Well-to-Well Domain Adaptation: Carrying a Fracture and Bedding Model Across Wells with Kriging and Stratigraphic Correlation

A borehole-image detector trained on one set of wells is, by default, a single-well predictor. It reads the wall of the borehole it was shown and nothing beyond it, and when a new well arrives from an adjacent part of the same field its accuracy can fall off a cliff that has nothing to do with the model and everything to do with distribution shift. This whitepaper is a field account of closing that gap. Working with a Middle East NOC / carbonate operator we partnered with, across a multi-phase formation-evaluation engagement, we took a Detection-Transformer-derived fracture and bedding detector and made it travel — not by re-labelling every new well, but by treating well-to-well transfer as a domain-adaptation problem solved with geostatistics. The detector emits two engineered logs, a fracture-density curve and a bedding-density curve, computed at ten-centimetre resolution; 2D and 3D kriging with gaussian, linear and exponential variograms interpolate those logs between wells spaced roughly forty to eighty metres apart; and stratigraphic well-tops align the interpolation within a common formation rather than across geological apples and oranges. The result is a field-scale correlation engine: bedding density that correlates laterally between neighbouring wells, a productivity lift of roughly sixty percent and an interpretation-accuracy lift of roughly seventy-five percent against a ninety-five-percent target precision and ninety-percent stratigraphic-correlation success. We argue that the unit of subsurface AI is not the well. It is the field — and the engineering that gets you there is feature-space alignment, not a bigger model.

Quamer Nasim
ML Research Engineer

Tannistha Maiti
Senior AI Researcher

Tarry Singh
Founder & CEO

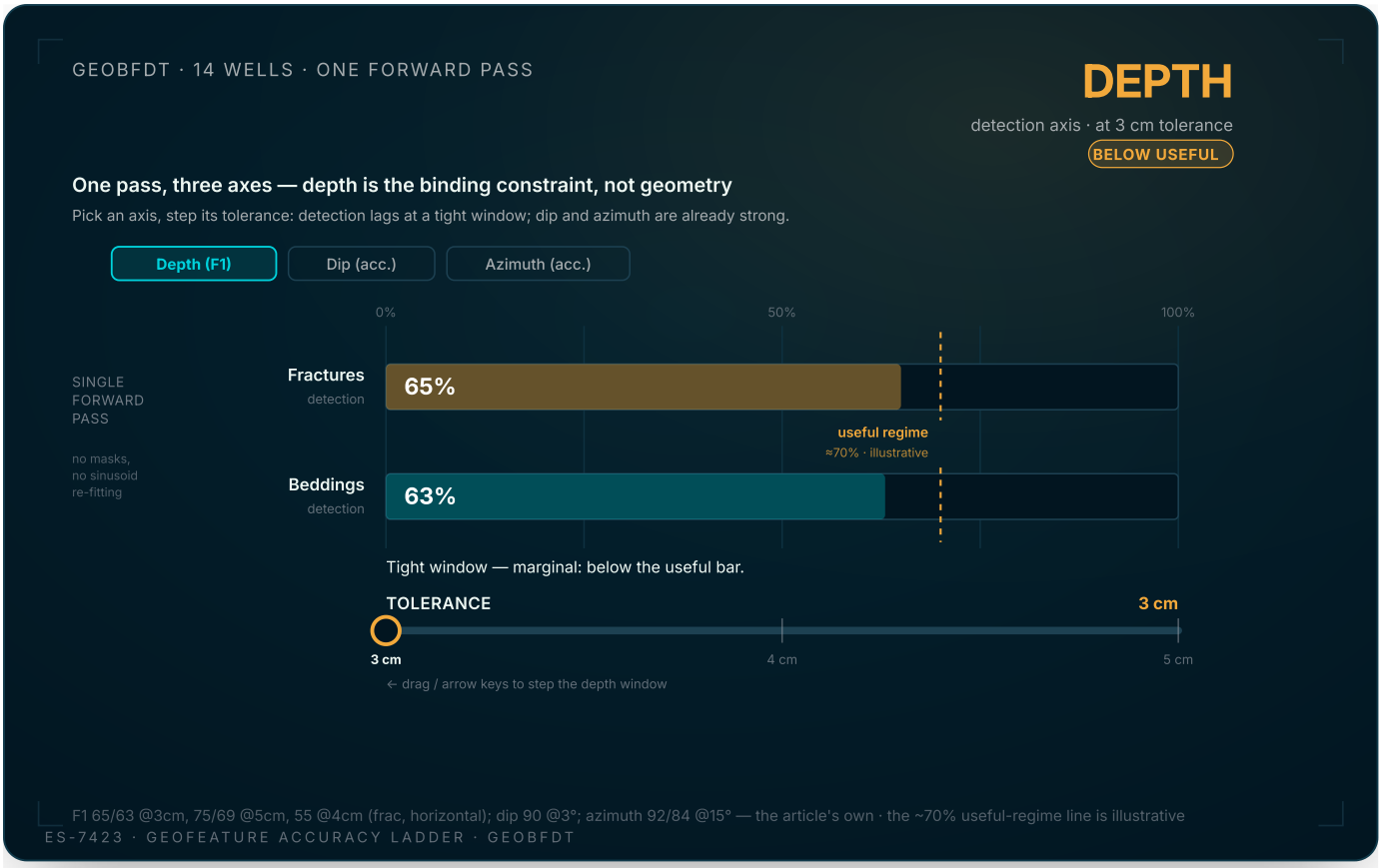
DOMAIN-ADAPTATION / WELL-TO-WELL-CORRELATION / KRIGING / GEOSTATISTICS / SUBSURFACE-AI / COMPUTER-VISION

CONTENTS

A model that reads a borehole wall has seen exactly one borehole. That is the quiet limitation behind almost every successful image-log AI pilot: it works beautifully on the wells it was trained and validated on, and then a new well arrives from the next fault block over and the numbers sag. The geoscientist who watches this happen is right to be suspicious — but usually wrong about the cause. The architecture did not get worse. The data did. The new well is drawn from a slightly different distribution of textures, dips, and tool responses, and a detector tuned to the old distribution is now extrapolating. In machine-learning terms this is out-of-distribution generalisation; in subsurface terms it is the difference between a single-well predictor and a field tool. This whitepaper is about how we crossed that line — turning a per-well fracture and bedding detector into something that correlates a whole field — for a Middle East NOC / carbonate operator we partnered with, and why the engineering that mattered was geostatistical alignment rather than a heavier network. It draws on the same body of subsurface-AI work we have run with operators across the Middle East and the United States, but every number below is pinned to the one carbonate engagement where we built and validated the well-to-well pipeline end to end.

The detector is the easy half

Before a model can travel between wells it has to work on one. The supervised core here is a Detection-Transformer-derived computer-vision model that frames borehole-feature picking as end-to-end set prediction: a ResNet backbone reads a patch of the high-resolution borehole-image (micro-resistivity) log, a transformer encoder-decoder attends across it, and parallel heads classify each detected sinusoid as a fracture or a bedding plane and regress its depth, dip, and azimuth. There are no anchor boxes and no non-maximum suppression; the model emits a fixed set of object queries and a bipartite matching loss assigns them to ground-truth picks. This is the same architectural family we have described elsewhere, and on its own it is a genuinely strong picker — at a three-degree tolerance the model places dip correctly for the large majority of true positives, and that geometric fidelity is the raw material everything downstream depends on.



GeoBFDT emits the whole (class, depth, dip, azimuth) tuple in one forward pass — but the three axes are not equally hard. Detection along the depth axis is the binding constraint: at a tight 3 cm window fracture F1 is only ~65% (beddings ~63%) and only clears the useful regime for structural work once tolerance loosens to 5 cm (~75% / ~69%); horizontal wells hold ~55% at 4 cm. The geometric axes the interpreter actually fits sinusoids for are already strong at tight tolerance — dip ~90% at 3°, azimuth ~92% (fractures) / ~84% (beddings) at 15°. Pick an axis and step its tolerance: depth is the lever you loosen, dip and azimuth are already past the line. All accuracies and tolerances are the article's own; the dashed ~70% 'useful regime' line is an illustrative reading aid (the article names no exact F1 cutoff).

The accuracy ladder above is the foundation, and it is worth being precise about what it does and does not buy you. It tells you that on a well the model has effectively seen — same field, same tool, distribution it was trained against — the per-feature geometry is trustworthy at engineering tolerances. It tells you nothing about the well next door. A picker this good is still, architecturally, a function of one borehole's pixels. The interpretation it produces is a vertical strip of fractures and beddings at the wellbore and zero information one centimetre into the rock. To make it a field tool you have to do two things the detector cannot do by itself: turn its per-feature picks into a continuous signal that can be compared across wells, and then carry that signal through the rock between the wells. Neither is a modelling problem. Both are engineering.

Step one: engineer the features the detector cannot

The first move is to stop treating the model's output as a list of picks and start treating it as a measurement that can be logged like any other curve. From the detector's fracture and bedding picks we computed two new well-log curves the operator had never had before: a **fracture-density log** and a **bedding-density log**, each a count of detected features per depth interval. We tested several interval sizes — five, ten, and fifty centimetres — and settled on a ten-centimetre grid, smoothed with a rolling-average kernel of five samples for fractures and ten for the slower-varying bedding signal. The counts themselves were tabulated at a tenth-of-a-metre resolution so the curve had real vertical structure rather than coarse bins.

This is the unglamorous, decisive step. A density log is a feature-engineering artefact: it projects a high-dimensional, irregular set of object detections into a single continuous variable that lives in the same coordinate system across every well in the field. You cannot krige a list of fractures. You can krige a density curve. The choice of grid size and kernel is a bias-variance decision made explicitly and recorded — too fine and the curve is noise, too coarse and the lateral structure that makes wells correlatable is smoothed away. Getting it right is what makes the next step possible at all.

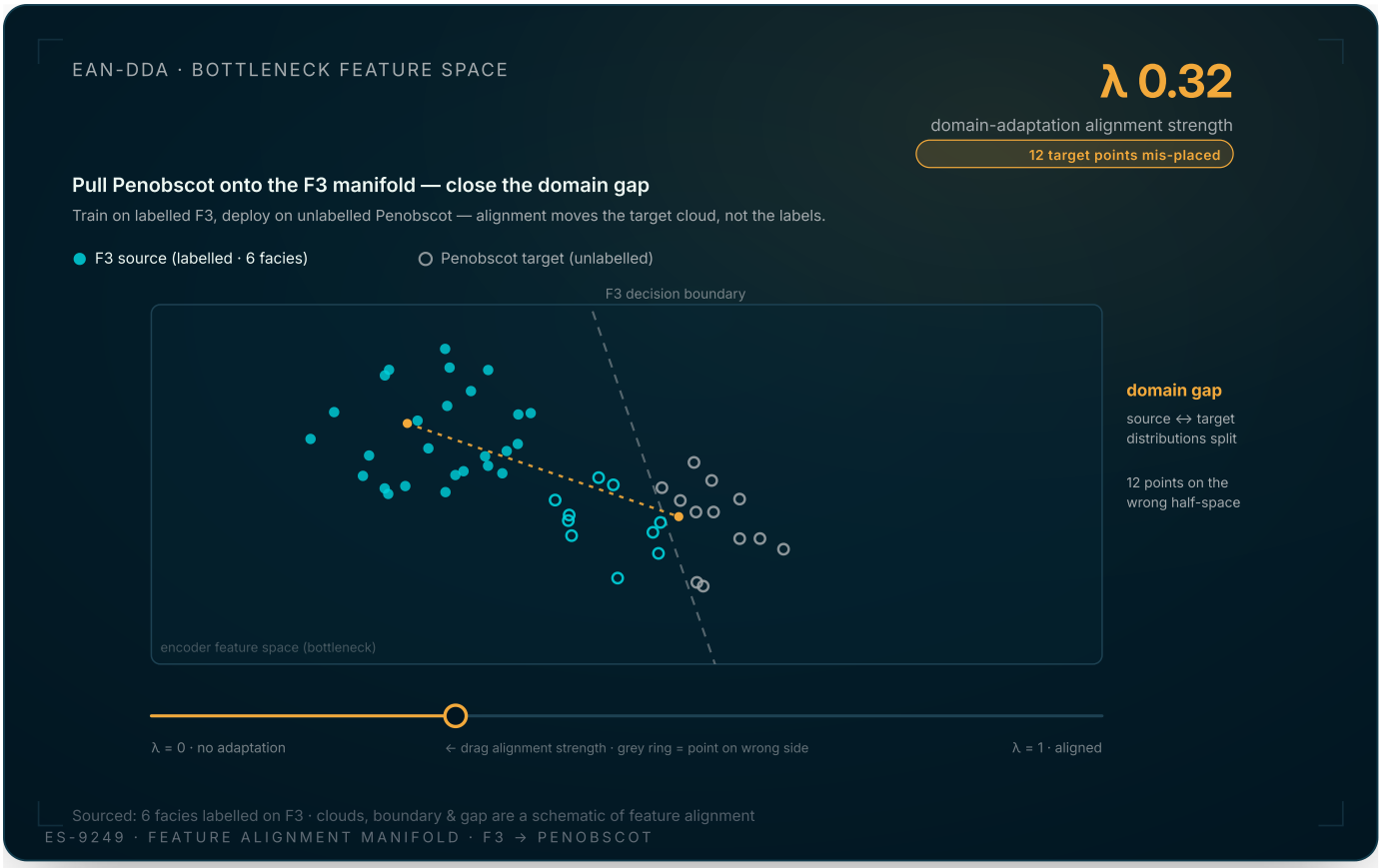
> **DENSITY IS THE TRANSFERABLE REPRESENTATION**
>
> A fracture pick is local to one borehole; a fracture-density log is a field variable. The reason well-to-well transfer works here is not that the detector generalises — it is that we converted its non-transferable output (a set of per-well detections) into a transferable one (a continuous density curve at a fixed ten-centimetre grid). Domain adaptation in this programme is mostly the discipline of choosing a representation that is comparable across the domains you want to bridge.

Step two: align the feature space across wells with kriging

With a density log per well, the well-to-well problem becomes a spatial-interpolation problem: given the same engineered feature measured at a handful of wells roughly forty to eighty metres apart, estimate it everywhere between them. The tool for this is kriging —

best-linear-unbiased spatial prediction — and the engineering substance is in the variogram, the function that encodes how quickly the density signal decorrelates with distance. We fitted and compared gaussian, linear, and exponential variogram models, and ran both 2D kriging (interpolating a density surface across the well locations) and 3D kriging (carrying the interpolation through depth and topography), correlating across three neighbouring wells at a time as the working unit. At the spacings in play the linear kernel was the workhorse for the closest pairs.

Framed in machine-learning language, this is domain adaptation by feature-space alignment. Each well is a domain. The kriged surface is the manifold on which all the wells' density features are made to live, so that a feature measured in one well constrains the estimate in its neighbours rather than standing alone. The interactive below makes the geometry concrete: an unaligned target distribution sits off the source manifold and predictions degrade; alignment pulls the target onto the shared manifold without re-labelling it. That is exactly what kriging does to the density logs — it moves the neighbouring wells' feature estimates onto a common surface so the field, not the well, becomes the unit of inference.



The mechanism behind EAN-DDA, in one move. At the encoder bottleneck, labelled F3 (source) and unlabelled Penobscot (target) patches map into a feature space. Train on F3 alone and the two clouds sit apart — the domain gap — so the F3 decision boundary cuts through the target cloud and facies get mis-read. The deep-domain-adaptation alignment loss pulls the target cloud onto the source manifold: drag the alignment strength λ from 0 to 1 and the orange gap collapses while grey target rings turn teal as they cross to the correct side of the boundary. Only '6 facies classes labelled on F3' is the article's own number; the point clouds, decision boundary, gap, and the live mis-placed-point count are a schematic of feature-distribution alignment — no benchmark metrics are shown.

The mechanics are worth stating precisely, because the variogram choice is where a careless implementation goes wrong. Kriging estimates the density value at an unsampled location as a weighted linear combination of the values at the sampled wells, and it chooses the weights to minimise estimation variance subject to unbiasedness. The weights are not inverse-distance heuristics; they are solved from the spatial-covariance structure encoded in the variogram, which measures how the squared difference between two density readings grows with the lag separating them.

ORDINARY-KRIGING ESTIMATE: A WEIGHTED SUM OF THE DENSITY AT SAMPLED WELLS, WEIGHTS SOLVED FROM THE VARIOGRAM, NOT FROM RAW DISTANCE.

Formula | LaTeX

Copy LaTeX

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i \, Z(x_i), \quad \text{with} \quad \sum_{i=1}^{n} \lambda_i = 1$$

The shape of the variogram model is the engineering decision that matters. A gaussian variogram assumes the density field is very smooth near the origin and is appropriate when bedding fabric varies slowly; a linear model is the safe workhorse at the closest well spacings, where there is too little data to justify a curved fit; an exponential model sits between, decorrelating faster than gaussian. We fitted and compared all three rather than defaulting to one, because the wrong variogram does not merely add noise — it produces confident interpolations with the wrong spatial texture, smoothing real lateral structure away or inventing structure that is not there. Choosing the model is a held-out validation exercise, not an aesthetic one.

The pay-off is geological, not just numerical. When we kriged the bedding-density logs across a cluster of nearby wells within a single carbonate formation, the bedding signal correlated laterally — two of the wells tracked each other closely while a third carried its high-density interval in the middle of the section, a structure a geologist can read straight off the surface and tie to deposition. Fracture density behaved differently: it stayed local and did not correlate cleanly between wells, which is itself a result worth having, because it tells the operator that fracturing in this field is a near-wellbore phenomenon to be mapped well by well, while bedding is a field-scale fabric that can be predicted into un-drilled rock. Knowing which of your features are correlatable and which are not is most of the value of a correlation engine.

Step three: correlate within stratigraphy, not across it

Kriging a density surface across wells only means something if the wells are aligned in geological time as well as in map space. Interpolating bedding density from the top of one well into the middle of another, when those depths belong to different formations, produces a confident and wrong answer. So the third engineering layer is stratigraphic correlation:

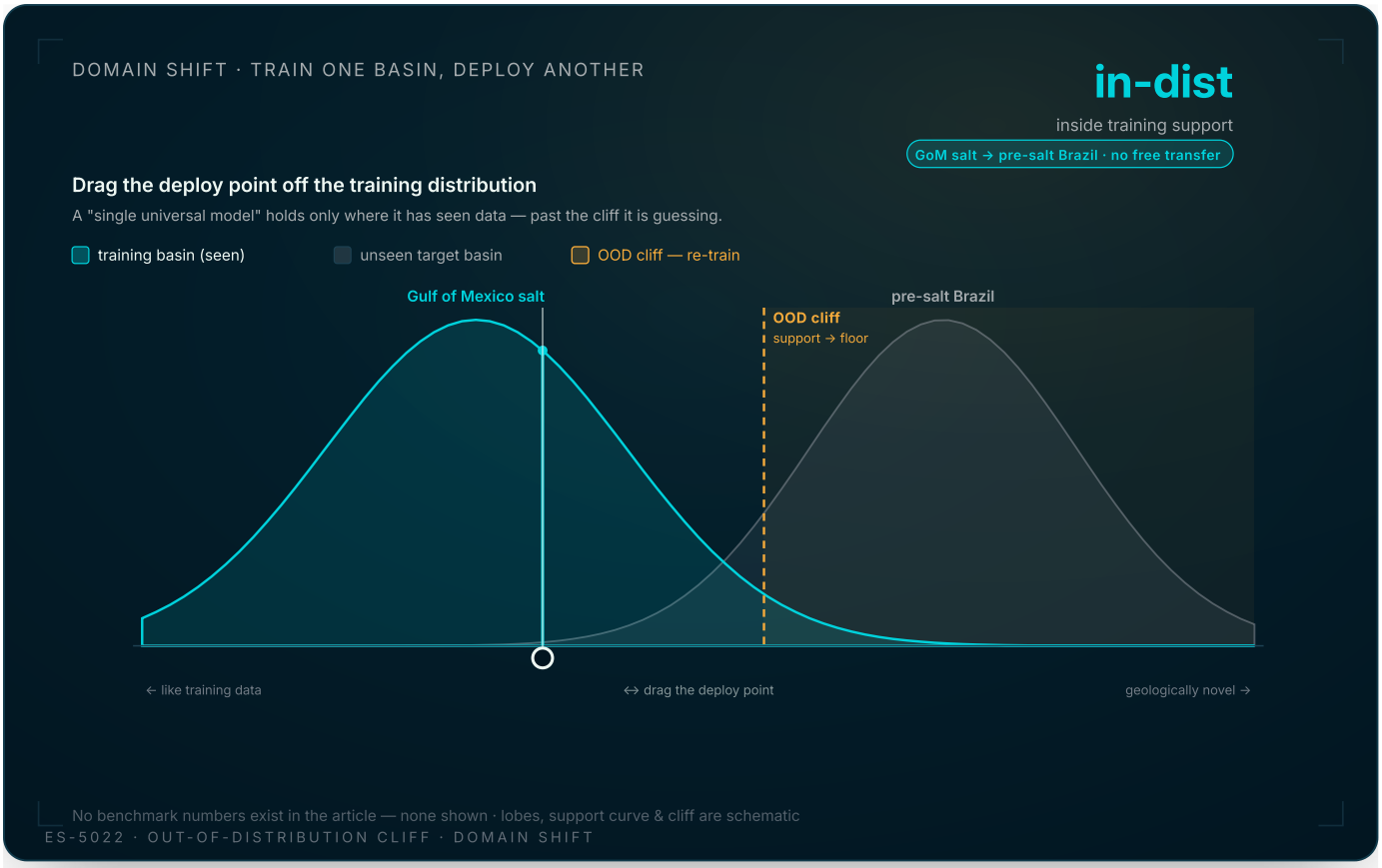
every well carries a column of formation tops, encoded as a small set of stratigraphic identifiers, and the kriging is performed within a common formation rather than across the whole borehole. We did all of the well-to-well analysis inside the carbonate formation where the fracture and bedding picks were most systematic, contoured the density surfaces at a fixed thirty-metre vertical interval across the ten-to-eleven wells in the working set, and used the well-tops to ensure every contour compared like with like.

This is the geomatics-and-software-engineering spine of the system, and it is easy to underrate. A well-tops table, a formation-identifier scheme, a consistent depth-to-SI-units conversion, contour generation at a fixed interval, rose diagrams of dip direction per formation — none of it is deep learning, and all of it is what makes the deep learning add up to a field model instead of a stack of unrelated single-well runs. The correlation engine is a pipeline: detector to density log to variogram to kriged surface to stratigraphically-gated contour, with the stratigraphy as the alignment key that keeps the whole thing geologically honest.

> **THREE ALIGNMENTS, NOT ONE**
>
> Well-to-well transfer needed alignment in three spaces at once. Feature alignment: the detector's picks become a common density log. Spatial alignment: kriging puts neighbouring wells' density on a shared interpolation surface. Stratigraphic alignment: formation tops ensure the surface compares the same geological interval across wells. Drop any one and the correlation is either non-comparable, non-spatial, or non-geological. The engineering is in holding all three.

The cliff this avoids

It is worth being explicit about the failure mode all of this is built to prevent, because it is the default outcome when an operator takes a per-well model to a new well and hopes. Distribution shift in subsurface data is real and abrupt: a new well from an unseen part of the field can sit outside the support the model was trained on, and accuracy does not degrade gracefully — it falls off a cliff. The honest response to that cliff is not to pretend a bigger backbone will climb it. It is to detect when you are on the wrong side of it and to have an adaptation step that recovers performance without a full re-label.



Out-of-distribution generalisation is the single most under-reported failure mode in published subsurface results. A model trained on one basin's distribution (Gulf of Mexico salt, teal) has no support over a geologically distinct one (pre-salt Brazil, grey). Drag the deployment point across the feature axis: while it sits inside the training lobe the model is interpolating; once it crosses the orange OOD cliff — where training support falls to the floor — the model is extrapolating and the 'single universal model' promise fails, so re-training is required. The lobes, support curve and cliff position are schematic illustrations of the distribution-shift mechanism — the article sources no benchmark numbers, so none are shown.

The cliff diagram above is the argument for treating well-to-well as adaptation rather than inference. A model deployed naively past the edge of its training support is operating in the red zone, and no amount of confidence calibration changes that — the support simply is not there. Our adaptation path is geostatistical: rather than retrain blind, we extend the trained detector's reach by kriging its engineered logs into the new well's neighbourhood, anchored on shared stratigraphy, and reserve full re-labelling for genuinely new formations where no amount of interpolation will substitute for new ground truth. That distinction — interpolate within support, re-label across it — is the operating policy that keeps the system both accurate and affordable.

What the adaptation buys

The point of all this engineering is a measurable change in how the operator's geoscientists spend their time and how much of the field they can see. Against the per-well baseline, the well-to-well capability delivered a productivity improvement on the order of **sixty percent** and an interpretation-accuracy improvement on the order of **seventy-five percent**, measured against a **ninety-five-percent target precision** for the picks the system carries between wells and a **ninety-percent stratigraphic-correlation success** rate for tying density structure to the right formation across the field. Those are not marginal numbers. They are the difference between an interpretation tool that runs one well at a time and a correlation engine that turns a cluster of wells into a continuous, predictive picture of bedding fabric and fracturing between them — directly usable for directional and infill-drilling decisions in un-drilled rock.

The companion productivity lift is the AutoFrac and AutoVug pickers themselves, which interpret roughly **five times faster** than manual workflows on the wells they cover. Well-to-well adaptation compounds that: the manual baseline does not just interpret each well by hand, it has no mechanism to predict between wells at all, so the field-scale view the kriging engine produces is not a faster version of an existing task — it is a task the operator could not previously perform.

The training-data realism behind the numbers

A field-scale claim is only as good as the wells behind it, and honesty about that is part of the engineering. The well-to-well density-log analysis was computed from thirteen vertical wells in the production phase — a small number by any machine-learning standard, and a reminder that in proprietary subsurface domains the binding constraint is always wells, never compute. To pressure-test the correlation machinery itself independently of that scarcity, we also exercised the well-to-well pipeline against a large public proxy: the FORCE 2020 machine-learning-competition dataset of one hundred eighteen wells from the Norwegian Sea, dividing each well into patches of seven hundred data points for kink-and-marker detection. The proxy is not the operator's carbonate — the formations and tool strings differ — but it let us validate that the correlation and patch-segmentation logic behaved sensibly at a well count an order of magnitude larger than the proprietary field offered, before trusting it on the thirteen wells that actually mattered.

The discipline this reflects is the same one that governs the rest of the programme: every reported correlation is pinned to the well count, the formation, the grid size, and the variogram that produced it. A kriged surface is a model output like any other, and it is reproducible only if the interpolation parameters travel with it.

Productionising the adaptation: the retrain loop

A correlation engine that only its builders can run is not an asset; it is a dependency. The well-to-well capability was therefore wired into the same production lifecycle as the rest of the programme rather than left as a notebook a data scientist re-executes by hand. Each of the engineered logs, the variogram fits, and the kriged surfaces is an addressable, versioned artefact, so that "the bedding-density correlation across these three wells" is a reproducible object pinned to a dataset version, a grid size, and a variogram model — not a figure someone regenerated from memory. When a new well arrives, the operator's own engineers run the same pipeline: pick with the detector, compute the density logs on the fixed grid, fit the variogram against the new well's neighbourhood, krige within the relevant formation, and contour. The pipeline is the deliverable, and the pipeline is what makes the sixty-and-seventy-five-percent gains durable rather than a one-time demonstration.

This is where the out-of-distribution policy becomes an operational control rather than a slogan. Interpolation handles new wells that fall inside the trained detector's support — same formation, same tool family, comparable texture — and runs in minutes without a geoscientist re-labelling anything. A genuinely new formation, or a tool response the detector has never seen, trips the re-label branch: the new well becomes labelled training data, the detector is retrained with the dataset version incremented, and the kriging is re-run on the refreshed density logs. Routing each new well to the correct branch — interpolate or re-label — is the single most consequential operating decision the field team makes, and it is made on evidence (does this well sit inside the support?) rather than on hope. That evidence-based routing is the difference between a system that degrades silently and one that flags its own blind spots.

The retrain loop also closes the data flywheel that a single-well model never has. Every well the operator interprets produces fresh density logs, which tighten the variograms, which sharpen the kriged surfaces, which make the next well's interpolation more reliable. A per-well predictor cannot compound; a field-scale correlation engine compounds with every well drilled, because each new well is simultaneously a prediction target and a new control

point on the shared interpolation surface. The engineering that makes this safe — versioned artefacts, a reproducible pipeline, an explicit support-aware routing policy — is exactly what lets the operator's team own and extend the system after the build team leaves.

What good correlation looks like

For a geophysicist, reservoir engineer, or geomodeller evaluating a well-to-well AI capability — building one, buying one, or trying to trust one — the questions that separate a field tool from a glorified single-well demo are not about the network:

- Does the system emit an engineered, continuous feature that is comparable across wells — a density log on a fixed grid — rather than a non-transferable list of per-well picks?
- Is inter-well interpolation done with a fitted variogram whose model (gaussian, linear, exponential) and well spacing are recorded, not a black-box surface?
- Is every correlation gated by stratigraphic well-tops so the system compares the same formation across wells, never across geological time?
- Is there an explicit policy for the out-of-distribution cliff — interpolate within support, re-label across it — rather than a hope that the model generalises?
- Are the field-scale claims pinned to the well count, formation, grid size, and variogram that produced them, so a colleague can reproduce the kriged surface?

If the answers are yes, the operator owns a correlation engine and the unit of their subsurface AI is the field. If the answers are no, they own a very good single-well predictor and a quietly false sense of how far it travels.

WHAT THIS WHITEPAPER ARGUES

1. A per-well borehole-image detector is a single-well predictor by default; carrying it across wells is a domain-adaptation problem, not a bigger-model problem.
2. The transferable representation is an engineered fracture-density and bedding-density log at a fixed ten-centimetre grid — you cannot krige a list of picks, but you can krige a density curve.
3. Kriging with fitted gaussian/linear/exponential variograms aligns neighbouring wells' density features onto a shared interpolation surface across wells ~40-80 m apart; stratigraphic well-tops keep the correlation within one formation.
4. Bedding density correlates laterally between nearby wells (a field-scale fabric); fracture density stays local — knowing which features are correlatable is most of the value.
5. The capability lifted interpretation productivity ~60% and accuracy ~75% against a 95% target precision and 90% stratigraphic-correlation success, turning a per-well model into a field-scale correlation engine for directional and infill drilling.

References

Krige, 1951 D.G. Krige. A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa (1951). The originating work on the best-linear-unbiased spatial prediction now bearing his name.

Matheron, 1963 G. Matheron. Principles of Geostatistics. Economic Geology, 58 (1963). The formalisation of the variogram and kriging as a theory of regionalised variables. https://doi.org/10.2113/gsecongeo.58.8.1246

Carion et al., 2020 N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko. End-to-End Object Detection with Transformers (DETR). ECCV 2020. The set-prediction detection architecture underlying the fracture and bedding picker. https://arxiv.org/abs/2005.12872

Bormann et al., 2020 P. Bormann, P. Aursand, F. Dilib, P. Dischington, M. Manral. FORCE 2020 Well Log and Lithofacies Dataset for Machine Learning Competition. Zenodo (2020). The 118-well Norwegian Sea dataset used as a public well-to-well correlation proxy. https://doi.org/10.5281/zenodo.4351156

Wang & Deng, 2018 M. Wang, W. Deng. Deep Visual Domain Adaptation: A Survey. Neurocomputing (2018). Framing of feature-space alignment as the mechanism of cross-domain transfer. https://arxiv.org/abs/1802.03601

EARTHSCAN

Well-to-Well Domain Adaptation: Carrying a Fracture and Bedding Model Across Wells with Kriging and Stratigraphic Correlation

Published November 2024. Generated 2026-07-22.
https://www.earthscan.io/whitepapers/well-to-well-domain-adaptation-fracture-bedding-kriging