Bayesian design of synthetic biological systems : Supplementary Information

Chris Barnes, Daniel Silk, Xia Sheng and Michael P.H. Stumpf

Contents

1 Approximate Bayesian Computation : ABC SMC

Here we outline the background behind approximate Bayesian computation (ABC) and describe the ABC SMC algorithm [1], which is implemented in the software package ABC-SysBio [2]. ABC methods have been developed to infer posterior distributions in cases where likelihood functions are computationally intractable or too costly to evaluate. They replace the calculation of the likelihood with a comparison between observed and simulated data.

1.1 Background

Following the notation in [3], let $\theta \in \Theta$ be a parameter vector with prior $\pi(\theta)$ and $f(y|\theta)$ be the likelihood of the data $y \in \mathcal{D}$. In Bayesian inference we are interested in the posterior density

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{\int_\Theta f(y|\theta)\pi(\theta)d\theta}.$$

Now imagine the case where we cannot write down the likelihood in closed form but we can simulate from the data generating model. We can proceed by first sampling a parameter vector from the prior, $\theta^* \sim \pi(\theta)$, and then sampling a data vector, x^*, from the model conditional on θ^*, ie $x^* \sim f(x|\theta^*)$. This alone gives the joint density $\pi(\theta, x)$. To obtain samples from the posterior distribution we must condition on the data y and this is done via an indicator function, i.e.

$$\pi(\theta, x|y) = \frac{\pi(\theta)f(x|\theta)\mathbb{I}_{\mathcal{A}_y}(x)}{\int_{\mathcal{A}_y \times \Theta} \pi(\theta)f(x|\theta)dxd\theta},$$

where $\mathbb{I}_{\mathcal{B}}(z)$ denotes the indicator function and is equal to 1 for $z \in \mathcal{B}$. Here $\mathcal{A}_y = \{x \in \mathcal{D} : x = y\}$, so the indicator is equal to one when the simulated data and the observed data are identical. This forms a rejection algorithm, and in this instance the accepted θ^* are from the true posterior density $\pi(\theta|y)$.

For most models it is impossible to achieve simulations with outputs in the subset $\mathcal{A}_y$ and so an approximation must be made. This is the basis for ABC. In the first instance we can replace $\mathcal{A}_y$ by $\mathcal{A}_{y,\epsilon} = \{x \in \mathcal{D} : \rho(x, y) \leq \epsilon\}$ where $\rho : \mathcal{D} \times \mathcal{D} \to \mathbb{R}^+$ is a distance function comparing the simulated data to the observed data. We then have

$$\pi_\epsilon(\theta, x|y) = \frac{\pi(\theta)f(x|\theta)\mathbb{I}_{\mathcal{A}_{y,\epsilon}}(x)}{\int_{\mathcal{A}_{y,\epsilon} \times \Theta} \pi(\theta)f(x|\theta)dxd\theta},$$

where π_ϵ is an approximation to the true posterior distribution. The rationale behind ABC is that if ϵ is small then the resulting approximate posterior, π_ϵ, is close to the true posterior. Often, for complex models or stochastic systems, the subset $\mathcal{A}_{y,\epsilon}$ is still too restrictive. In these cases we can resort to comparisons of summary statistics. We now specify the subset $\mathcal{A}_{y,\eta,\epsilon} = \{x \in \mathcal{D} : \rho_S(x, y) \leq \epsilon\}$ where $\eta : \mathcal{D} \to \mathcal{S}$ is a summary statistic and the distance function now takes the form $\rho_S : \mathcal{S} \times \mathcal{S} \to \mathbb{R}^+$. We often write the marginal posterior distribution as $\pi(\theta|\rho(x^*, y) \leq \epsilon)$.

1.2 ABC SMC

The simplest ABC algorithm is known as the ABC rejection algorithm [4] and proceeds as follows

R1 Sample θ^* from $\pi(\theta)$.
R2 Simulate a dataset x^* from $f(x|\theta^*)$.
R3 If $\rho(x^*, y) \leq \epsilon$ accept θ^*, otherwise reject.
R4 Return to R1.

This gives draws from π_ϵ but can be very inefficient in high dimensional models or when the overlap between the prior and posterior distributions is small. One way to improve the efficiency of the rejection algorithm is to perform sequential importance sampling (SIS) [5]. In SIS, instead of sampling directly from the posterior distribution, sampling proceeds via a series of intermediate distributions. The importance distribution at each stage is constructed from a perturbed version of the previous population. This approach can be used in ABC and the resultant algorithm is known as ABC SMC [1]. Described here is a slightly modified version that automatically calculates the ϵ schedule and as such, only the final value, ϵ_T, needs be specified. To obtain N samples $\{\theta^1, \theta^2, \theta^3 ...., \theta^N\}$ (known as particles) from the posterior, defined as, $\pi(\theta|\rho(x^*, y) \leq \epsilon_T)$, proceed as follows

S1 Initialize $\epsilon = \infty$
Set the population indicator $t = 0$
S2.0 Set the particle indicator $i = 1$
S2.1 If $t = 0$, sample θ^{**} independently from $\pi(\theta)$
If $t > 0$, sample θ^* from the previous population $\{\theta^i_{t-1}\}$ with weights w_{t-1}.
Perturb the particle, $\theta^{**} \sim K_t(\theta|\theta^*)$ where K_t is the perturbation kernel.
If $\pi(\theta^{**}) = 0$, return to S2.1
Simulate a candidate dataset $x^* \sim f(x|\theta^{**})$.
If $\rho(x^*, y) > \epsilon$ return to S2.1
S2.2 Set $\theta^i_t = \theta^{**}$ and $d^i_t = \rho(x^*, y)$, calculate the weight as

$$w^i_t = \begin{cases} 1 & \text{if } t = 0 \\ \frac{\pi(\theta^i_t)}{\sum_{j=1}^N w^j_{t-1} K_t(\theta^i_t|\theta^j_{t-1})} & \text{if } t > 0 \end{cases}$$

If $i < N$, set $i = i + 1$, go to S2.1
S3 Normalize the weights.
Determine ϵ such that $Pr(d_t \leq \epsilon) = 0.9$.
If $\epsilon > \epsilon_T$, set $t = t + 1$, go to S2.0.

Here $K_t(\theta|\theta^*)$ is the component-wise random walk perturbation kernel that, in this study, takes the form $K_t(\theta^*|\theta) = \theta + U(-\delta, \delta)$ where $\delta = \frac{1}{2}\text{range}\{\theta_{t-1}\}$. The denominator in the weight calculation can be seen as the probability of observing the current particle given the previous population.

1.3 Model selection

In Bayesian inference comparison of a discrete set of models can be be performed using the marginal posterior. Consider the joint space defined by $(M, \theta) \in \mathcal{M} \times \Theta_{\mathcal{M}}$; Bayes theorem can then be written

$$\pi(M|y) = \frac{f(y|M)\pi(M)}{\int_{\mathcal{M}} f(y|M')\pi(M')dM'} = \frac{f(y|M)\pi(M)}{\sum_{\mathcal{M}} f(y|M')\pi(M')},$$

where $f(y|M)$, the marginal likelihood, can be written

$$f(y|M) = \int_{\Theta_{\mathcal{M}}} \pi(\theta|M) f(y|\theta, M) d\theta.$$

Therefore the posterior probability of a model is given by the normalized marginal likelihood which may or may not be weighted depending on whether the prior over models is informative or uniform respectively. It has recently been noted that model selection using summary statistics can be problematic because the summary statistic must be sufficient for the joint space, $\{M, \theta\}$, rather than just θ [6]. This is not a concern here since in all our examples we use the full data set with no summary or we *define* our posterior distributions through the summary statistics.

Model selection can be incorporated into the ABC framework by introducing the model indicator M and proceeding with inference on the joint space. For example, the ABC rejection algorithm with model selection [7] proceeds as follows

MR1 Sample M^* from $\pi(M)$.
MR2 Sample θ^* from $\pi(\theta|M^*)$.
MR3 Simulate a dataset x^* from $f(x|\theta^*, M^*)$.
MR4 If $\rho(x^*, y) \leq \epsilon$ accept (M^*, θ^*), otherwise reject.
MR5 Return to R1.

Once N samples have been accepted an approximation to the marginal posterior, $\pi(M = m|y)$, is given by

$$\pi(M = m|y) = \frac{\#\text{accepted m}}{N}.$$

Model selection can also be incorporated into the ABC SMC algorithm [8]. To obtain N samples $\{(M, \theta)^1, (M, \theta)^2, (M, \theta)^3 ...., (M, \theta)^N\}$ from the posterior, defined as, $\pi(M, \theta|\rho(x^*, y) \leq \epsilon_T)$, proceed as follows

MS1 Initialize $\epsilon = \infty$
Set the population indicator $t = 0$
MS2.0 Set the particle indicator $i = 1$
MS2.1 If $t = 0$, sample (M^{**}, θ^{**}) from the prior $\pi(M, \theta) = \pi(M)\pi(\theta|M)$.
If $t > 0$, sample M^* with probability $P_{t-1}(M^*)$ and perturb $M^{**} \sim KM_t(M|M^*)$.
Sample θ^* from the previous population $\{\theta(M^{**})_{t-1}\}$ with weights w_{t-1}.
Perturb the particle, $\theta^{**} \sim K_{t,M^{**}}(\theta|\theta^*)$ where $K_{t,M}$ is the perturbation kernel.
If $\pi(M^{**}, \theta^{**}) = 0$, return to MS2.1
Simulate a candidate dataset $x^* \sim f(x|M^{**}, \theta^{**})$.
If $\rho(x^*, y) > \epsilon$ return to MS2.1
MS2.2 Set $(M, \theta)_t^i = (M^{**}, \theta^{**})$ and $d_t^i = \rho(x^*, y)$, calculate the weight as

$$w_t^i(M_t^i, \theta_t^i) = \begin{cases} 1 & \text{if } t = 0 \\ \frac{\pi(M_t^i, \theta_t^i)}{S_1 S_2} & \text{if } t > 0 \end{cases}$$

where
$S_1 = \sum_{j \in \mathcal{M}} P_{t-1}(M_{t-1}^j) KM_t(M_t^i|M_{t-1}^j)$
and

$$S_2 = \sum_{k \in M_t^i = M_{t-1}} \frac{w_{t-1}^k K_{t,M^i}(\theta_t^i|\theta_{t-1}^k)}{P_{t-1}(M_t^i = M_{t-1})}$$

If $i < N$, set $i = i + 1$, go to MS2.1
S3 Normalize the weights.
Obtain the marginal model probabilities given by

$$P_t(M_t = m) = \sum_{k \in M_t^i = M_{t-1}} w_t^i(M_t^i, \theta_t^i)$$

Determine ϵ such that $Pr(d_t \leq \epsilon) = 0.9$.
If $\epsilon > \epsilon_T$, set $t = t + 1$, go to MS2.0.

There are two obvious additions to the algorithm when compared to parameter inference. The model kernel, KM_t, perturbs the resampled models using a multinomial distribution, and the additional term in the weight denominator accounts for the probability of observing the current model given the previous population.

2 Prior distribution

The prior distribution encodes our knowledge of the system and should be set according to known biochemical properties. However, often the kinetic parameters are not well known and can be very difficult or even impossible to measure *in vivo*. In these cases we make the prior distribution non informative by specifying a large range over possible, biophysically and biochemically plausible values. As more information becomes available, through experimental studies or otherwise, the prior can be updated to reflect our increased knowledge of the system. Interestingly, for some systems, our design method could help to constrain kinematic parameters where experimental data are unavailable.

3 The distance function and output tolerance

In system design we would rarely insist on achieving the true posterior distribution corresponding to $\epsilon = 0$, but would like to reach the objective within some tolerance. A theorem due to Wilkinson (2008) [9] states that if we assume that the data can be considered as

$$y = \eta(\hat{\theta}) + e,$$

where $\eta(\hat{\theta})$ is a draw from the model at the 'best' input and e is an additive, independent error, then the approximate posterior distribution, $\pi(\theta|\rho(x^*, y) \leq \epsilon)$ can be interpreted as the 'true' posterior $\pi(\hat{\theta}|y)$. While the independence assumption is not always true, this theorem provides some insight into the relationship between

the final ϵ value and the tolerance on our specified behavior. For example when using uniform kernels, as in this study, if our desired output behavior is a constant of 0.5 and we finish the inference at $\epsilon = 0.05$ our final trajectories will be distributed $U(0.45, 0.55)$ giving a tolerance of $\pm 10\%$ on the output behavior. This can be used when considering our desired output objectives. To achieve other error distributions, such as Gaussian errors, we can always explicitly specify the error model in the design objectives.

4 Deterministic models

Inference for deterministic models such as ordinary differential equations can be problematic since there is a one to one relationship between the parameter vector θ and the data set x. Therefore, in the absence of observational error, the posterior distribution resembles a delta function, $\delta(\theta - \hat{\theta})$ where $\hat{x} = f(\hat{\theta})$ is data 'closest' to y. An additional problem for ABC methods is that the minimum distance, $\rho(\hat{x}, y)$, is greater than zero [10]. However, in practice, observational data have associated experimental errors and when this is included explicitly in the model, the problem is resolved. In the case of systems design, we omit the explicit error model for clarity, but note that it could be included with assumptions on the form of the distribution.

5 Biochemical adaptation

5.1 Models

We used the same models as those used in [11], which are enzymatic reactions assuming Michaelis-Menten kinetics. Below we give the full models including cooperativity but the more specific case of no cooperativity is when the exponents, n_i, are set to one. Here A, B, C denote the concentrations of the active form of the species and $(1-A), (1-B), (1-C)$ the concentrations of the inactive form. Species E_i and F_i refer to background activating and deactivating enzymes respectively and are assumed to have a constant concentration of 0.5. The models were simulated in the range $0 \leq t \leq 200$.

Design 1

$$\begin{aligned}
\frac{dA}{dt} &= Ik_{IA}\frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}} + K_{IA}^{n_{IA}}} - F_A k_{FA}\frac{A^{n_{FA}}}{A^{n_{FA}} + K_{FA}^{n_{FA}}} \\
\frac{dB}{dt} &= Ck_{CB}\frac{(1-B)^{n_{CB}}}{(1-B)^{n_{CB}} + K_{CB}^{n_{CB}}} - F_B k_{FB}\frac{B^{n_{FB}}}{B^{n_{FB}} + K_{FB}^{n_{FB}}} \\
\frac{dC}{dt} &= Ak_{AC}\frac{(1-C)^{n_{AC}}}{(1-C)^{n_{AC}} + K_{AC}^{n_{AC}}} - Bk_{BC}\frac{C^{n_{BC}}}{C^{n_{BC}} + K_{BC}^{n_{BC}}}
\end{aligned}$$

Design 2

$$\begin{aligned}
\frac{dA}{dt} &= Ik_{IA}\frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}} + K_{IA}^{n_{IA}}} - F_A k_{FA}\frac{A^{n_{FA}}}{A^{n_{FA}} + K_{FA}^{n_{FA}}} \\
\frac{dB}{dt} &= E_B k_{EB}\frac{(1-B)^{n_{EB}}}{(1-B)^{n_{EB}} + K_{EB}^{n_{EB}}} - Ck_{CB}\frac{B^{n_{CB}}}{B^{n_{CB}} + K_{CB}^{n_{CB}}} \\
\frac{dC}{dt} &= Ak_{AC}\frac{(1-C)^{n_{AC}}}{(1-C)^{n_{AC}} + K_{AC}^{n_{AC}}} - Bk_{BC}\frac{C^{n_{BC}}}{C^{n_{BC}} + K_{BC}^{n_{BC}}} - F_C k_{FC}\frac{C^{n_{FC}}}{C^{n_{FC}} + K_{FC}^{n_{FC}}}
\end{aligned}$$

Design 3

$$\begin{aligned}
\frac{dA}{dt} &= Ik_{IA}\frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}} + K_{IA}^{n_{IA}}} - F_A k_{FA}\frac{A^{n_{FA}}}{A^{n_{FA}} + K_{FA}^{n_{FA}}} \\
\frac{dB}{dt} &= E_B k_{EB}\frac{(1-B)^{n_{EB}}}{(1-B)^{n_{EB}} + K_{EB}^{n_{EB}}} - Ck_{CB}\frac{B^{n_{CB}}}{B^{n_{CB}} + K_{CB}^{n_{CB}}} \\
\frac{dC}{dt} &= Bk_{BC}\frac{(1-C)^{n_{BC}}}{(1-C)^{n_{BC}} + K_{BC}^{n_{BC}}} - Ak_{AC}\frac{C^{n_{AC}}}{C^{n_{AC}} + K_{AC}^{n_{AC}}}
\end{aligned}$$

Design 4

$$\frac{dA}{dt} = Ik_{IA}\frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}} + K_{IA}^{n_{IA}}} - Bk_{BA}\frac{A^{n_{BA}}}{A^{n_{BA}} + K_{BA}^{n_{BA}}}$$
$$\frac{dB}{dt} = Ak_{AB}\frac{(1-B)^{n_{AB}}}{(1-B)^{n_{AB}} + K_{AB}^{n_{AB}}} - F_Bk_{FB}\frac{B^{n_{FB}}}{B^{n_{FB}} + K_{FB}^{n_{FB}}}$$
$$\frac{dC}{dt} = Ak_{AC}\frac{(1-C)^{n_{AC}}}{(1-C)^{n_{AC}} + K_{AC}^{n_{AC}}} - F_Ck_{FC}\frac{C^{n_{FC}}}{C^{n_{FC}} + K_{FC}^{n_{FC}}}$$

Design 5

$$\frac{dA}{dt} = Ik_{IA}\frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}} + K_{IA}^{n_{IA}}} - Bk_{BA}\frac{A^{n_{BA}}}{A^{n_{BA}} + K_{BA}^{n_{BA}}}$$
$$\frac{dB}{dt} = Ak_{AB}\frac{(1-B)^{n_{AB}}}{(1-B)^{n_{AB}} + K_{AB}^{n_{AB}}} - F_Bk_{FB}\frac{B^{n_{FB}}}{B^{n_{FB}} + K_{FB}^{n_{FB}}}$$
$$\frac{dC}{dt} = E_Ck_{EC}\frac{(1-C)^{n_{EC}}}{(1-C)^{n_{EC}} + K_{EC}^{n_{EC}}} - Ak_{AC}\frac{C^{n_{AC}}}{C^{n_{AC}} + K_{AC}^{n_{AC}}}$$

Design 6

$$\frac{dA}{dt} = Ik_{IA}\frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}} + K_{IA}^{n_{IA}}} - Bk_{BA}\frac{A^{n_{BA}}}{A^{n_{BA}} + K_{BA}^{n_{BA}}}$$
$$\frac{dB}{dt} = Ck_{CB}\frac{(1-B)^{n_{CB}}}{(1-B)^{n_{CB}} + K_{CB}^{n_{CB}}} - F_Bk_{FB}\frac{B^{n_{FB}}}{B^{n_{FB}} + K_{FB}^{n_{FB}}}$$
$$\frac{dC}{dt} = Ak_{AC}\frac{(1-C)^{n_{AC}}}{(1-C)^{n_{AC}} + K_{AC}^{n_{AC}}} - F_Ck_{FC}\frac{C^{n_{FC}}}{C^{n_{FC}} + K_{FC}^{n_{FC}}}$$

Design 7

$$\frac{dA}{dt} = Ik_{IA}\frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}} + K_{IA}^{n_{IA}}} - Bk_{BA}\frac{A^{n_{BA}}}{A^{n_{BA}} + K_{BA}^{n_{BA}}} - F_Ak_{FA}\frac{A^{n_{FA}}}{A^{n_{FA}} + K_{FA}^{n_{FA}}}$$
$$\frac{dB}{dt} = E_Bk_{EB}\frac{(1-B)^{n_{EB}}}{(1-B)^{n_{EB}} + K_{EB}^{n_{EB}}} - Ck_{CB}\frac{B^{n_{CB}}}{B^{n_{CB}} + K_{CB}^{n_{CB}}}$$
$$\frac{dC}{dt} = Ak_{AC}\frac{(1-C)^{n_{AC}}}{(1-C)^{n_{AC}} + K_{AC}^{n_{AC}}} - F_Ck_{FC}\frac{C^{n_{FC}}}{C^{n_{FC}} + K_{FC}^{n_{FC}}}$$

Design 8

$$\frac{dA}{dt} = Ik_{IA}\frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}} + K_{IA}^{n_{IA}}} - Bk_{BA}\frac{A^{n_{BA}}}{A^{n_{BA}} + K_{BA}^{n_{BA}}} - F_Ak_{FA}\frac{A^{n_{FA}}}{A^{n_{FA}} + K_{FA}^{n_{FA}}}$$
$$\frac{dB}{dt} = Ck_{CB}\frac{(1-B)^{n_{CB}}}{(1-B)^{n_{CB}} + K_{CB}^{n_{CB}}} - F_Bk_{FB}\frac{B^{n_{FB}}}{B^{n_{FB}} + K_{FB}^{n_{FB}}}$$
$$\frac{dC}{dt} = E_Ck_{EC}\frac{(1-C)^{n_{EC}}}{(1-C)^{n_{EC}} + K_{EC}^{n_{EC}}} - Ak_{AC}\frac{C^{n_{AC}}}{C^{n_{AC}} + K_{AC}^{n_{AC}}}$$

Design 9

$$\frac{dA}{dt} = Ik_{IA}\frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}} + K_{IA}^{n_{IA}}} - Bk_{BA}\frac{A^{n_{BA}}}{A^{n_{BA}} + K_{BA}^{n_{BA}}}$$
$$\frac{dB}{dt} = E_Bk_{EB}\frac{(1-B)^{n_{EB}}}{(1-B)^{n_{EB}} + K_{EB}^{n_{EB}}} - Ck_{CB}\frac{B^{n_{CB}}}{B^{n_{CB}} + K_{CB}^{n_{CB}}}$$
$$\frac{dC}{dt} = E_Ck_{EC}\frac{(1-C)^{n_{EC}}}{(1-C)^{n_{EC}} + K_{EC}^{n_{EC}}} - Ak_{AC}\frac{C^{n_{AC}}}{C^{n_{AC}} + K_{AC}^{n_{AC}}}$$

Design 10

$$\begin{aligned}
\frac{dA}{dt} &= Ik_{IA}\frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}}+K_{IA}^{n_{IA}}} - F_A k_{FA}\frac{A^{n_{FA}}}{A^{n_{FA}}+K_{FA}^{n_{FA}}} \\
\frac{dB}{dt} &= Ak_{AB}\frac{(1-B)^{n_{AB}}}{(1-B)^{n_{AB}}+K_{AB}^{n_{AB}}} - F_B k_{FB}\frac{B^{n_{FB}}}{B^{n_{FB}}+K_{FB}^{n_{FB}}} \\
\frac{dC}{dt} &= Bk_{BC}\frac{(1-C)^{n_{BC}}}{(1-C)^{n_{BC}}+K_{BC}^{n_{BC}}} - Ak_{AC}\frac{C^{n_{AC}}}{C^{n_{AC}}+K_{AC}^{n_{AC}}}
\end{aligned}$$

Design 11

$$\begin{aligned}
\frac{dA}{dt} &= Ik_{IA}\frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}}+K_{IA}^{n_{IA}}} - F_A k_{FA}\frac{A^{n_{FA}}}{A^{n_{FA}}+K_{FA}^{n_{FA}}} \\
\frac{dB}{dt} &= Ak_{AB}\frac{(1-B)^{n_{AB}}}{(1-B)^{n_{AB}}+K_{AB}^{n_{AB}}} - F_B k_{FB}\frac{B^{n_{FB}}}{B^{n_{FB}}+K_{FB}^{n_{FB}}} \\
\frac{dC}{dt} &= Ak_{AC}\frac{(1-C)^{n_{AC}}}{(1-C)^{n_{AC}}+K_{AC}^{n_{AC}}} - Bk_{BC}\frac{C^{n_{BC}}}{C^{n_{BC}}+K_{BC}^{n_{BC}}}
\end{aligned}$$

5.2 Distance

The two component distance metric was defined to be $\rho(x, \mathcal{O}) = \{E, S^{-1}\}$, where E and S are the adaptation efficiency and sensitivity defined by

$$\begin{aligned}
E &= \left|\frac{(O_2 - O_1)/O_1}{(I_2 - I_1)/I_1}\right| \\
S &= \left|\frac{(O_{peak} - O_1)/O_1}{(I_2 - I_1)/I_1}\right|,
\end{aligned}$$

where I_1, I_2 are the input values (here fixed at 0.5 and 0.6 respectively), O_1, O_2 are the output steady state levels before and after the input change and and O_{peak} is the maximal transient output level. The final population was defined to be $\epsilon = \{0.1, 1.0\}$.

5.3 Priors

The priors on the Michaelis-Menten rates were chosen to correspond to the parameter ranges used in the original study; $\log k \sim U(-1, 1)$ and $\log K \sim U(-3, 2)$ [11].

6 Bacterial two component systems

6.1 Models

The models we used were based on the ones found in [12]. All simulations were performed in the range $0 \leq t \leq 10$.
Orthodox system
We modeled the following reactions

$$\begin{aligned}
HK + S &\xrightarrow{k_1} HKp + S \\
HKp + RR &\xrightarrow{k_2} HK + RRp \\
HKp &\xrightarrow{k_3} HK \\
HK + RRp &\xrightarrow{k_4} HKp + RR \\
RRp &\xrightarrow{k_5} RR.
\end{aligned}$$

Additionally we assumed that the total concentration of $HK_{tot} = HK + HKp$ and $RR_{tot} = RR + RRp$ were equal to one. This resulted in the following ordinary differential equations

$$\frac{d[HK]}{dt} = k_2[HKp][RR] + k_3[HKp] - k_4[HK][RRp] - k_1[HK][S]$$
$$\frac{d[RRp]}{dt} = k_2[HKp][RR] - k_4[HK][RRp] - k_5[RRp].$$

Orthodox system
We labelled the occupied states of the phosphorelay as

	H1	D1	H2
HK_1	x	x	x
HK_2	o	x	x
HK_3	x	o	x
HK_4	x	x	o
HK_5	o	o	x
HK_6	o	x	o
HK_7	x	o	o
HK_8	o	o	o

where H1, D1 and H2 are the binding domains on the Histidine Kinase and x, o represent an empty, occupied domain respectively. We modeled the following reactions

$$HK + S \xrightarrow{k_1} HKp + S$$

Again we assumed that the total concentration of $HK_{tot} = \sum HK_i$ and $RR_{tot} = RR + RRp$ were equal to one. This resulted in the following ordinary differential equations

$$\frac{dHK_1}{dt} = k_4[HK_4][RR] + k_6[HK_3] - k_7[HK_1][RRp] + k_8[HK_2] - k1[HK_1][S]$$
$$\frac{dHK_2}{dt} = k_4[HK_6][RR] + k_6[HK_5] - k_7[HK_2][RRp] - k_8[HK_2] + k_1[HK_1][S] - k_2[HK_2]$$
$$\frac{dHK_3}{dt} = -k_3[HK_3] + k_4[HK_7][RR] + k_5[HK_4] - k_6[HK_3] - k_7[HK_3][RRp] + k_8[HK_5] - k_1[HK_3][S] + k_2[HK_2]$$
$$\frac{dHK_4}{dt} = k_3[HK_3] - k_4[HK_4][RR] - k_5[HK_4] + k_6[HK_7] + k_7[HK_1][RRp] + k_8[HK_6] - k_1[HK_4][S]$$
$$\frac{dHK_5}{dt} = -k_3[HK_3] + k_4[HK_8][RR] + k_5[HK_6] - k_6[HK_5] - k_7[HK_5][RRp] - k_8[HK_5] + k_1[HK_3][S]$$
$$\frac{dHK_6}{dt} = k_3[HK_5] - k_2[HK_6] - k_4[HK_6][RR] - k_5[HK_6] + k_6[HK_8] + k_7[HK_2][RRp] - k_8[HK_6] + k_1[HK_4][S]$$
$$\frac{dHK_7}{dt} = k_2[HK_6] - k_4[HK_7][RR] - k_6[HK_7] + k_7[HK_3][RRp] + k_8[HK_8] - k_1[HK_7][S]$$
$$\frac{dRRp}{dt} = k_4[RR]([HK_4] + [HK_6] + [HK_7] + [HK_8]) - k_7[RRp]([HK_1] + [HK_2] + [HK_3] + [HK_5]) - k_9[RRp]$$

6.2 Distance

The distance functions for input-output behaviors $\rho(x, \mathcal{O})_{1-4}$ were defined to be

$$\begin{aligned}
\rho(x,\mathcal{O})_1 &= \left\{\mathbb{H}(0)(\text{argmax}_t\, x_t - 2.0), \mathbb{H}(0)(\text{argmin}_t\, x_t - 4.0)\right\} \\
\rho(x,\mathcal{O})_2 &= \sqrt{\sum_t (x_t - 1.0)^2} \\
\rho(x,\mathcal{O})_3 &= \sqrt{\sum_t (x_t - 0.5)^2} \\
\rho(x,\mathcal{O})_4 &= \left\{\rho(x,\mathcal{O})_1, \mathbb{H}(0)(1 - \max x_t - 0.2), \mathbb{H}(0)(\min x_t - 0.2)\right\}
\end{aligned}$$

where $\mathbb{H}(0)$ is the Heaviside function ensuring the distance is positive.

6.3 Priors

The priors on all variables were distributed as $U(0, 1000)$.

7 Stochastic genetic toggle switch

7.1 Models

We modeled each toggle switch using a continuous time Markov jump process which obeys the chemical master equation. We neglected processes at the RNA level and just modeled at the protein level. This makes the models simpler while retaining all the relevant behavior. In all the following gA, gB represent the gene promotor for protein A, B respectively and are fixed at one copy. $A - gB, B - gA$ represent the bound transcription factors and S, R represent the switch and reset signals. Because the concentration of these are fixed they have the effect of removing A and B from the system respectively. The models were simulated in the range $0 \leq t \leq 200$

Design 1

$$
\begin{aligned}
gA &\xrightarrow{k_1} gA + A \\
A &\xrightarrow{k_2} \emptyset \\
gB &\xrightarrow{k_3} gB + B \\
B &\xrightarrow{k_4} \emptyset \\
A + gB &\xrightarrow{k_5} A - gB \\
B + gA &\xrightarrow{k_6} B - gA \\
A - gB &\xrightarrow{k_7} A + gB \\
B - gA &\xrightarrow{k_8} B + gA \\
S + A &\xrightarrow{k_9} S - A \\
R + B &\xrightarrow{k_{10}} R - B
\end{aligned}
$$

Design 2

Here, in addition to the species in design 1, we have introduced the P species, fixed to be one copy, which ensures only A or B can be bound at any one time

$$
\begin{aligned}
gA &\xrightarrow{k_1} gA + A \\
A &\xrightarrow{k_2} \emptyset \\
gB &\xrightarrow{k_3} gB + B \\
B &\xrightarrow{k_4} \emptyset \\
A + gB + P &\xrightarrow{k_5} A - gB \\
B + gA + P &\xrightarrow{k_6} B - gA \\
A - gB &\xrightarrow{k_7} A + gB + P \\
B - gA &\xrightarrow{k_8} B + gA + P \\
S + A &\xrightarrow{k_9} S - A \\
R + B &\xrightarrow{k_{10}} R - B.
\end{aligned}
$$

Design 3
Here we have the same reactions in design 1 but include two extra reactions for the decay of the bound proteins

$$A - gB \xrightarrow{k_{11}} gB$$
$$B - gA \xrightarrow{k_{12}} gA.$$

Design 4
Here we have the same reactions in design 1 but include two extra reactions for the binding /unbinding of the proteins A and B

$$A + B \xrightarrow{k_{11}} A - B$$
$$A - B \xrightarrow{k_{12}} A + B.$$

7.2 Distance

The two component distance metric was defined to be $\rho(x, \mathcal{O}) = \{d_1, d_2\}$,

$$\begin{aligned} d_1 &= \sqrt{\frac{\sum_{t\in\alpha}(x_t - y)^2}{n_\alpha}} \\ d_2 &= \sqrt{\frac{\sum_{t\in\beta}(x_t - 0)^2}{n_\beta}}, \end{aligned}$$

where x_t is the number of protein B at time t, y is the target (here fixed at 40), $\alpha = \{t : 30 < t \leq 60\}$, $\beta = \{t : 0 < t \leq 10 \text{ and } 80 < t \leq 100\}$, $n_\alpha = \#\alpha$ and $n_\beta = \#\beta$. The final population was defined to be $\epsilon = \{7.0, 0.05\}$.

7.3 Priors

The priors for production, binding and interaction rates were distributed as $U(0, 50)$ and the priors for the degradation rates were given $U(0, 5)$ distributions.

8 Robust oscillator design

8.1 Analysis

Biochemical oscillations are increasingly being implemented in various synthetic systems [13, 14, 15, 16]. A recent study by Tsai *et al* [17] compared the ability of five small networks to achieve oscillations. The five designs are shown in Figure 10A where each node represents the active form of a protein, edges represent enzymatic reactions and thicker edges represent increased feedback strength. We applied our Bayesian design methodology to the original problem and further investigated the ability of these designs to achieve particular amplitude-frequency values.

Figure 10B shows the posterior probability for each model to achieve limit cycle behavior induced by a Hopf bifurcation. The addition of the negative feedback loop in models 4 and 5 does not improve the ability to achieve oscillations. We find that the addition of a positive feedback loop on species A in models 2 and 3 increases the ability of the system to achieve limit cycle behavior, but no significant increase in the posterior probability is provided by increasing the feedback strength. This is in conflict with the original study that found that model 3 outperformed model 2 [17]. Our approach does sample parameter space predominantly in regions where the desired behavior is more likely, rather than entirely at random as was done in the previous study; on balance this suggests that the posterior probability for delivering robust oscillations is approximately the same for models 2 and 3.

More insight can be gained into this discrepancy by specifying a particular frequency and amplitude of the oscillator as the desired output behavior. Figures 10C and D show the model posterior probability after requiring an amplitude of 0.1 and a frequency of 1.0 Hz on species A and C respectively. The first thing to

note is that the model posterior is significantly different in the two cases. When the constraints are applied to species A, model 3 is favored with the increase in feedback strength *decreasing* the ability to reach the specified behavior. When the conditions are applied to species C (and species B by symmetry) we get a posterior that more resembles the original findings; that the increase in feedback strength does indeed increase the ability to reach the specified oscillations. Thus the posterior for the Hopf bifurcation behavior represents a sum over all possible oscillator characteristics; in a manner that is reminiscent of Bayesian model averaging.

If we examine the posterior distribution and the principal component analysis for model 2 to achieve Hopf bifurcations (Figures S5 and S6), we can see that the parameters k_1 and k_3, which are the strengths of the deactivating reactions on nodes A and B, are constrained to be similar in magnitude to k_5. We also see that within this model the feedback strength, k_7, does not affect the dynamics significantly. Here, and elsewhere, we can use the posterior distributions in order to gain insights into the sensitivity and robustness of the system to variations in parameters, irrespective of whether the system's dynamics are deterministic or stochastic: our ABC SMC framework allows us to extract such information on the fly as part of the sequential design process.

8.2 Models

We used the same models as those used in [17], simulated in the range $0 \leq t \leq 10$. Again A, B, C denote the concentrations of the active form of the species and $(1-A), (1-B), (1-C)$ the concentrations of the inactive form. The feedback is modeled using Michaelis-Menten kinetics but the conversion of inactive form into active form is assumed to have a constant rate.

Design 1

$$\begin{aligned}
\frac{dA}{dt} &= k_1(1-A) - \frac{k_2 C^{n_1}}{K_1^{n_1} + C^{n_1}} A \\
\frac{dB}{dt} &= k_3(1-B) - \frac{k_4 A^{n_2}}{K_2^{n_2} + A^{n_2}} B \\
\frac{dC}{dt} &= k_5(1-C) - \frac{k_6 B^{n_3}}{K_3^{n_3} + B^{n_3}} C
\end{aligned}$$

Designs 2 and 3

$$\begin{aligned}
\frac{dA}{dt} &= k_1(1-A) - \frac{k_2 C^{n_1}}{K_1^{n_1} + C^{n_1}} A + k_7(1-A)\frac{A^{n_4}}{K_4^{n_4} + A^{n_4}} \\
\frac{dB}{dt} &= k_3(1-B) - \frac{k_4 A^{n_2}}{K_2^{n_2} + A^{n_2}} B \\
\frac{dC}{dt} &= k_5(1-C) - \frac{k_6 B^{n_3}}{K_3^{n_3} + B^{n_3}} C
\end{aligned}$$

Designs 4 and 5

$$\begin{aligned}
\frac{dA}{dt} &= k_1(1-A) - \frac{k_2 C^{n_1}}{K_1^{n_1} + C^{n_1}} A - k_7 A\frac{A^{n_4}}{K_4^{n_4} + A^{n_4}} \\
\frac{dB}{dt} &= k_3(1-B) - \frac{k_4 A^{n_2}}{K_2^{n_2} + A^{n_2}} B \\
\frac{dC}{dt} &= k_5(1-C) - \frac{k_6 B^{n_3}}{K_3^{n_3} + B^{n_3}} C
\end{aligned}$$

8.3 Distance

For the direct Hopf bifurcation detection the distance metric was defined to be

$$\rho(x, \mathcal{O}) = \frac{\prod_i \mathrm{Re}[\lambda_i]}{\prod_i (1 - 0.99 \exp(-|\,\mathrm{Im}[\lambda_i]|))}$$

where λ_i is the i^{th} complex eigenvalue of the linearized system in the steady state. Here $\epsilon = 0$ represents the location in parameter space where a limit cycle emerges through a Hopf bifurcation [18]. The final population was at $\epsilon = 0.001$.

To investigate the ability to achieve particular amplitude-frequency values, the distance was defined as $\rho(x, \mathcal{O}) = \{d_1, d_2, d_3\}$, where

$$\begin{aligned} d_1 &= \sum_n |x_{t_0+nT} - x_{t_0+(n-1)T}| \\ d_2 &= |f_t - f| \\ d_3 &= |\max x_{t>t_0} - \min x_{t>t_0} - A_t|, \end{aligned}$$

and n is an integer, f_t is the target frequency, f is the frequency determined from the largest component of the Fourier spectrum, A_t is the target amplitude and t_0 is a cut to remove initial transients (= 2s). The final population was defined to be $\epsilon = \{0.05, 0.05, 0.05\}$.

8.4 Priors

The priors were chosen to correspond to parameter ranges used in the original study; $k_1 \sim U(0, 10)$, $k_2 \sim U(0, 1000)$, $k_3 \sim U(0, 10)$, $k_4 \sim U(0, 1000)$, $k_6 \sim U(0, 1000)$, $k_7 \sim U(0, 100)$, $k_7^{strong} \sim U(500, 600)$, $n_i \sim U(1, 4)$ and $K_i \sim U(0, 4)$ [17].

References

[1] Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface* 6:187–202.

[2] Liepe J, *et al.* (2010) ABC-SysBio–approximate Bayesian computation in Python with GPU support. *Bioinformatics* 26:1797–9.

[3] Moral PD, Doucet A, Jasra A (2009) An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Annals of Applied Statistics* .

[4] Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol* 16:1791–8.

[5] Moral PD, Doucet A, Jasra A (2006) Sequential monte carlo samplers. *J Roy Stat Soc B* 68:411–436.

[6] Robert CP, Cornuet JM, Marin JM, Pillai N (2011) Lack of confidence in abc model choice. *arXiv* stat.ME.

[7] Grelaud A, Robert CP, Marin JM (2009) ABC methods for model choice in gibbs random fields. *Cr Math* 347:205–210.

[8] Toni T, Stumpf MPH (2010) Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* 26:104–10.

[9] Wilkinson RD (2008) Approximate Bayesian Computation (ABC) gives exact results under the assumption of model error. *arXiv* 0811.3355v1.

[10] Toni T (2010) Approximate Bayesian Computation for parameter inference and model selection in systems biology. *PhD thesis Imperial College London* .

[11] Ma W, Trusina A, El-Samad H, Lim WA, Tang C (2009) Defining network topologies that can achieve biochemical adaptation. *Cell* 138:760–73.

[12] Kim JR, Cho KH (2006) The multi-step phosphorelay mechanism of unorthodox two-component systems in E. coli realizes ultrasensitivity to stimuli while maintaining robustness to noises. *Comput Biol Chem* 30:438–44.

[13] Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403:335–8.

[14] Stricker J, *et al.* (2008) A fast, robust and tunable synthetic gene oscillator. *Nature* 456:516–9.

[15] Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M (2009) A tunable synthetic mammalian oscillator. *Nature* 457:309–12.

[16] Purcell O, Savery NJ, Grierson CS, di Bernardo M (2010) A comparative analysis of synthetic genetic oscillators. *Journal of the Royal Society Interface* .

[17] Tsai TYC, *et al.* (2008) Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science* 321:126–129.

[18] Chickarmane V, Paladugu SR, Bergmann F, Sauro HM (2005) Bifurcation discovery tool. *Bioinformatics* 21:3688–90.

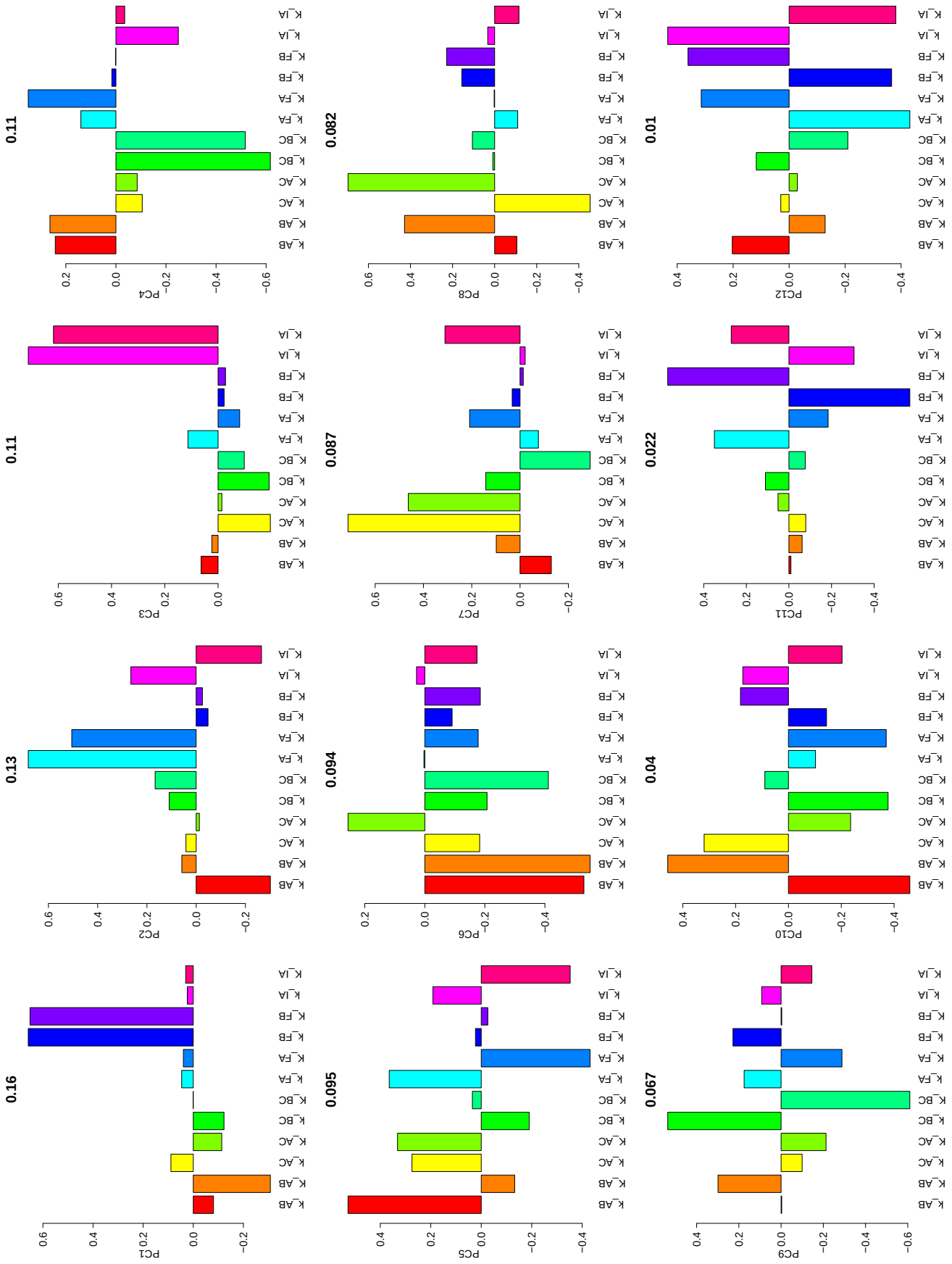

Figure 1: Biochemical adaptation: principal component analysis of the posterior distribution for model 11 in the case of no cooperativity.

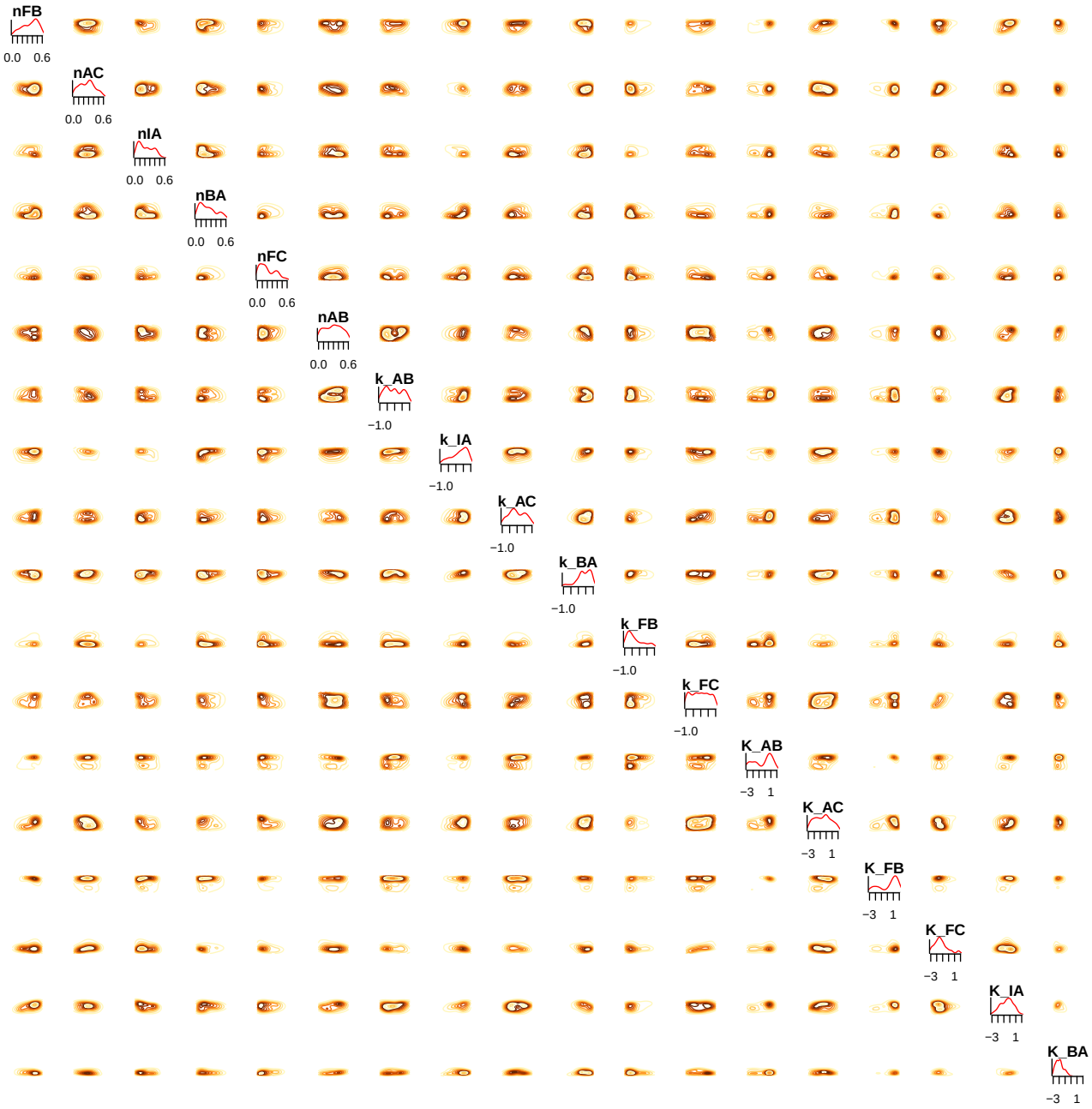

Figure 2: Biochemical adaptation: posterior distribution for model 4 in the case when cooperativity is included.

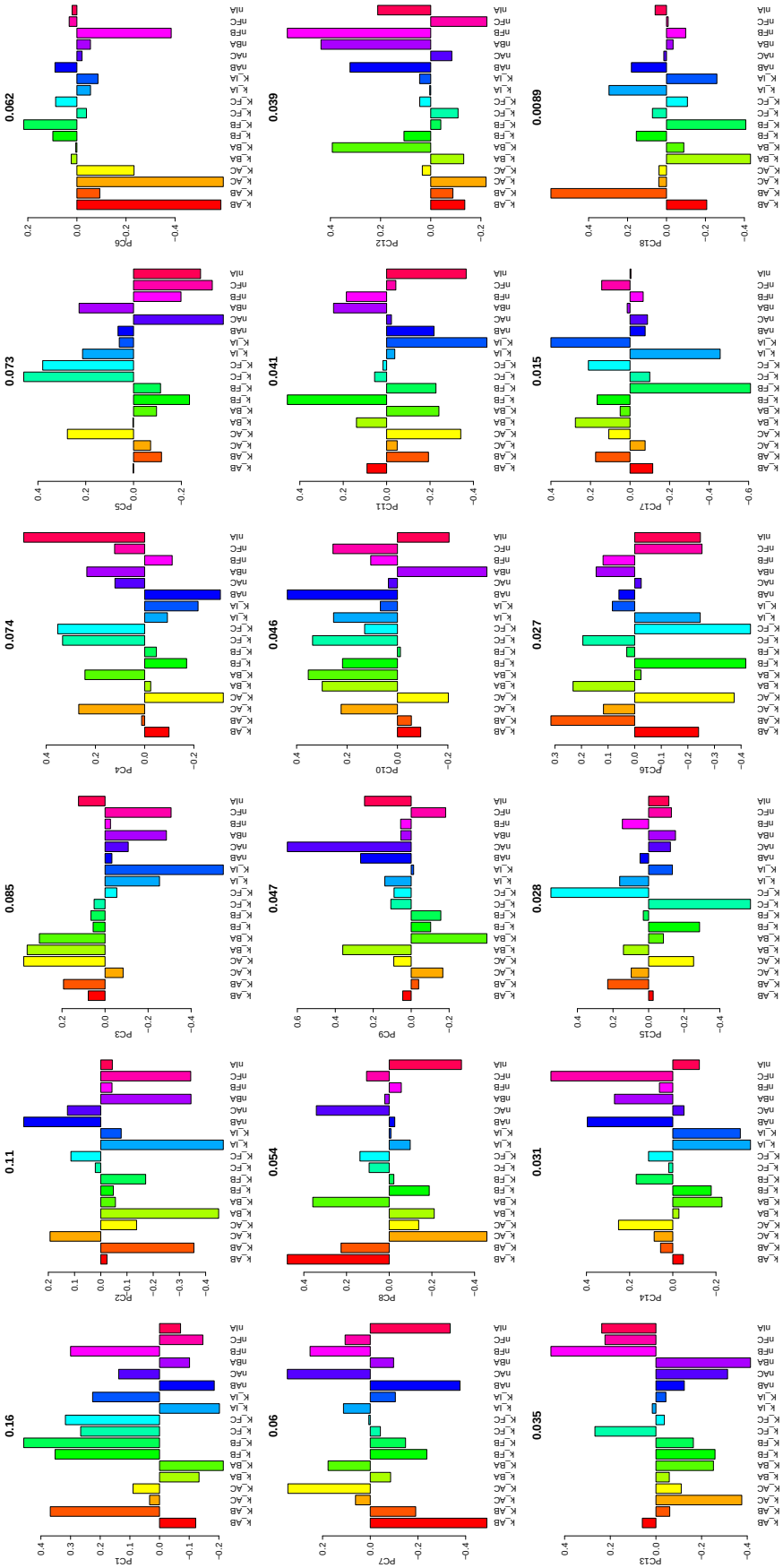

Figure 3: Biochemical adaptation: principal component analysis of the posterior distribution for model 4 in the case when cooperativity is included.

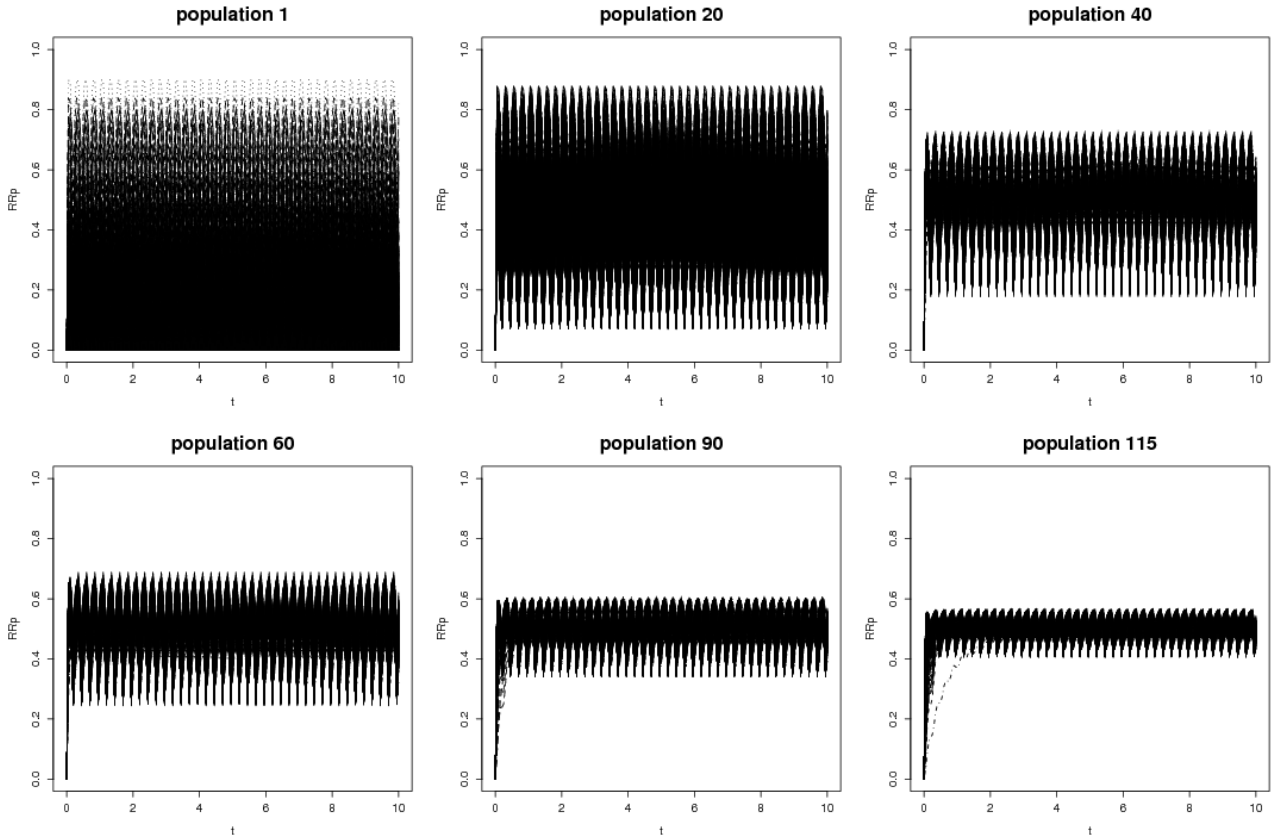

Figure 4: Two component systems: evolution to the noise reduction behavior.

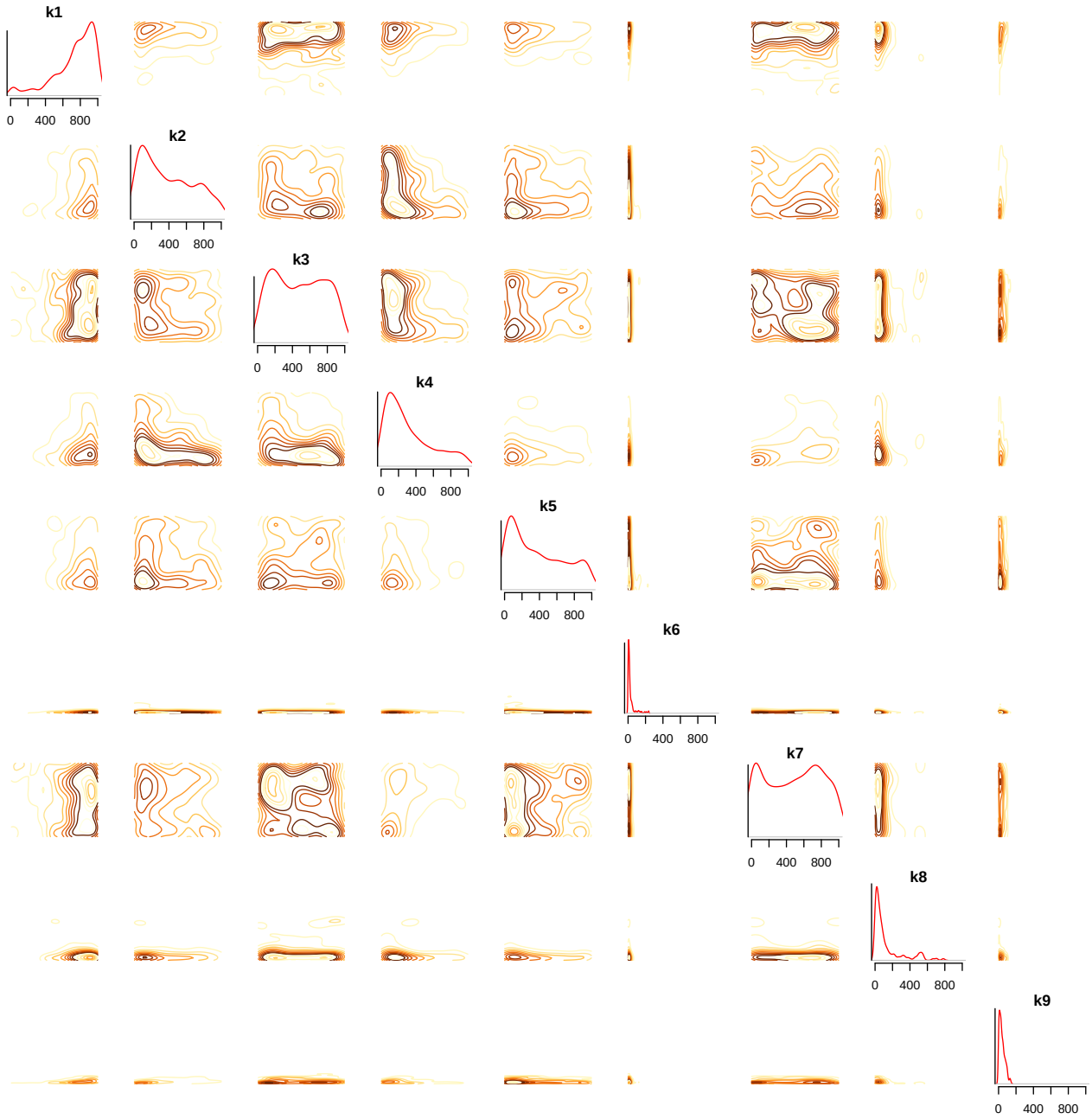

Figure 5: Two component systems: posterior distribution for the unorthodox system to achieve the noise reduction behavior.

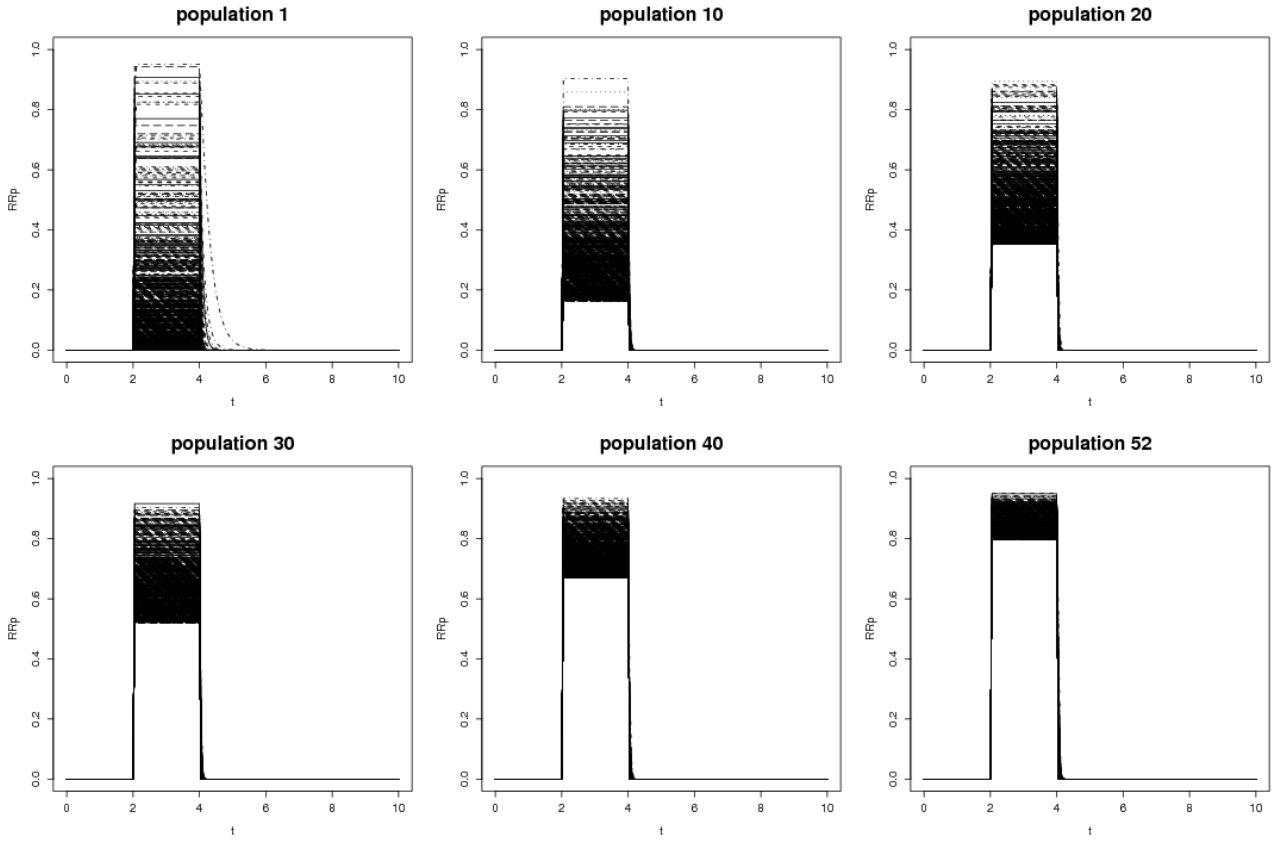

Figure 6: Two component systems: evolution to the signal reproduction behavior.

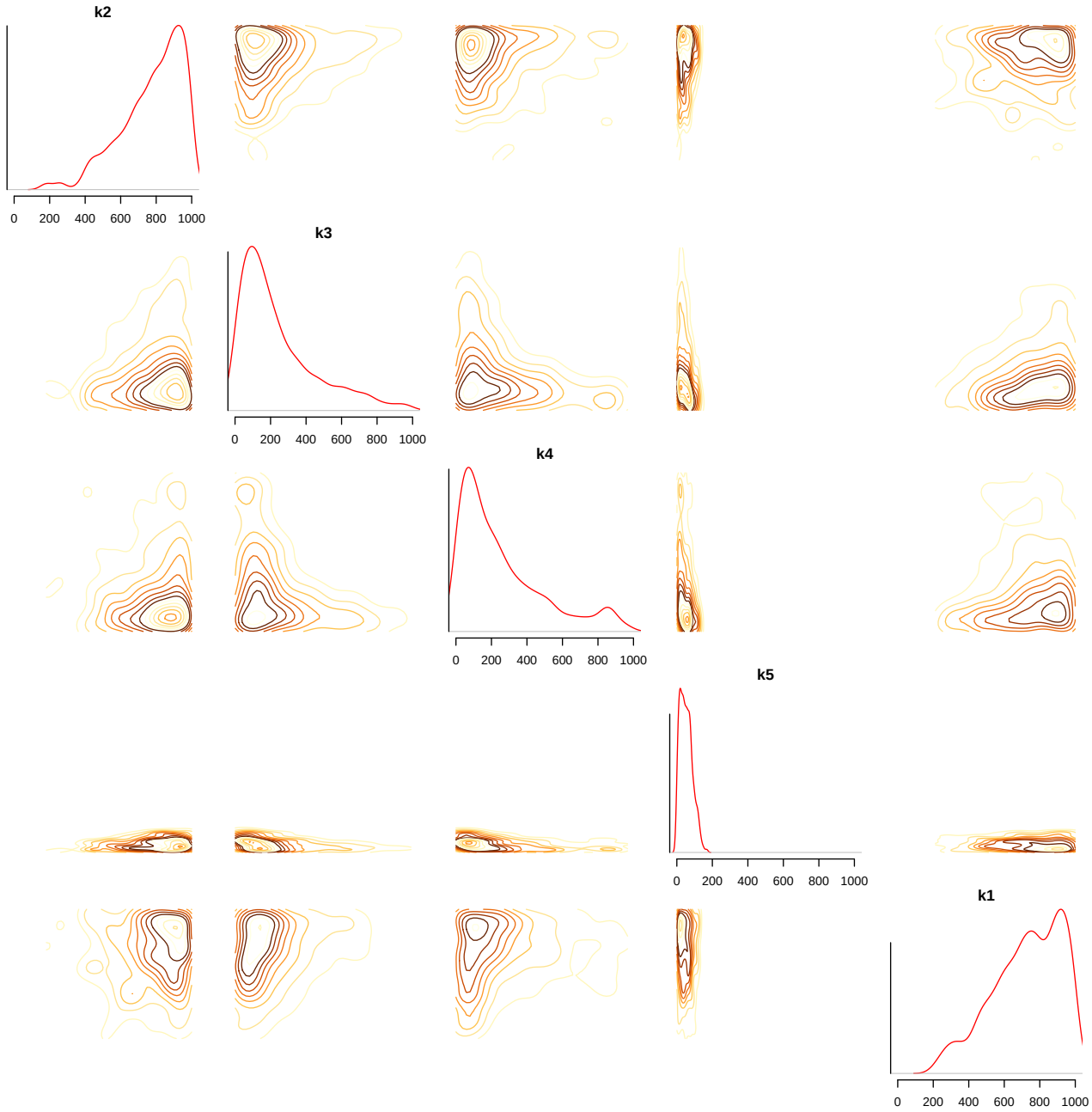

Figure 7: Two component systems: posterior distribution for the orthodox system to achieve the signal reproduction behavior.

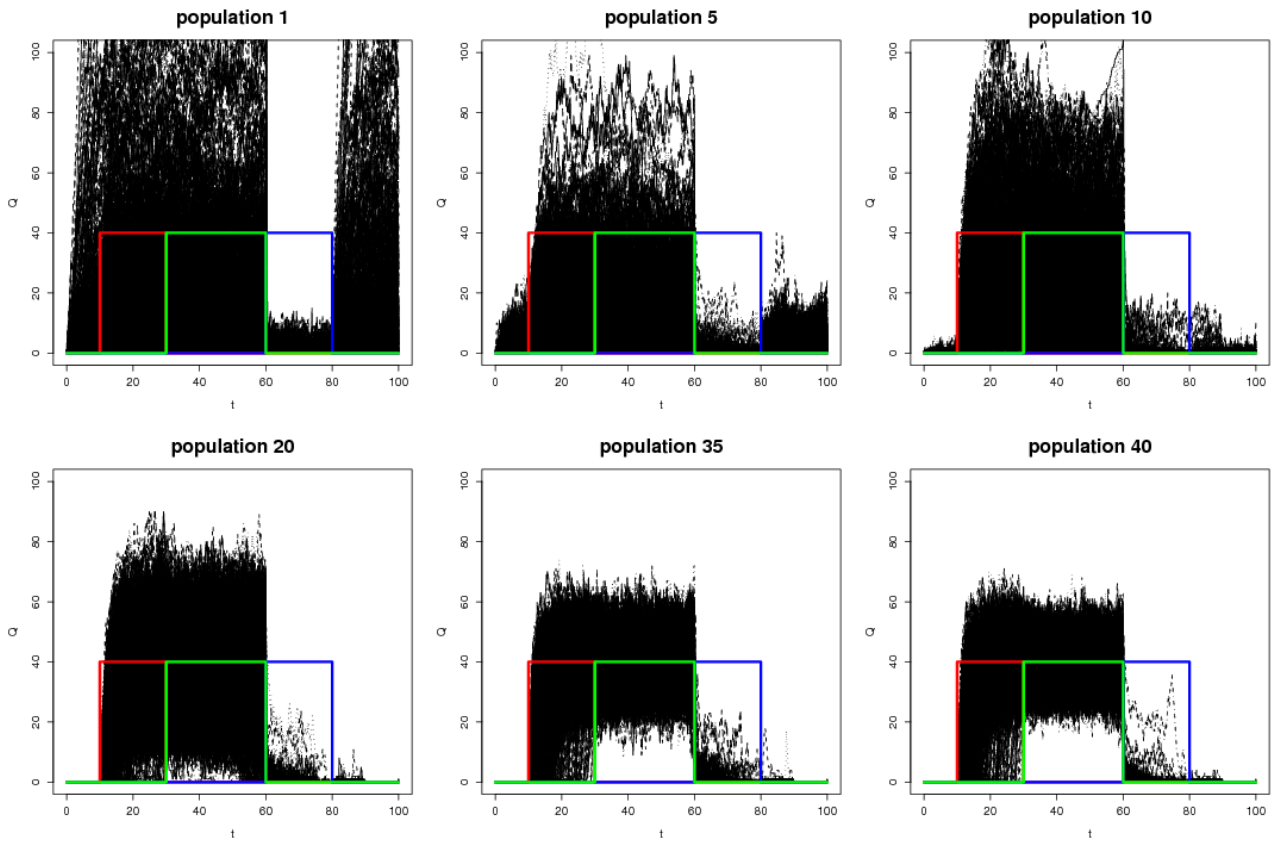

Figure 8: Stochastic genetic toggle switch: evolution to the desired behavior.

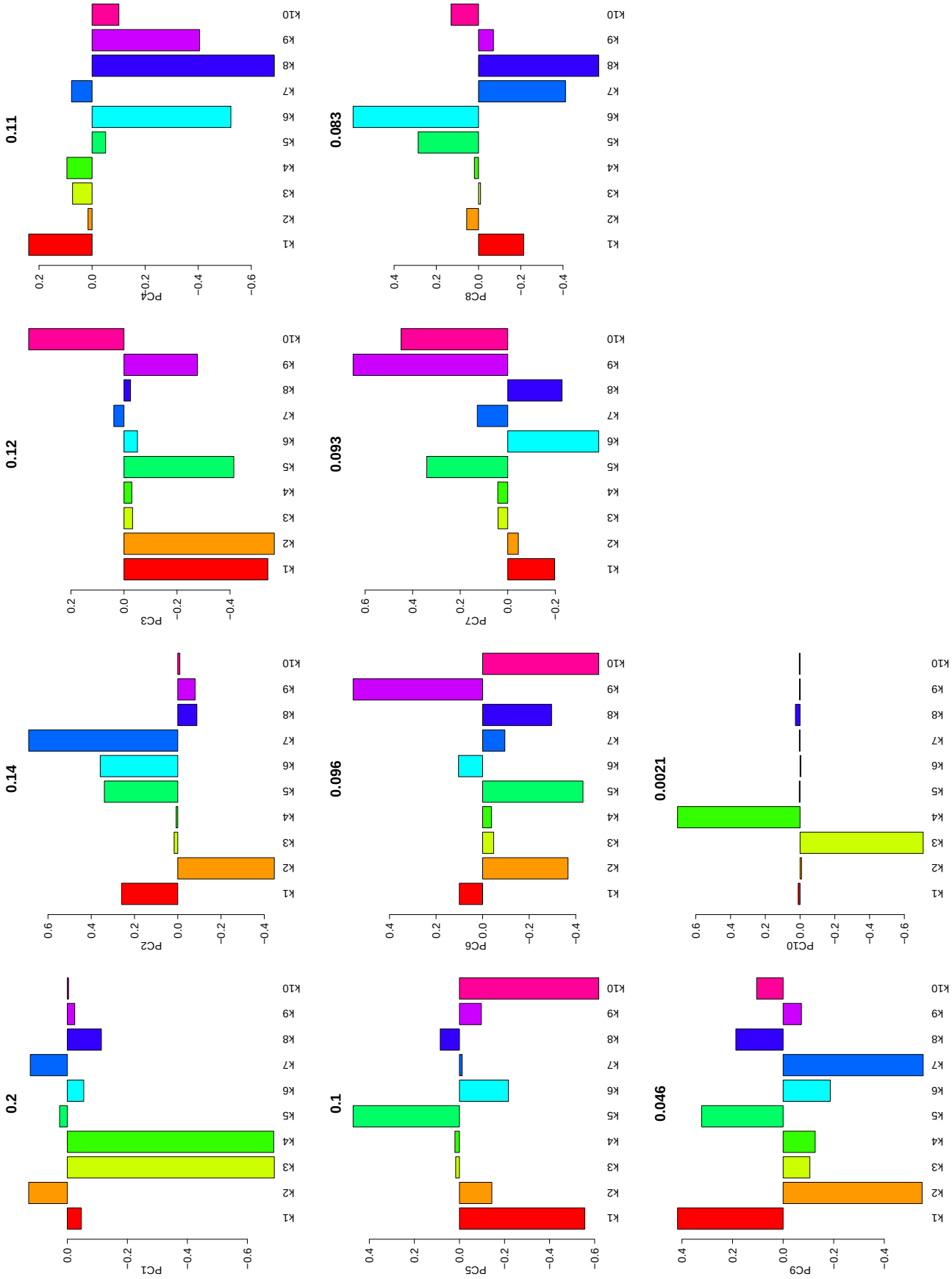

Figure 9: Stochastic genetic toggle switch: principal component analysis of the posterior distribution for model 2.

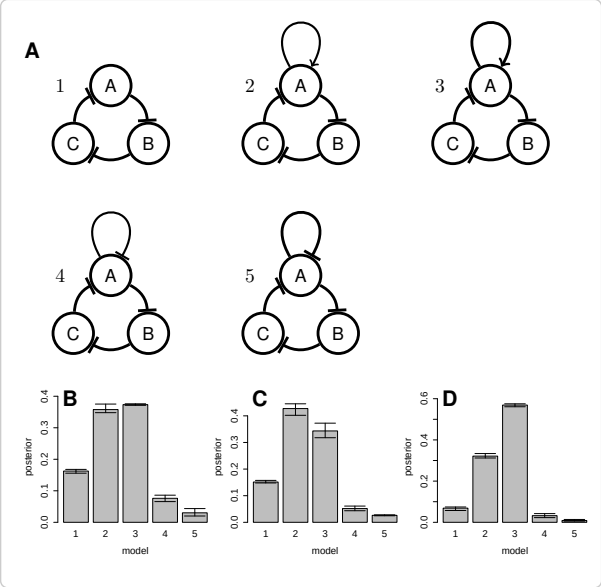

Figure 10: Robust oscillator models. A) 5 oscillator models. Model 1 is a loop of repressive enzymatic reactions. Models 2 and 3 have an additional positive feedback loop on node A with the feedback strength stronger in model 3 (represented by the thicker loop). Models 4 and 5 have an additional negative feedback loop on node A with the feedback strength stronger in model 5. B) Posterior probability for achieving Hopf bifurcation type limit cycle oscillations. C) Posterior probabilities for species A achieving oscillations with amplitude of 0.1 and a frequency of 1Hz. D) Posterior probabilities for species C achieving oscillations with amplitude of 0.1 and a frequency of 1Hz. The error bars in panels B, C and D indicate the variability in the marginal model posteriors over three separate runs.

Figure 11: Robust oscillator design: posterior distribution for model 2 to achieve limits cycles via a hopf bifurcation.

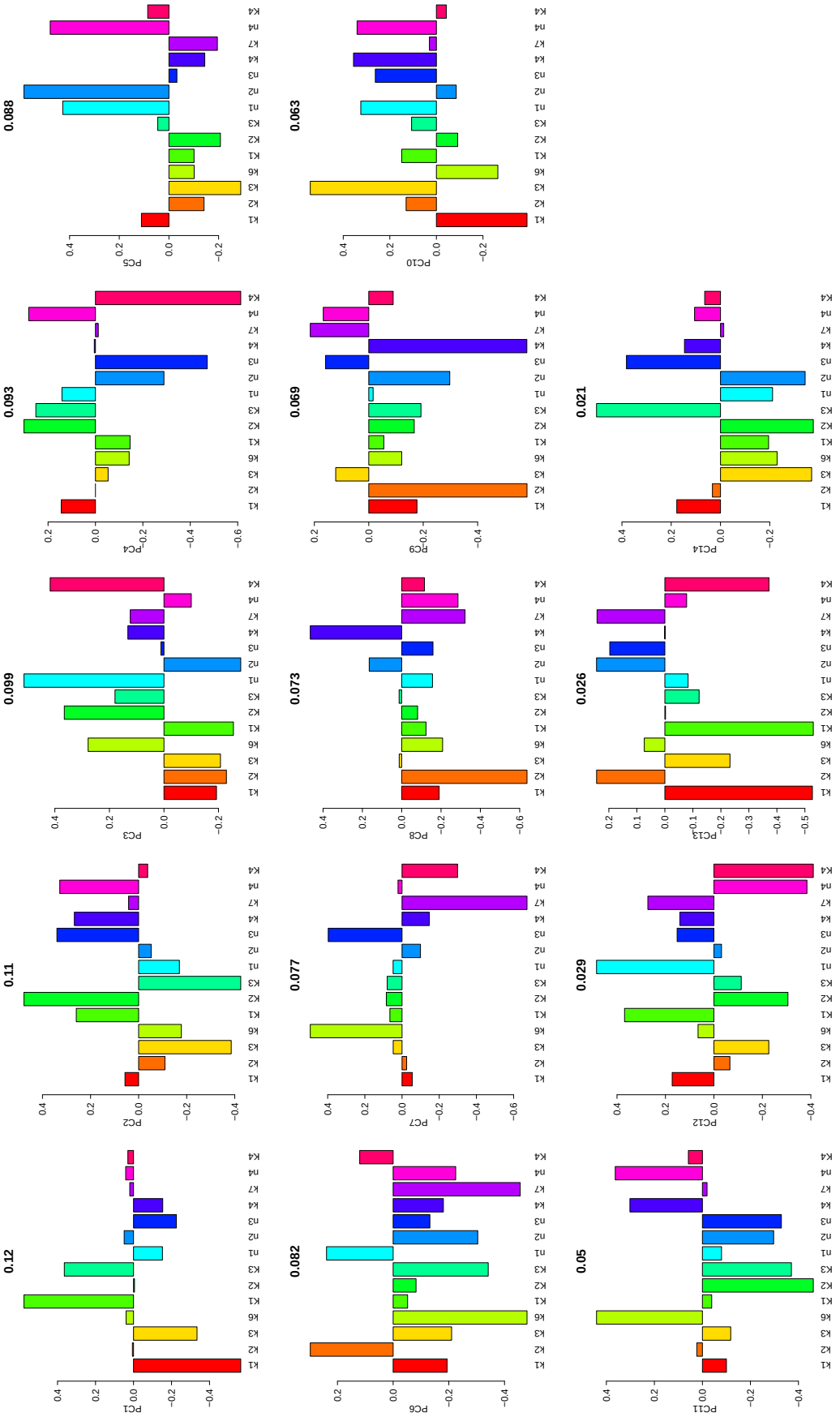

Figure 12: Robust oscillator design: principal component analysis of the posterior distribution for model 2 to achieve limits cycles via a hopf bifurcation.