Supplementary Information for

Practical sampling of constraint-based models: Optimized thinning boosts CHRR performance

Johann F. Jadebeck 1,2, Wolfgang Wiechert 1,2 and Katharina Nöh 1,*

1 Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, 52425 Jülich, Germany
2 Computational Systems Biotechnology (AVT.CSB), RWTH Aachen University, 52062 Aachen, Germany

*Correspondence: k.noe@fz-juelich.de
## 1 Benchmark models

Table A: **Benchmark models** for tuning and validation of the thinning guideline. Overview of models and simplices and their usage. The number of constraints gives the number of non-redundant inequalities after preprocessing with PolyRound \cite{1} as described in the main text (cf. Sec. Materials and models).

<table>
<thead>
<tr>
<th>Model</th>
<th>Type</th>
<th># reactions</th>
<th># effective polytope dimensions $d$</th>
<th>#constraints $n_{in}$</th>
<th>Usage</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>e_coh_core</td>
<td>GEM</td>
<td>95</td>
<td>24</td>
<td>36</td>
<td>Training</td>
<td>\cite{2}</td>
</tr>
<tr>
<td>iS312</td>
<td>GEM</td>
<td>519</td>
<td>46</td>
<td>137</td>
<td>Training</td>
<td>\cite{3}</td>
</tr>
<tr>
<td>iAB_RBC_283</td>
<td>GEM</td>
<td>469</td>
<td>130</td>
<td>188</td>
<td>Training</td>
<td>\cite{4}</td>
</tr>
<tr>
<td>iND750</td>
<td>GEM</td>
<td>1,266</td>
<td>180</td>
<td>273</td>
<td>Training</td>
<td>\cite{4}</td>
</tr>
<tr>
<td>iAT_PLT_636</td>
<td>GEM</td>
<td>1,008</td>
<td>289</td>
<td>593</td>
<td>Training</td>
<td>\cite{5}</td>
</tr>
<tr>
<td>iYS1720</td>
<td>GEM</td>
<td>3,357</td>
<td>738</td>
<td>962</td>
<td>Training</td>
<td>\cite{6}</td>
</tr>
<tr>
<td>RECON1</td>
<td>GEM</td>
<td>3,741</td>
<td>931</td>
<td>2,120</td>
<td>Training</td>
<td>\cite{7}</td>
</tr>
<tr>
<td>Yeast8 (v8.6.2)</td>
<td>GEM</td>
<td>4,063</td>
<td>1,108</td>
<td>1,760</td>
<td>Validation</td>
<td>\cite{8}</td>
</tr>
<tr>
<td>eeYeast8 (v8.3.4)</td>
<td>GEM</td>
<td>8,143</td>
<td>3,419</td>
<td>6,295</td>
<td>Validation</td>
<td>\cite{10}</td>
</tr>
<tr>
<td>Recon3D</td>
<td>GEM</td>
<td>10,600</td>
<td>4,861</td>
<td>11,195</td>
<td>Validation</td>
<td>\cite{11}</td>
</tr>
<tr>
<td>simplex_64D</td>
<td>Simplex</td>
<td>64</td>
<td>65</td>
<td></td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>simplex_256D</td>
<td>Simplex</td>
<td>256</td>
<td>257</td>
<td></td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>simplex_1024D</td>
<td>Simplex</td>
<td>1,024</td>
<td>1,025</td>
<td></td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>simplex_2048D</td>
<td>Simplex</td>
<td>2,048</td>
<td>2,049</td>
<td></td>
<td>Training</td>
<td></td>
</tr>
</tbody>
</table>
2 Measured $ESS/t$

Figure A: Measured $ESS/t$. Double logarithmic plot of absolute sampling efficiencies as measured by $ESS/t$ for selected thinning constants $\tau$. $ESS$ is determined by taking the minimum of the flux-specific values, i.e., $\min\{ESS_i\}_{i=1,...,D}$. Note that CHRRT for simplices achieves a higher absolute performance compared to sampling GEMs with similar dimensions. As an example, the best thinning constant $\hat{\tau}$ for simplex_{2048D} is comparable to the best thinning constant for the iAT_PLT_636 model, which has 582 dimensions. Comparing the absolute $ESS/t$ across various problems demonstrates that the effective dimensionality of the polytope is one, but not the only component determining sampling complexity.
3 Measured $ESS$

Figure B: Measured $ESS$. Double logarithmic plot of measured $ESS$ for selected thinning constants $\tau$. In almost all cases, the simulations were set up such that the minimum required $ESS = 400$ for all fluxes was achieved. Notable exceptions are a thinning constant of approximately $10^4$ for Recon1, where the ESS was so close to 400 that it is unlikely to bias the results. Furthermore, the convergence was clearly achieved in the interesting regions of the thinning constant, see Fig A.
4 MCMC convergence diagnostics

Figure C: Measured rank-normalized $\hat{R}$. A value close to one indicates convergence. In the case of Recon3D, the value is slightly above the threshold of 1.01 suggested in [12]. However, since the ESS is high (around 750), it is unlikely that this has a sizable impact on our results.
5 Flux-specific \( ESS \)

\( ESS \) values are given as the minimal value over all flux parameters, i.e., \( \min \{ ESS_i \} \). Here, we report the flux-specific \( ESS \) values for selected models. Note that the measured \( ESS \) does not directly map to sampling efficiency, because the different thinning constants \( \tau \) were not run equally long. Flux-specific \( ESS \) plots for the remaining models in Table S.1 are available in our code repository (https://jugit.fz-juelich.de/IBG-1/ModSim/Fluxomics/chrrt).

![Overview flux-wise ESS for e_coli_core](image)

Figure D: Flux-specific \( ESS \) a small core model of \( E. coli \) (\( e\_coli\_core \)) for different thinning constants \( \tau \). With increasing thinning constant, the \( ESS \) equalizes across the fluxes in the sense that the difference between the highest and lowest \( ESS \) is reduced.

![Overview flux-wise ESS for iJO1366](image)

Figure E: Flux-specific \( ESS \) for the genome-scale \( iJO1366 \) \( E. coli \) model for different thinning constants \( \tau \). We see that the \( ESS \) values move closer together with increasing thinning constant. In this case, this equalizing of the \( ESS \) is not as visible because all displayed thinning constants \( \tau \) are far from unthinned.
Figure F: **Flux-specific ESS for the *Recon3D* model for \( \tau = \frac{d^2}{6} \), as recommended by our thinning guideline.** Fluxes do not mix equally well. However, there is no visible outlier, in the sense that no single flux mixes substantially better or worse than any other flux.

Figure G: **Flux-specific ESS for *ecYeast8* for different thinning constants \( \tau \).** As the thinning constant increases, the ESS is equalized.
6 Selected flux distributions

We here show exemplary pair-plots for selected fluxes from different pathways. Further pair-plots for all benchmarked models are available in our code-repository [https://jugit.fz-juelich.de/IBG-1/ModSim/Fluxomics/chrrt]. Additionally, the selection of fluxes to include in the pair-plot can be altered in the jupyter-notebook.

Figure H: Selection of flux-distributions for the e.coli_core model. The biomass flux strongly correlates with the phosphate exchange flux EX_pi_e.

Figure I: Selection of flux-distributions for the iJO1366 model. In contrast to the results with the core model in Fig H, the biomass flux is not strongly correlated with the phosphate exchange EX_pi_e in this E. coli model.
Figure J: Selection of flux-distributions for the *Recon3D* model.

Figure K: Selection of flux-distributions *ecYeast8*. 
Reference


