quantil | analytics

Operations Research and AI for Surface Operations in Oil & Gas

From Development Planning to Drilling, Production, Logistics and Procurement

Álvaro Riascos Villegas
Partner & Co-Director, Quantil · Professor of Economics, Universidad de los Andes

King Fahd University of Petroleum & Minerals · Invited Lecture

Scope of this Lecture

Six deployed systems along the operational chain of a producing field, drawn from a portfolio of nearly ninety projects delivered to a national oil company.

I	**Integrated Development Planning**	Large-scale scenario generation under a multi-objective optimization framework
II	**Drilling Optimization**	Rig-state classification and causal estimation of borehole quality
III	**Production Surveillance**	Deep temporal segmentation of production time series
IV	**Energy Consumption Forecasting**	Ensemble and recurrent models by field and business unit
V	**Logistics Optimization**	Multimodal network design under cost, time or emissions criteria
VI	**Procurement Optimization**	Integer programming for service allocation under bundled discounts

Not covered: financial and commodity risk management.

Integrated Development Planning

Deciding which wells to drill, in what order, under which surface constraints — and under different objectives.

Integrated Development Planning — The Problem

An integrated development plan aligns field objectives with corporate priorities across interdependent sectors, regions and stakeholder groups.

- The plan must be sustainable: it has to account for interdependencies between sectors, regions and stakeholder groups rather than optimizing a single field in isolation.
- The plan must reduce environmental impact: the environmental footprint of field operations is a design variable, not a reporting obligation.
- The plan must optimize production and return: net present value is maximized subject to the surface constraints that actually bind — treatment facilities, transport, energy.
- Before the system, plan construction was a manual, expert-driven exercise. The combinatorial space of drilling sequences, rig counts and facility constraints was never explored; a small number of hand-built scenarios were compared, and the best of those was adopted.

The formal problem

Choose a drilling schedule — which wells, with how many rigs, in which order — that maximizes the net present value of the field and the valuation of its certified reserves, while minimizing energy consumption and carbon footprint, subject to treatment, transport and energy capacity constraints, and under uncertainty in the production profile of every candidate well.

Integrated Development Planning — Decision Architecture

Six stages, from well-level production forecasts to a ranked set of admissible development plans.

1	2	3	4	5	6
Inputs	**Decisions**	**Constraints**	**Monetization**	**Economic model**	**Decision support**
Wells to be drilled and estimated production forecasts, well by well	Drilling rigs: how many, and the timing of each drilling campaign	Treatment facilities, transport capacity, energy availability	Parameterization of OPEX and scheduled CAPEX	Economic evaluation; NPV distribution under uncertainty	Methodical selection and visualization of admissible plans

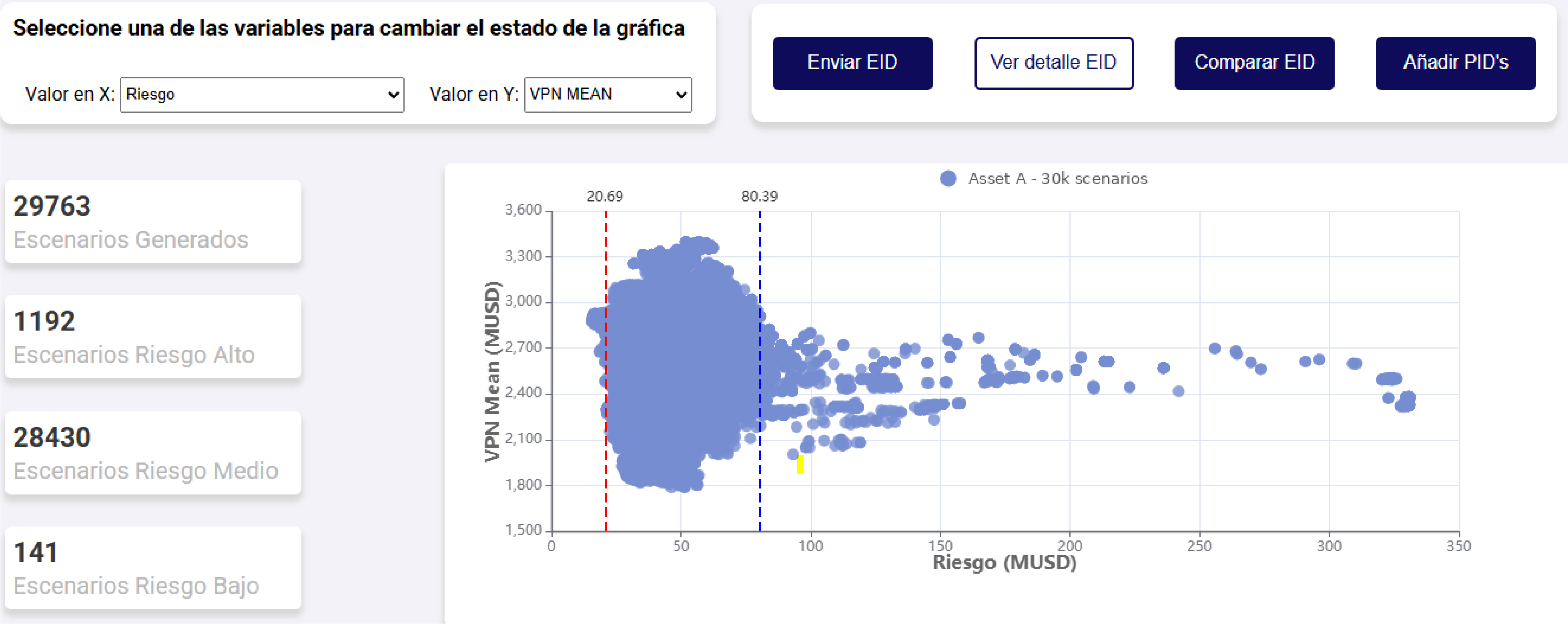
Engineering basis (stages 1–4) · Economic evaluation and decision criteria (stages 5–6)

What the system produces

- Economic evaluation of assets and field contracts, and evaluation of exploratory prospects
- Explicit visualization of the constraint scheme, and early identification of fluid imbalance
- Quantification of environmental impact, and quantification of uncertainty in every valuation

Integrated Development Planning — Scenario Space

Each point is one admissible development plan, evaluated on risk and expected net present value. The optimizer searches this cloud; the engineer selects within it.



Generated scenarios for a single asset, classified by risk band. Axes: risk (MUSD) against mean net present value (MUSD). The dashed lines delimit the risk tolerance bands used to filter admissible plans.

Integrated Development Planning — Results

The system is in production and its benefits are certified by the client's own audit process.

+4,000

Emulated production profiles, generated to replace reservoir simulation inside the optimization loop

+100k

Simulations and economic valuations executed under the distributed scheme

+13

Assets positively impacted by the system across 2023, 2024 and 2025

7,909

Engineering hours avoided in the execution of integrated development plans

14.35

MUSD — certified benefits of the product, 2025

1.82

MUSD — certified benefits of the product, 2023–2024

579

MUSD — value optimized across development plans under the tool, 2023–2024

II

Drilling Optimization

Well construction generates the densest operational data in the industry — and almost none of it is experimental.

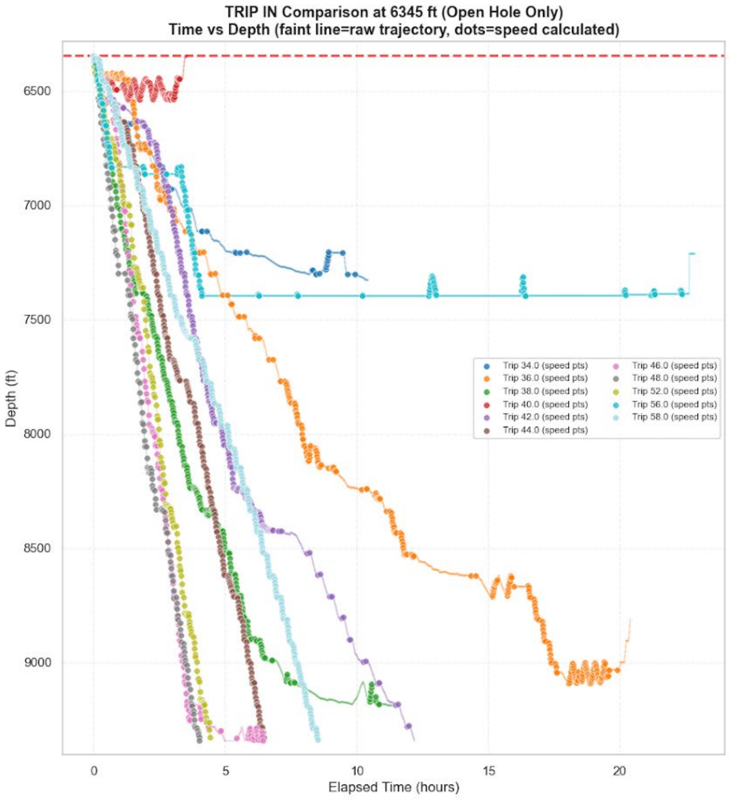
Drilling — From Raw Records to a Quality Index

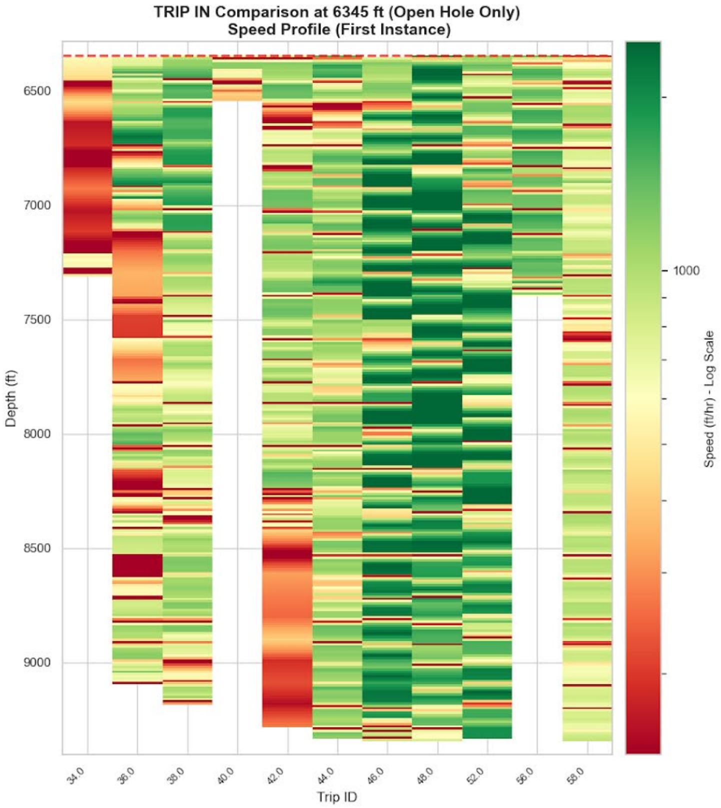
Four stages: integration, operational labelling, construction of an outcome variable, and causal estimation.

1	**Data ingestion and processing**	Integration, cleansing and standardization of operational, drilling and geological data across historical campaigns. This is the stage that determines whether anything downstream is estimable at all.
2	**Operational labeller**	A Rig State classification model learned from historical records, which segments continuous sensor streams into the discrete operations an engineer would recognize — trip in, trip out, rotating, reaming.
3	**Borehole quality index**	A formal definition of the outcome variable. Without an agreed, computable index there is no dependent variable, and the entire exercise reduces to visualization.
4	**Causal model**	Estimation of the effect of operational variables — drilling parameters, fluids, bottom-hole assembly — on borehole quality, enabling optimization decisions based on impact rather than correlation.

Drilling — Operational Evidence

Comparable operations across campaigns, once the rig-state labeller has made them comparable: identical section, identical depth, widely different execution.





What the labeller makes visible

- Left: elapsed time against depth for each trip-in operation at a common depth in open hole, with the calculated speeds superimposed.
- Right: the corresponding speed profile by trip identifier, on a logarithmic scale.
- The operations are physically comparable — same section, same depth, same hole — yet the execution time differs by a factor of several across trips.
- That dispersion is operational rather than geological, which is precisely what makes it a legitimate object of optimization. Without the rig-state classifier, none of these operations could have been isolated from the continuous sensor record in the first place.

Trip-in comparison at a common depth, open hole only.

Drilling — Why the Model Must Be Causal

Drilling parameters are chosen by engineers in response to conditions. The observational association between parameter and outcome is therefore confounded by design.

The identification problem

A driller reduces weight on bit when the formation is difficult. Difficult formations also produce poorer boreholes. The naive regression of quality on weight on bit therefore recovers the selection rule of the driller, not the physical effect of the parameter. Acting on that estimate would invert the correct decision.

Double Machine Learning

The effect of each treatment variable on the borehole quality index is estimated conditioning on drilling covariates, with flexible learners used for the nuisance functions and orthogonalization protecting the treatment-effect estimate from regularization bias. The specification identifies effect heterogeneity rather than a single average effect.

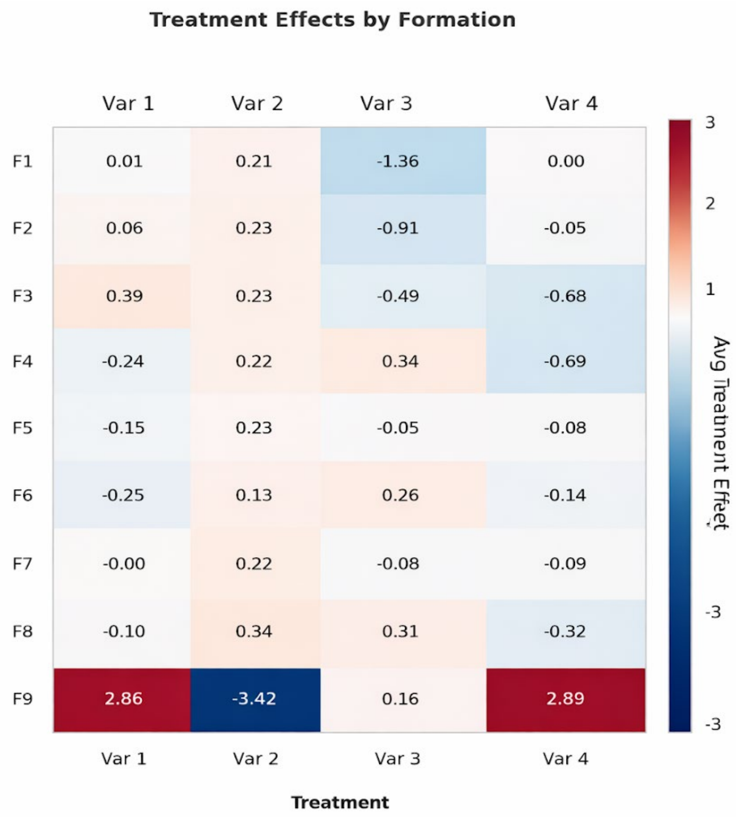
Validation protocol applied

- Hyperparameter optimization of the nuisance learners, performed independently of the treatment-effect stage
- Technical validation of the estimated effects against the operational record
- Interpretive analysis of the results with drilling engineers, since an effect that contradicts physics is a specification error and not a discovery

The practical consequence: the model answers "if we change this parameter, what happens to hole quality?" — which is the question the driller actually asks — rather than "which parameters accompany good holes?"

Drilling — Heterogeneity of the Estimated Effects

The average treatment effect is not the useful object. The conditional effect by formation is.



Average treatment effect on the borehole quality index, by formation and treatment variable. Variable names withheld.

Reading the estimates

- Rows index formations; columns index treatment variables. Each cell is an average treatment effect on the borehole quality index, conditional on the formation.
- For most formations the effects are small and stable, which is itself an operationally valuable result: those parameters are not worth contesting.
- One formation exhibits effects an order of magnitude larger, and of opposite sign across two treatments. That is where operational attention and further instrumentation belong.
- A single pooled estimate would have averaged this structure away entirely.

III

Production Surveillance

Identifying stabilized production periods across thousands of wells — where the difficulty lies in the definition, not in the estimator.

Production Surveillance — The Problem

Decline analysis requires pseudo-stabilized periods. Identifying them was manual, inconsistent between analysts, and did not scale.

The objective

Develop deep learning models that automatically identify pseudo-stabilized periods in oil and gas well production time series, so that stability detection becomes more precise, more consistent between analysts, and scalable to the full well inventory — improving operational efficiency in reserves and decline work.

Constructing the labels

Pseudo-stabilized periods were defined jointly by statistical criteria — **low variability in production and in pressure** — and by expert knowledge. A pipeline was then automated to generate training, testing and validation datasets from those definitions, for thousands of wells.

The methodological difficulty

- The target is not observed. It is a construct, agreed between statistical criteria and engineering judgement, and the model can be no better than the agreement it is trained on.
- The label is an interval, not a point. Ordinary classification metrics are therefore inadequate; the problem is closer to segmentation than to detection.
- Two analysts labelling the same well will not produce identical intervals — which places a ceiling on any achievable score that has nothing to do with the model.

Production Surveillance — Two Formulations

The problem was posed twice.

Business rules

Candidate zones identified directly from the statistical stability criteria. The approach is interpretable and requires no training, but proved highly complex to tune and delivered only moderate metrics. It was retained as a baseline and explicitly not adopted as the main solution.

Temporal segmentation with deep learning

The **series is treated as an image**: sequences are stacked and a deep segmentation model classifies each time step as stabilized or not. Evaluation uses intersection-over-union and F1 rather than accuracy, since the object being recovered is an interval. This formulation delivered the best performance.

Reported performance

0.46

F1, oil series, full well inventory

0.35

Jaccard index, oil series, full well inventory

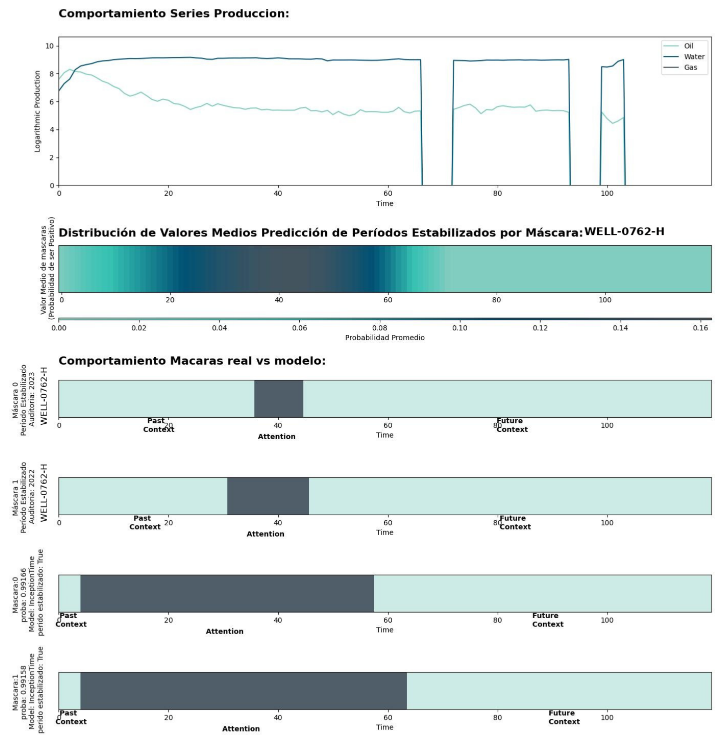
0.56

F1, best-performing individual fields

These are moderate figures and are presented as such. They are nonetheless the most promising option identified, and they are obtained against labels that carry substantial analyst-to-analyst variability. The operational gain is consistency and scale across thousands of wells, not the elimination of expert review.

Production Surveillance — Model against Analyst

A single well: production history, predicted stabilization probability, and the comparison between analyst-assigned and model-assigned intervals.



Predicted stabilization masks against audited intervals, single well.

Panel by panel

- Top — logarithmic oil, water and gas production for the well. The two interruptions are shut-in periods, and any procedure that mistakes them for stability is worthless.
- Centre — the mean predicted probability, by time step, of belonging to a stabilized period. This is the model output before any thresholding decision is taken.
- Bottom — the audited intervals assigned by the analyst for two production years, followed by the intervals recovered by the segmentation model for the same well.
- The model locates the region correctly but systematically proposes wider intervals than the analyst. That is a bias in the width of the recovered interval, which the Jaccard index penalizes and accuracy does not register at all — and it is the whole reason the evaluation metric was chosen before the model.
- Operationally, a wider interval is not necessarily wrong: it is a candidate for expert review rather than a replacement for it. The system was designed to shortlist, not to decide.

IV

Energy Consumption Forecasting & Optimal Dispatch

The operation's own energy demand, forecast by field and by business unit,
across three planning horizons. Forecast is an input for optimal dispatch of (auto) generating units

Energy Consumption Forecasting

Predicting energy consumption across the operator's business units and fields from historical hydrocarbon variables and forward projections of those fluids.

Problem

Forecast energy consumption for each business unit and field, using historical hydrocarbon production variables together with forward projections of the same fluids. The horizon requirement is not single: short, medium and long-term forecasts serve different planning decisions and are evaluated separately.

Pipeline

Each time a new version of the input file is deposited in the data lake, a job is triggered automatically. It performs preprocessing — outlier removal and interpolation — then trains the candidate models, then computes evaluation metrics. The selected model is written back for production use.

Model family

XGBoost, LSTM, an ensemble of the two, and a linear regression retained as a baseline. Selection is by mean absolute error and mean squared error, computed separately for scenarios with and without water — two operating regimes whose consumption behaviour differs materially.

Why this case belongs in a lecture on (dispatch) optimization

- Energy consumption is a constraint and an objective term in the development planning problem of Section I. Forecasting it is not an end in itself; it closes that model.
- The regime split — with and without water — is a modelling decision taken from process knowledge, not from cross-validation. It is the kind of decision that separates a deployed model from a benchmark entry.
- Model selection is automated and retriggered on every data refresh, keeping the forecast valid as the operating condition of the field changes.

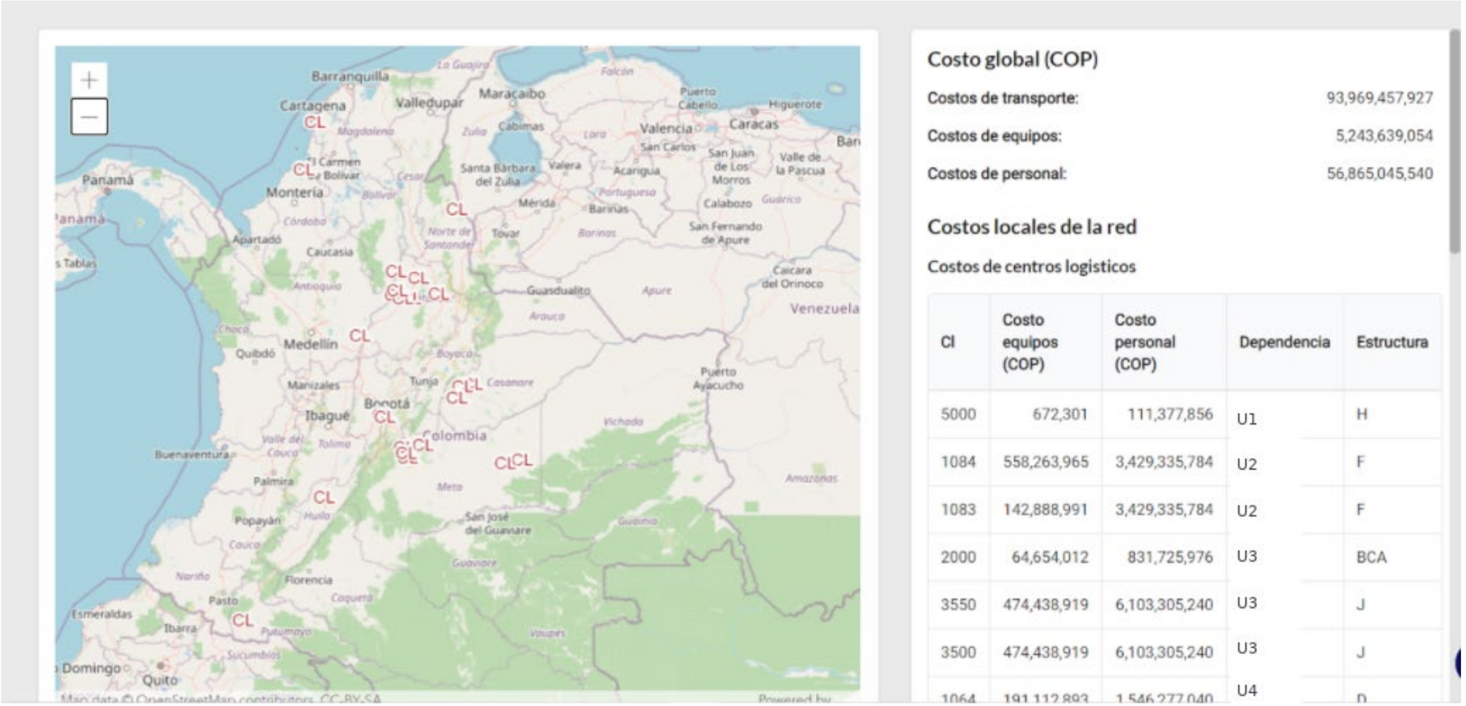
V

Logistics Optimization

Moving equipment and materials to the field across a multimodal network — and choosing what the network should be optimal for.

Logistics — The Hub Location Problem

A multimodal transport network over road, river and rail, in which the location of a distribution hub determines the operator's negotiating position with its suppliers.



- The graph models a multimodal transport network: road, river and rail arcs coexist and are substitutable at different cost and time.
- Locating an optimal hub allows tariffs to be renegotiated with suppliers from 80 municipalities in order to serve 52 logistics centres.
- Alternative routes between any given origin and destination are enumerated and evaluated rather than assumed.

Deployed tool: logistics centres over the national network, with global and local cost decomposition into transport, equipment and personnel components.

Logistics — One Network, Three Objective Functions

The cost of a path can be defined in monetary terms, in time, or in emissions. The network is unchanged; the optimal solution is not.

Monetary cost

Tariffs, equipment and personnel costs across each arc of the network. The conventional objective, and the one against which supplier contracts are written.

Transit time

Elapsed time across the route. Binding whenever equipment availability, rig scheduling or workover sequencing depends on delivery date rather than on delivery cost.

CO_2 emissions

Emissions attributable to each arc by transport mode. River and rail arcs that are uncompetitive on time become competitive here, which changes the optimal hub.

The methodological point

A network optimized for one criterion is not approximately optimal for another. Multimodal substitution means that the ranking of routes — and therefore the location of the hub — reorders as the criterion changes. The value of the model is not that it produces the optimum, but that it makes the trade-off between the three criteria explicit and quantified, so that the organization chooses the criterion deliberately rather than inheriting it from whichever cost happened to be measured.

This is a hard mathematical problem.

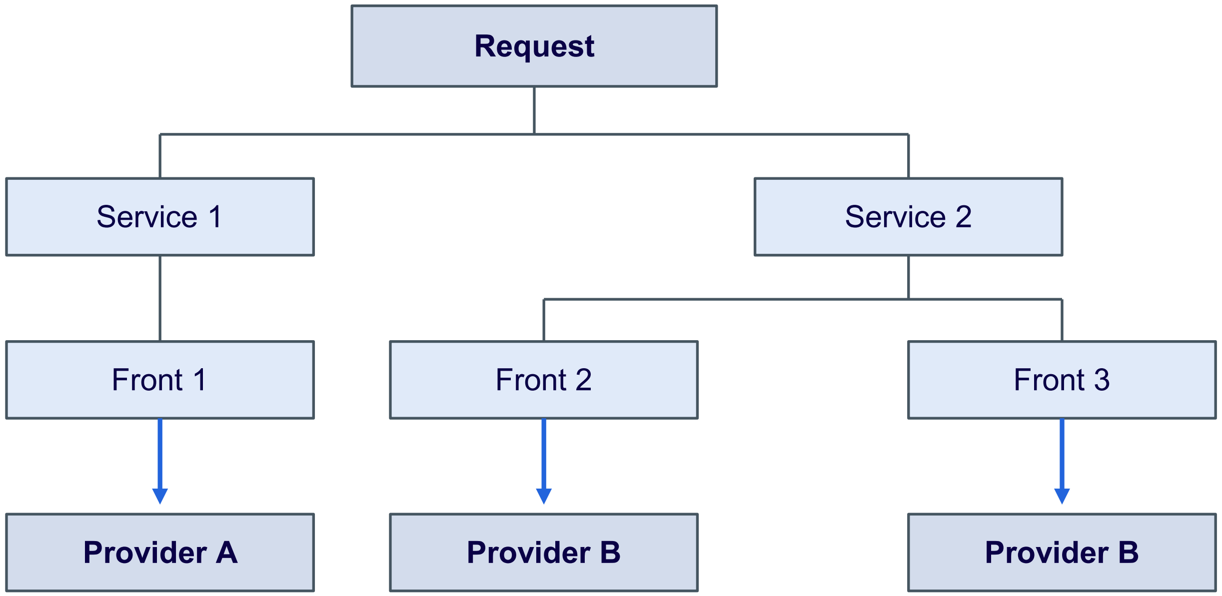
VI

Procurement Optimization

Allocating oilfield service providers across geographic fronts
under discount structures that make the problem genuinely combinatorial.

Procurement — The Combinatorial Structure

For each service required at each geographic front, several providers are admissible. Their prices are not separable across the assignment.



Why the problem is not separable

- Each provider quotes a base cost per service, and then two distinct discount structures on top of it.
- Service bundling: a discount that applies when several services are awarded to the same provider at the same front.
- Front bundling: a discount that applies when the same service is awarded to one provider across several fronts.
- Because both discounts depend on the assignment as a whole, the cost of assigning any single front–service pair depends on every other assignment. Greedy or pair-by-pair selection is therefore not merely suboptimal — it is ill-defined.

One request decomposes into services, each required at one or more geographic fronts; each front–service pair is assigned to a provider.

Practice before the model

There was no adequate optimization mechanism for assigning providers. A small number of possible combinations were enumerated by hand, and the cheapest among them was selected. The quality of the outcome depended on which combinations a buyer happened to think of.

Procurement — Formulation and Result

An integer program that evaluates the complete combinatorial space, and returns a ranked set of admissible allocations.

The model

A mathematical optimization model capable of considering all possible combinations of services, fronts, providers and discounts. Implemented as an integer program, it computes automatically the allocations that minimize total cost, with both bundling discounts represented explicitly in the objective rather than applied after the fact.

The deployment

The solution was integrated into the organization's existing system, which now computes the 45 best allocation options within the tool the buyers already use. The model replaced the previous selection logic with no change to the user interface — the workflow was preserved and only its content was substituted.

Results

45

Ranked allocation options returned per request, rather than a single recommendation

> 10,000

USD in savings generated per request in the runs performed, from correct application of the bundling discounts

An instructive detail

In several cases even the forty-fifth ranked option outperformed what the previous manual procedure had obtained. The gap was therefore not marginal: manual enumeration was not searching near the optimum at all. Returning a ranked set also preserves the buyer's ability to apply criteria the model does not represent — supplier reliability, contractual history, local capacity.

Three Observations that Recur Across the Portfolio

The cases differ in method. They do not differ in what determined whether they succeeded.

1 **The binding constraint is data engineering, not modelling**

In every case the first stage was integration, cleansing and standardization of operational records that were never designed to be analysed jointly. Where that stage was done well the modelling stage was routine; where it was weak, no estimator recovered the loss. In the development planning and drilling cases the labelling and integration work exceeded the modelling work by a wide margin, and it is the part that is never reported in a paper.

2 **The formulation of the objective matters more than the choice of algorithm**

Whether the objective includes carbon footprint, whether the logistics criterion is cost or time or emissions, whether a stabilized period is defined by variance in production or in pressure, whether a discount is inside the objective or applied afterwards — each of these decisions changed the solution more than any change of estimator would have. Formulation is where the engineering judgement lives.

3 **The systems that endure are the ones that fit the existing workflow**

The procurement model was adopted because it returned a ranked set inside the interface the buyers already used. The development planning system was adopted because it produced the artefact the asset teams were already required to deliver. A model that requires the organization to reorganize itself in order to use it will, in general, not be used.

Implications and the Frontier

What this portfolio suggests for engineers entering the industry, and where the open problems are.

For the engineer entering the industry

- The scarce skill is not the ability to fit a model. It is the ability to state an operational problem as a well-posed mathematical one, with an objective the organization recognizes as its own.
- Causal reasoning is not an academic refinement in this setting. Every operational parameter in a producing field was chosen by someone in response to conditions, which makes confounding the default and not the exception.
- Domain knowledge enters as specification, not as commentary: the regime split in the energy models, the stability criteria in the decline models, the constraint set in the planning model.

Where the technical frontier lies

- Causal identification with observational operational data at scale: credible estimation of heterogeneous effects when treatment assignment is an unrecorded human decision rule.
- Targets that are constructs rather than observations — stabilized periods, borehole quality — where the label carries irreducible analyst variability and the evaluation metric must reflect it.
- Genuinely multi-objective operation: production, cost, energy and emissions optimized jointly on the same network and the same plan, rather than reconciled after the fact.
- Optimization under uncertainty at industrial scale, where the deliverable is a characterized frontier of admissible plans rather than a point solution.

Thank you

Questions and discussion

Álvaro Riascos Villegas · Partner & Co-Director, Quantil

www.quantil.co · alvaro.riascos@quantil.com.co