Supplement to

A scalable moment-closure approximation for large-scale biochemical reaction networks

Atefeh Kazeroonian [1,2,3,*], Fabian J. Theis [1,2] and Jan Hasenauer [1,2]

[1]Helmholtz Zentrum München - German Research Center for Environmental Health, Institute of Computational Biology, 85764 Neuherberg, Germany

[2]Technische Universität München, Center for Mathematics, Chair of Mathematical Modeling of Biological Systems, 85748 Garching, Germany.

[3]Technische Universität München, Fakultät für Medizin, Institut für Medizinische Mikrobiologie, Immunologie und Hygiene, 81675 München, Germany.

[*]To whom correspondence should be addressed.

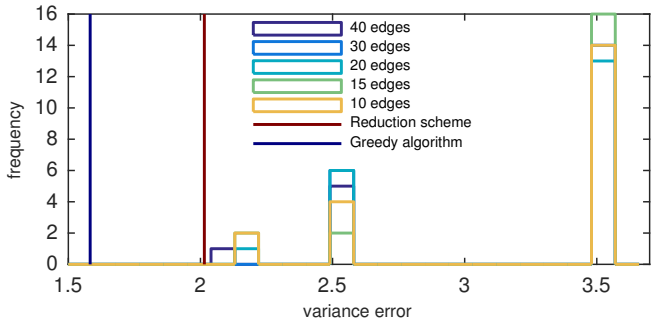
1 Comparison of the s2MA with structure-based, random and greedy selection of covariances

To corroborate our hypothesis that a structure-based selection of covariances is appropriate, we compared the proposed selection scheme to random selection for the simulation of the chain of monomolecular reactions (Supplement Figure 3). For random selection, the index set $I^{(\delta)}$ in Eq. (3) was defined as the union of index pairs corresponding to (1) the variances, $\{(i,i)|i=1,\dots,n\}$, and (2) a random sample drawn from $\{(i,j)\in\{1,\dots,n\}^2|i<j\}$ without replacement. For various numbers of covariances, we sampled the distribution of the error in the variances,

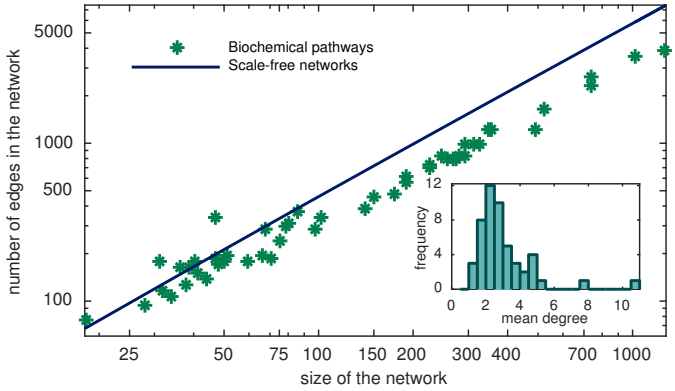
$$\sum_{i=1}^{n}\int_0^T \left(C_{ii}^{\text{rand}}(t)-C_{ii}^{\text{2MA}}(t)\right)^2 dt,$$

in which $C_{ii}^{\text{rand}}(t)$ denotes the time-dependent variance of species i calculated for a random set of covariances. The comparison with the error for the s2MA-1 showed that the structure-based selection achieved a lower approximation error than random selection, even if random selection was allowed to describe a larger number of the covariances (Supplement Figure 1). This implies that the direct dependency, as defined in the *Approach* section, is a good proxy for the relevance of a covariance for a good approximation.

To assess the sub-optimality of the structure-based selection of covariances, we compared it to a greedy approach. The greedy approach started with an ODE which merely includes the evolution equations for means and variances, i.e., Eq. (3) with $I^{(\delta)}=\{(i,i)|i=1,\dots,n\}$, and sequentially added one covariance. In each iteration, all possible choices for the additional covariance were considered and the covariance which resulted in the strongest decrease of the approximation error was included. This procedure was repeated until the ODE had the same size as the s2MA-1. The final model derived using this greedy approach possessed – as expected – a slightly lower approximation error than the s2MA-1 (Supplement Figure 1). The improvement of approximation accuracy, however, came with a substantial computational burden, which might not be feasible for large-scale biochemical reaction networks. Furthermore, the greedy approach is simulation-based and results generally depend on parameter values.



Supplement Figure 1: **Comparison of structure-based covariance selection, random selection and the greedy approach for the chain of monomolecular reactions ($n = 10$).** The integrated error in the variance is shown for the s2MA-1 and reduced MA with randomly/greedy-based selected sets of covariances. For random selection, the frequencies of 20 samples are depicted for various numbers of sample covariances.



Supplement Figure 2: **Number of edges in the simulated pathways.** The number of edges in the simulated pathways is compared to the number of edges in scale-free networks of the same sizes. Parameter γ in Section 4.1.1 is set to 2 to calculate the upper bound on the number of edges in scale-free networks. (inset) The distribution of the average degree in the simulated pathways.

2 Comparison of the published biochemical reaction networks to scale-free networks of same sizes

We calculated the size of the s2MA-1 (that is the number of edges in the network) for scale-free networks of the same size as the studied pathways. Supplement Figure 2 shows that for large networks, the s2MA-1 of scale-free networks is larger than the s2MA-1 of the studied biological pathways. These results suggest that the scale-free assumption can provide a safe upper bound for the connectivities/degree distribution in biochemical reaction networks. Also, to verify the local connectivities assumption, we calculated the average degree of a node in the pathways. Supplement Figure 2 (inset) illustrates that, independently of the size of the network, the average degree hardly exceeds 10.

3 List of published pathways

The Supplement Table 1 provides the list of published biochemical networks that are used for the scalability analysis of the s2MA.

4 Network motifs

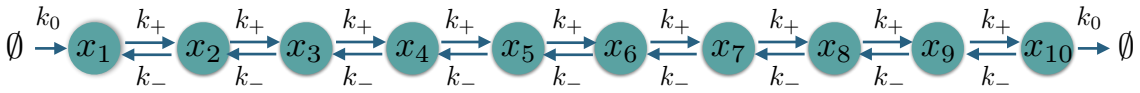
Illustrations of the considered *chain of monomolecular reactions* and *sequence of bimolecular reactions with a hub* are provided by Supplement Figures 3 and 4, respectively. The parameter values and initial conditions used for the simulation are listed in Supplement Tables 2 and 3.

References

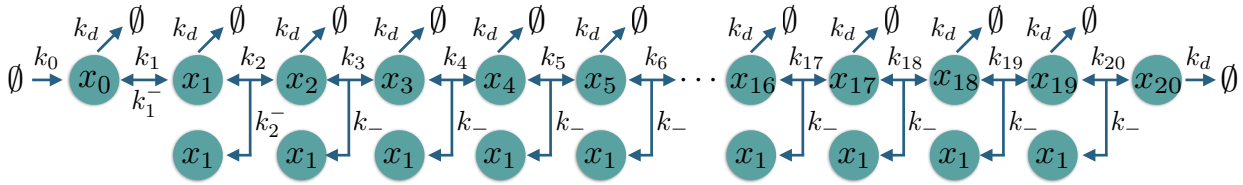
Bachmann, J., Raue, A., Schilling, M., Böhm, M. E., Kreutz, C., Kaschek, D., Busch, H., Gretz, N., Lehmann, W. D., Timmer, J., and Klingmüller, U. (2011). Division of labor by dual feedback regulators controls JAK2/STAT5 signaling over broad ligand range. *Mol. Syst. Biol.*, **516**(7).

Supplement Table 1: Published models used for the scalability analysis of s2MA.

	Name	**Identifier**	**Source**
1	Androgen receptor (AR) signalling pathway	NetPath_2	NetPath database
2	B Cell Receptor (BCR) signalling pathway	NetPath_12	NetPath database
3	Brain-derived neurotrophic factor (BDNF) signalling pathway	NetPath_76	NetPath database
4	Corticotropin-realising hormone (CRH) signalling pathway	NetPath_129	NetPath database
5	Epidermal growth factor receptor (EGFR1) signalling pathway	NetPath_4	NetPath database
6	Fibroblast growth factor-1 (FGF1) signalling pathway	NetPath_134	NetPath database
7	Gastrin signalling pathway	NetPath_154	NetPath database
8	Hedgehog signalling pathway	NetPath_10	NetPath database
9	Interleukin-2 (IL-2) signalling pathway	NetPath_14	NetPath database
10	Interleukin-3 (IL-3) signalling pathway	NetPath_15	NetPath database
11	Interleukin-4 (IL-4) signalling pathway	NetPath_16	NetPath database
12	Interleukin-5 (IL-5) signalling pathway	NetPath_17	NetPath database
13	Interleukin-6 (IL-6) signalling pathway	NetPath_18	NetPath database
14	Interleukin-7 (IL-7) signalling pathway	NetPath_19	NetPath database
15	Interleukin-9 (IL-9) signalling pathway	NetPath_20	NetPath database
16	Interleukin-10 (IL-10) signalling pathway	NetPath_132	NetPath database
17	Interleukin-11 (IL-11) signalling pathway	NetPath_147	NetPath database
18	Kit Receptor signalling pathway	NetPath_6	NetPath database
19	Leptin signalling pathway	NetPath_22	NetPath database
20	Notch signalling pathway	NetPath_3	NetPath database
21	Prolactin signalling pathway	NetPath_56	NetPath database
22	Receptor activator of nuclear factor kappa-B ligand (RANKL) signalling pathway	NetPath_21	NetPath database
23	T Cell Receptor (TCR) signalling pathway	NetPath_11	NetPath database
24	Transforming growth factor beta (TGF-beta) receptor signalling pathway	NetPath_7	NetPath database
25	Tumor necrosis factor (TNF) alpha signalling pathway	NetPath_9	NetPath database
26	Thyroid-stimulating hormone (TSH) signalling pathway	NetPath_23	NetPath database
27	Thymic stromal lymphopoietin (TSLP) signalling pathway	NetPath_24	NetPath database
28	TIE2/TEK signalling pathway	NetPath_138	NetPath database
29	Wnt signalling pathway	NetPath_8	NetPath database
30	Carbon metabolism	BIOMD0000000051	Biomodels database
31	E. coli metabolic adaptation	BIOMD0000000244	Biomodels database
32	Influenza virus replication	BIOMD0000000463	Biomodels database
33	TNF signalling network	BIOMD0000000407	Biomodels database
34	Yeast pheromone pathway	BIOMD0000000032	Biomodels database
35	Degradation of beta-catenin by the destruction complex	R-HSA-195253.1	Reactome database
36	DNA replication	R-HSA-69306	Reactome database
37	Interferon gamma signalling pathway	R-HSA-877300	Reactome database
38	Interferon alpha/beta signalling pathway	R-HSA-909733	Reactome database
39	Cholesterol biosynthesis	R-HSA-191273	Reactome database
40	DAG and IP3 signalling	R-HSA-1489509	Reactome database
41	Growth hormone receptor signalling	R-HSA-982772	Reactome database
42	Inositol phosphate metabolism	R-HSA-1483249	Reactome database
43	Integrin alphaIIb beta3 signalling	R-HSA-354192	Reactome database
44	ISG15 antiviral mechanism	R-HSA-1169408	Reactome database
45	Meiotic recombination	R-HSA-912446	Reactome database
46	Peroxisomal lipid metabolism	R-HSA-390918	Reactome database
47	PIP3 activates AKT signalling	R-HSA-1257604	Reactome database
48	RAF-independent MAPK1/3 activation	R-HSA-112409	Reactome database
49	RAF/MAP kinase cascade	R-HSA-5673001	Reactome database
50	JAK2/STAT5 signalling pathway	-	Bachmann *et al.* (2011)



Supplement Figure 3: **The schematic of the chain of monomolecular reactions**



Supplement Figure 4: **Schematic of the sequence of bimolecular reactions with a hub.**

Supplement Table 2: Parameter values and initial conditions used in the simulation of the chain of monomolecular reactions.

k_0	k_+	k_-	$[x_1](0)$	$[x_2](0)$	$[x_3](0)$	$[x_4](0)$	$[x_5](0)$	$[x_6](0)$	$[x_7](0)$	$[x_8](0)$	$[x_9](0)$	$[x_{10}](0)$
0.5	1	0.2	10	0	0	0	0	0	0	0	0	0

Supplement Table 3: Parameter values and initial conditions used in the simulation of the sequence of bimolecular reactions with a hub.

k_0	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	k_{11}	k_{12}	k_{13}	k_{14}
100	5	1	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.2	0.21	0.22	0.23

k_{15}	k_{16}	k_{17}	k_{18}	k_{19}	k_{20}	k_1^-	k_2^-	k_-	k_d	$[x_0](0)$	$[x_1](0)$ to $[x_{29}](0)$
0.24	0.25	0.26	0.27	0.28	0.29	0.1	0.01	0.001	0.001	50	0