTI anisotropic test dataset D_T when trained using a simplified isotropic rock physics model. The RMSE for porosity prediction is 6.76×10^{-2}, and for saturation prediction, it is 5.78×10^{-1}. These relatively high RMSE values indicate that the network's performance significantly decreases when an isotropic model is applied to a VTI anisotropic context. The elevated RMSE is primarily due to discrepancies in AVO interpretation between the training dataset, which assumes isotropic behavior, and the test dataset, which incorporates the VTI anisotropic effects of ϵ and δ on near and far AVO signatures (see Section 2.5). These substantial prediction errors underscore the importance of employing an appropriate rock physics model when analyzing complex reservoirs.

6.4.3 Case 3: Training a Neural Network with Uncalibrated VTI Rock Physics Modeled Data

I examine a scenario where my geological model includes a VTI anisotropic shale layer over the main reservoir, but lacks precise knowledge of the Thomsen

anisotropy parameters ϵ and δ necessary for calibrating my rock physics model's elastic properties to the reservoir. To address this, I generate 30,000 training instances with varying combinations of ϵ and δ. For ϵ, values are sampled from a uniform distribution ranging from 0 to 0.35, and for δ, the range is -0.2 to 0.3. These ranges are based on the maximum and minimum values observed for shale in rocks, as reported by Thomsen (1986). I then assess the trained network's accuracy by predicting porosity and saturation of sandstone in a test dataset D_T that has fixed Thomsen parameter values in the shale layer.

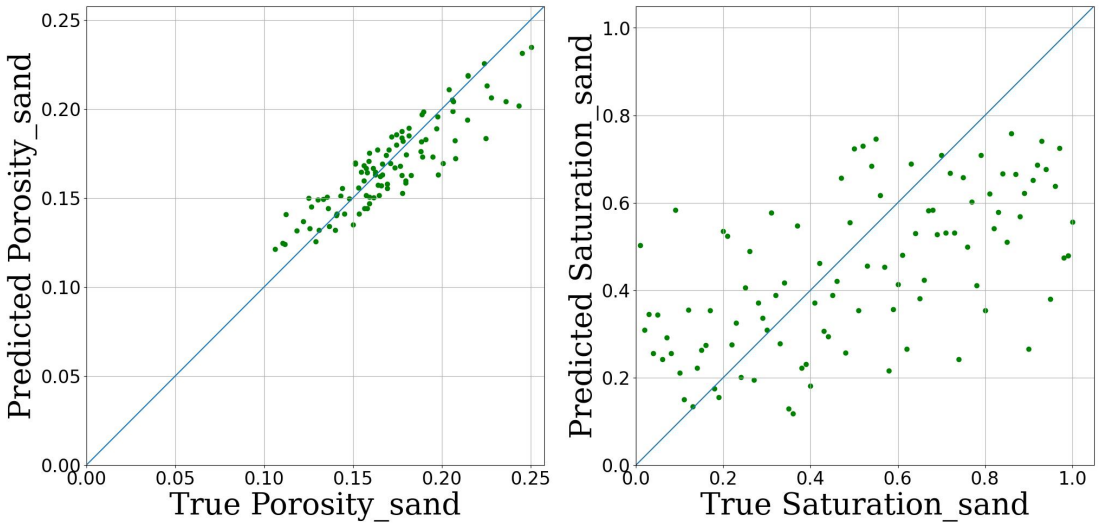


Figure 6.5: True vs. predicted porosity and saturation on the VTI test dataset D_T, using a network trained on wide range of ϵ and δ values.

Figure 6.5 presents the network's prediction results on the VTI anisotropic test dataset D_T when trained across a broad range of Thomsen parameters ϵ and δ. The RMSE for porosity and saturation are 1.47×10^{-2} and 2.39×10^{-1}, respectively. Utilizing the correct VTI physics model during training generally results in lower RMSE values compared to using an incorrect isotropic physics assumption (see Section 6.4.2). However, training an uncalibrated model across a wide range of Thomsen parameters leads to significant increases in prediction RMSE—240% for porosity and 165% for saturation—compared to training on a calibrated modeled dataset (see Section 6.4.4). This spike in RMSE is primarily due to multiple interpretations of the AVO curve influenced by varying ϵ and δ values, which

directly impact the modeled AVO signatures (see Section 2.5). These findings underscore the critical importance of employing the appropriate physics model and accurate calibration parameters to effectively characterize geological models.

6.4.4 Case 4: PGNN Training with Uncalibrated Rock Physics Models and Limited Field Data

In this section, I explore a scenario often encountered in practical settings, where the correct rock physics model is utilized but lacks proper calibration due to the unavailability of advanced field measurements, such as anisotropic parameters. My objective is to implement the PGNN training workflow to improve the prediction of porosity and saturation for a sandstone reservoir using these uncalibrated rock physics models.

I assume the existence of 20 basic field measurements that include porosity and saturation, alongside seismic gathers mirroring the reservoir's seismic responses. This data forms my field dataset D_F, with porosity and saturation as labels and features derived from seismic gathers as described in Section 6.3. 24 different combinations of D_F are created to test the robustness of my PGNN training.

I then utilize the uncalibrated rock physics model from Case 3 as D_M in my PGNN training. The success of this PGNN training is assessed by evaluating how well the network predicts sandstone properties in the test dataset D_T.

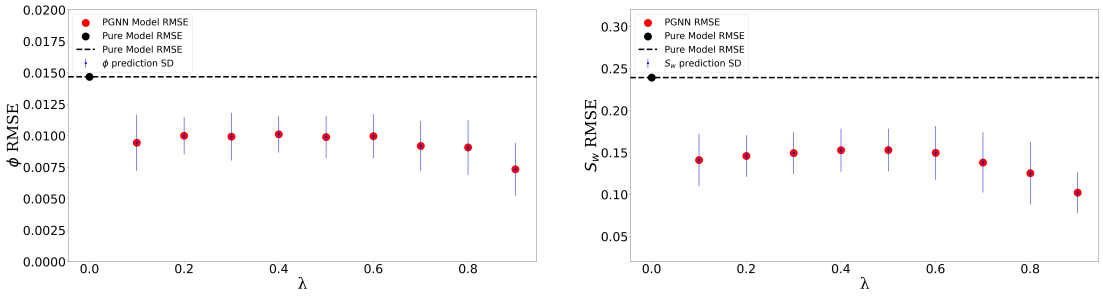


Figure 6.6: Average network predictions for porosity and saturation on the test dataset D_T using PGNN workflow at regularly sampled λ values.

Figure 6.6 presents the averaged prediction results for porosity and saturation on the VTI anisotropic test dataset D_T across different λ values using PGNN training. The results demonstrate an overall improvement in network predictions for both porosity and saturation as λ varies. The best performance is achieved at $\lambda = 0.9$, yielding an RMSE of 7.33×10^{-3} for porosity and 1.02×10^{-1} for saturation. This represents a reduction of 50% in porosity RMSE and 57% in saturation RMSE compared to the network trained on a wide range of ϵ and δ values (see Section 6.4.3).

Figure 6.7 illustrates an example of PGNN predictions at $\lambda = 0.9$, with an average saturation prediction RMSE of 1.27×10^{-1}. The figure highlights an overall improvement in saturation predictions, particularly at higher saturation levels, while lower saturation levels show less improvement. This discrepancy may be due to non-linear velocity-saturation relationships at lower saturations, as observed in Figure 6.1. Porosity predictions exhibit less variance when training is focused on values of 0.18 and below. However, at higher porosity values, accuracy decreases, potentially due to the increased impact of fluid on bulk density, which complicates the relationship between porosity and extracted seismic amplitudes (see Equation 2.8). These findings highlight the significance of the PGNN workflow in enhancing prediction accuracy with the implicit information about Thomsen parameters within the 20 basic field measurements.

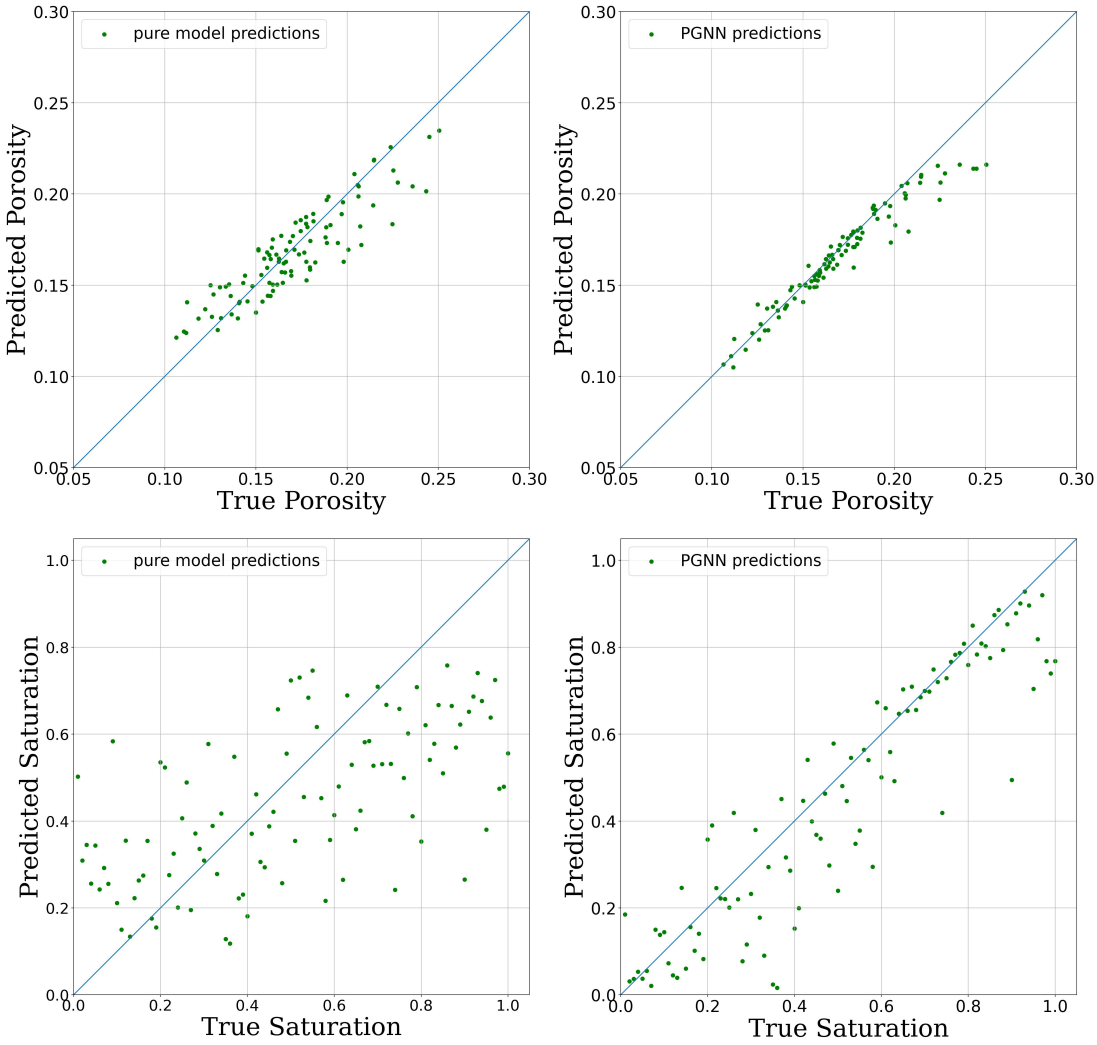


Figure 6.7: True vs. predicted porosity and saturation on the VTI test dataset D_T, using PGNN workflow at $\lambda = 0.9$.

6.5 Further Analysis

In the upcoming sections, I explore how increasing or decreasing the number of instances within the field dataset D_F influence PGNN's prediction accuracy using the above VTI numerical example. Additionally, I examine the effects of having minor variance in the unknown parameter(s) across the field, on the prediction accuracy of PGNN training workflow. I replicate such a scenario by introducing minor variances in the Thomsen modeling parameter when modeling the instances within D_F.

6.5.1 Impact of Field Dataset Size on PGNN Training in VTI Anisotropy

In field operations, the quantity of available measurements can fluctuate during the development stages. Consequently, these variations are likely to influence the prediction accuracy of the PGNN workflow. This section aims to assess how the number of measurements in the field dataset D_F affects PGNN's predictive accuracy. The analysis will follow the geological and rock physics model setup from case 4 (referenced in Section 6.4.4). I explore two scenarios: one with a limited dataset of 10 field measurements in D_F, and another with an expanded dataset containing 40 field measurements.

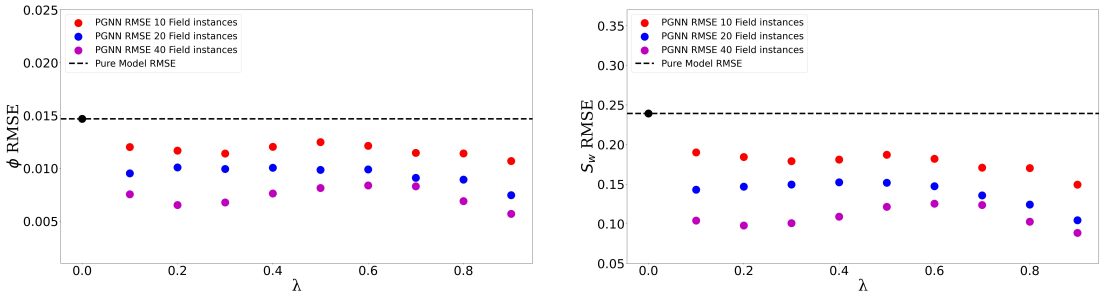


Figure 6.8: Average network predictions for porosity and saturation on the test dataset D_T using PGNN workflow at regularly sampled λ values with field dataset size of 10, 20 and 40.

Figure 6.8 demonstrates the impact of field dataset size (D_F) on PGNN prediction accuracy on a test dataset (D_T). Notably, the RMSE reductions for porosity and saturation predictions consistently improved as the size of the field dataset (D_F) increased. For the 10-instance dataset, porosity RMSE decreased by 21% and saturation RMSE by 31%. With a 20-instance dataset, these reductions were even greater, at 50% for porosity and 57% for saturation. Doubling the number of instances within D_F from 20 to 40 yielded a reduction in RMSE of 65% and 61% for porosity and saturation, respectively. These results highlight the network's ability to leverage measured data to implicitly infer reservoir anisotropy and improve predictions.

Figure 6.9 further shows decreasing standard deviation in PGNN predictions with increasing D_F size, reflecting better weight calibration and constrained predictions. This stability underscores the network's capability to integrate field data and account for anisotropic Thomsen parameters, enhancing prediction reliability for similar unseen datasets. Figure 6.10 presents examples of saturation predictions at $\lambda = 0.9$ for the various field dataset sizes, further illustrating the impact of field data quantity on the accuracy of PGNN training outcomes.

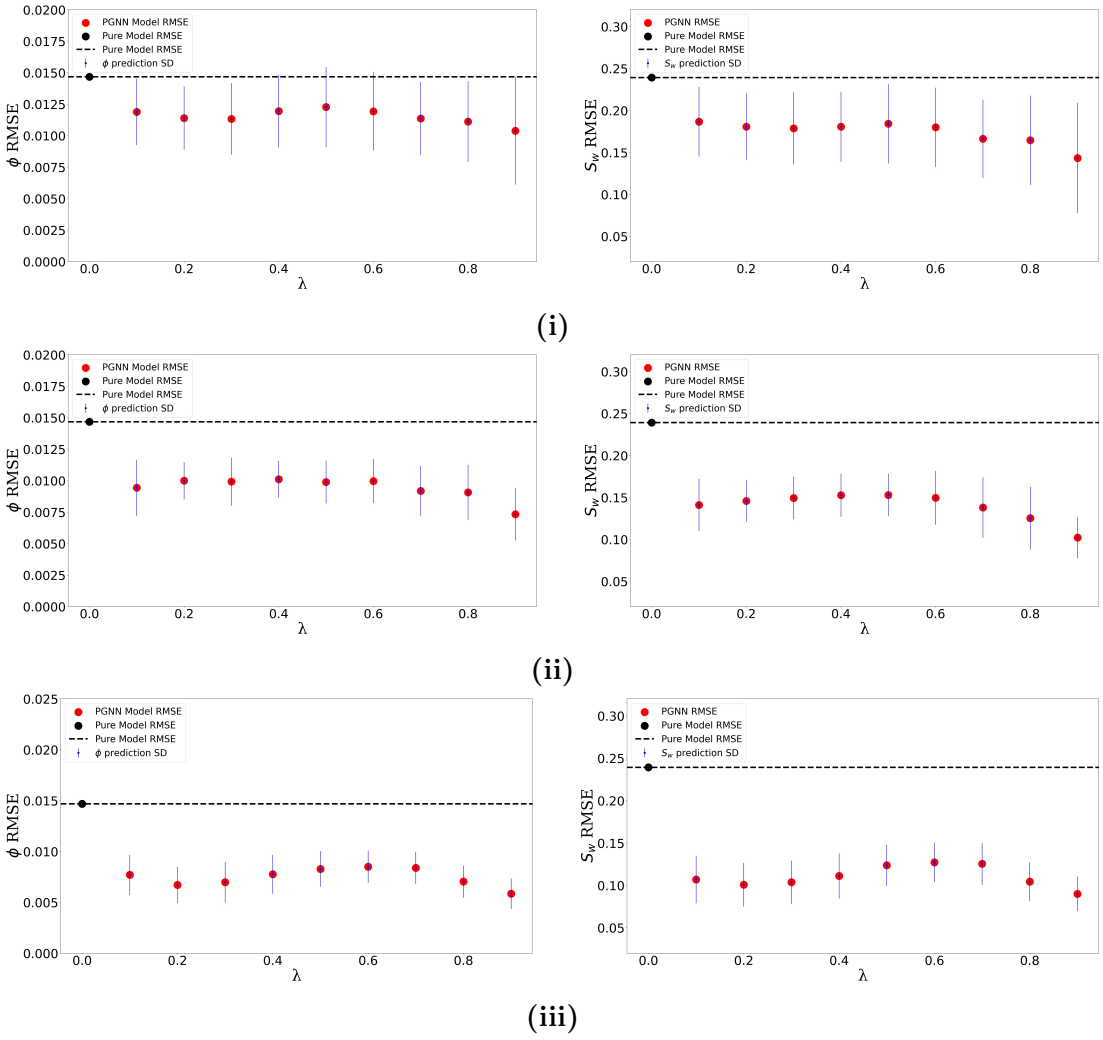


Figure 6.9: (**i**) PGNN prediction on test dataset D_T, when trained with $D_F = 10$ instances.(**ii**) PGNN prediction on test dataset D_T, when trained with $D_F = 20$ instances.(**iii**) PGNN prediction on test dataset D_T, when trained with $D_F = 40$ instances.

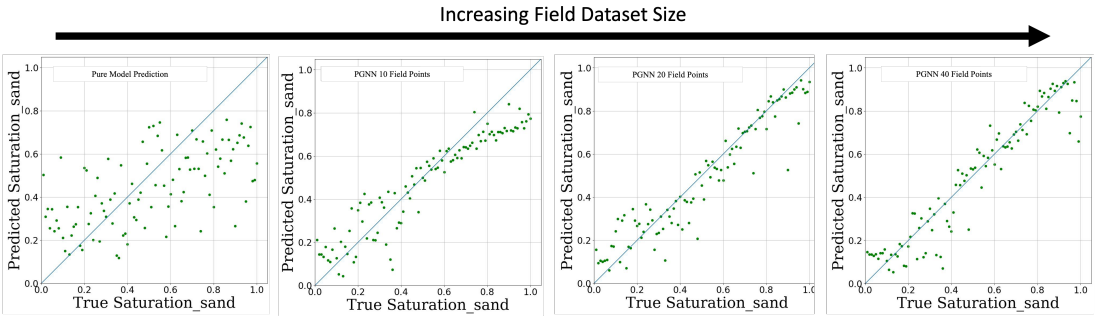


Figure 6.10: True vs. predicted porosity and saturation on the VTI test dataset D_T, using PGNN workflow at $\lambda = 0.9$ with varying D_F dataset sizes of 10, 20 and 40.

6.5.2 Impact of Variability in Unknown Parameter(s) on PGNN Training in VTI Anisotropy

In my examination of the Thomsen parameters ϵ and δ, I initially assumed their consistency across all instances within the field dataset D_F. While this assumption might be valid, real-world measurements often show slight variations in these parameters. To understand the impact of these variations, I introduced deviations in ϵ and δ values during the seismic gather modeling for D_F. I created two datasets, D_F^* and D_F^{**}, where ϵ and δ were sampled from normal distributions with the same means but different standard deviations. For D_F^*, the mean values for ϵ and δ are 0.1 and 0.2 respectively, with standard deviations of 0.01 and 0.02. For D_F^{**}, the means are the same, but the standard deviations are 0.03 and 0.04 for ϵ and δ, respectively. By keeping all other parameters constant across the 24 seeds of D_F, I was able to precisely assess the impact of these variances on the PGNN's predictive capabilities.

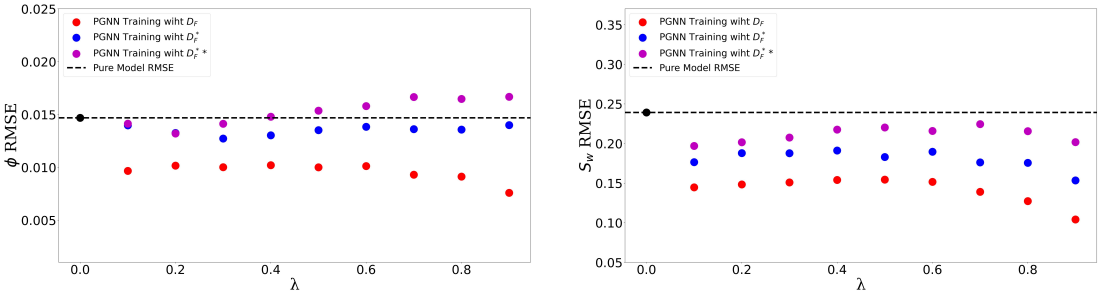


Figure 6.11: Comparison of PGNN training with fixed and variable unknown parameters ϵ and δ within D_F.

Figure 6.11 presents the average RMSE for porosity and saturation predictions on the unseen dataset D_T. Training with D_F^* at $\lambda = 0.9$ reduced saturation RMSE by 35% and porosity RMSE by 5% compared to pure model training. In contrast, using D_F^{**} resulted in a 16% decrease in saturation RMSE but a 15% increase in porosity RMSE at the same λ. Overall, introducing variability in the unknown Thomsen parameters degraded prediction accuracy because the network struggled to account for the varying relationships between seismic responses and reservoir properties.

Additionally, the optimal λ shifted towards lower values that place more weight on the modeled dataset D_M as parameter variability increased, especially affecting porosity predictions. High variability in ϵ and δ likely caused the network to overfit the field dataset by failing to capture the underlying relationships between seismic responses, Thomsen parameters, and reservoir properties.

Figure 6.12 shows that the standard deviation of predictions increases with higher λ values and greater parameter variability, supporting the notion of overfitting at these λ ranges. To mitigate overfitting, increasing the number of field data observations or restricting measurements to a geologically homogeneous zone could be effective strategies.

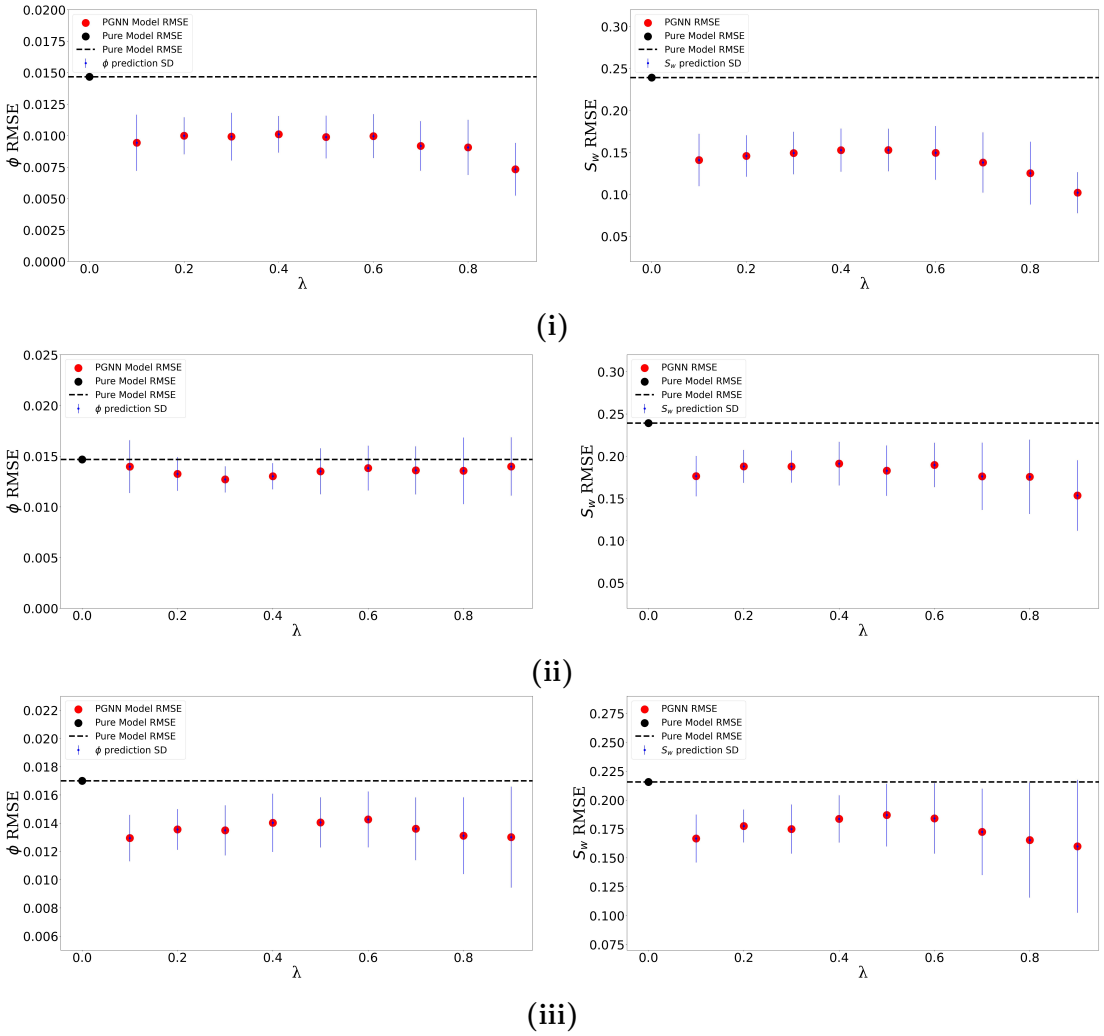


Figure 6.12: (**i**) Average Root Mean Square Error (RMSE) for porosity and saturation across 24 seeds of D_F: (**i**) without variance in ϵ and δ; (**ii**)with small variance in D_F^*; (**iii**) and with high variance in D_F^{**}.

6.6 Discussion

In practical scenarios, AVO modeling exercises are typically conducted to verify the accuracy of the applied rock physics model. A correct implementation should reveal a distinct discrepancy between the AVO signatures from isotropic rock physics modeling and seismic gathers with VTI anisotropic behaviors. Consequently, encountering such a scenario in real-world applications is unlikely. However, should this occur, the results from case 2 indicate that a network trained using isotropic AVO assumption would generate significant errors in predicting porosity and saturation when a VTI shale is on top of sandstone.

A more realistic scenario involves using the correct rock physics model but lacking sufficient measurements for proper calibration, as demonstrated in case 3. The results here indicate suboptimal predictions for both porosity and saturation. This issue arises mainly due to the extensive range of possible combinations of Thomsen parameters ϵ and δ, which significantly influence the AVO curve used in network training. As illustrated in Figure 6.2, even with the same sandstone reservoir model, the AVO response can vary greatly due to differences in ϵ and δ.

The PGNN training workflow demonstrated a notable improvement in predicting porosity and saturation, enhancing accuracy by 50% and 57% respectively. This was achieved using only 20 field measurements of porosity and saturation, along with their corresponding AVO curves. A key benefit of the PGNN approach is its ability to function without explicitly defining the Thomsen parameters ϵ and δ for calibrating the training dataset. Instead, it capitalizes on the neural network's capability to extract and interpret complex patterns between reservoir properties, Thomsen parameters and extracted seismic attributes, enabling it to implicitly optimize predictions based on both field-derived and modeled seismic features.

The availability of field measurements significantly influences the performance of the workflow. My findings reveal that increasing the number of field measurements enhances prediction accuracy and decreases variance, as shown in Figure 6.9. This improvement stems from the broader range of feature-label patterns provided to the neural network, aiding in better generalization to new datasets. The results also indicate that the optimal λ value remains stable across various field dataset sizes, suggesting the linear relationships between porosity-velocity and saturation-velocity are relatively straightforward to map, even with fewer data points. Therefore, with limited field data, it is still feasible to put more emphasis on the field dataset (D_F) during training. However, it is important to note that the modeled dataset D_M, even at a lower weight of $\lambda = 0.9$, is crucial for training the network to correctly map seismic AVO signatures to porosity and

saturation. Figure 6.13 showcases the average prediction from 24 seeds of D_F, using a standard neural network trained solely on a 20-instance field dataset D_F, to further illustrate this point.

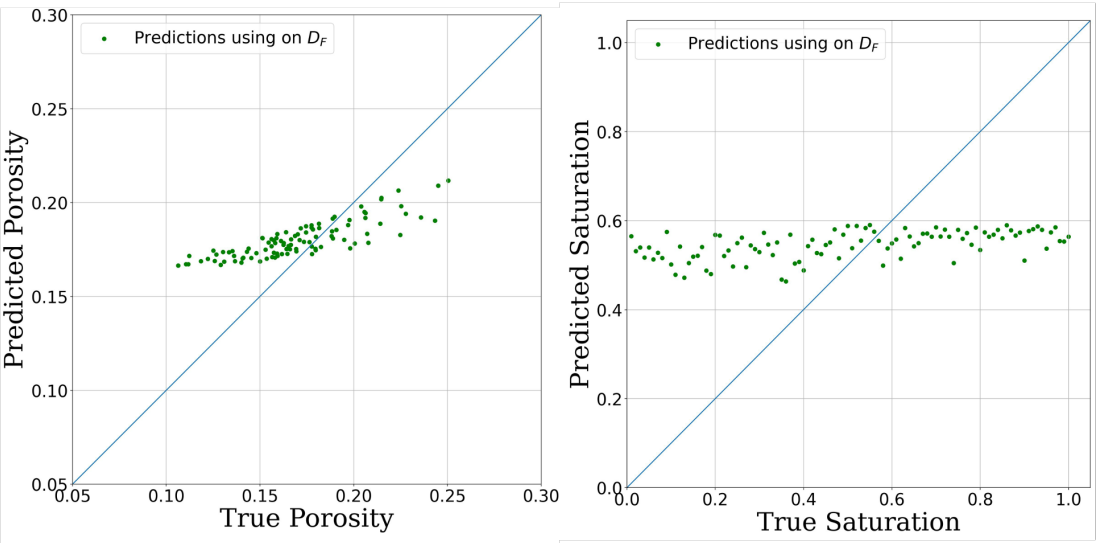


Figure 6.13: True vs. predicted porosity and saturation on the VTI test dataset D_T, using only the 20 instances within D_F.

The assessment of the PGNN training workflow with minor variations in Thomsen parameters ϵ and δ indicates that the workflow still improves prediction results, albeit with slightly reduced accuracy compared to scenarios with fixed parameters. The introduction of these variations resulted in increased variability in porosity and saturation predictions across 24 seeds of the field dataset D_F. This variability arises as the network adjusts its weights to account for measurements with ϵ and δ values deviating slightly from the fixed field value. However, as demonstrated in Figures 6.11 and 6.12, prediction accuracy is expected to decrease and variance to increase with greater variance in D_F. Consequently, in scenarios where significant variance in Thomsen parameters is anticipated across different sections of the field, employing separate PGNN models tailored to each area may ensure more accurate and consistent predictions.

7 Chapter 7: Discussion And Future Work

7.1 Discussion

Rock physics models are essential for reservoir characterization, as they link seismic attributes to reservoir properties. Traditionally, these models transform extracted elastic properties into reservoir characteristics, a process that is straightforward in isotropic, homogeneous reservoirs with basic log and core measurements. However, characterizing complex reservoirs is challenging due to the need for advanced measurements to calibrate their response, which can limit broader application.

This thesis explores the potential to accurately characterize complex reservoirs without fully defining all parameters or relying on expensive advanced measurements. It posits that variations in key model parameters affect elastic properties, which are then captured in seismic responses. Neural networks offer a promising solution by analyzing seismic amplitudes and implicitly calibrating their weights to properly map seismic extracted amplitudes to reservoir properties using a training dataset that contains different combinations of reservoir properties and their corresponding seismic response.

A major obstacle is often the lack of sufficient training data. This thesis proposes a method to mitigate this by using uncalibrated rock physics models combined with limited field data. The effectiveness of this approach in predicting reservoir properties greatly depends on a deep understanding of the reservoir and the availability of field data.

My approach focuses on understanding how variations in specific calibration parameters, such as Thomsen parameters ϵ and δ, affect seismic responses. By training the network with a variety of seismic attributes generated from an uncalibrated rock physics dataset, it learns to adjust its weights to capture the

influence of these parameters on the correlation between seismic attributes and predicted reservoir properties. This training allows the network to bias its weights towards certain parameter configurations using the hyperparameter λ, tailored to match field-derived seismic attributes, without explicitly defining the unknown parameters. The degree of this bias is influenced by the number of available measurements, relationship between attributes and reservoir properties and noise levels.

Optimizing the λ value in my PGNN workflow, particularly without a dedicated blind dataset, presents complexities. However, guidelines can be established for its selection. The λ range falls into three categories: model-biased, equal-contribution, and field-biased training. In scenarios with limited measurements, favoring the modeled dataset D_M (i.e., $\lambda \leq 0.3$) is recommended due to the risk of network overfitting and poor generalization on unseen data when high λ values are used, as noted in Section 6.5.2. As field measurements in D_F increase, I advise gradually shifting λ to prioritize D_F. The ideal λ value, correlating with the number of field measurements, is challenging to determine. A practical solution is to conduct a pilot study using fully synthetic data, varying the number of D_F instances and observing their impact on a synthetic blind dataset D_T. This approach helps establish a consensus on the most effective λ values relative to the volume of data in D_F.

Another factor influencing the selection of the hyperparameter λ is the distribution of reservoir properties within D_F. If D_F samples a narrow or constrained range of reservoir property values, more weight should be assigned to the modeled dataset (D_M). This allows the network to train on a broader range of potential values, thereby minimizing overfitting (see Section 5.6.2).

Similarly, if there is significant variation in the unknown parameters within the measured field dataset D_F, assigning more weight to the modeled dataset D_M

can help reduce the risk of overfitting. In such cases, the network may struggle to accurately calibrate its weights to account for the impact of the unknown parameters on the relationship between seismic responses and reservoir properties. A potential remedy for high variability within the measured data is to segment the data into smaller, geologically homogeneous field datasets or to acquire additional field measurements (see Section 6.5.2).

PGNN uses cross-validation with the dataset D_F to determine network convergence, as detailed in Section 5.5. To tackle potential underfitting of network weights due to this convergence approach, I tested transfer learning, which proved effective in improving network training, particularly at higher λ values (see Section 5.7.4).

An alternative method involves adjustments to the custom loss function or the neural network design. For instance, incorporating Gassmann's equations to verify that the predicted bulk and shear moduli fall within theoretical limits can be beneficial (Han and Batzle, 2004). Another approach is to integrate predicted reservoir properties into a seismic forward modeling exercise. This process generates seismic gathers, which can be compared with the original gathers used for feature extraction. The misfit between these gathers can serve as an additional component in the loss function, potentially enhancing the model's precision and adaptability (Biswas et al., 2019; Alfarraj and AlRegib, 2019). However, the extent to which modeling gathers using uncalibrated or simplified rock physics models can achieve accuracy remains uncertain.

Geoscientists often calibrate rock physics models by comparing observed seismic data with various model parameter combinations (Avseth et al., 2005). For instance, in my numerical example on Thomsen parameters, the 20 field measurements could be used in a least-squares inverse problem to identify optimal ϵ and δ values or in a grid search to find the combination that minimizes the misfit be-

tween modeled and field-extracted AVO curves. However, it has been noted that such searches often yield a large number of combinations due to the band-limited nature of seismic data (Sen, 1995), and grid searches can be resource-intensive depending on the parameter(s) bound and search resolution (Syarif et al., 2016).

My approach presents two significant advantages over traditional least-square optimization and grid search methods. First, neural networks are adept at uncovering intricate and nuanced patterns within extracted features. These patterns may not be immediately obvious from seismic gathers, but can greatly enhance the accuracy of reservoir property predictions. Second, training neural networks on both modeled and field datasets creates a model capable of managing random noise and processing artifacts in field data. These elements might be overlooked when relying exclusively on purely modeled rock physics data.

Additionally, while inverse methods for solving anisotropic model parameters may offer quicker solutions, my workflow implementation is notably straightforward. The most time-intensive aspect is the initial design and setup of a neural network that can effectively process seismic features and output corresponding reservoir properties. Once this foundation is established, modifying the loss function and executing the workflow for a designated λ value or range becomes a relatively straightforward process. This simplicity and the network's ability to discern complex patterns make my approach particularly advantageous for seismic data interpretation and reservoir property prediction.

The training time for my workflow is subject to variability, influenced by several key factors. In the context of a dispersive rock physics model, the training duration typically averages 3 hours per field iteration, covering 10 different values of the hyperparameter λ. In comparison, the application of PGNN to examples involving Thomsen parameters averages 4 hours per training session. This variance in training time can be attributed to the complexity of the relationship

between reservoir and elastic properties; more intricate relationships necessitate longer training periods. Additionally, the introduction of noise impacts training duration. While moderate noise levels extend the training time, excessively high noise can lead to early termination of training, often resulting in subpar prediction accuracy. The computing environment also plays a significant role. All my tests were conducted on an Apple MacBook Pro with an Apple M1 Max chip and 64 GB RAM. Sequential processing tends to speed up training, whereas running multiple tests in parallel can prolong it, especially as the number of concurrent tests increases. Moreover, higher λ values generally lead to longer training times when the dataset D_F robustly maps the relationship between elastic and reservoir properties. In contrast, if D_F comprises a limited number of instances or an uneven distribution of data, the training process concludes more rapidly. Therefore, the overall training time for PGNN is not a constant but varies depending on the specifics of the model, the data's noise level, computing resources, and the characteristics of the dataset in use.

Nevertheless, combining my method with grid search could potentially yield even better results. An initial grid search might narrow down the parameter search space, which can then be integrated into the PGNN training to further refine network predictions, as discussed in Section 5.7.1.

It is important to note that the proposed physics-guided neural network (PGNN) training method achieved higher accuracy compared to networks trained using uncalibrated rock physics models. However, it did not perform as well as networks trained with data from calibrated rock physics models, as discussed in Section 4.5.2 for dispersive models and Section 6.4.1 for VTI anisotropic reservoirs. The discrepancy in accuracy can be attributed to the network's need to fine-tune its weights against less representative training instances from the uncalibrated dataset that do not accurately reflect field reservoir properties. This misalignment often results in reduced prediction accuracy on field-extracted data.

Ideally, networks should be trained on calibrated models to maximize accuracy. However, this thesis presents an alternative approach that offers more accurate predictions when calibration of models is not feasible or when the available field dataset is insufficient for neural network training.

7.2 Future Work

There are multiple avenues for advancing and validating my PGNN workflow in future applications. The most straightforward method involves applying it to real field datasets, although this approach faces several challenges. A high SNR is crucial for dependable predictions. Additionally, seismic forward modeling must accurately replicate actual field conditions to prevent misinterpretations of reservoir properties, a concern underscored by Das et al. (2019). When dealing with seismic data from merged surveys, it's essential to analyze and normalize amplitude differences, which can be achieved through advanced seismic data processing techniques. A deep understanding of the field reservoir is necessary to correctly select a rock physics model that accurately represents its elastic properties. My research on dispersive rock physics models, as detailed in chapters 4 and 5, demonstrates potential effectiveness in scenarios such as CO_2 storage and heavy oil fields, where experimental settings have observed seismic dispersion (Batzle et al., 2006a; Behura et al., 2007; Falcon-Suarez et al., 2018).

A second strategy involves refining the neural network design and custom loss functions by exploring more complex neural architectures for enhanced accuracy. For example, deeper convolutional networks, typically used in image processing, might facilitate the extraction of more intricate patterns. Additionally, refining the custom loss function, by integrating physical constraints like Voigt or Reuss bounds or incorporating a forward modeling equation to seismically model predicted properties, could improve the calibration of critical parameters such as Thomsen anisotropy parameters (Karpatne et al., 2017b; Biswas et al., 2019).

Further, the workflow's adaptability could be tested across different geophysical inversion challenges such as fracture characterization, CO_2 sequestration, Total Organic Carbon (TOC) content analysis, and exploration in shale oil plays. Each scenario requires sophisticated rock physics modeling and precise seismic data interpretation, areas where my methods are well-equipped to manage complex challenges.

Moreover, while the PGNN workflow primarily focuses on directly predicting reservoir properties from seismic attributes, critical parameters like Thomsen parameters could also be predicted for applications such as anisotropic NMO analysis and migration, potentially enhancing these methods. Although current examples bypass the explicit prediction of these parameters, adapting the training approach to include them is feasible. Such predictions would rely on the network's accuracy for reservoir parameters, suggesting a potential use of my method as a pre-processing prediction step to identify key parameters critical for imaging or seismic processing.

Research by Moyano et al., 2011 illustrates how predicted reservoir properties vary depending on the chosen rock physics model and its assumptions. I propose that my workflow could mitigate such biases by training the network on multiple datasets generated from different rock physics models and using field data to reconcile prediction discrepancies. This approach could reduce bias in reservoir property predictions due to model selection and assumptions, although further analysis is required to validate this hypothesis.

8 Conclusion

I have replicated Wu et al.'s (2014) work on predicting gas saturation using neural networks, allowing the sensitivity of unknown parameters to be assessed. Wu et al. (2014) assumed that τ, which is the inverse of the frequency at which maximum attenuation occur due to fluid movement, is known. I have demonstrated how prediction accuracy declines as uncertainty around τ increases. For instance, when τ is sampled across a wide range of potential values, saturation prediction errors increase by 330%, and porosity prediction errors rise by 85% compared to a network trained with a known τ value. This emphasizes the importance of accurately identifying τ for improving the prediction accuracy of saturation and porosity.

I demonstrated that network predictions with an unknown τ can be improved by incorporating a small set of field measurements. For instance, 40 porosity and saturation measurements with fixed but unknown τ values, combined with their seismic attributes, constrained network predictions through a physics-guided neural network training workflow, reducing porosity RMSE by 40% and saturation RMSE by 33% compared to a network trained across a wide range of τ values without constraints. Furthermore, prediction accuracy correlated non-linearly with the number of available instances: increasing data points from 40 to 60 led to a 44% improvement in saturation prediction, while maintaining similar porosity RMSE. Conversely, reducing the data set to 20 instances lowered porosity and saturation RMSE to 14% and 15%, respectively.

The same physics-guided neural network (PGNN) training workflow was used to enhance predictions from a VTI layer with unknown parameters ϵ and δ. I showed that network predictions deteriorate by 240% and 165% for porosity and saturation predictions, respectively, if ϵ and δ are unknown. With the PGNN training workflow, a limited set of well measurements can significantly improve predictions. For instance, 20 porosity and saturation measurements with their

attributes, despite having fixed but unknown ϵ and δ Thomsen parameters, constrained network predictions and reduced predictions RMSE by 50% for porosity and 57% for saturation, compared to the network trained on a wide range of ϵ and δ. Furthermore, prediction accuracy directly non-linearly with the number of available measurements: doubling the field dataset reduced the prediction RMSE by 65% for porosity and 61% for saturation, while reducing it to 10 points reduced predictive power to 21% and 31% for porosity and saturation, respectively. I have also shown that increasing variations within the unknown parameters across field measurements also deteriorates prediction accuracy of the workflow.

I found that λ, which controls the influence/weight of the uncalibrated modeled dataset and the field dataset in the PGNN training, is affected by the number of field data points, the distribution of reservoir properties, and the variability of unknown parameters. Higher λ values correlate with an increasing number of field measurements, as shown in Chapter 5, where the optimal λ for porosity rose from 0.3 to 0.5 as field data expanded from 20 to 60 measurements. A uniform distribution of reservoir properties leads to higher λ, while skewed distributions require lower λ values for better generalization, as shown in Chapter 5. Variability in unknown parameters also shifts λ to lower values. For instance, in Chapter 6, high variability in ϵ and δ caused overfitting at high λ values, lowering the optimal value for porosity estimates from 0.9 to 0.2. These findings underscore the importance of pilot studies to assess these factors and tailor λ for accurate predictions using limited field measurement.

A Appendix: Derivation of Chapman Microstructural Poroelastic Model

Chapman et al. (2002) derived a formulation to calculate the effective bulk and shear moduli, K_{eff} and μ_{eff}, respectively at different frequencies (ω) assuming wet saturation, random connected cracks with unified radius size and spherical shape pores. The model $K_{eff}(\omega)$ and μ_{eff} are defined as:

$$\begin{aligned} K_{eff}(w) =& K - \varepsilon \left\{ \frac{4(3\lambda+2\mu)(\lambda+2\mu)}{(\mu(\lambda+\mu)}[1-3A(\omega)] - 4\pi r A(\omega) \right\} - \\ & \phi \left\{ \frac{3\lambda+2\mu}{4\mu} \left[\frac{\lambda+2\mu}{3\lambda+2\mu} + B(\omega) \right] - 3B(\omega) \right\}, \end{aligned} \tag{A.1}$$

$$\begin{aligned} \mu_{eff}(w) =& \mu - \frac{16}{45}\varepsilon \frac{1}{1+K_c} \frac{\mu(\lambda+2\mu)}{3\lambda+4\mu} \left(K_c + \frac{1}{1+i\omega\tau} \right) \\ & - \frac{32}{45}\varepsilon \frac{\mu(\lambda+2\mu)}{3\lambda+4\mu} - \phi \frac{15\mu(\lambda+2\mu)}{9\lambda+14\mu}. \end{aligned} \tag{A.2}$$

K, μ and λ are the bulk, shear and Lame parameters for the solid matrix of the reservoir. K_f is the fluid bulk modulus. ϕ is the total porosity. r is the aspect ratio of the cracks, a is the crack radius, ε is the crack density and τ is the timescale parameter. $A(\omega)$ and $B(\omega)$ are the frequency dependent constants defined as :

$$A(\omega) = \frac{\frac{i\omega\tau}{3(1+K_c)} - \gamma' i\omega\tau + \frac{1}{\gamma}(1+i\omega\gamma\tau)\left[\frac{1}{3(1+K_c)} + \gamma'\right]}{1 + i\omega\tau + \frac{1}{\gamma}(1+i\omega\tau)}, \tag{A.3}$$

$$B(\omega) = \frac{(1+i\omega\tau)\left[\frac{1}{3(1+K_c)} + \gamma'\right] - i\omega\tau\left[\frac{1}{3(1+K_c)} - \gamma'\right]}{1 + i\omega\gamma\tau + \gamma(1+i\omega\tau)}. \tag{A.4}$$

K_c and K_p are the crack and pore space compressibility parameters respectively, and they are described as:

$$K_c = \frac{\pi\mu r(\lambda+\mu)}{K_f(\lambda+2\mu)}, \tag{A.5}$$

$$K_p = \frac{4\mu}{3K_f}. \tag{A.6}$$

K_f represents the fluid bulk modulus of the pore fluid. γ and $\gamma^{'}$ are dimensionless parameters described as:

$$\gamma = \frac{9\phi(\lambda+\mu)(1+K_p)}{16\epsilon(\lambda+2\mu)(1+K_c)}, \tag{A.7}$$

$$\gamma^{'} = \gamma\frac{\lambda+2\mu}{(3\lambda+2\mu)(1+K_p)}. \tag{A.8}$$

All model parameters can be measured from basic laboratory experiments except for the time-scale parameter τ. τ can be approximated by:

$$\tau \approx \left(1+\frac{\sigma_c}{K_f}\right)\frac{\eta_f}{\zeta}, \tag{A.9}$$

ζ is the characteristic grain size length and σ_c is written as follows:

$$\sigma_c = \frac{\pi\mu r}{2(1-\nu)}, \tag{A.10}$$

ν is the Poisson's ratio of the solid mineral. ζ is a fundamental property of the rock that can be measured through lab experiments (Dvorkin et al., 1995). If τ is assumed known, then we can solve for ζ and calculate new $\tau^{'}$ at any fluid saturation by calculation new fluid bulk modulus $K_f^{'}$ and viscosity $\eta_f^{'}$ using equations 2.4 and 2.5 respectively.

References

Aggarwal, C. C. (2018). *Neural networks and deep learning*. Springer.

Ahmed, U., Crary, S. F., & Coates, G. R. (1991). Permeability estimation. The various sources and their interrelationships. *Journal of Petroleum Technology*, *43*(5).

Aki, K., & Richards, P. G. (1980). *Quantitative Seismology, Theory and Methods Volume I and Volume II* (Vol. 68).

Aleksandrov, K. S. (1961). The elastic properties of rock forming minerals II: layered silicates. *Bulletin USSR Academy of science, Geophysics Series 9*, 1165–1168.

Alfarraj, M., & AlRegib, G. (2019). Semi-supervised learning for acoustic impedance inversion. *SEG Technical Program Expanded Abstracts*.

Alkhalifah, T., & Larner, K. (1994). Migration error in transversely isotropic media. *Geophysics*, *59*(9).

Alkhalifah, T., & Tsvankin, I. (1995). Velocity analysis for transversely isotropic media. *Geophysics*, *60*(5).

Araya-Polo, M., Jennings, J., Adler, A., & Dahlke, T. (2018). Deep-learning tomography. *The Leading Edge*, *37*(1).

Avseth, P., Skjei, N., & Mavko, G. (2012). Rock Physics Modelling of Stress Sensitivity in Patchy Cemented Sandstones.

Avseth, P., Dvorkin, J., Mavko, G., & Rykkje, J. (2000). Rock physics diagnostic of north sea sands: Link between microstructure and seismic properties. *Geophysical Research Letters*, *27*(17), 2761–2764.

Avseth, P., Mukerji, T., & Mavko, G. (2005). *Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk*. Cambridge university press.

Avseth, P., Mukerji, T., Mavko, G., & Dvorkin, J. (2010). Rock-physics diagnostics of depositional texture, diagenetic alterations, and reservoir heterogeneity in high-porosity siliciclastic sediments and rocks - A review of selected models and suggested work flows. *Geophysics*, *75*(5).

Ba, L. J., & Frey, B. (2013). Adaptive dropout for training deep neural networks. *Advances in Neural Information Processing Systems 26*.

Bakulin, A., Grechka, V., & Tsvankin, I. (2000). Estimation of fracture parameters from reflection seismic data-Part I: HTI model due to a single fracture set. *Geophysics*, *65*(6).

Banik, N. C. (1984). Velocity anisotropy of shales and depth estimation in the North Sea basin. *Geophysics*, *49*(9).

Batzle, M., Hofmann, R., & Han, D. H. (2006a). Heavy oils - Seismic properties. *The Leading Edge*, *25*(6).

Batzle, M. L., Han, D.-H., & Hofmann, R. (2006b). Fluid mobility and frequency-dependent seismic velocity — Direct measurements. *Geophysics*, *71*(1), N1–N9.

Bebis, G., & Georgiopoulos, M. (1994). Feed-forward neural networks. *IEEE Potentials*, *13*(4), 27–31.

Behura, J., Batzle, M., Hofmann, R., & Dorgan, J. (2007). Heavy oils: Their shear story. *Geophysics*, *72*(5).

Benardos, P. G., & Vosniakos, G. C. (2007). Optimizing feedforward artificial neural network architecture. *Engineering Applications of Artificial Intelligence*, *20*(3).

Bianchini, M., & Scarselli, F. (2014). On the complexity of neural network classifiers: A comparison between shallow and deep architectures. *IEEE Transactions on Neural Networks and Learning Systems*, *25*(8).

Biot, M. A. (1956). Theory of Propagation of Elastic Waves in a Fluid-Saturated Porous Solid. II. Higher Frequency Range. *The Journal of the Acoustical Society of America*, *28*(2), 179–191.

Bishop, C. M. (1994). Neural networks and their applications. *Review of Scientific Instruments*, *65*(6), 1803–1832.

Bishop, C. M. (1995). Training with Noise is Equivalent to Tikhonov Regularization. *Neural Computation*, *7*(1), 108–116.

Biswas, R., Sen, M. K., Das, V., & Mukerji, T. (2019). Prestack and poststack inversion using a physics-guided convolutional neural network. *Interpretation*, *7*(3), SE161–SE174.

Blangy, J. P. (1994). AVO in transversely Isotropic media—An overview. *Geophysics*, *59*(5), 775–781.

Boitnott, G. N., Broadhead, M. K., & Keho, T. H. (2011). Laboratory measurements of modulus dispersion in sandstone at seismic frequencies. *SEG Technical Program Expanded Abstracts*, *30*(1).

Born, W. T. (1941). The Attenuation constant of earth materials. *Geophysics*, *6*(2).

Bosch, M., Mukerji, T., & Gonzalez, E. F. (2010). Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review.

Bredesen, K., Jensen, E. H., Johansen, T. A., & Avseth, P. (2015). Quantitative seismic interpretation using inverse rock physics modelling. *Petroleum Geoscience*, *21*(4).

Bridle, J. S. (1990). Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition. In *Neurocomputing*.

Brownlee, J. (2018). What is the Difference Between a Batch and an Epoch in a Neural Network? *Machine Learning Mastery*, (July).

Cadoret, T., Mavko, G., & Zinszner, B. (1998). Fluid distribution effect on sonic attenuation in partially saturated limestones. *Geophysics*, *63*(1).

Caruana, R., Lawrence, S., & Giles, L. (2001). Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. *Advances in Neural Information Processing Systems*.

Caspari, E., Müller, T. M., & Gurevich, B. (2011). Time-lapse sonic logs reveal patchy CO2 saturation in-situ. *Geophysical Research Letters*, *38*(13).

Cauchy, A.-L. (1847). Méthode générale pour la résolution des systèmes d'équations simultanées. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, *25*(2).

Chapman, M. (2009). Modeling the effect of multiple sets of mesoscale fractures in porous rock on frequency-dependent anisotropy. *Geophysics*, *74*(6), D97–D103.

Chapman, M., Liu, E., & Li, X. Y. (2006). The influence of fluid-sensitive dispersion and attenuation on AVO analysis. *Geophysical Journal International*, *167*(1).

Chapman, M., Zatsepin, S. V., & Crampin, S. (2002). Derivation of a microstructural poroelastic model. *Geophysical Journal International*, *151*(2), 427–451.

Connolly, P. (1999). Elastic impedance. *The Leading Edge*, *18*(4).

Danek, T., Andrzej Leśniak, & Anna Pita. (2010). *Numerical modeling of seismic wave propagation in selected anisotropic media* [Doctoral dissertation, Wydawnictwo Instytutu Gospodarki Surowcami Mineralnymi i Energia PAN].

Das, A., & Batzle, M. (2008). Modeling studies of heavy oil - In between solid and fluid properties. *The Leading Edge*, *27*(9).

Das, V., & Mukerji, T. (2020). Petrophysical properties prediction from prestack seismic data using convolutional neural networks. *Geophysics*, *85*(5).

Das, V., Pollack, A., Wollner, U., & Mukerji, T. (2019). Convolutional neural network for seismic impedance inversion. *Geophysics*, *84*(6).

Dewhurst, D. N., & Siggins, A. F. (2006). Impact of fabric, microcracks and stress field on shale anisotropy. *Geophysical Journal International*, *165*(1).

Dewhurst, D. N., Siggins, A. F., Sarout, J., Raven, M. D., & Nordgard-Bolas, H. M. (2011). Geomechanical and ultrasonic characterization of a Norwegian Sea shale. *Geophysics*, *76*(3).

Domenico, S. N. (1975). Effect of brine-gas mixture on velocity in an unconsolidated sand reservoir. *World Petroleum Congress Proceedings*, *1975-May*.

Dvorkin, J., Gutierrez, M. A., & Grana, D. (2014). *Seismic Reflections of Rock Properties*. Cambridge University Press.

Dvorkin, J., Mavko, G., & Nur, A. (1995). Squirt flow in fully saturated rocks. *Geophysics*, *60*(1), 97–107.

Dvorkin, J., Nolen-Hoeksema, R., & Nur, A. (1994). The Squirt-flow mechanism: Macroscopic description. *Geophysics*, *59*, 428–438.

Falcon-Suarez, I., Papageorgiou, G., Chadwick, A., North, L., Best, A. I., & Chapman, M. (2018). CO2-brine flow-through on an Utsira Sand core sample: Experimental and modelling. Implications for the Sleipner storage field. *International Journal of Greenhouse Gas Control*, *68*.

Frankle, J., & Carbin, M. (2019). The lottery ticket hypothesis: Finding sparse, trainable neural networks. *7th International Conference on Learning Representations, ICLR 2019*.

Frenay, B., & Verleysen, M. (2014). Classification in the Presence of Label Noise: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*, *25*(5), 845–869.

Gassmann, F. (1951). Uber die Elastizitat poroser Medien, Vier der Natur. Gesellshaft in Zurich. (96), 1–23.

Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*. O'Reily Media Inc.

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In Y. W. Teh & M. Titterington (Eds.), *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256, Vol. 9). PMLR.

González, E. F., Mukerji, T., Mavko, G., & Michelena, R. J. (2003). Near and far offset P-to-S elastic impedance for discriminating fizz water from commercial gas. *The Leading Edge*, *22*(10).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

Grana, D., Verma, S., Pafeng, J., Lang, X., Sharma, H., Wu, W., McLaughlin, F., Campbell, E., Ng, K., Alvarado, V., Mallick, S., & Kaszuba, J. (2017). A rock physics and seismic reservoir characterization study of the Rock Springs Uplift, a carbon dioxide sequestration site in Southwestern Wyoming. *International Journal of Greenhouse Gas Control*, *63*.

Grechka, V. (2009). *Applications of Seismic Anisotropy in the Oil and Gas Industry*. EAGE.

Grechka, V., & Mateeva, A. (2007). Inversion of P-wave VSP data for local anisotropy: Theory and case study. *Geophysics*, *72*(4).

Gurevich, B., Makarynska, D., de Paula, O. B., & Pervukhina, M. (2010). A simple model for squirt-flow dispersion and attenuation in fluid-saturated granular rocks. *Geophysics*, *75*(6), N109–N120.

Han, D. H., & Batzle, M. L. (2004). Gassmann's equation and fluid-saturation effects on seismic velocities. *Geophysics*, *69*(2).

Han, D.-H., & Batzle, M. (2002). Fizz water and low gas-saturated reservoirs. *The Leading Edge*, *21*(4), 395–398.

Han, D.-h., & Batzle, M. (2005). Diagnosis of "fizz-gas" and gas reservoirs in deep-water environment. *9th International Congress of the Brazilian Geophysical Society & EXPOGEF, Salvador, Bahia, Brazil, 11-14 September 2005*, 1324–1327.

Hawkins, D. M. (2004). The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences*, *44*(1).

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580.*

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1).

Hooke, R. (1678). De Potentia Reflitutiva ,or of Spring. Explaining the Power of Springing Bodies. In *Of spring: Explaining the power of springing bodies.*

Hsu, K., Schoenberg, M., & Walsh, J. (1991). Anisotropy from polarization and moveout. *1991 SEG Annual Meeting.*

Hudson, J. A., Liu, E., & Crampin, S. (1996). The mechanical properties of materials with interconnected cracks and pores. *Geophysical Journal International*, *124*(1), 105–112.

Hunter, D., Yu, H., Pukish, M. S., Kolbusz, J., & Wilamowski, B. M. (2012). Selection of proper neural network sizes and architectures-A comparative study. *IEEE Transactions on Industrial Informatics*, *8*(2).

Jin, P., Zhang, X., Chen, Y., Huang, S. X., Liu, Z., & Lin, Y. (2022). Unsupervised Learning of Full-waveform Inversion: Connecting CNN and Partial Differential Equation in a Loop. *ICLR 2022 - 10th International Conference on Learning Representations.*

Jin, Z., Chapman, M., Wu, X., & Papageorgiou, G. (2017). Estimating gas saturation in a thin layer by using frequency-dependent amplitude versus offset modelling. *Geophysical Prospecting*, *65*(3).

Johnston, J. E., & Christensen, N. I. (1995). Seismic anisotropy of shales. *Journal of Geophysical Research*, *100*(B4).

Jones, L. E. A., & Wang, H. F. (1981). Ultrasonic velocities in Cretaceous shales from the Williston basin. *Geophysics*, *46*(3).

Karlik, B., & Olgac, A. (2010). Performance analysis of various activation functions in generalized MLP architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, *1*(4), 111–122.

Karpatne, A., Watkins, W., Read, J., & Kumar, V. (2017a). Physics-guided Neural Networks (PGNN): An Application in Lake Temperature Modeling. *arXiv preprint arXiv:1710.11431*, *2*.

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., & Kumar, V. (2017b). Theory-guided data science: A

new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, *29*(10).

Karsoliya, S. (2012). Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture. *International Journal of Engineering Trends and Technology*, *3*(6).

Kim, J., Nam, M. J., & Matsuoka, T. (2013). Estimation of CO2 saturation during both CO2 drainage and imbibition processes based on both seismic velocity and electrical resistivity measurements. *Geophysical Journal International*, *195*(1).

Kim, K. Y., Wrolstad, K. H., & Aminzadeh, F. (1993). Effects of transverse isotropy on P-wave AVO for gas sands. *Geophysics*, *58*(6), 883–888.

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Klimentos, T., & McCann, C. (1990). Relationships among compressional wave attenuation, porosity, clay content, and permeability in sandstones. *Geophysics*, *55*(8).

Lawrence, S., Giles, C. L., & Tsoi, A. C. (1997). Lessons in neural network training: overfitting may be harder than expected. *Proceedings of the National Conference on Artificial Intelligence*, 540–545.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, *343*(6176), 1203–1205.

Li, J., & Li, W. (2018). A quantitative seismic prediction technique for the brittleness index of shale in the Jiaoshiba Block, Fuling shale gas field in the Sichuan Basin. *Natural Gas Industry B*, *5*(1).

Liu, E., Zelewski, G., Lu, C. P., Reilly, J., & Shevchek, Z. J. (2010). Seismic fracture prediction using azimuthal AVO analysis in a Middle East carbonate field: Workflow and mitigation of overburden effects. *Society of Exploration Geophysicists International Exposition and 80th Annual Meeting 2010, SEG 2010*.

Liu, Z., & Sun, Z. (2015). New brittleness indexes and their application in shale/clay gas reservoir prediction. *Petroleum Exploration and Development*, *42*(1).

Lu, P., Morris, M., Brazell, S., Comiskey, C., & Xiao, Y. (2018). Using generative adversarial networks to improve deep-learning fault interpretation networks. *The Leading Edge*, *37*(8), 578–583.

Maleki, S., Moradzadeh, A., Riabi, R. G., Gholami, R., & Sadeghzadeh, F. (2014). Prediction of shear wave velocity using empirical correlations and artificial intelligence methods. *NRIAG Journal of Astronomy and Geophysics*, *3*(1), 70–81.

Masters, D., & Luschi, C. (2018). Revisiting Small Batch Training for Deep Neural Networks. *arXiv preprint arXiv:1804.07612*.

Maultzsch, S., Chapman, M., Liu, E., & Li, X. Y. (2003). Modelling frequency-dependent seismic anisotropy in fluid-saturated rock with aligned fractures: implication of

fracture size estimation from anisotropic measurements. *Geophysical Prospecting*, *51*(5), 381–392.

Mavko, G., & Bandyopadhyay, K. (2009). Approximate fluid substitution for vertical velocities in weakly anisotropic VTI rocks. *Geophysics*, *74*(1).

Mavko, G., & Mukerji, T. (1998). Bounds on low-frequency seismic velocities in partially saturated rocks. *Geophysics*, *63*(3).

Mavko, G., Mukerji, T., & Dvorkin, J. (1998). Rock physics handbook cover: Tools for seismic analysis in porous media. In *The rock physics handbook*.

Mavko, G., Mukerji, T., & Dvorkin, J. (2020). *The Rock Physics Handbook*. Cambridge university press.

Moyano, B., Jensen, E. H., & Johansen, T. A. (2011). Improved quantitative calibration of rock physics models. *Petroleum Geoscience*, *17*(4), 345–354.

Müller, T. M., Gurevich, B., & Lebedev, M. (2010). Seismic wave attenuation and dispersion resulting from wave-induced flow in porous rocks - A review.

Murphy, W. F. (1982). Effects of partial water saturation on attenuation in Massilon sandstone and Vycor porous glass. *The Journal of the Acoustical Society of America*, *71*(6), 1458–1468.

Murphy, W. F., Winkler, K. W., & Kleinberg, R. L. (1984). Frame modulus reduction in sedimentary rocks: The effect of adsorption on grain contacts. *Geophysical Research Letters*, *11*(9), 805–808.

O'Connell, R. J., & Budiansky, B. (1977). Viscoelastic properties of fluid-saturated cracked solids. *Journal of Geophysical Research: Solid Earth*, *82*(36), 5719–5735.

Oppenheim, A. V., & Schafer, R. W. (1998). Discrete Time Signal Processing 2nd Edition.

Papageorgiou, G., & Chapman, M. (2017). Wave-propagation in rocks saturated by two immiscible fluids. *Geophysical Journal International*, *209*(3).

Payne, S. S., Worthington, M. H., Odling, N. E., & West, L. J. (2007). Estimating permeability from field measurements of seismic attenuation in fractured chalk. *Geophysical Prospecting*, *55*(5).

Röth, G., & Tarantola, A. (1994). Neural networks and inversion of seismic data. *Journal of Geophysical Research: Solid Earth*, *99*(B4), 6753.

Rüger, A. (1996). *Reflection coefficients and azimuthal avo analysis in anisotropic media* [Doctoral dissertation, Colorado School of Mines].

Russell, B. (2019). Machine learning and geophysical inversion - A numerical study. *The Leading Edge*, *38*(7).

Sarout, J., Delle Piane, C., Nadri, D., Esteban, L., & Dewhurst, D. N. (2015). A robust experimental determination of Thomsen's δ parameter. *Geophysics*, *80*(1).

Sarout, J., & Guéguen, Y. (2008). Anisotropy of elastic wave velocities in deformed shales: Part 1 - Experimental results. *Geophysics*, *73*(5).

Sbar, M. L. (2000). Exploration risk reduction: An AVO analysis in the offshore Middle Miocene, Central Gulf of Mexico. *The Leading Edge*, *19*(1).

Schmidt-Hieber, J. (2020). Nonparametric regression using deep neural networks with relu activation function. *Annals of Statistics*, *48*(4).

Schoenberg, M., & Protazio, J. (1990). "Zoeppritz" rationalized, and generalized to anisotropic media. *The Journal of the Acoustical Society of America*, *88*(S1), S46–S46.

Schön, J. H. (2015). Physical Properties of Rocks: Fundamentals and Principles of Petrophysics (2nd edition). *Developments in Petroleum Science (Elsevier)*, *65*.

Sen, M. K. (1995). Global Optimization Methods in Geophysical Inversion. *Advances in Exploration Geophysics*, *4*.

Sengupta, M., Mavko, G., & Mukerji, T. (2003). Quantifying subresolution saturation scales from time-lapse seismic data: A reservoir monitoring case study.

Simard, P. Y., Steinkraus, D., & Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, *2003-January*.

Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, *97*.

Smith, L. N. (2018). Disciplined Approach To Neural Network. *arXiv*.

Smith, T. M., Sondergeld, C. H., & Rai, C. S. (2003). Gassmann fluid substitutions: A tutorial. *Geophysics*, *68*(2).

Stoffer, D. S., & Bloomfield, P. (2000). Fourier Analysis of Time Series: An Introduction. *Journal of the American Statistical Association*, *95*(452).

Sun, J., Niu, Z., Innanen, K. A., Li, J., & Trad, D. O. (2020). A theory-guided deep-learning formulation and optimization of seismic waveform inversion. *Geophysics*, *85*(2), R87–R99.

Svozil, D., Kvasnička, V., & Pospíchal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, *39*(1).

Syarif, I., Prugel-Bennett, A., & Wills, G. (2016). SVM Parameter Optimization using Grid Search and Genetic Algorithm to Improve Classification Performance. *Telkomnika*, *14*(4).

Thomsen, L. (1986). Weak elastic anisotropy. *Geophysics*, *51*(10).

Thomsen, L. (1993). Weak anisotropic reflections. In *Offset dependent reflectivity . theory and practice of avo analysis*.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1).

Torrey, L., & Shavlik, J. (2010). Transfer Learning. In *Handbook of research on machine learning applications and trends* (pp. 242–264). IGI Global.

Tsvankin, I., & Thomsen, L. (1994). Nonhyperbolic reflection moveout in anisotropic media. *Geophysics*, *59*(8).

Tsvankin, I. (2012). *Seismic Signatures and Analysis of Reflection Data in Anisotropic Media* (3rd). Society of Exploration Geophysicists.

Tsvankin, I., Gaiser, J., Grechka, V., der Baan, M., & Thomsen, L. (2010). Seismic anisotropy in exploration and reservoir characterization: An overview. *Geophysics*, *75*(5).

Tsvankin, L. (1996). P-wave signatures and notation for transversely isotropic media: An overview.

Uzair, M., & Jamil, N. (2020). Effects of Hidden Layers on the Efficiency of Neural networks. *Proceedings - 2020 23rd IEEE International Multi-Topic Conference, INMIC 2020*.

Vernik, L., & Nur, A. (1992). Ultrasonic velocity and anisotropy of hydrocarbon source rocks. *Geophysics*, *57*(5).

Walsh, J., Sinha, B., Plona, T., Miller, D., & Ammerman, M. (2007). Derivation of anisotropy parameters in a shale using borehole sonic data. *EG International Exposition and Annual Meeting*.

Wang, Y. (2015). Frequencies of the Ricker wavelet. *Geophysics*, *80*(2).

Wang, Z. (2002). Seismic anisotropy in sedimentary rocks, part 2: Laboratory data. *Geophysics*, *67*(5), 1423–1440.

Wang, Z., & Nur, A. M. (1988). Velocity dispersion and the "local flow" mechanism in rocks. *SEG Technical Program Expanded Abstracts*, 548–550.

White, J. E. (1975). Computed seismic speeds and attenuation in rocks with partial gas saturation. *Geophysics*, *40*(2), 224–232.

Winkler, K., & Nur, A. (1979). Pore fluids and seismic attenuation in rocks. *Geophysical Research Letters*, *6*(1).

Winkler, K. W. (1985). Dispersion analysis of velocity and attenuation in Berea sandstone. *Journal of Geophysical Research: Solid Earth*, *90*(B8), 6793.

Winkler, K. W. (1986). Estimates of velocity dispersion between seismic and ultrasonic frequencies. *Geophysics*, *51*(1), 183–189.

Wood, A. B., & Lindsay, R. B. (1956). *A Textbook of Sound* (3rd).

Wright, J. (1987). The effects of transverse isotropy on reflection amplitude versus offset. *Geophysics*, *52*(4).

Wu, B., Meng, D., Wang, L., Liu, N., & Wang, Y. (2020). Seismic Impedance Inversion Using Fully Convolutional Residual Network and Transfer Learning. *IEEE Geoscience and Remote Sensing Letters*, *17*(12).

Wu, C., Ba, J., Zhang, L., & Carcione, J. M. (2022). Estimation of the squirt-flow length based on crack Properties in tight sandstones. *Lithosphere*, *2021*(Special 3).

Wu, M., Chen, D., & Chen, G. (2012). New spectral leakage-removing method for spectral testing of approximate sinusoidal signals. *IEEE Transactions on Instrumentation and Measurement*, *61*(5).

Wu, X., Chapman, M., Li, X.-Y., & Boston, P. (2014). Quantitative gas saturation estimation by frequency-dependent amplitude-versus-offset analysis. *Geophysical Prospecting*, *62*(6), 1224–1237.

Xiao, C. (2006). *A comparison of different methods for estimating Thomsen anisotropy parameters* [Doctoral dissertation, Calgary].

Yang, F., & Ma, J. (2019). Deep-learning inversion: A next-generation seismic velocity model building method. *Geophysics*, *84*(4).

Yilmaz, Ö. (2001). *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data* (Vol. 10).

Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.

Zhu, F., Gibson, R. L., Watkins, J. S., & Yuh, S. H. (2000). Distinguishing fizz gas from commercial gas reservoirs using multicomponent seismic data. *The Leading Edge*, *19*(11).