

METHODOLOGY ARTICLE

Open Access

Efficient characterization of high-dimensional parameter spaces for systems biology

Eliás Zamora-Sillero^{1,2,3*}, Marc Hafner^{1,3,4}, Ariane Ibig^{2,3}, Joerg Stelling^{2,3} and Andreas Wagner^{1,3,5,6}

Abstract

Background: A biological system's robustness to mutations and its evolution are influenced by the structure of its viable space, the region of its space of biochemical parameters where it can exert its function. In systems with a large number of biochemical parameters, viable regions with potentially complex geometries fill a tiny fraction of the whole parameter space. This hampers explorations of the viable space based on "brute force" or Gaussian sampling.

Results: We here propose a novel algorithm to characterize viable spaces efficiently. The algorithm combines global and local explorations of a parameter space. The global exploration involves an out-of-equilibrium adaptive Metropolis Monte Carlo method aimed at identifying poorly connected viable regions. The local exploration then samples these regions in detail by a method we call multiple ellipsoid-based sampling. Our algorithm explores efficiently nonconvex and poorly connected viable regions of different test-problems. Most importantly, its computational effort scales linearly with the number of dimensions, in contrast to "brute force" sampling that shows an exponential dependence on the number of dimensions. We also apply this algorithm to a simplified model of a biochemical oscillator with positive and negative feedback loops. A detailed characterization of the model's viable space captures well known structural properties of circadian oscillators. Concretely, we find that model topologies with an essential negative feedback loop and a nonessential positive feedback loop provide the most robust fixed period oscillations. Moreover, the connectedness of the model's viable space suggests that biochemical oscillators with varying topologies can evolve from one another.

Conclusions: Our algorithm permits an efficient analysis of high-dimensional, nonconvex, and poorly connected viable spaces characteristic of complex biological circuitry. It allows a systematic use of robustness as a tool for model discrimination.

Background

High-throughput experimental technologies have allowed biology to generate huge amounts of data. The enormity of these data sets permits a systemic view of the cell [1]. In this new framework mathematical models are immensely useful as compact representations of data [2], and as highly structured hypotheses that include underlying mechanisms of the processes under study. These models often consist of large systems of ordinary differential equations that govern the kinetics of proteins, mRNAs, and small molecules.

Mathematical modeling in biology faces several challenges that arise from uncertainty about relevant parameters. For example, the chemical reactions and the

corresponding kinetic equations governing any one biological system are only partially known [3,4]. Also, finding accurate numerical values for model parameters is virtually impossible, because many biochemical parameters cannot be measured directly. In addition, evolutionary processes can cause parameters to vary on evolutionary time scales, yet preserve system function. Thus, even a perfect mathematical model of an individual system might have limitations in describing other individuals of the same population that are sufficiently diverse genetically or epigenetically. In sum, it is often of limited use to identify a single best set of parameters for any one biochemical system. However, one can focus on a viable parameter *space* instead. This viable space is a subset of a space of biochemical parameters, where a model maintains a desirable behavior. Values of these parameters must lie inside

* Correspondence: e.zamora@bioc.uzh.ch

¹Department of Biochemistry, University of Zurich, Zurich, Switzerland
Full list of author information is available at the end of the article

the boundaries of this viable space for every organism in a population.

The investigation of viable spaces is closely linked to the analysis of robustness in biology. We here define robustness as the persistence, under perturbations, of a behavior that is characteristic for a system [5]. When focusing on robustness to changes in biochemical parameters that define system behavior, a biological system's robustness is a reflection of the topology and size of its viable space [6,7]. The volume of the viable space indicates the "amount" of parameter combinations that allow a system's desired behaviour. A small viable volume forces a precise tuning of biochemical parameters. In contrast, a large viable volume allows a system to successfully face changes in environmental conditions, because its parameters can change, sometimes by orders of magnitude, without impairing its function. Hence, robustness is associated with larger viable volumes.

The geometry of viable spaces also plays an important role in a system's robustness. Geometries that permit moderate parameter fluctuations without leaving the viable volume enhance robustness. In evolutionary terms, different ways of performing the same function - for instance, by conserved pathways with homologous yet different proteins [8] - can be traced back to a common ancestor and are thus "reachable" from each other [9]. A connected viable volume improves a system's evolvability and allows neutral evolutionary trajectories that may drive the system towards viable parameter points with high local robustness. Therefore, the robustness of a biological system can be a reflection of the geometry and size of its viable space.

A final motivation to characterize viable spaces comes from model building itself. As we pointed out above, some relevant components and interactions in cellular networks are typically unknown. It follows that the structure of mathematical models describing these networks contains uncertainties. These uncertainties may lead to qualitatively different models that match experimental observations equally well. In this case, robustness can be used as a tool to discriminate between more and less plausible models. Everything else being equal, a model can be considered superior if it is more robust than other plausible models [5,8].

The use of robustness for model discrimination raises the problem of how to measure robustness. Most robustness analyses in the literature are local (e.g. see [10-12] and references therein). They use a specific set of parameters, and their results do not reflect model behavior under all possible viable parameter sets. Some nonlocal approaches alter one or two parameters, and use bifurcation analysis to characterize the regions of a parameter space with similar qualitative model behavior [8,13-18].

These methods have serious limitations whenever multiple parameters have unknown values, which is usually the case. To address these limitations, a third group of techniques [7,19] use "glocal" approaches [20]. In a first "global" step of their analysis, these techniques obtain a sample of parameters from the viable space, and then, in a "local" analysis, they study the local robustness around every element of this set. In this way, they compute non-local measures of robustness, but they also face the problem of acquiring a large and statistically representative sample of viable parameter points. Therefore, they need efficient global methods to sample the viable space.

The main challenges for global methods typically result from the fact that parameter spaces can have many dimensions and a complex geometry, about which one has little prior knowledge. To characterize a viable space, some authors perform uniform sampling of the whole parameter space to identify regions where a model displays the desired behavior [8,21-25]. Determining this behavior typically involves integration of the model equations, which can become computationally very expensive when done for large samples. Even more fundamentally, the "curse of dimensionality" [26] makes the fraction of the whole parameter space occupied by viable parameters decrease exponentially with increasing dimension, i.e., increasing number of parameters. Therefore, "brute force" uniform sampling becomes quickly infeasible as model complexity increases. To avoid this problem, Hafner *et al.* [20] developed an algorithm that explores a parameter space by iterative Gaussian sampling. Briefly, in every iteration, this method determines the mean value and the covariance matrix of the identified viable points in parameter space to guide further sampling. However, the algorithm is only efficient when the viable region is convex and when enough viable points are found in each iteration.

Here, we propose an algorithm that overcomes these limitations. Specifically, it can efficiently characterize nonconvex and poorly connected viable spaces. The algorithm consists of two steps, namely a coarse-grained sampling of the viable space, which in turn delivers starting points for a finer-grained exploration. The sampled points also define a domain for subsequent volume computations by Monte Carlo integration, and for acquisition of a large set of uniformly distributed viable points. After describing the algorithm, we analyse a synthetic test problem involving a nonconvex and poorly connected viable space. This analysis will show that in high dimensional spaces our algorithm converges faster and identifies a larger proportion of the viable space than uniform sampling and Hafner's method. Moreover, in contrast to uniform sampling and Hafner's algorithm, whose performances scale exponentially with

the number of dimensions, our algorithm's performance scales linearly with the number of dimensions. Subsequently, we illustrate an application of our method to a biochemical circuit. To this end, we focus on a simplified model of biochemical oscillators with positive and negative feedback loops [27,28], in order to investigate the contributions of individual control loops to the robustness of oscillations in a narrow range of frequencies. Our algorithm allows us to characterize the non-convex viable space of this model. In spite of the model's simplicity, the geometry of this space shows well known properties of circadian oscillators. Specifically, it indicates that model topologies with an essential negative feedback loop and a nonessential positive feedback loop provide the most robust fixed period oscillations, as has been observed in different models of circadian oscillators [19,29-32]. In addition, the connectedness of the model's viable space suggests that biochemical oscillators with varying topologies can evolve from one another.

Methods

Viable regions

Given a model that involves d parameters, we define a parameter space as

$$\Theta^d = \Theta_1 \times \Theta_2 \times \dots \times \Theta_d, \quad (1)$$

where Θ_i is the interval of the real numbers \mathbb{R} for which the parameter θ_i is defined. We call the d -tuple $\theta = (\theta_1, \theta_2, \dots, \theta_d) \in \Theta^d$ a parameter point. It represents a configuration of the biochemical parameters involved in the model

(Figure 1). In addition, each parameter point has an associated value of a cost function

$$E(\theta) : \Theta^d \rightarrow \mathbb{R}^+, \quad (2)$$

that reflects how well a model produces a behavior under consideration. For a given θ , the lower the value of $E(\theta)$ the better the model behaves.

A parameter point θ is viable if it fulfills the condition

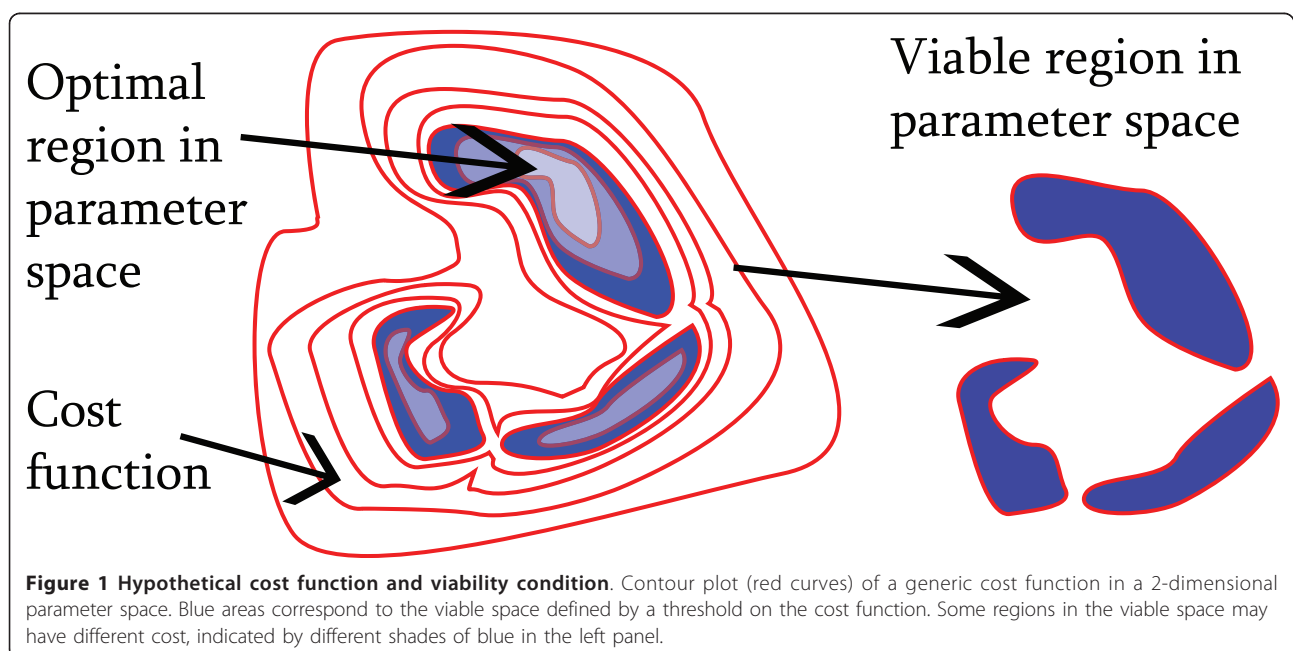
$$E(\theta) < E_0, \quad E_0 > 0, \quad (3)$$

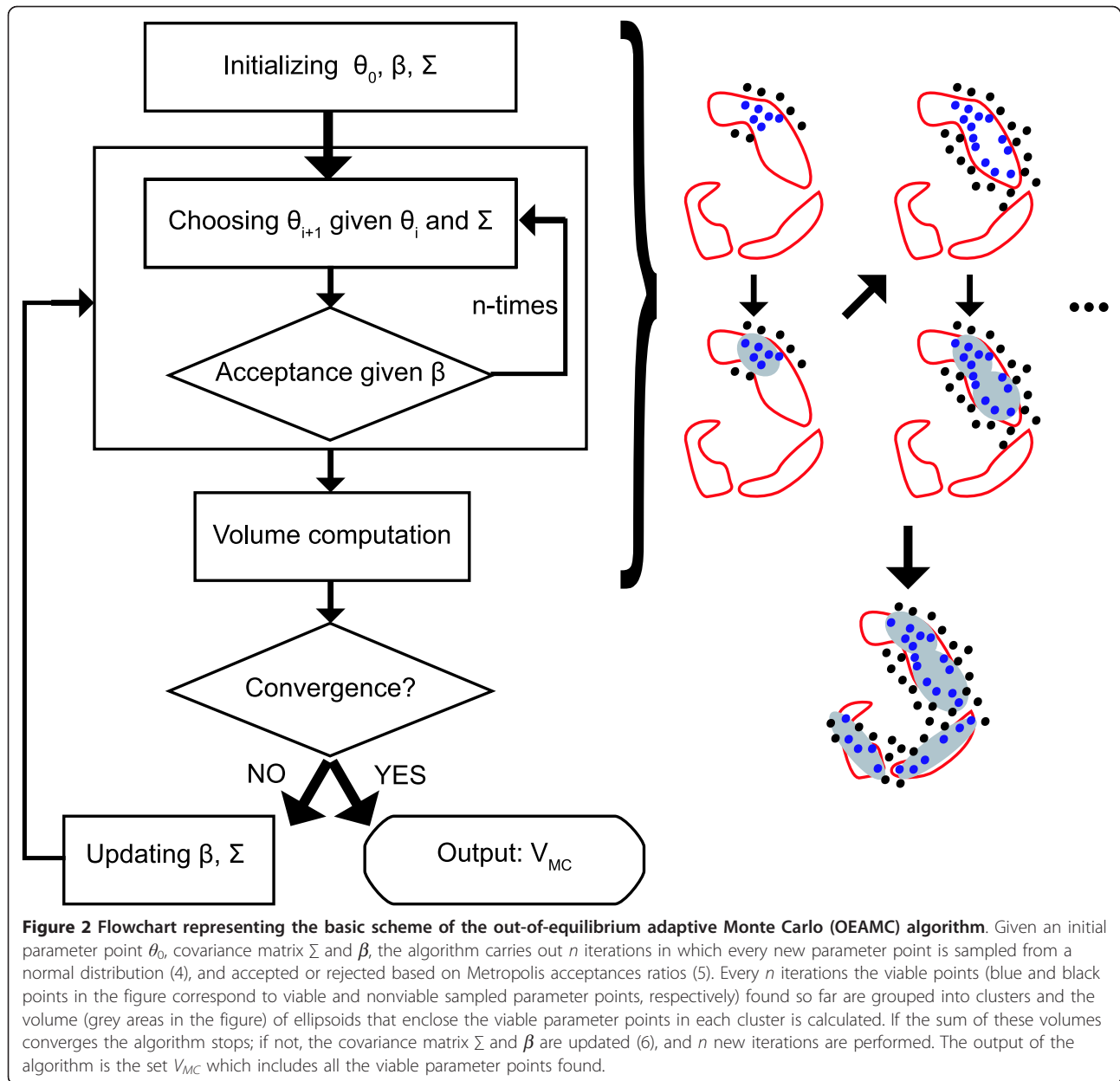
that is, if the cost function does not exceed some positive threshold E_0 . For example, θ may imply a system behavior that allows an organism to survive or reproduce. The subset of parameter points $\theta \in \Theta^d$ for which (3) holds comprises the viable space [2,20].

Out-of-equilibrium adaptive Monte Carlo sampling

We next describe our coarse-grained, global exploration of the viable space via an out-of-equilibrium adaptive Metropolis Monte Carlo sampling (OEAMC) (Figure 2).

The Metropolis algorithm was initially introduced to analyse thermodynamic systems [33]. However, it can also be applied to systems like those we study here. To do so, one must identify the parameter space Θ^d and the cost function $E(\theta)$ with a state space and with the energy of a thermodynamic system, respectively [34]. Moreover a parameter β has to be introduced in order to mimic the inverse of the temperature. This parallelism has been widely used in simulated annealing [35] and Metropolis Monte Carlo sampling [36-41].





This analogy allows us to use an adaptive selection probability with covariance matrix Σ

$$g(\theta_i \rightarrow \theta) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp \left[-\frac{1}{2} (\theta - \theta_i) \Sigma^{-1} (\theta - \theta_i)' \right], \quad (4)$$

in order to propose the transitions between parameter points, and Metropolis adaptive acceptance ratios

$$A(\theta_i \rightarrow \theta) = \begin{cases} \exp[-\beta (E(\theta) - E(\theta_i))], & \text{if } E(\theta) - E(\theta_i) > 0, \\ 1, & \text{otherwise,} \end{cases} \quad (5)$$

to accept or not those transitions.

Given β and Σ , the exploration starts from a known viable parameter point θ_0 . Then, from the current θ_0 a new θ is constructed by sampling the distribution (4) centred on θ_0 . If $E(\theta) < E(\theta_0)$, the new θ is automatically accepted and becomes θ_1 . In contrast, if $E(\theta) > E(\theta_0)$, θ is accepted with a probability $\exp[-\beta (E(\theta) - E(\theta_0))]$, in which case it becomes θ_1 . If θ is rejected, then $\theta_1 = \theta_0$. This scheme is repeated for a predefined number of iterations n .

After n iterations the algorithm determines whether OEAMC sampling must stop. To do so, the viable parameter points found so far are divided into a predefined

number of clusters. Then, OEAMC calculates the ellipsoids with minimum volume that enclose the points in each cluster and computes the sum of all ellipsoids volumes. The algorithm stops when the volume of all ellipsoids converges or when a maximum number of iterations is reached. If either of these criteria are met, OEAMC sampling terminates. Otherwise, n more iterations are carried out after updating β and Σ according to

$$\beta = \begin{cases} b\beta, & \text{if } f_v = 0, \\ \beta, & \text{if } 0 < f_v \leq f_0, \\ \beta/b, & \text{if } f_v > f_0, \end{cases} \quad (6)$$

$$\Sigma = \begin{cases} s\Sigma, & \text{if } f_a > f_u \\ \Sigma, & \text{if } f_l < f_a \leq f_u, \\ \Sigma/s, & \text{if } f_a < f_l, \end{cases}$$

where f_v and f_a are the proportions of sampled viable parameter points and accepted transitions calculated over the last n iterations, respectively. The parameters b , s are larger than one and must be specified by the user. Equation (6) implies the following procedure. When Monte Carlo sampling is mainly confined to a viable region ($f_v > f_0$), β decreases and the frequency of accepted transitions increases. If this makes the frequency of accepted transitions larger than an upper limit ($f_a > f_u$), the covariance matrix Σ will become larger and the method will sample broader regions. In contrast, when the method has not found any viable parameter point ($f_v = 0$), β increases and the frequency of accepted transitions decreases in order to force the algorithm to sample regions with lower cost function. If this frequency falls below a lower limit ($f_a < f_l$), Σ decreases to maintain the desired frequency of accepted transitions. The end product of OEAMC is the set V_{MC} of all the viable parameter points that it found.

Several differences of OEAMC to existing approaches are worth noting. First, OEAMC does not increase β continuously from values near zero to values much larger than the maximum of the cost function, as in simulated annealing (see [42,43] and references therein). Furthermore, OEAMC does not utilize β as an “extra” stochastic parameter, an idea used in tempering approaches (see [44,45]). In addition, it does not diminish the adaptation of Σ over time, as equilibrium adaptive Monte Carlo sampling does (see [45,46] and references therein). In contrast, OEAMC automatically adapts both β and Σ during the whole sampling in order to obtain high and low frequencies of accepted transitions and viable parameter points, respectively. The objective of OEAMC is not to find a point close to the global optimum of the cost function, as in the case of simulated annealing, or to obtain a Markov chain with a specified equilibrium distribution, as in the case of

equilibrium adaptive Monte Carlo sampling or simulated tempering. Instead, it aims to acquire a (potentially biased) sample of parameter points distributed all over the viable space.

Multiple ellipsoid-based sampling

The OEAMC samples the viable space at low resolution. Thus, it is necessary to introduce a method that uses the viable points already found by OEAMC to explore the viable space in detail. A novel method we call multiple ellipsoid based sampling (MEBS) (Figure 3) carries out this fine-grained exploration of the viable space.

The use of an ellipsoid to bound viable regions in search spaces has been known for decades (see [47] and references therein). However, nonconvex viable regions are not accurately bounded by a single ellipsoid [48]. The problem is specially difficult in high dimensional spaces, where the “curse of dimensionality” forces the volume of the bounding ellipsoid to be much larger than the volume of the nonconvex bounded object of interest. The probability of “hitting” this object by sampling uniformly inside a bounding ellipsoid becomes negligible as the number of dimensions increases. To overcome this problem, MEBS iteratively constructs ellipsoids that start firstly from viable points already found by OEAMC, and then also by points found by MEBS. These ellipsoids change their centres and orientations in order to enclose multiple nearly convex viable regions and to cover the whole viable space as tightly as possible.

The j -th ellipsoid expansion starts by selecting a viable parameter point $\theta_{v,j}$ in an adaptive way (see the Additional File 1 for details). In the first ellipsoid expansions the starting point will typically be a viable point obtained from OEAMC. This point defines $2d$ (d denotes the dimension of the parameter space) viable parameter points that are placed near the intersection between the boundary of the viable region and the straight lines parallel to the axes of the Cartesian coordinate system that pass through $\theta_{v,j}$ (see Additional File 1 for a more detailed description). Then MEBS constructs an ellipsoid L_j^i . If $i = 0$, L_j^0 is the minimum volume ellipsoid that encloses the $2d$ viable points near the boundary of the viable space. If $i \neq 0$, L_j^i is the minimum volume ellipsoid that encloses the set of viable points V_j^i which comprises the viable points found after the iteration i of the j -th ellipsoid expansion. From this ellipsoid L_j^i , the MEBS creates a new ellipsoid S_j^i that has the same orientation as L_j^i , but the lengths of its axes are multiplied by a scaling parameter g_i . Then the algorithm uniformly samples a predefined number of parameter points n from this ellipsoid S_j^i . The union of the set of viable points in S_j^i with V_j^i then gives V_j^{i+1} .

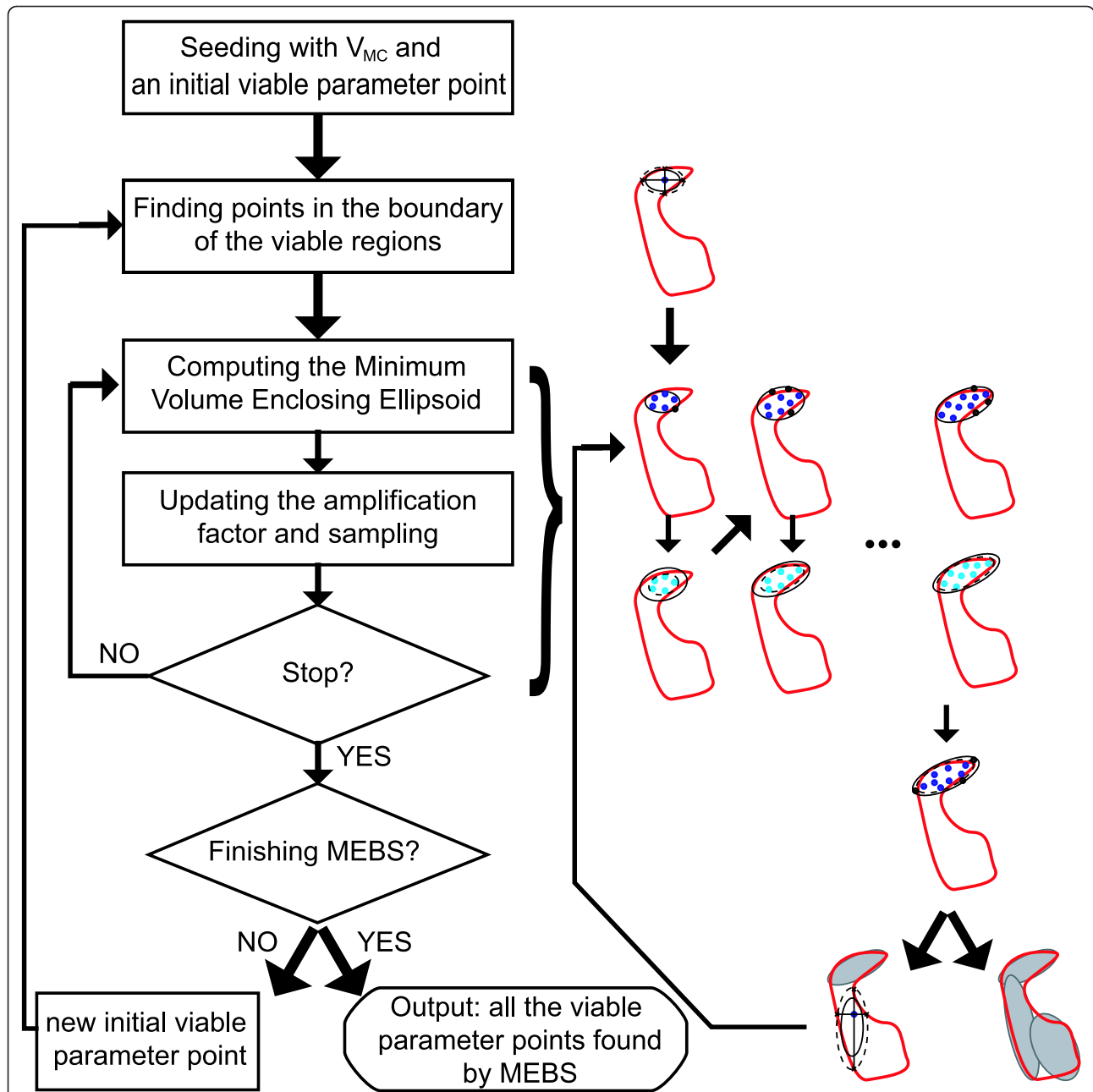


Figure 3 Flowchart for the multiple ellipsoid-based sampling (MEBS) procedure. Given V_{MC} , the set of viable parameter points found by OEAMC, and an initial viable parameter point, the method finds viable parameter points near the boundary of the viable region. Then, it calculates the minimum volume enclosing ellipsoid (MVEE) that encloses those viable parameter points and samples inside an ellipsoid with the same orientation but smaller axes. In the figure, the ellipsoids inside of which sampling is carried out are represented by solid curves; dark blue and black points correspond to viable and nonviable points found in the last sampling, respectively; the MVEE ellipsoids are represented by dashed curves. After the sampling step just described, the method again calculates the MVEE of the viable points found so far (light blue points in the figure), and samples inside a scaled ellipsoid with the same orientation but larger axes (7). If the scaling factor tends to one, or a fixed number of iterations is reached, the initial exploration finishes. If this does not happen the method calculates the MVEE of the viable parameter points found and performs a new uniform sampling inside a new scaled ellipsoid. At the end of every new ellipsoid expansion, the algorithm checks if MEBS must stop, which occurs if the algorithm does not find any new viable points in viable nonexplored regions (grey ellipsoids). If MEBS does not stop, it carries out another ellipsoid expansion starting from a different viable parameter point. The result of the MEBS is the set of the viable parameter points found during all the ellipsoid expansions.

The selection of the scaling parameter g_i is critical for the performance of the algorithm. We define it as:

$$g_i = \begin{cases} g_0 < 1, & \text{if } i = 0 \\ g_1 > 1, & \text{if } i = 1, \\ g_{i-1} + \frac{(g_{i-1} - 1)}{p}, & \text{if } |V_j^i| - |V_j^{i-1}| > nb_w, i > 1, \\ g_{i-1} - \frac{(g_{i-1} - 1)}{p}, & \text{if } |V_j^i| - |V_j^{i-1}| < nb_l, i > 1, \\ g_{i-1}, & \text{otherwise.} \end{cases} \quad (7)$$

where V_j^i indicates the number of elements in the set and b_l , b_w and $p < 1$ are parameters for lower and upper bounds, and for axis scaling, respectively.

The rationale behind equation (7) is as follows: Points in L_j^0 lie near the boundary of the viable space. In high dimensional spaces the ‘‘curse of dimensionality’’ may cause a large proportion of this ellipsoid volume to be filled by nonviable points. Setting $g_0 < 1$ forces S_j^0 to be smaller than L_j^0 . This makes it more likely that S_j^0 contains a larger proportion of viable parameter points, which will lead to a larger set V_j^0 . To explore a larger elliptic region around $\theta_{v,j}$, the method then performs a second iteration with $g_1 > 1$. All subsequent iterations depend on the number of viable points found in the last iteration ($|V_j^i| - |V_j^{i-1}|$). Specifically, when this number is larger than some upper limit nb_w , the scaling parameter grows by a factor $1/p > 1$ to explore larger domains of parameter space. When the difference ($|V_j^i| - |V_j^{i-1}|$) is below some lower limit nb_l - only few additional viable points have been found in the last iteration - shrinking the axes allows an efficient exploration of smaller regions. Thus, viable parameter points found in previous iterations guide and define the ellipsoid where the next sampling is carried out.

The j -th ellipsoid expansion started from $\theta_{v,j}$ finishes when g_i converges to one or after a fixed number of iterations is reached. The output is $V_{e,j}$, a set of sampled viable points that contains the $2d$ viable parameter points found near the boundary of the viable space, and the set of viable parameter vectors V_j^i updated in the last iteration.

Then, the MEBS initiates a $j+1$ -th ellipsoid expansion. The new initial point $\theta_{v,j+1}$, is chosen from the set composed by V_{MC} and the union of $V_{e,k}$, $k = 1 \dots j$, that is, the set of viable points obtained after OEAMC exploration and previous ellipsoid expansions, respectively. To explore regions that have not yet been sampled, we preferentially select a $\theta_{v,j+1}$ that is far away from the average of all previous starting points $\theta_{v,k}$, $k = 1 \dots j$ (see Additional File 1 for details).

At the end of each ellipsoid expansion, the algorithm determines if MEBS should stop. To do so, the viable parameter points found so far $\{V_{MC}, V_{e,1}, V_{e,2}, \dots, V_{e,j}, V_{e,j+1}\}$ are divided into a predefined number of clusters. Then, MEBS calculates the ellipsoids with minimum volume that enclose the points grouped in each cluster and computes the sum of all ellipsoids volumes. The algorithm stops when the sum of the volume of all ellipsoids converges, or when a maximum number of ellipsoid expansions is reached. The final result of MEBS is the set of viable parameter points $\{V_{MC}, V_{e,1}, V_{e,2}, \dots, V_{e,j}, V_{e,j+1}\}$.

Volume computation and acquisition of a large set of uniformly distributed viable parameter points

The end result of OEAMC and MEBS is a set of viable parameter points that can be used for a variety of purposes. Specifically, this set allows us to obtain simultaneously a large set of uniformly distributed viable points and an estimate of the viable volume Vol_v . (Note that the set of viable points obtained by OEAMC and MEBS is not an uniform sample from the viable space).

To calculate Vol_v we must evaluate the integral

$$\text{Vol}_v = \int_{\Theta^d} f(\theta) d\theta, \quad (8)$$

$$f(\theta) = \begin{cases} 1, & \text{if } E(\theta) < E_0, \\ 0, & \text{if } E(\theta) \geq E_0. \end{cases}$$

Given N parameter points uniformly sampled in Θ^d , the Monte Carlo integration theorem [49] implies that the volume (8) can be estimated by

$$\text{Vol}_v = \int_{\Theta^d} f(\theta) d\theta \simeq \text{Vol}_{\Theta^d} \langle f \rangle, \quad (9)$$

$$\langle f \rangle = \frac{1}{N} \sum_{i=1}^N f(\theta_i),$$

where Vol_{Θ^d} is the volume of the entire parameter space. If the error is Gaussian distributed, the standard deviation of the volume estimator is given by

$$\Delta \text{Vol}_v = \text{Vol}_{\Theta^d} \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N}}, \quad (10)$$

$$\langle f^2 \rangle = \frac{1}{N} \sum_{i=1}^N f^2(\theta_i).$$

Thus, if a high proportion of the N sampled parameter points is viable, the Monte Carlo integration in Θ^d will estimate the viable volume accurately.

This approach is usually sufficient to carry out viable volume estimations in low-dimensional spaces [8,21-25]. However, the ‘‘curse of dimensionality’’ poses a specific

problem when this technique is applied to high-dimensional parameter spaces. To calculate the viable volume (9) and to obtain a large set of uniformly distributed viable parameters efficiently, one cannot simply sample over the entire parameter space, because doing so would be too inefficient. It would be much better to perform a uniform sampling over a subspace $W \in \Theta^d$ that encloses the viable space as “tightly” as possible. This subspace will typically be much smaller than the entire space ($\text{Vol}_W \ll \text{Vol}_{\Theta^d}$).

To construct such a subspace (Figure 4), we build on the ideas already present in the algorithm developed by Hafner *et al.* [20]. The first step consists of using the set of viable parameter points V_t that comprises the viable points already found by OEAMV and MEBS (the letter t stands for *total*). To make Vol_v and Vol_W as similar as possible, Hafner’s method encloses the set of viable parameter points V_t into a single box with a smaller volume than the entire space. However, in many dimensions the volume of a nonconvex viable space may be much smaller than the volume of its enclosing box. To overcome this limitation we define the subspace W via a family of ellipsoids that cover the viable space locally (do not confuse with the ellipsoid based exploration of the viable space described above). To determine these ellipsoids we group the set of viable parameter points V_t into k clusters, and compute the ellipsoid with minimum volume that encloses the viable points grouped in every cluster (see Additional File 1 for details).

In this procedure, the subspace W is composed of the points of the parameter space enclosed by the k ellipsoids

$$W = \{\theta \in \Theta^d \mid \theta \in \bigcup W_i\} \quad i = 1, 2, \dots, k, \quad (11)$$

where W_i is the region of the parameter space enclosed by the i -th ellipsoid. In general, the k ellipsoids may intersect, so the viable volume in W may be smaller than the sum of the viable volumes in W_i . To avoid the resulting inaccuracy in volume estimation, we introduce a new integrand

$$f_i(\theta) = \begin{cases} 0, & \text{if } \theta \in \bigcup W_j, \quad j = 1, 2, \dots, i-1, \\ 0, & \text{if } \theta \notin \Theta^d, \\ f(\theta), & \text{otherwise.} \end{cases} \quad (12)$$

This integrand evaluates the parameter points in the ellipsoid intersections only once. Therefore, by sampling N parameter points uniformly from W (11) and by using (9), we can estimate the viable volume (8) as

$$\text{Vol}_v \simeq \int_W f(\theta) d\theta = \sum_{i=1}^k \int_{W_i} f_i(\theta) d\theta \simeq \sum_{i=1}^k \text{Vol}_{W_i} \langle f_i \rangle, \quad (13)$$

$$\sum_{i=1}^k m_i = N,$$

where m_i is the number of parameter vectors sampled inside W_i .

This approach of covering the viable region with ellipsoids can reduce the sampling volume dramatically, and thus increase the proportion of viable parameter points sampled in W far beyond that in the entire space Θ^d . This means that the viable volume can be calculated more accurately, and larger sets of viable parameter points can be sampled uniformly.

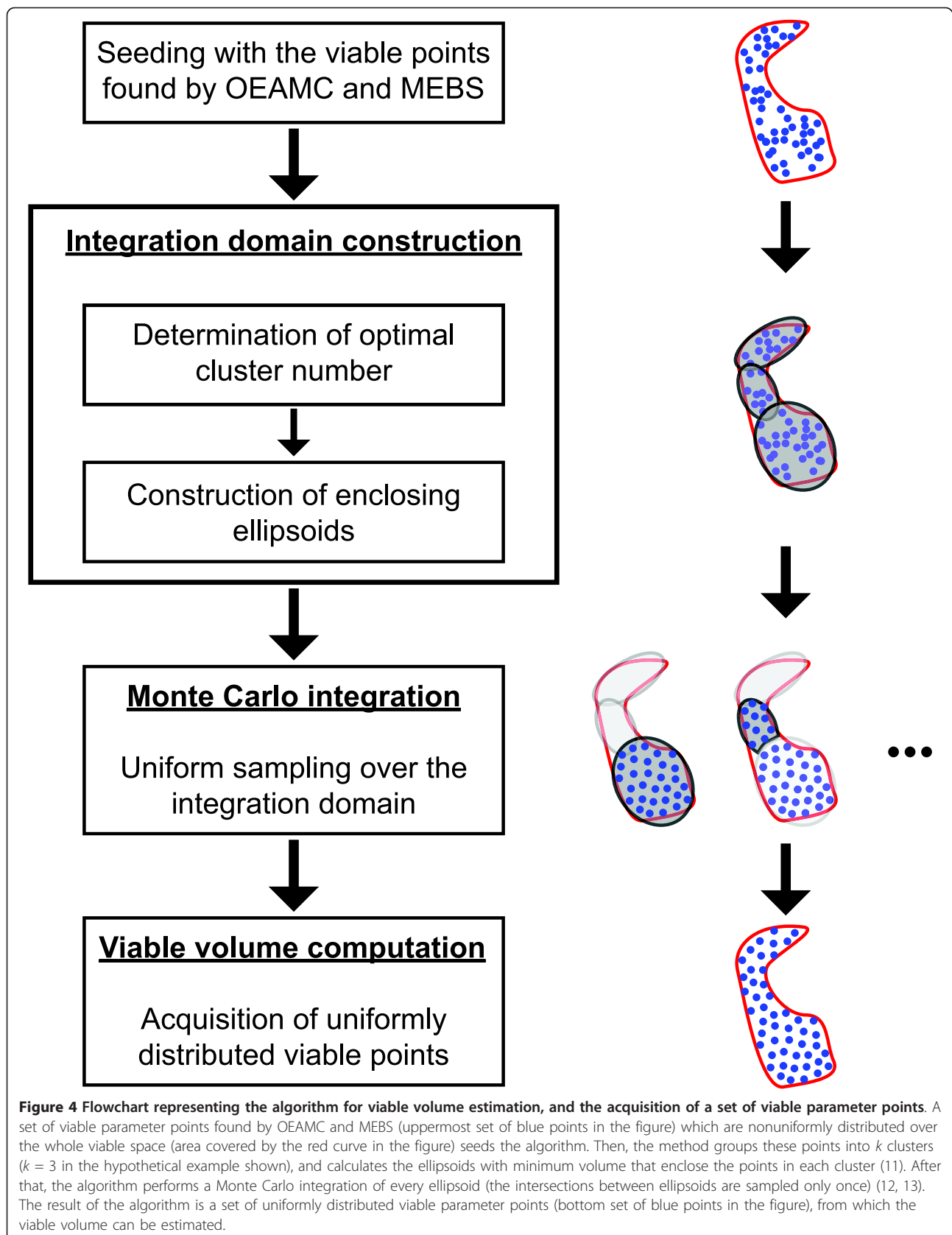
We caution that in practice, one can never be certain that the whole viable space is contained in the integration domain W that our approach (or any other approach) determines. The agreement between the actual viable volume from expression (8) and the estimated viable volume (13) depends on the proportion of the viable volume that is enclosed in W . The subspace W is defined by the set of viable parameter points V_t found by OEAMC and MEBS; therefore, the success of the volume estimation hinges on whether the previous exploration of parameter space found many viable points throughout the viable space. An implementation of our algorithm in MATLAB is available as the package HYPERSPACE from <http://www.ieu.uzh.ch/wagner/software> and <http://www.csb.ethz.ch/tools/index>.

Results and Discussion

A two-step algorithm for sampling of parameter spaces

The algorithm we propose starts from the definition of a viability condition and of a cost function (Figure 1). Depending on the biological model considered, the viability condition may include stability of a specific steady state, bistability [50], oscillations whose period lies in a given interval [20,24], the production of specific gene expression patterns [22], and many others. The cost function measures how closely the model’s behavior matches the viability condition.

The first step of the algorithm consists of a global coarse-grained exploration of the viable space by an out-of-equilibrium adaptive Monte Carlo (OEAMC) sampling of the entire parameter space (Figure 2). Following a thermodynamic analogy used by simulated annealing [35] and Metropolis Monte Carlo sampling [36-41], we identify the parameter space and the cost function with the state space and the energy, respectively, of a thermodynamic system that is in contact with a thermal bath with variable temperature. The objective of OEAMC is to identify viable regions in the parameter space by adjusting the “temperature” and the length of the jumps through the parameter space. Briefly, OEAMC adapts the “temperature” and jump lengths to force a finite but small frequency of sampled viable parameter points, and a high proportion of accepted transitions to new parameter points. This helps OEAMC not to “get lost” in the parameter space, but at the same time lets it “travel”



through nonviable regions where the cost function may have moderately high values. Thus, this procedure allows OEAMC to visit and sample from regions of the viable space that may be poorly connected to each other.

The low frequency of sampled viable parameter points forces OEAMC to explore the viable space at low resolution. To characterize the viable space in greater detail, it is necessary to define its borders more precisely, and to gain insight into its local geometry. In a second step, we therefore carry out a fine-grained exploration of the viable regions already identified through OEAMC, using a technique we call multiple ellipsoid-based sampling (MEBS) (Figure 3). This technique performs a local exploration of the parameter space by sampling from ellipsoids (an approach that is widely used in search algorithms, see [47] and references therein) that change their centres and expand or shrink their axes to enclose different regions of the viable space in which viable points are found. To cover locally nonconvex and/or poorly connected viable spaces, different ellipsoid expansions start from parameter points far away from each other (see Methods and Additional File 1).

The end result of OEAMC and MEBS is a set of viable parameter points that can be used for a variety of purposes. One of them is to define the integration domain in which a Monte Carlo integration estimates the volume of the viable space. (Note that the set of viable points obtained by OEAMC and MEBS is not an uniform sample from this space, and cannot be used directly for this purpose). We define this domain as the union of multiple ellipsoids - different from those used in MEBS sampling - that are constructed by grouping the viable parameter points into clusters, and by determining the ellipsoid with minimum volume that encloses the viable points in each of the clusters (Figure 4). This integration domain thus designed can cover nonconvex and high dimensional viable spaces "tightly". That is, the proportion of viable parameter points in this new integration domain is much higher than in the whole parameter space. By sampling viable points uniformly within this domain, we can compute the volume of a viable space. We reasoned that our procedure would allow us to reduce the computational effort in estimating a viable volume substantially. We will show in the next section that this is indeed the case. More generally, the large set of uniformly distributed viable parameter points that our method can generate permits us to characterize not only the size, but also the topology of a viable space. It also allows us to connect the robustness of a biological system to the geometrical properties of its viable space. Furthermore, this large set of viable parameters opens the possibility for a "glocal" analysis [20], in which the global characterization is supplemented by a local analysis around every viable parameter point. Thus, our algorithm can be used together

with a local robustness measurement (e.g., that proposed by Dayarian *et al.* [7]) to get insight into the distribution of a model's robustness in a viable space.

Efficient sampling of high-dimensional spaces

In a first test problem, we estimated the volume of a nonconvex region defined by either one single or two tangent multidimensional spherical shells (Figure 5). We chose this study system to analyze the efficiency of our method as a function of the geometry and dimension of a viable space, because here the viable volume can be calculated analytically.

We define the parameter space as $\Theta^d = \Theta_1 \times \Theta_2 \times \dots \times \Theta_d$, where $\Theta_i = [-10, 10]$, $i = 1, 2, \dots, d$. The cost function and the viability condition are given by

$$E_n(\theta) = \min_j \left| \|\theta - c_j\| - \frac{r_e + r_i}{2} \right|, \quad E_n \leq \frac{r_e - r_i}{2}, \quad (14)$$

$$j = 1, 2, \dots, n, \quad \|c_j - c_{j-1}\| = 2r_e,$$

where c_j is a point in Θ^d and r_e and r_i are two scalars that fulfill $r_e > r_i$ (in all our numerical tests $r_e = 0.5$ and $r_i = 0.3$).

When $n = 1$ (single spherical shell test case), the lines of constant cost are multidimensional spheres centred on c_1 (Figure 5-b). The (degenerate) global minimum of the cost function occurs in the multidimensional sphere centred on c , and with radius $\frac{r_e+r_i}{2}$ (Figure 5-a, b). The viability condition is fulfilled by the parameter points that lie in the region enclosed by two multidimensional spheres with centre c and radii r_i and r_e , respectively.

For $n = 2$ (two tangent spherical shells test case), the cost function has its degenerate global minimum in two multidimensional spheres centered on c_1 and c_2 , respectively, with radius $\frac{r_e+r_i}{2}$ (Figure 5-c, d); the viable parameter points lie in the inner region of two tangent multidimensional spherical shells with internal radii r_i , external radii r_e and centers c_1 and c_2 , respectively.

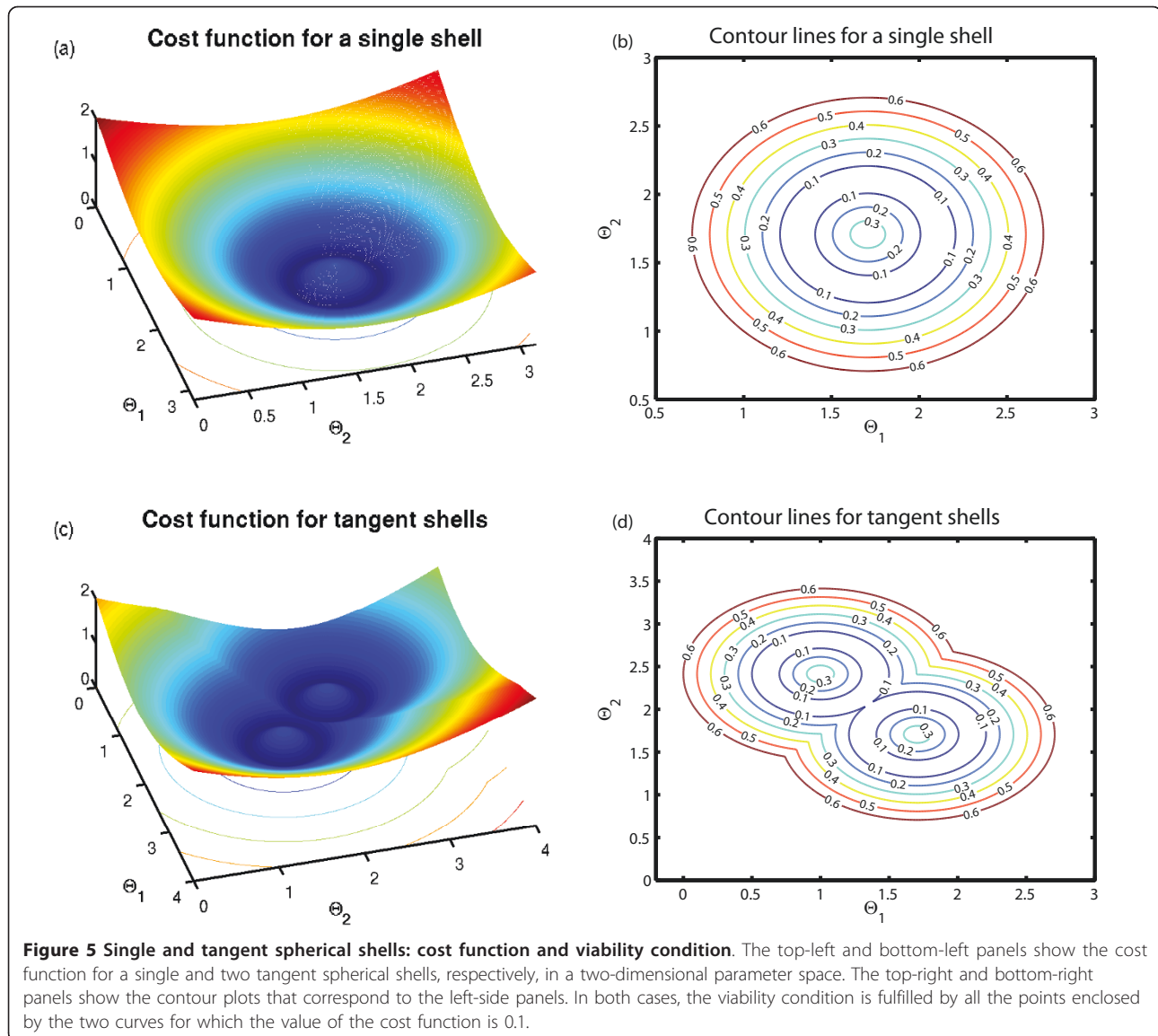
The volume filled by the viable region can be computed analytically as:

$$\text{Vol}_{v,t} = nC_d \left(r_e^d - r_i^d \right),$$

$$C_d = \begin{cases} 1, & \text{if } d = 0, \\ 2, & \text{if } d = 1, \\ \frac{2\pi}{d} C_{d-2}, & \text{otherwise,} \end{cases} \quad (15)$$

where C_d is the volume of a d - dimensional hypersphere with radius 1.

We now compare the performance of (i) MEBS and OEAMC alone, (ii) both of them together, (iii) uniform sampling, and (iv) the method proposed by Hafner *et al.* [20] based on Gaussian sampling (see the Additional File 1 for details). For the single spherical shell test case, MEBS and OEAMC alone, and the combination of both

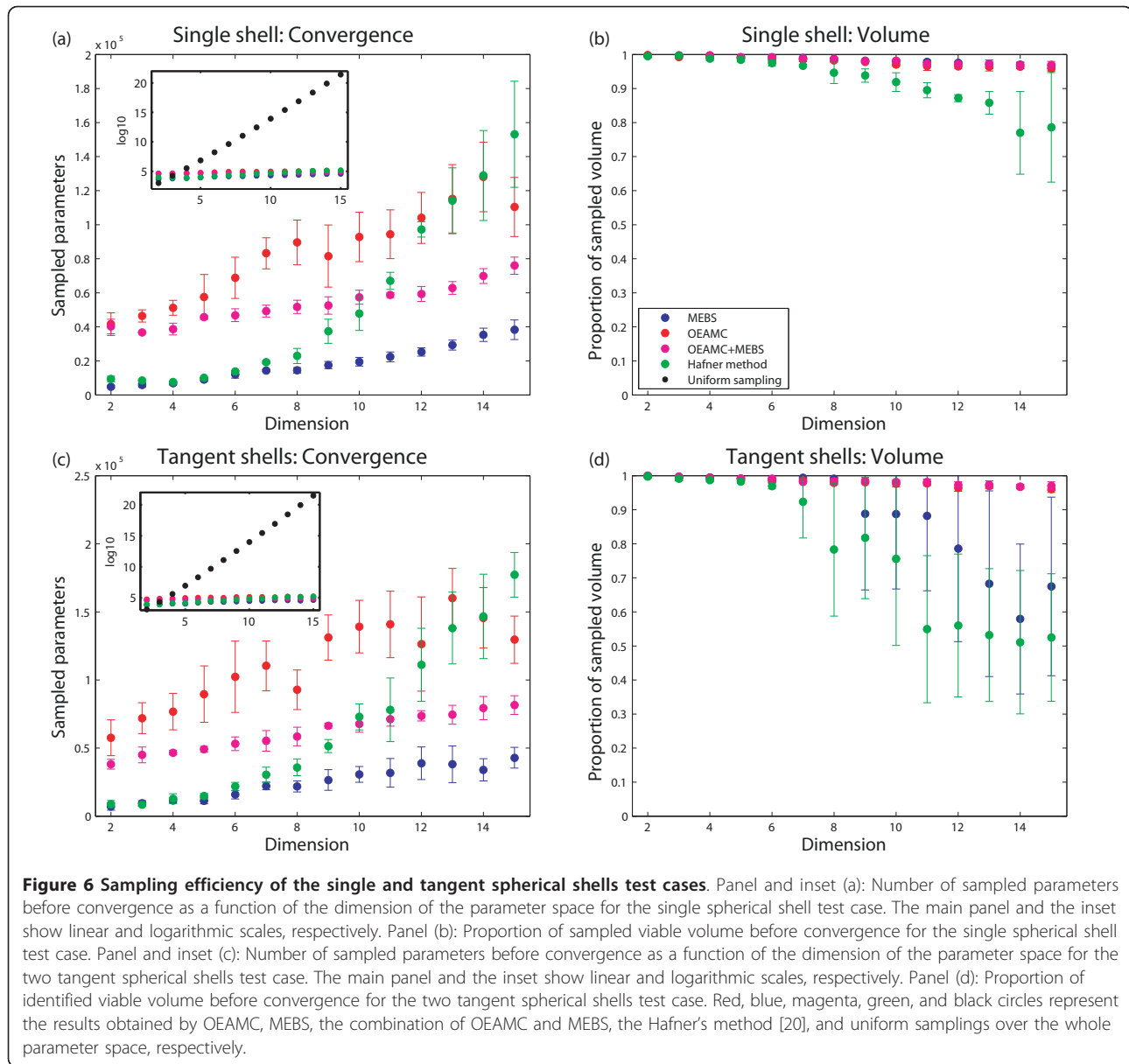


methods can identify the viable regions and obtain a good estimate of the viable volumes for dimensions up to $d = 15$ (Figure 6-b). Specifically, for all dimensions we studied they sample more than 95 per cent of the whole viable volume before converging. In addition, for this test case MEBS alone is much more efficient than OEAMC or a combination of both (Figure 6-a). Specifically, MEBS converges after sampling substantially fewer parameter points, because the frequency of viable points sampled by OEAMC is comparatively small, and OEAMC thus needs more sampling to estimate the viable volume to a given accuracy. For example, to achieve the same accuracy of volume estimation in $d = 15$ dimensions, MEBS uses 3-fold less samples than the OEAMC, and 2-fold less samples than the combination of both methods. In this first test case, the viable space,

albeit nonconvex, is well-connected. This permits a ready exploration of the space by ellipsoid expansions - efficient “travel” of ellipsoids inside the viable volume is possible.

MEBS, OEAMC, and their combination are much more efficient than uniform sampling of the parameter space. For instance, at $d = 15$ dimensions, “brute force” sampling uses 17 orders of magnitude more sampling points to estimate the viable volume (Figure 6-a inset).

The Gaussian sampling carried out by Hafner’s method *et al.* does not permit to identify in detail the borders of the viable volume for high dimensional spaces. Therefore, this technique can not estimate viable volumes in high dimensional spaces with precision (Figure 6-b). Moreover, in high dimensional spaces the tiny proportion of the whole parameter space filled by the viable volume



forces this technique to sample a large number of viable points before converging (Figure 6-a). For example in $d = 15$ dimensions, Hafner’s method uses 4-fold more samples than MEBS and underestimates the viable volume by 25 percent.

For the test case of two tangent spherical shells, MEBS and Hafner’s method often fail to “find” half of the viable volume in high dimensions (Figure 6-d). For example, in 14 dimensions, only 25 percent of the explorations carried out by MEBS and Hafner’s method find both shells. The two methods share the same limitation: the inability of sampling a point from the second shell, when starting from a random parameter point in the first shell. To find the second shell starting from the

first shell, MEBS and Hafner’s method must sample from an ellipsoid or from a Gaussian distribution, respectively, both of which must cover viable regions from both shells. However, both also include nonviable parameter points. In high dimensions the fraction of viable points becomes very small, and the probability of finding a viable point from the second shell is very low.

In contrast, OEAMC alone, and the combination of both OEAMC and MEBS sample the viable regions well (Figure 6-d). Specifically, for up to $d = 15$ dimensions, they estimate the viable volume with an error smaller than a 5 percent. Importantly, the combination of both OEAMC and MEBS is more efficient than OEAMC alone (Figure 6-c). For instance, to achieve the same accuracy

of volume estimation in $d = 15$ dimensions, the combination of MEBS and OEAMC used approximately 2-fold smaller samples than the OEAMC alone (and 17 order of magnitude smaller samples than uniform sampling).

The key for the success of the combination of OEAMC and MEBS is the complementary nature of their individual strengths. OEAMC does not need many sampled points to find two poorly connected regions. For example, in our two shell test case, it always hit both shells before sampling 25000 parameter in $d = 15$ dimensions. However, its low frequency of sampled viable points forces it to sample excessively many parameter points in order to explore a viable region in detail. In contrast, the bottleneck for the MEBS procedure is the discovery of a viable region - the second spherical shell in our example - that is poorly connected to a region that it already explored. Once such a region has been discovered by OEAMC, MEBS is able to sample from it efficiently, even if the region is nonconvex.

In sum, the combination of OEAMC and MEBS explores nonconvex and poorly connected viable regions in high dimensional parameter spaces more efficiently and accurately than either method alone and than other methods we evaluated. In addition, for both test cases the number of parameter points sampled by the combination of OEAMC and MEBS scales linearly with the number of dimensions (Figure 6-a and Figure 6-c). This suggests that for a given fixed complexity of the viable space, the computational effort needed by our method scales linearly with the dimensionality of the parameter space. This property makes our method suitable to explore high dimensional viable spaces.

Model of a biochemical oscillator with two feedback loops

The viable space of a realistic model of a biological system is in general unknown. Therefore, it is necessary to get an estimate of the viable volume through uniform sampling in order to check the performance of our method. However, complex models may have tiny and complex viable spaces that make it infeasible to get such an estimate. This hampers the use of biological models with realistic complexity to characterize our algorithm. To illustrate the application of our method and to check its performance with a biological model, we therefore used a very simplified biological model containing only 12 parameters that permits us to compare the results of our method with the uniform sampling of the parameter space.

This model describes a biochemical oscillator introduced by Hafner *et al.* [51]. It mimics the basic architecture of biological oscillators, such as cardiac pacemaker cells [52], intracellular calcium oscillations [53], cell cycle [27,54], and circadian clocks [55]. The model comprises two feedback loops (Figure 7) and it contains 12 individual

parameters and 5 state variables which correspond to the concentrations of different proteins. Briefly, in this model a protein R is expressed, phosphorylated and degraded. Protein R can also auto-phosphorylate. In the positive feedback loop, the phosphorylated form R_p acts as a kinase for protein Z whose active state Z_p increases the auto-phosphorylation rate of R . This kind of positive loop is a basic mechanism behind substrate-depletion oscillators. An example is the maturation promoting factor (MPF) oscillator involved in the cell division cycle of frog eggs [56]. The negative feedback loop is composed of three steps: R_p acts as kinase for an intermediate protein X . Its phosphorylated form X_p phosphorylates a second protein Y , whose phosphorylated state Y_p increases the degradation rate of R . Such negative feedback has been proposed as a basis for oscillations in many biological systems (see [27,28] for reviews).

The dynamics of the concentrations of the proteins R and R_p follow mass action kinetics [57]

$$\begin{aligned} \dot{[R]} &= \tilde{k}_1 - p([Z_p]) [R], \\ \dot{[R_p]} &= p([Z_p]) [R] - n([Y_p]) [R_p], \end{aligned} \quad (16)$$

where $p([Z_p])$ and $n([Y_p])$ respectively, reflect the effects of a positive and a negative feedback loops

$$\begin{aligned} p([Z_p]) &= \tilde{k}_2 + \tilde{k}_{11} [Z_p], \\ n([Y_p]) &= \tilde{k}_3 + \tilde{k}_{12} [Y_p]. \end{aligned} \quad (17)$$

In contrast, the concentrations of X_p , Y_p , and Z_p are governed by Michaelis-Menten kinetics [57]

$$\begin{aligned} \dot{[X_p]} &= \frac{\tilde{k}_4 [R_p] ([X_T] - [X_p])}{\tilde{k}_{10} + ([X_T] - [X_p])} - \frac{\tilde{k}_5 [X_p]}{\tilde{k}_{10} + [X_p]}, \\ \dot{[Y_p]} &= \frac{\tilde{k}_6 [X_p] ([Y_T] - [Y_p])}{\tilde{k}_{10} + ([Y_T] - [Y_p])} - \frac{\tilde{k}_7 [Y_p]}{\tilde{k}_{10} + [Y_p]}, \\ \dot{[Z_p]} &= \frac{\tilde{k}_8 [R_p] ([Z_T] - [Z_p])}{\tilde{k}_{10} + ([Z_T] - [Z_p])} - \frac{\tilde{k}_9 [Z_p]}{\tilde{k}_{10} + [Z_p]}, \end{aligned} \quad (18)$$

where $[X_T]$, $[Y_T]$, and $[Z_T]$ denote the total concentration of X , Y , and Z , respectively. For the sake of simplicity, we normalize all concentrations to one, i.e., $[X_T] = [Y_T] = [Z_T] = 1$.

The combination of active positive and negative feedback loops creates oscillators with a tunable frequency, and a robust amplitude [30]. These features make the negative plus positive loop oscillator suitable for systems like beating hearts and cell cycles. Here, we focused on oscillations in a narrow range of frequencies such as those produced by circadian clocks, and used the model to study the robustness of the oscillation period to parameter variations.

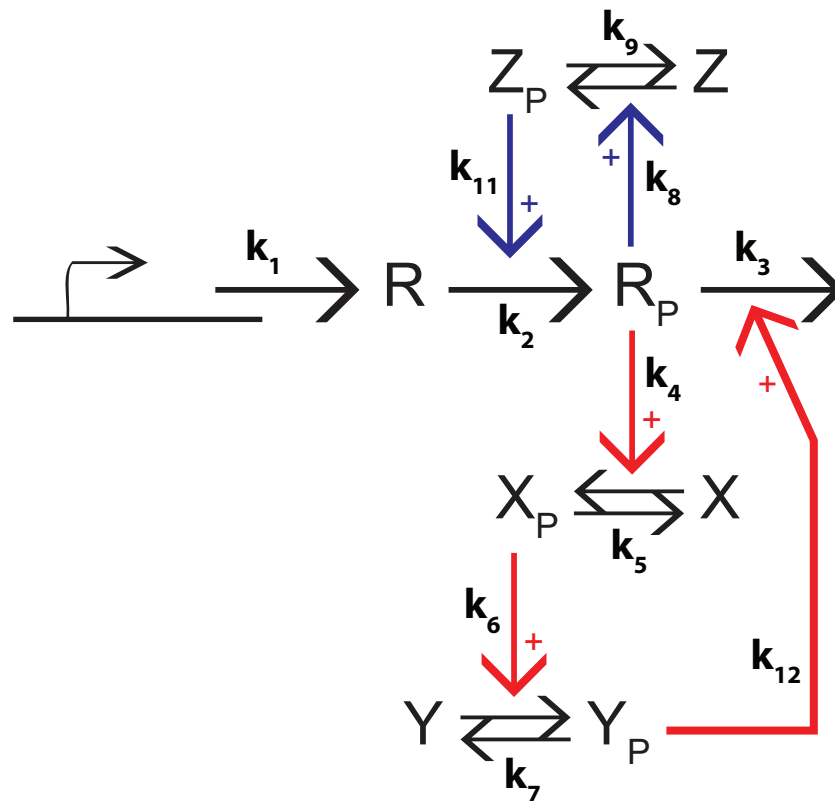


Figure 7 Reaction diagram of the model of a simplified biochemical oscillator with two feedback loops proposed by Hafner et al. [51]. The protein R is produced at a constant rate k_1 and its phosphorylated state R_p is produced at a rate, k_2 . The phosphorylated protein Z_p modulates this phosphorylation rate by means of a positive feedback loop (blue diagram in the figure). In addition, R_p is degraded with a rate, k_3 that depends on the phosphorylated protein Y_p by means of a negative feedback loop (red diagram in the figure).

To explore broad ranges of parameters values we work in a logarithmic domain in which the logarithm of individual parameters are constrained as follows

$$\begin{aligned} k_i &= \log(\tilde{k}_i), \\ k_i &\in [-4, 2], \quad i = 1, 2, \dots, 10, \\ k_i &\in [-7, 2], \quad i = 11, 12. \end{aligned} \quad (19)$$

Together, these ranges define the 12-dimensional parameter space $\Theta^{12} = k_1 \times k_2 \times \dots \times k_{12}$. We use the cost function

$$E_m(\theta) = \begin{cases} [(T_{R_p}(\theta) - 1)/0.1]^2, & \text{if } R_p \text{ oscillates,} \\ \infty, & \text{otherwise,} \end{cases} \quad (20)$$

where $T_{R_p}(\theta)$ is the period of the oscillations of R_p for a parameter point $\theta = (k_1, k_2, \dots, k_{12})$. The minimum of this cost function is attained by parameter vectors for which $T_{R_p}(\theta) = 1$.

Finally, we introduced the viability condition

$$E_m \leq 1, \quad (21)$$

meaning that a parameter point θ is viable if it causes R_p to oscillate with a period in the narrow interval $[0.9, 1.1]$.

To explore the viable space we carried out an OEAMC sampling followed by a MEBS. The viable parameter points obtained during this exploration are shown in Figure 8, which displays the 12-dimensional parameter space through six two-dimensional projections. The blue and red points, acquired by MEBS and OEAMC, respectively, occur in similar regions of the parameter space. This shows that the MEBS explored in detail the viable regions previously visited by OEAMC, just as for our spherical shells test case. The combination of OEAMC and MEBS revealed the nonconvexity of the viable space and its implications for the model function. Specifically, we note the viable region in Figure 8-f, which is composed of two approximately rectangular or bar-like regions that, together, form a nonconvex shape resembling an inverted L. Parts of these regions define topologies in which a single feedback loop produces the oscillations. More precisely, the left part of the horizontal bar corresponds to viable parameter points for which k_{12}

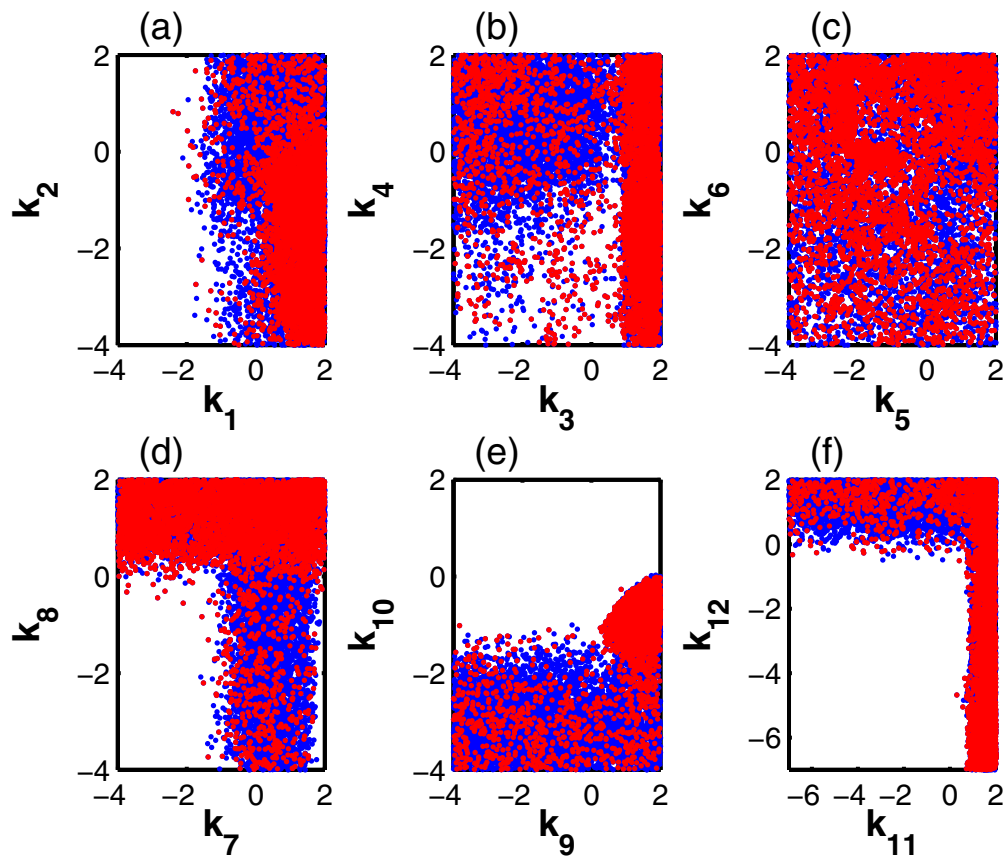


Figure 8 Exploration of the viable space for the oscillator with two feedback loops. Panels show projections of the 12-dimensional parameter space of the oscillator model onto six two-dimensional spaces corresponding to different parameter pairs. Red and blue points correspond to the viable parameter vectors found by OEAMC and MEBS, respectively.

is large and k_{11} small. In this region, only the negative feedback loop is active. Conversely, the bottom part of the vertical bar consists of viable parameter points for which k_{12} is small and k_{11} high. It corresponds to architectures where only the positive feedback loop is active (see Figure 7).

In a next step, we performed a Monte Carlo integration (see Methods and Additional File 1 for details) to estimate the viable volume. The integration domain is defined by using the viable points obtained by the OEAMC and MEBS explorations. This domain is approximately 630-times smaller than the whole parameter space. After uniformly sampling over the integration domain we obtained 3595 viable points, and estimated a viable volume of $\text{Vol}_v = 8.3 \cdot 10^4 \pm 2 \cdot 10^3$. To validate this estimate, we uniformly sampled over the whole parameter space with the same number of points we used in the OEAMC, MEBS, and integration parts of our algorithm. Only 9 of these points were viable, leading to a viable volume estimate of $\text{Vol}_v = 8.1 \cdot 10^4 \pm 2.7 \cdot 10^4$. The two estimates are very

similar, but the estimation obtained through uniform sampling has an uncertainty one order of magnitude larger than the one calculated through our method. In addition, we uniformly sampled $4 \cdot 10^7$ points from the whole parameter space to compare the distributions of every single viable parameter. The results showed that the distributions of each of the 12 parameters obtained through our method and the extensive brute force sampling are very similar (Figure S1).

In sum, our method yields an accurate characterization of the viable space for this complex twelve-dimensional system at much higher efficiency than brute-force approaches. Specifically, by using the same number of sampling points it carries out a 13 times more accurate estimation of the viable volume, and obtains 400 times more uniformly distributed viable points.

Robustness of positive and negative feedback loops

The sample of the viable space we obtained suggests a clear distinction between two oscillatory regimes, one

driven by a positive and the other driven by a negative feedback loops. We next discuss these regimes, as an illustration of the type of analyses that our method enables.

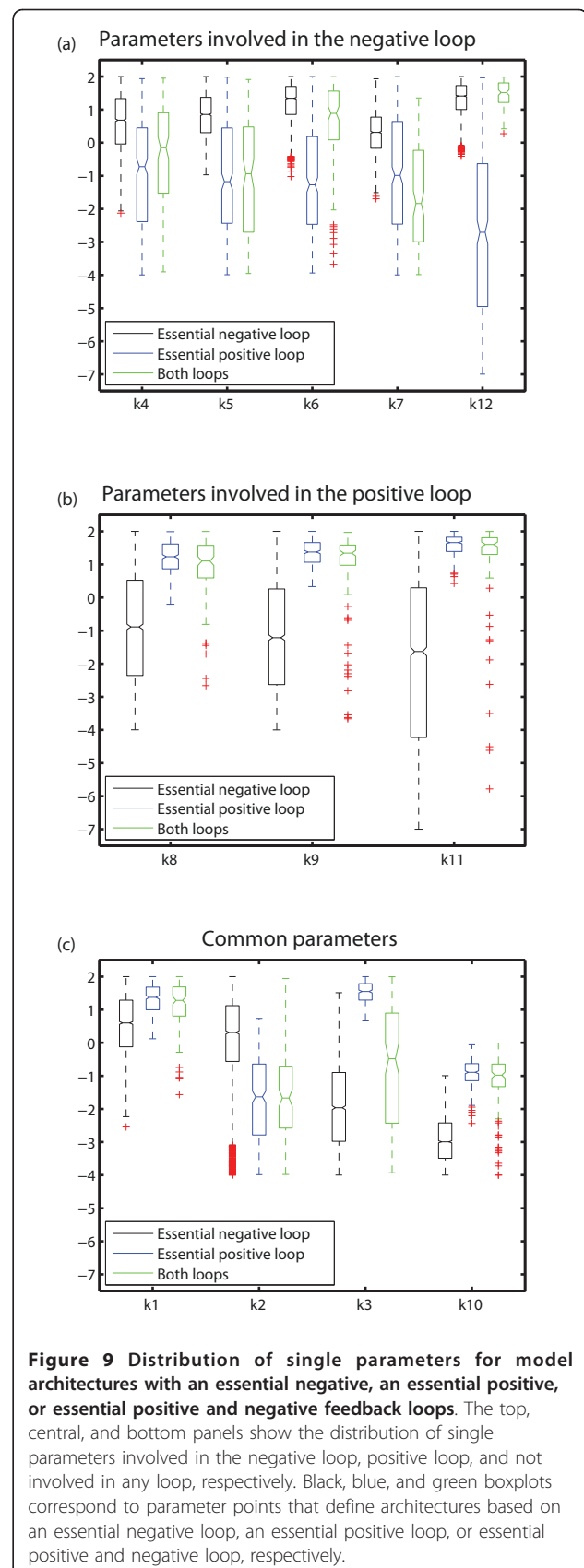
The many viable parameter points we found allowed us to characterize key properties of model architectures with individual or combined feedback loops via the geometry of the viable space. For this purpose, we classified each of the viable points into one of the following categories:

- Essential negative feedback loop: The model keeps fulfilling the viability condition (21) after removing the positive loop, or after substituting this loop with a higher activation rate of R_p (see Additional File 1).
- Essential positive feedback loop: The model keeps fulfilling the viability condition (21) after removing the negative loop or substituting this loop with a higher degradation rate of R_p (see Additional File 1).
- Essential positive and negative feedback loops: No loop can be removed or substituted by a higher activation or degradation rate without violating the viability condition (21).

We found that model architectures for which the negative feedback loop is essential occupy the vast majority (86%) of the viable space we sampled. In contrast, significantly fewer parameter combinations lead to viable oscillations based on an essential positive loop (10%), or on a combination of essential positive and negative feedback loops (4%).

If a single loop is essential, the parameters mainly responsible for this loop will be constrained. These are parameters k_8, k_9, k_{11} for the positive loop, and parameters $k_4, k_5, k_6, k_7, k_{12}$ for the negative loop (Figure 7). Figures 9-a and 9b illustrate these constraints. For example, in Figure 9-a, black coloring indicates to what extent parameters involved in the negative loop are constrained if this loop is essential, blue coloring indicates these constraints if only the positive loop is essential, and green coloring indicates these constraints if both loops are essential. Clearly, parameters involved in the negative loop can vary to a lesser extent if this loop is essential than when it is not essential. Analogous observations can be made for parameters involved in the positive loop (Figure 9-b).

A comparison of Figures 9-a and 9b also shows that parameters involved in the negative and positive feedback loops are constrained to different extents. Specifically, negative loop parameters can vary over broader intervals when the negative loop is essential than positive loop parameters can when this loop is essential. In addition, the parameters that do not form part of any loop (k_1, k_2, k_3, k_{10}) are more constrained in architectures with essential positive feedback loop than in



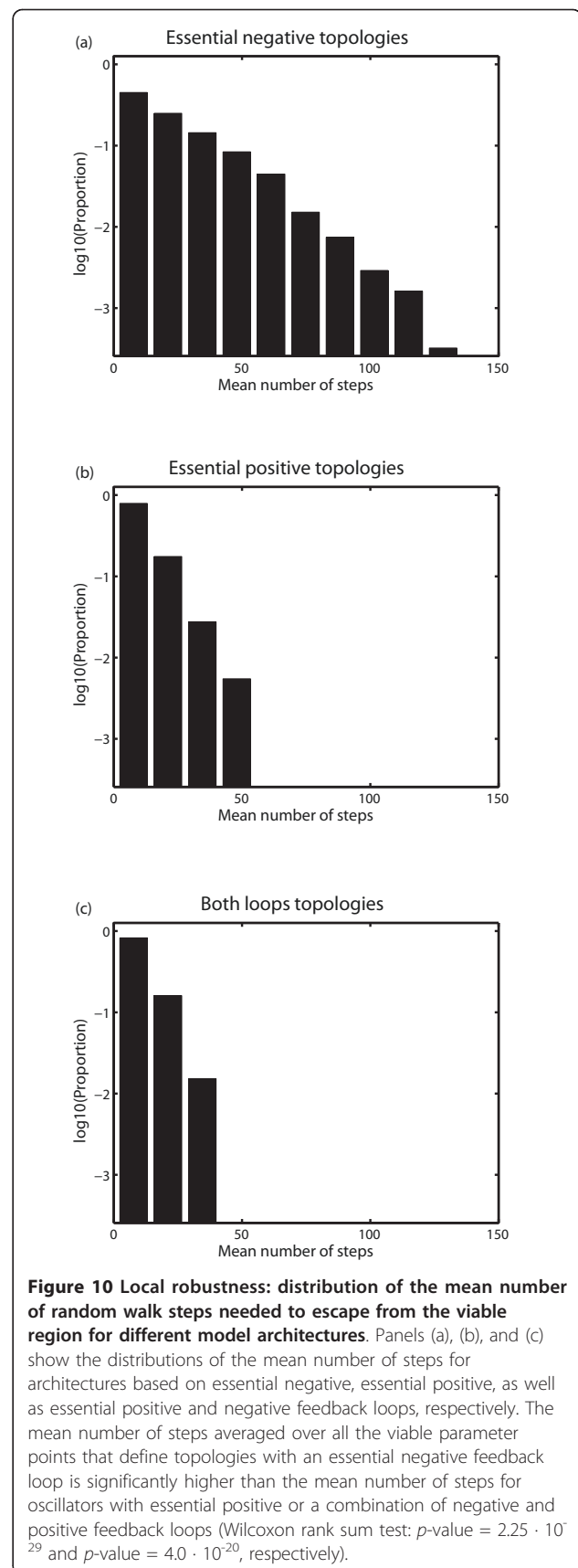
topologies with an essential negative feedback loop (Figure 9-c).

Taken together, these observations imply that model architectures based on a negative loop fill more of the viable space, and allow individual parameters to vary more broadly than architectures based on positive feedback loops. In other words, model topologies based on an essential negative feedback loop are more robust than topologies with essential positive loops, or topologies with both essential positive and negative loops.

To further explore this aspect of robustness, we used the method proposed by Dayarian *et al.* [7] which estimates the number of steps that a random walk needs to escape from the viable space. Briefly, we started ten random walks from every viable parameter point. Each new point in a random walk was selected from an independent Gaussian distribution centred on the previous parameter point and with a diagonal covariance matrix with standard deviations $\sigma = 0.01$. We followed every random walk until it arrived at a nonviable parameter point, and recorded the number of steps it had taken to reach this nonviable point. We used this number of steps as an indicator of local robustness around such parameter point. The mean number of steps before exiting the viable region was higher if the starting point corresponded to an architecture with a negative loop than to an architecture with an essential positive loop, or to a combination of essential positive and negative loops (Figure 10). Moreover, the distribution of the number of steps for the negative feedback architectures has a long tail (Figure 10-a). Specifically, two times more steps may be needed to leave the viable space than for the other two architectures (Figure 10-b, c). Hence, also in terms of local properties revealed by this approach, architectures with an essential negative feedback loop are significantly more robust than other topologies.

In addition, we found that adding a positive (not necessarily essential) loop to a model architecture based on a negative feedback loop further increases robustness and the allowable range of parameter variation. Figure 11-a already hints at this observation, because it shows that the largest density of viable parameter points occurs in regions of parameter space where both k_{11} and k_{12} are high. These parameters are important for the positive and negative feedback loops, respectively. In regions with the most viable parameter points both feedback loops are active and at least one of these loops is essential.

Further analysis corroborates this observation. In architectures with an essential negative feedback loop, the mean value of the parameter k_{11} , which controls the strength of the positive feedback loop, is significantly higher (p -value = $2.0 \cdot 10^{-27}$; Wilcoxon signed rank test) than the centre of the interval in which k_{11} is defined. In other words, the randomly sampled architectures with an essential negative feedback loop preferentially occur in



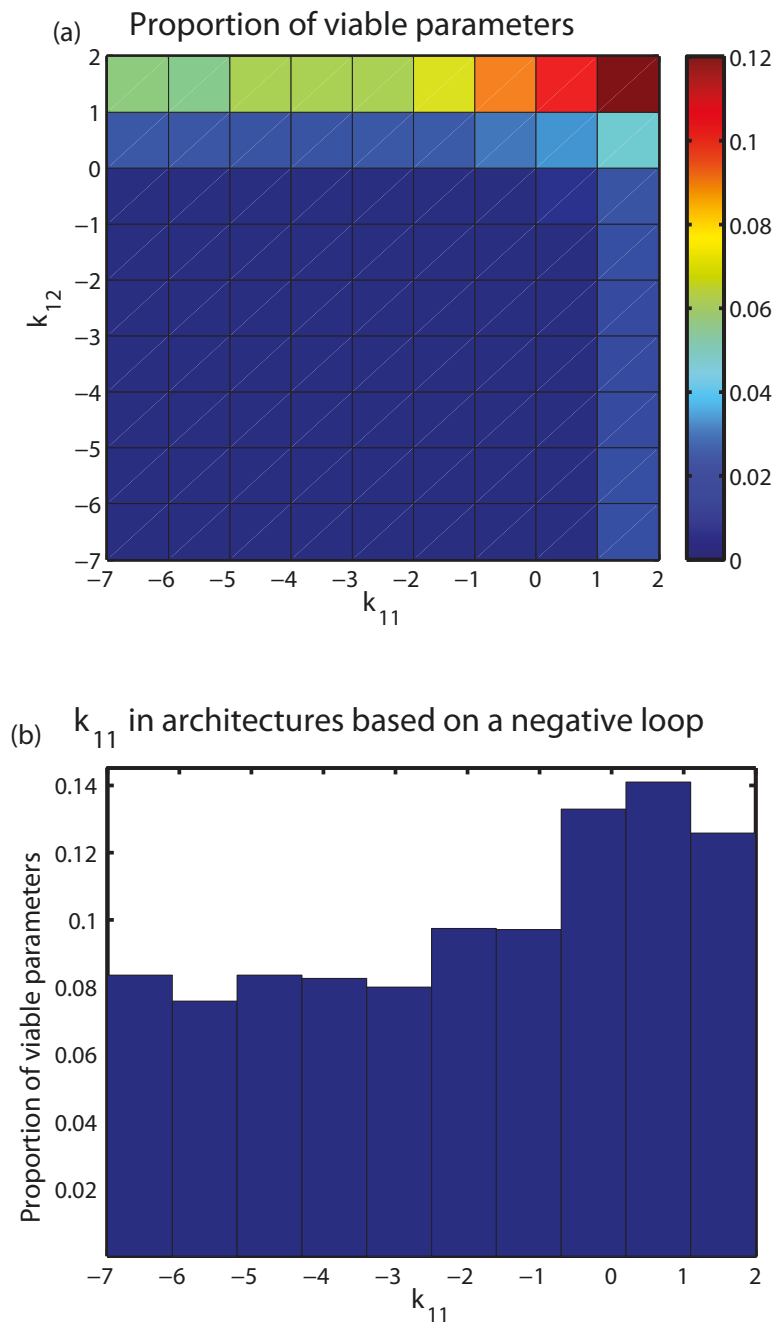


Figure 11 Distribution of viable parameter points in the k_{11} k_{12} plane. (a) proportion of viable parameter points found through Monte Carlo integration in every bin of the k_{11} k_{12} plane. The highest density of viable parameter points appears in configurations for which k_{11} and k_{12} are high; that is, model architectures in which both feedback loops are present (although one of them may not be essential). (b) proportion of viable parameter points which define architectures based on a negative feedback loop as a function of k_{11} ; that is, as a function of the single parameter that controls the strength of the positive feedback loop. The mean value of the parameter k_{11} is significantly higher (p -value = $2.0 \cdot 10^{-27}$ Wilcoxon signed rank test) than the centre of the interval in which k_{11} is defined. The density of viable parameter points increases with the value of the parameter k_{11} .

regions of parameter space where a positive loop is also active. Moreover, the density of viable parameter points increases with the value of the parameter k_{11} (Figure 11-b). Thus, a higher strength of the positive feedback loop

increases the number of parameter combinations that gives rise to viable oscillations.

Taken together, these observations suggest that an added nonessential positive feedback loop gives a

negative-loop-based model oscillator access to more viable parameter points. In the Additional File 1 we perform a similar analysis with a more complex model of a mammalian circadian oscillator. For this more realistic model we also observe that the circadian oscillations can be generated by a single negative feedback loop, whereas an additional positive feedback loop increases the robustness of the oscillations.

Connectivity of the viable space

The connectivity of the viable space indicates to what extent different model architectures with the same behavior can change into one another through small changes in individual parameters, as might occur on evolutionary time scales.

To study this connectivity, we chose a set of viable points in which each of the three basic model architectures we consider are represented. For every pair of parameter points, we defined a straight line connecting them, and identified a set of three points that subdivide the line into four equally long segments (we also subdivided the line into 5, 6, 7, and 8 equally long segments, obtaining qualitatively identical results). We then asked whether each of these points was located in the viable space. If so, it may be possible to connect the two parameter points by a straight line that lies entirely in the viable space. Based on this information, we defined a graph whose nodes are the set of viable parameter points. Two nodes are connected by an edge if the entire straight line between the nodes does not leave the viable space. Such an edge reflects the existence of potential evolutionary paths from one to the other node (parameter point) that does not leave the viable space. We find that this graph has one large connected component that comprises 95 percent of all nodes. This observation, together with our earlier analysis (Figure 8-f) shows that most of the viable space forms a nonconvex connected body with possible evolutionary trajectories that maintain the same behaviour and that connect qualitatively different system topologies through small changes in individual parameters.

The connected component contains nodes associated with all three basic architectures, but these three kinds of nodes are not equally likely to be connected to each other. Specifically, nodes (viable points) corresponding to model topologies with essential negative feedback loops are only connected to themselves, and to nodes with essential positive and negative feedback loops. Similarly, nodes that define topologies with essential positive feedback loops are only connected to themselves and to nodes with essential positive and negative feedback loops. Potential evolutionary trajectories that connect model architectures based on essential positive feedback loop and essential negative feedback loop, need to pass through configurations for which both loops are essential.

Overall, the global geometry of the viable space shows that model topologies based on an essential negative feedback loop are more robust than other architectures. Essential negative feedback allows the individual parameters to span larger intervals than essential positive feedback. Moreover, our local analysis reveals that topologies based on an essential negative feedback loop sustain the most change before losing viability. Successive small parameter changes can transform oscillators with an essential positive feedback loop into oscillators with an essential negative feedback loop, or vice versa. To do so, requires an intermediary stage in which both loops are essential.

Conclusions

In biological systems, the diversity of biochemical parameter values that can lead to similar behavior makes it useful to introduce the concept of a viable space in which a biological system maintains a given function. The algorithm we present here allows an efficient exploration and characterization of such a viable space in systems with many parameters. It involves a global coarse grained identification of viable regions, followed by detailed local explorations of these regions. The global part of our algorithm can find viable regions that may be poorly connected. In the local part, the viable regions discovered in the global part are explored in detail. The exploration of the viable space allows us to identify a (typically nonconvex) subspace of the whole parameter space in which the proportion of viable parameter points is much higher than in the whole space. Knowledge of this subspace can dramatically reduce the number of samples needed to characterize the viable space. It also permits us to acquire a large number of uniformly distributed viable parameter points. The advantages of our method are especially dramatic in high-dimensional parameter spaces. It allows us to explore high dimensional nonconvex and poorly connected viable regions more efficiently and accurately than iterative Gaussian sampling [20] or uniform sampling of the entire parameter space [21-25]. Moreover, in the test problems we studied, the number of sampled parameters necessary to estimate the volume of the viable space to a given accuracy scales exponentially with the number of dimensions for Gaussian and uniform sampling, whereas it scales linearly for our algorithm. This suggests that for a given fixed complexity of the viable space, the computational effort of our method scales linearly with the dimensionality of the parameter space. This allows our method to explore high dimensional viable spaces efficiently.

An intrinsic limitation of our approach is imposed by the potential increase of the viable space's geometric complexity, when the dimension of the parameter space also increases. That is, increasing the dimensionality may cause the emergence of more poorly connected

viable regions, which can exponentially increase the minimum number of iterations needed to identify all poorly connected viable regions and to sample them thoroughly. A second potential limitation concerns the identification of unconnected viable regions that are far from each other. The finite sampling frequency of viable parameter points required in the global exploration prevents one from “getting lost” in high dimensional spaces, but it may not allow the algorithm to travel across the wide nonviable region that may separate two viable regions far from each other. A third limitation includes that values for the parameters involved in the global and local explorations steps need to be chosen judiciously. These parameters include the maximum frequency of sampled viable points, bounds for the frequency of accepted iterations, and scaling factors for ellipsoid expansions.

Efficient sampling of the viable space allows one to accurately estimate the viable volume to assess model robustness, to study the topology of the viable space, and to carry out a “glocal” analysis [20], in which the global characterization of the viable space is supplemented by a local analysis. To illustrate how our method enables insights into the working of a biological system, we studied simple model of a biochemical oscillator with positive and negative feedback loops that involves 12 parameters [51]. We focused our attention on oscillations in a narrow range of frequencies such as those produced by circadian clocks, and used the model to study the robustness of the oscillation period to parameter variations. When characterizing the viable space composed by parameters for which the model oscillates in a narrow period interval, our method was 13 times more accurate in estimating the viable volume than uniform brute-force sampling. In addition, it obtained 400 times more uniformly distributed viable points.

We showed that the viable space of this oscillator forms a nonconvex connected body in which three classes of parameter points exist. They correspond to model architectures where the negative feedback loop, the positive feedback loop, or both loops are essential for fixed period oscillations. We also found that topologies with an essential negative feedback loop provide more robust fixed period oscillations than those based on an essential positive loop. Moreover, the addition of a non-essential positive feedback loop to a model with an essential negative feedback loop increases the number of parameter combinations that give rise to viable oscillations, and it therefore increases the robustness of fixed period oscillations. In spite of the model’s simplicity, these results are consistent with well known structural properties of circadian oscillators: they typically rely on positive and negative feedback loops [58-60], the negative feedback alone is sufficient for fixed period oscillations

[61-65], and the positive feedback loop increases the robustness of the oscillations to parameter changes [19,29-32]. These results reinforce the use of robustness as a tool for model discrimination [5,19]. Specifically, we observed that among the three model architectures that permit viable oscillations, the basic topology of circadian oscillators in nature coincides with the most robust one formed by an essential negative feedback loop and a non essential positive feedback loop.

In summary, we have introduced an efficient algorithm that explores and characterizes the often tiny regions of a parameter space in which a model displays a desired behavior. We have applied our method to a biological model, but it is not restricted to such systems. It is suitable for all models with many parameters whose values are not well constrained by experimental data. Its spectrum of applications ranges from systems biology [66] all the way down to atomic physics [67].

An implementation of our algorithm in MATLAB is available as the package HYPERSPACE from <http://www.ieu.uzh.ch/wagner/software> and <http://www.csb.ethz.ch/tools/index>.

Additional material

Additional file 1: Supplementary Information for “Efficient Characterization of High-Dimensional Parameter Spaces for Systems Biology”.

This document shows additional technical information about: • The calculation of minimum volume enclosing ellipsoids involved in OEAMC, MEBS, and the construction of the integration domain. • The determination of the number of clusters involved in the construction of the integration domain. • The acquisition of viable parameter points near the boundary of the viable space involved in the MEBS. • The choice of starting points for new ellipsoid expansions involved in MEBS. • The exploration and volume calculation of spherical shells. • The exploration and volume calculation of the viable space associated to biochemical oscillator model. • Characterization of the viable space of a model of the mammalian circadian oscillator with two feedback loops.

Acknowledgements

We would like to acknowledge support through SNF grants 315200-116814, 315200-119697, and 315230-129708, as well as through the YeastX project of SystemsX.ch. EZS wants to thank Eric Hayden, Karthik Raman, and Adrian Lopez Garcia de Lomana for a careful reading of the manuscript and revealing discussions.

Author details

¹Department of Biochemistry, University of Zurich, Zurich, Switzerland. ²Department of Biosystems Science and Engineering, ETH Zurich, Zurich, Switzerland. ³Swiss Institute of Bioinformatics, Lausanne, Switzerland. ⁴School of Computer and Communication, EPFL, Lausanne, Switzerland. ⁵The Santa Fe Institute, Santa Fe, New Mexico, USA. ⁶Department of Biology, University of New Mexico, Albuquerque, New Mexico, USA.

Authors’ contributions

Project planning: EZS, JS, AW. Development of the theory: EZS. Conceived and designed the experiments: EZS, MH, JS, AW. Performed the experiments: EZS. Analyzed the data: EZS, MH, JS, AW. Contributed reagents/materials/analysis tools: EZS, MH, AI. Creation of the figures: EZS, MH. Wrote the paper: EZS, JS, AW. All authors read and approved the final manuscript.

Received: 20 January 2011 Accepted: 15 September 2011
Published: 15 September 2011

References

- Ideker T, Galitski T, Hood L: **A new approach to decoding life: Systems biology.** *Annual Review of Genomics and Human Genetics* 2001, **2**:343-372.
- Palsson BØ: *Systems biology: Properties of Reconstructed Networks* New York: Cambridge University Press; 2006.
- Goodsell D: *Our Molecular Nature: The Body's Motors, Machines and Messages* New York: Copernicus; 1963.
- Goodsell D: *The Machinery of Life* New York: Springer-Verlag; 1993.
- Stelling J, Sauer U, Szallasi Z, Doyle FJ III, Doyle J: **Robustness of cellular functions.** *Cell* 2004, **118**:675-685.
- Chaves M, Sengupta A, Sontag E: **Geometry and topology of parameter space: investigating measures of robustness in regulatory networks.** *J Math Biol* 2009, **59**:315-358.
- Dayarian A, Chaves M, Sontag E, Sengupta A: **Shape, Size, and Robustness: Feasible Regions in the Parameter Space of Biochemical Networks.** *PLoS Comp Biol* 2009, **5**:e1000256.
- Moronashi M, Winn A, Borisuk M, Bolouri H, Doyle J, Kitano H: **Robustness as a Measure of Plausibility in Models of Biochemical Networks.** *J theor Biol* 2002, **216**:19-30.
- Wagner A: *Robustness and Evolvability in Living Systems* Princeton: Princeton University Press; 2005.
- Gonze D, Halloy J, Goldbeter A: **Robustness of circadian rhythms with respect to molecular noise.** *PNAS* 2002, **99**:673-678.
- Gonze D, Goldbeter A: **Circadian rhythms and molecular noise.** *Chaos* 2006, **16**:26110.
- Zak DE, Stelling J, Doyle FJ III: **Sensitivity analysis of oscillatory (bio) chemical systems.** *Computers and Chemical Engineering* 2005, **29**:663-673.
- Leloup JC, Goldbeter A: **Chaos and biorhythmicity in a model for circadian oscillations of the per and tim proteins in drosophila.** *J theor Biol* 1999, **198**:445-459.
- Ma L, Iglesias PA: **Quantifying robustness of biochemical network models.** *BMC Bioinformatics* 2002, **3**:38.
- Leloup JC, Goldbeter A: **Modeling the mammalian circadian clock: Sensitivity analysis and multiplicity of oscillatory mechanisms.** *J theor Biol* 2004, **230**:541-562.
- Battotokh D, Tyson JJ: **Bifurcation analysis of a model of the budding yeast cell cycle.** *Chaos* 2004, **14**:653-661.
- Stelling J, Gilles ED, Doyle FJ III: **Robustness properties of circadian clock architectures.** *PNAS* 2004, **101**:13210-13215.
- Tsumotoa K, Yoshinaga T, Iida H, Kawakami H, Aihara K: **Bifurcations in a mathematical model for circadian oscillations of clock genes.** *J theor Biol* 2006, **239**:101-122.
- Saithong T, Painter KJ, Millar AJ: **Consistent Robustness Analysis (CRA) Identifies Biologically Relevant Properties of Regulatory Network Models.** *PLoS ONE* 2010, **5**:e15589.
- Hafner M, Koeppel H, Hasler M, Wagner A: **"Glocal" Robustness Analysis and Model Discrimination for Circadian Oscillators.** *PLoS Comput Biol* 2009, **5**:e1000534.
- Barkai N, Leibler S: **Robustness in simple biochemical networks.** *Nature* 1997, **387**:913-917.
- von Dassow G, Meir E, Munro EM, Odell GM: **The segment polarity network is a robust developmental module.** *Nature* 2000, **406**:188-192.
- Blüthgen N, Herzog H: **How robust are switches in intracellular signaling cascades?** *J theor Biol* 2003, **225**:293-300.
- Wagner A: **Circuit topology and the evolution of robustness in two-gene circadian oscillators.** *PNAS* 2005, **102**:11775-11780.
- Zhang J, Yuan Z, Xiong H, Zhou T: **Architecture-Dependent Robustness and Bistability in a Class of Genetic Circuits.** *Biophysical Journal* 2010, **99**:1034-1042.
- Powell W: *Approximate Dynamic Programming: Solving the Curses of Dimensionality* Hoboken: Wiley; 2007.
- Tyson JJ, Chen KC, Novak B: **Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell.** *Curr Opin Cell Biol* 2003, **5**:221-231.
- Novak B, Tyson JJ: **Design principles of biochemical oscillators.** *Nat Rev Mol Cell Biol* 2008, **9**:981-991.
- Cheng P, Yang Y, Liu Y: **Interlocked feedback loops contribute to the robustness of the Neurospora circadian clock.** *Proc Natl Acad Sci USA* 2001, **98**:7408-7413.
- Tsai T, Choi YS, Ma W, Pomeroy JR, Tang C, Jr JEF: **Robust, Tunable Biological Oscillations from Interlinked Positive and Negative Feedback Loops.** *Science* 2008, **321**:126-129.
- Saithong T, Painter KJ, Millar AJ: **The Contributions of Interlocking Loops and Extensive Nonlinearity of Circadian Clock Models.** *PLoS ONE* 2010, **5**: e13867.
- Trane C: **Robustness Analysis of Intracellular Oscillators with Application to the Circadian Clock.** *PhD thesis, Royal Institute of Technology, Automatic Control Lab* 2009.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E: **Equation of state calculation by fast computing machines.** *J Chem Phys* 1953, **21**:1087-1092.
- Landau L, Lifshitz E: *Fisica Estadística* Barcelona: Reverte; 1969.
- Kirkpatrick S, Gelatt C, Vecchi M: **Optimization by simulated annealing.** *Science* 1983, **220**:671-680.
- Newman JEM, Barkema GT: *Monte Carlo Methods in Statistical Physics* New York: Oxford University Press; 1999.
- Battotokh D, Asch DK, Case ME, Arnold J, Schüttler HB: **An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of Neurospora crassa.** *PNAS* 2002, **99**:16904-16909.
- Yu Y, Dong W, Altimus C, Tang X, Griffith J, Morello M, Dudek L, Arnold J, Schüttler HB: **A genetic network for the clock of Neurospora crassa.** *PNAS* 2007, **104**:2809-2814.
- Dong W, Tang X, Yu Y, Nilsen R, Kim R, Griffith J, Arnold J, Schüttler HB: **Systems Biology of the Clock in Neurospora crassa.** *PLoS ONE* 2008, **3**: e3105.
- Brown KS, Sethna JP: **Statistical mechanical approaches to models with many poorly known parameters.** *Phys Rev E* 2003, **68**:021904.
- Brown KS, Hill CC, Calero GA, Myers CR, Lee KH, Sethna JP, Cerione RA: **The statistical mechanics of complex signaling networks: nerve growth factor signaling.** *Phys Biol* 2004, **1**:184-195.
- Ingber L: **Adaptive simulated annealing (ASA): Lessons learned.** *Control and Cybernetics* 1996, **25**:33-54.
- Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp J: **Systems biology: Parameter estimation for biochemical models.** *FEBS Journal* 2009, **276**:886-902.
- Marinari E, Parisi G: **Simulated tempering: A new Monte Carlo scheme.** *Europhysics Letters* 1992, **19**:451-458.
- Geyer C, Thompson E: **Annealing Markov chain Monte Carlo with applications to ancestral inference.** *J Amer Statistical Assoc* 1995, **90**:909-920.
- Andrieu C, Thoms J: **A tutorial on adaptive MCMC.** *Stat Comput* 2008, **18**:343-373.
- Khachiyan L: **Rounding of polytopes in the real number model of computation.** *Mathematics of Operations Research* 1996, **21**:307-320.
- Schwaab M, Biscaia EC, Monteiro J, Pinto J: **Nonlinear parameter estimation through particle swarm optimization.** *Chem Eng Sci* 2008, **63**:1542-1552.
- Press WH, Teukolsky SA, Vetterling WS, Flannery BP: *Numerical Recipes in C* New York: Cambridge University Press; 1992.
- Eissing T, Allgöwer F, Bullinger E: **Robustness properties of apoptosis models with respect to parameter variations and intrinsic noise.** *IEE Proc-Syst Biol* 2005, **152**:221-228.
- Hafner M, Koeppel H, Wagner A: **Evolution of feedback loops in oscillatory systems.** *Third International Conference on Foundations of Systems Biology in Engineering*, (arXiv:1003.1231v1 [q-bio.QM]) 2009.
- Hodgkin AL, Huxley F: **A quantitative description of membrane current and its application to conduction and excitation in nerve.** *J Physiol* 1952, **117**:500-544.
- Meyer T, Stryer L: **Molecular model for receptor-stimulated calcium spiking.** *PNAS* 1988, **85**:5051-5055.
- Pomenenring JR, Sontag ED, Ferrell JE Jr: **Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2.** *Nat Cell Biol* 2003, **5**:346-351.
- Vilar J, Kueh H, Barkai N, Lieber S: **Mechanism of noise-resistance in genetic oscillators.** *PNAS* 2002, **99**:5988-5992.

56. Novak B, Tyson J: Numerical analysis of a comprehensive model of M-phase control in *Xenopus* oocyte extracts and intact embryos. *J Cell Sci* 1993, **106**:1153-1168.
57. Klipp E, Herwig R, Kowald A, Wierling C, Lehrach H: *Systems biology in practice* Weinheim: Wiley-vch; 2006.
58. Lee K, Loros J, Dunlap JC: Interconnected feedback loops in the *Neurospora* circadian system. *Science* 2000, **289**:107-110.
59. Gallego M, Virshup DM: Post-translational modifications regulate the ticking of the circadian clock. *Nat Rev Mol Cell Biol* 2007, **8**:139-148.
60. Rust MJ, Markson JS, Lane WS, Lane DS, Fisher DS, O'Shea EK: Ordered Phosphorylation Governs Oscillation of a Three-Protein Circadian Clock. *Science* 2007, **318**:809-812.
61. Bunger MK, Wilsbacher LD, Moran SM, Clendenin C, Radcliffe LA, Hogenesch JB, Simon MC, Takahashi JS, Bradfield CA: **Mop3 is an essential component of the master circadian pacemaker in mammals.** *Cell* 2000, **103**:1009-1017.
62. Smolen P, Baxter D, Byrne JH: Modeling circadian oscillations with interlocking positive and negative feedback loops. *The Journal of Neuroscience* 2001, **21**:6644-6656.
63. Smolen P, Baxte D, Byrne JH: A reduced model clarifies the role of feedback loops and time Delays in *Drosophila* circadian oscillator. *J Biophys* 2002, **86**:2786-2802.
64. Becker-Weimann S, Wolf J, Herzel H, Kramer A: Modeling Feedback loops of the Mammalian Circadian Oscillator. *J Biophys* 2004, **87**:3023-3034.
65. Locke JCL, Southern MM, Kozma-Bognar L, Hibberd V, Brown PE, Turner MS, Millar AJ: Extension of a genetic network model by iterative experimentation and mathematical analysis. *Molecular Systems Biology* 2005, **1**.
66. 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**:e189.
67. Mortensen JJ, Kaasbjerg K, Frederiksen SL, Nørskov JK, Sethna JP, Jacobsen KW: Bayesian Error Estimation in Density-Functional Theory. *Phys Rev Lett* 2005, **95**:216401.

doi:10.1186/1752-0509-5-142

Cite this article as: Zamora-Sillero et al.: Efficient characterization of high-dimensional parameter spaces for systems biology. *BMC Systems Biology* 2011 **5**:142.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

