Downloaded 07/30/24 to 65.132.165.43. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/tle42110730.1

PRESIDENT'S PAGE

Digital twins in the era of generative AI

This month's author:
Felix J. Herrmann
Georgia Institute of Technology,
Seismic Laboratory for Imaging and Modeling

The industry is experiencing significant changes due to artificial intelligence (AI) and the challenges of the energy transition. While some view these changes as threats, recent advances in AI offer unique opportunities, especially in the context of "digital twins" for subsurface monitoring and control. IBM defines a digital twin as "a virtual representation of an object or system that spans its lifecycle, is updated from real-time data, and uses simulation, machine learning and reasoning to help decision-making." In this column, I will explore these concepts and their significance in addressing the challenges of underground monitoring and control, which are vital for cost-effective risk management and optimized underground resource production and storage. Furthermore, I hope to illustrate how digital twins serve as a platform to integrate the seemingly disparate and siloed fields of geophysics and reservoir engineering.

Digital twins are commonly used in manufacturing, healthcare, and environmental monitoring, but their application in subsurface resource management is still developing. To achieve sustainability goals in our industry, we must create digital twins that can confidently make decisions based on diverse monitoring data such as time-lapse well and seismic data. Subsurface complexities and reservoir heterogeneity demand a systematic approach to quantify uncertainty when using digital twins for production optimization or storage risk mitigation. This includes improved understanding of assurance of containment and conformance of injected CO_2, an important topic of this issue's special section on carbon management. Meeting these challenges requires digital twins to make statistical inferences from multimodal monitoring data. Instead of treating subsurface CO_2 saturation as deterministic, digital twins should infer probability distributions for the saturation conditioned on observed data. Additionally, they should understand how these distributions evolve as CO_2 plumes develop and new monitoring data become available.

Reservoir monitoring systems struggle to capture uncertainty in a principled way due to the large problem sizes, the complexity of the nonlinear relationships between reservoir properties, multiphase flow, and the seismic response. However, it can be argued that the root issue is that our simulators are ill-suited for statistical inference. To address this, digital twins can benefit from recent breakthroughs in generative AI and simulation-based inference (SBI). This raises the question of

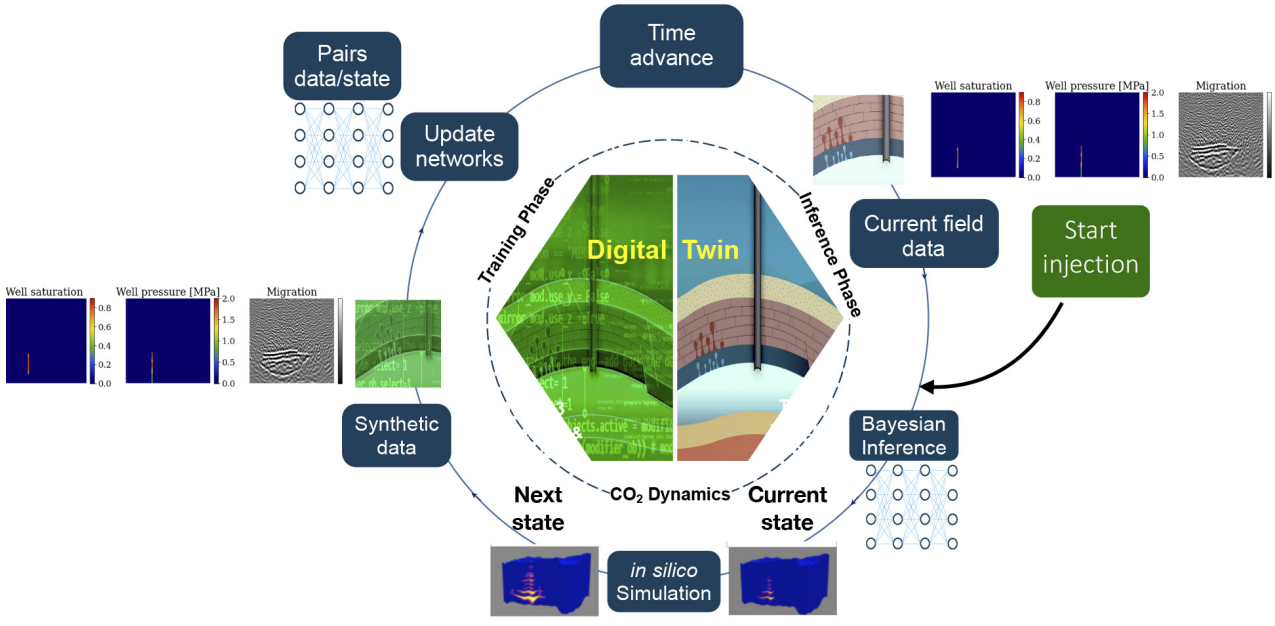

Figure 1. Digital twin for geologic carbon storage driven by CO_2 saturation and pressure at the well and by imaged seismic.

how digital twins can utilize generative AI. Deep generative networks, akin to advanced denoisers, can be trained to transform Gaussian noise into realistic samples of a specific distribution, whether it's images or CO_2 plumes. Moreover, this generative process can be conditioned on various data types, including geophysical data. In a physics-based context, SBI enables domain experts such as geophysicists and reservoir engineers to conduct principled statistical inference on field data by training deep networks on physics-based computer simulations. These principles of SBI will be demonstrated in a prototype digital twin for underground-storage monitoring, described in the following.

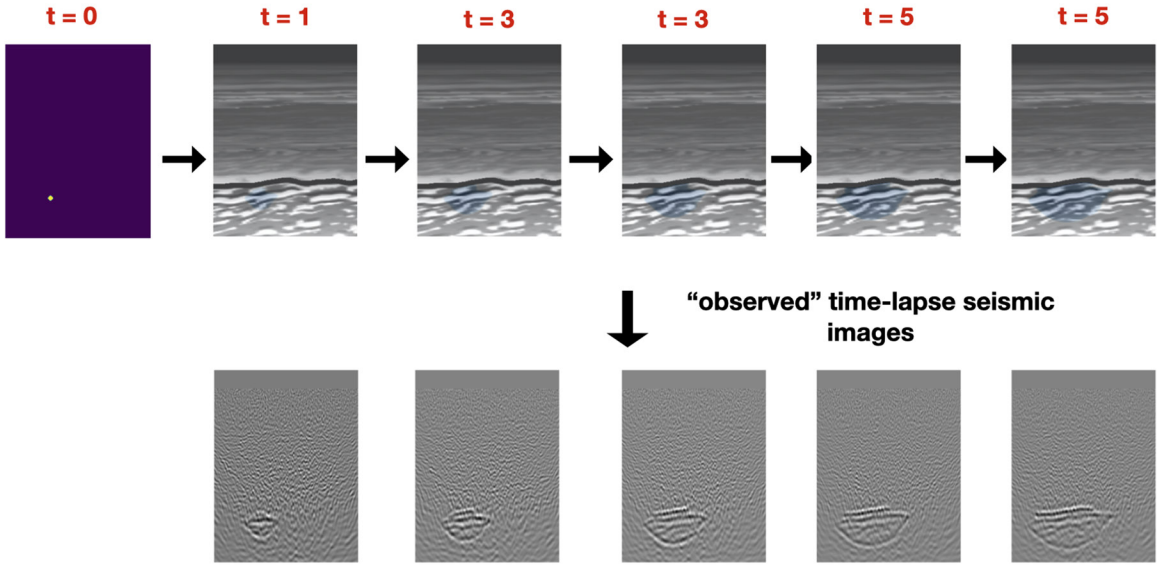

Figure 2. Ground-truth simulations for the CO_2 saturation and time-lapse seismic differences (bottom row) for a randomly sampled initial condition (plotted on the top left) and reservoir properties with developing CO_2 plumes superimposed.

Downloaded 07/30/24 to 65.132.165.43. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/tle42110730.1

Downloaded 07/30/24 to 65.132.165.43. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/tle42110730.1

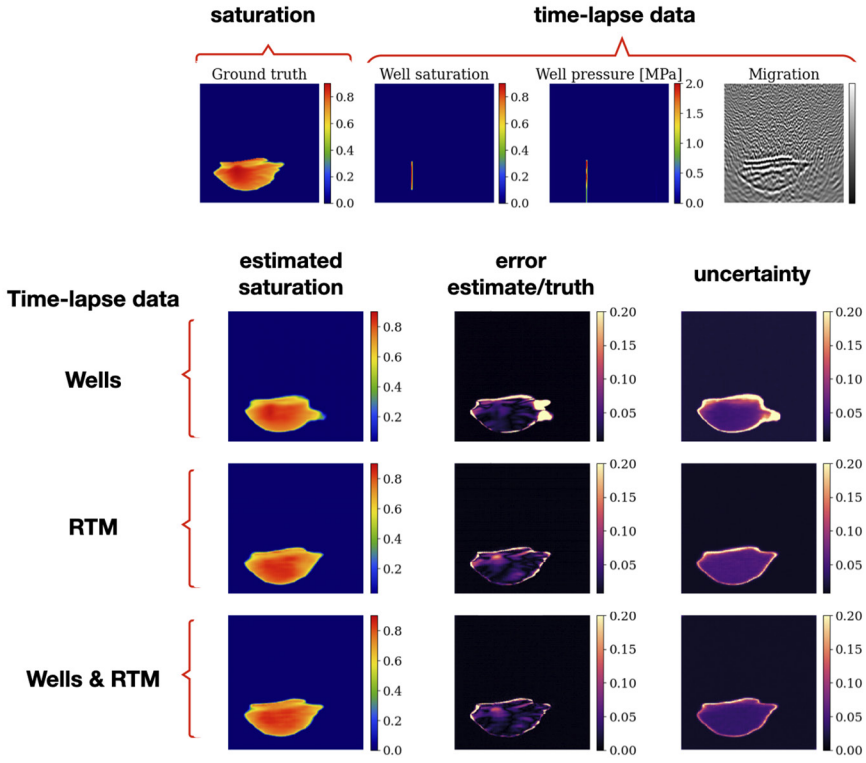

Figure 3. Example of inference by the first prototype of a digital twin for underground-storage monitoring. Top row: Ground-truth CO_2 saturation, corresponding saturation/pressure at injection well, and migrated seismic time-lapse difference image from noisy shot data (S/N 8 dB). Bottom rows: estimates for the state, the error between estimate and ground truth, and estimated uncertainty. These inferences are done on well data alone, seismic data alone, and well plus seismic data.

To enable SBI for dynamic systems, a recursive scheme is proposed (see Figure 1). In this scheme, digital twins are trained on simulations representing their state, the CO_2 saturation in this case, and observable data (well and/or seismic). Once training is complete, the system's state is inferred when time-lapse field data become available. The recursive process involves drawing samples for the state from a previous time step and using them as input for a reservoir simulator to obtain samples for the current state. These state samples are then "observed" at wells or imaged from seismic data. During the *training phase*, the digital twin's networks are trained using paired samples of the state and multimodal observations. After training, during the *inference phase*, the digital twin's networks generate state samples conditioned on new data collected in the field. This process is repeated for all time steps to cover the entire lifecycle of a CO_2 storage project.

To demonstrate the digital twin's recursive neural training in a saline aquifer, five synthetic time-lapse surveys are created for a CO_2 injection project. These surveys span 2000 days with an annual injection rate of 1.4 million tons of CO_2. Each survey involves eight shots recorded by 200 receivers. Noisy shot data (signal-to-noise ratio [S/N] of 8 dB) undergo reverse time migration, producing time-lapse difference images used as input for the digital twin (see Figure 2). To prepare the digital twin for inferences as ground-truth data become available, the recursion begins by drawing 256 random samples for the initial state. Computer simulations generate synthetic state samples and observations for the first time step based on the previous state. These pairs train the digital twin's neural network to draw 256 new samples for the state, conditioned on the observed ground-truth data for the first time step. This process repeats five times, resulting in inferences from the first digital twin prototype included in Figure 3. Conclusions from this prototype demonstrate its ability to estimate plume CO_2 saturations, remaining close to ground truth due to conditioning with observed data. The use of seismic data enhances plume estimation compared to estimates from well data alone, and combining both data types yields the best results, with minimal errors and reduced state sample uncertainty.

While these results, obtained by running for a day on an NVIDIA V100 GPU, are preliminary, some important observations can be made. Most notably, the digital twin's inference, based solely on probabilistic initial conditions and reservoir properties, produces estimates close to the actual state when conditioned on time-lapse data. Additionally, a comparison between the recursive training/inference scheme and IBM's digital twin definition reveals that this early prototype aligns with the concept of a "virtual representation (by deep neural networks) of a system spanning its lifecycle, updated from real-time data, and utilizing simulation and machine learning."

What's next? Besides extending to 3D and including geochemistry and geomechanics in the dynamics, the current digital twin for geologic storage monitoring lacks the ability "to reason and make optimized decisions." Ongoing research at the Center for Machine Learning for Seismic Industry Partners Program (ML4Seismic) aims to equip the digital twin with reasoning capabilities and the capacity to make optimized decisions. This includes causal "what if" reasoning, production optimization such as maximizing CO_2 injectivity while avoiding fracture pressure, and implementing statistically robust methods for detecting anomalous flow (e.g., leakage).

Our experience with this first digital twin prototype has shown that it can be a collaborative platform for practitioners from various fields to train the digital twin's neural networks. Generative AI's capability to comprehend natural language and physics-based simulations offers new possibilities for creating advanced digital twins to manage complex underground resource production and storage. It is hoped that this column might encourage the industry to explore digital twin development for sustainable subsurface management, fostering interdisciplinary collaborations similar to those described later in this issue's special section. TLE