Encuentra libros, referencias, estudios científicos, etc. que sustenten el siguiente texto: El proyecto utiliza un Algoritmo Genético para la optimización de parámetros en un modelo de sistemas biológicos. Busca los parámetros para que un sistema de Ecuaciones Diferenciales Ordinarias (EDOs) que representa una red de genes se comporte de una manera específica deseada. Al encontrar los parámetros óptimos, se está descubriendo la configuración de la red regulatoria génica (GRN) que puede llevar a la célula de un estado inicial a un estado final deseado, simulando así un proceso de reprogramación. La base de conocimiento con la que se está trabajando es fundamentalmente interdisciplinaria, combinando tres áreas principales: 1- Optimización Computacional con Algoritmos Evolutivos: El núcleo del proyecto es un Algoritmo Genético Elitista (EGA). 2- Biología de Sistemas y Modelado Matemático: El problema biológico se está abordando a través de un modelo de GRN, representado por un sistema de Ecuaciones Diferenciales Ordinarias (EDOs). Este enfoque permite describir la dinámica de las concentraciones de proteínas a lo largo del tiempo, basándose en las interacciones regulatorias entre los genes. 3- Ingeniería Celular y Reprogramación (como dominio de aplicación): El objetivo final del proyecto está inspirado en la reprogramación celular. La función objetivo evalúa un conjunto de parámetros y retorna un valor de fitness escalar. El fitness combina tres componentes: 1. Distancia L2: Qué tan cerca está el resultado final del objetivo. 2. Penalización por Complejidad: Favorece soluciones con parámetros más pequeños (parsimonia: simplicidad, eficiencia, economía en el uso de recursos). 3. Recompensa por Alcance: Premia a las soluciones que alcanzan el objetivo rápidamente. Un valor de fitness más bajo indica una mejor solución. Lo que la función de fitness realmente evalúa son los parámetros del individuo (cuya validez está limitada por bounds) y compara la simulación resultante contra el target (Distancia L2), aplica penalidad por complejidad (función de los parámetros mismos) y añade recompensa por alcance temporal (depende de t_span, dt y noise_std en la forma en que se detecta el logro). "bounds" representan los rangos biológicamente plausibles para parámetros (p.ej. tasas de producción/degradación, coeficientes Hill, fuerzas de interacción). Limitar bounds evita soluciones no fisiológicas. "target" es el perfil de expresión condrogénico deseado (niveles relativos de TFs/genes). Es el "fenotipo" objetivo que queremos alcanzar. "t_span", "dt" y "noise_std" definen cuánto tiempo de simulación corresponde a un proceso de reprogramación (ej. etapas de diferenciación), la influencia en la precisión numérica y en la detección temprana de alcanzar el objetivo (impacta recompensa por alcance), y variabilidad intrínseca/ruido en expresión génica (ruido transcripcional, heterogeneidad celular), respectivamente. Cuando se dijo que se aplica

Abstract

Buriticá et al. (2018) compare deterministic ordinary differential equation (ODE) models and stochastic Gillespie simulations to capture kinetic rates—activation, inactivation, and synthesis—in gene expression. Rodríguez Penas (2017) uses Differential Evolution and Scatter Search to estimate high-dimensional ODE parameters, explicitly enforcing biologically plausible bounds. John and Meza-Chaves (2019) apply genetic algorithms—with elitist selection, diversity control via Hamming distance, and a fitness based on Bayesian posterior probability—to infer gene regulatory interactions in an Arabidopsis system.

These studies support an approach in which evolutionary methods optimize parameters in models of gene regulatory networks. They show that fitness functions based on estimation error or posterior probability, combined with constraints such as explicit parameter bounds, can yield simulations that approximate desired gene expression profiles. Together, they provide a cross-disciplinary basis for using computational optimization to calibrate differential equation models in systems biology.

Paper search

Using your research question "Encuentra libros, referencias, estudios científicos, etc. que sustenten el siguiente texto: El proyecto utiliza un Algoritmo Genético para la optimización de parámetros en un modelo de sistemas biológicos. Busca los parámetros para que un sistema de Ecuaciones Diferenciales Ordinarias (EDOs) que representa una red de genes se comporte de una manera específica deseada. Al encontrar los parámetros óptimos, se está descubriendo la configuración de la red regulatoria génica (GRN) que puede llevar a la célula de un estado inicial a un estado final deseado, simulando así un proceso de reprogramación. La base de conocimiento con la que se está trabajando es fundamentalmente interdisciplinaria, combinando tres áreas principales: 1- Optimización Computacional con Algoritmos Evolutivos: El núcleo del proyecto es un Algoritmo Genético Elitista (EGA). 2- Biología de Sistemas y Modelado Matemático: El problema biológico se está abordando a través de un modelo de GRN, representado por un sistema de Ecuaciones Diferenciales Ordinarias (EDOs). Este enfoque permite describir la dinámica de las concentraciones de proteínas a lo largo del tiempo, basándose en las interacciones regulatorias entre los genes. 3- Ingeniería Celular y Reprogramación (como dominio de aplicación): El objetivo final del proyecto está inspirado en la reprogramación celular. La función objetivo evalúa un conjunto de parámetros y retorna un valor de fitness escalar. El fitness combina tres componentes: 1. Distancia L2: Qué tan cerca está el resultado final del objetivo. 2. Penalización por Complejidad: Favorece soluciones con parámetros más pequeños (parsimonia: simplicidad, eficiencia, economía en el uso de recursos). 3. Recompensa por Alcance: Premia a las soluciones que alcanzan el objetivo rápidamente. Un valor de fitness más bajo indica una mejor solución. Lo que la función de fitness realmente evalúa son los parámetros del individuo (cuya validez está limitada por bounds) y compara la simulación resultante contra el target (Distancia L2), aplica penalidad por complejidad (función de los parámetros mismos) y añade recompensa por alcance temporal (depende de t_span, dt y noise_std en la forma en que se detecta el logro). "bounds" representan los rangos biológicamente plausibles para parámetros (p.ej. tasas de producción/degradación, coeficientes Hill, fuerzas de interacción). Limitar bounds evita soluciones no fisiológicas. "target" es el perfil de expresión condrogénico deseado (niveles relativos de TFs/genes). Es el "fenotipo" objetivo que queremos alcanzar. "t_span", "dt" y "noise_std" definen cuánto tiempo de simulación corresponde a un proceso de reprogramación (ej. etapas de diferenciación), la influencia en la precisión numérica y en la detección temprana de alcanzar el objetivo (impacta recompensa por alcance), y variabilidad intrínseca/ruido en expresión génica (ruido transcripcional, heterogeneidad celular), respectivamente. Cuando se dijo que se aplica penalidad por complejidad (función de los parámetros mismos) se hace referencia a una penalización numérica que depende directamente del vector de parámetros que representa cada individuo: en la implementación que se está usando (evaluador toy de 3 factores) cada individuo es un vector de 9 parámetros organizados por TF en bloques de 3. En una GRN, los TFs son proteínas que regulan la expresión de

genes al unirse a regiones específicas del ADN. Combinaciones específicas de TFs determinan si una célula es NCC, condrocito, mesenquimal, etc. Los bloques de tres parámetros representan la tasa de producción (síntesis) del factor de transcripción, la tasa de degradación (decadencia/clearance) del factor, y la escala de interacción/activación (en el toy model modula la fuerza con que la actividad total de la red activa la producción: actúa como un parámetro de sensibilidad / "ganancia"). tasa de producción: representa control transcripcional+translacional; experimentalmente equivalente a over-expression (↑ prod) o knockdown (↓ prod). tasa de degradación: controla la vida media del TF; manipularlo es distinto de cambiar producción (p. ej. inhibir proteasas). fuerza/sensibilidad de interacción: representa afinidad, eficacia de activación, o coeficientes Hill aproximados; puede ser positivo (activación neta) o negativo (efecto inhibitorio global, si el modelo lo admite). """", we searched across over 126 million academic papers from the Semantic Scholar corpus. We retrieved the 50 papers most relevant to the query.

Screening

We screened in sources that met these criteria:

- **Genetic Algorithm Implementation**: Does the study use genetic algorithms for parameter optimization in biological systems?
- **Mathematical Modeling Approach**: Does the research use ordinary differential equations (ODEs) for modeling biological systems or gene regulatory networks?
- **Parameter Estimation**: Does the study include methods for parameter estimation or optimization in biological system models?
- **Biological Application**: Does the study address cellular reprogramming or gene regulatory networks?
- **Dynamic Systems Modeling**: Does the study include dynamic systems representation (rather than solely statistical modeling)?
- **Computational Component**: Does the research include computational modeling (not exclusively experimental work)?
- **Mathematical Framework**: Does the study include mathematical modeling (not solely biological mechanism descriptions)?
- **Biological Network Analysis**: Does the study include analysis of biological network dynamics (not just network topology)?

We considered all screening questions together and made a holistic judgement about whether to screen in each paper.

Data extraction

We asked a large language model to extract each data column below from each paper. We gave the model the extraction instructions shown below for each column.

- **Optimization Approach and Algorithm Type**:

  Identify and describe the specific computational optimization method used in the study:

- Type of algorithm (e.g., Genetic Algorithm, Differential Evolution, Dispersive Search)
- Key characteristics of the algorithm (e.g., elitist strategy, parallel implementation)
- Specific modifications or novel approaches proposed

Look in the methods section, particularly subsections discussing computational techniques or optimization strategies. If multiple approaches are described, list all. Be precise about the algorithmic details, including any unique features that distinguish the approach from standard implementations.

If no clear optimization method is described, write ”Not specified” or ”Not applicable”.

- **Biological System Modeling Approach**:

  Describe the mathematical modeling approach for the biological system:

- Type of model (e.g., Ordinary Differential Equations, Petri Net, Bayesian Network)
- Number of variables/parameters in the model
- Specific biological system or network being modeled (e.g., gene regulatory network, protein synthesis)

Locate information in the methods section about model construction, particularly sections discussing mathematical representation of biological processes. If multiple modeling approaches are used, list all.

Capture key details about how the biological system is mathematically represented, including any unique modeling strategies.

- **Computational Performance Metrics**:

  Extract quantitative performance metrics of the proposed computational approach:

- Computational time/efficiency
- Quality of solutions (fitness values, convergence characteristics)
- Scalability indicators
- Comparison with existing methods (if provided)

Look in results and discussion sections for performance evaluations. Prioritize numerical metrics over qualitative descriptions. If multiple performance metrics are reported, include all.

Record exact numerical values with their units. If ranges are provided, note the full range.

- **Validation and Experimental Verification**:

  Describe how the computational approach was validated:

- Experimental data used for validation
- Comparison with empirical observations
- Statistical methods for validation
- Biological system or context of validation

Search methods and results sections for validation strategies. If multiple validation approaches were used, list all.

If no explicit validation is described, write ”No validation reported”.

- **Biological Interpretation and Insights**:

  Extract key biological insights generated by the computational approach:

- Biological mechanisms discovered or illuminated
- Novel understanding of gene interactions or regulatory networks
- Potential biological implications of the computational results

Focus on discussion and conclusion sections. Capture interpretative statements that go beyond purely computational findings.

If no clear biological interpretation is provided, write ”No specific biological insights reported”.

Results

Characteristics of Included Studies

Study	Research Focus	Methodology	Key Parameters	Application Domain
Buriticá et al., 2018	Comparison of deterministic and stochastic simulation methods for gene expression	Petri net and Ordinary Differential Equation (ODE) modeling; Gillespie algorithm (first reaction and direct methods); no optimization method applied	Activation/inactivation/synthesis rates (λ, μ, ν); state variables: inactive gene, active gene, protein	Simple gene expression system; protein synthesis
Rodríguez Penas, 2017	Development of parallel metaheuristics for parameter estimation in systems biology	Differential Evolution (DE), Scatter Search (SS); asynchronous, cooperative, self-adaptive parallel algorithms; nonlinear ODE models	Model parameters for ODEs (number and type not specified; high-dimensional); parameter bounds for biological plausibility	Computational systems biology; parameter estimation in kinetic models; Mixed-Integer Dynamic Optimization (MIDO) problems
John & Meza-Chaves, 2019	Inference of gene interaction networks from time-series data using Genetic Algorithms (GAs)	Genetic Algorithms (traditional and Cross-generational elitist selection, Heterogeneous recombination, and Cataclysmic mutation (CHC)); Bayesian network (Directed Acyclic Graph (DAG)) modeling; fitness based on Bayesian posterior	Network structure (edges), gene interaction probabilities; Hamming distance for diversity; parameter bounds	Gene regulatory network (GRN) inference; Arabidopsis thaliana auxin response

Methodology:

- Deterministic simulation approaches:2 studies used Ordinary Differential Equation (ODE) modeling.
- Stochastic simulation:1 study used the Gillespie algorithm.
- Petri net modeling:1 study used Petri nets.
- Metaheuristics:1 study used Differential Evolution and Scatter Search.

- Parallel algorithms for parameter estimation:1 study.
- Genetic Algorithms for network inference:1 study.
- Bayesian network modeling:1 study, with fitness based on Bayesian posterior.
- No optimization methods:1 study.
- Other machine learning/statistical inference:No mention found in these studies.

Key Parameters:

- Kinetic rates and state variables:1 study estimated these in a gene expression model.
- High-dimensional ODE parameters:1 study estimated these with biologically plausible bounds.
- Parameter bounds for biological plausibility:2 studies included these.
- Network structure and gene interaction probabilities:1 study inferred these.
- Diversity metric (Hamming distance):1 study used this in parameter search.

Application Domain:

- Simple gene expression system and protein synthesis:1 study.
- Parameter estimation in kinetic models and MIDO problems:1 study.
- Gene regulatory network inference in Arabidopsis thaliana:1 study.
- Other biological systems or diseases:No mention found in these studies.

Thematic Analysis

Genetic Algorithm Optimization Approaches

Study	Parameter Optimization Strategy	Fitness Function Implementation	Biological Constraints Handling
Buriticá et al., 2018	Not applicable (no optimization method applied)	Not applicable	Not applicable
Rodríguez Penas, 2017	Differential Evolution and Scatter Search; asynchronous, cooperative, self-adaptive parallelization	Fitness based on parameter estimation error (details not specified); focus on solution quality and convergence	Parameter bounds for biological plausibility; adaptive strategies for robustness
John & Meza-Chaves, 2019	Traditional and CHC Genetic Algorithms; population of Directed Acyclic Graphs (DAGs); Hamming distance for diversity	Fitness: Bayesian posterior probability of model fit to data; multi-run consistency	Parameter bounds; mutation only when diversity is lost; avoids crossover between similar individuals

- Parameter optimization strategies:Reported in 2 of 3 studies; both used evolutionary algorithms (Differential Evolution, Scatter Search, or Genetic Algorithms). One study used asynchronous, cooperative, self-adaptive parallelization, and another used diversity strategies such as Hamming distance.

- Fitness function implementation:Described in 2 studies: one used parameter estimation error, and one used Bayesian posterior probability of model fit to data. No mention found of the fitness function in 1 study.
- Biological constraints handling:Addressed in 2 studies, both using parameter bounds for biological plausibility. One study also used adaptive strategies for robustness, and another used diversity strategies (mutation only when diversity is lost, avoiding crossover between similar individuals). No mention found of biological constraints handling in 1 study.

Mathematical Modeling of Gene Regulatory Networks

Study	ODE System Representation	Dynamic Behavior Analysis	Parameter Sensitivity
Buriticá et al., 2018	Petri net and Ordinary Differential Equations (ODEs) for gene expression; stochastic simulation via Gillespie algorithm	Comparison of deterministic (ODE) and stochastic (Gillespie) dynamics; analysis of state distributions	Sensitivity analysis via parameter variation; Analysis of Variance (ANOVA) for significance
Rodríguez Penas, 2017	Nonlinear ODEs for dynamic biological systems; high-dimensional models	Focus on parameter estimation and model calibration; scalability to large systems	Adaptive strategies to enhance robustness; statistical analysis of parameter effects
John & Meza-Chaves, 2019	Bayesian networks (Directed Acyclic Graphs) for gene interactions; not ODE-based	Dynamic inference from time-series data; model fit assessed via Bayesian posterior	Diversity maintenance in Genetic Algorithm; validation on engineered and real data

- ODE System Representation:
  - ODE-based models: 2 of 3 studies.
    - Of these, 1 also used Petri nets and stochastic simulation (Gillespie algorithm).
    - 1 used nonlinear, high-dimensional ODEs.
  - Non-ODE approach (Bayesian networks/DAGs): 1 of 3 studies.
- Dynamic Behavior Analysis:
  - Deterministic vs stochastic comparisons: 1 study.
  - State distribution analysis: 1 study.
  - Parameter estimation and model calibration: 1 study.
  - Scalability analysis: 1 study.
  - Dynamic inference from time-series data: 1 study.
  - Bayesian model fit assessment: 1 study.
- Parameter Sensitivity:
  - Sensitivity analysis via parameter variation: 1 study.
  - Analysis of Variance (ANOVA) or other statistical tests for parameter significance: 1 study.
  - Adaptive strategies to enhance robustness: 1 study.
  - Statistical analysis of parameter effects: 1 study.
  - Genetic Algorithm diversity maintenance: 1 study.

  - Validation of parameter sensitivity on engineered and real data: 1 study.

No mention found of studies combining all three modeling approaches (ODEs, Petri nets, and Bayesian networks) in a single framework. Each study reported some form of parameter sensitivity analysis, but the specific methods varied.

Biological Applications and Validation

Study	Cell Reprogramming Application	Model Validation Approach	Biological Relevance of Results
Buriticá et al., 2018	Not addressed	Comparison with literature (Goss & Peccoud, 1998); sensitivity analysis; Analysis of Variance (ANOVA)	Highlights importance of stochasticity in gene expression; deterministic models insufficient for population variability
Rodríguez Penas, 2017	Not addressed	Benchmarking on parameter estimation and Mixed-Integer Dynamic Optimization (MIDO) problems; statistical analysis	Demonstrates computational feasibility for large-scale biological models; no mention found of direct biological insights
John & Meza-Chaves, 2019	Not directly addressed, but relevant to gene regulatory network (GRN) inference	Validation on engineered signals and experimental Arabidopsis data; comparison with known relationships	Identifies known and novel gene interactions in auxin response; potential for broader gene regulatory network discovery

- Cell reprogramming:Not directly addressed in any of the three studies; two studies did not address it, and one study was relevant to gene regulatory network inference but did not directly address reprogramming.
- Model validation approaches:
  - Comparison with literature, sensitivity analysis, and Analysis of Variance (ANOVA): 1 study.
  - Benchmarking on parameter estimation and Mixed-Integer Dynamic Optimization (MIDO) problems, with statistical analysis: 1 study.
  - Validation on engineered signals and experimental Arabidopsis data, and comparison with known relationships: 1 study.
- Biological relevance:
  - Importance of stochasticity in gene expression and the insufficiency of deterministic models for population variability: 1 study.
  - Computational feasibility for large-scale biological models, but no mention found of direct biological insights: 1 study.
  - Identification of known and novel gene interactions in auxin response, with potential for broader gene regulatory network discovery: 1 study.

Implementation Considerations

Study	Parameter Types	Value Ranges	Biological Significance	Optimization Impact
Buriticá et al., 2018	Activation, inactivation, synthesis rates	Not specified; varied in sensitivity analysis	Reflects gene activation/inactivation and protein synthesis	Affects simulation outcomes; not used for optimization
Rodríguez Penas, 2017	ODE model parameters (varied, high-dimensional)	Bounded for biological plausibility	Represents kinetic rates, regulatory strengths	Parameter bounds critical for feasible solutions; adaptive strategies improve robustness
John & Meza-Chaves, 2019	Network structure (edges), interaction probabilities	Parameter bounds applied; Hamming distance for diversity	Encodes gene-gene regulatory relationships	Bounds and diversity maintenance prevent overfitting and ensure biological plausibility

- Parameter Types:
  - Kinetic or ODE model parameters (activation/inactivation/synthesis rates, or general ODE parameters): 2 studies.
  - Network structure (edges) and interaction probabilities as parameters: 1 study.
- Value Ranges:
  - Explicit parameter bounds for biological plausibility: 2 studies (one also used diversity maintenance).
  - Parameters varied in sensitivity analysis but no explicit bounds specified: 1 study.
- Optimization Impact:
  - Parameter bounds (and, in one case, diversity maintenance) were critical for feasible or robust optimization: 2 studies.
  - Parameters not used for optimization, only for simulation and sensitivity analysis: 1 study.

No mention found of studies using unbounded parameters for optimization. All studies that performed optimization emphasized the importance of parameter bounds or diversity constraints.

References

D. Buriticá, Bárbara Valeria Mejía Bohórquez, Lina Rojas, and Hector Leandro Sáenz Castro. "Análisis Estocástico de Un Sistema Génico Simple Para La Síntesis de Una Proteína Implementando Los Métodos de Gillespie." *Revista Cuarzo*, 2018.

David J. John, and Kenneth David Meza-Chaves. "Algoritmos Genéticos Como Mecanismo Para Modelar Interacciones Genéticas Utilizando Información Observada En Distintos Tiempos." *Revista Tecnología En Marcha*, 2019.

David Rodríguez Penas. "Optimization in Computational Systems Biology via High Performance Computing Techniques," 2017.