Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Learning-Driven Autonomy Pipeline for Horizontal Wellbore Geology Prediction

Lakshman Mahto

lakshman@bitmesra.ac.in

Birla Institute of Technology, Mesra

Research Article

Keywords: horizontal wellbore, geosteering, gamma ray, true vertical thickness, typewell, residual CVAE, sparse attention, Random Fourier Features, kernel mean embedding, GR-score, uncertainty-aware fusion, autonomous steering.

Posted Date: July 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9902401/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Learning-Driven Autonomy Pipeline for Horizontal Wellbore Geology Prediction

Lakshman Mahto
Quantitative Economics and Data Science
Birla Institute of Technology Mesra, Ranchi-835215, India
Corresponding author: Lakshman Mahto
Email: lakshman@bitmesra.ac.in

May 14, 2026

Abstract

Horizontal wellbore geosteering requires continuous estimation of subsurface position, lithologic confidence, and steering risk from incomplete, noisy, and heterogeneous measurements. These measurements commonly include measured depth, trajectory coordinates, gamma ray logs, interpreted formation surfaces, and typewell references. This paper presents a learning-driven autonomy pipeline for horizontal wellbore geology prediction using a real horizontal-well dataset and an associated typewell reference. The horizontal-well dataset contains 5,278 measured-depth stations, while the typewell dataset contains 1,296 samples. The prediction-start point is inferred from the first missing `TVT_input` value and occurs at row 1,442, corresponding to MD = 12,909 ft.

The proposed pipeline integrates geometry-based true-vertical-thickness (TVT) prediction, tree-ensemble baselines, typewell gamma ray interpolation, GR-score lithologic calibration, multimodal Random Fourier Feature kernel mean embeddings (RFF-KME), conditional variational autoencoder (CVAE) forecasting, causal block-sparse attention, local uncertainty-aware fusion, facies-state diagnostics, and steering logic. The empirical results indicate that ridge regression is the strongest pure numerical TVT predictor for the studied well, suggesting that TVT is primarily controlled by trajectory geometry and formation-surface relationships. However, ridge regression alone does not provide lithologic confidence, calibrated uncertainty, facies-shift detection, or steering guidance. Therefore, the recommended architecture uses ridge regression as a deterministic geometric backbone and trains a residual sparse-attention CVAE to learn probabilistic corrections around the ridge baseline.

The resulting framework combines geometry-driven TVT prediction, lithologic agreement through GR-score, probabilistic residual forecasting, uncertainty-aware fusion, and conservative steering. The proposed approach provides a practical foundation for autonomous geosteering systems that require not only accurate TVT estimates but also interpretable uncertainty and decision support.

Keywords: horizontal wellbore; geosteering; gamma ray; true vertical thickness; typewell; residual CVAE; sparse attention; Random Fourier Features; kernel mean embedding; GR-score; uncertainty-aware fusion; autonomous steering.

1 Introduction

Horizontal well geosteering is a sequential decision-making problem under geological uncertainty. At each measured-depth station, a drilling advisory or autonomous steering system must estimate

whether the wellbore remains within the intended target interval, whether the observed gamma ray (GR) response is lithologically consistent with the predicted stratigraphic position, and whether the next operational decision should be to continue drilling, steer upward, steer downward, or hold for additional evidence.

In practice, such decisions are made from limited real-time measurements, incomplete formation knowledge, and noisy log responses. Gamma ray logs are central to many geosteering workflows because GR measurements are commonly available in real time and provide a useful proxy for lithologic variation and stratigraphic correlation. However, GR interpretation is not unique. Similar GR values may occur in different lithologies, typewell signatures may not fully represent lateral facies variation, and local noise can create misleading pointwise matches. Consequently, reliable horizontal-well interpretation requires a workflow that combines trajectory geometry, GR behavior, typewell information, uncertainty estimation, and conservative steering logic.

The dataset considered in this study consists of a horizontal-well file and an associated typewell file. The horizontal-well file includes measured depth (MD), spatial coordinates, gamma ray, interpreted TVT, formation surfaces, and `TVT_input`. The typewell file contains TVT, GR, and geology labels. The prediction-start point is inferred from the first missing `TVT_input`, yielding row 1,442 at MD = 12,909 ft. The objective is not only to predict TVT but also to produce an operational state containing TVT uncertainty, lithologic confidence, predicted layer, favorable-layer probability, facies-state diagnostics, and steering recommendation.

The proposed pipeline follows a hybrid physics-learning strategy. Ridge regression provides a stable geometry-controlled TVT backbone; random forest, extra trees, and gradient boosting provide nonlinear tabular baselines; typewell GR interpolation and GR-score provide lithologic consistency; CVAE forecasting provides probabilistic future trajectories; sparse attention provides causal selective memory over recent multimodal context; and uncertainty-aware fusion combines residual error, model spread, GR variability, and GR-score mismatch into a steering confidence state.

The key empirical observation is that ridge regression is the strongest pure TVT predictor for the attached well. This indicates that TVT is primarily controlled by trajectory geometry and formation-surface relationships. Nevertheless, ridge regression alone is insufficient for autonomous geosteering because it does not quantify lithologic agreement, detect persistent facies shifts, provide calibrated uncertainty, or produce steering actions. The final architecture therefore uses ridge regression as the deterministic backbone and trains a residual sparse-attention CVAE to model structured corrections around the ridge baseline.

The main contributions of this work are as follows:

1. A complete learning-driven autonomy pipeline for horizontal wellbore geology prediction that integrates geometric regression, lithologic calibration, probabilistic residual forecasting, uncertainty-aware fusion, and steering diagnostics.

2. A GR-score formulation that compares horizontal-well GR with typewell GR interpolated at predicted TVT, thereby providing a soft lithologic confidence signal.

3. A residual CVAE formulation in which the deep model predicts

$$r_i = TVT_i - \widehat{TVT}_i^{ridge}$$

   rather than full TVT, improving stability and generalization.

4. A causal block-sparse attention design that provides interpretable local-to-global context aggregation while preserving real-time validity.

5. A local uncertainty-aware fusion scheme that combines model reliability, model spread, GR variability, and GR-score mismatch.

6. A comprehensive set of diagnostic simulations and plots, including TVT prediction comparison, residual diagnostics, uncertainty calibration, local fusion weights, GR-score analysis, sparse-attention masking, rolling performance, layer-wise error, and facies-state evolution.

2 Related Work

2.1 Geosteering and GR-Based Structural Interpretation

Automated geosteering has traditionally relied on comparing horizontal-well log signatures with typewell, pilot-well, or offset-well references. GR-based structural interpretation attempts to infer stratigraphic position by mapping measured-depth log behavior into a stratigraphic coordinate system. These methods are valuable because GR logs are often among the most accessible real-time measurements in horizontal wells. However, GR-based geosteering remains sensitive to typewell representativeness, local facies variation, layer-thickness changes, structural discontinuities, and log noise.

True stratigraphic thickness and pattern-matching methods seek to align horizontal-well GR signatures with vertical or typewell signatures. Although useful, such methods commonly rely on assumptions of layer continuity, limited faulting, and stable lithologic response. The present work extends GR-based geosteering by coupling GR evidence with geometry-driven TVT prediction, probabilistic residual learning, uncertainty-aware fusion, and conservative steering logic.

2.2 Regularized Regression and Tree-Ensemble Baselines

Ridge regression is a classical regularized linear model that stabilizes least-squares estimation under multicollinearity by imposing an L_2 penalty [Hoerl and Kennard, 1970]. This is particularly relevant for horizontal wellbore data because MD, coordinates, vertical position, inclination proxies, and formation-surface offsets are often strongly correlated.

Tree-ensemble methods such as random forests [Breiman, 2001], extremely randomized trees [Geurts et al., 2006], and gradient boosting [Friedman, 2001] provide nonlinear alternatives for tabular prediction. These methods can capture feature interactions and are often strong baselines in applied machine learning. However, tree ensembles may extrapolate poorly beyond the calibration interval, which is important in geosteering because prediction often occurs forward in measured depth.

In the present dataset, ridge regression strongly outperforms the tree baselines for pure TVT prediction, suggesting that the interpreted TVT is largely geometry controlled. This motivates using ridge as the deterministic backbone and reserving deep learning for residual correction, uncertainty estimation, and contextual diagnostics.

2.3 Random Fourier Features and Kernel Mean Embeddings

Random Fourier Features approximate shift-invariant kernels through randomized finite-dimensional feature maps [Rahimi and Recht, 2007]. This enables kernel-like nonlinear representation without explicitly forming large kernel matrices. Kernel mean embeddings represent distributions in reproducing kernel Hilbert spaces and are widely used for distributional representation learning [Smola et al., 2007, Muandet et al., 2017].

In this work, Random Fourier Feature kernel mean embeddings (RFF-KME) summarize local distributions of geometry, GR/log behavior, and stratigraphic-surface context within rolling windows. This allows the model to represent local texture and variability rather than relying only on pointwise measurements.

2.4 GRU, Attention, Transformers, and Sparse Attention

Gated recurrent units (GRUs) provide efficient recurrent sequence modeling for variable-length representations [Cho et al., 2014]. Attention mechanisms allow sequence models to learn which input positions are most relevant to a prediction target [Bahdanau et al., 2015]. The transformer architecture generalizes attention-based sequence modeling through multi-head self-attention [Vaswani et al., 2017].

Full self-attention has quadratic complexity in sequence length. Sparse attention restricts attention to selected local, block, or global positions to improve scalability [Child et al., 2019, Beltagy et al., 2020]. Linear and kernelized attention approximate full attention to reduce computational complexity [Katharopoulos et al., 2020, Choromanski et al., 2021]. In the proposed pipeline, causal block-sparse attention is used because it is interpretable, geologically structured, and real-time safe.

2.5 Conditional Variational Autoencoders and Probabilistic Forecasting

Variational autoencoders provide latent-variable generative modeling through an evidence lower bound objective [Kingma and Welling, 2014]. Conditional variational autoencoders extend this framework to structured conditional prediction [Sohn et al., 2015]. CVAEs are suitable for multi-step forecasting because latent sampling allows multiple plausible future trajectories.

The proposed residual CVAE models future ridge residuals and GR, rather than full TVT. This is motivated by the empirical dominance of ridge regression and by the residual-learning principle: when a strong baseline exists, a probabilistic model should focus on the structured residual rather than relearning the full signal.

2.6 Uncertainty Calibration and Filtering

Kalman filtering provides a recursive state-space framework for filtering and prediction under uncertainty [Kalman, 1960]. Gaussian processes provide probabilistic regression with posterior predictive uncertainty [Rasmussen and Williams, 2006]. Conformal prediction provides distribution-free uncertainty intervals under exchangeability assumptions [Vovk et al., 2005, Shafer and Vovk, 2008]. In the present pipeline, uncertainty is estimated through validation residual error, model spread, CVAE sampling, GR variability, and GR-score mismatch.

3 Problem Formulation and Assumptions

Let the horizontal well be represented by

$$\mathcal{H} = \{(MD_i, X_i, Y_i, Z_i, GR_i, \mathbf{S}_i, TVT_i)\}_{i=1}^{N}, \tag{1}$$

where MD_i is measured depth, (X_i, Y_i, Z_i) are spatial coordinates, GR_i is gamma ray, $\mathbf{S}_i$ denotes interpreted formation surfaces, and TVT_i is interpreted true vertical thickness. In the attached horizontal-well dataset, $N = 5278$.

Let the typewell be

$$\mathcal{T} = \{(TVT_j^{type}, GR_j^{type}, Layer_j^{type})\}_{j=1}^{M}, \tag{2}$$

where $M = 1296$ for the attached typewell dataset.

At current station t, the future forecasting target is

$$\mathbf{Y}_{t+1:t+H} = \{(TVT_{t+k}, GR_{t+k})\}_{k=1}^{H}. \tag{3}$$

The implemented sequence settings are

$$W_{raw} = 6, \qquad L_{seq} = 8, \qquad H = 6.$$

The autonomy output at station i is

$$\widehat{TVT}_i, \quad TVT_{q05,i}, \quad TVT_{q95,i}, \quad GRScore_i, \quad Layer_i, \quad P(Favorable_i), \quad Action_i. \tag{4}$$

The main assumptions are:

1. TVT is strongly influenced by wellbore geometry and formation-surface offsets.
2. GR provides lithologic evidence but can be noisy, locally shifted, or nonrepresentative.
3. Typewell GR is useful but imperfect.
4. Model disagreement and GR/typewell mismatch should increase uncertainty.
5. Real-time steering should become conservative when lithologic confidence is low or TVT uncertainty is high.

4 Preliminaries

4.1 Ridge Baseline

Ridge regression solves

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \|\mathbf{y} - X\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2 \right\}. \tag{5}$$

The ridge TVT prediction is denoted by

$$\widehat{TVT}_i^{ridge} = f_{ridge}(\mathbf{x}_i). \tag{6}$$

The ridge residual is

$$r_i = TVT_i - \widehat{TVT}_i^{ridge}. \tag{7}$$

4.2 GR-Score

The expected typewell GR at predicted TVT is obtained by interpolation:

$$GR^{type}(\widehat{TVT}_i) = \mathrm{Interp}(TVT^{type}, GR^{type}; \widehat{TVT}_i). \tag{8}$$

The GR residual is

$$\Delta GR_i = GR_i^{hor} - GR^{type}(\widehat{TVT}_i). \tag{9}$$

The GR-score is defined as

$$\boxed{GRScore_i = \exp\left[-\left(\frac{\Delta GR_i}{s_{GR}} \right)^2 \right]}, \qquad s_{GR} = 30. \tag{10}$$

Thus, a 30 API mismatch gives $GRScore_i \approx e^{-1} = 0.368$. GR-score is a soft lithologic confidence score rather than a hard layer classifier.

4.3 CVAE Objective

The CVAE models

$$p_\theta(\mathbf{y} \mid \mathbf{c}) = \int p_\theta(\mathbf{y} \mid \mathbf{z}, \mathbf{c}) p_\theta(\mathbf{z} \mid \mathbf{c})\, d\mathbf{z}, \tag{11}$$

with conditional prior $p_\theta(\mathbf{z} \mid \mathbf{c})$ and recognition model $q_\phi(\mathbf{z} \mid \mathbf{y}, \mathbf{c})$. The training loss is

$$\mathcal{L} = \mathcal{L}_{rec} + \beta D_{KL}\big(q_\phi(\mathbf{z} \mid \mathbf{y}, \mathbf{c}) \| p_\theta(\mathbf{z} \mid \mathbf{c})\big). \tag{12}$$

For the residual CVAE,

$$\mathcal{L}_{rec} = \lambda_r \|r - \widehat{r}\|^2 + \lambda_{GR} \|GR - \widehat{GR}\|^2. \tag{13}$$

5 Method

5.1 Feature Engineering

The geometry modality is

$$\mathbf{x}_i^{geo} = [MD_i, X_i, Y_i, Z_i, HD_i, dZ_i, Inc_i, Az_i, \kappa_i]. \tag{14}$$

The GR/log modality is

$$\mathbf{x}_i^{log} = [GR_i, GR_i^{MA}, GR_i^{STD}, \nabla GR_i, \nabla^2 GR_i, m_i^{GR}]. \tag{15}$$

The stratigraphic modality is

$$\mathbf{x}_i^{strat} = [\mathbf{S}_i, Z_i - \mathbf{S}_i]. \tag{16}$$

5.2 RFF-KME Multimodal Tokens

For modality m, an MLP encoder produces

$$\mathbf{z}_{s,r}^m = f_m(\mathbf{x}_{s,r}^m). \tag{17}$$

Random Fourier Features are computed as

$$\phi_m(\mathbf{z}) = \sqrt{\frac{2}{D}} \cos((W^m)^T \mathbf{z} + b^m). \tag{18}$$

The window embedding is

$$\mathbf{e}_s^m = \frac{1}{W_{raw}} \sum_{r=1}^{W_{raw}} \phi_m(\mathbf{z}_{s,r}^m). \tag{19}$$

The fused token is

$$\mathbf{f}_s = [\mathbf{e}_s^{geo} \| \mathbf{e}_s^{log} \| \mathbf{e}_s^{strat}]. \tag{20}$$

5.3 Sparse-Attention Residual CVAE

Full attention is

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{21}$$

with complexity $O(L^2 d)$. Causal block-sparse attention uses

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}} + M\right)V, \tag{22}$$

where $M_{ij} = -\infty$ if token j is in the future or outside the allowed sparse neighborhood. The effective cost is approximately $O(LBd)$, where $B \ll L$ is the block size.

The residual CVAE predicts

$$p_\theta(r_{t+1:t+H}, GR_{t+1:t+H} \mid c_t). \tag{23}$$

The reconstructed TVT is

$$\widehat{TVT}_{t+k}^{resCVAE} = \widehat{TVT}_{t+k}^{ridge} + \widehat{r}_{t+k}. \tag{24}$$

5.4 Uncertainty-Aware Fusion

Given component predictions $\widehat{TVT}_i^{(m)}$, the fused prediction is

$$\widehat{TVT}_i^{fused} = \sum_m w_{m,i}\widehat{TVT}_i^{(m)}. \tag{25}$$

If model uncertainty estimates are available, precision weights are

$$w_{m,i} = \frac{1/\sigma_{m,i}^2}{\sum_j 1/\sigma_{j,i}^2}. \tag{26}$$

Otherwise, validation RMSE defines global reliability weights:

$$w_m = \frac{1/RMSE_m^2}{\sum_j 1/RMSE_j^2}. \tag{27}$$

The fused spread is

$$Spread_i = \sqrt{\sum_m w_{m,i}\left(\widehat{TVT}_i^{(m)} - \widehat{TVT}_i^{fused}\right)^2}. \tag{28}$$

The fused uncertainty is

$$\sigma_i^{fused} = \sqrt{\sigma_{resid}^2 + Spread_i^2}\left(1 + \lambda_{noise}\frac{GRSTD_i}{30} + \lambda_{GR}(1 - GRScore_i)\right). \tag{29}$$

The 90% prediction interval is

$$TVT_{q05,i} = \widehat{TVT}_i^{fused} - 1.645\sigma_i^{fused}, \qquad TVT_{q95,i} = \widehat{TVT}_i^{fused} + 1.645\sigma_i^{fused}. \tag{30}$$

5.5 Favorability and Steering

The predicted layer is obtained by typewell interval lookup. Favorability is computed as

$$P(Favorable_i) = I(Layer_i \in \mathcal{L}_{target}) \cdot GRScore_i \cdot \exp\left(-\frac{\sigma_i}{50}\right). \tag{31}$$

Low GR-score or high TVT uncertainty pushes the decision logic toward conservative actions such as holding or acquiring more GR.

6 Algorithm

Algorithm 1 Learning-Driven Autonomy for Horizontal Wellbore Geology Prediction

Require: Horizontal well data $\mathcal{H}$, typewell data $\mathcal{T}$, W_{raw}, L_{seq}, horizon H

Ensure: Fused TVT, uncertainty interval, GR-score, predicted layer, favorability, facies state, steering action

Load and sort the horizontal well by MD and the typewell by TVT.
Infer prediction start from the first missing `TVT_input`.
Interpolate missing GR and construct geometry, GR/log, and stratigraphic features.
Train ridge and tree baselines on the pre-start interval.
Compute ridge residual $r = TVT - \widehat{TVT}^{ridge}$.
Interpolate typewell GR at predicted TVT and compute GR-score.
Build rolling multimodal tensors.
Encode modalities and apply RFF-KME pooling to obtain fused tokens.
Encode fused tokens using GRU-attention or causal block-sparse attention.
Train residual CVAE on future $[r, GR]$ targets.
Reconstruct TVT as ridge prediction plus CVAE residual.
Fuse component TVT predictions using uncertainty-aware weights.
Inflate uncertainty using residual scale, model spread, GR variability, and GR-score mismatch.
Map fused TVT to typewell layer and compute favorability.
Detect persistent facies shift using rolling GR-score and residual consistency.
Output steering recommendation.

7 Complexities and Generalization

Ridge inference costs $O(P)$ per station, where P is the number of engineered features. RFF-KME tokenization costs $O(LWD)$, where L is sequence length, W is raw-window length, and D is the RFF dimension. GRU context encoding costs approximately $O(Ld^2)$. Full attention costs $O(L^2d)$, whereas block-sparse attention costs approximately $O(LBd)$, where $B \ll L$. Fusion costs $O(M)$ per station, where M is the number of component models.

For multiwell generalization, sequences must be constructed within each well, train/validation/test splits must be performed by well ID, and normalization parameters must be computed only from training wells. Well and typewell embeddings can augment the context vector:

$$c_{aug} = [c \| e_{well} \| e_{typewell}]. \tag{32}$$

Global tokens may be encouraged to specialize into structural, lithologic, uncertainty, and boundary roles using auxiliary heads and diversity regularization.

8 Experimental Setup and Results

8.1 Dataset and Prediction-Start Definition

The experimental study uses one horizontal-well dataset and one associated typewell dataset. The horizontal-well dataset contains 5,278 rows, and the typewell dataset contains 1,296 rows. Each horizontal-well row corresponds to a measured-depth station and includes trajectory coordinates, gamma ray, formation-surface information, interpreted TVT, and the `TVT_input` field.

The prediction-start point is defined as the first row where `TVT_input` becomes unavailable. This occurs at row 1,442, corresponding to MD = 12,909 ft. All rows before this point are treated as the calibration interval, while the remaining rows form the forward prediction interval. This choice mimics the operational geosteering setting in which manual or interpreted TVT is available up to a certain depth, after which the model must predict TVT and steering state from available measurements.

The sequence-modeling configuration uses $W_{raw} = 6$ raw samples per local window, $L_{seq} = 8$ recent windows as the context length, and a forecast horizon of $H = 6$ future steps. This rolling construction is used to generate multimodal sequential examples for the GRU-attention CVAE and the residual sparse-attention CVAE design.

8.2 Simulation 1: Data Quality Control and Feature Engineering

The first simulation stage performs data quality control and feature construction. Missing GR values are interpolated and an explicit missing-GR indicator is retained. Geometry features include horizontal distance, coordinate differences, vertical displacement, inclination proxy, azimuth proxy, and curvature proxy. GR/log features include rolling GR means, rolling GR standard deviations, GR gradient, and second-order GR gradient. Stratigraphic features include available formation surfaces and offsets between the wellbore vertical coordinate and each surface.

This simulation prepares the complete multimodal feature set used by both the tabular baselines and the sequence models.

8.3 Simulation 2: Ridge Regression Baseline

The ridge regression baseline is trained on the pre-start interval and then evaluated on the forward prediction interval. Ridge regression acts as the deterministic geometric backbone of the pipeline. The key result is that ridge regression is the strongest pure numerical TVT predictor, indicating that TVT in the studied well is primarily controlled by geometry and surface-relative features.

The ridge prediction also defines the residual target for the residual CVAE:

$$r_i = TVT_i - \widehat{TVT}_i^{ridge}.$$

This residual formulation is central to the proposed learning architecture.

8.4 Simulation 3: Tree-Ensemble Baselines

Random forest, extra trees, and gradient boosting models are trained as nonlinear tabular baselines. These models are included to test whether nonlinear feature interactions improve TVT prediction

beyond the ridge model. Although these methods are flexible, the experiments show that they underperform ridge on the forward interval. This suggests that the forward prediction task requires stable geometric extrapolation, which ridge handles more reliably than the tree models in this dataset.

8.5 Simulation 4: GR-Match and Typewell GR Interpolation

The typewell GR curve is interpolated in TVT space. For a predicted TVT value, the expected typewell GR is obtained by

$$GR^{type}(\widehat{TVT}_i) = \mathrm{Interp}(TVT^{type}, GR^{type}; \widehat{TVT}_i).$$

This interpolation is used to compute GR residuals and GR-score. The GR-match diagnostic evaluates whether the horizontal-well GR pattern agrees with the typewell reference. Because GR is nonunique and sensitive to facies changes, GR-match is treated as a lithologic diagnostic rather than a dominant TVT predictor.

8.6 Simulation 5: GR-Score Lithologic Calibration

GR-score is computed as

$$GRScore_i = \exp\left[-\left(\frac{GR_i^{hor} - GR^{type}(\widehat{TVT}_i)}{30}\right)^2\right].$$

A high GR-score indicates strong lithologic agreement between the horizontal-well GR and the typewell GR expected at the predicted TVT. A low GR-score indicates possible typewell mismatch, facies shift, boundary crossing, or noisy GR.

The GR-score is used in two ways. First, it directly reduces favorable probability when lithologic agreement is weak. Second, it can inflate uncertainty through a factor of the form

$$1 + \lambda_{GR}(1 - GRScore_i).$$

8.7 Simulation 6: GRU-Attention CVAE

The GRU-attention CVAE uses multimodal RFF-KME tokens as input. Each token summarizes local geometry, GR/log behavior, and stratigraphic context. The GRU encodes the sequence of tokens, and an attention layer learns which recent windows are most relevant to the forecast. The CVAE then predicts future TVT and GR probabilistically.

The experiment indicates that direct full-TV CVAE forecasting is weaker than ridge regression for pure TVT prediction. This motivates the residual CVAE formulation, where the model predicts corrections to the ridge baseline instead of full TVT.

8.8 Simulation 7: Residual Sparse-Attention CVAE Design

The residual sparse-attention CVAE replaces the GRU-attention context encoder with a causal block-sparse attention module. The model predicts future ridge residuals and GR:

$$p_\theta(r_{t+1:t+H}, GR_{t+1:t+H} \mid c_t).$$

The final TVT prediction is reconstructed as

$$\widehat{TVT}_{t+k}^{resCVAE} = \widehat{TVT}_{t+k}^{ridge} + \widehat{r}_{t+k}.$$

Sparse attention allows the model to attend to local blocks, previous neighboring blocks, and global tokens while preventing future leakage. This makes the model suitable for real-time geosteering.

8.9 Simulation 8: Uncertainty-Aware Fusion

The uncertainty-aware fusion simulation combines ridge, tree baselines, GR-match diagnostics, and residual-CVAE-style predictions into a single fused TVT estimate. Fusion weights are based on validation reliability or model precision. Local modifiers reduce model influence when local evidence is weak. For example, GR-match influence is reduced when GR-score is low, and tree-model influence is reduced when model spread is high.

The fused uncertainty combines validation residual scale, model spread, local GR variability, and GR-score mismatch. This provides a TVT uncertainty interval that can be used by the steering logic.

8.10 Simulation 9: Facies-State and Steering Diagnostics

Persistent facies-state diagnostics are computed from rolling GR-score, GR residual magnitude, and GR residual sign consistency. A single low GR-score may reflect noise, but persistent low GR-score with a consistent residual sign may indicate a facies shift, boundary ambiguity, or typewell mismatch.

Steering recommendations are generated from fused TVT, uncertainty, predicted layer, favorability probability, and GR-score. Low uncertainty and high GR-score support continuation, whereas low GR-score or high TVT uncertainty pushes the recommendation toward conservative actions such as holding or acquiring more GR.

8.11 Simulation 10: Rolling Performance and Calibration Analysis

Rolling performance analysis evaluates how TVT error, GR-score, and uncertainty vary along measured depth. This simulation identifies intervals where performance degrades and determines whether degradation is associated with low GR-score or high uncertainty. Calibration analysis groups predictions by predicted uncertainty and compares predicted uncertainty with observed error and empirical coverage.

8.12 Quantitative Model Comparison

Table 1 reports the diagnostic model comparison. The main full simulation reported ridge RMSE near 0.6014 ft, and the regenerated diagnostic pass yields a comparable ridge-dominant result. This confirms that ridge regression should be retained as the deterministic TVT backbone.

8.13 Description of Simulation Plots

Figure 1 provides a composite overview of the horizontal well. It includes the GR curve, manual and predicted TVT, the prediction-start marker, the uncertainty band, and the plan-view well trajectory. This plot verifies that the prediction problem is aligned with the measured-depth trajectory and provides a global visual summary of the workflow.

Table 1: Diagnostic model comparison. Ridge regression provides the strongest pure TVT prediction and is therefore used as the deterministic backbone for residual learning.

Model	RMSE_ft	MAE_ft	Bias_ft	R2
ridge	0.5876	0.5129	-0.4628	0.9867
gr_match	0.5876	0.5129	-0.4628	0.9867
uncertainty_fused	0.6389	0.5575	-0.5133	0.9842
extra_trees	5.3922	4.7932	-4.7932	-0.1220
random_forest	7.9903	7.0386	-7.0386	-1.4637
gradient_boosting	19.5486	17.0164	-17.0083	-13.7465

Figure 2 compares model RMSE values on the holdout interval. Ridge regression dominates the numerical TVT prediction task, confirming that the well's TVT is strongly geometry controlled.

Figure 3 shows the ridge residual, fused prediction error, model spread, and fused TVT uncertainty. This plot is used to evaluate whether residual errors are spatially structured and whether predicted uncertainty increases in difficult intervals.

Figure 4 shows GR-score, GR residual, fused TVT, and the TVT uncertainty band. This plot evaluates whether lithologic mismatch coincides with uncertainty growth or possible facies-state changes.

Figure 5 evaluates uncertainty calibration by grouping predictions according to predicted σ. The plot compares predicted uncertainty with observed absolute error and empirical 90% coverage. This diagnostic identifies whether uncertainty intervals are underconfident, overconfident, or approximately calibrated.

Figure 6 shows local adaptive fusion weights along measured depth. This plot illustrates how the fusion mechanism changes model trust spatially. Ridge generally receives high weight, while other models contribute locally depending on model spread and lithologic confidence.

Figure 7 compares GR-score with TVT uncertainty. This plot emphasizes that lithologic confidence and positional uncertainty are distinct calibration signals. A prediction may be geometrically confident but lithologically inconsistent, or lithologically plausible but positionally uncertain.

Figure 8 shows the causal block-sparse attention mask. Tokens are allowed to attend to the same block, previous neighboring blocks, and global tokens, but not to future tokens. This confirms that the sparse-attention design is compatible with real-time geosteering.

Figure 9 presents rolling RMSE, rolling GR-score, and normalized uncertainty along measured depth. This plot identifies intervals where model performance deteriorates and helps determine whether error growth is associated with low GR-score or high uncertainty.

Figure 10 shows the distribution of absolute TVT error grouped by predicted typewell layer. This diagnostic evaluates whether certain geological intervals are more difficult to predict than others.

Figure 11 shows facies-state labels along measured depth, colored by GR-score. This plot summarizes persistent lithologic mismatch and identifies regions where the system should reduce steering aggressiveness.

9 Discussion

The experiments show that ridge regression is the strongest pure numerical TVT predictor, confirming that the attached well's TVT is highly geometry controlled. However, ridge alone does not

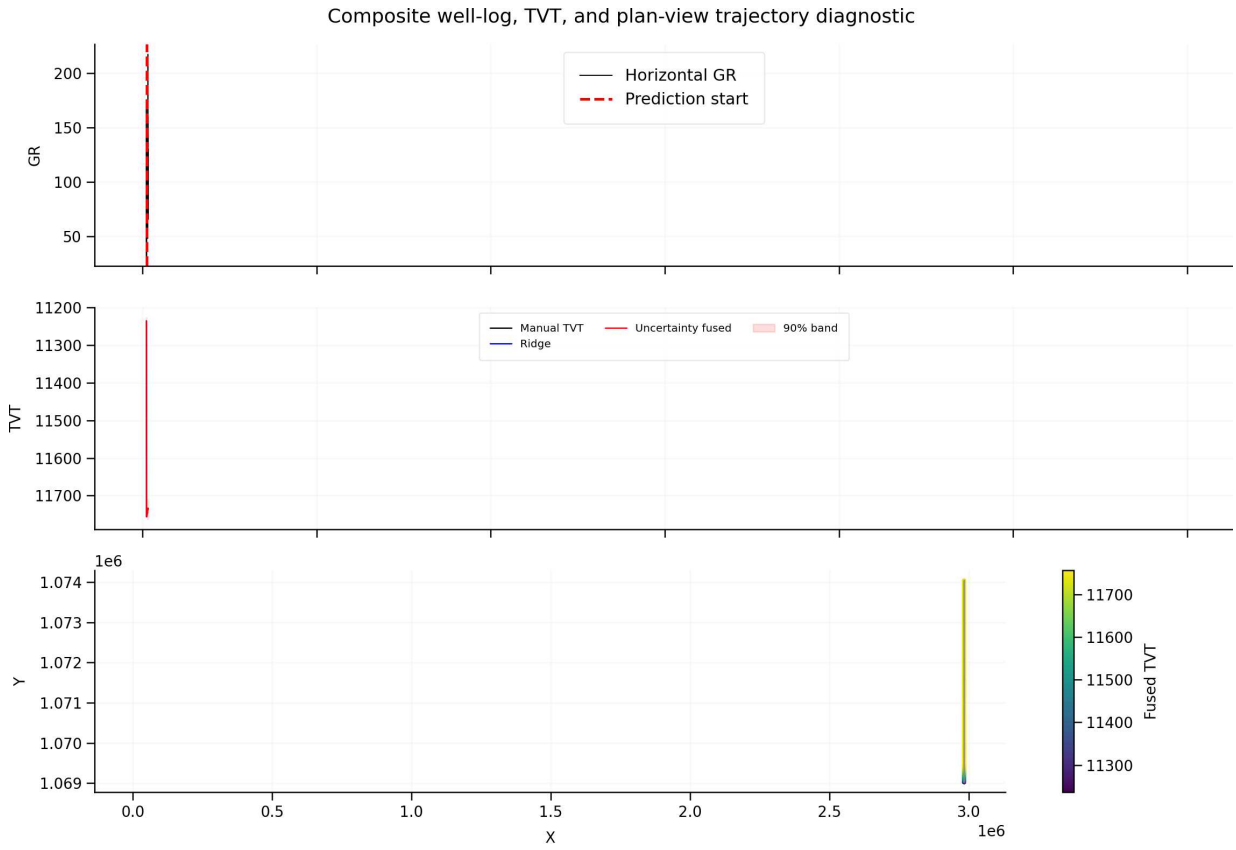


Figure 1: Composite diagnostic showing horizontal GR, manual and predicted TVT, prediction start, uncertainty band, and plan-view trajectory.

provide lithologic confidence, uncertainty-aware steering, or facies-state diagnostics. GR-score addresses this limitation by comparing horizontal-well GR with interpolated typewell GR at predicted TVT. Low GR-score reduces favorable probability and can inflate uncertainty.

The residual CVAE formulation improves over full-TV CVAE because the CVAE does not need to learn the entire TVT trajectory from scratch. Ridge captures the dominant geometric trend, while the CVAE learns structured deviations, future residual uncertainty, and GR-related context. Sparse attention improves residual forecasting by providing causal, interpretable, selective memory over local and global wellbore context. Same-block attention captures local GR texture, neighboring-block attention captures transitions, and global tokens capture longer-range structural or lithologic state.

Uncertainty-aware fusion is useful but has limitations. It can underestimate uncertainty if all component models share the same bias. Heuristic GR-score inflation factors must be calibrated carefully. Low model spread indicates model agreement, not necessarily geological correctness. Therefore, calibration should be evaluated by uncertainty bins, GR-score bins, predicted layers, facies-state intervals, and rolling measured-depth windows.

The diagnostic plots demonstrate that TVT prediction, lithologic confidence, and uncertainty should be interpreted jointly. A high-confidence steering recommendation should require both low TVT uncertainty and high GR-score. Conversely, low GR-score or high uncertainty should trigger conservative steering behavior.

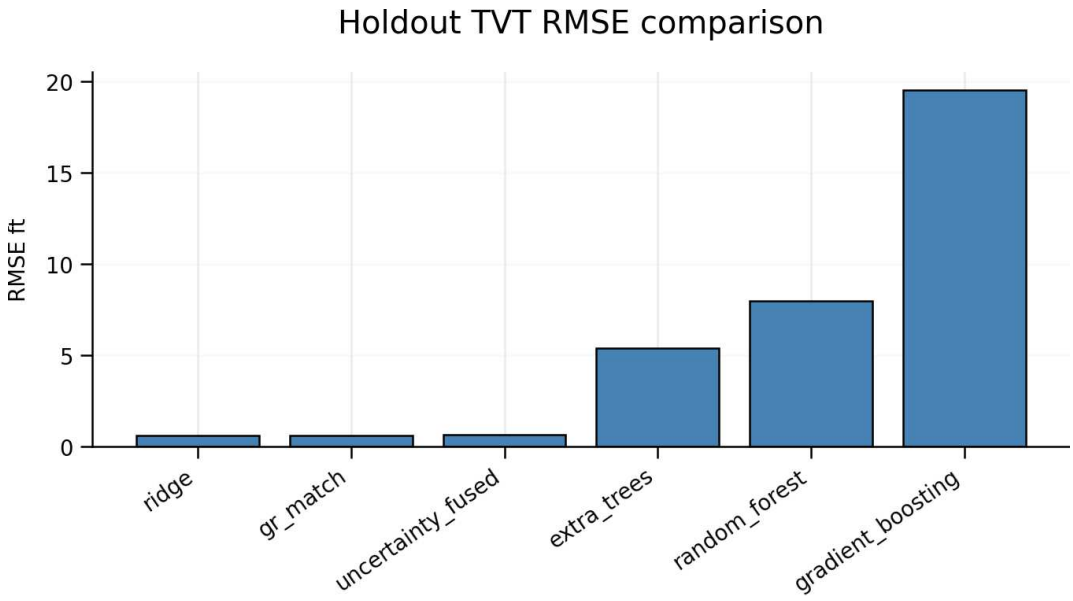


Figure 2: Holdout TVT RMSE comparison. Ridge is the dominant numerical TVT backbone.

10 Conclusion

This paper presents a learning-driven autonomy pipeline for horizontal wellbore geology prediction. The strongest empirical finding is that ridge regression is the most accurate pure TVT predictor for the attached dataset, while the autonomy layer adds decision-critical lithologic confidence, uncertainty quantification, facies-state detection, and steering logic.

The recommended architecture is a ridge baseline combined with causal sparse-attention residual CVAE, GR-score or DTW-based lithologic confidence, uncertainty-aware fusion, and conservative steering. Future work should validate the framework on multiwell datasets, incorporate well and typewell embeddings, calibrate uncertainty using Bayesian or conformal methods, replace pointwise GR-score with windowed or DTW-based lithologic confidence, and deploy optimized sparse or linear attention for long-context field-scale geosteering.

Declarations

Funding

A Funding Declaration has been added to the manuscript. The authors declare that no specific funding was received for this study.

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

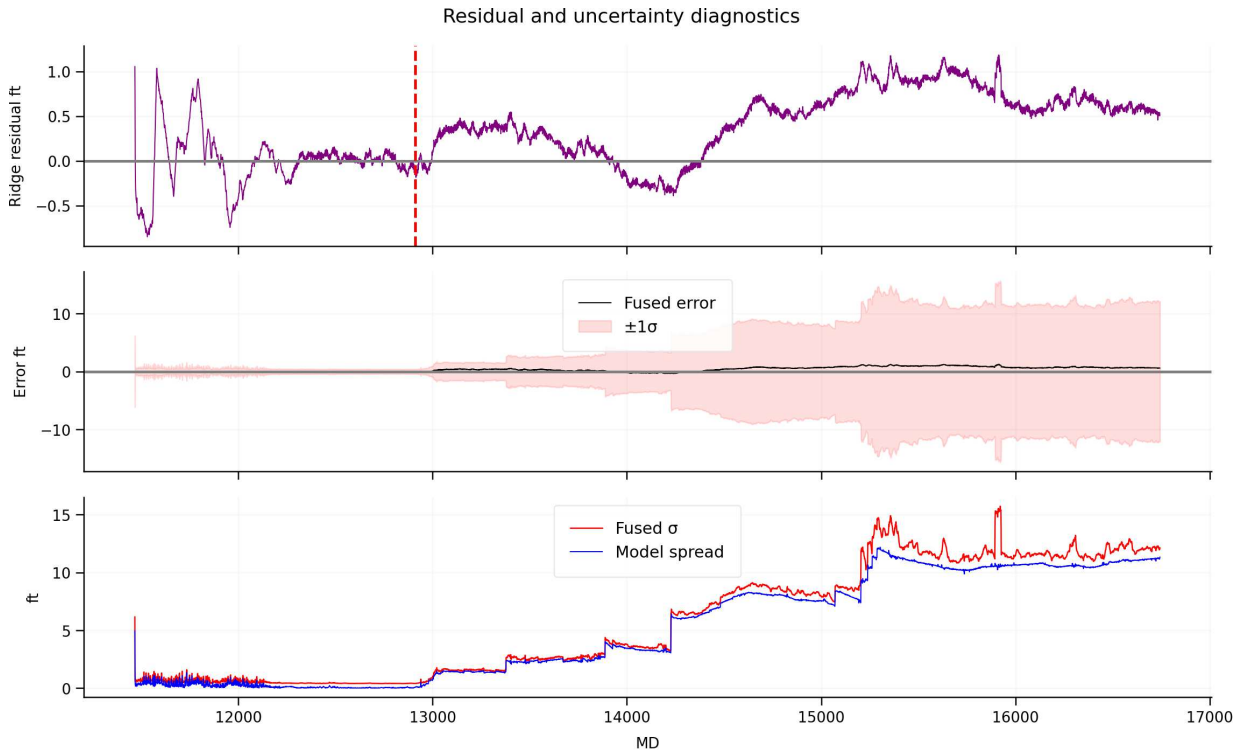


Figure 3: Residual and uncertainty diagnostics, including ridge residual, fused prediction error, model spread, and fused TVT uncertainty.

Availability of Data and Materials

The manuscript will be uploaded in an editable format. If the LaTeX version is submitted, the LaTeX source file along with all figures and tables will be included in a compressed zip file.

Competing Interests

The author declares no competing interests.

Author Contributions

Lakshman Mahto conceptualized the study, developed the methodology, performed the analysis, prepared the manuscript, and approved the final version.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *International Conference on Learning Representations*, 2015.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*, 2020.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

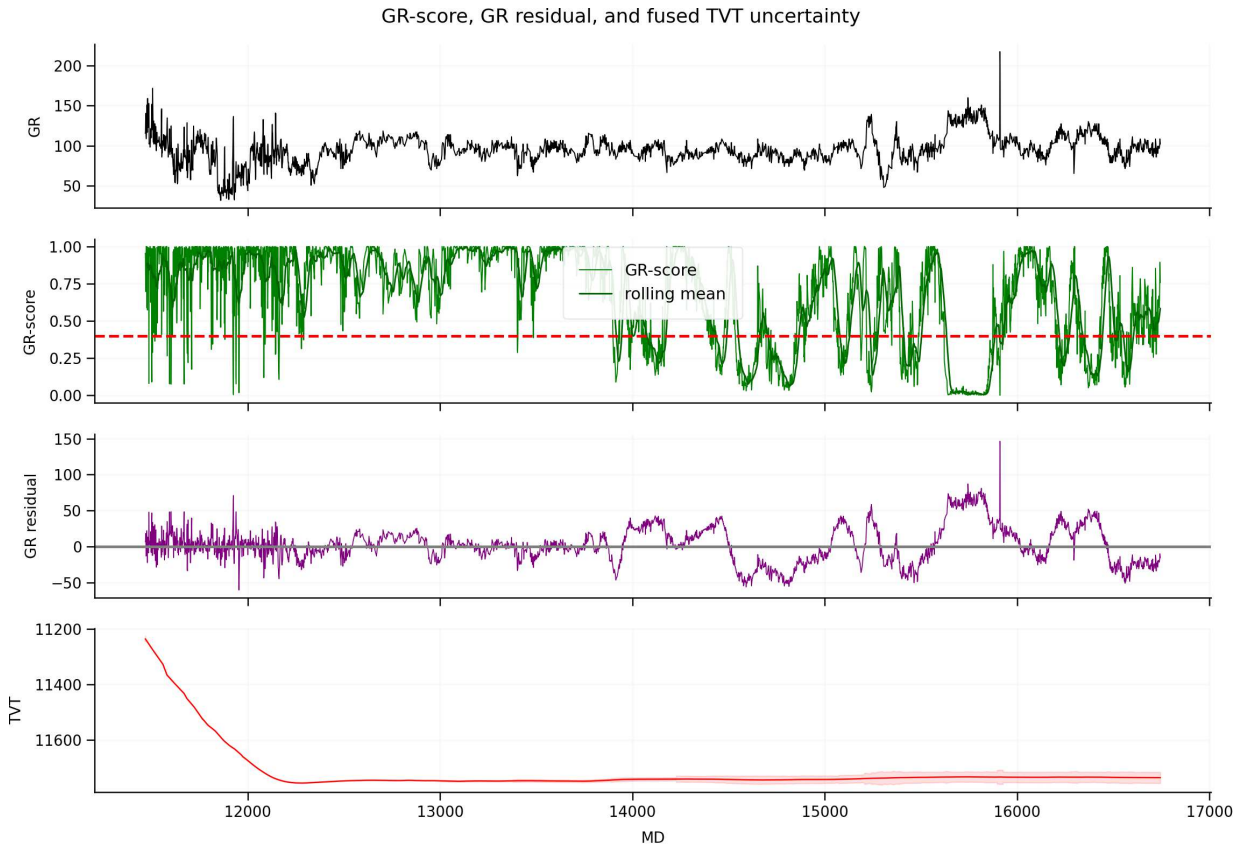


Figure 4: GR-score dashboard showing GR, GR-score, GR residual, fused TVT, and uncertainty band.

Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. Generating long sequences with sparse transformers. In *arXiv preprint arXiv:1904.10509*, 2019.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1724–1734, 2014.

Krzysztof Choromanski, Valerii Likhosherstov, David Dohan, Xingyou Song, Andreea Gane, Tamas Sarlos, Peter Hawkins, Jared Davis, Afroz Mohiuddin, Lukasz Kaiser, David Belanger, Lucy Colwell, and Adrian Weller. Rethinking attention with performers. In *International Conference on Learning Representations*, 2021.

Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.

Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine Learning*, 63(1):3–42, 2006.

Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

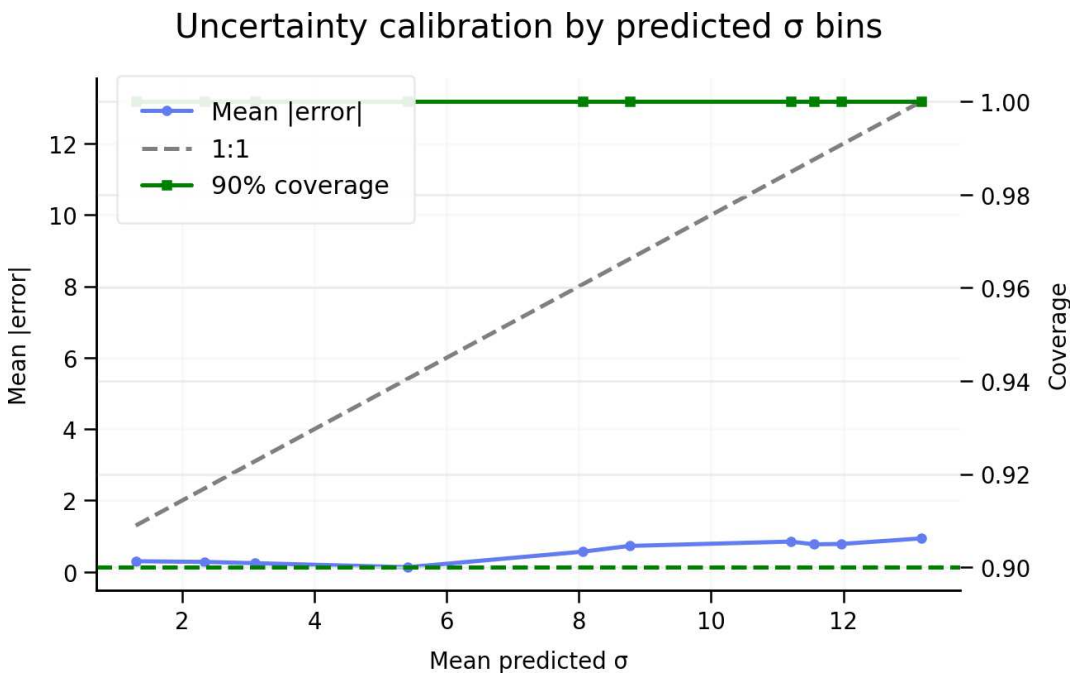


Figure 5: Uncertainty calibration by predicted σ bins.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

Angelos Katharopoulos, Apoorv Vyas, Nikolaos Pappas, and François Fleuret. Transformers are rnns: Fast autoregressive transformers with linear attention. In *Proceedings of the 37th International Conference on Machine Learning*, pages 5156–5165, 2020.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014.

Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, and Bernhard Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning*, 10(1–2):1–141, 2017.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems*, volume 20, 2007.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9:371–421, 2008.

Alex Smola, Arthur Gretton, Le Song, and Bernhard Schölkopf. A hilbert space embedding for distributions. In *Algorithmic Learning Theory*, pages 13–31. Springer, 2007.

Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

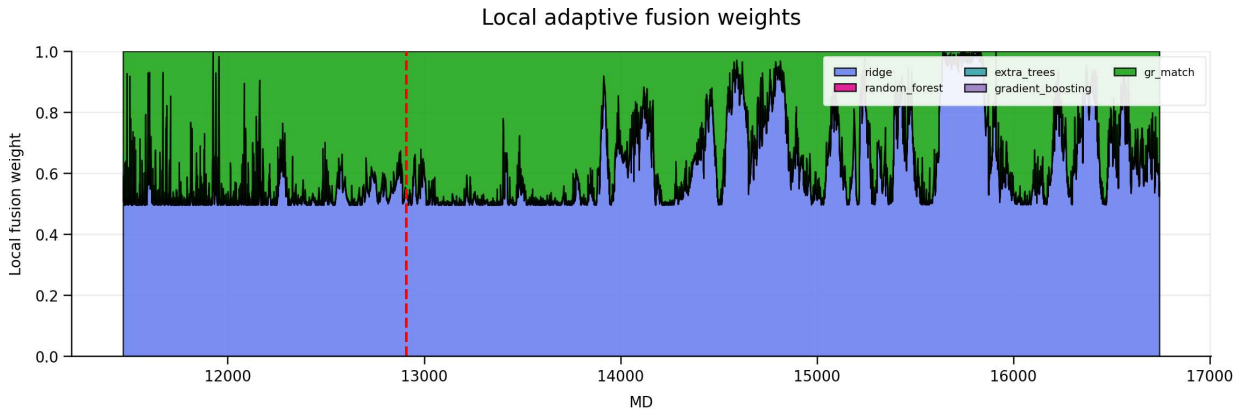


Figure 6: Local adaptive fusion weights along measured depth.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.

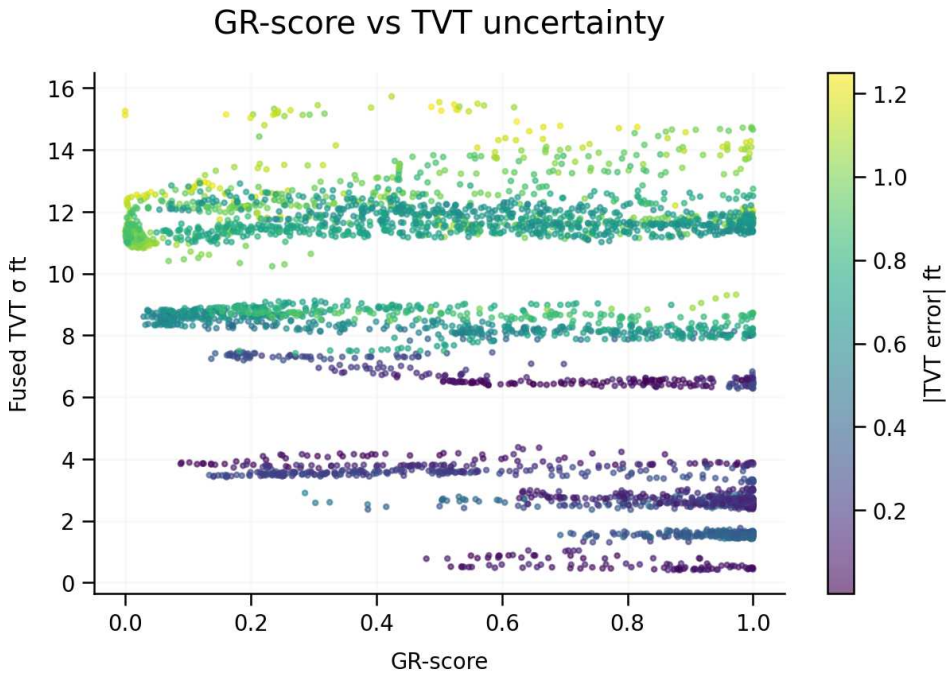


Figure 7: GR-score versus TVT uncertainty. Lithologic confidence and positional uncertainty are distinct but interacting calibration signals.

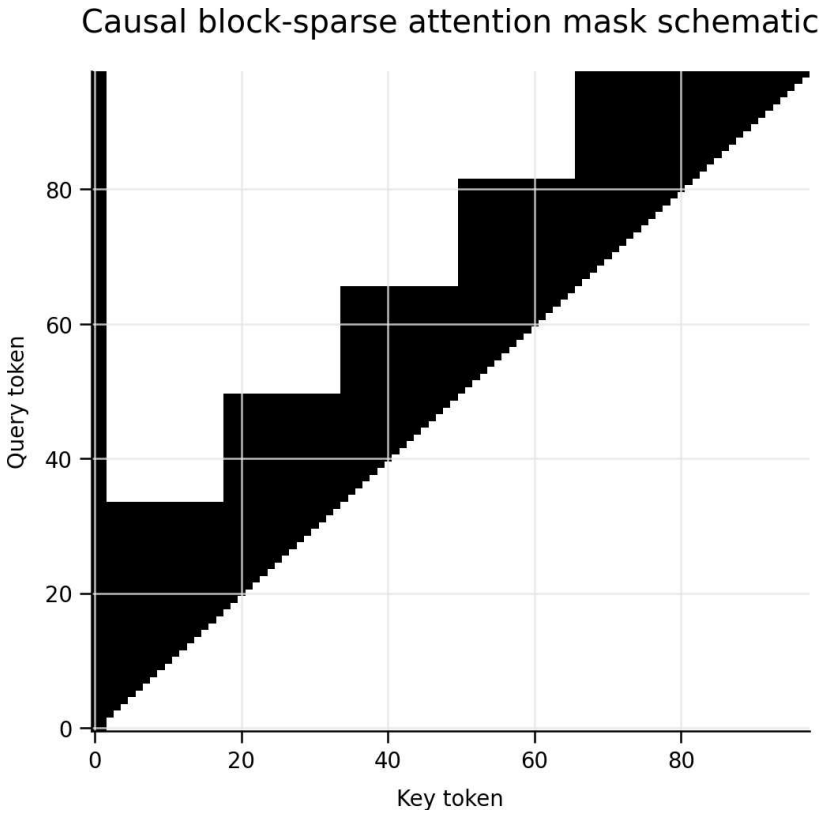


Figure 8: Causal block-sparse attention mask schematic. Tokens attend to the same block, the previous block, and global tokens, but not to future tokens.

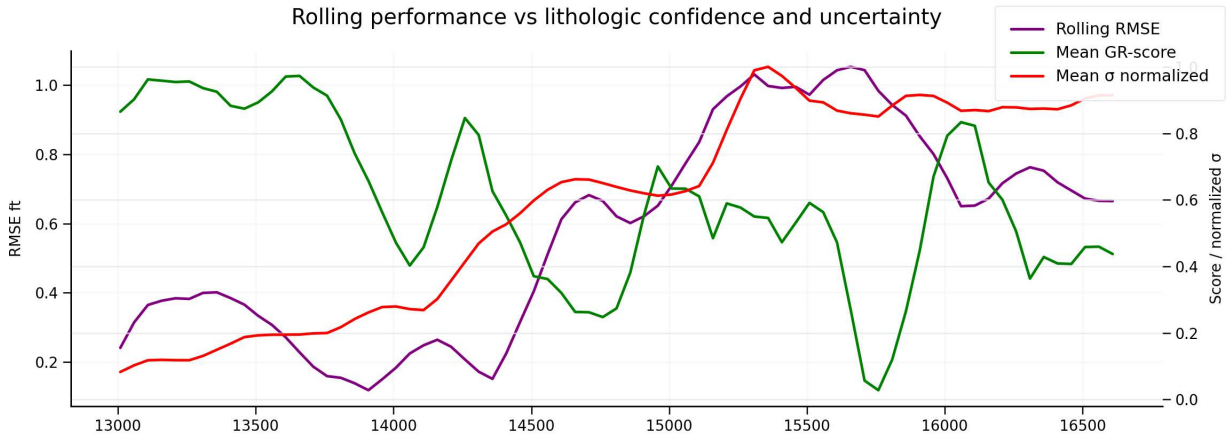


Figure 9: Rolling performance dashboard comparing rolling RMSE, rolling GR-score, and normalized uncertainty.

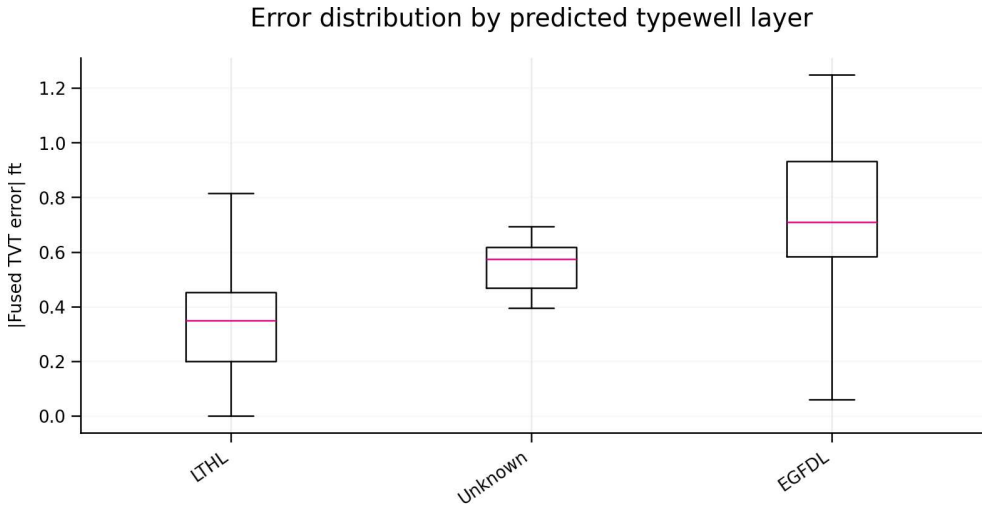


Figure 10: Error distribution by predicted typewell layer.

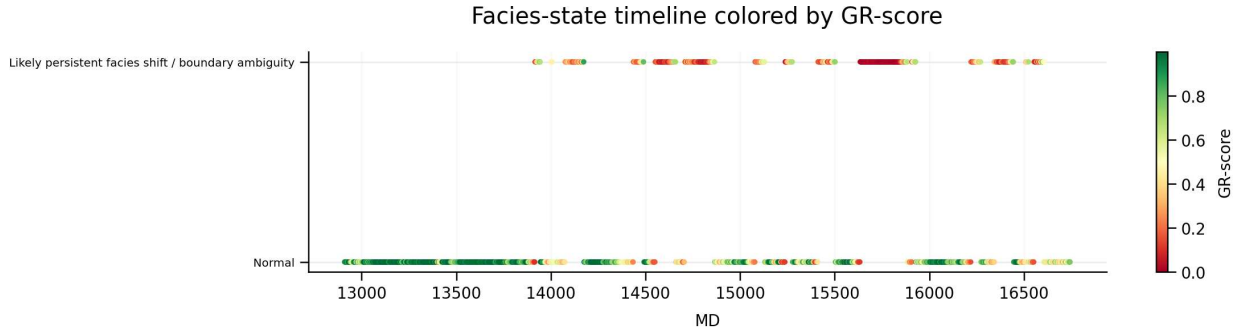


Figure 11: Facies-state timeline colored by GR-score.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- diagnosticmetrics.csv
- uncertaintycalibrationbins.csv
- rollingperformance.csv
- diagnosticpredictions.csv