# A geology-anchored, two-stage workflow for basin-scale well log normalization using 3D statistical volumes

*Patricia E. Rodrigues*[1], Reinaldo J. Michelena[1], Keyla Gonzalez[2], and Alejandro A. Valenciano[2]*
*[1]SeisPetro Geoconsulting, [2]TGS*

**Summary**

This work introduces a geology-anchored, two-stage workflow for large-scale log normalization designed to remove non-geologic variability while preserving stratigraphic and facies variations. The method integrates a systematically selected reference dataset with fully automated three-dimensional (3D) percentile-based normalization, enabling consistent basin-wide results at unprecedented scale across tens of thousands of wells. In the Midland Basin, where logs are affected by multiple casing intervals, different service companies, tool vintages, and sharp lateral facies transitions, traditional normalization approaches, focused on narrow target zones or neighboring wells, often fail to maintain geological and statistical consistency. Our workflow addresses these challenges through two stages: (1) reference normalization of high-quality wells using P10-P90 percentiles, and (2) generation of calibrated 3D statistical volumes that act as normalization engines for all remaining wells. Applied to more than 6,600 wells, the process achieved 98% successful normalization and was subsequently extended to over 20,000 additional wells using the same 3D volumes and rules, requiring only hours of additional computation once the volumes were built. The results show improved stratigraphic continuity, removal of casing-related artifacts, and clearer regional trends. Beyond Gamma Ray logs, the approach provides a scalable framework for normalizing other log types and preparing reliable inputs for large-scale property and facies modeling, seismic-log foundation models, and machine-learning workflows.

**Introduction**

Log normalization is a critical step in basin-scale reservoir characterization and in modern data integration workflows that combine well, seismic, and production data. Its objective is to remove non-geologic variability from well logs while preserving true stratigraphic and facies variations. Without consistent normalization, subsequent petrophysical modeling, seismic calibration, or machine learning efforts are compromised by inconsistent inputs.

In mature basins like the Midland Basin, decades of exploration have produced logs acquired by different service companies, with variable tools, calibrations, and depth corrections. Changes in borehole size and multiple casing strings frequently alter log responses with depth. These effects, compounded by incomplete coverage, create a patchwork of data that masks genuine geological trends (Figure 1). The Gamma Ray (GR) log, widely used as a proxy for lithologic variability, is particularly sensitive. Normalization is therefore not merely data preparation but an interpretive process that underpins regional correlations and multi-modal machine learning models.

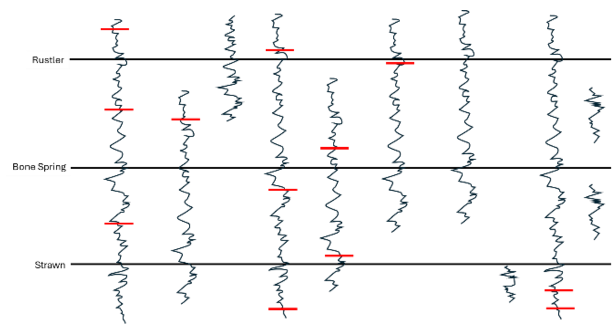


Figure 1. Examples of data coverage heterogeneity and casing locations (red segments) in GR logs across the Midland Basin.

Over past decades, several authors established the conceptual foundation for well log normalization. Early workflow guidelines emphasized calibration within limited stratigraphic intervals (Shier, 2004). Field-scale applications demonstrated its importance for petrophysical consistency but remained constrained by manual implementation and modest well populations (Aguirre and Antelo, 2001). Millard et al. (2019) summarized approaches aimed at preserving 2D regional geological trends. Commercial utilities such as Petra provide systematic normalization tools (IHS Markit, 2020), yet these remain largely interval-based and dependent on local reference wells. Collectively, these methods are limited in scalability when applied to thousands of wells with variable depth coverage, casing, and tool vintage.

To overcome these challenges, we developed a semi-automated, two-stage workflow combining geological consistency with statistical rigor. The first stage systematically selects and normalizes high-quality wells using two-dimensional percentile statistics along stratigraphic surfaces. These reference wells and interpreted tops define the framework for the second stage: a fully automated 3D volumetric normalization based on percentile volumes computed from the reference dataset. This transforms normalization from a well-by-well correction into a stratigraphy-guided, basin-scale process.

The workflow was applied to over 6,600 wells across approximately 3,900 square miles of the Midland Basin, producing a geologically consistent GR-normalized dataset suitable for integration with well, core, and seismic data.

## Data and challenges

The dataset includes Gamma Ray logs of variable depth coverage, resolution, and quality, reflecting decades of acquisition under diverse operational and environmental conditions. Figure 1 illustrates the two principal challenges needed to be addressed before a regional normalization could be achieved:

*Data coverage heterogeneity* Not all wells sample the same stratigraphic intervals. Some wells extend only through shallow formations, while others reach deep Paleozoic targets. The inconsistency in log depth coverage introduces lateral data gaps and makes it difficult to establish basin-wide trends.

*Multiple casing points and borehole effects*: Many wells contain several casing points (red segments in Figure 1) that affect the GR response. The change in borehole diameter and tool environment alters the measured signal, often producing step changes or abrupt amplitude variations unrelated to lithology.

Together, these factors introduce strong vertical and lateral heterogeneity that can obscure geologic trends and complicate the definition of a regional normalization model. In addition, sharp facies transitions and the presence of a regional carbonate ramp structure further increase variability. Any successful normalization method must therefore distinguish geological changes from acquisition artifacts while maintaining stratigraphic fidelity.

## Normalization workflow

The normalization process follows a two-stage design, progressing from an initial selection and 2D normalization of reference wells phase to a fully automated volumetric normalization. The goal is to balance quality control and scalability.

*Stage 1: Reference 2D statistical normalization.* The first step was to identify a subset of wells suitable to serve as a reference population for normalization. From an initial 6,826 wells (Figure 2a), approximately 2,757 were selected based on data completeness (vertical coverage), location (to minimize spatial clustering), and minimal amount of casing points (Figure 2b). Wells with more than three casing points or very small vertical coverage were excluded at this stage.

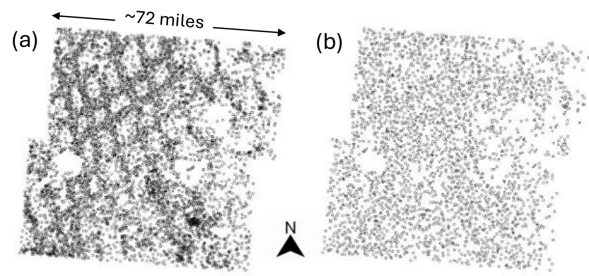


Figure 2. (a) Spatial distribution of all wells (6826) and (b) selected subset for reference normalization (2757).

Percentile-based statistics (P10, P90) of GR values were computed along key stratigraphic surfaces and used to adjust the logs within each interval. In this study, two stratigraphic surfaces were used to model the statistical trends across the area: Rustler and the deeper Bone Spring. Figure 3 shows an example of the P90 trend across the Bone Spring surface. The process to generate 2D trends of P10 and P90 automatically grids the surface to the desired resolution, removes outliers, and smooths the results. These maps are then used to produce a set of normalized reference wells that capture the geologic and statistical variability of the basin while removing noise. This set of reference wells is used as input for the next stage.

*Stage 2: 3D volume normalization.* Once the reference wells were normalized, they were used to generate three-dimensional percentile volumes (P10, P90) within a stratigraphic grid constructed based on key formation markers (Rustler, Bone Spring, Lower Spraberry Shale, Wolfcamp, Strawn, and Devonian Carbonate). These markers were modeled across all wells from interpreted tops in a subset of control wells. Figure 4a shows the grid used to generate the 3D statistical trends. After the grid is built, normalized logs are converted to blocked, upscaled logs at the resolution of the grid. Each block contains the extracted percentile (P10 or P90) from the original log data within the corresponding layer (Figure 4b).

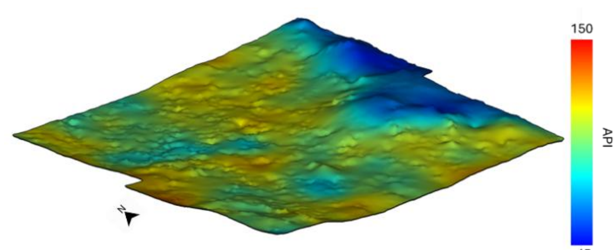


Figure 3. P90 trend along Bone Spring surface

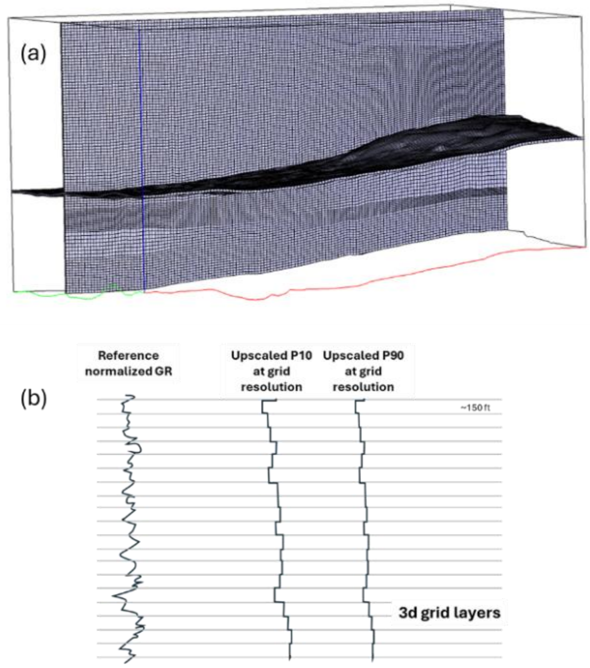


Figure 4. (a) Stratigraphic grid and (b) calculation of upscaled statistics at grid resolution using reference wells. Grid cell dimentions: 2500 ft x 2500 ft x ~150ft.

The next step was to interpolate the blocked P10 and P90 logs from the reference set across the entire 3D grid. The result of the P90 interpolation is shown in Figure 5. During 3D normalization, each interval between casing points or missing log segments required representative low and high statistics for the normalization equation. Analysis showed that the average P10 and P90 values extracted from the grid did not match the true corresponding percentiles within those intervals, as is expected when statistical measures are computed over already-aggregated statistics. To account for this well-known effect, the P25 and P75 of the grid distributions were used instead, ensuring accurate scaling and preservation of the reference data's statistical behavior.

The resulting P10/P90 volumes serve as normalization engines for the remaining wells. Each well is processed automatically: its measured GR values are scaled relative to the corresponding P10/P90 values from the 3D volumes at each stratigraphic level. This process effectively maps every GR measurement into a geologically consistent, statistically balanced framework. The method also handles wells with partial or discontinuous data, as the volumetric reference allows interpolation through missing log intervals.

In regions with structural ramps or abrupt facies changes, local smoothing parameters were adapted to prevent over-flattening of the data. The combination of percentile-based scaling and adaptive smoothing ensured both continuity and preservation of fine-scale variability.

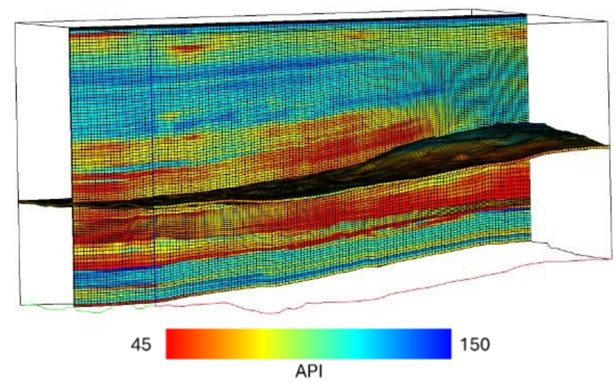


Figure 5. Interpolated upscaled P90 along stratigraphy.

**Results**

The workflow was applied to over 6,600 wells covering approximately 3,900 square miles. Of these, 98 percent were successfully normalized; the remaining wells were excluded due to incomplete or corrupted data. The entire process was completed with minimal manual intervention once the reference set was established.

Figure 6a  shows the interpolated original GR volume along the stratigraphic grid before normalization. The data shows significant lateral and vertical inconsistencies, bright and muted patches caused by casing effects, tool variations, and incomplete log coverage. Figure 6b demonstrates the improved continuity of stratigraphic markers and the enhanced visibility of regional trends, after applying the two-stage normalization and a 3×3 median filter along stratigraphy.

The normalized dataset aligns more closely with known regional markers such as the Bone Spring and Strawn intervals and exhibits a smooth transition across the carbonate ramp (transition from high to low GR indicated by the arrow). Along the ramp and elsewhere, the method preserves geological contrast while removing spurious variability, providing a clean, geologically meaningful GR dataset suitable for integration into regional modeling and data analytics.

**Discussion**

The combination of geologically guided statistical interpolations and statistical automation proved key to achieving both accuracy and scalability. Traditional normalization methods typically focus on limited target zones or rely on local correlations between nearby wells, which makes them prone to error propagation and inconsistent results across regions of variable data quality. By contrast, the use of volumetric percentile statistics ensures that every well is tied to the same regional statistical framework, maintaining reproducibility even with varying input data density and vertical coverage.

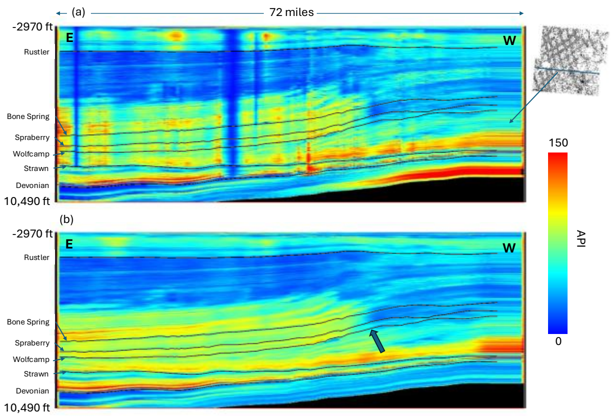


Figure 6. (a) Original GR log interpolated along stratigraphic grid. (b) Interpolated GR log after 3D normalization and 3x3 median filter along stratigraphy. Location of the cross-section relative to well locations is shown in the upper right. Cell dimensions: 2500 ft x 2500 ft x 4 ft.

Equally important is the conceptual shift from per-well or map-based normalization to volume-based normalization. The 3D percentile volumes encapsulate the basin's geologic and statistical behavior and can be reused for other datasets or extended to additional wells. When used on a larger dataset of ~20,000 additional wells (for a total of ~26,600 wells), the same 3D volumes and normalization rules produced consistent and geologically reasonable results without additional manual adjustment. This demonstrates the scalability of the method.

Because the first stage of reference normalization is supervised, the approach avoids a major pitfall of fully automated normalizations: the blind propagation of tool or noise-related problems. Once the high-quality reference framework is established, the second stage operates fully automatically, making it feasible to normalize tens of thousands of wells at basin scale, requiring only few additional hours of computation.

The 3D volumetric framework also opens new possibilities for integration with other data types. Since the normalization volumes are defined within a stratigraphic grid, they are directly compatible with seismic-derived attributes and can be used as conditioning inputs in foundation models that combine seismic and log data (Lasscock et al., 2025). In addition, these same 3D volumes can be sampled along horizontal well trajectories, enabling consistent extraction of stratigraphic and statistical information that can be directly integrated with production data analysis. In this sense, the normalized logs become a critical component of a multi-modal learning environment, ensuring that the model learns from geological variations rather than acquisition noise and biases.

## Conclusions

This study presents a new, geology-anchored approach to large-scale log normalization that merges systematically selected well control with automated volumetric processing. The main innovations include:

- The generation of 3D statistical volumes (P10, P90 → P25, P75) from a systematically normalized reference dataset, providing a consistent basis for basin-wide normalization.
- A semi-automated workflow in which systematic well selection ensures geological reliability, while volumetric normalization enables scalability to tens of thousands of wells in a short amount of time.
- Successful application to over 6,600 wells in the Midland Basin, with subsequent generalization to 20,000 additional wells (for a total of ~26,600 wells).
- Preservation of stratigraphic and facies variability despite strong heterogeneity in data quality, casing, and depth coverage.

The method shows that log normalization can evolve from a local correction task into a fully scalable, geologically meaningful operation. By encoding both statistical and stratigraphic relationships within 3D, stratigraphy guided volumes, this workflow establishes a reliable foundation for basin-scale characterization, reservoir modeling, and the emerging generation of integrated foundation models that couple seismic and well data.

## Acknowledgements

Thanks to TGS for permission to publish this work. Thanks also to Ben Lasscock for providing the well tops for different intervals from his machine-learning approach to automatic tops interpretation.