

# SCIENTIFIC REPORTS



OPEN

## Impact of measurement noise, experimental design, and estimation methods on Modular Response Analysis based network reconstruction

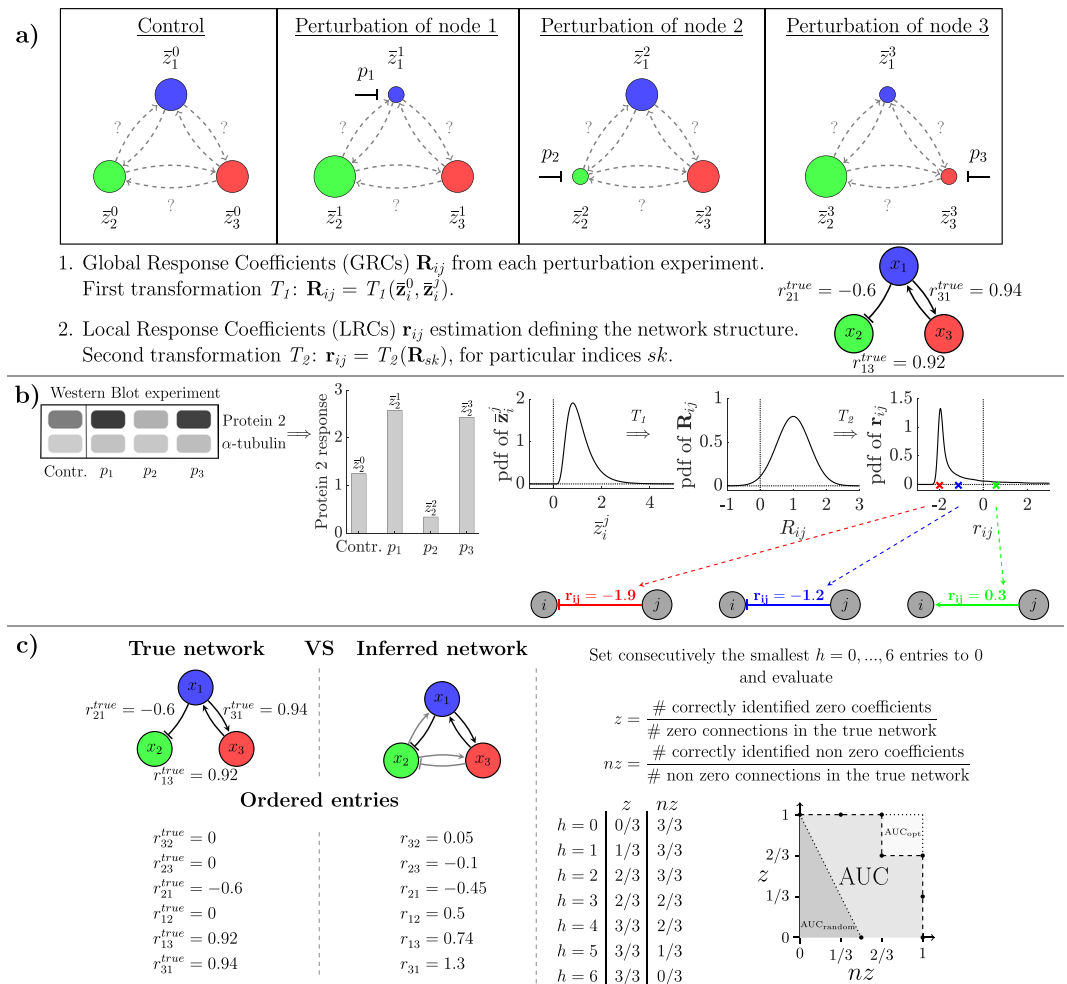
Caterina Thomaseth<sup>1</sup>, Dirk Fey<sup>2</sup>, Tapesh Santra<sup>2</sup>, Oleksii S. Rukhlenko<sup>2</sup>, Nicole E. Radde<sup>1</sup> & Boris N. Kholodenko<sup>2,3</sup>

Modular Response Analysis (MRA) is a method to reconstruct signalling networks from steady-state perturbation data which has frequently been used in different settings. Since these data are usually noisy due to multi-step measurement procedures and biological variability, it is important to investigate the effect of this noise onto network reconstruction. Here we present a systematic study to investigate propagation of noise from concentration measurements to network structures. Therefore, we design an *in silico* study of the MAPK and the p53 signalling pathways with realistic noise settings. We make use of statistical concepts and measures to evaluate accuracy and precision of individual inferred interactions and resulting network structures. Our results allow to derive clear recommendations to optimize the performance of MRA based network reconstruction: First, large perturbations are favorable in terms of accuracy even for models with non-linear steady-state response curves. Second, a single control measurement for different perturbation experiments seems to be sufficient for network reconstruction, and third, we recommend to execute the MRA workflow with the mean of different replicates for concentration measurements rather than using computationally more involved regression strategies.

Advanced experimental techniques have facilitated our mechanistic understanding of intracellular processes in the last decades. However, the problem of network reconstruction from experimental data remains a challenging task, for which many different approaches have been suggested. Among those, Modular Response Analysis (MRA) has been proven successful in many applications<sup>1–4</sup>. MRA uses steady-state data of experiments in which each node of a network is perturbed successively (see Fig. 1a). These steady-state data are transformed into quantitative pairwise interaction strengths, denoted Local Response Coefficients (LRCs), which define the network structure. This is done in a two-step process, in which first concentration measurements are transformed into Global Response Coefficients (GRCs), which are then used to calculate the LRCs. MRA is an elegant method that gives reliable results in case that concentrations can be accurately measured and measurement noise can be neglected<sup>5</sup>.

MRA is however often applied in settings in which we have to deal with real noisy experimental data and few replicates, for example when using Western blotting to investigate signalling pathways, as exemplified in<sup>1,2</sup>. In these studies the authors addressed the issue of measurement noise by using statistical approaches, like Monte Carlo simulations or maximum likelihood, to estimate interaction strengths and respective uncertainties. Despite the extensive usage of MRA, the effect of noise in the input data on network reconstruction is not completely understood. Recent developments include statistical reformulations of the MRA that have been suggested to address the issue of noisy and sparse/insufficient data<sup>6</sup>. A further extension combines the classical deterministic

<sup>1</sup>University of Stuttgart, Institute for Systems Theory and Automatic Control, Stuttgart, Germany. <sup>2</sup>University College Dublin, Systems Biology Ireland, UCD School of Medicine, Dublin, Ireland. <sup>3</sup>Department of Pharmacology, Yale University School of Medicine, New Haven, CT, USA. Caterina Thomaseth and Dirk Fey contributed equally. Nicole E. Radde and Boris N. Kholodenko jointly supervised this work. Correspondence and requests for materials should be addressed to N.E.R. (email: [nicole.radde@ist.uni-stuttgart.de](mailto:nicole.radde@ist.uni-stuttgart.de))



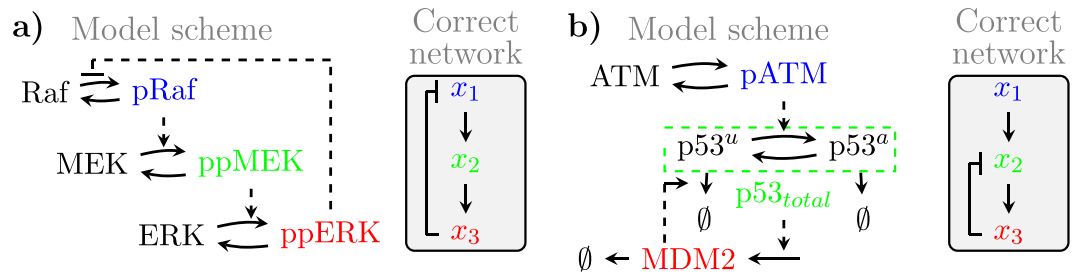
**Figure 1.** Variability and noise in measurements affect network reconstruction via Modular Response Analysis. **(a)** MRA workflow for a three-node network. After subsequent perturbation of all nodes and quantification of concentration fold changes, LRCs  $r_{ij}$  are calculated via a two-step non-linear transformation. **(b)** One exemplary realization (one replicate) of the noisy measurement for protein 2 is evaluated in all experimental conditions (left part). These values are interpreted as samples of the corresponding distributions, described by the probability density function (pdf of  $\bar{z}_i^j$ ) arising from the measurement noise. On the right we describe how network reconstruction is affected by the variability of the measured protein concentrations in terms of propagation of distributions from the measurements  $\bar{z}_i^j$  and  $\bar{z}_i^0$  via the GRCs  $\mathbf{R}_{ij}$  to the LRCs  $r_{ij}$ . **(c)** Performance evaluation of inferred network structures is done by using the assessment method presented in<sup>9</sup>, which compares the inferred structure with a reference structure. Similar to ROC analysis, the Area Under the Curve (AUC) serves as a normalized measure for the fit quality, and varies from the average value 0.25 in the random case (darker grey triangle) up to the optimal value 1 for a correctly identified network (whole square).

MRA framework with advanced nonparametric single-cell data resampling to discriminate between direct and indirect connectivities<sup>7</sup>.

Despite the broad literature tackling the issue of experimental noise, a comprehensive and systematic study on how network inference via MRA is affected by noise is still required. In particular, some of the studies do not consider noise propagation from the input data to the estimated LRCs, but start directly with the GRCs<sup>8</sup>. Furthermore, a realistic statistical characterization of the MRA variables (measured data, GRCs and LRCs) and a robustness analysis in a general experimental setup are also still missing.

In this study we develop a statistical framework to analyse noise propagation by transformations of input into output distributions (Fig. 1b). By comparing different experimental and computational strategies in an *in silico* study, we derive recipes for experimentalists and modellers regarding an optimal MRA workflow design. In particular, we investigate (1) how non-linear transformations and mathematical approximations of the MRA framework affect noise propagation; (2) the influence of perturbation strength, control strategy and number of replicates on the uncertainty of the estimated interactions; (3) the effects of different estimation methods on the performance of the network inference problem.

To evaluate the resulting network structure, we apply a performance evaluation method that was proposed in<sup>9</sup> and is schematically depicted in Fig. 1c. It works similar to a Receiver Operating Characteristic (ROC) and



**Figure 2.** Two test-bed models. Shown are the reaction kinetic schemes (left) and the arising network structure (right) for (a) the MAPK system; (b) the p53 system. The coloured nodes (blue, green, red) indicate the measured species, which define the states of the models.

its Area Under the Curve (AUC value) for evaluating the performance of a classifier but additionally takes correctness of the sign of an inferred interaction into account. A correctly identified network has an AUC value of 1 (whole square), the random case corresponds on average to an AUC value of 0.25 (darker grey triangle).

Results are given for two test-bed examples of well-known signalling pathways, a model for the mitogen-activated protein kinase (MAPK) pathway and a model for the tumour suppressor protein p53. These models show very different non-linear properties regarding their steady-state behaviour in dependence of perturbation strengths. Our results show that large perturbations and few technical replicates, combined with a simple control strategy and a basic estimation method, lead to an optimal bias-variability trade-off of the estimated pairwise interactions and also give robust results regarding network reconstruction.

## Results

**MAPK and p53 test-bed models with complementary dynamic behaviours.** We used two test-bed examples to investigate noise propagation and impacts of the experimental design on the MRA estimates. The models are similar in that they both consist of three states (nodes) with two positive and one negative interactions (Fig. 2). However, both models feature very different equations, dynamics, and non-linear properties (Suppl. Fig. S1), allowing us to judge the generality of our results, and to investigate the impacts of moderate and strong non-linearities.

A model of signal transduction of the MAPK pathway upon EGF stimulation is illustrated in Fig. 2a. It consists of a three-tiered cascade of phosphorylation-dephosphorylation cycles in which pRaf phosphorylates and thereby activates MEK, which then activates ERK, which negatively feeds back to Raf<sup>10</sup>. Both MEK and ERK require phosphorylation at two sites to become fully active, which is for simplicity assumed to happen in a single reaction step for both proteins<sup>11</sup>. Variables  $x_1$ ,  $x_2$  and  $x_3$  represent protein activities. All reactions are modelled using Michaelis-Menten type equations (Suppl. Fig. S1a). This system exhibits moderate non-linearity for the chosen parameters (Suppl. Fig. S1a).

The p53 model (Fig. 2b) is based on the core signalling system of the DNA damage response<sup>12</sup>. Here, ATM activates p53 by phosphorylation and protein stabilisation, p53 activates MDM2 by inducing gene expression, and MDM2 mediates a negative feedback loop to p53 by promoting p53 degradation<sup>13</sup>. In contrast to the MAPK system, the p53 system exhibits a strong degree of non-linearity (Suppl. Fig. S1b), including three ultra-sensitive Hill type equations for the reaction kinetics (Suppl. Fig. S1b).

**Problem formulation.** MRA is a mathematical approach to reveal interaction strengths from steady-state perturbation data of a dynamic network of interacting modules. In our examples each module consists of a single protein, and we refer to them as nodes<sup>5</sup>. Considering the network at equilibrium, pairwise interaction strengths between nodes are characterized by quantifying the immediate change of the activity of one node of the network caused by a small change of another node, whereas the rest of the network is unaffected. Local Response Coefficients (LRCs) express these local effects among  $N$  nodes and are defined as the fractional change of the steady-state concentration of node  $i$  ( $\bar{x}_i$ ) with respect to that of node  $j$  ( $\bar{x}_j$ ), while keeping all other nodes  $\bar{x}_k$ ,  $k \neq i, j$ , at a constant level,

$$\text{LRCs: } r_{ij}^{\text{true}} = \frac{\partial \ln \bar{x}_i(\bar{x}_j, \bar{x}_k)}{\partial \ln \bar{x}_j}, \quad \bar{x}_k = \text{const}, \quad k \neq i, j, \quad i \neq j, \quad i, j = 1, \dots, N. \quad (1)$$

These LRCs describe pairwise interactions between nodes when they act in isolation and are not directly accessible. A perturbation of one parameter  $p_j$  which specifically affects the activity of node  $j$ , spreads over the whole network, thus generating a global change of the equilibria of all nodes. This global change can be quantified from fold change measurements of concentrations relative to the unperturbed system. Formally, Global Response Coefficients (GRCs) are defined as the total derivative of the logarithm of the steady-state variables ( $\ln \bar{x}_i$ ) with respect to the perturbed parameter ( $p_j$ ) (see exemplary network in Fig. 1a),

$$\text{GRCs: } R_{ij}^{\text{true}} = \frac{d \ln \bar{x}_i(p_j)}{d p_j} = \frac{1}{\bar{x}_i(p_j)} \frac{d \bar{x}_i(p_j)}{d p_j}, \quad i, j = 1, \dots, N. \quad (2)$$

The corresponding MRA equations<sup>5</sup>

$$\sum_{j=1, j \neq i}^n r_{ij}^{true} R_{jk}^{true} = R_{ik}^{true} \quad (3)$$

establish a mathematically exact relationship between the GRCs and LRCs and can be used to extract LRCs from GRCs.

In our *in silico* study we assume that the investigated dynamical system can be described by a true underlying noise-free Ordinary Differential Equation (ODE) model  $\dot{x} = f(x, p)$ , with state variables  $x = (x_1, \dots, x_N) \in \mathbb{R}_+^N$  and parameters  $p = (p_1, \dots, p_N) \in \mathbb{R}_+^N$ . The state variable  $x_i$  represents the activity of node  $i$ . The parameters  $p_j > 0, j = 1, \dots, N$ , are all equal to one in the nominal setting (control experiment), and can be varied to simulate the perturbation experiment affecting the corresponding node  $j$ . These parameters often affect preserved quantities such as total protein concentrations or production rates. The true LRCs  $r_{ij}^{true}, i \neq j$ , are obtained by calculating the normalized entries of the Jacobian matrix at steady state, as described in<sup>5</sup>,

$$r_{ij}^{true} = - \left( \frac{\partial f_i(x, p)}{\partial x_j} / \frac{\partial f_i(x, p)}{\partial x_i} \right) \cdot \left( \frac{x_j}{x_i} \right) \Bigg|_{ss}, \quad i \neq j, \quad i, j = 1, \dots, N. \quad (4)$$

When the underlying ODE system is not known, LRCs can be inferred from concentration measurements via two non-linear transformations. In a first transformation  $T_1$ , differential GRCs are estimated from the steady states obtained in the control experiment ( $\bar{x}_i(p_j) =: \bar{x}_i^0$ ) and respective steady states in the perturbation experiments ( $\bar{x}_i(p_j + \Delta p_j) =: \bar{x}_i^j$ ),

$$T_1: \quad R_{ij}^{true} \Delta p_j \approx \tilde{R}_{ij} = \frac{\bar{x}_i^j - \bar{x}_i^0}{\frac{1}{2}(\bar{x}_i^j + \bar{x}_i^0)} = 2 \cdot \frac{\bar{x}_i^j - \bar{x}_i^0}{\bar{x}_i^0 + \bar{x}_i^j}, \quad i, j = 1, \dots, N, \quad (5)$$

where we have approximated the derivative in (2) with finite differences and  $\bar{x}_i^0$  with the average of  $\bar{x}_i^0$  and  $\bar{x}_i^j$ . The  $N \cdot (N - 1)$  LRCs are then obtained via substituting these  $\tilde{R}_{ij}$  into equation (3), which corresponds to solving  $N$  linear systems with  $N - 1$  equations in  $N - 1$  independent variables each<sup>5,14</sup>,

$$\sum_{j=1, j \neq i}^n \tilde{r}_{ij} \tilde{R}_{jk} = \tilde{R}_{ik}, \quad k \neq i; \quad i, k = 1, \dots, N. \quad (6)$$

We note here that  $\Delta p_k$  cancels out since it appears as a factor on both sides in this system. Due to the approximation (5), the values  $\tilde{r}_{ij}$  obtained in this way are also an approximation of the true LRCs ( $r_{ij}^{true}$ ), and depend in particular on the perturbation strengths. In the following we will always consider  $\tilde{R}_{ij}$  directly and thus refer to this measure simply as GRC. A second non-linear transformation, defined as  $T_2$ , provides a solution for all coefficients  $\tilde{r}_{ij}$ . As shown in Suppl. S2 for  $N = 3$ , we can rewrite equation (6) as a linear system,

$$\mathbf{y}(\tilde{R}_{ij}) = A(\tilde{R}_{ij}) \cdot \mathbf{x}, \quad (7)$$

in which the vector  $\mathbf{x}$  contains all unknowns  $\tilde{r}_{ij}$ , while  $\mathbf{y}(\tilde{R}_{ij})$  and  $A(\tilde{R}_{ij})$  are functions of the GRCs. The solution of this linear equation system, assuming that  $A$  has linearly independent columns, is given by

$$T_2: \quad \mathbf{x} = (A^T A)^{-1} A^T \mathbf{y}. \quad (8)$$

The variables  $\bar{x}_i^j$ ,  $\tilde{R}_{ij}$  and  $\tilde{r}_{ij}$  are assumed to be continuous functions of the perturbation parameters  $p_j$ . For the following analysis, we refer to the difference

$$\Delta r_{ij}(p_j) = \left| \tilde{r}_{ij}(p_j) - r_{ij}^{true} \right|, \quad i \neq j, \quad i, j = 1, \dots, N \quad (9)$$

as intrinsic bias which results from the approximations (5) and (8).

Our two considered test-bed models for the MAPK and the p53 signalling pathways (see Suppl. S1) significantly differ in the courses of  $\Delta r_{ij}$  over a large range of perturbation strengths  $p_j$  and thus can be considered as complementary examples concerning the approximation quality (8) and the validity of results in dependence of the perturbation strengths  $p_j$ .

According to our statistical methodology, concentration measurements are described by random variables  $\bar{z}_i^{0,j}, i, j = 1, \dots, N$  (see Fig. 1a,b), whose distribution is a function of the noise-free steady-state values  $\bar{x}_i^{0,j}$ . We consider a realistic error model consisting of a multiplicative and an independent additive part, similar to those suggested in<sup>15</sup>,

$$\bar{z}_i^{0,j} = \bar{x}_i^{0,j} \cdot \eta + \varepsilon, \quad \eta \sim \log \mathcal{N}(0, \sigma_\eta^2), \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad i, j = 1, \dots, N. \quad (10)$$

The parameters  $\sigma_\eta$  and  $\sigma_\varepsilon$  denote the standard deviations of the proportional ( $\log \eta$ ) and additive ( $\varepsilon$ ) measurement errors, respectively,  $\bar{x}_i^{0,j}$  are the simulated noise-free steady-state concentrations in the control (0) and perturbed ( $j$ ) conditions, and  $\bar{z}_i^{0,j}$  the resulting random variables representing the noisy simulated data.

|               | True variable     | Noise-free approximation | Random variable   | Realization | Probability density function  |
|---------------|-------------------|--------------------------|-------------------|-------------|-------------------------------|
| Steady states | $\bar{x}_i^{0,j}$ | —                        | $\bar{z}_i^{0,j}$ | $z_i^{0,j}$ | $P_{z_i^{0,j}}(z_i^{0,j})$    |
| GRCs          | $R_{ij}^{true}$   | $\tilde{R}_{ij}$         | $\mathbf{R}_{ij}$ | $R_{ij}$    | $P_{\mathbf{R}_{ij}}(R_{ij})$ |
| LRCs          | $r_{ij}^{true}$   | $\tilde{r}_{ij}$         | $\mathbf{r}_{ij}$ | $r_{ij}$    | $P_{\mathbf{r}_{ij}}(r_{ij})$ |

**Table 1.** Notation.

We note here that due to the experimental procedure we often do not directly obtain concentrations but fold changes only. Thus, measurements  $\bar{z}_i^0$  and  $\bar{z}_i^j$  are realizations of random variables which are proportional to the real concentrations. As shown in Fig. 1b (left), we assume that measurements refer to signals detected via Western Blotting which have been normalized to a loading control. Without loss of generality, we neglect the proportionality factor specific for each blot. In fact, the GRCs calculated via equation (5) and hence the LRCs are independent of these factors, as long as the two samples  $\bar{z}_i^0$  and  $\bar{z}_i^j$  have been quantified in the same blot.

Due to the two non-linear transformations (5) and (8), the GRCs and LRCs are also random variables, which we express as  $\mathbf{R}_{ij}$  and  $\mathbf{r}_{ij}$  (see Fig. 1a,b). Given measurements of  $\mathbf{R}_{ij}$ , a solution of equation (7) is obtained by applying estimation methods. The simplest choice is to use Ordinary (Linear) Least Squares, whose solution has exactly the same form as equation (8). Changing the method corresponds to changing the operator  $T_2$ , which remains a non-linear function of the GRCs in all cases.

Since it is impossible to derive the distributions of the LRCs directly from the error model (10) of the measurements, we applied a Monte Carlo approach in which we used our error model to simulate experimental data and propagated these to respective LRCs via the transformations  $T_1$  and  $T_2$ .

Throughout the manuscript, bold letters indicate random variables and simple letters refer to realizations, as listed in Table 1, which collects the notation of all model quantities. To help the readers, we also collected a list of all abbreviations in Table 3 at the end of the Supplementary Material.

**Solving the MRA equations results in heavy-tailed distributions for the estimated LRCs.** We started our study by investigating the propagation of noise from the concentration measurements to the estimated LRCs as a basis to deduce strategies for an optimal design of experiments and estimation methods.

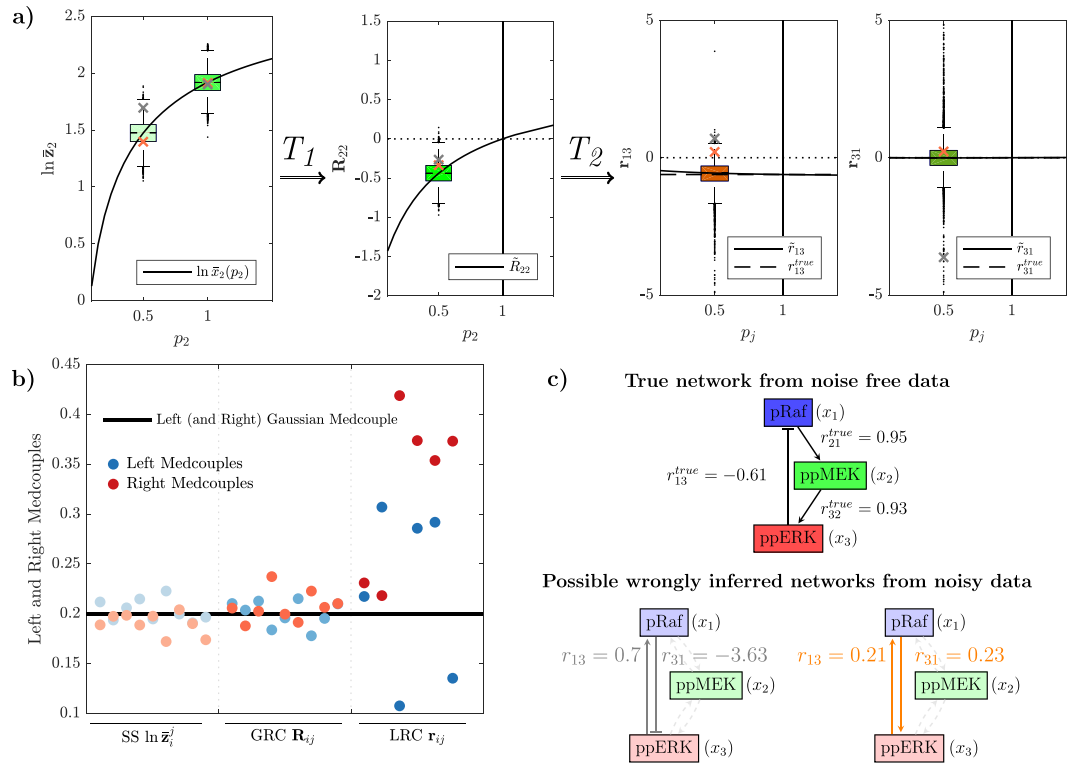
Therefore, we simulated the MAPK model with noise parameters that are in a biologically plausible range for Western blot data<sup>16,17</sup>. Exemplary results are shown in Fig. 3a, where resulting distributions are illustrated by box plots. While the variability of the resulting distributions of the GRCs is comparable to those of the inputs (Fig. 3a centre), we observe a much higher variability in the distributions of the LRCs (Fig. 3a right), which is mainly manifested in the number of outliers and the range covered by them. The complete set of distributions is given in the Supplementary material (Suppl. Figs S2–S4) and shows that these results are representative.

Driven by our analysis, we decided at that point to consider, besides standard measures for statistical dispersion such as interquartile ranges, also the amount of spread of the outliers, which is a measure for the degree of heavy-tailedness of an underlying distribution. Normalization of a signal obtained from a Western blot to a signal of a respective control experiment indeed corresponds to a transformation that may result in heavy-tailed distributions<sup>18</sup>. Since the tails of such distributions usually follow a power-law decay, the probability mass in the tails exceeds that of a Gaussian distribution, whose tails decay exponentially. As a consequence, samples from heavy-tailed distributions will contain more outliers which are spread over a larger range. A characteristic feature of heavy-tailed distributions is the fact that some or all moments do not exist. This severely impedes network reconstruction in our framework, since empirical estimators of moments are unstable due to the high occurrence of outliers. Empirical moments like the sample mean, the sample variance, or skewness and kurtosis, which are standard measures of asymmetry and tail-heaviness, do not provide meaningful estimates under these circumstances.

Thus, we decided to evaluate left and right medcouple (LMC and RMC)<sup>19</sup> as suitable measures of left and right tail weights. The medcouple function (see Methods) was proposed as an efficient measure of the asymmetry of a univariate continuous distribution alternative to the classical skewness estimator<sup>20</sup>. The medcouple applied to one single side of the distribution leads to LMC and RMC, which are monotonically increasing functions of tail-heaviness. They are robust to outliers, since they only depend on quantiles and hence are suitable for heavy-tailed distributions. LMC and RMC values are put into context by comparison with the respective values for a standard Gaussian and Cauchy distribution, which are 0.2 and 0.5, respectively.

LMC and RMC values for concentration measurements, GRCs and LRCs are depicted in Fig. 3b. LMCs and RMCs for the distributions of the measurement data and of the GRCs are comparable to those of a Gaussian distribution. Respective values for the distributions of the estimated LRCs are considerably larger, indicating that heavy-tailedness is mainly introduced by the transformation  $T_2$ . This increase might have severe consequences for network reconstruction, since it distorts estimation of moments of the LRCs such as the mean and the variance from samples. Evaluation of LMC and RMC values for the p53 test-bed model reveals similar results (Suppl. Fig. S5). Interestingly, MRA does not markedly affect the interquartile range (IQR) over the two transformations, which is a frequently used bulk-measure of variability (Suppl. Fig. S6).

We conclude that MRA amplifies the variability of the measurement noise in terms of degree of heavy-tailedness, while the IQR is not as much affected. Since heavy-tailedness is directly related to the occurrence of samples in the tails, which appear as outliers in the box plots, this impedes network reconstruction, as



**Figure 3.** Non-linear propagation of measurement error in the MAPK test-bed model. **(a)** The curves represent the dependencies of steady states (left), GRCs (centre) and LRCs (right) over the changing parameters  $p_j$ . Exemplary, on the left we show the distributions (boxplots) obtained from sampled noisy realizations of the steady state  $\bar{z}_2$  in the control experiment ( $p_2 = 1$ ) and in the 50% knockdown experiment of node 2 ( $p_2 = 0.5$ ). We generated  $n = 10,000$  realizations via Monte Carlo simulations from the noise model (10) with parameters  $\sigma_\eta = 0.1$  and  $\sigma_\varepsilon = 0.2$ . The ODE model with numerical values used for simulations is given in Suppl. S1. The variability in the measured steady states (left) translates into variability of the calculated GRCs (centre), which then translates into variability of the LRCs (right). Two sample points have been indicated by an orange and a grey cross and tracked during the transformations to illustrate consequences for network inference from individual samples. **(b)** Propagation of LMC and RMC values during the two-step transformation process reveals that heavy-tailedness is mainly introduced by the transformation  $T_2$ . Numerical values are given in Suppl. Table S1. **(c)** True network structure of the MAPK test-bed model as obtained via equation (4). As a comparison, also the two realizations of the LRCs  $r_{13}$  and  $r_{31}$  that result from the two tracked orange and grey sample points are shown.

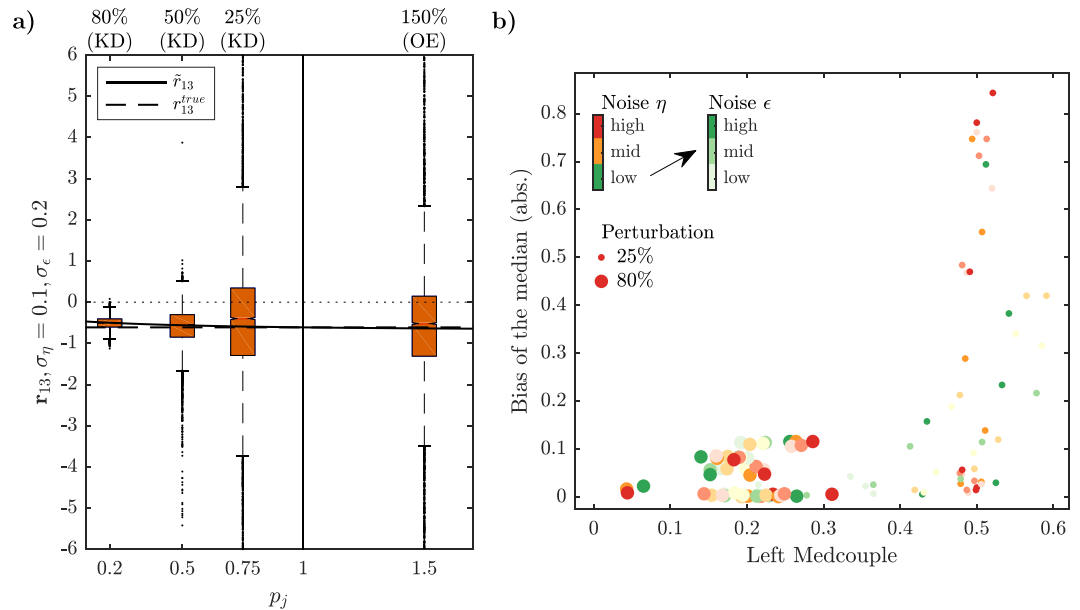
illustrated with two sample points indicated with orange and grey crosses and respective wrongly inferred network structures (Fig. 3c).

The question arises how we can optimize network reconstruction by influencing the distribution of the LRCs via experimental design and/or estimation procedures. In a first step we analyse how to best design the experiments regarding the choice of the perturbation strengths and the control strategy and subsequently investigate how to best handle multiple replicates.

**Large perturbations tend to improve the inference of pairwise node interactions.** Since the GRCs and LRCs are defined as derivatives, a precise approximation via finite differences theoretically requires infinitesimal small perturbations, which is not feasible in practice. Moreover, noise deteriorates estimation of derivatives particularly from small differences. The question arises whether we are able to define perturbation strengths that constitute a good trade-off. For the MAPK test-bed model we observe that the noise-free approximated solution for the LRCs  $\tilde{r}_{ij}$  is robust over a large range of perturbation parameters  $p_j$  and does not deviate much from the corresponding true value  $r_{ij}^{true}$  (see left of Fig. 3a and Suppl. Fig. S4). This is different for the p53 test-bed model (Suppl. Fig. S9) and might also not be the case for other systems, which we usually don't know a priori.

In order to answer our question, we compare the variability of the estimated LRCs resulting from different perturbation strengths. Therefore, we consider three knockdown experiments with downregulation of the 80%, 50% and 25% of the total protein concentrations with respect to the control experiment, and one overexpression experiment with 150% of total protein concentrations, resulting in a set of values for the perturbation strengths  $p_j \in \{0.2, 0.5, 0.75, 1.5\}$ .





**Figure 4.** Effects of different perturbation strengths on network reconstruction. **(a)** Boxplots of the estimated LRC  $r_{13}$  of the MAPK test-bed model, describing the negative feedback from ERK to Raf, for different perturbation strengths: 80%, 50%, 25% knockdowns (KD) and 150% overexpression (OE) of the total protein concentrations. **(b)** Absolute values of the bias of the median versus LMC values for the entire set of of LRC values obtained with large (80%) or small (25%) knockdown strengths of the total protein concentrations. These statistics are given for different noise levels  $\sigma_\eta \in \{0.05$  (green),  $0.1$  (yellow),  $0.2$  (red) $\}$  and  $\sigma_\epsilon \in \{0.1, 0.2, 0.5\}$  (indicated by increasing darkness).

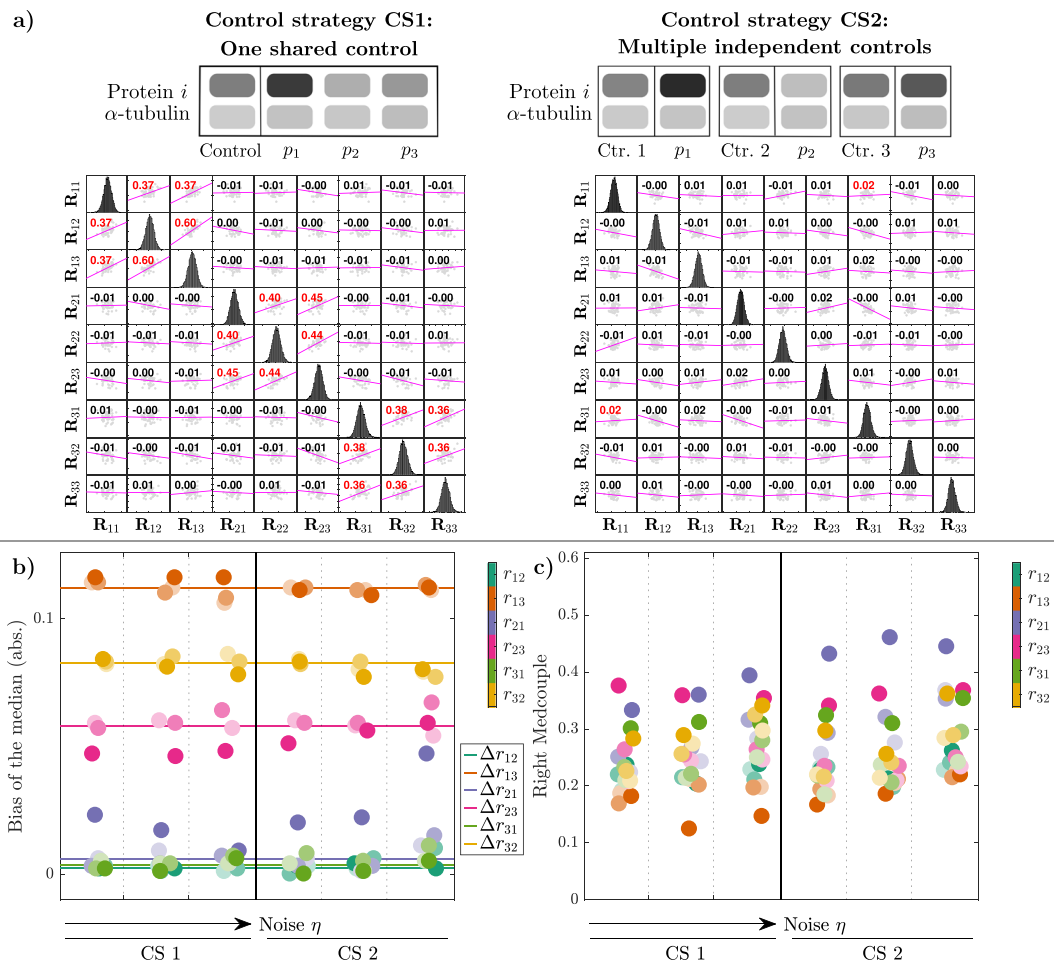
As can be seen in Fig. 4a and Suppl. Fig. S10, the distributions of the estimated coefficients differ significantly in the four scenarios. The spread of the estimated coefficients is smallest for  $p_j = 0.2$  and rapidly increases with decreasing perturbation strength, i.e. when  $p_j$  approaches one. The spread of the overexpression experiment is comparable to the 25% knockdown experiment, which is probably a result of the fact that we are in the saturated regime. We also observe a small and perturbation-dependent bias in the empirical estimate of the medians of all distributions.

To investigate the influence of the perturbation strength on accuracy and precision of the estimation more comprehensively, we collected values of the bias of the median and of the LMCs and RMCs for the 80% and the 25% knockdown experiments. Results are shown in Fig. 4b, where we have also visualized different noise levels  $\sigma_\eta$  and  $\sigma_\epsilon$  with different colours and corresponding different shades. We observe a bias-spread trade-off between large and small perturbations. A low bias and a low LMC can only be obtained with large perturbations (large dots), while small perturbations lead to higher LMC values.

Increasing noise levels affect the bias markedly only for the small perturbation (small dots), which is true for all coefficients (Suppl. Fig. S11). The influence of increasing noise levels on LMC and RMC values is visible but moderate in the 80% knockdown experiment, while in case of 25% knockdown a marked effect can only be seen for very small multiplicative noise  $\sigma_\eta$  (Suppl. Fig. S12). Intriguingly, in most of the cases these quantities behave non-monotonically with respect to noise. Increasing noise does not necessarily imply larger bias or medcouples, which is probably due to the non-linear transformations  $T_1$  and  $T_2$ . For the p53 test-bed model we observe similar trends (Suppl. Figs S13–S16), even though we observe a large bias of the median also for the large perturbation experiments here. From this analysis we conclude that larger perturbations are generally preferable, since they reduce the risk to infer erroneous network interactions.

**A simple control strategy is sufficient for the estimation of the LRCs.** The second component of the experiment design under investigation is the control strategy. Here we compare a single control for a node for all three perturbations (Fig. 5a left) versus individual controls for each perturbation (Fig. 5a right). The steady-state variable  $\tilde{x}_i^0$  of the control experiment appears in the GRCs  $\tilde{R}_{ij}$  of all perturbation experiments  $j = 1, \dots, N$  (equation (5)). Simulating the first control strategy thus translates into using the same realization of the random variable  $\tilde{z}_i^0$  to calculate the realizations  $R_{ij}$ ,  $j = 1, \dots, N$ , for fixed  $i$ , and results in block-wise positive correlations between the GRCs, as can be seen in the  $R_{ij}$  scatter plot matrix in Fig. 5a (left). These correlations disappear when performing multiple independent controls  $\tilde{z}_i^0$  for each perturbation experiment  $j$  (Fig. 5a right).

The choice of the control strategy determines the correlation among the coefficients  $R_{ij}$  with the same index  $i$ , but it is unclear whether it also has a marked effect on the LRCs  $r_{ij}$  and thus on network inference. In order to resolve this issue, we used the bias of the medians, the interquartile range (IQR), LMCs and RMCs as statistical measures of the distributions of the LRCs to compare the two control strategies. Since large perturbations have already turned out to be advantageous for MRA analysis, we simulated an 80% knockdown experiment and



**Figure 5.** Effects of two alternative control strategies. **(a)** The first strategy (CS1) considers one single control realization for the calculation of all samples  $R_{ij}$  with  $j = 1, 2, 3$ , while the second strategy (CS2) performs independent control experiments for each perturbation experiment. Corresponding correlations can be seen in the  $R_{ij}$  scatter plot matrices. **(b)** Absolute values of the bias of the medians of the resulting LRCs  $r_{ij}$  for both control strategies CS1 (left) and CS2 (right) in dependence of different noise levels  $\sigma_\eta \in \{0.05, 0.1, 0.2\}$  and  $\sigma_\varepsilon \in \{0.1, 0.2, 0.5\}$ . For every value of  $\sigma_\eta$ , i.e. for each specific column, the three shades of the LRCs correspond to the three (increasing) values of the standard deviation of the additive noise  $\sigma_\varepsilon$ . Lines indicate intrinsic bias values for each LRC. **(c)** Same illustration for the RMC values.

additionally also analysed the effect of increasing noise levels. As can be seen in Fig. 5b,c, we do not detect significant differences between the two control strategies with respect to the bias-spread trade-off. In fact, bias and RMCs behave similarly in the two cases. As before, we do not see a marked effect of increasing noise levels on the LRC distribution measures. These observations also generalize to the IQRs and the LMCs (Suppl. Fig. S18).

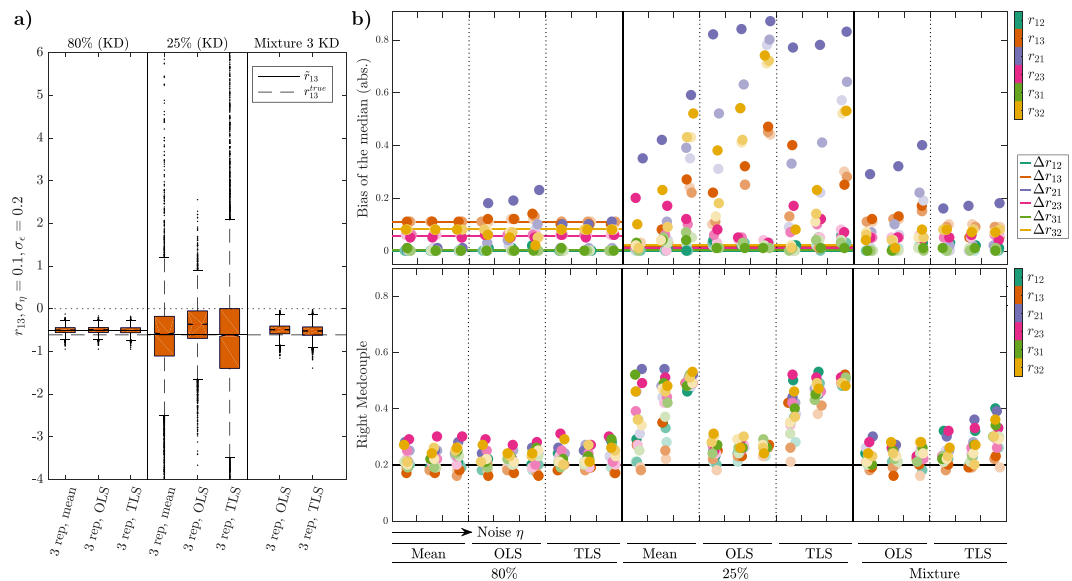
The horizontal lines shown in Fig. 5b represent the absolute values of the differences between the true LRCs and the LRCs resulting from the noise-free approximation, defined as  $\Delta r_{ij}$  (equation (9)). For realistic noise levels, as used here, the bias of the medians is centred around this corresponding intrinsic bias, showing that the main contribution to the bias is caused by the approximation (5) rather than by the measurement noise. As before, there is no clear monotonic relation visible between the considered statistics and the levels of additive and multiplicative noise, respectively. The p53 test-bed model behaves very similar in this analysis (Suppl. Figs S19 and S20).

Taken together, since we could not observe marked differences of the LRC statistics between the two control strategies in both models, we advise experimenters to use the first control strategy of taking a single control measurement for a node for all corresponding perturbations, since this requires less samples.

### Using MRA with replicate mean values tends to outperform linear regression techniques.

Generally, perturbation data contain several replicates of the same experiment. This raises the question of how to best handle these replicates during the MRA workflow. One solution is to calculate the mean over the replicates. Another, is the use of linear regression, for which several techniques have been suggested. The most common choice is to solve equation (7) by applying a least squares method, like Ordinary Least Squares (OLS) and Total Least Squares (TLS)<sup>8</sup> (see Methods). But whether regression, and if so which, is better than using the mean over all replicates remains unclear.





**Figure 6.** Effects of different estimation methods for the linear regression problem with multiple replicates. **(a)** Boxplots of the estimated LRC  $r_{13}$  of the MAPK model for two different perturbation strengths and different strategies to handle three replicates. First, the LRC is calculated by taking the mean values of the GRCs. Second, GRC replicates are taken individually into account and LRCs are obtained by solving OLS or TLS, respectively. The third column illustrates results from a mixture of measurements from three knockdown experiments with different perturbation strengths. **(b)** Absolute bias values of the estimated medians and RMC values for all LRCs and increasing levels of multiplicative and additive noise,  $\sigma_\eta \in \{0.05, 0.1, 0.2\}$  and  $\sigma_\epsilon \in \{0.1, 0.2, 0.5\}$ . For every value of  $\sigma_\eta$ , i.e. for a specific column, the three shades of the LRCs correspond to three (increasing) values of the standard deviation of the additive noise  $\sigma_\epsilon$ .

Therefore we aim to solve the question about which estimation method, combined with the proper experimental design and data normalization, allows the best results in the terms of accuracy, precision and robustness of the LRCs estimates.

We compare results obtained with three replicates, which is the typically required number in many biological studies. In our simulations we mimic replicates by drawing independent realizations  $z_i^{0,j}$ , providing different realizations of the GRCs  $R_{ij}$ . We considered the methods of taking the mean over the three obtained GRCs replicates and solving the linear regression problem (7); or determining GRC values for individual replicates and then applying either OLS or TLS from noisy values  $R_{ij}$ , delivering one estimate of the LRCs  $r_{ij}$ ,  $i, j = 1, \dots, N$ . Moreover, we consider yet another experimental approach, in which we take multiple sample data not by repeating the same perturbation experiment but by varying the perturbation strengths  $p_j$ ,  $j = 1, 2, 3$ . Our choice is to mix three realizations obtained using three different knockdown strengths: 80%, 50% and 25% KD of the total protein concentrations. The results of our analysis are summarized in Fig. 6.

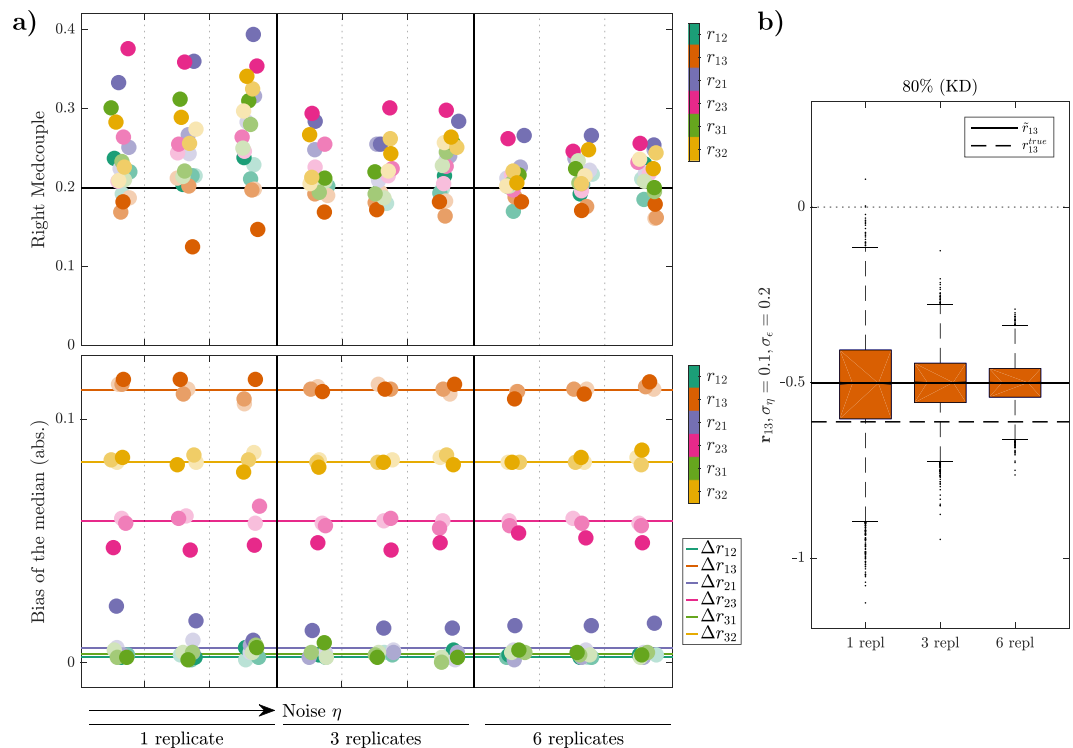
These results confirm that the experiment design with the large perturbation is superior compared to small perturbations or a combination of different perturbation strengths (see also Suppl. Figs S21 and S22). The considered measures for dispersion (LMC, RMC and IQR) are low and robust to noise for all three estimation approaches.

Interestingly, the mixture approach also delivers good results in terms of bias-spread trade-off (right part of Fig. 6, Suppl. Figs S21 and S22). As before, the OLS method results in a larger bias, but the dispersion measures are more sensitive to increasing noise if using TLS.

We can confirm that the experimental approach with the small perturbation strength delivers unsatisfactory results, leading to a high risk to reconstruct an erroneous network structure. Compared to the other two experimental designs, the bias is much larger and sensitive to noise with all three estimation methods: This holds especially true for the three non-zero coefficients  $r_{21}$ ,  $r_{32}$  and  $r_{13}$  (central part of Fig. 6b and Suppl. Fig. S22). The measures for dispersion are low for all coefficients only for very low noise and if using OLS.

Summarizing our results, we obtained the best estimation results in terms of accuracy, precision and robustness to noise by performing large perturbations and a simple control strategy. In terms of efficiency we recommend to use the simplest estimation method, which means to solve the regression problem (7) with the GRC means. In comparison, the mixture approach seems to be suboptimal in terms of bias-spread trade-off, but it might be beneficial for systems with higher non-linearities, as discussed in a later Subsection.

**Replicates increase precision, but not accuracy.** The choice of the number of replicates is another important question for experimental design because of the trade-off between the experimental effort and cost, and the quality of the inferred results<sup>21</sup>. We address this issue by investigating how much the estimation of the LRCs is improved by increasing the number of replicates. We compare results obtained with one, three and six replicates. For each Monte Carlo run, we proceed by taking the mean value of these GRCs to further calculate one



**Figure 7.** Effects of different numbers of replicates on network reconstruction. **(a)** RMCs and absolute bias values of the medians of the estimated distributions for the LRCs of the MAPK model for one, three and six replicates and different noise levels in the 80% KD perturbation experiments. Noise levels have been set to  $\sigma_\eta = \{0.05, 0.1, 0.2\}$  (different columns) and  $\sigma_\varepsilon = \{0.1, 0.2, 0.5\}$  (coded in different shades). The black line and the coloured lines indicate the RMC value for a normal distribution and the intrinsic bias values for each LRC, respectively. **(b)** Exemplary box plots of the LRC  $r_{13}$  for different numbers of replicates.

realization  $r_{ij}$  of the LRCs, which we have seen to be the most efficient estimation method, combined with large perturbations and the simple control strategy.

Results are depicted in Fig. 7. As expected, the precision of the estimation increases with the number of replicates, for both test-bed models. This manifests in a decrease of the considered measures for statistical dispersion, which are RMC and LMC values and the IQR, for all coefficients and noise levels and both test-bed models (Fig. 7a top, Suppl. Figs S26 and S28). In particular, RMC and LMC values converge to the value 0.2 of the standard Gaussian distribution. This effect is robust against increased multiplicative noise levels  $\eta$ .

In contrast, the biases in the medians are neither much affected by the number of replicates nor by the level of multiplicative noise, as can be seen in Fig. 7a (bottom) and Suppl. Fig. S28. In some cases increased additive noise  $\varepsilon$  (indicated by a darker shade of the coloured dots) leads to a larger bias, but not in a monotonic manner. As before, the medians rather coincide with the noise-free approximated values  $\tilde{r}_{ij}$  (see also Suppl. Figs S25 and S27), whose deviations from the true values result from the choice of a large perturbation, showing that the bias in the medians is again dominated by the error of the approximation (5).

Summarizing, increasing the number of replicates reduces the dispersion of the distribution and therefore increases precision, but the bias cannot be eliminated, which restricts the accuracy of the estimates. We consider three replicates to be a good bias-spread trade-off, since all RMC values decrease below 0.3 when going from a single measurement to three replicates, while the decrease is much less pronounced when going from three to six replicates. Thus we recommend to use at least three replicates, and to include more depending on how much experimental effort is acceptable.

**Non-linearity induces bias, but large perturbations are still required for precision.** In the MAPK model, the steady states show an approximately linear behaviour in dependence of the perturbation strengths in all cases (see Suppl. Fig. S1a), suggesting that the linear approximations (5) and (8) do not induce unduly large errors even for large perturbations. This was confirmed by our simulation results. When applying the MRA in practice, however, the course of the steady states of the system for varying perturbation strengths is not known, and it could also be highly non-linear. Do our recommendations and guidelines for an optimal performance of network reconstruction via MRA still hold true for such cases? In order to address this question, we used the p53 test-bed model as an example of a system whose steady states are non-linear functions of perturbation strengths (see Suppl. Fig. S1b). In this case the approximated LRCs  $\tilde{r}_{ij}$  are sensitive to the choice of the perturbation strength and it is not clear a priori whether they are a good approximation of the true values  $r_{ij}^{true}$  (compare Suppl. Figs S4 and S9).

We applied our MRA workflow to this test-bed model and performed the same analysis as before with the MAPK model. Summarizing, the results show that the most critical part is indeed the appearance of a large bias in the median of the distributions of the estimated  $r_{ij}$  if applying large perturbation experiments (see Suppl. Figs S23 (left parts) and S24a). This effect is related to the intrinsic bias  $\Delta r_{ij}(p_j)$  and cannot be reduced by an increase in the number of replicates (see Suppl. Fig. S28a).

Nevertheless, the goal is to estimate the correct network structure, and therefore it is important to minimize the dispersion of the distributions of the estimated  $r_{ij}$ . This holds especially if the intrinsic bias is significant for some of the LRCs, which is the case in the p53 example (Suppl. Fig. S23). In such cases it is necessary that the approximated LRCs  $\tilde{r}_{ij}$  have the same sign as the true values, leading to qualitatively correctly estimated interactions. The trend of the spread of the estimated distributions shows that in general the lowest dispersion is still obtained with the largest perturbation experiment, in a similar way for all three computational approaches (see Suppl. Fig. S24b–d). In all these cases this behaviour is robust to increasing noise levels.

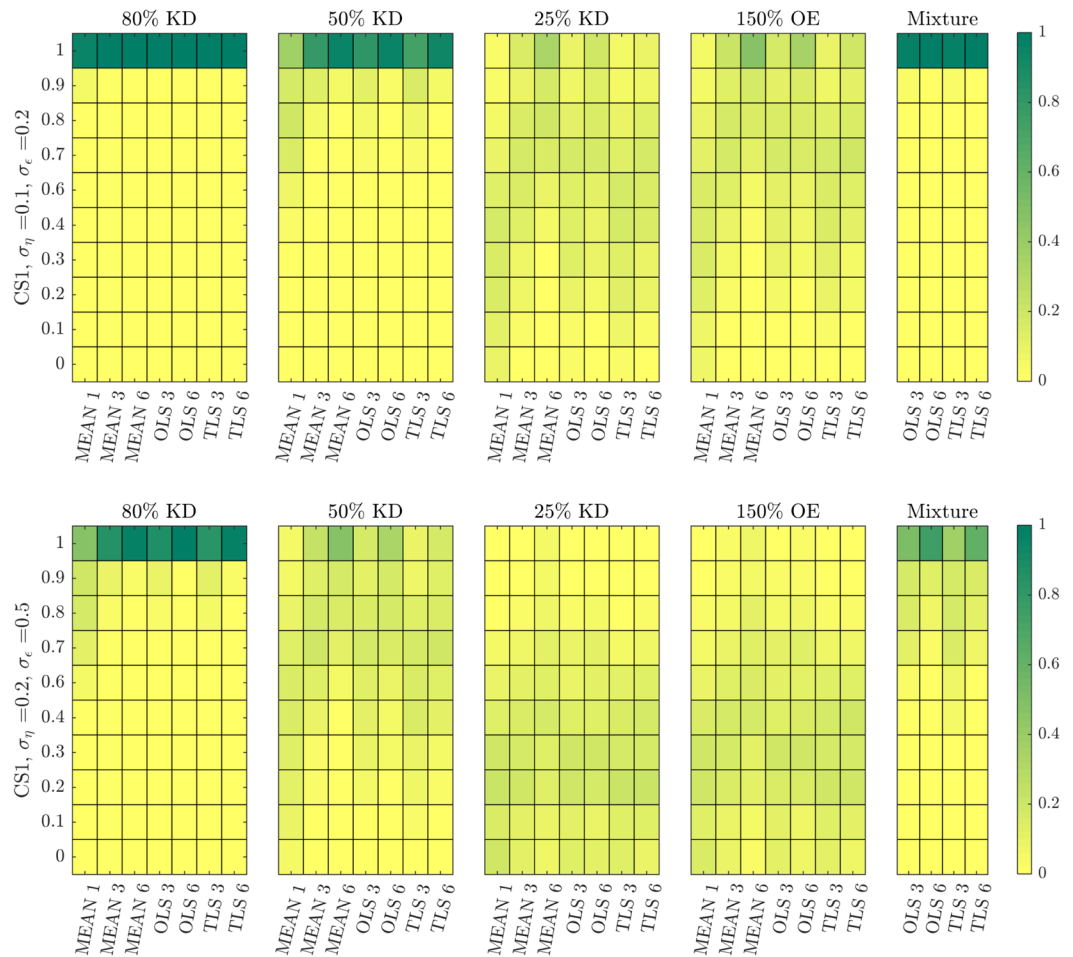
From these results we conclude that, due to the noise sensitivity, larger perturbations are generally still preferable, even for highly non-linear systems, since they reduce the risk to infer erroneous network interactions.

**Performance evaluation on the level of discrete network interactions corroborates our quantitative results.** So far, we have investigated the influence of different experiment designs, estimation methods and noise levels on the statistical properties of the estimated LRCs. We have in particular focused on the bias of the median and on LMC and RMC values as measures for accuracy and precision of the individual estimates. In a final analysis step we transfer these results onto network inference, where the set of inferred LRCs is used to decide upon the network structure. The simplest way to do this is to arrange all LRCs according to their absolute value and to define a threshold for an interaction to be present or not. Sensitivity and specificity can then be calculated for an inferred network by a comparison with the true or a reference network. Doing this with varying threshold values, the Area Under the Curve (AUC) value is then an aggregated measure for the overall performance of the inference method independent of the threshold parameter. For such an analysis, it is not sufficient to look at each LRC separately. Here we applied an assessment method proposed in<sup>9</sup>, which is similar to a receiver-operator analysis, but also takes the signs of the inferred interactions into account. Depending on the percentage of correctly identified interactions, a normalized measure for the fit quality is assigned to an inferred network structure (see Fig. 1c), which is 0 in the worst and 1 in the best case. This overall measure for fit quality was determined for the different scenarios considered before and the distribution of this measure was investigated by sampling  $n = 10,000$  network structures for each setting.

Results are shown in Fig. 8. Here, colour-coded empirical probability distributions of the discrete set of fit-quality values are shown for different settings. The first and second row depict results for intermediate and high noise levels, respectively. Different computational strategies and different perturbation strengths are compared. It can be seen that network inference works quite well in the 80% KD experiments for both noise levels and almost independent from the number of replicates and from the strategy to handle replicates. For intermediate noise levels, also the mixture method and the 50% KD perform very well, but are more sensitive to increasing noise levels. Both 25% KD and 150% OE perform worse in all scenarios. It can also be seen that results when considering replicates are not markedly different across almost all scenarios if averaging over the GRCs or when using OLS or TLS. We also compared these statistics across the two control strategies CS1 and CS2 (e.g. compare Suppl. Fig. S29a–d), which shows that the simple control strategy (CS1) is sufficient and there is no need to evaluate multiple independent control samples for each perturbation experiment. Similar results were obtained for the p53 model (Suppl. Fig. S30). Taken together, these results further confirm the conclusions drawn from the quantitative analyses in the previous sections: Firstly, due to noise, large perturbation are preferable, even for systems with a high degree of non-linearity. Secondly, it suffices to use a simple experimental strategy with one unperturbed control as reference for all perturbed conditions.

## Discussion

In this paper, we went through a comprehensive analysis of the effects of different experimental and estimation approaches for MRA on the goodness of network inference from noisy data, in terms of accuracy, precision and robustness. Our results led to some interesting findings. First, Monte Carlo simulations of concentration measurements with a realistic noise model for Western blot data clearly show a strong increase of heavy-tailedness, quantified in terms of LMC and RMC values, in the transformation from the GRCs to the LRCs, while respective values for concentration measurements and GRCs are quite similar (Fig. 3). This is a very relevant result, since heavy-tailedness deteriorates estimation of moments from samples, inducing as consequence a high risk of wrong outcomes for the network inference problem. In extreme cases, i.e. when certain moments are not defined, a stable estimation is not possible, even for large sample sizes. At least, this implies that concentrations and GRCs can be estimated more accurately from concentration measurements than LRCs. Second, for both test-bed models large perturbations are more favourable than smaller ones. Estimation of LRCs and hence network inference is much more accurate when using large perturbations. This is a non-trivial result, since estimation of the LRCs is done via a finite difference approximation of the GRCs in the MRA workflow, for which small differences would be beneficial in the absence of noise, since large differences imply an intrinsic bias. In particular, Fig. 4 shows a clear clustering of inferred LRCs according to the perturbation strengths: Results from 80% knockdown simulations show small biases and small LMC values, while 25% knockdowns show a much higher spread of bias values and consistently high LMC values. This leads to the clear advice to use large perturbations in the MRA workflow, even when the underlying model system features a considerable amount of non-linearity. Furthermore, regarding experiment design in the number of controls, our results indicate that a single control for different perturbation experiments, as often applied in practice, is sufficient (Fig. 5). While a single control causes correlations between GRCs, GRC marginals are not much affected. In particular, there is not much difference in the bias and RMC



**Figure 8.** Performance evaluation of all MRA settings for network inference obtained with the MAPK test-bed model. Empirical distributions of fit quality of the inferred networks for different experimental designs and computational strategies for intermediate (top row) and high (bottom row) noise levels.

values of inferred LRC values among the two control strategies. The bias values are dominated by the intrinsic bias, and this is true for all tested noise levels. Regarding the required number of measurements and the estimation method, we advice to use the mean of at least three replicates. The spread of inferred LRCs decreases monotonically with the number of replicates, while the bias of the median is dominated by the intrinsic bias (Figs 6 and 7). Finally, our conclusions also hold true for the overall network inference problem, as evaluated in this study by a normalized quality measure for a classification problem (Fig. 8). Our results in particular show that learning the network topology is possible with very high accuracy also for high noise levels in our setting with the 80% knockdown experiments and few replicates.

As pointed out in the introduction, the effect of noise and variability in the data used for MRA network inference had already been an issue in earlier studies<sup>1,8</sup>. In a later work<sup>22</sup>, the authors developed an advanced version of MRA, combining it with a Bayesian model selection algorithm, relaxing also the restriction of required number of perturbation experiments to equal the number of nodes of the network. However, none of these are comparable in considering propagation of noise from concentration measurements via GRCs and LRCs to network topology inference in a consistent stochastic framework with realistic noise assumptions. These studies also use Monte Carlo techniques, but start with i.i.d. normal distributions directly on the GRC values, and also completely neglect the effects of heavy-tailedness. They are also lacking concrete recommendations for experimental design and computational methodology regarding MRA based network inference.

As with all inference methods, our methodology has some limitations. The MRA framework itself assumes a continuous functional dependence between perturbation parameters and steady states of the system. This excludes for instance systems which exhibit bifurcations, as they appear for example in positive feedback systems which exhibit multi-stability. For those systems, the theory only holds as long as the perturbation does not induce a switch of the system to a different fixed point branch. It might be difficult to decide whether this is indeed the case in real settings, where the underlying dynamical system is not known. Furthermore, there might be potential for improvement regarding methodology to solve the regression problem to calculate the LRCs from the GRCs. Methods like feasible generalized least squares or Maximum Likelihood estimation might be beneficial in this respect. Finally, evaluation of our findings and recommendations in a setting with real experimental data is an open issue for the future.

## Methods

**Medcouple.** Given a set of  $n$  independent samples  $\{x_1, \dots, x_n\}$  from a continuous univariate distribution, with median  $m_n$ , the medcouple is defined as

$$MC = \text{med}_{x_i \leq m_n \leq x_j} h(x_i, x_j), \quad \text{with} \quad h(x_i, x_j) = \frac{(x_j - m_n) - (m_n - x_i)}{x_j - x_i}, \quad \forall x_j \neq x_i.$$

The kernel function  $h(x_i, x_j)$  measures the (normalized) difference between the distances of  $x_i$  and  $x_j$  to the median. The medcouple represents a robust measure of the asymmetry of a distribution, which can be computed also for distributions without finite moments, which is not the case for the classical skewness coefficient<sup>20</sup>. As robust measure of tail weight, the authors propose to apply the medcouple only to one single side of the distribution, leading to Left Medcouple (LMC) and Right Medcouple (RMC)<sup>19</sup>:

$$\text{LMC} = -\text{MC}(x < m_n) \quad \text{and} \quad \text{RMC} = \text{MC}(x > m_n).$$

The calculation of such quantities for all datasets in our study was performed with the MATLAB toolbox LIBRA<sup>23</sup>, developed by the same authors, which can be downloaded from <https://wis.kuleuven.be/stat/robust/LIBRA/LIBRA-home>.

**MRA estimation methods.** For the estimation problem, we have to solve equation (7),  $y = A \cdot x$ , which is a linear regression model, in the unknown variable  $x$ . Assuming no error in the regression variables, i.e. in the entries of the matrix  $A$ , and i.i.d. normal errors in the variable  $y$ , we obtain the well known ordinary least squares (OLS) solution, given in equation (8). However, this assumption is wrong, since the entries in the matrix  $A$  are also affected by noise, being samples of GRCs. One option is to consider error-in-variables models, such as total least squares (TLS), whose computation requires singular value decomposition and is presented in<sup>8</sup>.

## Data Availability

The datasets generated during and/or analysed in this study, as well as the Matlab code and functions and the SBML versions of the models, can be downloaded from <https://zenodo.org/record/1434792#.W6pBmC-B2MI>; or using the <https://doi.org/10.5281/zenodo.1434792>. The MAPK and p53 models are available in standard SBML format from the EMBL-EBI BioModels repository under accession numbers MODEL1809260001 and MODEL1809260002, respectively.

## References

- Santos, S. D. M., Verwee, P. J. & Bastiaens, P. I. H. Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nat. Cell Biol.* **9**, 324–330 (2007).
- Stelnic-Klotz, I. *et al.* Reverse engineering a hierarchical regulatory network downstream of oncogenic KRAS. *Mol. Syst. Biol.* **8**, 601 (2012).
- Gong, C., Zhang, Y., Shankaran, H. & Resat, H. Integrated analysis reveals that STAT3 is central to the crosstalk between HER/ErbB receptor signaling pathways in human mammary epithelial cells. *Mol. Biosyst.* **11**, 146–158 (2015).
- Speth, Z., Islam, T., Banerjee, K. & Resat, H. EGFR signaling pathways are wired differently in normal 184a115 human mammary epithelial and MDA-MB-231 breast cancer cells. *J. Cell Commun. Signal.* **11**, 341–356 (2017).
- Kholodenko, B. N. *et al.* Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. *Proc. Natl. Acad. Sci.* **99**, 12841–12846 (2002).
- Santra, T., Rukhlenko, O., Zhernovkov, V. & Kholodenko, B. N. Reconstructing static and dynamic models of signaling pathways using Modular Response Analysis. *Curr. Opin. Syst. Biol.* **9**, 11–21 (2018).
- Kang, T., Moore, R., Li, Y., Sontag, E. & Bleris, L. Discriminating direct and indirect connectivities in biological networks. *Proc. Natl. Acad. Sci.* **112**, 12893–12898 (2015).
- Andrec, M., Kholodenko, B. N., Levy, R. M. & Sontag, E. Inference of signaling and gene regulatory networks by steady-state perturbation experiments: structure and accuracy. *J. Theor. Biol.* **232**, 427–441 (2005).
- Bansal, M., Della Gatta, G. & Di Bernardo, D. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinforma.* **22**, 815–822 (2006).
- Kholodenko, B. N., Hancock, J. F. & Kolch, W. Signalling ballet in space and time. *Nat. Rev. Mol. Cell Biol.* **11**, 414 (2010).
- Kholodenko, B. N. Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.* **7**, 165 (2006).
- Purvis, J. E. *et al.* p53 dynamics control cell fate. *Sci.* **336**, 1440–1444 (2012).
- Fey, D., Kuehn, A. & Kholodenko, B. N. On the personalised modelling of cancer signalling. *IFAC-PapersOnLine* **49**, 312–317 (2016).
- Kholodenko, B. N. & Sontag, E. Determination of functional network structure from local parameter dependence data. *Web Arch. arXiv:physics/0205003* (2002).
- Kreutz, C. *et al.* An error model for protein quantification. *Bioinforma.* **23**, 2747–2753 (2007).
- Schilling, M. *et al.* Computational processing and error reduction strategies for standardized quantitative data in biological networks. *FEBS J.* **272**, 6400–6411 (2005).
- Taylor, S. & Posch, A. The design of a quantitative western blot experiment. *BioMed Res. Int.* **2014**, 361590 (2014).
- Thomaseth, C. & Radde, N. Normalization of western blot data affects the statistics of estimators. *IFAC-PapersOnLine* **49**, 56–62 (2016).
- Brys, G., Hubert, M. & Struyf, A. Robust measures of tail weight. *Comp. Stat. & Data Anal.* **50**, 733–759 (2006).
- Brys, G., Hubert, M. & Struyf, A. A robust measure of skewness. *Comp. & Graph. Stat.* **13**, 996–1017 (2004).
- Blainey, P., Krzywinsky, M. & Altman, N. Replication. *Nat. Methods* **11**, 879–880 (2014).
- Santra, T., Kolch, W. & Kholodenko, B. N. Integrating Bayesian variable selection with Modular Response Analysis to infer biochemical network topology. *BMC Syst. Biol.* **7**, 57 (2013).
- Verboven, S. & Hubert, M. Matlab library LIBRA. *Wiley Interdisciplinary Rev. Comput. Stat.* **2**, 509–515 (2010).



## Acknowledgements

C.T. and N.E.R. acknowledge funding from German Research Foundation via the Cluster of Excellence in Simulation Technology (EXC310/2). B.N.K. acknowledges support from the EU H2020 SmartNanoTox (Grant agreement No. 686098) and EU H2020 NanoCommons (Grant Agreement No. 731032). D.F. acknowledges support from EU H2020 COLOSSUS (Grant Agreement No. 754923). O.S.R. acknowledges support from the EU H2020 MSCA-IF SAMNets (Grant agreement No. 750688).

## Author Contributions

N.E.R. and B.N.K. designed the study together with C.T. and D.F. Simulations and implementation of computational methodology was done by C.T. with input from D.F. T.S. and O.S.R. provided additional support in the development of the methodology. Results were discussed and put into context among all authors. N.E.R. and B.N.K. supervised the project. The manuscript was mainly written by C.T., D.F., N.E.R. and B.N.K., with input from all authors. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-34353-3>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018