

Nishioka Algorithm: Biologically-Constrained Parameter Reduction for 5-Species Biofilm Model

TMCMC Parameter Estimation Project

February 2026

Abstract

This document describes the Nishioka Algorithm, 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.

Contents

| | | |
|----------|---|----------|
| 1 | Introduction | 3 |
| 2 | Biological Basis | 3 |
| 2.1 | Species in the Model | 3 |
| 2.2 | Interaction Network (Figure 4C) | 3 |
| 2.3 | Active Interactions | 4 |
| 2.4 | Absent Interactions (Locked) | 4 |
| 3 | Mathematical Formulation | 4 |
| 3.1 | Governing Equations | 4 |
| 3.2 | Symmetric Matrix Assumption | 4 |
| 3.3 | Parameter Vector Definition | 5 |
| 3.4 | Complete Parameter Mapping | 5 |
| 3.5 | Locked Parameter Indices | 5 |
| 3.6 | Prior Bounds | 6 |
| 3.7 | Effective Parameter Space | 6 |
| 4 | Experiment Conditions & Parameter Estimation | 6 |
| 4.1 | Parameter Locking Rules | 6 |
| 4.2 | Detailed Locking Logic | 6 |
| 4.3 | 4-Stage Sequential Estimation | 7 |
| 5 | Implementation | 7 |
| 5.1 | Core Module: <code>core/nishioka_model.py</code> | 7 |
| 5.2 | Estimation Script: <code>main/estimate_reduced_nishioka.py</code> | 8 |
| 6 | Comparison: Standard vs Nishioka | 8 |
| 7 | Advantages and Limitations | 8 |
| 7.1 | Advantages | 8 |
| 7.2 | Limitations | 9 |

| | | |
|----------|---|----------|
| 8 | Usage | 9 |
| 8.1 | Running the Estimation | 9 |
| 8.2 | Comparing with Standard Results | 9 |
| 8.3 | Output Files | 9 |
| 9 | Conclusion | 9 |

1 Introduction

The 5-species biofilm model [3] describes the dynamics of bacterial populations through an interaction matrix \mathbf{A} and decay vector \mathbf{b} . The estimation is performed using Transitional Markov Chain Monte Carlo (TMCMC) [1] coupled with a Time-Separated Stochastic Mechanics (TSM) solver [2]. The standard 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 Nishioka Algorithm 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 [3], the following interaction network was established:

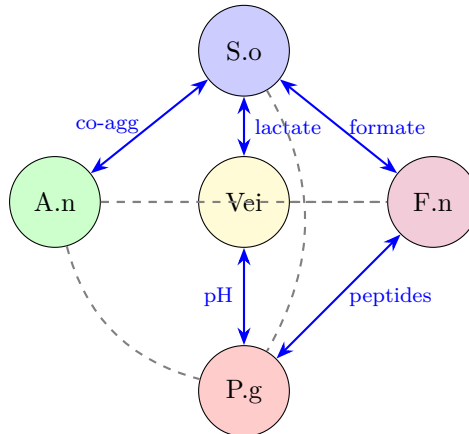


Figure 1: Species interaction network derived from Figure 4C. Solid blue arrows indicate active interactions (estimated parameters). Dashed gray lines indicate absent interactions (locked to zero).

2.3 Active Interactions

The following species pairs have direct biological interactions:

| 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.

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 Nishioka Algorithm:

| 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 Nishioka Algorithm. Numbers in parentheses are 0-indexed species IDs.

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 Symmetric Matrix Assumption

Critical assumption: The interaction matrix \mathbf{A} is symmetric:

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

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:

$$A_{02} = A_{20} = \theta_{10} \quad (\text{stored as } a_{13} \text{ in code}) \quad (3)$$

3.3 Parameter Vector Definition

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

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

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 ($A[i-1, j-1]$).

3.4 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 4).

3.5 Locked Parameter Indices

The Nishioka Algorithm defines the set of locked indices based on absent biological interactions:

$$\mathcal{L} = \{6, 12, 13, 16, 17\} \quad (5)$$

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

$$\theta_k = 0 \quad (\text{fixed, not estimated}) \quad (6)$$

3.6 Prior Bounds

The **base** prior distribution (for Commensal/Dysbiotic Static conditions) is:

$$\theta_k \sim \begin{cases} \text{Uniform}(0, 0) & \text{if } k \in \mathcal{L} \text{ (locked)} \\ \text{Uniform}(0, 1) & \text{if } k = 18 \text{ (Vei} \rightarrow \text{P.g, positive cooperation)} \\ \text{Uniform}(-1, 1) & \text{otherwise (free)} \end{cases} \quad (7)$$

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

$$\theta_{18} \sim \text{Uniform}(-3, -1) \quad (\text{Dysbiotic HOBIC only}) \quad (8)$$

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

3.7 Effective Parameter Space

The effective number of free parameters is:

$$n_{\text{free}} = 20 - |\mathcal{L}| = 20 - 5 = 15 \quad (9)$$

4 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.

4.1 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}).

| 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 5: Parameter estimation counts for each experimental condition.

4.2 Detailed Locking Logic

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

"Based on qPCR data showing *P. gingivalis* and *F. nucleatum* were undetectable or below detection limits (Heine et al., Table S8) [3]."

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 Nishioka locks ($N_{\text{locked}} = 5$).

"Metabolite accumulation in static culture limits the dynamic interactions observed in flow conditions (Heine et al., Discussion) [3]."

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 Nishioka 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) [3]."

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.

4.3 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 6: 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.

5 Implementation

5.1 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 Nishioka 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

```

5.2 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]

```

6 Comparison: Standard vs Nishioka

| Aspect | Standard | Nishioka |
|------------------------|----------------------|-------------------------|
| 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 7: Comparison of Standard and Nishioka parameter estimation approaches.

7 Advantages and Limitations

7.1 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

7.2 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

8 Usage

8.1 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 &
```

8.2 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
```

8.3 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 8: Output files from Nishioka estimation.

9 Conclusion

The Nishioka Algorithm provides a biologically grounded approach to parameter estimation in complex biofilm models. By leveraging experimental interaction data to constrain the parameter space, it achieves a significant reduction in computational complexity while improving the biological interpretability of the results. The reduction from 20 to 15 parameters mitigates overfitting and enhances the identifiability of the remaining active interactions. This method demonstrates the value of integrating domain knowledge into statistical inference frameworks for biological systems.

References

- [1] 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. Original TMCMC method.
- [2] Hendrik Geisler, Can Erdogan, Jonas Nagel, and Philipp Junker. A new paradigm for the efficient inclusion of stochasticity in engineering simulations: Time-separated stochastic mechanics. *Computational Mechanics*, 75(1):211–235, 2025. Comprehensive TSM framework.
- [3] N. Heine, K. Bittroff, S. P. Szafrński, et al. Influence of species composition and cultivation condition on peri-implant biofilm dysbiosis in vitro. *Frontiers in Oral Health*, 6:1649419, 2025. Local PDF: `tmcmc/Influence of species composition and cultivation condition on peri-implant biofilm dysbiosis in vitro.pdf`.