Computation Cost based on Variance Estimates

In the first section of this supporting information file, we derive an approximation of the variance of the upper bound inequality ((4) in Main Text) by approximating the variance of the square root of the product of two dependent random variables, see Lemma 1.2. In the second section, we present the technical details for the comparison of the proposed strategy with the coupling method applied to all the sensitivity indices (SIs) of the model, discussed in section “Computational Cost” in the Main Text.

1 Variance of the estimator of the sensitivity bound

In order to get confidence intervals for an estimator that consists of simpler estimators, e.g., the estimator of the sensitivity bound ((4) in Main Text), the variance of the complex estimator must be computed. The Delta method gives a way to express the variance of the complex estimator using the variances of the simpler ones. In the following discussion we denote by $\mathcal{N}(0, \Sigma)$ the normal distribution with mean 0 and covariance matrix $\Sigma$ and by $\overset{d}{\rightarrow}$ the convergence in distribution.

**Theorem 1.1** (Delta method). Suppose $\{X_n\}_{n=0}^\infty$ is a sequence of random vectors in $\mathbb{R}^K$ such that

$$\sqrt{n}(X_n - \theta) \overset{d}{\rightarrow} \mathcal{N}(0, \Sigma),$$

for some $\theta \in \mathbb{R}^K$ and $\Sigma \in \mathbb{R}^{K \times K}$. Let $g : \mathbb{R}^K \rightarrow \mathbb{R}$ and $J(\theta) = \nabla g(x)|_{x=\theta}$. Then,

$$\sqrt{n}(g(X_n) - g(\theta)) \overset{d}{\rightarrow} \mathcal{N}(0, J(\theta)^T \Sigma J(\theta)).$$

**Proof.** See Theorem 3.1 in [1].

The practical interpretation of the Delta method is that for large fixed $N$ the variance of $g(X_N)$ can be approximated by $\frac{1}{N} J(\theta)^T \Sigma J(\theta)$. After a Definition and an intermediate Lemma we will prove a Lemma that gives and approximation for the variance of the sensitivity bound ((4) in Main Text).

**Definition 1.1** (moments). For a random variable $X$ we define the moments,

$$\mu_k = \begin{cases} E[X], & k = 1 \\ E[(X - E[X])^k], & k > 1 \end{cases},$$

and the central moments

$$m_k = E[X^k], \quad k = 1, 2, \ldots.$$  

Moreover, for two random variables, $X$ and $Y$, we define the joint moments by,

$$\xi_{k,\ell} = E[(X - E[X])^k(Y - E[Y])^\ell], \quad k, \ell = 1, 2, \ldots.$$  

**Lemma 1.1.** Assume $\{X_i\}_{i=1}^N$ are independent and identically distributed random variables. Then the following are true:
1. an unbiased estimator for the mean, $\mu_1$, of distribution $f$ is given by

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i,$$

and holds that

$$\sqrt{N} (\bar{X} - \mu_1) \overset{d}{\to} N(0, \mu_2),$$

(6)

and holds that

$$\sqrt{N} (\bar{X} - \mu_1) \overset{d}{\to} N(0, \mu_2),$$

(7)

2. an unbiased estimator for the variance, $\mu_2$, of distribution $f$ is given by

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2,$$

(8)

with variance equal to

$$\sqrt{N} (S^2 - \mu_2) \overset{d}{\to} N(0, \mu_4 - \mu_2^2).$$

(9)

Proof. The proof of the first part is an application of the Central Limit Theorem and can be found in [1]. The second part can be proved using the Delta method and the details can be found in Section 3.1 of [1].

Lemma 1.2. Let $\{X_i\}_{i=1}^{N}$ and $\{Y_i\}_{i=1}^{N}$ be independent and identically distributed random variables. We assume that the central moments $\mu_1$ of $X$ and $\nu_1, \nu_2, \nu_3, \nu_4$ of $Y$ are finite and in particular $\mu_1, \nu_1 \neq 0$. Moreover, the joint moment $\xi_{1,2}$ is finite.

Let $\bar{X}$ be an estimator for $E[X_1]$ and $S_Y^2$ an estimator for $Var[Y_1]$, as defined in Lemma 1.1. Then, it holds that

$$\sqrt{N} \left( \sqrt{\bar{X} S_Y^2} - \sqrt{\mu_1 \nu_2} \right) \overset{d}{\to} N(0, \lambda),$$

(10)

where

$$\lambda = \frac{\nu_2}{4\mu_1} \mu_2 + \frac{\mu_1}{4\nu_2} (\nu_4 - \nu_2^2) + \frac{1}{2} \xi_{1,2}.$$  

(11)

Proof. For simplicity in the presentation we will work with the estimator $S^2 = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \bar{Y})^2$ instead of (8). The same result holds for the estimator (8), see the Example 3.2 in [1].

First, notice that $\sqrt{\bar{X} S_Y^2}$ is equal to $\phi(\bar{X}, \bar{Y}, Y^2)$ for $\phi(x, y, z) = \sqrt{x(y^2 - z)}$. Moreover, from the multivariate Central Limit Theorem, see [1], it holds that,

$$\sqrt{N} \left( \frac{\bar{X}}{\bar{Y}} \right) - \sqrt{N} \left( \frac{m_1}{n_1} \frac{n_2}{n_2} \right) \overset{d}{\to} N(0, \Sigma),$$

(12)

where the covariance matrix $\Sigma$ is equal to

$$\Sigma = E \left[ (X - m_1, Y - n_1, Y^2 - n_2)(X - m_1, Y - n_1, Y^2 - n_2)^T \right].$$

(13)

The Jacobian of $\varphi$ evaluated at $\theta = (m_1, n_1, n_2)$ is equal to

$$J(\theta) = \frac{1}{2\sqrt{\mu_1 \nu_2}} \left( \nu_2, 2\mu_1 \nu_1, -\mu_1 \right)^T.$$ 

(14)

Finally, use Theorem 1.1 to conclude equation (10) with $\lambda$ given by

$$\lambda = J(\theta)^T \Sigma J(\theta).$$

(15)

Equation (11) follows after basic algebraic manipulations. 

Figure A: The ratio of the variance of the coupling estimator ((12) in Main Text) to the variance of the bound estimator ((4) in Main Text) is presented here for all parameters of the EGFR model. The sum of these values contribute to the $G_1$ term in formula (22). This result indicates that the variance of the bound estimator is much less than that of the coupling estimator. The ratio of variance for Parameter 11 does not appear in the graph because the variance of the coupling estimator is equal to zero.

2 Technical details regarding the computational cost

In this section, we compare the computational cost of (a) the proposed strategy, involving estimation of the upper bound in the inequality ((4) in Main Text) and the subsequent estimation of just the most significant sensitivity indices ($\tilde{S}_{k,\ell m}$ (see (12) in Main Text), to (b) the estimation of the full sensitivity matrix, using the finite-difference coupling method for each one of the entries of the sensitivity matrix. The comparison will be done under the requirement that the relative confidence intervals of all estimators involved in both approaches will be of length equal to $\delta \ll 1$. The relative confidence interval is defined as the usual confidence interval, however the estimator of the standard deviation is replaced by a normalized standard deviation

$$\hat{\sigma} = \begin{cases} \frac{\sigma}{|\overline{x}|}, & |\overline{x}| > \epsilon \\ \hat{\sigma}, & \text{otherwise.} \end{cases}$$

where $\hat{\sigma}$ is the estimated standard deviation of the random variable $X$ and $\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$ is the estimator of the mean of $X$. With this definition, random variables with large mean values have standard deviation comparable with random variables with small mean values. $\epsilon$ plays the role of a separator between large and small values. In our study we used $\epsilon = 0.01$.

Next, we set notation before proceeding to the comparison of the estimators. The time averaged observable will be that of equation (1) in Main Text with $T = 10$ and $f_{\ell}(x) = x_\ell$, $\ell = 1,\ldots,N$, for the particular model considered here. The length of the relative confidence interval for the coupling estimator ((12) in Main Text) for the $k$–th parameter and the $\ell$–th observable, using $M_{k,\ell}$ samples, is equal to

$$L_{k,\ell}^c = 2\alpha \frac{V_{k,\ell}}{\sqrt{M_{k,\ell}}}, \quad k = 1,\ldots,K, \quad \ell = 1,\ldots,N,$$
where $V_{c,k,\ell}^c$ is the estimated normalized standard deviation, as defined in (16), using $M_{k,\ell}$ samples, and $\alpha = 1.96$ is the inverse of the cumulative normal distribution evaluated at 0.975 leading to a 95% confidence interval, see [2]. The superscript $c$ in (17) stands for the coupling method, see Section “Step 2: Finding the most sensitive parameters” in Main Text. Under the requirement that $L_{c,k,\ell}^c = \delta$ we get the minimum number of samples for every pair $(k, \ell)$$$
abla_{c,k,\ell} = \left\lceil \left( \frac{2\alpha}{\delta} V_{c,k,\ell}^c \right)^2 \right\rceil,$$$ (18)

where $[x]$ is the ceiling function defined by $[x] = \min\{n \in \mathbb{Z} : n \geq x\}$. In the coupling method, as a variant of the SSA method [3], for every perturbed parameter we compute the SIs of all species. Hence, the minimum number of trajectories needed to be produced by this method is determined by the species with the highest variance. Thus, we define the quantity

$$M_{c}^c = \max_{1 \leq \ell \leq N} M_{c,k,\ell}^c.$$ (19)

Let $M_{c}^c$ be the sum of $M_{c,k}^c$ over all $k$, i.e. the total number of samples required by the coupling method such that all estimated SIs will have a relative confidence interval equal to $\delta$. This number will be compared to the minimum samples required by the method proposed in this paper under the requirement that all estimators’ relative confidence interval will be of the same length.

Working in the same way, we find the minimum number of trajectories required in order the relative confidence interval of the estimator of the sensitivity bound (SB) ((4) in Main Text) to be less or equal to $\delta$, $$M_{b,k,\ell}^b = \left\lceil \left( \frac{2\alpha}{\delta} V_{b,k,\ell}^b \right)^2 \right\rceil,$$$ (20)

where the superscript $b$ stands for “bound”. The statistical estimators of the SB are presented in supporting information File S2 and the variance of the estimator is computed using Lemma 1.2. In one trajectory we collect samples from both FIM and IAT, i.e. the minimum number of iterations will be determined by the bound of the sensitivity of the $\ell$–th species of the $k$–th parameter with the higher variance. Thus we define the quantity

$$M_{b}^b = \max_{1 \leq \ell \leq K} \max_{1 \leq k \leq K} M_{b,k,\ell}^b,$$$ (21)

which is the minimum number of samples required for the upper bound estimator to have a relative confidence interval of less or equal to $\delta$.

In order to compare the straightforward SI estimator ((12) in Main Text) with the strategy proposed here, we have to compare the numbers in (19) and (21). Note that in our strategy, after the first screening using the SB is performed, we have to estimate accurately the SIs using the coupling method. Thus the ratio of the total computational cost of the proposed strategy to the computational cost of coupling in terms of number of samples is,$$
\frac{G = \frac{M_{b}^b + \sum_{k \in S} M_{c,k}^c}{\sum_{k=1}^{K} M_{c,k}^c}}{1 + \sum_{k=1}^{K} \frac{M_{c,k}^c}{M_{b,k}^b}} + \frac{1}{1 + \sum_{k \in S} \frac{M_{c,k}^c}{M_{b,k}^b}} = G_1 + G_2,$$$ (22)

where $I$ and $S$ are the sets of indices of insensitive and sensitive parameters, respectively. Note that in the above equation we dropped the ceiling functions in (19) and (21) making the calculation easier and introducing a negligible error. The interpretation of $G_1$ is “how many times is the coupling faster than Step 1 of the proposed strategy”, i.e. the computation of the upper bound, and that of $G_2$ “how many times is the coupling applied to all sensitivity indices is faster than Step 2 of the proposed strategy”, i.e. neglecting SIs with small values. The ratio $1/G$ is the total speed-up of the proposed two–step strategy.
We argue that the estimation of the SB is much faster than that of the actual SIs using the coupling method for the same confidence interval and the same $\epsilon$ in the normalized standard deviation (16). This means that $G_1 \ll 1$ and the important term in (22) is $G_2$. In Figure 1, the ratio

$$G_{1,k} = \frac{M_c^k}{M_b}, \quad k = 1, \ldots, K$$

(23)

is plotted in logarithmic scale for the EGFR model described in Section “The EGFR model” in Main Text with $\epsilon = 0.01$ in definition (16). The $G_1$ term is equal to 0.0034 showing that the estimation of the SB needs about 300 times less samples than that of the computation of SIs using the coupling method. Finally, we can get an approximation to $G_2$ by arguing that $M_c^k$ is approximately the same for all parameters, $k = 1, \ldots, K$. Then, from (22), $G_2 \approx \frac{K'}{K}$, where $K'$ is the size of the set $S$, i.e., $K'$ is the number of sensitive parameters. Under the assumption that $G_1 \ll 1$ we get an approximation for the total computational gain,

$$G \approx \frac{K'}{K}.$$  

(24)

References

1. A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

2. L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2004.

3. D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, 22:403–434, 1976.