http://www.egejfas.org
Ege Journal of Fisheries and Aquatic Sciences, 43(4), 372-382 (2026)

Su Ürünleri Dergisi (2026)
DOI: 10.12714/egejfas.43.4.06

RESEARCH ARTICLE ARAŞTIRMA MAKALESİ

Toward satellite-based growth prediction for mussel aquaculture in Turkish coastal waters: A Dynamic Energy Budget model with an empirically calibrated food parameter

Türkiye kıyı sularında midye yetiştiriciliği için uydu temelli büyüme tahminine doğru: Besin parametresi ampirik olarak kalibre edilmiş bir Dinamik Enerji Bütçesi modeli

Cem Hasşerbetçi[1]* • **Gökhan Tunçelli[2]** • **İdil Can-Tunçelli[3]**

[1]Faculty of Aquatic Sciences, Department of Food Safety, Istanbul University, İstanbul, 34134, Türkiye
[2]Faculty of Aquatic Sciences, Department of Aquaculture, Istanbul University, İstanbul, 34134, Türkiye
[3]Faculty of Aquatic Sciences, Department of Seafood Processing Technology, Istanbul University, İstanbul, 34134, Türkiye

*Corresponding author: cemhasserbetci@istanbul.edu.tr
Received date: 08.06.2026 Accepted date: 21.09.2026

How to cite this paper:
Hasşerbetçi, C., Tunçelli, G., & Can-Tunçelli, İ. (2026). Toward satellite-based growth prediction for mussel aquaculture in Turkish coastal waters: A Dynamic Energy Budget model with an empirically calibrated food parameter. *Ege Journal of Fisheries and Aquatic Sciences, 43*(4), 372-382. https://doi.org/10.12714/egejfas.43.4.06

Abstract: The Mediterranean mussel *Mytilus galloprovincialis* is the principal cultured bivalve in Turkish coastal waters, and Dynamic Energy Budget (DEB) models forced with satellite ocean colour offer a low-cost route to growth prediction. In practice, published parameter sets underestimate shell growth in coastal settings, and the reason is unresolved. We tested a published covariation-method DEB parameterisation against ten grow-out experiments in Türkiye (43 cohort groups in the Aegean, the Dardanelles, the Marmara and the Black Sea), forced with reanalysis sea-surface temperature at culture depth and satellite Level-4 chlorophyll-a. Applied as published (X_K=2.1 µg L^{-1}) the model under-predicted at every experiment (RMSE 19.8 mm, bias -16.5 mm). Recalibrating the half-saturation constant X_K alone, with no biological parameter changed, reduced error to 12.7 mm (95% CI 8.1–17.9) under leave-one-experiment-out cross-validation with every data-derived choice refitted inside each fold, and out-performed a satellite-to-in-situ ratio correction by 4.5 mm. We then tested the measurement-scale explanation for that improvement and did not find support for it: at the one site where in-situ chlorophyll is resolved in time and the product observes the site itself the ratio varied three-fold between seasons, rescaling X_K is algebraically identical to rescaling the forcing, and at two sites the product returns no valid retrieval at the deployment at all. The corrected model is therefore an empirical forcing-context calibration, whose residual absorbs unmodelled food quality and culture-system effects, rather than a measurement-scale conversion. Its limits matter as much: a near-zero overall bias masks a mean absolute per-experiment bias of 9.6 mm, and no site-ranking skill was demonstrable against baselines using no chlorophyll or no model. Satellite-forced DEB can reproduce the central tendency of Turkish mussel growth after calibration, but does not yet resolve the between-site differences on which farm siting depends.

Keywords: Model calibration, half-saturation constant, remote sensing, growth prediction, ocean colour, *Mytilus galloprovincialis*

Öz: Akdeniz midyesi *Mytilus galloprovincialis*, Türkiye kıyı sularında yetiştiriciliği yapılan başlıca çift kabuklu tür olup, uydu okyanus rengi verileriyle zorlanan Dinamik Enerji Bütçesi (DEB) modelleri büyüme tahmini için düşük maliyetli bir yaklaşım sunmaktadır. Ancak pratikte, yayımlanmış parametre setleri kıyısal ortamlarda kabuk büyümesini düşük tahmin etmektedir ve bunun nedeni henüz açıklığa kavuşturulmamıştır. Bu çalışmada, kovaryasyon yöntemiyle elde edilmiş, daha önce yayımlanmış bir DEB parametrizasyonu, Türkiye'deki on büyütme denemesine (Ege, Çanakkale Boğazı, Marmara ve Karadeniz'de toplam 43 kohort grubu) karşı, yetiştiricilik derinliğinde re-analiz deniz yüzeyi sıcaklığı (SST) ve uydu tabanlı Seviye-4 (L4) klorofil-a ile zorlanarak test edilmiştir. Model yayımlandığı haliyle (X_K=2,1 µg L^{-1}) uygulandığında her denemede büyümeyi düşük tahmin etmiştir (RMSE 19,8 mm; yanlılık -16,5 mm). Hiçbir biyolojik parametre değiştirilmeden yalnızca yarı doygunluk sabiti X_K'nın yeniden kalibre edildiğinde, her katmanın içinde veriden türetilen bütün seçimlerin yeniden hesaplandığı bir-denemeyi-dışarıda-bırakma çapraz doğrulaması altında hata 12,7 mm'ye (%95 güven aralığı 8,1–17,9) düşmüş ve bu sonuç uydu-yerinde oran düzeltmesine göre 4,5 mm daha düşük hatayla daha iyi sonuç vermiştir. Bu iyileşmenin ölçüm ölçeğine dayalı açıklaması ayrıca sınanmıştır ve desteklenmediği görülmüştür. Yerinde klorofil ölçümlerinin zaman içinde çözümlenebildiği ve ürünün alanın kendisini gözlediği tek istasyonda oran mevsimler arasında üç kat değişmiştir. □'yı ölçeklemek ile girdiyi (zorlayıcı veriyi) ölçeklemek cebirsel olarak özdeştir; iki istasyonda ise ürün, yerleştirme süresince hiçbir geçerli geri kazanım değeri vermemektedir. Dolayısıyla düzeltilmiş model, bir ölçüm ölçeği dönüşümünden ziyade, kalıntısının modellenmemiş besin kalitesi ve yetiştirme sistemi etkilerini absorbe ettiği ampirik bir zorlama-bağlamı kalibrasyonudur. Modelin sınırları da en az bu bulgu kadar önemlidir: sıfıra yakın genel yanlılık, deneme başına 9,6 mm'lik ortalama mutlak yanlılığı gizlemektedir ve klorofil kullanmayan ya da model kullanmayan temel karşılaştırmalara karşı gösterilebilir bir saha sıralama becerisi tespit edilememiştir. Sonuç olarak uydu girdili DEB modelleri, kalibrasyon sonrasında Türkiye'deki midye büyümesinin genel eğilimini yeniden üretebilmekte, ancak yetiştiricilik yer seçiminin dayandığı sahalar arası farkları henüz çözememektedir.

Anahtar kelimeler: Model kalibrasyonu, yarı doygunluk sabiti, uzaktan algılama, büyüme tahmini, okyanus rengi, *Mytilus galloprovincialis*

INTRODUCTION

The Mediterranean mussel *Mytilus galloprovincialis* is Türkiye's most commercially important cultured bivalve, grown from the southern Aegean to the Black Sea across a wide range of hydrographic conditions (Acarlı et al., 2024; Çelik et al., 2015; Karayücel et al., 2003, 2010; Keskin et al., 2020; Keskin and Ekici, 2021; Lök et al., 2007; Şensurat-Genç et al.,

© Published by Ege University Faculty of Fisheries, İzmir, Türkiye

2021; Yildirim et al., 2024; Yıldırım and Ercan, 2024; Yıldız and Lök, 2005). The sector faces two persistent management challenges. First, site selection is still largely empirical: operators identify productive locations through trial deployments rather than predictive screening, which is slow and costly. Second, deployment timing and stocking decisions are made without quantitative growth forecasts, leaving producers unable to anticipate bottlenecks or to optimise rotation cycles.

Predictive growth models driven by freely available satellite data could address both challenges at low cost. Dynamic Energy Budget (DEB) theory (Kooijman, 2010) provides a mechanistic framework that converts sea-surface temperature (SST) and food availability, proxied by chlorophyll-a (CHL-a), into species-specific growth trajectories, and it has been applied widely to cultured bivalves including *M. galloprovincialis* (Hatzonikolakis et al., 2017; Monaco et al., 2019; Sarà et al., 2012). When satellite and reanalysis products supply the environmental forcing, the model can in principle generate basin-scale suitability maps and time-to-harvest estimates for any candidate site without prior field deployment (Palmer et al., 2020; Stechele et al., 2022; Thomas et al., 2011).

In practice, however, DEB models forced with satellite chlorophyll have consistently underestimated mussel growth in coastal settings, undermining confidence in their operational use. Previous studies have attributed this to parameter uncertainty or to site-specific environmental complexity (Hatzonikolakis et al., 2017; Monaco et al., 2019). A further candidate has received less attention. The half-saturation constant X_K of the Holling Type II functional response, which sets how sensitive modelled growth is to food, is typically calibrated against *in-situ* chlorophyll measurements yet applied to *satellite* chlorophyll retrievals that are systematically lower in coastal waters. Satellite ocean-colour algorithms are known to underestimate chlorophyll in turbid, nearshore Case-2 waters by factors of two to ten (Cherif et al., 2021; Gohin, 2011; Lacava et al., 2018), so a model forced this way may perceive the mussels as chronically food-limited even where in-situ conditions are adequate, and where the mismatch is severe it can depress the food-dependent maximum size toward the observed initial size (Palmer et al., 2020). Whether this mechanism accounts for the observed bias, and whether it can be separated from the other processes a food parameter might absorb, are the subject of this study.

Türkiye's coastal waters make these questions tractable. Within a few hundred kilometres they span the oligotrophic southern Aegean, the mesotrophic northern Aegean, the Dardanelles (a two-layer strait rather than a basin, with its own hydrography), the brackish Marmara and the brackish Black Sea, so a framework evaluated across this range covers the full extent of *M. galloprovincialis* aquaculture in the country. Summer SST at culture depth exceeds 25 °C at five of the ten experiments, reaching 27.6 °C at the warmest, which also constrains the parameter set: among published *M. galloprovincialis* parameterisations, only the covariation-method set of Monaco and McQuaid (Monaco and McQuaid, 2018) carries an upper thermal bound high enough to remain valid at these temperatures.

This study addresses three questions, all grounded in aquaculture practice. First, we quantify the magnitude and distribution of the bias incurred by a published parameter set applied without calibration, and ask whether that bias tracks the food forcing rather than the biological or thermal parameters. Second, we ask whether recalibrating X_K alone repairs the fit, under cross-validation at the level of whole experiments so that reported performance refers to sites the calibration has not seen. Third, we test the measurement-scale explanation directly: a ratio between satellite and in-situ chlorophyll is a scale conversion only if the same factor links the two at each point in time. The third question determines how the first two should be read, because an improvement in fit after adjusting a food parameter is equally consistent with that parameter absorbing salinity, seston quality, culture-system hydrodynamics or retrieval error.

MATERIALS AND METHODS

Study area and cohort data

The study domain is the coast of Türkiye, spanning the southern and northern Aegean, the Dardanelles, the Sea of Marmara and the southern Black Sea (Figure 1).

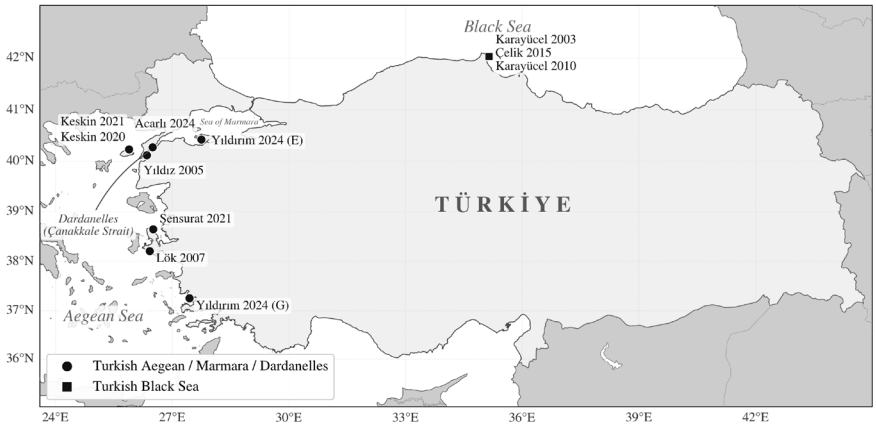

Figure 1. Study area and validation sites across the coast of Türkiye (Aegean, Marmara, Dardanelles and Black Sea)

The four systems differ in trophic status and hydrography: the southern Aegean is oligotrophic and optically clear; the Dardanelles is a two-layer strait carrying brackish Black Sea outflow above saline Mediterranean inflow rather than a basin in its own right; the Marmara is brackish and semi-enclosed (23.7–25.9 PSU (Yildirim et al., 2024)); and the Black Sea is the least saline and most strongly stratified (annual means 14.9–17.9 PSU at the three Sinop experiments (Çelik et al., 2015; Karayücel et al., 2003, 2010)). Salinity is taken from the cohort's own source publications and is used for description and grouping only, not as a model input.

Calibration units are distinct from that description, and are chosen for what the data can identify rather than for hydrographic similarity: *aegean_med* (7 experiments, 39 cohort groups) pools the Aegean, Dardanelles and Marmara sites, and *black_sea* (3 experiments, 4 cohort groups) covers the Sinop-area deployments. Alternative groupings, including a single group and a three-way split separating the Marmara, are reported as a sensitivity.

Ten published *M. galloprovincialis* growth experiments were digitised into a uniform cohort-group schema (initial length, final length, duration, culture system, coordinates, depth), totalling 43 cohort groups spanning spat to adult (initial lengths 3.2–70.2 mm), deployments of 8 to 22 months, and suspended systems (longline, raft, rope, net) as well as reef-based deployments (Table S3).

Inclusion required a deployment within the study domain, cohort tracking, a locatable site, and a deployment period within the satellite record; no experiment meeting these criteria was excluded. One further indexed experiment, Acarlı et al. (2024), is cross-sectional rather than cohort-tracking, and is retained only as a plausibility check contributing to no fitted quantity. Six of the ten involve figure digitisation, cross-checked against the source paper's prose, an audit that corrected the Erdek in-situ chlorophyll to 1.843 µg L^{-1} (Supplementary Methods S1, Tables S3–S4). Two sites recur: Mersin Bay (İzmir, Turkish Aegean) is small, enclosed and optically complex; Güllük Bay (Muğla, southern Aegean) pairs an offshore fish-farm station with a coastal control, providing an internal test of local enrichment. Culture systems are heterogeneous and not resolved by the model; where a source reports several at one site these are kept as separate cohort groups, so the variance stays visible in the residuals. Observational uncertainty is unevenly reported, 29 of 43 groups carrying a final-length SD (Table S3), so residuals are unweighted and the Results quantify what that costs.

DEB model and parameters

We use the post-metamorphosis formulation of the abj model for *M. galloprovincialis* (Monaco and McQuaid, 2018), one of the standard typified DEB models of the Add-my-Pet collection (Kooijman, 2014; Marques et al., 2018). It differs from the standard model in that between birth and metamorphosis the animal grows with *metabolic acceleration*, its surface-specific assimilation rate and energy conductance rising in proportion to structural length (Kooijman, 2014), which is the appropriate formulation for mytilids. Every cohort here begins well past metamorphosis, so the acceleration factor $S_M=L_j/L_b$ is constant throughout and is pre-applied to both rates (Supplementary Methods S5).

The three integrated state variables (reserve energy E, structural volume V, reproduction buffer E_R) follow the standard abj flux definitions (Kooijman, 2010; Monaco and McQuaid, 2018), with structural shrinkage disallowed ($\dot{V}\geq 0$) (Hatzonikolakis et al., 2017; Monaco and McQuaid, 2018), and the observable used for validation is physical shell length, $L=10\ V^{1/3}/\delta_M$ mm. Temperature dependence enters every rate through the five-parameter Arrhenius correction of Kooijman (2010), and feeding follows a Holling Type II functional response in CHL-a directly, $f(X)=X/(X+X_K)$, without carbon conversion, in line with prior satellite-DEB practice for this species (Monaco et al., 2019; Monaco and McQuaid, 2018; Palmer et al., 2020; Thomas et al., 2011). Spawning is not represented, which is immaterial for structural shell length but not for tissue weight; validation is therefore on length alone. Every cohort is initialised at full reserve density, an optimistic assumption tested in the Results. Full definitions are in Supplementary Methods S5, with the parameter table in Table S2.

The parameter set is adopted from Monaco and McQuaid (2018), the only published *M. galloprovincialis* parameterisation meeting three criteria essential for the waters of Türkiye: an upper thermal bound high enough for summer SST (T_H=306.1 K ≡33 °C); core parameters co-estimated by the covariation method (Lika et al., 2011) against species-specific data rather than inherited from *M. edulis* (Van der Veer et al., 2006); and independent field validation (MAPE = 10.3% for shell length at three South African sites). Four parameters come from elsewhere and each carries its provenance in Table S2: energy conductance, maturity maintenance and puberty threshold are gap-filled from the Add-my-Pet entry (Montalto et al., 2018), which shares authorship and estimation framework with the Monaco set, and the shape coefficient δ_M=0.23 is the morphometric estimate of Sarà et al. (2011); (see also Sarà et al., 1998), preferred over the covariation value because δ_M enters the length observable directly. The half-saturation constant X_K=2.1 µg L^{-1} (Sarà et al., 2012) was calibrated against in-situ CHL-a, and is the only parameter this study recalibrates.

One derived quantity is used as a diagnostic throughout. Growth at reserve equilibrium requires $L_{max}(f)>L$, where

$$L_{max}(f)=\frac{10\ \kappa\ \{\dot{p}_{Am}\}_{eff}}{[\dot{p}_M]\ \delta_M}\ f \approx 155.5\ f\ \text{mm}$$

for this parameter set. The threshold is independent of temperature, because $\kappa\dot{p}_C$ and $\dot{p}_M$ carry the same $TC(T)$ factor, which cancels. We evaluate it from the daily forcing series rather than from a site mean, as the fraction of deployment

days on which $L_{max}(f)<L_0$, and report its association with model error rather than treating it as a binary criterion (Figure S6, Supplementary Methods S3).

Environmental forcing

SST was extracted at mussel culture depth (1–25 m, site-specific) from the Copernicus Marine Mediterranean Sea Physics Reanalysis (E.U. Copernicus Marine Service Information (CMEMS), 2022b), and from the Black Sea Physics Reanalysis (E.U. Copernicus Marine Service Information (CMEMS), 2022a) for Sinop-area sites: NEMO-based coupled reanalyses at ∼4.2 and ∼2.5 km resolution, daily mean, with shallowest levels at 1.0 and 0.5 m. Depth-corrected rather than surface SST matters for summer predictions at moored sites, since surface SST exceeds culture-depth SST under summer stratification. Primary CHL-a forcing is the Copernicus Marine daily gap-free Level-4 ocean colour product (multi-sensor merged, regionally tuned, bias-corrected against in-situ databases): Mediterranean (E.U. Copernicus Marine Service Information (CMEMS), 2023b) at ∼1 km resolution, and the Black Sea-specific L4 product (E.U. Copernicus Marine Service Information (CMEMS), 2023a) for Sinop-area sites.

We retain the satellite product as primary forcing because it is independent of any biogeochemical model, is available operationally, and covers every site; a coupled biogeochemical reanalysis (Ciavatta et al., 2019; Teruzzi et al., 2018), which has no coverage at one site, is carried as a sensitivity run and compared at each product's own calibrated optimum (Figure S4).

CHL-a was averaged over a box centred on the deployment, of half-width 0.1° (approximately 20 km) by default, widened in steps to 1° where that box returned no valid retrieval (Supplementary Methods S2; per-experiment buffers and pixel counts in Table S5). At two sites that widening is not a technicality. The Dardanelles site of Yıldız and Lök (2005) rests on three valid pixels 12 km north-west of the deployment; and at Erdek no box up to 0.5° returns any valid retrieval, the Mediterranean L4 product not covering the Sea of Marmara, so the 1° box that succeeds draws all 226 of its valid pixels from 68 to 85 km west. These two experiments supply 18 of the 43 cohort groups, and their forcing is treated as spatially unmatched throughout: Erdek is excluded from the satellite-to-in-situ ratio entirely, and both are identified wherever site-level inference is drawn. Separately, 107 of the 382 days of the 1997 Sinop deployment predate SeaWiFS and were assigned that site's September climatological mean, a fill tested explicitly because it covers the period contributing most of that experiment's growth.

Correction schemes and X_K calibration

Four ways of reconciling the model with satellite forcing are evaluated in one framework. The ratio orientation is fixed once, for the whole paper, as $\alpha_i = X_i^{sat}/X_i^{in\text{-}situ}$, so $\alpha<1$ means the satellite reports lower concentrations than the in-situ instrument; X_i^{sat} is the mean of the satellite series at site i and $X_i^{in\text{-}situ}$ the annual-mean in-situ CHL-a reported by the source paper (Table S4). The schemes are: the published parameter unchanged; the global median α applied as $X_K^{sat} = \alpha\, X_K^{in\text{-}situ}$; a per-site α_i; and X_K fitted freely to growth data, as a single value and as one value per calibration group. Rescaling X_K by α is algebraically identical to dividing the forcing by α, since

$$\frac{X}{X+\alpha X_K} = \frac{X/\alpha}{(X/\alpha)+X_K},$$

so converting the measurement scale and calibrating the parameter denote one operation and cannot be distinguished by their effect on the model. The identity holds exactly only if α is constant at each time point, because f is evaluated daily and is non-linear; a ratio of long-term means does not in general preserve it, and that assumption is tested in the Results.

X_K is calibrated by exhaustive grid search over log-spaced values on [0.01,10] µg L^{-1} (312 points), resolving the optimum to better than 2.3% of its value. The objective is the mean absolute relative error (MARE) on final shell length, the DEB calibration convention (Lika et al., 2011), which prevents large-bodied groups from dominating the fit; a grid is preferred to a directed search because it returns the full MARE profile. Identifiability is assessed from that profile: the plateau is the set of X_K values scoring within 5% of the minimum, and an optimum is treated as bounded rather than estimated when the plateau reaches a grid boundary.

Validation and statistical analysis

Generalisation is assessed by leave-one-experiment-out cross-validation, folds being defined at the experiment level because cohort groups within an experiment share forcing, site, deployment and operator. Within each fold every data-derived quantity is recomputed from the training experiments alone, so each fold predicts a site that contributed nothing to the parameters used to predict it; the naive variant fitted once on the full cohort is reported alongside, so leakage is measured rather than assumed. Site ranking, being among the intended applications, is tested directly, as Spearman correlations against three counterfactuals: each site's own mean chlorophyll held constant, one constant chlorophyll everywhere, and a no-model arithmetic baseline. Full protocols are in Supplementary Methods S3.

Three metrics are reported jointly: RMSE (primary, in mm), bias (mean signed residual pred − obs, mm), and MARE (the calibration objective). Overall bias is reported alongside the mean absolute per-experiment bias, because opposite-signed site biases cancel in the overall figure.

Cohort groups are nested within experiments and are not independent. That dependence is quantified by a one-way random-effects intraclass correlation of the residuals, from which a design effect and an effective sample size n_{eff} are derived, and all resampling draws whole experiments rather

than cohort groups: an experiment-level block bootstrap for X_K, cluster bootstraps for intervals on RMSE and on rank correlations, and paired differences on shared resamples for comparisons between schemes.

The X_K calibration is repeated under three published thermal parameterisations: Monaco 2018 (T_A=10590 K, full 5-parameter Arrhenius), AmP mod_3 (Montalto et al., 2018) (T_A=14821 K, with T_{AL}≈0 effectively disabling cold suppression), and Van der Veer (Van der Veer et al., 2006) (T_A=5800 K, single-T_A Arrhenius with no thermal bounds). These are compared by the corrected Akaike information criterion (Burnham and Anderson, 2002) on residuals in shell length, counting as parameters only those fitted to growth data. Because the underlying likelihood assumes independent residuals and they are not, AICc is evaluated at n_{eff} as well as at the nominal n; paired Wilcoxon signed-rank tests (Wilcoxon, 1945) on squared residuals, with the Diebold–Mariano *t*-approximation (Diebold and Mariano, 1995) as a check, are computed at each thermal set's own calibrated X_K.

Software and reproducibility

The model is implemented in Python with NumPy and SciPy, the reserve–structure–reproduction ODE system being integrated with scipy.integrate.odeint, which uses the LSODA adaptive stiff solver, at relative tolerance 10^{-8}, absolute tolerance 10^{-10} and a maximum step of 1 day. The implementation was verified against the tabulated values of Monaco and McQuaid (2018) and the AmP entry (Montalto et al., 2018) before any validation run, and its predictions were reproduced on two independent software stacks differing in Python, NumPy and SciPy version. Solver configuration, seeding and the verification protocol are detailed in Supplementary Methods S4; the complete analysis, with an environment specification and a single-command reproduction script, is deposited as described under Data Availability.

RESULTS

Performance of the published parameter set

Forced with depth-corrected SST and satellite CHL-a under the published Monaco 2018 set with X_K=2.1 µg L^{-1}, the DEB model produces essentially zero growth across most of the cohort (Figure 2a): overall RMSE 19.81 mm, bias −16.54 mm across 43 groups from 10 experiments. The pattern is one-sided: every experiment under-predicts, and residual magnitude tracks the satellite CHL-a mean (r=-0.54 across the 43 groups). The only experiment in quantitative agreement is Güllük Bay (south Aegean; RMSE 3.62 mm, bias -3.34 mm); every other under-predicts by more than 10 mm, up to 30.9 mm (karayucel2010_sinop_raft).

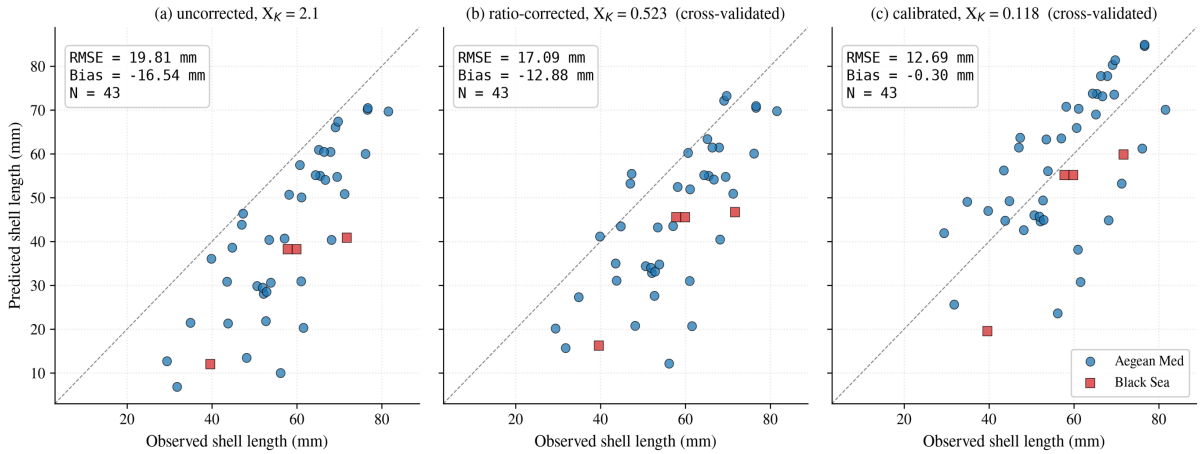

Figure 2. Predicted vs observed final shell length under three correction schemes: (a) the published parameter unchanged, X_K=2.1 µg L^{-1}; (b) the global ratio correction, X_K=0.523; (c) X_K calibrated on growth data, X_K=0.118. Panels (b) and (c) show cross-validated predictions, every point predicted by a model refitted without its own experiment. Circles are aegean_med, squares black_sea; the dashed line is 1:1.

The deficit is located in the food response. Evaluated on the time-varying forcing, the food-dependent ceiling $L_{max}(f)$ (Methods) falls below the initial size L_0, when f is evaluated at the site-mean chlorophyll, at 32 of 43 cohort groups at X_K=2.1 µg L^{-1}, at 9 of 43 under the ratio correction and at 1 of 43 at the calibrated value derived below (Table S6). Predicted growth at those cohorts is negligible rather than absent, about 2% of observation, transient blooms lifting f above the threshold. The ceiling does not, however, predict model error, the Spearman correlation between the fraction of days below threshold and relative error being -0.216 (95% cluster-bootstrap CI [-0.668,+0.274]) at the calibrated X_K; it is therefore reported as a property of the food response and not as a screen (Table S6).

Food-scale correction and calibration

For the experiments reporting in-situ CHL-a we computed $\alpha_i = X_i^{sat}/X_i^{in\text{-}situ}$ (Table S4). Erdek is excluded, its satellite series being drawn from the north Aegean rather than from the site. Over the remaining eight experiments the ratio spans two orders of magnitude, from 0.022 at Mersin Bay to 1.83 at Gökçeada, with median α=0.249, giving X_K^{sat}=0.523 µg L^{-1}.

Whether that spread reflects genuine site differences or the absence of any stable conversion can be tested only where in-situ observations are resolved in time, and the cohort supports one such test. At Güllük, in-situ CHL-a is reported by station and by season and the satellite series comes from a genuine local box; across those four station-by-season windows α spans 3.0-fold, and seasonally, being high when in-situ concentrations are low and low when they are high. A single multiplicative factor therefore does not describe the relationship even within one site and one year. Table 1 evaluates every correction scheme in one framework.

Table 1. Correction schemes evaluated in one framework on all 43 cohort groups. The first RMSE column scores the experiments used to choose X_K; the second refits every data-derived quantity inside each leave-one-experiment-out fold

How X_K is obtained	X_K used (µg L^{-1})	Parameters fitted to growth	RMSE on the experiments used to fit (mm)	RMSE on unseen experiments (mm)	Bias on unseen (mm)
Published value, no correction	2.1	0	19.81	19.81	-16.54
Satellite/in-situ ratio, one value for all sites	0.523	0	16.80	17.09	-12.88
Satellite/in-situ ratio, computed per site	varies by site	0	13.25	17.09	-12.88
Fitted to growth data, one value	0.118	1	11.26	12.69	-0.30
Fitted to growth data, one per group	0.130 / ≤ 0.01	2	11.25	12.16	-1.20

Fitting X_K outperforms the ratio correction by 4.54 mm (95% CI 1.19–7.72) on paired experiment resamples: the α-based schemes recover roughly half the deficit and retain a large negative bias, whereas free calibration removes the bias almost entirely. Cross-validation costs 1.40 mm (95% CI 0.53–2.94), so the in-sample figure of 11.26 mm is optimistic and 12.69 mm (95% CI 8.08–17.94) is the appropriate out-of-sample statement. The per-site α scheme has no out-of-sample advantage, becoming numerically identical to the global median once nested, because a held-out site contributes no in-situ measurement of its own. One calibrated value cannot be distinguished from two either, the paired difference between the single and per-group fits being -0.58 mm (95% CI -2.05 to +0.09), so the single value is reported as primary and the two-group fit as a sensitivity; a three-way split changes overall RMSE by 0.07 mm.

The calibrated value is well determined in the aggregate but not everywhere. On the aegean_med subset the MARE plateau runs from 0.088 to 0.164, interior to the search grid; for the four Black Sea cohort groups it runs to the grid floor, so the optimum is censored, the functional response being saturating at the observed Sinop chlorophyll levels for any sufficiently small X_K (Figure S3). This is a loss of identifiability rather than biological information (Lavaud et al., 2017), and the Black Sea value is reported as a bound, not an estimate. Regional intervals reflect the same imbalance: aegean_med 12.76 mm (95% CI 7.99–18.73, 7 experiments) against black_sea 11.95 mm (95% CI 3.84–20.03, 3 experiments).

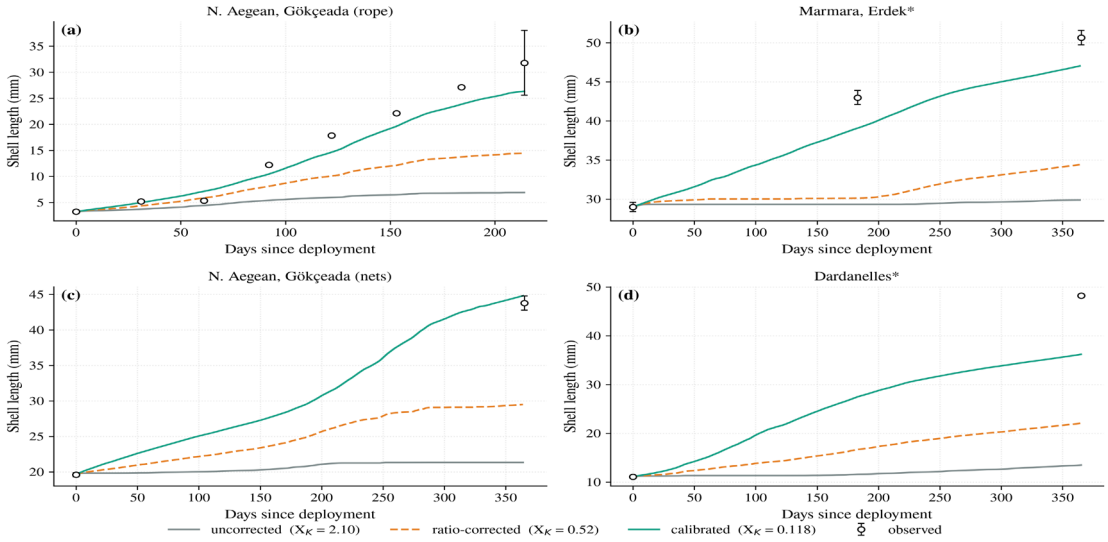

Figure 3. Simulated growth at four representative sites under the uncorrected parameter set, the global ratio correction and the calibrated X_K, against observed lengths (symbols, with SD where reported). Asterisks mark the two experiments whose forcing is regional rather than local (Methods). All 43 cohort groups are in Figure S2.

Distribution of error, and generalisation

Overall bias after calibration is -0.30 mm, but that near-zero value is produced by cancellation between sites rather than by accuracy at any of them: the mean absolute per-experiment bias is 9.55 mm, individual experiments range from -21.98 to +9.33 mm, and four of ten over-predict while six under-predict (Table S7). The design is also strongly clustered, the intraclass correlation of residuals within experiments being 0.78 and giving a design effect of 3.57, so the 43 cohort groups carry approximately the information of 12 independent observations, and all inference here uses that basis. Observational uncertainty does not account for the misfit, the median reported final-length SD being 0.90 mm against a median absolute model residual of 7.39 mm (Supplementary Methods S3).

One published experiment outside the cohort makes the spatial limit of the forcing explicit. At Pašman, Croatia (Peharda et al., 2007), three cohorts from one initial stock were grown at 0, 60 and 700 m from a fish farm; they differ by 5.13 mm in observed final length, 25% of the mean growth increment, yet receive the same prediction to three decimals, because at ∼1 km resolution they occupy a single pixel. The experiment enters no fitted quantity here and is reported only for this comparison.

The ability to rank sites by growth potential was tested directly, and none was demonstrable (Table S8). The calibrated model's site-level rank correlation with observed growth rate is ρ=0.297 (95% interval -0.615 to 0.926), and once the initial size and duration the model was given are controlled for, the correlation is negative in every configuration driven by real satellite chlorophyll. The only interval excluding zero belongs to the configuration supplied with no chlorophyll at all (0.603, CI 0.075–0.888), and a no-model arithmetic baseline ranks final length at 0.843 against the calibrated model's 0.722; all paired differences nonetheless span zero, so with ten experiments the design cannot establish a difference in either direction. The pattern is visible in the site means: Güllük carries the highest satellite chlorophyll and the second-slowest observed growth, Mersin the lowest chlorophyll and the fourth-fastest.

Sensitivity and robustness

Recalibrating X_K jointly at each of three published thermal sets (Figure S5, Table S1) leaves the fit essentially unchanged, and every statistic below is reported on the clustered basis established above. RMSE spans 0.6 mm across the 43 cohort groups, from 10.97 mm (Van der Veer) to 11.58 mm (AmP), and 9.28–11.00 mm when each of the ten experiments is weighted once, with calibrated X_K within 0.128–0.146. Paired tests comparing the sets *at their own calibrated values*, one loss per experiment rather than one per group, give p=0.70-0.92. Model selection agrees: at the effective sample size n_{eff}=12.0 all three sets fall within ΔAICc<2 of one another, Monaco against Van der Veer being 0.85, where the nominal n=43 would have given 3.03 and placed the two in different support bands (Burnham and Anderson, 2002). None of the three is therefore distinguishable on this cohort. A temperature-only sweep at fixed X_K confirms that food rather than T_A dominates predictions under satellite forcing (Figure S1), RMSE varying by 1.7 mm over $T_A \in$[3000,18000] K.

The calibrated X_K and the overall fit were tested against six analytical choices, each recomputed rather than re-scored (Table S10). Four are immaterial. The 1997 Sinop forcing fill, halved, doubled, replaced by the record mean or dropped with its experiment, and the initial reserve density, from full reserves down to 40%, leave the calibrated X_K unchanged and move overall error by at most 0.26 mm; how sites are grouped, across one, two or three calibration groups, moves it by at most 0.25 mm; and digitisation accuracy moves X_K by $\pm 10\%$, well inside its bootstrap interval. The remaining two are not.

Spatial extraction geometry is the largest by an order of magnitude, moving the calibrated X_K 3.5-fold, from 0.047 to 0.165, and overall error by 2.37 mm with a change in the sign of the bias (Table S9). The calibration is identical from the nearest valid pixel out to a 10 km box and becomes unstable at the 20 km box used here and beyond, the proximate cause being Güllük, an enclosed bay whose mean CHL-a falls from 1.00 to 0.19 µg L^{-1} as the box widens. Widening does not degrade the fit monotonically, which is consistent with the ranking result above: if the local pixel carried the growth-relevant signal, tightening the box would improve the fit, and it does not. Refitting without the two spatially unmatched experiments gives X_K=0.136 and 0.102 against 0.118 for the full cohort, all inside the full-cohort plateau. The chlorophyll product is the second: compared at each product's own calibrated optimum, on the nine experiments both cover, the biogeochemical reanalysis fits better by 1.42 mm, although the interval reaches the origin (95% CI 0.03–3.72) and the two products are therefore not clearly separated; at the uncorrected X_K=2.1 they are indistinguishable, that being a configuration at which both products fail. What the comparison displaces is the justification we gave for the choice of product, not the choice itself.

Three sites bound the result individually. Mersin Bay (Lök et al., 2007) retains 23.1 mm RMSE and -21.98 mm bias after calibration, with a satellite-to-in-situ ratio of 0.022, the lowest in the cohort by an order of magnitude; the single offshore Sinop cohort group of Çelik et al. (2015) retains 20.0 mm, combining cold water (mean 10.7 °C) with small initial size; and Erdek is not anomalous once its in-situ value is corrected.

DISCUSSION

What the calibrated parameter is

Recalibrating a single parameter repaired most of the deficit, and the parameter that mattered was the one describing food: neither the thermal parameters nor the biological set had to be touched. Taken with the raw model's one-sided under-prediction at every experiment, this locates the failure in how

the model perceives food rather than in how it converts food into growth.

It is tempting to read that as a statement about sensors. Published X_K values for this species in the Mediterranean do vary in a way that correlates with the forcing they were calibrated against: 2.1 μg L^{-1} against in-situ chlorophyll (Monaco and McQuaid, 2018; Sarà et al., 2012), 0.56–0.72 for Greek Aegean sites under a different base parameterisation (Hatzonikolakis et al., 2017), and 0.48–0.60 at typical satellite levels from a power-function formulation (Stamataki et al., 2020). Ours is lower still. Palmer et al. (2020) met the same problem in a *Crassostrea gigas* application and separated satellite bias correction from X_K calibration, explicitly to preserve the parameter's biological meaning.

That reading requires the two chlorophyll scales to be related by a constant, and the three findings reported above weigh against it, the last of them making the choice between those descriptions interpretive rather than mathematical. A power-law transform, which would have restored a genuine scale conversion, drives its exponent to the edge of the solvable range and scores 16.7 mm against 11.3 mm for calibrating X_K directly. The outcome is therefore better described as an empirical calibration against a particular forcing product than as the recovery of a biological constant, and the recommendation that follows is correspondingly narrow: X_K should be reported together with the chlorophyll product, processing level and spatial resolution against which it was calibrated. The present dataset is sufficient to propose this convention but not to establish it.

Confounded processes and regional limits

A food parameter fitted to growth outcomes is a sink for every process the model does not represent, and three are present here, all of which drive the fitted X_K in the same direction and so cannot be separated from one another. Salinity is the largest: no published *M. galloprovincialis* DEB parameterisation carries a salinity term, and none was added, so the depressed filtration and growth of the brackish Black Sea and Marmara sites (Çelik et al., 2015; Lavaud et al., 2017; Maar et al., 2015) has nowhere to go but into a lower X_K, which the model then reads as greater food sensitivity. Multipliers proposed for *M. edulis* (Maar et al., 2015) and *C. virginica* (Lavaud et al., 2017) have not been transferred to this species, and the salinities reported by our sources are period summaries, enough to test whether salinity defines useful calibration groups, which it does not (Table S10), but not enough to fit a mechanism. Food quality is likewise represented by chlorophyll alone, although ingestion depends on total seston and on the organic fraction filtered (Alunno-Bruscia et al., 2011; Filgueira et al., 2011). Culture system and hydrodynamics are unmodelled: Yıldız and Lök (2005) report roughly two-fold differences between rope and net systems at one site, which the formulation cannot resolve without variables describing food delivery at the rope scale.

Local enrichment is invisible to the forcing. Güllük (Yıldırım and Ercan, 2024) provides the one test the Turkish cohort permits, pairing an offshore fish-farm station with a coastal control that fall in one satellite pixel and so receive identical forcing; no systematic difference in residual appears there, whereas the Pašman cohorts differ by 5.13 mm under the same identical forcing. The limitation is not that enrichment invariably biases the calibration, but that the forcing provides no mechanism by which it could enter at all.

The four systems should not be read as a single Mediterranean water body, and the parameter set transferred into them was derived in none of them. The Dardanelles is a two-layer strait, so a mussel at culture depth may experience either layer, a distinction the single-layer forcing cannot make; the Marmara is brackish and semi-enclosed, and has since experienced severe mucilage events whose effect on both physiology and ocean-colour retrieval would be substantial, though the observations used here predate the 2021 event; and the Black Sea is the most stratified and least saline, with a distinctive seasonal plankton cycle (Oguz and Ediger, 2006). The calibration groups are therefore statistical units rather than hydrographic claims, and a parameter set estimated for Mediterranean populations carries an unquantified transfer error into brackish systems, confounded with everything else the fitted X_K absorbs.

The thermal sets are indistinguishable here because summer SST at culture depth reaches 27–28 °C only briefly, and over the range where most growth occurs the two Arrhenius-plus-bounds sets differ by about a quarter in the temperature correction. The three sets diverge far more at the cold end, by a factor of 5.9 at 5 °C, where the single-T_A Van der Veer form predicts much the higher rate; none of this propagates into shell length after recalibration. We retain Monaco and McQuaid (2018) on biological rather than statistical grounds, its upper thermal bound being the only published value with operational headroom for Türkiye's summer SST, whereas the single-T_A Van der Veer form (Van der Veer et al., 2006) has no upper bound at all. Warmer climate scenarios (Montalto et al., 2015; Montalto et al., 2016) would stress these parameters far harder.

Scope, limits and outlook

The calibrated model reproduces the magnitude of growth across the cohort, and that is the extent of the claim made for it. It does not reproduce the differences between sites, so no claim is made regarding site selection or harvest timing. The defensible applications are scenario-level: comparing growth under different deployment months, initial sizes or durations at a site whose forcing is trustworthy, and bounding expected size at harvest with an interval attached. The sub-kilometre discrimination that farm siting requires lies beyond what a 1 km ocean-colour product can support, as the Pašman cohorts show directly. The $L_{\max}(f)$ relation is a property of the DEB algebra rather than of this parameterisation, and being temperature-independent it explains why the thermal

parameters account for so little of the fit; but it does not support use as a pre-flight screen, flagging one of 43 cohorts at the calibrated X_K with an association with model error indistinguishable from zero. It remains transportable to other satellite-forced bivalve applications (Lavaud et al., 2021; Palmer et al., 2020; Rosland et al., 2009; Stechele et al., 2022; Thomas et al., 2011) as an explanation of why food dominates, though not as a decision rule.

The two sites that resist it do so for different reasons. At Mersin Bay, an enclosed Case-2 bay, the chlorophyll retrieval fails before the model receives it, so improved optics, whether regional algorithms, Sentinel-3 OLCI at 300 m or 100 m near-real-time products (Cherif et al., 2021; Gohin et al., 2020), rather than further calibration represent the path forward. Erdek is the more instructive case: a model forced with a gap-filled Level-4 product will always return a number, and nothing in the output distinguishes a value measured at the deployment from one carried in from 70 km away. Any application of this kind should report, per site, the extraction geometry used and the number of valid pixels behind it.

The limitations follow in decreasing order of their likely effect. Salinity, food quality and culture system are unmodelled and mutually confounded, so the calibration cannot be attributed to any one of them. The design is small and strongly clustered, which is why intervals rather than point estimates are given throughout. The satellite-to-in-situ comparison is not matched: no source reports a date-stamped in-situ value, half are seasonal extremes rather than means, and none state sampling depth or determination method, so direct matchup validation is not possible from these sources. Coastal pixel selection remains a substantial uncertainty, so the value reported here is conditional on the ±0.10° box. Observational uncertainty is small relative to model error and digitisation error well within the bootstrap interval, which is why residuals are unweighted; spawning is absent from the formulation, which is why validation is on length alone.

CONCLUSION

A published, biologically defensible DEB parameter set for *M. galloprovincialis* under-predicts shell growth at every one of ten grow-out experiments in Türkiye when forced with satellite chlorophyll, and the correction that repairs it is required in the food parameter rather than in the biological or thermal ones. Recalibrating the half-saturation constant alone recovers most of the deficit on experiments the calibration has not seen. That result is robust to the analytical choices tested here except two: the choice of chlorophyll product, and the spatial geometry over which it is extracted.

How the correction should be read is a separate question, and the present evidence does not support a measurement-scale account: the ratio between the two chlorophyll scales varies seasonally where it can be resolved, is undefined at the two sites the product does not observe, and rescaling the half-saturation constant is in any case algebraically the same operation as rescaling the forcing. The corrected parameter is therefore better described as an empirical property of a particular forcing product than as the same physiology viewed through a different sensor, and its residual may absorb salinity, seston quality, culture-system hydrodynamics and retrieval error in proportions this study cannot apportion. Reporting X_K together with the forcing product against which it was calibrated would make such values comparable between studies.

The limits matter as much as the correction. The model reproduces the central tendency of the cohort but not the differences between sites, and sub-kilometre siting decisions lie outside what a 1 km ocean-colour product can support. For Turkish mussel aquaculture, satellite-forced DEB modelling after calibration reproduces the magnitude of growth across the cohort, which is a precondition for scenario-level use rather than a demonstration of it: no scenario application is tested here, and the approach is not yet a tool for choosing between candidate sites. Closing that gap will require improved coastal ocean-colour retrievals, in-situ chlorophyll reported with dates, depths and methods, and an explicit salinity formulation for this species.

ACKNOWLEDGEMENTS AND FUNDING

We thank the Copernicus Marine Service for freely available ocean products and the Add-my-Pet community for curating the DEB parameters our gap-filling relied on. We are grateful to the authors of the validation studies, whose thorough reporting enabled an independent quantitative test. No specific funding was received.

AUTHORSHIP CONTRIBUTIONS

Cem Hasşerbetçi conceived the study, implemented the DEB model, performed the calibration, validation and resampling analyses, digitised the grow-out database, curated the environmental forcing, produced all figures and drafted the manuscript. Gökhan Tunçelli advised on Turkish aquaculture practice and production context, reviewed site selection and cohort interpretation, and contributed to the discussion. İdil Can-Tunçelli reviewed and edited the manuscript. All authors approved the final version.

CONFLICT OF INTEREST

The authors declare that there is no known financial or personal conflict that may affect the research reported in this article.

ETHICAL APPROVAL

This article does not contain any human or animal studies performed by the authors. All growth and environmental data were obtained from previously published sources and publicly available remote-sensing products, so no specific ethical approval was necessary.

DECLARATION OF AI USE

AI-assisted tools were used for language editing and manuscript formatting assistance during the preparation of this manuscript.

DATA AVAILABILITY

All growth observations were digitised from the ten primary publications listed in Table S3 and are available in those sources. The environmental forcing is drawn entirely from public Copernicus Marine Service reanalysis and ocean-colour products (E.U. Copernicus Marine Service Information (CMEMS), 2022b, 2022a, 2023b, 2023a), accessible through the Copernicus Marine Data Store.

The model implementation, the forcing extraction, the cohort database, the calibration and validation pipeline and every script that produced the figures and tables are deposited at Zenodo (https://doi.org/10.5281/zenodo.21990907).

REFERENCES

Acarlı, S., Vural, P., & Yıldız, H. (2024). Temporal variability in biochemical composition and morphometric parameters of cultivated Mediterranean mussels (*Mytilus galloprovincialis*) in Çanakkale, Türkiye. *Regional Studies in Marine Science*, *78*, 103759. https://doi.org/10.1016/j.rsma.2024.103759

Alunno-Bruscia, M., Bourlès, Y., Maurer, D., Robert, S., Mazurié, J., Gangnery, A., Goulletquer, P., & Pouvreau, S. (2011). A single bio-energetics growth and reproduction model for the oyster *crassostrea gigas* in six Atlantic ecosystems. *Journal of Sea Research*, *66*(4), 340–348. https://doi.org/10.1016/j.seares.2011.07.008

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.

Cherif, E. K., Mozetič, P., Francé, J., Flander-Putrle, V., Faganeli-Pucer, J., & Vodopivec, M. (2021). Comparison of in-situ chlorophyll-a time series and Sentinel-3 ocean and land color instrument data in Slovenian national waters (Gulf of Trieste, Adriatic Sea). *Water*, *13*(14), 1903. https://doi.org/10.3390/w13141903

Ciavatta, S., Kay, S., Brewin, R.J.W., Cox, R., Di Cicco, A., Nencioli, F., Polimene, L., Sammartino, M., Santoleri, R., Skákala, J., & Tsapakis, M. (2019). Ecoregions in the Mediterranean Sea through the reanalysis of phytoplankton functional types and carbon fluxes. *Journal of Geophysical Research: Oceans*, *124*(10), 6737-6759. https://doi.org/10.1029/2019JC015128

Çelik, M. Y., Karayücel, S., Karayücel, İ., Eyüboğlu, B., & Öztürk, R. (2015). Settlement and growth of the mussels (*Mytilus galloprovincialis*, Lamarck, 1819) on different collectors suspended from an offshore submerged longline system in the Black Sea. *Aquaculture Research*, *47*(12), 3765–3776. https://doi.org/10.1111/are.12827

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

E.U. Copernicus Marine Service Information (CMEMS). (2022a). *Black Sea Physics Reanalysis (BLKSEA_MULTIYEAR_PHY_007_004)*. Marine Data Store (MDS). https://doi.org/10.25423/CMCC/BLKSEA_MULTIYEAR_PHY_007_004

E.U. Copernicus Marine Service Information (CMEMS). (2022b). *Mediterranean Sea Physics Reanalysis (MEDSEA_MULTIYEAR_PHY_006_004)*. Marine Data Store (MDS). https://doi.org/10.25423/CMCC/MEDSEA_MULTIYEAR_PHY_006_004_E3R1

E.U. Copernicus Marine Service Information (CMEMS). (2023a). *Black Sea Ocean Colour Plankton MY L4 daily gap-free product (OCEANCOLOUR_BLK_BGC_L4_MY_009_154)*. Marine Data Store (MDS). https://doi.org/10.48670/moi-00304

E.U. Copernicus Marine Service Information (CMEMS). (2023b). *Mediterranean Sea Ocean Colour Plankton MY L4 daily gap-free product (OCEANCOLOUR_MED_BGC_L4_MY_009_144)*. Marine Data Store (MDS). https://doi.org/10.48670/moi-00300

Filgueira, R., Rosland, R., & Grant, J. (2011). A comparison of scope for growth (SFG) and dynamic energy budget (DEB) models applied to the blue mussel (*Mytilus edulis*). *Journal of Sea Research*, *66*(4), 403–410. https://doi.org/10.1016/j.seares.2011.04.006

Gohin, F. (2011). Annual cycles of chlorophyll-*a*, non-algal suspended particulate matter, and turbidity observed from space and in-situ in coastal waters. *Ocean Science*, *7*(5), 705–732. https://doi.org/10.5194/os-7-705-2011

Gohin, F., Bryère, P., Lefebvre, A., Sauriau, P.-G., Savoye, N., Vantrepotte, V., Bozec, Y., Cariou, T., Conan, P., Coudray, S., Courtay, G., Françoise, S., Goffart, A., Hernández Fariñas, T., Lemoine, M., Piraud, A., Raimbault, P., & Rétho, M. (2020). Satellite and in situ monitoring of Chl-a, turbidity, and total suspended matter in coastal waters: Experience of the year 2017 along the French coasts. *Journal of Marine Science and Engineering*, *8*(9), 665. https://doi.org/10.3390/jmse8090665

Hatzonikolakis, Y., Tsiaras, K., Theodorou, J. A., Petihakis, G., Sofianos, S., & Triantafyllou, G. (2017). Simulation of mussel *Mytilus galloprovincialis* growth with a dynamic energy budget model in Maliakos and Thermaikos Gulfs (Eastern Mediterranean). *Aquaculture Environment Interactions*, *9*, 371–383. https://doi.org/10.3354/aei00236

Karayücel, S., Çelik, M. Y., Karayücel, İ., & Erik, G. (2010). Growth and production of raft cultivated Mediterranean mussel (*Mytilus galloprovincialis* Lamarck, 1819) in Sinop, Black Sea. *Turkish Journal of Fisheries and Aquatic Sciences*, *10*(1), 9-17. https://doi.org/10.4194/trjfas.2010.0102

Karayücel, S., Karayücel, I., Erdem, M., Saygun, S., & Uyan, O. (2003). Growth and production in long-line cultivated Mediterranean mussel (*Mytilus galloprovincialis*) in Sinop, Black Sea. *Israeli Journal of Aquaculture – Bamidgeh*, *55*(3), 169–178. https://doi.org/10.46989/001c.20348

Keskin, İ., & Ekici, A. (2021). Effects of environmental factors and food availability in Northern Aegean Sea on the cultivation of Mediterranean mussels (*Mytilus galloprovincialis*). *Aquaculture Research*, *52*(1), 65–76. https://doi.org/10.1111/are.14870

Keskin, İ., Ekici, A., & Serdar, S. (2020). Determination of the growth performance of *Mytilus galloprovincialis* in nets at Gökçeada Island. *The European Zoological Journal*, *87*(1), 559-570. https://doi.org/10.1080/24750263.2020.1818856

Kooijman, S. A. L. M. (2010). *Dynamic energy budget theory for metabolic organisation* (3rd ed.). Cambridge University Press.

Kooijman, S. A. L. M. (2014). Metabolic acceleration in animal ontogeny: An evolutionary perspective. *Journal of Sea Research*, *94*, 128–137. https://doi.org/10.1016/j.seares.2014.06.005

Lacava, T., Ciancia, E., Di Polito, C., Madonia, A., Pascucci, S., Pergola, N., Piermattei, V., Satriano, V., & Tramutoli, V. (2018). Evaluation of MODIS–Aqua chlorophyll-a algorithms in the Basilicata Ionian coastal waters. *Remote Sensing*, *10*(7), 987. https://doi.org/10.3390/rs10070987

Lavaud, R., La Peyre, M. K., Casas, S. M., Bacher, C., & La Peyre, J. F. (2017). Integrating the effects of salinity on the physiology of the eastern oyster, *Crassostrea virginica*, in the northern Gulf of Mexico through a Dynamic Energy Budget model. *Ecological Modelling*, *363*, 221–233. https://doi.org/10.1016/j.ecolmodel.2017.09.003

Lavaud, R., La Peyre, M. K., Justic, D., & La Peyre, J. F. (2021). Dynamic Energy Budget modelling to predict eastern oyster growth, reproduction, and mortality under river management and climate change scenarios. *Estuarine, Coastal and Shelf Science*, *251*, 107188. https://doi.org/10.1016/j.ecss.2021.107188

Lika, K., Kearney, M. R., Freitas, V., Veer, H. W. van der, Meer, J. van der, Wijsman, J. W. M., Pecquerie, L., & Kooijman, S. A. L. M. (2011). The "covariation method" for estimating the parameters of the standard Dynamic Energy Budget model I: Philosophy and approach. *Journal of Sea Research*, *66*(4), 270-277. https://doi.org/10.1016/j.seares.2011.07.010

Lök, A., Acarlı, S., Serdar, S., Köse, A., & Yıldız, H. (2007). Growth and mortality of Mediterranean mussel *Mytilus galloprovincialis* Lam., 1819, in relation to size on longline in Mersin Bay, Izmir (Turkey – Aegean Sea). *Aquaculture Research*, *38*(8), 819–826. https://doi.org/10.1111/j.1365-2109.2007.01717.x

Maar, M., Saurel, C., Landes, A., Dolmer, P., & Petersen, J. K. (2015). Growth potential of blue mussels (*m. Edulis*) exposed to different salinities evaluated by a Dynamic Energy Budget model. *Journal of Marine Systems*, *148*, 48–55. https://doi.org/10.1016/j.jmarsys.2015.02.003

Marques, G. M., Augustine, S., Lika, K., Pecquerie, L., Domingos, T., & Kooijman, S. A. L. M. (2018). The AmP project: Comparing species on the basis of dynamic energy budget parameters. *PLOS Computational Biology*, *14*(5), e1006100. https://doi.org/10.1371/journal.pcbi.1006100

Monaco, C. J., & McQuaid, C. D. (2018). Applicability of Dynamic Energy Budget (DEB) models across steep environmental gradients. *Scientific Reports*, *8*(1), 16384. https://doi.org/10.1038/s41598-018-34786-w

Monaco, C. J., Porporato, E. M. D., Lathlean, J. A., Tagliarolo, M., Sarà, G., & McQuaid, C. D. (2019). Predicting the performance of cosmopolitan species: Dynamic energy budget model skill drops across large spatial scales. *Marine Biology*, *166*(2), 14. https://doi.org/10.1007/s00227-018-3462-4

Montalto, V., Helmuth, B., Ruti, P. M., Dell'Aquila, A., Rinaldi, A., & Sarà, G. (2016). A mechanistic approach reveals non-linear effects of climate warming on mussels throughout the Mediterranean Sea. *Climatic Change*, *139*(2), 293–306. https://doi.org/10.1007/s10584-016-1780-4

Montalto, V., Rinaldi, A., & Sarà, G. (2015). Life history traits to predict biogeographic species distributions in bivalves. *The Science of Nature*, *102*(9–10), 61. https://doi.org/10.1007/s00114-015-1313-4

Montalto, V., Rinaldi, A., Sarà, G., Kooijman, S. A. L. M., Monaco, C. J., & Augustine, S. (2018). *AmP Mytilus galloprovincialis, version 2018/09/17*. Add-my-Pet database. https://www.bio.vu.nl/thb/deb/deblab/add_my_pet/

Oguz, T., & Ediger, D. (2006). Comparison of in situ and satellite-derived chlorophyll pigment concentrations, and impact of phytoplankton bloom on the suboxic layer structure in the western Black Sea during May–June 2001. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *53*(17–19), 1923–1933. https://doi.org/10.1016/j.dsr2.2006.07.001

Palmer, S. C. J., Gernez, P. M., Thomas, Y., Simis, S., Miller, P. I., Glize, P., & Barillé, L. (2020). Remote sensing-driven Pacific oyster (*Crassostrea gigas*) growth modeling to inform offshore aquaculture site selection. *Frontiers in Marine Science*, *6*. https://doi.org/10.3389/fmars.2019.00802

Peharda, M., Župan, I., Bavčević, L., Frankić, A., & Klanjšček, T. (2007). Growth and condition index of mussel *Mytilus galloprovincialis* in experimental integrated aquaculture. *Aquaculture Research*, *38*(16), 1714–1720. https://doi.org/10.1111/j.1365-2109.2007.01840.x

Rosland, R., Strand, Ø., Alunno-Bruscia, M., Bacher, C., & Strohmeier, T. (2009). Applying Dynamic Energy Budget (DEB) theory to simulate growth and bio-energetics of blue mussels under low seston conditions. *Journal of Sea Research*, *62*(2-3), 49-61. https://doi.org/10.1016/j.seares.2009.02.007

Sarà, G., Kearney, M., & Helmuth, B. (2011). Combining heat-transfer and energy budget models to predict thermal stress in Mediterranean intertidal mussels. *Chemistry and Ecology*, *27*(2), 135-145. https://doi.org/10.1080/02757540.2011.552227

Sarà, G., Manganaro, A., Cortese, G., Pusceddu, A., & Mazzola, A. (1998). The relationship between food availability and growth in *Mytilus galloprovincialis* in the open sea (southern Mediterranean). *Aquaculture*, *167*, 1–15. https://doi.org/10.1016/S0044-8486(98)00281-6

Sarà, G., Reid, G. K., Rinaldi, A., Palmeri, V., Troell, M., & Kooijman, S. A. L. M. (2012). Growth and reproductive simulation of candidate shellfish species at fish cages in the Southern Mediterranean: Dynamic Energy Budget (DEB) modelling for integrated multi-trophic aquaculture. *Aquaculture*, *324-325*, 259-266. https://doi.org/10.1016/j.aquaculture.2011.10.042

Şensurat-Genç, T., Karadal, O., Oruç, A. Ç., Özgül, A., & Lök, A. (2021). Effects of seasonal changes and depth on growth parameters of the Mediterranean mussel (*Mytilus galloprovincialis*) on a shipwreck in the Eastern Mediterranean Sea. *Oceanological and Hydrobiological Studies*, *50*(4), 441–454. https://doi.org/10.2478/oandhs-2021-0038

Stamataki, N., Hatzonikolakis, Y., Tsiaras, K., Tsangaris, C., Petihakis, G., Sofianos, S., & Triantafyllou, G. (2020). Modelling mussel (*Mytilus* spp.) microplastic accumulation. *Ocean Science*, *16*(4), 927–949. https://doi.org/10.5194/os-16-927-2020

Stechele, B., Maar, M., Wijsman, J., Van der Zande, D., Degraer, S., Bossier, P., & Nevejan, N. (2022). Comparing life history traits and tolerance to changing environments of two oyster species (*Ostrea edulis* and *Crassostrea gigas*) through Dynamic Energy Budget theory. *Conservation Physiology*, *10*(1). https://doi.org/10.1093/conphys/coac034

Teruzzi, A., Bolzon, G., Salon, S., Lazzari, P., Solidoro, C., & Cossarini, G. (2018). Assimilation of coastal and open sea biogeochemical data to improve phytoplankton simulation in the Mediterranean Sea. *Ocean Modelling*, *132*, 46–60. https://doi.org/10.1016/j.ocemod.2018.09.007

Thomas, Y., Mazurié, J., Alunno-Bruscia, M., Bacher, C., Bouget, J.-F., Gohin, F., Pouvreau, S., & Struski, C. (2011). Modelling spatio-temporal variability of *Mytilus edulis* (L.) Growth by forcing a Dynamic Energy Budget model with satellite-derived environmental data. *Journal of Sea Research*, *66*(4), 308–317. https://doi.org/10.1016/j.seares.2011.04.015

Van der Veer, H. W., Cardoso, J. F. M. F., & Meer, J. van der. (2006). The estimation of DEB parameters for various Northeast Atlantic bivalve species. *Journal of Sea Research*, *56*(2), 107-124. https://doi.org/10.1016/j.seares.2006.03.005

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83. https://doi.org/10.2307/3001968

Yildirim, Ş., Erdem, Ö. A., Sabanci, F., Gökvardar, A., Serdar, S., & Ertan, A. (2024). Effects of different culture methods on growth, meat yield, and nutritional composition of Mediterranean mussels (*Mytilus galloprovincialis*), south side of Marmara Sea, Türkiye. *The European Zoological Journal*, *91*(2), 869-889. https://doi.org/10.1080/24750263.2024.2378829

Yıldırım, P., & Ercan, E. (2024). Effect of fish farming on bioecological characteristics of the Mediterranean mussels *Mytilus galloprovincialis* Lamarck, 1819 (Bivalvia: Mytilidae): Güllük Bay example, south Aegean Sea, Turkey. *Acta Zoologica Bulgarica*, *76*(2), 277–288. https://doi.org/10.71424/azb76.2.002595

Yıldız, H., & Lök, A. (2005). Growth and survival rates of different size classes of *Mytilus galloprovincialis* Lamarck, 1819 at two culture systems in Dardanelles. *Ege Journal of Fisheries & Aquatic Sciences*, *22*(1–2), 69–74.

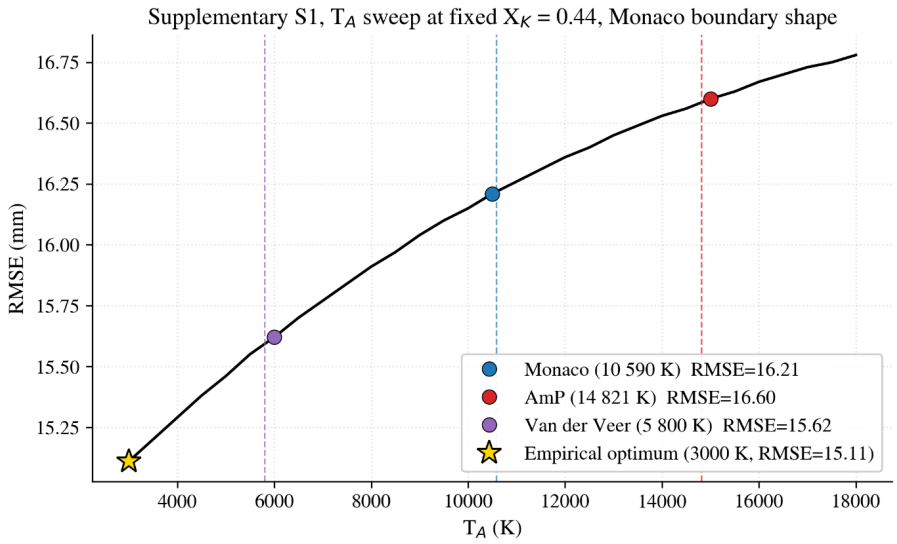

Figure S1. [Sensitivity analysis] T_A-only sweep at fixed X_K=0.456 µg L^{-1} (the ratio-corrected value as first derived; the ratio median was later revised to give 0.523 µg L^{-1}, which does not affect the shape of the response shown here), showing monotonic but near-flat dependence of RMSE on T_A across the literature range [3000,18000] K: RMSE rises monotonically from 15.11 to 16.78 mm, a span of 1.7 mm, which is why the food parameter rather than T_A is the one recalibrated. Cited in Results, Sensitivity and robustness.

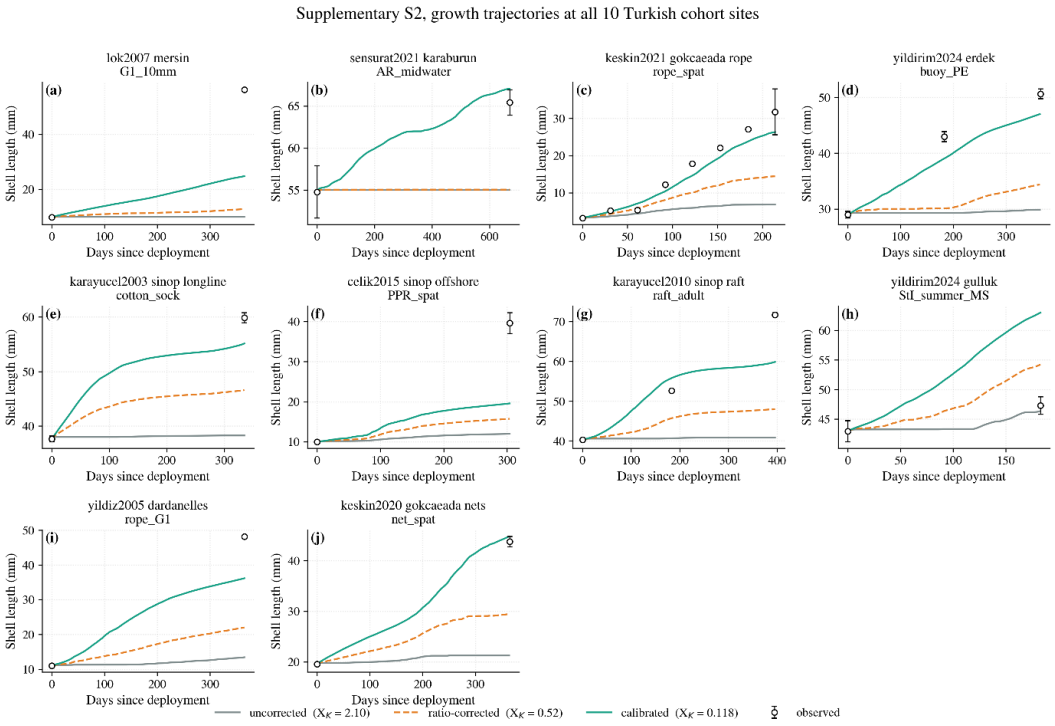

Figure S2. [Model output] Observed-versus-predicted growth trajectories at all ten cohort sites, one representative group each, under the three correction schemes (Monaco thermal set). Cited in Results, Food-scale correction and calibration.

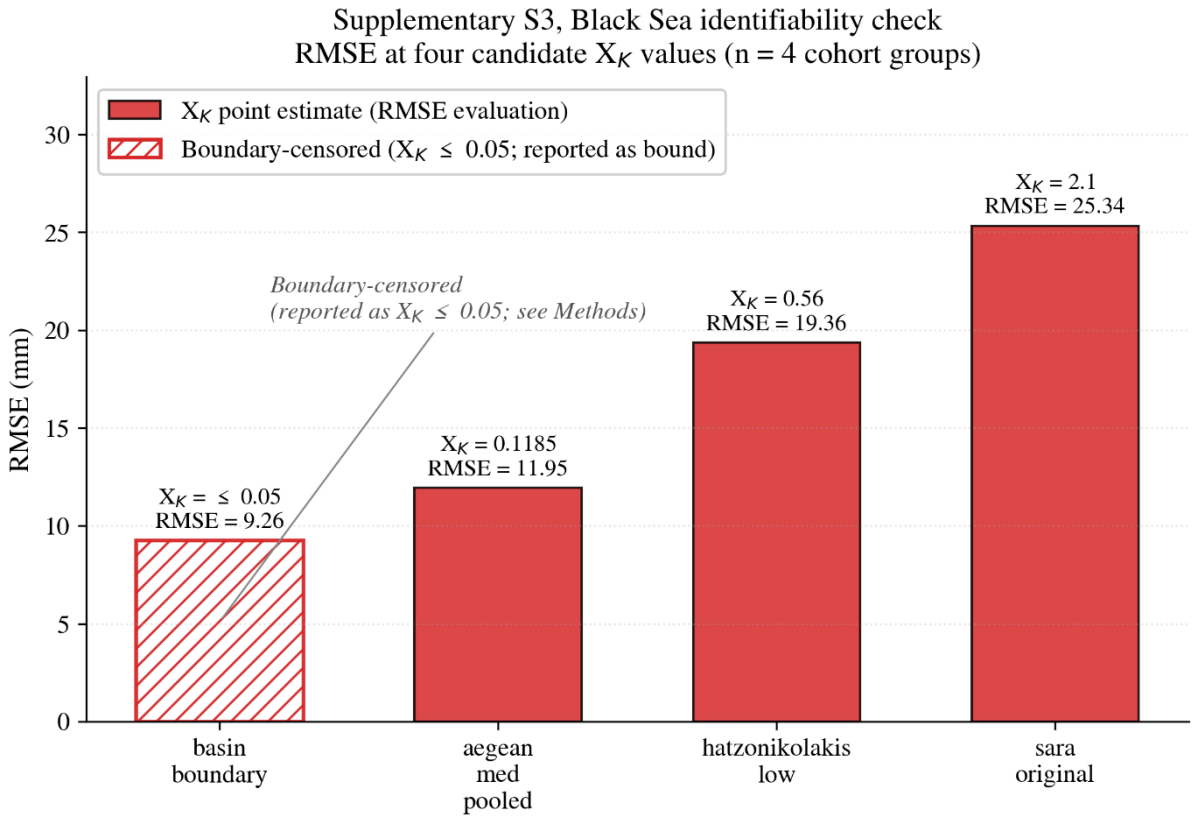

Figure S3. [Calibration] Black Sea X_K identifiability. RMSE versus X_K over [0.01,2.5] µg L^{-1}, showing that the optimum is censored at the grid floor and that fits below ~0.05 are indistinguishable: the data bound this parameter from above but do not estimate it. Cited in Results, Food-scale correction and calibration.

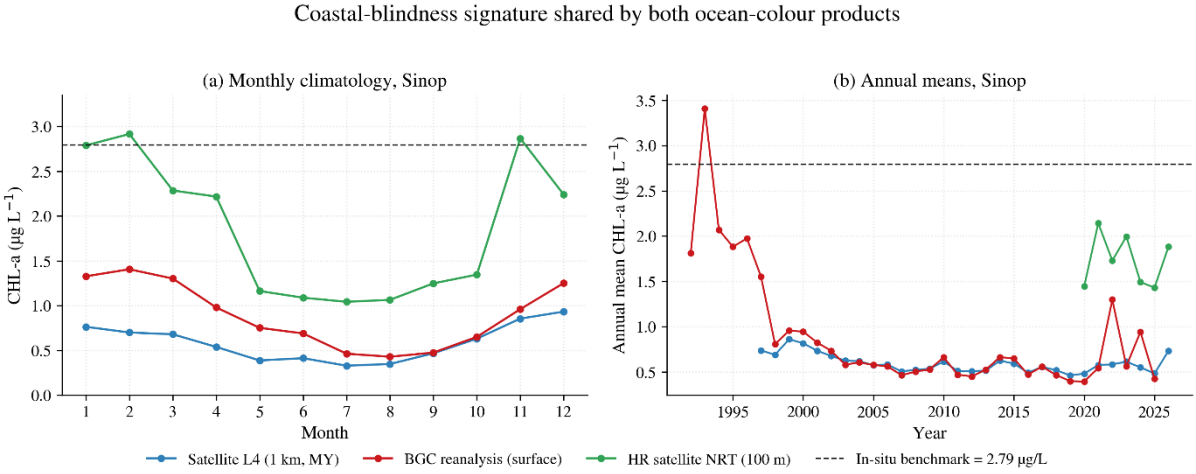

Figure S4. [Input data] CHL-a product comparison at Sinop (Black Sea): satellite L4 gap-free versus biogeochemical reanalysis, with in-situ literature values where available. Note that agreement between the two products at the uncalibrated parameter does not imply equivalence: once each is calibrated the reanalysis fits better (Results, Sensitivity and robustness). Cited in Methods, Environmental forcing.

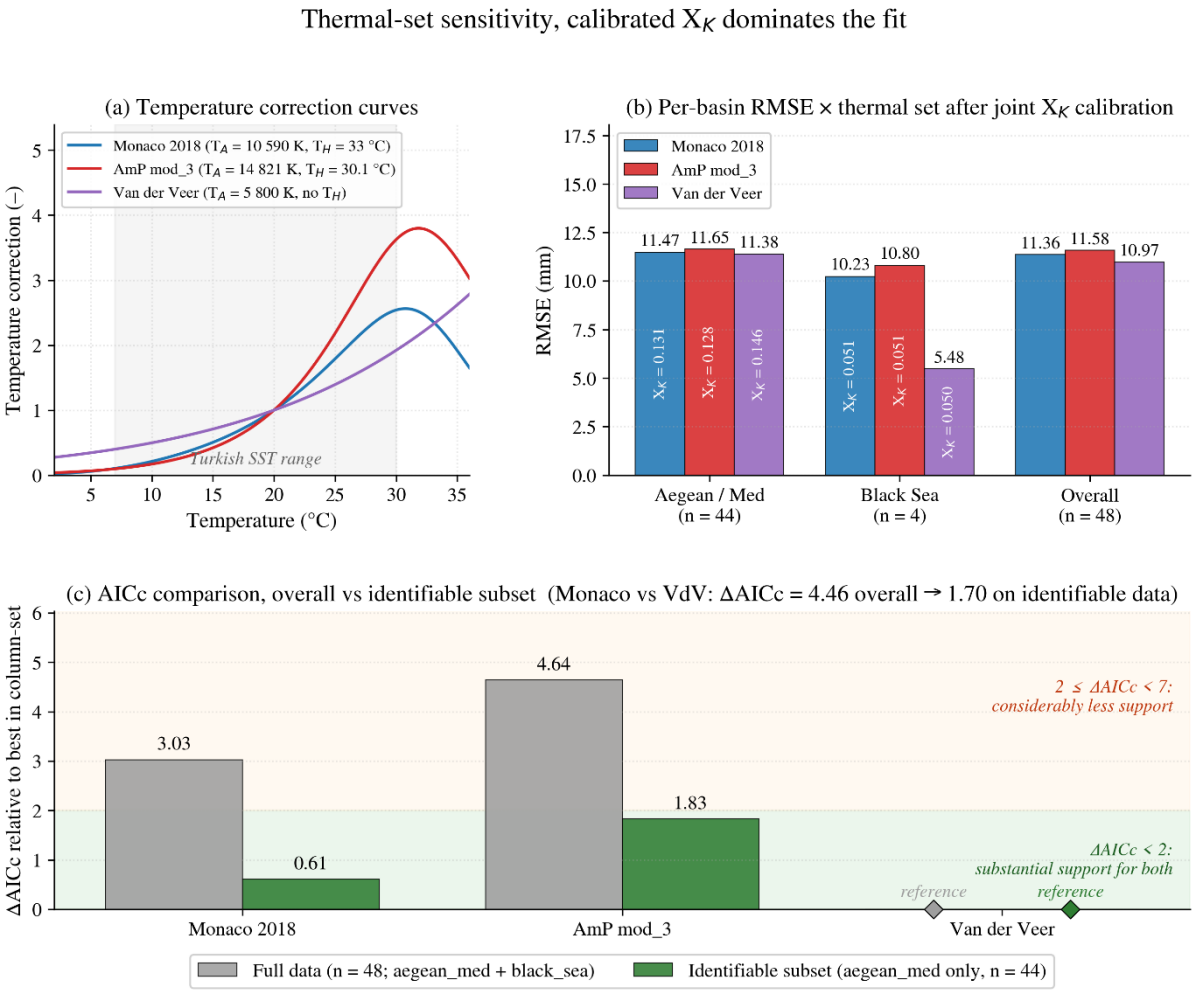

Figure S5. [Sensitivity analysis] Thermal-set sensitivity under joint recalibration of X_K. Left: calibrated X_K for Monaco 2018, AmP mod_3 and Van der Veer 2006, with overlapping bootstrap CIs for aegean_med. Right: AICc with the ΔAICc<2 band shaded. Note that the AICc plotted here is computed at the nominal n, which over-discriminates because the cohort groups are clustered (ICC = 0.78): the shaded band separates the sets, whereas at the effective sample size n_{eff}=12.0 all three fall inside it. Table S1 gives ΔAICc at both sample sizes, the Support verdict read at n_{eff}, and the full numerical results. Cited in Results, Sensitivity and robustness.

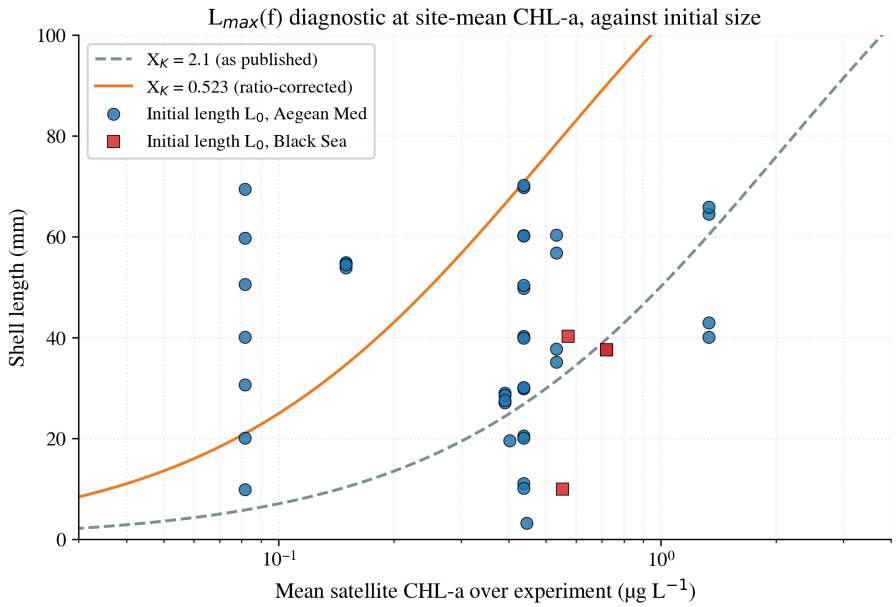

Figure S6. [Sensitivity analysis] L_{max}(f) diagnostic evaluated at site-mean satellite CHL-a, versus observed initial size, with contours at X_K=2.1 (dashed) and at the ratio-corrected X_K=0.523 µg L^{-1} (solid). Sites below a contour have L_{max}(f)<L_0 at that X_K. This is the site-mean form of the diagnostic and is shown for comparison only: it overestimates L_{max} by 9–17% relative to the daily calculation (Jensen's inequality), and the daily calculation in turn shows that cohorts below the threshold still grow, so the condition marks negligible predicted growth rather than blocked growth. See Results, Performance of the published parameter set. Cited in Methods, DEB model and parameters.

Table S1. [Sensitivity analysis] Thermal-set sensitivity under joint recalibration of X_K, every statistic at both the cohort-group and the experiment scale. X_K in µg CHL-a L^{-1}, RMSE in mm. At the effective sample size the three thermal sets are not distinguishable on this cohort. Notes below; cited in Results, Sensitivity and robustness.

	Van der Veer et al. (2006)	Monaco and McQuaid (2018)	Montalto et al. (2018)
T_A (K)	5800	10590	14821
X_K, aegean_med	0.146	0.131	0.128
X_K, black_sea (boundary-censored)	0.050	0.051	0.051
RMSE, aegean_med	11.38	11.47	11.65
RMSE, black_sea	5.48	10.23	10.80
RMSE, 43 cohort groups	10.97	11.36	11.58
RMSE, 10 experiments	9.28	10.76	11.00
ΔAICc, nominal n = 43	0.00	3.03	4.64
ΔAICc, n_{eff} = 12.0	0.00	0.85	1.30
ΔAICc, subset, nominal n = 39	0.00	0.61	1.83
ΔAICc, subset, n_{eff} = 8.5	0.00	0.14	0.40
Support at n_{eff}	ref.	substantial	substantial

Notes: The two RMSE rows are one estimator at two units of replication: the 43-group figure pools all cohort residuals, so an experiment counts once for each group it contributes; the 10-experiment figure counts each experiment once. Cohort groups are clustered within experiments, so ΔAICc is given at the nominal n and at the effective n_{eff} = 12.0, and the Support row is read at n_{eff} (Supplementary Methods S3). The subset rows drop the Black Sea, boundary-censored on either grid (Figure S3), leaving aegean_med as the only calibration group (k = 1). Every X_K is a per-group value: the paper's primary single X_K = 0.118 is not a column here, and the Monaco aegean_med optimum of 0.131 is the 0.130 of Table 1 at this grid resolution. All values from scripts/d5_cluster_inference.py.

Table S2. [Model parameters] DEB parameter set used in this study (post-metamorphosis abj model; operative rates include s_M = 10.37). All rates at T_{ref} = 293.15 K. The Source column gives the provenance of every value: the Monaco and McQuaid (2018) covariation set is primary, four parameters are gap-filled from the Add-my-Pet entry whose estimates share authorship and estimation framework with that set, δ_M is the morphometric estimate of Sarà et al. (2011) (0.23±0.04), and the remainder are derived. X_K is the only parameter this study recalibrates. The value shown is the published starting point, which Sarà et al. (2012) calibrated against in-situ CHL-a; here it is recalibrated against observed final shell length across the 43 cohort groups, under satellite CHL-a forcing, by minimising mean absolute relative error, giving X_K = 0.118 µg L^{-1} (Results). Rationale in Methods, DEB model and parameters.

Symbol	Value	Unit	Source
T_{ref}	293.15	K	convention
T_A	10590	K	Monaco and McQuaid (2018)
T_L	279.6	K	Monaco and McQuaid (2018)
T_H	306.1	K	Monaco and McQuaid (2018)
T_{AL}	22670	K	Monaco and McQuaid (2018)
T_{AH}	34540	K	Monaco and McQuaid (2018)
κ	0.47	–	Monaco and McQuaid (2018)
$[\dot{p}_M]$	10.27	J d^{-1} cm^{-3}	Monaco and McQuaid (2018)
$[E_G]$	3156	J cm^{-3}	Monaco and McQuaid (2018)
$\{\dot{p}_{Xm}\}$	9.42	J d^{-1} cm^{-2}	Monaco and McQuaid (2018)
a_e	0.80	–	Monaco and McQuaid (2018)
$\{\dot{p}_{Am}\}$	7.536	J d^{-1} cm^{-2}	derived: $a_e\{\dot{p}_{Xm}\}$
$\dot{v}$	0.023785	cm d^{-1}	Montalto et al. (2018), gap-fill
s_M	10.37	–	Monaco and McQuaid (2018)
$\{\dot{p}_{Am}\}_{eff}$	78.15	J d^{-1} cm^{-2}	derived: $s_M\{\dot{p}_{Am}\}$
$\dot{v}_{eff}$	0.2466	cm d^{-1}	derived: $s_M\dot{v}$
$[E_m]$	316.9	J cm^{-3}	derived: $\{\dot{p}_{Am}\}/\dot{v}$
δ_M	0.23	–	Sarà et al. (2011)
X_K	2.1	µg CHL-a L^{-1}	Sarà et al. (2012); Monaco and McQuaid (2018), recalibrated here
κ_R	0.95	–	Montalto et al. (2018), gap-fill
k_J	0.002	d^{-1}	Montalto et al. (2018), gap-fill
E_{Hp}	1587	J	Montalto et al. (2018), gap-fill
d_V	0.09	g cm^{-3}	Monaco and McQuaid (2018)
ρ_E	5.71×10^{-5}	g J^{-1}	Monaco and McQuaid (2018)

Table S3. [Input data] Validation cohort: ten published M. galloprovincialis growth experiments at sites in the Aegean, the Dardanelles, the Marmara and the Black Sea, totalling 43 cohort groups. The Code column gives the short identifier used in the results tables. †Cross-sectional monthly sampling of an existing farm (A24-C); used only as a plausibility range check and excluded from quantitative fits. ‡Gökçeada Island deployments at the same location in two different years with different suspension gear (nets in 2017–2018, rope in 2018). Cited in Methods, Study area and cohort data.

Code	Source	Region	Lat / Lon	Depth (m)	Period	Dur (mo)	*n*
L07-M	Lök et al. (2007)	Aegean	38.21 / 26.42	1	2001–2002	12	7
S21-K	Şensurat-Genç et al. (2021)	Aegean	38.65 / 26.51	25, 36	2018–2020	22	4
K20-GN	Keskin et al. (2020)‡	Gökçeada	40.23 / 25.89	–	2017–2018	12	1
K21-GR	Keskin and Ekici (2021)‡	Gökçeada	40.23 / 25.89	2.5	2018	8	1
Y05-D	Yıldız and Lök (2005)	Dardanelles	40.12 / 26.35	–	2002–2003	12	14
Y24-E	Yildirim et al. (2024)	Marmara	40.43 / 27.75	–	2016–2017	12	4
Y24-G	Yıldırım and Ercan (2024)	S. Aegean	37.26 / 27.45	8, 51	2013–2014	12	8
K03-SL	Karayücel et al. (2003)	Black Sea	42.03 / 35.15	13	1997–1998	11	2
K10-SR	Karayücel et al. (2010)	Black Sea	42.03 / 35.15	13	2005–2006	13	1
C15-SO	Çelik et al. (2015)	Black Sea	42.03 / 35.15	25–27	2008–2009	10	1
A24-C†	Acarlı et al. (2024)	Dardanelles	40.28 / 26.50	–	2019–2020	12	1

Where the growth values come from. Read from published tables: S21-K, Y05-D, A24-C. Read from a digitised figure: L07-M, K03-SL, K10-SR, K21-GR, Y24-E. Read from a figure whose values the source also states in prose: K20-GN. Read from a table and cross-checked against a figure: C15-SO, Y24-G. Six of the ten quantitatively fitted experiments therefore draw their growth values from a figure. Digitisation method, cross-checking and the propagation of digitisation error are in Supplementary Methods S1.

Table S4. [Input data] Per-experiment satellite-period mean CHL-a (X^{sat}), literature in-situ CHL-a ($X^{in\text{-}situ}$), ratio $\alpha = X^{sat}/X^{in\text{-}situ}$, food-limited maximum length $L_{max} = 155.5\ X^{sat}/(X^{sat} + X_K)$ at $X_K = 2.1$ µg L^{-1}, computed from the X^{sat} column of this table, and the observed initial length L_0 (range across the experiment's cohort groups, to the nearest mm). CHL-a in µg L^{-1}; lengths in mm; codes as in Table S3. The in-situ column is not homogeneous, and the Basis column states what each value is. Most entries are annual means as reported; the Y24-E value is the annual mean of the digitised monthly series (Methods, Study area and cohort data), not the midpoint of the reported annual range; A24-C is a single February observation. The median α = 0.249 is taken over the eight quantitatively fitted experiments that report in-situ CHL-a and have satellite coverage at the site. A24-C (cross-sectional, plausibility check only) is excluded from the median, and S21-K reports no in-situ CHL-a. Y24-E is excluded outright: correcting its in-situ value withdrew the claim that Erdek was a refractory site, and the later finding that the product does not observe the site withdraws its α altogether, moving the median from 0.217 to 0.249. Cited in Methods, Correction schemes and X_K calibration, and Results, Food-scale correction and calibration

Code	X^{sat}	$X^{in\text{-}situ}$	α	L_{max}	L_0	Basis of $X^{in\text{-}situ}$
L07-M	0.086	3.880	0.022	6.1	10–70	annual mean
K10-SR	0.561	3.070	0.183	32.8	40	annual mean
K03-SL	0.703	3.520	0.200	39.0	38	annual mean (log scale)
Y05-D	0.448	2.140	0.209	27.3	10–70	annual mean
C15-SO	0.517	1.790	0.289	30.7	10	annual mean
Y24-G	0.844	0.997	0.847	44.6	35–66	mean of 4 station × season means
K21-GR	0.424	0.320	1.326	26.1	3	multi-month mean
K20-GN	0.385	0.210	1.834	24.1	20	annual mean
Y24-E*	–	1.843	–	–	27–29	no valid satellite pixel at the site
A24-C†	0.33	2.19	–	21.1	53–69	single February observation
S21-K‡	0.143	–	–	9.9	~54	not reported

Median α over the eight experiments L07-M to K20-GN: 0.249, giving $X_K^{sat} = 0.523$. The last three rows are reported but excluded from the median. *Erdek. The in-situ value is the annual mean of the digitised monthly series, not the midpoint of the reported annual range. No α is quoted because the Mediterranean ocean-colour product returns no valid retrieval within 50 km of this site: the series used to force the model is a 1° box mean whose 226 valid pixels all lie 68–85 km west, in the north Aegean beyond the Dardanelles (Methods). A ratio between Marmara in-situ chlorophyll and north Aegean satellite chlorophyll is not a satellite-to-in-situ ratio for any site, so Erdek cannot enter the median. † Cross-sectional; excluded from the median, and no α is quoted because a February observation has no annual-mean counterpart. ‡Karaburun; no in-situ CHL-a reported.

Table S5. [Input data] Satellite CHL-a extraction metadata per experiment, for the forcing actually used in this study. Extraction is a box mean centred on the deployment, half-width 0.1° by default, widened in steps only where the default returns no valid retrieval on any day (Methods, Environmental forcing). "Valid days" counts days on which at least one pixel in the box carried a retrieval; "valid pixels" is the number contributing to the box mean. The last two rows are the two experiments whose forcing is not spatially matched to the site; together they supply 18 of the 43 cohort groups. Codes as in Table S3. Cited in Methods, Environmental forcing, and Results, Sensitivity and robustness.

Code	Lat / Lon	Buffer (°)	Valid days	Valid pixels	Mean CHL-a	Coverage at the site
L07-M	38.21 / 26.42	0.10	422	150	0.086	local
S21-K	38.65 / 26.51	0.10	731	177	0.143	local
K20-GN	40.23 / 25.89	0.10	395	150	0.385	local
K21-GR	40.23 / 25.89	0.10	275	167	0.424	local
Y24-G	37.26 / 27.45	0.10	426	202	0.844	local (enclosed bay; see note)
K03-SL	42.03 / 35.15	0.10	275	218	0.760	local
K10-SR	42.03 / 35.15	0.10	457	230	0.561	local
C15-SO	42.03 / 35.15	0.10	365	230	0.517	local
Y05-D	40.12 / 26.35	0.10	426	3	0.448	3 pixels, ~12 km NW at the strait mouth
Y24-E	40.43 / 27.75	1.00	395	226	0.400	none within 0.5°; 68–85 km W, north Aegean

"Mean CHL-a" here is the mean over the valid retrieval days counted in the previous column, before any gap filling, so it describes what the product delivered. It is the same quantity as X^{sat} in Table S4 at every experiment except K03-SL, where the 107 days assigned the September climatology (Supplementary Methods S2) are part of the series the model was actually forced with: 0.760 over the 275 observed days becomes 0.703 over all 382. Y24-E is likewise 0.400 here because that is what the 1° box returns, whereas Table S4 quotes no X^{sat} for it, the box not observing the site.

Güllük (Y24-G) is flagged because it is the site that drives the extraction-geometry sensitivity reported in Results, *Sensitivity and robustness*: its box-mean CHL-a falls from 1.00 to 0.19 µg L^{-1} as the half-width widens from 0.05° to 0.50°, as an enclosed bay's signal is diluted by surrounding open water. Its coverage at the default buffer is nonetheless genuinely local.

Table S6. [Model output] $L_{max}(f)$ diagnostic at three values of X_K: cohort groups for which $L_{max}(f) < L_0$ when f is evaluated at the site-mean chlorophyll, and, recomputed from daily forcing, the mean fraction of deployment days on which the same condition holds. The count collapses as X_K falls, and at the calibrated value the condition is essentially never met, which is why the diagnostic is reported as an explanation of why food dominates rather than as a screening rule. Flagged cohorts are not static: transient blooms lift f above the threshold, and their median simulated growth is 0.27 mm against a median observed growth of 13.1 mm, so predicted growth is negligible, about 2% of observation, rather than absent. The association with model error is also of the wrong sign for a screen, and two cohorts never flagged on any day still under-predict by 87% and 93%. Cited in Results, Performance of the published parameter set.

X_K (µg L^{-1})	Cohorts flagged (site-mean *f*)	Mean fraction of days below threshold (daily)
2.1 (as published)	32 / 43	0.75
0.523 (ratio-corrected)	9 / 43	0.37
0.118 (calibrated)	1 / 43	0.04

Table S7. [Model output] Cross-validated performance by experiment at the calibrated single X_K, ordered by RMSE. The overall bias of −0.30 mm is produced by cancellation between these rows, not by accuracy at any of them: the mean absolute per-experiment bias is 9.55 mm, four experiments over-predict and six under-predict. Five of the ten experiments contribute one or two cohort groups while one contributes fourteen, which is the clustering the effective sample size accounts for. Codes and sources as in Table S3. Cited in Results, Distribution of error, and generalisation.

Experiment	*n* groups	RMSE (mm)	Bias (mm)
L07-M (Lök et al., 2007)	7	23.14	−21.98
C15-SO (Çelik et al., 2015)	1	20.03	−20.03
K10-SR (Karayücel et al., 2010)	1	11.85	−11.85
Y24-G (Yıldırım and Ercan, 2024)	8	10.34	+9.33
Y05-D (Yıldız and Lök, 2005)	14	9.69	+7.77
S21-K (Şensurat-Genç et al., 2021)	4	7.32	+7.04
Y24-E (Yildirim et al., 2024)	4	6.70	−6.58
K21-GR (Keskin and Ekici, 2021)	1	6.17	−6.17
K03-SL (Karayücel et al., 2003)	2	3.84	−3.70
K20-GN (Keskin et al., 2020)	1	1.03	+1.03

Table S8. [Model output] Site-ranking validation. Spearman ρ between predicted and observed growth, with 95% cluster-bootstrap percentile intervals resampling whole experiments. Columns differ in the unit ranked and in what is controlled for; the "rate | L_0, dur" column removes the initial length and duration the model was given, and is therefore the column that isolates information the model added. The only interval excluding zero belongs to the configuration supplied with no chlorophyll information at all, and the no-model arithmetic baseline ranks final length above the calibrated model. Cited in Results, Distribution of error, and generalisation.

Configuration	site ρ (rate)	group ρ (rate)	group ρ (rate | L_0, dur)	group ρ (final length)
uncorrected	0.358 [−0.569, 0.958]	0.274 [−0.512, 0.886]	−0.226 [−0.450, 0.299]	0.794 [0.587, 0.928]
global α	0.345 [−0.567, 0.901]	0.317 [−0.340, 0.886]	−0.221 [−0.638, 0.455]	0.727 [0.505, 0.936]
calibrated	0.297 [−0.615, 0.926]	0.295 [−0.310, 0.867]	−0.206 [−0.682, 0.549]	0.722 [0.499, 0.947]
site-mean chlorophyll	0.285 [−0.613, 0.925]	0.230 [−0.330, 0.835]	−0.307 [−0.651, 0.365]	0.671 [0.461, 0.924]
constant chlorophyll	0.297 [−0.568, 0.962]	0.603 [0.075, 0.888]	0.139 [−0.154, 0.635]	0.867 [0.673, 0.932]
no model (arithmetic)	n/a	n/a	n/a	0.843 [0.646, 0.934]

The arithmetic baseline is L_0 + median rate × duration; it has no rankable rate signal by construction, so the rate columns are undefined for it. All paired differences between the calibrated model and these baselines span zero, so the model is not demonstrably *worse* either: with ten experiments the design lacks the power to establish a difference in either direction.

Table S9. [Sensitivity analysis] Calibration under six spatial extraction geometries, computed on the eight experiments (25 cohort groups) for which all six return data, so that the variants are compared over the same cohort. This is the largest analytical sensitivity in the suite: X_K spans a factor of 3.5, RMSE spans 2.4 mm, and the bias changes sign. The calibration is identical from the nearest single valid pixel out to a 10 km box and becomes unstable at the 20 km box used for the primary results and beyond. Cited in Results, Sensitivity and robustness.

Geometry	Median valid pixels	X_K	RMSE (mm)	Bias (mm)
nearest valid pixel	1	0.047	12.25	+1.29
±0.02° (~4 km)	7	0.047	12.22	+1.20
±0.05° (~10 km)	47	0.047	12.40	+1.24
±0.10° (~20 km, used here)	177	0.102	13.86	−4.46
±0.25° (~50 km)	1256	0.165	13.79	−6.61
±0.50° (~100 km)	4808	0.150	11.49	−4.62

The two experiments excluded from this comparison are the two whose forcing is not spatially matched to the site (Table S5), and they drop out for different reasons: at Erdek (Y24-E) no geometry returns a valid retrieval at any width, while at the Dardanelles (Y05-D) the three geometries tighter than 0.1° return nothing and only the 0.1° box and wider reach the three pixels 12 km away. Comparing all six variants therefore requires dropping both. Widening the box does not degrade the fit monotonically, which is consistent with the ranking result: if the local pixel carried the growth-relevant signal, tightening the box would improve the fit.

Table S10. [Sensitivity analysis] Sensitivity of the calibrated X_K and of the overall fit to six analytical choices, each recomputed rather than re-scored. The digitisation test perturbs the observations themselves, so its effect is reported on X_K alone; the extraction-geometry row uses the eight experiments every geometry can deliver (Table S9). The two choices that change a conclusion are given in full in Results, Sensitivity and robustness; the four that do not are summarised there. The grouping row is also the test the Discussion appeals to over salinity: the three-way split assigns the Black Sea, the Marmara and the remaining marine sites to separate calibration groups, and spending a third fitted parameter on that distinction makes the fit no better than the two groups actually used (11.19 mm against 11.12 mm for two groups and 11.37 mm for one). The salinities the source papers report are therefore enough to test salinity as a grouping factor, and it does not earn its parameter. Cited in Results, Sensitivity and robustness, and in Discussion, Confounded processes and regional limits.

What was varied	How far	Change in fitted X_K	Change in error (mm)	Changes the conclusion?
The 1997 Sinop forcing gap	fill halved, doubled, replaced by the record mean, or the experiment dropped	none	≤0.26	No
Initial reserve density	full reserves down to 40%	none	0.04	No
Digitisation accuracy	random perturbation up to 2 mm	±10%	not comparable	No
How sites are grouped	a single group; the two groups used here; a three-way split by salinity class separating the Marmara	0.010 to 0.136	≤0.25	No
Which chlorophyll product	satellite L4 against biogeochemical reanalysis, on the nine experiments both cover	0.136 → 0.092	1.42, reanalysis better, but the interval reaches zero	Justification only
Where chlorophyll is sampled	nearest valid pixel out to a 100 km box	0.047 → 0.165 (3.5×)	2.37, and the bias changes sign	Yes

Supplementary Methods

These sections give the procedural detail behind the Methods of the main text. Nothing here is a separate analysis: the main text states what was done and what was found, and these sections state exactly how, at the level needed to reproduce it from the deposited code.

S1. Cohort digitisation and provenance

Growth trajectories were transcribed into a uniform cohort-group schema (initial length, final length, duration, culture system, coordinates, depth). Six of the ten experiments required figure digitisation (Table S3). Where the source PDF contained vector chart geometry, values were recovered by reading the path coordinates directly and mapping them onto the published axis calibration, which is exact to the precision of the drawing. Where the figure was a raster image, values were read by calibrated pixel measurement against the axis ticks. In both cases the recovered series was cross-checked against any value the source paper states in its own prose, and a per-experiment verification report records the comparison.

That audit corrected one entry with consequences for the analysis. The in-situ chlorophyll tabulated for Erdek was the midpoint of a reported annual range rather than an annual mean, overstating it by a factor of 2.7. The monthly series was re-digitised from Figure S5 of the source, a vector chart permitting exact geometric extraction, and the extraction was verified against the three values that paper states in prose (9.75, 0.31 and 7.97 µg L^{-1}, recovered as 9.750, 0.314 and 7.971). The resulting annual mean, 1.843 µg L^{-1}, is the value used throughout, and it withdrew the published claim that Erdek was a refractory site.

Digitisation error is propagated rather than assumed negligible. Each digitised final length was perturbed by Gaussian noise at σ = 0.5, 1 and 2 mm, 200 draws per level, and the calibration recomputed on each draw. The calibrated X_K moves by approximately 10% at σ = 2 mm, well inside its experiment-level bootstrap interval.

S2. Forcing extraction

Chlorophyll was averaged over a box centred on the deployment coordinates, of half-width 0.1° by default. Where that box returned no valid retrieval on any day of the deployment, the half-width was widened in fixed steps to 0.15°, 0.25°, 0.5° and finally 1°, stopping at the first width that yielded a series. Only pixels not carrying a quality flag were included. Per-experiment results are in Table S5, and the sensitivity of the calibration to this choice is in Table S9.

One temporal gap required filling. Chlorophyll retrievals are unavailable before SeaWiFS became operational, leaving 107 of the 382 days of the 1997 Sinop deployment (28% of the record, 1 June to 15 September) without data. These days were assigned a constant 0.5557 µg L^{-1}, the September climatological mean at that site. Because the gap covers the period contributing most of that experiment's growth, four variants were tested: the fill scaled by 0.5 and by 2, replacement by the whole-record satellite mean, and dropping the experiment from the cohort altogether. None moves the calibrated X_K and all move overall RMSE by ≤0.26 mm.

S3. Statistical protocol

Cross-validation. Folds are defined at the experiment level. Within each fold, the training set is the nine remaining experiments and every data-derived quantity is recomputed from them alone: the per-site α values, their median, and X_K itself. The held-out experiment is then predicted with those quantities. Basin assignment is geographic rather than data-derived and is held fixed. The naive variant, in which the correction is fitted once on the full cohort and then scored on it, is reported alongside so that the leakage is measured rather than assumed; it is +1.40 mm (95% CI +0.53, +2.94).

Clustering. The intraclass correlation is the one-way random-effects ICC of model residuals with experiment as the grouping factor. The design effect is $1 + (\bar{m} - 1)\rho_I$ with $\bar{m}$ the mean cluster size, and $n_{\text{eff}} = n/\text{deff}$. For this cohort ρ_I = 0.779, deff = 3.57 and $n_{\text{eff}} \approx 12$ against a nominal n = 43.

Observational uncertainty. A final-length SD is reported for 29 of the 43 cohort groups; all 14 missing values belong to one experiment (Y05-D), so weighting residuals by their reported precision would silently drop a third of the cohort and the whole of its largest experiment. Residuals are therefore unweighted. The cost of that choice is small relative to the model's own error: across the 29 groups that do report one, the median observation SD is 0.90 mm against a median absolute model residual of 7.39 mm, and the correlation between the two is −0.457 (p = 0.013), so the model fits worst where measurement is most precise. Measurement noise therefore does not account for the misfit.

Resampling. All resampling draws whole experiments with replacement, never cohort groups. The interval on X_K is an experiment-level block bootstrap of 1000 iterations in which each drawn experiment contributes its cohort groups with weight equal to its multiplicity in that resample. Intervals on RMSE and on rank correlations are cluster bootstraps of 10 000 draws. Comparisons between schemes are computed as paired differences on shared resamples, so the comparison is not contaminated by between-resample variance.

Model selection. AICc is $n \ln(\text{RSS}/n) + 2k + 2k(k+1)/(n-k-1)$, the Gaussian log-likelihood with concentrated variance $\hat{\sigma}^2 = \text{RSS}/n$ on residuals in shell length (mm). The parameter count k is the number of parameters fitted to *growth* data: 1 for a

single X_K, 2 for the per-group fit, and 0 for the α schemes, which are derived from in-situ chlorophyll and cost no growth-data degrees of freedom. Because the likelihood assumes independent residuals and they are not, AICc is evaluated at n_{eff} as well as at the nominal n, and a conclusion is accepted only if it holds at both.

Paired tests. Wilcoxon signed-rank tests on squared residuals, with the Diebold–Mariano t-approximation as a robustness check. For the thermal-set comparison these are computed at each set's *own* calibrated X_K, not at a common value, so that the sets are compared at their respective optima.

Ranking. Spearman ρ between predicted and observed growth, at group and site level, against three counterfactuals: each site's own mean chlorophyll held constant across time, one constant chlorophyll everywhere (no between-site information), and a no-model arithmetic baseline, L_0 + median rate × duration. Ranking is evaluated on growth rate and on increment as well as on final length, because final length is largely determined by the initial length and duration supplied to the model and would otherwise credit it for information it was given. Full results in Table S8.

Farm proximity. Güllük is the one paired farm/control design inside the cohort: an offshore fish-farm station and a coastal control that fall in a single satellite pixel and therefore receive identical forcing. Residuals are compared across the four matched station-by-cohort pairs. No systematic difference appears, +9.36 mm at the farm station against +9.30 mm at the control, the observed lengths themselves differing by 0.02 mm on average. This is a null result on this cohort and not evidence that enrichment is immaterial in general: the comparison is reported because the forcing provides no mechanism by which such a difference could enter the model at all, which is the point made in the Discussion. The contrast is with Pašman, where three cohorts at 0, 60 and 700 m from a fish farm differ by 5.13 mm in observed final length under the same identical forcing.

S4. Implementation and reproducibility

The model is implemented in Python with NumPy and SciPy. The reserve–structure–reproduction ODE system is integrated with 'scipy.integrate.odeint', which uses the LSODA adaptive stiff solver, at relative tolerance 10^{-8}, absolute tolerance 10^{-10}, a maximum step of 1 day, and an internal step budget of 5000 per output interval. That budget is raised from the SciPy default of 500, which multi-year runs exhaust under stiffness; solves that still fail are recorded rather than silently dropped, and grid points at which any cohort fails to solve are excluded from scoring, because an objective averaging over whatever converged will prefer parameter values at which the worst-fitting cohorts fail. Random seeds are fixed per analysis script.

Before any validation run, the implementation was checked against the tabulated values of the source parameterisation and the Add-my-Pet entry for temperature correction, functional response, flux balance, post-metamorphic s_M scaling and reserve-density consistency ($[E_m] = \{\dot{p}_{Am}\}/\dot{v}$).

Model predictions were reproduced on two independent software stacks, Python 3.12 with NumPy 2.4 and SciPy 1.17 and Python 3.14 with NumPy 2.3 and SciPy 1.16, the two SciPy versions spanning changes to the ODE integrator implementations this model exercises. Predicted final shell lengths for all 43 cohort groups, at five values of X_K spanning the calibration range (0.01–2.1 µg L^{-1}), agreed between the stacks to within 5×10^{-6} mm, with a median difference of zero and a worst case of 63.216301 against 63.216306 mm; no cohort group differed by more than 10^{-3} mm. This is six orders of magnitude below the model's own error, so no result reported here depends on the environment that produced it. The fingerprint tests the integrator and the parameter set rather than re-deriving every analysis, which the deposited reproduction script performs once the forcing has been re-extracted. The complete analysis, model implementation, forcing extraction, cohort database, calibration, validation and figure code, together with an environment specification and a single-command reproduction script, is deposited as described under Data Availability in the main text.

S5. Model formulation

The three integrated state variables are reserve energy E, structural volume V and the reproduction buffer E_R, following the standard abj flux definitions. Assimilation is $\dot{p}_A = \{\dot{p}_{Am}\}_{\text{eff}}\, f(X)\, V^{2/3}\, \text{TC}(T)$ with $f(X) = X/(X + X_K)$ evaluated on the daily chlorophyll series; mobilisation, somatic maintenance and growth follow Kooijman (2010) with the κ-rule allocation, and maturity maintenance is paid from the $1 - \kappa$ branch with the surplus accumulating in E_R. Structural shrinkage is disallowed, V being clamped at zero rather than permitted to go negative under starvation, which is standard practice in DEB growth studies of this species and avoids an unphysical shell-length decrease in the observable.

Because every cohort begins past metamorphosis, the metabolic acceleration factor $s_M = L_j/L_b$ is constant and is pre-applied to the surface-specific assimilation rate and to the energy conductance, giving the operative rates $\{\dot{p}_{Am}\}_{\text{eff}}$ and $\dot{v}_{\text{eff}}$ of Table S2. Using the unaccelerated rates instead would make $L_{\max}(f)$ wrong by an order of magnitude (15.0 mm rather than 155.5 mm at $f = 1$), so the distinction is recorded explicitly here.

Temperature dependence enters every rate through the five-parameter Arrhenius correction, evaluated in the DEBtool *tempcorr* convention with species-specific T_A, T_L, T_H, T_{AL} and T_{AH} (Table S2). The same TC(T) multiplies both $\kappa\dot{p}_C$ and $\dot{p}_M$, which is why it cancels from the $L_{\max}(f)$ ceiling and why that ceiling is a food threshold rather than a thermal one.

Initial conditions are $V_0 = (\delta_M L_0/10)^3$, $E_0 = [E_m]\, V_0$ and $E_{R0} = 0$, so every cohort starts at scaled reserve density $e = 1$ irrespective of the food environment its spat originated from. This is deliberately optimistic: it gives the model the most favourable possible starting state, so that any under-prediction cannot be attributed to a depleted initial reserve. Reducing it to $e(0) = 0.4$ leaves the calibrated X_K unchanged and moves overall RMSE by 0.04 mm (Table S10).

The observable used for validation is physical shell length, $L = 10\, V^{1/3}/\delta_M$ in mm. Spawning is not represented: E_R accumulates continuously rather than discharging in discrete events. This is immaterial for structural shell length, which is what we validate on, but not for tissue weight, which is why no weight statistics are reported anywhere in this study.