

ARTICLE

Physics-guided well-family multi-expert ensemble model for in situ stress prediction using well logs

Naihao Liu^{1,2}, Yuwei Liu^{1,2*}, Xiwu Liu^{1,2}, Yongyi Huang³, and Xinyu Wang^{1,2}

¹State Key Laboratory of Shale Oil and Gas Enrichment Mechanisms and Efficient Development, Beijing, China

²Key Laboratory of Oil & Gas Reservoir Geophysics, SINOPEC, Beijing, China

³Department of Computer Science, College of Information Engineering, Northwest A&F University, Yangling, Shaanxi, China

Abstract

Accurate estimation of *in situ* stress from conventional well logs is challenging when stress measurements are available for only a few wells and blind wells differ in pressure regime, logged interval, or curve baseline. We propose a physics-guided well-family multi-expert ensemble for predicting maximum and minimum horizontal *in situ* stresses (SHMAX and SHMIN). Conventional logs are augmented with overburden and effective-stress proxies, pressure interactions, local curve descriptors, and well-level statistics. The model comprises Direct, depth-normalized, and Physics-residual extremely randomized trees (ExtraTrees) experts trained on 10 development wells. Development-well leave-one-well-out predictions determine the three-expert blend weights, whereas an absolute 54 MPa/1,600 m routing rule is frozen before final expert fitting and blind-well evaluation. Labels from the eight blind test wells are not used for feature construction, weight selection, expert fitting, route assignment, or prediction. In cumulative ablation, the sample-weighted average mean absolute error (MAE) decreases from 1.8437 MPa for raw logs plus ExtraTrees to 1.3776 MPa after physics-guided feature construction, 1.2564 MPa after multi-expert blending, and 0.9589 MPa after the complete fixed 54 MPa/1,600 m gate. The final SHMAX and SHMIN MAEs are 1.0284 and 0.8894 MPa, respectively; the average MAE is 0.9589 MPa, and the average R^2 is 0.9819. These results indicate complementary contributions from physics-guided representation, target-space diversification, and well-level routing; however, generalizability remains limited by the single-basin, 18-well dataset.

*Corresponding author:

Yuwei Liu
(liuyw.syky@sinopec.com)

Citation: Liu N, Liu Y, Liu X, Huang Y, Wang X. Physics-guided well-family multi-expert ensemble model for in situ stress prediction using well logs. *J Seismic Explor.* doi: 10.36922/JSE026270122

Received: July 1, 2026

Revised: August 22, 2026

Accepted: August 24, 2026

Published online: October 28, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Keywords: *In situ* stress prediction; ExtraTrees; Physics-guided learning; Multi-expert ensemble; Well-family gating

1. Introduction

In situ stress is required for borehole-stability analysis, hydraulic-fracture design, fault-reactivation assessment, and geomechanical evaluation during drilling and stimulation. These applications have been discussed in both foundational and recent studies.¹⁻⁵ Horizontal stress can be constrained by hydraulic-fracturing and leak-off tests, borehole breakouts, drilling-induced tensile fractures, and core measurements. These observations

are sparse, however, and seldom match the vertical sampling density of field-scale models.^{6,7} Continuous prediction therefore requires the available stress observations to be related to lithological, pressure, density, and acoustic logs.^{8,9} Well logs provide depth-continuous responses to burial, compaction, lithology, and pressure. Density is related to overburden loading, pore pressure (PP) controls effective loading, sonic and neutron-porosity logs respond to compaction and pore structure, and gamma ray (GR) provides lithological information. A useful model should retain these first-order physical relations while learning nonlinear log-stress associations from data.¹⁰⁻¹³

Mechanical earth models (MEMs) commonly combine density integration, PP, elastic properties, tectonic strain, and local calibration to constrain the stress field.^{14,15} This approach is preferred when complete logs and reliable calibration measurements are available. The problem considered here is more restricted. We seek to predict continuous stress curves in blind wells when the available wells have unequal logged intervals and different curve baselines. Under this condition, a data-driven model must preserve a plausible stress scale without assuming local stress anchors in every target well.

Tree ensembles and neural networks have been used to model nonlinear relations among well logs, reservoir properties, PP, and *in situ* stress.¹⁶⁻¹⁸ Related studies have examined missing-log reconstruction, basin-scale reservoir-property prediction, and pre-trained time-series models for transfer across heterogeneous wells.¹⁹⁻²³ Sparse labels and cross-well distribution shift remain coupled problems. A model may interpolate samples within familiar wells yet retain a systematic magnitude offset in an unseen well. Foundation-model pre-training and domain adaptation require large multi-basin datasets, which are unavailable for a specific study area with few wells, such as the present 18 wells used in this study. We therefore explore a lower-capacity alternative that combines interpretable predictors with transparent well-level routing. In addition, sparse stress observations constrain regional inference. Bayesian and statistical studies show that uncertainty depends on both the number and spatial distribution of stress measurements.^{24,25} This motivates a well-disjoint evaluation. Adjacent depth samples from one well cannot serve as independent evidence of transfer to another well, and target-well labels must remain outside feature construction, route selection, and model fitting.

We propose a physics-guided well-family-gated ensemble (PG-WFME) for predicting maximum horizontal *in situ* stress (SHMAX) and minimum horizontal *in situ* stress (SHMIN), which comprises physics-guided feature construction, three extremely randomized trees

(ExtraTrees) experts, development-stage leave-one-well-out (LOWO) weight selection, and a frozen label-free well-family gate. The study proceeds in four stages. First, physics-guided features expose overburden- and pore-pressure-controlled stress scales to the learner. Second, three ExtraTrees experts learn direct stress, depth-normalized stress growth, and residual correction in complementary target spaces. Third, aggregate LOWO predictions from development wells determine a fixed expert blend and a label-free gate before blind-well evaluation. Fourth, we report component ablation, threshold and coefficient sensitivity, paired well-level uncertainty, and measured runtime. The study tests a constrained ensemble under a small-well protocol; it does not claim basin-independent generalization.

2. Methods

2.1. Overall architecture

The PG-WFME workflow comprises physics-guided feature construction, three ExtraTrees experts, development-stage LOWO weight selection, and a frozen label-free well-family gate, as presented in Figure 1. The Direct expert predicts absolute SHMAX and SHMIN. The Gradient expert, retained as a concise branch name, predicts stress normalized by depth rather than a finite-difference derivative and then back-transforms the output to stress units. The Physics-residual expert estimates a correction to an empirical vertical stress–pore pressure (SV–PP) baseline. During blind-well inference, the frozen pipeline uses only conventional logs and derived descriptors. The blind-well SHMAX and SHMIN labels are separated from the input data before feature construction, excluded from all modeling and prediction procedures, and used only for final evaluation after the predicted curves and route assignments have been finalized.

Figure 1 is a compact architectural map rather than a fold-by-fold training diagram. The input block retains DEPTH, bulk density (RHOB), GR, neutron porosity (NPHI), compressional slowness (DT), and PP and expands them with physics-guided stress proxies, local gradients, curve-shape descriptors, and well-level statistics. The expert bank learns three complementary target spaces, i.e., absolute stress, depth-normalized stress, and residual correction relative to the SV–PP baseline. The final gate uses only mean PP and the complete logged interval. No stress label from a blind well enters feature construction, expert fitting, route assignment, or curve prediction.

Development-stage model selection uses 10 LOWO folds. In each fold, the 3 experts are reinitialized, fitted on 9 development wells, and evaluated on the entire held-out well. This produces 30 temporary model instances.

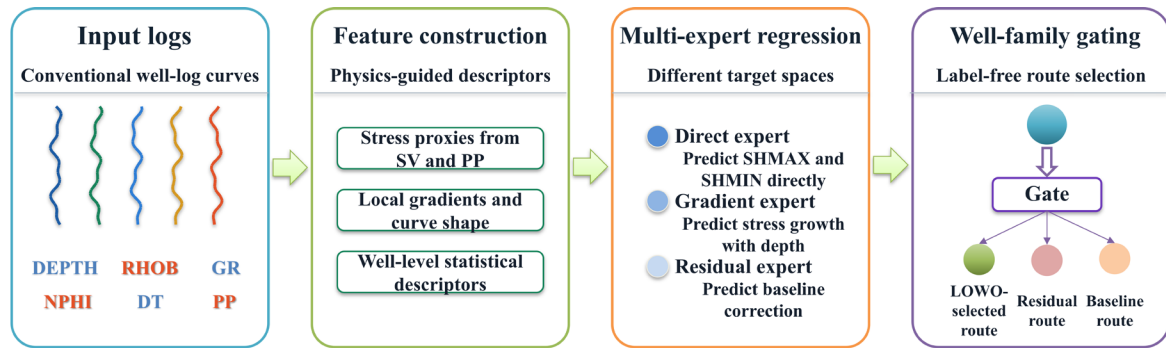


Figure 1. The proposed PG-WFME architecture from conventional logs through physics-guided feature construction, multi-expert regression, and label-free well-family routing. Development-stage LOWO selection and frozen blind-well inference are detailed in the text.

Abbreviations: DT: Compressional slowness; GR: Gamma ray; LOWO: Leave-one-well-out; NPHI: Neutron porosity; PP: Pore pressure; PG-WFME: Physics-guided well-family-gated ensemble; RHOB: Bulk density; SHMAX: Maximum horizontal *in situ* stress; SHMIN: Minimum horizontal *in situ* stress; SV: Vertical stress.

Their outputs are concatenated by expert type into three complete development-set out-of-fold (OOF) prediction sets, i.e., Direct OOF, Gradient OOF, and Residual OOF. The temporary model objects are not merged. The three OOF sets are used to select the blend weights, after which the temporary models are discarded, and the final experts are reinitialized.

2.2. Physics-guided feature construction

Feature construction is designed to make the machine-learning model consistent with basic geomechanical reasoning. The original logging suite is not discarded, because each curve has a distinct physical meaning in the stress-prediction task. RHOB is related to overburden load, PP controls the effective loading environment, DT and NPHI respond to compaction and pore-network characteristics, and GR helps distinguish lithological variations. These raw inputs are expanded into four groups of predictors, i.e., depth-normalized attributes, physics-guided stress proxies, local-shape descriptors, and well-level statistical descriptors. The depth-normalized attribute is denoted by $z_{d,i}$, it is derived from the measured depth z_i and is used to describe relative burial position as a model input. The purpose of this expansion is not to replace a MEM, but to provide the regression model with variables that are easier to interpret and more stable across wells than raw logs alone. This is particularly useful for tree-based models, because the split rules are learned from feature values rather than from an explicit physical equation. When physically meaningful quantities are supplied as inputs, the ensemble can use them as stable anchors while still learning nonlinear corrections from the data.

The main physics-guided attribute is a vertical-stress proxy (SV). In a full geomechanical workflow, SV is obtained by integrating the density log from the surface to

the depth of interest. Because the present dataset contains wells with different logged intervals, the available RHOB curve is converted into a scaled overburden proxy after unit normalization. This approximation does not serve as a complete MEM calculation. Rather, it anchors the tree-based regression to a reasonable stress scale before nonlinear corrections are learned from the data. The corresponding vertical-stress and effective-stress proxies are expressed as:

$$\sigma_{v,i} = \rho'_{b,i} g z_i, \quad g = 9.80665 \text{ m s}^{-2} \quad (1)$$

$$\sigma_{eff,i} = \sigma_{v,i} - P_{p,i} \quad (2)$$

where $\sigma_{v,i}$ is the vertical-stress proxy at measured depth z_i , $\rho'_{b,i}$ is the bulk-density log value at the same depth after unit normalization and scaling from raw RHOB, g is gravitational acceleration, z_i is the measured depth used in the overburden calculation, $\sigma_{eff,i}$ is the effective-stress proxy, and $P_{p,i}$ is the pore-pressure value at measured depth z_i . In this workflow, $\sigma_{v,i}$ is used as a scaled overburden proxy after unit normalization rather than as a full mechanical-earth-model integration. The measured-depth symbol z_i is therefore different from $z_{d,i}$, i.e., z_i enters the physical stress-proxy calculation, whereas $z_{d,i}$ is the normalized depth coordinate supplied to the machine-learning feature vector.

Additional composite attributes are calculated to describe pressure-lithology coupling that may be difficult to infer from any single log curve. These attributes include the effective-stress proxy, pressure ratios, and cross-log products or ratios. To help the tree experts distinguish long-wavelength geomechanical trends from local log variations, rolling averages and first-order vertical derivatives are computed over several window lengths. The rolling descriptors reduce the influence of isolated log spikes,

whereas the gradient descriptors retain local turning points and interval-scale changes. Well-level statistics, including the mean, standard deviation, and selected percentiles, are also calculated for each curve and assigned to all samples from the same well. These static descriptors give the model information about the broader well context, for example whether a blind well is relatively shallow, overpressured, thickly logged, or highly compacted, without using its *in situ* stress labels. In this sense, the feature set contains both sample-scale information and well-scale context, which is essential for cross-well transfer.

For sample i , the final predictor vector combines normalized depth, raw logs, physics-guided attributes, local curve-shape descriptors, and well-level statistics, presented as

$$\mathbf{x}_i = [z_{d,i}, \rho_{b,i}, GR_p, NPHI_p, DT_p, P_{p,i}, \sigma_{v,i}, \sigma_{eff,i}, \mathbf{r}_p, \mathbf{g}_p, \mathbf{w}_i]^T \quad (3)$$

where $\mathbf{x}_i$ denotes the assembled predictor vector at the i -th depth sample; $z_{d,i}$ is the depth-normalized coordinate derived from measured depth z_i and is used only as a model input; $\rho_{b,i}$, GR_p , $NPHI_p$, DT_p and $P_{p,i}$ are the conventional well-log inputs representing formation RHOB, gamma-ray lithological response, neutron-porosity-related pore-volume information, acoustic slowness associated with compaction and elastic stiffness, and pore-pressure condition, respectively; $\sigma_{v,i}$ and $\sigma_{eff,i}$ are physics-guided stress proxies derived from density- and pressure-related loading terms; $\mathbf{r}_i$ contains pressure ratios and cross-log interaction terms; $\mathbf{g}_i$ contains local gradient and curve-shape descriptors, including first-order depth derivatives, rolling-window statistics, and interval-scale morphology indicators; $\mathbf{w}_i$ contains well-level statistical descriptors, such as logged-interval length, well-wise means, standard deviations, and selected percentiles; and T denotes vector transposition. Figure 1 uses the compact symbols z_d , ρ_b , P_p , σ_v , σ_{eff} , $\mathbf{r}$, $\mathbf{g}$, and $\mathbf{w}$ to keep the workflow readable at manuscript scale, while the full sample-level notation is given in Equation 3.

2.3. Multi-expert regression

We use ExtraTrees as the common regressor because they can model nonlinear feature interactions in small tabular datasets while reducing the variance of a single tree through randomized ensemble averaging.^{26,27} All three experts receive the same predictor vector but use different supervised targets. The Direct expert predicts the two reference stress components. The Gradient expert learns stress divided by depth in kilometers and converts its output back to stress by multiplication with depth. The Physics-residual expert learns the difference between

measured stress and the empirical SV-PP baseline. Figure 2 expands the residual expert only; it does not depict the complete expert bank or the final gate.

The multi-expert design learns the same physical quantity in three target spaces. The Direct expert predicts the two stress components directly. The Gradient expert predicts depth-normalized stress and is inverse-transformed before combination. The Physics-residual expert predicts the residual relative to an empirical SV-PP baseline. In Figure 2, M randomized trees produce two-component outputs $\mathbf{h}_m(\mathbf{x}_{i-1})$, and their arithmetic mean is the residual correction. This is an intra-expert tree average, not an average over the three experts. Adding the correction to the baseline yields only the Physics-residual route candidate. The development-selected Blend and the frozen well-family gate operate afterward. The empirical baseline and residual target are written as:

$$\mathbf{y}_{base,i} = [0.70\sigma_{v,i} + 0.30P_{p,i}, 0.5\sigma_{v,i} + 0.45P_{p,i}]^T \quad (4)$$

$$\mathbf{r}_i = \mathbf{y}_i - \mathbf{y}_{base,i}$$

$$D_{res} = \{(\mathbf{x}_j, \mathbf{r}_j) \mid j \text{ in } D_{train}\}, \mathbf{f}_{res} = \text{ExtraTrees}(D_{res})$$

$$\hat{\mathbf{r}}_i = \mathbf{f}_{res}(\mathbf{x}_i) \quad (5)$$

$$\hat{\mathbf{y}}_i = \mathbf{y}_{base,i} + \hat{\mathbf{r}}_i = \mathbf{y}_{base,i} + \mathbf{f}_{res}(\mathbf{x}_i)$$

Here, Equation 4 defines an empirical two-component baseline for SHMAX and SHMIN at sample i . The two components correspond to the SHMAX and SHMIN baseline values, respectively. The coefficients 0.70/0.30 and 0.55/0.45 are prescribed stress-scale weights rather than calibrated mechanical constants; they keep the overburden- and pore-pressure-controlled stress scale explicit while leaving the nonlinear correction to the regression model. In Equation 5, the first line defines the residual target as the measured stress vector minus the empirical baseline. The residual training set D_{res} is then formed from labeled training-well samples, where each pair contains the feature vector $\mathbf{x}_j$ and the corresponding residual $\mathbf{r}_j$. Here, D_{res} denotes the residual training set used by the physics-residual expert, D_{train} denotes the labeled training-well sample set used for model fitting, and j indexes an individual labeled depth sample in this set. For each such sample, $\mathbf{x}_j$ is the log-derived feature vector, including conventional logging inputs, physics-guided stress proxies, local curve-shape descriptors, and well-level statistical descriptors, whereas $\mathbf{r}_i$ is the residual target calculated as $\mathbf{r}_i = \mathbf{y}_i - \mathbf{y}_{base,i}$. Thus, D_{res} contains only paired feature-residual samples constructed from labeled training wells and does not include blind-well target labels. $\mathbf{f}_{res}$ denotes the residual mapping learned by the physics-residual ExtraTrees expert from D_{res} . It is not an analytical

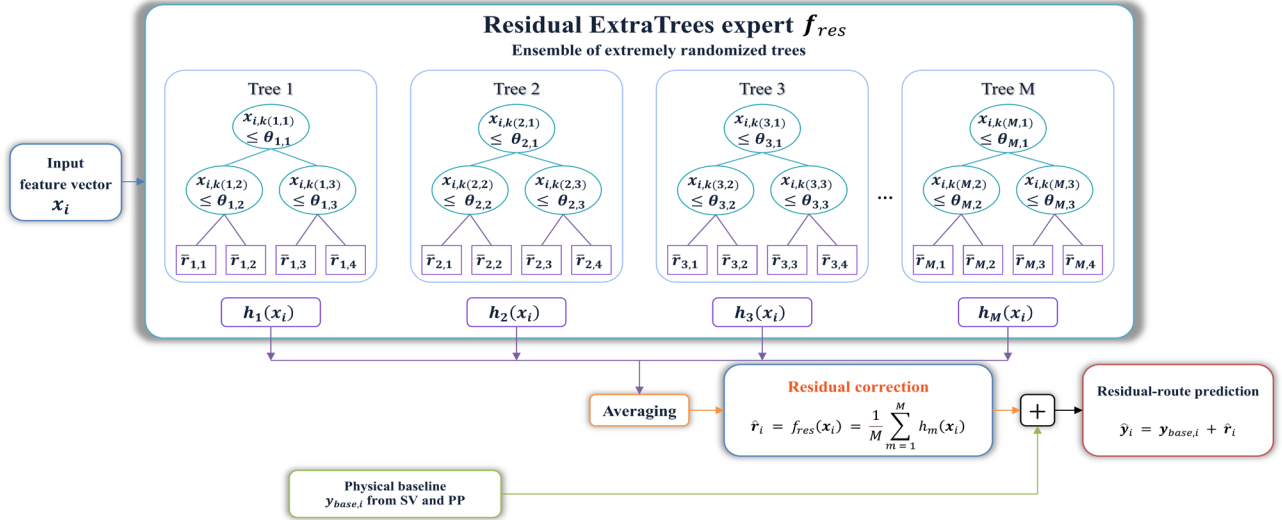


Figure 2. The structure of the Physics-residual ExtraTrees expert. Each tree returns a two-component SHMAX/SHMIN residual vector; their arithmetic mean is added to the empirical SV–PP baseline to form the residual-route prediction. Three-expert blending and well-family gating occur afterward. Abbreviations: PP: Pore pressure; SV: Vertical stress.

stress equation, but a data-driven nonlinear regression function trained to approximate the mapping from log-derived features to the baseline correction. During blind-well prediction, the true stress vector y_i and the residual target r_i are unavailable; the model only computes x_i and $y_{base,i}$ from well logs and uses the trained residual expert to produce $\hat{r}_i = f_{res}(x_i)$. The final predicted stress vector is obtained by adding this residual correction back to the empirical baseline, as shown in the last line of **Equation 5**.

2.4. Unlabeled well-family gating

We use a low-capacity gate with three candidate routes, i.e., the Direct expert, the development-selected three-expert Blend, and the Physics-residual expert. Before prediction, the model calculates the mean PP and complete logged interval from each blind well. The reported experiment uses the original absolute operating thresholds of 54 MPa and 1,600 m, as detailed below. These values are stored in the frozen configuration before final expert fitting and final metric computation. A 5×5 development-OOF grid, with PP thresholds of 50, 52, 54, 56, and 58 MPa and interval thresholds of 1,200, 1,400, 1,600, 1,800, and 2,000 m, is reported as a sensitivity analysis rather than as a source of retrospective test-well optimization. The fixed 54 MPa/1,600 m point gives a development LOWO equal-well mean absolute error (MAE) of 4.2252 MPa; the grid minimum is 4.1896 MPa at 56 MPa/1,800 m, a difference of 0.0357 MPa.

These thresholds should be reassessed for different well-log datasets and study areas.

$$M(w) = \begin{cases} \text{LOWO-weighted blend,} & \bar{P}_{p,w} \geq 54 \text{ MPa} \\ \text{Physics-residual route,} & \Delta z_w \geq 1600 \text{ m} \\ \text{Direct route,} & \text{otherwise} \end{cases} \quad (6)$$

The LOWO model selection is performed entirely within the 10-well development set. In each of 10 folds, 3 newly initialized ExtraTrees experts are fitted on nine wells and generate whole-well OOF predictions for the remaining well. Thus, LOWO trains $10 \times 3 = 30$ temporary model instances. Predictions are concatenated by expert type to form three complete development OOF sets rather than combining the 30 model objects. We searched 5,151 nonnegative Direct/Gradient/Physics-residual weight triples at a step of 0.01 under the sum-to-one constraint. This search minimizes sample-weighted average MAE across all development depth samples and selects 0.08/0.64/0.28. The absolute routing thresholds are then held fixed at 54 MPa and 1,600 m; development OOF scores are used to quantify their local sensitivity.

After the blend weights, routing rule, and hyperparameters are frozen, all 30 temporary LOWO models are discarded. Three final experts are reinitialized and fitted on all 10 development wells. Each blind test well then receives Direct, Gradient, and Physics-residual predictions, and the frozen weights produce a three-expert Blend. The gate selects only among Direct, Physics-residual, and Blend; Gradient is not a stand-alone final route and contributes only through Blend. Wells with mean PP of at least 54 MPa use Blend. Among the remaining wells,

intervals of at least 1,600 m use the Physics-residual route; all others use Direct. The selected route yields one two-component SHMAX/SHMIN prediction at each depth. Test-well labels are withheld from all preprocessing, fitting, routing, and prediction procedures and are accessed only for final evaluation.

2.5. Implementation details and statistical analysis

Experiments were conducted on 64-bit Windows 11 Pro. The workstation had an Intel Core i7-14650HX processor (16 physical cores and 24 logical processors), 32 GB RAM, and an NVIDIA GeForce RTX 5070 Ti Laptop GPU with 12 GB of system-reported video memory; the driver version was 610.88. PG-WFME uses scikit-learn ExtraTrees and runs on the CPU. LOWO folds used 180 trees per expert, and final models used 500 trees. All experts used the squared-error criterion, `max_depth = None`, `max_features = 0.80`, `bootstrap = False`, and `n_jobs = -1`. The Direct and Gradient experts used `min_samples_leaf = 5`, whereas the Physics-residual expert used `min_samples_leaf = 50`. Each fold and each final expert used a fixed random state. Feature construction required 1.2921 s, and fitting all three final experts required 36.7804 s. Generating the three candidate predictions for 75,711 samples from eight test wells required 0.6602 ± 0.0250 s over five runs; the rule-based gate added negligible cost. The comparison 1D U-Net was retrained under the same well-disjoint development/test split because the exact legacy run could not be recovered. It received DEPTH, RHOB, GR, NPHI, DT, and PP and returned paired SHMAX and SHMIN sequences. The encoder contained 32, 64, and 128 channels followed by a 256-channel bottleneck. Each block used two kernel-3 Conv1d-BatchNorm-ReLU layers; max pooling downsampled the sequence, transposed convolutions upsampled it, skip features were concatenated with the decoder, and a kernel-1 convolution produced two outputs. Training used 256-point windows with stride 16, inference stride 128, batch size 32, Adam with an initial learning rate of 0.0005, mean squared error on standardized targets, and seed 42. The candidate network was fitted on eight development wells, and two held-out development wells were used for epoch selection; the minimum validation loss occurred at epoch 3. The network was then reinitialized and fitted for three epochs on all 10 development wells before blind evaluation. Clipping and standardization statistics were fitted only on wells permitted at each stage. Our PG-WFME itself contained no convolutional layer; Conv1d appeared only in the 1D U-Net comparator.

We reported MAE, root mean square error (RMSE), and the coefficient of determination (R^2). Metrics were calculated separately for SHMAX and SHMIN, and the reported overall value was their arithmetic mean. Sample-

weighted metrics pooled all permitted depth samples, whereas mean per-well MAE first evaluated each well and then gave every well the same weight. The blend-weight search used sample-weighted average MAE. Development mean per-well MAE was used only for the gate-threshold sensitivity analysis. For model comparisons, we bootstrapped per-well average-MAE differences with 20,000 resamples and applied an exact paired sign-permutation test across the eight wells. Blind-well labels were used only for final evaluation, after all predictions and route assignments had been finalized. The evaluation metrics are defined as follows:

$$\begin{aligned} \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\ \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \\ R^2 &= 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \end{aligned} \quad (7)$$

where n is the number of evaluated depth samples, y_i is the measured *in situ* stress value, the predicted term is the final model output defined in **Equation 5**, and $\bar{y}$ is the mean measured *in situ* stress. MAE, RMSE, and R^2 were calculated separately for SHMAX and SHMIN and then averaged when an overall metric was reported.

3. Results

3.1. Study area and dataset

The retained dataset contains 18 wells from the southern Bohai Bay Basin, China. Ten development wells (BYP1, L19, Y172, Y176, Y177, Y180, Y184, Y282, Y285, and Y286) are used for expert fitting and LOWO selection. Eight wells (Y141, Y142, Y179, Y187, Y189, Y284, Y289, and YD302) form the blind test set. Each well included DEPTH, RHOB, GR, NPHI, DT, and PP. SHMAX and SHMIN labels from the development wells were used for supervised model fitting and selection, whereas those from the blind test wells were withheld until final evaluation. The split was well-disjoint; no SHMAX or SHMIN label from a blind test well contributed to preprocessing-parameter fitting, blend-weight selection, expert fitting, route assignment, or prediction. However, the unlabeled log data from each blind test well were used to construct the corresponding predictors and to calculate the label-free routing variables, including mean PP and logged-interval length. The dataset split and well assignments are summarized in [Table 1](#).

3.2. Well-family routing and ablation analysis

After freezing the experiment configuration, we calculated the mean PP and complete logged interval for each blind

well. Table 2 lists these gate variables together with mean GR and mean DT as supporting descriptors; mean GR and mean DT do not enter the gate. Y187 reaches the 54 MPa PP threshold and is assigned to the LOWO-weighted Blend. Y189, Y284, and Y289 enter the Physics-residual route because their complete intervals are at least 1,600 m. Y141, Y142, Y179, and YD302 use the Direct route.

The three routes are therefore active in the reported blind-well evaluation, i.e., four Direct wells, one Blend well, and three Physics-residual wells. Y187 has the highest mean PP (55.7145 MPa), whereas Y284 is assigned to the Physics-residual route through its 1,742.5000 m complete interval. Mean GR and mean DT document inter-well lithological

and compaction differences but do not influence routing. The route is assigned from unlabeled well-level descriptors before test labels are read for evaluation.

Table 3 reports a cumulative ablation under the same well-disjoint test set. Physics-guided features reduce the sample-weighted average MAE by 0.4661 MPa relative to raw-log ExtraTrees. The global three-expert Blend provides a further 0.1212 MPa reduction, and the fixed well-family gate lowers the error by another 0.2975 MPa. The corresponding average R^2 increases from 0.9218 to 0.9819. Because each row retains all preceding components, the table measures incremental system-level gain rather than isolated causal effects.

Table 1. Dataset split and wells used in the final evaluation

Subset	Wells	Role
Development data	BYP1, L19, Y172, Y176, Y177, Y180, Y184, Y282, Y285, Y286	Expert fitting and LOWO model selection
Blind test data	Y141, Y142, Y179, Y187, Y189, Y284, Y289, YD302	Final evaluation only

Abbreviation: LOWO: Leave-one-well-out.

Table 2. Gate variables and selected route for each blind test well

Well	Complete interval (m)	Mean PP (MPa)	Mean GR (API)	Mean DT (μ s/ft)	Selected route
Y141	409.7500	33.3922	60.2180	84.3950	Direct expert
Y142	1022.6250	31.7109	75.1790	89.4410	Direct expert
Y179	811.0000	52.5105	69.6920	83.5960	Direct expert
Y187	628.3750	55.7145	63.8840	86.7090	LOWO-weighted blend
Y189	1900.6250	50.5831	71.4740	85.8990	Physics-residual expert
Y284	1742.5000	48.1087	82.9260	84.2490	Physics-residual expert
Y289	2093.1250	37.7428	80.2200	92.6080	Physics-residual expert
YD302	861.5000	40.3712	66.8940	83.5430	Direct expert

Abbreviations: DT: Compressional slowness; GR: Gamma ray; LOWO: Leave-one-well-out; PP: Pore pressure.

Table 3. Component-wise ablation on the eight blind test wells

Configuration	SHMAX MAE	SHMAX RMSE	SHMIN MAE	SHMIN RMSE	Average MAE	Average R^2	Mean per- well MAE
A Raw logs + ExtraTrees	2.0245	2.8096	1.6628	2.4078	1.8437	0.9218	1.7795
B + physics-guided Direct expert	1.4605	1.8517	1.2947	1.6698	1.3776	0.9643	1.2645
C + global LOWO-weighted Blend	1.3989	1.6178	1.1140	1.2992	1.2564	0.9755	1.2298
D + fixed 54/1,600 m gate (full)	1.0284	1.3368	0.8894	1.1691	0.9589	0.9819	0.9601

Abbreviations: ExtraTrees: Extremely randomized trees; LOWO: Leave-one-well-out; MAE: Mean absolute error; RMSE: Root mean square error; SHMIN: Minimum horizontal *in situ* stress; SHMAX: Maximum horizontal *in situ* stress.

Table 4 presents one-dimensional slices through the 5×5 development-OOF sensitivity grid. The fixed 54 MPa/1,600 m operating point yielded a mean per-well MAE of 4.2252 MPa. The lowest grid score, 4.1896 MPa, occurred at 56 MPa/1,800 m; the difference was 0.0357 MPa, or about 0.9% of the minimum. The local variation was modest, so the table is interpreted as a robustness check. The 54 MPa/1,600 m pair remains the frozen rule for the reported rerun and is not presented as the numerical optimum of this grid.

Table 5 evaluates the empirical baseline coefficients using development-well OOF predictions. Development mean per-well MAE ranges from 3.9097 to 4.0379 MPa across the five settings. The nominal SV weights of 0.70 for SHMAX and 0.55 for SHMIN are retained as prescribed scale weights, not as calibrated mechanical constants or as the optimum of the sensitivity scan. The relatively narrow range indicates that the residual branch is not controlled by a single finely tuned coefficient pair.

3.3. Blind-well prediction performance

Figures 3 and 4 compare the measured and predicted SHMAX and SHMIN profiles for the eight blind test wells. The sample-weighted MAEs are 1.0284 MPa for SHMAX and 0.8894 MPa for SHMIN, with R^2 values of 0.9810 and 0.9829, respectively. Y141 has the lowest average MAE (0.3885 MPa), followed by Y284 (0.4585 MPa) and Y179 (0.5322 MPa). Y142 and YD302 have the largest average errors, at 1.8038 and 1.8775 MPa. Their deviations occur mainly as interval-scale level shifts, while the principal depth trends are retained.

The panels also show route-specific behavior. The Direct route preserves local stress levels well in Y141 and Y179. The LOWO-weighted Blend follows the main trend and short-scale changes of high-pressure well Y187, and the Physics-residual route reproduces the long-interval profiles of Y189, Y284, and Y289. Larger separations remain in parts of Y142 and YD302. Because routes and model settings are frozen before evaluation, the panels

Table 4. Gate-threshold sensitivity based on development-well OOF predictions

PP threshold (MPa)	Interval threshold (m)	Development LOWO mean-well MAE	Status
50	1,600	4.3056	Sensitivity point
52	1,600	4.2252	Sensitivity point
54	1,200	4.2490	Sensitivity point
54	1,400	4.2252	Sensitivity point
54	1,600	4.2252	Fixed operating point
54	1,800	4.1997	Sensitivity point
54	2,000	4.4827	Sensitivity point
56	1,600	4.2151	Sensitivity point
58	1,600	4.5493	Sensitivity point

Abbreviations: LOWO: Leave-one-well-out; MAE: Mean absolute error; OOF: Out-of-fold; PP: Pore pressure.

Table 5. Sensitivity of the empirical SV–PP baseline coefficients

SHMAX SV/PP weights	SHMIN SV/PP weights	Development LOWO mean-well MAE
0.60/0.40	0.45/0.55	4.0379
0.65/0.35	0.50/0.50	3.9568
0.70/0.30	0.55/0.45	4.0039
0.75/0.25	0.60/0.40	4.0133
0.80/0.20	0.65/0.35	3.9097

Abbreviations: LOWO: Leave-one-well-out; MAE: Mean absolute error; PP: Pore pressure; SHMAX: Maximum horizontal *in situ* stress; SHMIN: Minimum horizontal *in situ* stress; SV: Vertical stress.

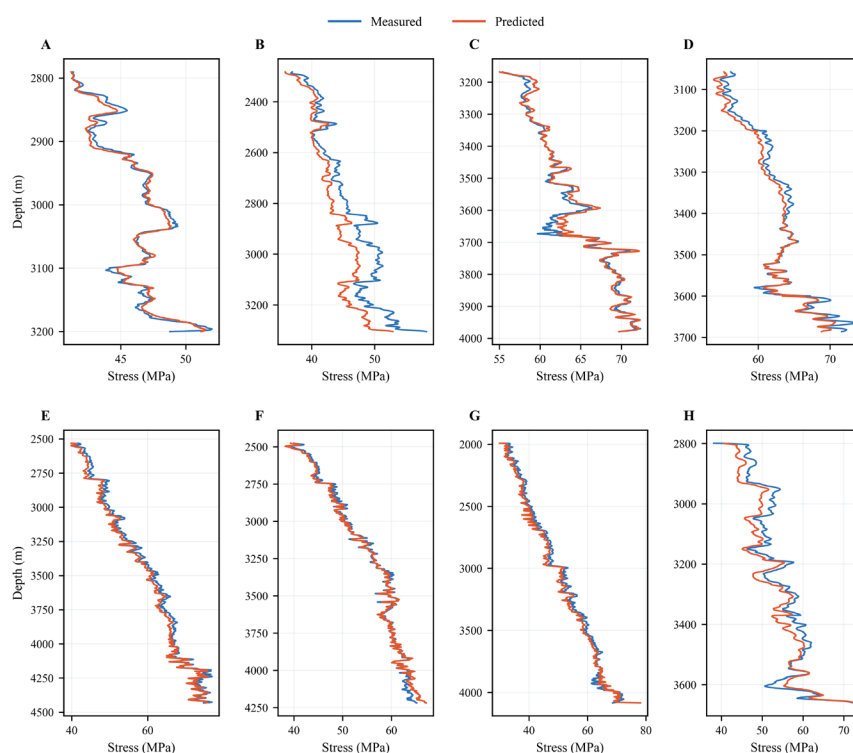


Figure 3. Predicted and measured SHMAX curves for the eight blind test wells. (A) Y141, (B) Y142, (C) Y179, (D) Y187, (E) Y189, (F) Y284, (G) Y289, and (H) YD302.

Abbreviation: SHMAX: Maximum horizontal *in situ* stress.

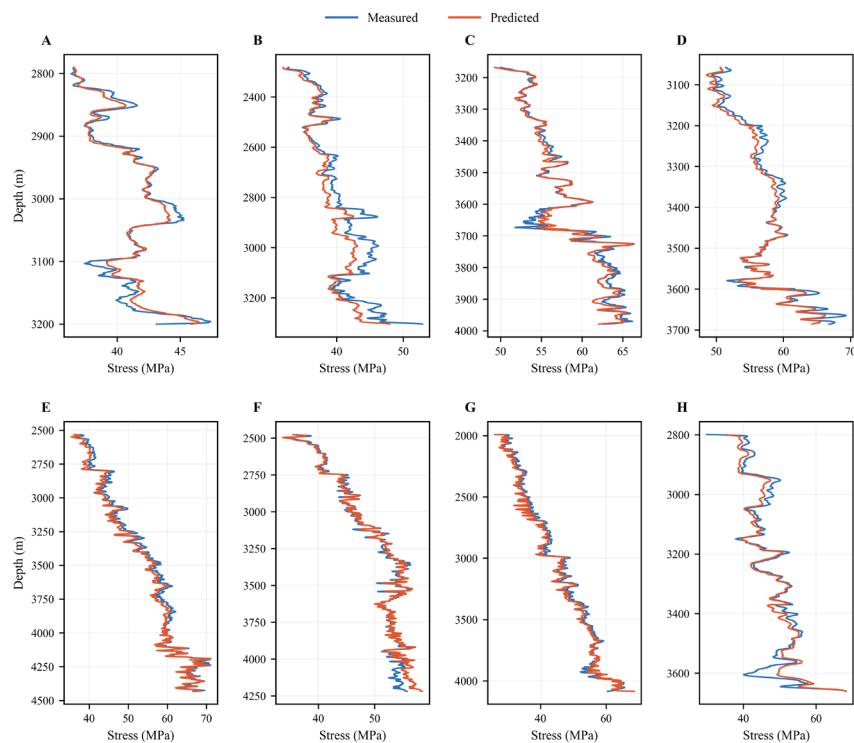


Figure 4. Predicted and measured SHMIN curves for the eight blind test wells. (A) Y141, (B) Y142, (C) Y179, (D) Y187, (E) Y189, (F) Y284, (G) Y289, and (H) YD302.

Abbreviation: SHMIN: Minimum horizontal *in situ* stress.

report the resulting curves without target-well anchoring or retrospective horizontal shifting.

Figure 5 enlarges the 2,820–3,180 m interval of Y141, the well with the lowest average MAE. Both stress components reproduce the main turning points and plateau changes without local calibration. The cross-plots in Figure 6 concentrate near the one-to-one line, consistent with sample-weighted MAEs of 1.0284 and 0.8894 MPa. Figure 7 shows mean errors of -0.7826 MPa for SHMAX and -0.5415 MPa for SHMIN, indicating a modest overall

underprediction. Together, these diagnostics show that the model captures the curve geometry while leaving a measurable low-side bias in several intervals.

The retrained one-dimensional U-Net (1D U-Net) gives sample-weighted SHMAX and SHMIN MAEs of 2.7528 and 2.2939 MPa and R^2 values of 0.8337 and 0.8619, yielding an average MAE of 2.5234 MPa and an average R^2 of 0.8478. The re-evaluated raw-log ExtraTrees baseline gives an average MAE of 1.8437 MPa and an average R^2 of 0.9218. These pooled depth-sample metrics are distinct

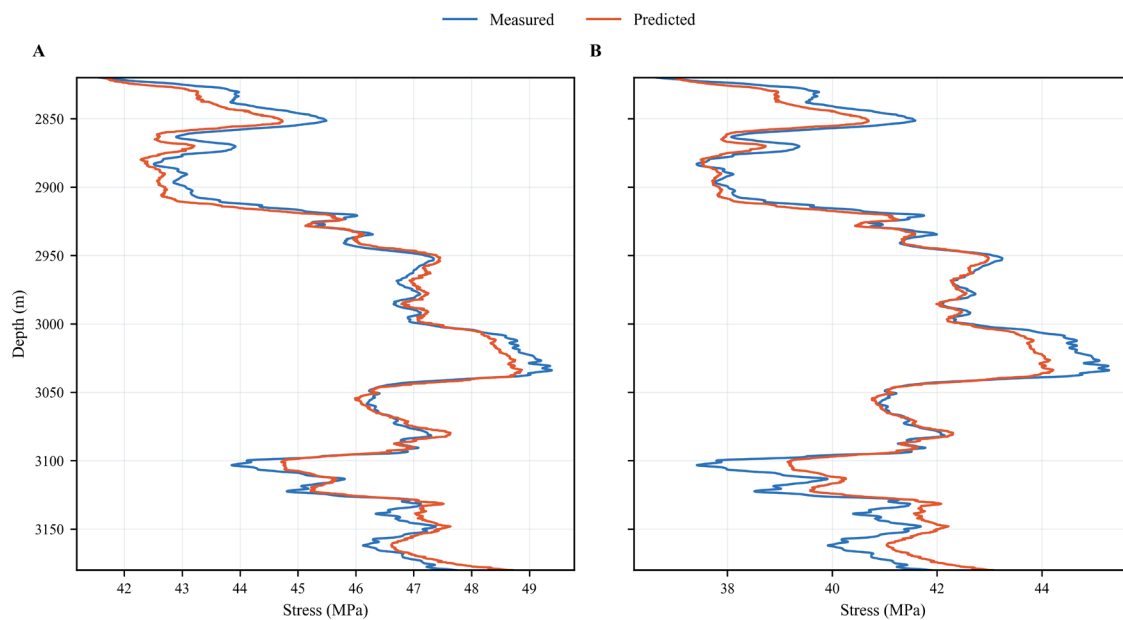


Figure 5. Local prediction details of well Y141 over 2,820–3,180 m for (A) SHMAX and (B) SHMIN. Abbreviations: SHMAX: Maximum horizontal *in situ* stress; SHMIN: Minimum horizontal *in situ* stress.

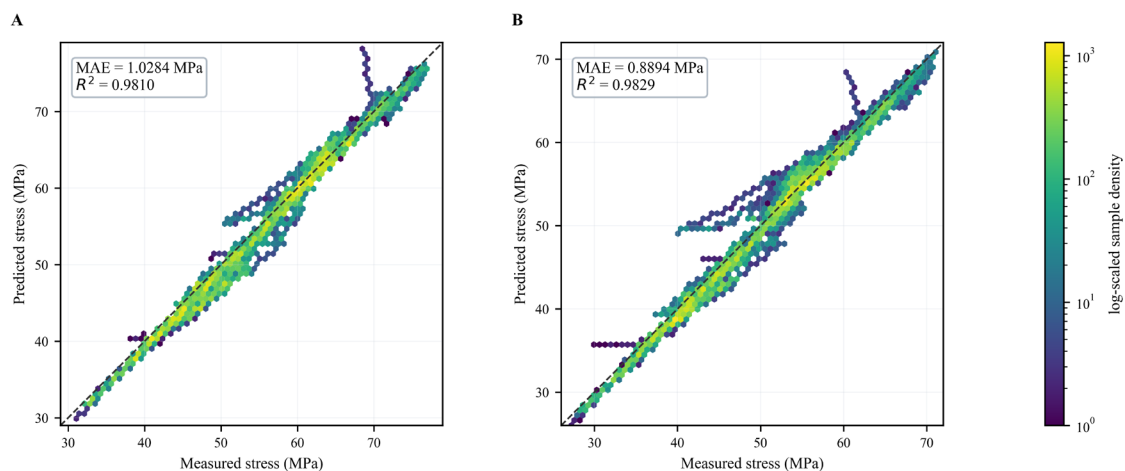


Figure 6. Cross-plots for the final fixed-gate model, (A) SHMAX and (B) SHMIN. Color represents log-scaled sample density. Abbreviations: MAE: Mean absolute error; SHMAX: Maximum horizontal *in situ* stress; SHMIN: Minimum horizontal *in situ* stress.

from equal-well statistics. Table 6 reports well-by-well U-Net/ExtraTrees/PG-WFME results. PG-WFME has a lower per-well average MAE than both comparators in seven of the eight wells. Relative to U-Net, the mean per-well reduction is 1.5445 MPa, with a 95% bootstrap confidence interval of 0.7658–2.3307 MPa and an exact paired sign-permutation p -value of 0.0156. Relative to raw-log ExtraTrees, the reduction is 0.8195 MPa (95% interval 0.2036–1.4881 MPa; $p = 0.0391$). These intervals and p -values summarize variation across eight wells and do not establish basin-wide significance.

4. Discussion

The ablation results show that cross-well magnitude offset cannot be removed by increasing pointwise regression capacity alone. Physics-guided descriptors provide the largest first-stage gain by exposing broad loading and pressure scales. The three target spaces then separate absolute stress, depth-normalized behavior, and residual correction. Finally, the well-family gate chooses among Direct, Blend, and Physics-residual candidates using only unlabeled well descriptors. This last step reduces average

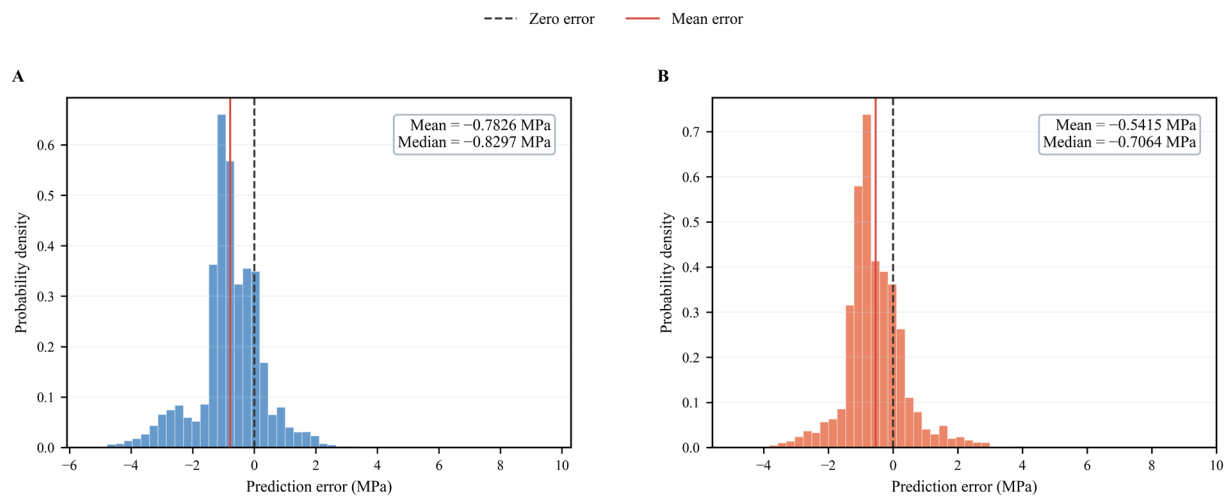


Figure 7. Prediction-error distributions for the final fixed-gate model, (A) SHMAX and (B) SHMIN. Error is defined as predicted minus measured stress. Abbreviations: SHMAX: Maximum horizontal *in situ* stress; SHMIN: Minimum horizontal *in situ* stress.

Table 6. Prediction accuracy for each blind test well

Well	SHMAX MAE (1D U-Net/ExtraTrees/ PG-WFME)	SHMIN MAE (1D U-Net/ExtraTrees/ PG-WFME)	SHMAX R^2 (1D U-Net/ExtraTrees/ PG-WFME)	SHMIN R^2 (1D U-Net/ExtraTrees/ PG-WFME)
Y284	4.1820/3.0432/0.4228	3.9411/2.8844/0.4941	0.2698/0.6395/0.9925	0.0353/0.5006/0.9792
Y141	2.6624/1.2795/0.3225	2.2323/1.0333/0.4544	0.7395/0.5745/0.9689	0.1993/0.7433/0.9395
Y179	3.2443/2.4565/0.4900	2.6289/2.2373/0.5745	0.0448/0.6062/0.9792	0.3283/0.6637/0.9724
Y187	2.5605/1.7226/0.7359	1.8481/1.4054/0.7353	0.4537/0.7414/0.9510	0.7139/0.8332/0.9525
Y189	2.4423/2.0960/0.9403	1.9522/1.6316/0.9168	0.8639/0.8997/0.9900	0.8986/0.9331/0.9877
YD302	3.2790/2.1884/2.1547	2.8387/1.6924/1.6003	0.1992/0.7168/0.7941	0.4128/0.7129/0.8371
Y142	1.3940/1.5347/2.1723	1.0668/0.9069/1.4353	0.8148/0.8465/0.6643	0.8680/0.9140/0.7641
Y289	2.1768/1.3517/1.0244	1.6234/1.0087/0.8876	0.9396/0.9731/0.9898	0.9609/0.9818/0.9902

Abbreviations: ExtraTrees: Extremely randomized trees; MAE: Mean absolute error; PG-WFME: Physics-guided well-family-gated ensemble; SHMAX: Maximum horizontal *in situ* stress; SHMIN: Minimum horizontal *in situ* stress.

MAE from 1.2564 MPa for a global Blend to 0.9589 MPa. Y179 illustrates the value of preserving an absolute-stress route, whereas Y187 is the single high-pressure well assigned to the LOWO-weighted Blend. The improvement is therefore associated with route specialization rather than with a universal preference for one expert.

In situ stress is a key geomechanical quantity controlling seismic velocity anisotropy, fracture development, fault reactivation, and the interpretation of seismic attributes in unconventional reservoirs. In practical seismic exploration workflows, sparse stress measurements at wells are commonly used to calibrate prestack inversion, azimuthal anisotropy analysis, fracture prediction, and seismic-geomechanical reservoir models. The proposed PG-WFME model provides SHMAX and SHMIN results at blind wells, which can serve as well-tie constraints or training and validation labels for seismic-driven stress prediction. Therefore, the contribution of this study is not to replace seismic data analysis, but to provide a physics-guided well-log-based stress prediction module that strengthens the borehole constraints required for seismic exploration and reservoir characterization. When seismic elastic attributes, prestack inversion results, or azimuthal anisotropy information are available, the predicted stress profiles can be integrated with seismic data to extend stress characterization from borehole locations to interwell and reservoir-scale seismic volumes.

External validity is the main limitation in this study. All 18 wells come from one basin, and no independent basin with the same six logs and paired SHMAX/SHMIN labels was available. The blind test therefore measures transfer within the retained inventory rather than basin-independent generalization. The fixed 54 MPa/1,600 m gate is intentionally low capacity, but its thresholds remain dataset-specific. The development sensitivity grid shows local stability, not universal optimality. The residual baseline also uses empirical scale weights rather than calibrated constitutive parameters. More independent wells, broader pressure regimes, and cross-basin evaluation are required before the routing rule can be interpreted as generally transferable.

The measured runtime is compatible with tabular field-scale inference. After fitting, the three candidate experts predict 75,711 depth samples in 0.6602 ± 0.0250 s on the reported CPU, and the rule-based gate adds negligible cost. The retrained 1D U-Net serves as a reproducible convolutional comparator, but PG-WFME itself is a machine-learning ensemble of ExtraTrees regressors and contains no convolutional layer. This distinction matters in the present limited-well setting, where explicit target decomposition and route control provide a transparent alternative to increasing network depth.

5. Conclusion

This study presents a PG-WFME framework for continuous prediction of SHMAX and SHMIN from conventional well logs. The method combines geomechanically interpretable feature construction, three complementary ExtraTrees experts, and a label-free well-family gate. We evaluated PG-WFME for continuous SHMAX and SHMIN prediction from six conventional logs. Ten development wells provide whole-well OOF predictions for selecting a 0.08/0.64/0.28 Direct/Gradient/Physics-residual Blend. In the final analysis, the prespecified 54 MPa/1,600 m gate was held fixed, the temporary LOWO models were discarded, and three final experts were refitted on all development wells before evaluation on the eight held-out test wells. All three routes were used during blind-test inference: four wells used Direct, one used the LOWO-weighted Blend, and three used Physics-residual prediction. The final sample-weighted MAEs were 1.0284 MPa for SHMAX and 0.8894 MPa for SHMIN, with an average MAE of 0.9589 MPa and an average R^2 of 0.9819. The cumulative ablation and profile diagnostics support the combined use of physics-guided representation, complementary target spaces, and low-capacity routing. Evaluation on additional independent wells and external basins remains necessary to establish generalizability.

Acknowledgments

None.

Funding

This research was supported by the Key Laboratory of Oil & Gas Reservoir Geophysics of Sinopec under the project “A New Technology for Predicting Geomechanical Parameters between Wells in Eastern Continental Shale Oil Reservoirs Driven by Seismic Waves.”

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Naihao Liu, Yuwei Liu

Investigation: Yongyi Huang, Xinyu Wang

Methodology: Naihao Liu, Yongyi Huang

Writing—original draft: Naihao Liu, Yongyi Huang, Xinyu Wang

Writing—review & editing: Yuwei Liu, Xiwu Liu

Availability of data

The data that support the findings of this study cannot be shared publicly because the authors do not have permission to distribute the dataset.

References

- Hubbert MK, Willis DG. Mechanics of hydraulic fracturing. *Trans AIME*. 1957;210:153-168.
doi: 10.2118/686-G
- Eaton BA. Fracture gradient prediction and its application in oilfield operations. *J Pet Technol*. 1969;21(10):1353-1360.
doi: 10.2118/2163-PA
- Amann F, Gischig V, Evans K, *et al*. The seismo-hydromechanical behavior during deep geothermal reservoir stimulations: open questions tackled in a decameter-scale in situ stimulation experiment. *Solid Earth*. 2018;9(1):115-137.
doi: 10.5194/se-9-115-2018
- Yin Z, Chang S, Zhang S, *et al*. High-resolution characterization of in-situ stress fields in coal reservoirs from pre-stack inversion of mechanical parameters. *J Seism Explor*. 2026;35(4):026140061.
doi: 10.36922/jse026140061
- Sharma RK, Chopra S. The importance of fracture toughness in the estimation of seismic anisotropy and stress orientation in shale formations. *J Seism Explor*. 2021;30(5):405-418.
doi: 10.36922/JSE81
- Haimson BC, Cornet FH. ISRM suggested methods for rock stress estimation—Part 3: hydraulic fracturing (HF) and/or hydraulic testing of pre-existing fractures (HTPF). *Int J Rock Mech Min Sci*. 2003;40(7-8):1011-1020.
doi: 10.1016/j.ijrmms.2003.08.002
- Moos D, Zoback MD. Utilization of observations of well bore failure to constrain the orientation and magnitude of crustal stresses: application to continental, Deep Sea Drilling Project, and Ocean Drilling Program boreholes. *J Geophys Res Solid Earth*. 1990;95(B6):9305-9325.
doi: 10.1029/JB095iB06p09305
- Xiao Y, Liang C, Zhu D, *et al*. Multi-scale geomechanical modelling of unconventional shale gas: the implication on assisting geophysics-geology-engineering integration. *Int J Energy Res*. 2024;2024:4145930.
doi: 10.1155/er/4145930
- Cheng J, Chen Y, Sun Z. Seismic prediction methods for evaluating in situ stress in tilted transversely isotropic and monoclinic media. *J Seism Explor*. 2025;34(1):60-80.
doi: 10.36922/JSE025190002
- Shiri Y, Moradzadeh A, Ghavami-Riabi R, *et al*. Integration of 2D seismic and well log data for petrophysical modeling and gas reserve estimation in appraisal state of petroleum exploration. *J Seism Explor*. 2012;21:231-246.
doi: 10.36922/JSE342
- Zhang J. Pore pressure prediction from well logs: methods, modifications, and new approaches. *Earth Sci Rev*. 2011;108(1-2):50-63.
doi: 10.1016/j.earscirev.2011.06.001
- Jin ZL, Garipov T, Volkov O, *et al*. Reduced-order modeling of coupled flow and quasistatic geomechanics. *SPE J*. 2020;25(1):326-346.
doi: 10.2118/193863-PA
- Xu K, Zhang G, Chen Y, *et al*. Geophysical prediction and evaluation of lacustrine shale oil and gas: Applications in the Fuxing area. *J Seism Explor*. 2026;35(3):026070030.
doi: 10.36922/jse026070030
- Obradors-Prats J, Rouainia M, Aplin AC, *et al*. Stress and pore pressure histories in complex tectonic settings predicted with coupled geomechanical-fluid flow models. *Mar Pet Geol*. 2016;76:464-477.
doi: 10.1016/j.marpetgeo.2016.03.031
- Alam J, Chatterjee R, Dasgupta S. Estimation of pore pressure, tectonic strain and stress magnitudes in the Upper Assam basin: a tectonically active part of India. *Geophys J Int*. 2019;216(1):659-675.
doi: 10.1093/gji/ggy440
- Ibrahim AF, Gowida A, Ali A, *et al*. Machine learning application to predict in-situ stresses from logging data. *Sci Rep*. 2021;11(1):23445.
doi: 10.1038/s41598-021-02959-9
- Delavar MR, Ramezanzadeh A. Pore pressure prediction by empirical and machine learning methods using conventional and drilling logs in carbonate rocks. *Rock Mech Rock Eng*. 2023;56(1):535-564.
doi: 10.1007/s00603-022-03089-y
- Mustafa A, Kelley M, Lu G, *et al*. An integrated machine learning workflow to estimate in situ stresses based on downhole sonic logs and laboratory triaxial ultrasonic velocity data. *J Geophys Res Mach Learn Comput*. 2024;1(4):e2024JH000318.
doi: 10.1029/2024JH000318
- Al-Fakih A, Kaka SI, Koeshidayatullah AI. Reservoir property prediction in the North Sea using machine learning. *IEEE Access*. 2023;11:140148-140160.
doi: 10.1109/ACCESS.2023.3336623
- Al-Fakih A, Ibrahim AF, Elkatatny S, Abdulraheem A. Estimating electrical resistivity from logging data for oil wells using machine learning. *J Pet Explor Prod Technol*. 2023;13(6):1453-1461.
doi: 10.1007/s13202-023-01617-2
- Al-Fakih A, Koeshidayatullah A, Mukerji T, Al-Azani S, Kaka SI. Well log data generation and imputation using sequence based generative adversarial networks. *Sci Rep*.

- 2025;15(1):11000.
doi: 10.1038/s41598-025-95709-0
22. Al-Fakih A, Li K. Estimation of bottom-hole temperature based on machine/deep learning. In: Proceedings of the International Petroleum and Petrochemical Technology Conference 2021. Springer; 2022:340-353.
doi: 10.1007/978-981-16-9427-1_33
23. Koeshidayatullah A, Al-Fakih A, Kaka SI. Toward basin-agnostic well log imputation and anomaly detection via a pre-trained time-series foundation model. *Energy Geosci.* 2026;7(2):100536.
doi: 10.1016/j.engeos.2026.100536
24. Feng Y, Harrison JP, Bozorgzadeh N. Uncertainty in in situ stress estimations: a statistical simulation to study the effect of numbers of stress measurements. *Rock Mech Rock Eng.* 2019;52:5071-5084.
doi: 10.1007/s00603-019-01891-9
25. Feng Y, Gao K, Mignan A, Li J. Improving local mean stress estimation using Bayesian hierarchical modelling. *Int J Rock Mech Min Sci.* 2021;148:104924.
doi: 10.1016/j.ijrmms.2021.104924
26. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5-32.
doi: 10.1023/A:1010933404324
27. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12:2825-2830.