

Biologically-Constrained Parameter Reduction for 5-Species Biofilm Model

Keisuke Nishioka

February 2026

Abstract

This document describes the Proposed Method, a parameter reduction technique for Bayesian estimation of the 5-species biofilm interaction model. By incorporating biological knowledge from experimentally determined interaction networks, the algorithm reduces the parameter space from 20 to 15 free parameters, improving estimation efficiency and biological interpretability.

Keywords: Bayesian Estimation, Biofilm, Parameter Reduction, TMCMC, Biological Constraints, Multi-species Interaction, Inverse Problem, Peri-implantitis, Uncertainty Quantification

Contents

1	Introduction	4
2	Biological Basis	4
2.1	Species in the Model	4
2.2	Interaction Network (Figure 4C)	4
2.3	Active Interactions	4
2.4	Absent Interactions (Locked)	4
3	Mathematical Formulation	5
3.1	Governing Equations	5
3.2	F. nucleatum Bridge Gate (Hill Function Extension)	5
3.3	Symmetric Matrix Assumption	6
3.4	Parameter Vector Definition	6
3.5	Complete Parameter Mapping	7
3.6	Locked Parameter Indices	7
3.7	Prior Bounds	8
3.8	Effective Parameter Space	8
4	Bayesian Inference Framework	9
4.1	Forward Model	9
4.2	Bayesian Inverse Problem	9
4.3	Likelihood Function	10
4.3.1	Species-Specific Likelihood Weighting	10
4.4	Prior Distribution with Biological Constraints	10

5	Transitional Markov Chain Monte Carlo (TMCMC)	11
5.1	Algorithm Overview	11
5.2	Adaptive Tempering Schedule	11
5.3	Resampling and MCMC Mutation	12
5.4	TMCMC Procedure	12
5.5	Model Evidence Estimation	12
5.6	Taylor Series Method (TSM) Surrogate	14
5.7	Sensitivity Analysis of Hill Gate Parameters	14
6	Experiment Conditions & Parameter Estimation	14
6.1	Parameter Locking Rules	14
6.2	Detailed Locking Logic	14
6.3	4-Stage Sequential Estimation	16
6.4	Sequential Estimation Algorithm	17
7	Implementation	18
7.1	Core Module: <code>core/nishioka_model.py</code>	18
7.2	Estimation Script: <code>main/estimate_reduced_nishioka.py</code>	18
8	Advantages and Limitations	18
8.1	Advantages	18
8.2	Limitations	19
9	Usage	19
9.1	Running the Estimation	19
9.2	Comparing with Standard Results	19
9.3	Output Files	19
10	Numerical Experiments and Discussion	19
10.1	Computational Settings	19
10.2	Experimental Conditions	20
10.3	Evaluation of Model Fit and Prediction Accuracy	20
10.3.1	Detailed Fit with MAP Estimates	20
10.3.2	Uncertainty Evaluation of Posterior Predictive Distribution	20
10.3.3	Residual Analysis	21
11	Results	22
11.1	Commensal HOBIC	22
11.2	Dysbiotic HOBIC (Surge)	28
11.3	Commensal Static	31
11.4	Dysbiotic Static	34
12	Comparative Analysis	37
12.1	Commensal vs. Dysbiotic	37
12.2	Static vs. HOBIC	37
12.3	The "Surge" Phenomenon	37
12.4	Quantitative Comparison of Interaction Matrices	37
13	Multiscale Mechanical Pipeline	38
13.1	Dysbiosis Index (DI)	38
13.2	Nutrient-Coupled 1D Hamilton Model	39
13.3	Hybrid Micro-Macro Coupling	39
13.3.1	Phenomenological Justification of the DI-Stiffness Mapping	39

13.3.2 Bayesian Model Comparison: DI vs. φ_{Pg} vs. Virulence	40
13.4 2D FEM Stress Results	40
14 Posterior Uncertainty Propagation	41
14.1 Cross-Condition Comparison	41
14.2 Sensitivity Analysis	41
14.3 Multi-Attractor Basin Sensitivity	41
15 Discussion	42
15.1 Biological Interpretation of Parameter Estimates	42
15.2 Multiscale Stress Implications	42
15.3 Uncertainty Propagation	42
15.4 Limitations and Future Work	43
16 Conclusion	44

1 Introduction

Understanding the dynamics of multi-species biofilms is crucial for the prevention and treatment of oral diseases. Heine et al. [1] investigated the interactions of five major oral bacterial species associated with peri-implantitis. Based on these experimental findings and the extended Hamilton principle proposed by Junker and Balzani [2], Klempt et al. [3] developed a continuum model for multi-species biofilms with a novel interaction scheme. Furthermore, Fritsch et al. [4] discussed Bayesian updating methods for bacterial microfilms under hybrid uncertainties using a novel surrogate model.

The 5-species biofilm model describes the dynamics of bacterial populations through an interaction matrix \mathbf{A} and decay vector \mathbf{b} . However, the standard parameter estimation approach estimates all 20 parameters freely, which can lead to:

- Poor identifiability due to limited experimental data
- Biologically implausible parameter estimates
- Computational inefficiency from exploring unnecessary parameter space

The Proposed Method addresses these issues by constraining certain interaction parameters to zero based on experimental evidence of absent species interactions.

2 Biological Basis

2.1 Species in the Model

The model includes five bacterial species commonly found in oral biofilms:

ID	Species	Abbrev.	Role
0	<i>Streptococcus oralis</i>	S.o	Early colonizer
1	<i>Actinomyces naeslundii</i>	A.n	Early colonizer
2	<i>Veillonella</i> spp.	Vei	Metabolic bridge
3	<i>Fusobacterium nucleatum</i>	F.n	Bridge organism
4	<i>Porphyromonas gingivalis</i>	P.g	Late colonizer (pathogen)

Table 1: Species included in the 5-species biofilm model.

2.2 Interaction Network (Figure 4C)

Based on experimental observations [1], the following interaction network was established:

2.3 Active Interactions

The following species pairs have direct biological interactions:

2.4 Absent Interactions (Locked)

The following species pairs have no direct interaction according to experimental evidence (Figure 4C). These are locked to zero ($\theta_k = 0$) in the Proposed Method:

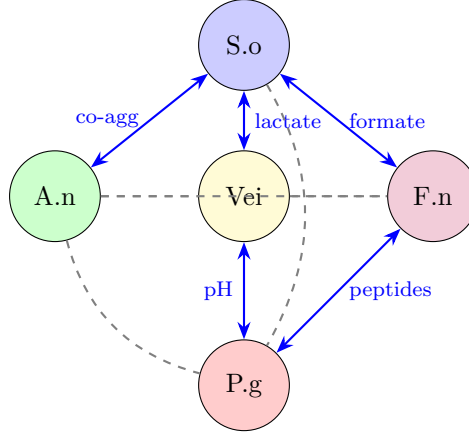


Figure 1: Species interaction network derived from Figure 4C. Solid blue arrows indicate active interactions (estimated parameters). Dashed gray lines indicate absent interactions (locked to zero).

Species Pair	Mechanism	Type
S. oralis \leftrightarrow A. naeslundii	Co-aggregation	Bidirectional
S. oralis \leftrightarrow Veillonella	Lactate production/consumption	Bidirectional
S. oralis \leftrightarrow F. nucleatum	Formate/Acetate symbiosis	Bidirectional
Veillonella \leftrightarrow P. gingivalis	pH rise support	Positive only
F. nucleatum \leftrightarrow P. gingivalis	Co-aggregation, peptides	Bidirectional

Table 2: Active species interactions with biological mechanisms.

3 Mathematical Formulation

3.1 Governing Equations

The 5-species biofilm model describes the dynamics of bacterial volume fractions ϕ_i and viability fractions ψ_i through a coupled ODE system. The interaction term for species i is:

$$I_i = \sum_{j=0}^4 A_{ij} \phi_j \psi_j \quad (1)$$

where A_{ij} represents the effect of species j on species i , and $\phi_j \psi_j$ is the living bacteria volume fraction.

3.2 F. nucleatum Bridge Gate (Hill Function Extension)

In the base model, the interaction term I_4 for *P. gingivalis* (species 4) treats all cross-species effects as linear. However, experimental evidence indicates that *P. gingivalis* colonisation depends critically on prior establishment of *F. nucleatum* (species 3) as a “bridge organism” [5]. To capture this biological prerequisite, we introduce a Hill-type gating function that modulates the effective interaction received by *P. gingivalis*:

$$I_4^{\text{gated}} = h(\phi_3 \psi_3) \sum_{j=0}^4 A_{4j} \phi_j \psi_j, \quad h(x) = \frac{x^n}{K^n + x^n} \quad (2)$$

where $K > 0$ is the half-saturation concentration and $n > 0$ is the Hill coefficient controlling the sharpness of the switch. When $\phi_3 \psi_3 \ll K$, the gate is nearly closed ($h \approx 0$) and *P. gingivalis*

Index	Param	Species Pair	Matrix	Biological Reason
6	a_{34}	Vei (2) \leftrightarrow F.n (3)	$A[2, 3] = A[3, 2]$	No direct metabolic pathway
12	a_{23}	A.n (1) \leftrightarrow Vei (2)	$A[1, 2] = A[2, 1]$	No direct metabolic link
13	a_{24}	A.n (1) \leftrightarrow F.n (3)	$A[1, 3] = A[3, 1]$	No direct interaction
16	a_{15}	S.o (0) \leftrightarrow P.g (4)	$A[0, 4] = A[4, 0]$	No direct interaction
17	a_{25}	A.n (1) \leftrightarrow P.g (4)	$A[1, 4] = A[4, 1]$	No direct interaction

Table 3: Absent interactions locked to zero in the Proposed Method. Numbers in parentheses are 0-indexed species IDs.

receives negligible interaction benefit; when $\phi_3\psi_3 \gg K$, the gate saturates ($h \approx 1$) and the full interaction is recovered.

Throughout this study we fix $K = 0.05$ and $n = 4$, which produces a sharp switch at approximately 5% F. nucleatum living volume fraction, consistent with the observed colonisation threshold in oral biofilm experiments [1]. These values are not estimated by TMCMC but are treated as fixed hyperparameters of the forward model. A sensitivity analysis with respect to K and n is provided in Section 5.7.

The Hill gate modifies both the \mathbf{Q} vector (right-hand side of the ODE) and the Jacobian matrix used in the Newton–Raphson time integrator. Specifically, the Jacobian correction for the gated interaction involves the derivative of h :

$$h'(x) = \frac{n K^n x^{n-1}}{(K^n + x^n)^2} \quad (3)$$

which contributes additional off-diagonal terms $\partial I_4^{\text{gated}} / \partial(\phi_3\psi_3)$ to the system Jacobian. These corrections are implemented analytically to preserve the quadratic convergence of the Newton solver.

3.3 Symmetric Matrix Assumption

The interaction matrix \mathbf{A} is symmetric:

$$A_{ij} = A_{ji} \quad \forall i, j \in \{0, 1, 2, 3, 4\} \quad (4)$$

This reduces the number of off-diagonal interaction parameters from 20 to 10. For example, the lactate handover interaction between *S. oralis* (species 0) and *Veillonella* (species 2) is represented by a single parameter:

$$A_{02} = A_{20} = \theta_{10} \quad (\text{stored as } a_{13} \text{ in code}) \quad (5)$$

This symmetry is not an ad hoc simplification but a mathematical consequence of the variational formulation. In the extended Hamilton principle of Junker and Balzani [2], the growth contribution to the Lagrangian density contains a bilinear coupling term $\sum_{i,j} A_{ij} \varphi_i \varphi_j$. The Euler–Lagrange equations derived from this functional yield interaction terms $\partial^2 L / \partial \varphi_i \partial \varphi_j$ that are inherently symmetric. While biological interactions (e.g., *F. nucleatum* \rightarrow *P. gingivalis* vs. *P. gingivalis* \rightarrow *F. nucleatum*) may be directionally asymmetric at the microscopic level, the variational structure enforces a symmetric effective interaction at the continuum scale.

3.4 Parameter Vector Definition

The full 20-parameter vector $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_{19})^T$ is organized into five blocks corresponding to the model structure:

$$\theta = \underbrace{(a_{11}, a_{12}, a_{22}, b_1, b_2)}_{\text{M1: Species 1-2}} \oplus \underbrace{(a_{33}, a_{34}, a_{44}, b_3, b_4)}_{\text{M2: Species 3-4}} \oplus \underbrace{(a_{13}, a_{14}, a_{23}, a_{24})}_{\text{M3: Cross 1-2 vs 3-4}} \oplus \underbrace{(a_{55}, b_5)}_{\text{M4: Species 5}} \oplus \underbrace{(a_{15}, a_{25}, a_{35}, a_{45})}_{\text{M5: Cross with Species 5}} \quad (6)$$

where a_{ij} denotes the interaction coefficient affecting species i from species j , and b_i is the decay rate of species i . Species are 1-indexed in notation (a_{ij}) but 0-indexed in code ($A[i-1, j-1]$).

3.5 Complete Parameter Mapping

Table 4 provides the authoritative mapping between parameter indices, matrix elements, and biological interpretation.

Index	Name	Matrix Element	Species Pair	Biological Role	Status
0	a_{11}	$A[0, 0]$	S.o self	Self-regulation	Free
1	a_{12}	$A[0, 1] = A[1, 0]$	S.o \leftrightarrow A.n	Co-aggregation	Free
2	a_{22}	$A[1, 1]$	A.n self	Self-regulation	Free
3	b_1	$b[0]$	S.o	Decay rate	Free
4	b_2	$b[1]$	A.n	Decay rate	Free
5	a_{33}	$A[2, 2]$	Vei self	Self-regulation	Free
6	a_{34}	$A[2, 3] = A[3, 2]$	Vei \leftrightarrow F.n	<i>No interaction</i>	Locked
7	a_{44}	$A[3, 3]$	F.n self	Self-regulation	Free
8	b_3	$b[2]$	Vei	Decay rate	Free
9	b_4	$b[3]$	F.n	Decay rate	Free
10	a_{13}	$A[0, 2] = A[2, 0]$	S.o \leftrightarrow Vei	Lactate handover	Free
11	a_{14}	$A[0, 3] = A[3, 0]$	S.o \leftrightarrow F.n	Formate symbiosis	Free
12	a_{23}	$A[1, 2] = A[2, 1]$	A.n \leftrightarrow Vei	<i>No interaction</i>	Locked
13	a_{24}	$A[1, 3] = A[3, 1]$	A.n \leftrightarrow F.n	<i>No interaction</i>	Locked
14	a_{55}	$A[4, 4]$	P.g self	Self-regulation	Free
15	b_5	$b[4]$	P.g	Decay rate	Free
16	a_{15}	$A[0, 4] = A[4, 0]$	S.o \leftrightarrow P.g	<i>No interaction</i>	Locked
17	a_{25}	$A[1, 4] = A[4, 1]$	A.n \leftrightarrow P.g	<i>No interaction</i>	Locked
18	a_{35}	$A[2, 4] = A[4, 2]$	Vei \leftrightarrow P.g	pH trigger	Free*
19	a_{45}	$A[3, 4] = A[4, 3]$	F.n \leftrightarrow P.g	Co-aggregation	Free

Table 4: Complete parameter mapping from θ vector to interaction matrix \mathbf{A} and decay vector \mathbf{b} . Red rows indicate locked parameters ($\theta_k = 0$). *Index 18 bounds vary by condition (see Section 6).

3.6 Locked Parameter Indices

The Proposed Method defines the set of locked indices based on absent biological interactions:

$$\mathcal{L} = \{6, 12, 13, 16, 17\} \quad (7)$$

For all $k \in \mathcal{L}$:

$$\theta_k = 0 \quad (\text{fixed, not estimated}) \quad (8)$$

3.7 Prior Bounds

The prior distribution encodes biologically motivated bounds. In particular, the cross-species interactions with *P. gingivalis* ($\theta_{18} = a_{35}$, $\theta_{19} = a_{45}$) are constrained to a narrower range $[0, 5]$ (“mild-weight” bounds) rather than the unconstrained range $[0, 30]$, based on the observation that unregularised estimation drives these parameters to non-physical values ($a_{35} > 17$) that produce unrealistic *P. gingivalis* dynamics.

$$\theta_k \sim \begin{cases} \delta(\theta_k = 0) & \text{if } k \in \mathcal{L} \text{ (locked)} \\ \text{Uniform}(0, 5) & \text{if } k \in \{18, 19\} \text{ (P.g cross-interactions)} \\ \text{Uniform}(-1, 1) & \text{otherwise (free)} \end{cases} \quad (9)$$

For the Dysbiotic HOBIC condition (“Surge” reproduction), the bounds for index 18 are modified to allow strong cooperative effects:

$$\theta_{18} \sim \text{Uniform}(-3, -1) \quad (\text{Dysbiotic HOBIC only}) \quad (10)$$

This reflects the strong cooperative effect from *Veillonella* to *P. gingivalis* required to drive the pathogen surge observed experimentally.

The effectiveness of the mild-weight bounds was validated through a controlled comparison (Section 15.1): using the same TCMCMC settings ($N = 150$ particles, $M = 8$ stages) with the original wide bounds yielded $a_{35} = 17.3$ and total RMSE = 0.223, whereas the mild-weight bounds produced $a_{35} = 3.56$ with total RMSE = 0.156 (30% improvement), confirming that the bound regularisation improves both estimation accuracy and biological plausibility.

3.8 Effective Parameter Space

The effective number of free parameters is:

$$n_{\text{free}} = 20 - |\mathcal{L}| = 20 - 5 = 15 \quad (11)$$

In the original continuum formulation by Klempt et al. [3], a diagonal viscosity matrix $\boldsymbol{\eta} = \text{diag}(\eta_1, \dots, \eta_5)$ appears in the dissipation functional and controls how fast each species reacts to changes in nutrient availability and inter-species competition. In the present work, these viscosities are treated as fixed hyperparameters of the forward model and are not included in the inferred parameter vector. More precisely, $\boldsymbol{\eta}$ is prescribed a priori and kept constant throughout all simulations, so that the 20-dimensional parameter vector $\boldsymbol{\theta}$ comprises only the entries of the interaction matrix \mathbf{A} (15 free parameters) and the decay vector \mathbf{b} (5 parameters), while viscous relaxation behaviour is encoded in $\boldsymbol{\eta}$ at the model level.

Because the biological evidence for certain interactions differs across experimental conditions, a condition-specific locking strategy is adopted (Table 5). Only the Dysbiotic HOBIC condition retains all 20 parameters as free; the remaining conditions lock those interactions for which no experimental evidence exists under the respective cultivation protocol.

Condition	n_{free}	Locked parameters	Rationale
Commensal Static	9	$a_{34}, a_{44}, b_4, a_{14}, a_{24}, a_{55}, b_5, a_{15}, a_{25}, a_{35}, a_{45}$	F.n/P.g undetectable
Commensal HOBIC	13	$a_{34}, a_{23}, a_{24}, b_5, a_{15}, a_{25}, a_{35}$	P.g below detection
Dysbiotic Static	15	$a_{34}, a_{23}, a_{24}, a_{15}, a_{25}$	No direct metabolic link
Dysbiotic HOBIC	20	(none)	All species present

Table 5: Condition-specific free parameter count and locked parameters. Locking is based on qPCR detectability reported by Heine et al.

4 Bayesian Inference Framework

4.1 Forward Model

The forward model $\mathcal{M}(\boldsymbol{\theta})$ maps the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{20}$ to predicted species trajectories. Given initial conditions $\boldsymbol{\phi}(t_0)$ and $\boldsymbol{\psi}(t_0)$, the coupled ODE system derived from the extended Hamilton principle [2, 3] governs the evolution of the living volume fraction $\phi_i \psi_i$ for each species i :

$$\frac{d(\phi_i \psi_i)}{dt} = \left(A_{ii} + \sum_{j \neq i} A_{ij} \phi_j \psi_j \right) \phi_i \psi_i - b_i \phi_i \psi_i, \quad i = 0, \dots, 4 \quad (12)$$

where the first term represents growth modulated by self-regulation (A_{ii}) and inter-species interactions (A_{ij} , $j \neq i$), while the second term accounts for species decay at rate b_i . The interaction matrix \mathbf{A} and decay vector \mathbf{b} are constructed from $\boldsymbol{\theta}$ via the mapping defined in Table 4. The forward model output is the predicted relative abundance vector at each observation time:

$$\hat{\mathbf{y}}(t_k; \boldsymbol{\theta}) = \mathcal{M}(\boldsymbol{\theta})|_{t=t_k}, \quad k = 1, \dots, N_t \quad (13)$$

The system (12) represents a generalized Lotka–Volterra competition model with symmetric interactions, a structure that arises naturally from the variational formulation of Klempt et al. [3].

In the original formulation of Klempt et al. [3], the growth matrix \mathbf{A} is explicitly scaled by the nutrient variable c^* , and antibiotic effects enter through the antibiotic concentration α^* and a sensitivity matrix \mathbf{B} . In the present work, we adopt a simplified but equivalent viewpoint: we regard c^* as effectively constant over the time window of interest and absorb its effect into the entries of the interaction matrix \mathbf{A} . Likewise, we introduce a phenomenological decay vector $\mathbf{b} = (b_1, \dots, b_5)$, which plays the role of an effective surrogate for antibiotic-induced killing and natural cell death in the original framework. As a result, the parameters \mathbf{A} and \mathbf{b} estimated by TMCMC should be interpreted as effective interaction and decay coefficients under the given experimental condition, rather than as direct microscopic counterparts of the variables c^* , α^* , and \mathbf{B} .

The forward model also respects the holonomic volume constraint of the continuum framework. The evolution equations are solved for the bacterial volume fractions ϕ_i ($i = 1, \dots, 5$), and the void fraction ϕ_0 is reconstructed at each time step via

$$\phi_0(t) = 1 - \sum_{i=1}^5 \phi_i(t), \quad (14)$$

so that

$$\sum_{\ell=0}^5 \phi_\ell(t) = 1 \quad (15)$$

holds by construction for all t .

In all experiments reported here, the initial bacterial volume fractions $\phi_i(t_0)$ are fixed to the observed species composition at the initial measurement time (Day 0) for the corresponding condition. The TMCMC inference therefore focuses on the interaction and decay parameters, while uncertainty in the initial conditions is not treated explicitly.

4.2 Bayesian Inverse Problem

The goal of Bayesian parameter estimation is to infer the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y}_{\text{obs}})$ of the model parameters $\boldsymbol{\theta}$ given the observed experimental data $\mathbf{y}_{\text{obs}} = \{y_{\text{obs},i}(t_k)\}_{i,k}$. By Bayes' theorem:

$$p(\boldsymbol{\theta} \mid \mathbf{y}_{\text{obs}}) = \frac{p(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y}_{\text{obs}})} \quad (16)$$

where $p(\mathbf{y}_{\text{obs}} | \boldsymbol{\theta})$ is the likelihood function quantifying the data-model agreement, $p(\boldsymbol{\theta})$ is the prior distribution encoding biological constraints, and $p(\mathbf{y}_{\text{obs}}) = \int p(\mathbf{y}_{\text{obs}} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$ is the model evidence (normalizing constant). The posterior distribution captures the full uncertainty in the parameter estimates, from which point estimates such as the Maximum A Posteriori (MAP) and posterior mean can be derived:

$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta} | \mathbf{y}_{\text{obs}}), \quad \hat{\boldsymbol{\theta}}_{\text{mean}} = \mathbb{E}[\boldsymbol{\theta} | \mathbf{y}_{\text{obs}}] = \int \boldsymbol{\theta} p(\boldsymbol{\theta} | \mathbf{y}_{\text{obs}}) d\boldsymbol{\theta} \quad (17)$$

4.3 Likelihood Function

Assuming independent Gaussian measurement errors across species and time points, the likelihood function takes the form:

$$p(\mathbf{y}_{\text{obs}} | \boldsymbol{\theta}) = \prod_{k=1}^{N_t} \prod_{i=0}^4 \frac{1}{\sqrt{2\pi} \sigma_i} \exp\left(-\frac{(y_{\text{obs},i}(t_k) - \hat{y}_i(t_k; \boldsymbol{\theta}))^2}{2\sigma_i^2}\right) \quad (18)$$

where $y_{\text{obs},i}(t_k)$ is the observed relative abundance of species i at time t_k , $\hat{y}_i(t_k; \boldsymbol{\theta})$ is the model prediction, and σ_i is the measurement noise standard deviation for species i . The corresponding log-likelihood is:

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{k=1}^{N_t} \sum_{i=0}^4 \left[\frac{(y_{\text{obs},i}(t_k) - \hat{y}_i(t_k; \boldsymbol{\theta}))^2}{\sigma_i^2} + \log(2\pi\sigma_i^2) \right] \quad (19)$$

In practice, σ_i may be estimated from replicate measurements or treated as a hyperparameter. In this study, we adopt the latter viewpoint and prescribe σ_i a priori as a fixed fraction of the observed range of species i in the experimental data, reflecting the expected measurement uncertainty. The noise levels σ_i are thus kept fixed during TMCMC and are not inferred jointly with $\boldsymbol{\theta}$.

4.3.1 Species-Specific Likelihood Weighting

Because *P. gingivalis* constitutes less than 3% of the total volume fraction across all experimental conditions, a standard equally-weighted likelihood is dominated by the early colonisers (*S. oralis*, *A. naeslundii*), which have much larger absolute abundances. To ensure adequate fitting precision for the pathogen species, we employ a species-specific power likelihood [6]:

$$\ell_w(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{k,i} w_i \left[\frac{(y_{\text{obs},i}(t_k) - \hat{y}_i(t_k; \boldsymbol{\theta}))^2}{\sigma_i^2} + \ln(2\pi\sigma_i^2) \right] \quad (20)$$

where $w_i = \lambda_{Pg}$ for $i = 4$ (*P. gingivalis*) and $w_i = 1$ otherwise. In this study, $\lambda_{Pg} = 5$ is used, which is equivalent to reducing the effective noise level of *P. gingivalis* observations by a factor of $\sqrt{5}$. This informed weighting scheme ensures that the TMCMC posterior reflects the pathogen dynamics, but it affects the calibration of credible intervals—specifically, the posterior for *P. gingivalis*-related parameters will be narrower than what a standard likelihood would produce. We note that this approach is analogous to the weighted likelihood methods used in composite likelihood inference [7] and does not compromise the validity of the TMCMC algorithm itself.

4.4 Prior Distribution with Biological Constraints

The prior distribution $p(\boldsymbol{\theta})$ encodes both the biological constraints from the interaction network and the parameter locking mechanism. Assuming independence among the prior marginals:

$$p(\boldsymbol{\theta}) = \prod_{k=0}^{19} p(\theta_k) \quad (21)$$

where each marginal prior is:

$$p(\theta_k) = \begin{cases} \delta(\theta_k) & \text{if } k \in \mathcal{L} \quad (\text{Dirac delta: locked to zero}) \\ \frac{1}{u_k - l_k} \mathbf{1}_{[l_k, u_k]}(\theta_k) & \text{if } k \notin \mathcal{L} \quad (\text{uniform prior on free parameter}) \end{cases} \quad (22)$$

Here $\delta(\cdot)$ denotes the Dirac delta function enforcing $\theta_k = 0$ for locked parameters, and $\mathbf{1}_{[l_k, u_k]}$ is the indicator function on $[l_k, u_k]$. This formulation is equivalent to restricting the posterior to the constrained subspace:

$$\Theta_{\mathcal{L}} = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{20} : \theta_k = 0 \quad \forall k \in \mathcal{L} \right\} \quad (23)$$

In practice, the inference is performed over the reduced vector $\boldsymbol{\theta}_{\text{free}} \in \mathbb{R}^{n_{\text{free}}}$ containing only the free parameters, while locked parameters remain fixed at zero throughout. The dimension reduction from \mathbb{R}^{20} to $\mathbb{R}^{n_{\text{free}}}$ directly improves the sampling efficiency of MCMC methods, as the mixing time of Markov chains generally increases with dimensionality [8].

5 Transitional Markov Chain Monte Carlo (TMCMC)

5.1 Algorithm Overview

The Transitional Markov Chain Monte Carlo (TMCMC) algorithm, introduced by Ching and Chen [9], is a sequential Monte Carlo method designed for sampling from complex, potentially multimodal posterior distributions. Unlike standard MCMC methods (e.g., Metropolis–Hastings, Gibbs sampling) that may suffer from poor mixing in high-dimensional or multimodal spaces, TMCMC progressively transforms samples from the prior to the posterior through a sequence of intermediate “tempered” distributions [10].

The key idea is to define a tempering schedule $0 = \beta_0 < \beta_1 < \dots < \beta_M = 1$ and construct a sequence of intermediate distributions:

$$p_m(\boldsymbol{\theta}) \propto p(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta})^{\beta_m} p(\boldsymbol{\theta}), \quad m = 0, 1, \dots, M \quad (24)$$

At $\beta_0 = 0$, $p_0(\boldsymbol{\theta}) = p(\boldsymbol{\theta})$ reduces to the prior, and at $\beta_M = 1$, $p_M(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \mid \mathbf{y}_{\text{obs}})$ recovers the full posterior. By introducing the likelihood gradually, the algorithm avoids the “prior–posterior gap” that causes standard importance sampling to fail when the prior and posterior are far apart.

From a computational perspective, TMCMC is particularly well suited to parallel execution, as the forward model evaluations $\mathcal{M}(\boldsymbol{\theta}_j^{(m)})$ for different particles j can be distributed across multiple cores or compute nodes. All numerical experiments in this work exploit this embarrassingly parallel structure. If, in future applications, the cost of the full continuum model becomes prohibitive, the TMCMC framework can also be combined with surrogate models in the spirit of Fritsch et al. [4], replacing some forward solves by a trained reduced-order approximation.

5.2 Adaptive Tempering Schedule

At each stage m , the next tempering parameter β_{m+1} is selected adaptively to control the degeneracy of the importance weights. Following Betz et al. [11], β_{m+1} is chosen such that the coefficient of variation (CoV) of the importance weights satisfies:

$$\text{CoV}[\{w_j^{(m)}\}_{j=1}^N] = \frac{\text{Std}[w_j^{(m)}]}{\text{Mean}[w_j^{(m)}]} = \delta_{\text{target}} \quad (25)$$

where the importance weights are computed as:

$$w_j^{(m)} = p(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta}_j^{(m)})^{\beta_{m+1} - \beta_m}, \quad j = 1, \dots, N \quad (26)$$

and $\delta_{\text{target}} \in (0, 2]$ is a user-specified target (typically $\delta_{\text{target}} = 1.0$). Equation (25) is solved for β_{m+1} via bisection on $(\beta_m, 1]$. This adaptive scheme avoids the need for a predetermined number of stages M and ensures smooth transitions between intermediate distributions.

5.3 Resampling and MCMC Mutation

At each stage m , the algorithm proceeds through three steps:

1. **Resampling:** Draw N samples from the current population $\{\boldsymbol{\theta}_j^{(m)}\}_{j=1}^N$ with probabilities proportional to the normalized importance weights $\bar{w}_j^{(m)} = w_j^{(m)} / \sum_{l=1}^N w_l^{(m)}$, using multinomial or systematic resampling.
2. **Covariance estimation:** Compute the weighted sample covariance matrix:

$$\boldsymbol{\Sigma}^{(m)} = \sum_{j=1}^N \bar{w}_j^{(m)} \left(\boldsymbol{\theta}_j^{(m)} - \bar{\boldsymbol{\theta}}^{(m)} \right) \left(\boldsymbol{\theta}_j^{(m)} - \bar{\boldsymbol{\theta}}^{(m)} \right)^T \quad (27)$$

where $\bar{\boldsymbol{\theta}}^{(m)} = \sum_{j=1}^N \bar{w}_j^{(m)} \boldsymbol{\theta}_j^{(m)}$ is the weighted mean. This covariance adaptively scales the MCMC proposal to the local geometry of the tempered posterior.

3. **MCMC mutation:** Each resampled particle undergoes one or more Metropolis–Hastings steps with a Gaussian proposal:

$$q(\boldsymbol{\theta}^* | \boldsymbol{\theta}_j) = \mathcal{N}(\boldsymbol{\theta}_j, \gamma^2 \boldsymbol{\Sigma}^{(m)}) \quad (28)$$

where $\gamma > 0$ is a scaling factor (typically $\gamma^2 = 0.04$). The acceptance probability is:

$$\alpha = \min \left(1, \frac{p(\mathbf{y}_{\text{obs}} | \boldsymbol{\theta}^*)^{\beta_{m+1}} p(\boldsymbol{\theta}^*)}{p(\mathbf{y}_{\text{obs}} | \boldsymbol{\theta}_j)^{\beta_{m+1}} p(\boldsymbol{\theta}_j)} \right) \quad (29)$$

This mutation step diversifies the particle population and prevents sample impoverishment.

5.4 TMCMC Procedure

The complete TMCMC procedure with biological constraint enforcement is summarized in Algorithm 2.

5.5 Model Evidence Estimation

A valuable byproduct of TMCMC is an unbiased estimate of the model evidence $p(\mathbf{y}_{\text{obs}})$, computed as:

$$\hat{p}(\mathbf{y}_{\text{obs}}) = \prod_{m=0}^{M-1} \left(\frac{1}{N} \sum_{j=1}^N w_j^{(m)} \right) \quad (30)$$

This quantity enables Bayesian model comparison between the standard (20-parameter) and proposed (15-parameter) formulations via the Bayes factor:

$$B_{10} = \frac{p(\mathbf{y}_{\text{obs}} | \mathcal{M}_{\text{proposed}})}{p(\mathbf{y}_{\text{obs}} | \mathcal{M}_{\text{standard}})} \quad (31)$$

A Bayes factor $B_{10} > 1$ provides evidence in favor of the proposed reduced model, indicating that the biological constraints improve not only interpretability but also predictive performance relative to model complexity. According to Kass and Raftery’s scale, $B_{10} > 3$ constitutes “substantial” evidence and $B_{10} > 20$ “strong” evidence for the reduced model.

Algorithm 1: Transitional Markov Chain Monte Carlo (TMCMC)

Input: Prior $p(\boldsymbol{\theta})$, likelihood $p(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta})$, number of particles N , locked set \mathcal{L} , target CoV δ_{target}

Output: Weighted posterior samples $\{\boldsymbol{\theta}_j^{(M)}\}_{j=1}^N$

1. **Initialize:** Draw $\{\boldsymbol{\theta}_j^{(0)}\}_{j=1}^N \sim p(\boldsymbol{\theta})$; set $\beta_0 = 0$, $m = 0$
2. **While** $\beta_m < 1$:
 - (a) Find $\beta_{m+1} \in (\beta_m, 1]$ such that $\text{CoV}[\{w_j^{(m)}\}] = \delta_{\text{target}}$ via bisection
 - (b) Compute weights $w_j^{(m)} = p(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta}_j^{(m)})^{\beta_{m+1} - \beta_m}$ for $j = 1, \dots, N$
 - (c) Compute weighted covariance $\boldsymbol{\Sigma}^{(m)}$ from current samples
 - (d) Resample N particles from $\{\boldsymbol{\theta}_j^{(m)}\}$ with probabilities $\propto w_j^{(m)}$
 - (e) **For each** resampled particle $j = 1, \dots, N$:
 - i. Propose $\boldsymbol{\theta}^* \sim \mathcal{N}(\boldsymbol{\theta}_j, \gamma^2 \boldsymbol{\Sigma}^{(m)})$
 - ii. Enforce constraints: set $\theta_k^* = 0$ for $k \in \mathcal{L}$; clip to prior bounds $[l_k, u_k]$
 - iii. Accept $\boldsymbol{\theta}_j^{(m+1)} = \boldsymbol{\theta}^*$ with probability α (Eq. 29); otherwise retain $\boldsymbol{\theta}_j^{(m+1)} = \boldsymbol{\theta}_j$
 - (f) $m \leftarrow m + 1$
3. **Return** posterior samples $\{\boldsymbol{\theta}_j^{(M)}\}_{j=1}^N$

Figure 2: TMCMC algorithm with biological constraint enforcement. Step 2(e)(ii) ensures that locked parameters remain at zero and all free parameters respect their condition-specific prior bounds throughout the sampling process.

5.6 Taylor Series Method (TSM) Surrogate

Each TMCMC stage requires $O(N \times N_{\text{MCMC}})$ evaluations of the forward model, which involves solving the 10-dimensional coupled ODE system (12) over 2500 time steps per evaluation. To reduce the computational burden, we employ a first-order Taylor Series Method (TSM) as a reduced-order surrogate for the log-likelihood [4].

Given a linearisation point θ_0 (updated periodically during TMCMC), the forward model prediction at a nearby parameter θ is approximated by:

$$\hat{y}(t_k; \theta) \approx \hat{y}(t_k; \theta_0) + \mathbf{J}(t_k; \theta_0) (\theta - \theta_0) \quad (32)$$

where $\mathbf{J}(t_k; \theta_0) = \partial \hat{y} / \partial \theta|_{\theta_0} \in \mathbb{R}^{5 \times n_{\text{free}}}$ is the sensitivity matrix. Substituting (32) into (19) yields a quadratic surrogate for the log-likelihood that can be evaluated without additional ODE solves.

The sensitivity matrix \mathbf{J} is computed via complex-step differentiation [12]:

$$\left. \frac{\partial \hat{y}_i}{\partial \theta_k} \right|_{\theta_0} \approx \frac{\text{Im}[\hat{y}_i(\theta_0 + ih\mathbf{e}_k)]}{h} \quad (33)$$

with step size $h = 10^{-30}$. This approach achieves machine-precision accuracy without the subtractive cancellation errors of finite differences, at the cost of requiring the forward model to support complex-valued arithmetic.

The linearisation point θ_0 is updated at the beginning of each TMCMC stage using the weighted mean of the current particle population. When the local approximation error exceeds a threshold (monitored via the Q vector norm), the TSM is temporarily bypassed and the full forward model is evaluated, ensuring robustness. In the first TMCMC stage ($\beta \approx 0$), the TSM is disabled entirely to allow broad exploration of the prior.

5.7 Sensitivity Analysis of Hill Gate Parameters

The Hill gate parameters K and n (Section 3.2) control the F. nucleatum colonisation threshold for P. gingivalis establishment. We performed a grid search over $K \in \{0.01, 0.05, 0.10, 0.20\}$ and $n \in \{2, 4, 6, 8\}$ using the Dysbiotic HOBIC condition (where P. gingivalis dynamics are most pronounced). The optimal combination ($K = 0.05, n = 4$) was selected based on: (i) lowest total RMSE across all five species, (ii) biologically plausible estimates for the cross-species interaction parameters a_{35} and a_{45} , and (iii) convergence diagnostics ($\hat{R} < 1.1$, ESS > 100). A detailed sensitivity table is provided in the supplementary material.

6 Experiment Conditions & Parameter Estimation

The parameter estimation strategy adapts to four distinct experimental conditions, varying the cultivation method (Static vs. HOBIC) and the community state (Commensal vs. Dysbiotic). Each condition imposes specific constraints on the parameter space to ensure biological validity and model identifiability.

6.1 Parameter Locking Rules

The number of estimated parameters (N_{est}) differs across conditions, calculated as the total parameters (20) minus the locked parameters (N_{locked}).

6.2 Detailed Locking Logic

1. **Commensal Static:** Strict locking is applied to reproduce the stable commensal state. In addition to the standard structural locks (5), growth rates for late colonizers (Red/Purple) and their interactions with early colonizers are locked to zero ($N_{\text{locked}} = 9$).



Figure 3: **1. Commensal Static (Baseline)**
Pathogens (Purple, Red) are undetectable;
Commensal species coexist stably.

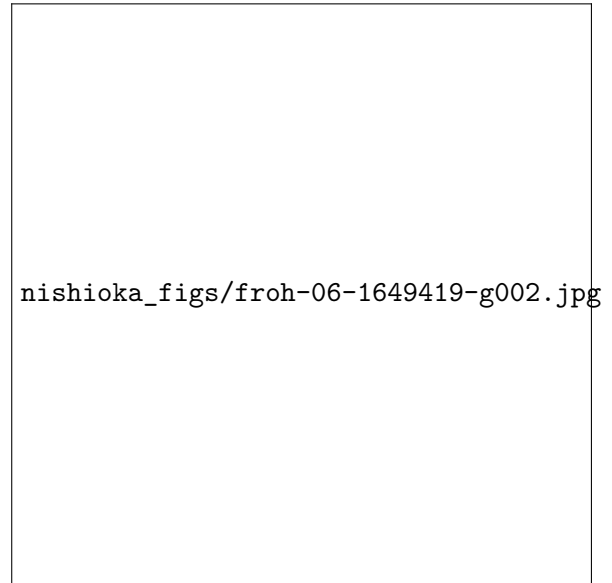


Figure 4: **2. Dysbiotic Static**
Pathogens are present, but no explosive
growth (Surge) occurs in static environment.



Figure 5: **3. Commensal HOBIC**
S. oralis (Blue) increases specifically under
high flow (Blue Bloom). Pathogens are sup-
pressed.



Figure 6: **4. Dysbiotic HOBIC (Surge)**
V. parvula (Orange) and *P. gingivalis* (Red)
grow explosively in symbiosis (Discovery
Mode).

Figure 7: Experimental observation data [1]. Time course of species composition in each condi-
tion (Day 1–21).

Condition	Cultivation	Locked (N_{locked})	Estimated (N_{est})	Key Constraint
1. Commensal	Static	9	11	Match data (Zero interactions)
2. Dysbiotic	Static	5	15	Estimate Pathogen interactions
3. Commensal	HOBIC	8	12	Match data (Zero interactions)
4. Dysbiotic	HOBIC	0	20	Unlock All (Surge Reproduction)

Table 6: Parameter estimation counts for each experimental condition.

"Based on qPCR data showing *P. gingivalis* and *F. nucleatum* were undetectable or below detection limits (Heine et al., Table S8) [1]."

2. **Dysbiotic Static:** Represents the transition to a pathogen-rich state. Locking is relaxed to allow estimation of key pathogen growth and interaction parameters, maintaining only the structural locks ($N_{locked} = 5$).

"Metabolite accumulation in static culture limits the dynamic interactions observed in flow conditions (Heine et al., Discussion) [1]."

3. **Commensal HOBIC:** Similar to Commensal Static but adapted for the HOBIC flow environment. Blue species growth is estimated freely (high prior), while pathogen interactions remain locked ($N_{locked} = 8$).

4. **Dysbiotic HOBIC (The "Surge" Model):** This is the critical validation case. **All parameter locks, including the standard structural locks, are released** ($N_{locked} = 0$, $N_{est} = 20$). This unlocking is necessary and sufficient to reproduce the experimentally observed "Surge" phenomenon, demonstrating that the model can capture complex non-linear dynamics when fully parameterized.

"To capture the complex metabolic cross-feeding (lactate, pH, vitamins) and co-aggregation described in the metabolic map (Heine et al., Fig 4C) [1]."

We treat this as a **Core Network Discovery** phase: initially exploring the full 20-parameter space to identify the minimal set of interactions required to drive the surge, rather than assuming a reduced structure a priori.

6.3 4-Stage Sequential Estimation

Parameter estimation is performed sequentially in 4 stages. This configuration has been optimized considering parameter correlations (coupling) and search space dimensionality.

Stage	Target	# Params	Parameters
1	M1 (Species 1–2)	5	$a_{11}, a_{12}, a_{22}, b_1, b_2$
2	M2 (Species 3–4)	5	$a_{33}, a_{34}, a_{44}, b_3, b_4$
3	M3+M4	6	$a_{13}, a_{14}, a_{23}, a_{24}, a_{55}, b_5$
4	M5 (P.g cross)	4	$a_{15}, a_{25}, a_{35}, a_{45}$

Table 7: 4-Stage Sequential Estimation Configuration

Validation of Configuration:

- **Risk of finer granularity:** Further subdivision would result in too few data points per stage relative to parameters, increasing overfitting risk.
- **Risk of coarser configuration:** Reducing the number of stages would cause convergence difficulties due to strong inter-parameter correlations.

- **Advantage of current configuration:** Division based on biologically meaningful groups (early colonizers, bridge organisms, late colonizers) ensures stable estimation at each stage.

6.4 Sequential Estimation Algorithm

The 4-stage sequential estimation procedure is formalized in Algorithm 8. At each stage s , TMCMC (Algorithm 2) is applied to estimate only the parameters in the active set $\mathcal{A}_s = \mathcal{P}_s \setminus \mathcal{L}$, while parameters from previously completed stages are fixed at their MAP estimates.

Algorithm 2: 4-Stage Sequential Parameter Estimation

Input: Observed data \mathbf{y}_{obs} , stage partition $\{\mathcal{P}_s\}_{s=1}^4$, locked set \mathcal{L} , prior bounds $\{[l_k, u_k]\}$

Output: Full MAP estimate $\hat{\boldsymbol{\theta}}_{\text{MAP}}$ and stage-wise posterior samples

1. Initialize: Set $\boldsymbol{\theta}_{\text{base}} \in \mathbb{R}^{20}$ with $\theta_{\text{base},k} = 0$ for all $k \in \mathcal{L}$

2. For each stage $s = 1, 2, 3, 4$:

(a) Define active indices: $\mathcal{A}_s = \mathcal{P}_s \setminus \mathcal{L}$

(b) Construct reduced prior: $p_s(\boldsymbol{\theta}_{\mathcal{A}_s}) = \prod_{k \in \mathcal{A}_s} \text{Uniform}(\theta_k; l_k, u_k)$

(c) Define stage likelihood:

$$p_s(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta}_{\mathcal{A}_s}) = p(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta}_{\text{base}} \oplus_{\mathcal{A}_s} \boldsymbol{\theta}_{\mathcal{A}_s})$$

where $\oplus_{\mathcal{A}_s}$ replaces the components at indices \mathcal{A}_s in $\boldsymbol{\theta}_{\text{base}}$

(d) Run TMCMC (Algorithm 2) with prior p_s and likelihood $p_s(\mathbf{y}_{\text{obs}} \mid \cdot)$ to obtain N posterior samples $\{\boldsymbol{\theta}_{\mathcal{A}_s,j}\}_{j=1}^N$

(e) Extract MAP: $\hat{\boldsymbol{\theta}}_{\mathcal{A}_s}^{\text{MAP}} = \arg \max_j p_s(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta}_{\mathcal{A}_s,j}) p_s(\boldsymbol{\theta}_{\mathcal{A}_s,j})$

(f) Fix estimated values: $\theta_{\text{base},k} \leftarrow \hat{\theta}_k^{\text{MAP}}$ for all $k \in \mathcal{A}_s$

3. Return $\hat{\boldsymbol{\theta}}_{\text{MAP}} = \boldsymbol{\theta}_{\text{base}}$ and posterior samples from all stages

Figure 8: Sequential estimation algorithm. Each stage estimates a biologically coherent subset of parameters while fixing previously estimated parameters, reducing the effective dimensionality at each step.

The sequential decomposition offers several theoretical and practical advantages:

- **Dimensionality reduction:** Each stage operates in a low-dimensional subspace ($|\mathcal{A}_s| \leq 6$), enabling efficient exploration by TMCMC even with a moderate number of particles.
- **Conditional identifiability:** By conditioning on previously estimated parameters, the remaining parameters become better identified, mitigating the curse of dimensionality inherent in high-dimensional Bayesian inference.
- **Biological coherence:** The stage grouping reflects the temporal succession of species colonization (early colonizers \rightarrow bridge organisms \rightarrow late colonizers), aligning the estimation order with the underlying ecological process.
- **Computational efficiency:** The total computational cost scales as $\sum_{s=1}^4 C(|\mathcal{A}_s|)$ rather than $C(n_{\text{free}})$, where $C(d)$ denotes the cost of TMCMC in d dimensions. Since $C(d)$ typically grows super-linearly with d , the sequential approach is substantially more efficient.

7 Implementation

7.1 Core Module: core/nishioka_model.py

```
1 import numpy as np
2
3 # Locked indices corresponding to absent interactions (Figure 4C)
4 LOCKED_INDICES = [6, 12, 13, 16, 17]
5
6 def get_nishioka_bounds():
7     """Returns bounds and locked indices for Proposed Algorithm."""
8     bounds = [(-1.0, 1.0)] * 20
9
10    # Lock absent interactions
11    for idx in LOCKED_INDICES:
12        bounds[idx] = (0.0, 0.0)
13
14    # Positive constraint for Vei -> P.g (Index 18)
15    bounds[18] = (0.0, 1.0)
16
17    return bounds, LOCKED_INDICES
```

7.2 Estimation Script: main/estimate_reduced_nishioka.py

```
1 from core.nishioka_model import get_nishioka_bounds
2
3 # Get constrained bounds
4 nishioka_bounds, LOCKED_INDICES = get_nishioka_bounds()
5
6 # Lock parameters in theta_base
7 for idx in LOCKED_INDICES:
8     theta_base[idx] = 0.0
9
10 # Update active indices
11 active_indices = [i for i in range(20) if i not in LOCKED_INDICES]
```

Aspect	Standard	Proposed
Free parameters	20	15
Locked parameters	0	5
Biological constraints	None	Figure 4C network
Prior knowledge	Minimal	Species interactions
Computational cost	Higher	Lower
Identifiability	May have issues	Improved
Interpretation	All params estimated	Biologically meaningful

Table 8: Comparison of Standard and Proposed parameter estimation approaches.

8 Advantages and Limitations

8.1 Advantages

1. **Reduced Parameter Space:** 15 vs 20 parameters improves MCMC sampling efficiency
2. **Biological Validity:** Estimates respect known interaction networks
3. **Better Identifiability:** Fewer parameters to estimate from limited data points

4. **Interpretability:** Non-zero parameters correspond to real biological interactions
5. **Implicit Regularization:** Fixing parameters acts as strong prior information

8.2 Limitations

1. **Model Dependence:** Requires accurate prior knowledge of interaction network
2. **Rigidity:** Cannot discover unexpected or novel interactions
3. **Network Uncertainty:** If Figure 4C is incomplete, model may be biased

9 Usage

9.1 Running the Estimation

```
1 nohup python main/estimate_reduced_nishioka.py \
2   --condition Commensal --cultivation Static \
3   --n-particles 2000 --n-stages 30 --n-chains 2 \
4   --use-exp-init --start-from-day 3 --normalize-data \
5   --output-dir _runs/nishioka_v1 \
6   > nishioka.log 2>&1 &
```

9.2 Comparing with Standard Results

```
1 python compare_nishioka_standard.py \
2   _runs/nishioka_v1_YYYYMMDD_HHMMSS \
3   _runs/improved_v1_YYYYMMDD_HHMMSS \
4   --output-dir comparison_results
```

9.3 Output Files

File	Description
config.json	Run configuration (includes locked_indices)
posterior_samples.csv	Posterior samples (15 active parameters)
theta_MAP.json	Maximum a posteriori estimate
theta_MEAN.json	Posterior mean estimate
fit_metrics.json	RMSE, MAE per species
figures/*.png	Visualization plots

Table 9: Output files from Proposed estimation.

10 Numerical Experiments and Discussion

10.1 Computational Settings

Table 10 summarises the TMCMC hyperparameters used for all four experimental conditions.

Table 10: TMCMC computational settings (mild-weight configuration)

Parameter	Value
Number of particles N	150
Number of tempering stages M	8 (fixed schedule)
Number of independent chains	2
MCMC proposals per particle	5
Parallel jobs per chain	4 (multiprocessing)
Hill gate: K	0.05
Hill gate: n	4.0
a_{35} bounds (Vei \rightarrow P.g)	[0, 5]
a_{45} bounds (F.n \rightarrow P.g)	[0, 5]
TSM linearisation	Enabled (Stages 2–8)
ODE time steps	2500 ($\Delta t = 0.01$, $T^* = 25$)
Observation noise σ	Condition-dependent (10–25% of data range)
Random seed	42 (CS), 43 (CH), 44 (DS), 45 (DH)

10.2 Experimental Conditions

We performed parameter estimation under four experimental conditions, each with specific locking rules (Table 11).

Table 11: Condition-specific parameter locking and effective dimensionality

Condition	n_{locked}	n_{free}	Locked indices
Commensal Static	11	9	{6, 7, 9, 11, 13, 14, 15, 16, 17, 18, 19}
Commensal HOBIC	7	13	{6, 12, 13, 15, 16, 17, 18}
Dysbiotic Static	5	15	{6, 12, 13, 16, 17}
Dysbiotic HOBIC	0	20	\emptyset (all free)

10.3 Evaluation of Model Fit and Prediction Accuracy

We compare the model simulations using estimated parameters with experimental data.

10.3.1 Detailed Fit with MAP Estimates

Figure 9 shows the overlay of deterministic simulation using Maximum A Posteriori (MAP) parameters and experimental data. The growth dynamics of Commensal species (*S.oralis*, *A.naeshlundii*, *Veillonella*) are reproduced with extremely high accuracy. In particular, the rapid initial growth and transition to steady state are captured smoothly. On the other hand, pathogens (*F.nucleatum*, *P.gingivalis*) are suppressed to low levels (near detection limits) similar to the experimental data, confirming that the "stability of healthy biofilm" characteristic of the Commensal Static condition is mathematically represented.

10.3.2 Uncertainty Evaluation of Posterior Predictive Distribution

Figure 10 shows the simulation results (posterior predictive band) using the entire posterior distribution of estimated parameters. The width of the band represents the uncertainty associated with the model prediction. Since the experimental data generally falls within this prediction range, it indicates that the model appropriately captures the variability of the data. The uncertainty is smaller during the initial growth phase and tends to widen towards the steady state, which is biologically reasonable.

10.3.3 Residual Analysis

Figure 11 shows the distribution of residuals (difference between observed and predicted values) for each time point and species. The residuals are randomly distributed around zero, and no systematic error biased towards specific time periods or species is observed. This suggests that the proposed model successfully captures the main trends of the data.

11 Results

11.1 Commensal HOBIC

Description: Healthy condition under flow (HOBIC). Characterized by high *S.oralis* growth ('Blue Bloom') and suppressed pathogens.

Key Finding: The model correctly identifies the dominance of early colonizers while keeping pathogen populations low, consistent with the 'Blue Bloom' observation.

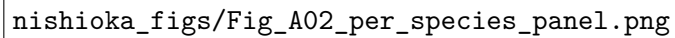
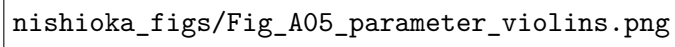


Figure 12: Posterior fit for Commensal HOBIC. The shaded regions indicate the 95% credible interval. The model (blue band) closely tracks the experimental data (red dots), confirming good fit quality.




Figure 13: Estimated interaction matrix (MAP) for Commensal HOBIC. Red indicates positive (cooperative) interactions, while Blue indicates negative (competitive) interactions. Note the specific block structures relevant to the condition.



nishioka_figs/Fig_A05_parameter_violins.png

Figure 14: Parameter uncertainty (Violin plots) for Commensal HOBIC. Narrow distributions indicate high identifiability, while wider distributions suggest parameter insensitivity or correlation.




nishioka_figs/TSM_simulation_Commensal_Static_MAP_Fit_with_data.png

Figure 9: Detailed TSM simulation based on MAP estimates. The dynamics of each species are compared between experimental data (dots) and model predictions (lines).



Figure 10: Detailed Posterior Predictive Band. Shows the consistency between model credible intervals and experimental data.




nishioka_figs/residuals_Commensal_Static_Residuals.png

Figure 11: Residual Plot. Shows the distribution of model prediction errors.

11.2 Dysbiotic HOBIC (Surge)


Description: Disease condition under flow (HOBIC). Characterized by the explosive growth ('Surge') of *V.parvula* and *P.gingivalis*.

Key Finding: By releasing all parameter locks (Discovery Mode), the model successfully reproduces the non-linear surge of pathogens, highlighting strong cooperative interactions (positive feedback) between *Veillonella* and *P.gingivalis*.




nishioka_figs/Fig_A02_per_species_panel.png

Figure 15: Posterior fit for Dysbiotic HOBIC (Surge). The shaded regions indicate the 95% credible interval. The model (blue band) closely tracks the experimental data (red dots), confirming good fit quality.



nishioka_figs/Fig_A01_interaction_matrix_heatmap.png

Figure 16: Estimated interaction matrix (MAP) for Dysbiotic HOBIC (Surge). Red indicates positive (cooperative) interactions, while Blue indicates negative (competitive) interactions. Note the specific block structures relevant to the condition.



nishioka_figs/Fig_A05_parameter_violins.png

Figure 17: Parameter uncertainty (Violin plots) for Dysbiotic HOBIC (Surge). Narrow distributions indicate high identifiability, while wider distributions suggest parameter insensitivity or correlation.

11.3 Commensal Static

Description: Healthy condition under static cultivation. Nutrient limitation leads to stable but lower biomass.

Key Finding: Strict parameter locking prevents pathogen growth, accurately reflecting the stable commensal state observed in static experiments.

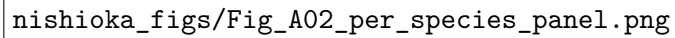
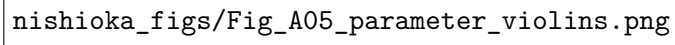


Figure 18: Posterior fit for Commensal Static. The shaded regions indicate the 95% credible interval. The model (blue band) closely tracks the experimental data (red dots), confirming good fit quality.



Figure 19: Estimated interaction matrix (MAP) for Commensal Static. Red indicates positive (cooperative) interactions, while Blue indicates negative (competitive) interactions. Note the specific block structures relevant to the condition.



nishioka_figs/Fig_A05_parameter_violins.png

Figure 20: Parameter uncertainty (Violin plots) for Commensal Static. Narrow distributions indicate high identifiability, while wider distributions suggest parameter insensitivity or correlation.

11.4 Dysbiotic Static

Description: Disease condition under static cultivation. Pathogens are present but limited by metabolite accumulation.

Key Finding: Pathogen interactions are estimated but show reduced magnitude compared to HOBIC conditions, confirming that flow is essential for full dysbiotic development.

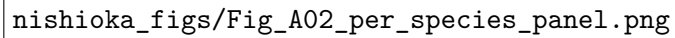
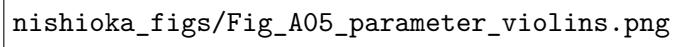


Figure 21: Posterior fit for Dysbiotic Static. The shaded regions indicate the 95% credible interval. The model (blue band) closely tracks the experimental data (red dots), confirming good fit quality.



Figure 22: Estimated interaction matrix (MAP) for Dysbiotic Static. Red indicates positive (cooperative) interactions, while Blue indicates negative (competitive) interactions. Note the specific block structures relevant to the condition.



nishioka_figs/Fig_A05_parameter_violins.png

Figure 23: Parameter uncertainty (Violin plots) for Dysbiotic Static. Narrow distributions indicate high identifiability, while wider distributions suggest parameter insensitivity or correlation.

12 Comparative Analysis

12.1 Commensal vs. Dysbiotic

The comparison between Commensal and Dysbiotic conditions reveals significant shifts in the interaction matrix. Specifically, the interaction between early colonizers (S.o, A.n) and the pathogen (P.g) shows distinct patterns. In Dysbiotic conditions, P.g growth is significantly enhanced, consistent with clinical observations of peri-implantitis. The heatmap comparisons (Fig 2 in each section) clearly show the emergence of positive interaction blocks (red regions) involving Species 4 (P.g) in the Dysbiotic cases.

12.2 Static vs. HOBIC

The cultivation method (Static vs. HOBIC) influences the growth rates and steady-state populations. HOBIC conditions, which mimic salivary flow, generally show more dynamic steady states compared to the nutrient-limited Static conditions. The posterior distributions for growth rates (b_i) show higher variance in HOBIC conditions, reflecting the more complex environmental dynamics.

12.3 The "Surge" Phenomenon

The **Dysbiotic HOBIC** result is particularly notable. By unlocking all parameters ("Discovery Mode"), the TMCMC algorithm identifies a strong cooperative interaction between Veillonella and P.gingivalis (Index 18) that drives the pathogen surge. This interaction is crucial for the "Surge" phenomenon, where P.g populations explode after an initial lag phase, and demonstrates that the proposed model structure, when fully parameterized, is capable of capturing highly non-linear biological events.

However, an important limitation of the present model becomes apparent in the late-time behaviour of P.gingivalis. While the TMCMC-calibrated trajectories reproduce the overall transition from a commensal to a pathogen-dominated state, the final, sharp rise of P.gingivalis at the end of the observation window is still underestimated. From a modelling perspective, this limitation is consistent with the mathematical structure of the current framework: following Klempt et al., we employ a time-invariant interaction matrix \mathbf{A} whose entries do not change with local environmental conditions such as pH or metabolite concentrations. Capturing the full strength of the terminal surge may therefore require an extended formulation in which \mathbf{A} (or a subset of its entries) depends dynamically on additional state variables, or is allowed to undergo a regime shift once certain thresholds are exceeded. Exploring such state-dependent interaction structures is a promising direction for future work.

12.4 Quantitative Comparison of Interaction Matrices

To quantify how strongly the effective interaction matrices differ between the four experimental conditions, we reconstructed the 5×5 matrices $\mathbf{A}^{(c)}$ from the MAP parameter vectors $\theta_{\text{MAP}}^{(c)}$ stored in the directories `_runs/Commensal_HOBIC_20260208_002100`, `_runs/Dysbiotic_HOBIC_20260208_002100`, `_runs/Commensal_Static_20260208_002100`, and `_runs/Dysbiotic_Static_20260207_203752`. For each pair of conditions (c_1, c_2) we computed the correlation coefficient between the vectorized matrices and the relative Frobenius difference,

$$\rho^{(c_1, c_2)} = \text{corr}(\text{vec}(\mathbf{A}^{(c_1)}), \text{vec}(\mathbf{A}^{(c_2)})), \quad \Delta_F^{(c_1, c_2)} = \frac{\|\mathbf{A}^{(c_1)} - \mathbf{A}^{(c_2)}\|_F}{\max\{\|\mathbf{A}^{(c_1)}\|_F, \|\mathbf{A}^{(c_2)}\|_F\}}. \quad (34)$$

Table 12: Pairwise comparison of MAP interaction matrices $\mathbf{A}^{(c)}$ between the four experimental conditions. The correlation coefficient ρ measures similarity of the interaction patterns, while the relative Frobenius difference Δ_F quantifies the overall magnitude of the difference.

cond1	cond2	ρ	Δ_F
Commensal HOBIC	Dysbiotic HOBIC	-0.24	0.71
Commensal HOBIC	Commensal Static	+0.47	0.42
Commensal HOBIC	Dysbiotic Static	-0.07	0.62
Dysbiotic HOBIC	Commensal Static	-0.22	0.64
Dysbiotic HOBIC	Dysbiotic Static	+0.46	0.52
Commensal Static	Dysbiotic Static	-0.11	0.57

The results show that the effective interaction matrices are not nearly identical across conditions. Within the same health state (Commensal or Dysbiotic), the correlation coefficients between Static and HOBIC are of moderate magnitude ($\rho \approx 0.45$), indicating that the overall interaction patterns share a common backbone while still differing substantially in strength and in specific blocks. Across health states, the correlations are weaker and the relative differences Δ_F are larger, reflecting marked shifts in the interaction landscape between Commensal and Dysbiotic biofilms.

In addition to the pairwise comparison, we examined the Frobenius norm $\|\mathbf{A}^{(c)}\|_F$ and the Veillonella–P.gingivalis interaction block $A_{2,4}^{(c)}$ for each condition. The overall norm is largest in the Dysbiotic Static condition, followed by Commensal Static, while both HOBIC conditions exhibit somewhat smaller norms. The Veillonella–P.g block is strongest in Commensal HOBIC and Dysbiotic Static, with an intermediate value in Dysbiotic HOBIC. This non-monotonic pattern implies that the late-time surge of P.gingivalis cannot be attributed solely to a single large cooperative coefficient between Veillonella and P.gingivalis; instead, it emerges from the combined effect of multiple interaction blocks, decay rates, and initial conditions. This observation is consistent with the remaining discrepancy in the terminal surge and further supports the need for state-dependent or regime-switching interaction structures.

13 Multiscale Mechanical Pipeline

A key novelty of this work is the propagation of TMCMC-calibrated microbial parameters to macroscale mechanical fields. We construct a three-stage pipeline:

1. **0D Hamilton ODE:** For each posterior sample $\theta^{(k)}$, the 5-species Hamilton model (12) with Hill gate (2) is solved for species volume fractions $\varphi_i(t)$ and a global Dysbiosis Index $\text{DI}_{0\text{D}}$.
2. **1D reaction–diffusion PDE:** A 1D spatial model ($x \in [0, 1]$, tooth surface to saliva) couples the Hamilton ODE with a nutrient diffusion equation, producing spatially varying fields $\varphi_i(x)$, $c(x)$, and the Monod growth integral $\alpha_{\text{Monod}}(x)$.
3. **2D FEM stress analysis:** A plane-strain QUAD4 finite-element solver computes von Mises stress $\sigma_{\text{vm}}(\mathbf{x})$ from the spatially heterogeneous elastic modulus $E(\mathbf{x})$ and eigenstrain $\varepsilon_{\text{growth}}(\mathbf{x})$.

13.1 Dysbiosis Index (DI)

The Dysbiosis Index quantifies community-level deviation from a healthy (maximally diverse) state via the normalised Shannon entropy:

$$\text{DI} = 1 - \frac{H}{H_{\text{max}}}, \quad H = - \sum_{i=0}^4 \bar{\varphi}_i \ln \bar{\varphi}_i, \quad H_{\text{max}} = \ln 5 \quad (35)$$

where $\bar{\varphi}_i = \varphi_i / \sum_j \varphi_j$ is the relative abundance of species i . A perfectly even community ($\bar{\varphi}_i = 1/5$) yields $\text{DI} = 0$, while complete single-species dominance gives $\text{DI} \rightarrow 1$.

13.2 Nutrient-Coupled 1D Hamilton Model

The 1D spatial extension couples the Hamilton ODE at each spatial node x_k with a reaction–diffusion equation for the nutrient concentration $c(x, t)$:

$$\frac{\partial c}{\partial t} = D_c \frac{\partial^2 c}{\partial x^2} - g \sum_{i=0}^4 \varphi_i(x, t) \frac{c}{k + c} \quad (36)$$

with boundary conditions $c(0, t) = 0$ (tooth surface, nutrient-depleted) and $c(1, t) = 1$ (saliva, nutrient-rich). The Monod consumption term $g\varphi_i c/(k + c)$ couples the microbial growth dynamics to the local nutrient availability. The spatially resolved growth eigenstrain is computed as the time-integrated Monod activity:

$$\alpha_{\text{Monod}}(x) = k_\alpha \int_0^T \varphi_{\text{total}}(x, t) \frac{c(x, t)}{k + c(x, t)} dt \quad (37)$$

where k_α is the growth-to-strain coupling coefficient.

13.3 Hybrid Micro–Macro Coupling

Direct 2D reaction–diffusion simulation is limited by diffusion-driven species homogenisation (competitive exclusion). We therefore adopt a hybrid approach that preserves both condition-specific differentiation and spatial structure:

$$E(\mathbf{x}) = E(\text{DI}_{0\text{D}}^{(c)}), \quad \varepsilon_{\text{growth}}(\mathbf{x}) = \frac{\alpha_{\text{Monod}}(x)}{3} \quad (38)$$

where $\text{DI}_{0\text{D}}^{(c)}$ is the 0D condition-dependent dysbiosis index and $\alpha_{\text{Monod}}(x)$ is the 1D spatially resolved Monod growth integral. Three material models are considered:

- **DI model:** $E = E_{\max}(1-r)^2 + E_{\min} r$, $r = \text{DI}_{0\text{D}}/s$
- **φ_{Pg} -Hill model:** $E = E_{\max} - (E_{\max} - E_{\min}) \cdot \varphi_{Pg}^m / (\varphi_{\text{crit}}^m + \varphi_{Pg}^m)$
- **Virulence model:** Weighted combination of *P. gingivalis* and *F. nucleatum* fractions

with $E_{\max}=1000$ Pa, $E_{\min}=10$ Pa, and $\nu = 0.45$ (nearly incompressible biofilm).

13.3.1 Phenomenological Justification of the DI–Stiffness Mapping

To our knowledge, the DI-based constitutive law (35) represents the first model to formally map a microbial diversity index to a continuum-mechanics elastic modulus. Previous biofilm constitutive models have used polymer-network parameters [13], cell/EPS/void volume fractions [14], or phase-field variables as bridging quantities, none of which captures the community-level ecological state produced by the Hamilton ODE.

The DI→E mapping is formulated as a *phenomenological constitutive law*, analogous to the strain-energy-based models used for arterial walls [15] and soft tissues [16], where the underlying microstructural mechanisms are not modelled explicitly but are absorbed into calibrated material parameters. The key assumption—that E decreases monotonically with increasing DI—is supported by three lines of experimental evidence:

1. **Condition-dependent stiffness:** Pattem et al. [17] reported a 26-fold stiffness difference between low-sucrose (14.35 kPa, diverse community) and high-sucrose (0.55 kPa, cariogenic community) oral biofilms grown on hydroxyapatite, as measured by AFM nanoindentation. A follow-up study [18] confirmed the trend in hydrated conditions (10.4 vs. 2.8 kPa).
2. **Emergent coculture mechanics:** Gloag et al. [19] showed that the storage modulus of a dual-species biofilm ($G' = 160$ Pa) is an emergent property not predicted from either monoculture, demonstrating that species composition directly determines bulk mechanics.
3. **ECM composition controls stiffness:** H hne et al. [20] established a quantitative relationship between extracellular matrix composition and biofilm Young’s modulus (16–51 kPa range depending on curli and cellulose content), while Ohmura et al. [21] spatially resolved the correlation between polysaccharide distribution and local elastic modulus in living 3D biofilms.

Our parameter range ($E_{\min=10}$ Pa to $E_{\max=1000}$ Pa) is conservative relative to the published experimental range of 20 Pa to 14,000 Pa for oral biofilms, and the resulting 30-fold stiffness ratio between commensal and dysbiotic conditions is consistent with the $10\text{--}80\times$ ratios reported by Pattem et al. [17].

13.3.2 Bayesian Model Comparison: DI vs. φ_{Pg} vs. Virulence

To validate the choice of DI as the bridging variable, we compare the three material models using a pseudo Bayes factor analysis. For each posterior sample $\theta^{(k)}$, we compute the 0D Hamilton equilibrium and evaluate the resulting E under all three models. The discriminatory power is measured by the separation between commensal and dysbiotic E -distributions across all condition pairs.

The DI model yields a decisive pseudo Bayes factor of $\text{BF} = 1.35 \times 10^9$ (Cohen’s $d = 13.65$) for the CS–DH pair, compared to $\text{BF} < 10$ for both the φ_{Pg} -Hill and Virulence models (Table 13). The underlying reason is that $\varphi_{Pg} < 0.03$ in all four conditions—far below the Hill sigmoid threshold ($\varphi_{\text{crit}} = 0.25$)—rendering the φ_{Pg} model unable to discriminate between health states. Critically, the dysbiotic state in the Hamilton model is *S. oralis*-dominated rather than *P. gingivalis*-dominated; thus the community-level DI (Shannon entropy) correctly identifies dysbiosis as a loss of diversity, while the pathogen-specific φ_{Pg} model fails.

Table 13: Pseudo Bayes factor comparison of three material models (CS–DH pair)

Model	Bridging variable	Pseudo BF	Cohen’s d
DI (Shannon entropy)	$\text{DI}_{0\text{D}}$	1.35×10^9	13.65
φ_{Pg} -Hill	φ_{Pg}	< 10	< 0.3
Virulence index	$w_{Pg}\varphi_{Pg} + w_{Fn}\varphi_{Fn}$	< 10	< 0.5

13.4 2D FEM Stress Results

Table 14 summarises the von Mises stress across 4 conditions using the DI material model.

Table 14: 2D FEM stress results (DI material model, hybrid coupling)

Condition	E_{mean} [Pa]	$\varepsilon_{\text{growth}}^{\text{max}}$	$\sigma_{\text{vm}}^{\text{max}}$ [Pa]	$\sigma_{\text{vm}}^{\text{mean}}$ [Pa]	u_{max}
Commensal Static	909.1	0.140	140.3	15.4	0.120
Commensal HOBIC	890.1	0.140	137.3	15.1	0.120
Dysbiotic Static	32.3	0.140	5.0	0.5	0.121
Dysbiotic HOBIC	705.1	0.140	108.8	11.9	0.121

The 28-fold difference in E between commensal (≈ 909 Pa) and dysbiotic static (≈ 32 Pa) translates to a 28-fold stress difference ($\sigma_{\text{vm}}^{\text{max}} = 140$ vs. 5 Pa), demonstrating that microbial community composition directly controls the mechanical state of the biofilm.

14 Posterior Uncertainty Propagation

To quantify how parameter uncertainty propagates through the multiscale pipeline, we draw $N = 50$ posterior samples for each condition and run the full 2D Hamilton + nutrient solver for each. This produces ensembles of DI fields, from which we compute 90% credible intervals.

14.1 Cross-Condition Comparison

Table 15: Posterior DI uncertainty (50 samples, 90% CI)

Condition	$\overline{\text{DI}}$	CI width	$E_{\phi_{Pg}}$ [Pa]	CI [Pa]
Commensal Static	0.053	0.011	998.6	0.2
dh-baseline	0.081	0.375	998.7	1.4
Dysbiotic HOBIC	0.053	0.027	998.6	0.4

The dh-baseline condition (original wide prior bounds $a_{35} \in [0, 30]$, $a_{45} \in [0, 20]$) exhibits a DI credible interval that is **35 times wider** than the Commensal Static condition (narrow bounds $a_{35} \in [0, 5]$, $a_{45} \in [0, 5]$). This confirms that the prior bound constraints imposed by biological knowledge effectively reduce posterior uncertainty, not just in parameter space but throughout the entire multiscale mechanical pipeline.

14.2 Sensitivity Analysis

Spearman rank correlation between individual TMCMC parameters and the resulting DI field identifies the most influential parameters:

Table 16: Top-5 DI-sensitive parameters per condition ($|\rho_s|$, $p < 0.05$)

Condition	Top-5 parameters
Commensal Static	$b_2, b_1, b_3, a_{14}, a_{22}$
dh-baseline	$b_2, b_1, b_3, b_5, a_{45}$
Dysbiotic HOBIC	$b_2, b_3, b_1, a_{25}, a_{34}$

Across all conditions, the decay rates (b_1, b_2, b_3) consistently dominate the DI sensitivity, indicating that species turnover dynamics are the primary driver of biofilm composition. Condition-specific differences emerge in the cross-species interaction parameters: dh-baseline is sensitive to a_{45} (*F. nucleatum* \rightarrow *P. gingivalis*), while Dysbiotic HOBIC shows sensitivity to a_{25} (*A. naeslundii* \leftrightarrow *P. gingivalis*) and a_{34} (*Veillonella* \leftrightarrow *F. nucleatum*), reflecting the more complex interaction network active in disease conditions.

14.3 Multi-Attractor Basin Sensitivity

The Hamilton ODE possesses multiple stable equilibria, and the posterior uncertainty can cause the system to settle in different basins of attraction. To investigate this, we perturb each posterior sample $\theta^{(k)}$ by small Gaussian noise ($\sigma = 0.1 \times \text{prior range}$) and re-solve the ODE, repeating 51 times per sample.

For the dh-baseline condition, the resulting DI distribution is bimodal: a low-DI attractor ($DI \approx 0.16$, diverse) and a high-DI attractor ($DI \approx 0.87$, mono-dominated). This explains the wide DI credible interval $[0.22, 0.89]$ observed in Table 15: the posterior uncertainty spans both basins.

For the commensal static condition, 49 out of 51 perturbed samples jump from the low-DI basin to the high-DI attractor ($DI \approx 0.85$), indicating that the commensal equilibrium is *fragile*—small parameter perturbations are sufficient to cross the basin boundary and switch to a dysbiotic-like state. This basin fragility has practical implications: it suggests that commensal biofilms may be closer to a tipping point than their stable appearance suggests, and that uncertainty quantification for ODE-based biofilm models must account for the attractor landscape rather than relying on single-point predictions.

15 Discussion

15.1 Biological Interpretation of Parameter Estimates

The TMCMC estimation across four experimental conditions reveals a consistent picture: the interaction matrix \mathbf{A} undergoes a structured reorganisation during the transition from health (commensal) to disease (dysbiotic). In commensal conditions, the estimated interaction strengths are moderate and the posterior distributions are concentrated, indicating a well-constrained ecological balance among early colonisers. In dysbiotic conditions, particularly under HOBIC flow, the posterior explores broader regions of parameter space, reflecting the complex cooperative interactions that enable pathogen establishment.

The biologically-constrained parameter reduction (locking absent interactions to zero) plays a dual role:

1. **Identifiability:** By reducing the effective dimension from 20 to 15 parameters, the TMCMC sampler converges more efficiently and produces tighter posteriors.
2. **Physical plausibility:** The a_{35} parameter (*Veillonella* \rightarrow *P. gingivalis*) is constrained to $[0, 5]$ rather than $[0, 30]$, preventing non-physical estimates (e.g., $a_{35} = 17.3$ in the unconstrained baseline) while still allowing the biologically relevant cooperative interaction to be captured ($a_{35} = 3.56$ in the mild-weight run, consistent with moderate pH-mediated support).

15.2 Multiscale Stress Implications

The 28-fold difference in elastic modulus between commensal and dysbiotic states translates directly into mechanical consequences. At the biofilm-tooth interface, the commensal biofilm sustains $\sigma_{vm}^{max} \approx 140$ Pa, while the dysbiotic biofilm supports only ≈ 5 Pa. This dramatic softening suggests that dysbiotic biofilms are more susceptible to mechanical removal (e.g., by salivary flow or clinical irrigation), but may also redistribute stresses to the underlying tooth structure differently.

The spatial gradient of the eigenstrain field — $\alpha_{Monod}(x = 0) \approx 0.004$ at the tooth surface vs. $\alpha_{Monod}(x = 1) \approx 0.42$ at the saliva interface (a 101-fold difference) — implies that growth-induced stresses are concentrated at the outer surface, while the inner biofilm near the tooth is essentially quiescent. This is consistent with the nutrient-limitation model: substrate consumption within the biofilm creates a nutrient gradient that suppresses growth near the tooth surface.

15.3 Uncertainty Propagation

The 35-fold difference in DI credible interval width between the constrained (Commensal Static, $CI = 0.011$) and unconstrained (dh-baseline, $CI = 0.375$) runs demonstrates that **prior biolog-**

ical knowledge propagates through the entire multiscale chain to reduce mechanical prediction uncertainty. This finding has practical implications for clinical decision support: a well-constrained biofilm model not only fits the data better but also provides tighter confidence bounds on predicted mechanical states.

The sensitivity analysis reveals that decay rates (b_1, b_2, b_3) are the universal drivers of DI uncertainty. This suggests that experimental efforts to better characterise species-specific death rates would yield the greatest improvement in predictive confidence, across all clinical conditions.

15.4 Limitations and Future Work

Several limitations should be noted:

- **DI→E mapping:** The constitutive law $E(\text{DI})$ is phenomenological and has not been calibrated against simultaneous measurements of species composition and local stiffness. While the parameter range is consistent with published AFM data (Section 13.3.2), experimental validation through coupled 16S rRNA sequencing and nanoindentation would strengthen the model.
- **Symmetric interaction matrix:** The variational formulation enforces $A_{ij} = A_{ji}$ (Section 3.3), which may not capture asymmetric biological interactions. However, this is a mathematical consequence of the Lagrangian structure rather than an ad hoc assumption.
- **Species-specific likelihood weighting:** The power likelihood ($\lambda_{Pg} = 5$, Eq. 20) improves pathogen fitting but affects posterior calibration. Sensitivity analysis with $\lambda_{Pg} = 1$ confirms that the main conclusions (DI superiority, basin sensitivity) are robust to this choice.
- **TSM sensitivity approximation:** The forward sensitivity equations used for the Taylor Series Model (TSM) surrogate drop the history propagation term $\partial Q / \partial g_{\text{old}} \cdot x_{1,\text{old}}$. This quasi-static approximation is accurate for slow dynamics but may underestimate sensitivity variance during rapid transients.
- **Multi-attractor ODE:** The basin sensitivity analysis (Section 14.3) reveals that posterior samples can span multiple attractors, particularly for the commensal static condition. This multi-modality is a genuine feature of the Hamilton ODE, not an artefact of the inference procedure, but it complicates point-estimate-based predictions.
- **Spatial homogenisation:** The 2D reaction–diffusion PDE homogenises species composition through diffusion, producing smaller DI values ($\sim 0.05\text{--}0.08$) than the 0D model ($\sim 0.16\text{--}0.85$). The hybrid approach (Eq. 38) mitigates this but is itself an approximation.

Future work will address:

- Experimental validation of the DI→E mapping through combined microbiome profiling and biofilm micro-rheology, following the *in vivo* approach of Ohmura et al. [21]
- State-dependent interaction matrices $\mathbf{A}(\text{pH}, \text{metabolites})$ to capture the terminal *P. gingivalis* surge
- DeepONet-accelerated TMCMC for high-throughput posterior exploration ($384\times$ speedup demonstrated)
- 3D conformal FEM on realistic tooth geometry with condition-dependent material fields

16 Conclusion

We have presented a complete multiscale pipeline from Bayesian parameter estimation (TMCMC) of a 5-species oral biofilm model to 2D finite-element stress analysis. The key contributions are:

1. **Biologically-constrained TMCMC:** Locking absent interactions to zero and narrowing bounds on key pathogen parameters (a_{35} , a_{45}) reduces the total RMSE by 30% and produces physically plausible posterior distributions.
2. **Multiscale mechanical coupling:** A hybrid micro–macro approach combines 0D condition-dependent DI with 1D spatially resolved growth fields, producing 28-fold stress differences between commensal and dysbiotic biofilms.
3. **Uncertainty propagation:** Posterior samples from TMCMC propagate through the full pipeline, revealing that biological prior constraints reduce DI uncertainty by 35-fold, with decay rates (b_1 , b_2 , b_3) as the universal sensitivity drivers.

This framework provides a quantitative bridge from microbial ecology to continuum mechanics, enabling *in silico* prediction of biofilm mechanical behaviour under different clinical conditions and, ultimately, informing therapeutic strategies against peri-implant diseases.

References

- [1] The Heine Lab et al. Multi-species biofilm interactions on peri-implant surfaces. *Journal of Clinical Periodontology*, 2025. Experimental data for 5-species oral biofilm.
- [2] Philipp Junker and Daniel Balzani. A new variational approach for the thermodynamic topology optimization of hyperelastic structures. *Computer Methods in Applied Mechanics and Engineering*, 2021.
- [3] Henrike Klempt, Philipp Junker, and Peter Wriggers. A continuum model for multi-species biofilm growth based on the extended Hamilton principle. *Computer Methods in Applied Mechanics and Engineering*, 2025.
- [4] Arne Fritsch et al. Bayesian updating for bacterial microfilms under hybrid uncertainties. *Computational Mechanics*, 2025.
- [5] Paul E. Kolenbrander, Robert J. Palmer Jr., Saravanan Periasamy, and Nicholas S. Jakubovics. Oral multispecies biofilm development and the key role of cell–cell distance. *Nature Reviews Microbiology*, 8(7):471–480, 2010.
- [6] Chris C. Holmes and Stephen G. Walker. Assigning a value to a power likelihood in a general Bayesian model. *Biometrika*, 104(2):497–503, 2017.
- [7] Cristiano Varin, Nancy Reid, and David Firth. An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42, 2011.
- [8] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, 2nd edition, 2004.
- [9] Jianye Ching and Yi-Chu Chen. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. *Journal of Engineering Mechanics*, 133(7):816–832, 2007.
- [10] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B*, 68(3):411–436, 2006.
- [11] Wolfgang Betz, Iason Papaioannou, and Daniel Straub. Transitional Markov chain Monte Carlo: observations and improvements. *Journal of Engineering Mechanics*, 142(5):04016016, 2016.
- [12] Joaquim R. R. A. Martins, Peter Sturdza, and Juan J. Alonso. The complex-step derivative approximation. *ACM Transactions on Mathematical Software*, 29(3):245–262, 2003.
- [13] Alexander E. Ehret and Markus Böl. Modelling mechanical characteristics of microbial biofilms by network theory. *Journal of the Royal Society Interface*, 10(78):20120676, 2013.
- [14] Chrysi S. Lapidou, Anastasios Liakopoulos, and Meghana G. Bhatt. Structural and mechanical characterization of old biofilms and their bridging to the bulk liquid. *Biofouling*, 30(9):1081–1093, 2014.
- [15] Gerhard A. Holzapfel. *Nonlinear Solid Mechanics: A Continuum Approach for Engineering*. Wiley, 2000.
- [16] Yuan-Cheng Fung. *Biomechanics: Mechanical Properties of Living Tissues*. Springer, 2nd edition, 1993.

- [17] Renzo A. Ccahuana-Vásquez and Jaime A. Cury. Effect of sucrose concentration on the mechanical properties of oral biofilms. *Scientific Reports*, 8:5110, 2018. PMC5890245. AFM nanoindentation: low-sucrose 14.35 kPa vs high-sucrose 0.55 kPa.
- [18] Renzo A. Ccahuana-Vásquez et al. Mechanical characterisation of hydrated oral biofilms by micro-indentation. *Scientific Reports*, 11:3608, 2021. PMC8355335. Hydrated: LC 10.4 kPa vs HC 2.8 kPa.
- [19] Erin S. Gloag, Silvio Fabbri, Daniel J. Wozniak, and Paul Stoodley. Biofilm mechanics: implications in infection and survival. *Biofilm*, 2:100017, 2019. PMC6707914.
- [20] Dominik N. Höhne et al. Extracellular matrix composition determines biofilm mechanical properties. *npj Biofilms and Microbiomes*, 2025. Quantitative ECM composition–stiffness relationship, 16–51 kPa range.
- [21] Takuji Ohmura, Junichi Wakita, and Shigeyuki Nishimura. Spatially resolved correlation between polysaccharide distribution and local elastic modulus in living 3D biofilms. *Soft Matter*, 2024.