Attractor Landscapes as a Model Selection Criterion in Data Poor Environments

Cole A. Lyman[1][0000−0001−7921−165X], Spencer Richman[1][0000−0002−6479−1687], Matthew C. Morris[1][0000−0002−1236−1255], Hongbao Cao[2], Antony Scerri[2], Chris Cheadle[2], and Gordon Broderick[1][0000−0002−3588−8252]

[1] Center for Clinical Systems Biology, Rochester General Hospital, Rochester, NY, USA
{Cole.Lyman,Gordon.Broderick}@rochesterregional.org
[2] In Silico Biology, Elsevier, Amsterdam, Netherlands

Abstract. Modeling of systems for which data is limited often leads to underdetermined model identification problems, where multiple candidate models are equally adherent to data. In such situations additional optimality criteria are useful in model selection apart from the conventional minimization of error and model complexity. This work presents the attractor landscape as a domain for novel model selection criteria, where the number and location of attractors impact desirability. A set of candidate models describing immune response dynamics to SARS-CoV infection is used as an example for model selection based on features of the attractor landscape. Using this selection criteria, the initial set of 18 models is ranked and reduced to 7 models that have a composite objective value with a p-value < 0.05. Additionally, the impact of pharmacologically induced remolding of the attractor landscape is presented.

Keywords: Multi-valued Logic Networks · Attractors

1 Introduction

Even though broad-spectrum data is becoming increasingly accessible, biological modeling methods designed to operate in data poor environments are still critically important. The ongoing SARS-CoV-2 pandemic offers a highly relevant example of this where the research community has been called upon to make informed predictions of appropriate interventions when very little to no data is available, in particular molecular level data. Data-poor environments often result in underdetermined model identification problems, i.e. models for which there are more degrees of freedom than data-imposed constraints, which generally leads to multiple candidate models with equivalent adherence to a given objective. Under such conditions, multiple equivalently optimal sets of model parameter values exist. It is important however to consider the specific criteria used to define optimality of a given model. Adherence of model predictions to available data is a standard optimality criterion. Additionally, the complexity of a model is often minimized concurrently to achieve the most parsimonious adherence to data [36]. While adherence to data and parsimony are important criteria for assessing the desirability of a model, in cases where there are a large number of equally optimal models, i.e. equally data adherent and equally parsimonious, additional criteria may be required to further discern biological plausibility.

We propose that the number, type, and location of dynamically stable attractors (i.e. the attractor landscape) constitute useful criteria for further assessing the optimality of a regulatory network model. This is based on the widely accepted assumption that attractors can be used to represent phenotypes in models of biological systems [3,7,16,17,31,35]. In this work, we focus on qualitative models of immune regulation [20,44] and present methods to exhaustively compute attractors using this framework. Even though this work focuses on qualitative regulatory models, the ideas presented herein can be applied to any dynamical model type and are especially useful in mathematically underdetermined settings.

To demonstrate the utility of leveraging the attractor landscape to support additional optimality criteria, 19 distinct candidate models of the immune regulatory response to SARS-CoV [26] are analyzed. All competing candidate models reproduce the experimental data to within 5% departure and 11 of these match the available data exactly (100% accuracy). Furthermore, all of the models have a similar structural and parametric efficiency (i.e. structural and functional parsimony). The attractor landscapes, that is stationary steady states and limit cycles, for these 19 models are presented. These are compared in terms of relative abundance,

location in state space, and oscillatory period along with their plausible biological counterparts. Moreover, the effects of potential repurposed drugs on these attractor landscapes are demonstrated suggesting that while some drugs serve to destabilize illness states, others also destabilize and make normal homeostatic regulation inaccessible.

1.1   Related Work

Identifying attractors in Boolean Networks has been shown to be an NP-complete problem in general [1, 50], but under certain conditions can be polynomially reduced to a satisfiability (SAT) problem [43]. Despite its NP-completeness, much work has been done to develop algorithms to efficiently characterize attractors. Some methods refine the state transition graph [2, 11, 21, 39], decompose the state transition graph into strongly connected components [25, 48], or reduce the model itself [45], thereby limiting the search space and increasing efficiency. An alternative approach to efficiently identify attractors is to exploit SAT [10, 14] and integer linear programming [32] methods. Moreover, some methods compute attractors by encoding the Boolean functions as binary decision diagrams [49, 51]. Finally, constraint satisfaction programming [9] has been utilized and is the paradigm upon which the current work is based.

Regardless of the multitude of methods proposed to identify attractors, to the best of the authors' knowledge this work is the first to use the landscape defined by such attractors as a criterion for model selection.

2   Methods

This section briefly describes how 19 candidate regulatory models of immune response to SARS-CoV infection were obtained, and then proceeds with formal definitions of the model parameters and corresponding attractors the models support. Next, an algorithm for exhaustive attractor identification via constraint satisfaction programming is proposed. Finally, the criteria for evaluating a model's performance is defined.

The network structure common to these 19 dynamic immune response models of SARS-CoV infection [26] consisted of 19 immune mediators (vertices) linked by 112 regulatory actions (edges) (see supplemental Fig. S1 for the network structure). These regulatory interactions were extracted from the Elsevier Knowledge Graph* (Elsevier, Amsterdam) [19] using the Pathway Studio* suite of tools [28] and were based on the automated text-mining [8, 29] of 2,653 published references. The system is modeled using a discrete dynamic network logic framework [20, 44]. The predicted dynamic behaviors supported by any set of decisional logic parameters enacting these models were constrained to reproduce publicly available experimental measurements from time-series experiments (GEO accession number GSE33267) where human Calu-3 lung adenocarcinoma cells were infected *in vitro* by SARS-CoV over 72 hours with regular sampling for transcriptomic sequencing [38]. This dataset was chosen because high quality transcriptomic data of SARS-CoV-2 was not available at the time of analysis and SARS-CoV is closely related to SARS-CoV-2 [34]. The departure from these discretized data as well as structural and functional model complexity were concurrently minimized.

2.1   Attractor Definition

Let A be the set of attractors that a given model supports, i.e. the attractor landscape, where $a \in A$ is an attractor. An attractor a is a set of states such that when a model reaches a state $a_i \in a$ the model will update to a state $a_j \in a$. A state is a vector of length m where each element corresponds to a discrete value for an entity (vertex) in the model.

It follows that the number of states in a ($|a|$) is the period of the attractor, with the special case of $|a| = 1$ where the attractor is not a limit cycle (or oscillatory attractor) and can be referred to as a fixed stationary point, stable state, or steady state. An attractor can be viewed as a loop in the state transition graph where a steady state is a self-loop and a limit cycle of period n is a cycle of n states within the state transition graph.

---

*Copyright © 2021 Elsevier Limited except certain content provided by third parties. Pathway Studio is a trademark of Elsevier Limited.

2.2   Attractor Search Constraint Satisfaction Problem

Constraint Satisfaction Problems (CSP) are a class of optimization problems consisting of models that are composed of free variables, usually discrete, such that the permitted assignment of these variables satisfies all of the stipulated constraints that define the problem. A CSP solution is found when all variables are assigned and all the constraints are met. Here, the attractors for a model are formulated as a CSP and are encoded using the MiniZinc language (version 2.4.3) [27, 41] and the Chuffed solver (version 0.10.4) [6] to discover solutions (e.g. attractors). Pseudo-code of the CSP is found in the supplemental Algorithm S1.

The first variable defined in the CSP is X, a $(n+1) \times m$ matrix that represents an attractor of period n where $n > 0$, m is the number of entities in the model, and where each row is a state within the attractor. The next variable is I, a $(n+1) \times m$ matrix that represents the row-wise corresponding images for the states in X. The image I specifies the discrete values that the entities should move towards.

Next, constraints are stipulated to ensure that the states in X evolve according to the dynamics of the parameterized model. Formally, this constraint is $\wedge_{i=1}^{n} X_{(i+1)*} = H(X_{i*}, I_{i*})$ where X_{i*} represents the ith row of X, I_{i*} represents the ith row of I, and the H function computes the next state. The next state $X_{(i+1)*}$ is determined by comparing the previous state X_{i*} to the image I_{i*} and updating entities according to the rules of the synchronous update scheme (see [5, 9, 36] for details on how the image computed and updates are performed in CSP). Also, the first and last states in X are constrained to be equal, thereby making X an attractor by definition. Formally, this constraint is $X_{1*} = X_{(n+1)*}$.

An additional constraint only applies to cyclic attractors where $n > 1$, and constrains the first state in the cycle to not be equal to all of the other states excluding the last state. Formally, it is $\wedge_{i=2}^{n} X_{1*} \neq X_{i*}$ and ensures that the steady states (attractors where $n = 1$) aren't classified as limit cycles; because without this constraint, all steady states repeated n times would be valid solutions. The last constraint also only applies when $n > 1$ and ensures that the states in X are sorted in increasing order according to their sum. This constraint simply breaks the symmetry in the search space to prevent the same limit cycle from creating duplicate solutions. Without this constraint, each limit cycle would have n solutions where each solution has one of the n states as X_{1*}. Formally, it is $X = \text{increasing}(X)$ where the increasing function orders the rows of X increasingly by the L1-norm of the row.

Due to a technical limitation of MiniZinc which only allows static sized arrays, the variable n is treated as a parameter to the CSP. This means that separate searches for each n must be performed. The occurrence of attractors of period n tends to follow a power-law distribution [13] (also see Fig. 3b) with longer periods being increasingly rare. Thus, the search for attractors starts with $n = 1$ and continues up to a pre-defined maximum period.

2.3   Pharmacologically Redirecting Attractors

Simulating how a drug will modify the attractor landscape is also accomplished as a CSP. First, the targets of a specific drug and the regulatory actions imparted to each of these targets are recovered based on data collected across multiple drug-action databases, including the Elsevier Reaxsys* database and integrated into the Elsevier Knowledge Graph*. In each case the targets are limited to those entities included in the model. Finally, an idealized drug action is applied such that it constrains the activation state of their corresponding targets in the network to a constant upregulated or downregulated state (assumption of continued administration) depending on that drug's specific mode-of-action.

For a drug that is agonistic to a target (i.e. the drug up-regulates the target), the state value for that target is constrained to be at a level greater than its minimum level. For example, if a drug agonizes entity $v_i \in V$, then $X_{*i} > 0$ where X_{*i} is the ith column of X. Conversely, in the case of a drug that antagonizes a target (i.e. the drug down-regulates the target), the state value for that target is constrained to be less than its maximum level. This is formally represented as $X_{*i} < \rho_i$ where $v_i \in V$ is the entity that the drug antagonizes and ρ_i is the maximum level for v_i. Note that this definition, when applied to a ternary target ($\rho_i = 2$), allows for the state to have a nominal value of 1 indicating a smaller dose of the drug.

---

*Copyright © 2021 Elsevier Limited except certain content provided by third parties. Pathway Studio is a trademark of Elsevier Limited.

2.4   Model Performance Criteria

The formal definitions of the criteria used to select models are as follows. First, the adherence to data O_d is calculated using (1) as the L1-norm $\| \cdot \|_1$ (Manhattan distance) of the difference between the transient data D and the model-predicted trajectories Tr where $|D| = |Tr| = L$ is the number of transient trajectories, m is the number of entities in the model, $Tr^i \in Tr$ is a $l_i \times m$ matrix of the trajectory corresponding to the data $D^i \in D$, and $\mathbf{tr}_{j*}^i$ ($\mathbf{d}_{j*}^i$) is the jth row of Tr^i (D^i) and $\mathbf{tr}_{jk}^i$ ($\mathbf{tr}_{jk}^i$) is the kth element of the jth row of Tr^i (D^i).

$$O_d = \sum_{i=1}^{L} \| \mathbf{tr}_{j*}^i - \mathbf{d}_{j*}^i \|_1 \tag{1}$$

Optimal adherence to data is quantified as a minimal deviation of measured and predicted values, expressed as a fraction of the maximum possible error $\frac{O_d}{P}$, which accounts for absent data represented by $\perp$ [36] and ρ_k is the maximum level for $v_k \in V$. The maximum level ρ_k is used because it represents the largest possible amount of deviation between measured and predicted values when data is present.

$$P = \sum_{i=1}^{L} \sum_{j=1}^{l_i} \sum_{k=1}^{m} \begin{cases} \rho_k & \mathbf{d}_{jk}^i \neq \perp \\ 0 & \mathbf{d}_{jk}^i = \perp \end{cases}$$

Additionally the model's structural complexity O_c should also be minimal and is computed here as the sum of the threshold values ω for all edges E, as illustrated in (2) (see [36] for further details on how complexity is calculated). Let $(i, j) \in E$ be an edge from $v_i \in V$ to $v_j \in V$ with a corresponding activation threshold of ω_{ij}.

$$O_c = \sum_{(i,j) \in E} \omega_{ij} \tag{2}$$

Turning now to the attractor space, the distance separating a given model's attractors from a set of biologically relevant reference stable states O_r should also be minimal. This is similarly calculated using a weighted sum of the L1-norm separating each reference state from the most proximal model-predicted attractors, as shown in (3). Let $(z_r, r) \in R$ be the set of reference stable states, e.g. stable health and persistent illness, where r is a reference stable state and z_r is the associated weight of significance and let $a \in A$ be an attractor in the set of attractors A for a given model. In this work the reference states are all equally weighted.

$$O_r = \sum_{(z_r, r) \in R} z_r \min(\{\| r - a \|_1 \mid \forall a \in A\}) \tag{3}$$

Third, we posit that the distribution of the periods corresponding to the cyclic attractors predicted by a given model should ideally favor higher frequencies, or in other words, be biased in favor of shorter periods. This translates to higher values of O_p, as defined in (4) where $|a| = n$ indicates that the attractor $a \in A$ has a period of n, $N = \{|a| \mid a \in A\}$ is the set of periods of the attractors in A, and $\lambda_n(A)$ is a function that returns the number of attractors in A with period n.

$$O_p = \sum_{n \in N} \frac{\lambda_n(A)}{n} \tag{4}$$

Finally the size of the basin, defined here (5) as the relative number of initial seed states that converge to that attractor, surrounding the non-pathological reference steady states O_f should be maximal. Let $F(a)$ be the relative frequency of convergence to attractor $a \in \hat{A}$ where $\hat{A} \subseteq A$ are attractors supported under non-pathological circumstances (e.g. no coronavirus present in the system) and where $\hat{R} \subseteq R$ are non-pathological reference attractors.

$$O_f = \sum_{a \in \hat{A}} \frac{F(a)}{\min(\{\| r - a \|_1 \mid \forall r \in \hat{R}\}) + 1} \tag{5}$$

2.5   Model Selection

To rank the desirability of 19 models studied here based on their dynamic properties and the stable phenotypes they support, the 5 separate objective values in Section 2.4 are reduced to one composite value using Principal Component Analysis (PCA) [47]. The raw objective values for each objective are first individually scaled such that they range from 0 to 1. Additionally, objective values that should be maximal in the most suitable solution (O_p and O_f) are transformed so that the direction of each objective value agree. Error O_d requires no scaling as it is already expressed as a percentage of the maximum error P. Structural complexity O_c is scaled in terms of the maximum complexity value. Likewise, the adherence to reference stable states O_r is scaled in terms of the maximum possible distance to the states. The reciprocal of the distribution of the periods O_p is taken $\frac{1}{O_p}$, because O_p value should be maximal. Lastly, the frequency of attractors most proximal to non-pathological reference steady states O_f are filtered using add-one smoothing and the reciprocal is taken $\frac{1}{O_f+1}$. This transformation is performed because the minimum of O_f is 0 and it should be maximized. Table 1 shows each objective value after their respective transformations.

Following a PCA, the score values along the first latent feature PC1 (which explains 35% of the variation) are examined using the Shapiro-Wilk test to assess departure from normality [37] resulting in a p-value of 0.6935, suggesting PC1 scores describing an aggregate of the objective values do not depart significantly from a normal distribution (see Fig. 2a for a Quantile-Quantile plot of the distribution). Accordingly, a one-sided Student's t-test [42] is performed to identify models having a significantly lower PC1 aggregate score. Table 1 lists the PC1 transformed objective values and the corresponding p-values for each model, where an asterisk indicates $\alpha \leq 0.05$. Furthermore, the PC1 transformed objective values with the corresponding Student's t-test p-values are illustrated graphically in Fig. 2b.

Table 1. Objective Values

Model ID	O_d	O_c	O_r	O_p	O_f	PC1 Transformed Objective	P-Value
18	0	0.7	0.019	0.19	0.83	-2.5	2.5e-08*
13	0	0.7	0.11	0.23	0.9	-1.5	3.6e-05*
17	0	0.7	0.019	0.2	0.96	-1.4	1e-04*
14	0	0.7	0.11	0.17	0.94	-1.3	0.00017*
19	0	0.7	0.11	0.12	0.97	-1.2	0.00051*
12	0	0.7	0.15	0.14	0.96	-1.2	0.00055*
15	0	0.7	0.019	0.26	1	-0.83	0.0066*
16	0	0.7	0.26	0.25	0.96	-0.52	0.051
11	0	0.8	0.056	0.1	0.97	-0.18	0.28
10	0	0.8	0.056	0.25	0.95	0.1	0.63
9	0	0.8	0.074	0.25	0.96	0.24	0.78
7	0.01	0.8	0.037	0.25	0.96	0.55	0.96
8	0.01	0.8	0.056	0.25	0.96	0.6	0.97
6	0.02	0.8	0.093	0.25	0.96	1.1	1
4	0.03	0.8	0.056	0.29	0.91	1.1	1
1	0.04	0.8	0.037	0.12	0.95	1.3	1
2	0.04	0.8	0.037	0.13	0.96	1.5	1
5	0.02	0.8	0.31	0.33	0.96	2	1
3	0.03	0.8	0.37	0.18	0.97	2.2	1

3   Results

3.1   Model Selection

The original 19 SARS-CoV candidate models, all adhered to the experimental data to within 5% error with 11 of these matching the data exactly. A closer examination of the underlying regulatory complexity shows

that 8 of these 11 models are less structurally complex (e.g. more parsimonious), where structural complexity is measured by the number and value of activation thresholds ω for each edge [36]. Conventional criteria of model error and parsimony would therefore leave us with 8 equally plausible models. We propose that these data-adherent parsimonious models could now be further ranked based on their attractor landscapes. We propose that such an extended model ranking scheme could not only favor 1. conventional adherence of model predictions to all available data, both steady state and transient, with minimal error and 2. structural and functional parsimony, but more importantly by also 3. biologically plausible predicted resting states that are maximally proximal to a set of stable *reference* states that are phenotypically relevant, 4. cyclic behaviors that favor shorter periods of oscillation, and finally 5. the accurate prediction of a reference non-pathological resting state with a broad basin of attraction as evidenced by high frequency of return after perturbation. As we require each model's attractor landscape to demonstrate these features, we first conduct an exhaustive search of available stable states ($n = 1$) and limit cycles ($1 < n \leq 20$) supported by each model. We find in this example that the limit cycles identified exhibit a maximum oscillatory period of $n = 4$. Details of the number of attractors with periods $1 \leq n \leq 4$ found for each model are listed in Supplemental Table S1.

Now in addition to experimental data, we might also be able to define a number of dynamically stable *reference* states that correspond to a specific set of persistent biological phenotypes of interest, both pathological and non-pathological e.g. healthy homeostasis and chronic illness. Recovery of these *reference* states is represented by the O_r objective. One could require that these *reference* states be recovered to within a user-specified tolerance as part of the attractor space supported by any feasible data-adherent model. Here we use 2 idealized phenotypes as reference states. First, we define a steady state of immune inactivation or immune quiescence to which the system should come to rest normally in the absence of an infectious or other challenge. This is idealized here as a state where all network entities exhibit a minimal level of activation, namely 0. In addition, we define a pathological state of persistent immune activation, namely cytokine storm, where network signaling mediators are expressed at high levels of activation. In this work, cytokine storm is modeled as the final state recorded in a 72-hour time course *in vitro* experiment, where viral titer had reached $\sim 10^7$ pfu/ml and immune signaling activity was persistently elevated with respect to control. The Manhattan distance of the closest stationary steady state ($n = 1$) supported by each model to each of these reference steady states respectively is shown in Supplemental Table S1.

All of the 8 parsimonious, data-exact models support an attractor that is identical to the reference cytokine storm state. In contrast, the minimum distance to the reference state of normal immune quiescence is more varied, ranging from a Manhattan distance of 1 to 14.

Criterion (3), O_r, was applied to these results, as a score consisting of a weighted sum of the minimal Manhattan distance to all reference steady states, with a lower score being more favorable. In this case the weights for each reference state are equal. We now have only 3 models where proximity to our set of reference stable attractors is equally well supported, namely models 18, 17, and 15. Under criterion (4), O_p, we propose that the distribution of cyclic attractors is biased in favor of short oscillatory periods e.g. circadian or faster. This is expressed as a weighted sum of the number of attractors divided by their respective periods of oscillation n, with a higher score being more desirable. Results in Table 1 (and Supplemental Table S1) show that out of these 3 models, model 18 fares best in this regard with $O_p = 0.19$, while model 15 does not support any oscillatory attractors and model 15 is biased towards supporting oscillatory behaviors of longer periods. This is illustrated further in Fig. 1 where the location of attractors for models 15 and 18 are presented. Each attractor exists in a 19-dimensional state space (19 state variables or network vertices in the model) that has been projected into 2-dimensions to facilitate visualization using Multidimensional Scaling [4] (implemented in Scikit-learn [15, 30]). In the case of oscillatory attractors ($n > 1$) arrows indicate the successor of each state.

Finally, under criterion (5), O_f, we require that a non-pathological resting state (if defined) be the preferred steady state attractor for any system expected to support normal healthy homeostasis. Here we represent the breadth of the basin of attraction by the frequency with which the system returns to a given steady state in 100,000 random start states. The basins of attraction for the steady state attractors ($n = 1$) for models 15, 17, and 18 are shown in Table 2. These results indicate that in the absence of active infection, roughly 95% of the random perturbations to model 18 come to rest within a Manhattan distance of 6 bits or better from the reference resting state of immune quiescence. Indeed, 18% of the simulations conducted with model 18 come to rest in an attractor (attractor 2) located only 1 bit away from idealized immune inactivation. This is in stark contrast to models 15 and 17 where perturbations achieve such a final resting
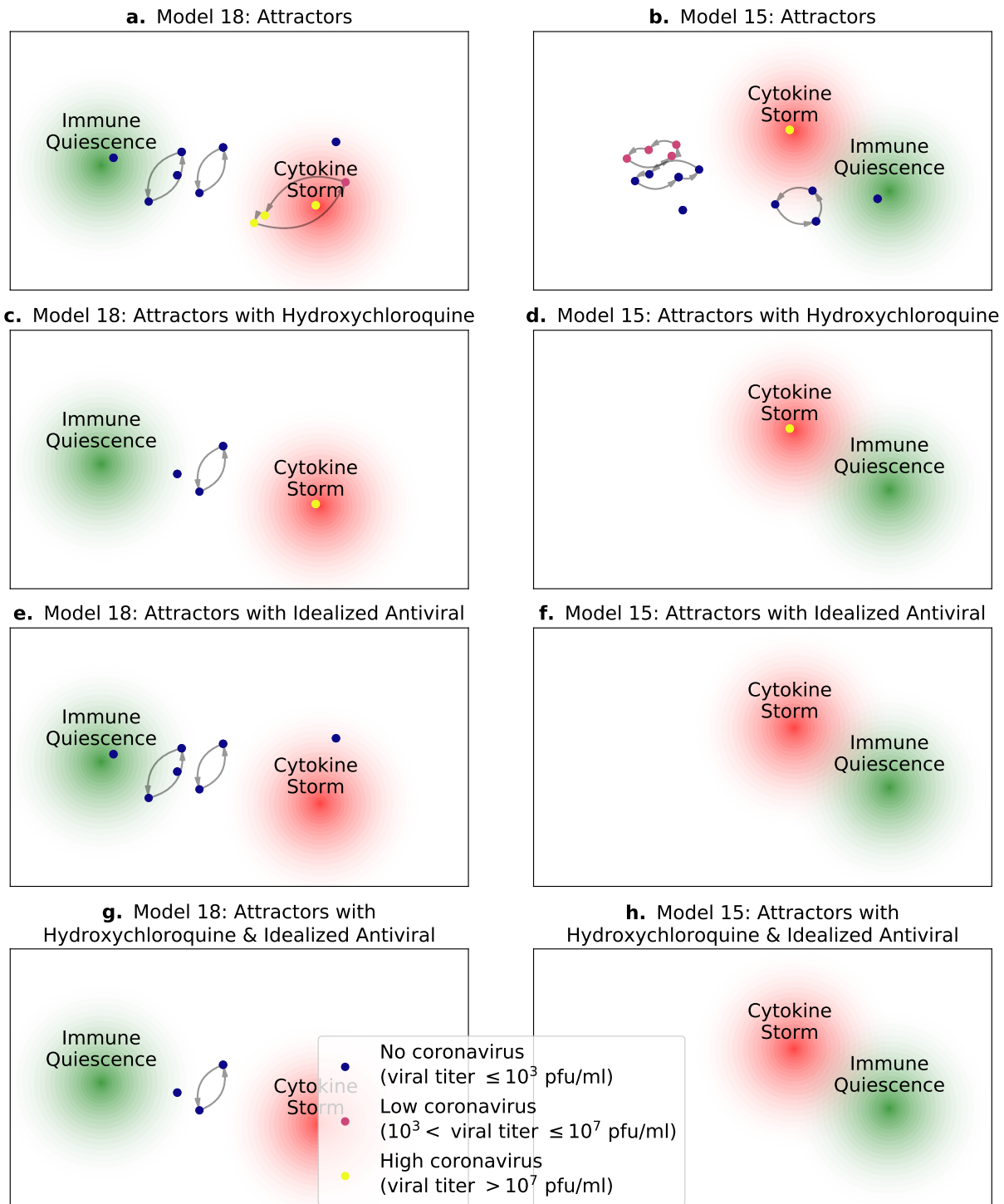
Fig. 1. Attractors for models 18 and 15 with an idealized antiviral and hydroxychloroquine redirecting the attractors. **a, b.** The attractors for models 18 and 15 respectively with no drugs constraining the attractors. **c, d.** The attractors for models 18 and 15 respectively with hydroxychloroquine redirection, notice how the attractors near immune quiescence are no longer accessible for both models. **e, f.** The attractors for models 18 and 15 respectively with an idealized antiviral. **g, h.** The attractors for models 18 and 15 respectively with hydroxychloroquine combined with the idealized antiviral. Generated using Matplotlib [18].

state proximity less than 1% of the time. This general proximity to a quiescent resting state is summarized in Table 2 by a composite score O_f consisting of the sum across all attractors predicted in the absence of active infection (no coronavirus, i.e. viral titer $\leq 10^3$ pfu/ml) of their respective frequencies of occurrence divided by the corresponding Manhattan distances to the reference immune quiescent state. Finally, all 5 objective function scores are range-adjusted and combined into a single overall aggregate desirability using the first latent score from a Principal Component Analysis (PCA), with a lower aggregate score being preferable.

Table 2. $n = 1$ Attractor Frequency

| Model ID	Coronavirus value	Attractor ID	Frequency	Distance Immune Quiescence	Distance Cytokine Storm	Normalized Immune Quiescence Score O_f
18	No coronavirus	1	1×10^{-3}	17	16	$\dfrac{1}{\left(\frac{1\times10^{-3}}{17+1} + \frac{0.18}{1+1} + \frac{0.77}{6+1}\right)+1} = 0.83$
		2	0.18	1	14	
		3	0.77	6	9	
		Remainder	0.05	-	-	
	High coronavirus	4	0.26	13	0	
		Remainder	0.74	-	-	
17	No coronavirus	1	3×10^{-5}	1	14	$\dfrac{1}{\left(\frac{3\times10^{-5}}{1+1} + \frac{0.01}{7+1} + \frac{0.99}{21+1}\right)+1} = 0.96$
		2	0.01	7	18	
		3	0.99	21	16	
	High coronavirus	4	0.03	13	0	
		5	0.97	22	13	
15	No coronavirus	1	3×10^{-5}	1	15	$\dfrac{1}{\left(\frac{3\times10^{-5}}{1+1} + \frac{0.02}{16+1}\right)+1} = 1$
		2	0.02	16	16	
		Remainder	0.98	-	-	
	High coronavirus	3	0.11	14	0	
		Remainder	0.89	-	-	

3.2   Pharmacologically Redirected Attractors

Applying an idealized antiviral and the drug hydroxychloroquine to models 18 and 15, produces the attractor landscapes shown in Fig. 1c-h. This idealized antiviral is assumed to completely eliminate viral load, and its effects are approximated here by constraining the coronavirus activation level to its lowest value, 0. Fig. 2 panels f and h show that the simulated antiviral is sufficient to disrupt all attractors supported by model 15 and that hydroxychloroquine alone leaves intact only the virally fueled cytokine storm. In contrast, simulations conducted with model 18 tell a somewhat more plausible story. In Fig. 1c when only hydroxychloroquine is applied to model 18, it succeeds in disrupting the limit cycle that orbits cytokine storm as well as a normally occurring steady state proximal to it, leaving the virally supported cytokine storm stationary steady state in place. One quiescent limit cycle is also collapsed but a stationary steady state and another limit cycle proximal to the idealized immune quiescent state remain intact. In Fig. 1e the application of the idealized antiviral alone to model 18 leaves intact all limit cycles and stationary points that were proximal to "healthy" immune quiescence. However, it disrupts and renders unavailable the cytokine storm stationary point and limit cycle, leaving only one stationary point in the vicinity of cytokine storm. Lastly, in Fig. 1g we show the attractors supported by model 18 when both hydroxychloroquine and the idealized antiviral are applied. This combination regimen eliminates the cytokine storm steady state as well as all previously proximal pathological attractors, leaving the "parking" attractors proximal to immune quiescence as accessible stable resting states.
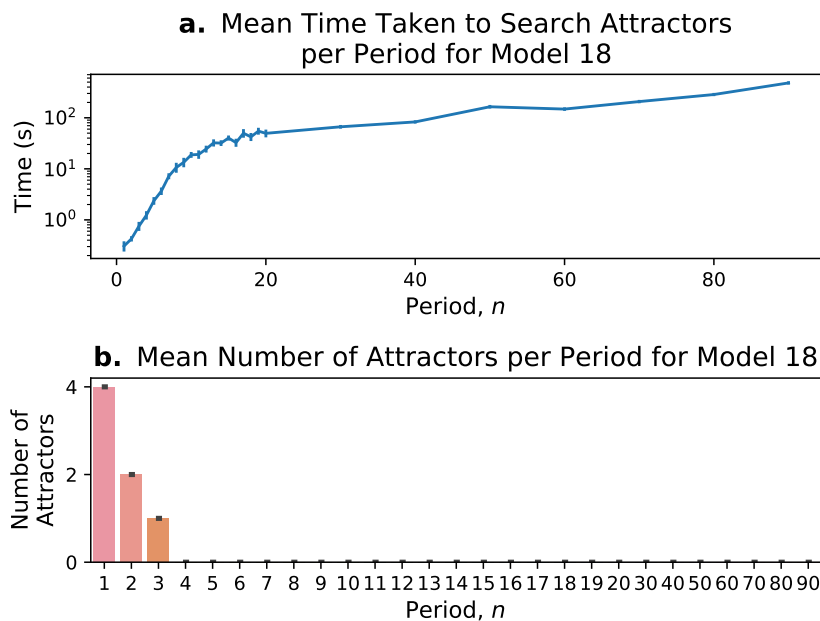
3.3   Attractor Search Time Complexity

Searching for attractors is an NP-Complete problem [1, 50], thus the time complexity of the algorithm is an important consideration. Fig. 3a illustrates the mean time (in seconds) taken to search for attractors using
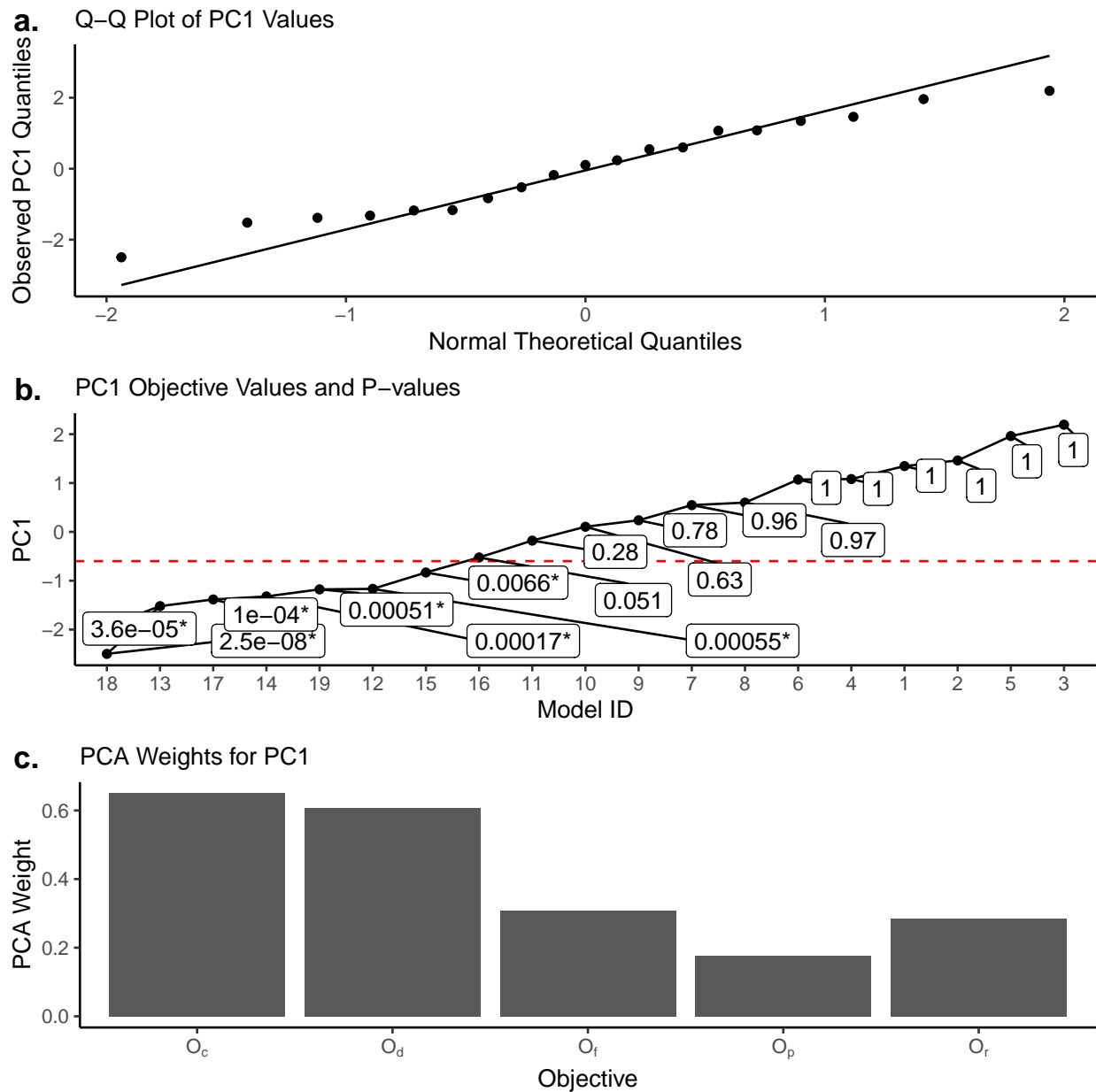
Fig. 2. Dimensionally reduced objective values using Principal Component Analysis (PCA). **a.** A Quantile-Quantile (Q-Q) plot comparing the distribution of Principal Component 1 (PC1) objective values to the Gaussian distribution. The distribution of PC1 transformed objective values, with a Shapiro-Wilk [37] p-value of 0.6935 suggesting that these values do not depart significantly from a normal distribution. **b.** PC1 objective values for each model with associated p-values for a single sample, one sided Student's t-test [42] where an asterisk indicates models with a significant p-value (< 0.05) and the dashed horizontal red line represents this significance threshold. **c.** The PCA weights for each objective. Generated using R [33], ggplot2 [46] and ggrepel [40].

Fig. 3. Time complexity and number of attractors per period. **a.** The mean time, in seconds, to search for attractors for model 18 repeated 10 times. Error bars represent the standard deviation for each period, which range from 1-20 and from 20-90 in increments of 10. **b.** The number of attractors identified for model 18 for each period, which follows a power-law distribution. Generated using Matplotlib [18].

model 18 in relation to the period of the attractor, n, where each search was repeated 10 times. Periods range from 1 to 20 (incrementing by 1) and from 20 to 90 (incrementing by 10) to provide a better sense of how the algorithm scales near-exponentially with regards to the period of the attractor. Despite the exponential scaling, searching for an attractor where $n = 90$ only takes approximately 8 minutes to determine that no such attractors exist. The whisker at each point is the standard deviation across the 10 repeats for that period. Fig. 3b illustrates the number of attractors found for that period, and shows that the attractors for model 18 follow a power-law distribution [13]. All experiments were run on a 2017 MacBook Pro with a 2.3 GHz Dual-Core Intel Core i5 processor and 8 GB of RAM.

4   Discussion

When selecting among competing models, we propose that in addition to adherence to reference data and model parsimony, the number, type, and location of attractors can offer important insight into biological plausibility. Moreover, we propose that the presence and periodicity of cyclic attractors is of special relevance to biological systems as these typically exhibit cyclic behaviors across multiple levels of resolution from the intracellular to entire organ systems [12]. Additionally, some illnesses have been shown to disrupt essential biorhythms which can be characterized as cyclic attractors, making such cycles of special therapeutic relevance.

It is interesting to note in the example presented here that hydroxychloroquine is predicted to not only disrupt the cyclic attractor proximal to cytokine store but also concurrently disrupt attractors nearest to immune quiescence. One might expect this behavior as strong immune mediators such as hydroxychloroquine aren't meant to be administered for long periods of time, nor would they be administered as the sole agent in the case of cytokine storm [24]. Such results support ongoing efforts by our group directed at the development of computational methods for design of therapeutic approaches that formally account for the recovery of such essential cyclic attractors or biorhythms.

Given the time complexity exhibited by the current approach the authors are optimistic that other more complex cycles can also be characterized using this method. For example, this method could be applied and

would be computationally feasible in the modeling of complex pulse-like oscillatory behavior to discover attractors such as those exhibited by the hypothalamic–pituitary–gonadal axis over the course of 30 days [23]. Furthermore, for less common biological phenomena that exhibit longer cycles, such as annual seasonal affective disorder [22], one could feasibly discover attractors with weekly measurements over the span of a year. In cases where an even larger number of states is needed to identify attractors efficiently the authors plan, as future work, to implement a flexible grid approach that allows for varying degrees of resolution in the attractor landscape.

5   Conclusion

This work presents a novel selection criterion of attractor landscapes that is useful in underdetermined modeling scenarios where models are equally optimal using conventional objective values. By combining conventional metrics, error and parsimony, with three attributes of the attractor landscape 18 models are ranked using a composite objective value. Out of the 18 original models, 7 models have a composite objective value that is significantly lower than the composite objective values of the other models. Using the top 2 ranking models, the effect of disrupting the attractor landscape with drug simulations is also shown. The attractor landscape is an important characteristic of network logic models and should thereby be included when selecting which models are suitable.

Acknowledgment

Mandatory Disclaimer

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of Defense.

References

1. Akutsu, T., Kuhara, S., Maruyama, O., Miyano, S.: A system for identifying genetic networks from gene expression patterns produced by gene disruptions and overexpressions. Genome Informatics **9**, 151–160 (1998). https://doi.org/10.11234/gi1990.9.151

2. Ay, F., Xu, F., Kahveci, T.: Scalable steady state analysis of boolean biological regulatory networks. PLoS ONE **4**(12), e7992 (Dec 2009). https://doi.org/10.1371/journal.pone.0007992

3. Baverstock, K.: A comparison of two cell regulatory models entailing high dimensional attractors representing phenotype. Progress in Biophysics and Molecular Biology **106**(2), 443–449 (Aug 2011). https://doi.org/10.1016/j.pbiomolbio.2011.01.002

4. Borg, I., Groenen, P.J.: Modern Multidimensional Scaling: Theory and Applications, chap. The Four Purposes of Multidimensional Scaling, pp. 3–18. Springer New York, New York, NY (2005). https://doi.org/10.1007/0-387-28981-X_1

5. Chaouiya, C., Remy, E., Mossé, B., Thieffry, D.: Qualitative analysis of regulatory graphs: A computational tool based on a discrete formal framework. In: Positive Systems, pp. 119–126. Springer Berlin Heidelberg (2003). https://doi.org/10.1007/978-3-540-44928-7_17, https://doi.org/10.1007/978-3-540-44928-7_17

6. Chu, G., de la Banda, M.G., Mears, C., Stuckey, P.J.: Symmetries, almost symmetries, and lazy clause generation. Constraints **19**(4), 434–462 (Mar 2014). https://doi.org/10.1007/s10601-014-9163-9, https://doi.org/10.1007/s10601-014-9163-9

7. Creixell, P., Schoof, E.M., Erler, J.T., Linding, R.: Navigating cancer network attractors for tumor-specific therapy. Nature Biotechnology **30**(9), 842–848 (Sep 2012). https://doi.org/10.1038/nbt.2345

8. Daraselia, N., Yuryev, A., Egorov, S., Novichkova, S., Nikitin, A., Mazo, I.: Extracting human protein interactions from MEDLINE using a full-sentence parser. Bioinformatics **20**(5), 604–611 (Jan 2004). https://doi.org/10.1093/bioinformatics/btg452, https://doi.org/10.1093%2Fbioinformatics%2Fbtg452

9. Devloo, V., Hansen, P., Labbé, M.: Identification of all steady states in large networks by logical analysis. Bulletin of mathematical biology **65**(6), 1025–1051 (2003). https://doi.org/10.1016/S0092-8240(03)00061-2, https://doi.org/10.1016/S0092-8240(03)00061-2

10. Dubrova, E., Teslenko, M.: A SAT-based algorithm for finding attractors in synchronous boolean networks. IEEE/ACM Transactions on Computational Biology Bioinformatics **8**(5), 1393–1399 (2011). https://doi.org/10.1109/tcbb.2010.20

11. Garg, A., Di Cara, A., Xenarios, I., Mendoza, L., De Micheli, G.: Synchronous versus asynchronous modeling of gene regulatory networks. Bioinformatics **24**(17), 1917–1925 (2008). https://doi.org/10.1093/bioinformatics/btn336, https://doi.org/10.1093/bioinformatics/btn336

12. Goldbeter, A.: Biological rhythms: Clocks for all times. Current Biology **18**(17), R751–R753 (Sep 2008). https://doi.org/10.1016/j.cub.2008.06.044

13. Greil, F., Bassler, K.E.: Attractor period distribution for critical boolean networks (2009), https://arxiv.org/abs/0911.2481

14. Guo, W., Yang, G., Wu, W., He, L., Sun, M.: A parallel attractor finding algorithm based on boolean satisfiability for genetic regulatory networks. PLoS ONE **9**(4), e94258 (2014). https://doi.org/10.1371/journal.pone.0094258

15. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del R'ıo, J.F., Wiebe, M., Peterson, P., G'erard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E.: Array programming with NumPy. Nature **585**(7825), 357–362 (Sep 2020). https://doi.org/10.1038/s41586-020-2649-2, https://doi.org/10.1038/s41586-020-2649-2

16. Hopfensitz, M., Müssel, C., Maucher, M., Kestler, H.A.: Attractors in boolean networks: a tutorial. Computational Statistics **28**(1), 19–36 (Apr 2012). https://doi.org/10.1007/s00180-012-0324-2

17. Huang, S., Ingber, D.E.: Shape-dependent control of cell growth, differentiation, and apoptosis: Switching between attractors in cell regulatory networks. Experimental Cell Research **261**(1), 91–103 (Nov 2000). https://doi.org/10.1006/excr.2000.5044

18. Hunter, J.D.: Matplotlib: A 2d graphics environment. Computing in Science & Engineering **9**(3), 90–95 (2007). https://doi.org/10.1109/MCSE.2007.55, https://doi.org/10.1109/MCSE.2007.55

19. Kamdar, M.R., Stanley, C.E., Carroll, M., Wogulis, L., Dowling, W., Deus, H.F., Samarasinghe, M.: Text snippets to corroborate medical relations: An unsupervised approach using a knowledge graph and embeddings. AMIA Summits on Translational Science Proceedings **2020**, 288 (2020), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7233036/

20. Kauffman, S.: Homeostasis and differentiation in random genetic control networks. Nature **224**(5215), 177–178 (Oct 1969). https://doi.org/10.1038/224177a0

21. Klarner, H., Siebert, H.: Approximating attractors of boolean networks by iterative CTL model checking. Frontiers Bioengineering Biotechnology **3** (2015). https://doi.org/10.3389/fbioe.2015.00130

22. McClung, C.A.: Circadian genes, rhythms and the biology of mood disorders. Pharmacology & Therapeutics **114**(2), 222–232 (May 2007). https://doi.org/10.1016/j.pharmthera.2007.02.003

23. McLachlan, R.I., Cohen, N.L., Dahl, K.D., Bremner, W.J., Soules, M.R.: Serum inhibin levels during the periovulatory interval in normal women: Relationships with sex steroid and gonadotrophin levels. Clinical Endocrinology **32**(1), 39–48 (Jan 1990). https://doi.org/10.1111/j.1365-2265.1990.tb03748.x, https://doi.org/10.1111/j.1365-2265.1990.tb03748.x

24. Meyerowitz, E.A., Vannier, A.G.L., Friesen, M.G.N., Schoenfeld, S., Gelfand, J.A., Callahan, M.V., Kim, A.Y., Reeves, P.M., Poznansky, M.C.: Rethinking the role of hydroxychloroquine in the treatment of COVID-19. The FASEB Journal **34**(5), 6027–6037 (Apr 2020). https://doi.org/10.1096/fj.202000919, https://doi.org/10.1096%2Ffj.202000919

25. Mizera, A., Pang, J., Qu, H., Yuan, Q.: Taming asynchrony for attractor detection in large boolean networks. IEEE/ACM Transactions on Computational Biology and Bioinformatics **16**(1), 31–42 (Jan 2019). https://doi.org/10.1109/tcbb.2018.2850901

26. Morris, M.C., Lyman, C.A., Richman, S., Cao, H.B., Cheadle, C., Broderick, G.: Predicting the immune response to repurposed drugs in coronavirus-induced cytokine storm. In: 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE). IEEE (Oct 2020). https://doi.org/10.1109/bibe50027.2020.00080, https://doi.org/10.1109%2Fbibe50027.2020.00080

27. Nethercote, N., Stuckey, P.J., Becket, R., Brand, S., Duck, G.J., Tack, G.: MiniZinc: Towards a standard CP modelling language. In: Principles and Practice of Constraint Programming – CP 2007, vol. 4741, pp. 529–543. Springer Berlin Heidelberg (2007). https://doi.org/10.1007/978-3-540-74970-7_38, https://doi.org/10.1007/978-3-540-74970-7_38

28. Nikitin, A., Egorov, S., Daraselia, N., Mazo, I.: Pathway studio–the analysis and navigation of molecular networks. Bioinformatics **19**(16), 2155–2157 (Oct 2003). https://doi.org/10.1093/bioinformatics/btg290

29. Novichkova, S., Egorov, S., Daraselia, N.: MedScan, a natural language processing engine for MEDLINE abstracts. Bioinformatics **19**(13), 1699–1706 (2003). https://doi.org/10.1093/bioinformatics/btg207, https://doi.org/10.1093/bioinformatics/btg207

30. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011), http://jmlr.org/papers/v12/pedregosa11a.html

31. Poret, A., Boissel, J.P.: An in silico target identification using boolean network attractors: Avoiding pathological phenotypes. Comptes Rendus Biologies **337**(12), 661–678 (Dec 2014). https://doi.org/10.1016/j.crvi.2014.10.002

32. Qiu, Y., Tamura, T., Ching, W.K., Akutsu, T.: On control of singleton attractors in multiple boolean networks: integer programming-based method. BMC Systems Biology **8**(Suppl 1), S7 (2014). https://doi.org/10.1186/1752-0509-8-s1-s7

33. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2021), https://www.R-project.org/

34. Rabaan, A.A., Al-Ahmed, S.H., Haque, S., Sah, R., Tiwari, R., Malik, Y.S., Dhama, K., Yatoo, M.I., Bonilla-Aldana, D.K., Rodriguez-Morales, A.J.: SARS-CoV-2, SARS-CoV, and MERS-CoV: a comparative overview. Infez Med **28**(2), 174–184 (2020), https://www.infezmed.it/index.php/article?Anno=2020&numero=2&ArticoloDaVisualizzare=Vol_28_2_2020_174

35. Schwab, J.D., Kühlwein, S.D., Ikonomi, N., Kühl, M., Kestler, H.A.: Concepts in boolean network modeling: What do they all mean? Computational and Structural Biotechnology Journal **18**, 571–582 (2020). https://doi.org/10.1016/j.csbj.2020.03.001

36. Sedghamiz, H., Morris, M., Craddock, T.J.A., Whitley, D., Broderick, G.: Bio-ModelChecker: Using bounded constraint satisfaction to seamlessly integrate observed behavior with prior knowledge of biological networks. Frontiers Bioengineering Biotechnology **7** (2019). https://doi.org/10.3389/fbioe.2019.00048, https://doi.org/10.3389/fbioe.2019.00048

37. Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality (complete samples). Biometrika **52**(3-4), 591–611 (Dec 1965). https://doi.org/10.1093/biomet/52.3-4.591, https://doi.org/10.1093/biomet/52.3-4.591

38. Sims, A.C., Tilton, S.C., Menachery, V.D., Gralinski, L.E., Schafer, A., Matzke, M.M., Webb-Robertson, B.J.M., Chang, J., Luna, M.L., Long, C.E., Shukla, A.K., Bankhead, A.R., Burkett, S.E., Zornetzer, G., Tseng, C.T.K., Metz, T.O., Pickles, R., McWeeney, S., Smith, R.D., Katze, M.G., Waters, K.M., Baric, R.S.: Release of severe acute respiratory syndrome coronavirus nuclear import block enhances host transcription in human lung cells. Journal of Virology **87**(7), 3885–3902 (Jan 2013). https://doi.org/10.1128/jvi.02520-12

39. Skodawessely, T., Klemm, K.: Finding attractors in asynchronous boolean dynamics. Advances in Complex Systems **14**(03), 439–449 (2011). https://doi.org/10.1142/s0219525911003098

40. Slowikowski, K.: ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2' (2021), https://CRAN.R-project.org/package=ggrepel, r package version 0.9.1

41. Stuckey, P.J., Feydy, T., Schutt, A., Tack, G., Fischer, J.: The MiniZinc challenge 2008–2013. AI Magazine **35**(2), 55 (Jun 2014). https://doi.org/10.1609/aimag.v35i2.2539, https://doi.org/10.1609/aimag.v35i2.2539

42. Student: The probable error of a mean. Biometrika pp. 1–25 (1908)

43. Tamura, T., Akutsu, T.: Detecting a singleton attractor in a boolean network utilizing SAT algorithms. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences **E92-A**(2), 493–501 (2009). https://doi.org/10.1587/transfun.e92.a.493

44. Thomas, R., d'Ari, R.: Biological feedback. CRC press, Boca Raton, FL, USA (1990)

45. Veliz-Cuba, A., Aguilar, B., Hinkelmann, F., Laubenbacher, R.: Steady state analysis of boolean molecular network models via model reduction and computational algebra. BMC Bioinformatics **15**(1) (2014). https://doi.org/10.1186/1471-2105-15-221

46. Wickham, H.: ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York (2016), https://ggplot2.tidyverse.org

47. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. Chemometrics and Intelligent Laboratory Systems **2**(1-3), 37–52 (Aug 1987). https://doi.org/10.1016/0169-7439(87)80084-9, https://doi.org/10.1016/0169-7439(87)80084-9

48. Yuan, Q., Mizera, A., Pang, J., Qu, H.: A new decomposition-based method for detecting attractors in synchronous boolean networks. Science of Computer Programming **180**, 18–35 (Jul 2019). https://doi.org/10.1016/j.scico.2019.05.001

49. Yuan, Q., Qu, H., Pang, J., Mizera, A.: Improving BDD-based attractor detection for synchronous boolean networks. Sci. China Information Sci. **59**(8) (2016). https://doi.org/10.1007/s11432-016-5594-9

50. Zhao, Q.: A remark on "Scalar equations for synchronous Boolean networks with biological Applications" by C. Farrow, J. Heidel, J. Maloney, and J. Rogers. IEEE Transactions on Neural Networks **16**(6), 1715–1716 (2005). https://doi.org/10.1109/tnn.2005.857944
51. Zheng, D., Yang, G., Li, X., Wang, Z., Liu, F., He, L.: An efficient algorithm for computing attractors of synchronous and asynchronous boolean networks. PLoS ONE **8**(4), e60593 (Apr 2013). https://doi.org/10.1371/journal.pone.0060593