2023

Advancing Carbon Sequestration through Smart Proxy Modeling: Leveraging Domain Expertise and Machine Learning for Efficient Reservoir Simulation

Iman Oraki Kohshour
West Virginia University, imanoraki@gmail.com

Advancing Carbon Sequestration through Smart Proxy Modeling: Leveraging Domain Expertise and Machine Learning for Efficient Reservoir Simulation

Iman Oraki Kohshour, BSc, MSc

Dissertation submitted

to the Statler College of Engineering and Mineral Resources

at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in

Petroleum and Natural Gas Engineering

Shahab D. Mohaghegh, Ph.D., Chair

Samuel Ameri, M.S., P.E.

Kashy Aminian, Ph.D.

Mohamed El Sgher, Ph.D.

Qingqing Huang, PhD.

Department of Petroleum and Natural Gas Engineering

Morgantown, West Virginia,
2023

Keywords: Smart Proxy Modeling in Reservoir
Engineering, Artificial Intelligence, Data-Driven, Neural Network, Big Data, CO_2 sequestration

ABSTRACT

Advancing Carbon Sequestration through Smart Proxy Modeling: Leveraging Domain Expertise and Machine Learning for Efficient Reservoir Simulation

Iman Oraki Kohshour

Geological carbon sequestration (GCS) offers a promising solution to effectively manage extra carbon, mitigating the impact of climate change. This doctoral research introduces a cutting-edge Smart Proxy Modeling-based framework, integrating artificial neural networks (ANNs) and domain expertise, to re-engineer and empower numerical reservoir simulation for efficient modeling of CO_2 sequestration and demonstrate predictive conformance and replicative capabilities of smart proxy modeling.

Creating well-performing proxy models requires extensive human intervention and trial-and-error processes. Additionally, a large training database is essential to ANN model for complex tasks such as deep saline aquifer CO_2 sequestration since it is used as the neural network's input and output data. One major limitation in CCS programs is the lack of real field data due to a lack of field applications and issues with confidentiality.

Considering these drawbacks, and due to high-dimensional nonlinearity, heterogeneity, and coupling of multiple physical processes associated with numerical reservoir simulation, novel research to handle these complexities as it allows for the creation of possible CO_2 sequestration scenarios that may be used as a training set. This study addresses several types of static and dynamic realistic and practical field-base data augmentation techniques ranging from spatial complexity, spatio-temporal complexity, and heterogeneity of reservoir characteristics. By incorporating domain-expertise-based feature generation, this framework honors precise representation of reservoir overcoming computational challenges associated with numerical reservoir tools.

The developed ANN accurately replicated fluid flow behavior, resulting in significant computational savings compared to traditional numerical simulation models. The results showed that all the ML models achieved very good accuracies and high efficiency. The findings revealed that the quality of the path between the focal cell and injection wells emerged as the most crucial factor in both CO_2 saturation and pressure estimation models. These insights significantly contribute to our understanding of CO_2 plume monitoring, paving the way for breakthroughs in investigating reservoir behavior at a minimal computational cost.

The study's commitment to replicating numerical reservoir simulation results underscores the model's potential to contribute valuable insights into the behavior and performance of CO_2 sequestration systems, as a complimentary tool to numerical reservoir simulation when there is no measured data available from the field. The transformative nature of this research has vast implications for advancing carbon storage modeling technologies. By addressing the computational limitations of traditional numerical reservoir models and harnessing the synergy between machine learning and domain expertise, this work provides a practical workflow for efficient decision-making in sequestration projects.

ACKNOWLEDGMENTS

I express my deepest gratitude to all who contributed significantly to the completion of this dissertation:

To my esteemed PhD academic advisor, **Dr. Shahab D. Mohaghegh**: Your mentorship has been a beacon of wisdom, illuminating both my dissertation and my entire academic journey. Your unwavering support and insightful guidance have profoundly shaped not only the work presented here but have also been a transformative influence in my personal and professional growth. Your impact extends far beyond the pages of this dissertation, leaving an indelible mark on my life.

My sincere thanks extend to the members of my dissertation committee, **Prof. Sam Ameri**, **Dr. Kashy Aminian**, **Dr. Mohamed El Sgher, and Dr. Qingqing Huang**, for their invaluable insights and constructive feedback. I would also like to thank **Dr. Mehrdad Zamirian** for his valuable feedback throughout the early stages of this project's construction phase. I am appreciative of the engaging discussions with my colleagues peers at the WVU's Laboratory for Engineering Application of Data Science (**LEADS**) (Maher, Marwan, Ayo, Yvon, Gabe, Mehrdad, Amir, Sina, Andrew, Arya, Ali, Ansa, Afeez, Yasir, and Munirah), which provided an intellectually stimulating environment for me. Special thanks to Statler College of Engineering and Mineral Resources (**Dianne Stewart, Sarah Lowery**), Statler College Hall's Fellowship Committee, and the PNGE Office Manager (**Adriane Matheny**), Other helpful/ resourceful Statler's College directors and staff (Dean Pedro J. Mago, Cerasela Zoica Dinu, Robin Hissam, Jessica Queener, Nancie Yost, Kathleen Cullen, Kolin Brown), and Statler's IT staff. I also acknowledge the financial support from **Prof. Sam Ameri** and **Dr. Shahab D. Mohaghegh**, enabling me to dedicate time and resources to this research. I extend my gratitude to the Computer Modeling Group, Inc, Intelligent Solutions Inc., and Python developers. for their crucial contributions to this research. I'd like to emphasize the significant impact of **Intelligent Solutions Inc.**, particularly their innovative Data-Driven Modeling Softwares, **IMprove™** and **IMagine™**. These advanced tools, provided by Intelligent Solutions Inc., played a pivotal role in shaping the success of this study. Their contribution not only elevated the research capabilities but also added sophistication to the methodology, enhancing the overall quality and depth of the findings. Intelligent Solutions Inc.'s expertise and support were instrumental.

One of the greatest things about studying at WVU is enriching experience, primarily attributable to the vibrant community and the pristine natural surroundings. Chillin' in the Appalachian Mountains has been a blast – endless adventures from hikes to laid-back boat rides. It's kinda funny how diving into the wilds of WV, hilly terrain of Morgantown's streets, and diving into large datasets, and finding a global minimum during model training share the same vibe – both unexplored territories that totally shaped my academic and personal journey. My run at WVU has been a rollercoaster of learning, teaming up on projects, and making awesome friends. And off the academic grind, whether it was hitting the ping-pong table, playing charades, sand volleyball, or camping out, it's been nothing but good vibes. So, major props to Almost Heaven!

To My Alaska Friends: Thank you to Dr. Abhijit Dandekar, Dr. Shirish Patil, Dr. Mohabbat Ahmadi, Bahareh Barati, Sasan Ghanbari, Dr. Obadare Awoleke, Dr. Catherine Hanks, Dr. Obadare Awoleke, Carol Holz, Sue Wolfe, Grant Shimer, Nicole, Dr. Reza Oskui (RIP), Dr.

Samson Ning, Vahid Nourpour, Yasin Hajizadeh (also in Texas), Asaad Abdollazadeh, Mohammad Shahvali, Jeff Judkins, Dr. Ahmad Jamili, Soodabeh Esmaili, Zeeshan Shaikh, Dr. Jo Mongrain, Rijo Simon, Sumanth, Chidi Igbokwe, Kiran, Sagar, Trupti, Vaibhav Raj, Bahram Novruzaliyev,Chinmay Shukla, Ashish, Karan, Hrishikesh, Prachi, Vivek, Sumanth, Vasil, Ibrahim Adebiyi, Behnam Zanganeh (in Texas as well), Jim Flynn, Masoud Hajizadeh, Bob Jennings, Hugh Nicholson, Jenny Sutton, Maduabuchi Pascal Umekwe, Santosh Patil, Lyazzat DeMallie, John DeMallie, Vanessa Bergstedt (in Texas as well), Vahid Atashbari, SPE Fellowship Award Committee, etc…
#NanookNation

To My Wyoming Friends: Thank you to Mahdi Khishvand, Mohammadhadi Hajilou, Amir Hossein Alizadeh, Mohammad Heshmati, Arsalan Zolfaghari (also in Kansas), Greg Marsh, Bahareh Nojabaei, Peyruz Gasimov, Pejman Tahmasebi, Serveh Kamrava, Morteza Akbarabadi, Vahideh Mirchi, Soheil Saraji, Mohammad Sabti, Nijat Hakimov (also in Kansas), Dr. Peyman Pourafshary, Dr. Jianchao Cai, Michele Ury Foist, Chigozie Aniemena, Ali Darzi, Bingjun, Marwan Hafez, Dr. Mohamed Mostafa, Mohammad(Arash) Norouzzadeh, Mohammad Piri, Morteza Dejam, Arash Hassanzadeh, Richard Parma, Dr. Ali Ghalambor etc…

To My Kansas Friends: Thank you to Dr. Reza Barati, Dr. Masoud Kalantari, Dr. Jyun Syung Tsau, Mohammad Kazemi, Esmail Ansari, Sherifa Cudjoe, Joel Tetteh, Yousef Saeed, Nazirah Mohd, Siyan Liu, Ahmad Ghuloum, Mubarak Alhajeri, Mustafa Alhubail, Rana Al Homoud, Qinwen “Kelly” Fu, Siyan Liu, Bao Jia, Kamran Aghaee, Hooman Hosseinin, Martha Hele Kehr, etc…

To My Texas Friends: Thank you to Javed Iqbal, Usman Ahmed, Nathan Meehan, Yinghui Li, George Vassilellis, Surya Rajan, Kaveh Ahmadi, Hasan Shojaei, Mojtaba Ardali, Erhan Eren, Milad Saidian, Michael Dunkel, Arash Shadravan, Richard Clayton, Mahdi Sadeghi, Oscar Nosrati, Ardeshir Tajvari, Mossy Alavi, K M Hadi Hadipour, Malek Hadipour, Faith Roointan, Martin Wysocki, Sunny Iranpour, Mahdi Parak, Saeed Khorsandi, Mojtaba Pordel, Arash Shadravan, Mo Izadi, Sadegh, Café Caspian, Abouzar Mirzaei-Paiaman, Sachin Ghorpade, Meaghan Yorro, etc…

Last but definitely not the least, To My West Virginia Friends: Thank you to Shahab, Narges Jan, Ghazal, Peter, Luna, Sam, Dr. Kashy, Dr. Mohamed El Sgher, Dr. Qingqing Huang, Ali, Mozhdeh, Diba, Mehrdad, Arya, Maher, Amir, Sina, Yvon, Andrew, Marwan, Mohanned, Abdul, Omar, Saif, Foad Haeri, Yemi Akinkugbe, Vida Gholami, Ali Tarokh, Carla Lynx, Faegheh Javadi, Josh Dalton, Mohannad Elhamod, Dr. Dengliang Gao, Evan Loflin, Dr. Vasudevan Jagannathan, Awadh ALJassasi, Mohanned, Ravi, Shivam, Dr. Li Pang, Dr. Nishant Sharma, Dr. Benjamin Silverberg, Dr. Saira George, Nicolas Fernandez, Andrey Lukyanenko, Majid Torof, Sharon Flanery, Bevin VanGilder, Dr. Shikha Sharma, Mohammad Akyash, Ali Zafari, Asiye Zeytonli, Atefeh Khoshkhahtinat, Banafsheh Adami, and Minou Khayami for making my journey and stay more pleasant. #Let’sGo!

I commemorate this accomplishment in honor of **Dr. Mohabbat Ahmadi**, a true friend who walked alongside me during my journey that commenced in Alaska. Rest In Peace, Mohabbat!

I extend my deepest gratitude to **My Family**, who have been the steadfast pillars of support and inspiration throughout this journey. Their boundless understanding and the sacrifices they have made are the bedrock upon which I have built my academic pursuits. **My Father and Mother** have been the guiding lights in my life, encouraging me to persevere even when the path seemed daunting. This work is a tribute to their unwavering faith and the nurturing environment they created, which was instrumental in bringing this dream to fruition.

Specials thanks to Dr. Jamshid Moghadasi, Dr. Mohammad Takassi, Dr. Bahram Habibnia, Dr. Hossein Ghobadi, Dr. Abdolnabi Hashemi, Dr. Bahman Moselmi, Dr. Riyaz Kharrat, Dr. Reza Shadizadeh, Dr. Ahmad Shariati, Dr. Gholamreza Rashed, Dr. Khalil Shahbazi, Dr. Siavash Ashouri, Dr. Shahin Kord, Dr. Abbas Helalizadeh, Dr. Masoud Bahrami, Dr. Mohsen Pourghafar, Shaban Vardak, Mr. Moradi, whom I learned a lot during my undergraduate studies at PUT in Ahwaz.

Finally, I turn to the words of the Persian poet Rumi, whose wisdom resonates with the essence of this journey: "Out beyond ideas of wrong-doing and right-doing, there is a field. I'll meet you there…".

I am deeply grateful to the United States for providing an environment that has fostered my intellectual growth and shaped me into a better individual. This journey has been enriched by the opportunities and diverse experiences offered by this great nation, for which I am profoundly thankful.

Above all, utmost appreciation to the **Almighty God** for the divine intervention in this academic endeavor; I would not have come this far without his blessing.

Table of Contents

List of Figures

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Problem Statement

Growing evidence supports the idea that climate and the cycles of greenhouse gases are closely linked. Carbon dioxide (CO_2), primarily emitted by human activities, is a major contributor to the greenhouse effect and global warming (IEA 2016). The primary source of anthropogenic CO_2 is the combustion of fossil fuels, which currently dominates global energy production. As there is a widely accepted need to limit the concentration of greenhouse gases, especially CO_2, in the atmosphere, a crucial advancement in the near-to-medium-term solution involves a technique that can reduce CO_2 emissions while still allowing the use of fossil fuels. One proposed solution for this is Carbon Capture and Sequestration (CCS). Geologic CO_2 storage offers a promising solution for reducing global CO_2 emissions **(Figure 1. Commercial-scale** integrated carbon capture and storage (CCS) projects in the USA While there are uncertainties, the vast storage capacity of geologic formations, particularly deep saline formations, suggests that they can accommodate a significant portion of anthropogenic CO_2 **(Table 1).**

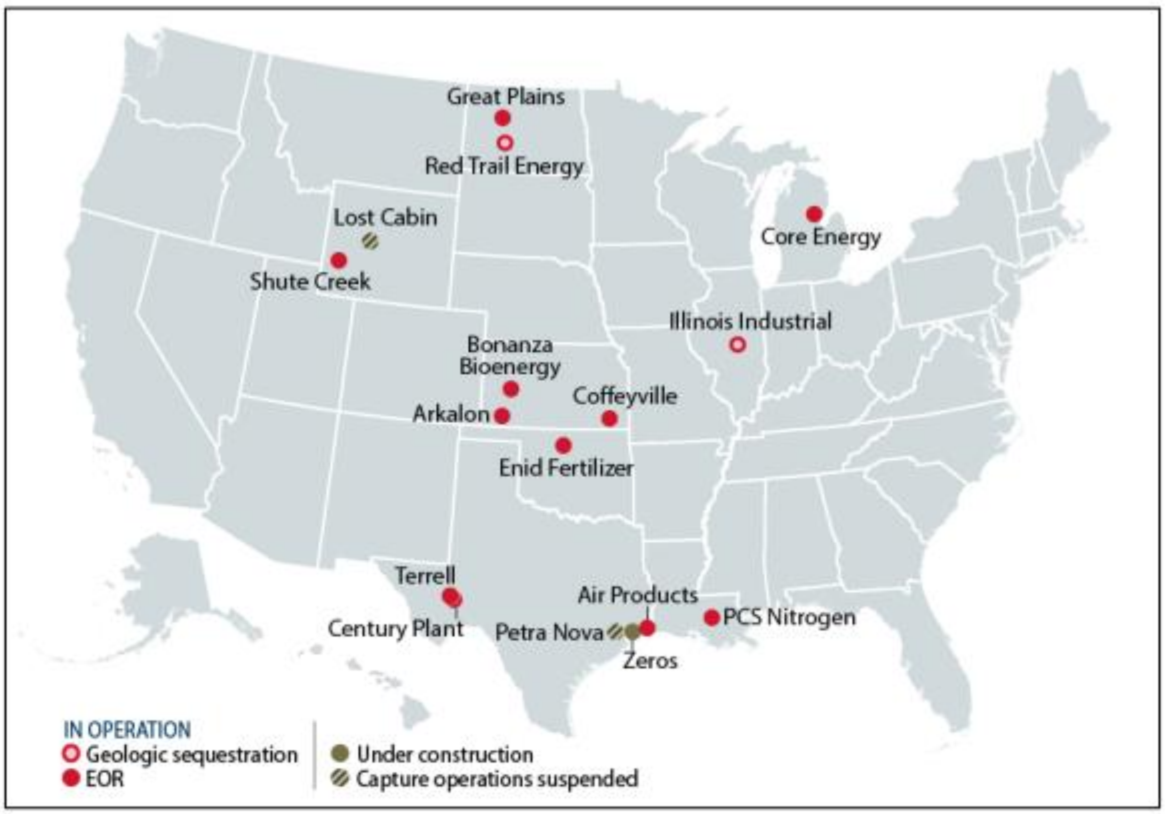


Figure 1. Commercial-scale integrated carbon capture and storage (CCS) projects in the USA (CRS, 2022)

Table 1. Estimates of US CO_2 sequestration (CRS, 2022)

	Oil and Natural Gas Reservoirs	Unmineable Coal	Saline Formations	Total
Low Estimate	186	54	2,379	2,618
Medium Estimate	205	80	8,328	8,613
High Estimate	232	113	21,633	21,978

The storage capacity for CO_2 considering all the sedimentary basins globally is potentially very large, ranging from 2,618 Gt to 21,978 Gt. In the United States alone, the estimated total storage capacity ranges between about 2.6 trillion and 22 trillion metric tons of CO_2. These estimates represent physical restraints on storage, not considering economic or regulatory constraints. The low, medium, and high estimates correspond to calculated probabilities of exceedance of 90%, 50%, and 10%, respectively. When CO_2 is injected into a reservoir, it forms a plume that displaces some of the existing brine (Chadwick et al. 2019; Zhu et al. 2019). This injected CO_2 can either partially or entirely dissolve in the native brine. Additionally, the dynamics of capillary pressure gradients oriented towards the injection point, along with structural and mineral trapping mechanisms play a significant role over extended periods, often spanning decades or centuries, in the storage process **(Figure 2).** The effectiveness of these mechanisms can vary throughout the storage period. Some mechanisms, like structural trapping, may work immediately, while others, like mineral trapping, may take longer to achieve significant impact. The structural trapping depends on the presence of impermeable barriers, while mineral trapping depends on specific geochemical reactions.

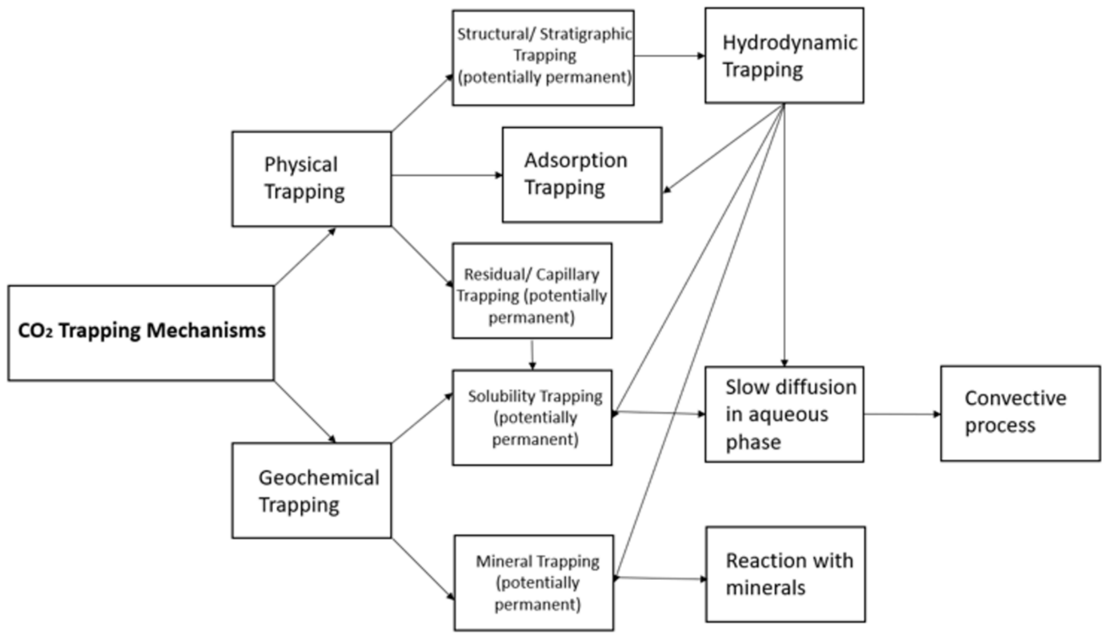


Figure 2. Different CO_2 trapping mechanisms during the geological storage process (modified after Ajayi et al. 2019)

However, for CCS to be a practical strategy for carbon management, a key issue that must be tackled is the risk of CO_2 leakage. In addressing this concern, a critical aspect of evaluating potential sites for geologic carbon sequestration involves determining the pressures associated with CO_2 storage and the extent of the CO_2 plume across the subsurface formation. In the context of carbon sequestration, employing subsurface modeling emerges as the most rapid and cost-effective method to mitigate project risks. This modeling approach should allow for real-time or near real-time) identification of optimal locations, methods, and the quantification of carbon storage capacities across different reservoirs. Saline aquifers, particularly those characterized by their elevated salinity levels, have emerged as a pioneering and innovative reservoir for carbon sequestration.

Conformance, a measure of how closely a product or system adheres to a specified standard, is a critical requirement for CO_2 storage. To demonstrate conformance, storage site operators must demonstrate consistency between predictive models of reservoir performance and monitoring observations. Conformance serves as an indicator that storage processes are well understood and enhances the reliability of long-term predictions. Satisfactory conformance is essential throughout the operational phase and, most importantly, at the end, when responsibility for the site is transferred from the operator to the national authority.

Demonstrating conformance can be challenging due to the inherent limitations in geological understanding, modeling software, and the resolution of monitoring tools. It is virtually impossible to achieve a perfect and unique match between observed and modeled behavior. The objective of this paper is to establish that conformance can be demonstrated by demonstrating that predictive modeling capability systematically improves over time as monitoring data is gradually acquired **(Figure 3).**

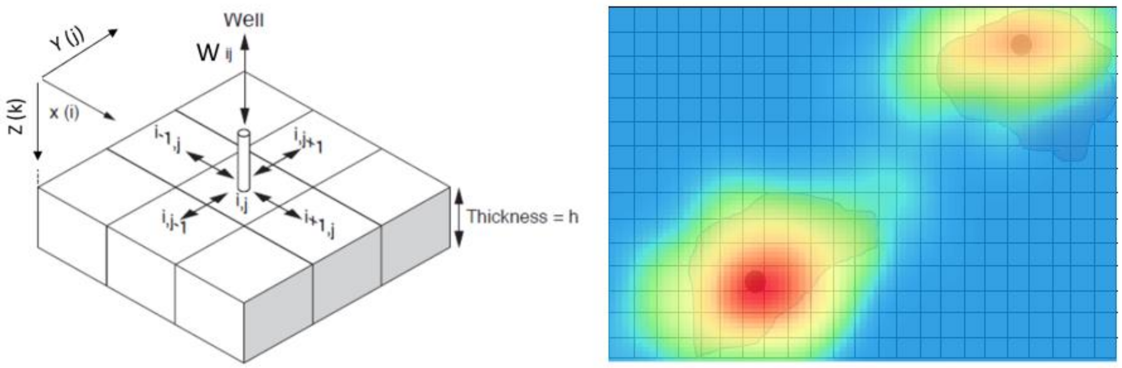


Figure 3. Representation of pressure and CO_2 saturation monitoring (From Alabboodi 2021)

Several CGS projects are currently underway or planned, as shown in **Table 2.** It shows that there are a number of large-scale CCS projects in operation or under construction around the world. These projects capture CO_2 from a variety of sources, including natural gas processing, synthetic gas production, fertiliser production, hydrogen production, and iron and steel production. The CO_2 is then stored in geological formations, such as saline aquifers, depleted oil and gas reservoirs, and unmineable coal seams.

Over the past two decades, there has been a substantial expansion in the array of techniques available for monitoring CO_2, both in deep subsurface and on the surface. The development of numerous shallow monitoring methods has been closely linked to environmental impact assessments, reflecting societal apprehensions about potential near-surface leakage (IEA 2016).

Pressure and saturation monitoring are essential components of a comprehensive risk management strategy for geological sequestration projects. They play crucial roles in plume tracking, validating disposal process physics, ensuring storage security, and meeting regulatory compliance. Integration of monitoring data into reservoir models enhances accuracy and aids in understanding subsurface fluid flow dynamics.

Table 2. Large-scale CCS projects in operation or under construction (data from EIA 2021)

Project	Start Year	Project Type	Tonnes of CO_2 injected per year	Primary storage type	Project Status
Val Verde Natural Gas Plants	1972	Natural gas processing	1.3	EOR	Operational
Enid Fertilizer CO_2-EOR Project	2000	Fertiliser production	0.7	EOR	Operational
Shute Creek Gas Processing Facility	1986	Natural gas processing	7	EOR	Operational
Sleipner CO_2 Storage Project	1996	Natural gas processing	0.9	Dedicated	Operational
Great Plains Synfuel Plant and Weyburn-Midale Project	2000	Synthetic gas	3	EOR	Operational
In Salah CO_2 Storage*	2004	Natural gas processing	1	Dedicated	Operational
Snøhvit CO_2 Storage Project	2008	Natural gas processing	0.7	Dedicated	Operational
Century Plant	2010	Natural gas processing	8.4	EOR	Operational
Air Products Steam Methane Reformer EOR Project	2013	Hydrogen production	1	EOR	Operational
Coffeyville Gasification Plant	2013	Fertiliser production	1	EOR	Operational
Lost Cabin Gas Plant	2013	Natural gas processing	0.9	EOR	Operational
Petrobras Lula Oil Field CCS Project	2013	Natural gas processing	0.7	EOR	Operational
Boundary Dam Carbon Capture and Storage Project	2014	Power generation	1	EOR	Operational
Quest Canada	2015	Hydrogen production	1	Dedicated	Operational
Uthmaniyah CO_2 EOR Demonstration Project	2015	Natural gas processing	0.8	EOR	Operational
Abu Dhabi CCS Project	2016	Iron and steel production	0.8	EOR	Operational
Illinois Industrial Carbon Capture and Storage Project	2017	Chemical production	1	Dedicated	Under Construction
Kemper County Energy Facility	2016	Power generation	3	EOR	Operational
Petra Nova Carbon Capture Project	2016	Power generation	1.4	EOR	Operational
Alberta Carbon Trunk Line ("ACTIL") with Agrium CO_2 stream	2017	Fertiliser production	0.3	EOR	Under Construction
ACTIL with North West Sturgeon Refinery CO_2 stream	2017	Oil refining	1.2	EOR	Under Construction
Gorgon Carbon Dioxide Injection Project	2017	Natural gas processing	3.4	Dedicated	Under Construction

1.1.1 Pressure Monitoring:

Pressure monitoring is a crucial element in the comprehensive risk management approach for geological sequestration projects and for assessing the overall behavior and containment of the injected CO_2 within the formation. Changes in pressure offer information about the behavior of CO_2 within the aquifer for detecting any deviations from the expected pressure profile, which could indicate areas of concern, such as leaks or uneven distribution of CO_2. It becomes particularly essential during the permitting process and public acceptance phase of CO_2 sequestration initiatives. The monitoring process serves various purposes, including tracking the plume's location, validating disposal process physics, and ensuring storage security by detecting potential leaks in abandoned wells or through aquifer seals. Regulatory requirements also necessitate ongoing monitoring to prevent CO_2 leakage into shallow natural resources like groundwater, thereby safeguarding local populations. Additionally, pressure monitoring provides valuable feedback to reservoir simulation studies. Commonly collected pressure and temperature data can be integrated into reservoir models, enhancing accuracy and aiding in the estimation of unknown parameters through history matching.

1.1.2 Saturation Monitoring:

Monitoring carbon dioxide (CO_2) saturation is a pivotal aspect of comprehensive risk management and assessment in geological sequestration projects. Similar to pressure monitoring, saturation monitoring is integral to the permitting process and public acceptance of CO_2 sequestration initiatives as it provides information about the spatial extent and pattern of CO_2 saturation and any unexpected fluid movement. The monitoring process serves multiple purposes, including tracking the spatial distribution and understanding the behavior of the injected CO_2 plume. Validating the physics of the disposal process and ensuring the secure storage of CO_2, especially concerning leak detection in abandoned wells or through aquifer seals, are critical objectives. Additionally, regulatory compliance mandates continuous monitoring to guarantee that CO_2 does not leak into shallow natural resources like groundwater, mitigating potential risks to local populations.

Saturation monitoring complements reservoir simulation studies by providing essential feedback and by identifying areas where the injected CO_2 may not be distributing as expected, allowing for adjustments to injection strategies to optimize storage efficiency. Commonly obtained saturation data, alongside pressure and temperature data, can be integrated into reservoir models. The superior attributes of saline aquifers in comparison to depleted natural gas reserves lie in their capacity to dissolve carbon within the expansive aqueous environment, as opposed to merely storing it, as is the case with depleted wells **(Figure 4).**

1.1.3 New and Emerging Monitoring Techniques:

Various monitoring techniques are employed based on project-specific needs, showcasing the diversity of approaches in different initiatives worldwide. **Table 3** shows a concise overview of various monitoring technologies used in Carbon Capture and Storage (CCS) projects. It demonstrates the diversity of monitoring approaches, including seismic, electromagnetic, geochemical, and satellite-based methods. The diverse set of monitoring technologies in the table aligns with the need for robust and diverse data inputs for CO_2 sequestration modeling. The application of advanced technologies, such as Fiber-Optic sensing coupled with Artificial Intelligence and Machine Learning, as demonstrated in the study by Aboaba (2022), could offer a transformative solution. This approach which is also based on AI-based monitoring could be used to allow for continuous, real-time monitoring of CO_2 saturation and pressure in saline aquifers, addressing the limitations of snapshot-based conventional methods and providing valuable insights for optimized sequestration strategies.

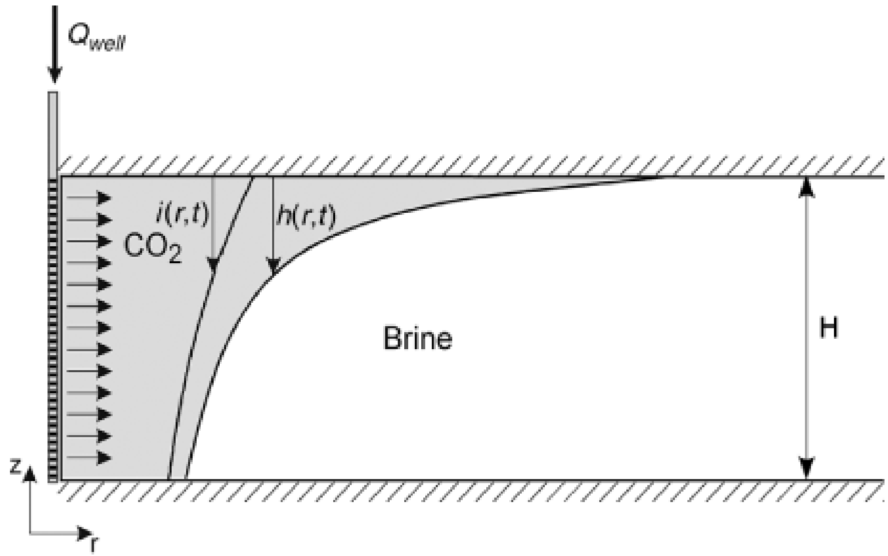


Figure 4. Schematic of the CO_2 injection front, with thickness of the CO_2 denoted by h(r,t), and the drying front, with thickness denoted by i(r,t). (Nordbotten and Celia 2006).

Mawalkar et al. (2021) conducted a comprehensive study as part of the Midwest Regional Carbon Sequestration Partnership (MRCSP) large-scale CO_2 storage test in a depleted Niagaran pinnacle reef oil reservoir in Northern Michigan. The research, an integral component of the Michigan Basin Large-Scale Injection Project, aimed to assess the technical potential and dynamics of carbon capture, utilization, and storage. The study employed a Distributed Temperature Sensing (DTS) system in conjunction with a five-level behind-casing pressure sensing array to monitor the injection of approximately 101,000 metric tons of CO_2 between January 2017 and December 2018. The results highlighted the versatility of DTS data, demonstrating its utility in warmback analysis for identifying CO_2 migration patterns and corroborating zonal isolation within the wellbore. This study contributes practical applications of DTS technology in the context of CO_2 injection, offering insights into reservoir dynamics for enhanced oil recovery (EOR) strategies.

1.2 Research Objective

To address challenges related to modeling and monitoring of CO_2 pressure and saturation, we propose the use of AI/ML-based data-driven proxy models, specifically artificial neural networks, to predict CO_2 pressure and saturation under several complexity schemes. Employing advanced methodologies, including data augmentation, the research aims to create a practical workflow empowering reservoir engineering and management teams to make informed decisions, particularly in CO_2 sequestration projects. In the realm of CO_2 sequestration, numerical reservoir models, built upon geological models, inherently harbor uncertainty due to the unknown ground truth. This uncertainty prompts the introduction of uncertainty quantification in reservoir simulation. Given the proven efficiency of the Smart Proxy Model in accurately simulating reservoir conditions at a reduced computational cost, it could be applied to expedite uncertainty

Table 3.Various monitoring technologies used in Carbon Capture and Storage (CCS) projects

Monitoring Technology	Advantages	Reference
3D seismic	Provides a three-dimensional image of geological structures and CO_2 plume migration	Ringrose, P.S., et al. (2013)
4D seismic	Offers benefits for overburden imaging and time-lapse responses with an improved acquisition plan	Ringrose, P.S., et al. (2013)
Microseismic	Useful for monitoring geomechanical response to injection	Oye, V., et al. (2013)
Vertical seismic profiling	Provides valuable information on geological structure details	Götz, J., et al. (2018)
Gravimetry	Beneficial for evaluating formation fluids density and CO_2 plume	Kabirzadeh, H., et al. (2017)
Cross-hole electromagnetic	Advantageous for detection and monitoring of CO_2 location	Carcione, J.M., et al. (2012)
Pressure and temperature monitoring	Provides direct information for evaluating reservoir stability	Liebscher, A., et al. (2013)
Geochemical sampling	Crucial for establishing a useful baseline in groundwater hydrology	Boreham, C., et al. (2011)
Soil and gas sampling	Provides more data on natural CO_2 variations and associated fluctuations	Ringrose, P.S., et al. (2013)
Tracers	A valuable, cost-effective method for monitoring CO_2 origin at wells and in the storage complex	Ringrose, P.S., et al. (2013)
Atmospheric monitoring	Useful for identifying anomalies above the natural baseline	Etheridge, D., et al. (2011)
Microbiology	Offers valuable data on biogeochemical processes affecting CO_2 diffusion.	Morozova, D., et al. (2011)
Core analysis	Essential for obtaining petrophysical data and rock mechanical properties	Ringrose, P.S., et al. (2013)
Satellite monitoring	Valuable, cost-effective data for onshore CO_2 injection operation	Etheridge, D., et al. (2011)
Distributed temperature sensing technology	Provides high-resolution information on CO_2 migration in the reservoir	Mawalkar, S., et al. (2019)

quantification analysis and simulate CO_2 injection into underground formations under different operational and geological settings. This study aims to create a practical workflow for reservoir engineering and management teams, particularly involved in CO_2 sequestration projects. The outlined augmentation workflow involves generating diverse geological realizations, designing varying numbers of wells, implementing CO_2 injection under different operational constraints, and varying injection times, and other static and dynamic uncertainties. There are five phases of complexity that are designed and incorporated into this study. They range from changing the static and dynamic reservoir characteristics, varying the number and location of injection wells, varying time of injection between the injection wells, and changing the injection well patterns in the aquifer.

1.3 Dissertation structure

In the absence of sufficient subsurface CO_2 sequestration data, this research created realistic and yet complex data sets models at a resolution of each grid-cell, specifically tailored to research and applied objectives, facilitating the exploration of various operational scenarios and dynamic conditions.

Chapter 2 introduces numerical reservoir simulation, the history of such technology, limitations and benefits. This chapter also lays the required background for building the simulation model, and the data extraction methods used in the remaining of this study.

Chapter 3 introduces artificial neural network relevant to this study. This chapter introduces the concept of AI and ML techniques for developing effective smart proxy models. the artificial neural network as one of the AL algorithms, its design and architecture, and its role in the domain-based modeling of CO_2 sequestration projects. Methods such as training, calibration, and blind validation are crucial for achieving expected results. It also provides literature reviews, highlighting their applications.

Chapter 4 focuses on smart proxy modeling, data collection, complex phases definitions, methodology, detailing the procedures and steps undertaken to design, and select geological realizations data for each phase of complexity, preparation of spatio-temporal database for accurate predictions by the developed smart proxy. The chapter ends by feature generation and is focused on the role of domain expertise in the AI-based modeling of CO_2 sequestration projects.

Chapter 5 provides the results for the blind datasets. The blind dataset is that part of the data which has never been seen by the smart proxy model during its training process. Diverse datasets are utilized to produce comprehensible results, presented through figures.

Chapter 6 provides a summary of the entire research study, outlines key conclusions, and offers recommendations for future research in similar domains. It aims to enhance and guide future research efforts on the critical topic studied, contributing to the improvement of research endeavors in the future.

Chapters 7 and 8 provide the Appendix and the list of references, respectively.

Chapter 2 – Numerical Reservoir Simulation

2.1 Introduction to Numerical Reservoir Simulation

The ultimate goal of constructing a numerical reservoir simulation model is to achieve a process known as "history matching." History matching is a critical and indispensable step in reservoir engineering and modeling because it allows engineers and geoscientists to fine-tune the model's parameters and characteristics to closely replicate the actual historical performance of the reservoir, including production rates, pressure changes, and other observable data. This matching process serves as a validation tool, ensuring that the numerical model accurately represents the physical behavior of the reservoir. Achieving a successful history match is essential for gaining confidence in the model's predictive capabilities. By comparing the model's simulated historical performance with real-world data, engineers can make informed decisions regarding reservoir management, optimize production strategies, and plan for future development with a high degree of reliability and accuracy. In essence, history matching is the linchpin that bridges the gap between theoretical modeling and practical reservoir management, making it a crucial objective in the field of reservoir engineering. The rate of return on investment (ROI) for subsurface energy technologies can vary significantly depending on various factors, including the specific technology, project scale, market conditions, regulatory environment, and overall project execution **(Figure 5).**

Figure 5. The ultimate objective of any multi-disciplinary reservoir study is value addition

When evaluating the ROI for subsurface energy technologies several key points must be taken into consideration:

Technology Efficiency and Performance: The efficiency and performance of the subsurface energy technology play a crucial role in determining the ROI. Technologies that can extract or utilize subsurface energy resources with higher efficiency and effectiveness tend to offer better financial returns. Factors such as energy conversion efficiency, operational reliability, and productivity directly impact the financial viability of the technology.

Resource Potential and Availability: The quantity and quality of the subsurface energy resource, such as oil, gas, geothermal heat, or underground storage capacity, significantly influence the ROI. Larger and higher-quality resources generally offer greater revenue potential. The accessibility and availability of the resource, including factors like depth, permeability, and reservoir characteristics, also impact the economic feasibility and ROI of the technology.

Market Conditions and Pricing: The prevailing market conditions and energy prices play a crucial role in determining the financial viability of subsurface energy technologies. Fluctuations in energy prices, supply-demand dynamics, and market regulations can impact revenue generation and profitability. A favorable market with stable or increasing energy prices can enhance the ROI, while volatile or declining prices may pose challenges to the financial success of the project.

Capital Investment and Operational Costs: The capital investment required to deploy and operate the subsurface energy technology, as well as ongoing operational costs, significantly affect the ROI. Higher capital costs may lead to a longer payback period and lower overall returns. However, advancements in technology, economies of scale, and operational efficiencies can help reduce costs and improve the ROI.

Project Scale and Duration: The scale and duration of the subsurface energy project also impact the ROI. Larger projects with significant resource potential and longer operational lifespans often have the potential for higher returns, as they can generate sustained revenue over an extended period. However, smaller-scale projects with quicker payback periods may offer more immediate returns.

Regulatory and Policy Environment: The regulatory and policy environment, including government incentives, tax benefits, and support mechanisms, can influence the financial attractiveness and ROI of subsurface energy technologies. Favorable policies that encourage the adoption and development of clean energy sources, promote energy efficiency, or provide financial incentives can enhance the ROI by reducing project costs or improving revenue streams.

Project Risks and Uncertainties: Subsurface energy projects are associated with various risks and uncertainties, including geological uncertainties, technical challenges, environmental considerations, and regulatory risks. Accurate assessment and mitigation of these risks are essential to safeguard the ROI. Comprehensive risk analysis, contingency planning, and adequate project management strategies can help minimize risks and enhance the financial success of the project.

2.1.1 Definition and Purpose

Numerical reservoir simulation (NRS) is a computational technique employed to model the complex dynamics of fluid flow within subsurface reservoirs. At its core, it leverages mathematical equations and computational algorithms to simulate the behavior of fluids, such as carbon dioxide, as they move through porous rock formations beneath the Earth's surface. By integrating geological, physical, and fluid flow data, numerical reservoir simulation provides a comprehensive understanding of the intricate processes governing subsurface fluid movements. This method plays a pivotal role in the field of reservoir engineering, aiding researchers and practitioners in making informed decisions related to resource extraction, environmental impact assessments, and, significantly, carbon sequestration initiatives.

Furthermore, reservoir simulation plays a crucial role in enhancing our comprehension of the physical and chemical processes influencing subsurface fluid flow. This understanding is essential for mitigating environmental risks, optimizing injection parameters, and ensuring the long-term success and safety of carbon sequestration initiatives. As numerical reservoir simulation continues to evolve, its application in carbon sequestration becomes increasingly sophisticated, allowing for more accurate predictions and informed decision-making in the pursuit of sustainable and effective carbon management strategies. The reservoir simulation process entails the creation of both static and dynamic models. The static model encapsulates the reservoir's geological properties, such as porosity, permeability, and rock type, while the dynamic model simulates the fluid flow over time, incorporating injection/ production history, fluid properties, and reservoir management strategies. Upon integration of these models, the simulator partitions the reservoir into millions of grid blocks **Figure 6,** each representing a small

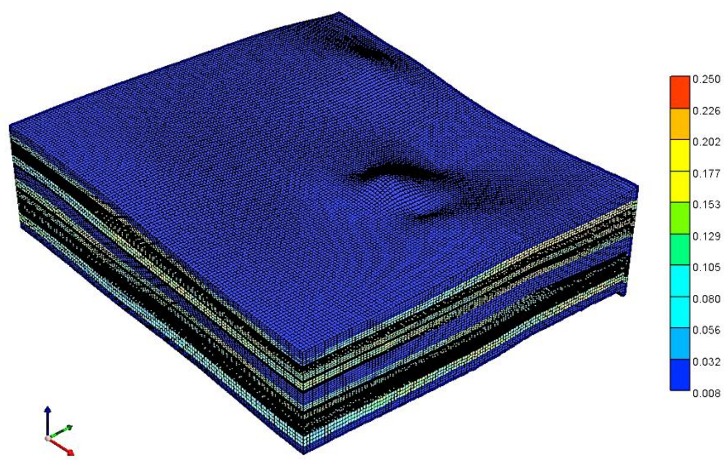


Figure 6. A reservoir model with millions of grid blocks

segment of the reservoir. The simulation then calculates the pressure, and fluid saturation within each block at each time step, providing a comprehensive understanding of the reservoir's behavior.

Reservoir simulation models are therefore crucial for optimizing production, evaluating enhanced oil recovery techniques, and assessing the impact of drilling new wells. They serve as a critical tool for the oil and gas industry, enabling informed decision-making that maximizes the recovery of valuable hydrocarbons.

2.1.2 Relevance to Carbon Sequestration

In the context of carbon sequestration, the primary objectives of employing numerical reservoir simulation are manifold. One key goal is to model and predict the behavior of injected carbon dioxide within subsurface formations, assessing its migration, distribution, and interaction with geological structures over time. Simulation models enable researchers to evaluate the effectiveness of various injection strategies, optimizing the sequestration process to achieve maximal carbon storage. Additionally, these models contribute to understanding the potential risks associated with sequestration, including the identification of potential leakage pathways and the assessment of long-term storage stability.

Positioned at the forefront of carbon management strategies, sequestration relies on the meticulous injection and storage of carbon dioxide within subsurface reservoirs. The significance of numerical reservoir simulation becomes particularly pronounced in this context, providing a virtual laboratory to assess various injection strategies, predict the migration patterns of carbon dioxide **(Figure 7),** and evaluate the long-term stability of storage formations. By simulating the dynamic interplay between geological structures

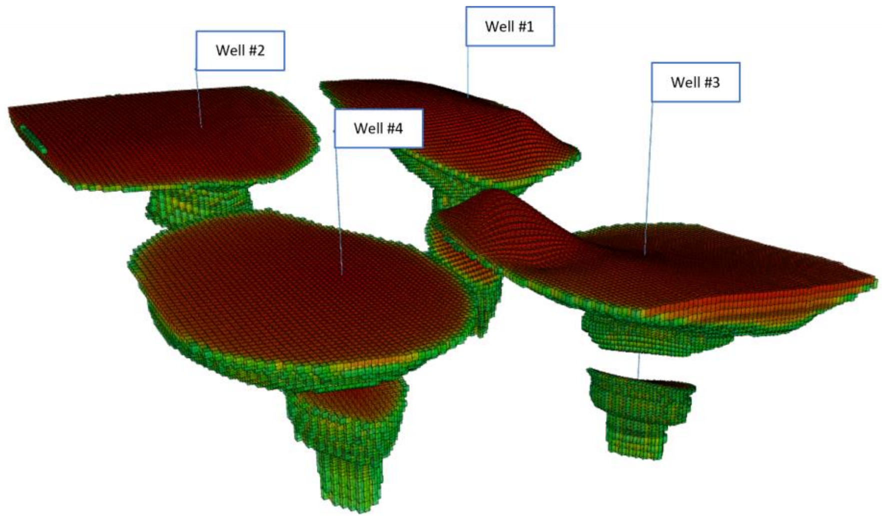


Figure 7. A depiction of CO_2 plumes in an aquifer model with four injection wells

and fluid flow, this tool guides informed decision-making, enriching the efficiency and reliability of carbon sequestration efforts.

The nexus between numerical reservoir simulation and the effective management of carbon sequestration projects is intrinsic to its capacity to inform decisions at every stage of project development. From the initial planning and design phase to the continual monitoring and optimization of injection operations, simulation models serve as invaluable guides. Accurate predictions regarding the behavior of injected carbon dioxide empower the optimization of injection parameters, minimizing environmental risks while maximizing storage capacity. This connection underscores the indispensable role of simulation models in comprehending subsurface fluid flow dynamics and propelling the efficiency and efficacy of carbon sequestration endeavors. By leveraging mathematical equations and computational algorithms, numerical simulation models provide a holistic understanding of how fluids, particularly carbon dioxide in the context of sequestration, traverse porous rock formations beneath the Earth's surface. This enhanced understanding extends to the characterization of reservoir behavior under diverse conditions, allowing researchers and practitioners to explore a spectrum of scenarios. Simulation models enable the examination of fluid migration, pressure distribution, and interactions with geological structures, shedding light on the complex interplay of factors influencing subsurface dynamics. Through this process, numerical reservoir simulations have become an invaluable tool for unraveling the complexities of fluid behavior in reservoirs, offering insights essential for informed decision-making in carbon sequestration projects.

2.1.3 Optimization of Carbon Sequestration:

The ability to predict and evaluate the consequences of different strategies empowers decision-makers with the knowledge needed to enhance the overall efficiency and success of carbon sequestration projects. Through systematic optimization facilitated by numerical reservoir simulation, the potential for mitigating environmental impact and ensuring the sustainability of carbon sequestration initiatives is significantly amplified.

2.2 History of Numerical Reservoir Simulation

2.2.1 Early Developments

Early developments in numerical reservoir simulation mark a transformative phase in the evolution of subsurface modeling, precipitating advancements that continue to shape contemporary reservoir engineering practices **(Figure 8).** The origins of numerical reservoir simulation trace back to pioneering efforts in the mid-20th century.

During the 1950s to 1970s, the evolution of numerical reservoir simulation was marked by significant progress in modeling techniques, with a focus on two or three dimensions, simple geometry, and the simulation of black oil fluid behavior. In this era, reservoir simulation models transitioned from simplistic one-dimensional representations to more realistic two or three-dimensional frameworks, allowing engineers to capture the spatial distribution of reservoir properties and fluid flow dynamics more accurately.

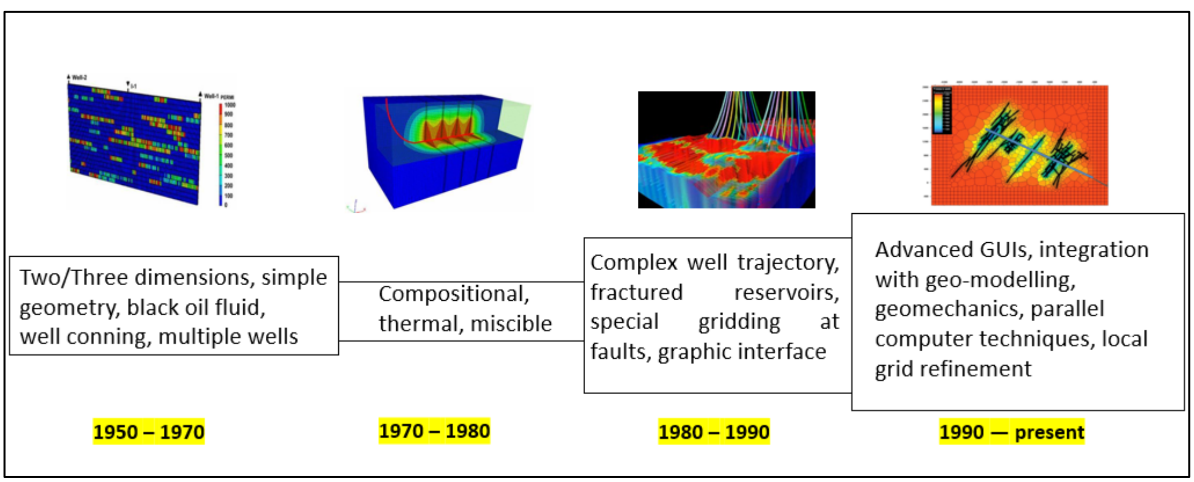


Figure 8. History of Reservoir Simulation in a nutshell

In numerical simulation and modeling, the functional relationships utilized in the equation mentioned earlier incorporate fundamental principles such as the law of conservation of mass, Darcy's law (or Fick's law of diffusion when applicable), thermodynamics, and energy conservation (especially when modeling thermal recovery processes). These functional relationships are considered to be accurate, deterministic, and unalterable. Consequently, if the production results obtained from numerical simulation and modeling do not align with the observations (measurements) collected from the actual field, it leads us to the conclusion that there may be inaccuracies or shortcomings in the characterization of reservoir properties, represented by the static model. Therefore, it becomes necessary to make adjustments to the reservoir characteristics in order to achieve a match between the simulated and observed data.

This represents the widely accepted approach that has been consistently followed for several decades. There is broad consensus regarding the validity and application of this technology. However, it's important to note that this functional framework has evolved significantly from its origins, which were based on simpler principles in the early days of reservoir simulation, primarily involving single-phase flow and Darcy's law. It has since evolved into a much more intricate system of relationships, allowing for the modeling of greater complexities within reservoirs, including multi-phase flow, dual porosity formulations, compositional considerations, integration with geomechanics and surface facilities, and more. These relationships are expected to continue evolving as our understanding of these physical phenomena deepens.

The capabilities of Artificial Intelligence & Data Mining (AI&DM) in pattern recognition can serve various roles in assisting engineers and geoscientists in the development of more efficient and effective reservoir simulation models. To provide context for these new AI-based workflows, let's briefly summarize reservoir simulation and modeling as a process that ultimately represents production from a field (comprising multiple wells) as a function of reservoir and fluid properties, operational constraints, and other variables, using the following formulation:

$$q = f(x1, x2, ..., xn, y1, y2, ..., yn, w1, w2, ..., wn)$$

Where:

- **q** = production from the reservoir
- **x1, x2, ..., xn** = reservoir fluid characteristics
- **y1, y2, ..., yn** = operational constraints (such as drilling new wells, water injection, well shutdowns, adjusting surface facility capacity, etc.)
- **w1, w2, ..., wn** = other parameters (e.g., well configurations, completion techniques, etc.)
- **f** = functional relationship

The equation above simply states that production from a field is modeled using a set of functional relationships that account for reservoir fluid characteristics, operational constraints, and other variables like well configurations and completion techniques. This formulation is applicable to both numerical reservoir simulation and AI-based modeling. In both of these modeling techniques, the goal is to model production as a function of reservoir-fluid properties, well characteristics, and operational constraints. The primary distinction between these two approaches lies in the philosophy regarding our understanding of the underlying phenomenon (fluid flow in porous media) and the assumptions made during the modeling process **(Table 4).**

Table 4. Main difference between numerical reservoir simulation & modeling and AI-based reservoir modeling (from Shahab Mohaghegh)

	Numerical Models	**AI-Based Models**
Reservoir Characteristics	Uncertain: • Measurements • Interpretations (subject to modification during the history matching)	Uncertain: • Measurements • Interpretations (subject to modification during the history matching)
Functional Relationships	Certain: • Conservation of Mass • Darcy's Law (unchanged during the history matching	Uncertain: • Relationship between reservoir characteristics and production or injection (subject to modification during the history matching)

In the context of history matching for numerical reservoir simulation models, where the functional relationships remain constant and unalterable (indicating that our current comprehension of physical phenomena is sufficient and doesn't require modification regardless of the reservoir being modeled), engineers focus on adjusting reservoir characterization, such as permeability, to achieve a reasonable match. Reservoir characterization is represented through a geo-cellular (static) model developed by geoscientists, and it often involves interpretations and uncertain values. Engineers feel comfortable making adjustments to these parameters in order to achieve a satisfactory match. It's important to note that this approach is not criticized but rather explained to highlight the distinctions between traditional methods and AI-based reservoir modeling.

Sattari et al. (2023) conducted a comprehensive literature review addressing the constraints of traditional numerical reservoir simulation in unconventional shale plays. The study proposes the integration of data-driven artificial intelligence (AI) models to overcome limitations associated with **simplifying assumptions**, time constraints, and data quality. The research compares traditional reservoir simulations with AI-based Shale Analytics, aiming to provide guidelines for incorporating AI models into reservoir simulation, improve efficiency, and optimize production strategies in unconventional shale plays. The findings enhanced the understanding of reservoir simulation processes, particularly in the context of the Marcellus Shale and shale gas production.

The unique strength of SPM lies in its capacity to faithfully duplicate pressure and saturation distributions throughout the reservoir at the grid block level and each time step without compromising the physics or resolution of the original numerical simulation model.

Quintero's (2022) investigated the application of Rate Transient Analysis (RTA) in unconventional shale reservoirs. The research primarily focused on the methodology that combines assumptions with history matching to create reservoir models. The study utilized the workflow provided by IHS Harmony, a well-known RTA model software, to demonstrate potential user bias and the variability in model results across multiple wells. The analytical models, including the Blasingame Fracture Typecurve, Agarwal Fracture Typecurve, and Wattenbarger Typecurve, served as guides in the analysis. Subsequently, analytical models such as URM (Unconventional Reservoir Module) Superposition Time and FMB (Flowing Material Balance) were employed to derive drainage area and matrix permeability for the numerical model. Finally, the Multiphase Numerical Model tool was used to achieve a model match by iterating on fracture half-length and dimensionless fracture conductivity. A significant emphasis was placed on the presence of **user bias**, highlighting cases where reservoir engineering assumptions, such as fracture efficiency, resulted in models of comparable accuracy despite significant variations.

As the industry grappled with the challenges of managing multiple wells in a reservoir, the 1950s to 1970s witnessed the inclusion of multiple well representations in simulation models. This allowed for the assessment of well interference, production strategies, and the overall optimization of field development plans. While the computational resources were limited compared to today, these early efforts laid the foundation for the subsequent advancements in reservoir simulation, setting the stage for the more sophisticated models used in modern reservoir engineering.

These seminal investigations laid the groundwork for subsequent developments in reservoir simulation methodologies. Concurrently, technological innovations such as the advent of digital computers played a pivotal role in facilitating the computational complexities inherent in simulating fluid flow within geological formations (Aziz and Settari, 1979). Key milestones emerged as numerical reservoir simulation transitioned from its nascent stages to a more sophisticated and comprehensive discipline. The introduction of the pioneering three-dimensional reservoir simulator, ECLIPSE, by Schlumberger in the early 1970s marked a watershed moment, offering enhanced capabilities for modeling complex geological structures and fluid interactions. This milestone heralded a new era in reservoir simulation, enabling engineers to consider the intricacies of three-dimensional reservoir behavior in their analyses. The subsequent integration of compositional modeling techniques further expanded the scope of reservoir simulation, allowing for a more accurate representation of multiphase fluid behavior and compositional variations within subsurface formations. These breakthroughs set the stage for the development of modern simulation techniques, underscoring the iterative nature of advancements in numerical reservoir simulation that continuously refines our understanding of subsurface fluid flow dynamics (Aziz and Settari, 1979).

2.2.2 Evolution and Technological Advances

- The evolution of NRS has been marked by a transformative journey through the decades, reflecting the continuous quest for enhanced understanding and computational capabilities. In the early stages, reservoir simulation primarily involved rudimentary mathematical models that aimed to capture basic fluid flow behaviors in subsurface formations. These initial models often relied on simplified representations of reservoir properties, emphasizing fundamental principles of fluid dynamics. Over the years, advancements in computational technology and reservoir engineering led to the development of more sophisticated numerical methods capable of simulating complex geological structures and fluid interactions. The transition from one-dimensional to three-dimensional models marked a significant milestone, allowing for a more accurate representation of reservoir heterogeneity and improving the fidelity of simulation results. As computing power increased and algorithms became more refined, numerical reservoir simulation evolved into a robust tool capable of simulating intricate multiphase fluid flow phenomena, providing invaluable insights for reservoir management and optimization.

- The field of NRS has witnessed remarkable technological advancements and innovations that have revolutionized the way reservoir engineers approach complex fluid dynamics problems. The integration of advanced mathematical models, such as compositional fluid models and black oil models, has enabled more accurate representation of phase behavior phenomena, essential for simulating processes like enhanced oil recovery (EOR) or geological carbon sequestration. The utilization of high-performance computing (HPC) techniques, including multi-threading, multi-core processing, and cloud computing, has significantly accelerated simulation times and enabled the handling of large and complex reservoir models.

The evolution of numerical reservoir simulation has been driven by a number of technological advances, including:

- Increased computational power: The development of more powerful computers has made it possible to run more complex and detailed reservoir simulation models.
- Advanced mathematical techniques: The development of new mathematical techniques, such as adaptive time stepping and unstructured grids, has improved the accuracy and efficiency of reservoir simulation models.
- Improved understanding of reservoir physics: The development of a better understanding of reservoir physics, such as multiphase flow and rock-fluid interactions, has led to the development of more accurate and realistic reservoir simulation models.
- Advancements in data acquisition and interpretation: The availability of more accurate and detailed geological, geophysical, and well log data has made it possible to create more realistic reservoir models.

These technological advances have made numerical reservoir simulation an essential tool for the oil and gas industry, enabling the development of more efficient and profitable production strategies. The complexity of reservoir system management and the cycles of complexification in numerical reservoir simulation reflect the iterative and evolving nature of technology, model development, and the understanding, history matching and troubleshooting of subsurface reservoirs. The historical trajectory often involves the invention and application of technology, leading to the creation of complex models to represent the intricate dynamics of fluid flow within reservoirs. As these complex models are employed, they inevitably raise new questions and challenges, creating a need for more advanced technologies and methodologies to address the emerging complexities.

The initial phase of technology invention and application marks a period of simplicity, where early reservoir models capture the fundamental aspects of fluid flow. However, as the industry delves deeper into understanding reservoir behavior and optimizing production, these models become more intricate, incorporating additional parameters and features to enhance accuracy and realism. This phase of complexification results from the continuous pursuit of improved representations of subsurface processes. With the deployment of these advanced technologies, new questions and uncertainties emerge. The increased complexity of models necessitates a deeper understanding of the underlying reservoir physics and geology. Researchers and engineers are compelled to address these uncertainties, leading to the development of even more sophisticated technologies, methodologies, and models. This cyclical process continues as each advancement begets new challenges. The pursuit of accuracy, efficiency, and comprehensive reservoir management often leads to an ongoing cycle of innovation, complexity, and the quest for answers. Each iteration of complexification is driven by the desire to bridge gaps in knowledge and improve the predictive capabilities of numerical reservoir simulation models **(**

Figure 9).

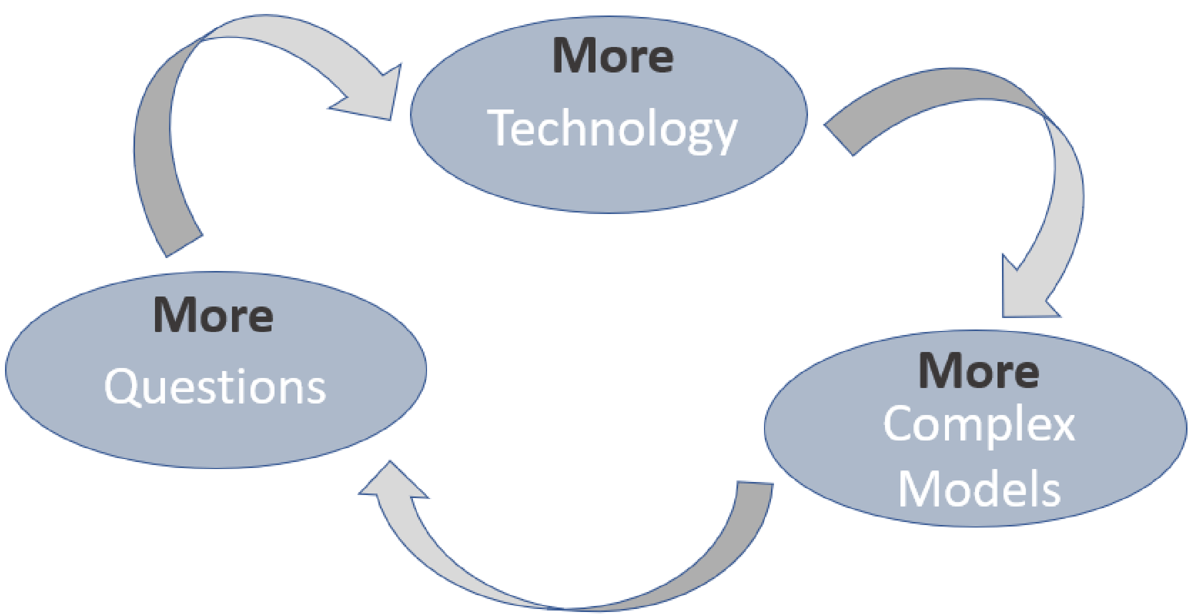


Figure 9. Cycles of complexification

In AI-based reservoir modeling, instead of keeping the functional relationships constant, they are allowed to change, in addition to the possibility of modifying reservoir characteristics. In other words, the approach avoids rigid, deterministic, and inflexible functional relationships between production and reservoir characteristics. Instead, it seeks the functional relationship that generates observed production from the reservoir using a set of measured reservoir characteristics through AI and data mining-based pattern recognition technology. Reservoir characteristics can still be modified if a different set of measurements is believed to be more accurate. Once a set of reservoir characteristics that geoscientists find reasonably reliable is identified, they are not altered during the history matching process. Instead, the functional relationships are adjusted until a match is achieved.

2.3 Limitations of Numerical Reservoir Simulation

2.3.1 Assumptions and Simplifications

NRS, while a powerful tool, operates within a framework of assumptions and simplifications that influence the accuracy and scope of its predictions. One primary assumption lies in the homogeneity of reservoir properties, wherein the model often treats the subsurface grids as uniform in terms of porosity, permeability, and rock type. As the refinement of the grids is increased, the computational cost is also significantly increases. The temporal and spatial discretization inherent in NRS introduces another layer of assumptions. The model has to divide the reservoir into discrete grid blocks, assuming uniform properties within each block. The time-stepping approach, while crucial for computational efficiency, assumes that reservoir conditions change uniformly across the entire block during each time step. These discretization assumptions may lead to an oversimplification of transient effects, impacting the precision of predictions, especially in areas with rapid changes in fluid dynamics or geological properties. Understanding these inherent assumptions is crucial for interpreting simulation results and acknowledging the trade-off between computational efficiency and the complexity of real-world reservoir dynamics. This simplification is a necessity to streamline computational complexity but can lead to deviations from real-world reservoir heterogeneity.

2.3.2 Computational Challenges

In numerical reservoir simulation, the question is not whether, but how and how much. The complexity of the questions being asked, and the amount and reliability of the data available, must determine the sophistication of the system to be used. In NRS, the fundamental inquiry shifts from a binary consideration of whether a phenomenon occurs to a nuanced exploration of how and to what extent it manifests. This shift is driven by the recognition that subsurface fluid flow within reservoirs is a highly intricate process influenced by numerous factors. The emphasis on "how" delves into the mechanisms and intricacies of fluid behavior, acknowledging that it is not merely about the presence or absence of phenomena, but about understanding the underlying processes governing fluid movement. Moreover, the consideration of "how much" extends beyond the binary nature of earlier questions. Instead of merely determining the occurrence or non-occurrence of events, reservoir simulation seeks to quantify the magnitude, impact, and distribution of various factors influencing fluid flow. This quantitative approach is crucial for accurate predictions and optimal decision-making in reservoir management.

The complexity of the questions being posed in NRS is intimately linked to the wealth of available data and the reliability of that data **(Figure 10).** As questions become more intricate and detailed, demanding a deeper understanding of reservoir behavior, the sophistication of the simulation system must align with these complexities (Celia and Nordbotten, 2011; Kohshour et al. 2013). It's an acknowledgment that a one-size-fits-all approach is insufficient in reservoir simulation; the system's sophistication should be tailored to the specific questions at hand and the quality of data accessible for analysis.

Therefore, the choice of reservoir simulation models and methodologies is intricately tied to the intricacy of the questions posed, creating a dynamic relationship where advancements in technology and data acquisition enable more sophisticated inquiries, and, in turn, the need for more sophisticated simulation systems emerges. This iterative process underscores the continuous evolution and refinement of reservoir simulation to meet the ever-growing demands of understanding and managing subsurface reservoirs.

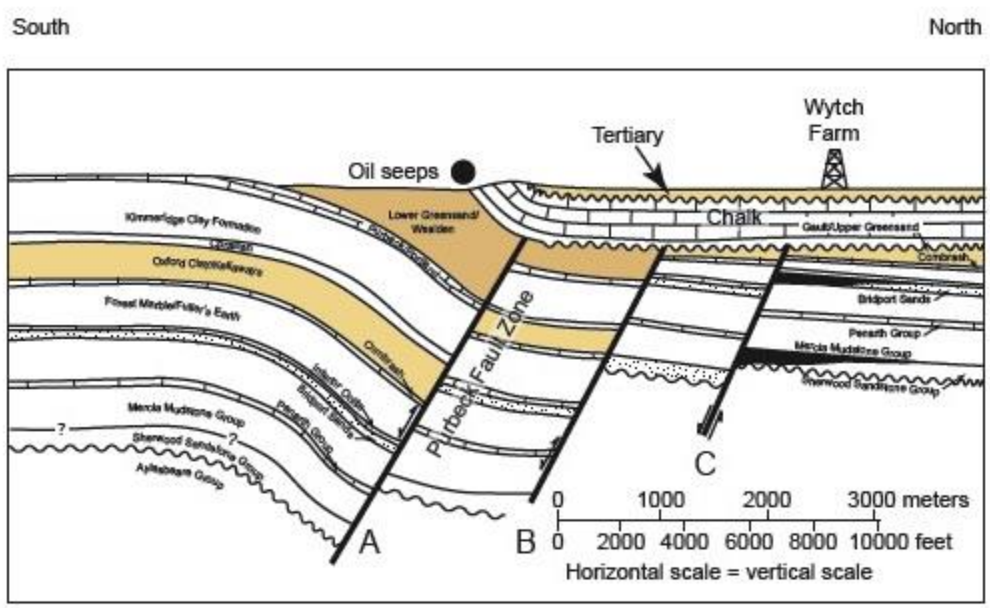


Figure 10. A reservoir system can get as complex as we want (Underhill, 1998)

Numerical reservoir simulations, renowned for their capability to model intricate fluid dynamics in subsurface reservoirs, confront formidable computational challenges that have become increasingly evident with the growing complexity and scale of contemporary models. One primary challenge stems from the computational demand posed by the sheer size and resolution of reservoir models. As these models evolve to depict more realistic geological features, such as intricate fault systems and heterogeneous rock properties, the number of grid cells within the simulation domain significantly escalates. This increase, often reaching millions or even tens of millions of cells, necessitates substantial computational resources and processing power. The utilization of high-performance computing (HPC) systems becomes imperative to navigate the intricacies of these large-scale simulations, raising concerns about accessibility and affordability for researchers and practitioners.

Another significant computational challenge is associated with simulating multiphase fluid flow in porous media. Reservoirs frequently exhibit complex behaviors involving multiple phases, such as oil, gas, and water, each with distinct properties and interactions. Precisely capturing the dynamics of these multiphase flows requires solving intricate sets of partial differential equations, introducing a high level of computational complexity. The quest for efficient and accurate numerical algorithms capable of handling multiphase flow phenomena remains a forefront challenge in advancing reservoir simulation capabilities.

Furthermore, the integration of different modeling components, such as coupling reservoir simulation with geomechanical or thermal processes, adds another layer of complexity. These interdisciplinary simulations, aimed at capturing the coupled behavior of various subsurface processes, demand sophisticated coupling algorithms and compatible grid structures. Achieving numerical stability and accuracy in these coupled simulations presents a formidable challenge, often requiring the development of novel numerical techniques. As the industry strives to model reservoirs more comprehensively, addressing these computational challenges becomes imperative for ensuring the reliability and efficiency of numerical reservoir simulations, especially in the context of advancing carbon sequestration initiatives.

2.3.3 Uncertainties and Model Calibration

2.3.3.1 Reservoir Properties:

Numerical reservoir simulation models is inherently challenged by uncertainties in key reservoir properties, including porosity, permeability, and fluid saturations. These uncertainties stem from the limited spatial resolution of data acquisition methods, the complexity of subsurface geological formations, and the inherent variability of rock properties. As a consequence, accurately characterizing these properties becomes a formidable task. The impact of uncertain reservoir properties on simulation outcomes is profound, influencing the reliability of predictions related to fluid flow, pressure distribution, and the effectiveness of carbon sequestration. Researchers and practitioners must grapple with these uncertainties, necessitating advanced methodologies to quantify, analyze, and mitigate their effects for robust reservoir management decisions.

2.3.3.2 Challenges of Model Calibration:

Model calibration is a crucial step in the numerical reservoir simulation process, aimed at aligning simulated outputs with observed field data. However, this process is beset with challenges, particularly when dealing with the complexity of subsurface systems. The interplay of various reservoir parameters, coupled with limited and often indirect data, makes calibration a non-trivial task. Identifying an optimal set of parameters that faithfully represents the geological and fluid flow characteristics of the reservoir requires sophisticated algorithms and computational approaches. The challenge intensifies in the context of carbon sequestration, where the stakes are high, and uncertainties in reservoir behavior can have significant environmental and economic consequences. Addressing these challenges demands a multi-faceted approach that combines advanced optimization techniques, machine learning algorithms, and a thorough understanding of the underlying geological processes.

2.3.3.3 Inherent Uncertainty in Predicting Subsurface Behavior:

The subsurface environment is inherently complex and dynamic, contributing to the intrinsic uncertainty associated with predicting subsurface behavior in reservoir simulation. The intricate interplay of geological heterogeneity, fluid-rock interactions, and the injection and migration of carbon dioxide introduces multifaceted uncertainties. While numerical models strive to capture these complexities, the dynamic nature of subsurface processes and the incomplete knowledge of reservoir properties pose formidable challenges. Uncertainties in predicting subsurface behavior and conformance control are magnified during long-term simulations, such as those associated with carbon sequestration projects, where the system's response to injection strategies needs to be forecasted over extended periods. Effectively managing and communicating these uncertainties is vital for instilling confidence in simulation results and informing decision-makers in the carbon sequestration domain. Uncertainties in reservoir properties pose challenges to numerical reservoir simulation, with calibration complexities further compounded by the intricate nature of subsurface systems. Acknowledging and addressing these uncertainties are imperative for enhancing the reliability and relevance of simulation outcomes, especially in the context of carbon sequestration initiatives where accurate predictions are pivotal for effective and responsible reservoir management.

Due to the multi-physics, non-linear, and multi-scale nature of these processes, numerical simulation is the primary approach used to solve mass and energy conservation equations for these applications. These numerical simulations are often very time consuming and computationally intensive since they require fine spatial and temporal discretization to accurately capture the flow processes (Doughty, 2010, Wen and Benson, 2019).

Several studies reported in these literatures have conducted reservoir simulation to evaluate the feasibility of CO_2 sequestration in a saline aquifer. Numerical simulations are essential to the understanding of the long-term geological storage of CO_2. Modelling and simulations can be used to predict where CO_2 is likely to flow, to interpret the volume and spatial distribution of CO_2 under storage conditions, and to optimize the injection operation. Jiang (Jiang 2011) analyzed the state of the art of physical modelling and numerical simulation of CO_2 dispersion. Pasala et al. (2003) implemented a simulation study to analyze the faults that affect CO_2 sequestration in a saline aquifer. Kumar et al. (2004) presented the results of a sequestration project to quantify estimates of the most important CO_2 storage mechanisms under realistic physical conditions. Hesse et al. (2006) studied the influence of factors, such as saline aquifer size, formation angle, capillary pressure, and residual trap, on CO_2 sequestration to optimize the injected CO_2 volume. Lee (2009) carried out a simulation to evaluate the feasibility of CO_2 injection into the aquifer and the potential for its leakage to the surface using a compositional simulator. However, as mentioned, numerical simulations are complex and time-consuming processes because they use iterative procedures to obtain a certain operating condition. The simulation also requires massive exploration data such as geological and geophysical data.

2.4 Introduction to Building the Simulation Model

2.4.1 Initial Data Collection:

The process of building a numerical reservoir simulation model begins with collecting extensive geological, physical, and fluid flow data from the subsurface reservoir. This includes data on rock properties, porosity, permeability, fluid characteristics, and any historical production data.

2.4.2 Static Reservoir Model Construction:

Based on the collected data, a static reservoir model is constructed. This involves creating a three-dimensional representation of the reservoir's geological properties, including porosity, permeability, and rock types. This static model serves as the foundational framework for subsequent dynamic simulations. The reservoir is discretized into a grid comprising numerous cells or grid blocks. This grid generation process involves defining the spatial resolution of the model and ensuring that it captures the essential geological features of the reservoir.

2.4.3 Developing the Dynamic Reservoir Model

2.4.3.1 Dynamic Model Setup:

Building upon the static model, a dynamic reservoir model is developed. This model incorporates additional parameters such as fluid properties, production history, and reservoir management strategies. It provides a time-dependent simulation of fluid flow within the reservoir.

2.4.3.2 Numerical Algorithms:

Selection of appropriate numerical algorithms is crucial for solving the partial differential equations governing fluid flow. Finite difference, finite element, or finite volume methods are commonly employed, each with its advantages and considerations.

2.5 Data Extraction from Numerical Reservoir Simulations

2.5.1 Execution of Simulations:

Once the simulation model is set up, the dynamic reservoir model is executed using numerical solvers. The simulation calculates fluid flow, pressure, and other parameters across the grid over time.

2.5.2 Output Data Analysis:

The output data, including spatial and temporal distributions of pressure, phase saturation, and other relevant parameters, are then analyzed. This analysis provides insights into the reservoir's performance under different conditions.

2.6 Description of Base Case Model

A 3D homogeneous aquifer with a constant-rate injector was simulated for the base case. The initial phase of this research involved creating a foundational numerical reservoir simulation model designed to replicate CO_2 injection into a saline formation. A comprehensive depiction of subsurface geology was generated for reservoir simulation modeling, considering the distribution of petrophysical properties such as porosity and permeability. The reservoir models were crafted using Computer Modeling Group (CMG) software. The planned reservoir model aimed to represent a hypothetical, heterogeneous reservoir with spatial variations in properties both between and within layers. Key structural elements of the reservoir model, including topography, bottom structure, and layer thickness, were derived from a previously created history match model at West Virginia University (WVU) on the Citronelle field, a saline reservoir situated in Mobile County, Alabama, USA (Haghighat 2013). The thickness exhibited variability within and across layers, introducing diverse levels of heterogeneity into the reservoir simulation model. The model comprises a grid of 125 x 125 blocks in the X and Y directions, while its geological structure encompasses 65 layers in the Z direction. Among these layers, 51 consist of sand, interbedded with 14 shale or impermeable layers strategically positioned at the top, middle, and bottom of the reservoir model. Layers 5 to 28 define the Upper Aquifer, and layers 35 to 61 define the Lower Aquifer within the simulation model, as shown in **Figure 11. Figure 12** provide visual representations of the reservoir structure and well locations.

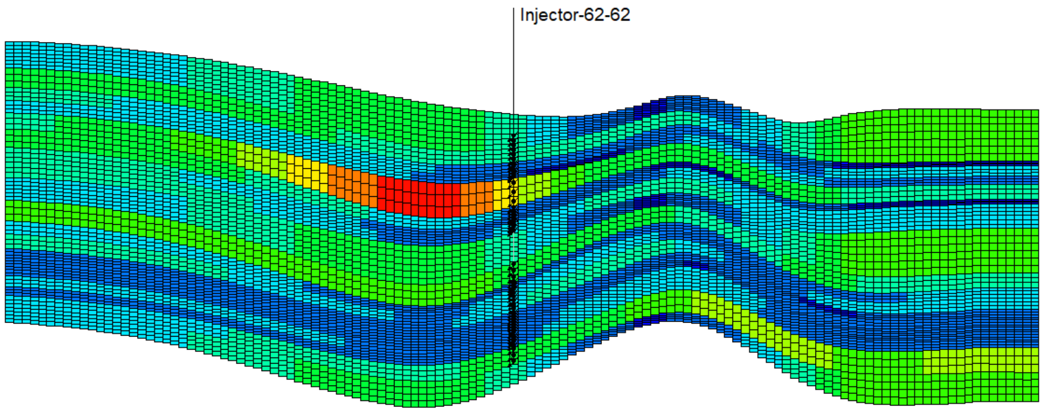


Figure 11. Cross-sectional view for the reservoir model. The position of the shale and sand layers can also be seen in grey and white, respectively

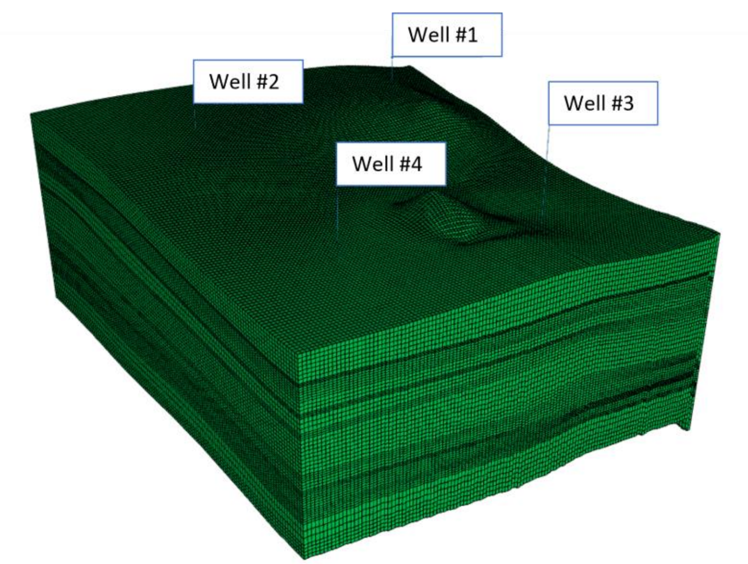


Figure 12. 3D View for the Reservoir Model Geometry and the relative position of each injection well. Note each layer's thickness is different across the depth

2.7 Development of Geological Reservoir Realizations

In the creation of the Smart Proxy Model, the utilization of Artificial Neural Network (ANN) algorithms necessitates extensive data training to comprehend intricate fluid patterns within the numerical reservoir simulation. Focusing on two critical variables, namely (1) Porosity and (2) Permeability distribution, the design of the numerical reservoir simulations delves into exploring these parameters in greater detail. Employing Gaussian Geostatistical Simulations, diverse meaningful and controlled porosity distributions were generated for the additional geological realizations. Consequently, alongside the initial geological model (base model), multiple geological realizations (comprising different number of variations, based on which phase of the study) were crafted, featuring distinct properties such as porosity and permeability. Despite sharing a common overarching porosity distribution pattern with a consistent range of values, these geological realizations exhibit unique and diverse distributions. The objective behind these models was to emulate a conceivable, heterogeneous reservoir characterized by spatial variations in reservoir attributes across different layers.

Figure 13 and Figure 14 provide visual representations of the porosity distribution maps for layers number 5 and 50 across all 20 realizations employed in phase 5. Notably, in the northwest region of the reservoir, all realizations generally display a high porosity distribution. However, each realization presents a distinctive porosity distribution, showcasing the variability inherent in the models. Obtaining directed permeability measurements is a challenging and costly endeavor compared to the relatively inexpensive methods for measuring porosity data. Due to this cost disparity, porosity-permeability correlations are frequently employed to estimate permeability. In an analysis of the porosity and permeability data from Haghighat (2013), it was identified that the reservoir model encompasses a minimum of four distinct permeability groups. These classifications, termed 'Very Conductive Perm,' 'Conductive Perm,' 'Average Perm,' and 'Tight Perm,' were established to represent varying degrees of rock permeability. To generate diverse yet meaningful permeability data, the porosity-permeability correlations were utilized to assign permeability values to each grid cell in the model. Specifically, in the upper section of the Upper Aquifer, a highly conductive permeability correlation was applied to the top eight sand layers, while an average conductive permeability correlation was designated for the lower 16 sand layers. Analogously, the layers within the Lower Aquifer formation were categorized into Conductive Perm and Tight Perm correlations. Consequently, the complex 65-layer model described earlier can be streamlined into two primary reservoir intervals. These intervals consist of the top portion with a high-permeability upper layer and an intermediate-permeability middle layer, and the bottom portion featuring a single relatively low-permeability layer. These are interspersed with three shale formations characterized by very low permeability, serving as barriers to vertical gas migration.

The porosity data ranged from 0.01 to 0.0.25, and the permeability data ranged from 1e-006 to 544 md across all the realizations and in all different phases of the research. However, their distribution are significantly different than each other. To have a seal for the CO_2, the shale layers were assigned a low fixed value of 0.01 for their porosity and very low 1e-006 md for their permeability values. The permeability in the I and J directions was assumed to be the same. However, based on experience in geological formations, it was decided to use a permeability anisotropy (Kv/Kh). By plotting porosity and the derived permeability data, at least 4 rock types were observed. Therefore, 4 static rock types, each consisting of a certain number of layers, could be created (Alabboodi, 2021) as shown in **Table 5.**

The interplay between different phases in porous media significantly influences fluid flow, a phenomenon captured by the concept of relative permeability. To ensure accurate reservoir simulation outcomes, it is crucial to have appropriate values for relative permeability. However, in the context of CO_2 sequestration in saline aquifers, a relatively new area of research, no laboratory studies using actual cores have been conducted thus far to obtain these curves. To overcome this limitation, saturation and relative permeability tables from existing literature, specifically for CO_2-water systems, were utilized (Bennion and Bachu, 2005). The tables employed in this project are derived from studies on the deep Basal Cambrian Sandstone aquifer situated in the Wabamun Lake area southwest of Edmonton, Alberta, Canada. The resulting relative permeability curves, crucial for the reservoir simulation, are illustrated in **Figure 15.**

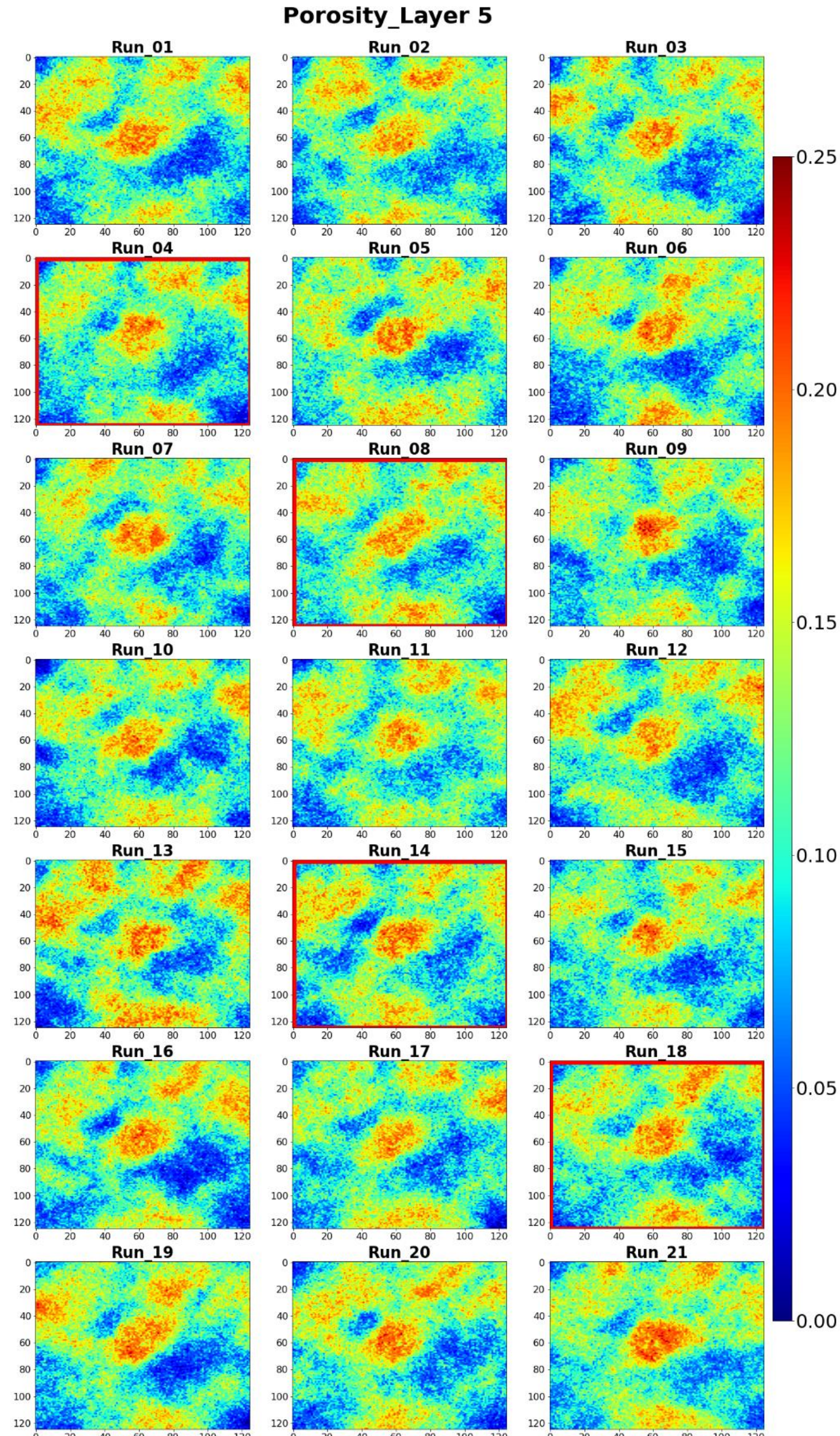


Figure 13. Porosity Distribution of layer# 5 for the 20 realizations (for demonstration: Phase 5 of the research). The highlighted realizations in red square are the blind realizations

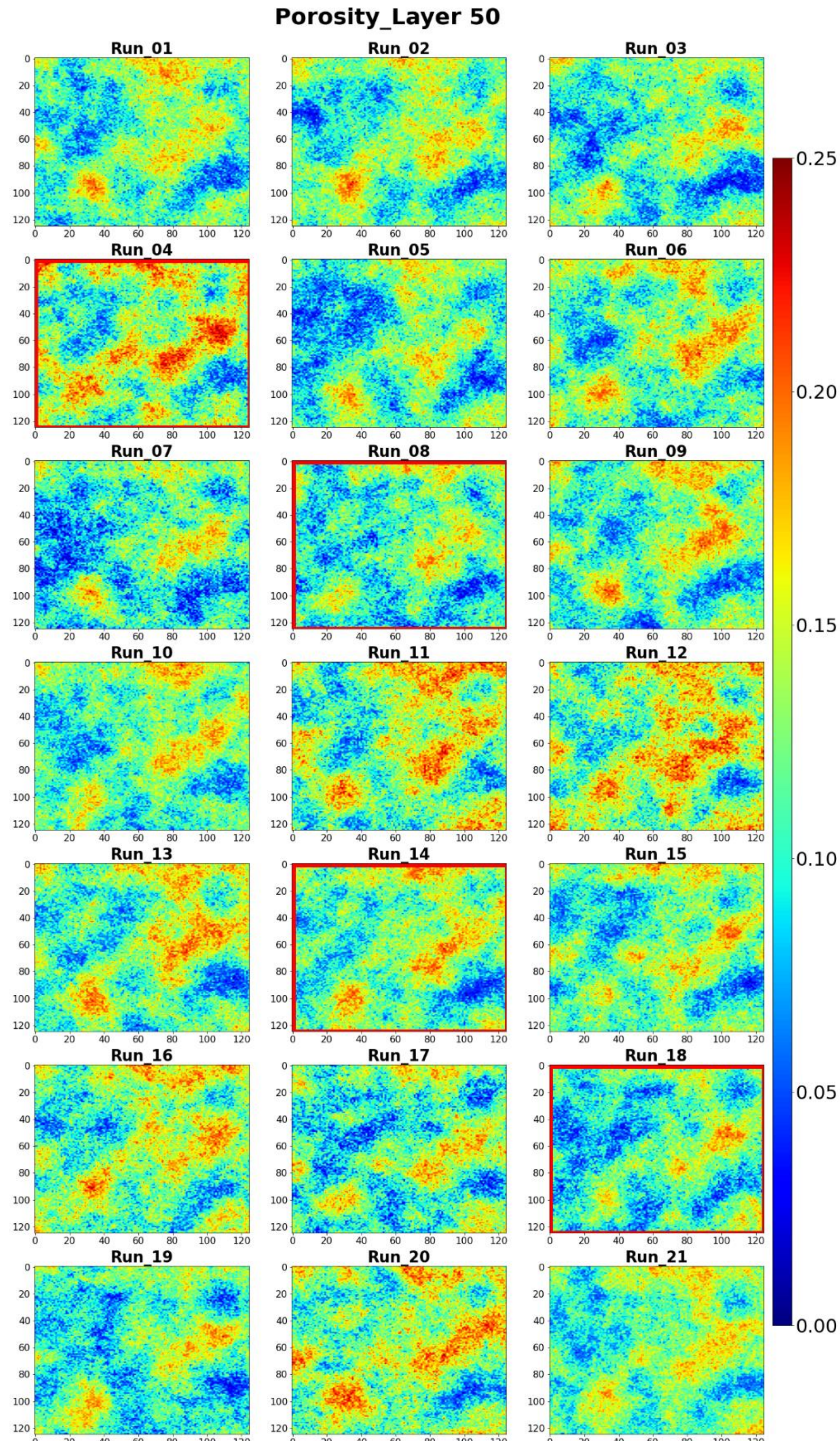


Figure 14. Porosity Distribution of layer# 50 for the 20 realizations (for demonstration: Phase 5 of the research). The highlighted realizations in red square are the blind realizations

Table 5. Classification and generation of permeability based on porosity

Permeability Classification	Data Ranges
Very conductive	Min — 9.9564 * EXP (21.749 * Porosity)
Conductive	9.9564 * EXP (21.749 * Porosity) — 0.826 * EXP (28.18 * Porosity)
Average	0.826 * EXP (28.18 * Porosity) — 0.646 * EXP (21.871 * Porosity)
Tight	0.2533 * EXP (22.369 * Porosity) — Max

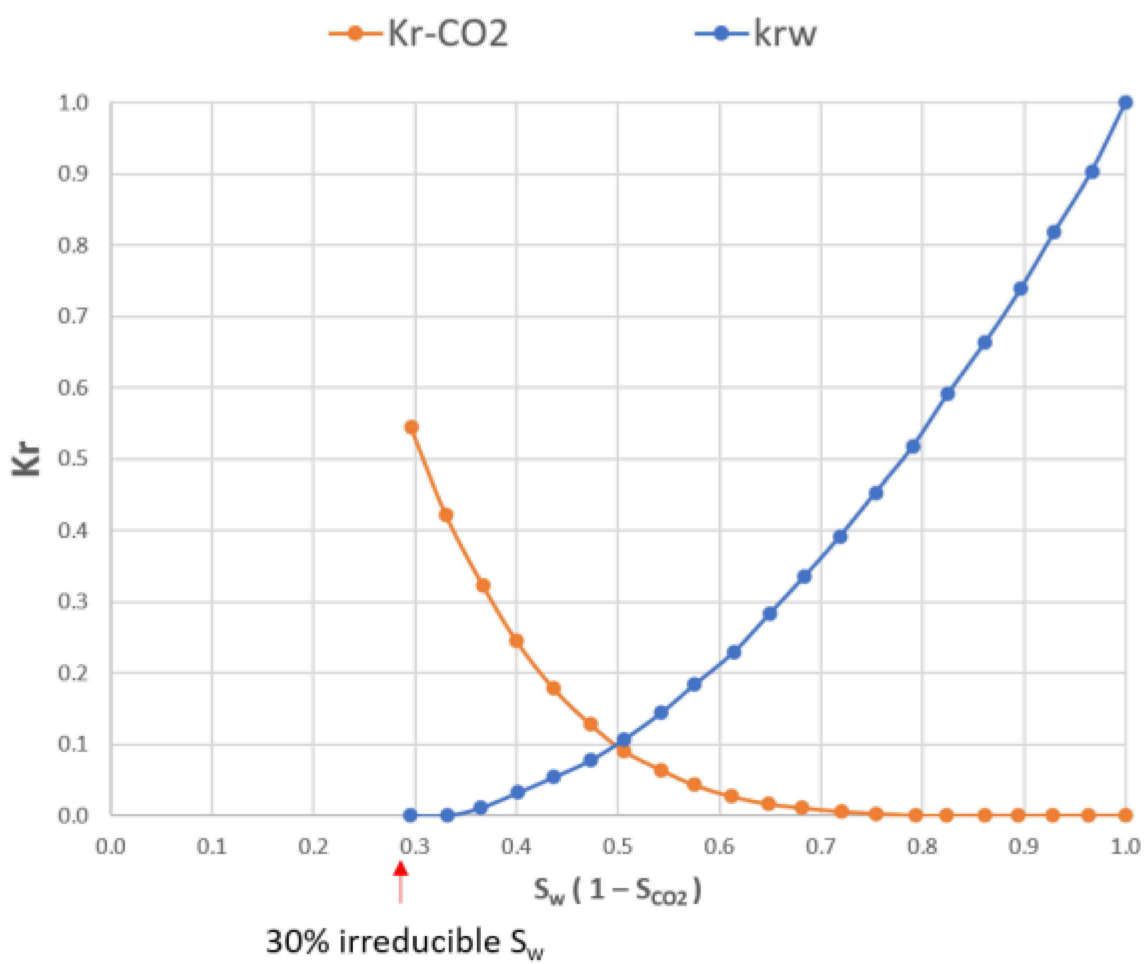


Figure 15. Relative permeability curve used in the simulation study

Model parameter data for the three very tight interbedded sealing intervals (on the top, bottom, and between two reservoir formations) are not shown. The saline formation was assumed to be a close flow boundary reservoir. Volume modifiers of 1E+5 were used in the edge blocks to reduce the effect of the lateral boundaries. It is important to note that these parameters, including the relative permeability curve, were kept constant across the 20 realizations. The only data that was different across each realization is porosity, permeability and transmissibility.

Table 6 contains additional model parameters and constraints not elaborated upon. These listed properties represent the average reservoir characteristics of the two previously described formations. However, specific data for the three tightly interbedded sealing intervals (located at the top, bottom, and between two reservoir formations) is not provided. The saline formation is assumed to act as a reservoir with a close flow boundary. To minimize the impact of lateral boundaries, volume modifiers of 1E+5 are applied to the edge blocks. It's crucial to highlight that, across all other geological realizations, these parameters, including the relative permeability curve, are held constant. The only variables that differ among realizations are porosity, permeability, injection wells, their patterns, and transmissibility.

Table 6. Reservoir Parameters and Properties

Parameter	Value	Parameter	Value
Porosity	3D distribution	Water Density (lb/ft^3)	62
Permeability (md)	3D distribution	Water Viscosity (cp)	0.26
Temperature (°F)	230	Water Compressibility (1/psi)	3.20E-06
Salinity (ppm)	100,000	Datum Depth (ft)	9416
Residual Gas Saturation	0.22	Pressure Reference (psi)	4,370
Irreducible Water Saturation	0.3	Injection well BHP (psi)	5.500
Fracture Pressure Gradient (psi/ft)	0.7		

2.8 Numerical Reservoir Simulation Runs:

The numerical reservoir simulation in this study utilized the GEM simulator developed by the Computer Modeling Group (CMG). GEM is a sophisticated compositional simulator equipped with a general equation of state, incorporating features such as dual porosity, CO_2 injection and sequestration, miscible gases, volatile oil, gas condensate, horizontal wells, well management, complex phase behavior, and more. In this investigation, the CO_2 module of the simulator was specifically employed to replicate CO_2 injection and sequestration into an aquifer formation, along with considerations for aqueous phase chemical reactions, mineral precipitation, and dissolution. Modeling CO_2 storage in saline aquifers involves solving component transport equations, equations for thermodynamic equilibrium between gas and the aqueous phase, and geochemistry equations accounting for reactions between aqueous species and mineral precipitation/dissolution.

Two methods are employed to solve the coupled system of equations: the sequential solution method and the simultaneous solution method. The sequential solution involves solving flow and chemical equilibrium equations individually and consecutively, with iterations applied between the two until convergence is achieved. In contrast, the simultaneous solution method, also known as the fully-coupled approach, solves all equations simultaneously using Newton's method. The GEM simulator employs the fully-coupled methodology for modeling CO_2 storage in saline aquifers, utilizing the equation of state compositional and greenhouse gas simulator with geochemical options for solubility, residual gas, and mineral trapping (Nghiem 2004).

For each phase of the research, a certain number of geological realizations were designed. For example, for phase 1 of the research, from the foundational reservoir model, a total of 63 simulation runs were generated for the Smart Proxy Modeling process. Within this set, 30 geological realizations have been specifically earmarked for training purposes, while an additional 9 geological realizations are reserved for blind validation. Each of these simulation runs boasts distinct porosity and permeability maps, providing a diverse array of geological conditions for analysis. In remaining phases of the research, 20 or 21 realizations were generated with 4 realizations as the blind dataset realizations and the remaining for train, and calibration and validation datasets.

It's worth noting that, despite the variability in porosity and permeability, the sequence of rock types along the reservoir layers remains consistent across all simulation runs, maintaining a uniform arrangement from the top to the bottom of the reservoir. Additionally, the relative permeability curve employed in these simulations is standardized, ensuring uniformity across the entire set of geological models. This standardized approach enhances the comparability of results and contributes to the robustness of the Smart Proxy Modeling process.

The thickness of model grids is not uniform, and changes based on the location of the grid. **Figure 16** shows a general visualization about non-uniform grid thickness model.

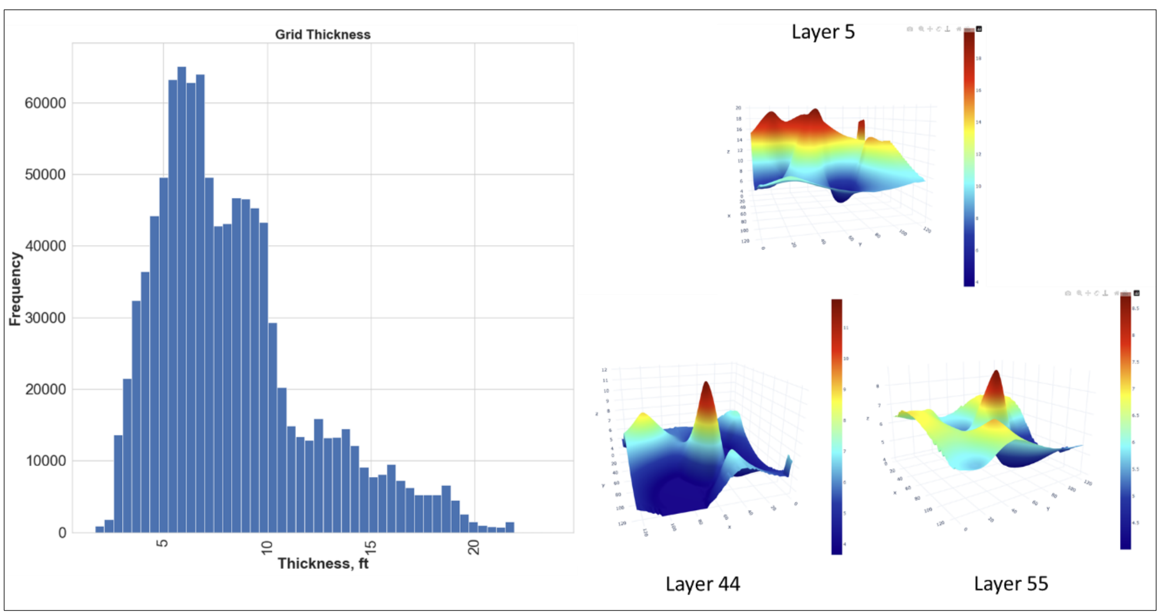


Figure 16. Grid thickness frequency and variations across several layers

2.8.1 CO_2 Injection Design (Number of Injector, Injection Time):

The reservoir simulation model features four vertical injection wells, perforated and completed across all 51 sand layers within the Upper and Lower Aquifer formations. The injection scheme adheres to two primary and secondary constraints: well-level primary Bottom Hole Pressure and group-level secondary Max Injection Rate. The operational constraints for CO_2 injection well are set at 1.5 million cubic feet/day (standard condition at the surface), equivalent to 2.48 MT/Year, and 5,500 psi, respectively. The maximum allowable BHP of 5,500 psi is determined based on an assumed 10% buffer from the reservoir fracture pressure. The BHP constraints are designed with the "CONT REPEAT" action, meaning that if a constraint is violated, the simulation will repeat the same time step, switching to the second entered constraint to avoid any violation. Group-level constraints for max injection rate are designed with only "CONT," allowing the simulation to continue if the constraint is violated, changing to the second constraint for the next time step. Injection commenced on January 1, 2020, and will continue for 30 years until January 1, 2050. Post-injection monitoring is designed to span from January 1, 2050, to January 1, 2320, covering a period of 270 years. The overall injection and post-injection scheme extends over 300 years **(Figure 17).** The focus of this study is on a time step 30 years time into the injection **(Figure 18).**

Figure 17. Injection and post-injection timesteps used in the simulation study

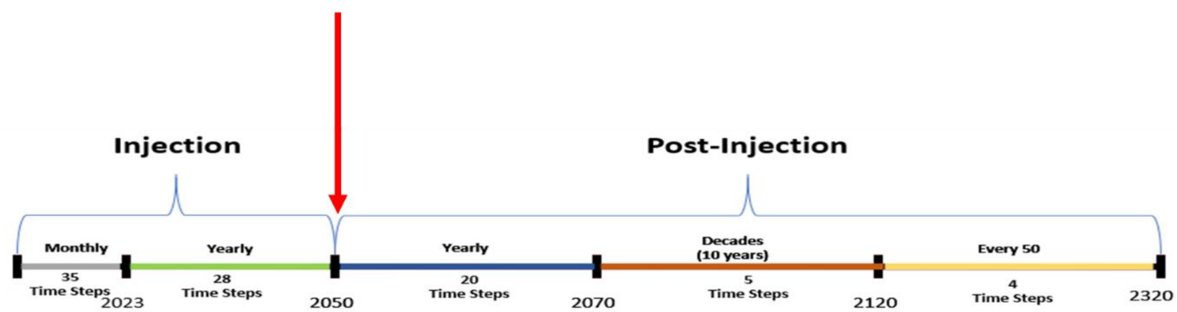


Figure 18. The prediction was focused on a timestep 30 years after the start of the injection

2.9 Summary

In this chapter, we embarked on exploration of numerical reservoir simulation, unraveling its significance in understanding subsurface fluid dynamics and its pivotal role in advancing carbon sequestration behavior. We initiated our journey by defining numerical reservoir simulation as a computational technique crucial for modeling fluid flow within subsurface reservoirs. The primary objectives within the context of carbon sequestration were delineated, emphasizing the role of simulation models in predicting carbon dioxide behavior, optimizing injection strategies, and assessing associated risks. Our exploration then delved into the history of numerical reservoir simulation, tracing its early developments and key milestones that laid the foundation for modern simulation techniques. The evolution of this technology was scrutinized, highlighting its relevance to carbon sequestration initiatives and the efficient management of sequestration projects. The chapter also shed light on the computational challenges faced by numerical reservoir simulations, acknowledging issues related to large-scale models, multiphase fluid flow, and the integration of interdisciplinary components.

The discussion on uncertainties and model calibration underscored the complexities in predicting subsurface behavior, emphasizing the need for advanced technologies to address these challenges. As we navigated through the benefits of numerical reservoir simulation, its role in improving the understanding of subsurface dynamics, optimizing injection well location in carbon sequestration, and supporting decision-making and risk management in sequestration projects became apparent.

2.11 Transition to Smart Proxy Modeling Part 1

In anticipation of the forthcoming chapter, this section marks a pivotal shift in focus towards the realm of artificial intelligence (AI) and machine learning (ML), specifically on artificial neural networks (ANNs). This transition signifies a harmonious amalgamation of reservoir simulation expertise with the state-of-the-art capabilities embedded in ANNs. will delve into the principles, applications, and advancements of this cutting-edge approach, showcasing its transformative potential in the realm of carbon sequestration and subsurface fluid dynamics.

In the realm of geological CO_2 sequestration, the veracity and reliability of numerical simulation models are frequently challenged by the absence of quantifiable accuracy metrics linked to direct observations, which would serve as the definitive ground truth. The deficiency in such validation conditions introduces a potential problem, as it may lead to uninformed decision-making in the context of GCS projects. An illustrative example of the repercussions of this challenge is evident in historical cases where early numerical models underestimated the migration of CO_2 plumes (Chadwick and Noy 2015). The subsequent refinement achieved through the process of history matching underscored the critical role of uncertainty quantification. The concept of history matching encounters unique challenges compared to traditional applications in oil and gas production fields. Unlike hydrocarbon reservoirs, GCS (Geological Carbon Storage) sites often lack a substantial history of reservoir behavior, making the conventional history matching approach less straightforward. The closed-loop nature of traditional reservoir management also makes it extremely prone to uncertainties and inaccuracies in predictions, particularly when extrapolating future reservoir performance. This necessitates the development of more adaptive, data-driven models that can integrate limited historical data with real-time monitoring to enhance predictive accuracy and reliability in GCS operations **(Figure 19)**.

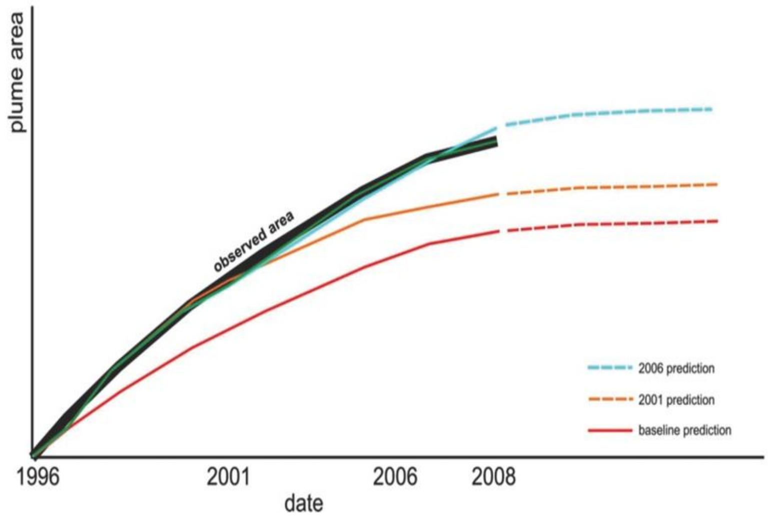


Figure 19. Numerical simulation *predictive modeling is a progressive path. Based on 1996, 2001, and 2006 datasets compared with an observed performance measure (from Chadwick and Noy 2015)*

Within the closed-loop reservoir management framework for geological CO_2 sequestration, history matching becomes a critical and often tedious task. History matching involves adjusting the parameters of the geological models to replicate past field observations and performance data, ensuring that the simulated behavior aligns closely with the actual reservoir conditions. Engineers engage in an iterative process of tweaking various model parameters, such as permeability, porosity, and fluid properties, to achieve a satisfactory match between simulation results and historical data. This iterative adjustment is a time-consuming and labor-intensive process, as it requires a deep understanding of the subsurface geological formations and the dynamic behavior of injected CO_2 within those formations.

The storage of CO_2 in geological formations is subjected to many uncertainties. Simulating geological CCS is extremely difficult due of the large range of time and length scales involved. The numerical treatment of nonlinearity, as well as the discretization of space and time, are significant computing challenges. The fundamental difficulties in the discretization and numerical solution of discretized partial differential equations are accuracy, stability, and computational speed. In light of the variations observed among various modern simulation methodologies, as outlined in the benchmark study conducted by Nordbotten et al. (2012), this presents a matter of apprehension. The benchmark analysis further reveals that, even with simplified geological parameters, the existing models exhibit significant sensitivity to the underlying physical assumptions **(Figure 20).**

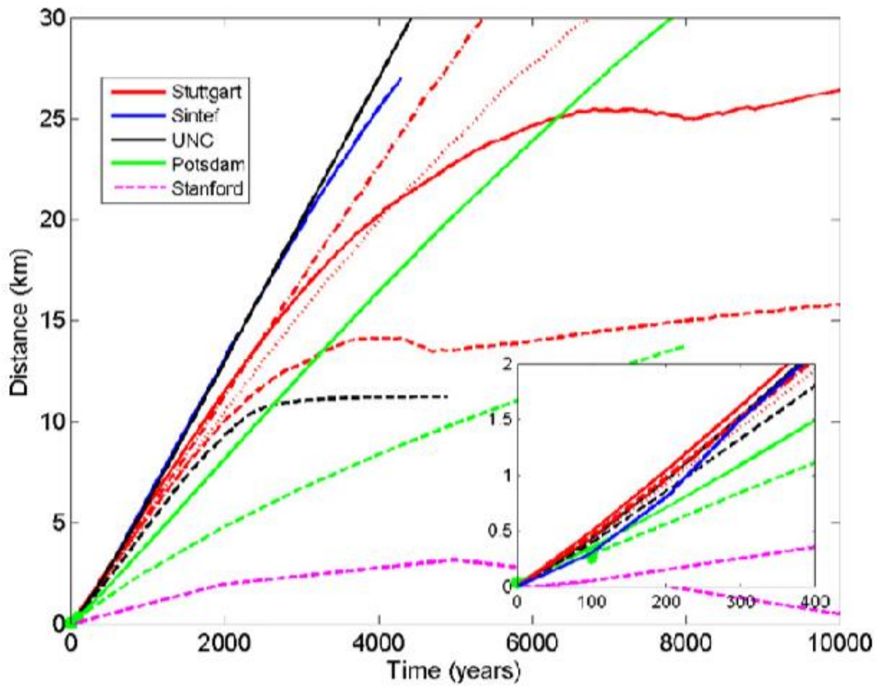


Figure 20. Various participating groups are distinguished by distinct colors, and diverse line styles signify different runs conducted within the confines of a single group (Nordbotten et al. 2012)

The figure includes data from different participating groups, with different line styles representing different runs within the same group, and different colors representing different participating groups. The solid lines refer to the common base case. The figure also includes an inset that shows the large range of variability in the results already at the end of the injection phase, with a spread of 2.5 km or 1.5 miles between the different model predictions, which continues increasing throughout the post-injection period.

A myriad of numerical tools, each employing distinct algorithms, has been applied globally, encompassing TOUGH2, ECLIPSE, CMG's GEM, CO2-PENS, NUFT, TRANSTOUGH, MODFLOW, FLOTRAN, novel reservoir monitoring, modeling, and simulation (NORMS), as well as MATLAB reservoir modeling tools (MRST), among others (Ennis-King and Paterson, 2007; Pruess and Spycher, 2007; Celia et al., 2015; Møll Nilsen et al., 2015; Teletzke and Lu, 2013; Rezk and Foroozesh, 2019; Wen and Benson, 2019; Urych and Smoliński 2019). The holistic modeling of fluid properties across multiple phases, CO_2 plume behavior, pressure dynamics, and reactive-transport processes, coupled with mechanical interactions across various temporal and spatial scales, heavily relies on factors like storage mechanisms, detailed geological models, cross-scaling of geological properties, upscaling methodology, and result interpretation with comparatively lesser emphasis on specific numerical modeling algorithms (Nordbotten et al., 2012) **(Figure 21. The** current industry workflow to model GCS (Ajayi et al, 2019) Law et al. (2004) examined the outputs of five simulators against a benchmark scenario concerning CO_2 storage in coalbeds. Similarly, Class et al. (2009) undertook a benchmark study using different simulators to tackle issues related to CO_2 storage in geological formations. These benchmark studies also highlighted that simulation outcomes for any storage problem vary depending on the chosen simulator, heavily influenced by the employed numerical methods and the physics of the processes incorporated. They recommended that selecting a simulator should be based on the specific physical processes being targeted to achieve the most accurate results.

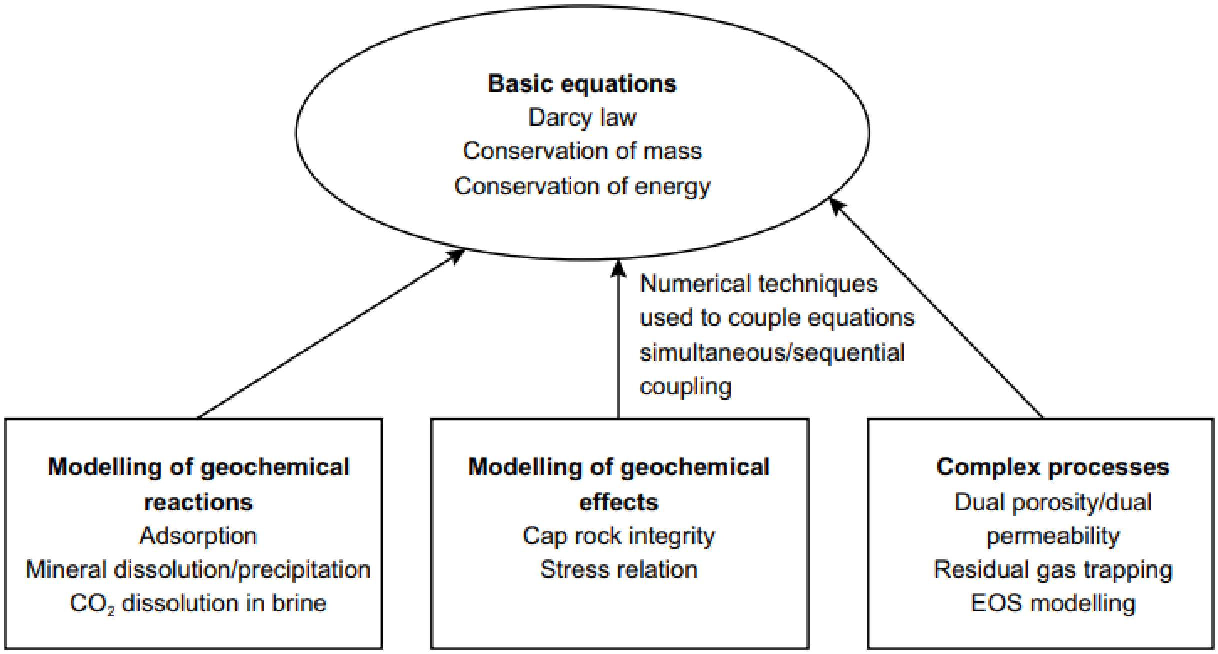


Figure 21. The current industry workflow to model GCS (Ajayi et al, 2019)

While uncertainty modeling, utilizing statistical data and stochastic tools, enhances CO_2 capacity predictions to a certain extent, it falls short in offering a comprehensive account of overall uncertainties in a practical time frame. The essential role of history matching, involving time-lapse monitoring, comes to the fore in enhancing predictions concerning the long-term performance of the site and the behavior of CO_2 underground (Nordbotten et al. 2012). Furthermore, comprehensive site characterizations and experiments conducted at various scales, ranging from pore to site scale, divulge fundamental parameters influencing the storage process. These parameters are intricately linked with characteristics and upscaling methodologies at smaller scales, including pore geometry, capillary pressure, rock and fluid properties, interfacial tension, wettability, molecular diffusion, hydrodynamic dispersion, water salinity, surface minerals, as well as mineralization and precipitation processes (Middleton et al. 2012).

Effectively addressing uncertainties in numerical simulation necessitates geological models and gridding that reflects cross-scaling of complex geological properties through site characterization. It also requires models that account for complex properties involving multiple phases of fluid, algorithms reflecting various trapping mechanisms, and heterogeneous reservoir properties (Middleton et al., 2012). Despite the utilization of existing numerical simulation tools, even for highly idealized problems, fundamental questions concerning CO_2 storage lack conclusive predictions with satisfactory accuracy (Nordbotten et al. 2012).

The challenge lies in striking a balance between achieving a good match with historical data and ensuring that the adjusted model accurately represents the true geological characteristics of the reservoir. Engineers must constantly navigate the trade-off between model complexity and computational efficiency, as excessively complex models may hinder the practicality of simulations. The iterative process of history matching, where engineers continually adjust the parameters of the geological models to align with observed data, often leads to modifications of the original models crafted by geologists. Such workflows demand a meticulous approach, with engineers constantly changing the models until a satisfactory match is attained **(**Error! Reference source not found.**).**

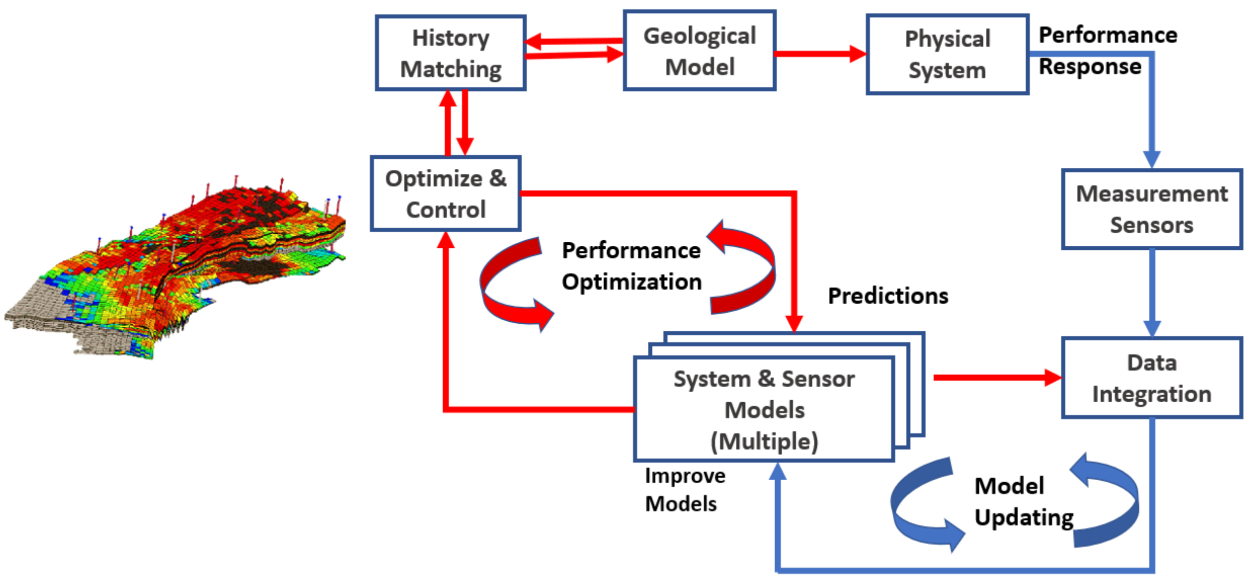


Figure 22. Traditional closed-loop reservoir management cycle

The scarcity of historical data in GCS necessitates a reevaluation of the history matching paradigm. Instead of relying solely on historical production records, the emphasis shifts towards incorporating available data related to the injection and monitoring phases of CO_2. This may include data on injection rates, pressure responses, and monitoring data obtained during the injection and post-injection periods. Given the limited historical context, the history matching process in GCS becomes an exercise in constraining the numerical simulation model to match observed data obtained during the operational phase. Engineers typically adjust model parameters, such as permeability, porosity, or other relevant properties, to achieve a satisfactory alignment between simulated and observed data. However, the challenge lies in the inherent ambiguity and non-uniqueness of the matching process. The uniqueness challenge in history matching is exacerbated in GCS due to the complex and variable nature of saline subsurface conditions. Engineers face the task of exploring various combinations of parameters to replicate observed behavior, introducing a level of subjectivity and variability in the process. This non-uniqueness is a fundamental characteristic of GCS history matching, and engineers must exercise expertise and judgment to arrive at a plausible and scientifically sound match.

Moreover, the time-consuming nature of the reservoir simulation, uncertainty assessment, and history matching process in GCS further complicates the endeavor. The iterative adjustments to simulation parameters and subsequent model run demand a significant computational effort. This underscores the importance of adopting efficient and advanced techniques, such as machine learning and artificial intelligence, to streamline the history matching process and enhance the reliability of the resulting numerical models. This singular case serves as a compelling demonstration of the pressing need for a systematic and reliable tool capable of providing real-time answers for diverse scenarios. One in which avoids the possibility of unexpected or divergent future outcomes. At site closure, predictive models should be sufficiently robust, and uncertainties sufficiently understood, to effectively rule out the possibility of significant adverse future outcomes.

Traditional reservoir management methods have relied on assumptions and statistical algorithms to characterize reservoir properties between wells **(**Error! Reference source not found.**).** The imaginary side of reservoir simulation involves incorporating assumptions, models, and parameters that are not directly measured but are essential for simulating and predicting reservoir behavior. This includes factors such as reservoir heterogeneity, fluid properties, rock properties, and various other reservoir characteristics. This approach has limitations in accuracy and adaptability. It's important to note that the distinction between the real side (measured data) and the imaginary side (incorporating additional factors) is a conceptual framework rather than a literal division within the simulation process. Both sides work together to create a comprehensive understanding of the reservoir system and facilitate decision-making in reservoir engineering.

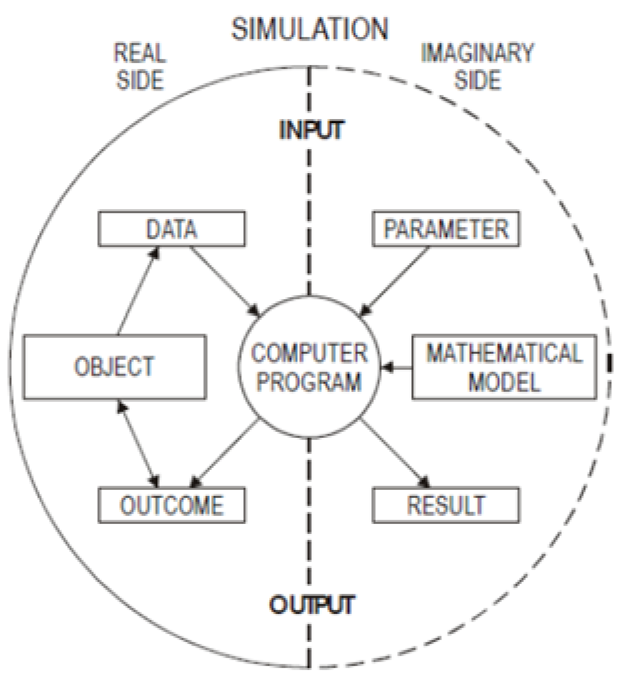


Figure 23. Real side and imaginary side of numerical reservoir simulation (M. R. Islam, et al. 2016)

The fate of many complex projects is often determined by the ability of the multidisciplinary teams to balance between their instinctive pull towards complexification (in search of presumed accuracy) and the desire to meet project deadlines **(**Error! Reference source not found.**).**

Simulation of CO_2 storage tends to be more challenging than traditional simulations because of the complex interactions among phase change, composition, and reservoir heterogeneity, necessitating the use of highly efficient computational algorithms (Jiang 2011).

Figure 24. The major factors in a modeling assignment with focus on speed, accuracy, flexibility, domain decision, and cost

A novel AI-based reservoir management workflow would offer a paradigm shift by employing smart proxy modeling, real-time data integration, and artificial intelligence forecasting to construct a fact-based geological model. This innovative workflow should offer precision in characterizing reservoir properties and optimizing operational strategies **(Figure 25).** The overriding objective of a multidisciplinary reservoir study is value addition.

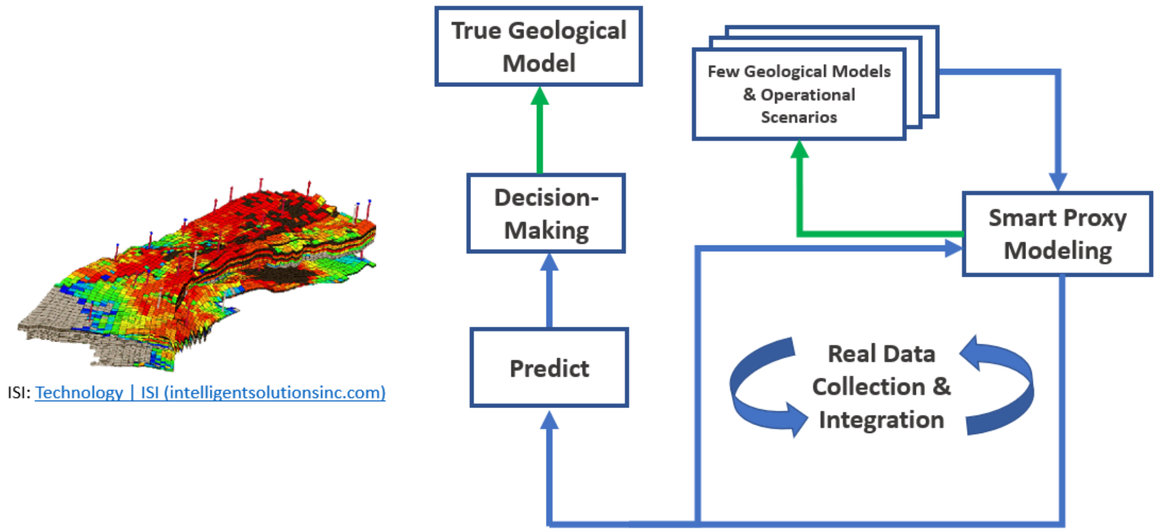


Figure 25. Proposed workflow for reservoir management decision-making with dynamic model updating

The methodological approach applied to model CO_2 sequestration into aquifers and conduct history matching is characterized by a meticulously designed and comprehensive workflow. The process unfolds in a step-by-step progression that reflects the scientific rigor inherent in such an endeavor. The journey begins with Problem Formulation and Data Collection, where the precise objectives of the CO_2 sequestration initiative are defined. This phase involves the meticulous collection of geological, reservoir, and fluid data, while also addressing key uncertainties and recognizing potential sources of sparsity in measurement data. Moving forward, Geological and Reservoir Characterization involves the development of intricate geological models that encapsulate the nuanced complexities of the aquifer system. Reservoir properties, including porosity, permeability, and caprock integrity, are characterized, and available data and expert insights are integrated into foundational geological models. Model Setup and Simulation follow, where a numerical reservoir simulator is implemented to model CO_2 injection and migration. Injection scenarios, rates, and operational parameters are defined, and baseline simulations are conducted to prognosticate reservoir behavior over temporal horizons. These models are trained using historical data derived from numerical simulations, and their accuracy is rigorously validated against any operational scenarios. The process proceeds to Initial History Matching, involving a comparative analysis between results from numerical simulations and smart proxy predictions. Pivotal parameters within the geological and reservoir models are adjusted to align with observed data, with iterations refining models for enhanced precision. Uncertainty Quantification follows, wherein thorough assessments are conducted to gauge the impact of parameter variations. Real-Time Data Integration is implemented through monitoring systems capturing ongoing data during CO_2 injection, with continuous model updates based on real-time assimilation. Dynamic Model Updating employs adaptive algorithms to refine geological and reservoir models dynamically. The most realistic geological model is updated through the integration of newly acquired measurement data, contributing to augmented predictive accuracy. Decision-Making and Optimization leverage updated models to optimize injection strategies and operational decisions. Sensitivity analyses are conducted to discern critical parameters influencing performance. This structured approach ensures a systematic progression, emphasizing adaptability and accuracy throughout the CO_2 sequestration modeling and history matching process.

This workflow accentuates the amalgamation of smart proxy models, real-time data assimilation, and continuous model updating to effectively contend with uncertainties and sparse measurement data inherent in CO_2 sequestration initiatives.

The enhanced smart proxy modeling approach in this study focuses on the first part of this proposed workflow and adeptly addresses the spatial complexity, spatio-temporal intricacies, and heterogeneity inherent in reservoir characteristics. By integrating domain-expertise-based feature generation, this framework ensures a precise representation of reservoir models at the resolution of each grid-cell. It effectively optimizes CO_2 sequestration strategies while navigating computational challenges associated with conventional numerical reservoir tools. The utilization of data from numerical reservoir simulation sets the stage for creating controlled and reproducible datasets tailored specifically to the research objectives. This approach facilitates the exploration of various scenarios and conditions, offering invaluable insights that significantly contribute to our understanding of CO_2 sequestration processes. Furthermore, it paves the way for breakthroughs in investigating reservoir behavior, all achieved at a minimal computational cost.

The justifications presented for employing smart proxy models in the absence of real measured data for every grid in a reservoir for CO_2 sequestration seem reasonable. Here's a breakdown of the key points:

1. Data Completeness: Using data generated from numerical reservoir simulation allows for a more comprehensive dataset, covering a wider range of grid points and time steps. This can enhance the understanding of the reservoir's behavior, even in areas without measured data.
2. Data Consistency: Employing data from the numerical reservoir simulation ensures consistency in terms of format, quality, and reliability. This eliminates discrepancies that may arise when combining multiple data sources.
3. Representative Trends: The data generated from the numerical reservoir simulation can still capture important trends, patterns, and relationships within the reservoir. Smart proxy models and machine learning algorithms can leverage this data to make predictions and approximate behaviors, providing valuable insights and reasonably representing the reservoir's behavior.
4. Computational Efficiency: Smart proxy models are often simpler and computationally efficient compared to numerical reservoir simulations. By utilizing these models, the computational burden can be significantly reduced, enabling faster evaluations, scenario analysis, and real-time decision-making.
5. Calibration and Validation: The data generated from the numerical reservoir simulation can be used for calibration and validation, comparing the simulation results with available field data. This helps assess the accuracy and reliability of the numerical model and validates the use of simulation data as input for smart proxy models.
6. Continuous Learning and Improvement: Smart proxy models and machine learning algorithms can continuously learn and improve over time. By incorporating feedback and real measured data as it becomes available, the models can adapt and refine their predictions, gradually improving their accuracy and performance.

These justifications highlight the potential benefits of utilizing smart proxy models when real measured data is limited. However, it's important to note that the reliability and accuracy of the smart proxy models at the **GRID LEVEL** depend on the quality and representativeness of the numerical simulation data as of now, as well as the chosen machine learning techniques and model calibration. The developed smart proxy modeling goal and domain-based feature engineering in this study was to replicate the results generated by the numerical reservoir simulation and therefore, this approach aligns with the overarching objective of providing a viable alternative when confronted with limited measured data in Carbon Capture and Storage (CCS) projects. The reliability of the smart proxy models hinges on meticulous consideration of the simulation data's quality and its representativeness of actual reservoir conditions. Rigorous calibration of machine learning techniques plays a pivotal role in fine-tuning the smart proxy models, ensuring their accuracy and predictive power. The goal-oriented feature engineering adopted in this study is designed to enhance the model's domain-specific capabilities, emphasizing its relevance to the unique challenges posed by CO_2 sequestration in deep saline aquifers. **Numerical reservoir simulation (NRS) models are presently not being used to their full potential (Mohaghegh 2022).** The deployment of this smart proxy models could serve as a strategic solution to the data limitations inherent in CCS projects. The study's commitment to replicating numerical reservoir simulation results underscores the model's potential to contribute valuable insights into the behavior and performance of CO_2 sequestration systems, as a complimentary tool to numerical reservoir simulation when there is no measured data available from the field offering a pragmatic avenue for decision-making and risk assessment in the absence of extensive measured data.

Traditionally, any modeling attempt, once constructed, is limited to predicting outcomes based on the specific conditions of the simulated reservoir, particularly if the underlying physics of the numerical simulation remain unchanged. For example, if a model is created for a reservoir model at a pressure of 4000 psia, it won't be suitable for making estimations under different pressure conditions. A new model would need to be developed for these altered conditions.

The development and efficacy of Smart Proxy Models (SPM), particularly when focused on grid-based approach highlight a critical aspect of reservoir simulation: the handling of complex, variable conditions. While the spatio-temporal database forms the backbone of ANN training in SPM, it also underscores the need for a nuanced approach to modeling geological complexities and dynamic variations in change in the operational schemes and complexities. This leads us into an important facet of my research, which is provided in the next section.

2.12 Design of Complexity

2.12.1 Phase 1: Applying Rotation to Realizations

Geological heterogeneity is a multi-scale complex and sparsely sampled and inherently uncertain. It is the key control on fluid flow in a reservoir and influences engineering and management decisions **(Figure 26).**

Motivation:

- *Uncertainty in interpretation:* Introduces variability by rotating each realization, simulating geological heterogeneity.
- *NN Adaptation:* Enables the ANN to recognize and understand the impact of geological variations, improving its predictive accuracy.

Advantages and Comparisons:

- *Computational Efficiency:* ANN, once trained on rotated realizations, provides faster predictions compared to rerunning simulations with rotated scenarios.
- *Understanding Anisotropy:* Insights gained help in understanding how anisotropy affects CO_2 movement, guiding reservoir engineers in real-world scenarios.

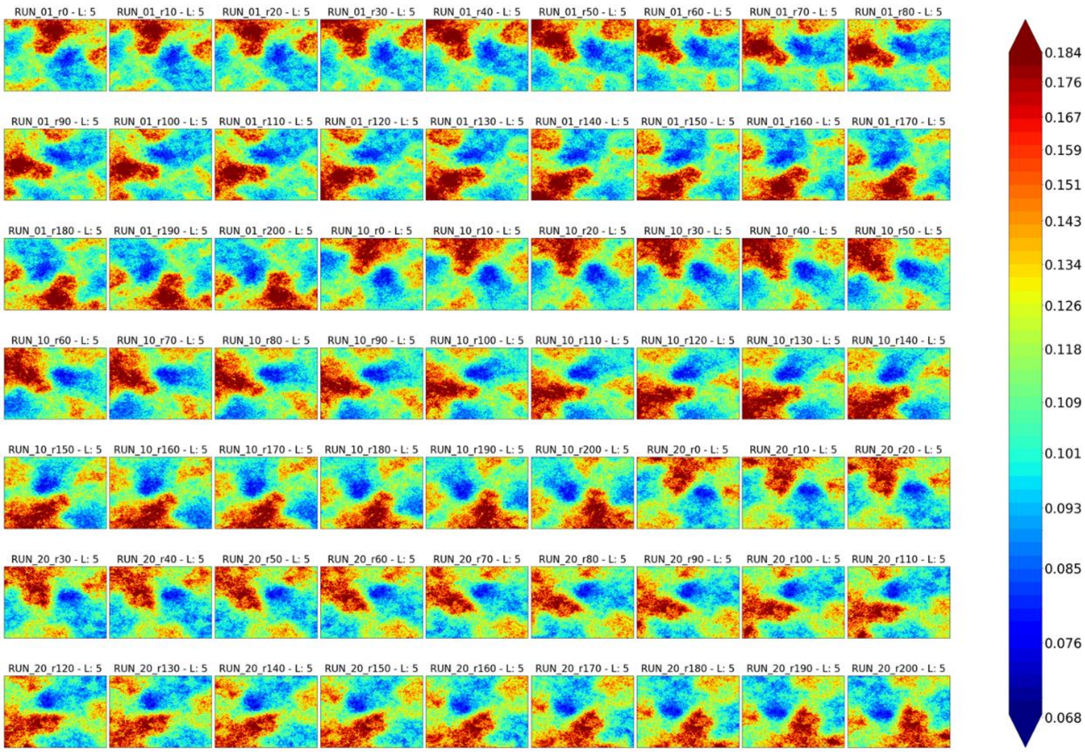


Figure 26. Porosity distribution for Phase 1 of this study for different geological realizations for Layer #5 of the aquifer

2.12.2 Phase 2: Varying Well Number and Location

Despite significant advancements in reservoir characterization and modeling techniques, achieving a perfect match between predictive models and real-world observations remains an elusive goal. An alternative approach to establishing conformance is to demonstrate the progressive refinement of predictive models as new monitoring data accumulates. This approach highlights the robustness of the geological model and modeling assumptions, indicating that additional data leads to continuous model improvement rather than radical overhauls **(Figure 27).**

Motivation:

- *Operational Realism:* Replicates scenarios where well numbers and locations might change due to operational adjustments or optimization efforts.
- *NN Generalization:* NN learns to generalize from scenarios with varying well numbers, enhancing its adaptability.

Advantages and Comparisons:

- *Resource Optimization:* ANN efficiently handles diverse well configurations, reducing the need for extensive numerical simulations.
- *Risk Identification:* Highlights potential risks associated with changes in well numbers, aiding in risk assessment and mitigation planning.

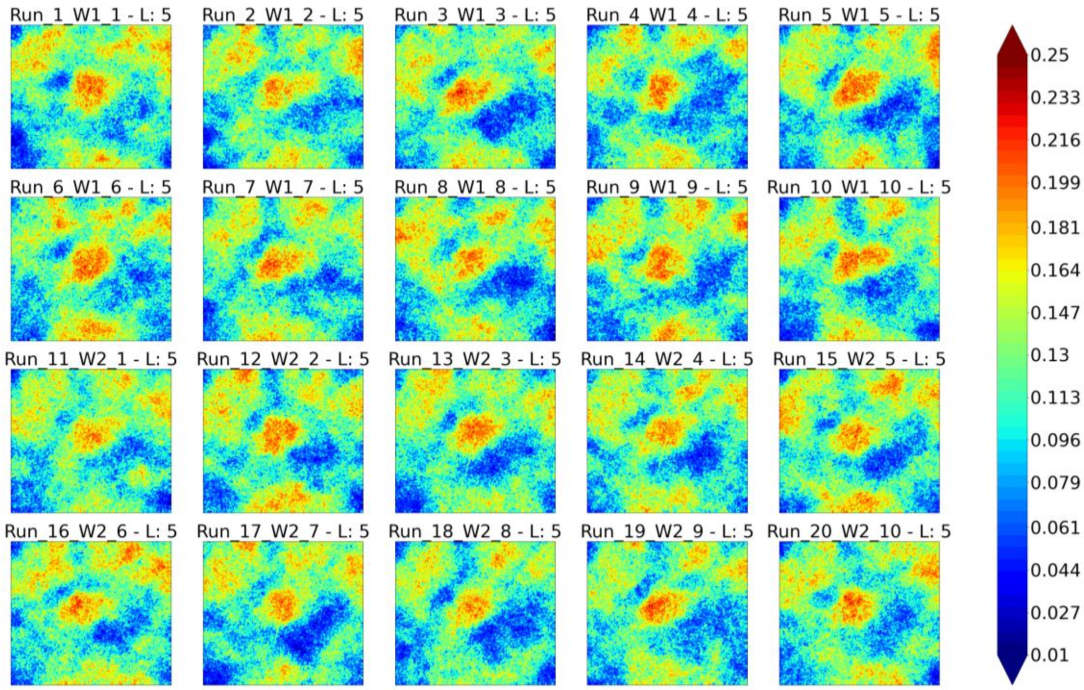


Figure 27. Porosity distribution for Phase 2 of this study for different geological realizations for Layer #5 of the aquifer

2.12.3 Phase 3: Changing Well Locations within Realization

The longevity of free CO_2 entrapment underground poses uncertainties about its ultimate fate and necessitates prolonged monitoring, driving up costs. To address these concerns, expediting CO_2 dissolution and minimizing free CO_2 concentrations in the subsurface presents an effective solution. Cameron and Durlofsky (2012) harnessed the Hooke–Jeeves direct search algorithm to optimize CO_2 injection well placements and rates, seeking to minimize mobile CO_2 within the CCS system over a 1000-year timeframe. Their results demonstrated that the fraction of mobile CO_2 could be reduced from 0.220 to 0.072 under optimal conditions, highlighting the critical role of well location optimization to the potential for significant pressure increases caused by excessive CO_2 injection, which can restrict injection and potentially reactivate preexisting faults, the removal of resident brine from storage formations is considered a solution to mitigate such pressure buildup. An essential challenge in implementing extraction wells is the identification of optimal well locations, which can be achieved through reservoir simulation but is computationally intensive.

In this phase, since the geological realization and reservoir characteristics is the same across different realizations, and only the well locations changes, the porosity distribution for each geological realization is shown in an area 25 by 25 grids around each injection well in each realization **(Figure 28). Figure 29** also shows the entire porosity distribution for one realization layer 5, and relative location of each injection well in the aquifer.

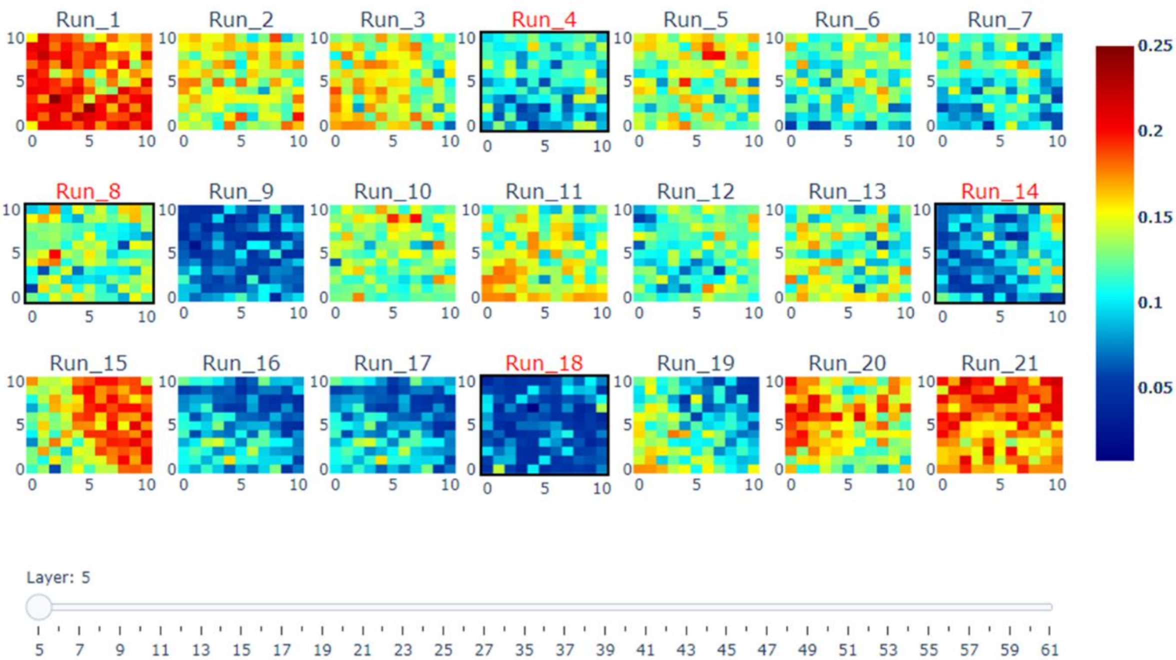


Figure 28. Porosity distribution for Phase 3 of this study for different 25 by 25 grids around different well locations in the same geological realization for Layer #5 of the aquifer. The blind data is highlighted in black squares

Motivation:

- *Localized and Global Uncertainty:* Focuses on introducing uncertainty at a localized level, mimicking challenges in precisely placing wells.
- *NN Adaptation:* Trains ANN to understand the impact of localized changes, improving its ability to predict outcomes.

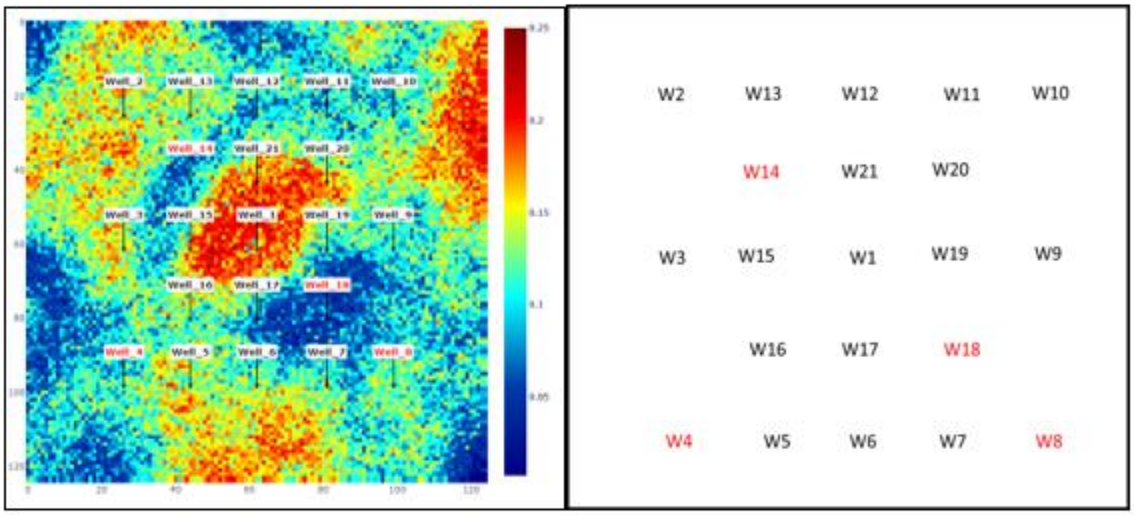


Figure 29. Porosity distribution for Phase 3 of this study for different geological realizations for Layer #6 of the aquifer

Advantages and Comparisons:

- *Localized Decision Support:* Provides insights into the localized effects of well placement changes, aiding in decision-making at specific reservoir areas.
- *Adaptability to Small-Scale Changes:* ANN, having learned from localized changes, becomes adept at predicting outcomes for smaller adjustments.

The porosity distribution of several layers of the reservoir model is shown in **Figure 30 through 33.**

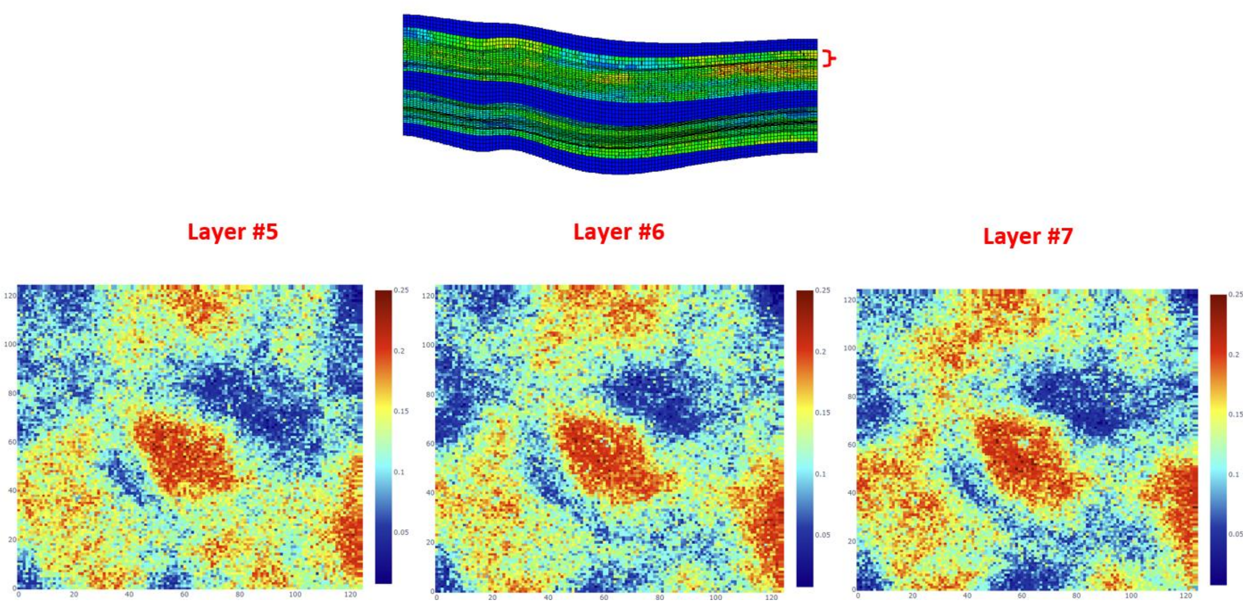


Figure 30. Distribution of porosity for Layers #5, 6, and 7 of the model

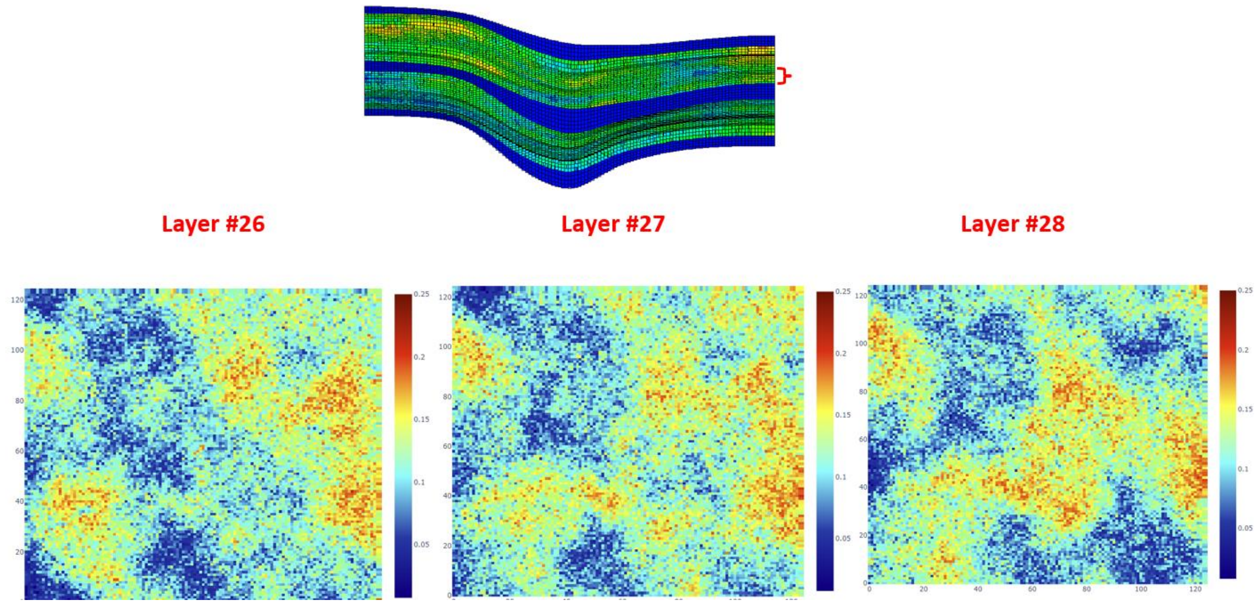


Figure 31. Distribution of porosity for Layers #26, 27, and 28 of the model

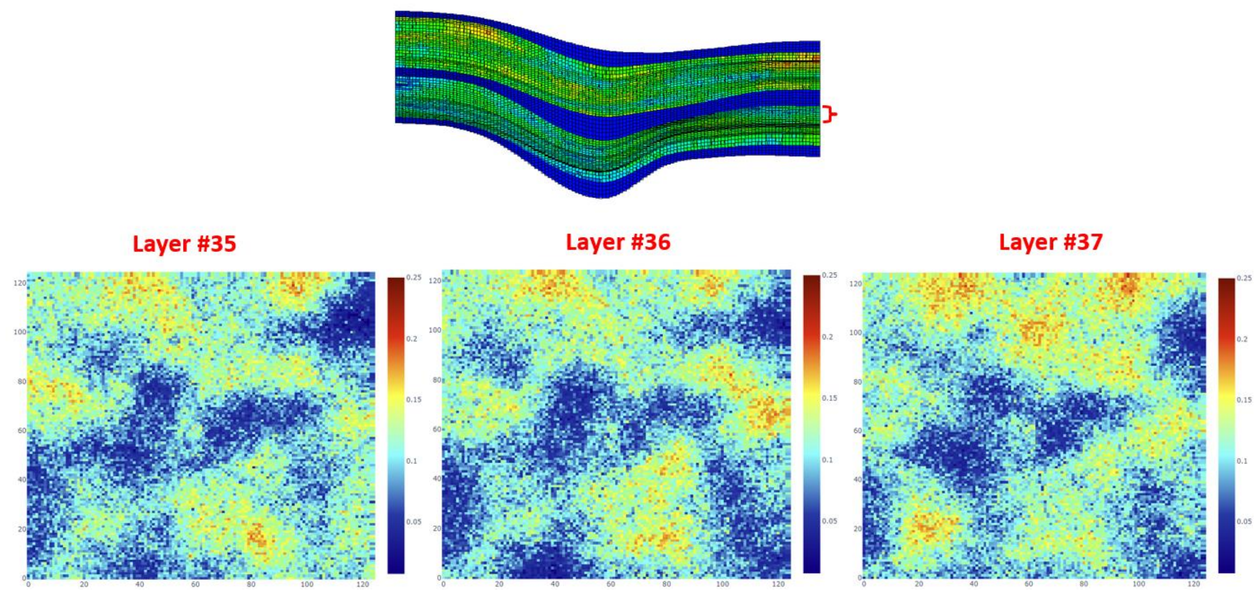


Figure 32. Distribution of porosity for Layers #35, 36, and 37 of the model

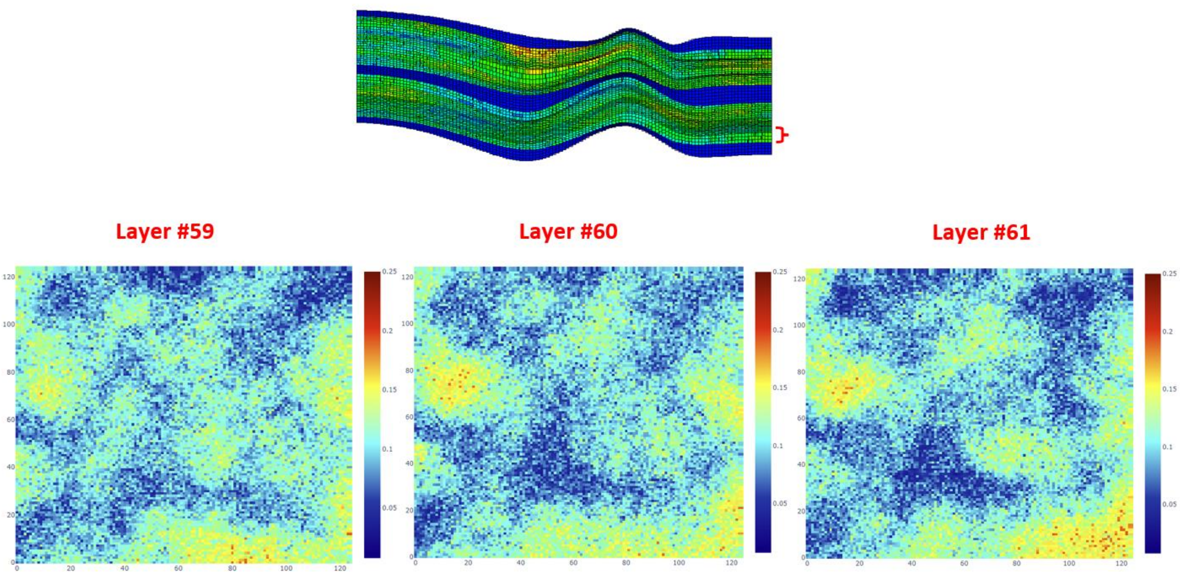


Figure 33. Distribution of porosity for Layers #35, 36, and 37 of the model

The distribution of pressure data for all the geological realizations is also presented in **Figure 34.** As it can be seen, the count of grid cells with different range of pressures is significantly different from one realization to the other.

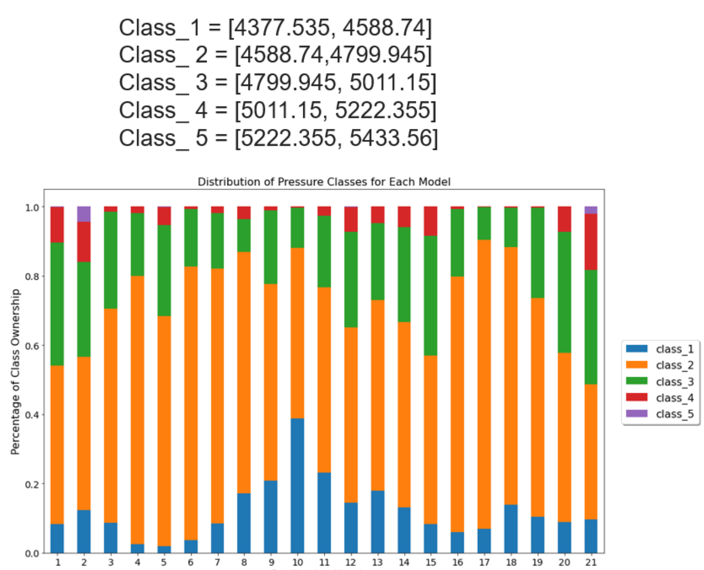


Figure 34. Distribution of actual pressure data for all the realizations in Phase 3

2.12.4 Phase 4: Modifying Wells in Different Realizations

Motivation:

- *Inter-Realization Dynamics:* Explores how changes in well configuration across different geological realizations influence system behavior **(Figure 35).**

- *NN Generalization:* Trains ANN to generalize from scenarios with different well configurations, enhancing its applicability.

Advantages and Comparisons:

- *Interplay Understanding:* ANN captures the interplay between geological variations and well modifications, offering a holistic understanding.
- *Reduced Computational Burden:* ANN efficiently handles diverse well configurations without requiring separate numerical simulations for each case.

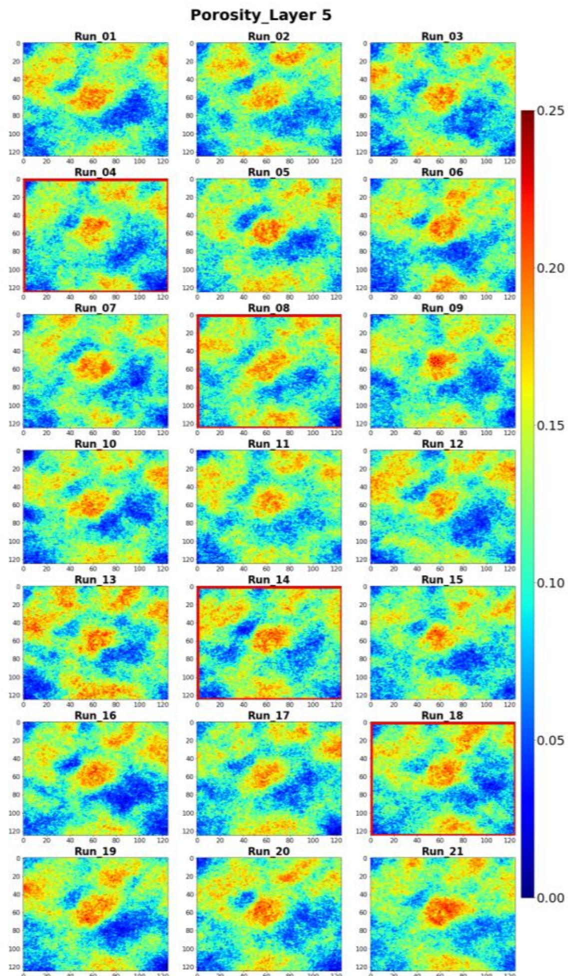


Figure 35. Porosity distribution for Phase 4 of this study for different geological realizations for Layer #5 of the aquifer

2.12.5 Phase 5: Adjusting Well Locations within Same Realization

Motivation:

- *Fine-Tuning Locally:* Investigates the impact of localized well adjustments within a consistent geological setting **(Figure 36).**
- *NN Adaptation:* Trains ANN to adapt to smaller, localized changes, enhancing its precision.

- *Inter-well connectivity:* Well to well connectivity carry implicit information about the reservoir geology.

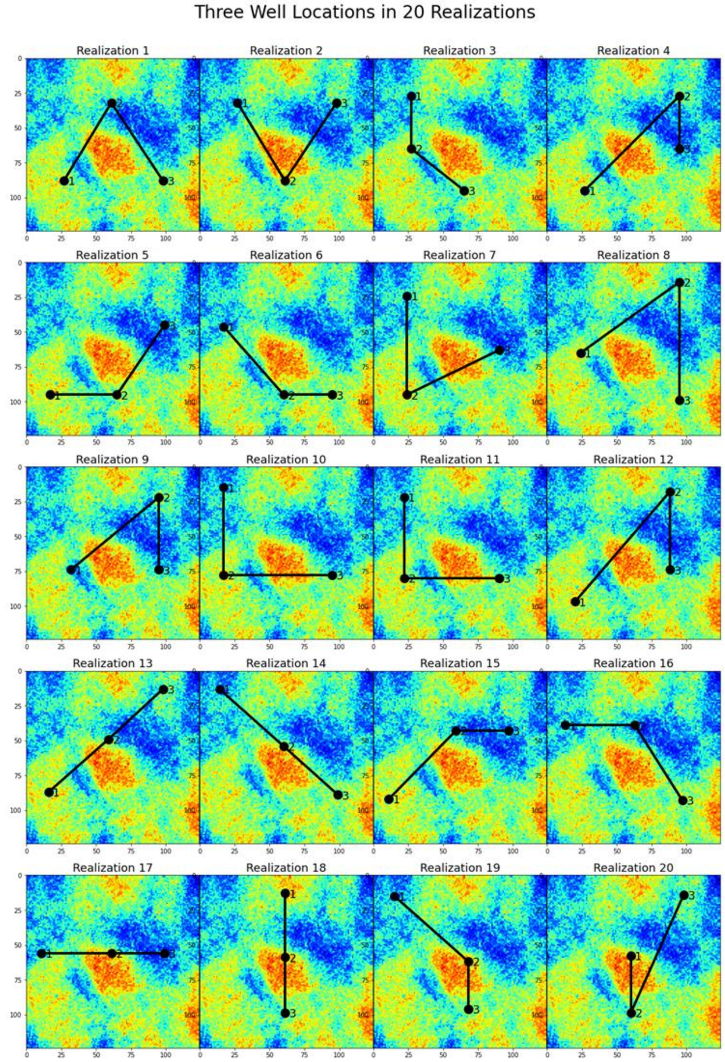


Figure 36. Porosity distribution for Phase 5 of this study for different geological realizations for Layer #5 of the aquifer

Advantages and Comparisons:

- *Localized Precision:* ANN becomes proficient in predicting outcomes for scenarios involving localized well adjustments and different injection well patterns.
- *Resource Savings:* ANN, once trained, provides predictions without the need for extensive numerical simulations for localized changes.

Each phase contributes uniquely to the overall understanding of CO_2 sequestration. The use of ANN, especially after extensive feature engineering, ensures efficient, adaptable, and accurate predictions, surpassing the computational and predictive capabilities of traditional numerical reservoir simulations. The combination of phased research and ANN utilization stands as a robust approach for advancing our understanding and modeling of CO_2 sequestration scenarios.

Chapter 3 – Introduction to Artificial Intelligence and Machine Learning

3.1 Overview of Artificial Intelligence

Artificial Intelligence (AI) encompasses the emulation of human intelligence in computer systems, enabling them to analyze, model, and make decisions. Machine Learning (ML), a subset of AI, involves the use of algorithms that allow computers to learn from data without explicit programming. ML algorithms utilize open computer algorithms to identify patterns in data and make predictions or decisions based on the knowledge acquired through the learning process. Artificial Intelligence (AI) represents a transformative paradigm in computer science, aiming to endow machines with cognitive abilities that mimic human intelligence. The field encompasses a broad spectrum of techniques, ranging from rule-based systems to complex, learning-enabled models. At its core, AI seeks to enable machines to perform tasks that typically require human intelligence, such as problem-solving, pattern recognition, language understanding, and decision-making **(Figure 37).**

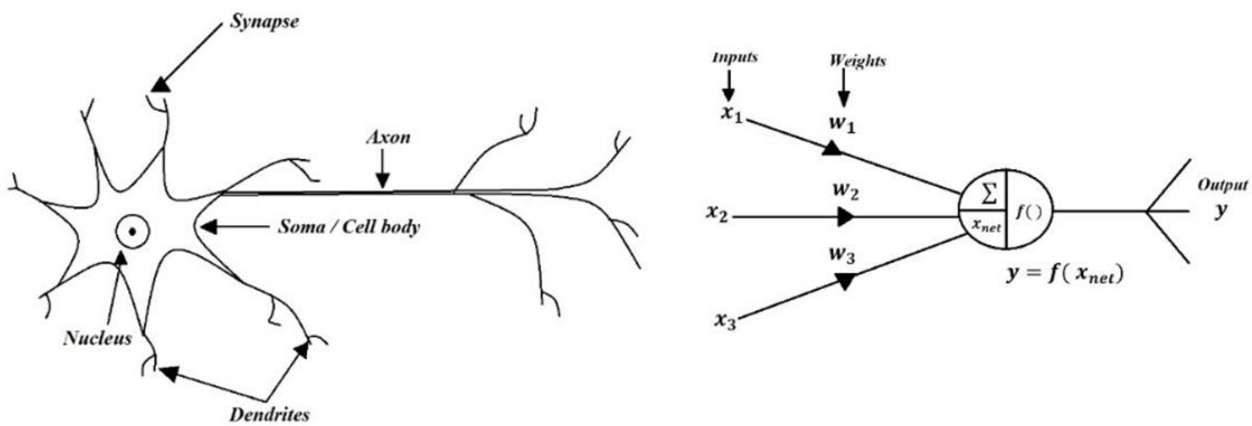


Figure 37. Schematics of biological and artificial neural network

3.2 Evolution of Artificial Intelligence

The roots of AI can be traced back to the mid-20th century when pioneers like Alan Turing and John McCarthy laid the foundation for the field. Early AI systems were rule-based and relied on explicit programming to emulate human reasoning **(Figure 38).** However, progress was constrained by limited computational power and a lack of sufficient data. The resurgence of interest in AI in recent decades can be attributed to advancements in hardware, the accumulation of vast datasets, and breakthroughs in machine learning algorithms. Machine learning, a subset of AI, has emerged as a dominant approach, allowing systems to learn from data and improve their performance over time.

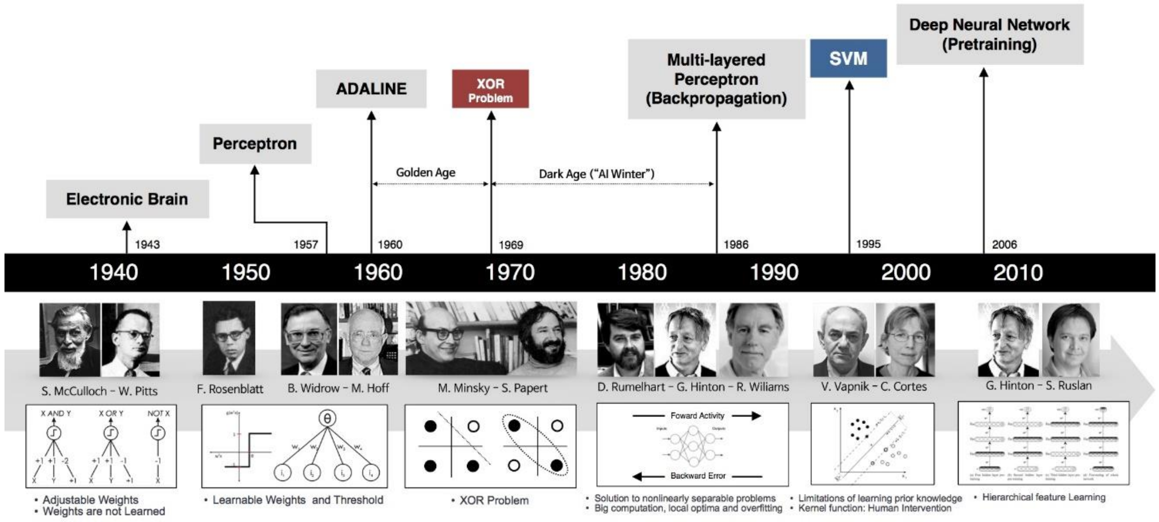


Figure 38. Evolution of artificial neural network from their beginning to present (From Patel and Thakkar, 2020)

3.3 Fundamentals of Machine Learning

3.3.1 Supervised Learning

Supervised learning forms the bedrock of many machine learning applications. In this paradigm, the algorithm is trained on labeled datasets, where input-output pairs are provided. The model learns to map inputs to corresponding outputs, enabling it to make predictions on unseen data.

3.3.2 Unsupervised Learning

Unsupervised learning deals with unlabeled data, seeking to identify patterns and structures inherent in the dataset. Clustering and dimensionality reduction are common techniques, fostering a deeper understanding of the underlying data distribution.

3.3.3 Reinforcement Learning

Reinforcement learning introduces an interactive learning approach, where an agent learns to make decisions by interacting with an environment. Through a system of rewards and penalties, the agent refines its strategy over time.

3.4 Applications of Artificial Intelligence and Machine Learning

The applications of AI and ML span diverse domains, revolutionizing industries and enhancing problem-solving capabilities. From natural language processing and computer vision to healthcare diagnostics and financial predictions, AI technologies are at the forefront of innovation.

3.5 Challenges and Future Directions

Despite remarkable progress, AI and ML face challenges such as interpretability, bias, and ethical considerations. The pursuit of more explainable models and robust ethical frameworks is crucial for the responsible development and deployment of AI technologies. Future directions could involve the integration of AI with other technologies like edge computing, quantum computing, and continued exploration of interdisciplinary applications.

3.6 Evolution of Artificial Neural Networks

3.6.1 Early Foundations (1940s-1950s)

The conceptual groundwork for Artificial Neural Networks can be traced back to the work of Warren McCulloch and Walter Pitts in the 1940s. They proposed a mathematical model of a simplified neuron, inspired by the workings of the human brain. However, early ANNs faced limitations due to the lack of computational resources and data.

3.6.2 Perceptron and Limitations (1950s-1960s)

In the late 1950s and 1960s, Frank Rosenblatt introduced the perceptron, a simple neural network architecture capable of binary classification. Despite initial enthusiasm, perceptron had significant limitations. They could only learn linearly separable functions, limiting their applicability to complex real-world problems.

3.6.3 Neural Network Winter (1970s-1980s)

The 1970s and 1980s witnessed a period known as the "neural network winter" as interest waned due to the realization of the limitations of single-layer perceptron. Theoretical critiques, such as the perceptron's inability to solve XOR-like problems, led to a decline in funding and interest in neural network research.

3.6.4 Backpropagation Resurgence (1980s-1990s)

The resurgence of interest in neural networks occurred with the development of the backpropagation algorithm in the 1980s. This breakthrough allowed the training of multi-layer perceptrons, overcoming the limitations of single-layer models. The discovery of efficient training algorithms facilitated the development of deeper and more complex neural networks. When applying the Gradient Descent (GD) algorithm to Artificial Neural Networks (ANNs), the first step in the process is to feed the ANN's training phase with both input and output data. The inputs enter the ANN and move through a number of layers as they are processed more and more to create a predicted output. To identify errors, this anticipated output is then contrasted with the desired, actual output. These errors, which are passed back through the ANN in a process called backpropagation, are essential to the learning process. To lower these errors, the ANN's weights and biases are changed during backpropagation. Until the errors are reduced to a predetermined tolerance level or a predetermined number of iterations is reached, this iterative process is carried out. One distinctive feature of gradient descent computations is the use of first-order derivatives. To be more precise, it uses the error function's first-order derivative to determine the error space's minimum error value. The gradient must be calculated at each iteration (referred to as iteration 't'). With each iteration, the gradient essentially directs the ANN toward making predictions that are more accurate by providing a direction and magnitude for the weight and bias adjustments. The GD algorithm's ability to train ANNs effectively depends on this methodical, iterative refinement.

3.6.5 Rise of Connectionism and Deep Learning (2000s-Present)

The early 2000s saw a paradigm shift with the advent of connectionism and the rebranding of neural networks as "deep learning." Increasing computational power, the availability of massive datasets, and innovations in model architectures, such as convolutional and recurrent neural networks, fueled the rapid evolution of deep learning.

3.6.6 Deep Learning Revolution (2010s-Present)

The 2010s marked an epochal transformation in the landscape of deep learning, ushering in a profound revolution that transcended the boundaries of artificial intelligence. This transformative era bore witness to groundbreaking achievements across diverse domains, including image and speech recognition, natural language processing, and AI's triumph in mastering complex board games. Iconic successes such as AlexNet, which redefined image classification, and AlphaGo, a milestone in game-playing AI, vividly demonstrated the immense potential of deep neural networks. Additionally, this transformative period witnessed significant advancements in image compression techniques, exemplified by Khoshkhahtinat et al. (2023) and Zafari et al. (2023), stand as testaments to the continual evolution of deep learning and its diverse applications. In parallel, natural language processing saw remarkable strides with the advent of colossal language models like OpenAI's ChatGPT and Google's Bard, redefining human-computer interactions. These language models have unlocked a realm of possibilities, enabling machines to generate remarkably human-like text and engage in intricate dialogues. These advancements, explored in this literature review, represent not just milestones in AI but also crucial stepping stones toward innovative applications across numerous domains, including the core focus of our research **(Figure 39).**

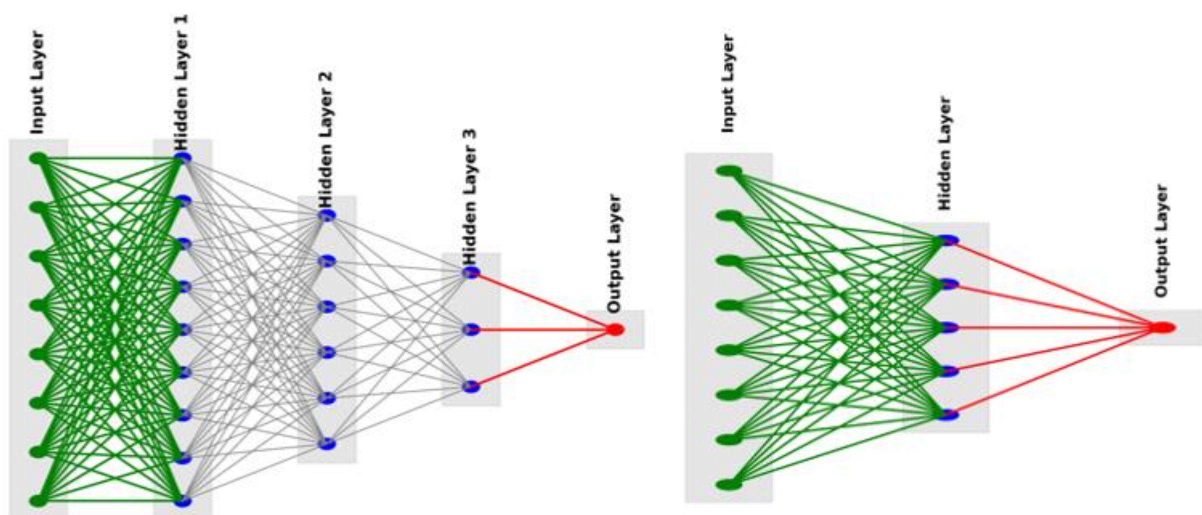


Figure 39. Schematics of ANN model with several (left) and single-hidden layer (right)

Application of artificial neural network in the oil and gas industry started as early as 1990s (Mohaghegh et al. 1991, Mohaghegh et al. 1994, Mohaghegh and Ameri 1995, White et al. 1995, Mohaghegh et al. 2000, Mohaghegh et al. 2000a, Mohaghegh et al. 2000b, Mohaghegh et al. 2000c, Mohaghegh et al. 2005).

3.7 Current Landscape and Future Directions

Today, Artificial Neural Networks are integral to various applications, from image and speech recognition to autonomous vehicles and healthcare diagnostics. The field continues to evolve with ongoing research in explainability, transfer learning, and the integration of neural networks with other AI techniques. Within the realm of the oil and gas industry, a diverse array of data is systematically gathered from both surface and subsurface environments to gain insights into the hydrocarbon potential. Predominantly, sensors emerge as pivotal tools in the collection of vast datasets. The effective analysis of this extensive data necessitates meticulous plotting and technical intervention. To facilitate this process, machine learning methods have proven invaluable by establishing relationships between input variables and predicting corresponding outputs. Notably, these methods achieve this without interference in the physical behavior of the system.

Given the colossal volume of data associated with the oil and gas sector, the task of correlating this information is intricate. The complexity arises from the multifaceted nature of the processes involved. This article delves into the utilization of Artificial Neural Networks (ANN) within the industry, a technology that deals with numerous input and output signals wherein synaptic weights play a crucial role. In the ANN model, the summation of the product of inputs and their respective weights is passed through a transfer function, ultimately yielding the output of the layer.

Learning from data has emerged as a prolific area of exploration across numerous engineering disciplines due to the exponential increase in data volume, surpassing human cognitive capabilities to decipher information and identify patterns within associated datasets. While Machine Learning (ML) exhibits promise in reservoir simulation applications, it is imperative to acknowledge and address certain limitations. Foremost among these concerns is the availability and quality of data, as ML models rely heavily on large volumes of high-quality data for effective training and optimal performance. In reservoir simulation ML applications, the data used for model training is typically derived from conventional simulations conducted offline, ensuring considerable data volume and freedom from noise.

In the context of individual simulation runs, the primary goal is to build ML models that streamline overall simulation runtime by rapidly determining cell-specific parameters. Proxy models, such as those predicting pressure and saturation, replace the need for non-linear solvers, enhancing efficiency in complex phase behavior calculations. In the history matching of a reservoir model, a computationally expensive process, proxy models can optimally calibrate uncertain parameters, achieving a good match between calculated and observed production data. Reservoir simulation software is continually evolving to meet the demands of large data management, despite the increasing availability of computational power. However, these simulations still face challenges in terms of speed and robustness, entailing high computational costs. To address these issues, there is a need for more time-efficient tools capable of providing fast and accurate predictions resembling real reservoir performance within an acceptable error margin.

3.8 ANN Architecture:

In the context of a Multi-Layer Perceptron (MLP) Hyperparameters in a neural network are the external configuration settings that are not learned from the data but are set prior to the training process. In this architecture, the hyperparameters define the neural network's structure and guide the optimization process during training. The choice of activation function in the hidden layer, such as the commonly used Rectified Linear Unit (ReLU), also falls under hyperparameter selection. The ReLU activation introduces non-linearity to the model, enabling it to learn complex relationships within the data. Additionally, regularization techniques like dropout, which involves randomly "dropping out" neurons during training to prevent overfitting, are essential hyperparameters in fine-tuning the MLP's generalization capabilities. Striking the right balance in setting hyperparameters is a crucial aspect of optimizing the MLP's performance for accurate and efficient learning.

Number of Epochs: The number of epochs represents the total number of times the entire training dataset is processed by the neural network. It influences how well the model learns from the data.
Activation Functions: Activation functions introduce non-linearity to neural networks, allowing them to learn complex patterns. Common activation functions include:
ReLU (Rectified Linear Unit): ReLU is widely used and returns zero for negative inputs and the input value for positive inputs, promoting sparsity and faster convergence.
tanh (Hyperbolic Tangent): tanh squashes the input values between -1 and 1, offering stronger gradients during training, especially for centered data.
Sigmoid: Sigmoid transforms inputs to values between 0 and 1, making it suitable for binary classification problems.
Batch Size: Batch size represents the number of training examples utilized in one iteration. Larger batch sizes can improve training speed but may require more memory.
Hidden Layer and Neurons: In a neural network, hidden layers are intermediary layers between the input and output layers. They transform inputs using weights during training. Neurons (or nodes) within hidden layers process input data using weights, activation functions, and biases, contributing to the network's capacity to learn complex patterns.
Optimizers: Optimizers are algorithms that adjust the weights and biases of a neural network during training to minimize the error. Adam (Adaptive Momentum Estimation) optimizer combines ideas from RMSprop and momentum, adapting the learning rates for each parameter individually.
Learning Rate (lr): The learning rate controls the size of the steps taken during optimization. It influences the convergence speed, with smaller values potentially leading to more accurate models but slower training.
lr_reduce_patience: The number of epochs with no improvement after which the learning rate is reduced. This is often used in combination with a learning rate schedule.
Min_lr: The minimum learning rate is the lower bound for the learning rate during training. It helps stabilize training dynamics and is often used with learning rate schedules.
Earlystop_patience: The number of epochs with no improvement after which the training is stopped. This helps prevent overfitting and ensures the model generalizes well.
Earlystop_min_delta: The minimum change in the monitored quantity (e.g., validation loss) to qualify as an improvement. It prevents premature stopping due to minor fluctuations.

Table 7. General hyperparameters for the Pressure ANN **and Table 8.** General hyperparameters for the Saturation ANN show the general hyperparameters used for pressure and saturation ANN models.

Table 7. General hyperparameters for the Pressure ANN

Hyperparameter	Value
Number of Epochs	10,000
Number of Hidden Layers	1
Neurons in Hidden Layer	1,000
Learning Rate	0.001
Batch Size	20,000
Activation_1	'ReLU'
Activation_2	'tanh'
Min_lr	0.00001
lr_reduce_patience	10
earlystop_patience	50
earlystop_min_delta	0.000001
Optimizer	Adam

Table 8. General hyperparameters for the Saturation ANN

Hyperparameter	Value
Number of Epochs	10,000
# Hidden Layer (HL)	1
# Neurons in the HL	1,000
Learning Rate	0.001
Batch Size	20,000
Activation_1	tanh
Activation_2	Sigmoid
Min_lr	0.00001
lr_reduce_patience	10
earlystop_patience	50
earlystop_min_delta	0.000001
Optimizer	Adam

Architectural considerations like the depth and width of a neural network play a crucial role in model design. Deeper networks have the potential to capture more abstract and hierarchical features, although they must be carefully managed to avoid overfitting. Balancing network depth and width based on problem complexity and available data helps strike a harmonious equilibrium between representational capacity and generalization. When it comes to feature hierarchies, tailoring the network architecture to the specific problem domain, for instance, employing convolutional layers for spatial hierarchies in image data or recurrent layers for temporal hierarchies in time series data, enhances the likelihood of learning representations that generalize effectively. Moreover, ensuring model interpretability is essential, and this can be achieved by incorporating techniques like attention mechanisms or visualization methods, facilitating a deeper understanding of the learned representations and decision-making processes within the model.

3.9 Transition to Smart Proxy Modeling Part 2:

Building on the AI and ANN insights outlined earlier, our focus now shifts to smart proxies in CO_2 sequestration projects. This section delves into the transformative impact of smart proxies, where domain expertise converges with cutting-edge machine learning. Representing a paradigm shift, smart proxies tackle computational hurdles and uncertainties entrenched in conventional reservoir simulations. In the upcoming chapter, dissect smart proxies' intrinsic attributes. Explore their potential in mitigating challenges, from computational bottlenecks to uncertainties, reshaping our approach to CO_2 sequestration dynamics. This transition guides readers toward understanding how intelligent proxies redefine CO_2 sequestration simulations, ensuring heightened precision and efficacy in addressing this critical environmental concern.

Numerous data-driven models aiming to optimize brine extraction in CO_2 sequestration have been proposed that leverage machine learning (ML) techniques like neural networks (Musayev et al. 2023). Artificial neural networks (ANNs) have become a crucial tool in data-driven simulations due to their remarkable ability to approximate any measurable function, known as the universal approximation property (Hornik et al., 1989). Their versatility extends across various scientific and engineering domains, showcasing their potential in nonlinear universal approximation and data assimilation (Shanmuganathan, 2016; Abiodun et al., 2018). Wang et al. (2021) introduced an integrated method that combines fluid flow predictions with direct simulation, significantly reducing computation time while maintaining accuracy. Alakeely and Horne (2022) explored the effectiveness of generative deep learning techniques in predicting multiphase flow profiles for new wells in uncharted locations using historical production data and a variational autoencoder algorithm. Dong et al. (2022) proposed a deep reinforcement learning approach for automating curve matching in well test interpretation, leveraging the double deep Q-network. Despite these remarkable advancements, challenges persist in the conventional data-driven approach. Firstly, it is often viewed as a "black box," lacking the incorporation of physical insights, which can lead to predictions that defy physical consistency (Karniadakis et al., 2021). Secondly, the robustness of data-driven models may be limited, particularly in long-term predictions, as they struggle to generalize beyond the parameter space of the training dataset.

Recently, various data-driven models have been developed for predicting the behavior of carbon storage sites during and after injection, utilizing machine learning techniques. Among these, the application of Physics-Informed Neural Networks (PINNs), as explored in studies by Han et al. (2023), Yan et al. (2023), Raissi et al. (2019), Yang and Karniadakis (2020), Zhang and Karniadakis (2020), and Wang et al. (2021), presents notable challenges in reservoir simulation. These techniques, along with hybrid models or hybrid physics-informed data-driven neural network (HPDNN) (Wan et al. 2023, offer potential advantages, such as their ability to handle complex data. However, there are limitations to consider. Firstly, while PINNs enhance interpretability compared to purely data-driven models, they may not achieve the same level of transparency and interpretability as traditional physics-based simulations. This can make understanding the decision-making process of the model more challenging. Additionally, unlike the approach proposed in this study, their primary focus is not on pattern recognition within the data. Secondly, in scenarios involving large-scale computational domains or production wells, PINNs may encounter difficulties due to limited labeled data availability. Although the introduction of decomposition techniques aims to mitigate this issue, it may not entirely eliminate the challenge of data scarcity. Additionally, the hybrid training methods, which combine physical knowledge (PDEs) and sampled data, needs to balance the loss functions associated with these different aspects accurately. This balance is crucial for training efficiency and the model's effectiveness. Also, PINNs must demonstrate robust generalization capabilities to accurately predict outcomes in unseen scenarios and conditions, a task that remains an ongoing challenge. PINN models’ precision diminishes when employed in the context of time-dependent Partial

Differential Equations (PDEs) for extended durations (Meng et al. 2020, Wang and Perdikaris, 2023, Wang et al. 2021). Moreover, considering the physical aspect, the amplification of natural convection calls for high-resolution simulations to adequately represent the rapid spatial and temporal fluctuations in flow phenomena. This, in turn, results in computational expenses that can become prohibitive, as highlighted by Ajayi et al. (2019). Other models require transformation in the spatio-temporal coordinates of the input data which is not transparent (Honghui et al. 2023). Finally, the implementation of PINNs demands a profound understanding of the physics of the problem as it involves the intricate development of appropriate loss functions that faithfully represent the underlying physical processes, contributing to the complexity of their application in reservoir simulation. This could make understanding the decision-making process of these models more difficult. Additional methods include conditional deep convolutional generative adversarial network (cDC-GAN) used to forecast the migration of CO_2 plumes in heterogeneous reservoirs (Zhong et al., 2019). (Wang et al. 2021). No matter what kind of modeling is used, the claimed data-driven models aim to create a connection between input and simulated output by attempting to approximate the dynamic system within the constraints of the training set. The models reliance on simplified models that may not fully capture the complexity of real-world reservoir systems. The simplified models presented are based on idealized assumptions and may not fully account for reservoir heterogeneity. A large training dataset is required to achieve a reasonably accurate and flexible model, which means that many simulations must be run in order to train the model. Consequently, overfitting the training data may make the model's predictions vulnerable and therefore not align with the expected or desired governing physical principles. Additionally, and most importantly, Evaluating the application performance of the HPDNN model in different field scenarios, especially with varying noise factors, is challenging and critical to assess the model's real-world applicability.

Moreover, investigators like Shokouhi and colleagues (2021) have utilized advanced deep learning methodologies, including Long Short-Term Memory (LSTM) and Multilayer Perceptron Neural Networks (MLPNN), for the creation of predictive models for injection simulations. These models are designed to predict the evolving patterns of CO_2 saturation and fluid pressure fields in carbon capture and storage (CCS) systems over time and space. Most recently, Aslam et al. 2023 developed a coarse–grid network model (CGNet) reduced-order model proxy model.

ANNs can be designed with architectures that inherently handle variability and uncertainty. This may involve using larger networks with more parameters to allow for greater flexibility in learning from diverse input data. Robust architectures can adapt to variations in input conditions. Handling uncertainty in input data, especially when dealing with geological properties or initial conditions, is crucial for developing robust and reliable Artificial Neural Network (ANN) models. Uncertainty quantification during training is a critical aspect of training Artificial Neural Networks (ANNs) to recognize and account for uncertainties.

Data augmentation is a technique used to increase the size of a training dataset by generating new samples from the existing ones. This technique can help the training process of a neural network in several ways:

1. Prevent overfitting: By increasing the size of the training dataset, data augmentation can help prevent overfitting, which occurs when a model becomes too complex and starts to memorize the training data instead of learning general patterns.
2. Improve generalization: By generating new samples that are similar to the original ones, data augmentation can help the model learn more robust and generalizable features that can be applied to unseen data.
3. Increase diversity: By introducing variations in the training data, data augmentation can help the model learn to recognize patterns that are invariant to certain transformations, such as rotation, scaling, or translation.

Overall, data augmentation can help improve the performance of a neural network by providing more diverse and representative training data, which can lead to better generalization and more accurate predictions on unseen data (Akyash et al. 2021).

3.9.1 Phases of Complexity an Uncertainty:

3.9.2 Increasing Training Dataset Size:

- Data augmentation is employed to artificially expand the effective size of the training dataset. Instead of relying solely on the original data, the training set is enriched with augmented samples. This is particularly beneficial when the original dataset is limited, as it provides the model with a more extensive and diverse set of examples.

3.9.3 Random Transformations:

Random transformations involve applying various modifications or distortions to the input data. These transformations are random and diverse, introducing variability into the training process. Common transformations include:

- **Rotation:** Rotating the input data by a random angle.
- **Translation:** Shifting the position of the input data horizontally or vertically.
- **Scaling:** Changing the size or scale of the input data.
- **Flipping:** Mirroring the input data horizontally or vertically.
- **Zooming:** Altering the zoom level of the input data.

3.9.4 Robustness to Variations:

By exposing the ANN to different variations of the input data, the model becomes more robust. It learns to recognize patterns and features regardless of transformations, mimicking real-world scenarios where data can exhibit variations due to different conditions, perspectives, or measurement inaccuracies.

3.9.5 Generalization to Uncertainties:

Data augmentation assists the model in generalizing well to uncertainties and unexpected variations in unseen data. Since the training set now includes diverse representations of the input space, the model is less likely to be overly sensitive to specific features present in the original data.

3.9.6 Mimicking Real-World Conditions:

In many real-world situations, data may exhibit variations due to factors such as sensor noise, environmental changes, or measurement errors. Data augmentation simulates these variations during training, preparing the model for challenges encountered in practical applications. In reservoir modeling, augmented data could represent variations in geological properties, initial conditions, or other factors influencing the behavior of the reservoir.

3.9.7 Adaptation to Limited Data:

In scenarios where collecting a large amount of diverse data is challenging or expensive, data augmentation becomes a crucial tool. It enables the model to learn from a more extensive range of scenarios even when the original dataset is relatively small.

3.9.8 Implementation in Training Pipeline:

During the training phase, augmented samples are generated on-the-fly by applying random transformations to the input data before each iteration. This dynamic augmentation ensures that the model encounters a different set of examples in each training step, preventing overfitting to the original data.

The following phases involve introducing various levels of complexity and uncertainty into the design.

Phase 1: In this phase, the objective is to apply rotation to each realization. The geological realization remains the same, and the well locations and numbers are fixed.
Phase 2: The objective of this phase is to vary the number and location of wells in each geological realization. The geological realizations are different but based on the same 10x10 grid. The well locations are fixed, but the well numbers can be up to 2. Delayed injection and inter-well interpolation are also considered.
Phase 3: This phase focuses on changing well locations within the same geological realization. The objective is to introduce uncertainty by adjusting the well locations. Only one well is involved, and there is no delayed injection or inter-well interpolation.
Phase 4: The objective of this phase is to modify wells in different geological realizations. The geological realizations are different, and the well locations are fixed. Up to 3 wells can be involved, and delayed injection is considered. However, there is no inter-well interpolation
Phase 5: In this phase, the objective is to adjust well locations within the same geological realization. The geological realization remains the same, but the well locations are different. Three wells are involved, and delayed injection is considered. There is no inter-well interpolation.

These phases which are summarized in **Table 9.** Different phases of complexity and uncertainty used in this study which provide a framework for introducing different levels of complexity and uncertainty into the design of CO_2 sequestration scenarios. The specific details of each phase can be used to inform the development of smart proxy modeling for predicting the outcomes of these scenarios. These phased approaches systematically introduce varying levels of complexity and uncertainty into the design of CO_2 sequestration scenarios. From maintaining geological consistency to altering well configurations, each phase contributes distinctive elements. Smart proxy modeling, guided by the specifics of each phase, can adeptly predict outcomes, providing a nuanced understanding of the interplay between geological factors and operational variations in CO_2 sequestration projects.

Table 9. Different phases of complexity and uncertainty used in this study

Phase	Objective	Geological Realization	Well Locations	Well Numbers	Delayed Injection	Inter-well Interpolation
Phase 1	Apply rotation to each realization	Different	Fixed/ Same	4/ Same	No	No
Phase 2	Vary the number and location of wells in each geological realization	Different But Same 10x10 grid around Injection Wells	Fixed/ Same	Up to 2	Yes	Yes
Phase 3	Change well locations within the same geological realization	Same	Different	1	No	No
Phase 4	Modify wells in different geological realizations	Different	Fixed/ Same	Up to 3	Yes	No
Phase 5	Adjust well locations within the same geological realization	Same	Different	3	Yes	No

- **Geological Realization:** Whether the geological model remains the same or changes.
- **Well Locations:** Whether the locations of wells change and, if so, how.
- **Well Numbers:** The number of wells involved.
- **Delayed Injection:** Whether there is a delay in injection timing of different wells.
- **Inter-well Interpolation:** Whether porosity and permeability are interpolated between wells.

The phased approach to CO_2 sequestration scenario exploration serves various critical needs and yields several benefits. Firstly, it allows us to consider geological variations, well configurations, and operational conditions, providing a holistic view of potential scenarios. Secondly, to proactively complexities of each phase, the model was systematically introduced operational or reservoir characteristic changes, aiding in identifying potential risks and facilitating the development of contingency plans. Phased scenarios provide decision-makers with detailed insights into the impacts of different geological and operational choices, empowering them to make informed choices based on a spectrum of possibilities. Additionally, the refined predictive accuracy achieved through phased modeling, accounting for different layers of complexity, contributes to more reliable outcomes. Finally, phased approaches lay the groundwork for smart proxy modeling, offering insights into the design and training requirements for intelligent proxies in CO_2 sequestration.

Chapter 4 – Smart Proxy Modeling

4.1 Introduction:

The ultimate goal of developing an Artificial Neural Network (ANN)-based smart proxy model is to ensure its validation against blind or unseen datasets. This validation process is of paramount importance as it assesses the model's generalization ability, its capability to make accurate predictions on data it has never encountered before. Validating the smart proxy model against unseen datasets is crucial for several reasons. First, it confirms that the model has not merely memorized the training data but has genuinely learned underlying patterns and relationships. Second, it demonstrates the model's reliability and robustness in making predictions for real-world scenarios, which often involve data variations and unforeseen patterns. Third, it instills confidence in the model's predictive performance, making it a trustworthy tool for decision-making and applications beyond the training dataset. Ultimately, the validation against blind datasets is the litmus test that ensures the ANN-based smart proxy model's effectiveness and usefulness in practical, real-world applications.

4.2 Definition and Purpose

Smart Proxy Modeling (SPM) is characterized as an instantaneous replication of a reservoir simulation model. These models integrate neuro-fuzzy systems that are interconnected and trained to understand fluid flow behavior from reservoir simulation models, enabling real-time reproduction of results with high accuracy. SPMs can be tailored for various simulation models, such as black oil, compositional, and dual porosity models, offering significantly faster run times compared to conventional reservoir simulation models, which may take hours or days.

Inputs to SPMs encompass static elements (reservoir characteristics, boundary conditions, well configurations) and dynamic components (extracted from simulation model runs), creating an integrated spatio-temporal database. Outputs include pressure or rate profiles at wells and pressure/saturation distribution over time at each grid block.

SPMs are trained with a minimal number of simulation runs through innovative spatio-temporal data management inspired by fluid flow physics. This involves careful data generation during the construction and analysis stages of the reservoir model, reducing simulation run requirements.
Expertise in reservoir engineering is crucial for successful AI-related projects like SPM development, requiring an understanding of how AI learns, differentiation of its features from statistics, and effective communication of ideas to AI through data. SPMs also serve as powerful tools for data mining analysis of numerical reservoir simulation models, uncovering valuable information within extensive resources.

AI & ML technologies offer distinct approaches beyond traditional statistics, particularly in modeling the physics of fluid flow in porous media comprehensively. Acknowledging the simplicity of certain tasks addressed by AI and ML does not diminish the importance of using these technologies but clarifies the nature of problems being solved and distinguishes algorithmic approaches from other problem-solving methods.

The traditional approach to modeling and solving engineering problems involves careful observation of the physical phenomena, identification of parameters, understanding their interactions, and often starting with fundamental physics principles. This approach leads to the development of mathematical equations representing the modeled phenomenon. The complexity of these equations determines whether analytical or numerical methods are used to find solutions.

1. **Analytical Method:** Involves simplifying the model to reach a quick and responsive solution. Often used in well testing, this method uses radial cylindrical coordinates for single well solutions.
2. **Numerical Method:** Utilizes Cartesian coordinates for multiple wells or full-field models. The model is solved by approximating the solution, commonly employed in numerical reservoir simulation.

Bahrami et al. (2022) summarizes different classification of proxy modeling which are used in the literature under different names **(Figure 40).** Physics-driven models are entirely based on physical laws, ideal for well-understood systems but limited in complex, poorly-understood scenarios, which is the case for most of the time. Physics-inspired models loosely incorporate physical principles, offering flexibility but potentially sacrificing accuracy. Hybrid models blend physics-based, statistics and data-driven approaches such as MFM, or ROM, aiming to balance accuracy with computational feasibility but mostly inaccurate when applied to validation cases as hybridization can never be well-structured or the underlying physics is poorly represented.

Domain-expertise driven ANNs, like those developed here, integrate domain-specific knowledge, enhancing accuracy, relevance, and interpretability, and reducing overfitting. These models leverage neural networks' pattern recognition capabilities while ensuring outputs align with domain-specific understanding, making them particularly suitable for applications where domain knowledge is essential for success.

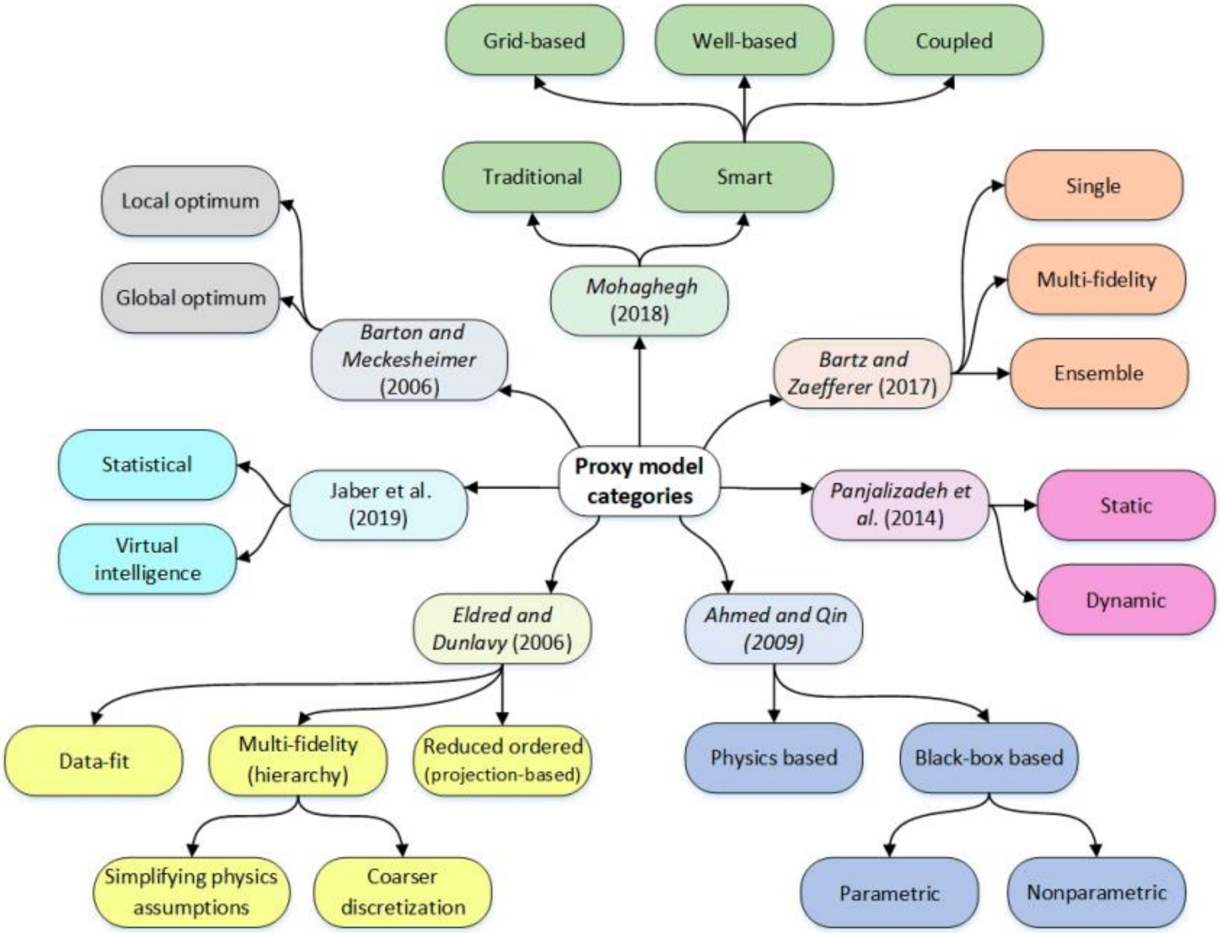


Figure 40. Different classification of proxy modeling in the literature (from Bahrami et al. 2022)

Smart Proxy Modeling for Numerical Reservoir Simulation (NRS) can be realized through two distinct methodologies: Well-based Smart Proxy Modeling and Cell-based Smart Proxy Modeling The specific approach chosen for developing Smart Proxy Modeling in NRS is determined by the goals and objectives that need to be achieved. Classification can also be based on the desired output level (grid-based model, well-based model, and fully-coupled model), based on the type of field the technology is being applied to (green fields vs. brown fields), or based on their functionality (models built for history matching purposes or models for predictive and field development purposes). SPMs are mainly classified based in the main source of the data used to develop the spatio-temporal database that forms the foundation of the model. If the source of the spatio-temporal database is a numerical reservoir simulation model, then the model will be called a Cell-based Smart Proxy Modeling. If the source of the spatio-temporal database is actual field data (historical production data, well logs, cores, well test, seismic attributes, etc.) then the model will be called a Top-Down Model (TDM). Many papers have been published in recent years that demonstrate the applicability of Top-Down Modeling in building reservoir simulation models for many different types of reservoirs from tight gas formations, to shale plays to sandstone and finally naturally fractured prolific carbonate reservoirs of Gulf of Mexico and the Middle East

(Kalantari and Mohaghegh 2009, Kalantari and Mohaghegh 2010, Grujic, 2010, Zargari, 2010, Mohaghegh, 2010). The TDM is however, not the focus of this study.

To develop well-based Smart Proxy Models, data from every production and injection well featured in the Numerical Reservoir Simulation (NRS) are utilized. This approach hinges on the diffusivity equation, a critical element in NRS, which facilitates the physical modeling of fluid flow through porous media. This process inherently considers the spatial and temporal dimensions, which significantly influence the activities and results of the wells. Therefore, the compilation of these models necessitates a comprehensive collection of spatial and temporal data from the NRS for each individual well. Data from all production and injection wells included in the NRS are used to create well-based Smart Proxy Models. This method considers the diffusivity equation, a key component of NRS, which allows for the physical modeling of fluid flow in porous media where spatial and temporal aspects impact well activities and outcomes. Thus, the creation of these models requires space and temporal data from the NRS for every well.

In traditional history matching of numerical reservoir simulation models, the focus is primarily on modifying reservoir characterization, such as permeability, to achieve a reasonable match with observed data. This approach is based on the assumption that the functional relationships between production and reservoir characteristics are constant and unchangeable due to a good understanding of the underlying physical phenomena. Engineers feel comfortable adjusting these reservoir characteristics, represented by a static geo-cellular model developed by geoscientists, as it contains interpretations and uncertain values. In contrast, AI-based reservoir modeling introduces modifications to the traditional assumptions. Instead of maintaining constant and deterministic functional relationships, AI-based models allow these relationships to change. This flexibility extends to the possibility of modifying reservoir characteristics as well. Rather than relying on fixed relationships, AI-based models leverage pattern recognition technology to seek functional relationships between production and reservoir characteristics. Once a set of reservoir characteristics is identified and accepted by geoscientists, it remains unchanged during the history matching process, while the functional relationships are adjusted until a match is achieved. AI-based reservoir modeling departs from the conventional use of first-principle physics, opting for a pattern recognition approach inspired by physics. Instead of explicitly formulating physics, a spatio-temporal database is developed, capturing changes in pressure over space and time. This database serves as the foundation for training a predictive model by adjusting the free parameters that represent interconnections between parameters. The goal is to deduce the physics from observations implicitly. This shift in approach aligns with the concept of data-intensive science, the fourth paradigm, where massive data collection and model building based on data are central. The five major steps in an AI-based reservoir modeling project involve the development of a spatio-temporal database, simultaneous training and history matching of the reservoir model, designing field development strategies using fuzzy pattern recognition, sensitivity analysis, and uncertainty quantification, and finally, application of the model in predictive mode to address reservoir management questions. In this context, the approach diverges from building explicit physics-based models and instead draws inspiration from how humans navigate complex situations without relying on such models. Humans excel in controlling complex machinery and solving intricate problems through observation and pattern recognition. AI-based reservoir modeling seeks to replicate this innate human ability. Rather than directly applying physics principles in their first-principle and explicit form, AI-based modeling leverages physics, particularly our scientific understanding of fluid flow through porous media, as a source of inspiration. This inspiration is used to construct a repository of clever observations, essentially a customized spatio-temporal database. This database serves as the foundation for training a predictive model.

During the training process, the model modifies the free parameters that govern the connections between various parameters, allowing it to adapt and improve its performance over time. As the algorithm continues to learn from the data, it converges to a state where it can effectively mimic the behavior of a hydrocarbon reservoir. In essence, instead of explicitly formulating the physics, AI-based reservoir modeling aims to deduce the underlying physics from observations, adopting an implicit and data-driven approach.

4.3 Case Studies of Smart Proxy Modeling:

An early surrogate reservoir model (called SPM) was developed in 2006 by Mohaghegh (2006), and it faithfully represented a sophisticated full-field reservoir model. As described by Mohaghegh and others in (2012a and c), these surrogate models were then used for geological uncertainty analysis in a number of genuine fields in Saudi Arabia. Mohaghegh et al. made additional changes in 2012b and 2015, dividing SPM into two groups: well-based and grid-based. Well-based SPM concentrates on well-level analysis, whereas grid-based SPM is utilized for numerical model analysis at the grid block level. As mentioned in Mohaghegh et al. (2012b), the grid-based technique has been utilized in a number of CO_2 sequestration projects, and the well-based approach has been applied to production scheduling optimization in a field in the United Arab Emirates (Mohaghegh et al. 2015). Furthermore, the use of SPM spread to new fields like enhanced oil recovery (EOR) and history matching. He and associates combined SPM and differential evolution (DE) in 2016 to enable automated history matching. A SPM was created in 2016 by Alenezi and Mohaghegh to model and forecast the dynamics of a reservoir that has been filled with water. In 2018, Mohaghegh investigated the application of SPM as a storage method for CO_2-EOR. Additionally, SPM was used by Parada and Ertekin in 2012 to develop a new screening method for four different improved oil recovery (IOR) methods, including waterflooding, miscible injection of CO_2 and N2, and steam injection.

In a case study for ADNOC (Abu Dhabi National Oil Company), the effectiveness of Smart Proxy Modeling was demonstrated in optimizing oil production in a mature field with over 160 wells, particularly addressing the challenge of balancing oil output with water cut control. The study focused on peripheral water injection and imposed production restrictions for individual wells. Using a well-based Smart Proxy Model, ADNOC was able to evaluate and prioritize wells based on their potential to boost oil production without significantly increasing water cut. This approach facilitated rapid simulation and analysis, leading to the implementation of a rate relaxation program in 2006 for selected wells. Follow-up "look-back" studies in 2010 validated the Smart Proxy Model's accuracy and efficacy in enhancing field management decisions and optimizing production, as reported by Mohaghegh (2022). In another case study for ARAMCO (Mohaghegh et al. 2012, Mohaghegh 2022), the potential for oil production in a greenfield with carbonate reservoirs was explored using a detailed reservoir simulation model. This model, featuring 1.4 million active cells and 62 wells, was used to develop a Smart Proxy Model through various simulation runs under different operational constraints. The process involved building a comprehensive spatio-temporal database to reflect the reservoir's fluid behavior, identifying key performance indicators for oil and gas production, and implementing dynamic allocation of reservoir volume across wells. The Smart Proxy Model was trained and validated to accurately predict outcomes under unencountered conditions. Its success in replicating reservoir simulation results highlights its effectiveness in sensitivity analysis, uncertainty quantification, and optimizing field development strategies, especially in fields with limited data and high uncertainties.

Cell-based Smart Proxy Modeling is used in Computational Fluid Dynamics (CFD) and Numerical Reservoir Simulation (NRS) in a similar way, even though CFD is more common in mechanical, chemical, and other engineering domains than in reservoir simulation. In general, this approach uses machine learning and artificial intelligence (AI) to build precise AI-based proxy models for a variety of numerical simulation models. To create Smart Proxy Models for numerical simulations utilized in domains such as meteorological forecasting, the same methodology can be expanded upon. This chapter will first explain how cell-based Smart Proxy Modeling is applied in NRS, and then it will look at some recent NRS and CFD (Ansari et al. 2019) case studies. Cell-based Smart Proxy Modeling in Numerical Reservoir Simulation (NRS) entails building an AI-based model that records the pressure and saturation for every grid block in the NRS at every time step. There are two stages to this kind of modeling implementation: prior to and following history matching. During the pre-history matching stage, a cell-based Smart Proxy Model for the original NRS is created. This model makes use of the original geological model that geologists and geophysicists contributed for reservoir modeling. It is well known that this first numerical simulation of the reservoir usually does not produce a satisfactory history match. After developing a Smart Proxy Model for such a Numerical Reservoir Simulation (NRS)—which includes phases like training, calibration, and validation—a single deployment of this Smart Proxy Model can deliver results with over 95% accuracy for reservoir simulation. This accuracy pertains to parameters like reservoir pressure and saturation for every cell at every time-step. Remarkably, such a deployment can be completed within minutes, even on standard desktops and laptops.

4.4 Petroleum Data Analytics:

Petroleum Data Analytics involves the application of AI and ML in petroleum engineering problem-solving. It integrates expertise in petroleum engineering and AI/ML. The success of this technology depends on recognizing the differences between engineering and non-engineering problem-solving and understanding how AI/ML differs from traditional statistical analysis.

4.4.1 Characteristics of Petroleum Data Analytics Experts:

1. **Domain Expertise:** Profound knowledge and experience in petroleum and geo-science-related areas (Drilling, Reservoir, Completion, Production & Facilities). Avoiding misjudgments on how AI/ML should be incorporated is crucial.

2. **AI & ML Practice Expertise:** Understanding AI & ML algorithms contributing to engineering problem-solving, including artificial neural networks, fuzzy set theory, and evolutionary computation. Studying the biology of the human brain aids in comprehending AI & ML approaches.

3. **Practicing AI & ML:** Gaining expertise in the application of AI & ML in engineering problem-solving. Distinguishing between traditional engineering problem-solving and AI & ML approaches is vital.

4.5 Traditional Approach vs. AI & ML:

Traditional approaches begin by building an understanding of the physics of the phenomenon, using fundamental equations, and solving complex mathematical models analytically or numerically. AI & ML approaches simulates the human brain, using algorithms to learn from data. It doesn't rely on fundamental physics equations but learns patterns and correlations directly from data. Becoming a proficient Petroleum Data Analytics expert involves combining domain expertise in petroleum engineering with a deep understanding and practice of AI & ML techniques, differentiating between traditional and AI-based problem-solving approaches **(Figure 41).**

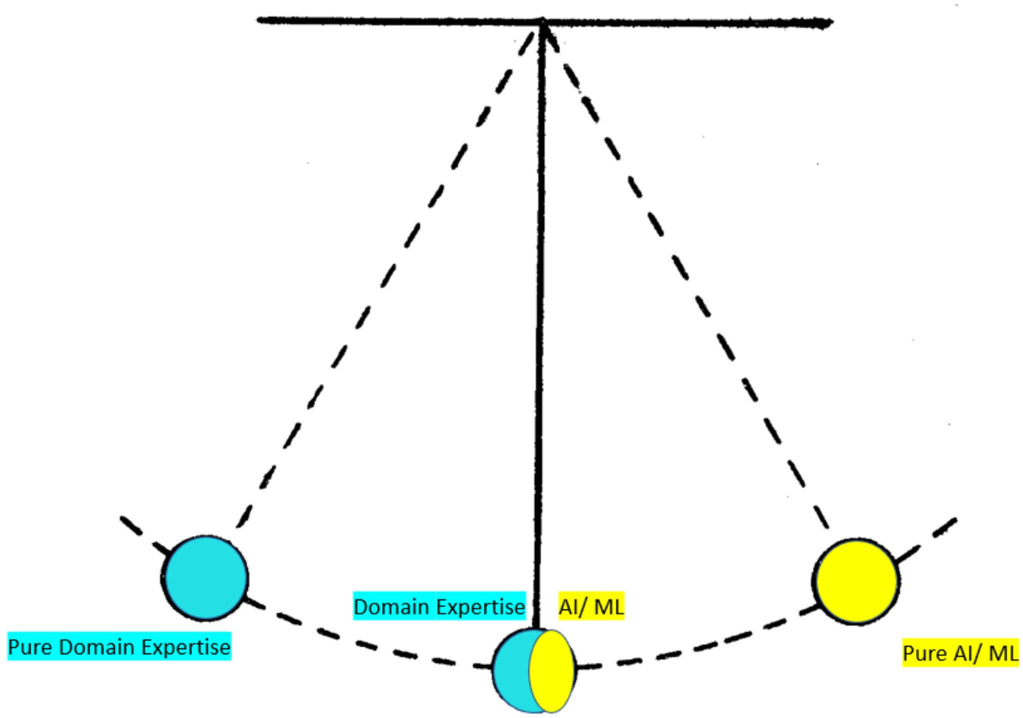


Figure 41. Striking a balance between domain expertise and AI/ML

4.6 Smart Proxy Models: Applications and Previous Works

In the current era of advanced computer technologies, the storage and processing of large volumes of data, accessible from remote locations via computer networks, have become possible. Reservoir simulation, a branch of reservoir engineering, employs computer models to predict fluid flow through porous media, such as oil, water, and gas. Initiated as early as 1954 with radial gas-flow simulations, computational advancements have significantly enhanced reservoir simulation, particularly in terms of model size and resolution.

The landscape of reservoir simulation users and their roles has expanded with the emergence of high-speed computers and the electronic revolution. Reservoir modeling, once sophisticated and expensive, has transformed into a practical toolbox, accessible even on a student's laptop. Despite its established status, reservoir simulation continues to evolve, playing a crucial role in modern reservoir development decisions. Recent technological and software advances prompt consideration of automated data analysis tools, with machine learning emerging as a prominent candidate.

Smart Proxy Modeling, also known as surrogate reservoir modeling, represents a computationally efficient alternative to numerical simulation, generating complete results for a Numerical Reservoir Simulation in a matter of seconds (Mohaghegh 2018). While Smart Proxy Models may not precisely replicate numerical simulation outcomes, the presented outputs exhibit an acceptable range of error, making them considerably valuable due to their swift processing compared to the time-intensive nature of numerical models (Amini et al. 2014; Alenezi and Mohaghegh 2017; Mohaghegh 2015; Mohaghegh et al. 2015; Gholami et al, 2019). These models, classified as well-based or grid-based depending on their purpose, have seen extensive application in well-based scenarios but limited exploration in dynamic grid-based proxy modeling for complex simulations like CO_2 sequestration into saline aquifers.

Numerous case studies have validated the effectiveness of Smart Proxy Modeling, with applications ranging from shale gas estimation to conventional oil fields and CO_2 sequestration projects (Shahkarami et al. 2014; Mohaghegh 2014; Mohaghegh 2011; Mohaghegh et al. 2009a; Amini et al. 2014; Jalali et al. 2009). The well-based Smart Proxy Models aim to simulate reservoir responses at well locations, while the grid-based counterparts allow users to replicate dynamic reservoir parameters at any desired time or location (Mohaghegh et al. 2012; Shahkarami et al. 2014; Gholami et al. 2019; Alenezi and Mohaghegh 2017). Notably, the technology's success has been demonstrated in a large oil field in Saudi Arabia, capturing pressure and saturation changes with high accuracy (Kalantari et al. 2011; Mohaghegh et al. 2012). In a CO_2 sequestration context, a grid-based Smart Proxy Model showcased minimal error in predicting gas saturation at the grid level compared to numerical models (Amini et al. 2014). Additionally, Smart Proxy Modeling has proven computationally more feasible than traditional numerical reservoir simulations, as demonstrated in optimizing reservoir injection strategies (Gholami 2014).

In a specific field application in Scurry County, Texas, Smart Proxy Modeling at the grid block level was implemented using a cascade training and validation method. This innovative approach allowed the smart proxy model to autonomously derive dynamic data sources, enhancing its adaptability to changing conditions (Alenezi and Mohaghegh 2017). The complexity of production performance and geological characterization in the SACROC unit field made it an ideal testing ground, requiring multiple reservoir simulation scenarios for smart proxy model construction. Key geological parameters and simulation results were used to create a spatial-temporal database, and the trained model exhibited high accuracy when validated on blind runs. Another grid-based Smart Proxy Model was developed to construct well-production profiles, providing valuable insights into field performance with precision (Alenezi 2017).

Most recently, Alabboodi (2021) conducted a study that developed Dynamic Smart Proxy models using Artificial Neural Networks (ANN) for Geologic Carbon Dioxide Storage (GCS) simulations for all the timesteps from 2020-01-01 to 01-01-2320, maintaining a fixed number of injection wells and locations across all geological realizations. The study demonstrated high accuracy in predicting both pressure and CO_2 saturation, significantly streamlining the computational process of subsurface modeling in GCS operations. In another study and dataset, Al Nuaimi (2022) conducted a study on the application of Smart Proxy Models (SPM) using Artificial Intelligence (AI) and Machine Learning (ML) to simulate and monitor CO_2 injection into saline aquifers for climate change mitigation. The study's goal was to replicate pressure and saturation results of a numerical reservoir simulation model (CMG) over a period of 10 years of CO_2 injection and 190 years post-injection, using four injectors. Utilizing 46 geological realizations and extensive data, the AI model accurately predicted pressure distribution and CO_2 plume behavior, demonstrating potential as a competitive alternative to traditional numerical reservoir simulators in petroleum engineering and CO_2 sequestration. Al Nuaimi (2022) conducted a study on using Smart Proxy Models (SPM) with AI and ML for simulating CO_2 injection into saline aquifers, despite the challenge of a small plume dataset. The study aimed to replicate the pressure and saturation results of a numerical reservoir model over 10 years of injection and 190 years post-injection. To enhance the model's learning process with the limited data, various techniques were explored, successfully predicting pressure distribution and CO_2 plume behavior and highlighting SPM's potential in petroleum engineering and CO_2 sequestration.

In the context of this study, a grid-based proxy model is developed to rapidly generate complete results for a numerical reservoir simulation (CMG's GEM) over multiple time steps, focusing on CO_2 injection and post-injection phases. Employing a phased complexity approach, this Smart Proxy Model predicts pressure distribution and CO_2 saturation. Smart Proxy Models have evolved over time, contributing to various reservoir scenarios. Initially applied at the well level, Mohaghegh extended the concept to grid-block levels, capturing pressure and saturation changes across large reservoirs with precision. Smart Proxy Models have been validated in studies ranging from shale gas deposits to CO_2 sequestration projects, showcasing their accuracy and efficiency.

4.7 Unique Features and Integration

Smart Proxy Models distinguish themselves by relying on artificial intelligence and data mining. They replicate traditional reservoir simulations accurately, achieving high speeds compared to conventional numerical simulators. The integration of reservoir engineering, modeling, and machine learning positions smart proxy modeling as a comprehensive solution for various geological scenarios. The advancement of proxy models, especially with the introduction of smart proxy modeling, has transformed the landscape of reservoir modeling and simulation. These techniques, backed by machine learning and data mining, offer efficient and accurate alternatives, addressing challenges associated with traditional methods. Despite challenges, the evolution of these models indicates a promising future in optimizing reservoir engineering procedures and enhancing our understanding of fluid flow behavior in porous media.

4.8 SPM developed in this study:

The original data for this study was a adjusted to be used in this study and is collected from Citronetelle aquifer located in Mobile County, Alabama **(Figure 42).** The CO_2 injection is performed in a deep saline formation (Paluxy) at a deptjh of 9400 ft. After designing numerical reservoir simulation runs, the next step involved running these simulations and extracting essential data. This data was used to create a spatio-temporal database, forming the foundation for the smart proxy model. The twenty simulation runs generated a substantial amount of data, exceeding 1.25 TB.

Building a reliable Artificial Neural Network (ANN) model required domain knowledge to prepare the dataset effectively. Understanding the fluid's physics and flow behavior in the reservoir model was crucial for creating a suitable dataset for the Smart Proxy. Reservoir parameters were thoroughly analyzed at each time step and layer, and additional features were generated to enhance Smart Proxy Models' accuracy. The dataset preparation was a crucial step in developing a SPM, emphasizing the importance of data integrity for SPM output reliability.

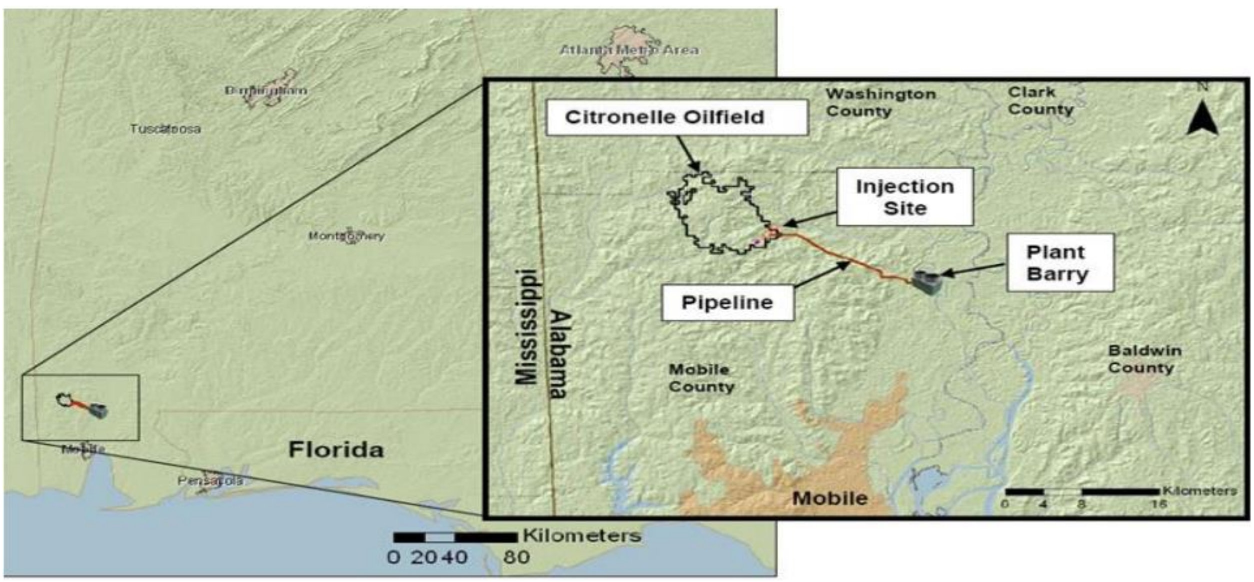


Figure 42. Location of the saline aquifer (NETL 2015)

4.8.1 Data Processing:

Raw data from numerical reservoir simulation runs were processed and translated into a readable form. Python was employed to read simulation outputs/inputs, extract relevant data, and store it in a specific format. Choosing input parameters for training a neural network from assimilated variables required careful consideration **(Figure 43).** Feature engineering was employed to select and transform important variables from raw data, providing domain-specific information to machine learning algorithms. Feature engineering aimed to improve ML algorithms' performance by imparting reservoir engineering knowledge.

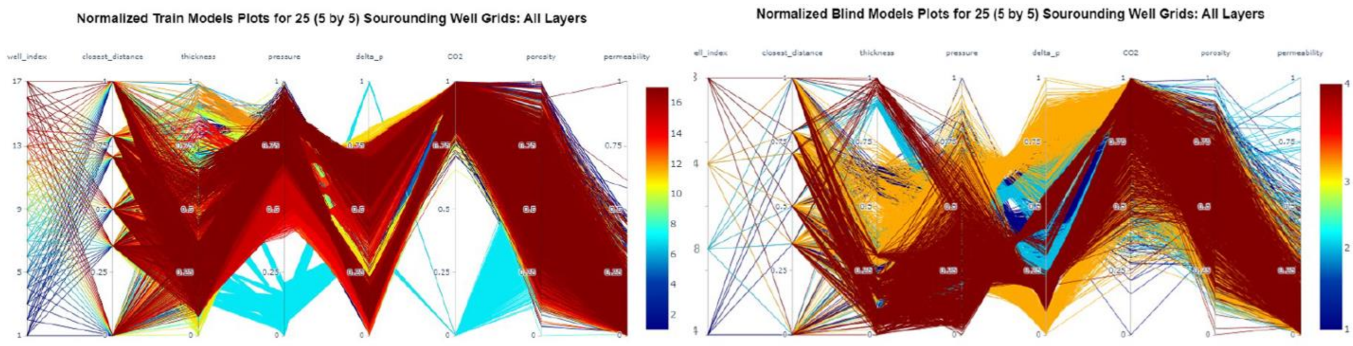


Figure 43. Quality check parallel plots for train and blind data

4.8.2 Spatio-Temporal Dataset Construction:

The spatio-temporal dataset, containing both space and time parameters, was constructed by collecting geological features, generated features, and some simulation inputs and outputs. This dataset taught the Smart Proxy principles of fluid flow in porous media and the complexities of a heterogeneous reservoir. Each numerical reservoir simulation run generated over a million records for each time-step, resulting in a spatio-temporal database with more than 115 million records. The dataset's records represented features or attributes, and the dataset was partitioned into training, calibration, and validation sets.

It's vital to recognize that a smart proxy is a type of data-driven model, built by analyzing collected data. Therefore, the accuracy and reliability of the spatio-temporal database are crucial; if it contains incorrect data, the smart proxy will learn incorrectly and yield unsatisfactory results.

4.8.3 Data Partitioning:

Development data for each Surrogate Reservoir Model (SRM) was randomly divided into training, calibration, and validation sets. The training set, comprising 80% of the data, was used to train the neural network. The calibration set monitored the neural network's performance during training, while the validation set, constituting 10%, assessed the trained model's generalization ability. A portion of reservoir simulation runs was reserved as blind runs for additional evaluation **(Figure 44).**

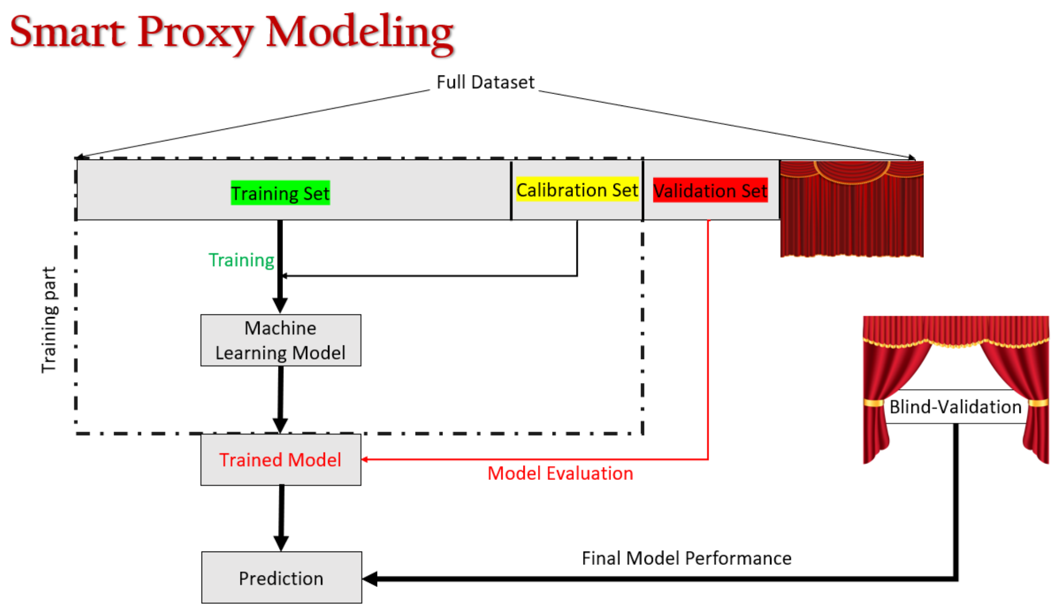


Figure 44. Data partitioning of the Smart Proxy Model

4.8.4 Structure and Topology of Artificial Neural Network:

The structure and topology of the Artificial Neural Network (ANN) were crucial parameters, with fully connected neural networks being successfully employed. The three-layer neural network, consisting of an input layer, a hidden layer, and an output layer, was used in this study. Hyperparameters, such as learning rate, number of neurons in hidden layers, and activation function, were tuned based on the dataset's size and training observations.

4.8.5 Training Process of Artificial Neural Network:

The training process involved passing the entire training dataset through the neural network (epoch) to adjust synaptic weights based on back-propagation of errors. The training process aimed to improve the network's performance until a predetermined threshold was reached. Hyperparameters, including learning rate and activation function, played a significant role in training performance.

4.8.6 Error Measurements:

Error measurements were crucial for validating the smart proxy model, comparing the output with the numerical simulator's output. Different error calculation formulas were used based on the type of output data, such as pressure or CO_2 saturation. Error percentages were calculated to assess the precision of the smart proxy model.

The validation of the proposed intelligent proxy model involves the use of a validation dataset, and it is crucial to assess the model's accuracy relative to a blind set. In this study, precision is gauged by calculating the disparity between the output from the numerical simulator and that of the intelligent proxy model. The evaluation of errors is conducted at each grid block, considering that the smart proxy generates output at each of these blocks. Depending on the nature of the output data, distinct error calculation formulas are applied. Specifically, for pressure output, the Absolute Error Percentage is determined using the formula:

Mean Error Percentage= [(absolute (Artificial Neural Network Output - Numerical Simulator Output)) / Numerical Simulator Output] * 100

The approach differs for CO_2 saturation data due to the unique characteristics of the data, where values range between 0 and 1. Consequently, the error formula for CO_2 saturation is given by:

Absolute Error Percentage = [(absolute (Artificial Neural Network Output - Numerical Simulator Output))] * 100

Furthermore, given the significance of the CO_2 plume extension in carbon sequestration, the plume's shape is visually presented in the dissertation. CO_2 values are categorized into "0" and "1" to illustrate the shape, designating cells with less than 10 percent CO_2 saturation as "0" and those with more than 10 percent as "1." This categorization is applied to both CMG and SPM results. The error in this context is computed using the formula:

Absolute Error (or Prediction Accuracy)=absolute (Artificial Neural Network Output - Numerical Simulator Output)Absolute Error (or Prediction Accuracy)=absolute (Artificial Neural Network Output - Numerical Simulator Output)

4.8.7 Validation of the trained ANN with validation dataset:

After training, the smart proxy model was validated using a separate validation dataset not used during training. The validation dataset helped determine the model's precision in predicting new datasets.

4.8.8 Smart Proxy Deployment:

Once the training and validation processes were completed, the smart proxy models were calibrated and validated internally. The deployment process involved using blind validation datasets, consisting of numerical simulations not seen during training. The smart proxy models were connected to forecast outputs for new time steps, and their performance was assessed using blind validation runs.

This comprehensive process aimed to develop accurate and reliable Smart Proxy Models for reservoir simulation, integrating domain knowledge and machine learning techniques. The Smart Proxy Model, a machine learning model driven by data, exhibits the capability to emulate the outcomes of a sophisticated reservoir simulation model at each step swiftly and with high precision. Furthermore, the developed Smart Proxy model operates on a grid-based framework, specifically tailored to reproduce pressure and CO_2 saturation values for every grid block within the reservoir simulation. Its versatility extends to various applications, encompassing assisted history matching, uncertainty analysis and quantification, as well as production/injection optimization. The development of the Smart Proxy Model involves the training of algorithms within Artificial Neural Networks (ANN) of the Smart Proxy on extensive datasets to comprehend the intricate patterns of fluid behavior in the numerical reservoir simulation. This chapter is intended to provide a concise overview of the steps undertaken in the development of the dynamic Smart Proxy for this project. The workflow adopted for this research is delineated in **Figure 45,** and each step in this chapter is briefly elucidated.

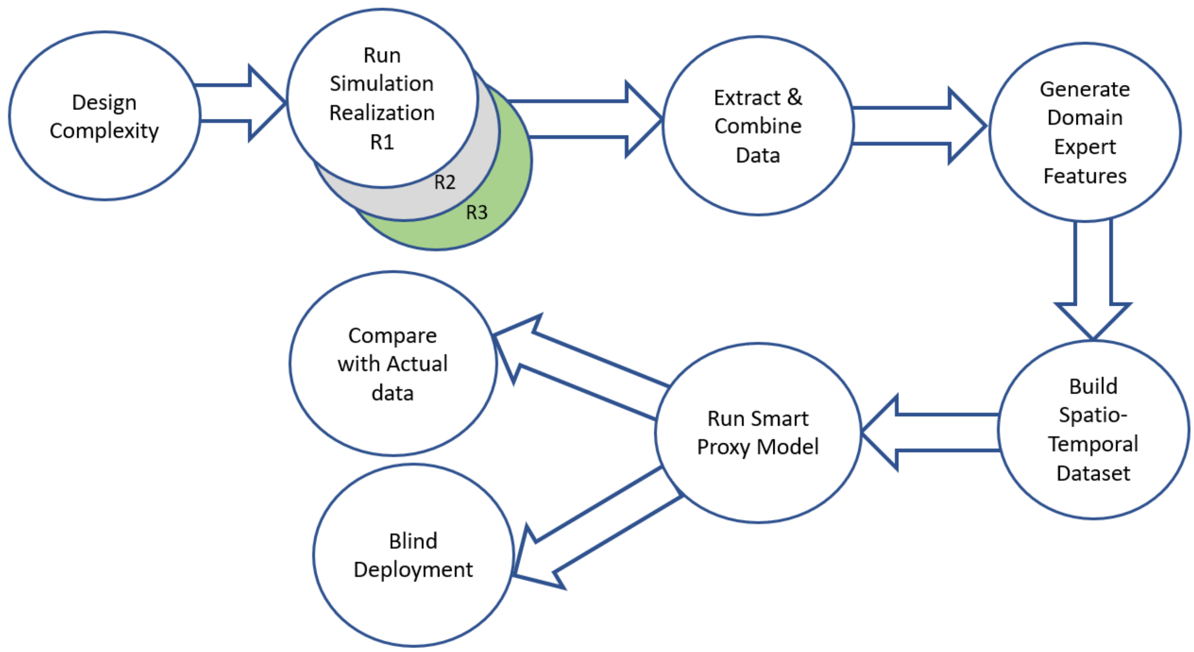


Figure 45. The used workflow for SPM in this study

4.8.9 Data Preparation for Artificial Neural Network (ANN)

After designing the numerical reservoir simulation runs in the preceding chapter, the subsequent step involved running them and extracting crucial data. These extracted data, stemming from twenty simulation runs designed to develop the smart proxy, amounted to an extensive datasets. To construct a robust Artificial Neural Network (ANN) model, domain knowledge played a vital role in preparing the dataset. This preparation is essential to unveil the problem's structure for the machine learning algorithm. Therefore, a deep understanding of the fluid's physics and flow behavior in the reservoir model became paramount. In this project, a meticulous analysis of reservoir parameters at each time step and layer was conducted to comprehend the interactions among parameters. Additionally, important features were generated to enhance Smart Proxy Model accuracy, making the dataset pivotal in Smart Reservoir Modeling (SRM). The reliability of SRM's output hinges on the integrity of its input, emphasizing the importance of accurate and non-faulty data for positive outcomes. Beyond data mining expertise, constructing an SRM necessitates a profound understanding of reservoir engineering.

4.8.10 Data Processing

Numerous raw data from the numerical reservoir simulation runs were processed and translated into a readable format. This involved Python reading the simulation output/input, extracting useful data, and storing it in a specific format. Choosing the input parameters for training a neural network from the assimilated variables in the database is a non-trivial task. The database typically contains numerous parameters that could be potential inputs for neural networks. This includes static and dynamic characteristics and equivalent data for multiple offset wells, all stored as columns in a flat file for input to train data-driven models. The selection of the best input data for ANN models is explored further in the next section. The base ANN was constructed in three layers, an input layer, a hidden layer, and an output layer. The input layer for the pressure model contains some 124 selected parameters or features. These features included static and dynamic properties relevant to the scope of the smart proxy models. The hidden layer has 1,000 nodes and there was one output in the output layer. A list of the selected parameters that were used for one of the runs is presented in **Figure 46.**

Index	Feature	Max)	Min
0	Sample Number	20.0	1.0
1	Grid Cell ID	953125.0	62501.0
2	i	125.0	1.0
3	j	125.0	1.0
4	k	61.0	5.0
5	x	1291155.0	1258096.6
6	y	11285000.0	11251941.6
7	porosity	0.25	0.01
8	permeability	574.47	0.28
9	thickness	23.65	1.7
10	top	10091.4	9286.87
11	bottom	10109.0	9296.46
12	paydepth	10100.2	9291.67
13	Distance to East	33191.7	133.3
14	Distance to West	33191.7	133.3
15	Distance to North	33191.7	133.3
16	Distance to South	33191.7	133.3
17	Distance to Top Seal	244.06	1.58
18	Distance to Bottom Seal	243.75	1.78
19	pr_initial	4789.02	4293.91
20	pr_t	5422.42	4453.62
21	sg_t	0.71	0.0
22	pr_t_dt	5420.2	4452.61
23	sg_t_dt	0.71	0.0
24	dp	22.57	0.0
25	closest Dist 1	23378.86	0.0
26	closest Dist 2	200000.0	4915.86
27	inj_well_based_bhp_1	5500.0	4838.67
28	inj_well_based_bhp_2	5500.0	4838.67
29	inj_well_based_rate_1	59666800.0	0.0
30	inj_well_based_rate_2	59666800.0	0.0
31	inj_cell_based_bhp_1	5630.26	4614.1
32	inj_cell_based_bhp_2	5630.26	4614.1
33	inj_cell_based_rate_1	8889190.0	0.0
34	inj_cell_based_rate_2	8889190.0	0.0
35	porosity_E	0.25	0.0
36	permeability_E	574.47	0.0
37	thickness_E	23.65	0.0
38	top_E	10073.9	0.0
39	pr_initial_E	4778.28	0.0
40	porosity_W	0.25	0.0
41	permeability_W	574.47	0.0

Index	Feature	Max)	Min
42	thickness_W	23.65	0.0
43	top_W	10091.4	0.0
44	pr_initial_W	4789.02	0.0
45	porosity_N	0.25	0.0
46	permeability_N	574.47	0.0
47	thickness_N	23.65	0.0
48	top_N	10091.4	0.0
49	pr_initial_N	4789.02	0.0
50	porosity_S	0.25	0.0
51	permeability_S	574.47	0.0
52	thickness_S	23.65	0.0
53	top_S	10091.4	0.0
54	pr_initial_S	4789.02	0.0
55	porosity_T	0.25	0.0
56	permeability_T	574.47	0.0
57	thickness_T	23.65	0.0
58	top_T	10091.4	0.0
59	pr_initial_T	4789.02	0.0
60	porosity_B	0.25	0.0
61	permeability_B	574.47	0.0
62	thickness_B	23.65	0.0
63	top_B	10091.4	0.0
64	pr_initial_B	4789.02	0.0
65	porosity_NT	0.25	0.0
66	permeability_NT	574.47	0.0
67	thickness_NT	23.65	0.0
68	top_NT	10073.9	0.0
69	pr_initial_NT	4778.28	0.0
70	porosity_NB	0.25	0.0
71	permeability_NB	574.47	0.0
72	thickness_NB	23.65	0.0
72	thickness_NB	23.65	0.0
74	pr_initial_NB	4789.02	0.0
74	pr_initial_NB	4789.02	0.0
75	porosity_NW	0.25	0.0
76	permeability_NW	574.47	0.0
77	thickness_NW	23.65	0.0
78	top_NW	10091.4	0.0
79	pr_initial_NW	4789.02	0.0
80	porosity_NE	0.25	0.0
81	permeability_NE	574.47	0.0
82	thickness_NE	23.65	0.0

Index	Feature	Max)	Min
83	top_NE	10091.4	0.0
84	pr_initial_NE	4789.02	0.0
85	porosity_ST	0.25	0.0
86	permeability_ST	574.47	0.0
87	thickness_ST	23.65	0.0
88	top_ST	10073.9	0.0
89	pr_initial_ST	4778.28	0.0
90	porosity_SB	0.25	0.0
91	permeability_SB	574.47	0.0
92	thickness_SB	23.65	0.0
93	top_SB	10091.4	0.0
94	pr_initial_SB	4789.02	0.0
95	porosity_SW	0.25	0.0
96	permeability_SW	574.47	0.0
97	thickness_SW	23.65	0.0
98	top_SW	10091.4	0.0
99	pr_initial_SW	4789.02	0.0
100	porosity_SE	0.25	0.0
101	permeability_SE	574.47	0.0
102	thickness_SE	23.65	0.0
103	top_SE	10091.4	0.0
104	pr_initial_SE	4789.02	0.0
105	porosity_BW	0.25	0.0
106	permeability_BW	574.47	0.0
107	thickness_BW	23.65	0.0
108	top_BW	10091.4	0.0
109	pr_initial_BW	4789.02	0.0
110	porosity_BE	0.25	0.0
111	permeability_BE	574.47	0.0
112	thickness_BE	23.65	0.0
113	top_BE	10091.4	0.0
114	pr_initial_BE	4789.02	0.0
115	porosity_TW	0.25	0.0
116	permeability_TW	574.47	0.0
117	thickness_TW	23.65	0.0
118	top_TW	10073.9	0.0
119	pr_initial_TW	4778.28	0.0
120	porosity_TE	0.25	0.0
121	permeability_TE	574.47	0.0
122	thickness_TE	23.65	0.0
123	top_TE	10073.9	0.0
124	pr_initial_TE	4778.28	0.0

Figure 46. A list of base selected features as input to the neural network a particular SPM mode

4.9 Model Content vs. Complexity:

Upscaling, or homogenization, is substituting a heterogeneous property region consisting of fine grid cells with an equivalent homogeneous region made up of a single coarse-grid cell with an effective property value. Upscaling is performed for each of the cells in the coarse grid and for each of the grid properties needed in the reservoir flow-simulation model. In reservoir simulation, the question is not whether, but how and how much. The complexity of the questions being asked, and the amount and reliability of the data available, must determine the sophistication of the system to be used. High-fidelity simulations in science and engineering are computationally expensive and time-prohibitive for quick iterative use cases, from design analysis to optimization **(Figure 47 and Figure 48).**

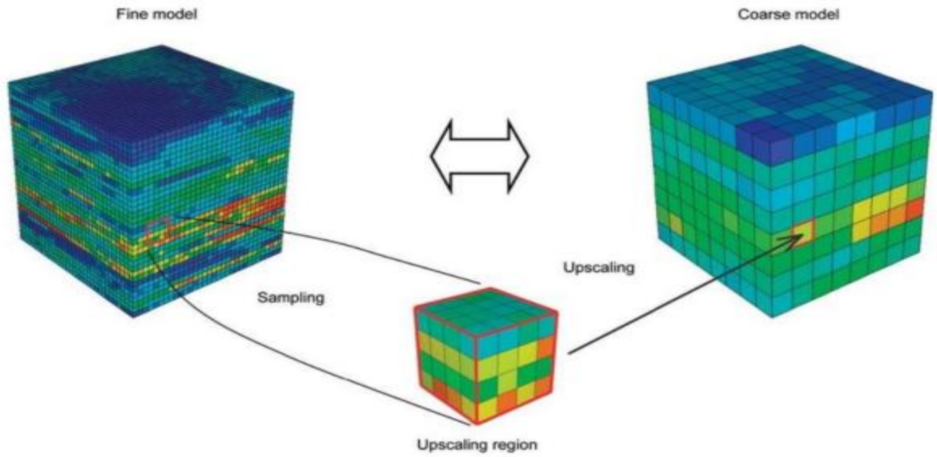


Figure 47. Model content versus complexity

Figure 48. Analogy of model content versus complexity in real life

4.9.1 Feature Engineering

In machine learning modeling, feature engineering is the process of leveraging domain expertise to select and transform the most crucial variables from raw data. The aim is to enhance the performance of machine learning algorithms by incorporating reservoir engineering knowledge from numerical simulation model data. Feature engineering in this project provides features or information to machine learning, aiding in the understanding of fluid flow mechanics in porous media. Neural networks are not trained on all parameters included in the database, and limiting the number of parameters used to design a neural network is advisable. This approach involves extracting key features related to reservoir characteristics and fluid flow. Static attributes are feature-engineered, including Tier Features and Distance-related Features, discussed in subsequent sections.

It's important to recognize the differences between how machines "learn" and conventional statistical methods in the context of Artificial Intelligence (AI) and Machine Learning (ML). It's a common misconception to mix up AI & ML with traditional statistics when one thinks that machine learning is just about feeding data to ML algorithms.

While traditional statistics has been practiced for more than 120 years, the field of artificial intelligence and machine learning as it exists today started about 35 years ago, in the mid-1980s. There are key distinctions between the two fields. The main goal of traditional statistics is to use algorithms to apply gathered data. But more is needed for the engineering application of AI & ML, which is a major topic of this book. Effective Machine Learning requires expert "teaching". This means that, in contrast to learning that depends only on data input, the learning process of AI & ML in engineering problem-solving is greatly enhanced by guided instruction and expertise. This distinction emphasizes how crucial it is to comprehend the breadth and complexity of AI & ML, particularly when using them for purposes other than simple data analysis and producing outcomes that go beyond surface-level analysis (Mohaghegh 2022).

4.9.1.1 Static Reservoir Features

Static features encompass data that remains constant over time, such as reservoir characteristics for focal cells and neighboring cells, including porosity, permeability, thickness, grid top, and cell locations.

4.9.1.2 Dynamic Reservoir Features

Dynamic features change over time and originate from two domains: well domain data (injection rates and well bottomhole pressure) and grid-block domain data (reservoir pressure and CO_2 saturation).

4.9.1.3 Tier System Features

A Tier System is introduced to capture the influence of surrounding cells on the pressure and saturation of each focal cell. This involves tiers 1 and 2, which contribute to building the dataset for the ANN Model. Tier 1 includes 6 faced contact cells, while tier 2 comprises 12-line contact cells, incorporating features of 18 surrounding cells as input data to the ANN Model. **Figure 49** illustrates tier 1 and tier 2, emphasizing their significance in the approach.

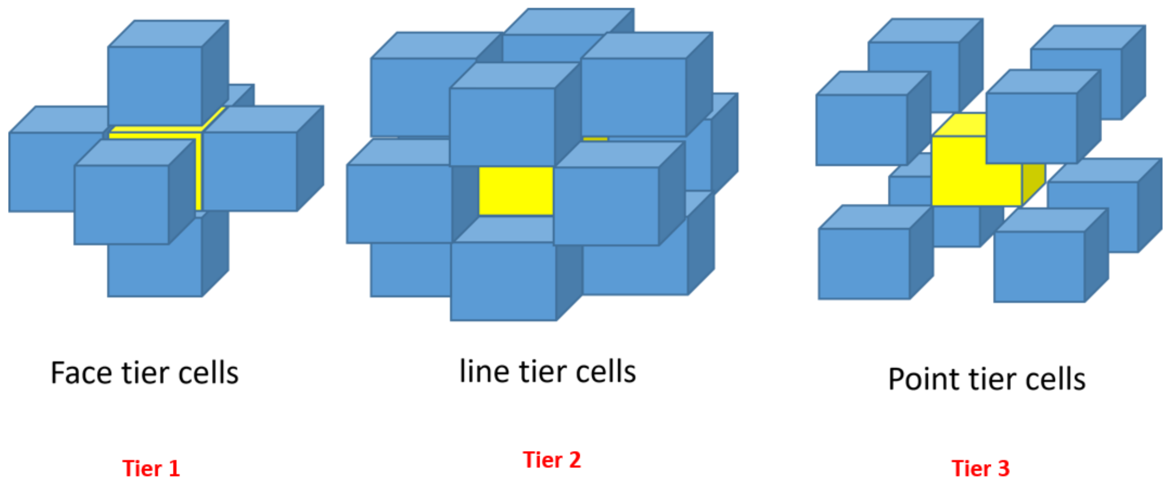


Figure 49. Focal grid block with its neighboring blocks

4.9.1.4 Coordinate and Distance-related System Features

To furnish the Artificial Neural Network (ANN) model with comprehensive insights into the behavior of each cell within the system, it is imperative to uniquely determine and employ the location of each cell as input to the ANN Model. The location of each cell in the reservoir simulation model can be expressed by computing the distance from each cell to various reservoir model boundaries, encompassing distances to the top, bottom, east, west, north, and south boundaries. Additionally, the three indices (i, j, and k) assigned to each cell serve to identify its location. Furthermore, a distinctive number (cell ID) assigned to each cell in the reservoir simulation model is incorporated as input to the ANN, facilitating the identification of the unique location for each cell.

In tandem with establishing distances between each cell's position and the reservoir model boundaries, it becomes essential to acquaint the Neural Network (NN) with the relative location of each cell concerning the injection wells, based on their proximity order (distance to the 1st closest injector, distance to the 2nd closest injector, and so forth). This pertinent information is introduced as an additional feature to enlighten the NN about the relative significance of each cell's location concerning the CO_2 injection source. Similarly, the porosity index, introduced through formulations like FCL, FCR, FCN, and FCS, can provide valuable insights into the spatial variations of porosity. By calculating the porosity index based on neighboring cells, it captures the relative differences in porosity, highlighting the heterogeneity within the reservoir. This information can enhance the understanding of how porosity varies in different directions around a focal cell for cells in its tier 1 category. For instance, FCL reflects the porosity contrast between the focal cell and its left-side neighbor. **(50 through Figure 52).**

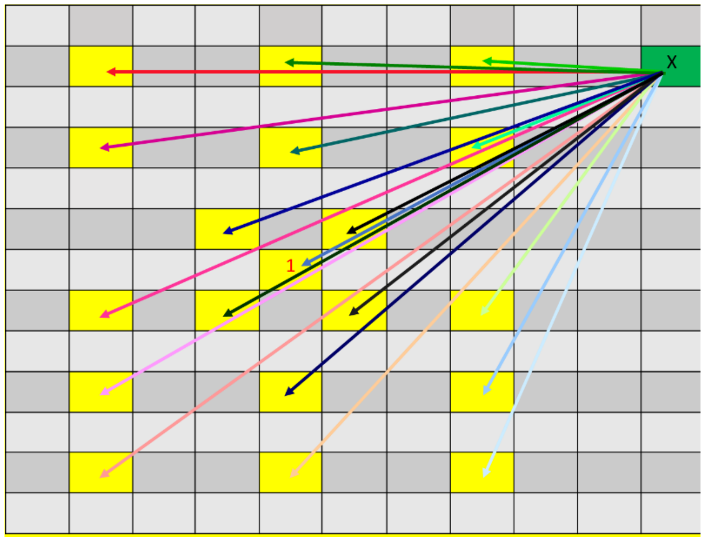


Figure 50. Distance to closest injection well

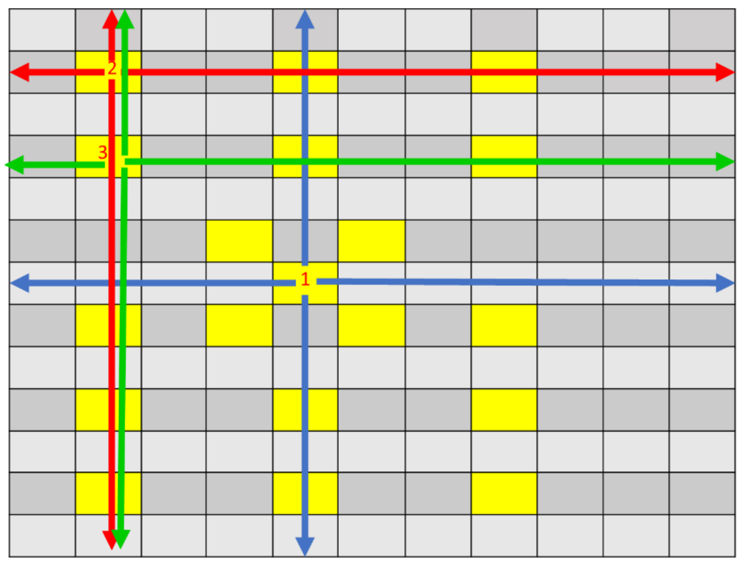


Figure 51. Injection well distances from each boundary

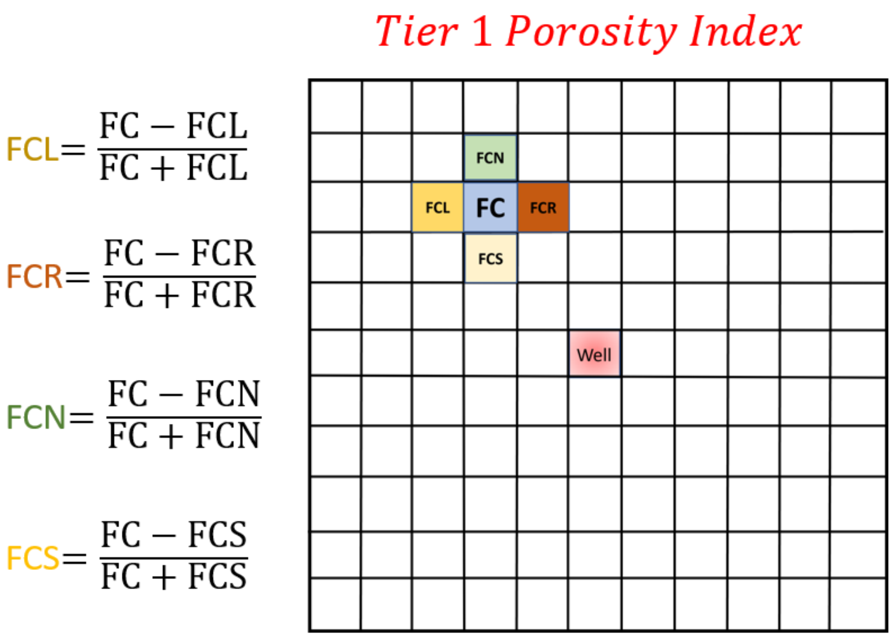


Figure 52. Tier 1 Porosity Index

4.10.1.5 Features related to the quality of path between focal cells and each injection well:

Quality of Flow Path to Different Injectors: Quality of the path is a valuable feature in the context of feature engineering for an Artificial Neural Network (ANN) in reservoir modeling. This feature considers both the distance and porosities of every grid block lying between the focal grid and the injection wells **(Figure 53).** It can enhance the understanding of fluid flow in the reservoir in several aspects:

1. Distance Consideration:
   - The distance between a focal grid block and injection wells plays a crucial role in determining the flow path. Longer distances generally mean higher resistance to fluid flow.
   - Feature engineering includes incorporating the distance to each injector as an input feature. This information helps the ANN understand the spatial relationship between the focal grid block and each injection well.
   - The "Quality of Flow Path" considers not only the absolute distances but also the relative distances to different injectors. This is essential as the impact of each injector on the flow path may vary based on its proximity.
2. Porosity Influence:
   - Porosity is a fundamental reservoir property that affects fluid flow. Higher porosities generally imply better fluid mobility, facilitating a more efficient flow path.
   - Including porosity information in the feature engineering process allows the ANN to learn how the quality of the flow path is influenced by variations in porosity along the path.

- The feature can be designed to capture the porosity values of each grid block between the focal grid and injection wells, providing insights into the permeability characteristics of the flow path.

3. Comprehensive Flow Path Evaluation:
   - By combining distance and porosity information, the "Quality of Flow Path" feature offers a holistic evaluation of the flow path's effectiveness from the focal grid block to different injectors.
   - This feature enables the ANN to discern not only the spatial aspects of the flow path but also how variations in porosity contribute to the overall quality of the path.
4. Importance for Fluid Dynamics:
   - In reservoir engineering, understanding the quality of flow paths is critical for optimizing injection strategies, managing reservoir sweep efficiency, and predicting fluid distribution.
   - The feature becomes especially relevant for enhanced oil recovery or CO_2 sequestration projects, where efficient fluid flow paths contribute to the success of injection and recovery processes.

The generated features are presented in **Figure 54 through Figure 61.**

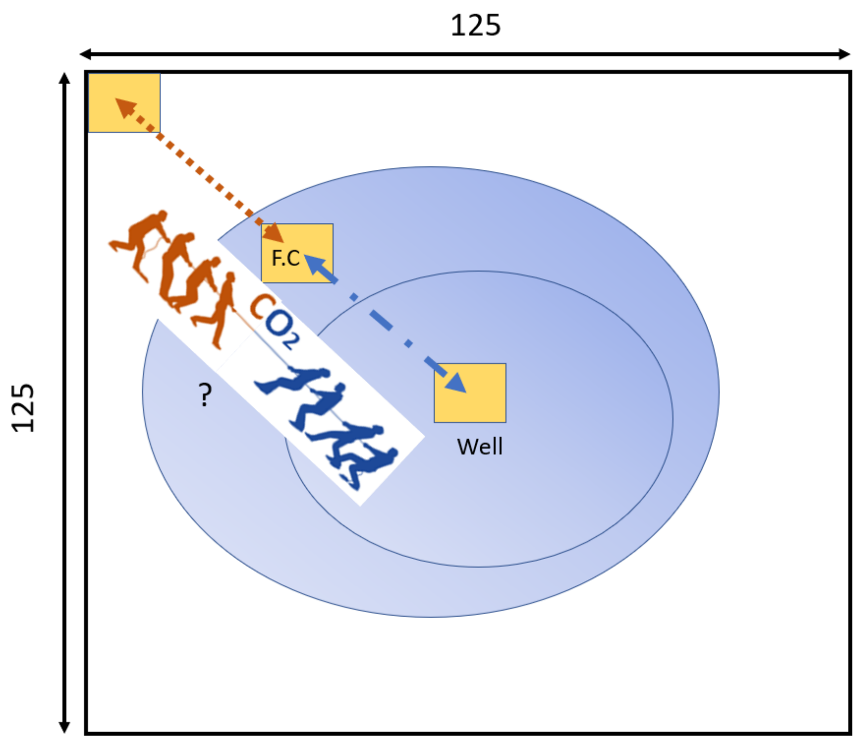


Figure 53. The quality of path between focal cell and CO_2 injection well

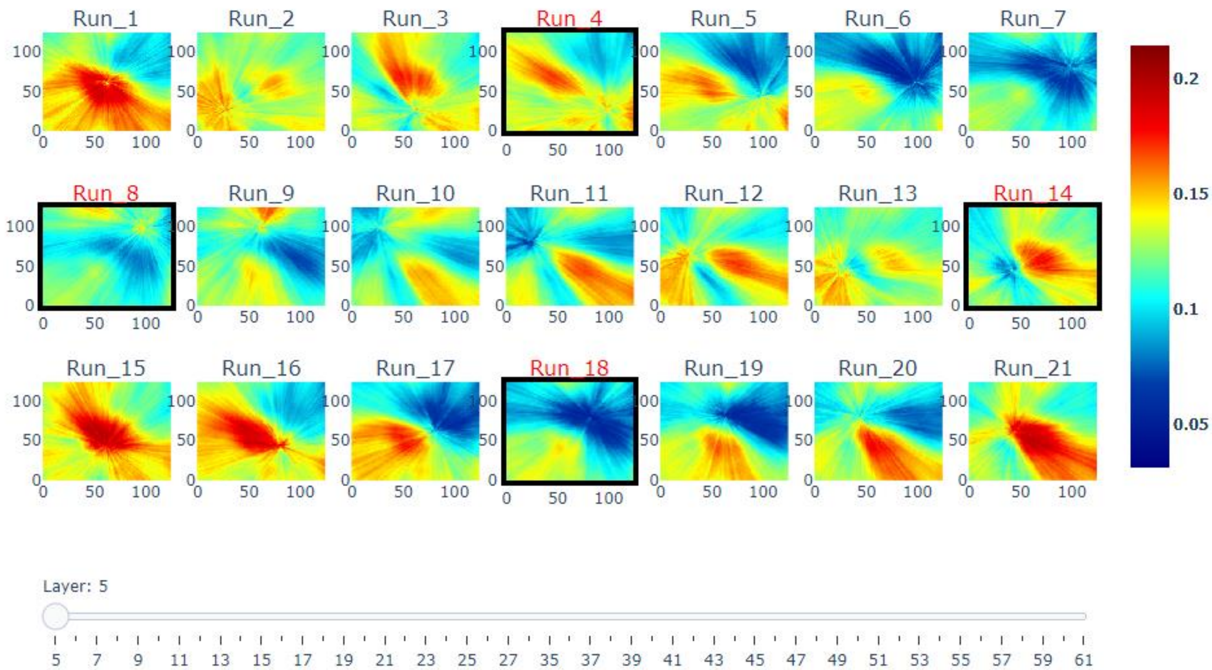


Figure 54. Quality of Flow Path to Different Injectors from every focal grid for Layer 5

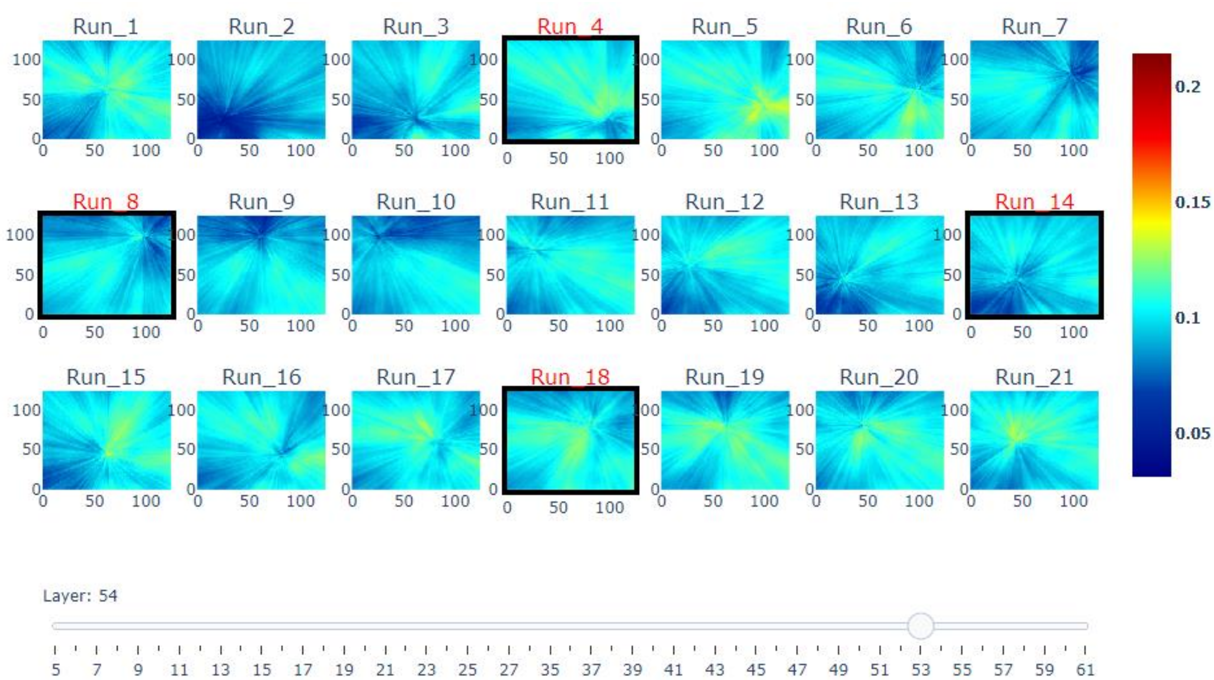


Figure 55. Quality of Flow Path to Different Injectors from every focal grid for Layer 54

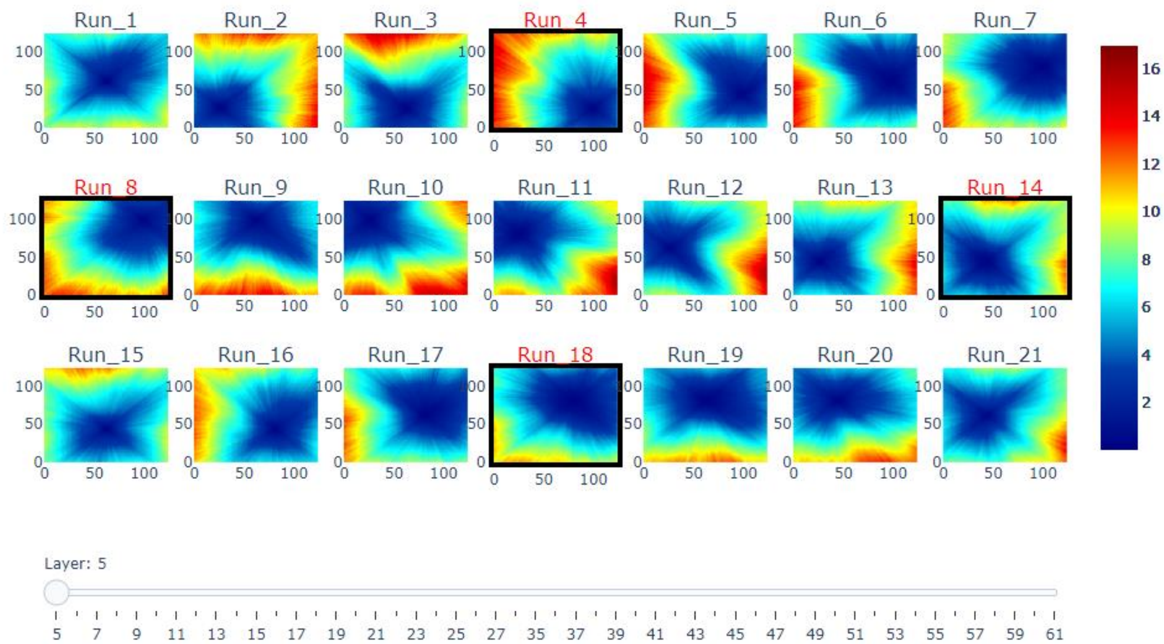


Figure 56. Another method for Quality of Flow Path to Different Injectors from every focal grid for Layer 5

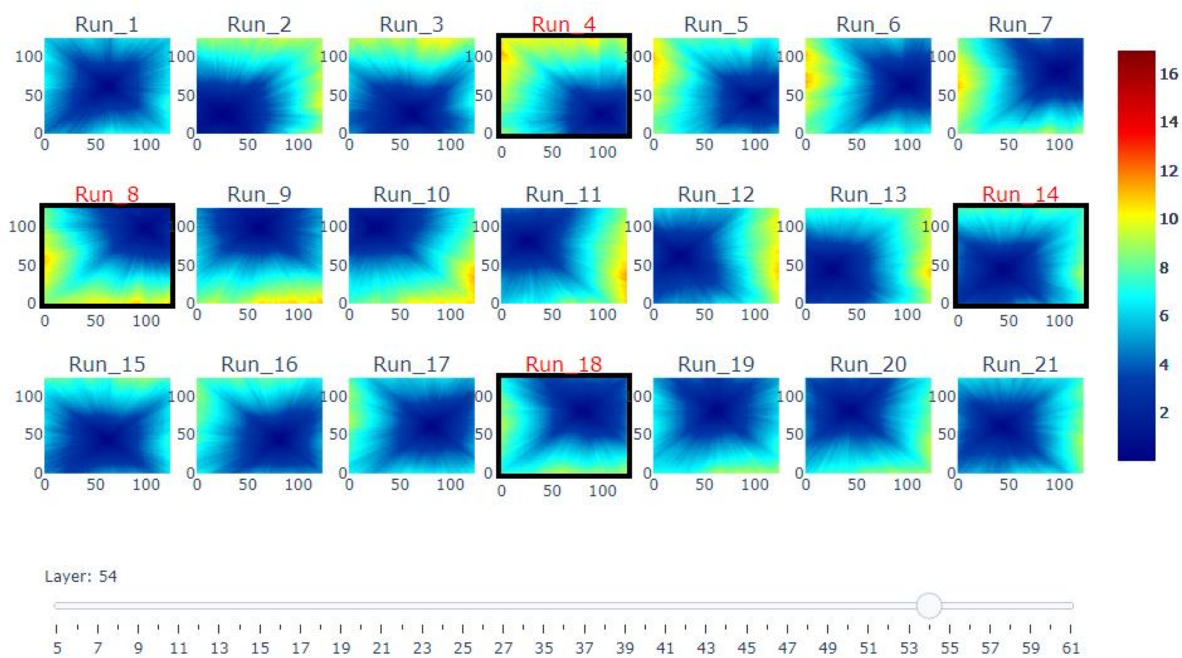


Figure 57. Another method for Quality of Flow Path to Different Injectors from every focal grid for Layer 54

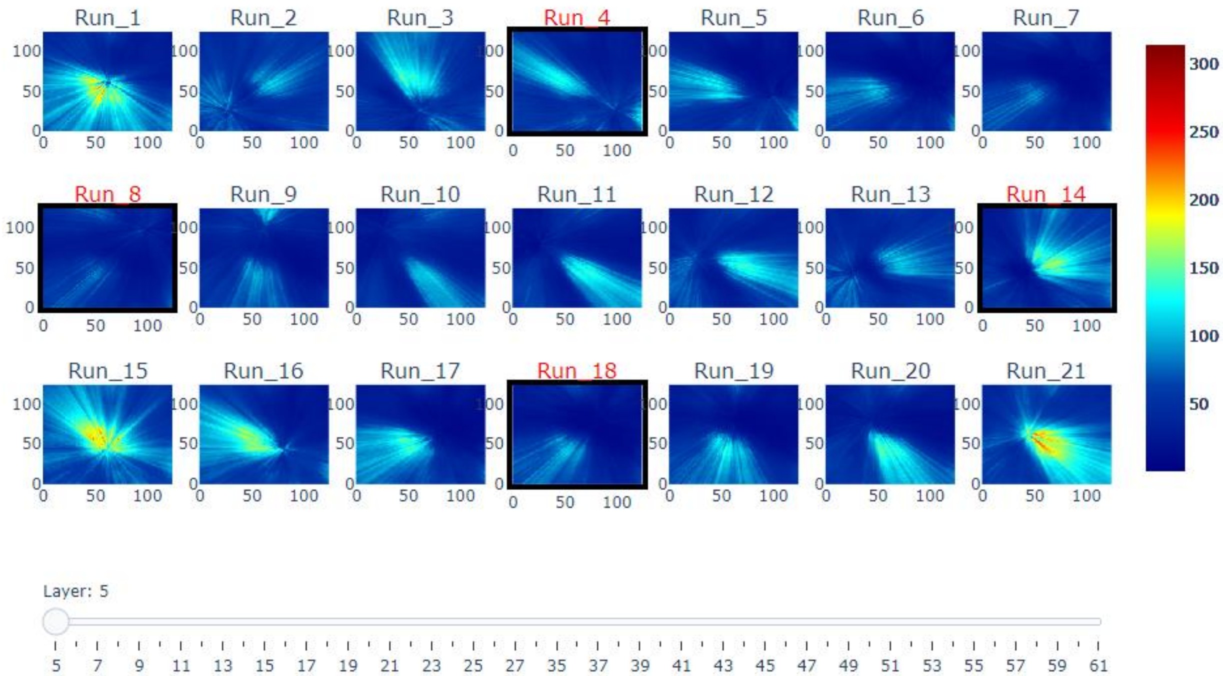


Figure 58. Another method for Quality of Flow Path to Different Injectors from every focal grid for Layer 5

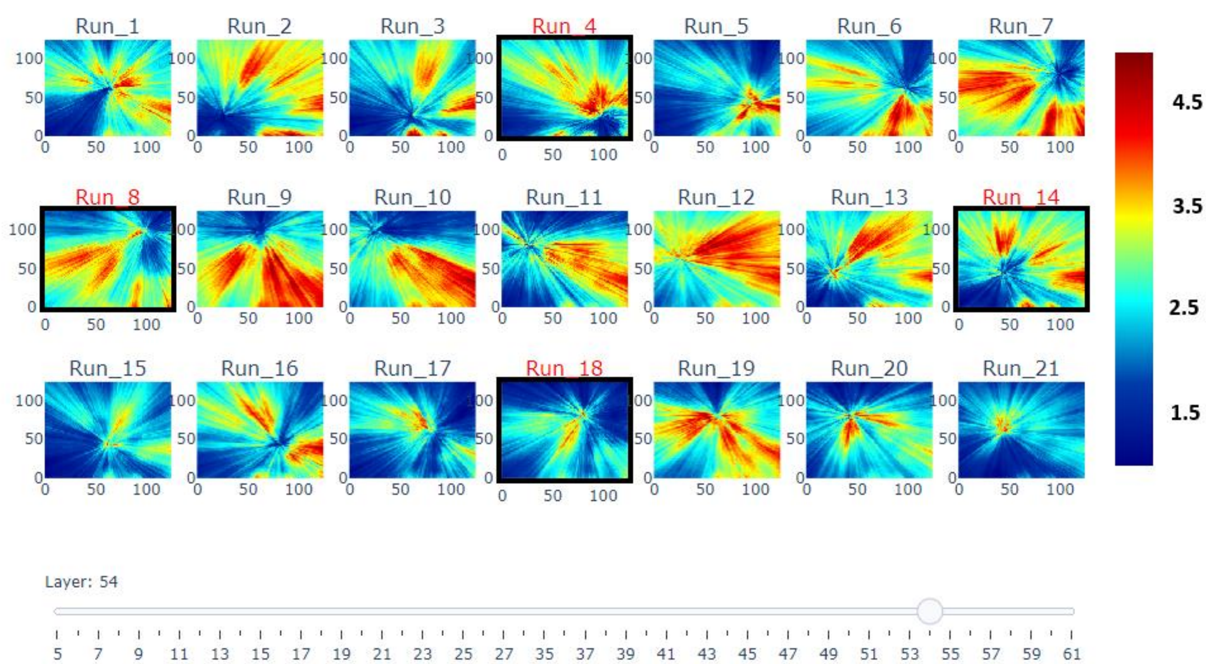


Figure 59. Another method for Quality of Flow Path to Different Injectors from every focal grid for Layer 54

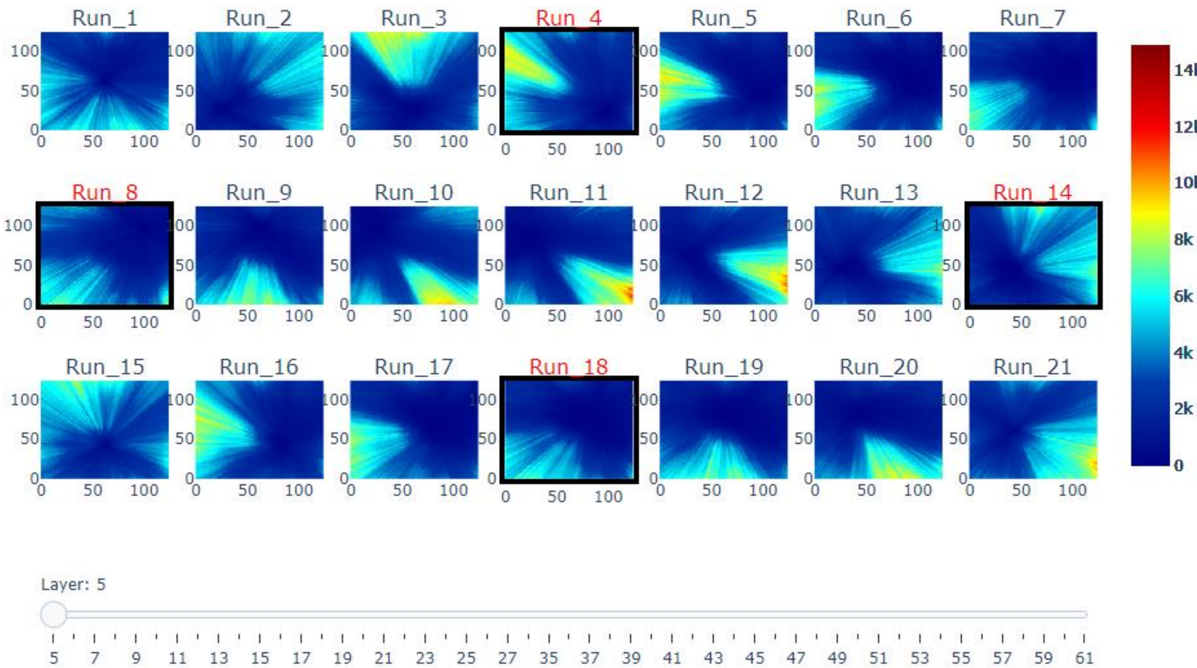


Figure 60. Another method for Quality of Flow Path to Different Injectors from every focal grid for Layer 5

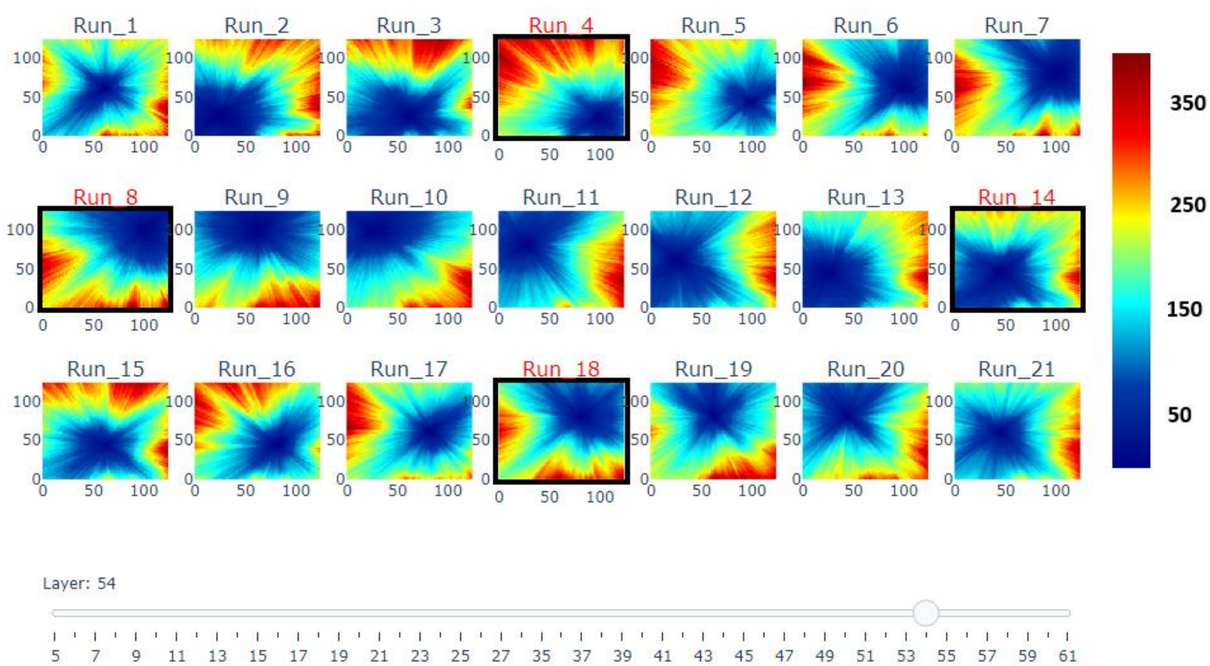


Figure 61. Another method for Quality of Flow Path to Different Injectors from every focal grid for Layer 54

Quality of Flow Path to Different Injectors feature to include tiered grid cells surrounding the focal grid is a comprehensive approach that can significantly enhance the understanding of fluid flow patterns in a reservoir **(Figure 62).** Incorporating this expanded feature for different tiers can be beneficial in several aspects:

1. **Focal Grid (Tier 0):**
   - Provides insights into the immediate flow path from the focal grid to injection wells, considering both distance and porosities of the grid blocks in between.
   - Essential for understanding the direct impact of injection wells on the focal grid and optimizing fluid flow within its immediate vicinity.
2. **Tier 1 (Face-Adjacent Grids):**
   - Expands the analysis to grid cells sharing a face with the focal grid (south, north, left, right, east, west).
   - Offers a broader perspective on how neighboring grid blocks influence the flow path, considering their immediate connectivity with the focal grid.
3. **Tier 2 (Line-Adjacent Grids):**
   - Encompasses grid cells sharing a line with the focal grid (north-east, south-west, north-west, south-east) in the same layer, layer above, and layer below.
   - Captures the influence of grid cells that may not share a face but are still in close proximity, contributing to the overall flow dynamics.
4. **Tier 3 (Point-Adjacent Grids):**
   - Includes grid cells sharing only a corner point with the focal grid in the same layer, layer above, and layer below.
   - Considers the impact of grid cells that are farther away but still have a spatial relationship with the focal grid, contributing to a more complete understanding of fluid flow patterns.

Key Advantages:

- **Comprehensive Spatial Insight:** This extended feature provides a multi-tiered spatial perspective, allowing the ANN to learn how flow paths are influenced not only by direct neighbors but also by cells in varying degrees of proximity.
- **Vertical Consideration:** Including layers above and below ensures a three-dimensional understanding of the reservoir, vital for scenarios where fluid movement may occur vertically.
- **Optimized Feature Set:** By incorporating distance and porosity information for multiple tiers, the feature set becomes more robust, capturing a wider range of spatial interactions.

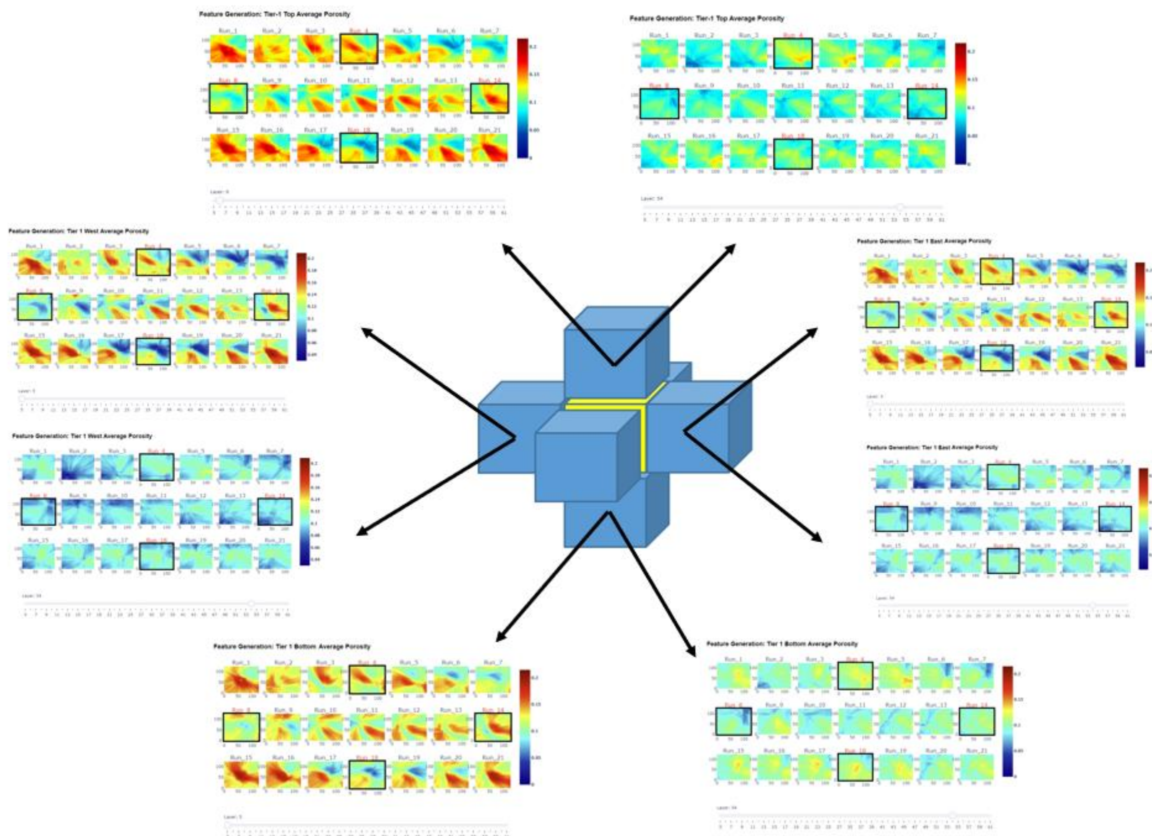


Figure 62. Quality of Flow Path to Different Injectors from every focal grid for Tier 1 and 2 grids

4.9.1.6 Tortuosity and Transmissibility:

The grain packing provides situations, such that the connected pores do not arrange linearly, but rather randomly. This random arrangement of the connected pore spaces guarantees that, fluids do flow in a non-linear direction in the porous media. Tortuosity can be changed by variations in porosity, grain packing, grain size, grain shape and the sorting **(Figure 63 and Figure 64).** The transmissibility between two blocks is the measure of how easily fluids flow between them. The mathematical expression for two phase flow between grid block i and (i+1), for water **(Figure 65):**

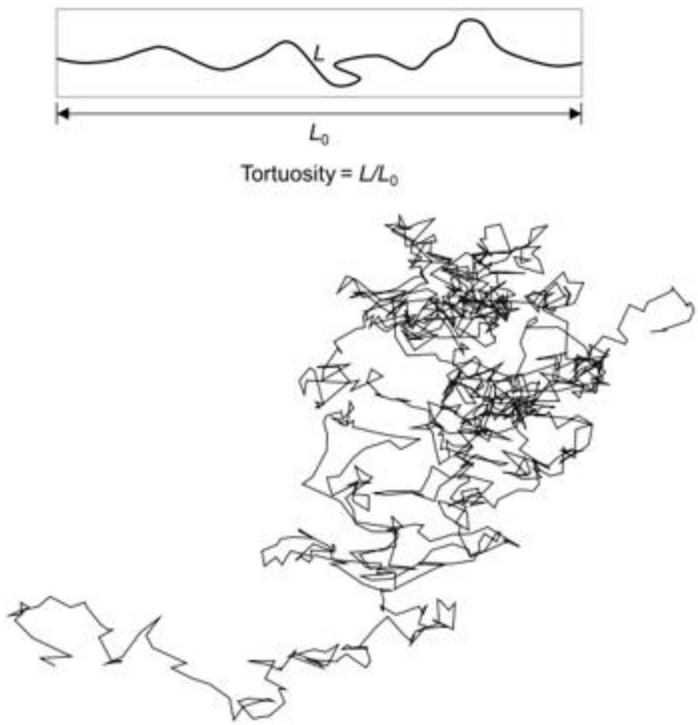


Figure 63. Schematic representation of tortuosity between two points

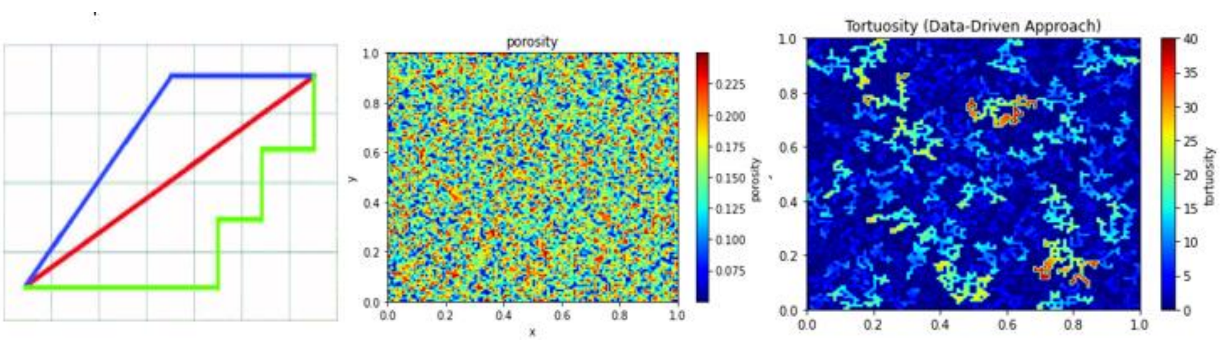


Figure 64. How tortuosity between every grid cells can be communicated from a random porosity data and visualized in terms of tortuosity

$$Q_w = (kA)_{i+1/2}\left(\frac{k_{rw}}{\mu_w\beta_w}\right)_{i+1/2}\frac{(P_{i+1}-P_i)}{\Delta x}$$

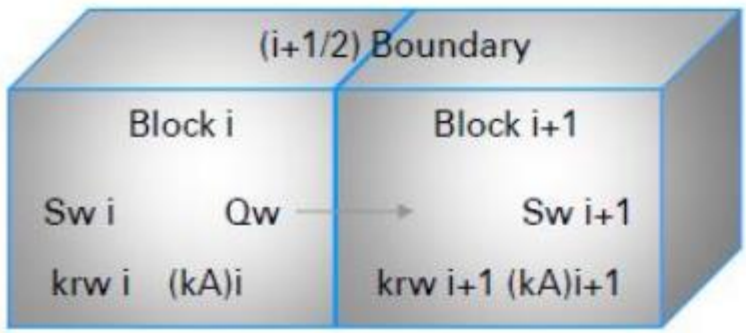


Figure 65. communication between two adjacent blocks

4.9.1.7 Connected Neighbors Count:

This feature takes into account the interconnected nature of porosity between different cells and the focal grid cell, providing insights into the spatial relationships and connectivity within the reservoir **(Figure 66).** By including this connectivity feature, you can capture the interconnectedness of grids in the reservoir. Grids with a higher count of connected neighbors may have a higher degree of connectivity and can play a significant role in fluid flow and pressure transmission. This information can provide valuable insights for predicting the pressure of the reservoir after CO_2 injection. Here's an explanation of how this feature can enhance the understanding of porosity distribution:

1. Interconnected Porosity Distribution:
    - Porosity in a reservoir is not isolated to individual grid cells; it often exhibits interconnected patterns where the porosity of one cell influences that of its neighboring cells.
    - The "Connected Neighbors Count" feature quantifies the number of neighboring grid cells that share interconnected porosity relationships with the focal grid cell. This count reflects the extent to which porosity values are spatially correlated.
2. Spatial Connectivity Analysis:
    - Porosity is a property that tends to exhibit spatial continuity, meaning neighboring cells often have similar porosity values.
    - The feature provides a measure of how well-connected the focal grid cell is to its immediate neighbors in terms of porosity. A higher count indicates a greater degree of spatial connectivity in porosity distribution.
3. Influence on Fluid Flow:
    - Interconnected porosity patterns are crucial for understanding fluid flow behavior in the reservoir. Fluid tends to follow paths of least resistance, and porosity connectivity influences the preferential flow paths.

- The "Connected Neighbors Count" feature contributes to the ANN's ability to learn how porosity connectivity impacts fluid movement within the reservoir.

4. Quantification of Porosity Variability:
   - Different reservoir zones may exhibit varying degrees of porosity connectivity. For instance, certain areas may have highly interconnected porosity, while others may show more isolated patterns.
   - The feature allows the ANN to quantify and differentiate between regions with distinct porosity connectivity characteristics, aiding in the creation of a more nuanced reservoir model.
5. Integration with Other Features:
   - When combined with other features related to porosity, such as average porosity, standard deviation, or spatial gradients, the "Connected Neighbors Count" enriches the feature set by providing a complementary perspective on porosity distribution.

The goal is to calculate the connectivity and tortuosity values for each grid in the reservoir using the following steps:

1. Determine the connectivity:
   - For each grid, check if its permeability and porosity values exceed a threshold.
   - If the values exceed the threshold, consider the grid connected to its neighboring grids.
   - Use the connected components algorithm to determine the connectivity of the pore space.
2. Calculate tortuosity:
   - For each connected grid, use Dijkstra's algorithm (Dijkstra 1959) or an alternative method to find the shortest path lengths to all other connected grids.
   - Determine the longest shortest path length from the current grid.
   - Calculate the tortuosity as the ratio of the longest shortest path length to the shortest path length from the current grid to itself.

The calculate_tortuosity() function developed in this study would take the porosity and permeability arrays, as well as the source and destination coordinates and points. It creates a graph based on the porosity and permeability values, finds the shortest path between the source and destination nodes using Dijkstra's algorithm, and calculates the tortuosity as the ratio of the actual path length to the shortest possible length. One way to do this is to use a Dijkstra algorithm. The Dijkstra algorithm is a graph search algorithm that finds the shortest path between two points in a graph. In this case, the graph would be the porosity and permeability arrays. The Dijkstra algorithm would start at point A and then explore all of the neighboring points that have a porosity and permeability value greater than a certain threshold. The algorithm would continue exploring neighboring points until it reached point B. The points that the algorithm explored would be the points that the flow passed from point A to point B. In summary, the assigned value of tortuosity to a single grid represents how tortuous the path is from that grid to any other connected grid. The distance between grids is measured in terms of the shortest path length, which is the minimum number of grid-to-grid steps required to reach the destination.

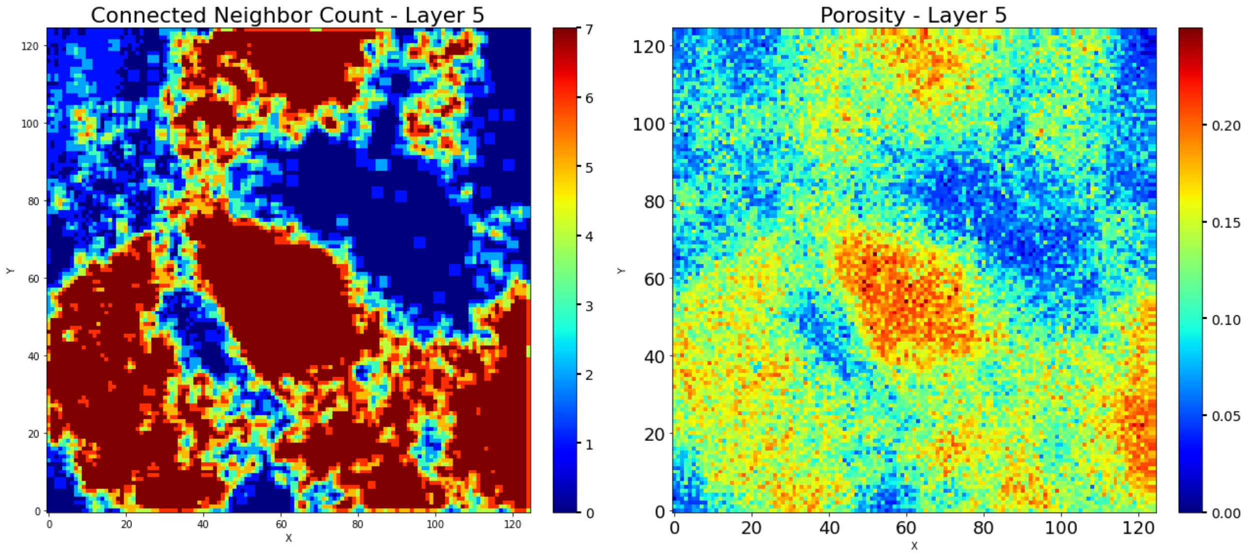


Figure 66. How Connected Neighbors Count between every grid cells can be communicated and visualized for Smart Proxy Modeling

4.9.1.8 Relative Permeability to CO_2:

This feature considers the ease of flow between different cells and the focal grid cell, providing insights into the permeability characteristics of the reservoir **(Figure 67).** This feature is designed to contribute to the understanding of fluid flow dynamics:

1. Quantifying CO_2 Flow Characteristics:
    - Relative permeability is a crucial parameter in reservoir engineering that quantifies the ease with which fluids, in this case, CO_2, can flow through porous media. It is a measure of the fraction of the total permeability available for a particular fluid phase.
    - The "Relative Permeability to CO_2" feature provides a quantitative assessment of how easily CO_2 can flow through the porous structure of each grid cell, considering the unique characteristics of the reservoir.

2. Influence on Fluid Distribution:
    - CO_2 injection and distribution play a key role in enhanced oil recovery and carbon capture and storage projects. The feature captures the local variations in relative permeability, offering insights into how CO_2 is likely to distribute within the reservoir.
    - Variations in relative permeability can impact the preferential pathways and areas where CO_2 is more likely to migrate.
3. Spatial Connectivity of CO_2 Pathways:
    - Fluid flow in a reservoir is influenced by the spatial connectivity of permeable pathways. The feature provides information on how well-connected the focal grid cell is to its neighbors in terms of facilitating CO_2 flow.
    - A high "Relative Permeability to CO_2" in a specific grid cell indicates a more efficient pathway for CO_2 movement, contributing to a better understanding of fluid flow patterns.

4. Integration with Flow Dynamics:
    - When integrated into the ANN alongside other features related to fluid flow, such as pressure gradients, porosity, and reservoir geometry, the "Relative Permeability to CO_2" feature enriches the model's understanding of the complex interactions governing fluid movement.

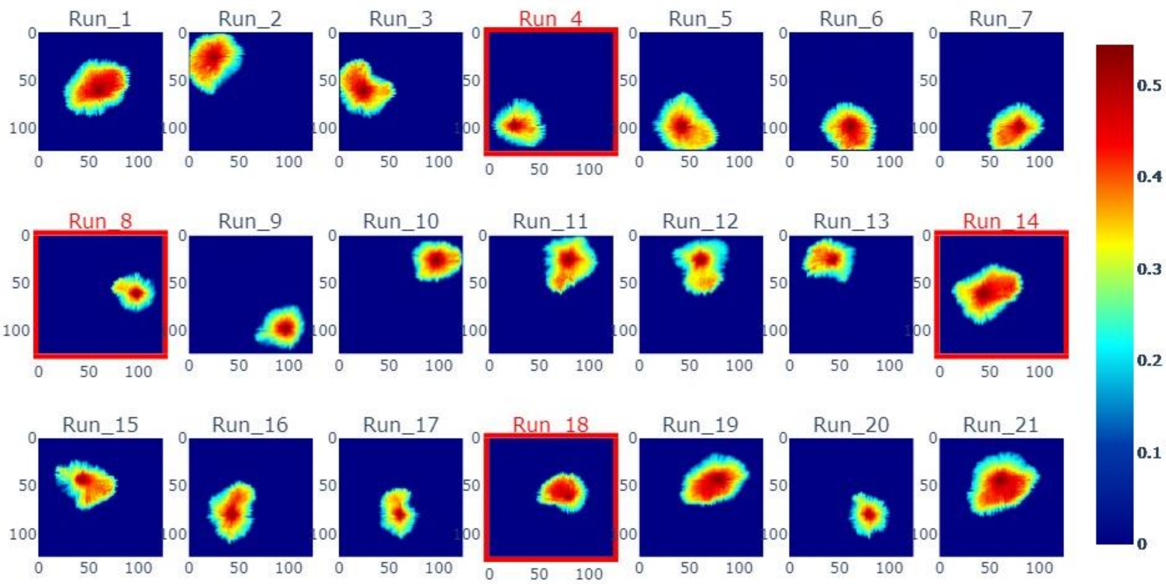


Figure 67. How relative permeability to CO_2 between every grid cells can be communicated and visualized for Smart Proxy Modeling

4.9.1.9 Angle of Permeability Reduction:

This feature takes into account the fluid flow communication between different cells and the focal grid cell, providing insights into how the permeability reduction angle influences fluid flow dynamics **(Figure 68).** This feature could contribute to a better understanding of fluid flow in the reservoir in the following ways:

1. Permeability Reduction Impact:
    - It quantifies the directional reduction in permeability from the focal grid cell to its neighboring cells. Permeability reduction is often encountered in reservoirs due to factors such as compaction, clay swelling, or fluid-rock interactions.
    - This feature provides information about the extent and directionality of the reduction in fluid flow pathways, which is crucial for understanding the impact on overall reservoir performance.
2. Directional Flow Constraints:
    - Fluid flow in a reservoir is influenced by the permeability distribution and the pathways available for flow. The feature captures the directional constraints imposed by permeability reduction, indicating how fluid flow communication is restricted or redirected between neighboring cells.
    - Understanding directional constraints is vital for predicting how injected fluids or natural reservoir fluids will migrate through the reservoir.

3. Influence on Connectivity:
   - The feature contributes to assessing the connectivity of the focal grid cell with its neighbors in terms of altered permeability pathways. It offers insights into how fluid flow communication may be preferentially directed along certain angles or directions.
   - Connectivity variations due to permeability reduction angles impact the efficiency and predictability of fluid flow within the reservoir.
4. Integration with Reservoir Structure:
   - When integrated into the ANN alongside other features related to reservoir structure, such as porosity, permeability magnitude, and geological formations, the feature enriches the model's understanding of the spatial relationships governing fluid flow.
   - It helps the model learn how changes in permeability directionality influence the overall fluid flow behavior.

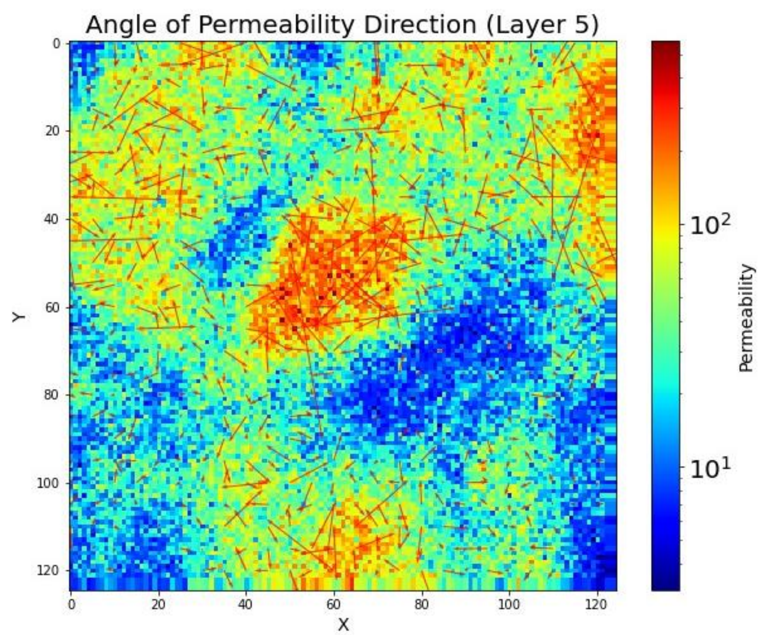


Figure 68. How Angle of Permeability Reduction between every grid cells can be communicated and visualized for Smart Proxy Modeling

4.9.1.10 Fuzzy Clustering of Porosity:

This feature takes into account the different classes of porosity and their representation in the reservoir between different cells and the focal grid cell **(Figure 69 and Figure 70).** This attribute might play a key role in advancing our understanding of reservoir dynamics by:

1. Classifying Porosity Variability:
   - Fuzzy clustering of porosity involves categorizing porosity values into different classes or clusters based on their similarities. This feature provides a representation of the variability in porosity classes across the reservoir.
   - By considering the different classes of porosity, the feature captures the diverse nature of reservoir rock properties and their impact on fluid flow.

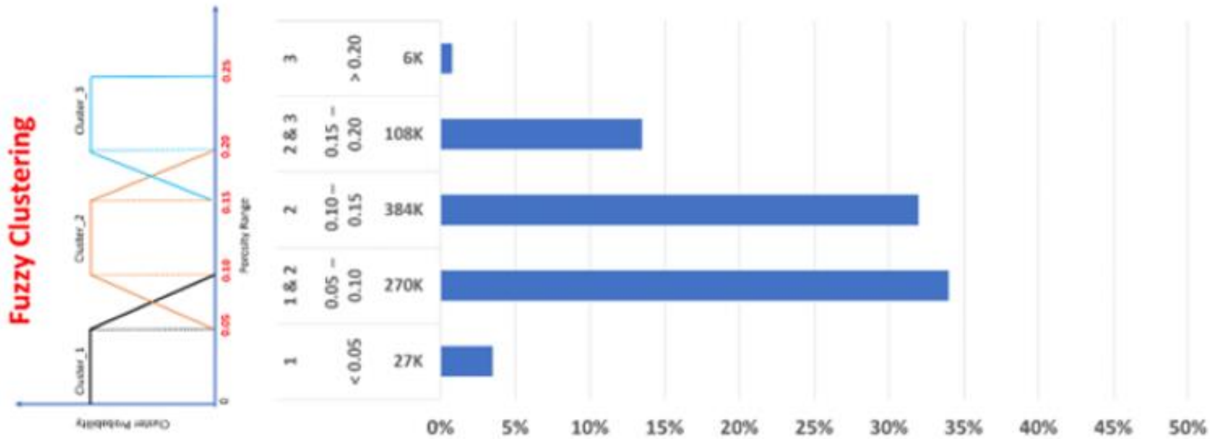
2. Quantifying Heterogeneity:
    - Reservoir heterogeneity, especially in terms of porosity variations, significantly influences fluid flow behavior. The fuzzy clustering feature quantifies the heterogeneity by identifying distinct porosity classes within and around the focal grid cell.
    - Understanding the spatial distribution of porosity classes helps in predicting how fluid flow pathways may vary throughout the reservoir.
3. Impact on Fluid Flow Connectivity:
    - Different porosity classes may have varying permeabilities and fluid storage capacities. The feature contributes to understanding how the fuzzy clustering of porosity influences fluid flow connectivity between the focal grid cell and its neighbors.
    - It helps in identifying zones with similar porosity characteristics, which may exhibit cohesive fluid flow behavior.
4. Integration with Rock Typing:
    - Porosity clustering is closely related to rock typing, where rocks with similar porosity characteristics are grouped together. When integrated into the ANN alongside other features related to rock typing, this feature enriches the model's understanding of how different rock types contribute to fluid flow.
    - Rock typing based on fuzzy clustering enhances the model's ability to differentiate between reservoir zones with distinct porosity characteristics.

4.9.1.11 CO_2 Arrival time to each grid:

This feature takes into account the permeability and porosity of the rock, the viscosity and density of the CO_2, the injection rate, fluid flow communication between different cells and the focal grid cell from the injection wells, providing insights into the timing of CO_2 arrival at various locations within the reservoir **(Figure 71).** This feature may offer insight on its contribution to enhancing comprehension of reservoir dynamics by:

1. Temporal Aspect of Fluid Flow:
    - CO_2 arrival time provides a temporal dimension to fluid flow patterns. It represents the time taken for injected CO_2 to reach specific grid cells within the reservoir.
    - Including this temporal information allows the ANN to capture not only spatial variations but also the dynamic nature of CO_2 propagation over time.
2. Identification of Preferential Flow Paths:
    - The feature helps identify preferential pathways for CO_2 migration within the reservoir. Grid cells that experience earlier CO_2 arrival times may indicate more efficient or dominant flow paths.
    - Understanding these preferential flow paths is crucial for optimizing injection strategies and managing reservoir performance.
3. Quantification of Dispersion and Mixing:
    - CO_2 arrival times across different grid cells provide information on the dispersion and mixing of injected CO_2. Variations in arrival times reflect the complex interactions and mixing of fluids within the reservoir.
    - The feature contributes to the model's ability to simulate and predict the spread of CO_2 across different regions of the reservoir.

4. Impact on Reservoir Sweep Efficiency:
   - Reservoir sweep efficiency is a critical factor in enhanced oil recovery (EOR) and carbon capture and storage (CCS) projects. CO_2 arrival time helps assess how effectively injected CO_2 sweeps through the reservoir, reaching different regions.
   - Optimizing reservoir sweep efficiency is essential for maximizing resource recovery or ensuring successful CO_2 sequestration.
5. Integration with Well Management Strategies:
   - Well management strategies, such as controlling injection rates and locations, can significantly influence CO_2 arrival times. This feature allows the ANN to learn how changes in well management practices impact the timing of CO_2 arrival at specific grid cells.
   - It supports the optimization of injection strategies for achieving desired reservoir performance.
6. Dynamic Learning Capability:
   - Including CO_2 arrival time as a feature enables the ANN to adapt to changing reservoir conditions over time. The model learns from historical data to predict how CO_2 arrival times may evolve based on varying operational parameters.
   - The dynamic nature of this feature enhances the model's predictive capabilities in response to changing reservoir dynamics.
7. Validation of Reservoir Connectivity:
   - CO_2 arrival time serves as a validation metric for assessing reservoir connectivity. Consistent arrival times in well-connected regions and variations in poorly connected areas provide insights into reservoir connectivity patterns.
   - The feature supports the identification of reservoir compartments and potential barriers to fluid flow.



Fuzzy Clustering **Total Cells ~ 800K**

Cluster Number	Cluster Porosity Range	Cell Count	Cell Percentage
1	< 0.05	27K	3.5 %
1 & 2	0.05 – 0.10	270K	34 %
2	0.10 – 0.15	384K	32 %
2 & 3	0.15 – 0.20	108K	13.5 %
3	> 0.20	6K	0.79 %

porosity	FCM_1	FCM_2	FCM_3	i	j	k
0.145300	0.000000	1.000000	0.00000	101.0	1.0	5.0
0.163154	0.000000	0.736920	0.26308	102.0	1.0	5.0
0.104900	0.000000	1.000000	0.00000	103.0	1.0	5.0
0.123852	0.000000	1.000000	0.00000	104.0	1.0	5.0
0.121333	0.000000	1.000000	0.00000	105.0	1.0	5.0
0.087525	0.249498	0.750502	0.00000	106.0	1.0	5.0
0.140704	0.000000	1.000000	0.00000	107.0	1.0	5.0
0.134200	0.000000	1.000000	0.00000	108.0	1.0	5.0
0.126601	0.000000	1.000000	0.00000	109.0	1.0	5.0
0.118612	0.000000	1.000000	0.00000	110.0	1.0	5.0
0.105579	0.000000	1.000000	0.00000	111.0	1.0	5.0

Figure 69. Fuzzy clustering of porosity

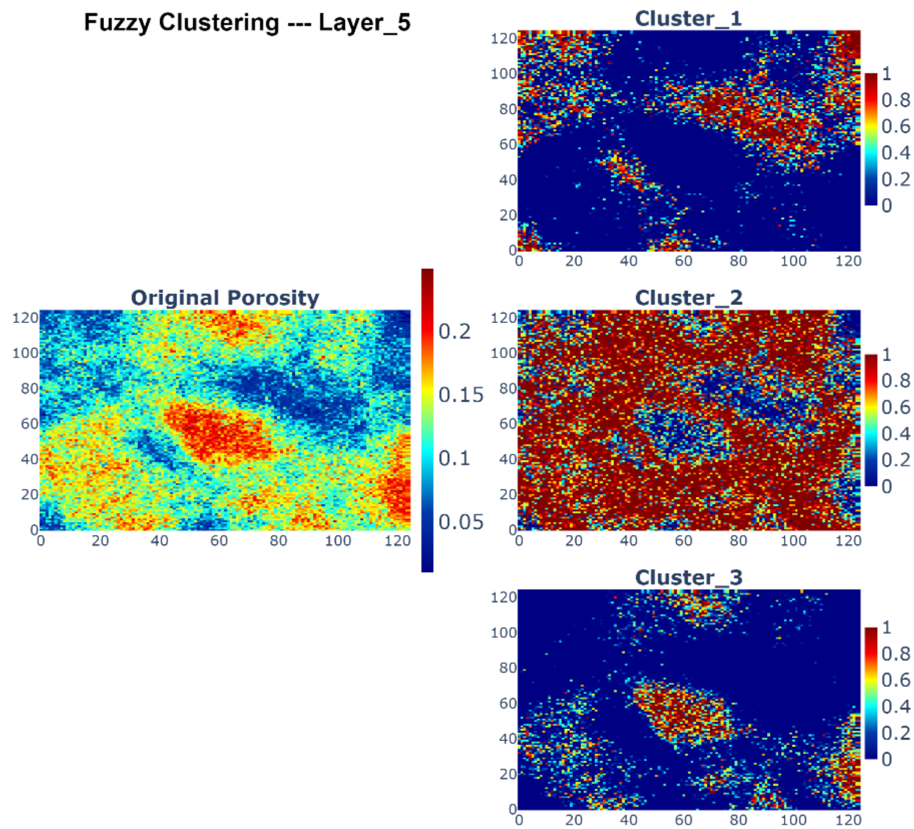


Figure 70. How Fuzzy Clustering of Porosity between every grid cells can be communicated and visualized for Smart Proxy Modeling

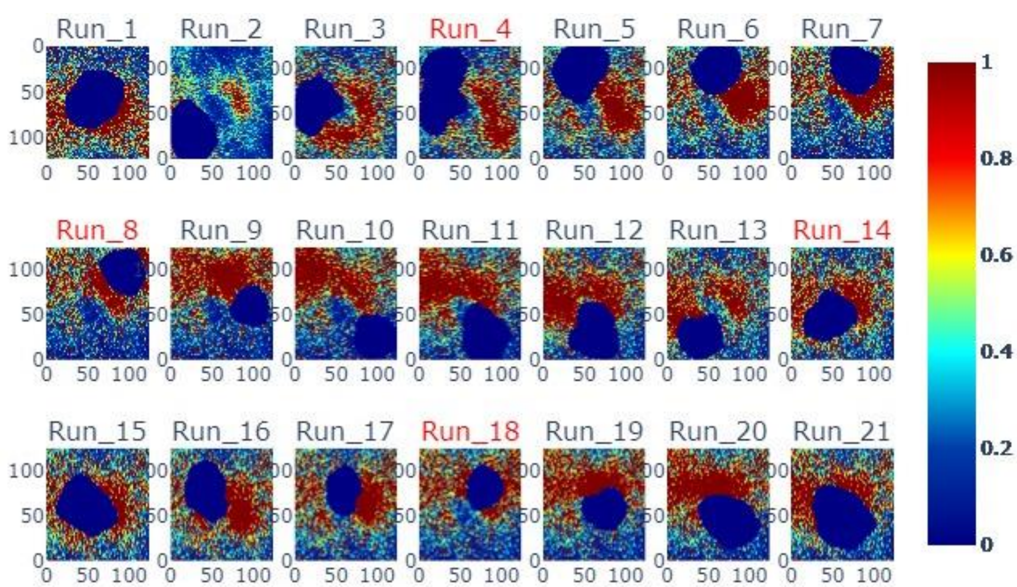


Figure 71. How CO_2 Arrival time to each grid can be communicated and visualized for Smart Proxy Modeling

4.10 Not all features work for all cases:

It's crucial to acknowledge that the effectiveness of features may be problem-specific, and their utility can vary based on the unique characteristics and nature of each reservoir problem. Reservoirs exhibit diverse geological features, and the appropriateness of certain features depends on the specific challenges or properties of the reservoir in question. It's important to consider the following points in your explanation:

1. **Reservoir-Specific Nature:** Reservoirs vary significantly in terms of geological formations, structures, and fluid flow behaviors. The unique nature of each reservoir problem may necessitate different types and qualities of features for effective characterization.
2. **Tailoring Features to Reservoir Characteristics:** The choice of features, including porosity indices, should be tailored to the specific reservoir characteristics. Features that work well for one reservoir problem may not be universally applicable to all scenarios.
3. **Problem-Specific Challenges:** Discuss the challenges inherent to the particular reservoir problem you're addressing. Certain features, including porosity indices, may address specific challenges in understanding porosity distribution, but their applicability might be limited to the context of those challenges.
4. **Geological Heterogeneity:** The geological heterogeneity within reservoirs and how this heterogeneity may influence the relevance and effectiveness of certain features. Features that capture fine-scale variations may be more suitable for some reservoirs than others.
5. **Validation and Adaptation:** The importance of validating the chosen features against observed data and reservoir characteristics. Additionally, be open to adapting or refining feature selections based on the performance observed during the modeling process.

While the introduced features, may offer valuable insights for a specific reservoir problem, their success is contingent on careful consideration of the reservoir's unique attributes. Flexibility and adaptability in feature selection are crucial to address the diverse challenges presented by different reservoirs.

CHAPTER 5: RESULTS AND DISCUSSIONS

5.1 Review:

In this chapter, the focus is solely on a specific time step, precisely 30 years after the commencement of injection on 01-01-2050. The primary objective is to showcase the outcomes of blind runs generated by the Smart Proxy Model (SPM). These blind runs, constituting new simulation instances not utilized during the training phase, were executed on the developed Smart Proxy models. The results were generated at the grid block level and for the selected time step in the reservoir simulation model. To assess the predictive capabilities and robustness of the Smart Proxy, a comparison is drawn against results obtained from CMG simulations for layer of the reservoir model.

Visual representations, including two-dimensional distribution maps for each layer aid in assessing the quality of results from the Smart Proxy and numerical reservoir simulation. With a total of 51 sand layers in different geological realizations and several phases of complexity, each phase of complexity an extensive dataset comprising 204 plots, only for the blind cases. For conciseness, this chapter exclusively presents the results of only one of the blind realizations in layer 5, and layer 51, while additional outcomes could be found in the Appendix.

It is important to emphasize that all neural networks employed and developed in this study were configured with a single-layer structure, and no extensive hyper-parameter tuning was conducted to elaborate on the network's depth. The primary emphasis of this research was directed towards conveying the fundamental principles of fluid flow within the domain of petroleum engineering to the neural network.

5.2. Phase 1 Results:

In Phase 1 of the study, the focus is on applying rotation to realizations as a means of addressing the multi-scale complexity, sparsity, and inherent uncertainty associated with geological heterogeneity, which serves as a primary control factor for fluid flow in a reservoir and significantly influences engineering and management decisions. The motivation behind this phase includes the goal of advancing realism by introducing variability through the rotation of each realization, thereby simulating real-world geological heterogeneity. This approach aims to enhance the adaptability of neural networks (NN) by enabling them to recognize and understand the impact of geological variations, ultimately improving predictive accuracy. The advantages and comparisons of this approach include enhanced computational efficiency, as a trained NN on rotated realizations provides faster predictions compared to rerunning simulations with rotated scenarios. Furthermore, the insights gained from this process contribute to a better understanding of anisotropy and its effects on CO_2 movement, providing valuable guidance for reservoir engineers in real-world scenarios.

5.2.1 Results of Reservoir Pressure Distribution for Blind Datasets:

High accuracy is demonstrated by the developed Smart Proxy in predicting and replicating reservoir pressure distribution results for blind runs at the selected time step. With an error margin of less than 5%, the results are in constant alignment with CMG results over all the reservoir layers **(Fig. 72 through Fig. 75).** Each row represents the result for one time step; left plot is CMG, middle plot is Smart Proxy, and far right plot is the percentage error for the pressure case and absolute error for the saturation cases.

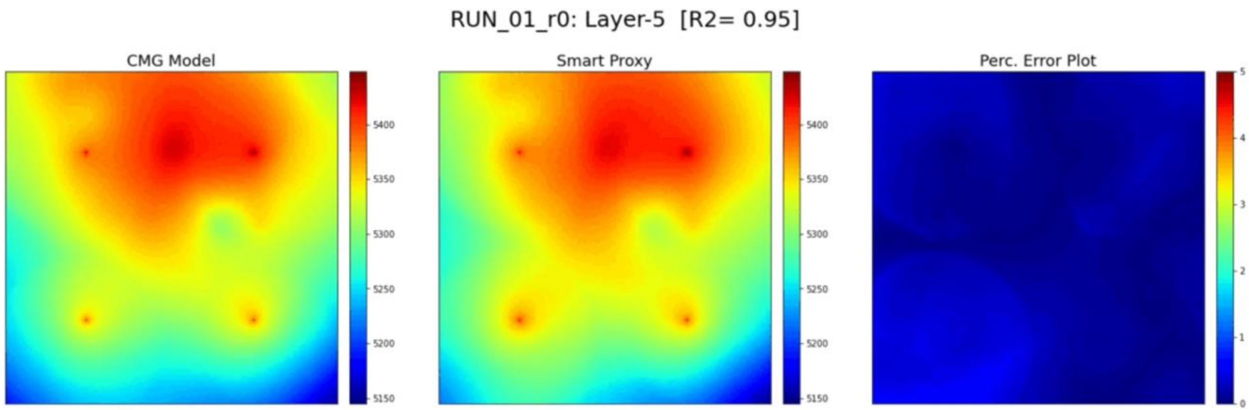


Figure 72. Pressure results for layer #5 for one of the blind models used in Phase 1

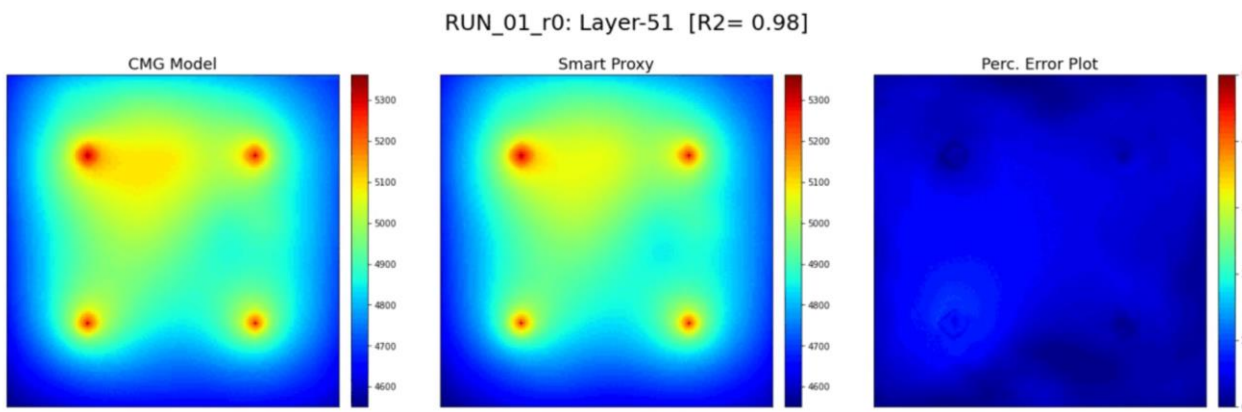


Figure 73. Pressure results for layer #51 for one of the blind models used in Phase 1

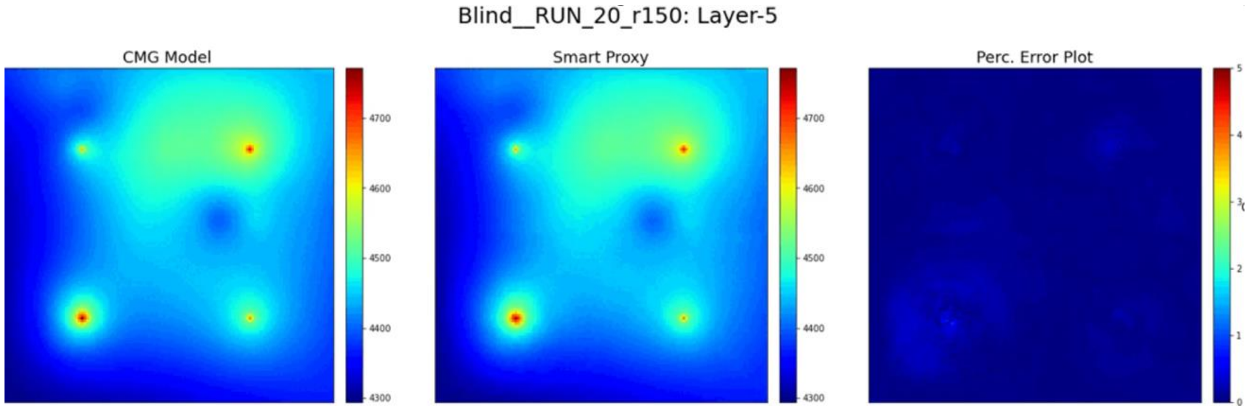


Figure 74. Pressure results for layer #5 for another blind model used in Phase 1

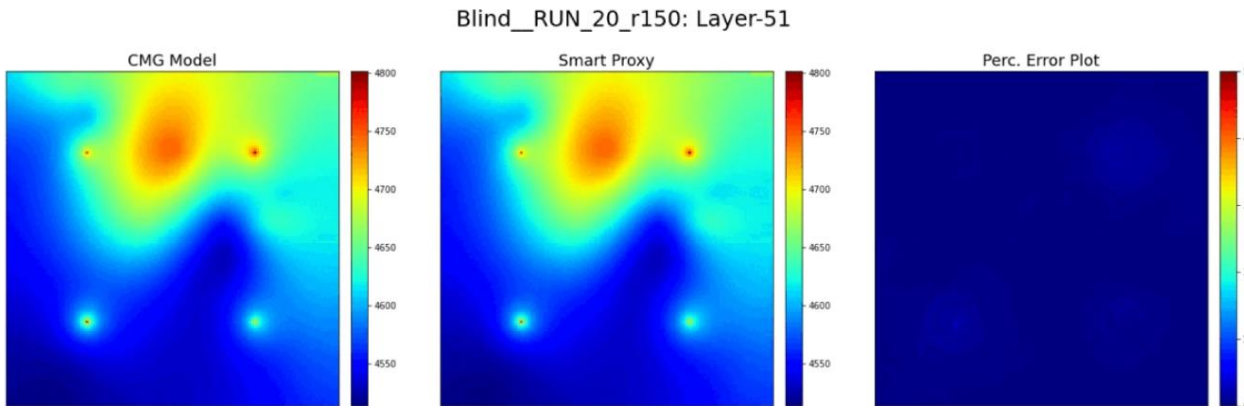


Figure 75. Pressure results for layer #51 for another blind model used in Phase 1

5.2.2 Results of Reservoir Saturation Distribution for Blind Datasets:

High accuracy is demonstrated by the developed Smart Proxy in predicting and replicating CO_2 saturation distribution results for blind runs at the selected time step. With an error margin of less than 10%, the results are in constant alignment with CMG results over all the reservoir layers **(Fig. 76 and Fig. 77).** Each row represents the result for one time step; left plot is CMG, middle plot is Smart Proxy, and far right plot is the percentage error for the pressure case and absolute error for the saturation cases.

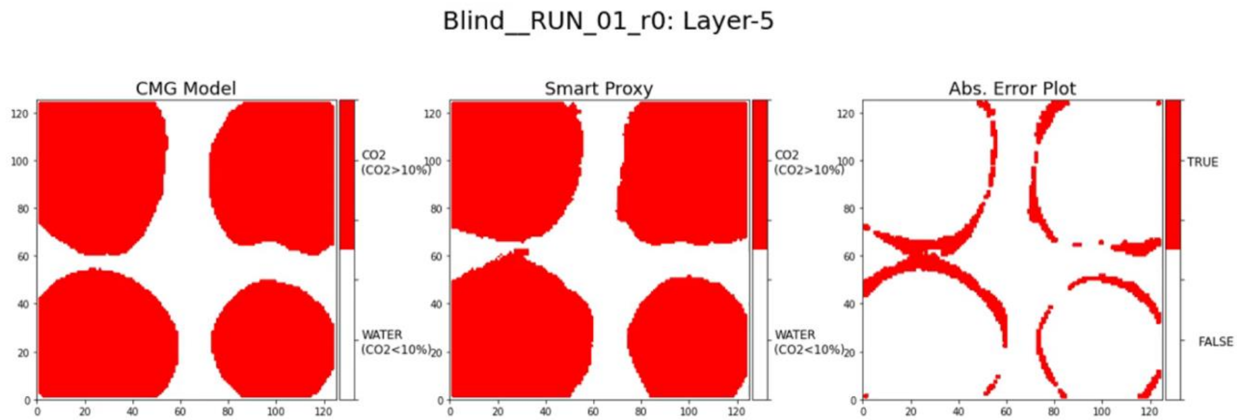


Figure 76. Binary visualization. Saturation results for layer #5 for another blind model used in Phase 1

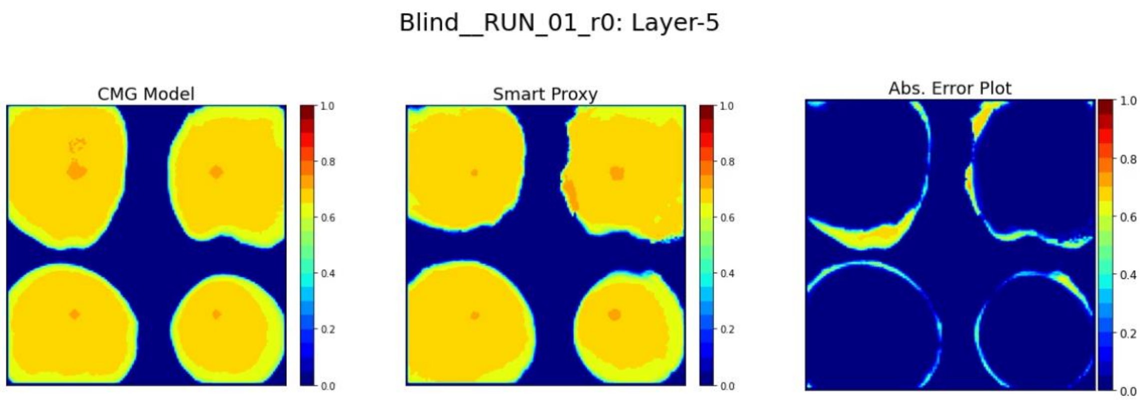


Figure 77. Discrete visualization. Saturation results for layer #5 for another blind model used in Phase 1

5.2.3 Effect of the hidden layers on the performance of the model on the blind dataset:

In an artificial neural network (ANN) study for CO_2 pressure prediction, three architectures were compared:

- Single Hidden Layer ANN with 1000 neurons.
- 3 Hidden Layer ANN with 1000, 600, and 200 neurons.
- 5 Hidden Layer ANN with 1000, 800, 600, 400, and 200 neurons.

The results, presented in **Fig 78**, indicate that models with additional hidden layers exhibit slightly better performance on blind datasets compared to the single hidden layer ANN. However, this improvement comes with a slightly higher computational cost. The trade-off between enhanced performance and increased computational requirements is evident in the study's findings. Given this comparison, for the remaining part of the study, the single hidden layer ANN was utilized.

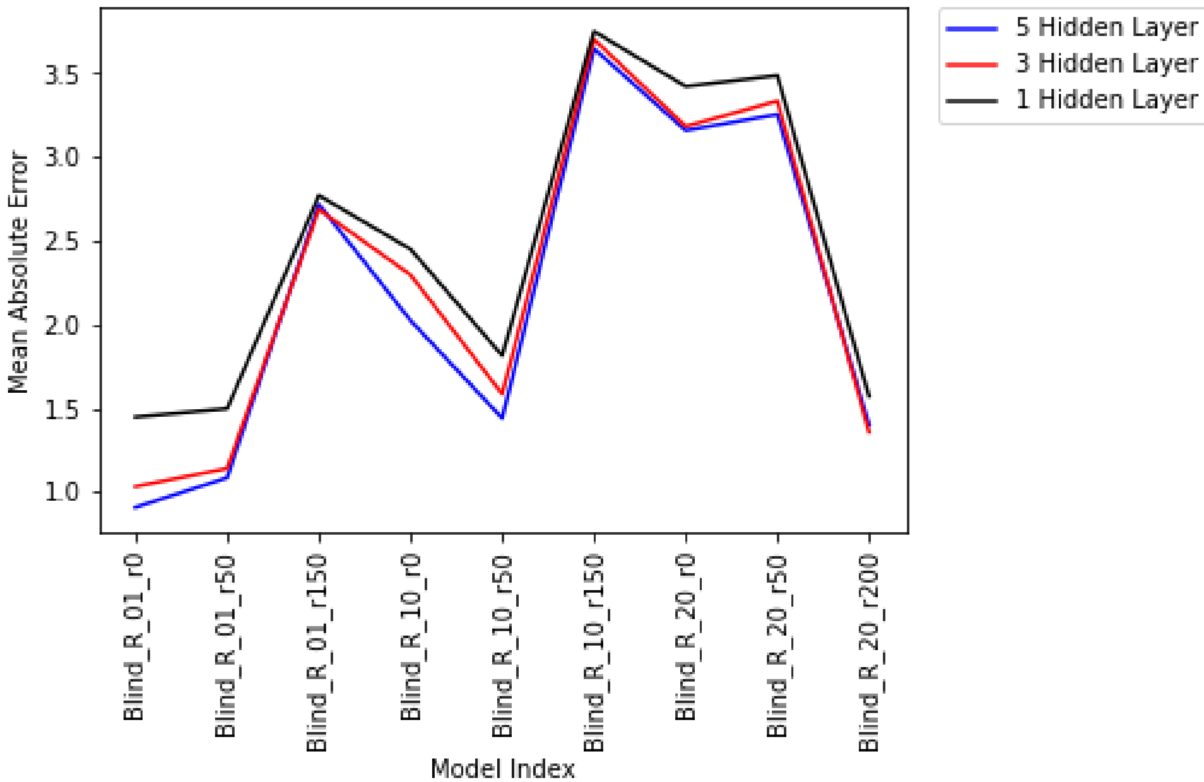
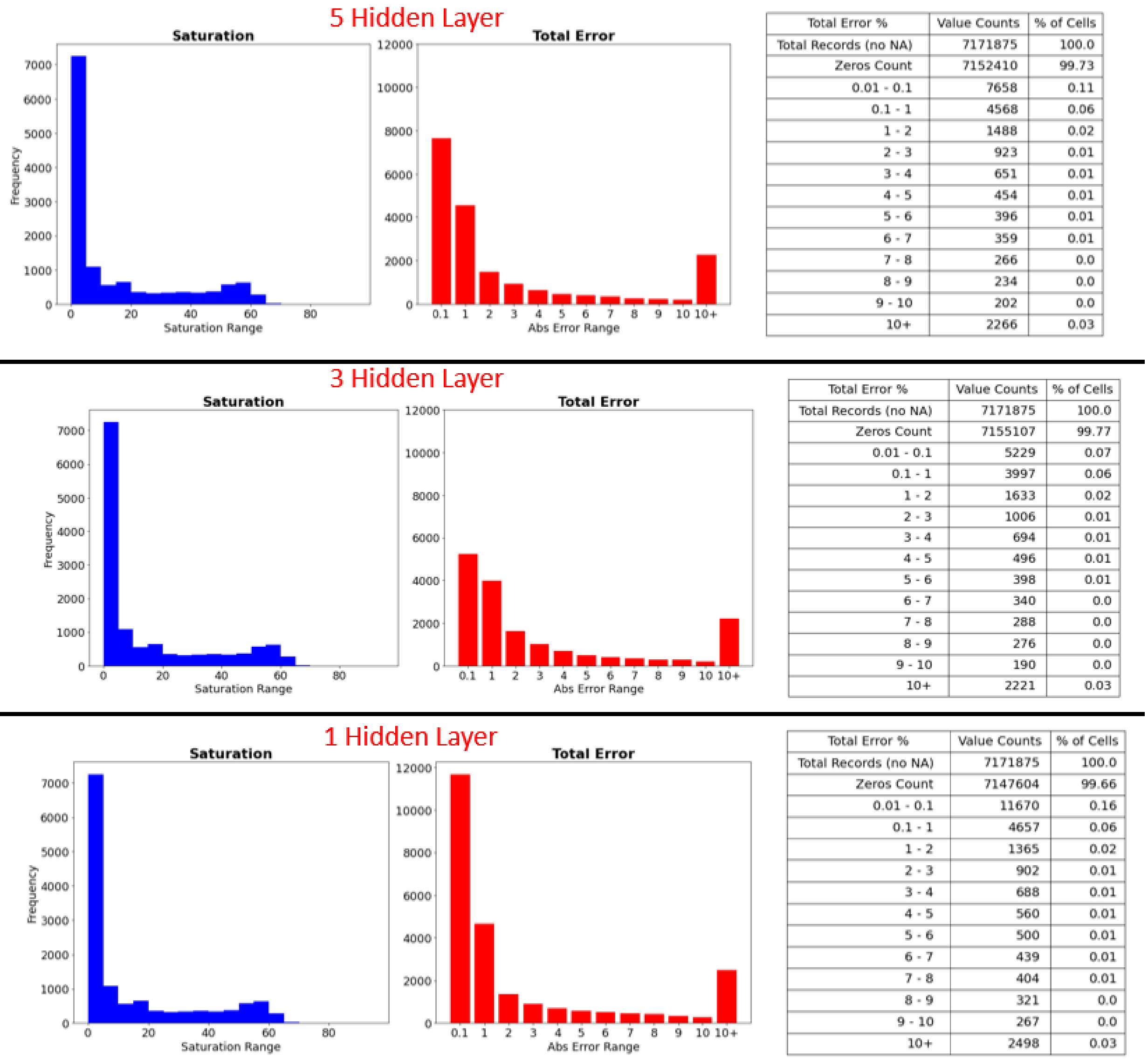


Figure 78. Comparing the impact of changing the number of hidden layers on prediction for all the blind realizations

Fig. 79 also shows the distribution of output error and their ranges with respect to the number of grid cells influenced by the error for different number of hidden layers. The findings showed that for the CO_2 saturation, much of the improvements for deeper ANN architecture is directly tied to the cells with reduced amount of CO_2 saturation in them. Such behavior specifically in the absolute error range of 0.01 to 0.1, could be attributed to the varying capacity of these models to capture and represent complex relationships within the data. The model with a single hidden layer may struggle to capture intricate patterns and nuanced variations in the data, particularly in regions where the CO_2 saturation levels are moderately close. This limitation could result in a broader distribution of errors across a larger number of grid cells within the specified error range. Deeper architectures, with multiple hidden layers, have a greater capacity to learn hierarchical representations of features. This enhanced capacity allows them to capture finer details and complex relationships within the data. In the specified error range, these deeper models may exhibit a more concentrated distribution of errors, with fewer grid cells having errors falling within this range but potentially achieving higher precision in those cells.



5 Hidden Layer

Total Error %	Value Counts	% of Cells
Total Records (no NA)	7171875	100.0
Zeros Count	7152410	99.73
0.01 - 0.1	7658	0.11
0.1 - 1	4568	0.06
1 - 2	1488	0.02
2 - 3	923	0.01
3 - 4	651	0.01
4 - 5	454	0.01
5 - 6	396	0.01
6 - 7	359	0.01
7 - 8	266	0.0
8 - 9	234	0.0
9 - 10	202	0.0
10+	2266	0.03

3 Hidden Layer

Total Error %	Value Counts	% of Cells
Total Records (no NA)	7171875	100.0
Zeros Count	7155107	99.77
0.01 - 0.1	5229	0.07
0.1 - 1	3997	0.06
1 - 2	1633	0.02
2 - 3	1006	0.01
3 - 4	694	0.01
4 - 5	496	0.01
5 - 6	398	0.01
6 - 7	340	0.0
7 - 8	288	0.0
8 - 9	276	0.0
9 - 10	190	0.0
10+	2221	0.03

1 Hidden Layer

Total Error %	Value Counts	% of Cells
Total Records (no NA)	7171875	100.0
Zeros Count	7147604	99.66
0.01 - 0.1	11670	0.16
0.1 - 1	4657	0.06
1 - 2	1365	0.02
2 - 3	902	0.01
3 - 4	688	0.01
4 - 5	560	0.01
5 - 6	500	0.01
6 - 7	439	0.01
7 - 8	404	0.01
8 - 9	321	0.0
9 - 10	267	0.0
10+	2498	0.03

Figure 79. Distribution of output error percentage accounted for by the number of grid cells for one of the blind dataset in the ANN CO_2 model

5.3. Phase 2 Results:

Despite notable progress in reservoir characterization and modeling techniques, achieving a perfect alignment between predictive models and real-world observations remains challenging. An alternative strategy for establishing conformance involves showcasing the continuous refinement of predictive models as new monitoring data accumulates, emphasizing the robustness of geological models and modeling assumptions. This approach suggests that additional data contributes to ongoing model improvement rather than necessitating radical overhauls. The motivation behind this strategy lies in achieving operational realism by replicating scenarios where well numbers and locations may change due to operational adjustments or optimization efforts. It also facilitates neural network (NN) generalization, allowing the NN to adapt effectively to scenarios with varying well numbers. The advantages and comparisons of this approach include resource optimization, as the NN efficiently handles diverse well configurations, thereby reducing the need for extensive numerical simulations. Additionally, it aids in risk identification by highlighting potential risks associated with changes in well numbers, contributing to effective risk assessment and mitigation planning.

5.3.1 Results of Reservoir Pressure Distribution for Blind Datasets:

The pressure results for Blind realization are presented in this section. It must be noted that the first four layers, the layers 29 to 34 and 62 to 65 are shale barriers, and the remaining 51 layers of the reservoir are the actual reservoir layers. With an error margin of less than 5%, the results are in constant alignment with CMG results over all the reservoir layers **(Fig. 80 and Fig 81).**

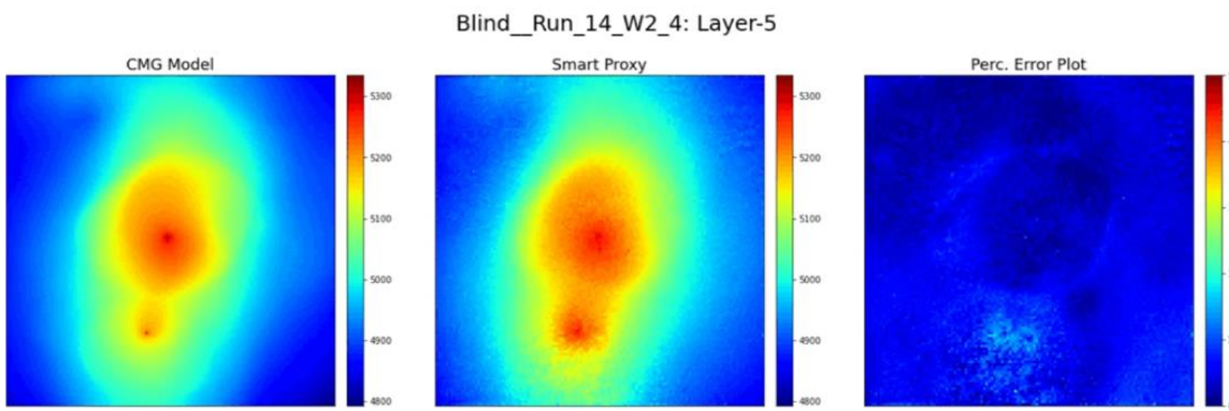


Figure 80. Pressure results for layer #5 for one of the blind models used in Phase 2

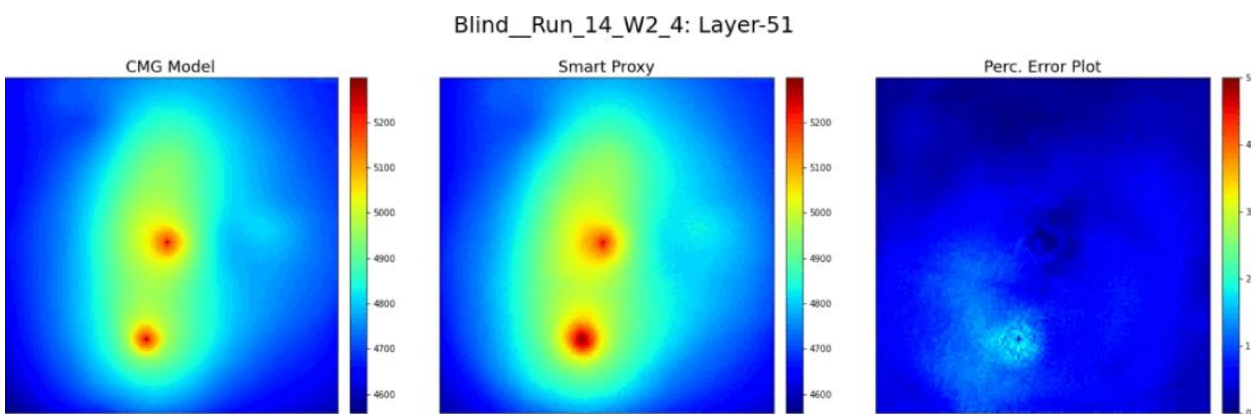


Figure 81. Pressure results for layer #51 for one of the blind models used in Phase 2

5.3.2 Results of Reservoir Saturation Distribution for Blind Datasets:

With an error margin of less than 10%, the results are in constant alignment with CMG results over all the reservoir layers **(Fig. 82 through Fig. 84).**

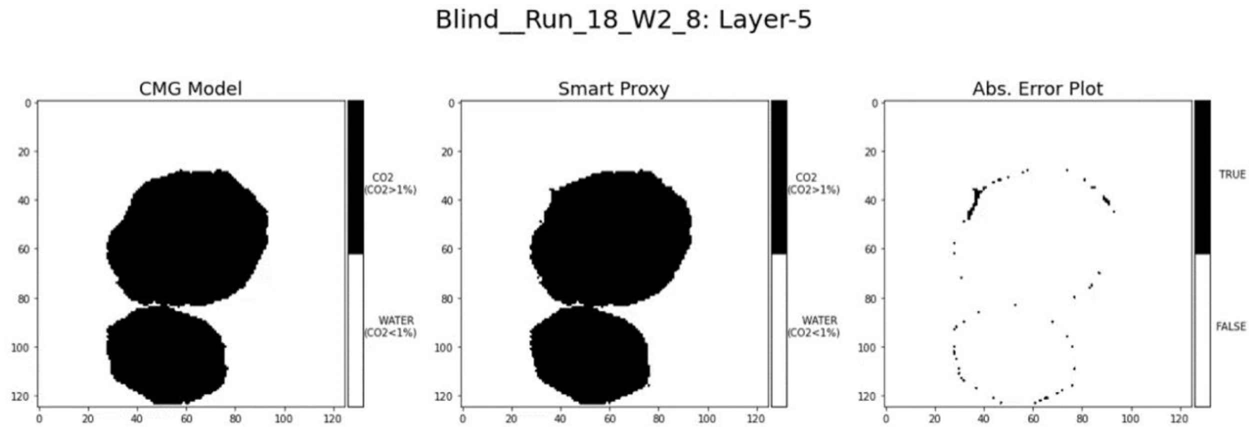


Figure 82. Binary visualization. Saturation results for layer #5 for one of the blind models used in Phase 2

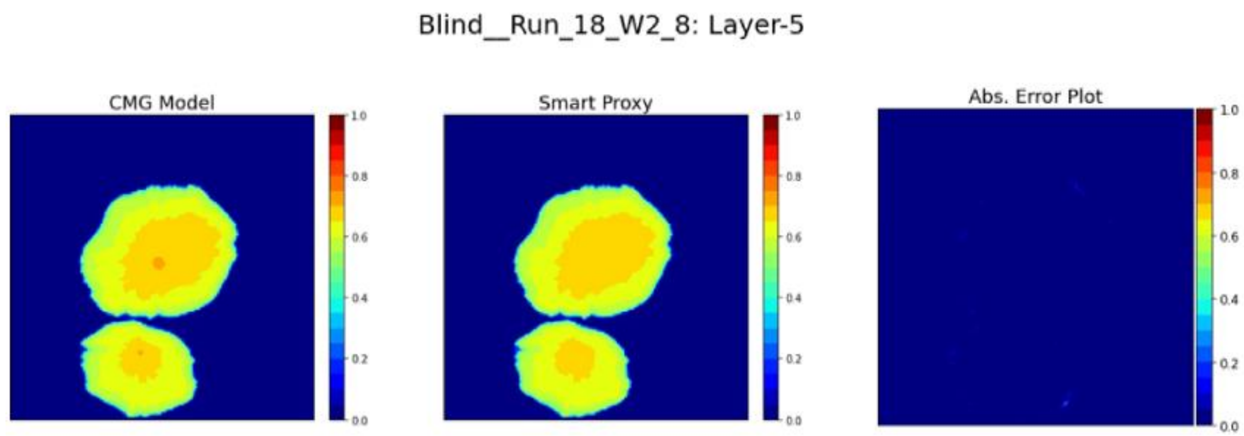


Figure 83. Discrete visualization. Saturation results for layer #5 for one of the blind models used in Phase 2

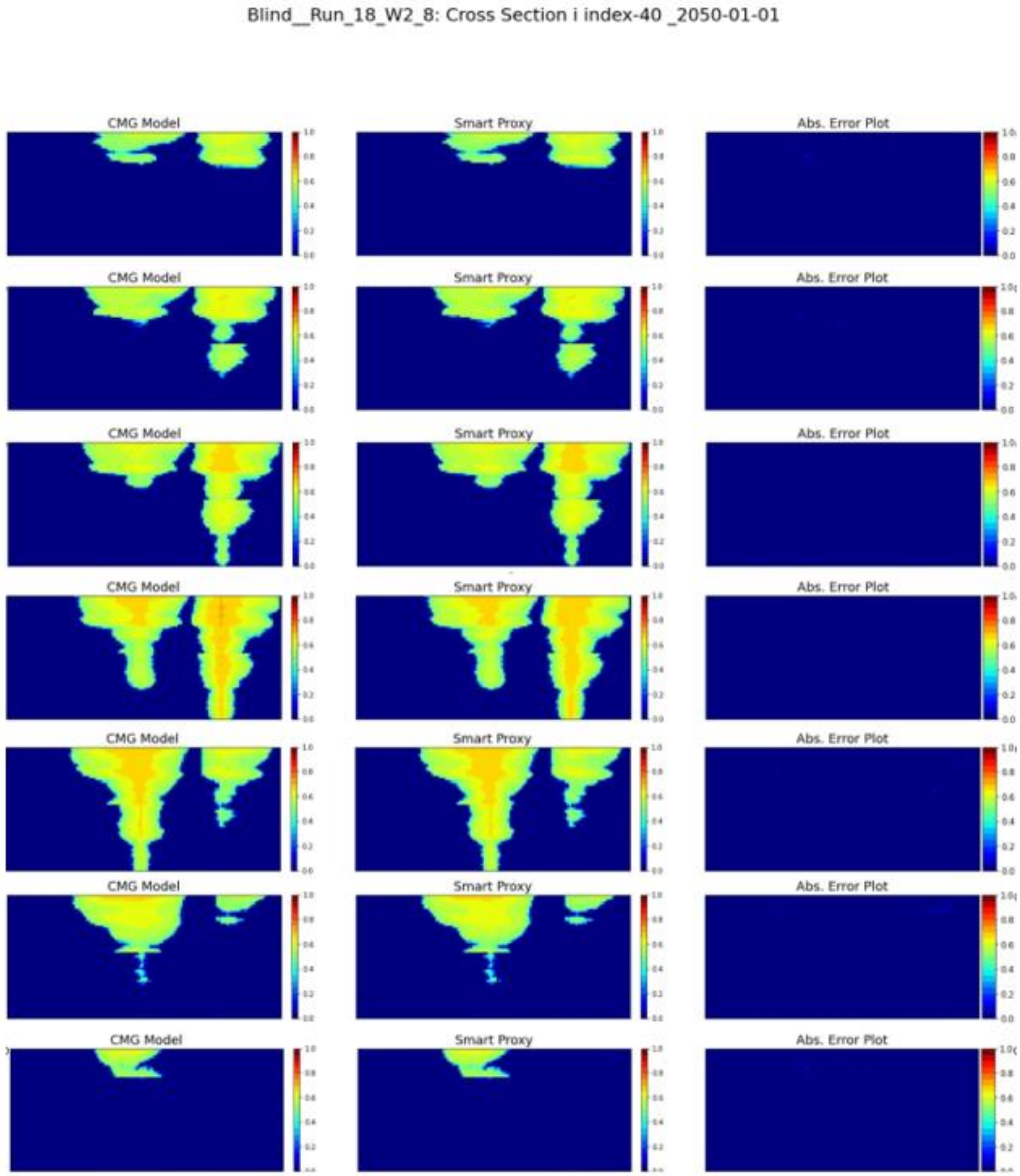


Figure 84. Cross sectional view of Saturation results for JK 2D of various slices for one of the blind models used in Phase 2

5.4. Phase 3 Results:

In Phase 3, the focus is on addressing uncertainties surrounding the long-term entrapment of free CO_2 underground, necessitating extended monitoring and escalating costs. To tackle these concerns, expediting CO_2 dissolution and minimizing free CO_2 concentrations in the subsurface emerges as an effective solution. The motivation behind this phase lies in introducing uncertainty at a localized level, mirroring challenges in precisely placing wells, and adapting neural networks (NN) to understand the impact of such localized changes, thereby improving predictive capabilities. The advantages and comparisons include providing insights into the global effects of well placement changes for enhanced decision-making in specific reservoir areas, and the NN's adaptability to predicting outcomes for smaller adjustments after learning from localized changes.

5.4.1 Results of Reservoir Pressure Distribution for Blind Datasets:

The pressure results for Blind realization are presented in this section. It must be noted that the first four layers, the layers 29 to 34 and 62 to 65 are shale barriers, and the remaining 51 layers of the reservoir are the actual reservoir layers. High accuracy is demonstrated by the developed Smart Proxy in predicting and replicating reservoir pressure distribution results for blind runs at the selected time step. With an error margin of less than 5%, the results are in constant alignment with CMG results over all the reservoir layers **(Fig. 85 and Fig. 86).**

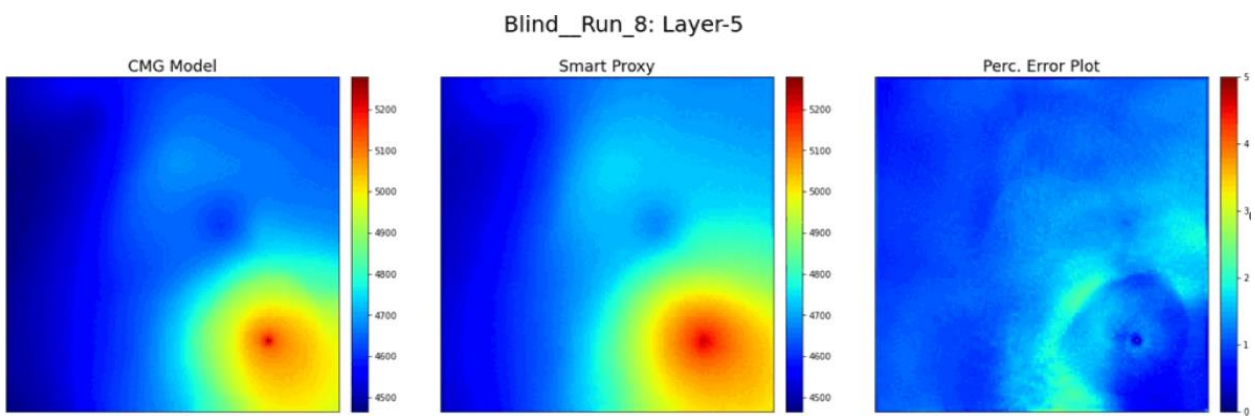


Figure 85. Pressure results for layer #5 for one of the blind models used in Phase 3

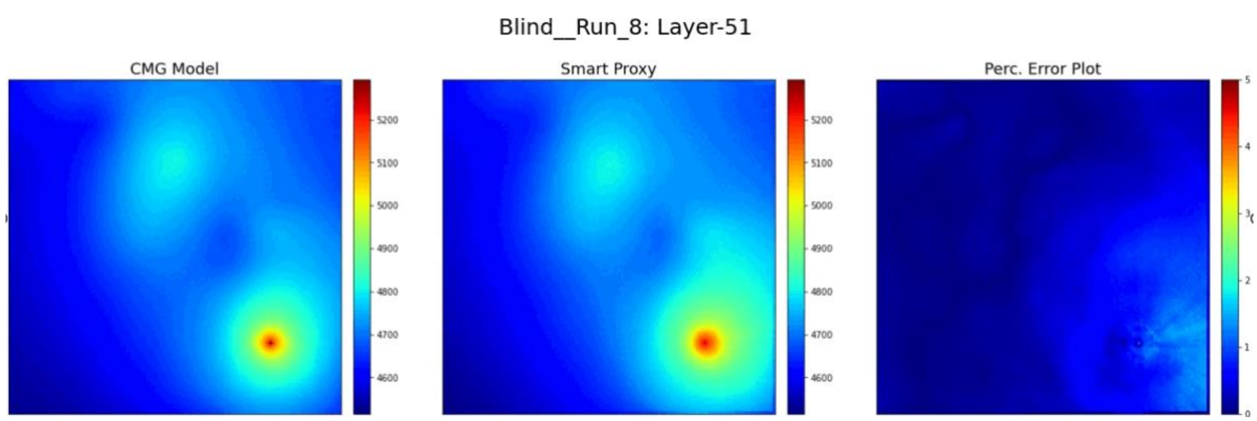


Figure 86. Pressure results for layer #51 for one of the blind models used in Phase 3

5.4.2 Results of Reservoir Saturation Distribution for Blind Datasets:

With an error margin of less than 10%, the results are in constant alignment with CMG results over all the reservoir layers **(Fig. 87 and Fig 88).**

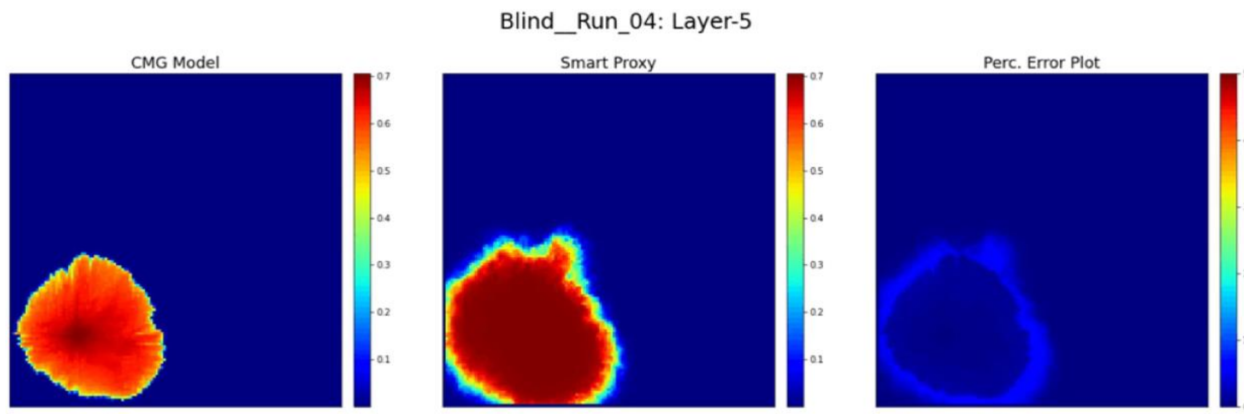


Figure 87. Binary visualization. Saturation results for layer #5 for one of the blind models used in Phase 3

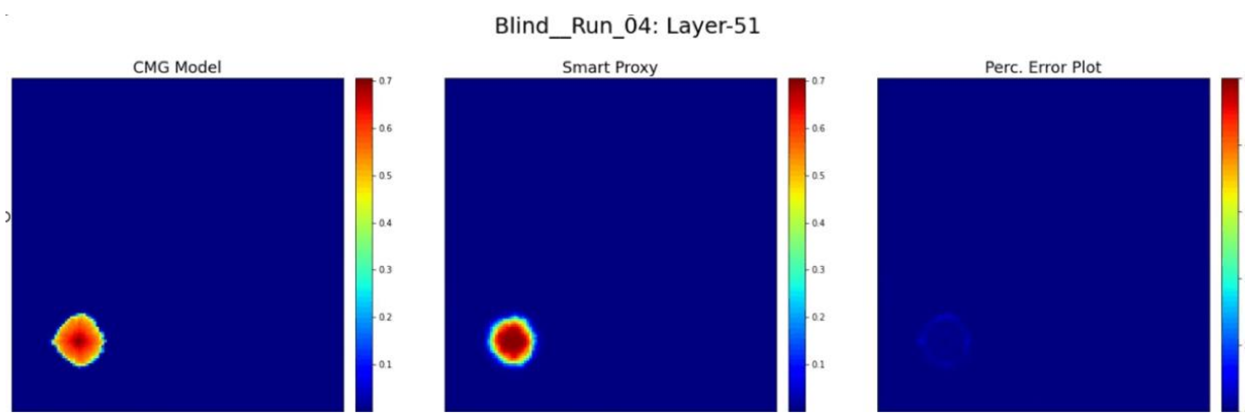


Figure 88. Discrete visualization. Saturation results for layer #51 for one of the blind models used in Phase 3

5.5. Phase 4 Results:

In Phase 4, the emphasis is on modifying wells across different geological realizations, motivated by the exploration of how changes in well configuration influence system behavior. The goal is to understand the dynamics between geological variations and well modifications, providing a holistic view of their interplay. This phase also contributes to training the neural network (NN) to generalize from scenarios with diverse well configurations, thereby enhancing its applicability. The advantages and comparisons include the NN's ability to efficiently capture the interplay between geological variations and well modifications, offering a comprehensive understanding, while simultaneously reducing the computational burden by handling diverse well configurations without requiring separate numerical simulations for each case.

5.5.1 Results of Reservoir Pressure Distribution for Blind Datasets:

In this section, we showcase the pressure outcomes for the Blind realization. It is crucial to highlight that the shale barriers encompass the first four layers, layers 29 to 34, and 62 to 65, while the remaining 51 layers constitute the actual reservoir layers. The Smart Proxy developed exhibits notable accuracy in predicting and reproducing the reservoir pressure distribution results for blind simulations at the specified time step. The achieved precision is evident, with an error margin consistently below 10%, indicating a consistent alignment with CMG results across all reservoir layers **(Fig. 89 and Fig. 90).**

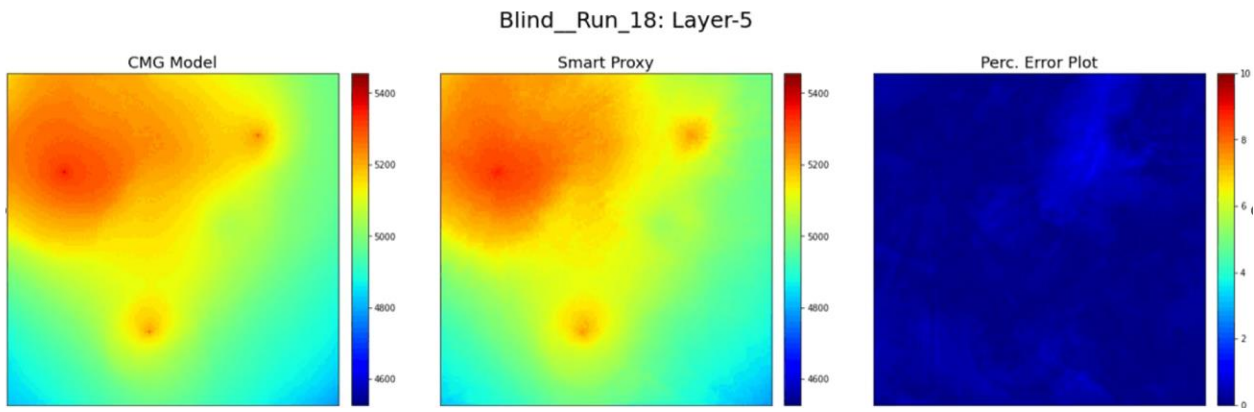


Figure 89. Pressure results for layer #5 for one of the blind models used in Phase 4

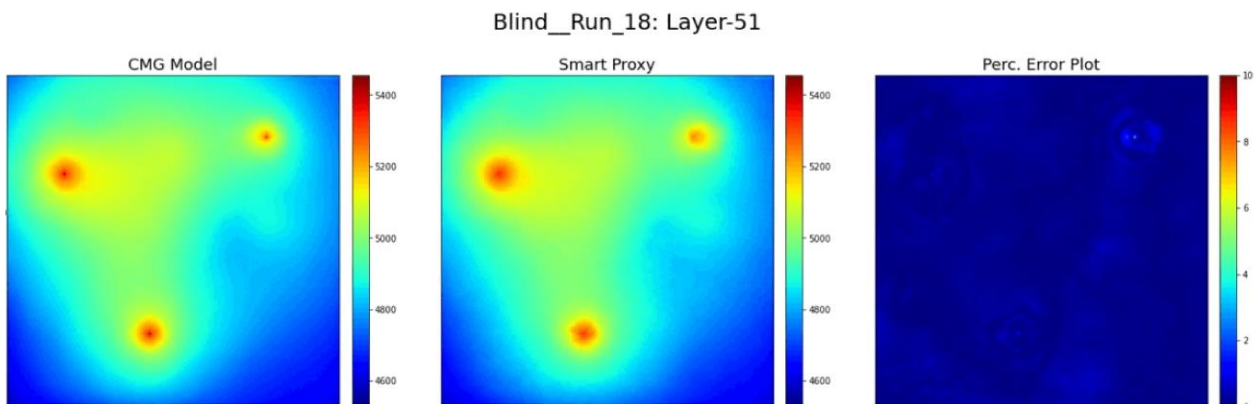


Figure 90. Pressure results for layer #51 for one of the blind models used in Phase 4

5.5.2 Results of Reservoir Saturation Distribution for Blind Datasets:

With an error margin of less than 10%, the results are in constant alignment with CMG results over all the reservoir layers **(Fig. 91 and Fig. 92).**

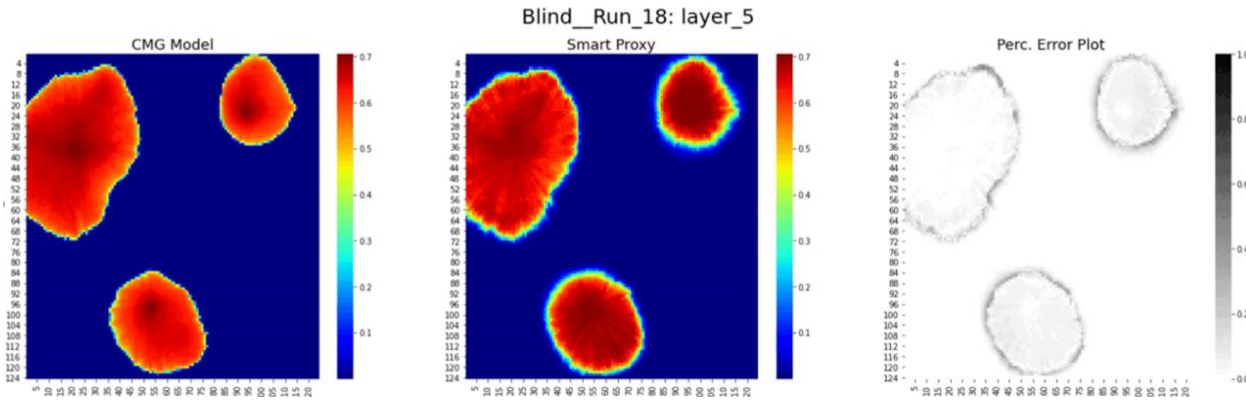


Figure 91. Binary visualization. Saturation results for layer #5 for one of the blind models used in Phase 4

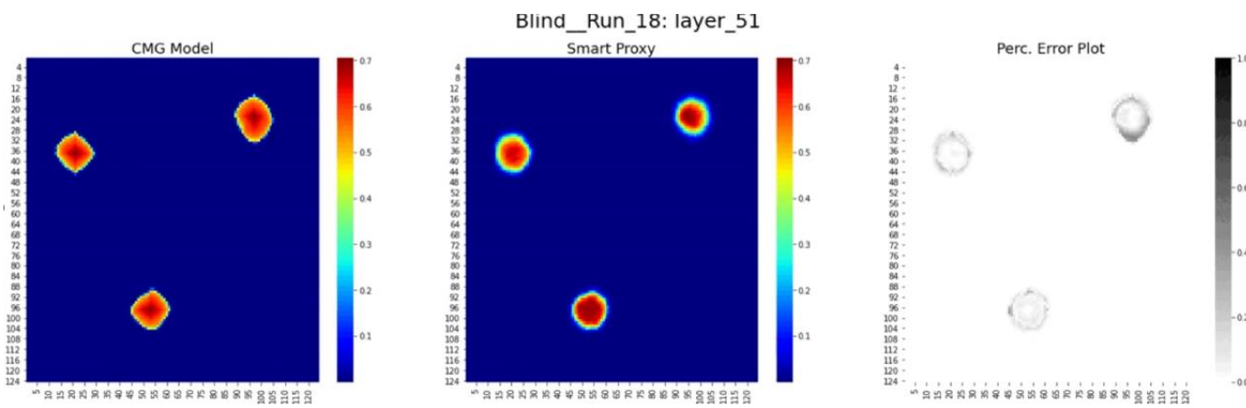


Figure 92. Discrete visualization. Saturation results for layer #51 for one of the blind models used in Phase 4

5.6. Phase 5 Results:

In Phase 5, the focus shifts to adjusting well locations within the same geological realization, motivated by the exploration of the impact of localized well adjustments within a consistent geological setting. This phase aims to investigate the fine-tuning of well locations locally and trains the neural network (NN) to adapt to smaller, localized changes, ultimately enhancing its precision. The advantages and comparisons of this phase include the NN's proficiency in predicting outcomes for scenarios involving localized well adjustments and different injection well patterns, providing localized precision. Additionally, resource savings are achieved as the NN, once trained, can provide predictions without the need for extensive numerical simulations for localized changes.

Fig. 93 show the location and pattern of different injection wells overlaid on a porosity distribution for layer 5. The highlighted boxes are the blind realization that their results are presented in the following figures.

5.6.1 Results of Reservoir Pressure Distribution for Blind Datasets:

The pressure results for Blind realization are presented in this section. It must be noted that the first four layers, the layers 29 to 34 and 62 to 65 are shale barriers, and the remaining 51 layers of the reservoir are the actual reservoir layers. High accuracy is demonstrated by the developed Smart Proxy in predicting and replicating reservoir pressure distribution results for blind runs at the selected time step. With an error margin of less than 5%, the results are in constant alignment with CMG results over all the reservoir layers **(Fig. 94 through Fig. 97).**

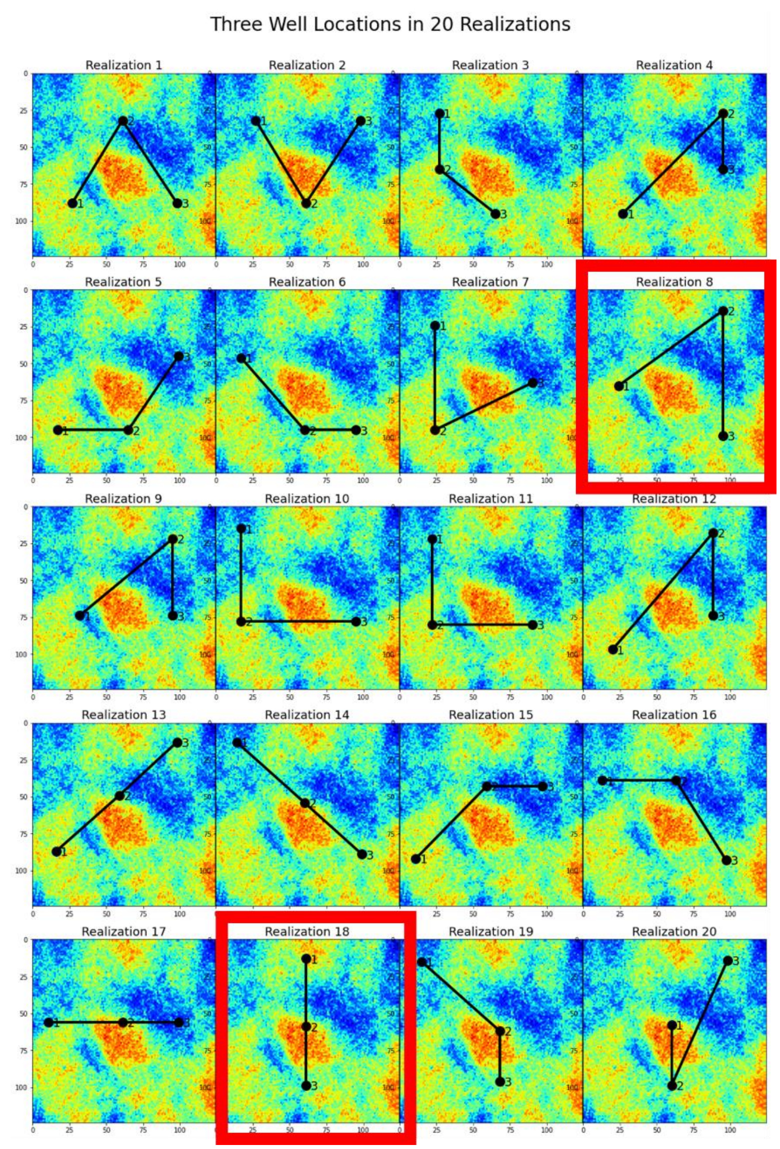


Figure 93. Different well pattern configuration used in phase 5 of this study

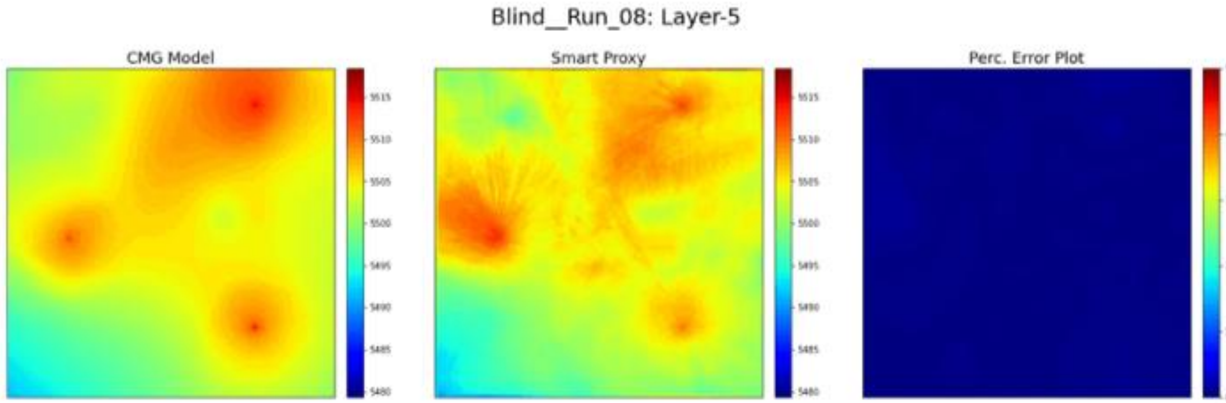


Figure 94. Pressure results for layer #5 for one of the blind models used in Phase 5

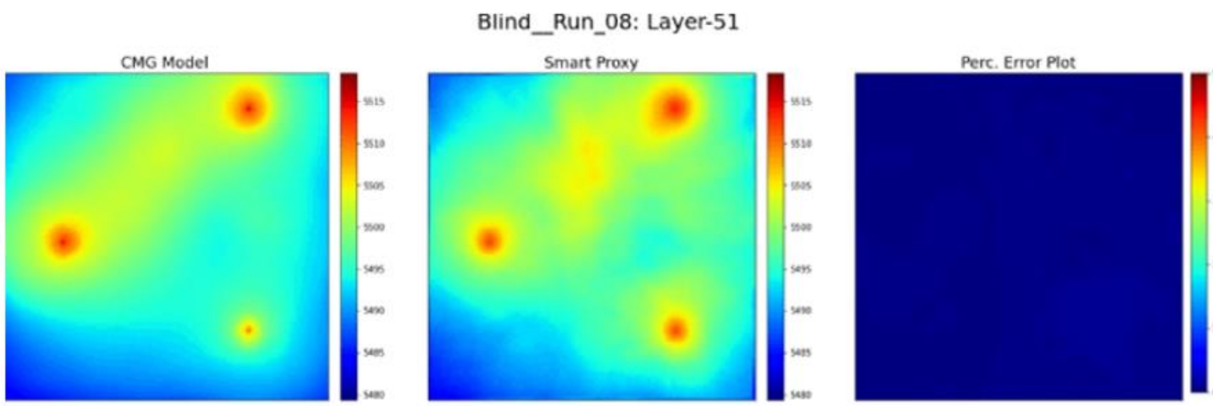


Figure 95. Pressure results for layer #51 for one of the blind models used in Phase 5

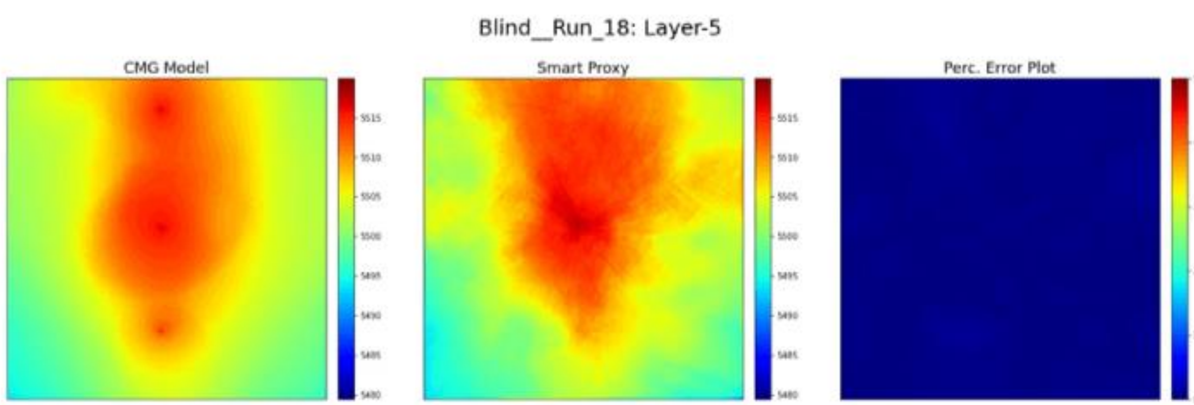


Figure 96. Pressure results for layer #5 for another blind models used in Phase 5

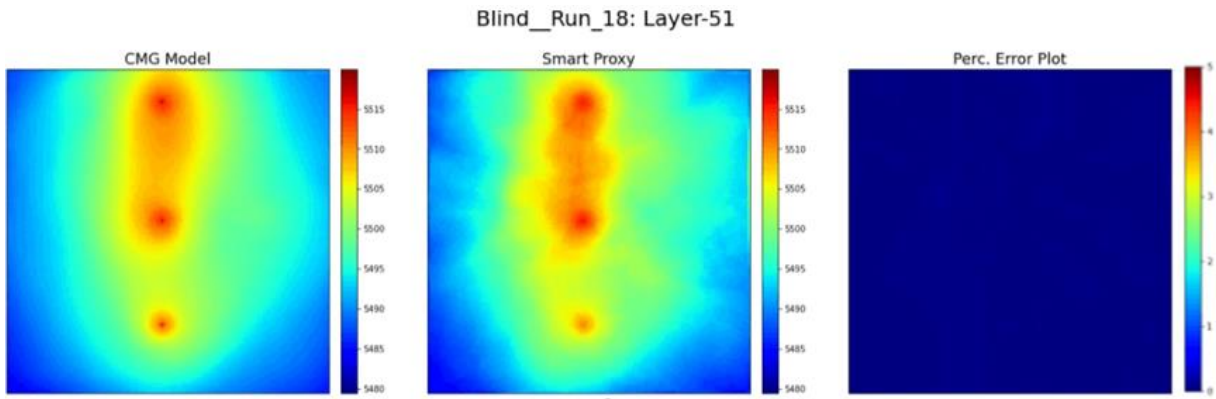


Figure 97. Pressure results for layer #51 for another blind models used in Phase 5

5.6.2 Results of Reservoir Saturation Distribution for Blind Datasets:

With an error margin of less than 10%, the results are in constant alignment with CMG results over all the reservoir layers **(Fig. 98 through Fig 101).**

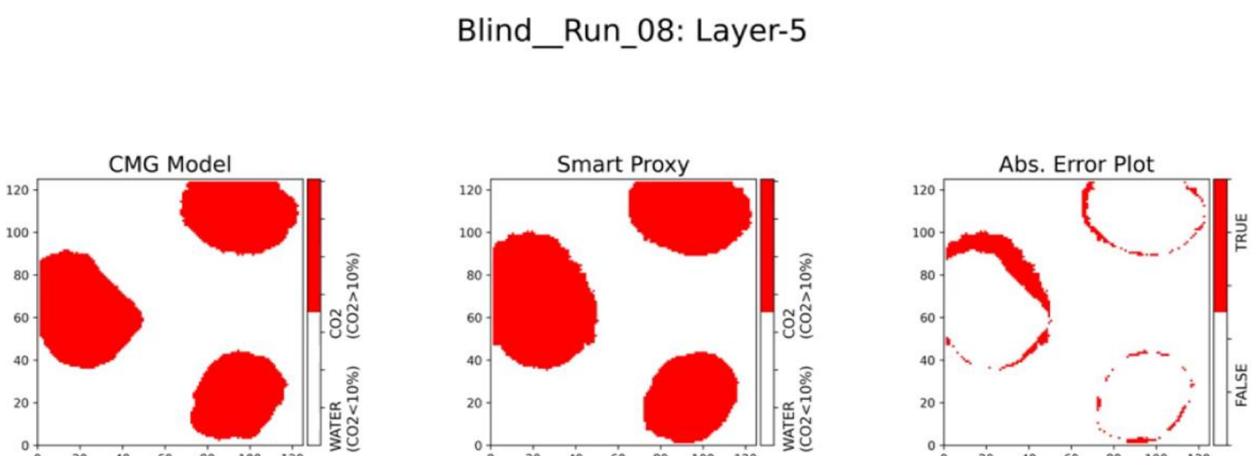


Figure 98. Binary visualization. Saturation results for layer #5 for one of the blind models used in Phase 5

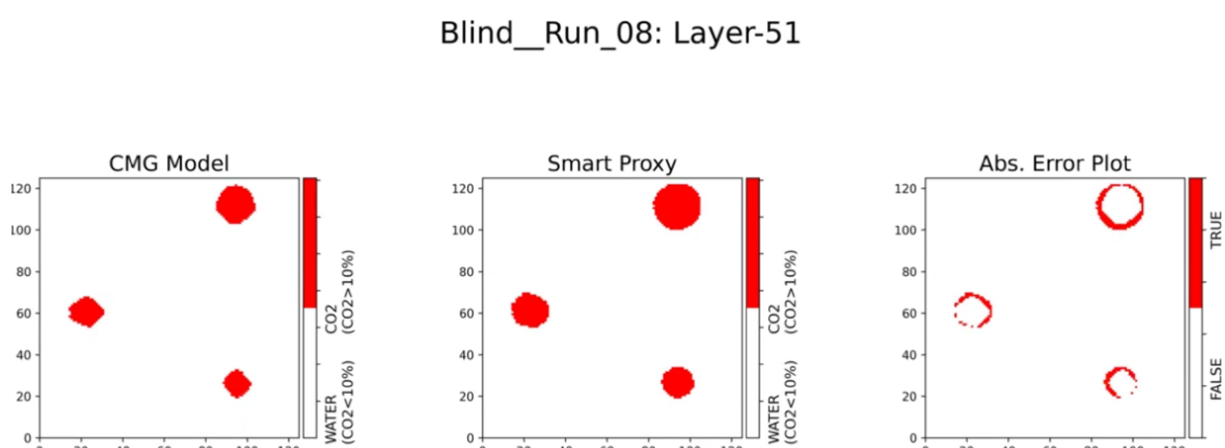


Figure 99. Discrete visualization. Saturation results for layer #51 for one of the blind models used in Phase 5

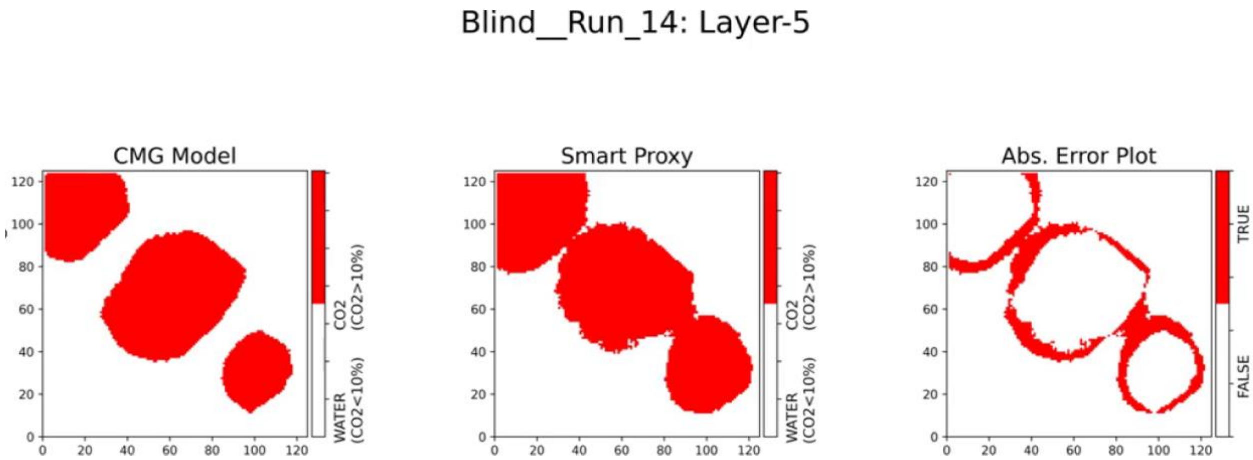


Figure 100. Binary visualization. Saturation results for layer #5 for another blind models used in Phase 5

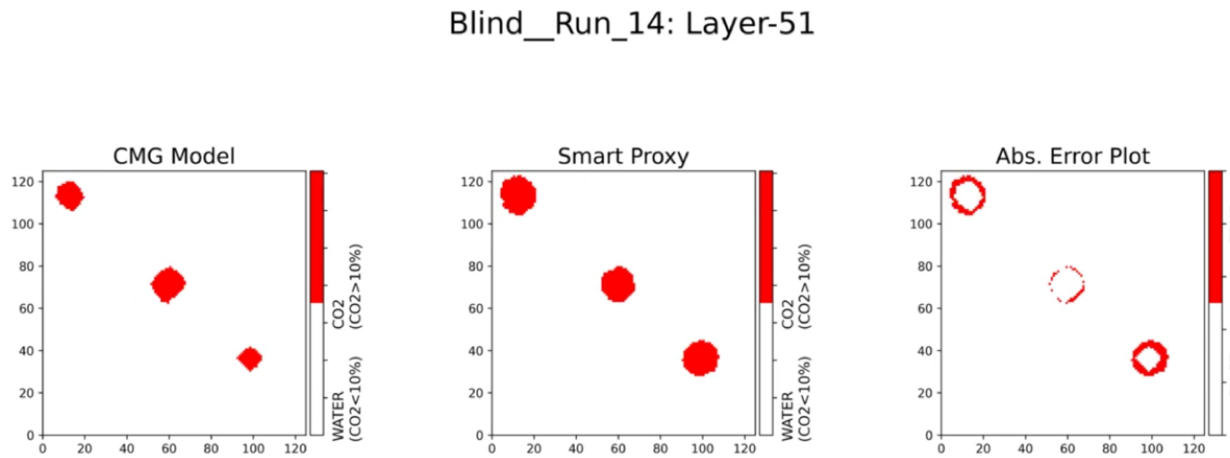


Figure 101. Discrete visualization. Saturation results for layer #51 for another blind models used in Phase 5

As evidenced by the reservoir pressure results from blind validation run #5, the developed Smart Proxy Model excels in faithfully reproducing the outcomes of numerical reservoir simulations with remarkable accuracy across various layers and at distinct time steps. This attests to the successful training of the Artificial Neural Network (ANN) algorithm, which has demonstrated its capacity to generalize pressure pattern predictions when confronted with entirely new simulation runs, such as blind runs. In the context of a single time step, specifically the concluding phase of the 30-year injection period, the initial time step involves the assignment of initial pressure values to every grid cell (approximately 800,000 relevant cells in each realization). The ANN-pressure model efficiently captures the pressure pattern by leveraging the consistent range and magnitude of pressure values across all grid cells. This effectiveness is enhanced as the neural network receives input not only from the initial conditions but also from the pressure values at preceding time steps (~800,000 data points in each realization).

It is important to mention that, as the study aimed to assess the spread of the CO_2 plume from the injection wells, a categorical visualization method was employed to distinguish areas with CO_2 presence from those unaffected by the plume. A 10% threshold was applied to delineate the two categories, and plots using the conventional visualization approach (without the 10% threshold) were also produced and appended for reference.

The varying performance of the smart proxy model across different reservoir layers, particularly in the upper layers and for specific blind datasets **(Fig. 101),** can be attributed to several factors related to feature engineering, data characteristics, and the intricacies of reservoir dynamics. Here's a detailed examination of these contributing factors:

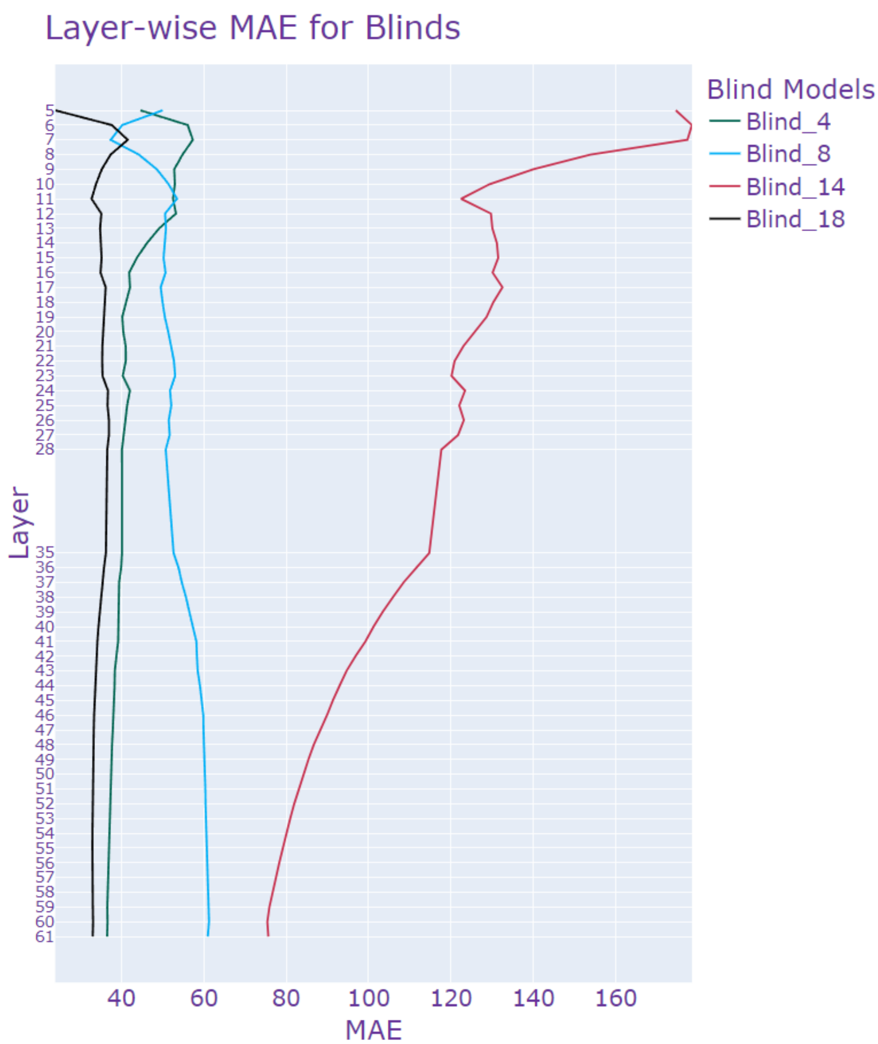


Figure 102. Behavior of the Smart Proxy Modeling based on the layer-wise error contribution in the model

5.7 Input Parameter Ranking:

When observing the weights assigned to the hidden layer based on input features, it becomes evident that certain input features exhibit greater variability in terms of parameter weights than others **(Figure 104).**

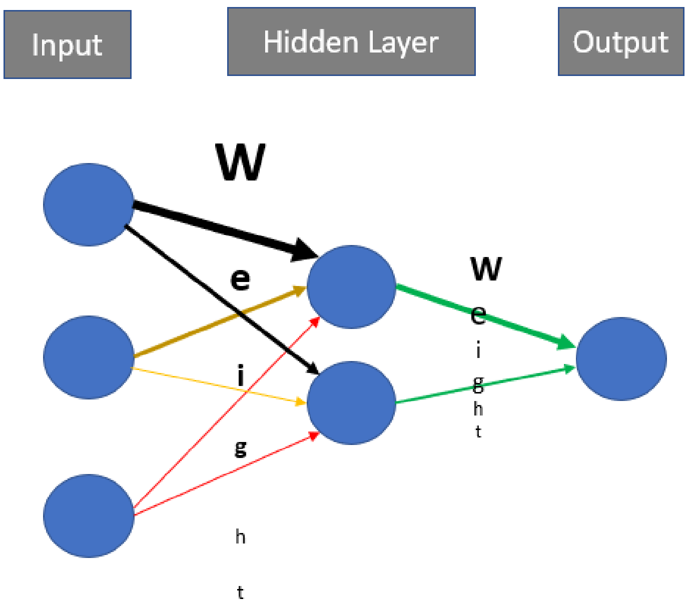


Figure 103. Input to Hidden and Hidden to Output Weights

Certain features in the study's Spatio-Temporal database, such as distance to injector and the quality of path to the injector well, significantly enhance the Smart Proxy Model's training and prediction process **(Fig. 104).** This is based on the weights assigned to each feature in the model. Features with larger absolute weights are considered more important.

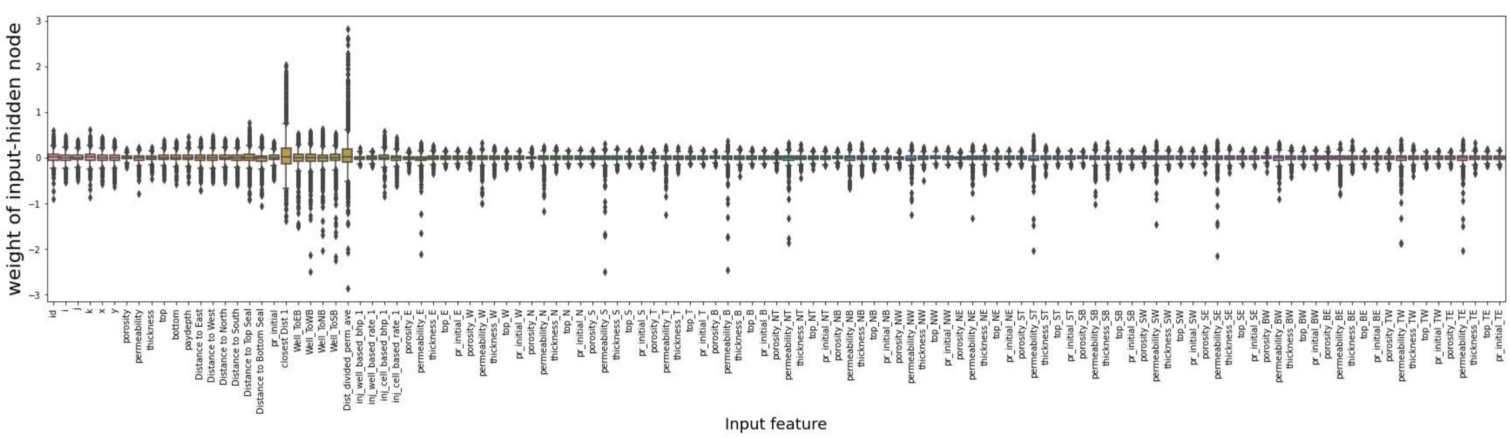


Figure 104. Input parameters ranking of the Smart Proxy Modeling based on their contribution during the training process

CHAPTER 6: SUMMARY, CONCLUSION AND FUTURE RESEARCH

6.1 Summary:

This study provided a Framework for Complexity and Uncertainty in CO_2 Sequestration Scenarios. Each phase targets specific objectives, incorporating varying levels of complexity and uncertainty.

Rotation Application (Phase 1): Objective: Apply rotation to each realization to introduce variability.

- Geological realization remains constant.
- Well locations and numbers are fixed.
- Focus on variability through rotation.

Well Configuration Variation (Phase 2): Objective: Vary the number and location of wells in different geological realizations.

- Different geological realizations based on the same 10x10 grid.
- Fixed well locations with variable well numbers (up to 2).
- Consideration of delayed injection and inter-well interpolation.

3. Well Location Adjustment (Phase 3): Objective: Change well locations within the same geological realization to introduce uncertainty.

- Single well involved.
- No delayed injection or inter-well interpolation.

4. Well Modification (Phase 4): Objective: Modify wells in different geological realizations to assess the impact.

- Different geological realizations with fixed well locations.
- Up to 3 wells involved.
- Consideration of delayed injection without inter-well interpolation.

5. Well Adjustment (Phase 5): Objective: Adjust well locations within the same geological realization.

- Geological realization remains constant.
- Different well locations (three wells involved).
- Consideration of delayed injection without inter-well interpolation.

These phased approaches systematically introduce varying levels of complexity and uncertainty into the design of CO_2 sequestration scenarios. From maintaining geological consistency to altering well configurations, each phase contributes distinctive elements. Smart proxy modeling, guided by the specifics of each phase, adeptly predicts outcomes, providing nuanced insights into the interplay between geological factors and operational variations in CO_2 sequestration projects. This systematic framework ensures a comprehensive exploration of scenarios for improved predictive modeling. In any CO_2 sequestration initiative, it is imperative to monitor both the extent of the CO_2 plume and the pressure distribution throughout the reservoir at various locations. Throughout the injection process, the Smart Proxy exhibited spatial and temporal variations in the generated CO_2 saturation or plume. Notably, while the pressure smart proxy model could accurately replicate the outcomes of the numerical reservoir simulation, the CO_2 Smart Proxy encountered challenges in replicating CO_2 levels during early time steps, primarily due to a dearth of data on CO_2, especially originating from the initial timestep. As a consequence, the Smart

Proxy's accuracy in predicting the edges of the plume was slightly compromised. This, coupled with the constraint of a limited number of training records available to train the neural network, led to formation of such errors, the extent of the error was reduced for the lower aquifer layers.

Smart Proxy Models play a pivotal role in expediting the generation of pressure and saturation results in CO_2 sequestration endeavors, thereby significantly reducing computational demands while upholding model accuracy. This efficiency proves particularly beneficial in handling the complexities and uncertainties inherent in reservoir characteristics, well pattern and location optimization, and data quality. Notably, the time required for such tasks is markedly diminished compared to traditional numerical reservoir simulations.

In the scope of this study, five distinct levels of complexities were addressed, with each phase contributing uniquely to the holistic understanding of CO_2 sequestration and operational uncertainties. The application of Smart Proxy Models extends beyond mere computational expediency; it proves instrumental in conducting uncertainty analyses, thereby facilitating the scalable development and deployment of CO_2 sequestration projects.

Optimal storytelling is a fundamental component of impartial data research. Employing suitable visualizations allows for the identification of patterns and anomalies in the data, facilitating precise conclusions and interpretations. The study highlights the significance of certain features within the Spatio-Temporal database, such as the distance to the injector and the quality of the path to the injector well. These features significantly enhance the training and prediction processes of the Smart Proxy Model.

In essence, the research offers a practical workflow incorporating Data-Driven and Machine Learning techniques tailored for reservoir engineering and management in the context of CO_2 sequestration projects. By systematically addressing subsurface problems, this workflow contributes to the advancement of efficient and informed decision-making in CO_2 sequestration endeavors.

6.2 Conclusion:

The main conclusions drawn from this study can be summarized as below:

- Smart Proxy Models expedite the generation of pressure and saturation results in CO_2 sequestration, reducing computational requirements while maintaining model accuracy. For the pressure model, the results were great and well below an average error of less than 5%. For the saturation, the model performed well and the results on average were below 10%. Most of the error in the saturation models was concentrated on the edges of the plume in which training data may have not provided enough information to the ANNs about CO2 saturation data.
- By training the models with a range of input data that covers the uncertainty in both space and time, Smart Proxy can be applied to uncertainty analysis, facilitating the scaling up of CO_2 sequestration development and deployment projects. Complexities and uncertainties in reservoir characteristics, well patterns and location optimization, and data quality can be handled and managed in fraction of time compared to numerical reservoir simulation. Once trained, the proxy models provide faster predictions and analysis compared to running the full

numerical simulations. This efficiency is particularly beneficial for scenarios that require quick decision-making or repeated simulations and more extensive exploration of different reservoir scenarios, sensitivity analysis, and optimization studies within a reasonable time frame.

- Data-driven smart proxy models have the ability to capture and learn complex relationships and patterns from the numerical simulation data. While the training data comes from the numerical simulations, the proxy models can generalize and make predictions for reservoir scenarios not explicitly included in the training dataset. This generalization capability allows for the exploration of different reservoir conditions and production strategies beyond what has been simulated.
- The efficient replication of fluid flow behavior through machine learning models results in substantial computational savings compared to traditional numerical simulation models. By identifying and emphasizing critical factors, the research minimizes computational costs while maximizing the value of insights. This approach streamlines decision-making processes by focusing resources on the most impactful aspects of CO_2 plume monitoring. This not only accelerates research processes but also makes these models more accessible and practical for real-world applications.
- Specific attributes created within the Spatio-Temporal database in this research, such as the quality of the path between injection wells and focal cells, demonstrated a more substantial impact on improving the training of the Smart Proxy Model compared to other attributes. Recognizing that the generation of these attributes requires expertise in reservoir engineering and an understanding of the data's environmental context, it becomes evident that incorporating domain knowledge can greatly enhance the precision of Data-Driven Models related to subsurface applications.
- Augmenting the dataset with variations in spatial complexity, spatio-temporal dynamics, and reservoir heterogeneity allows the model to learn a more comprehensive set of features. This enhances the model's ability to generalize to diverse conditions, making it more robust and applicable to real-world scenarios. This study addressed 5 different levels of complexities. Each phase contributes uniquely to the overall understanding of CO_2 sequestration and operational uncertainties.
- Data-driven smart proxy models offer flexibility in handling different reservoir scenarios and operational conditions. Once trained, the models can be easily adapted to new data or variations in reservoir properties. This adaptability allows for the exploration of multiple what-if scenarios and the evaluation of alternative production strategies without the need to rerun time-consuming numerical simulations.
- Data-driven smart proxy models can complement numerical reservoir simulations by offering a different perspective and additional insights. The proxy models may capture reservoir behaviors and dynamics that are not explicitly captured in the numerical simulations. Their ability to identify non-intuitive patterns or correlations in the training data can provide valuable insights that might have been overlooked by traditional reservoir engineering approaches. This study presents a systematic and hands-on workflow that can be readily employed by reservoir engineering and management teams. It offers a toolbox of advanced Data-Driven and Machine Learning techniques, enabling informed decision-making for various subsurface challenges. This is particularly valuable for addressing issues related to CO_2 sequestration projects, which are anticipated to gain increased prominence in the coming years.

- Among the five phases, phase 5 is more practical, but also more challenging. The complexity factors associated with this phase included:

  1. **Geological Consistency with Well Adjustment:** Phase 5 maintains geological consistency while adjusting well locations. This introduces complexities as the geological framework remains constant, requiring the smart proxy model to discern subtle changes attributed solely to well adjustments.
  2. **Multiple Well Involvement:** In Phase 5, three wells are involved, increasing the spatial interaction and potential interference between wells. This complexity challenges the smart proxy model to accurately capture the dynamic behavior resulting from multiple well adjustments within the same geological realization.

  The challenges that the ANN faced during the training could be categorized as:

  1. **Sensitivity to Well Adjustment:** Smart proxy models may struggle to precisely capture the impact of well adjustments on CO_2 saturation within a consistent geological setting. The subtle changes introduced by adjusting well locations demand a high sensitivity that may be challenging for the model.
  2. **Interactions Among Multiple Wells:** The involvement of three wells adds complexity due to potential interference and interactions. The smart proxy model needs to discern the nuanced effects of each well adjustment and their combined influence on CO_2 saturation.
  3. **Dependency on Training Data:** Smart proxy models heavily rely on training data, and the limited geological variation in Phase 5 may result in challenges. The model may struggle to generalize well adjustments' effects, especially if the training data lacks diversity in similar scenarios.
  4. It must be noted that data availability and resolution at the edges of the CO_2 plume are often limited. Sparse data points and lower resolution in these areas can lead to challenges in accurately capturing the intricacies of fluid behavior, resulting in discrepancies between modeled and observed values.

- In the realm of advancing carbon sequestration through Smart Proxy Modeling, it becomes evident that as the complexity of the systems rises, traditional modeling approaches lose their precision, and precision alone loses its meaningfulness. This is particularly true in the context of numerical reservoir simulation, where the intricate interplay of high-dimensional nonlinearity, heterogeneity, and complex physical processes challenges conventional methods. As this research journey unfolded, we find ourselves navigating through a landscape characterized by the scarcity of real-world field comprehensive data and the necessity for extensive feature generation in constructing effective proxy models incorporating both quantitative insights and domain expertise. In this intricate puzzle, the quote by **Lotfi A. Zadeh** (1921-2017) resonates profoundly: **'As complexity rises, precise statements lose meaning and meaningful statements lose precision.'** It underscores the fundamental shift towards embracing the complexities inherent in modeling deep saline aquifer CCS programs, driving us to explore innovative data augmentation techniques. By expanding our dataset through the creation of diverse and realistic CO_2 sequestration scenarios, we bridge the gap between precision and meaning, harnessing the power of Smart Proxy Modeling and machine learning to revolutionize our approach to reservoir simulation for efficient carbon sequestration.

6.3 Future Work Recommendation:

This study provides the following recommendations:

1. Explore the effectiveness of noise injection techniques to enhance model robustness. Introduce controlled noise into training data or model parameters to simulate real-world uncertainties. Evaluate how noise injection mitigates overfitting and improves the model's ability to handle variations in the data. Based on the level of performance of the model, the specific problem, objective, the amount of data, and number of geological realizations, adjust the level of complexity in the model.
2. Systematically benchmark SPMs under various overfitting scenarios. This involves deliberately introducing scenarios with limited data or high noise levels to assess how well the models generalize. Analyze the impact of overfitting on different reservoir types and operational conditions.
3. Fine-Tuning for Well Interactions: Implementing more fine-tuning strategies that specifically address the interactions among multiple wells. This can involve refining the model's sensitivity to subtle variations and optimizing its ability to predict CO_2 saturation under complex well configurations.
4. Advanced Feature Engineering Based on Domain Knowledge: Explore more sophisticated feature engineering techniques tailored to reservoir characteristics. Identify key parameters that significantly influence reservoir behavior and develop innovative ways to represent them in the model. This could involve leveraging domain-specific knowledge to enhance feature selection and extraction.
5. Refinement of Model Complexity: Investigate methods to refine the complexity of SPMs in alignment with available data. This involves assessing whether increasing model complexity enhances predictive accuracy or if there's an optimal point beyond which additional complexity doesn't yield substantial benefits. Consider incorporating advanced machine learning techniques or hybrid models to strike the right balance.
6. Cross-Validation Strategies: Develop and employ sophisticated cross-validation strategies tailored to SPMs. Explore techniques such as k-fold cross-validation or leave-one-out cross-validation to assess the model's performance across various data subsets. This can provide more reliable estimates of model generalization.
7. Hyperparameter Tuning: Systematically tune hyperparameters of SPMs to find optimal configurations that balance model complexity and generalization. Use techniques like grid search or randomized search to explore the hyperparameter space and identify settings that mitigate overfitting while optimizing predictive accuracy.
8. Training with more Noisy Data: ANNs can be trained using data that includes noise and variations representative of the uncertainties in the real-world input conditions. By exposing the model to diverse and noisy data during training, it learns to generalize better and accommodate uncertainties in unseen input conditions.
9. Sensitivity Analysis: Sensitivity analysis involves evaluating how changes in input variables impact the model's output. ANNs can be subjected to sensitivity analyses to understand which input features contribute more to uncertainty. This knowledge can guide the model in assigning appropriate weights to different inputs.
10. Transfer Learning: Transfer learning involves training an ANN on a related task before fine-tuning it for the specific CO_2 sequestration prediction task. This approach can leverage

knowledge gained from handling uncertainties in related domains and adapt the model to the specific uncertainties in geological or initial condition data.

11. The selection of the most suitable Machine Learning (ML) model poses a challenge, requiring a deep understanding of the specific reservoir engineering problem at hand. The complexity and dataset characteristics of each problem must be considered from different modeling aspects. Despite the success demonstrated in accelerating reservoir simulation and improving history matching results, the adoption of ML methods in commercial products within the oil and gas industry remains limited. The industry's hesitancy is not attributed to the effectiveness of these methods but rather to challenges related to synthetic training data generation. Generalizing synthetic data generation, especially in enterprise environmental conditions, poses a non-trivial task requiring collaboration between domain expert engineers and ML experts.
12. The application of ML in subsurface reservoir simulations represents a transformative shift, offering a powerful tool to understand and predict complex reservoir behaviors. ML's ability to handle multi-dimensional data enhances predictive accuracy, uncovering complex patterns and relationships that traditional methodologies may overlook. Continuous learning and improvement, coupled with challenges such as data quality and model interpretability, underscore the dynamic landscape of ML applications in reservoir engineering. While the oil and gas industry's adoption may be gradual, ongoing efforts to automate processes under enterprise conditions are crucial for realizing the full potential of ML.
13. In this work, we have used numpy.memmap files to handle the large datasets. It is recommended to consider using efficient data storage formats, such as HDF5, which provide compression and efficient data access capabilities. These formats are designed for handling large datasets and can speed up the data loading process. Considerations for neural network training include prioritizing memory efficiency, especially when iterating over the entire dataset in mini-batches to avoid exhausting system memory. For datasets with a straightforward structure where loading chunks of data suffices, numpy.memmap could be sufficient. However, for more complex data structures, HDF5 is a good choice. In distributed or parallel computing environments, HDF5 and Parquet formats may provide advantages due to their support for parallel processing. Additionally, if storage space is a concern, formats like HDF5, with their compression capabilities, can be beneficial.
14. The effectiveness of developed models is dependent on the quality and relevance of the input features. While the introduction of additional features enhanced the model's overall performance, the challenge lies in identifying features that accurately capture the nuances of CO_2 behavior in each reservoir layer. The upper layers, due to their complex geological structure and heterogeneities, may require specific features that were not adequately addressed enough.
15. In the course of our current investigation, I have made noteworthy strides in understanding the three-dimensional aspects of fluid flow by exploring neighboring grids in multiple directions, including the tier 1, 2, and 3 systems. While our study has shed light on these components, there exists a valuable opportunity to delve deeper into features that more comprehensively capture the intricate 3D nature of the fluid flow, particularly emphasizing the characteristics of the formation layers and their anisotropy. To enhance the richness of our analysis and provide a more nuanced perspective, future research should prioritize the examination of features that distinctly exhibit 3D characteristics or more directional special

inputs. This involves a meticulous exploration of the interplay between different layers, considering their anisotropic behavior. By focusing on such features, we could gain a more holistic understanding of the fluid dynamics within the formation.

16. Spatial Variability: The geological properties of reservoirs exhibit significant spatial variability across different layers. Even with the grid cell-based approach employed in this study, the inherent heterogeneity in the upper layers may necessitate more refined or localized features that capture the unique geological characteristics specific to those layers. Conventional feature engineering techniques may not be sufficient to fully represent these fine-scale variations.
17. Sensitivity to Input Resolution: The smart proxy model's sensitivity to input resolution could differ across layers. While it may effectively replicate results in the lower layers, the upper layers might require a higher resolution or more granular representation to account for localized variations in geological properties and CO_2 behavior. This sensitivity to resolution could be attributed to the more complex and heterogeneous nature of the upper layers.
18. Training Data Limitations: The distribution and quality of the training data can significantly impact the model's generalization capabilities. If the training data for the upper layers is imbalanced or lacks diversity in representing upper layer scenarios, the model may struggle to generalize effectively, especially when faced with blind datasets that exhibit different data characteristics. This could explain the observed discrepancies in the upper layers.
19. Overfitting Concerns: Overfitting, where the model learns the training data too well and fails to generalize to unseen data, can also contribute to performance discrepancies. If the developed ANN adapted too closely to the training data characteristics, it could overfit and perform poorly on unseen data, particularly in the upper layers where data characteristics may differ. Addressing overfitting through careful selection of training data is essential.
20. As a future work recommendation, it is essential to expand the scope of this study to encompass multi-time step dynamic prediction and modeling for all time steps under the same/ different or more advanced data augmentation and complexity schemes, both in cascading and non-cascading modes. While the current focus is on a specific time step, precisely 30 years after the commencement of injection on January 1, 2050, the reservoir behavior and performance over the entire injection and post-injection period merit comprehensive investigation. This expansion would involve developing predictive models and simulations that account for the evolving dynamics and responses of the reservoir at various time intervals, ensuring a thorough understanding of long-term behavior, potential challenges, and optimization opportunities. By considering all time steps within the injection and post-injection scheme, we could make better use of the temporal features and gain valuable insights into the reservoir's behavior over extended periods, aiding in more robust decision-making and sustainable management practices.
21. In addition to the findings and future work recommendations, it is essential to emphasize the significance of adhering to best practices to ensure fair and unbiased data science assessments. These best practices encompass a range of critical elements, including:
    1. **Vigilance Against Biases:** Remaining vigilant regarding various biases, such as statistical biases considering their potential impact at each stage of the research.
    2. **Bias Audits:** Conducting routine bias audits on both datasets and algorithms to identify and mitigate potential sources of bias.

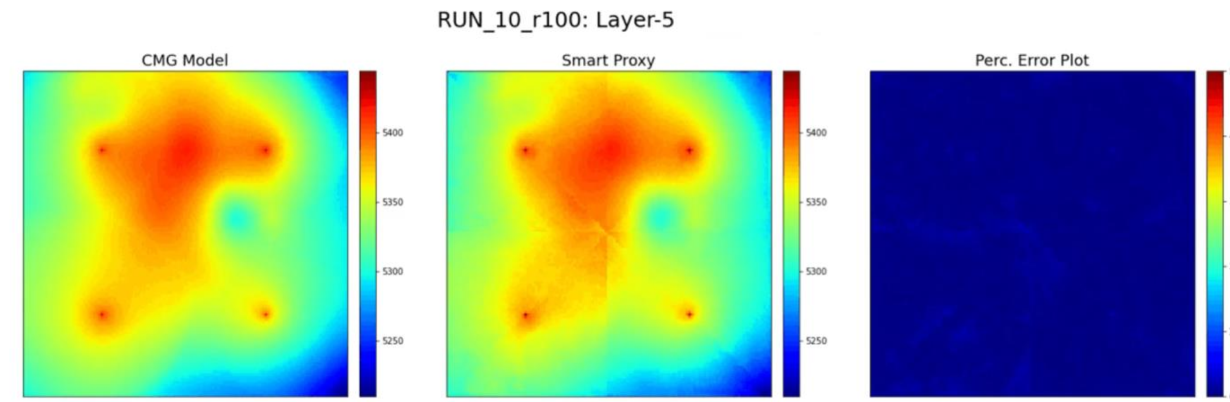
3. **Metrics Assessment:** Prior to incorporating fairness and bias metrics into the overall assessment process, conducting a thorough assessment of these metrics to ensure their relevance and appropriateness in alignment with the problem and application.

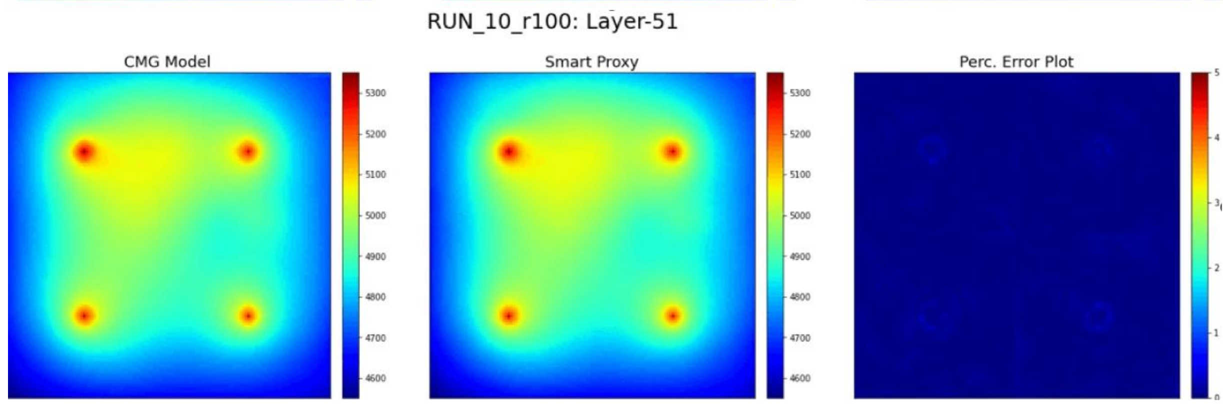
In AI engineering, the data's quality and quantity used for developing models significantly influence the effectiveness of AI-based models. The approach of employing AI and Machine Learning in engineering involves leveraging actual measurements and real physics data to create models of physical phenomena, rather than solely relying on mathematical equations. Historically, the modeling of physics has depended on the engineers' and scientists' comprehension of the phenomena at hand. As this understanding evolves, so does the nature of the mathematical equations used in modeling. The data's quality and quantity in constructing AI models play a crucial role in determining the presence of biases within these models. The aim of AI Ethics is to assess this data to pinpoint potential biases, whether intentional or not, in the models. Traditional engineering biases, including major assumptions and simplifications, can seep into AI applications. This often occurs when data derived from mathematical equations is merged with actual field data, and this amalgamated data is used in AI and Machine Learning algorithms to model physics, leading to what are known as "Hybrid Models." In the realm of AI and Machine Learning in engineering, such models can indicate a lack of thorough and scientific comprehension of these technologies. There is a growing concern that some entities using AI in engineering might be incorporating significant human biases, especially when they cannot develop an unbiased AI model. These human biases in engineering are largely connected to the formulation of mathematical equations for resolving physics-based challenges. This perspective suggests that the integration of biases in AI engineering applications is often linked to insufficient scientific understanding of how AI should be applied to model physical phenomena (Mohaghegh 2024).

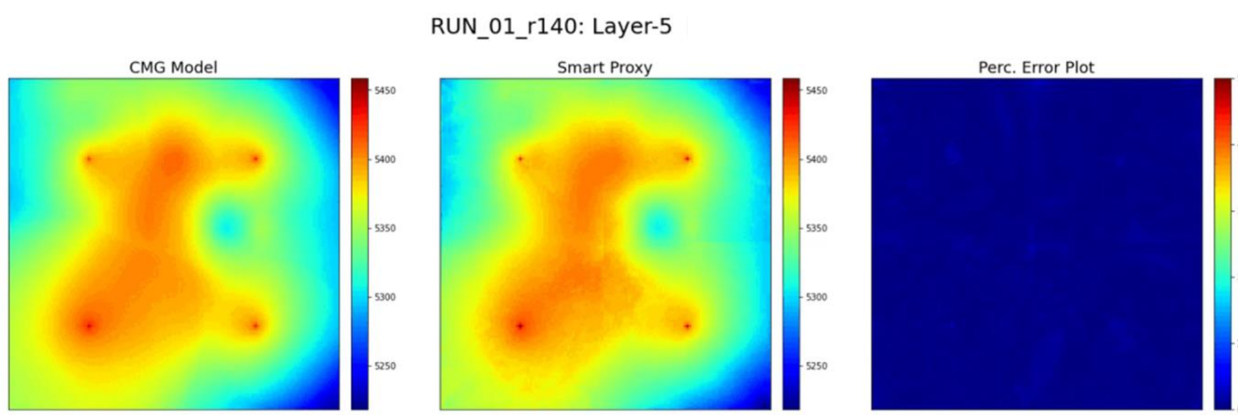
CHAPTER 7: APPENDIX

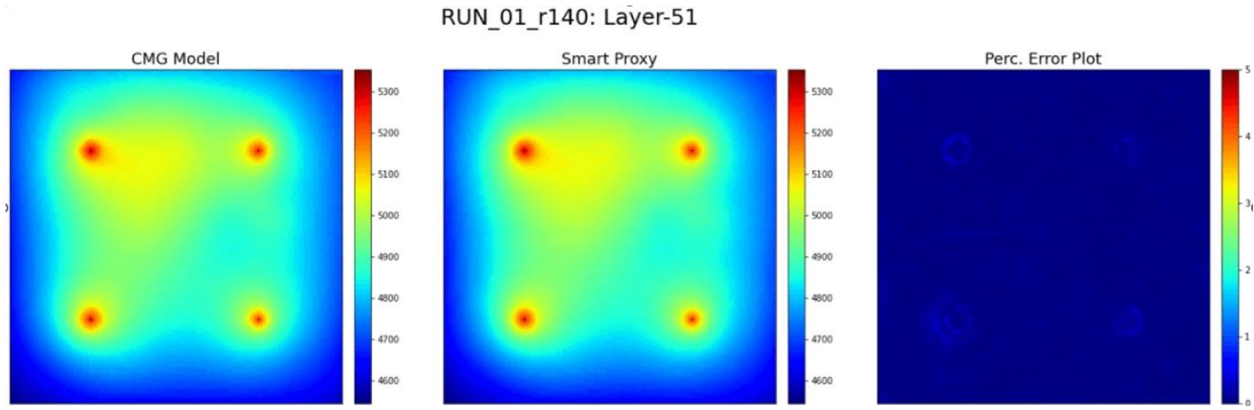
Pressure Results:

7.1 Phase 1 Pressure Trains:

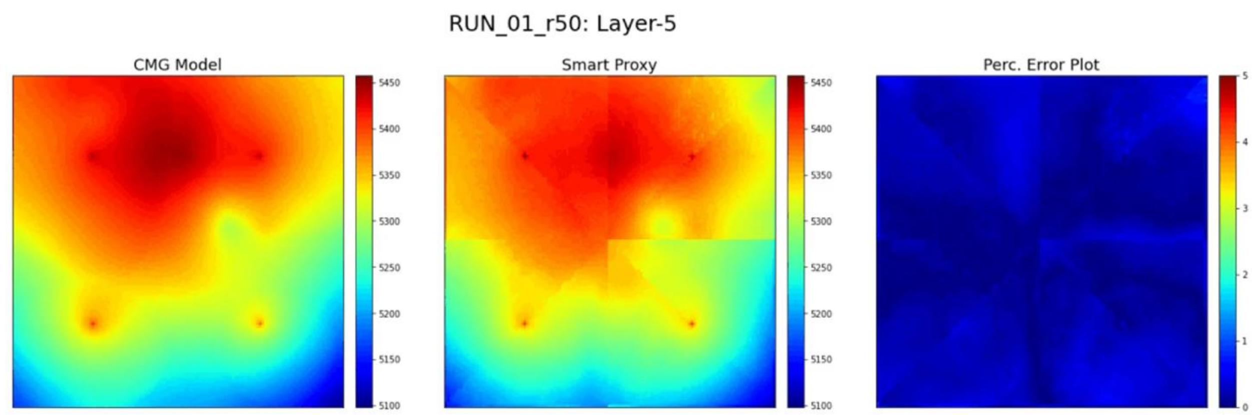


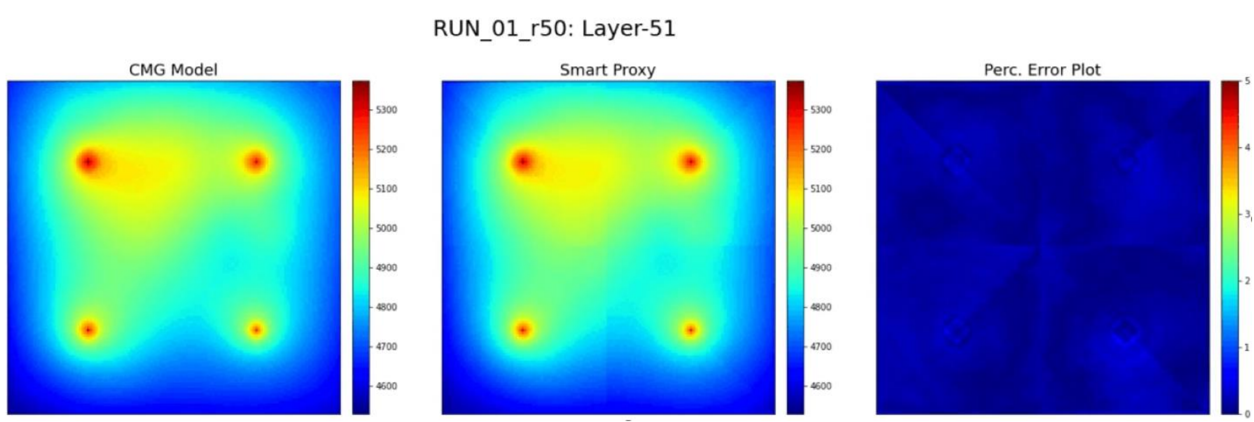


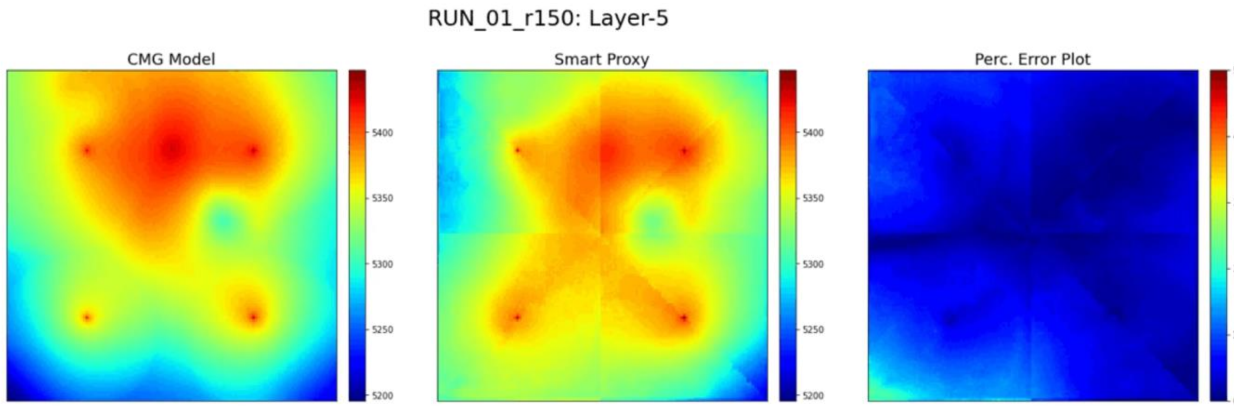


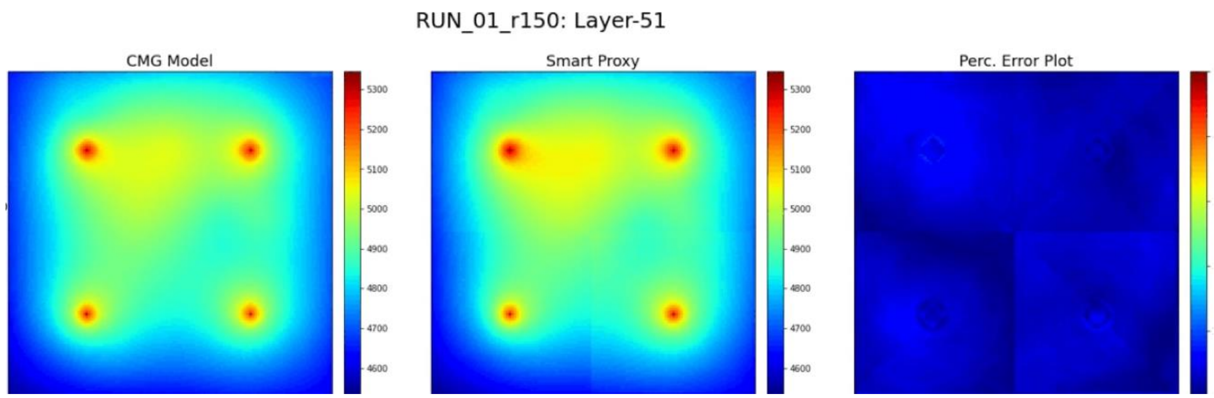


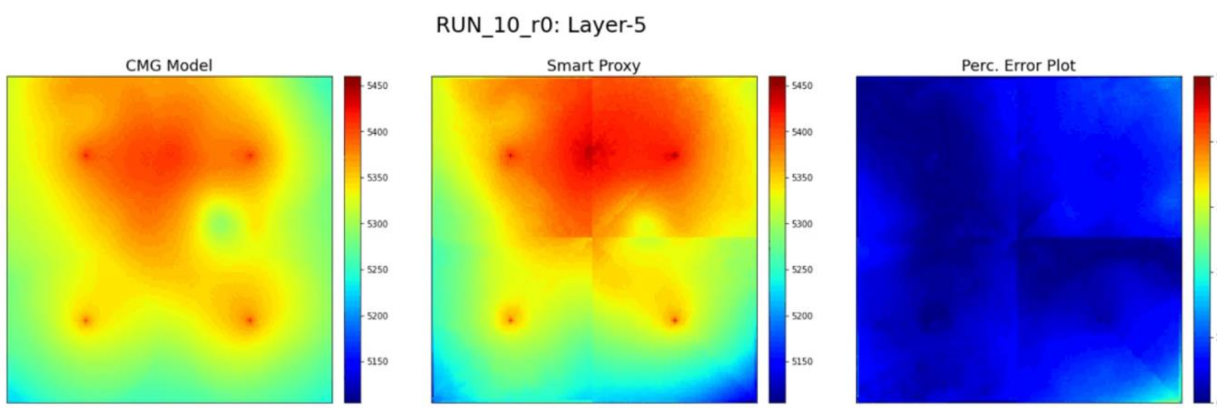
7.2 Other Phase 1 Pressure Blinds:

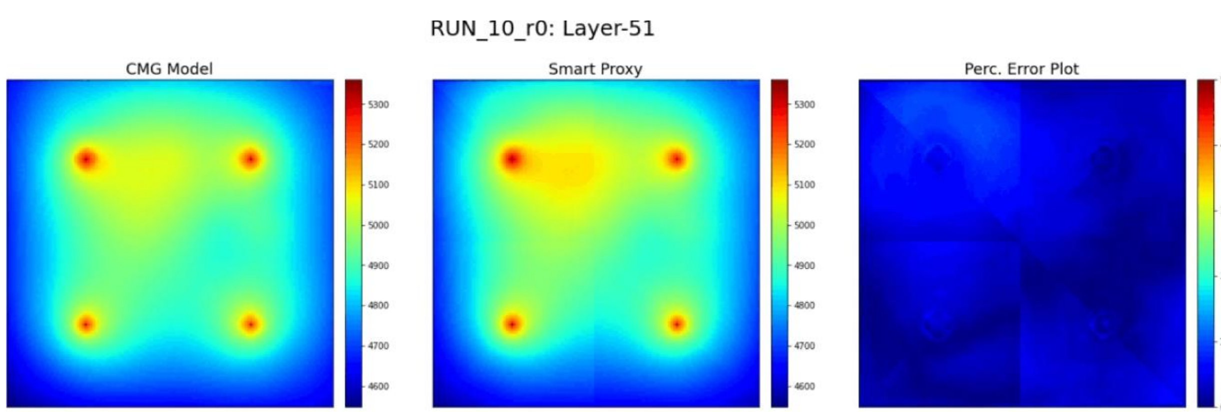


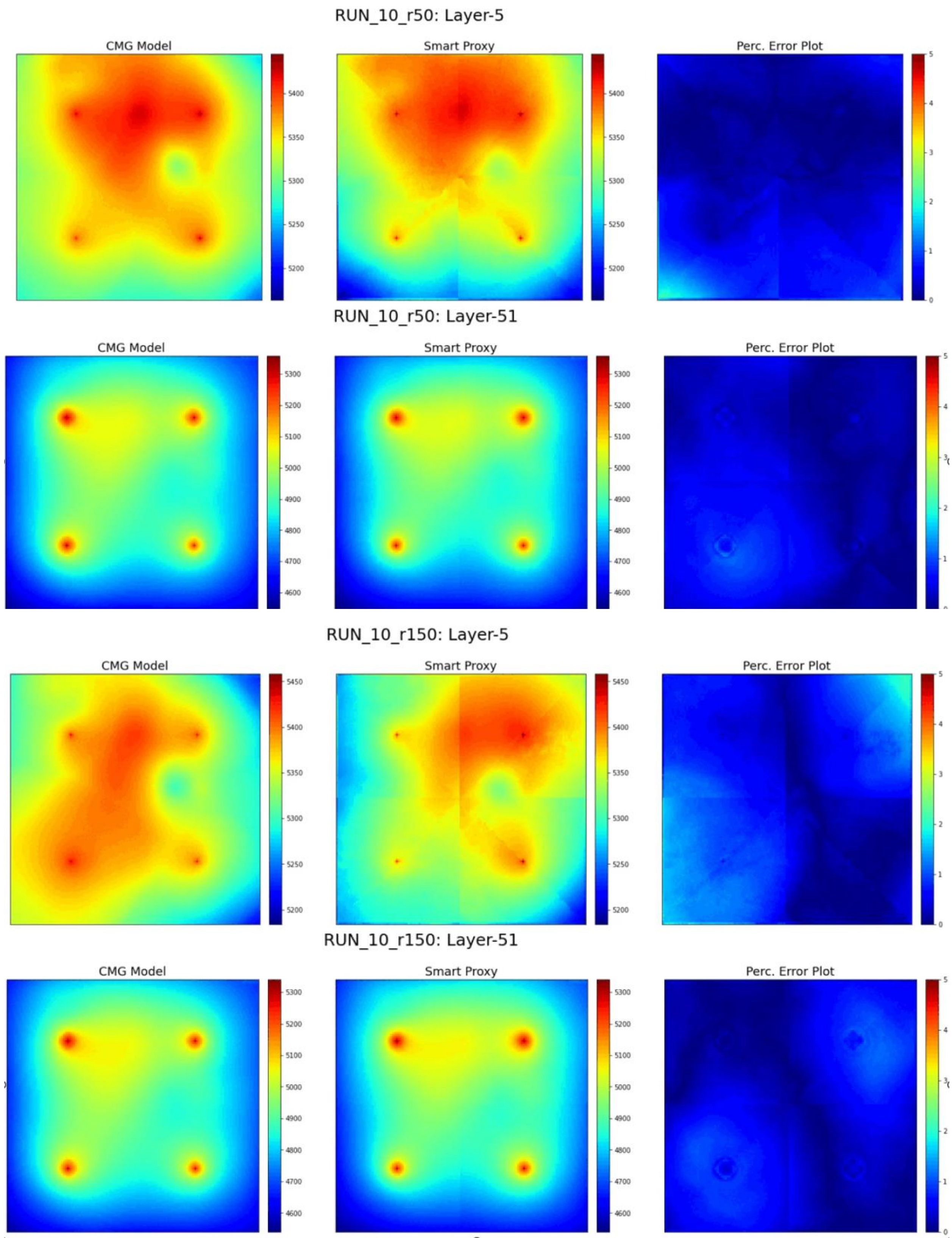


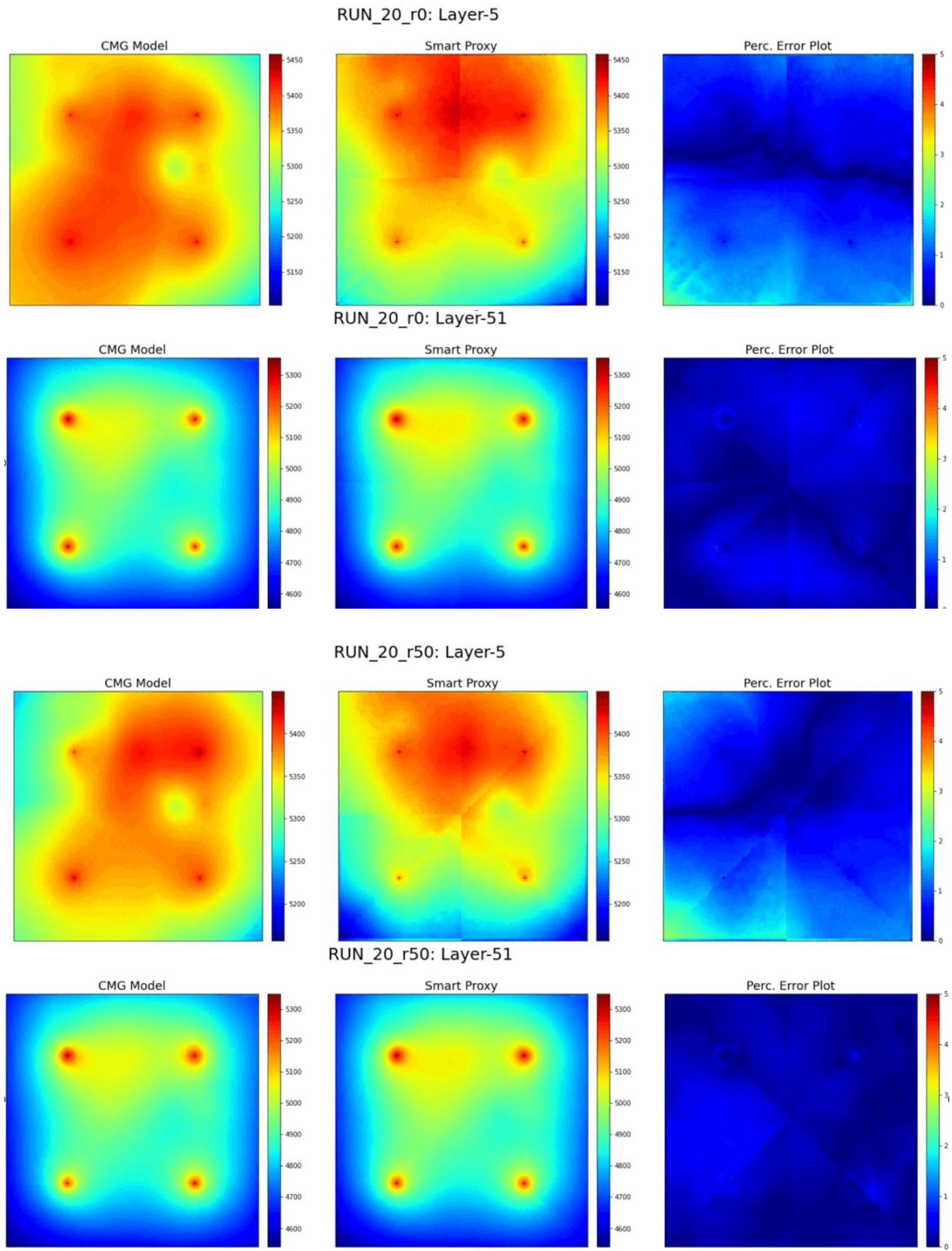

RUN_01_r150: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot

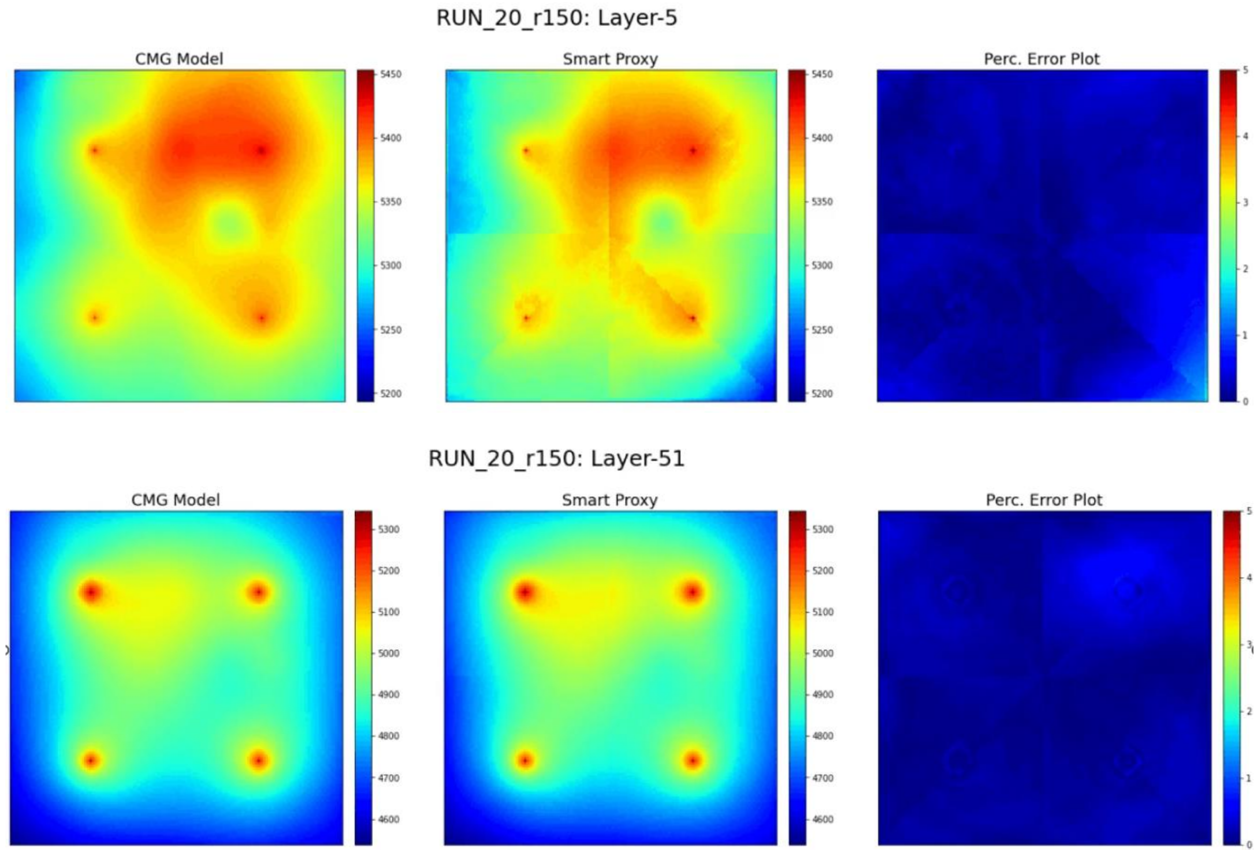

RUN_01_r150: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot


RUN_10_r0: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot

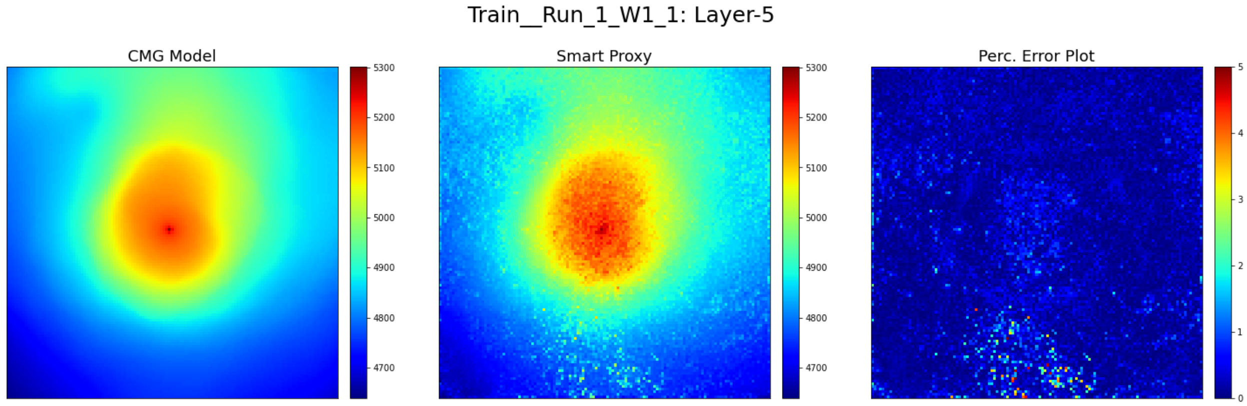

RUN_10_r0: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot

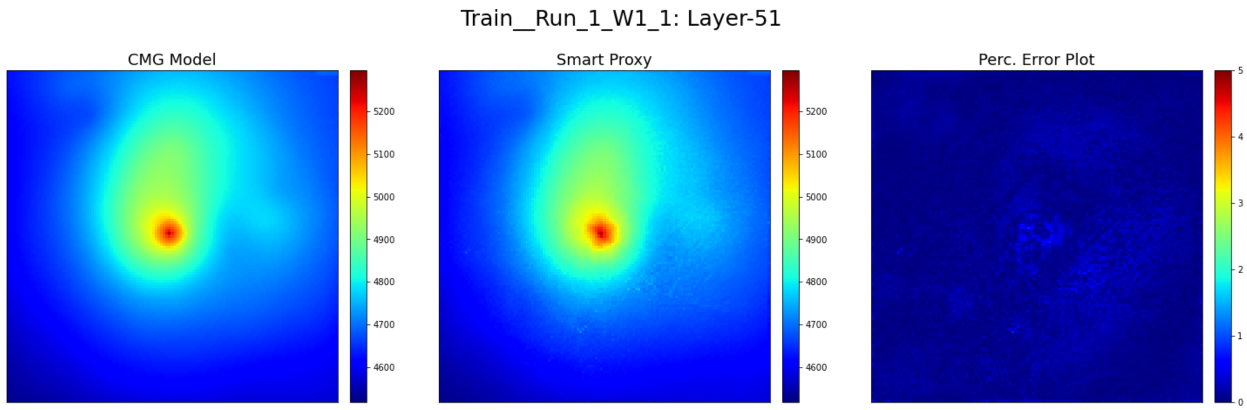

RUN_10_r50: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot
RUN_10_r50: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot
RUN_10_r150: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot
RUN_10_r150: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot


RUN_20_r0: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot
RUN_20_r0: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot
RUN_20_r50: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot
RUN_20_r50: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot

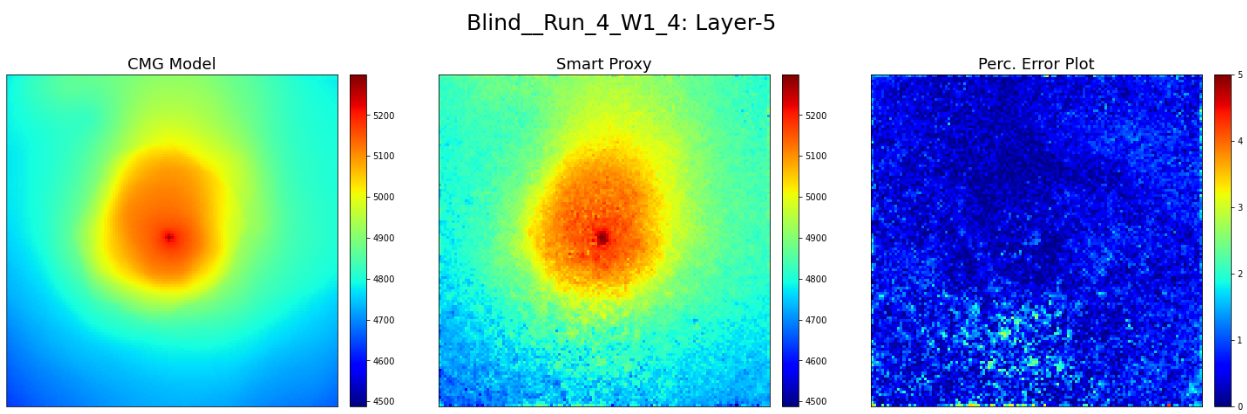

RUN_20_r150: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot
RUN_20_r150: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot

7.3 Phase 2 Pressure Trains:

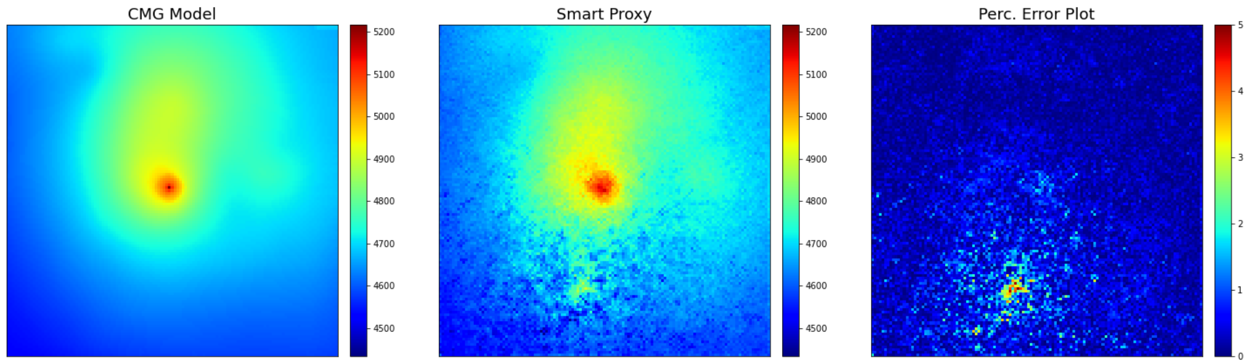




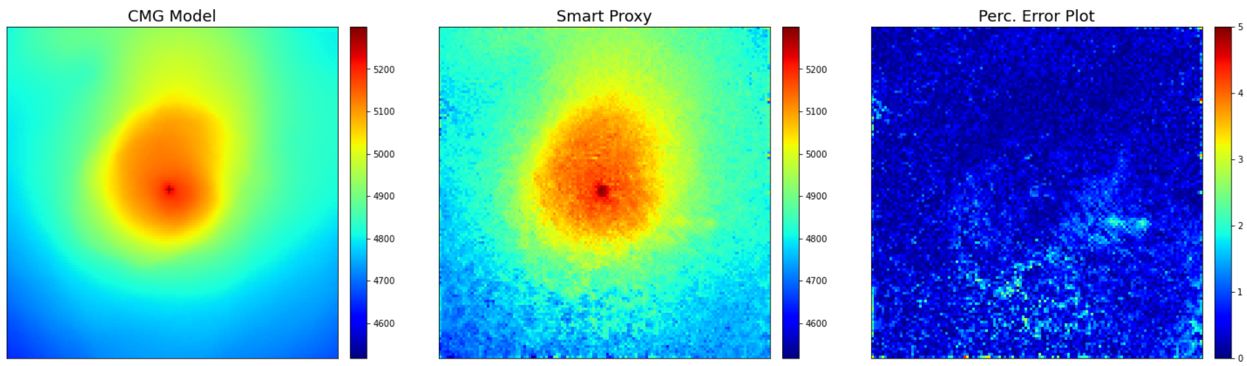
7.4 Other Phase 2 Pressure Blinds:



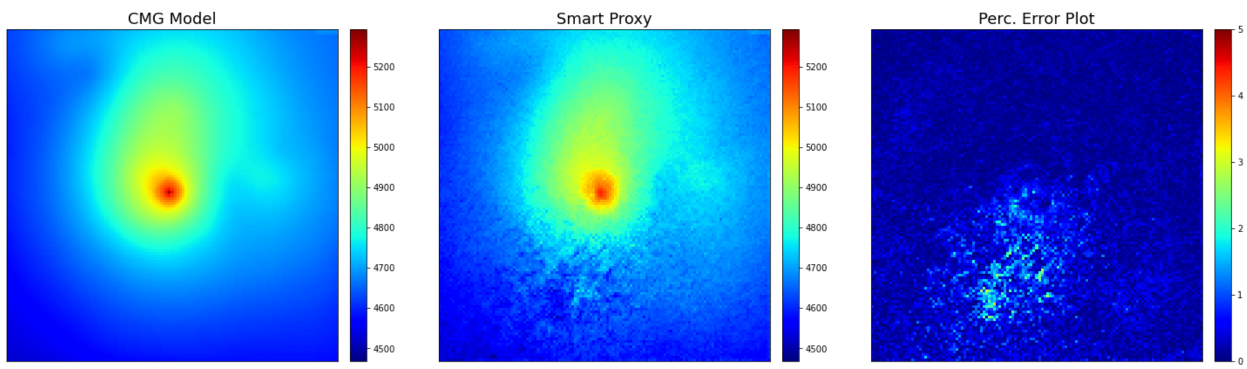
Blind__Run_4_W1_4: Layer-51



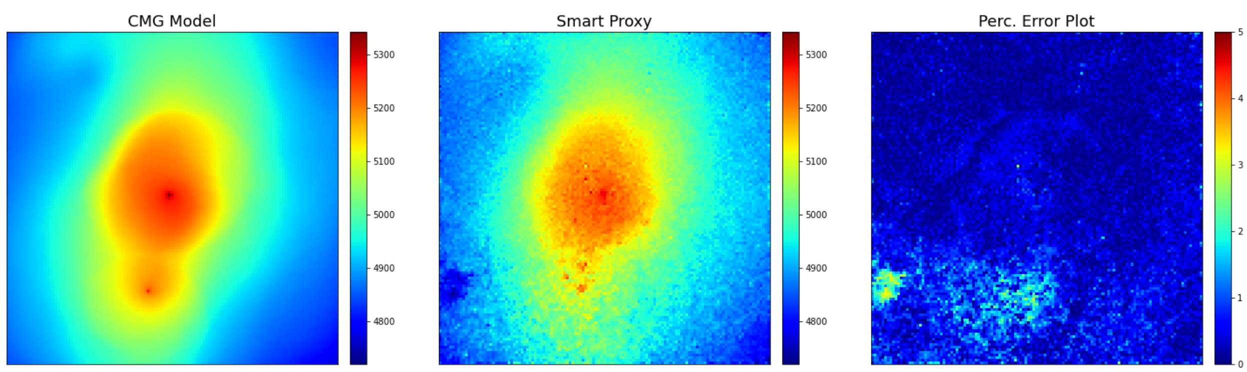
Blind__Run_8_W1_8: Layer-5

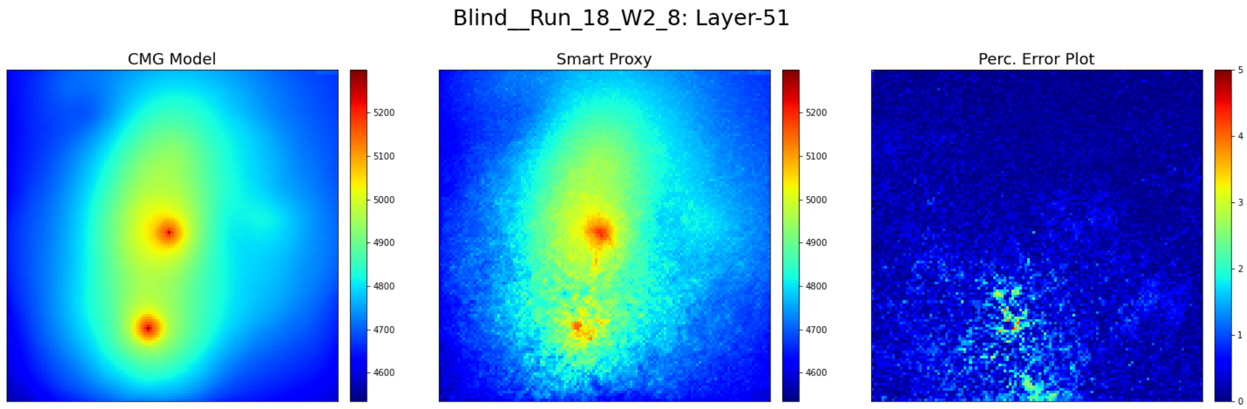


Blind__Run_8_W1_8: Layer-51

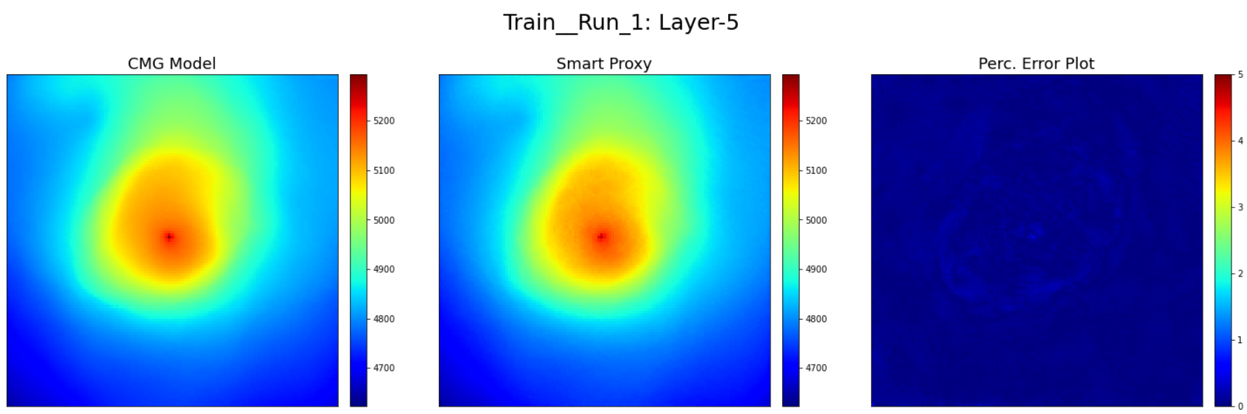


Blind__Run_18_W2_8: Layer-5

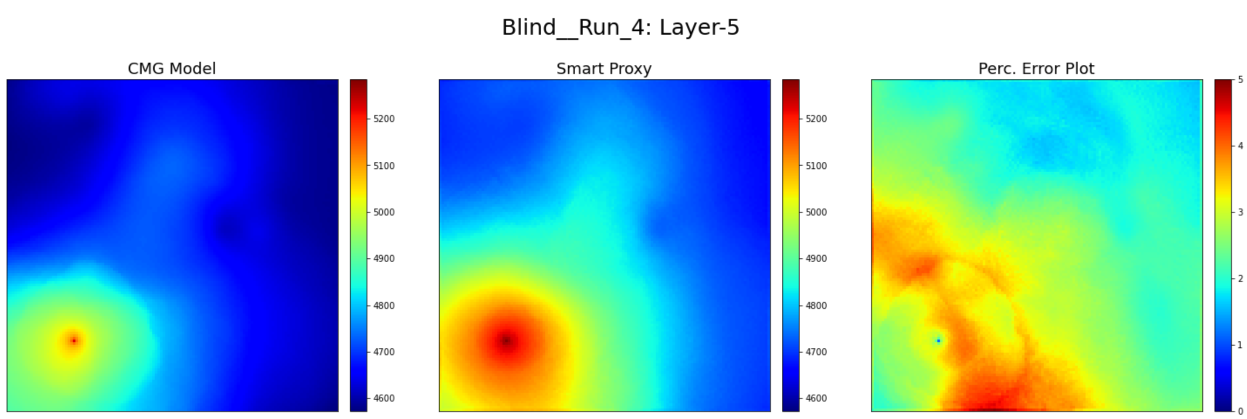


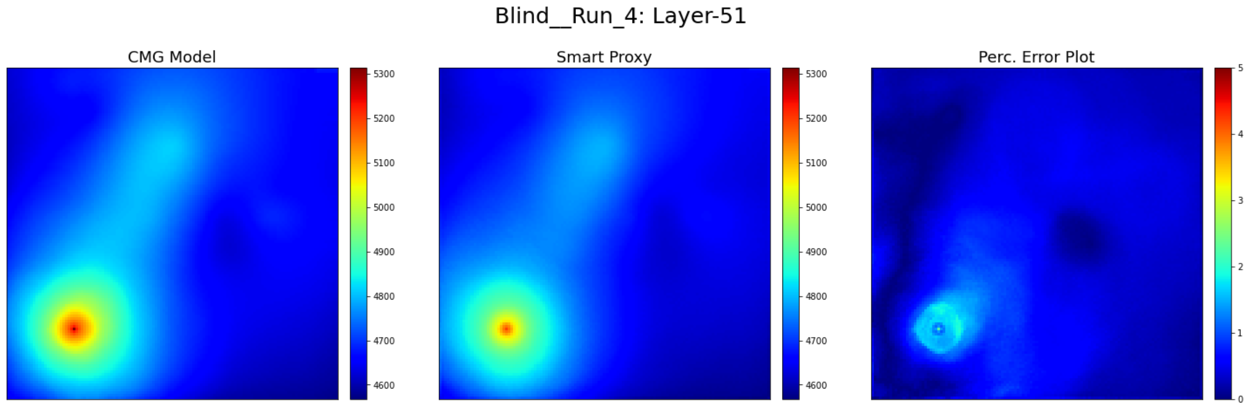


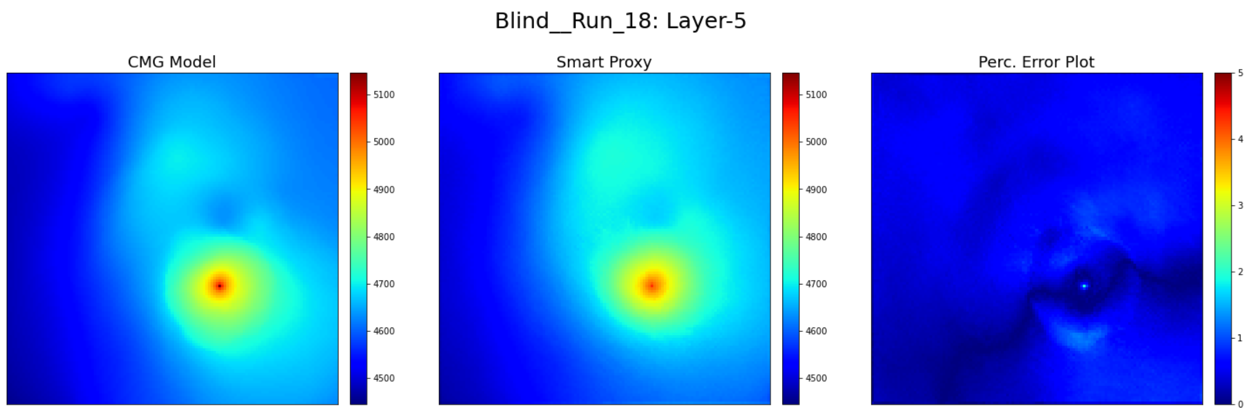
7.5 Phase 3 Pressure Trains:

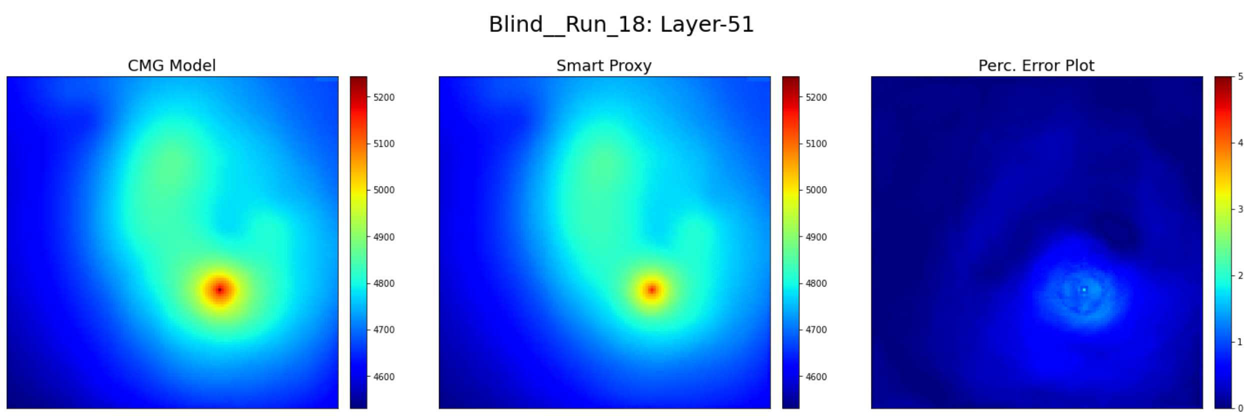


7.6 Other Phase 3 Pressure Blinds:

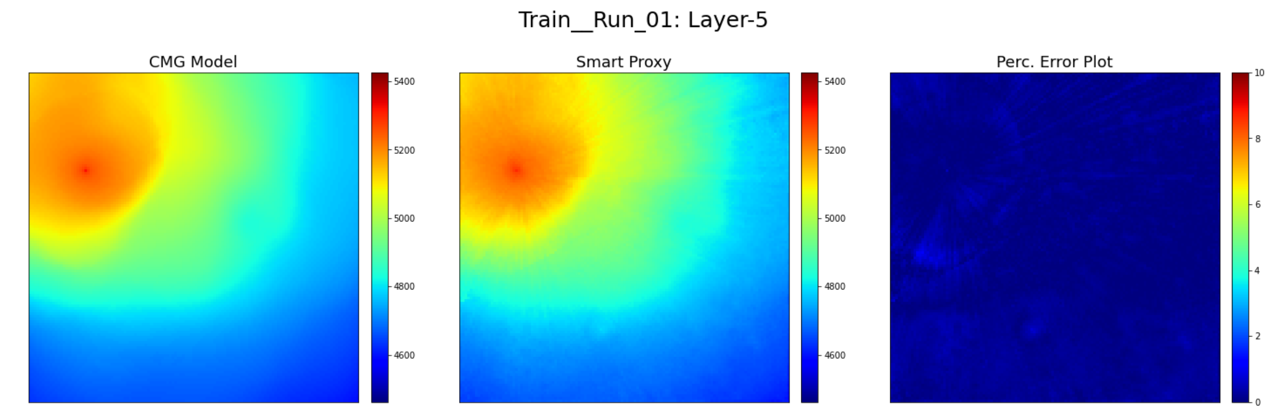


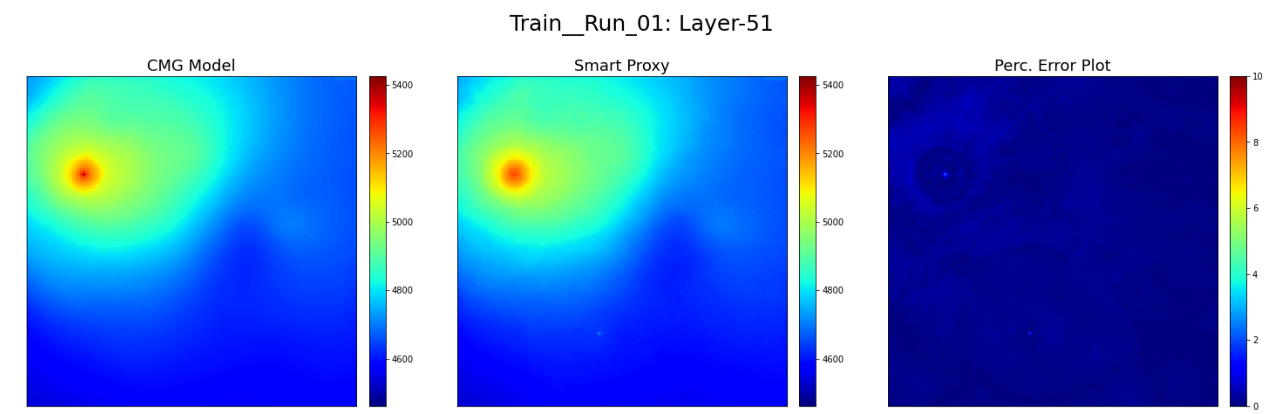

Blind__Run_4: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot
5300
5200
5100
5000
4900
4800
4700
4600
5
4
3
2
1
0

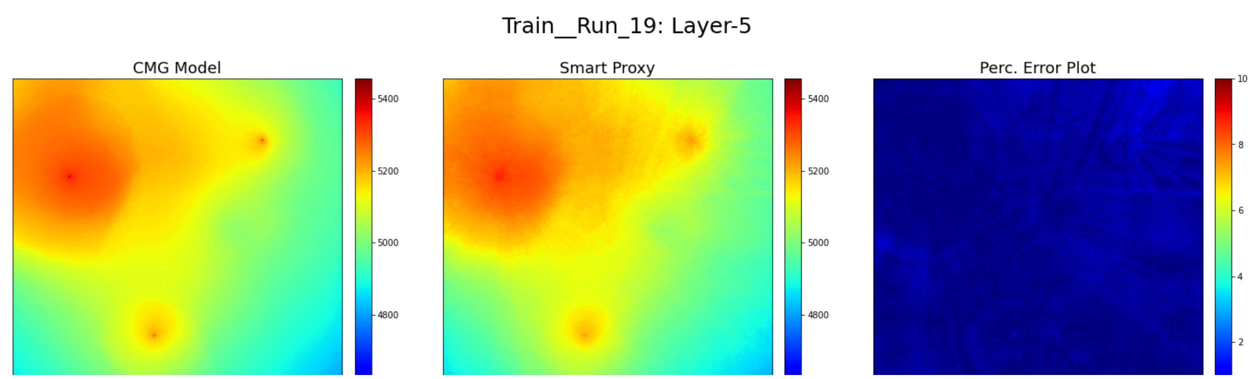

Blind__Run_18: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot
5100
5000
4900
4800
4700
4600
4500
5
4
3
2
1
0

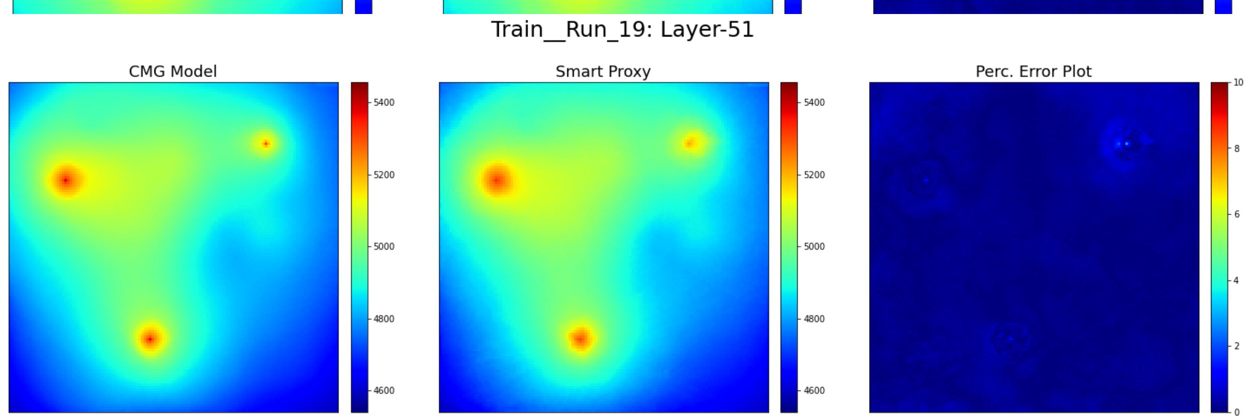

Blind__Run_18: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot
5200
5100
5000
4900
4800
4700
4600
5
4
3
2
1
0

7.7 Phase 4 Pressure Trains:

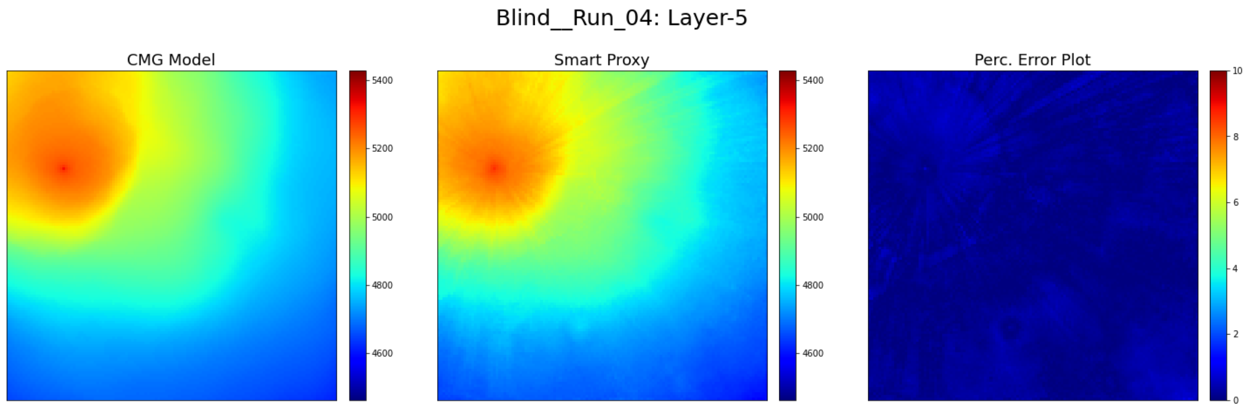


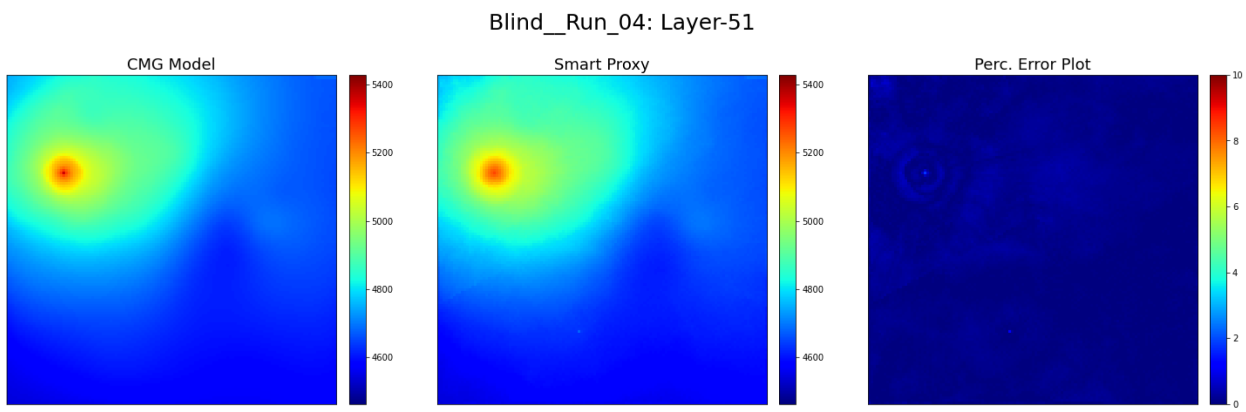


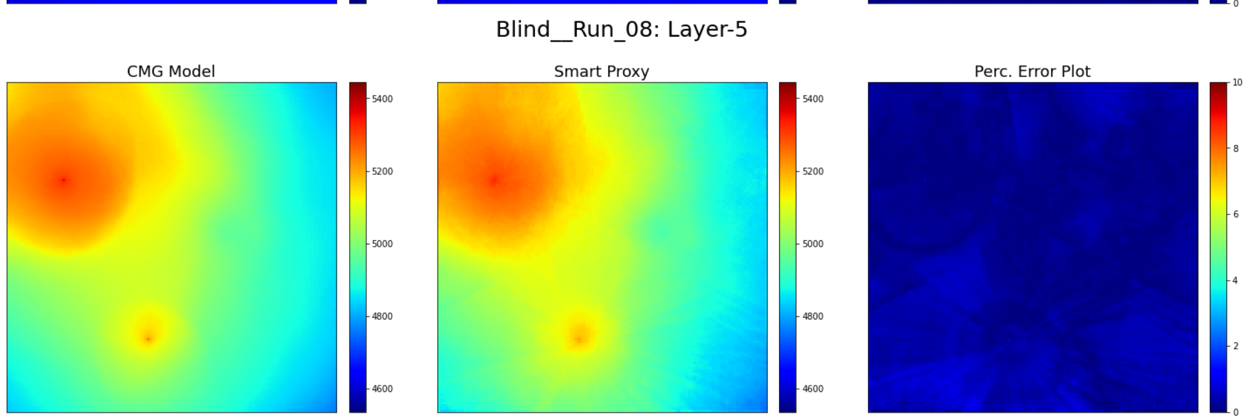


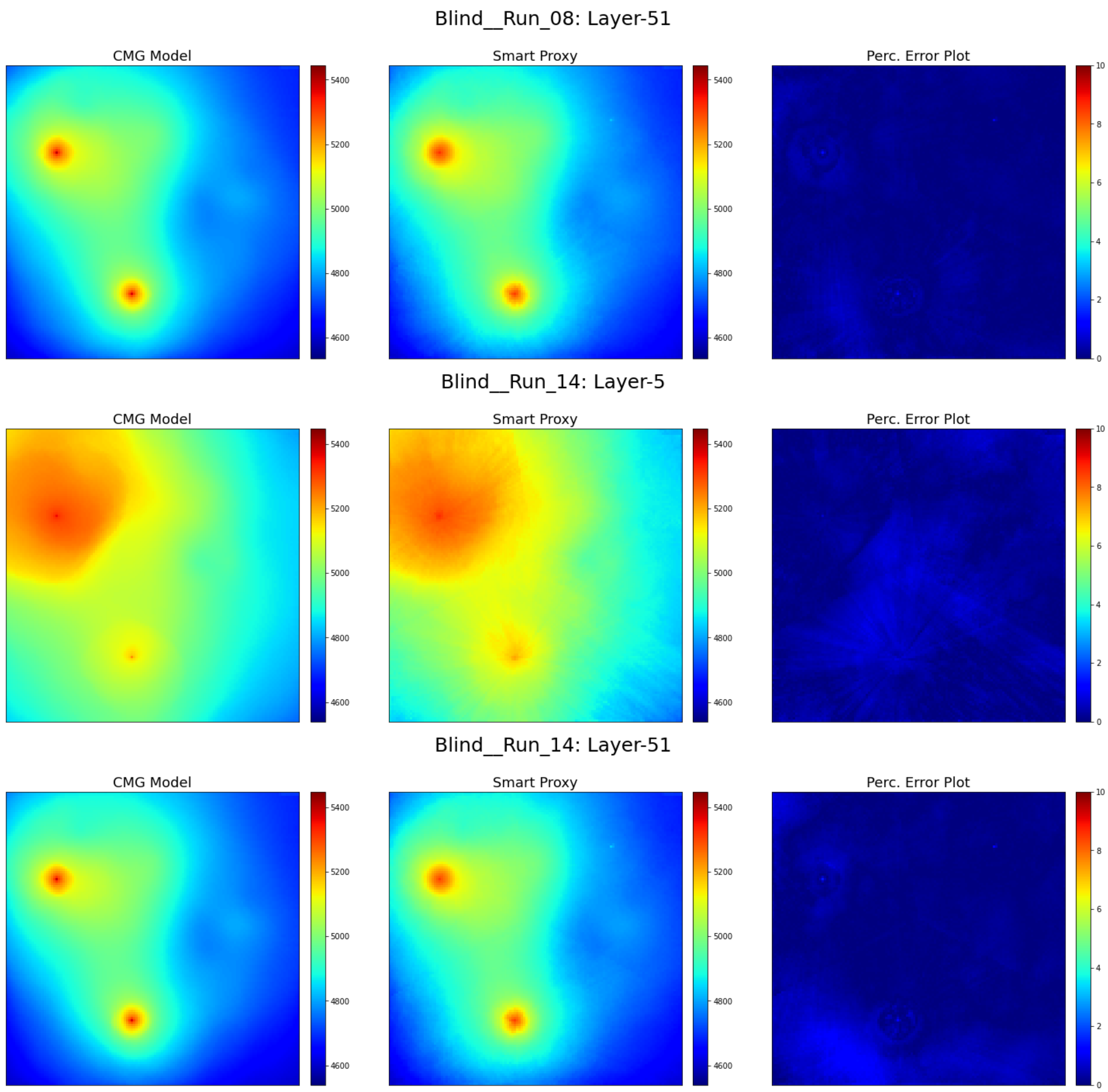


7.8 Other Phase 4 Pressure Blinds:

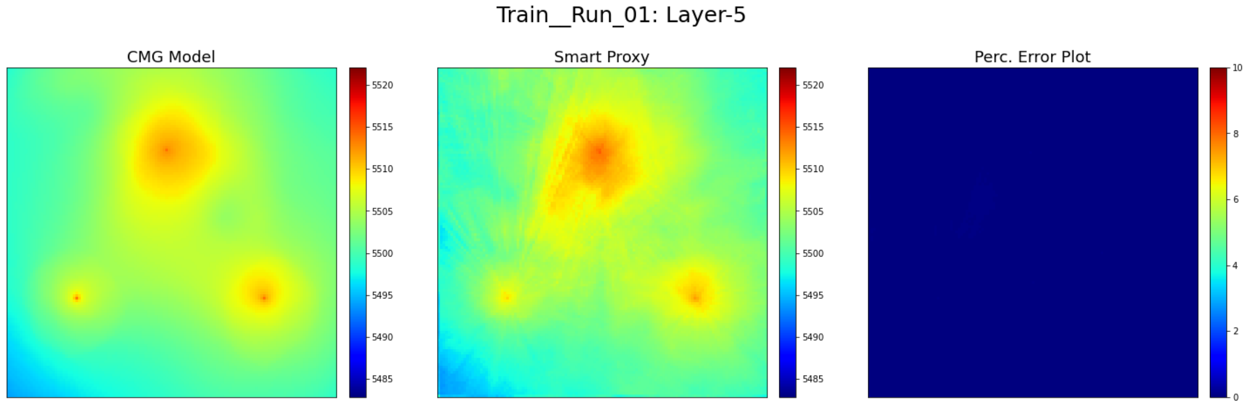


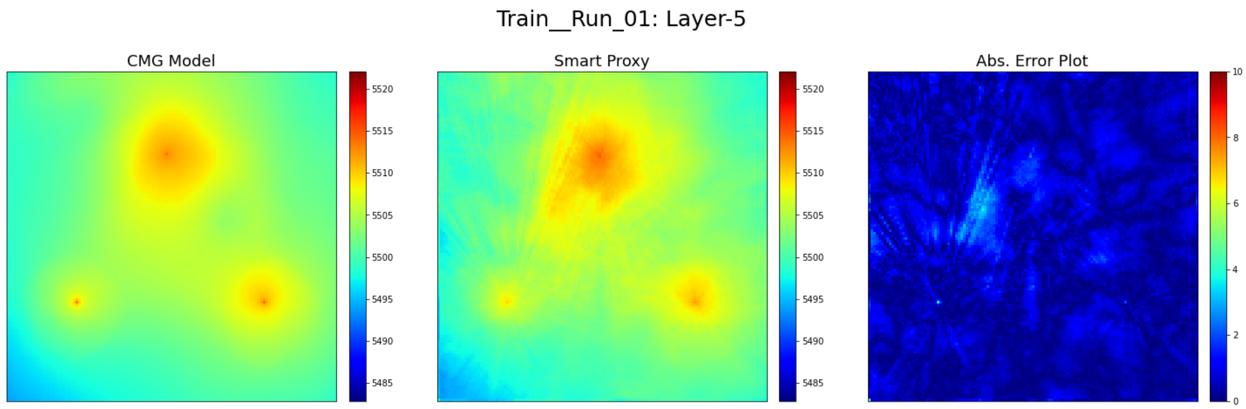


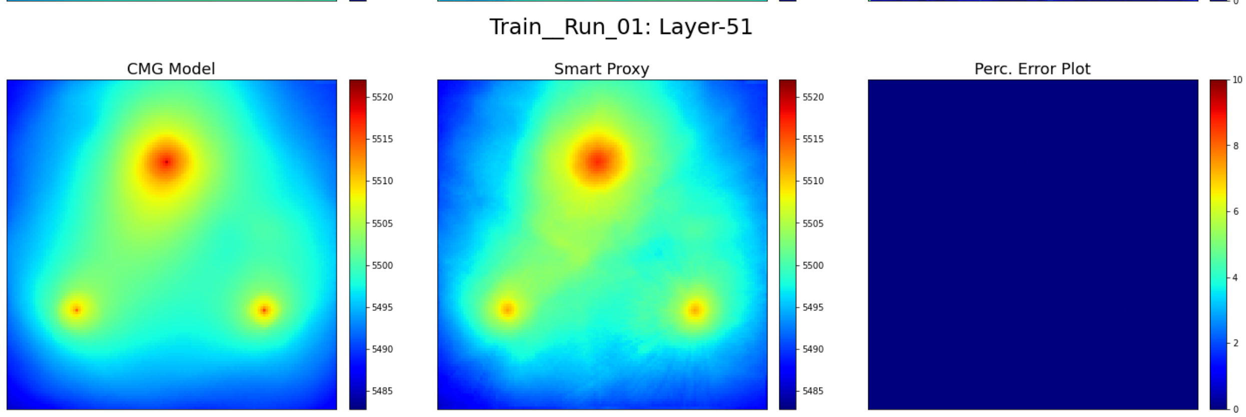


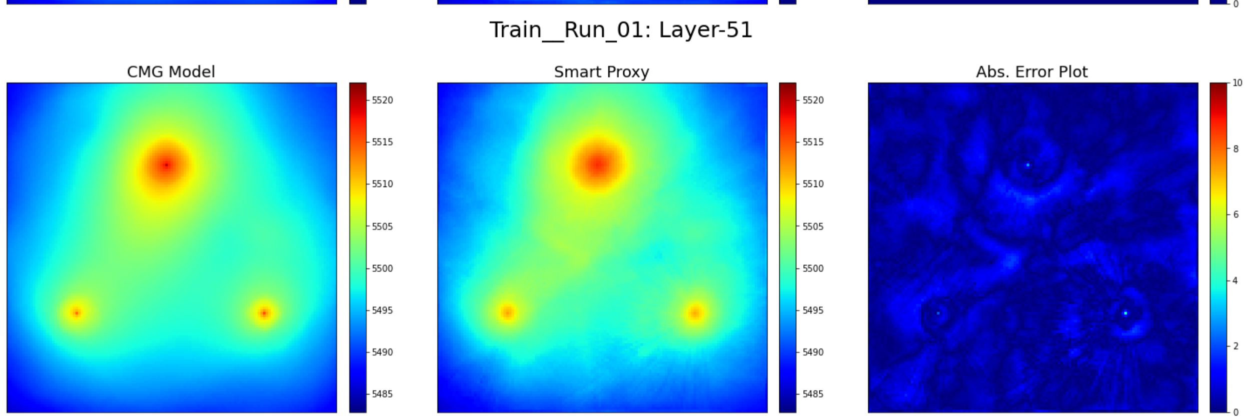

Blind__Run_08: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot
Blind__Run_14: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot
Blind__Run_14: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot
5400
5200
5000
4800
4600
10
8
6
4
2
0

7.9 Phase 5 Pressure Trains:

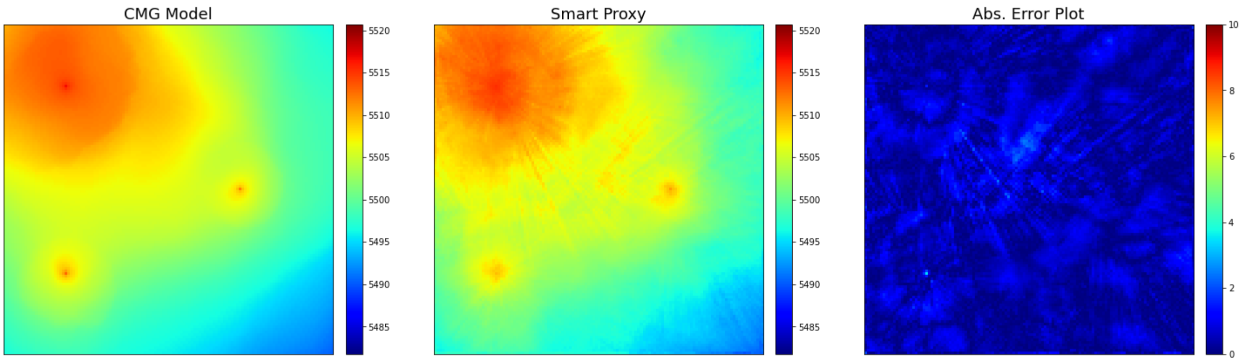




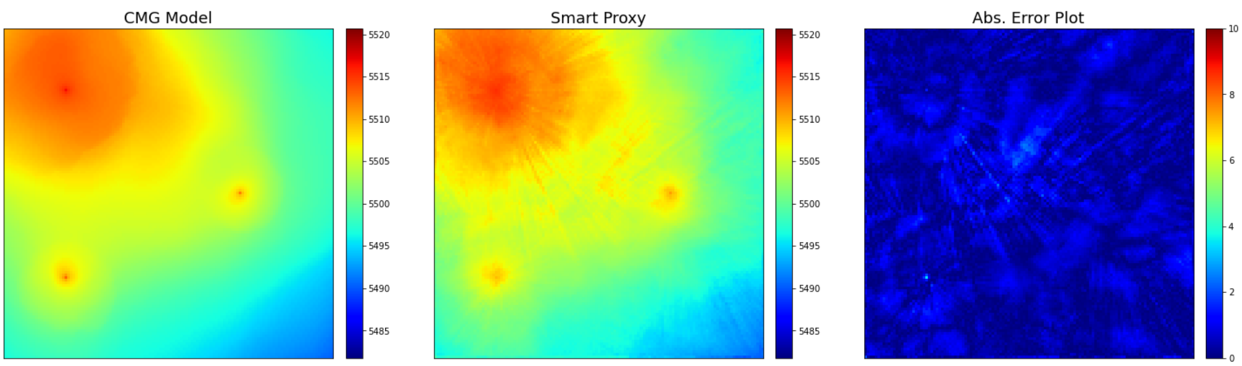




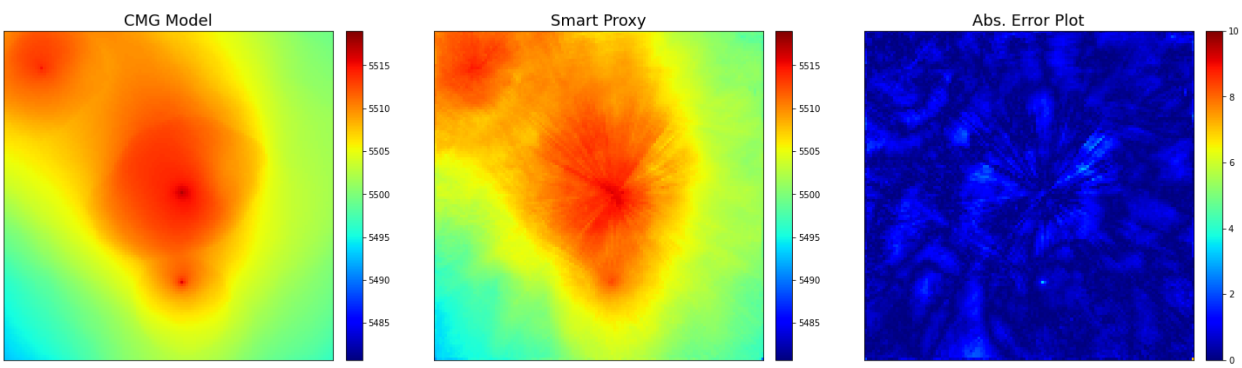
Train__Run_07: Layer-5



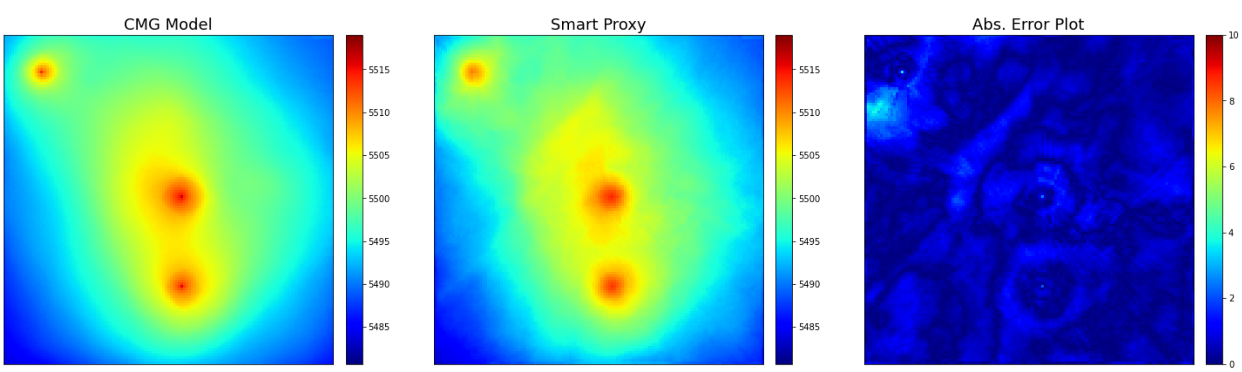
Train__Run_07: Layer-5

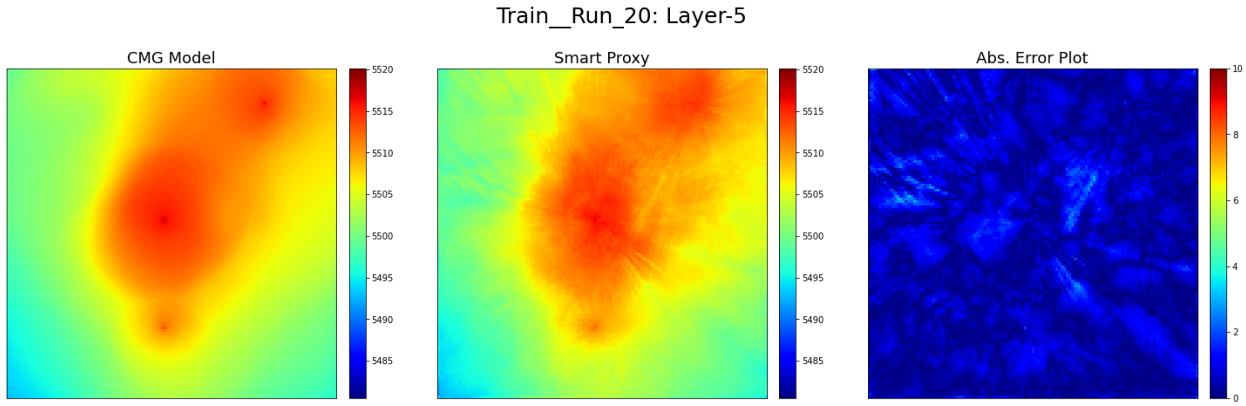


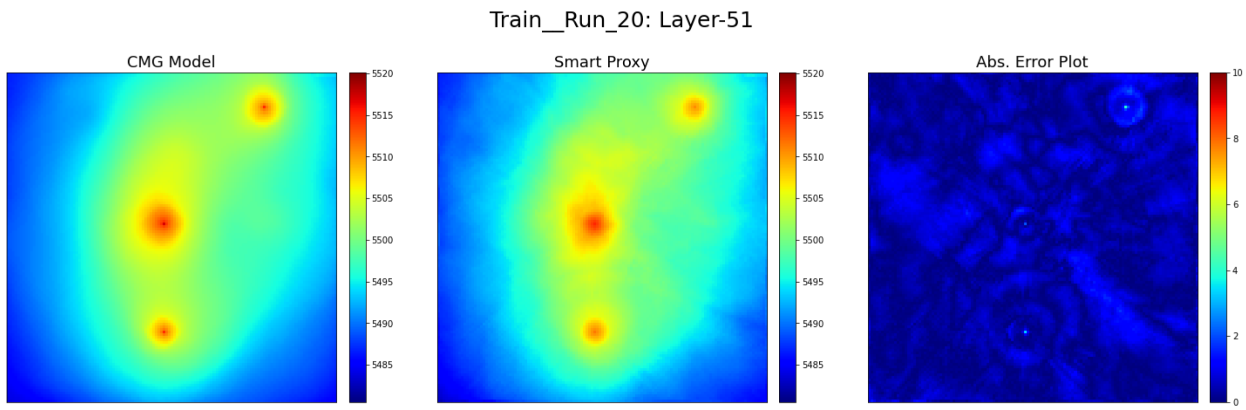
Train__Run_19: Layer-5



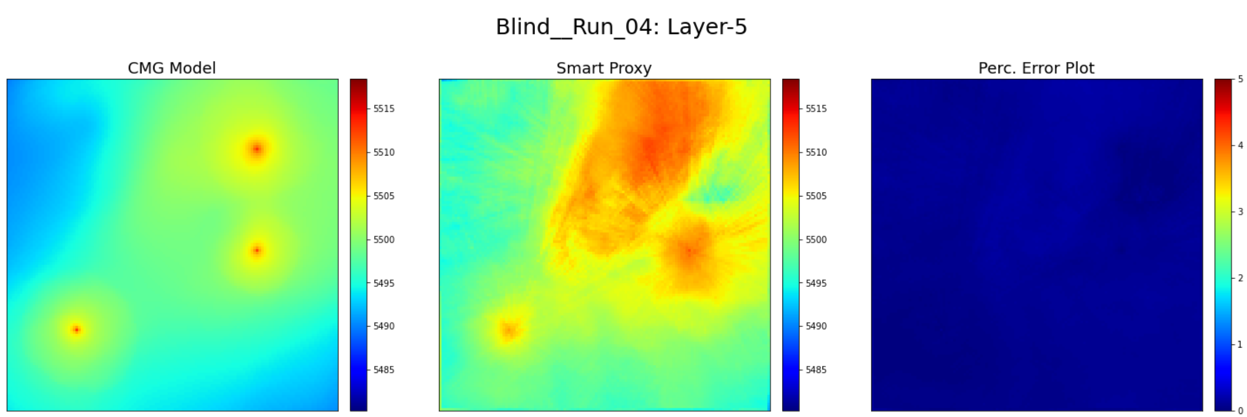
Train__Run_19: Layer-51

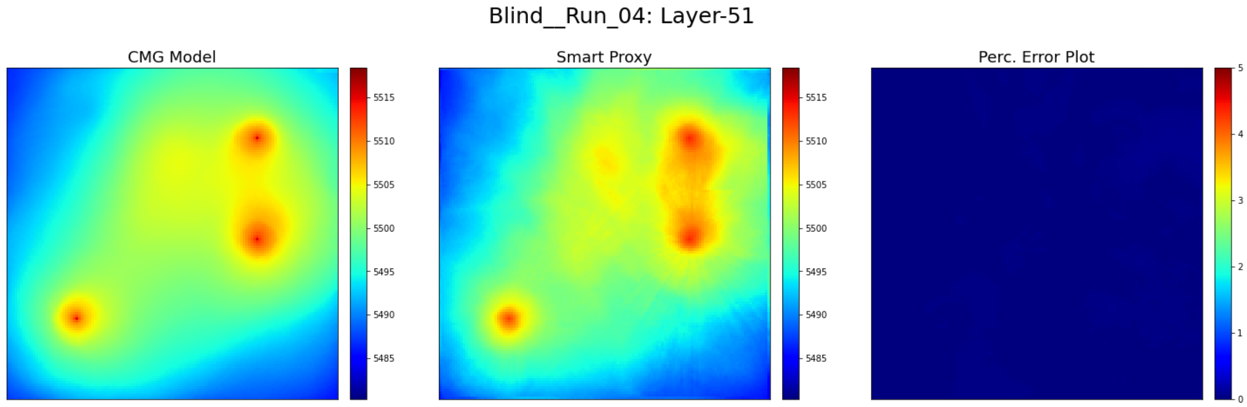


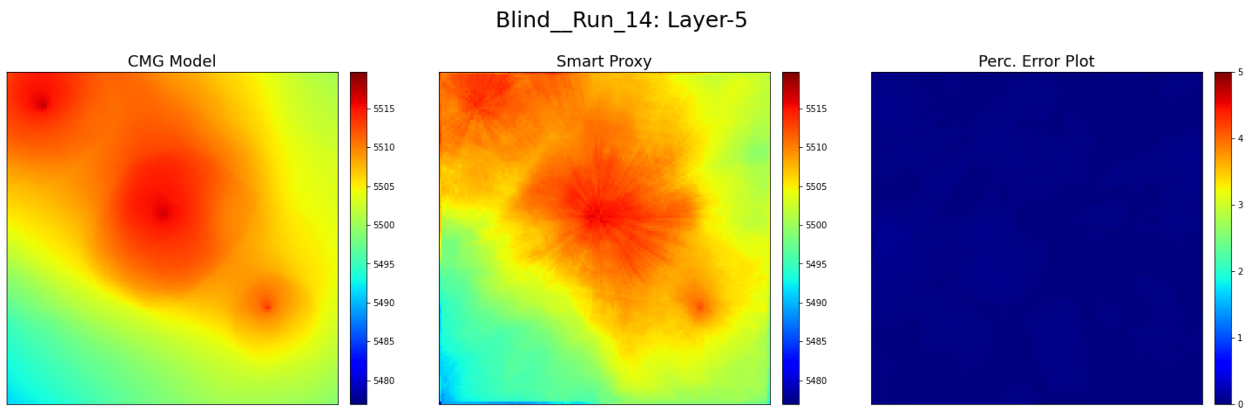


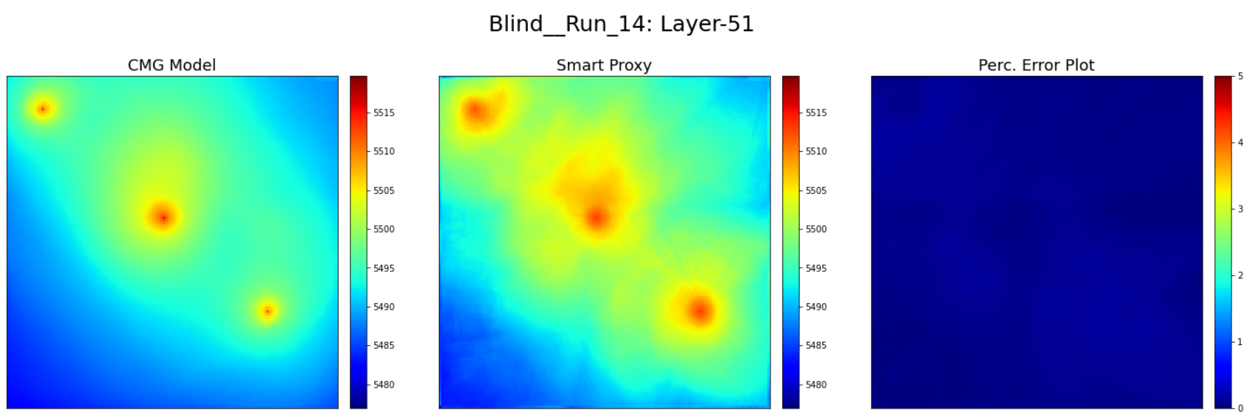


7.10 Other Phase 5 Pressure Blinds:



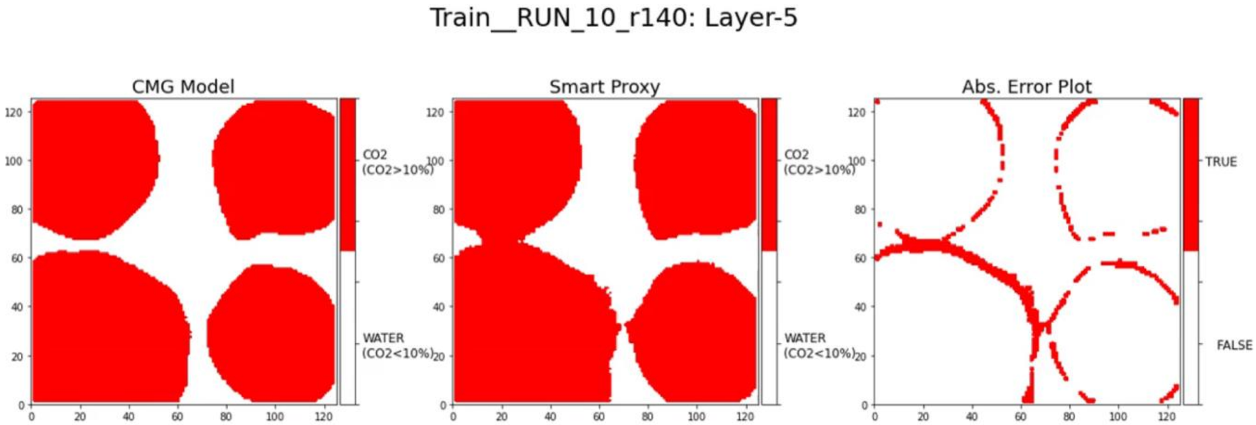

Blind__Run_04: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot
5515
5510
5505
5500
5495
5490
5485
5
4
3
2
1
0

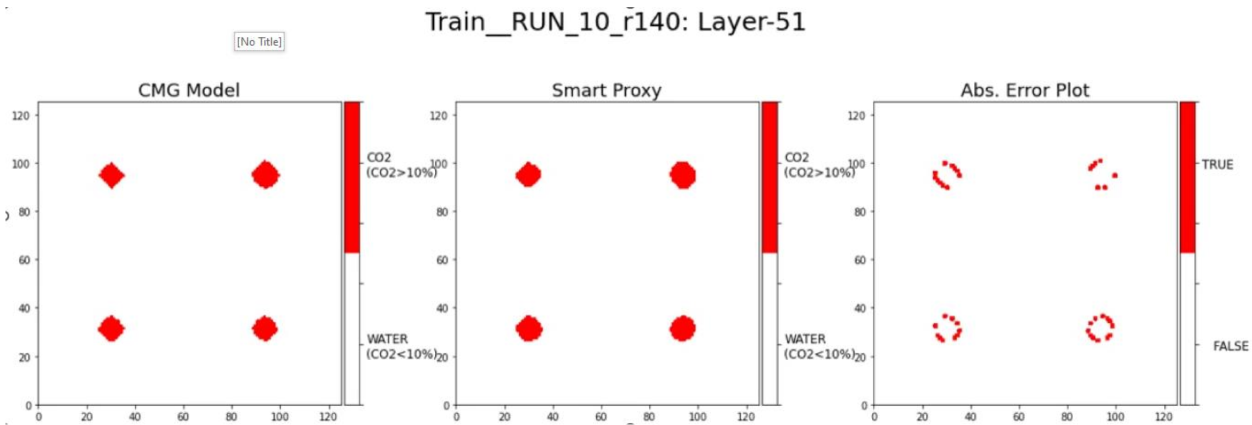

Blind__Run_14: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot
5515
5510
5505
5500
5495
5490
5485
5480
5
4
3
2
1
0

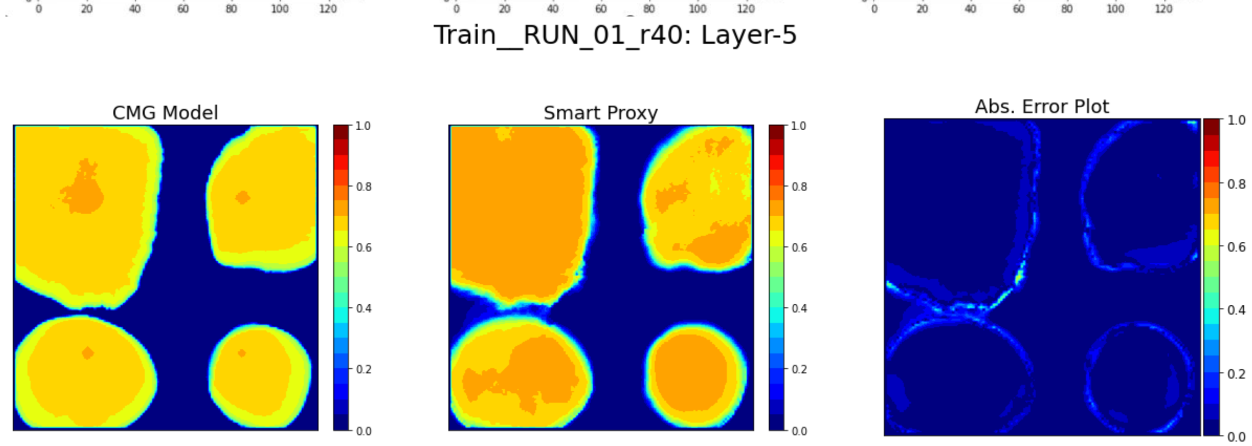

Blind__Run_14: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot
5515
5510
5505
5500
5495
5490
5485
5480
5
4
3
2
1
0

Saturation Results:

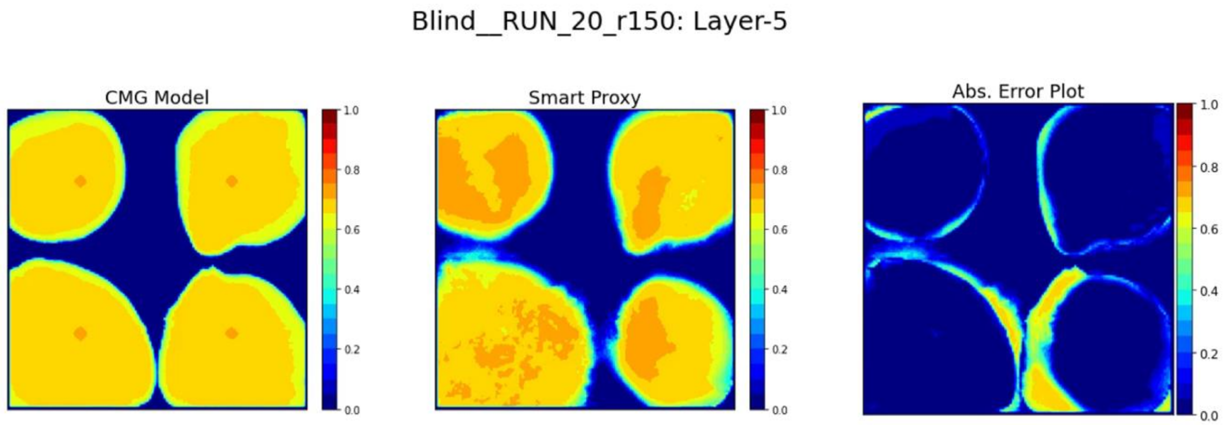
7.11 Phase 1 Saturation Trains:

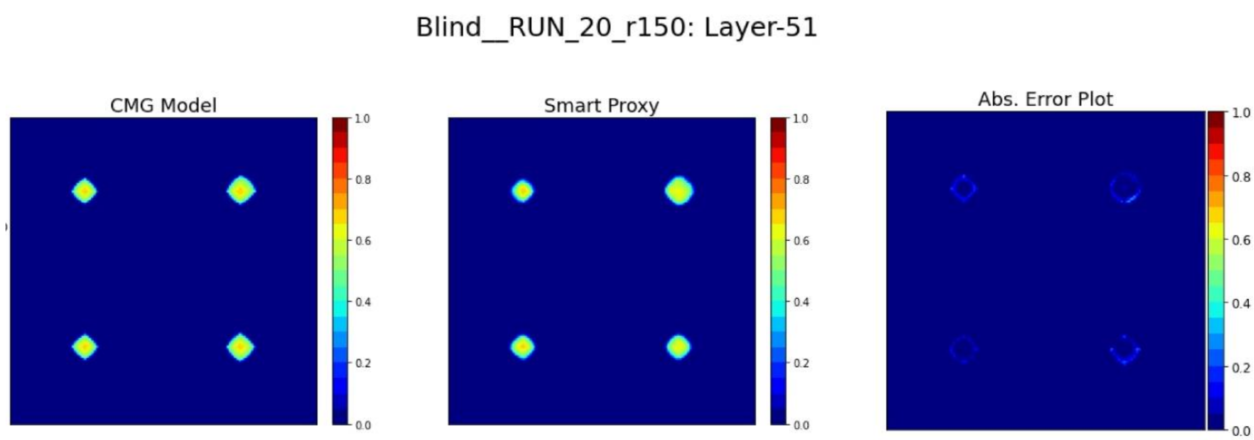




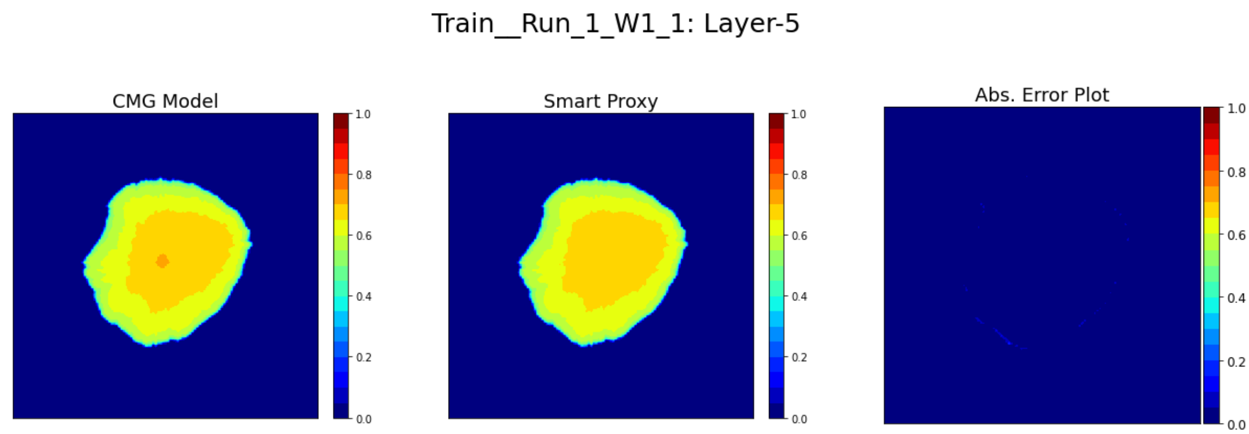


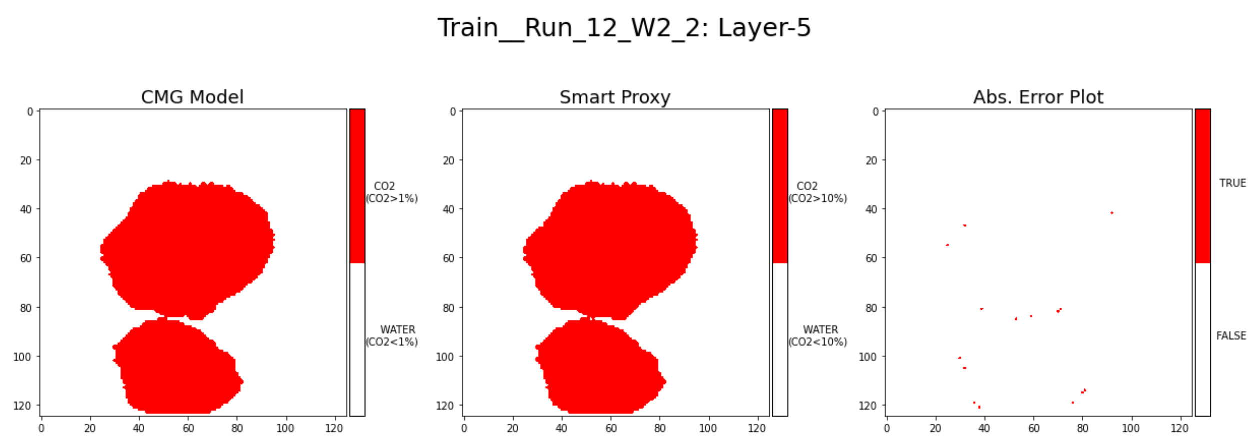
7.12 Other Phase 1 Saturation Blinds:

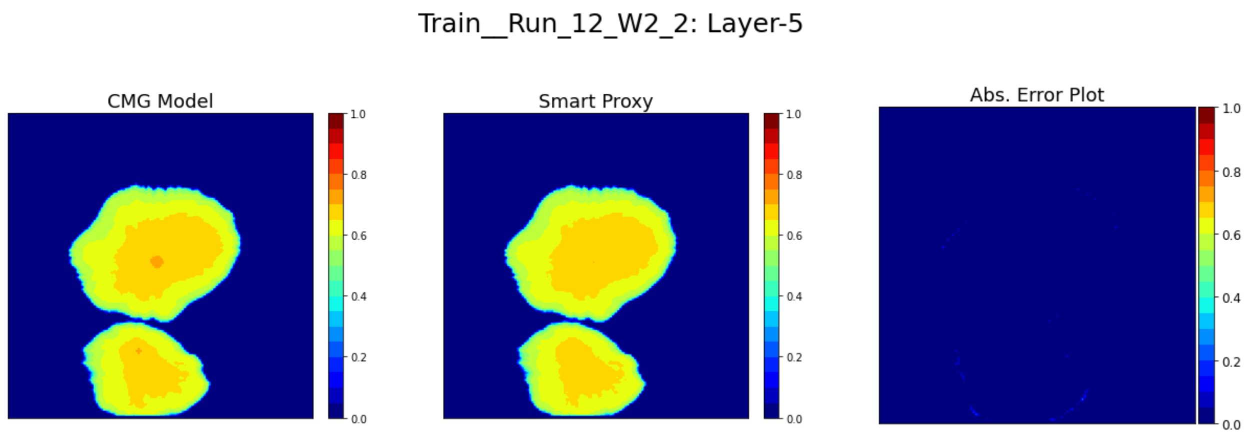




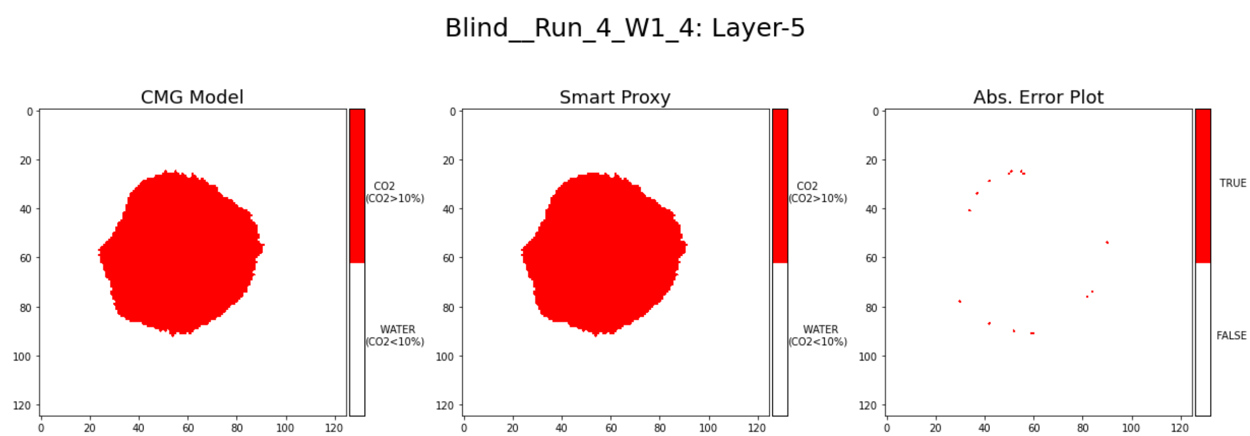
7.13 Phase 2 Saturation Trains:

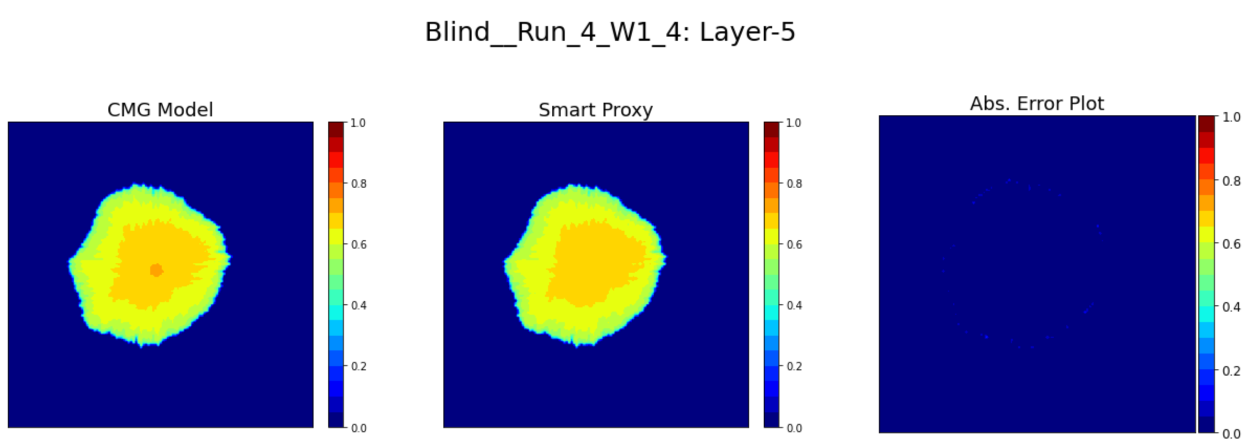




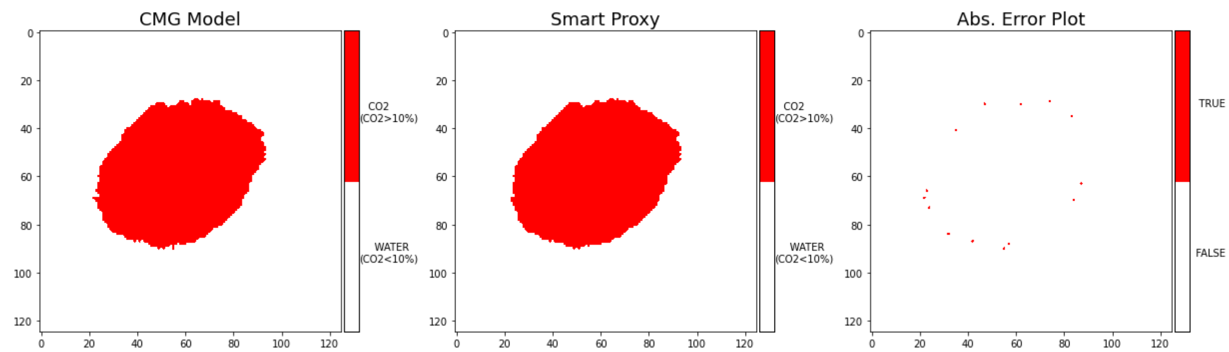


7.14 Other Phase 2 Saturation Blinds:

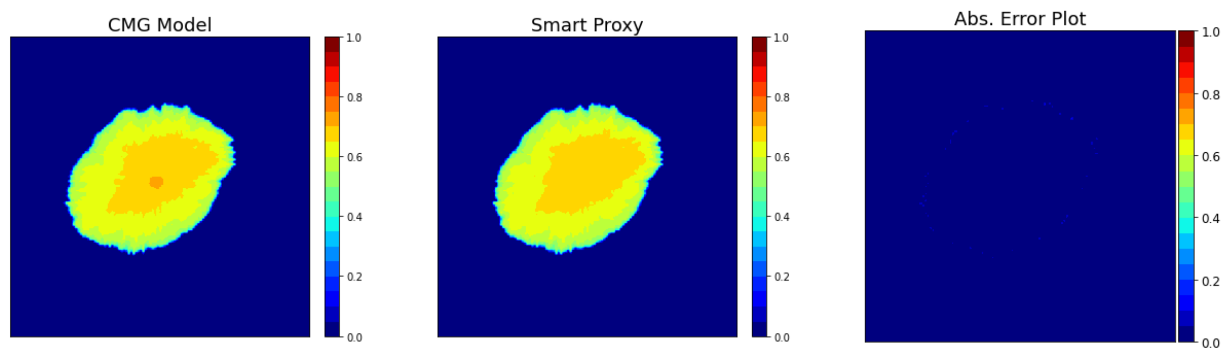




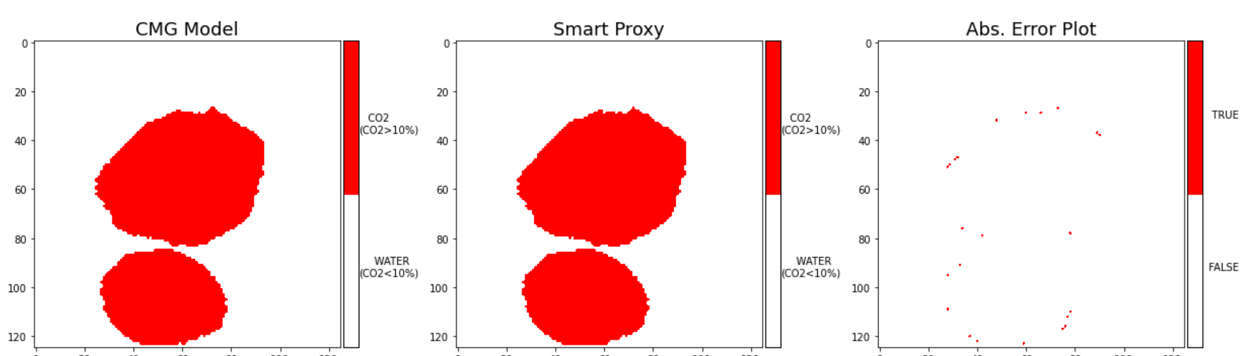
Blind__Run_8_W1_8: Layer-5



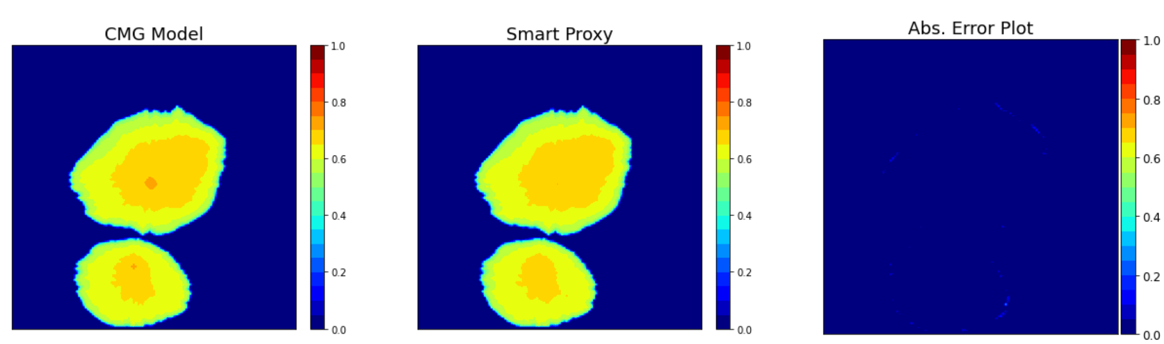
Blind__Run_8_W1_8: Layer-5



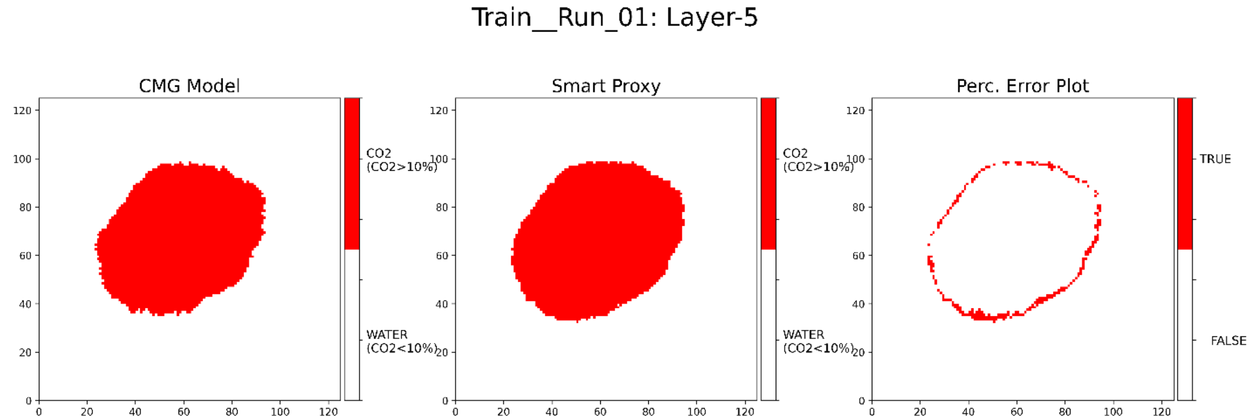
Blind__Run_14_W2_4: Layer-5

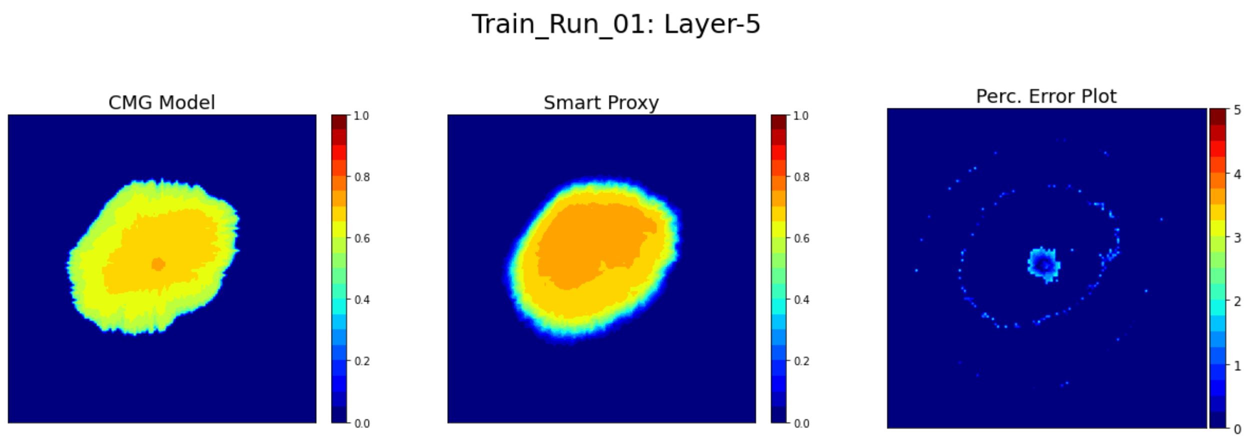


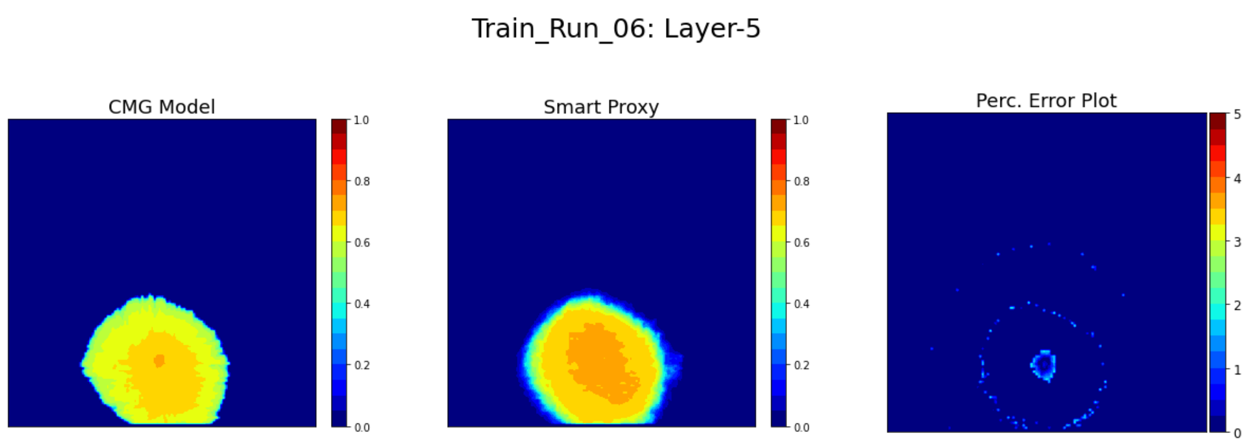
Blind__Run_14_W2_4: Layer-5

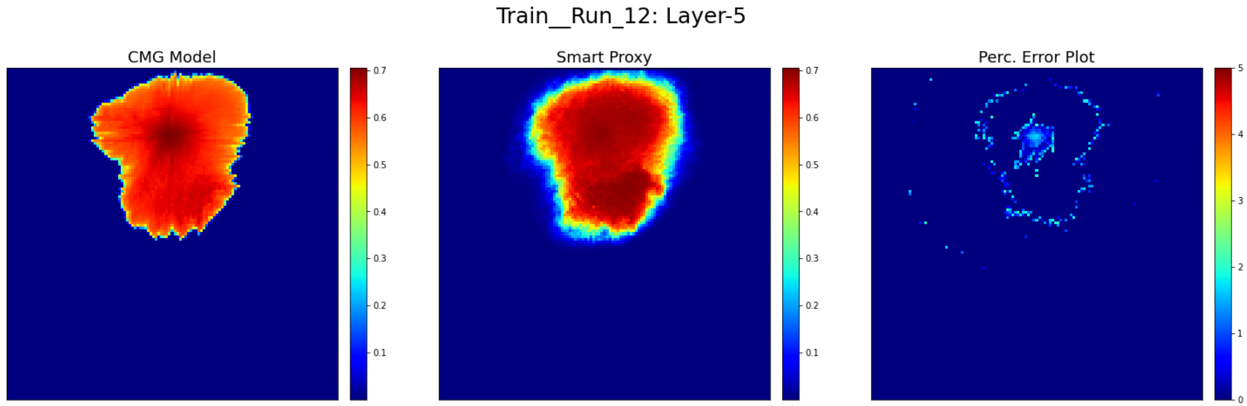


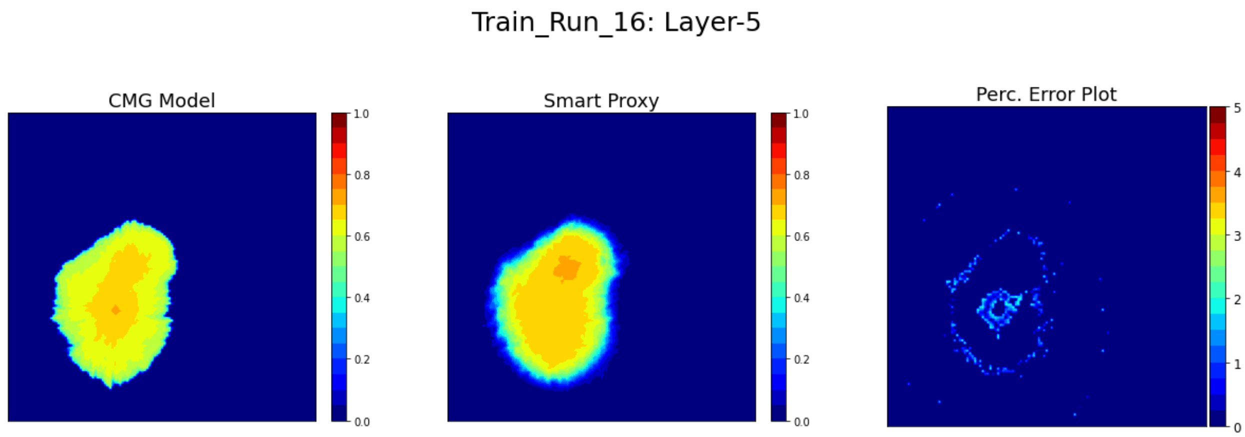
7.15 Phase 3 Saturation Trains:

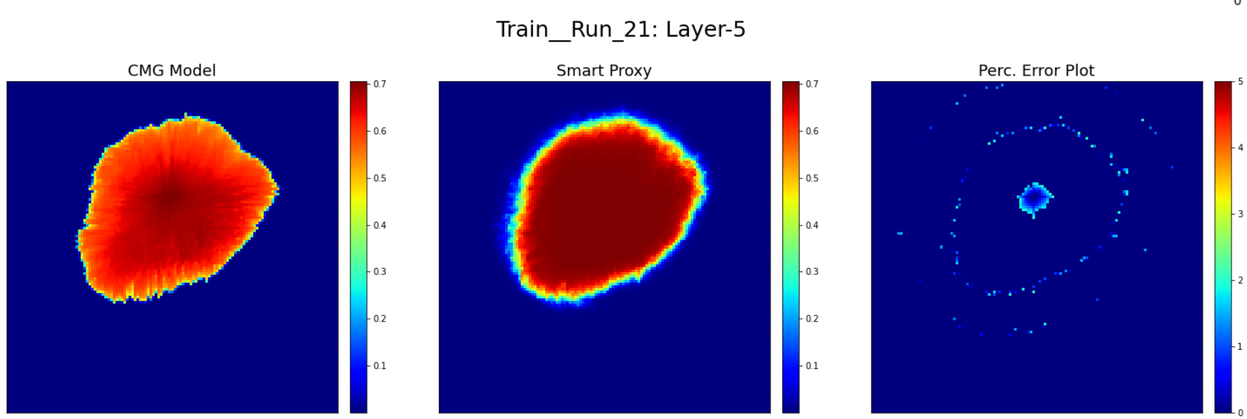




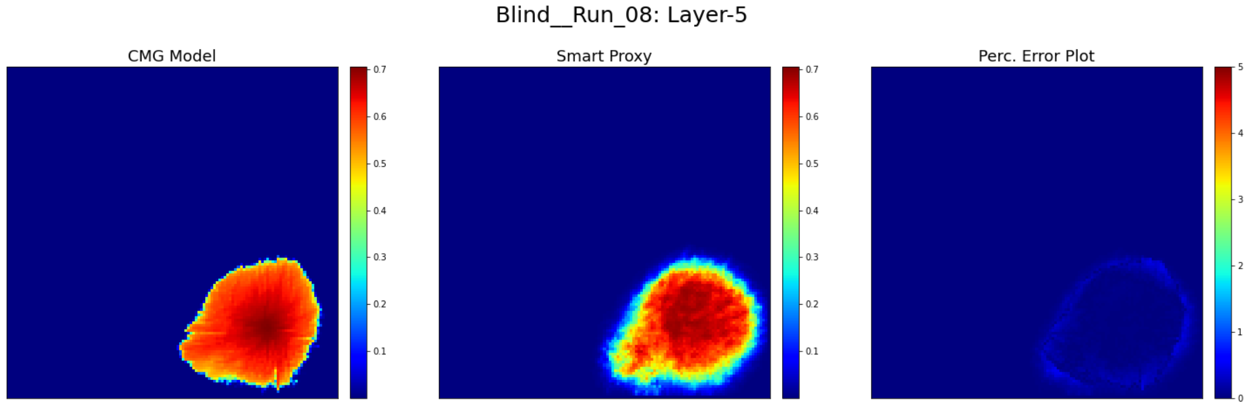


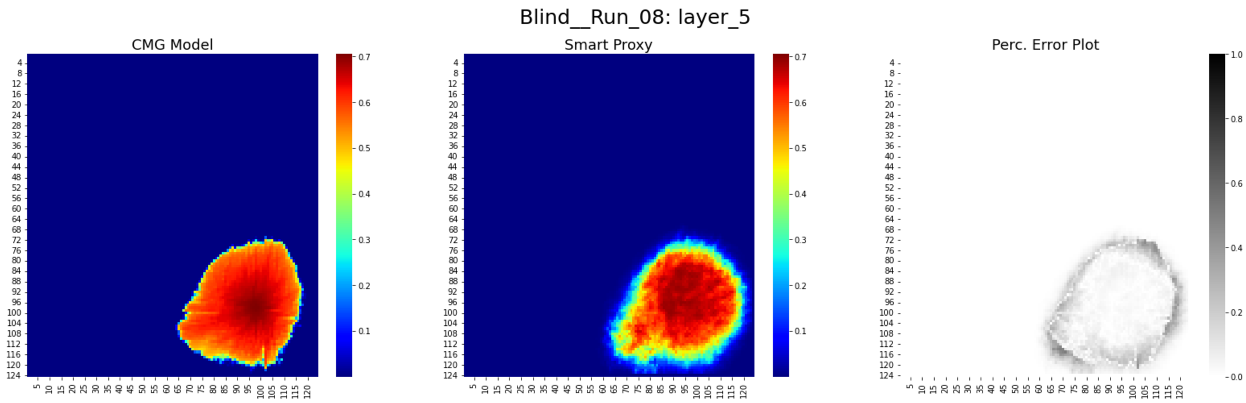

Train__Run_12: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot

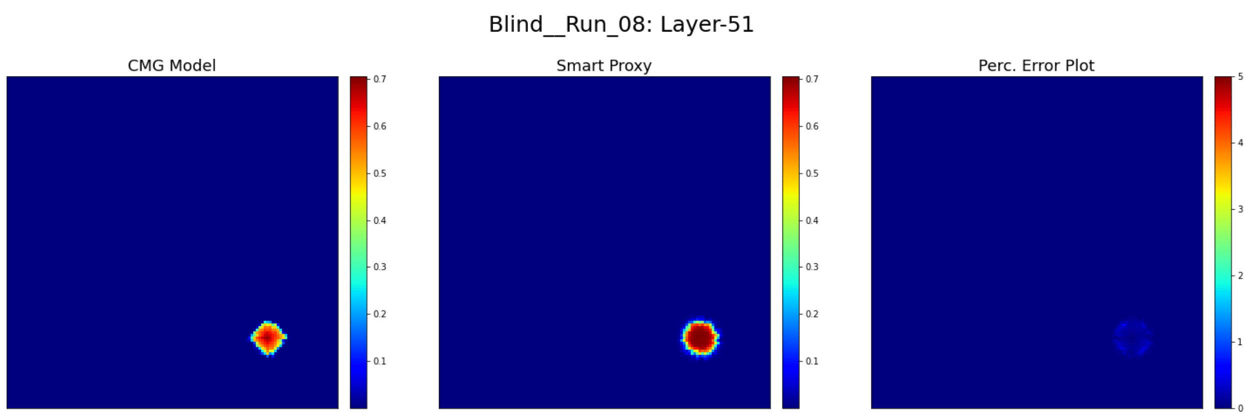

Train_Run_16: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot

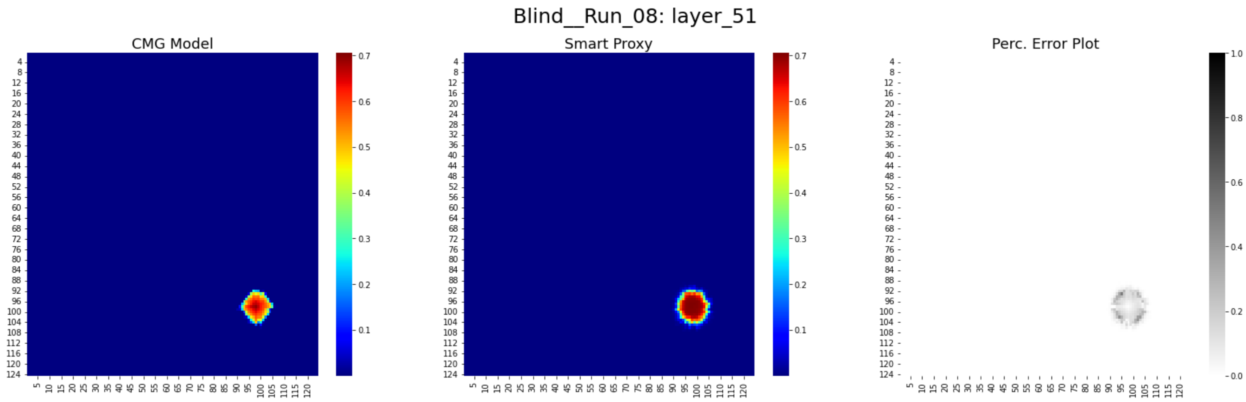

Train__Run_21: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot

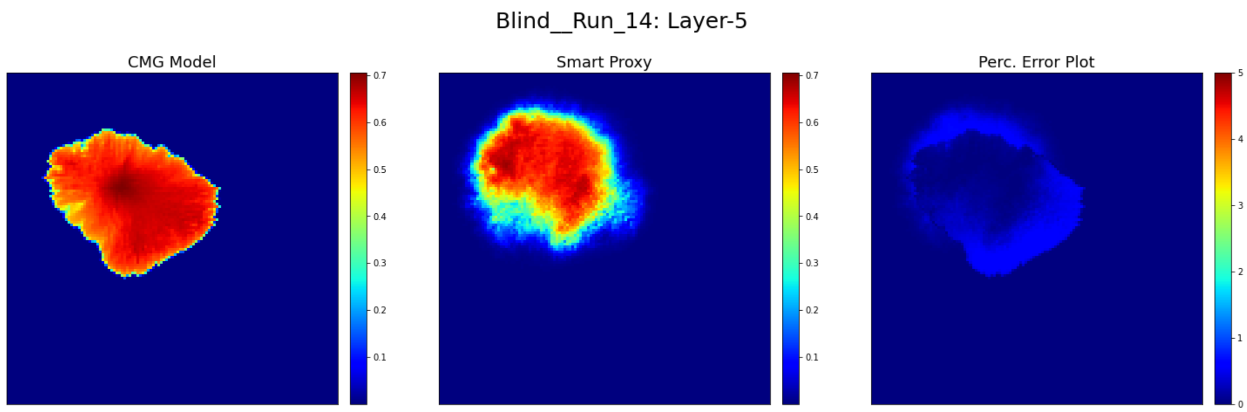
7.16 Other Phase 3 Saturation Blinds:

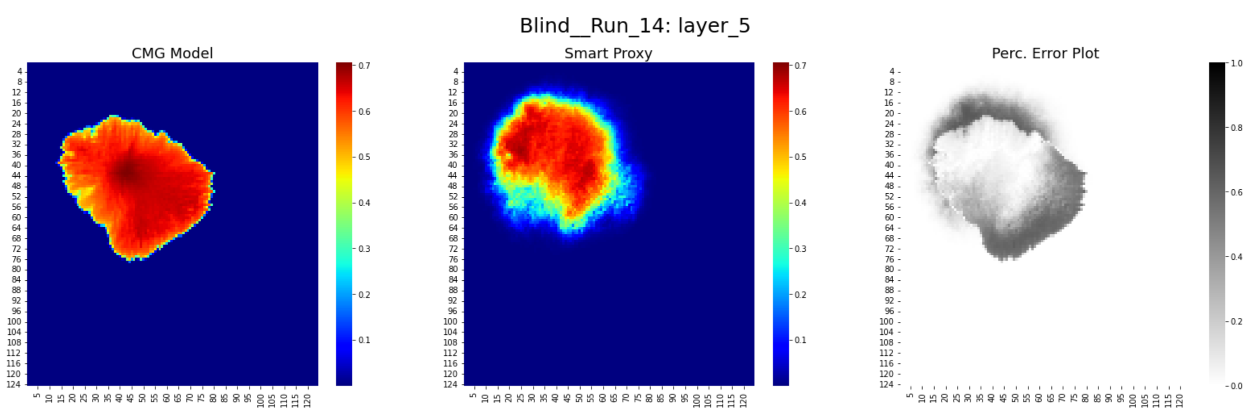


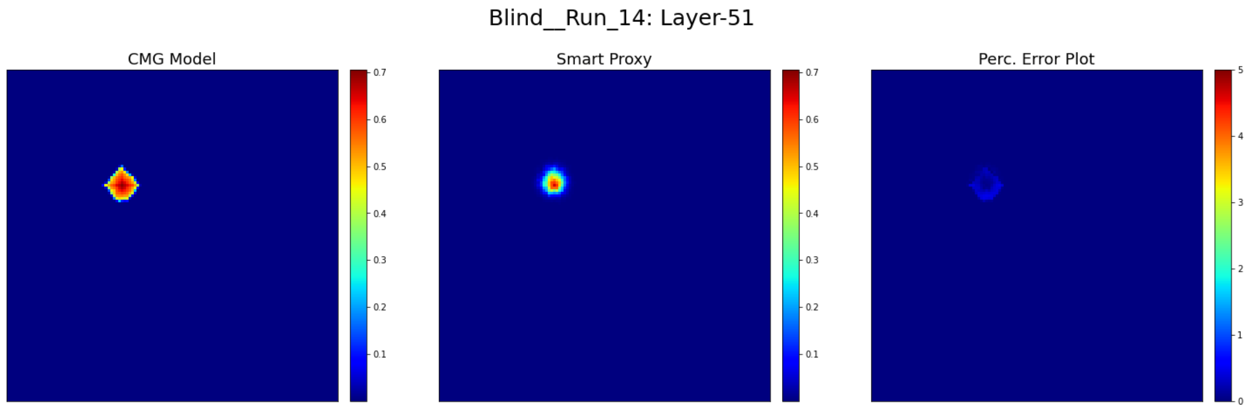


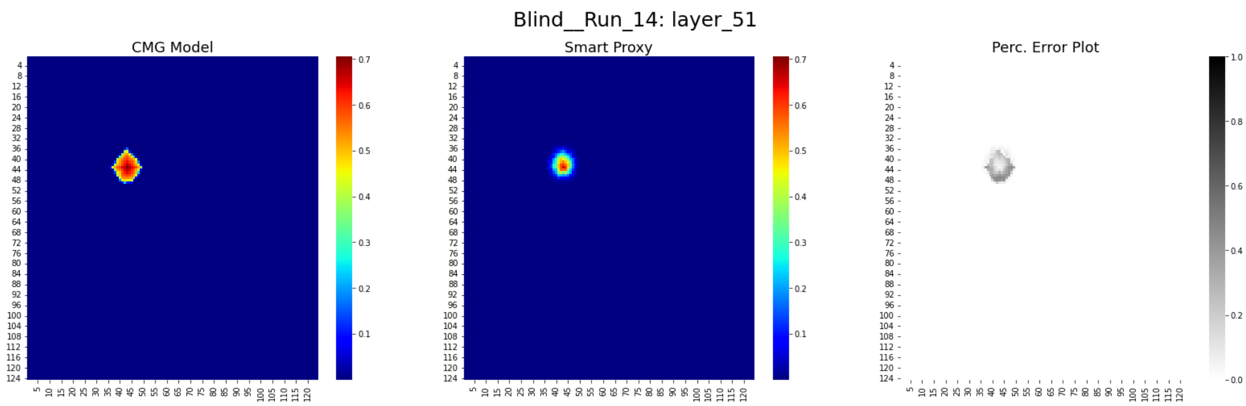



Blind__Run_08: layer_51
CMG Model
Smart Proxy
Perc. Error Plot

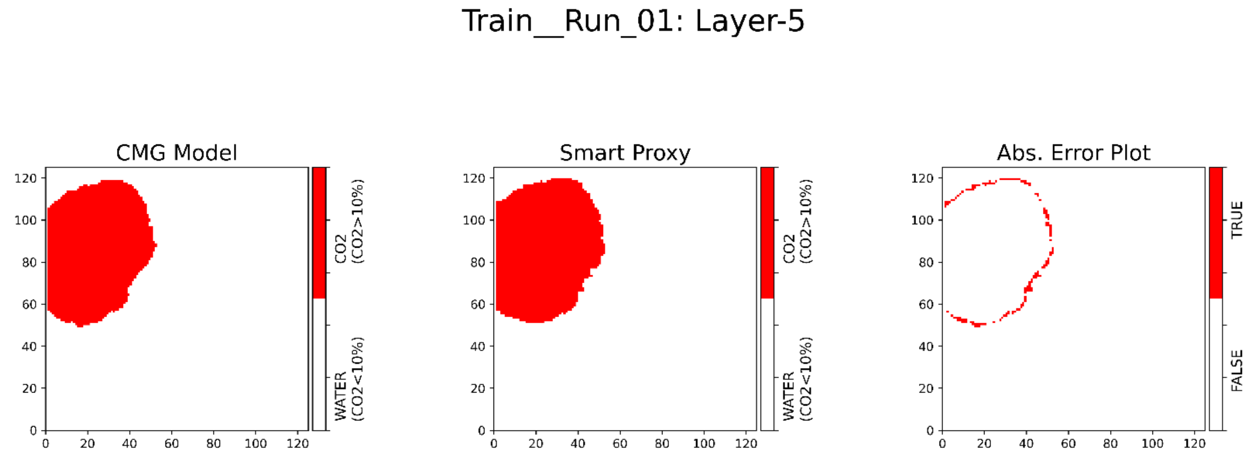

Blind__Run_14: Layer-5
CMG Model
Smart Proxy
Perc. Error Plot

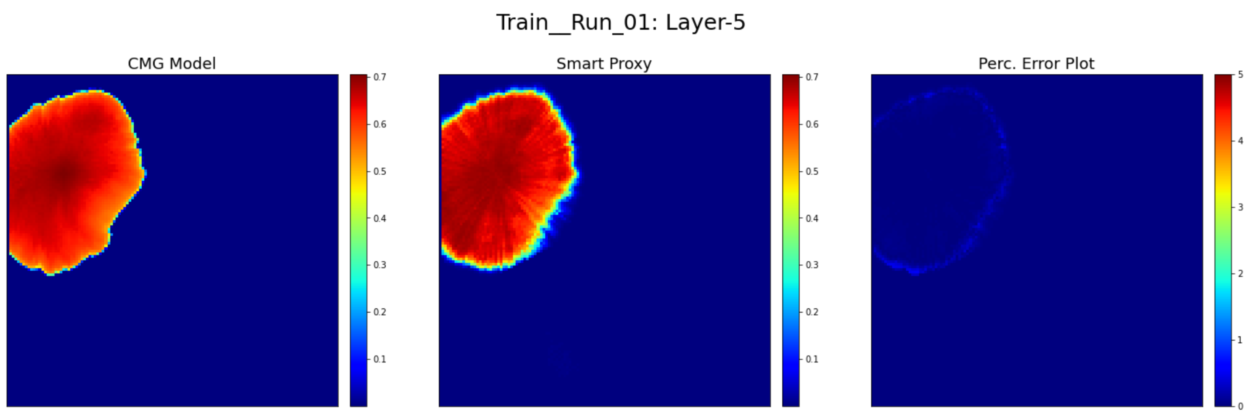

Blind__Run_14: layer_5
CMG Model
Smart Proxy
Perc. Error Plot

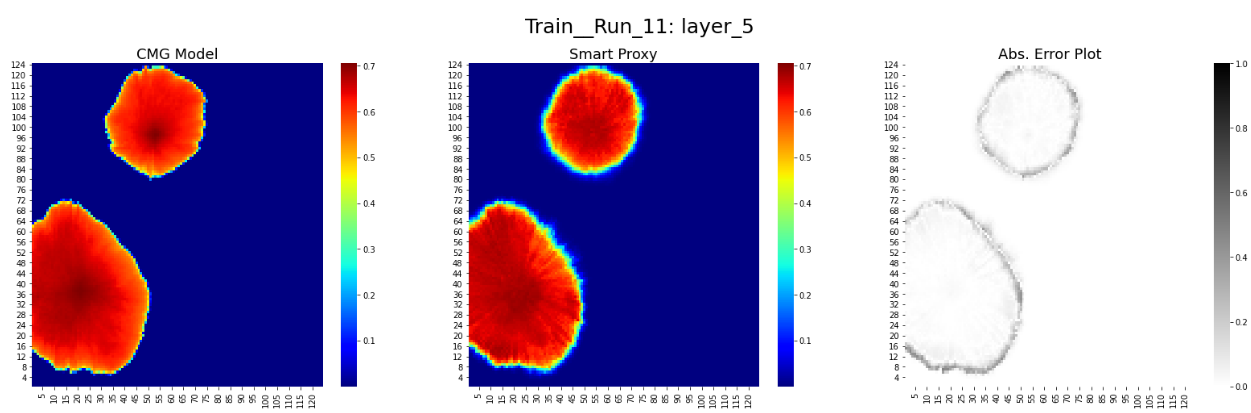

Blind__Run_14: Layer-51
CMG Model
Smart Proxy
Perc. Error Plot

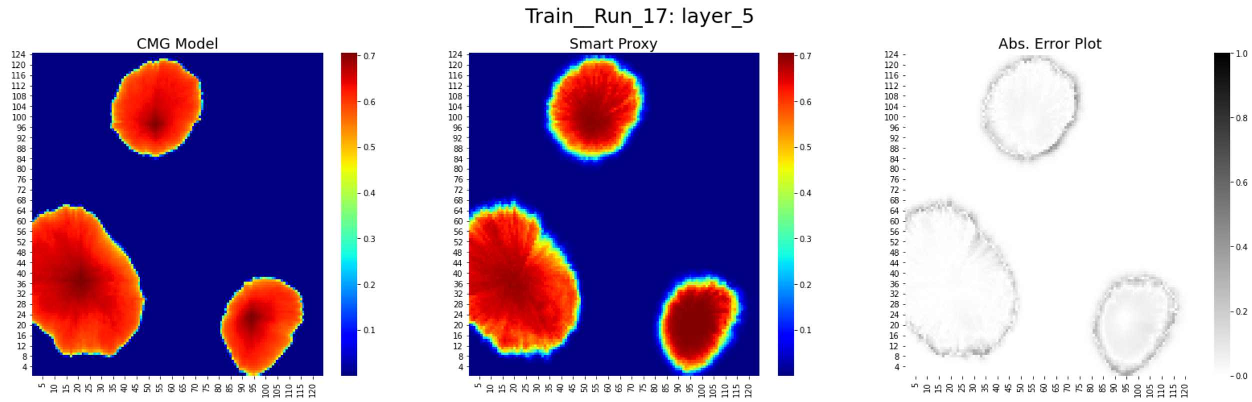

Blind__Run_14: layer_51
CMG Model
Smart Proxy
Perc. Error Plot

7.17 Phase 4 Saturation Trains:

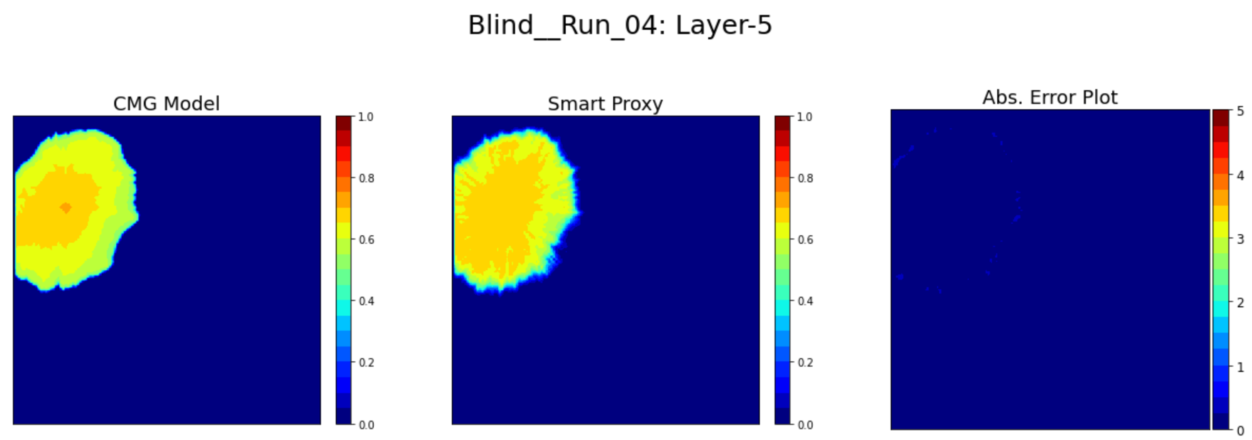


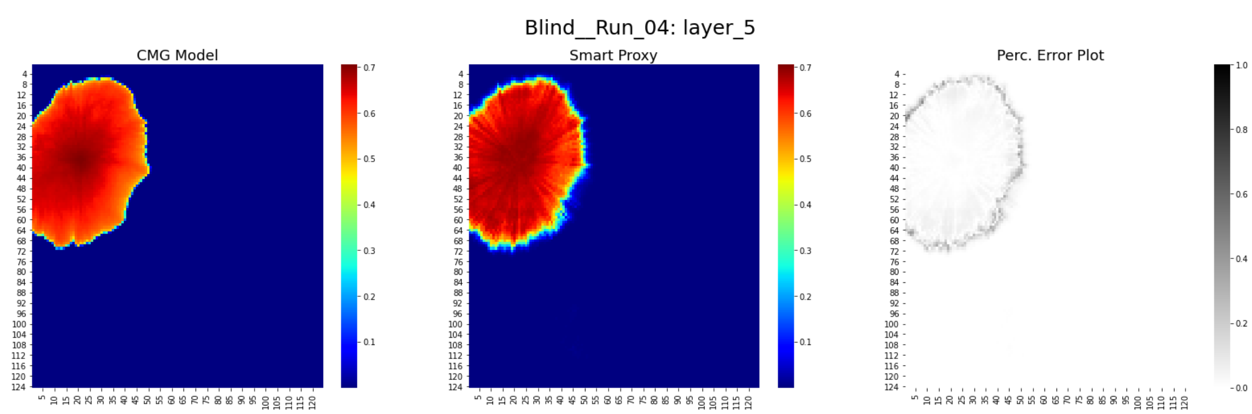




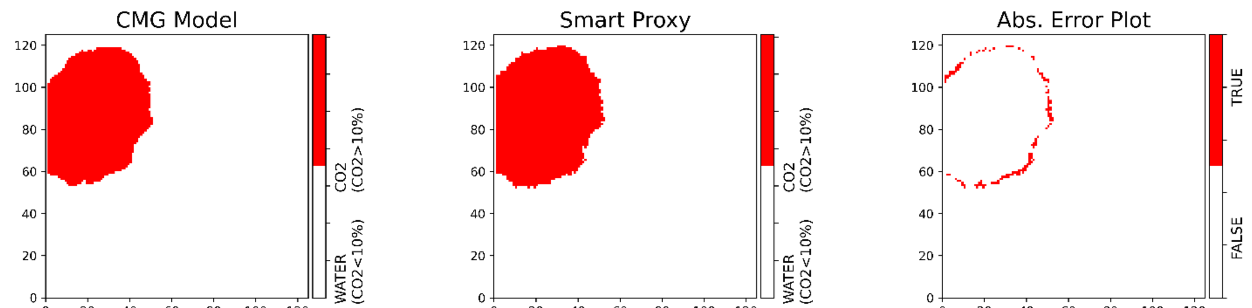


7.18 Other Phase 4 Saturation Blinds:

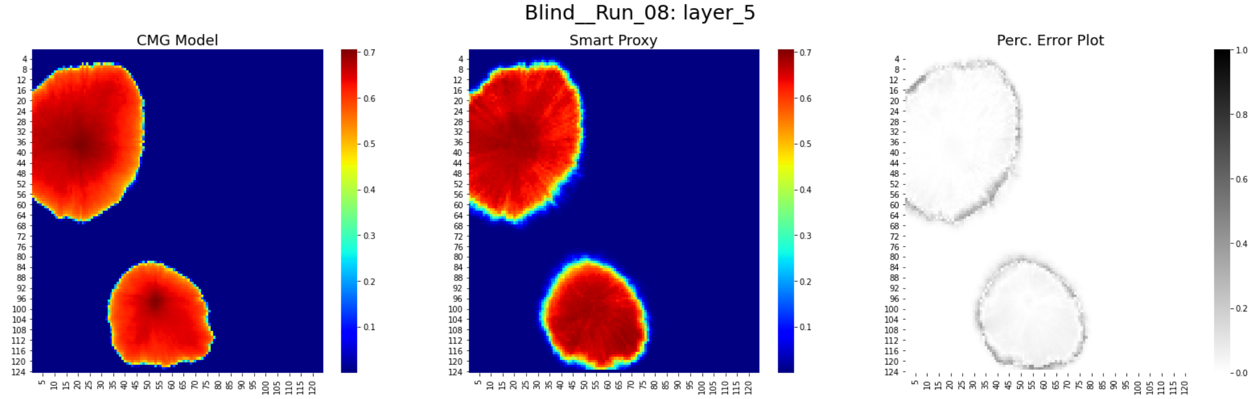




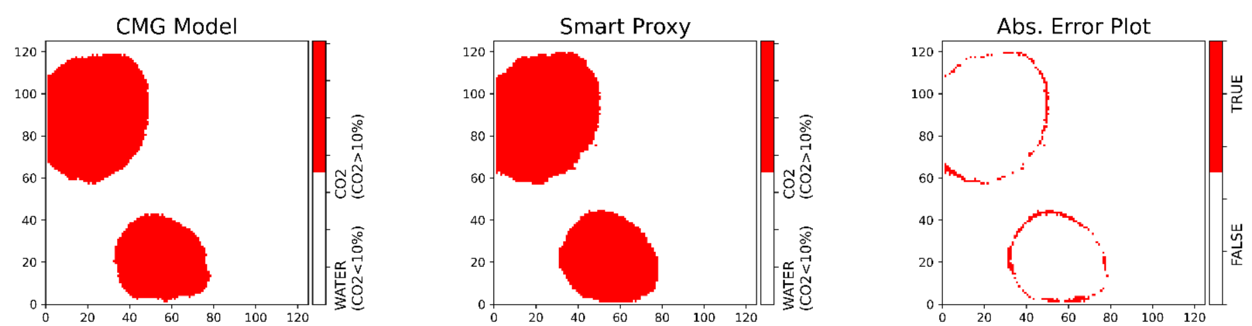
Blind__Run_04: Layer-5

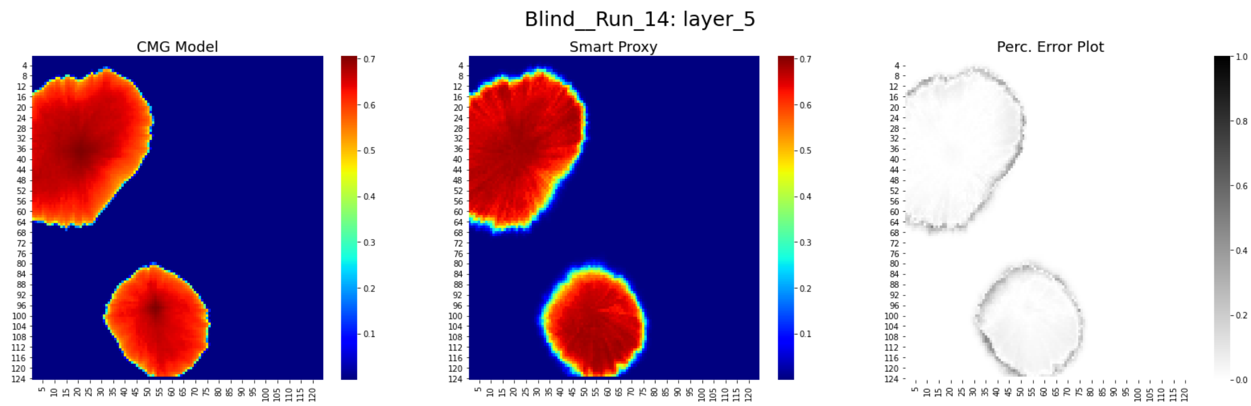


Blind__Run_08: layer_5

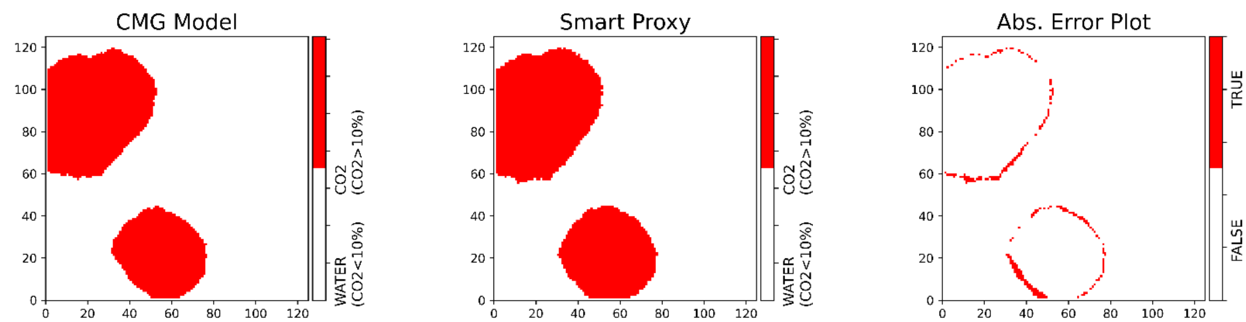


Blind__Run_08: Layer-5



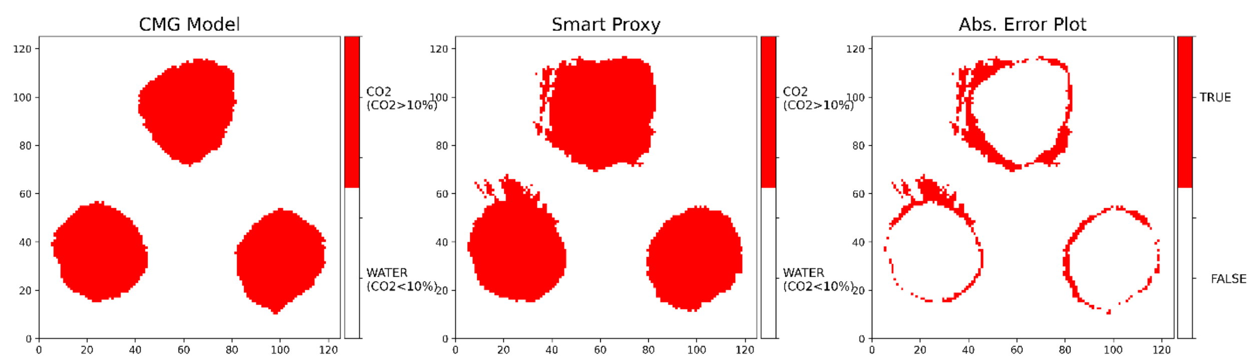


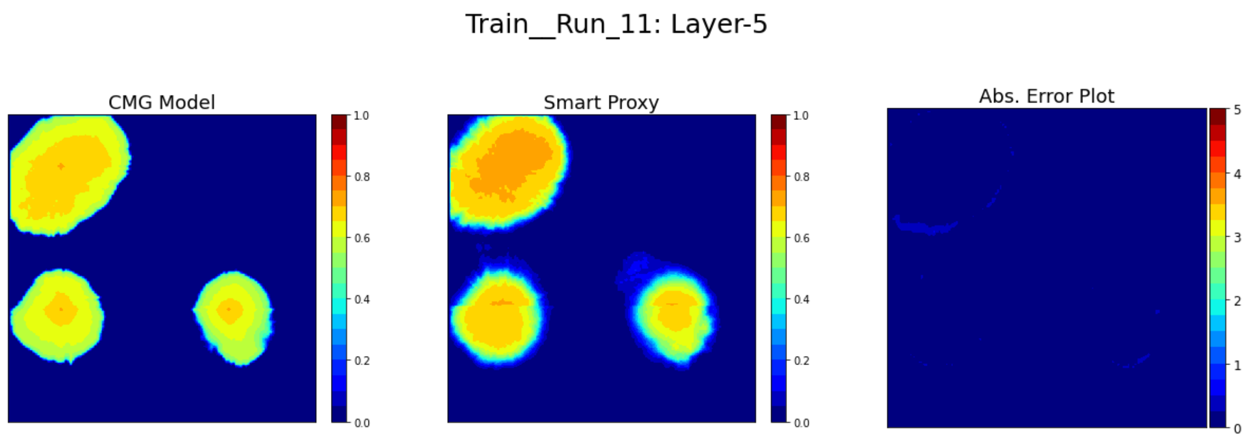
Blind__Run_14: Layer-5

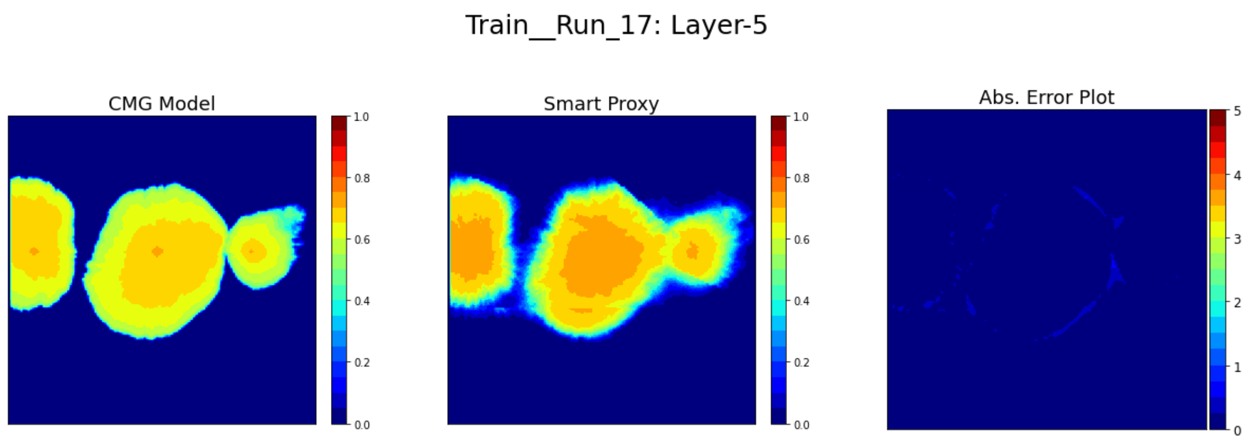


7.19 Phase 5 Saturation Trains:

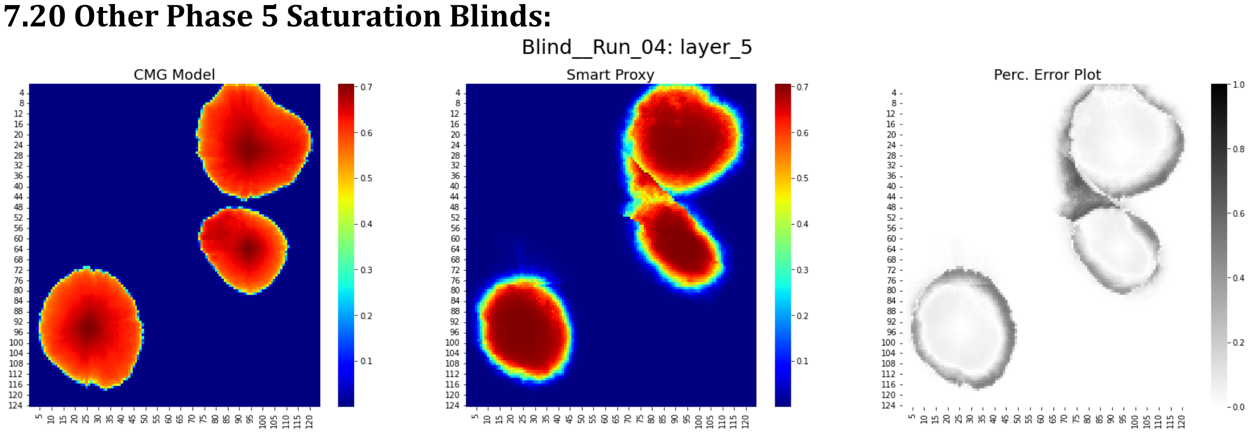
Train__Run_01: Layer-5



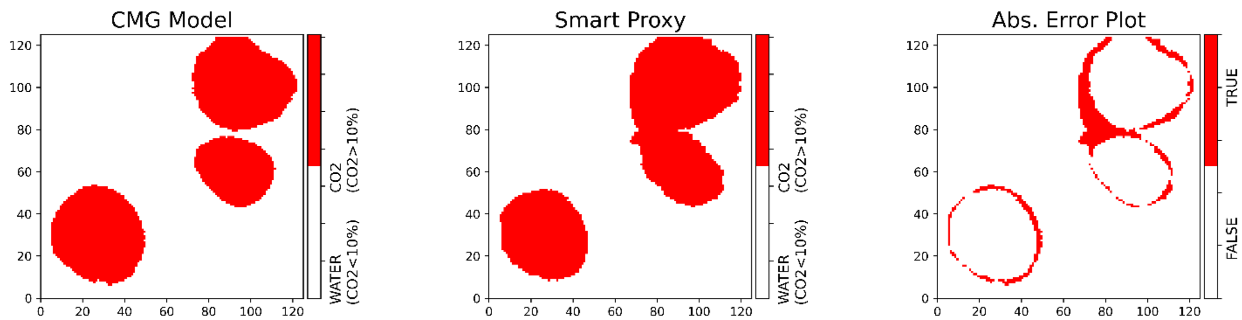


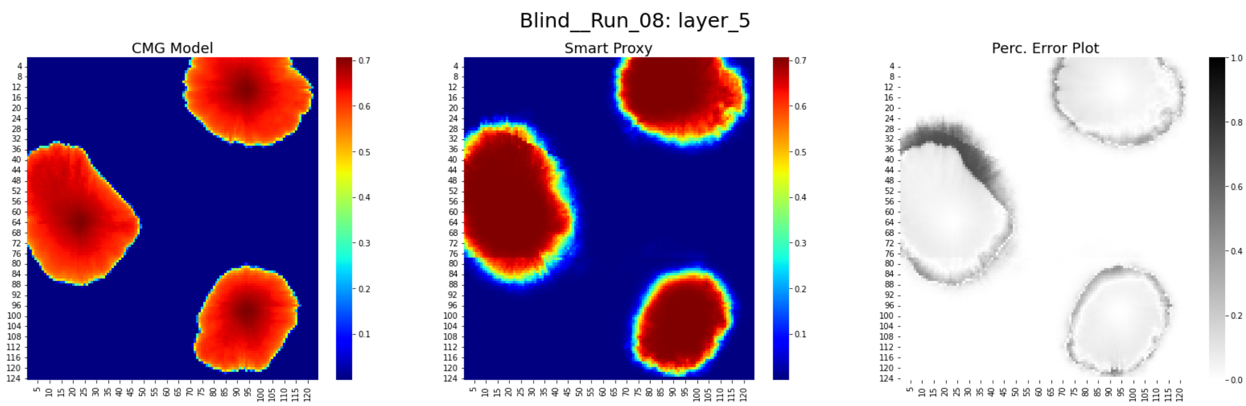


7.20 Other Phase 5 Saturation Blinds:

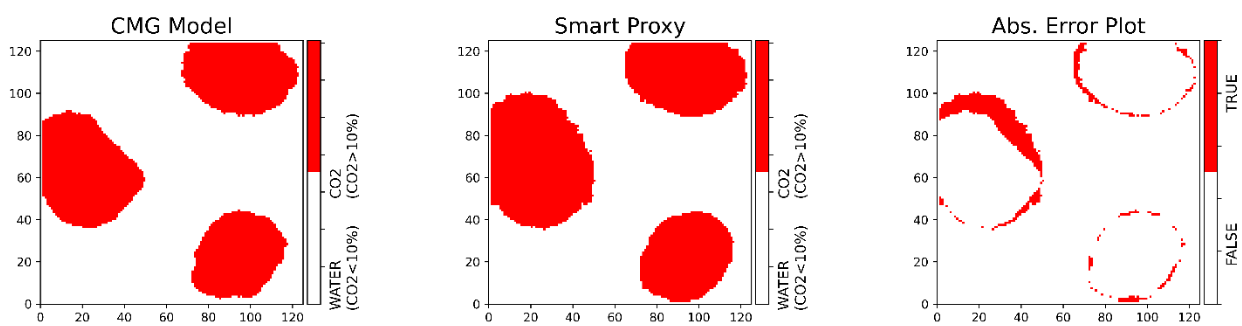


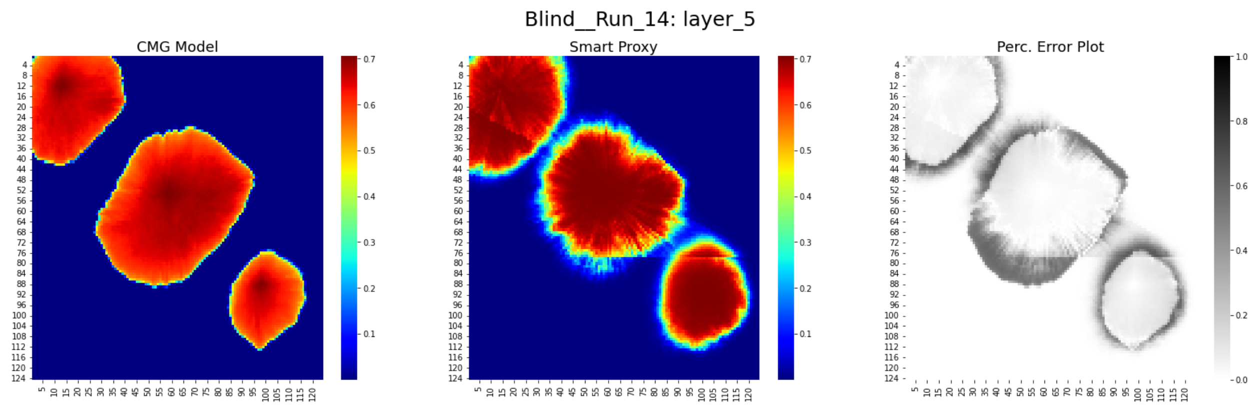
Blind__Run_04: Layer-5

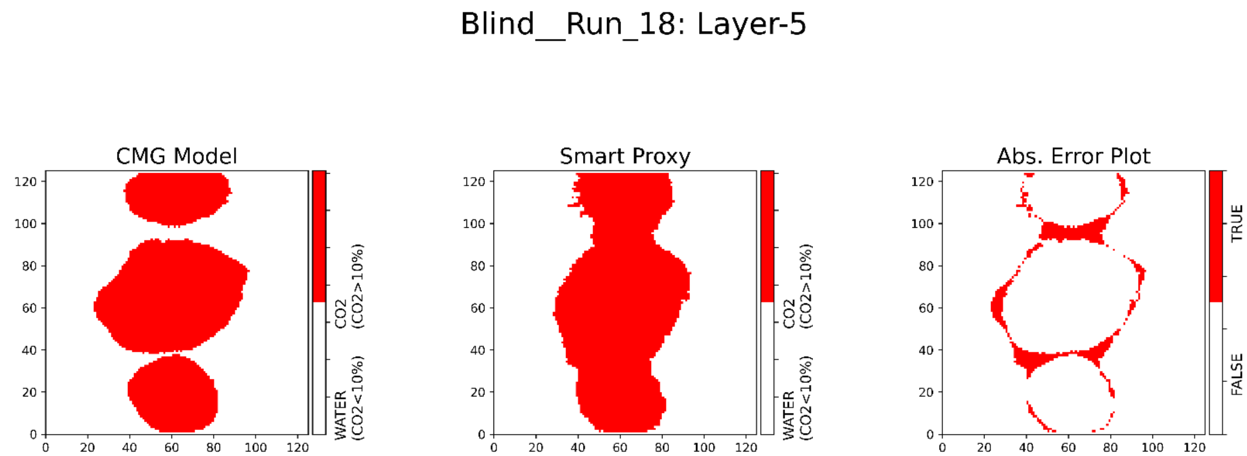


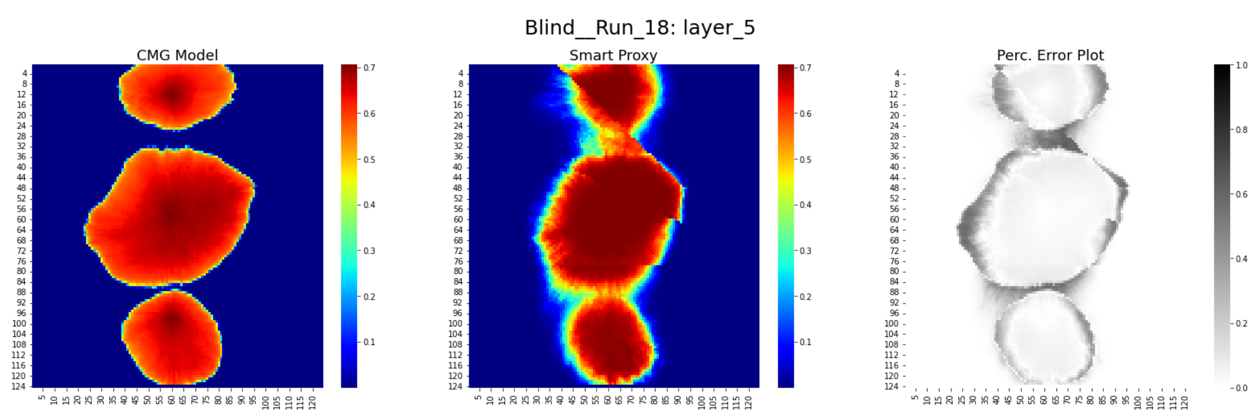


Blind__Run_08: Layer-5




Blind__Run_14: layer_5
CMG Model
Smart Proxy
Perc. Error Plot


Blind__Run_18: Layer-5
CMG Model
Smart Proxy
Abs. Error Plot
CO2
(CO2>10%)
WATER
(CO2<10%)
TRUE
FALSE


Blind__Run_18: layer_5
CMG Model
Smart Proxy
Perc. Error Plot

CHAPTER 8: REFERENCES

Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A., Arshad, H., 2018. State-of-the-art in artificial neural network applications: a survey. Heliyon 4, e00938. https://doi.org/10.1016/j.heliyon.2018.e00938

Alnuaimi, Marwan Mohammed, "APPLICATION OF ARTIFICIAL INTELLIGENCE FOR CO_2 STORAGE IN SALINE AQUIFER (SMART PROXY FOR SNAP-SHOT IN TIME)" (2022). Graduate Theses, Dissertations, and Problem Reports. 11278. https://researchrepository.wvu.edu/etd/11278

Ansari, Amir, "Reservoir Simulation of the Volve Oil field using AI-based Top-Down Modeling Approach" (2023). *Graduate Theses, Dissertations, and Problem Reports*. 11970. https://researchrepository.wvu.edu/etd/11970

Alenezi, Faisal, "A Data-Driven Smart Proxy Model for a Comprehensive Reservoir Simulation" (2017). *Graduate Theses, Dissertations, and Problem Reports*. 5055. https://researchrepository.wvu.edu/etd/5055

Aboaba, Ayodeji Luke, "Machine Learning Based Real-Time Quantification of Production from Individual Clusters in Shale Wells" (2022). *Graduate Theses, Dissertations, and Problem Reports*. 11269.
https://researchrepository.wvu.edu/etd/11269

Alabboodi, Maher Jasim, "Dynamic Data-Driven Smart Proxy Modeling For Numerical Reservoir Simulation" (2021). *Graduate Theses, Dissertations, and Problem Reports*. 10235. https://researchrepository.wvu.edu/etd/10235

Akyash, M., H. Mohammadzade, and H. Behroozi, “Dtw-merge: A novel data augmentation technique for time series classification,” arXiv preprint arXiv:2103.01119, 2021.

Ajayi, T., Gomes, J. S., & Bera, A. (2019). A review of CO_2 storage in geological formations emphasizing modeling, monitoring and capacity estimation approaches. Petroleum Science, 16, 1028-1063.

Alenezi, F.; Mohaghegh, S. A Data-Driven Smart Proxy Model for a Comprehensive Reservoir Simulation. In Proceedings of the 4th Saudi International Conference on Information Technology (Big Data Analysis) (KACSTIT), Riyadh, Saudi Arabia, 6–9 November 2016; pp. 1–6.

Alenezi, Faisal, and Shahab Mohaghegh. "Developing a smart proxy for the SACROC waterflooding numerical reservoir simulation model." SPE western regional meeting. OnePetro, 2017.

Amini, S. , Mohaghegh, S. D., Gaskari, R.. , and G. S. Bromhal. "Pattern Recognition and Data-Driven Analytics for Fast and Accurate Replication of Complex Numerical Reservoir Models at the Grid Block Level." Paper presented at the SPE Intelligent Energy Conference & Exhibition, Utrecht, The Netherlands, April 2014.

Ansari, A.; Mohaghegh, S.D.; Shahnam, M.; Dietiker, J.F. Modeling average pressure and volume fraction of a fluidized bed using data-driven smart proxy. Fluids 2019, 4, 123

Aslam, Billal, Yan, Bicheng, Tariq, Zeeshan, Krogstad, Stein, and Knut-Andreas Lie. "Rapid Simulation and Optimization of Geological CO2 Sequestration Using Coarse Grid Network Model." Paper presented at the SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition, Jakarta, Indonesia, October 2023. doi: Link

Aziz, Khalid and Antonín Settari. “Petroleum Reservoir Simulation.” (1979). DOI:10.1016/c2018-0-04535-1

Bahrami, P.; Sahari Moghaddam, F.; James, L.A. A Review of Proxy Modeling Highlighting Applications for Reservoir Engineering. Energies 2022, 15, 5247. Link

Benson, S. 2006. Monitoring Carbon Dioxide Sequestration in Deep Geological Formations for Inventory Verification and Carbon Credits. Paper SPE 102833 presented at the SPE Annual Technology Conference and Exhibition, San Antonio, Texas, 24-27 September.

Bennion, Brant, and Stefan Bachu. "Relative permeability characteristics for supercritical CO2 displacing water in a variety of potential sequestration zones." SPE Annual Technical Conference and Exhibition. OnePetro, 2005.

Benson, S. 2006. Monitoring Carbon Dioxide Sequestration in Deep Geological Formations for Inventory Verification and Carbon Credits. Paper SPE 102833 presented at the SPE Annual Technology Conference and Exhibition, San Antonio, Texas, 24-27 September.

Burkardt, John, et al. "Reduced order modeling of complex systems." (2003).

Bachu, Stefan. "CO_2 storage in geological media: Role, means, status and barriers to deployment." Progress in energy and combustion science 34.2 (2008): 254-273.

Bennion, Brant, and Stefan Bachu. "Relative permeability characteristics for supercritical CO_2 displacing water in a variety of potential sequestration zones." SPE Annual Technical Conference and Exhibition. OnePetro, 2005.

Burkardt, John, et al. "Reduced order modeling of complex systems." (2003). Cardoso, M.; Dulrofsky, L. 2009. Use of Reduced Order Modeling Procedures for Production Optimization. SPE J. 2009, 15, 426–435.

Cameron D.A. and Durlofsky L.J. 2012. Optimization of well placement, CO2 injection rates, and brine cycling for geological carbon sequestration. International Journal of Greenhouse Gas Control 10: 100 – 112.

Cardoso, M.A.. A., and L.J.. J. Durlofsky. "Use of Reduced-Order Modeling Procedures for Production Optimization." SPE J. 15 (2010): 426–435. doi:https://doi.org/10.2118/119057-PA

Celia, M. A., & Nordbotten, J. M. (2011). How simple can we make models for CO2 injection, migration, and leakage?. Energy Procedia, 4, 3857-3864.

Chadwick, R. A., Williams, G. A. & Falcon-Suarez, I. Forensic mapping of seismic velocity heterogeneity in a CO2 layer at the Sleipner CO2 storage operation, North Sea, using time-lapse seismics. Int. J. Greenh. Gas. Con. 90, 102793. Link (2019).

Chadwick, R.A. and Noy, D.J. (2015), Underground CO_2 storage: demonstrating regulatory conformance by convergence of history-matched modeled and observed CO_2 plume behavior using Sleipner time-lapse seismics. Greenhouse Gas Sci Technol, 5: 305-322. Link

Chen, Han, Hector Klie, and Qiqi Wang. "A Black-Box Stencil Interpolation Method to Accelerate Reservoir Simulations." SPE Reservoir Simulation Symposium. OnePetro, 2013.
CRS, 2022. R4490 Congressional Research Service, Link

Class H, Ebigbo A, Helmig R, Dahle HK, Nordbotten JM, Celia MA, et al. A benchmark study on problems related to CO_2 storage in geologic formations. Comput Geosci. 2009;13(4):409–34. https ://doi.org/10.1007/s10596-009-9146-x

CRS, 2022. R4490 Congressional Research Service, https://crsreports.congress.gov/product/pdf/R/R44902

Czernichowski-Lauriol, Isabelle, et al. "Geochemical interactions between CO 2, pore-waters and reservoir rocks." Advances in the geological storage of carbon dioxide. Springer, Dordrecht, 2006. 157-174.

Dijkstra, E. W.1959. A Note on Two Problems in Connexion with Graphs. Numerische Mathematik1 (1): 269–271.

Domingos, Pedro. "A few useful things to know about machine learning." Communications of the ACM 55.10 (2012): 78-87.

Dong, P., Chen, Z.-M., Liao, X.-W., Yu, W., 2022. A deep reinforcement learning (DRL) based approach for well-testing interpretation to evaluate reservoir parameters. Petrol. Sci. 19, 264e278. Link.

Doughty, C., 2010. Investigation of CO2 plume behavior for a large-scale pilot test of geologic carbon storage in a saline formation. Transp Porous Med. [Link](https://doi.org/ 10.1007/s11242-009-9396-z).

Du, H., Zhao, Z., Cheng, H., Yan, J., & He, Q. (2023). Modeling density-driven flow in porous media by physics-informed neural networks for CO_2 sequestration. Computers and Geotechnics, 159, 105433.

Ebrom, D., Buddery, D., Watts, G., and Taylor, B. (2006). Monitoring activities at In Salah. In 8th International Conference on Greenhouse Gas Control Technologies, Trondheim Norway.

IEA. 2006. International Carbon Dioxide Emissions and Carbon Intensity Database http://www.eia.doe.gov/emeu/international/carbondioxide.html.

IEA, 2016. 20 Years of Carbon Capture and Storage Accelerating Future Deployment https://iea.blob.core.windows.net

EIA, 2021. CCUS facilities in operation by application, 1980-2021, https://www.iea.org/data-and-statistics/charts/ccus-facilities-in-operation-by-application-1980-2021

Ennis-King, Jonathan and Lincoln Paterson. "Coupling of geochemical reactions and convective mixing in the long-term geological storage of carbon dioxide." International Journal of Greenhouse Gas Control 1 (2007): 86-93.

Gaganis, V.; Marinakis, D.; Samnioti, A. A soft computing method for rapid phase behavior calculations in fluid flow simulations. J. Pet. Sci. Eng. 2021, 205, 108796.

Gholami, V.; Mohaghegh, S. and Maysami, M. 2019. Smart Proxy Modeling of SACROC CO_2-EOR. Fluids 2019, 4(2), 85; https://doi.org/10.3390/fluids4020085.

Gholami, Vida, and S. Mohaghegh. "Modeling analysis of CO_2 Sequestration in Saline Formation Using Artificial Intelligence Technology." 2012 International Pittsburgh Coal Conference, Pittsburgh, PA. 2012.

Gholami, Vida, "On the optimization of CO_2-EOR process using surrogate reservoir model" (2014). Graduate Theses, Dissertations, and Problem Reports. 5668. https://researchrepository.wvu.edu/etd/5668

Gholami, Vida. " On the optimization of CO_2-EOR process using surrogate reservoir model " Order No. 3672856 West Virginia University, 2014. Ann Arbor: ProQuest. Web. 30 Nov.2021.

Gupta, D. 2020. Fundamentals of Deep Learning – Activation Functions and When to Use Them? shorturl.at/sNR03

Grujic, O., Mohaghegh, S.D., Bromhal, G. (2010). "Fast Track Reservoir Modeling of Shale Formations in the Appalachian Basin. Application to Lower Huron Shale in Eastern Kentucky." SPE 139101, Proceedings, 2010 SPE Eastern Regional Conference & Exhibition. Morgantown, West Virginia. 12-14 October 2010.

Gunter, W.D., Wong, S., Cheel, D.B., and Sjostrom G. 1998. Large CO_2 Sinks: Their Role in the Mitigation of Greenhouse Gases from an International, National (Canadian) and Provincial (Alberta) Perspective. Appl Energy 61(4):209-227.

Gunter, W.D. and Perkins, E. 2001. Geochemical Monitoring of CO_2 Enhanced Oil Recovery. Proc. of the NETL Workshop on Carbon Sequestration Science, http://www.netl.doe.gov/.

Haghighat, S. Alireza, Mohaghegh, Shahab D., Gholami, Vida , and David Moreno. "Production Analysis of a Niobrara Field Using Intelligent Top-Down Modeling." Paper presented at the SPE Western North American and Rocky Mountain Joint Meeting, Denver, Colorado, April 2014. doi: https://doi.org/10.2118/169573-MS.

Haghighat, Seyed Alireza, "Monitoring the Integrity of CO_2 Storage Sites Using Smart Field Technology" (2014). *Graduate Theses, Dissertations, and Problem Reports*. 5732. https://researchrepository.wvu.edu/etd/5732

Haghighat, S. A., et al. "Reservoir Simulation of CO_2 Sequestration in Deep Saline Reservoir, Citronelle Dome, USA." Twelfth Annual Conference on Carbon Capture, Utilization and Sequestration. 2013

He, J., Sætrom, J., Durlofsky, L.J., 2011a. Enhanced linearized reduced-order models for subsurface flow simulation. Journal of Computational Physics 230, 8313–8341.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Network. 2, 359e366. https://doi.org/10.1016/ 0893-6080(89)90020-8.

Hovorka, S. D., Benson, S. M., and Doughty, C. K. 2006. Measuring Permanence of CO_2 Storage in Saline Formations‒the Frio Experiment. Geosciences 13 (2): 105-121.

Haghighat, S. A., et al. "Reservoir Simulation of CO_2 Sequestration in Deep Saline Reservoir, Citronelle Dome, USA." Twelfth Annual Conference on Carbon Capture, Utilization and Sequestration. 2013.

Haghighat, S. A., Mohaghegh, S. D., Borzouie, N., Moreno, D., Shahkarami, A., & A. (2013, May). Reservoir Simulation of CO_2 Sequestration in Deep Saline Reservoir, Citronelle Dome, USA. In Twelfth Annual Conference on Carbon Capture, Utilization and Sequestration.

Haghighat, S. Alireza, and Shahab D. Mohaghegh. "Interpretation of real-time pressure measurements to detect CO_2 leakage." SPE Annual Technical Conference and Exhibition. OnePetro, 2015.

Han, J.-X., L. Xue, Y.-S. Wei et al. (2023). Physics-informed neural network-based petroleum reservoir simulation with sparse data using domain decomposition. Petroleum Science, https://doi.org/10.1016/j.petsci.2023.10.019.

Haykin, S. 2009. Neural Networks and Learning Machines, 3rd edn. Englewood Cliffs, NJ: Prentice Hall. He, Jincong. Reduced-order modeling for oil-water and compositional systems, with application to data assimilation and production optimization. Stanford University, 2013.

Islam M.R., Hossain M., Moussavizadegan S., Mustafiz S., Abou-Kassem J. Advanced petroleum reservoir simulation: Towards developing reservoir emulators, (2nd ed.), Scrivener Publishing & WILEY, Salem, Massachusetts (2016).

IPCC. 2008. Special Report on Carbon Dioxide Capture and Storage, Metz, B., Davidson, O., de Coninck, H., Loos, M., Meyer, L.A., editors. Cambridge University Press: New York.

Jalali, Jalal, and Shahab D. Mohaghegh. "Reservoir Simulation and Uncertainty Analysis of Enhanced CBM Production Using Artificial Neural Networks." Paper presented at the SPE Eastern Regional Meeting, Charleston, West Virginia, USA, September 2009. doi: https://doi.org/10.2118/125959-MS.

Jiang X. A review of physical modelling and numerical simulation of long-term geological storage of CO_2. Appl Energy. 2011;88(11):3557–66. https://doi.org/10.1016/j.apene rgy.2011.05.004.

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L., 2021. Physics-informed machine learning. Nat. Rev. Phys. 3, 422e440. https://doi.org/ 10.1038/s42254-021-00314-5.

Khazaeni, Y. and Mohaghegh, S.D. (2010). "Intelligent Time Successive Production Modeling." SPE 132643, Proceedings, SPE Western Regional Conference & Exhibition. Anaheim, California. 27-29 May 2010.

Kim, Ki-Joong, Jeffrey T. Culp, James E. Ellis, and Matthew D. Reeder, 2022. Real-Time Monitoring of Gas-Phase and Dissolved CO_2 Using a Mixed-Matrix Composite Integrated Fiber Optic Sensor for Carbon Storage Application. Environmental Science & Technology 2022 56 (15), 10891-10903. DOI: 10.1021/acs.est.2c02723

Kikuta, K., Hongo, S., Tanase, D. and Ohsumi, T. 2005. Field Test of CO_2 Injection in Nagaoka, Japan. Proc. of the 7th International Conference on Greenhouse Gas Control Technologies, Oxford 2: 1367-1372.

Kalantari-Dahaghi, A., and S. D Mohaghegh. "Numerical Simulation and Multiple Realizations for Sensitivity Study of Shale Gas Reservoir." Paper presented at the SPE Production and Operations Symposium, Oklahoma City, Oklahoma, USA, March 2011. doi: https://doi.org/10.2118/141058-MS.

Kalantari-Dahaghi, A., and Shahab D. Mohaghegh. "Numerical simulation and multiple realizations for sensitivity study of shale gas reservoir." SPE Production and operations symposium. OnePetro, 2011.

Kalantari-Dahaghi, A.M. and Mohagegh, S.D. (2009). "Top-Down Intelligent Reservoir Modeling of New Albany Shale." SPE 125859, Proceedings, 2009 SPE Eastern Regional Conference & Exhibition. Charleston, West Virginia.

Kalantari-Dahaghi, A.M., Mohaghegh, S.D., and Khazaeni, Y. (2010). "New Insight into Integrated Reservoir Management using Top-Down, Intelligent Reservoir Modeling Technique; Application to a Giant and Complex Oil Field in the Middle East." SPE 132621, Proceedings, 2010 SPE Western Regional Conference & Exhibition. Anaheim, California. 27-29 May 2010.

Khuri, André I., and Siuli Mukhopadhyay. "Response surface methodology." Wiley Interdisciplinary Reviews: Computational Statistics 2.2 (2010): 128-149.

Khoshkhahtinat, A., Zafari, A., Mehta, P. M., Akyash, M., Kashiani, H., & Nasrabadi, N. M. (2023). Multi-Context Dual Hyper-Prior Neural Image Compression. arXiv preprint arXiv:2309.10799.

Kumar, A., M. Noh, G.A. Pope, K. Sepehrnoori, S. Bryant, L.W. Lake Reservoir simulation of CO_2 storage in deep saline aquifers Proceedings of SPE/DOE symposium on improved oil recovery, Tulsa, Oklahoma, April 17–21 (2004)

Kohshour, I. O., Ahmadi, M., and C.. Hanks. "Uncertainty Assessment in Geologic Modeling and Sensitivity Analysis of Static and Dynamic Models in Umiat: A Frozen Shallow Oil Accumulation in National Petroleum Reserve of Alaska." Paper presented at the SPE Western Regional & AAPG Pacific Section Meeting 2013 Joint Technical Conference, Monterey, California, USA, April 2013. doi: https://doi.org/10.2118/165341-MS

Law DH-S, van der Meer L, Gunter W. Comparison of numerical simulators for greenhouse gas storage in coalbeds, Part IV; History match of feld micro-pilot test data. In: The 7th international conference on greenhouse gas control technologies, Vancouver BC, Canada, September 5–9; 2004. p. 2239–42

Lee, Y.S. 2009. A simulation study on the analysis of CO_2-water flow for CO_2 sequestration in deep saline aquifer Ph.D. thesis Hanyang University, Seoul (2009), p. 95

Luke Aboaba, A. 2022. “Machine Learning Based Real-Time Quantification of Production from Individual Clusters in Shale Wells” (2022). Graduate Theses, Dissertations, and Problem Reports. 10235.

M. Hesse, H.A. Tchelepi, F.M. Orr Scaling analysis of the migration of CO_2 in saline aquifers Proceeding SPE annual technical conference and exhibition, San Antonio, Texas, September 24–27 (2006)

Maroto-Valer, M. Mercedes, ed. Developments and innovation in carbon dioxide (CO_2) capture and storage technology: Carbon dioxide (CO_2) storage and utilisation. Elsevier, 2010.

Miles, N., Davis, K., and Wyngaard, J. 2005. Detecting Leaks from CO_2 Reservoirs Using Micrometeorological Methods, Carbon Dioxide Capture for Storage in Deep Geologic Formations. Results from the CO_2 Capture Project, Volume 2: Geologic Storage of Carbon Dioxide with Monitoring and Verification, S.M. Benson, editor. Elsevier Science: London. 1031-1044.

Mirzabozorg, Arash. "Incorporation of Engineering Knowledge in History Matching, Optimization, and Uncertainty Assessment Frameworks with Application to the SAGD Process". PhD thesis, University of Calgary, 2015.

Mawalkar, Sanjay, Gupta, Neeraj, & Burchwell, Andrew. *Distributed Temperature Sensing (DTS) to Monitor CO2 Migration in an Enhanced Oil Recover Field in Northern Michigan*. United States. https://doi.org/10.18141/1755349

Michael, K. et al. Geological storage of CO_2 in saline aquifers: A review of the experience from existing storage operations. Int. J. Greenh. Gas. Con. 4, 659–667. https://doi.org/10.1016/j.ijggc.2009.12.011 (2010).

Middleton, R. S., Keating, G. N., Viswanathan, H. S., Stauffer, P. H., & Pawar, R. J. (2012). Effects of geologic reservoir uncertainty on CO2 transport and storage infrastructure. International Journal of Greenhouse Gas Control, 8, 132-142.

Mohaghegh, S.D. (2024). Artificial Intelligence for Science and Engineering Applications (1st ed.). CRC Press. https://doi.org/10.1201/9781003369356

Mohaghegh, S. D., and R. Gaskari. "An intelligent system's approach for revitalization of brown fields using only production rate data." (2009): 89-106.

Mohaghegh, S. D., et al. "Smart proxy: an innovative reservoir management tool; Case study of a giant mature oilfield in the UAE." Abu Dhabi International Petroleum Exhibition and Conference. OnePetro, 2015.

Mohaghegh, S. D., Gaskari, R.. Maysami, M.. and Y. Khazaeni. "Data-Driven Reservoir Management of a Giant Mature Oilfield in the Middle East." Paper presented at the SPE Annual Technical Conference and Exhibition, Amsterdam, The Netherlands, October 2014. doi: https://doi.org/10.2118/170660-MS.

Mohaghegh, Shahab D., Amini, Shohreh , Gholami, Vida , Gaskari, Razi , and Grant Bromhal. "Grid-Based Surrogate Reservoir Modeling (SRM) for Fast Track Analysis of Numerical Reservoir Simulation Models at the Grid block Level." Paper presented at the SPE Western Regional Meeting, Bakersfield, California, USA, March 2012. doi: https://doi.org/10.2118/153844-MS.

Mohaghegh, Shahab, Shohreh Amini, Vida Gholami, Razi Gaskari, Grant S Bromhal, et al. Grid-based surrogate reservoir modeling (srm) for fast track analysis of numerical reservoir simulation models at the gridblock level. In SPE Western Regional Meeting. Society of Petroleum Engineers, 2012.

Mohaghegh, S., Abdulla, F., Abdou, M., and Gaskari, R. 2015. Smart Proxy: An Innovative Reservoir Management Tool; Case Study of a Giant Mature Oilfield in the UAE. 2015.

Mohaghegh, S. Smart proxy modeling for numerical reservoir simulations - big data analytic in e and p, 2015.

Mohaghegh, Shahab D "Top-Down Intelligent Reservoir Modeling (Tdirm); A New Approach In Reservoir Modeling By Integrating Classic Reservoir Engineering With Artificial Intelligence & Data Mining Techniques".AAPG,2009

Mohaghegh, Shahab D. "Converting detail reservoir simulation models into effective reservoir management tools using SRMs; case study–three green fields in Saudi Arabia."International Journal of Oil, Gas and Coal Technology 7.2 (2014): 115-131.

Mohaghegh, Shahab D. "Quantifying uncertainties associated with reservoir simulation studies using a surrogate reservoir model." SPE Annual Technical Conference and Exhibition. OnePetro, 2006.

Mohaghegh, Shahab D. "Reservoir simulation and modeling based on pattern recognition." SPE Digital Energy Conference and Exhibition. OnePetro, 2011.

Mohaghegh, Shahab D., et al. "Application of surrogate reservoir model (SRM) to an onshore green field in Saudi Arabia; case study." North Africa Technical Conference and Exhibition. OnePetro, 2012.

Mohaghegh, Shahab D., et al. "Application of well-based surrogate reservoir models (SRMs) to two offshore fields in Saudi Arabia, case study." SPE Western Regional Meeting. OnePetro, 2012.

Mohaghegh, Shahab D., et al. "Development of Surrogate Reservoir Model (SRM) for fast track analysis of a complex reservoir." International Journal of Oil, Gas and Coal Technology 2.1 (2009): 2-23.

Mohaghegh, Shahab D., et al. "Development of Surrogate Reservoir Model (SRM) for fast track analysis of a complex reservoir." International Journal of Oil, Gas and Coal Technology 2.1 (2009): 2-23.

Mohaghegh, Shahab D., et al. "Grid-Based Surrogate Reservoir Modeling (SRM) for Fast Track Analysis of Numerical Reservoir Simulation Models at the Grid block Level." SPE Western Regional Meeting. OnePetro, 2012.

Mohaghegh, Shahab D., et al. "Uncertainty analysis of a giant oil field in the middle east using surrogate reservoir model." Abu Dhabi International Petroleum Exhibition and Conference. OnePetro, 2006.

Mohaghegh, Shahab Dean. "Reservoir simulation and modeling based on artificial intelligence and data mining (AI&DM)." Journal of Natural Gas Science and Engineering 3.6 (2011): 697-705.

Mohaghegh, Shahab. Data-driven Analytics for the Geological Storage of CO_2. CRC Press, 2018.

Mohaghegh, S.D. (2022). Smart Proxy Modeling: Artificial Intelligence and Machine Learning in Numerical Simulation (1st ed.). CRC Press. https://doi.org/10.1201/9781003242581

Mohaghegh, S. D. AI-Based Simulation: An Alternative to Numerical Simulation and Modeling. Intelligent Solutions, Inc. and West Virginia University: https://www.academia.edu/102748513/AI_Based_Simulation_An_Alternative_to_Numerical_Simulation_and_Modeling

Mohaghegh, Shahab, 2018. Data-Driven Analytics for the Geological Storage of CO_2Data-Driven Analytics for the Geological Storage of CO_2. First edition, CRC Press, DOI: https://doi.org/10.1201/b21913.

Mohaghegh, S.D. (2010). "Top-Down, Intelligent Reservoir Model." European Geological Union General Assembly (Vol. 12, EGU2010-233, 2010). Vienna, Austria. May 2010.

Mohaghegh, S.D., Liu, J, Gaskari, R., Maysami, M., and Olukoko, O. (2012). "Application of Surrogate Reservoir Model (SRM) to an Onshore Green Field in Saudi Arabia; Case Study." SPE 151994, Proceedings, 2012 the SPE North Africa Technical Conference and Exhibition held in Cairo, Egypt, 20 - 22 February 2012.

Mohaghegh S (2000a) Virtual-intelligence applications in petroleum engineering: part 1—artificial neural networks. J Petrol Technol 52(09):64–73

Mohaghegh S (2000b) Virtual-intelligence applications in petroleum engineering: part 2—evolutionary computing. J Petrol Technol 52(10): 40–46

Mohaghegh S (2000c) Virtual-intelligence applications in petroleum engineering: part 3—fuzzy logic. J Petrol Technol 52(11):82–87

Mohaghegh S (2005) Recent developments in application of artificial intelligence in petroleum engineering. J Petrol Technol 57(04):86– 91

Mohaghegh S, Ameri S (1995). Artificial neural network as a valuable tool for petroleum engineers. paper SPE, 29220.

Mohaghegh S, Popa A, & Ameri S. (1999). Intelligent systems can design optimum fracturing jobs. Paper presented at the SPE Eastern Regional Conference and Exhibition.

Mohaghegh S, Popa A, Ameri S (2000) Design optimum frac jobs using virtual intelligence techniques. Comp Geosci 26(8):927–939

Mohaghegh, S., Platon, V., and S. Ameri. "Candidate Selection for Stimulation of Gas Storage Wells Using Available Data With Neural Networks and Genetic Algorithms." Paper presented at the SPE Eastern Regional Meeting, Pittsburgh, Pennsylvania, November 1998. doi: https://doi.org/10.2118/51080-MS

Mohaghegh, Shahab, Arefi, Reza, Ameri, Samuel, and M. H. Hefner. "A Methodological Approach for Reservoir Heterogeneity Characterization Using Artificial Neural Networks." Paper presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, September 1994. doi: https://doi.org/10.2118/28394-MS

Mohaghegh, Shahab, Reeves, Scott, and David Hill. "Development of an Intelligent Systems Approach for Restimulation Candidate Selection." Paper presented at the SPE/CERI Gas Technology Symposium, Calgary, Alberta, Canada, April 2000. doi: https://doi.org/10.2118/59767-MS

Møll Nilsen, H., Lie, K.-A., and Andersen, O. (2015). Analysis of CO2 Trapping Capacities and Long-Term Migration for Geological Formations in the Norwegian North Sea Using MRST-Co2lab. *Comput. Geosciences* 79, 15–26. doi:10.1016/j.cageo.2015.03.001

Meng, X., Li, Z., Zhang, D., & Karniadakis, G. E. (2020). PPINN: Parareal physics-informed neural network for time-dependent PDEs. Computer Methods in Applied Mechanics and Engineering, 370, 113250.

Musayev, K., Shin, H., Nguyen-Le, V., 2023. Optimization of CO_2 injection and brine production well placement using a genetic algorithm and artificial neural network-based proxy model. Int. J. Greenh. Gas Control127, 103915. https://doi.org/10.1016/j.ijggc.2023.103915

NETL, 2015. Southeast Regional Carbon Sequestration Partnership Citronelle Project Carbon Storage R&D Project, https://www.netl.doe.gov/sites/default/files/event-proceedings/2015/carbon%20storage/proceedings/08-18_08_SECARB_Citronelle_Hill_Thomas_Trautz.pdf

Nghiem, Long, et al. "Modeling CO_2 storage in aquifers with a fully-coupled geochemical EOS compositional simulator." SPE/DOE symposium on improved oil recovery. OnePetro, 2004.

Nordbotten, J.M., and M.A. Celia. 2006b. An improved analytical solution for interface upconing around a well. Water Resources Research 42, W08433. doi: 10.1029/2005WR004738.

Nordbotten J., Flemisch B., Gasda S., Nilsen H., Fan Y., Pickup G., Wiese B., Celia M., Dahle , Eigestad G., Pruess K. Uncertainties in practical simulation of CO_2 storage Int. J. Greenh. Gas Control, 9 (2012), pp. 234-242

Patel, Priyanka and Amit Thakkar. “The upsurge of deep learning for computer vision applications.” *International Journal of Electrical and Computer Engineering* 10 (2020): 538-548.

Pasala, S.M., C.B. Forster, S.J. Lim, M.D. Deo Simulating the impact of faults on CO_2 sequestration and enhanced oil recovery in sandstone aquifers Proceedings of society of petroleum engineer annual technical conference and exhibition, Denver, Colorado, October 5–8 (2003)

Pruess, K., & Spycher, N. (2007). ECO2N–A fluid property module for the TOUGH2 code for studies of CO2 storage in saline aquifers. Energy conversion and management, 48(6), 1761-1767.

Quintero, Gabriel, "Quantitative Analysis of Rate Transient Analysis in Unconventional Shale Gas Reserviors" (2022). *Graduate Theses, Dissertations, and Problem Reports*. 11225. https://researchrepository.wvu.edu/etd/11225

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686-707.

Rezk, M. G., & Foroozesh, J. (2019). Study of convective-diffusive flow during CO2 sequestration in fractured heterogeneous saline aquifers. Journal of Natural Gas Science and Engineering, 69, 102926.

Shanmuganathan, S., 2016. Artificial neural network modelling: an introduction. In: Shanmuganathan, S., Samarasinghe, S. (Eds.), Artificial Neural Network Modelling. Springer International Publishing, Cham, pp. 1e14.

Shahkarami, Alireza, et al. "Artificial intelligence (AI) assisted history matching." SPE western North American and Rocky Mountain joint meeting. OnePetro, 2014.

Shahkarami, Alireza, et al. "Modeling pressure and saturation distribution in a CO_2 storage project using a Surrogate Reservoir Model (SRM)." Greenhouse Gases: Science and Technology 4.3 (2014): 289-315.

Sattari, A.M. 2023. "Comparative Analysis of Artificial Intelligence and Numerical Reservoir Simulation in Marcellus Shale Wells" (2023). Graduate Theses, Dissertations, and Problem Reports. 10235.

Shahkarami, Alireza, et al. "Artificial intelligence (AI) assisted history matching." SPE western North American and Rocky Mountain joint meeting. OnePetro, 2014.

Shahkarami, Alireza , Mohaghegh, Shahab D., Gholami, Vida , and Grant Bromhal. "Application of Artificial Intelligence and Data Mining Techniques for Fast Track Modeling of Geologic Sequestration of Carbon Dioxide – Case Study of SACROC Unit." Paper presented at the SPE Digital Energy Conference and Exhibition, The Woodlands, Texas, USA, March 2015. doi: https://doi.org/SPE-173406-MS.

Shokouhi P, Kumar V, Prathipati S, Hosseini SA, Giles CL, Kifer D. Physics-informed deep learning for prediction of CO_2 storage site response. J Contam Hydrol 2021;241:103835. https://doi.org/10.1016/j.jconhyd.2021.103835.

Teletzke, G. F., & Lu, P. (2013). Guidelines for reservoir modeling of geologic CO2 storage. Energy Procedia, 37, 3936-3944.

Thanh, V., H., Sugai, Y. & Sasaki, K. Application of artificial neural network for predicting the performance of CO_2 enhanced oil recovery and storage in residual oil zones. Sci Rep 10, 18204 (2020). https://doi.org/10.1038/s41598-020-73931-2

Underhill, J. R., and R. Stoneley. 1998 Introduction to the development, evolution and petroleum geology of the Wessex Basin. In Development, evolution and petroleum geology of the Wessex Basin, ed. J. R. Underhill, special publication 133, pp. 1–18. London: The Geological Society.

Urych, T., and Smoliński, A. (2019). Numerical Modeling of CO2 Migration in Saline Aquifers of Selected Areas in the Upper Silesian Coal Basin in Poland. *Energies* 12, 3093. doi:10.3390/en12163093

van Doren, J.F.M., Markovinovic´, R., Jansen, J.D., 2006. Reduced-order optimal control of water flooding using proper orthogonal decomposition. Computational Geosciences 10, 137–158.

Voskov, D.V.; Tchelepi, H. Comparison of nonlinear formulations for two-phase multi-component EoS based simulation. J. Pet. Sci. Eng. 2012, 82–83, 101–111.

Wang, S., & Perdikaris, P. (2023). Long-time integration of parametric evolution equations with physics-informed deeponets. Journal of Computational Physics, 475, 111855.

Wang, S., Wang, H., & Perdikaris, P. (2021). On the eigenvector bias of Fourier feature networks: From regression to solving multi-scale PDEs with physics-informed neural networks. Computer Methods in Applied Mechanics and Engineering, 384, 113938.

Wang, P.; Stenby, E.H. Compositional simulation of reservoir performance by a reduced thermodynamic model. Comput. Chem. Eng. 1994, 18, 75–81.

Wang, Y.D., Chung, T., Armstrong, R.T., Mostaghimi, P., 2021. ML-LBM: predicting and accelerating steady state flow simulation in porous media with convolutional neural networks. Transp Porous Med. 138, 49e75. https://doi.org/ 10.1007/s11242-021-01590-6

Wen G., Benson S.M. 2019. CO_2 plume migration and dissolution in layered reservoirs (2019) International Journal of Greenhouse Gas Control, 87, pp. 66-79.

White, A. C., Molnar, D., Aminian, K., Mohaghegh, S., Ameri, S., and P. Esposito. "The Application of ANN for Zone Identification in a Complex Reservoir." Paper presented at the SPE Eastern Regional Meeting, Morgantown, West Virginia, September 1995. doi: https://doi.org/10.2118/30977-MS

Wilson, M. and Monea, M. 2004. IEA GHG Weyburn CO_2 Monitoring and Storage Operation Summary Report 2000 - 2004. Petroleum Technology Research Centre, Regina.

Wilson, K. C., and Louis J. Durlofsky. "Computational optimization of shale resource development using reduced-physics surrogate models." SPE Western Regional Meeting. OnePetro, 2012.

Wu, Xiaohui, et al. "Method and System for Rapid Model Evaluation Using Multilevel Surrogates." U.S. Patent Application No. 13/392,035.

Wu, Xiaohui, et al. "Method and System for Rapid Model Evaluation Using Multilevel Surrogates." U.S. Patent Application No. 13/392,035.

Yan, B., Li, C., Tariq, Z., & Zhang, K. (2023). Estimation of heterogeneous permeability using pressure derivative data through an inversion neural network inspired by the fast marching method. Geoenergy Science and Engineering, 211982.

Zargari, S. and Mohaghegh, S.D. (2010). "Field Development Strategies for Bakken Shale Formation." SPE 139032, Proceedings, SPE Eastern Regional Conference & Exhibition. Morgantown, West Virginia. 12-14 October 2010.

Yang, K., & Karniadakis, G. E. (2020). Physics-informed neural networks for multi-fidelity modeling. Journal of Computational Physics, 402, 109009.

Zhang, Y., & Karniadakis, G. E. (2020). Physics-informed neural networks for high-dimensional surrogate modeling and uncertainty quantification without full-field simulations. Journal of Computational Physics, 402, 109009.

Wang, J., Zhang, Y., & Karniadakis, G. E. (2021). Physics-informed neural networks for high-dimensional partial differential equations. Journal of Computational Physics, 429, 110010.

Zhi Zhong, Alexander Y. Sun, Hoonyoung Jeong, 2019. Predicting CO_2 Plume Migration in Heterogeneous Formations Using Conditional Deep Convolutional Generative Adversarial Network, Water Resources Research, Volume:55, Issue:7, Page(s):5830-5851, John Wiley and Sons, Ltd 2019, doi:https://doi.org/1029/2018WR024592

Zhang, Y. Modeling Uncertainty and Risk in Carbon Capture and Storage. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2013a

Zhang, Y.; Sahinidis, N. Developing Surrogate Models for CO_2 Sequestration Using Polynomial Chaos Expansion; NRAP-TRS-III-004- 2013c; NRAP Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Morgantown, WV, USA, 2013; p. 20.

Zhang, Yan, and Nikolaos V. Sahinidis. "Uncertainty quantification in CO_2 sequestration using surrogate models from polynomial chaos expansion." Industrial & Engineering Chemistry Research 52.9 (2013): 3121-3132.

Zafari, A., Khoshkhahtinat, A., Mehta, P. M., Nasrabadi, N. M., Thompson, B. J., Kirk, M. S., & da Silva, D. (2023). Multi-spectral Entropy Constrained Neural Compression of Solar Imagery. arXiv preprint arXiv:2309.10791.

Zhang, Y.; Pau, G. Reduced-Order Model Development for CO_2 Storage in Brine Reservoirs; NRAP-TRS-III-005-2012; NRAP Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Morgantown, WV, 2012; p 20. DOI: 10.18141/1432691.

Zhang, Y.; Sahinidis, N. Uncertainty Quantification in CO_2 Sequestration Using Surrogate Models from Polynomial Chaos Expansion. Ind. Eng. Chem. Res. 2013b, 52, 3121–3132.

Zhang, Y.; Sahinidis, N. Developing Surrogate Models for CO_2 Sequestration Using Polynomial Chaos Expansion; NRAP-TRS-III-004- 2013; NRAP Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Morgantown, WV, USA, 2013c; p. 20.

Zhu, T., Ajo-Franklin, J., Daley, T. M. & Marone, C. Dynamics of geologic CO2 storage and plume motion revealed by seismic coda waves. Proc. Acad. Nat. Sci. 116, 2464–2469. https://doi.org/10.1073/pnas.1810903116 (2019).