Uncertainty-Aware Well Placement: Simulator-Verified Dual-Network Reinforcement Learning Approach meets Particle Filters

Hibat Errahmen Djecta[1,2,*][0009−0005−9417−8364], Sergey Alyaev[1][0000−0002−2105−2067], Kristian Fossum[1], Reidar B.Bratvold[2], Ressi Bonti Muhammad[2], and Apoorv Srivastava[3]

[1] NORCE Norwegian Research Centre As, Bergen, Norway
[*] hidj@norceresearch.no
[2] University of Stavanger, Stavanger, Norway
[3] Stanford University, Stanford, CA, USA

Abstract. Geosteering, the art of navigating wells to maximize the reservoir resources, is fraught with challenges of geological uncertainty and the relentless pace of real-time operations. In this paper, we present a novel framework that integrates Particle Filters (PF) for probabilistic subsurface interpretation with a Dual-Network Deep Reinforcement Learning (DRL) model for adaptive decision-making in geosteering operations. The PF component quantifies subsurface uncertainties, providing a probabilistic interpretation of geological boundaries, while the DRL model leverages this information to generate optimal steering decisions. This synergy ensures robust trajectory planning that dynamically adapts to real-time geological changes. The framework incorporates key features, such as target-line alignment to maintain wellbore proximity to reservoir zones and dog-leg severity constraints to ensure operational feasibility. Extensive verification in an industry-standard environment accessed via an API demonstrates the model's ability to accurately track reservoir boundaries, predict gamma-ray values, and optimize well trajectories. The results highlight significant improvements over traditional geosteering approaches and standard DRL-based methods in terms of reservoir contact, decision-making efficiency, and trajectory accuracy, even in low-data scenarios. The proposed framework provides a scalable and robust solution for quantifying uncertainties in real-time geosteering, paving the way for informed operational decisions improving value-creation and drilling effciency.

Keywords: Deep Dual Reinforcement Learning · Geosteering · Particle Filters · Uncertainty Modeling · Reservoir Optimization.

1 Introduction

Real-time trajectory optimization in uncertain environments is a key challenge across various engineering disciplines, often likened to tasks such as autonomous

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14

driving. Yet the geosteering problem in drilling operations can be even more complex, as it involves adjusting well trajectories in a largely unobservable subsurface. The aim is to maximize monetary value by optimizing reservoir contact, minimizing costs, and mitigating hazards, all while dealing with indirect and potentially noisy data streams. Traditional geosteering practices have generally depended on manual interpretation of real-time measurements—methods that can be time-consuming, prone to human error, and difficult to scale to today's higher drilling demands. Recent advances in artificial intelligence (AI) and computational science have motivated the development of automated, data-driven geosteering workflows that seek to address these issues more systematically.

Kullawan et al. [5] pioneered a formalized decision-analytic framework for geosteering allowing the balancing multiple objectives such as reservoir contact, cost control, and drilling risk. The framework combined greedy optimization and Bayesian treatment of boundaries linking geological uncertainties to decision-making process. A follow-up paper [6] implemented Discretized Stochastic Dynamic Programming (DSDP) for fixed-thickness reservoir improving the value of the well by up to 31% compared to the original approach. Building on these approaches, Alyaev et al. [2] introduced a Decision Support System (DSS) that combined Ensemble Kalman Filtering (EnKF) with simplified dynamic programming, thus offering reproducible, effective decisions under uncertainty. Later, Alyaev et al. [1] integrated Generative-Adversarial-Networks (GAN) geological model into the DSS, extending decision-making to complex geology.

In parallel with these contributions, Reinforcement Learning (RL) has become a promising paradigm for sequential geosteering decisions. Muhammad et al. [9] introduced Deep Q-Networks (DQN) for geosteering that outperformed earlier model-based strategies, including greedy optimization and DSDP for test cases from [5, 6] - all without the implementation complexities and limitations. Muhammad et al. [10] coupled DQN with Particle Filters (PF) to model the uncertainty of hidden reservoir boundaries, verifying the RL advantages in more general case. Recently, the authors adapted PF+RL method to a realistic Geosteering-World-Cup (GWC) environment, called the Pluralistic robot [8]. The robot outperformed most human experts in a post-GWC-2021 synthetic test with noiseless measurements, even though the PF was setup to handle noise. Under these idealistic conditions the robot compared favorable against experts in the competition proving effectiveness of RL when the probabilistic estimation is accurate.

Despite this progress, several notable gaps persist. All of these geosteering methods are verified in controlled synthetic environments that may not capture the full range of real drilling uncertainties. Although the Pluralistic robot demonstrated a successful integration of RL and PF, it lacked a deeper exploration of more stable dual-network designs or testing across diverse, real-time operational constraints.

In this paper, we propose a new geosteering robot that enhances the Pluralistic framework by incorporating a Dual-Network Deep Reinforcement Learning (DRL) architecture [15] together with Particle Filters. Through a comparative

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14

analysis with the original Pluralistic robot we demonstrate that by separating value and advantage streams, this design promotes more stable policy learning and reduces the risk of converging to suboptimal decisions in complex drilling environments. Moreover, we link our approach to the ROGII Solo API [11] to enable testing on high-fidelity simulator and in real workflows within ROGII StarSteer. The industrial simulator simplifies testing with different geological scopes and includes industry-standard models for measurement noise and various drilling constraints.

In the next section, we describe our dual-network RL and PF integration, alongside the StarSteer environment (Section 2). We then present performance evaluations (Section 3), followed by a discussion of operational implications and future enhancements (Section 4). Finally, we conclude with a summary of our findings and propose directions for extending this approach to more complex geosteering problems (Section 5).

2 Methodology

In this study, we propose an advanced geosteering framework that integrates Particle Filters for probabilistic subsurface interpretation with a Dual-Network Deep Reinforcement Learning model for sequential decision-making. This approach explicitly accounts for uncertainties in geological parameters while optimizing well placement over multiple drilling steps.

2.1 System Architecture

Figure 1 provides an overview of the proposed framework. At each time step:

1. Real-time measurements (In our case: gamma-ray logs, trajectory data) are acquired from the drilling environment.
2. The PF assimilates these observations to update a posterior distribution of the subsurface state, quantifying uncertainties in reservoir boundaries and well orientation.
3. A dual-network DRL agent processes the PF's probabilistic state estimates and outputs a steering command (target-line shift).
4. The command is executed in the drilling simulator (StarSteer), leading to new measurements at the next time step.

This sequential loop continues until reaching the target depth or other operational objectives.

2.2 Particle Filter (PF)

A Particle Filter tracks multiple discrete boundary configurations, represented by particles, each characterized by a boundary offset and an angle. Angle increments are sampled from an approximate proposal distribution defined as a kernel density estimator (KDE) built from a reference set of angle increments,

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14

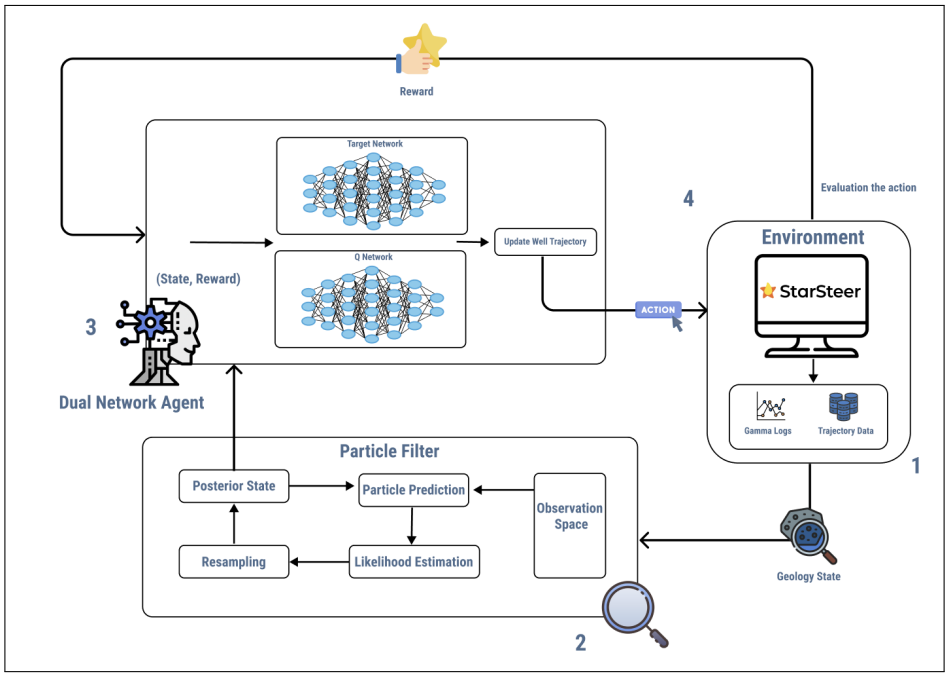


Fig. 1: High-level architecture illustrating the sequential data flow among the PF, Dual-Network DRL, and the drilling environment. Uncertainties in geological parameters are handled by the PF, while the DRL agent selects optimal actions.

thereby capturing realistic changes between drilling steps. As new gamma-ray logs become available, the PF updates each particle's weight and periodically resamples, yielding a posterior distribution of reservoir boundary positions under uncertainty.

Sequential Bayesian Update Let the state s_t^i of the i^{th} particle at time t be

$$s_t^i = \left(\text{offset}_t^i,\ \text{angle}_t^i\right),$$

where offset_t^i indicates how far above or below a reference depth the boundary lies, and angle_t^i encodes the local dip. The PF models the evolution , as described in [4], from time $t-1$ to t by

$$s_t^i = f\left(s_{t-1}^i\right) + \epsilon,$$

where ϵ is a stochastic term modeling geological variability (sampled from a KDE of angle increments). Once a new gamma-ray measurement o_t is obtained, each particle's weight w_t^i is updated according to Bayes' rule:

$$w_t^i \ \propto\ w_{t-1}^i \ \times\ p\left(o_t \mid s_t^i\right) \ \times\ p\left(s_t^i \mid s_{t-1}^i\right), \tag{1}$$

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14

where w_{t-1}^i is the weight of the i^{th} particle at time $t-1$, $p(o_t \mid s_t^i)$ quantifies how well the particle's predicted log matches o_t, and $p(s_t^i \mid s_{t-1}^i)$ is the prior transition probability for evolving from s_{t-1}^i. Here, o_t is the observed gamma-ray log at time t, which is generated by the environment as a function of the current subsurface configuration. Although the measurement is provided externally, its functional dependence on the true state (which, in turn, is influenced by both the previous state s_{t-1}^i and the applied control u_t) is implicitly captured in the likelihood function $p(o_t \mid s_t^i)$.

A normalization step follows so that $\sum_{i=1}^{N} w_t^i = 1$, the posterior distribution of states is updated accordingly.

Uncertainty Quantification and Measurements When many particles have negligible weights, resampling discards those low-weight interpretations and replicates more plausible ones. Representing the posterior distribution as weighted particles $\left(s_t^i, w_t^i\right)$ allows calculation of:

$$\text{Mean}(s_t) = \sum_{i=1}^{N} w_t^i \, s_t^i, \quad \text{Var}(s_t) = \sum_{i=1}^{N} w_t^i \left(s_t^i - \text{Mean}(s_t)\right)^2,$$

where N is the total number of particles. These metrics reveal both the most likely boundary configuration and the spread uncertainty around it.

To better replicate realistic log fluctuations, a correlation-based noise term can be added to the observed gamma-ray values:

$$\eta(\ell) = \big(\nu * \kappa\big)(\ell),$$

where ℓ is the log sample index, $\nu(\ell)$ is a vector of independent random draws, and κ is a Gaussian-like filter kernel. During training, the correlation scale is set to 2, meaning adjacent points within roughly two samples influence each other (depends on $\ell \pm 2$), producing moderately smooth fluctuations in the logs. By iteratively updating particle states and weights in tandem with these correlated measurements, the PF delivers a faster-than-real-time probabilistic interpretation of the boundary under geological uncertainty. During testing, the synthetic measurements are produced by the StarSteer simulator which uses a different noise model unknown to the robot.

2.3 Deep Reinforcement Learning (DRL) with a Dual-Network Architecture

Geosteering poses a high-dimensional, uncertain, and sequential decision-making challenge, where actions have long-term consequences. A standard Deep Q Network often struggles with the sparse and delayed rewards typical of geosteering tasks. We mitigate these issues using a dual-network DQN [13] that addresses these obstacles by decomposing the action-value function into separate value and advantage components, enhancing both learning stability and efficiency. In the following, we first define our action and state spaces.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14

Action and State Spaces

State Space: In the geosteering robot, each state draws on both probabilistic boundary estimates and real-time drilling information. A PF provides multiple particles describing possible boundary offsets and angles at or near the current drill-bit depth. Rather than using an entire segment, the robot samples a subset of these particles (the top five by weight), yielding around ten parameters (two per particle) that capture geological uncertainty. Additional drilling context—inclination in this case (i.e., the actual measured angle of the well relative to vertical), further expand the state vector. Under typical settings, the overall state vector comprises around 266 components, reflecting both the subsurface uncertainty and the drilling context at the current depth.

Action Space: Steering actions are discretized according to drilling phases. During the landing phase, the agent adjusts the wellbore angle ($\Delta\theta$) within a range of -10° to $+10^\circ$ in 0.5-degree increments to steer the well toward the reservoir zone. In the horizontal drilling phase, the agent selects from a set of discrete vertical adjustments that shift the planned well trajectory upward or downward by small increments (typically between -6 and $+6$ units). These vertical adjustments are intended to keep the well within the productive zone while preventing excessively abrupt directional changes. Such rapid changes in direction are constrained by the dog-leg severity (DLS) limit—a safety parameter that restricts the maximum rate of change in the wellbore's direction to protect the drilling assembly. Because the impact of these actions may only be observed several steps later, the agent's decision-making algorithm must be robust to delayed feedback.

Dual-Network Q-Value Decomposition A dual-network Deep Q-Network (DQN) functions as the agent's brain, receiving the current state s as input and outputting the optimal steering action a. This dual-network architecture decomposes the action-value function $Q(s, a)$ into a state-value $V(s)$ and an advantage $A(s, a)$ [13], as defined below:

$$Q(s, a) = V(s) + A(s, a) - \frac{1}{|\mathcal{A}|} \sum_{a' \in \mathcal{A}} A(s, a'),$$

where $V(s)$ captures the inherent desirability of the drilling state s, and $A(s, a)$ measures the relative benefit of taking action a compared to the average action in state s. The set $\mathcal{A}$ comprises all feasible steering adjustments, and $|\mathcal{A}|$ denotes the total number of possible actions.

By segregating $Q(s, a)$ into value and advantage components, the dual-network DQN mitigates overestimation biases and promotes stable learning, as illustrated in Figure 2. The agent processes state transitions (s, a, r, s') stored in an experience replay buffer, which enables random sampling and breaks temporal correlations in the training data. Here, r represents the reward signal received by the agent for taking action a in state s, guiding it toward optimal geosteering decisions. Additionally, the policy network's parameters θ are learned from these past state transitions through gradient-based updates, and a separate state-value

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14

network, updated periodically with the policy network's parameters θ^-, provides consistent Q-value targets, further enhancing learning stability. Meanwhile, the advantage network estimates the relative benefit of each action in a given state, allowing the model to distinguish between more and less advantageous choices. This architecture empowers the DRL agent to effectively handle the long-horizon and high-dimensional challenges inherent in geosteering tasks. By leveraging PF-derived states that encapsulate geological uncertainty, the agent learns geosteering policies that balance immediate operational constraints with long-term reservoir objectives, ensuring robust decision-making in complex and uncertain subsurface environments.

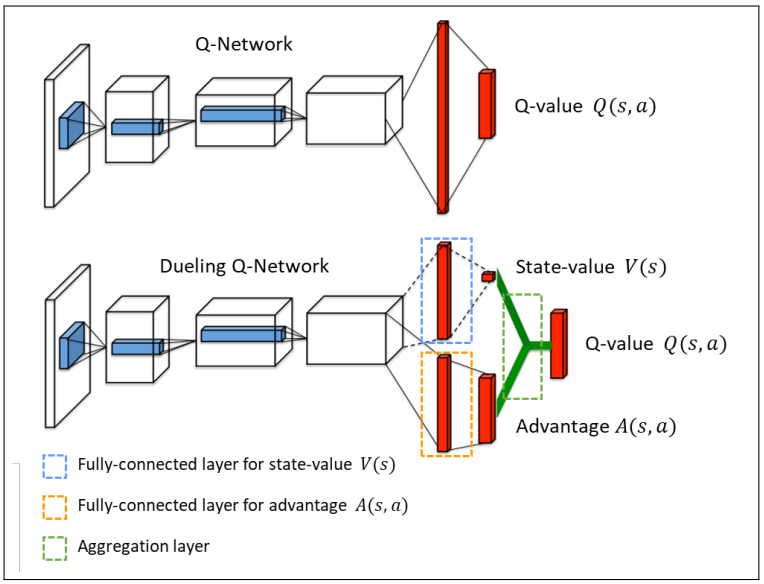


Fig. 2: Illustration of the dual-network (bottom) versus a single-stream architecture (top). In the dual-network approach, the network splits into two separate streams for estimating the value function *(left branch)* and the advantage function *(right branch)*, then recombines these outputs to produce the final Q-values. This decomposition can lead to faster convergence and more stable learning, particularly in high-dimensional geosteering tasks with sparse or delayed rewards [15].

Reward Formulation: Building on previous geosteering research [8], we define a reward function that balances both reservoir placement and operational constraints [12]. Concretely, the agent receives:

– Positive reward proportional to reservoir contact, i.e., how much of the drilled interval remains within the target layer or pay zone.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14

- Penalties for excessive dog-leg severity, which encourage smoother wellbore trajectory adjustments.
- Penalties for drilling out of zone, reflecting lost production potential and elevated operational risks.
- Minor or shaping penalties to maintain stable inclination/azimuth transitions, discouraging abrupt or extreme steering changes.

A typical formulation is given by Eq. (2):

$$r_t = \underbrace{w_c \cdot \text{contact}_t}_{\text{zone contact}} - \underbrace{w_d \cdot \text{dogleg}_t}_{\text{dog-leg penalty}} - \underbrace{w_o \cdot \text{out\_of\_zone}_t}_{\text{out-of-zone penalty}} - \underbrace{w_s \cdot \text{steer\_change}_t}_{\text{shaping penalty}}. \tag{2}$$

where t is the time step at which the reward is computed. In this expression, contact_t tracks how effectively the well trajectory stays within the reservoir, dogleg_t measures dog-leg severity, out_of_zone_t captures any excursions outside the target zone, and steer_change_t is a minor shaping term penalizing abrupt or excessive steering changes. The coefficients w_c, w_d, w_o, and w_s govern each term's relative importance. By adjusting these weights, we can emphasize reservoir contact or operational smoothness as needed, ensuring that the agent's drilling strategy balances short-term positioning goals with long-term trajectory stability.

Temporal Difference (TD) Learning: We train the agent using an off-policy Q-learning method with experience replay [7]. At each time step t, the environment transitions from state s_t to state s' after the agent selects an action a_t, yielding a reward r_t. These transitions (s_t, a_t, r_t, s') are stored for later randomly sampled updates. The replay buffer enables effectively offline training, helping to reduce correlation in consecutive samples. The agent's parameters are updated by minimizing the Temporal Difference (TD) error

$$\delta_t = r_t + \gamma \max_{a'} Q\big(s', a'; \theta^-\big) \; - \; Q\big(s_t, a_t; \theta\big), \tag{3}$$

where s' denotes the next state, a' is a candidate action in s', γ is the discount factor, and θ^- are parameters of the slowly updated target network. This off-policy approach handles potentially delayed or sparse geosteering rewards and leverages a separate target network to stabilize Q-value estimates [14]. By iteratively reducing δ_t, the agent converges on a policy that maximizes long-term return under geological uncertainty.

2.4 Integration of PF and Dual-Network DRL in a Sequential Loop

Algorithm 1 summarizes how PF outputs and dual-network DRL are coupled in a sequential geosteering system for testing. At each time step, the PF refines its posterior distribution of possible subsurface configurations in light of new measurements. The already trained DRL agent observes these PF outputs—along

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14

with drilling data such as the previous action—and selects a steering command to maximize the long-term expected reward.

Algorithm 1 Integration of Particle Filter (PF) and Dual-Network DRL in Geosteering Environment

Require: Real-time measurements o_t, initial particle set $\{s_0^i, w_0^i\}_{i=1}^N$, trained dual-network DRL agent, environment $\mathcal{E}$
Ensure: Steering actions a_t and updated trajectory
Initialize:
  $\{s_0^i, w_0^i\}$ (particle states & weights)
  environment $\mathcal{E}$ with initial subsurface model
  operational constraints (e.g. target-line geometry, dog-leg severity)
  replay buffer D for DRL (optional if training is on)
for each time step $t = 1, 2, \ldots, T$ **do**
  Step 1: Receive Observation

$$o_t \leftarrow \mathcal{E}.\text{get\_observation}()$$

  Step 2: PF Update
  for each particle $i = 1, \ldots, N$ **do**
    Predict $s_t^i = f(s_{t-1}^i, u_t) + \epsilon$
    Update $w_t^i \propto w_{t-1}^i \times p(o_t \mid s_t^i) \times p(s_t^i \mid s_{t-1}^i)$
  end for
  Normalize & Resample
  Step 3: Construct DRL State
    gather top N_{eff} particles by weight
    combine these with o_t and the previous action to form the DRL state
  Step 4: DRL Action Selection

$$a_t = \arg\max_{a \in \mathcal{A}} Q(\text{State}, a; \theta)$$

  $\mathcal{E}$.apply_action(a_t)
  Step 5: Reward Computation (Optional Training)
    $r_t \leftarrow \mathcal{E}$.compute_reward()
    log transition (s_t, a_t, r_t, s_{t+1}) in D
  if training is on **then**
    sample mini-batch from D, update $Q(\cdot; \theta)$ via TD error
  end if
end for

Each iteration thus involves gathering new measurements, using the PF to sample from the posterior distribution of subsurface configurations based on refined geological knowledge, constructing a DRL state that encodes the most likely scenarios, and selecting a steering action that maximizes the Q-value. Although the algorithm supports on-demand re-training of the dual-network parameters θ, in our current approach, the agent uses a pre-trained policy at

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14

inference time. Future extensions will consider real-time adaptation of θ once more data is accrued, enabling the robot to respond dynamically to changing subsurface conditions.

3 Experimental setup and results

The results of the proposed framework are presented in this section, offering a detailed comparison of the Normal DRL Robot and the Dual DRL Robot based on multiple evaluation metrics and visualizations.

The experiments were conducted on a high-performance computing system equipped with a 13th Gen Intel® Core™ i7-13800H × 20 processor, NVIDIA Corporation / Mesa Intel® Graphics (RPL-P) with 32GB VRAM, and running on Ubuntu 22.04. During training, a KDE was utilized to sample the geological environment, enabling the robot to learn and adapt effectively within a controlled setup. The training process, which lasted four hours, utilized several critical parameters, which are summarized in Table 1. These parameters were optimized to achieve robust and reliable performance.

For testing, the StarSteer simulator was employed via Solo API [11] to communicate new placements and extract feedback from real subsurface interactions, allowing the robot to operate in a dynamic and realistic setting. Such setup assures unbiased evaluation process of the developed method. To our knowledge this is the first publication of automated geosteering results in StarSteer apart from internal automatic geosteering developments [3].

Table 1: Training Parameters

Parameter	**Value**
Number of Episodes	20,000
Learning Rate	0.0005
Discount Factor (γ)	0.95
Batch Size	64
Number of Particles in Particle Filter (Training)	256
Number of Particles in Particle Filter (Testing)	2064
Replay Buffer Size	50,000
Episods before replacements	1000
Epsilon Decay Rate	0.995
Minimum Epsilon	0.01
Target Network Update Frequency	100 steps

Figure 3a highlights the progression of training rewards for both models. The Dual DRL Robot demonstrates significantly faster convergence and consistently higher cumulative rewards compared to the Normal DRL Robot. This clearly indicates its superior learning capabilities and improved performance.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14

Figure 3a also depicts the mean absolute error (MAE) progression for trajectory predictions. The Dual DRL Robot achieves lower MAE values throughout training, underscoring its ability to accurately minimize trajectory prediction errors and outperform the Normal DRL Robot in terms of prediction reliability.

The epsilon decay progression is shown in Figure 3b, which illustrates the transition from exploration to exploitation for both robots. The Dual DRL Robot benefits from an optimized epsilon decay schedule, enabling a quicker and more efficient policy convergence. Conversely, the Normal DRL Robot retains higher exploration for a prolonged duration, resulting in slower overall progress.

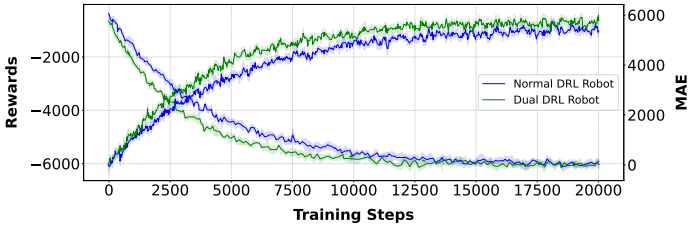
The boundary prediction performance, as illustrated in Figure 3c, compares the differences between true and predicted boundaries over measured depth (MD). The Dual DRL Robot produces smoother and more accurate boundary predictions in most cases. However, occasional spikes reveal instability in certain scenarios, which highlights potential areas for further improvement. The Normal DRL Robot, while more consistent, provides less accurate predictions overall. Notably, the spikes observed in the Dual DRL Robot's predictions coincide with regions of rapid geological transitions, suggesting the need for enhanced boundary prediction mechanisms or incorporating geological priors.

The plot in Figure 4 offers a comprehensive view of the Dual DRL Robot's performance in trajectory tracking, geological boundary prediction, and gamma-ray estimation. These results are benchmarked against the true geologic boundaries and real gamma-ray logs derived from a geosteering solution performed by experts at ROGII. In the upper panel, the trajectory (red solid line) closely follows the real boundaries (black solid line), whereas the best-predicted boundaries (dashed blue line) indicate the robot's ability to approximate subsurface features. Meanwhile, the colored dashed lines from the PF capture uncertainty in more geologically complex regions.

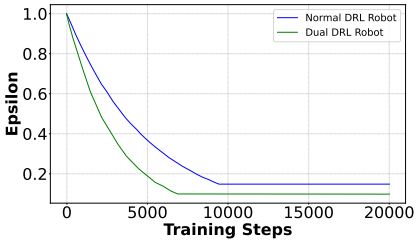
In the lower panel, the yellow line (predicted gamma-ray values) and the green line (real gamma-ray logs, as provided by the ROGII-based solution) track the corresponding subsurface changes. At around MD 3750 ft, a slight misalignment between the predicted and real boundaries in the upper plot coincides with a broader spread in particle-filter predictions for the gamma-ray curve, signaling increased uncertainty. This correlation underlines how gamma-ray measurements dynamically reflect changes in boundary geometry and trajectory adjustments across the two panels. In spite of local uncertainties, the Dual DRL Robot demonstrates effective alignment with the real geologic features, showing that gamma-ray predictions serve as a real-time indicator of subsurface variability captured in both the trajectory and boundary estimations.

These results collectively confirm the effectiveness of the Dual DRL Robot in achieving faster convergence, higher rewards, and more accurate predictions. Its advanced architecture and optimized training approach provide significant advantages over the Normal DRL Robot, reinforcing its potential in geosteering applications. Moreover, the combination of reward stability, predictive accuracy, and real-time adaptability positions the Dual DRL Robot as a promising tool for reducing operational uncertainties in complex drilling scenarios.

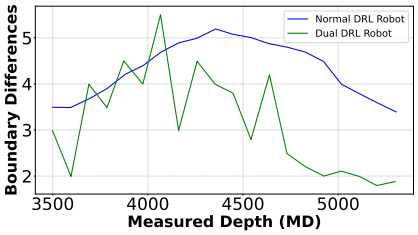
ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14



(a) Training rewards and MAE progression for Normal and Dual DRL Robots.



(b) Epsilon decay progression for Normal and Dual DRL Robots.



(c) Boundary differences between true and predicted over measured depth.

Fig. 3: Comparison of metrics for Normal and Dual DRL Robots. Each subfigure highlights a distinct aspect of model performance.

4 Discussion

The results demonstrate the Dual DRL Robot's superiority over the Normal DRL Robot with faster convergence, higher rewards, and more accurate trajectory and boundary predictions. This is due to its dual-network architecture, which stabilizes learning and reduces overfitting, along with PF's probabilistic state estimates for handling subsurface uncertainties. These features enable more reliable and precise steering, crucial for optimal reservoir contact. An optimized epsilon decay strategy accelerates policy convergence, improving learning efficiency over the Normal DRL Robot. However, occasional boundary instabilities suggest the need for adaptive constraints or regularization. Further testing across varied geological conditions could enhance robustness and mitigate predictive inconsistencies. The dual-network and PF integration increase computational demands, posing challenges for real-time deployment. Optimizing efficiency through parallel processing or model compression is essential.

The Dual DRL Robot advances geosteering by integrating DRL with uncertainty quantification. Refining its computational performance and predictive consistency will further strengthen its practical impact.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14

5 Conclusion

This paper presents a novel framework integrating a Dual DRL Robot with a PF for geosteering, tested in an environment external to training. Compared to earlier RL Robots, the Dual DRL Robot achieved faster convergence, higher rewards, and greater accuracy in trajectory and boundary predictions.

The dual-network architecture improved training stability and reduced overfitting, while the PF quantified subsurface uncertainties through probabilistic state estimates. An optimized epsilon decay strategy further enhanced learning efficiency, enabling accurate alignment with geological boundaries and adaptation to dynamic drilling conditions.

Challenges remain, including occasional boundary misalignments and computational demands for real-time deployment. Future work could incorporate additional geological features, optimize efficiency, and explore robustness across diverse geological settings.

In summary, the Dual DRL Robot significantly advances geosteering by combining DRL with uncertainty quantification, enhancing decision-making and operational efficiency. Addressing existing limitations will further improve its practical applicability in drilling operations.

Acknowledgments

H.E. Djecta, S. Alyaev, K. Fossum, and R.B. Bratvold acknowledge the support from the project DISTINGUISH (Decision support using neural networks to predict geological uncertainties when geosteering), funded by Aker BP, Equinor, and the Research Council of Norway (RCN PETROMAKS2 project no. 344236).

R.B. Muhammad acknowledges the support from the Center for Research-based Innovation DigiWells: Digital Well Center for Value Creation, Competitiveness and Minimum Environmental Footprint (NFR SFI project no. 309589), funded by Aker BP, ConocoPhillips, Equinor, Harbour Energy, Petrobras, TotalEnergies, Vår Energi, and the Research Council of Norway.

The authors thank ROGII Inc. for providing the academic licenses for Solo Cloud and StarSteer and the relevant training.

Statement on AI-generated text

The authors employed OpenAI's ChatGPT to refine their initial drafts and then carefully revised the AI-generated text to ensure it accurately represented their views and insights.

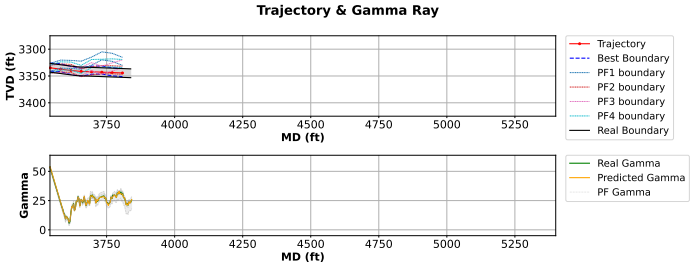
References

1. Alyaev, S., Fossum, K., Djecta, H., Tveranger, J., Elsheikh, A.: Distinguish workflow: a new paradigm of dynamic well placement using generative machine learning.

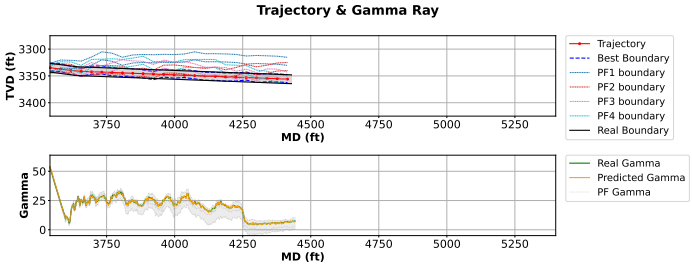
ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14

In: ECMOR 2024. vol. 2024, pp. 1–16. European Association of Geoscientists & Engineers (2024)
2. Alyaev, S., Suter, E., Bratvold, R.B., Hong, A., Luo, X., Fossum, K.: A decision support system for multi-target geosteering. Journal of Petroleum Science and Engineering **183**, 106381 (Dec 2019). https://doi.org/10.1016/j.petrol.2019.106381, http://dx.doi.org/10.1016/j.petrol.2019.106381
3. Denisenko, I.D., Kuvaev, I.A., Uvarov, I.B., Kushmantzev, O.E., Toporov, A.I.: Automated geosteering while drilling using machine learning. case studies. In: SPE Russian Petroleum Technology Conference? p. D023S009R004. SPE (2020)
4. Glasauer, S.: Chapter 1 - sequential bayesian updating as a model for human perception. In: Ramat, S., Shaikh, A.G. (eds.) Mathematical Modelling in Motor Neuroscience: State of the Art and Translation to the Clinic. Gaze Orienting Mechanisms and Disease, pp. 3–18. Progress in Brain Research, Elsevier (2019)
5. Kullawan, K., Bratvold, R., Bickel, J.: A decision analytic approach to geosteering operations. SPE Drilling & Completion **29** (03 2014). https://doi.org/10.2118/167433-PA
6. Kullawan, K., Bratvold, R., Bickel, J.: Sequential geosteering decisions for optimization of real-time well placement. Journal of Petroleum Science and Engineering **165**, 90–104 (2018). https://doi.org/https://doi.org/10.1016/j.petrol.2018.01.068, https://www.sciencedirect.com/science/article/pii/S0920410518300809
7. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M.A., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D.: Human-level control through deep reinforcement learning. Nature **518**, 529–533 (2015), https://api.semanticscholar.org/CorpusID:205242740
8. Muhammad, R.B., Cheraghi, Y., Alyaev, S., Srivastava, A., Bratvold, R.B.: Geosteering robot powered by multiple probabilistic interpretation and artificial intelligence: Benchmarking against human experts. SPE Journal pp. 1–15 (01 2025). https://doi.org/10.2118/218444-PA, https://doi.org/10.2118/218444-PA
9. Muhammad, R.B., Alyaev, S., Bratvold, R.B.: Optimal sequential decision-making in geosteering: A reinforcement learning approach (2025), https://arxiv.org/abs/2310.04772
10. Muhammad, R.B., Srivastava, A., Alyaev, S., Bratvold, R.B., Tartakovsky, D.M.: High-precision geosteering via reinforcement learning and particle filters (2024), https://arxiv.org/abs/2402.06377
11. Rogii Inc.: Solo REST API Documentation (2025), https://api.solo.cloud/, accessed: 2025-02-11
12. Shelton, C.: Balancing multiple sources of reward in reinforcement learning. In: Leen, T., Dietterich, T., Tresp, V. (eds.) Advances in Neural Information Processing Systems. vol. 13. MIT Press (2000)
13. Sikchi, H., Zheng, Q., Zhang, A., Niekum, S.: Dual rl: Unification and new methods for reinforcement and imitation learning (2024), https://arxiv.org/abs/2302.08560
14. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. The MIT Press, second edn. (2018), http://incompleteideas.net/book/the-book-2nd.html
15. Wang, Z., Schaul, T., Hessel, M., van Hasselt, H., Lanctot, M., de Freitas, N.: Dueling network architectures for deep reinforcement learning (2016), https://arxiv.org/abs/1511.06581

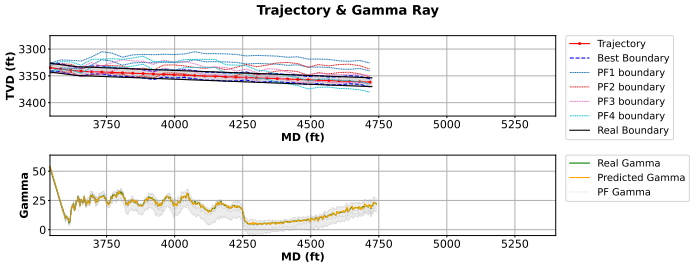
ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14



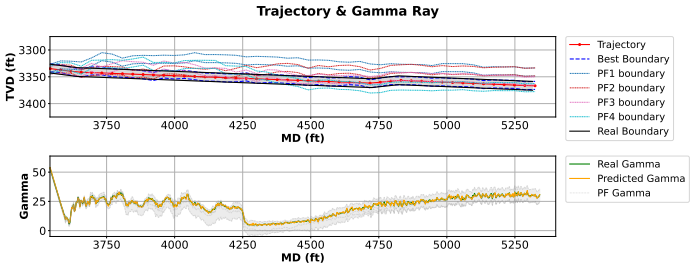
(a) Trajectory predictions at the first time slot, showing the comparison between predicted and actual values.



(b) Gamma-ray predictions at the second time slot, highlighting real-time interaction with StarSteer.



(c) Trajectory and gamma-ray predictions at the third time slot, showing alignment between predicted and actual values.



(d) Final trajectory and gamma-ray prediction progression at the fourth time slot, demonstrating interaction with StarSteer.

Fig. 4: Trajectory and gamma-ray prediction progression at different time slots with real-time interaction with StarSteer.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_14