Supplemental Material: Data-Driven Discovery of Mechanistic Ecosystem Models with LLMs

October 17, 2025

S1 Curated Literature Collection

The local document collection used in this case study was carefully curated to provide comprehensive coverage of marine ecosystem modeling approaches, with particular focus on COTS-coral dynamics and management interventions. The collection encompasses several key research areas:

- Ecosystem Modeling Frameworks: [12] established foundational principles for ecosystem approaches to fisheries, while [13] introduced Models of Intermediate Complexity for Ecosystem assessments (MICE). [2] explored optimal model complexity levels.

- COTS Management and Ecology: [14] provided a comprehensive thirty-year review of COTS research. [9] developed models for COTS outbreak management, while [16] analyzed corallivore culling impacts under bleaching scenarios.

- Ecological Regime Shifts: [1] investigated predator-driven regime shifts in marine ecosystems. [11] provided insights into ecological tipping points through ecosystem modeling.

- Management Interventions: [3] examined large-scale interventions on the Great Barrier Reef. [15] explored harvest control implications using MICE models.

- Model Application Guidelines: [5] provided critical guidelines for adapting ecosystem models to new applications. [7] demonstrated multispecies production model applications for analyzing ecological and fishing effects.

- Integrated Systems: [8] and [10] explored integrated multi-trophic aquaculture modeling, providing insights into coupled biological systems. [17] analyzed trade-offs in seaweed farming between food production, livelihoods, marine biodiversity, and carbon sequestration benefits.

These papers were selected based on their direct relevance to COTS population dynamics, coral reef ecology, and ecosystem modeling approaches. The collection provided both specific parameter values and broader ecological context for model development.

S2 RAG Architecture Implementation

The Retrieval-Augmented Generation (RAG) system facilitates parameter search and extraction from scientific literature. The system employs two primary search strategies: a local search of user-curated documents and a comprehensive web search. For local search, the system uses ChromaDB as a persistent vector store to maintain an indexed collection of scientific papers and technical documents specifically curated by research teams for their ecological systems. These documents are processed into semantic chunks of approximately 512 tokens with small overlaps to preserve context while enabling precise retrieval of relevant information.

The parameter search process begins with the generation of enhanced semantic descriptions for each parameter. These descriptions are crafted to improve search relevance by capturing the ecological and mathematical context in which the parameters are used. The system first searches the user-curated local documents using embeddings generated through Azure OpenAI's embedding service. When necessary, it extends to web-based sources through two channels: querying the Semantic Scholar database for highly-cited papers in biology, mathematics, and environmental science, and conducting broader literature searches through the Serper API to capture additional relevant sources.

The search results from both local and web sources are processed through an LLM to extract numerical values. The system applies consistent validation across both search pathways, identifying minimum and maximum bounds, ensuring unit consistency, and validating source reliability. When direct parameter values are not found in either the local collection or web sources, the system defaults to the initial estimates from the coding LLM. All extracted information, including parameter values, valid ranges, and complete citation details, is stored in a structured JSON database for reproducibility and future reference.

The RAG system implements automatic retry mechanisms when initial searches fail to yield usable results. Each retry attempt follows a structured progression: first accessing the curated local collection through ChromaDB queries, then expanding to Semantic Scholar for peer-reviewed literature, and finally utilizing Serper API for broader scientific content. This progressive broadening of scope, while maintaining focus on ecologically relevant sources, ensures robust parameter estimation even in cases where direct measurements are sparse in the literature.

S3 AI Prompts Used in Model Development

The development of the model relied on several carefully crafted prompts to guide the artificial intelligence system. These prompts were designed to ensure numerical stability, proper likelihood calculation, and clear model structure. The following sections detail the exact prompts used at each stage of model development.

S3.1 Initial Model Creation

The initial model creation utilized a comprehensive prompt that emphasized three key aspects of model development. The prompt used for model initialization was:

```
Please create a Template Model Builder model for the following topic
    :[PROJECT_TOPIC]. Start by writing intention.txt, in which you
    provide a concise summary of the ecological functioning of the
    model. In model.cpp, write your TMB model with the following
    important considerations:

1. NUMERICAL STABILITY:
- Always use small constants (e.g., Type(1e-8)) to prevent division
    by zero
- Use smooth transitions instead of hard cutoffs in equations
- Bound parameters within biologically meaningful ranges using
    smooth penalties rather than hard constraints

2. LIKELIHOOD CALCULATION:
- Always include observations in the likelihood calculation, don't
    skip any based on conditions
- Use fixed minimum standard deviations to prevent numerical issues
    when data values are small
- Consider log-transforming data if it spans multiple orders of
    magnitude
- Use appropriate error distributions (e.g., lognormal for strictly
    positive data)

3. MODEL STRUCTURE:
- Include comments after each line explaining the parameters (
    including their units and how to determine their values)
- Provide a numbered list of descriptions for the equations
- Ensure all important variables are included in the reporting
    section
```

```
- Use '_pred' suffix for model predictions corresponding to '_dat' observations
```

S3.2 Parameter Enhancement

To enhance parameter descriptions for improved semantic search capabilities, the following prompt was employed:

```
Given a mathematical model about [PROJECT_TOPIC], enhance the semantic descriptions of these parameters to be more detailed and searchable. The model code shows these parameters are used in the following way:

[MODEL_CONTENT]

For each parameter below, create an enhanced semantic search, no longer than 10 words, that can be used for RAG search or semantic scholar search.
```

S3.3 Model Improvement

For iterative model improvements, the system utilized this prompt:

```
Improve the fit of the following ecological model by modifying the equations in this TMB script. Only make ONE discrete change most likely to improve the fit. Do not add stochasticity, but you may add other ecological relevant factors that may not be present here already.

You may add additional parameters if necessary, and if so, add them to parameters.json. Please concisely describe your ecological improvement in intention.txt and then provide the improved model.cpp and parameters.json content.
```

S3.4 Error Handling Prompts

For compilation errors, the system used this prompt:

```
model.cpp failed to compile. Here's the error information:

[ERROR_INFO]

```

```
Do not suggest how to compile the script
```

For data leakage issues, the system employed this detailed prompt:

```
Data leakage detected in model equations. The following response variables cannot be used to predict themselves:

To fix this:
1. Response variables ([RESPONSE_VARS]) must be predicted using only:
   - External forcing variables ([FORCING_VARS])
   - Other response variables' predictions (_pred variables)
   - Parameters and constants
2. Each response variable must have a corresponding prediction equation
3. Use ecological relationships to determine how variables affect each other

For example, instead of:
  slow_pred(i) = slow * growth_rate;
Use:
  slow_pred(i) = slow_pred(i-1) * growth_rate * (1 - impact_rate * cots_pred(i-1));

Please revise the model equations to avoid using response variables to predict themselves.
```

For numerical instabilities, the system used an adaptive prompt that became progressively more focused on simplification after multiple attempts:

```
The model compiled but numerical instabilities occurred. Here's the error information:

[ERROR_INFO]

[After 2+ attempts: Consider making a much simpler model that we can iteratively improve later.]
Do not suggest how to compile the script
```

S3.5 NPZ Case Study - Recovering Equations

The model implementation can be compared to the original NPZ equations from[4]:

$$\frac{dN}{dt} = \underbrace{-\frac{V_m N P}{k_s + N}}_{\text{nutrient uptake}} + \underbrace{\gamma(1-\alpha)\frac{gP^2Z}{k_g + P^2} + \mu_P P + \mu_Z Z^2}_{\text{recycling}} + \underbrace{S(N_0 - N)}_{\text{mixing}}$$

$$\frac{dP}{dt} = \underbrace{\frac{V_m N P}{k_s + N}}_{\text{growth}} - \underbrace{\frac{gP^2Z}{k_g + P^2}}_{\text{grazing loss}} - \underbrace{\mu_P P}_{\text{mortality}} - \underbrace{SP}_{\text{mixing}}$$

$$\frac{dZ}{dt} = \underbrace{\alpha\frac{gP^2Z}{k_g + P^2}}_{\text{growth (assimilation)}} - \underbrace{\mu_Z Z^2}_{\text{mortality}} - \underbrace{SZ}_{\text{mixing}}$$

Our generated model captures several key ecological processes from the original system:

1. Nutrient uptake by phytoplankton following Michaelis-Menten kinetics
2. Quadratic zooplankton mortality
3. Nutrient recycling through zooplankton excretion
4. Environmental mixing effects

For evaluating the ecological characteristics of generated models against the NPZ reference model, the system employed a 4-level ordinal scoring system that compares each model component to both the ground truth equations and recognized alternate formulations from the ecological literature. The evaluation system assessed nine ecological characteristics organized by equation: nutrient uptake, recycling, and mixing (dN/dt); phytoplankton growth, grazing loss, mortality, and mixing (dP/dt); and zooplankton growth and mortality (dZ/dt).

The scoring rubric used for all evaluations was:

```
Scoring rubric per characteristic (choose exactly one category):
- 3 = TRUTH_MATCH
    The mathematical structure is equivalent to the TRUTH model (
   modulo variable names,
    syntax, factor grouping, and coefficient naming). Quote the
   exact snippet that matches.
- 2 = ALTERNATE
    The implementation matches one of the alternates enumerated in
   the literature catalog,
```

```
    even if not identical to TRUTH. Name the family (e.g., "
   Michaelis-Menten uptake",
    "Ivlev grazing with threshold", "linear mortality", "Droop quota
   ").
- 1 = SIMILAR_NOT_LISTED
    The implementation plays the same ecological role and is
   mathematically similar
    (e.g., another saturating curve or plausible closure) but is not
     represented in TRUTH
    or alternates list.
- 0 = NOT_PRESENT_OR_INCORRECT
    The ecological component is missing or cannot be identified.
```

The alternate formulations catalog was based on[6] and included:

- Phytoplankton light response: linear, saturating (Michaelis-Menten, exponential, tanh), and photo-inhibiting forms
- Nutrient uptake: Michaelis-Menten, Liebig minimum limitation, Droop quota models
- Zooplankton grazing: linear, saturating with threshold, Holling/Ivlev type, acclimating forms
- Mortality terms: linear and quadratic (density-dependent) for both phytoplankton and zooplankton

Each characteristic was assigned a weight based on its contribution to its parent equation: the three nutrient equation components each had weight 0.333, the four phytoplankton components each had weight 0.25, and the two zooplankton components each had weight 0.5. The aggregate ecological score was calculated as the weighted sum of individual scores, then normalized to a 0-1 scale by dividing by the maximum possible score.

S3.5.1 Validation of Scoring System

To validate the ecological characteristics scoring system, we tested it on the ground truth NPZ model itself (evaluating the model against its own equations). This test confirmed that the scoring system could correctly identify and score all nine ecological characteristics when they were present in their canonical forms.

The validation results demonstrated perfect performance:

- All nine characteristics received scores of 3 (TRUTH_MATCH)

- Raw total score: 8.997 (out of maximum 9.0, with small rounding due to floating point arithmetic)
- Normalized total score: 1.0000 (perfect score on 0-1 scale)
- Zero extra components identified (correctly recognized model contained only canonical NPZ processes)

The LLM evaluator correctly identified each ecological mechanism in the ground truth model, providing detailed explanations such as "algebraically identical to the TRUTH NPZ model" and specifically noting the presence of "Michaelis-Menten style nutrient limitation multiplied by a light/self-shading term for phytoplankton growth" and "a saturating $P^2/(\mu^2+P^2)$ (Hill/Type-III-like) grazing formulation." This validation confirmed that the scoring system could reliably distinguish between different levels of ecological fidelity, from exact matches to the ground truth through recognized alternates to novel formulations, providing a robust framework for assessing LEMMA-generated models.

S4 NPZ Validation

S4.1 Best Performing NPZ Model

S4.1.1 Model Description

The following model represents our framework's attempt to recover the NPZ dynamics from[4]. The model aims to capture three key components:

- Nutrient uptake and recycling
- Phytoplankton growth and mortality
- Zooplankton predation and dynamics

S4.1.2 Model Intention

```
\section{Ecological Intention}

A key modification was made to incorporate direct nutrient recycling
     from zooplankton grazing activity. In marine systems,
    zooplankton feeding is often inefficient, with a significant
    portion of consumed phytoplankton being released as dissolved
    nutrients rather than being assimilated into biomass or entering
```

the detritus pool. This "sloppy feeding" process creates an important feedback loop where grazing can stimulate new primary production through rapid nutrient recycling.

The recycling efficiency is temperature-dependent, reflecting how metabolic rates and feeding mechanics vary with temperature. This creates an adaptive feedback where warmer conditions lead to both increased grazing pressure and faster nutrient recycling, better capturing the coupled nature of predator-prey interactions in planktonic systems.

The modification introduces a direct pathway from grazing to dissolved nutrients, complementing the slower recycling through the detritus pool. This better represents the multiple timescales of nutrient cycling in marine food webs and helps explain how high productivity can be maintained even under intense grazing pressure.

S4.1.3 Model Implementation

NPZ Model: Parameter and Equation Tables

Parameter summary

Symbol	Units	Meaning	Init. value	Bounds	Source	Literature (citekey)
log_mu_max	day^{-1} (log scale)	Log of maximum phytoplankton growth rate at reference conditions (day^{-1}).	-0.02	[-0.22, 0.18]	literature	Yes (LitNPZ_log_mu
log_K_N	g C m^{-3} (log scale)	Log of half-saturation constant for nutrient uptake (g C m^{-3}).	-3.00	[-6.91, 0.00]	literature	Yes (LitNPZ_log_K_
I	W m^{-2}	Mean photosynthetically active irradiance proxy over the modeled period.	150.00	[0.00, 500.00]	initial estimate	No
log_K_I	W m^{-2} (log scale)	Log of light half-saturation constant for photosynthesis (W m^{-2}).	4.32	[0.00, 5.70]	literature	Yes (LitNPZ_log_K_
log_g_max	day^{-1} (log scale)	Log of maximum zooplankton grazing rate per unit Z biomass (day^{-1}).	-0.69	[-3.00, 0.69]	literature	Yes (LitNPZ_log_g_
log_K_G	g C m^{-3} (log scale)	Log of P half-saturation constant for grazing functional response (g C m^{-3}).	-2.30	[-6.91, 0.00]	literature	Yes (LitNPZ_log_K_
h_grazing	dimensionless	Holling type III shape exponent ($h \geq 1$).	2.00	[1.00, 3.00]	literature	Yes (LitNPZ_h_grazi
logit_e_Z	dimensionless (logit scale)	Logit of zooplankton assimilation efficiency ($e_Z \in (0,1)$); $e_Z = 0.5$ at value 0.	0.00	—	literature	Yes (LitNPZ_logit_e_
log_m_P	day^{-1} (log scale)	Log of phytoplankton linear mortality rate (day^{-1}).	-3.00	[-6.91, -1.20]	literature	Yes (LitNPZ_log_m_
log_m_Z	day^{-1} (log scale)	Log of zooplankton linear mortality rate (day^{-1}).	-3.51	[-6.91, -1.20]	literature	Yes (LitNPZ_log_m_
log_gamma_Z	(g C m^{-3})$^{-1}$ day^{-1} (log scale)	Log of zooplankton quadratic self-limitation coefficient ((g C m^{-3})$^{-1}$ day^{-1}).	-4.61	[-9.21, -1.61]	initial estimate	No
logit_r_P	dimensionless (logit scale)	Logit of fraction of P mortality that is remineralized to N (0..1).	0.85	—	literature	Yes (LitNPZ_logit_r_
logit_r_Z	dimensionless (logit scale)	Logit of fraction of Z mortality that is remineralized to N (0..1).	0.85	—	literature	Yes (LitNPZ_logit_r_
log_ex_Z	day^{-1} (log scale)	Log of zooplankton excretion rate to nutrients (day^{-1}).	-4.61	[-13.82, -1.61]	initial estimate	No
log_k_mix	day^{-1} (log scale)	Log of vertical mixing rate driving nutrients toward $N_\star$ (day^{-1}).	-3.91	[-13.82, -0.69]	initial estimate	No
$N_\star$	g C m^{-3}	Deep/source nutrient concentration towards which mixing relaxes the system.	0.30	[0.00, 2.00]	initial estimate	No
log_q10	dimensionless (log scale)	Log of Q10 temperature scaling factor (dimensionless), typical $Q10 \approx 2$.	0.66	[0.61, 0.71]	literature	Yes (LitNPZ_log_q10
T_C	deg C	Ambient temperature used for Q10 scaling (deg C).	15.00	[0.00, 35.00]	initial estimate	No
T_ref	deg C	Reference temperature for Q10 scaling (deg C).	15.00	[0.00, 35.00]	literature	Yes (LitNPZ_T_ref)
log_k_rem	day^{-1} (log scale)	Log of detritus remineralization rate to nutrients (day^{-1}).	-2.30	[-4.61, 0.00]	conceptual addition	No
log_k_sink	day^{-1} (log scale)	Log of detritus sinking/export rate out of mixed layer (day^{-1}).	-4.61	[-13.82, 0.00]	conceptual addition	No
log_sigma_N	log-scale SD	Log of observation SD for N on the log scale.	-2.30	[-5.00, 2.00]	initial estimate	No
log_sigma_P	log-scale SD	Log of observation SD for P on the log scale.	-2.30	[-5.00, 2.00]	initial estimate	No
log_sigma_Z	log-scale SD	Log of observation SD for Z on the log scale.	-2.30	[-5.00, 2.00]	initial estimate	No

S5 CoTS Model Convergence

S5.1 Model Evolution and Convergence

The evolutionary process exhibited consistent refinement across generations, with measurable improvements in model performance. On average, populations reached their best-performing individual within 6.9 generations, and the mean improvement frequency across all populations was 38.0%. Figure 1 shows the distribution of successful, culled, and broken models across generations. Notably, two populations achieved convergence below the target threshold, representing 9.5% of all populations. Performance varied significantly across populations. The fastest-converging population reached an optimal objective value of 0.0035 in just 3 generations, while others required up to 13 generations. This population also demonstrated a high improvement rate of -0.655 and a consistent improvement frequency of 50%. In contrast, several populations showed minimal or no improvement, with some failing to converge within the allotted iterations.

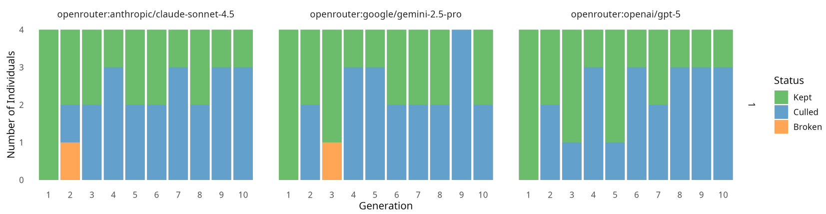

Figure 1: Evolution of model performance during the genetic algorithm optimization process. Each generation represents an iteration of model development, where models are evaluated and classified into three categories: the best performers according to the NMSE objective value (kept, green), those that are numerically stable but outcompeted (culled, blue), and those that failed due to numerical instability, data leakage, or syntax errors (broken, orange). The vertical axis shows the count of new models in each category per generation, while rows represent independent replicates using different LLM configurations. Gemini-2.5-Pro was included in the analysis but did not produce successful runs for some populations.

S5.2 Numerical Stability and Optimization

Numerical stability varied across LLM configurations, with runtime and generation time metrics reflecting differences in optimization efficiency. The GPT-5 configuration showed moderate efficiency, with an average generation time of 12.0 minutes (SD = 13.0). The Claude Sonnet 4.5 configuration had longer generation times, averaging 71.2 minutes (SD = 155.2), though this includes variability from a small number of outlier populations. In contrast, the Gemini-2.5-Pro configuration demonstrated the fastest generation cycles, averaging 4.1 minutes per generation (SD = 0.54), though it exhibited lower convergence rates and higher instability in some cases. Figure 2 illustrates the distribution of iteration counts required for successful model convergence across LLMs. Most models converged within 4 to 7 iterations, with some outliers requiring up to 11 iterations.

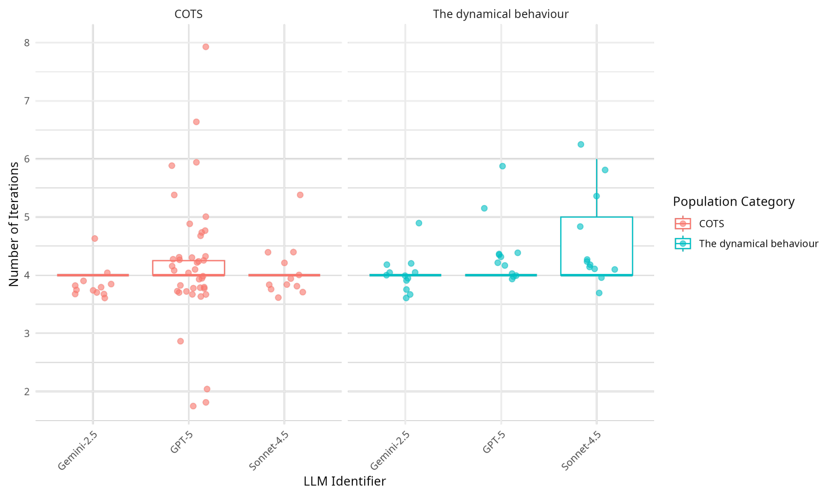

Figure 2: Distribution of iteration counts for successful model instances by LLM configuration. The boxplot excludes cases that reached maximum iterations or remained numerically unstable.

S6 Comparative Analysis of Best-Performing Models

Before presenting the full code for each model, we analyze the key differences between the best-performing models to understand their ecological approaches and mathematical structures.

S6.1 Detailed Ecological Mechanisms

Mechanism (anchored to Human `CoTSmodel_v4.cpp`)	Human v4	Google Gemini 2.5 Pro	OpenAI GPT–5	Anthropic Claude Sonnet 4.5
COTS state structure / life history	*Age-structured (3 classes)*: $N_{t+1,1} = N_{t,0}e^{-M_0}$; $N_{t+1,2} = N_{t,1}e^{-fM_1} + N_{t,2}e^{-fM_2}$; recruits form $N_{t+1,0}$. Age-specific mortality $M_a = M_{\text{cots}}+\lambda/(1+a)$.	*Single compartment* C_t (no explicit ages). Growth from consumption + modifiers.	*Two-stage* (juveniles J and adults C): maturation $m_J J$; juvenile mortality $\mu_J J$; adult mortality $(\mu_C + \gamma_C C)C$.	*Single compartment* with logistic-like adult growth plus additive recruitment pulse; no age classes.
COTS stock–recruitment	*Beverton–Holt (BH) from spawners*: $R_{t+1} = \dfrac{\alpha(N_{t+1,2}/K_{\text{sp}})}{\beta + (N_{t+1,2}/K_{\text{sp}})}$; with α, β derived from slope steepness h and R_0.	*Not BH*: recruitment implicit via $\big(e_F \cdot \text{Cons}_F + e_S \cdot \text{Cons}_S\big) \times$ Allee $\times$ temperature Gaussian (no explicit SR function).	*BH-like taper on adults*: $\text{Rec} = \alpha_{\text{rec}}\,[C^{\phi}/(1 + C/C_{\text{sat}})] \cdot f_{\text{Allee}}(C) \cdot f_T(\text{SST}) \cdot f_{\text{food}} + \text{imm}_{t-1}$.	*Additive pulse recruitment*: $\text{recruit\_pulse} = R_{\max} \cdot \sigma\big(\text{fav} - \tau\big) \cdot \text{fav}$, added to adult change; adult growth also logistic-like.
COTS immigration & interannual pulses	*Background immigration* + year-specific deviations: $N_{t+1,0} = (R_{t+1} + \text{Imm}_{\text{CoTS}})\exp(\varepsilon_{t+1} + \sigma^2/2)$; ε_t applied in specified years.	Exogenous additive series $\text{cotsimm}(t)$ added directly to C_t dynamics.	Exogenous series cotsimm_{t-1} added to recruitment (lagged).	Used twice: (i) normalized within favorability index, and (ii) multiplicative $\text{immigration\_boost} = 1 + \text{effect} \cdot \text{imm}$ on adult growth.
COTS mortality and its drivers	*Baseline + age-dependent*, modulated by	*No age structure*; mortality = survival Allee	*No age structure*; adult mortality	*No age structure*; baseline + density-

Human (name)	**Role / Units**	**Gemini 2.5 Pro**	**OpenA**
`Mcots`	Baseline instantaneous COTS mortality (yr^{-1})	$m_{C,\max}$ (low-density mortality scale); $m_{C,\text{dd}}$ (DD mortality)	μ_C (bas
`lam`	Age-dependence of mortality, $M_a = M_{\text{cots}} + \lambda/(1+a)$	— (no age structure)	— (two-
`ptil`	Fraction of M attributable to fast-coral availability (mortality modulation)	— (no coral-modulated M)	— (no c
`h` (BH steepness)	Shapes SR via $(R_0, h) \to \alpha, \beta$	— (no BH SR; growth from consumption)	$C_{\text{sat,rec}}$
`R0`	Recruitment at unexploited state (for SR derivation)	— (no explicit R_0)	α_{rec} (sca

Human (name)	Role / Units	Gemini 2.5 Pro
`p1f`, `p1m`	Per-COTS coral loss coefficients (fast/slow), units of % cover $\cdot$ (ind^{-1} yr^{-1})	a_F, a_S (attack), h (handling) with efficiencies e_F, e_S for COTS growt
`p2f`, `p2m`	Logistic saturation vs COTS density (shape of Q_F, Q_M)	Denominator $1 + a_F hF + a_S hS$ (resource-based saturation; no explic
`switchSlope`	Controls prey switching: $\rho = \exp(-\text{switchSlope} \cdot F/K)$	— (no explicit ρ)

Table 3: Predation and prey-switching parameters. Human model uses COTS-density logistic saturation (via $p2$) and an explicit exponential switching function; LLMs use Holling multi-prey forms with handling and (in GPT–5) Type-III exponents; Claude adds an explicit preference term.

Human (name)	Role / Units	Gemini 2.5 Pro	OpenAI GPT–5	Claude Sonnet 4.5
`K`	Shared coral carrying capacity (% cover scale)	K_{coral} (`log_K_coral` parameterized)	K_{tot}	K_{coral} (`log_K_coral`)
`rf`, `rm`	Intrinsic regrowth (fast/slow), yr^{-1}	r_F, r_S (`log_r_F`, `log_r_S`)	rF, rS	$r_{\text{fast}}, r_{\text{slow}}$ (`log_r_fast`, `log_r_slow`)
`Cf_init`, `Cm_init`	Initial coral state (fraction of K)	Init from first data row (no param)	$F0, S0$	Init from first data row (no param)

Table 4: Coral demography and space limitation. All models use a shared carrying capacity for two coral groups; GPT–5 exposes explicit initial-state parameters.

Human (name)	**Role / Units**	**Gemini 2.5 Pro**	**OpenAI GPT-**
`Eta_f`, `Eta_m`	Logistic bleaching slope (fast/slow)	k_{bleach} (common steepness)	β_{bleach} (growth r
`M_SST50_f`, `M_SST50_m`	50% bleaching SST (fast/slow)	$T_{\text{bleach},F/S}$	$T_{\text{opt,bleach}}$ (singl
`Ble_imp_f`, `Ble_imp_m`	Optional impulse bleaching toggles	—	—
`SST0_f`, `SST0_m`	Coral Gaussian performance optima (fast/slow)	— (no Gaussian growth multiplier; bleaching only)	$T_{\text{opt,bleach}}$ used i
`SST_sig_f`, `SST_sig_m`	Coral Gaussian performance widths (fast/slow)	—	—
— (none; human has no COTS thermal term)	—	$T_{\text{opt,rec}}, \beta_{\text{rec}}$ (Gaussian on recruitment)	`log_temp_opt`,

Table 5: Temperature effects. Human model: Gaussian performance on coral growth + logistic bleaching; LLMs: Gemini logistic bleaching and Gaussian window for COTS reproduction; GPT-5 heat reduction on coral growth

COTS–Coral Model: Parameter and Equation Tables

Implementation details (optimization stability)

- **Soft bounds**: Parameter ranges are enforced via a smooth quadratic penalty with weight $w_{\text{pen}} = 10^{-3}$ (not hard constraints).
- **Non-negativity and % clamping**: States use a smooth positive-part function for $x_+ \approx \max(0, x)$; coral % is smoothly clamped to $[0, 100]$.
- **Logit transform safety**: The logit of % cover uses a small ϵ to avoid $0/100$ singularities.
- **Likelihood SD floors**: Observation SDs use a floor (≥ 0.05) for numerical stability.
- **Food term default**: If no external driver is provided, $f_{\text{food}} = 1$ (neutral), leaving recruitment unaffected by food.

Parameter summary

Symbol	Units	Meaning	Init. value	Bounds	Source	Citation
C0	ind m^{-2}	Initial adult COTS density at first time step	0.1	[0.0, 50.0]	initial estimate	No
J0	ind m^{-2}	Initial juvenile COTS pool at first time step	0.1	[0.0, 50.0]	initial estimate	No
F0	% cover	Initial fast coral (Acropora) cover at first time step	30.0	[0.0, 100.0]	initial estimate	No
S0	% cover	Initial slow coral (Faviidae/Porites) cover at first time step	30.0	[0.0, 100.0]	initial estimate	No
alpha_rec	ind m^{-2} yr^{-1}	Recruitment productivity scaling controlling outbreak potential	1.0	[0.0, 10.0]	initial estimate	No
phi	dimensionless	Fecundity density exponent shaping recruitment curvature	1.5	[1.0, 3.0]	initial estimate	No
k_allee	m^2 ind^{-1}	Steepness of smooth Allee effect on recruitment	2.0	[0.01, 20.0]	initial estimate	No
C_allee	ind m^{-2}	Allee density where mating success increases rapidly	0.2	[0.0, 5.0]	initial estimate	No
C_sat_rec	ind m^{-2}	Adult density scale for stock–recruitment taper (Beverton–Holt)	2.0	[0.01, 50.0]	improvement	No
muC	yr^{-1}	Baseline adult COTS mortality rate	0.6	[0.0, 3.0]	initial estimate	No
gammaC	m^2 ind^{-1} yr^{-1}	Density-dependent mortality coefficient	0.5	[0.0, 10.0]	initial estimate	No
mJ	yr^{-1}	Annual maturation fraction from juvenile to adult	0.5	[0.0, 1.0]	initial estimate	No
muJ	yr^{-1}	Annual proportional mortality of juvenile COTS	0.5	[0.0, 1.0]	initial estimate	No
T_opt_rec	degC	Optimal SST for COTS recruitment success	26.5	[20.0, 34.0]	literature	Yes[3,9,14]
beta_rec	$degC^{-2}$	Curvature of Gaussian temperature effect on recruitment	0.2	[0.0, 2.0]	initial estimate	No
K_food	units of food_dat	Half-saturation constant for larval food limitation on recruitment	1.0	[0.001, 1000.0]	improvement	No
T_opt_bleach	degC	SST threshold where bleaching stress starts impacting coral	32.65	[20.0, 34.0]	literature	Yes[3,16]
beta_bleach	dimensionless	Multiplier controlling growth reduction under heat stress	0.5	[0.0, 5.0]	initial estimate	No
m_bleachF	yr^{-1} $degC^{-1}$	Additional proportional loss of fast coral per °C above threshold	0.2	[0.0, 2.0]	initial estimate	No
m_bleachS	yr^{-1} $degC^{-1}$	Additional proportional loss of slow coral per °C above threshold	0.1	[0.0, 2.0]	initial estimate	No
rF	yr^{-1}	Intrinsic regrowth rate of fast coral	0.5	[0.0, 2.0]	literature	Yes[3,14]
rS	yr^{-1}	Intrinsic regrowth rate of slow coral	0.2	[0.0, 2.0]	literature	Yes[3,14]
K_tot	% cover	Total carrying capacity for combined coral cover	70.0	[10.0, 100.0]	literature	Yes[3,16]
aF	yr^{-1} $\%^{-\eta_F}$ m^2 ind^{-1}	Encounter/attack parameter on fast coral	0.02	[0.0, 1.0]	initial estimate	No
aS	yr^{-1} $\%^{-\eta_S}$ m^2 ind^{-1}	Encounter/attack parameter on slow coral	0.01	[0.0, 1.0]	initial estimate	No
etaF	dimensionless	Shape exponent for fast coral	1.5	[1.0, 3.0]	initial estimate	No
etaS	dimensionless	Shape exponent for slow coral	1.2	[1.0, 3.0]	initial estimate	No
h	yr $\%^{-1}$	Handling/satiation scaler controlling saturation	0.02	[0.0, 1.0]	initial estimate	No
qF	dimensionless	Efficiency converting fast coral feeding into % cover loss	0.8	[0.0, 1.0]	literature	Yes[16]
qS	dimensionless	Efficiency converting slow coral feeding into % cover loss	0.5	[0.0, 1.0]	literature	Yes[16]
sigma_cots	log-space SD	Observation/process error SD for COTS	0.5	[0.01, 2.0]	initial estimate	No
sigma_fast	logit-space SD	Observation/process error SD for fast coral	0.3	[0.01, 2.0]	initial estimate	No
sigma_slow	logit-space SD	Observation/process error SD for slow coral	0.3	[0.01, 2.0]	initial estimate	No

Process and observation equations

Name	Equation	Description
Allee effect	$f_{\text{Allee}}(C) = \frac{1}{1+\exp\{-k_{\text{allee}}(C-C_{\text{allee}})\}}$	Smooth logistic Allee effect on recruitment
Recruitment temperature	$f_{T,\text{rec}}(\text{SST}) = \exp\{-\beta_{\text{rec}}(\text{SST} - T_{\text{opt,rec}})^2\}$	Gaussian peak temperature modifier
Food modifier	$f_{\text{food}} = \frac{\text{food}}{K_{\text{food}}+\text{food}}$ $(= 1$ if no driver$)$	Monod saturation; neutral if driver missing
Stock term	$s(C) = \frac{C^{\phi}}{1+C/C_{\text{sat,rec}}}$	Beverton–Holt taper at high adult density
Juvenile recruitment	$\text{Rec}_t = \alpha_{\text{rec}}\, s(C_{t-1})\, f_{\text{Allee}}(C_{t-1})\, f_{T,\text{rec}}(\text{SST}_{t-1})\, f_{\text{food}} + \text{IMM}_{t-1}$	Recruitment with environmental modifiers and immigration
Adult mortality	$\text{Mort}_{\text{adult}} = (\mu_C + \gamma_C C)\, C$	Baseline + density-dependent adult mortality
Juvenile flows	$\text{Mat} = m_J J,\;\; \text{Mort}_{\text{juv}} = \mu_J J$	Maturation and juvenile mortality
Adult update	$C_t = \max\{0,\; C_{t-1} + \text{Mat} - \text{Mort}_{\text{adult}}\}$	Nonnegative adult state update
Juvenile update	$J_t = \max\{0,\; J_{t-1} + \text{Rec}_t - \text{Mat} - \text{Mort}_{\text{juv}}\}$	Nonnegative juvenile state update
Heat stress	$\text{heat} = \max\{0,\; \text{SST} - T_{\text{opt,bleach}}\}$	Bleaching temperature exceedance
Bleach growth mod	$g_{\text{bleach}} = \exp\{-\beta_{\text{bleach}}\, \text{heat}\}$	Growth reduction under heat stress
Space limitation	$\text{space} = 1 - \frac{F+S}{K_{\text{tot}}}$	Shared coral space carrying capacity
Coral growth	$G_F = r_F F\, \text{space}\, g_{\text{bleach}},\;\; G_S = r_S S\, \text{space}\, g_{\text{bleach}}$	Fast/slow coral intrinsic regrowth
Bleaching losses	$B_F = m_{\text{bleachF}}\, \text{heat}\, F,\;\; B_S = m_{\text{bleachS}}\, \text{heat}\, S$	Additional proportional bleaching losses
Functional response denom	$\text{denom} = 1 + h(a_F F^{\eta_F} + a_S S^{\eta_S})$	Type II/III multi-prey denominator
Feeding losses	$\text{Cons}_F = q_F \frac{a_F F^{\eta_F} C}{\text{denom}},\;\; \text{Cons}_S = q_S \frac{a_S S^{\eta_S} C}{\text{denom}}$	COTS consumption of fast/slow coral
Coral updates	$F_t = F + G_F - \text{Cons}_F - B_F,\;\; S_t = S + G_S - \text{Cons}_S - B_S$	Fast/slow coral cover updates
Observation: COTS	$\log Y^{(C)} \sim \mathcal{N}(\log C,\, \sigma^2_{\text{cots}})$	Lognormal observation with Jacobian in NLL
Observation: fast	$\text{logit}_{\%}(Y^{(F)}) \sim \mathcal{N}(\text{logit}_{\%}(F),\, \sigma^2_{\text{fast}})$	Normal on logit-% cover
Observation: slow	$\text{logit}_{\%}(Y^{(S)}) \sim \mathcal{N}(\text{logit}_{\%}(S),\, \sigma^2_{\text{slow}})$	Normal on logit-% cover

References

[1] Blamey, L. K., Plagányi, E. E., and Branch, G. M. (2014). Was overfishing of predatory fish responsible for a lobster-induced regime shift in the Benguela? *Ecological Modelling*, 273:140–150.

[2] Collie, J. S., Botsford, L. W., Hastings, A., Kaplan, I. C., Largier, J. L., Livingston, P. A., Plagányi, E., Rose, K. A., Wells, B. K., and Werner, F. E. (2016). Ecosystem models for fisheries management: finding the sweet spot. *Fish and Fisheries*, 17:101–125.

[3] Condie, S. A., Anthony, K. R. N., Babcock, R. C., Baird, M. E., Beeden, R., Fletcher, C. S., Gorton, R., Harrison, D., Hobday, A. J., Plagányi, E. E., and Westcott, D. A. (2021). Large-scale interventions may delay decline of the Great Barrier Reef. *Royal Society Open Science*, 8:201296.

[4] Edwards, A. M. and Brindley, J. (1999). Zooplankton mortality and the dynamical behaviour of plankton population models. *Bulletin of mathematical biology*, 61(2):303–339.

[5] Essington, T. E. and Plagányi, E. E. (2014). Pitfalls and guidelines for “recycling” models for ecosystem-based fisheries management: evaluating model suitability for forage fish fisheries. *ICES Journal of Marine Science*, 71:118–127.

[6] Franks, P. J. (2002). Npz models of plankton dynamics: their construction, coupling to physics, and application. *Journal of Oceanography*, 58(2):379–387.

[7] Gamble, R. J. and Link, J. S. (2009). Analyzing the tradeoffs among ecological and fishing effects on an example fish community: A multispecies (fisheries) production model. *Ecological Modelling*, 220:2570–2582.

[8] Hadley, S., Wild-Allen, K., Johnson, C., and Macleod, C. (2015). Modeling macroalgae growth and nutrient dynamics for integrated multi-trophic aquaculture. *Journal of Applied Phycology*, 27:901–916.

[9] Morello, E. B., Plagányi, E. E., Babcock, R. C., Sweatman, H., Hillary, R., and Punt, A. E. (2014). Model to manage and reduce crown-of-thorns starfish outbreaks. *Marine Ecology Progress Series*, 512:167–183.

[10] Oca, J., Cremades, J., Jiménez, P., Pintado, J., and Masaló, I. (2019). Culture of the seaweed Ulva ohnoi integrated in a Solea senegalensis recirculating system: influence of light and biomass stocking density on macroalgae productivity. *Journal of Applied Phycology*, 31(4):2461–2467.

[11] Plagányi, E., Ellis, N., Blamey, L., Morello, E., Norman-Lopez, A., Robinson, W., Sporcic, M., and Sweatman, H. (2014a). Ecosystem modelling provides clues to understanding ecological tipping points. *Marine Ecology Progress Series*, 512:99–113.

[12] Plagányi, E. E. (2007). Models for an ecosystem approach to fisheries.

[13] Plagányi, E. E., Punt, A. E., Hillary, R., Morello, E. B., Thébaud, O., Hutton, T., Pillans, R. D., Thorson, J. T., Fulton, E. A., Smith, A. D. M., Smith, F., Bayliss, P., Haywood, M., Lyne, V., and Rothlisberg, P. C. (2014b). Multispecies fisheries management and conservation: tactical applications using models of intermediate complexity. *Fish and Fisheries*, 15:1–22.

[14] Pratchett, M., Caballes, C., Wilmes, J., Matthews, S., Mellin, C., Sweatman, H., Nadler, L., Brodie, J., Thompson, C., Hoey, J., Bos, A., Byrne, M., Messmer, V., Fortunato, S., Chen, C., Buck, A., Babcock, R., and Uthicke, S. (2017). Thirty Years of Research on Crown-of-Thorns Starfish (1986–2016): Scientific Advances and Emerging Opportunities. *Diversity*, 9:41.

[15] Punt, A. E., MacCall, A. D., Essington, T. E., Francis, T. B., Hurtado-Ferro, F., Johnson, K. F., Kaplan, I. C., Koehn, L. E., Levin, P. S., and Sydeman, W. J. (2016). Exploring the implications of the harvest control rule for Pacific sardine, accounting for predator dynamics: A MICE model. *Ecological Modelling*, 337:79–95.

[16] Rogers, J. G. D. and Plagányi, E. E. (2022). Culling corallivores improves short-term coral recovery under bleaching scenarios. *Nature Communications*, 13:2520.

[17] Spillias, S., Cottrell, R. S., Layton, C., O'Brien, K. R., and McDonald-Madden, E. (2024). Having our kelp and eating it too: Minimizing trade-offs from seaweed farming. *Journal of Cleaner Production*, 448:141150.