

SOFTWARE

Open Access



# JuPOETs: a constrained multiobjective optimization approach to estimate biochemical model ensembles in the Julia programming language

David M. Bassen<sup>2</sup>, Michael Vilkhovoy<sup>1</sup>, Mason Minot<sup>1</sup>, Jonathan T. Butcher<sup>2</sup> and Jeffrey D. Varner<sup>1\*</sup> 

## Abstract

**Background:** Ensemble modeling is a promising approach for obtaining robust predictions and coarse grained population behavior in deterministic mathematical models. Ensemble approaches address model uncertainty by using parameter or model families instead of single best-fit parameters or fixed model structures. Parameter ensembles can be selected based upon simulation error, along with other criteria such as diversity or steady-state performance. Simulations using parameter ensembles can estimate confidence intervals on model variables, and robustly constrain model predictions, despite having many poorly constrained parameters.

**Results:** In this software note, we present a multiobjective based technique to estimate parameter or models ensembles, the Pareto Optimal Ensemble Technique in the Julia programming language (JuPOETs). JuPOETs integrates simulated annealing with Pareto optimality to estimate ensembles on or near the optimal tradeoff surface between competing training objectives. We demonstrate JuPOETs on a suite of multiobjective problems, including test functions with parameter bounds and system constraints as well as for the identification of a proof-of-concept biochemical model with four conflicting training objectives. JuPOETs identified optimal or near optimal solutions approximately six-fold faster than a corresponding implementation in Octave for the suite of test functions. For the proof-of-concept biochemical model, JuPOETs produced an ensemble of parameters that gave both the mean of the training data for conflicting data sets, while simultaneously estimating parameter sets that performed well on each of the individual objective functions.

**Conclusions:** JuPOETs is a promising approach for the estimation of parameter and model ensembles using multiobjective optimization. JuPOETs can be adapted to solve many problem types, including mixed binary and continuous variable types, bilevel optimization problems and constrained problems without altering the base algorithm. JuPOETs is open source, available under an MIT license, and can be installed using the Julia package manager from the JuPOETs GitHub repository

**Keywords:** Ensemble modeling, Multiobjective optimization, Julia

\*Correspondence: [jdv27@cornell.edu](mailto:jdv27@cornell.edu)

<sup>1</sup>Department of Chemical and Biomolecular Engineering, Cornell University, 14853 Ithaca, NY, USA

Full list of author information is available at the end of the article

## Background

Ensemble modeling is a promising approach for obtaining robust predictions and coarse grained population behavior in deterministic mathematical models. It is often not possible to uniquely identify all the parameters in biochemical models, even when given extensive training data [1]. Thus, despite significant advances in standardizing biochemical model identification [2], the problem of estimating model parameters from experimental data remains challenging. Ensemble approaches address parameter uncertainty in systems biology and other fields like weather prediction [3–6] by using parameter families instead of single best-fit parameter sets. Parameter families can be selected based upon simulation error, along with other criteria such as diversity or steady-state performance. Simulations using parameter ensembles can estimate confidence intervals on model variables, and robustly constrain model predictions, despite having many poorly constrained parameters [7, 8]. There are many techniques to generate parameter ensembles. Battogtokh et al., Brown et al., and later Tasseff et al. generated experimentally constrained parameter ensembles using a Metropolis-type random walk [3, 5, 9, 10]. Liao and coworkers developed methods to generate ensembles that all approach the same steady-state, for example one determined by fluxomics measurements [11]. They have used this approach for model reduction [12], strain engineering [13, 14] and to study the robustness of non-native pathways and network failure [15]. Maranas and coworkers have also applied this method to develop a comprehensive kinetic model of bacterial central carbon metabolism, including mutant data [16]. We and others have used ensemble approaches, generated using both sampling and optimization techniques, that have robustly simulated a wide variety of signal transduction processes [9, 10, 17–19], neutrophil trafficking in sepsis [20], patient specific coagulation behavior [21], uncertainty quantification in metabolic kinetic models [22] and to capture cell to cell variation [23]. Further, ensemble approaches have been used in synthetic biology to sample possible biocircuit configurations [24]. Thus, ensemble approaches are widely used to robustly simulate a variety of biochemical systems.

Identification of biochemical models requires significant training data perhaps taken from diverse sources. These real-world data sets often contain intrinsic conflicts resulting from, for example, the use of different cell lines, different measurement technologies, different reagent vendors or lots, uncontrollable experimental artifacts or general cross laboratory variability. Parameter ensembles that optimally balance these inherent conflicts lead to more robust model performance. Multiobjective optimization is an ensemble generation technique that naturally balances conflicts in noisy training data [25].

Multiobjective optimization has been used to identify signal transduction models [18, 23], for the design of synthetic circuits [24], to design the folding behaviors of novel RNAs [26], to design bioprocesses [27], and to understand bacterial adaptation [28]. Thus, it is a widely used approach for a variety of biochemical applications. Previously, we developed the Pareto Optimal Ensemble Technique (POETs) algorithm to address the challenge of competing or conflicting training objectives. POETs, which integrates simulated annealing (SA) and multiobjective optimization through the notion of Pareto rank, estimates parameter ensembles which optimally trade-off between competing (and potentially conflicting) experimental objectives [29]. However, the previous implementation of POETs, in the Octave programming language [30], suffered from poor performance and was not configurable. For example, Octave-POETs does not accommodate user definable objective functions, bounds and problem constraints, cooling schedules, different variable types e.g., a mixture of binary and continuous design variables or custom diversity generation routines. Octave-POETs was also not well integrated into a package or source code management (SCM) system. Thus, upgrades to the approach containing new features, or bug fixes were not centrally managed.

## Implementation

In this software note, we present an open-source implementation of the Pareto optimal ensemble technique in the Julia programming language (JuPOETs). JuPOETs takes advantage of the unique features of Julia to address many of the shortcomings of the previous implementation. Julia is a cross-platform, high-performance programming language for technical computing that has performance comparable to C but with syntax similar to MATLAB/Octave and Python [31]. Julia also offers a sophisticated compiler, distributed parallel execution, numerical accuracy, and an extensive function library. Further, the architecture of JuPOETs takes advantage of the first-class function type in Julia allowing user definable behavior for all key aspects of the algorithm, including objective functions, custom diversity generation logic, linear/non-linear parameter constraints (and parameter bounds constraints) as well as custom cooling schedules. Julia's ability to naturally call other languages such as Python or C also allows JuPOETs to be used with models implemented in a variety of languages across many platforms. Additionally, Julia offers a built-in package manager which is directly integrated with GitHub, a popular web-based Git repository hosting service offering distributed revision control and source code management. Thus, JuPOETs can be adapted to many problem types, including mixed binary and continuous variable types, bilevel problems

and constrained problems without altering the base algorithm, as was required in the previous POETs implementation.

**JuPOETs optimization problem formulation**

JuPOETs solves the  $\mathcal{K}$ -dimensional constrained multiobjective optimization problem:

$$\min_{\mathbf{p}} \begin{cases} O_1(\mathbf{x}(t, \mathbf{p}), \mathbf{p}) \\ \vdots \\ O_{\mathcal{K}}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}) \end{cases} \quad (1)$$

subject to the model equations and constraints:

$$\begin{aligned} \mathbf{f}(t, \mathbf{x}(t, \mathbf{p}), \dot{\mathbf{x}}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}) &= \mathbf{0} \\ g_1(t, \mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}) &\geq 0 \\ &\vdots \\ g_C(t, \mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}) &\geq 0 \end{aligned}$$

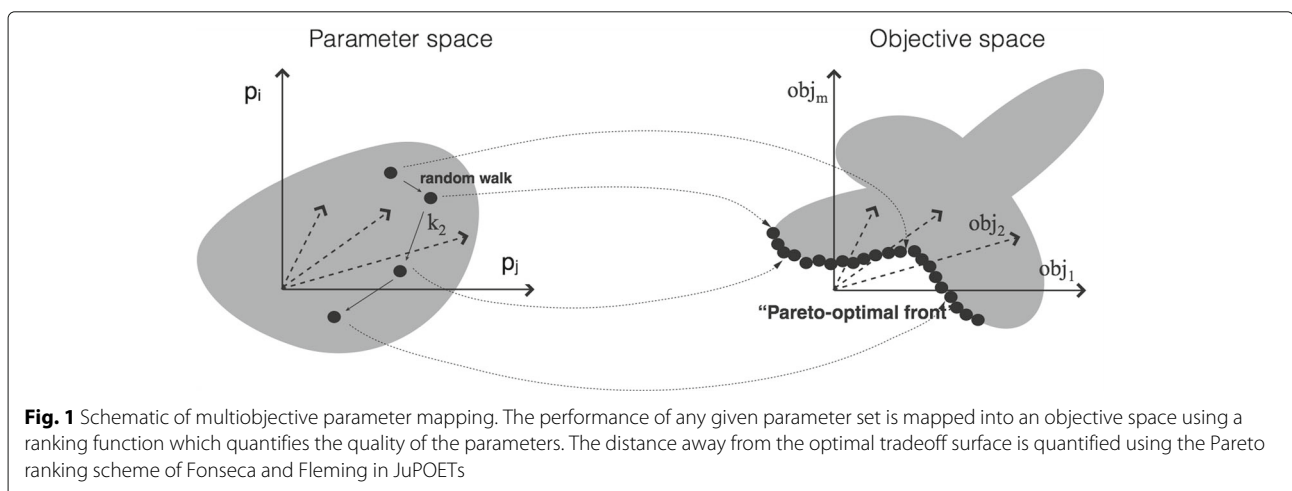
and parameter bound constraints:

$$\mathcal{L} \leq \mathbf{p} \leq \mathcal{U}$$

The quantity  $O_j$  denotes the  $j^{th}$  objective function ( $j = 1, 2, \dots, \mathcal{K}$ ), typically the sum of squared errors for the  $j^{th}$  data set for biochemical modeling applications. The terms  $\mathbf{f}(t, \mathbf{x}(t, \mathbf{p}), \dot{\mathbf{x}}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p})$  denote the system of model equations (e.g., differential equations, differential algebraic equations or linear/non-linear algebraic equations) where  $\mathbf{p}$  denotes the decision variable vector e.g., unknown model parameters ( $\mathcal{D} \times 1$ ). In typical biochemical modeling applications, the model equations  $\mathbf{f}(\cdot)$  are a system of continuous real-valued non-linear differential equations that comprise a kinetic model, but other types of models e.g., stoichiometric models are also common. The quantity  $t$  denotes time,  $\mathbf{x}(t, \mathbf{p})$  denotes the model state (with an initial state  $\mathbf{x}_0$ ), and  $\mathbf{u}(t)$  denotes an

input vector. The decision variables (e.g., kinetic parameters) can be subject to bounds constraints, where  $\mathcal{L}$  and  $\mathcal{U}$  denote the lower and upper bounds, respectively as well as  $\mathcal{C}$  problem specific constraints  $g_i(t, \mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}), i = 1, \dots, \mathcal{C}$ . The decision variables  $\mathbf{p}$  are typically real-valued kinetic constants, or metabolic fluxes in the case of stoichiometric models. However, other variables types e.g., binary or categorical decision variables can also be accommodated.

JuPOETs integrates simulated annealing (SA) [32] with Pareto ranking to estimate decision variables on or near the optimal tradeoff surface between competing objectives (Fig. 1 and Algorithm 1). A tradeoff surface defines the best possible performance for every conflicting objective, such that an increase in the performance of one objective does not decrease the performance of at least one other objective. Pareto rank is a scalar measure of distance away from the optimal tradeoff surface (low rank is near the surface, while higher ranks are progressively further away). Thus, the central idea underlying POETs is a mapping between the value of the objective vector evaluated at  $\mathbf{p}_{i+1}$  (decision variable guess at iteration  $i + 1$ ) and the scalar Pareto rank (Fig. 1). Traditional simulated annealing uses a scalar performance value e.g., simulation error to make a probabilistic decision to keep or reject a set of decision variables; decision variables with better performance are always accepted, while those with worse performance are sometimes accepted depending upon a parameter called the temperature. On the other hand, JuPOETs makes this same decision using the Pareto rank instead of a single performance objective. The problem of estimating biochemical model parameters from experimental data is typically posed as an error minimization problem over continuous real-valued decision variables (model parameters) subject to the model equations. A parameter set  $\mathbf{p}_{i+1}$  lies along the optimal tradeoff surface if no other parameter guess leads to decreased error for



---

**Algorithm 1:** Pseudo-code for the JuPOETs run-loop. The user must specify the objective function and an initial parameter guess. The user can optionally specify the neighbor, acceptance, cooling and refinement functions (or use the default implementations). The rank archive  $\mathcal{R}$ , solution archive  $\mathcal{S}$  and objective archive  $\mathcal{O}$  are initialized from the initial guess. The initial guess (potentially following a single objective local refinement step) is perturbed in the neighbor function, which generates a new solution whose performance is evaluated using the user supplied objective function. The new solution and objective values are then added to the respective archives and ranked using the builtin rank function. If the new solution is accepted (based upon a probability calculated with the user supplied acceptance function) it is added to the solution and objective archive. This solution is then perturbed during the next iteration of the algorithm. However, if the solution is not accepted, it is removed from the archive and discarded. The temperature is adjusted using the user supplied cooling function after each  $\mathcal{I}$  iterations. When JuPOETs terminates, the parameter solution archive  $\mathcal{S}$ , objective archive  $\mathcal{O}$  and rank archive  $\mathcal{R}$  are returned to the caller.

---

**input** : User specified objective function, and initial guess ( $\mathcal{D} \times 1$ ). User can also specify custom neighbor, acceptance, cooling and refinement functions or use the default functions provided.

**Output:** Rank archive  $\mathcal{R}$  ( $\mathcal{A} \times 1$ ), parameter solution archive  $\mathcal{S}$  ( $\mathcal{D} \times \mathcal{A}$ ) and objective archive  $\mathcal{O}$  ( $\mathcal{K} \times \mathcal{A}$ ), where  $\mathcal{A}$  denotes the number of accepted solutions

```

1 initialize:  $\mathcal{R}$ ,  $\mathcal{S}$  and  $\mathcal{O}$  using initial guess  $\mathbf{p}_0$ ;
2 initialize:  $T \leftarrow 1.0$ ;
3 initialize:  $T_{min} \leftarrow 1/10000$ ;
4 initialize: Maximum number of steps per temperature  $\mathcal{I}$ ;

// Call to local refinement function (single objective problem)
5  $\mathbf{p}_o \leftarrow \text{user-function::refinement}(\mathbf{p}_o)$ ;

6 while  $T > T_{min}$  do
7    $i \leftarrow 1$ ;
8   while  $i < \mathcal{I}$  do
9     // Generate a new parameter solution using user neighbor function
10     $\mathbf{p}_{i+1} \leftarrow \text{user-function::neighbor}(\mathbf{p}^*)$ ;
11    // Evaluate  $\mathbf{p}_{i+1}$  using user objective function
12     $\mathbf{o}_{i+1} \leftarrow \text{user-function::objective}(\mathbf{p}_{i+1})$ ;
13    Add  $\mathbf{p}_{i+1}$  to solution archive  $\mathcal{S}$ ;
14    Add  $\mathbf{o}_{i+1}$  to objective archive  $\mathcal{O}$ ;
15    // Calculate Pareto rank of solutions in  $\mathcal{O}$  using builtin rank function
16     $\mathcal{R} \leftarrow \text{builtin-function::rank}(\mathcal{O})$ ;
17    // Accept  $\mathbf{p}_{i+1}$  into the archive with user defined probability
18     $\mathcal{P} \leftarrow \text{user-function::acceptance}(\mathcal{R}, T)$ ;
19    if  $\mathcal{P} > \text{rand}$  then
20      // Update the best solution with  $\mathbf{p}_{i+1}$ 
21       $\mathbf{p}^* \leftarrow \mathbf{p}_{i+1}$ ;
22      prune  $\mathcal{S}$ ,  $\mathcal{R}$  and  $\mathcal{O}$  of all solutions above a rank threshold;
23    else
24      Remove  $\mathbf{p}_{i+1}$  from solution archive  $\mathcal{S}$ ;
25      Remove  $\mathbf{o}_{i+1}$  from error archive  $\mathcal{O}$ ;
26    end
27     $i \leftarrow i + 1$ ;
28  end
29  // Update  $T$  using the user cooling function
30   $T \leftarrow \text{user-function::cooling}(T)$ ;
31 end

```

---

every objective. JuPOETs calculates the performance of a candidate parameter set  $\mathbf{p}_{i+1}$  by calling the user defined objective function; `objective` takes a parameter set as an input, evaluates the model equations, and using this solution, returns the  $\mathcal{K} \times 1$  objective vector. Candidate parameter sets are generated by the user supplied `neighbor` function; the default implementation of `neighbor` is a random perturbation, however other perturbation logic can be implemented by the user. The error vector associated with  $\mathbf{p}_{i+1}$  is ranked using the builtin Pareto rank function, by comparing the error at iteration  $i + 1$  to the error archive  $\mathcal{O}_i$  (all error vectors up to iteration  $i$  meeting a ranking criterion). Parameter sets on or near the optimal trade-off surface between the objectives have a rank equal to 0 (no other current parameter sets are better). These rank zero parameter sets define the Pareto optimal group for the ensemble, wherein Pareto optimality is defined as a parameter set not being dominated by any other sets within the ensemble. Sets with increasing non-zero rank are progressively further away from the optimal trade-off surface. Thus, a parameter set with a rank = 0 is *better* in a trade-off sense than rank > 0. We implemented the Fonseca and Fleming ranking scheme in the builtin rank function [33]:

$$\text{rank}(\mathcal{O}_{i+1}(\mathbf{p}_{i+1}) \mid \mathcal{O}_i) = r \tag{2}$$

where rank  $r$  is the number of parameter sets that dominate (are better than) parameter set  $\mathbf{p}_{i+1}$ , and  $\mathcal{O}_{i+1}(\mathbf{p}_{i+1})$  denotes the objective vector evaluated at  $\mathbf{p}_{i+1}$ . We used the Pareto rank to inform the SA calculation. The parameter set  $\mathbf{p}_{i+1}$  was accepted or rejected by the SA at each iteration, by calculating an acceptance probability  $\mathcal{P}(\mathbf{p}_{i+1})$ :

$$\mathcal{P}(\mathbf{p}_{i+1}) \equiv \exp\{-\text{rank}(\mathcal{O}_{i+1}(\mathbf{p}_{i+1}) \mid \mathcal{O}_i) / T\} \tag{3}$$

where  $T$  is the simulated annealing temperature; the temperature provides control over how strictly decreasing Pareto rank is enforced. As  $\text{rank}(\mathcal{O}_{i+1}(\mathbf{p}_{i+1}) \mid \mathcal{O}_i) \rightarrow 0$ , the acceptance probability moves toward one, ensuring that we explore parameter sets along the Pareto surface. Occasionally, (depending upon  $T$ ) a parameter set with a high Pareto rank is accepted by the SA allowing a more diverse search of the parameter space. However, as  $T$  is reduced as a function of iteration count (using the `cooling` function), the probability of accepting a high-rank set decreases. Parameter sets could also be accepted by the SA but *not* permanently archived in  $\mathcal{S}_i$ , where  $\mathcal{S}_i$  is the solution archive. Only parameter sets with rank less than or equal to a threshold (rank  $\leq 4$  by default) are included in  $\mathcal{S}_i$ , where the archive is re-ranked and filtered after accepting every new parameter set. Parameter bounds were implemented in the `neighbor` function as box constraints, while problem specific constraints were implemented in `objective` using a penalty method:

$$O_i + \lambda \sum_{j=1}^{\mathcal{C}} \min\{0, g_j(t, \mathbf{x}(t, \mathbf{p}), \mathbf{u}(t, \mathbf{p}))\} \quad i = 1, \dots, \mathcal{K} \tag{4}$$

where  $\lambda$  denotes the penalty parameter ( $\lambda = 100$  by default). However, because both the `neighbor` and objective functions are user defined, different constraint implementations are easily defined.

To use JuPOETs, the user specifies the `neighbor`, `acceptance`, `cooling` and `objective` functions along with an initial decision variable guess. Default implementations of the `neighbor`, `acceptance` and `cooling` functions can be used directly, or they can be overridden by user defined logic. However, the user must provide an implementation of the `objective` function and provide an initial decision variable guess. Lastly, if the user is operating JuPOETs in hybrid mode, then a `refinement` function pointer must also be specified. Hybrid mode temporarily switches the search from a multiobjective to a single objective problem, where the sum of the objective functions can be used to update the best (or initial) parameter guess. The specific hybrid mode search logic is up to the user; by default hybrid mode is off, and the default `refinement` implementation is simply a pass through function. However, we have shown previously that POETs operated in hybrid mode (where the single objective problem used a pattern search approach) had better performance than POETs alone [29]. Thus, hybrid mode is generally recommended for most applications. In addition, there are several user configurable parameters that can be adjusted to control the performance of JuPOETs: `maximum_number_of_iterations` controls the number of iterations per temperature (default 20); `rank_cutoff` controls the upper rank bound on the solution archive (default 5); `temperature_min` controls the minimum temperature after which JuPOETs returns the error and solution archives (default 0.001); `show_trace` controls the level of output shown to the user (default true). After the completion of the run, JuPOETs returns the parameter solution archive  $\mathcal{S}$ , objective archive  $\mathcal{O}$  and rank archive  $\mathcal{R}$ . The parameter solution archive  $\mathcal{S}$  contains is an  $\mathcal{D} \times \mathcal{A}$  array, where  $\mathcal{A}$  denotes the number of solutions in the archive when JuPOETs terminated. On the other hand, the objective archive  $\mathcal{O}$  is an  $\mathcal{K} \times \mathcal{A}$  array containing the performance values for each objective corresponding the columns of  $\mathcal{S}$ . Lastly, JuPOETs returns the rank archive  $\mathcal{R}$  which is an  $\mathcal{A} \times 1$  array of Pareto ranks corresponding to the columns of  $\mathcal{S}$ . One technical note, if JuPOETs is run from multiple starting locations, and the archives from each of these runs is combined into a single collective archive, the combined parameter rank archive may become invalid. In these cases, it is required to re-rank the parameter sets

using the built-in rank function to produce a collective parameter ranking.

**Results and discussion**

JuPOETs identified optimal or nearly optimal solutions significantly faster than Octave-POETs for a suite of multiobjective algebraic test problems (Table 1). The algebraic test problems were constrained non-linear functions with bound constraints and additional non-linear constraints on the decision variables in one case. The problems had up to three-dimensional continuous real-valued decision vectors, and each case had two objective functions. The wall-clock time for JuPOETs and Octave-POETs was measured for 10 independent trials for each of the test problems. The same `cooling`, `neighbor`, `acceptance`, and `objective` logic was employed between the implementations, and all other parameters were held constant. For each test function, the search domain was partitioned into 10 segments, where an initial parameter guess was drawn from each partition. The number of search steps for each temperate was  $\mathcal{I} = 10$  for all cases, and the cooling parameter was  $\alpha = 0.9$ . On average, JuPOETs identified optimal or near optimal solutions for the suite of test problems six-fold faster (60s versus 400s) than Octave-POETs (Fig. 2). JuPOETs produced the characteristic tradeoff curves for each test problem, given both decision variable bound and problem constraints (Fig. 3). Thus, JuPOETs estimated an ensemble of solutions to constrained multiobjective algebraic test problems significantly faster than the current Octave implementation. Next, we tested JuPOETs on a proof-of-concept biochemical model identification problem.

JuPOETs estimated an ensemble of biochemical model parameters that were consistent with the mean of synthetic training data (Fig. 4). Four synthetic training data sets were generated from a prototypical biochemical network consisting of 6 metabolites and 7 reactions (Fig. 4, inset right). We considered a common case in which

the same extracellular measurements of  $A_e, B_e, C_e$  and cellmass were made on four hypothetical cell types, each having the same biological connectivity but different performance. Network dynamics were modeled using the hybrid cybernetic model with elementary modes (HCM) approach of Ramkrishna and coworkers [34]. In the HCM approach, metabolic networks are first decomposed into a set of elementary modes (EMs) (chemically balanced steady-state pathways, see [35]). Dynamic combinations of elementary modes are then used to characterize network behavior. Each elementary mode is catalyzed by a pseudo enzyme; thus, each mode has both kinetic and enzyme synthesis parameters. The proof of concept network generated 6 EMs, resulting in 13 model parameters (continuous real-valued decision variables). The synthetic training data was generated by randomly varying these parameters.

The general form of the biochemical test problem was given by:

$$\min_{\mathbf{p}} (O_1, \dots, O_{\mathcal{K}}) \tag{5}$$

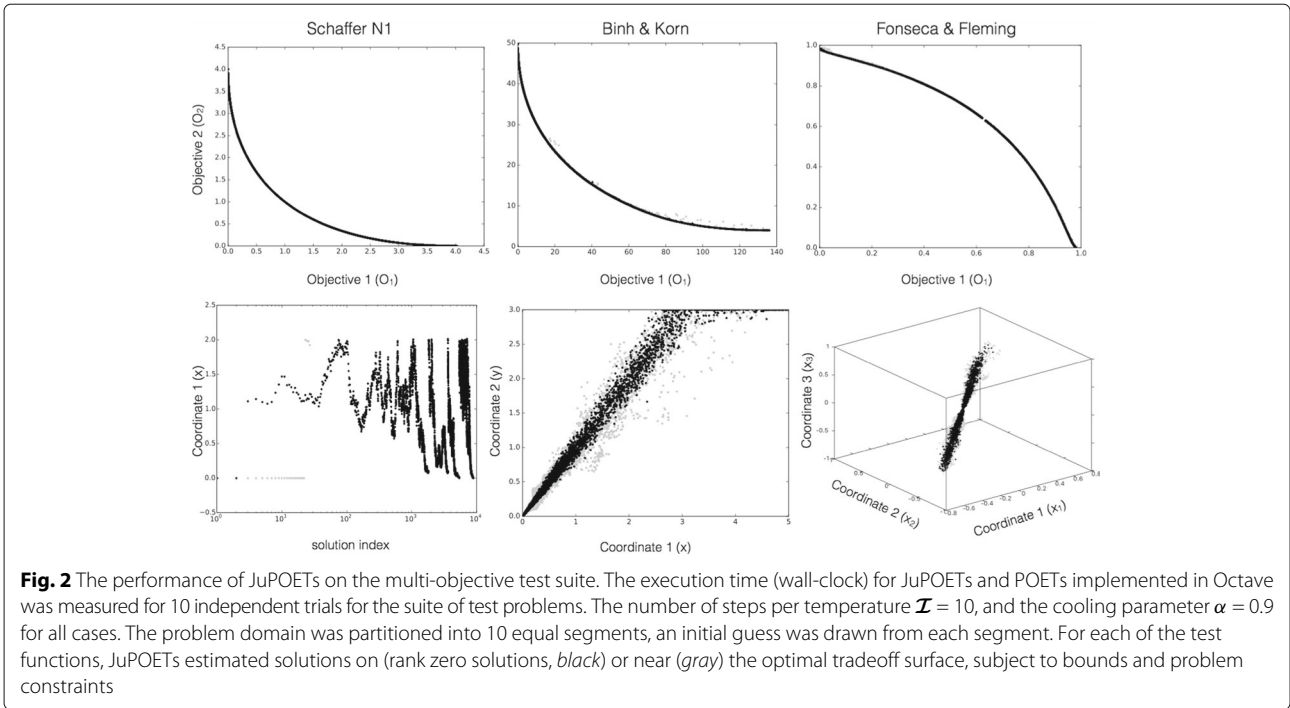
subject to model and bounds constraints. We considered four training data sets ( $\mathcal{K} = 4$ ), each of which contained time-series measurements of  $A_e, B_e, C_e$  and cellmass. Each objective  $O_j$ ,  $j = 1, \dots, \mathcal{K}$  quantified the squared difference between the simulated ( $x_i$ ) and measured extracellular species abundance ( $y_i$ ) in the  $j^{th}$  data set:

$$O_j = \sum_i \sum_{\tau} (x_i(\tau) - y_i(\tau))^2 \quad j = 1, \dots, \mathcal{K} \tag{6}$$

where,  $i$  denotes the species index and  $\tau$  denotes the time index. The abundance of extracellular species  $i$  ( $x_i$ ), the pseudo enzyme  $e_l$  (catalyzes flux through mode  $l$ ), and cellmass were governed by the model equations:

**Table 1** Multi-objective optimization test problems. We tested the JuPOETs implementation on three two-dimensional test problems, with one-, two- and three-dimensional parameter vectors. Each problem had parameter bounds constraints, however, on the Binh and Korn function had additional non-linear problem constraints. For the Fonesca and Fleming problem,  $N = 3$

Name	Dimension	Function	Domain	Constraints
Schaffer function	1	$O_1(x) = x^2$ $O_2(x) = (x - 2)^2$	$-10 \leq x \leq 10$	
Binh and Korn function	2	$O_1(x, y) = 4x^2 + 4y^2$ $O_2(x, y) = (x - 5)^2 + (y - 5)^2$	$0 \leq x \leq 5$ $0 \leq x \leq 3$	$g_1(x, y) = (x - 5)^2 + y^2 \leq 25$ $g_2(x, y) = (x - 8)^2 + (y + 3)^2 \leq 7.7$
Fonseca and Fleming function	3	$O_1(x_i) = 1 - \exp\left(-\sum_{i=1}^N \left(x_i - \frac{1}{\sqrt{N}}\right)^2\right)$ $O_2(x_i) = 1 - \exp\left(-\sum_{i=1}^N \left(x_i + \frac{1}{\sqrt{N}}\right)^2\right)$	$-4 \leq x_i \leq 4$	

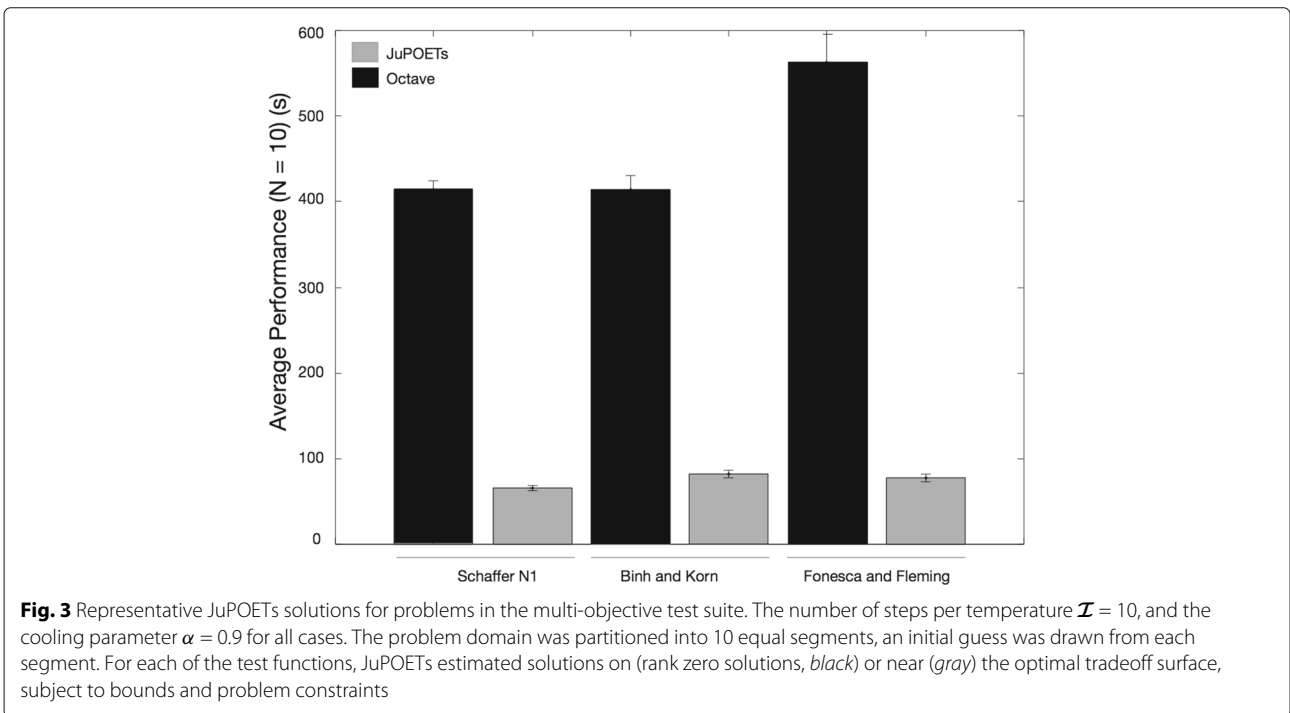


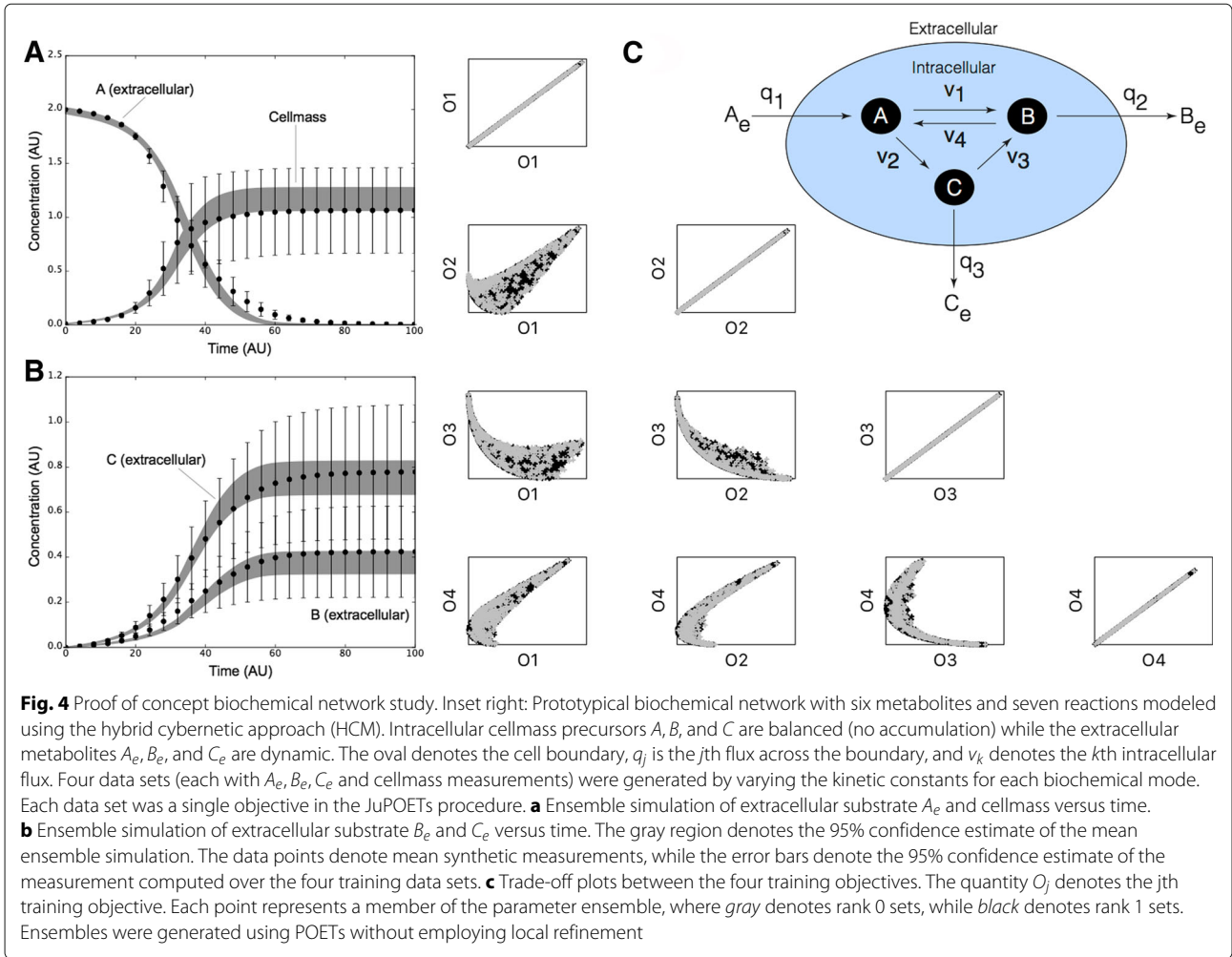
$$\frac{dx_i}{dt} = \sum_{j=1}^{\mathcal{R}} \sum_{l=1}^{\mathcal{L}} \sigma_{ij} z_{jl} q_l(\mathbf{e}, \mathbf{p}, \mathbf{x}) c \quad i = 1, \dots, \mathcal{M}$$

$$\frac{de_l}{dt} = \alpha_l + r_{El}(\mathbf{p}, \mathbf{x}) u_l - (\beta_l + r_G) e_l \quad l = 1, \dots, \mathcal{L}$$

$$\frac{dc}{dt} = r_G c$$

where  $\mathcal{R}$  and  $\mathcal{M}$  denote the number of reactions and extracellular species in the model and  $\mathcal{L}$  denotes the number of elementary modes. The quantity  $\sigma_{ij}$  denotes the stoichiometric coefficient for species  $i$  in reaction  $j$  and  $z_{jl}$  denotes the normalized flux for reaction  $j$  in mode  $l$ . If  $\sigma_{ij} > 0$ , species  $i$  is produced by reaction  $j$ ; if  $\sigma_{ij} < 0$ ,





species  $i$  is consumed by reaction  $j$ ; if  $\sigma_{ij} = 0$ , species  $i$  is not connected with reaction  $j$ . Extracellular species, cellmass and pseudo-enzyme were subject to the initial conditions  $\mathbf{x}(t_0) = \mathbf{x}_0$ ,  $c(t_0) = c_0$  and  $e_l = 0.5$ , respectively. The term  $q_l(\mathbf{e}, \mathbf{p}, \mathbf{x})$  denotes the specific uptake/secretion rate for mode  $l$  where  $\mathbf{e}$  denotes the pseudo enzyme vector,  $\mathbf{p}$  denotes the unknown kinetic parameter vector (decision variables),  $\mathbf{x}$  denotes the extracellular species vector, and  $c$  denotes the cell mass;  $q_l(\mathbf{e}, \mathbf{p}, \mathbf{x})$  is the product of a kinetic term ( $\bar{q}_l$ ) and a control variable governing enzyme activity. Flux through each mode was catalyzed by a pseudo enzyme  $e_l$ , synthesized at the regulated specific rate  $r_{E,l}(\mathbf{p}, \mathbf{x})$ , and constitutively at the rate  $\alpha_l$ . The term  $u_l$  denotes the cybernetic variable controlling the synthesis of enzyme  $l$ . The term  $\beta_l$  denotes the rate constant governing non-specific enzyme degradation, and  $r_G$  denotes the specific growth rate through all modes. The specific uptake/secretion rates and the specific rate of enzyme synthesis were modeled using saturation kinetics. The specific growth rate was given by:

$$r_G = \sum_{l=1}^{\mathcal{L}} z_{\mu l} q_l(\mathbf{e}, \mathbf{p}, \mathbf{x})$$

where  $z_{\mu l}$  denotes the growth flux  $\mu$  through mode  $l$ . The control variables  $u_l$  and  $v_l$ , which control the synthesis and activity of each enzyme respectively, were given by:

$$u_l = \frac{z_{s_l} \bar{q}_l}{\sum_{l=1}^{\mathcal{L}} z_{s_l} \bar{q}_l} \tag{7}$$

and

$$v_l = \frac{z_{s_l} \bar{q}_l}{\max_{l=1, \dots, \mathcal{L}} z_{s_l} \bar{q}_l} \tag{8}$$

where  $z_{s_l}$  denotes the uptake flux of substrate  $s$  through mode  $l$ . Each unknown kinetic parameter was continuous and real-valued, and subject to bounds constraints:  $\mathcal{L} \leq \mathbf{p} \leq \mathcal{U}$ .

JuPOETs produced an ensemble of approximately  $\dim \mathcal{S} \simeq 13,000$  parameter sets that captured the mean



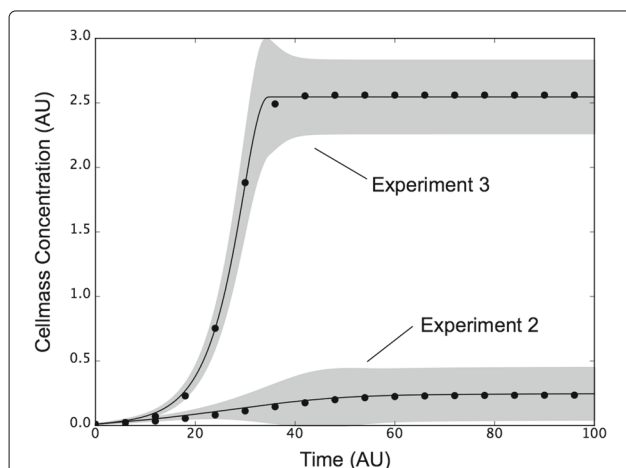
of the measured data sets for extracellular metabolites and cellmass (Fig. 4a and b). JuPOETs minimized the difference between the simulated and measured values for extracellular metabolites  $A_e$ ,  $B_e$ ,  $C_e$  and cellmass, where the residual for each data set was treated as a single objective (leading to four objectives). The 95% confidence estimate produced by the ensemble was consistent with the mean of the measured data, despite having significant uncertainty in the training data. JuPOETs produced a consensus estimate of the synthetic data by calculating optimal trade-offs between the training data sets (Fig. 4c). Multiple trade-off fronts were visible in the objective plots, for example between data set 3 ( $O_3$ ) and data set 2 ( $O_2$ ). Thus, without a multiobjective approach, it would be challenging to capture these data sets as fitting one leads to decreased performance on the other. However, the ensemble contained parameter sets that described each data set independently (Fig. 5). Thus, JuPOETs produced an ensemble of parameters that gave the mean of the training data for conflicting data sets, while simultaneously estimating parameter sets that performed well on each individual objective function.

Currently, JuPOETs does not consider parameter identifiability when constructing parameter ensembles. Although JuPOETs produces parameter estimates that give model performance similar to the training data, we do not have strict statistical confidence that the *true* parameter values are contained within the ensemble. However, despite this, ensembles produced by POETs can be predictive [18, 23]. Thus, JuPOETs produces a collection of

parameters that are constrained by the performance of the model, and not by specific hypotheses regarding the individual values of the raw model parameters. Of course, knowledge of specific parameter values, or the relationship between parameter combinations, can be used to inform the search through either bounds or problem specific constraints (for example, as demonstrated in the first example problem).

## Conclusions

In this software note, we presented JuPOETs, a multiobjective technique to estimate parameter ensembles in the Julia programming language. JuPOETs is open source, and available for download under an MIT license from the JuPOETs GitHub repository at <https://github.com/varnerlab/POETs.jl>. We demonstrated JuPOETs on a suite of algebraic test problems, and a proof-of-concept ODE based biochemical model. While JuPOETs outperformed (and was significantly more flexible) than the previous Octave implementation, there are several areas that could be explored further. First, JuPOETs should be compared with other multiobjective evolutionary algorithms (MOEAs) to determine its relative performance on test and real world problems. Many evolutionary approaches e.g., the non-dominated sorting genetic algorithm (NSGA) family of algorithms, have been adapted to solve multiobjective problems [36, 37]. However, since there is a lack of open source Julia implementations of these alternative approaches, we did not benchmark the relative performance of JuPOETs in this note. One advantage that JuPOETs may have when compared to a strictly evolutionary approaches, is the inclusion of a local refinement step (hybrid mode), which temporarily reduces the problem to a single objective formulation. Previously, POETs run in hybrid mode led to better convergence on a proof-of-concept signal transduction model compared to the same approach without the hybrid refinement step [29]. Other hybrid multiobjective methods have also been shown to be more efficient than evolutionary approaches alone, for a variety of biochemical optimization problems [24, 38]. Thus, there are several different algorithms that we can use to benchmark, and improve the performance of JuPOETs, after we implement them in Julia. Another strategy to improve the performance of JuPOETs is to reduce the number (or cost) of function evaluations that are required to obtain optimal or near optimal solutions. For example, in many real world parameter estimation problems, the bulk of the execution time is spent evaluating the objective functions. One strategy to improve JuPOETs performance could be to optimize surrogates [39], while another would be parallel execution of the objective functions. Currently, JuPOETs serially evaluates the objective function vector. However, parallel evaluation of the objective functions e.g., using the `parallel` Julia



**Fig. 5** Experiment to experiment variation captured by the ensemble. Cellmass measurements (points) versus time for experiment 2 and 3 were compared with ensemble simulations. The full ensemble was sorted by simultaneously selecting the top 25% of solutions for each objective with rank  $\leq 1$ . The best fit solution for each objective (line)  $\pm 1$ -standard deviation (gray region) for experiment 2 and 3 brackets the training data despite significant differences the training values between the two data sets

macro or other techniques, could be implemented without significantly changing the JuPOETs run loop. Taken together, JuPOETs demonstrated improved flexibility, and performance over POETs in parameter identification and ensemble generation for multiple objectives. JuPOETs has the potential for widespread use due to the flexibility of the implementation, and the high level syntax and distribution tools native to the Julia programming language.

### Availability and requirements

JuPOETs is open source, available under an MIT software license. The JuPOETs source code is freely available from the JuPOETs GitHub repository at <https://github.com/varnerlab/POETs.jl>. All samples used in this study are included in the `sample/biochemical` and `sample/test_functions` subdirectories of the JuPOETs GitHub repository. JuPOETs can be run on all common.

**Operating system environments:** (Linux, Mac OS, Windows).

### Acknowledgements

We gratefully acknowledge Ani Chakrabarti, Russell Gould and Kathy Rogers for their input and suggestions regarding new features to include into JuPOETs. We also gratefully acknowledge the suggestions from the anonymous reviewers to improve this manuscript and JuPOETs.

### Funding

This study was supported by an award from the National Science Foundation (NSF CBET-0955172) and the National Institutes of Health (NIH HL110328) to J.B. and by a National Science Foundation Graduate Research Fellowship (DGE-1144153) to D.B. Lastly, J.V. was supported by an award from the US Army and Systems Biology of Trauma Induced Coagulopathy (W911NF-10-1-0376).

### Authors' contributions

JV developed the software presented in this study. MM and MV developed the proof-of-concept biochemical model. The manuscript was prepared and edited for publication by DB, JB and JV. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Consent for publication

Not applicable.

### Ethics approval and consent to participate

Not applicable.

### Author details

<sup>1</sup>Department of Chemical and Biomolecular Engineering, Cornell University, 14853 Ithaca, NY, USA. <sup>2</sup>Department of Biomedical Engineering, Cornell University, 14853 Ithaca, NY, USA.

Received: 8 June 2016 Accepted: 16 December 2016

Published online: 25 January 2017

### References

- Gadkar KG, Varner J, Doyle FJ. Model identification of signal transduction networks from data using a state regulator problem. *Syst Biol* (Stevenage). 2005;2(1):17–30.
- Gennemark P, Wedelin D. Benchmarks for identification of ordinary differential equations from time series data. *Bioinformatics*. 2009;25(6):780–6. doi:10.1093/bioinformatics/btp050.
- Battogtokh D, Asch DK, Case ME, Arnold J, Shüttler HB. An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of *Neurospora crassa*. *Proc Natl Acad Sci USA*. 2002;99(26):16904–9.
- Kuepfer L, Peter M, Sauer U, Stelling J. Ensemble modeling for analysis of cell signaling dynamics. *Nat Biotechnol*. 2007;25(9):1001–6. doi:10.1038/nbt1330.
- Brown KS, Sethna JP. Statistical mechanical approaches to models with many poorly known parameters. *Phys Rev E*. 2003;68:021904–19.
- Palmer TN, Shutts GJ, Hagedorn R, Doblas-Reyes FJ, Jung T, Leutbecher M. Representing model uncertainty in weather and climate prediction. *Ann Rev Earth Planetary Sci*. 2005;33:163–93.
- Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*. 2007;3(10):1871–8. doi:10.1371/journal.pcbi.0030189.
- Song S, Varner J. Modeling and analysis of the molecular basis of pain in sensory neurons. *PLoS ONE*. 2009;4:6758–72.
- Tasseff R, Nayak S, Salim S, Kaushik P, Rizvi N, Varner JD. Analysis of the molecular networks in androgen dependent and independent prostate cancer revealed fragile and robust subsystems. *PLoS ONE*. 2010;5(1):8864. doi:10.1371/journal.pone.0008864.
- Tasseff R, Nayak S, Song SO, Yen A, Varner JD. Modeling and analysis of retinoic acid induced differentiation of uncommitted precursor cells. *Integr Biol (Camb)*. 2011;3(5):578–91. doi:10.1039/c0ib00141d.
- Tran LM, Rizk ML, Liao JC. Ensemble modeling of metabolic networks. *Biophys J*. 2008;95(12):5606–17. doi:10.1529/biophysj.108.135442.
- Tan Y, Rivera JGL, Contador CA, Asenjo JA, Liao JC. Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux. *Metab Eng*. 2011;13(1):60–75. doi:10.1016/j.ymben.2010.11.001.
- Contador CA, Rizk ML, Asenjo JA, Liao JC. Ensemble modeling for strain development of L-lysine-producing *Escherichia coli*. *Metab Eng*. 2009;11(4–5):221–33. doi:10.1016/j.ymben.2009.04.002.
- Tan Y, Liao JC. Metabolic ensemble modeling for strain engineers. *Biotechnol J*. 2012;7(3):343–53. doi:10.1002/biot.201100186.
- Lee Y, Lafontaine Rivera JG, Liao JC. Ensemble modeling for robustness analysis in engineering non-native metabolic pathways. *Metab Eng*. 2014;25:63–71. doi:10.1016/j.ymben.2014.06.006.
- Khodayari A, Zomorodi AR, Liao JC, Maranas CD. A kinetic model of *Escherichia coli* core metabolism satisfying multiple sets of mutant flux data. *Metab Eng*. 2014;25:50–62. doi:10.1016/j.ymben.2014.05.014.
- Luan D, Zai M, Varner JD. Computationally derived points of fragility of a human cascade are consistent with current therapeutic strategies. *PLoS Comput Biol*. 2007;3(7):142. doi:10.1371/journal.pcbi.0030142.
- Song SO, Varner J. Modeling and analysis of the molecular basis of pain in sensory neurons. *PLoS ONE*. 2009;4(9):6758. doi:10.1371/journal.pone.0006758.
- Nayak S, Siddiqui JK, Varner JD. Modelling and analysis of an ensemble of eukaryotic translation initiation models. *IET Syst Biol*. 2011;5(1):2. doi:10.1049/iet-syb.2009.0065.
- Song SO, Song SOK, Hogg J, Peng ZY, Parker R, Kellum JA, Clermont G. Ensemble models of neutrophil trafficking in severe sepsis. *PLoS Comput Biol*. 2012;8(3):1002422. doi:10.1371/journal.pcbi.1002422.
- Luan D, Szlam F, Tanaka KA, Barie PS, Varner JD. Ensembles of uncertain mathematical models can identify network response to therapeutic interventions. *Mol Biosyst*. 2010;6(11):2272–86. doi:10.1039/b920693k.
- Andreozzi S, Miskovic L, Hatzimanikatis V. iSCHRUNK—in silico approach to characterization and reduction of uncertainty in the kinetic models of genome-scale metabolic networks. *Metab Eng*. 2016;33:158–68. doi:10.1016/j.ymben.2015.10.002.
- Lequieu J, Chakrabarti A, Nayak S, Varner JD. Computational modeling and analysis of insulin induced eukaryotic translation initiation. *PLoS Comput Biol*. 2011;7(11):1002263. doi:10.1371/journal.pcbi.1002263.
- Otero-Muras I, Banga JR. Multicriteria global optimization for biocircuit design. *BMC Syst Biol*. 2014;8:113. doi:10.1186/s12918-014-0113-3.
- Handl J, Kell DB, Knowles J. Multiobjective optimization in bioinformatics and computational biology. *IEEE/ACM Trans Comput Biol Bioinform*. 2007;4(2):279–92. doi:10.1109/TCBB.2007.070203.
- Taneda A. Multi-objective optimization for RNA design with multiple target secondary structures. *BMC Bioinformatics*. 2015;16(1):280. doi:10.1186/s12859-015-0706-x.

27. Sendin J, Otero-Muras I, Alonso AA, Banga J. Improved Optimization Methods for the Multiobjective Design of Bioprocesses. *Ind Eng Chem Res.* 2006;45:8594–603.
28. Angione C, Lió P. Predictive analytics of environmental adaptability in multi-omic network models. *Sci Rep.* 2015;5:15147. doi:10.1038/srep15147.
29. Song SO, Chakrabarti A, Varner JD. Ensembles of signal transduction models using pareto optimal ensemble techniques (poets). *Biotechnol J.* 2010;5(7):768–80. doi:10.1002/biot.201000059.
30. Eaton JW, Bateman D, Hauberg S. GNU octave version 3.0.1 manual: a high-level interactive language for numerical computations. North Charleston: CreateSpace Independent Publishing Platform; 2009.
31. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: A fresh approach to numerical computing. *arXiv CoRR.* abs/1411.1607. Ithaca: Cornell University; 2014.
32. Kirkpatrick S, Gelatt Jr CD, Vecchi MP. Optimization by simulated annealing. *Science.* 1983;220(4598):671–80. doi:10.1126/science.220.4598.671.
33. Fonseca CM, Fleming PJ. Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization. *arXiv CoRR* Publisher. In: Proceedings of the 5th International Conference on Genetic Algorithms. Ithaca: Cornell University; 1993. p. 416–23.
34. Kim J, Varner J, Ramkrishna D. A hybrid model of anaerobic e. coli gjt001: Combination of elementary flux modes and cybernetic variables. *Biotechnol Prog.* 2008;24(5):993–1006. doi:10.1002/btpr.73.
35. Schuster S, Fell DA, Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol.* 2000;18(3):326–2. doi:10.1038/73786.
36. Kalyanmoy D, Pratap A, Agarwal S, Meyarivan T. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Trans Evol Comp.* 2002;6:182–97.
37. Huband S, Hingston P, Barone L, While L. A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Trans Evol Comp.* 2006;10:477–506.
38. Sendin JOH, Otero-Muras I, Alonso AA, Banga JR. Improved optimization methods for multiple objective design of bioprocesses. *Ind Eng Chem Res.* 2006;45:8594–603.
39. Booker AJ, Dennis JE, Frank PD, Serafini DB, Torczon V, Trosset MW. A rigorous framework for optimization of expensive functions by surrogates. *Struct Optim.* 1999;17:1–13.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

