What weather routing actually buys

A geometry / seasonal / forecast decomposition of optimizer fuel–ROI across ocean regimes, under ERA5+GLORYS replay

S. Laignel (Weather Routing & Ships Analytics)

2026-05-25 Confirmatory study against a pre-registered protocol. All work on branches; nothing deployed.

Abstract

Weather-routing vendors quote headline fuel savings (“we save $X\%$”). We show that the honest number is much smaller, and we decompose it into three components with sharply different robustness and commercial honesty: a *geometry* term (route shortening, free from any shortest-path tool), a *structural–seasonal* term β (the climatologically-knowable corridor and quasi-stationary current-riding, capturable *without* a forecast), and a *tactical* term α (the marginal fuel saved by reacting to the specific departure’s synoptic weather — the only part a real-time service adds). The estimator is built on a paired counterfactual: the same voyage flown through the *same* realized weather field (ERA5 reanalysis for wind/waves, CMEMS GLORYS for surface currents), benchmarked against a *constraint-respecting geodesic* rather than the ship’s planned route. Over a pre-registered four-route basket including a trade-wind negative control, the headline operational saving (vs. the planned route) is dominated by geometry and β; the tactical α — measured in hindcast, hence an *upper bound* — is small, and on the North Atlantic crossing it is *negative*: a fixed seasonal-prior route beats re-optimising for each departure’s forecast. The negative control manufactures no saving. The defensible weather-routing ROI is a few percent of fuel, lives mostly in route design and season, and is materially smaller than the marketed headline once measured against the right baseline.

1 Problem

A widely-quoted figure holds that weather routing saves on the order of a few to ten percent of voyage fuel. Two ambiguities make such figures hard to trust. First, the *baseline*: savings are routinely reported against a least-distance (great-circle) path or a company-planned route, so ordinary route-shortening is silently credited to “weather skill”. Second, *foresight*: most academic case studies and vendor demonstrations optimise on reanalysis or analysis fields, i.e. with the true future weather in hand (“hindcast”), which is an upper bound — the operationally realisable saving, with forecast error, is strictly lower. This study addresses both: it benchmarks against a constraint-respecting geodesic (removing route-design), and it separates the forecast-free structural value from the forecast-dependent tactical value, labelling the latter explicitly as a hindcast upper bound.

2 Methodology

2.1 Evaluation rig

Each voyage is evaluated in-process inside a throwaway container built from the production image, seeded with a SQLite copy of the production-calibrated mid-range tanker model (engine-log calibrated, 139 entries). The weather is sourced by *replay*: 10 m wind, significant wave height and SST from the Google ARCO ERA5 Zarr (0.25°, hourly), and surface currents (u_o, v_o) from

CMEMS GLORYS (`cmems_mod_glo_phy-cur_anfc_0.083deg_P1D-m`, 0.083°, daily). Currents enter the per-leg speed-over-ground and the fuel is rescaled to the SOG time basis, so currents affect both transit time and fuel. Twelve monthly departures (the first of each month of 2025, 0 UTC) are run per route at a fixed calm speed of 13 kt, laden, fixed engine setting (Mode 1). The optimiser is the isochrone engine with a distance-scaled time step (converging on all routes; see §3.5).

2.2 The constraint-respecting geodesic baseline

The reference is *not* the ship's planned route, which departs from the great circle for reasons unrelated to weather (hard law: coast-distance minima, traffic separation; soft law: master's prudence). Crediting the optimiser for undoing those is double-counting. We therefore build the *constrained geodesic*: a land-masked $A^\star$ on a 0.5° grid whose node validity uses the optimiser's own land mask (so both respect the *same* hard-law constraints — *constraint symmetry*), with cost equal to the great-circle (haversine) distance and an admissible heuristic, followed by a shore-aware string-pulling pass that removes grid metrication. The result is the shortest navigable path between origin and destination.

2.3 The geometry / β / α decomposition

Four nested baselines, each a valid "no-X-routing" reference, are flown through each month m's realized weather and currents (a paired design — each departure is its own control):

- $F_{\text{plan}}(m)$ — the planned route (the master's soft-law corridor);
- $F_{\text{geo}}(m)$ — the constraint-respecting geodesic (shortest legal path);
- $F_{\text{prior}}(m)$ — the *season-prior* route: the geodesic-Fréchet medoid of that season's spot-optimal routes, flown through month m (the climatological optimum);
- $F_{\text{spot}}(m)$ — the spot-optimal (isochrone) route for month m (the hindcast ceiling).

The total routing gain decomposes additively into three differences of fuels, each measured on the same month's weather:

$$G(m) = F_{\text{plan}}(m) - F_{\text{geo}}(m) \qquad \text{geometry / route-design,} \tag{1}$$

$$\beta(m) = F_{\text{geo}}(m) - F_{\text{prior}}(m) \qquad \text{structural–seasonal + current-riding,} \tag{2}$$

$$\alpha(m) = F_{\text{prior}}(m) - F_{\text{spot}}(m) \qquad \text{tactical / forecast-reactive,} \tag{3}$$

$$\underbrace{F_{\text{plan}}(m) - F_{\text{spot}}(m)}_{\text{total gain}} = G(m) + \beta(m) + \alpha(m). \tag{4}$$

Each component is reported in tonnes and as a percentage of $F_{\text{plan}}(m)$, with bootstrap 95% confidence intervals (10 000 resamples, seed 20 260 524) over the twelve departures, and by season. The interpretation is the spine of the study:

- **Geometry** G is a one-time route-design correction. A shortest-path tool delivers it; it is not weather routing.
- β is the value of being on the climatologically better corridor and riding the quasi-stationary boundary current (Agulhas, Gulf Stream). It is *knowable a priori* from routeing charts and a current climatology — it does *not* require a forecast.
- α is the only forecast-dependent component: the marginal fuel from reacting to *this* departure's synoptic weather over the seasonal prior. It is what real-time routing services actually sell.

2.4 Route basket and negative control

The basket was pre-registered before any run: **Agulhas** (Southern-Ocean storms + the Agulhas current), a **North Atlantic** crossing (mid-latitude depressions + the Gulf Stream), a **South-Atlantic trade-wind** leg as a **negative control** (benign trades, no dominant current — where a correct method must show ≈ 0 weather gain), and a short coastal leg (deferred). The negative control is the falsification test: if the optimiser "saves" fuel where there is no weather to exploit, the method is biased.

2.5 Honesty controls and validity threats

1. **Perfect foresight.** Replay gives the optimiser the true future, so α is an *upper bound*; the operational α (optimiser on the departure forecast, scored on realized weather) is lower and requires a forecast archive aligned to historical dates (specced, not run — we do not fabricate forecasts).
2. **In-sample β.** The season-prior route is the medoid of *this* year's optima, fit on the data the residual is measured on; this makes β slightly optimistic and α slightly conservative. The clean estimator uses a multi-year reanalysis climatology (specced follow-on).
3. **Currents.** Surface currents are included via GLORYS; the engine models current set/drift. (An earlier wind/wave-only run, without currents, materially understated routing value on the current-dominated routes — see §3.4.)
4. **Calibration window, grid resolution, optimiser feasibility, coarse-route geometry** are each reported (§3, §3.5). The numeric coast-distance floor is inert in this rig (the high-resolution coastline service was unavailable; land avoidance uses a 1 km raster), which relaxes both baselines symmetrically.

3 Results

3.1 Seasonal dispersion (Layer 1)

Baseline fuel on the fixed route across the twelve departures isolates pure weather-on-a-fixed-path. A weather-severity index (significant wave height dominant, plus headwind and wind speed) explains the dispersion strongly: $R^2 = 0.88$ (Agulhas), 0.97 (North Atlantic), 0.65 (the benign control). Routing's value scales with the severity a route is exposed to — it earns its keep in the stormy tail, not the median voyage.

3.2 The decomposition

Table 1 reports G, β, α (current-aware) as a percentage of planned-route fuel, with bootstrap CIs.

Table 1: Geometry / β / α decomposition of the routing gain (current-aware; % of planned-route fuel; bootstrap 95% CI). α is a hindcast upper bound.

Route	Geometry G	β (seasonal+current)	α (forecast, UB)	Total
Agulhas (storm+current)	11.9%	+1.1%	0.9% $[0.03, 1.85]$	14.0%
North Atlantic (Gulf Stream)	5.5%	−1.4%	0.7% $[0.30, 1.19]$	4.8%
Trade-wind (neg. control)	weather term	−0.2% $[-0.34, -0.02]$	(CI through/below 0 — unbiased)	

The pattern is consistent across regimes: the total gain is dominated by G and β — both forecast-free — while α is small. On the North Atlantic, α is *negative* ($[0.30, 1.19]$): a fixed seasonal-prior route *beats* re-optimising for each month's forecast, because the optimiser's storm-dodging detours add more distance than they save in hindcast.

3.3 Structural vs. tactical geometry

On Agulhas and the control the optimised routes collapse to a single corridor at the model-grid resolution (within-season $\approx$ across-season Fréchet spread), i.e. the optimum is structurally predictable from climatology. The North Atlantic genuinely swings ($\sim 6\times$ the grid width) — real tactical routing — but those swings do not pay off (negative α).

3.4 Currents: the wind/wave-only run understated value

An initial run used ERA5 only (no currents) and found the optimiser barely beating the geodesic on the current-dominated routes. That was a current-blind artefact: with GLORYS currents, the favorable Agulhas current lowers fuel along the (near-coast) shortest path, while the North Atlantic crossing fights an adverse current on the planned route, which a current-aware optimiser routes around. The decomposition makes the point precise: on Agulhas the structural term moved from $\beta = 0.1\%$ (current-blind, essentially zero) to $\beta = +1.1\%$ (current-aware), while the tactical term was unchanged ($\alpha \approx 1.0\%$ current-blind, 0.9% current-aware). Currents are *quasi-stationary* — they enter β (structural, knowable from a current climatology), not α — so including them *strengthens* the "structural > tactical" conclusion rather than overturning it: the current-route value the wind/wave-only run missed is forecast-free.

3.5 Production stability and charter evidence

The optimiser converges on 12/12 months for every basket route (after a distance-scaled time-step fix; it was 2/8 before). Under a waypoint stress test ($6 \rightarrow 200$ corridor waypoints) it produces a self-consistent route with no self-intersections or land crossings and bounded memory. Per-passage Beaufort/Douglas distributions and the fraction of passage time above charter exclusion thresholds (Beaufort 5 / Douglas 4) are produced as a charter-party evidence sheet per route.

4 Discussion

The honest reframing the evidence supports: weather routing is *a cheap, robust seasonal-structural core (route design G + climatology/current β) plus a small, tail-concentrated tactical premium α that real-time services sell* — and the marketed headline conflates all three against a flattering baseline. G and β are free (a shortest-path tool and a routeing/current chart deliver them); only α is genuinely chargeable, and α is small, hindcast-inflated, and on at least one major commercial route negative. This does not say routing is worthless — it beats no routing, and the tail value is real — but it relocates the value away from the real-time spot-optimisation that is marketed.

5 Limitations and follow-ons

The operational (forecast-error) α — the genuinely realisable number — requires an archived operational forecast aligned to historical departures and is specced, not run. The clean out-of-sample β requires a multi-year current+weather climatology. The forecast-only Cython engines were not replay-benched. These are the next steps; none is expected to raise α above its hindcast upper bound reported here.

Reproducibility

Pre-registered protocol, harness, constrained geodesics, per-route datasets and this article are committed on a branch; the decomposition is computed by `decompose_abg.py` from the per-route fuel records; all randomness is seeded (20 260 524).