

Supplementary Material for “Tracking SARS-CoV-2 genomic variants in wastewater sequencing data with *LolliPop*”

Code availability

All code used to produce the results presented in this article is available on:
<https://github.com/cbg-ethz/LolliPop/releases/tag/v0.1.0>

Supplementary 1

Let b be the vector of variant relative abundances, π be the vector of expected fractions of mutated reads given a design matrix X , i.e. $\pi(b) = Xb$. The Hessian matrix of the deconvolution likelihood model is found as follows:

$$\begin{aligned} \frac{\partial}{\partial b} \ell(\pi) &= \frac{\partial}{\partial \pi(b)} \ell(\pi(b)) \frac{\partial}{\partial b} \pi(b) \\ \Rightarrow \frac{\partial^2}{\partial b^2} \ell(\pi(b)) &= \frac{\partial}{\partial b} \left\{ \frac{\partial}{\partial \pi(b)} \ell(\pi(b)) \right\} \frac{\partial}{\partial b} \pi(b) + \frac{\partial^2}{\partial b^2} \pi(b) \frac{\partial}{\partial \pi(b)} \ell(\pi(b)) \\ &= \frac{\partial}{\partial b} \frac{\partial}{\partial \pi(b)} \ell(\pi(b)) \frac{\partial}{\partial b} \pi(b) = \left(\frac{\partial}{\partial \pi(b)} \frac{\partial}{\partial b} \ell(\pi(b)) \right)^\top \frac{\partial}{\partial b} \pi(b) \\ &= \left(\frac{\partial}{\partial \pi(b)} \frac{\partial}{\partial \pi(b)} \ell(\pi(b)) \frac{\partial}{\partial b} \pi(b) \right)^\top \frac{\partial}{\partial b} \pi(b) = \left(\frac{\partial}{\partial b} \pi(b) \right)^\top \left(\frac{\partial^2}{\partial \pi(b)^2} \ell(\pi(b)) \right)^\top \frac{\partial}{\partial b} \pi(b) \end{aligned}$$

$\sim \pi(b) = Xb \Rightarrow \frac{\partial^2}{\partial b^2} \pi(b) = 0$
 $\sim \frac{\partial^2}{\partial \pi(b)^2} \ell(\pi(b))$

The Fisher information matrix is then

$$I_b(b) = E \left[- \frac{\partial^2}{\partial b^2} \ell(\pi) \right] = E \left[\left(\frac{\partial}{\partial b} \pi \right)^\top \left(- \frac{\partial^2}{\partial \pi^2} \ell(\pi) \right)^\top \frac{\partial}{\partial b} \pi \right] = X^\top \text{diag} \left\{ \pi_m (1 - \pi_m) \right\}_{m=1, \dots, M}^{-1} X$$

Supplementary Figures

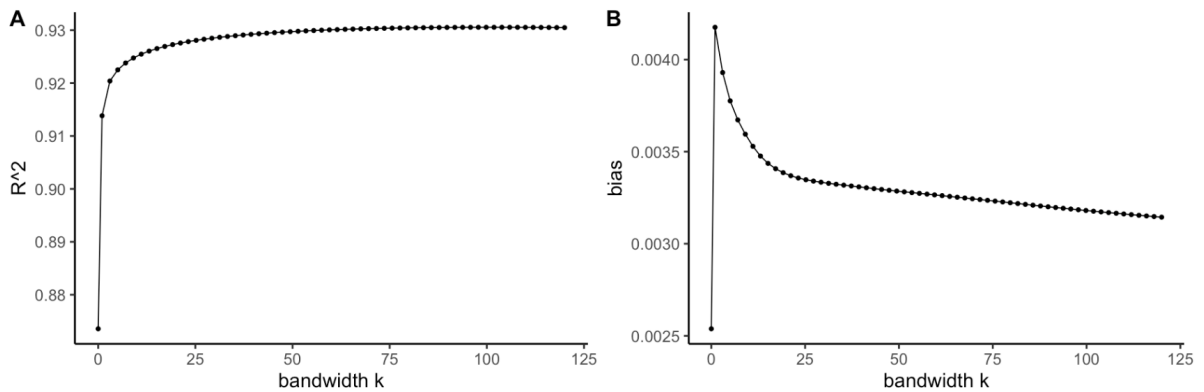


Figure S1: Goodness of fit (A) and bias (B) of the MSE-based kernel deconvolution of the wastewater NGS, as a function of the kernel bandwidth κ . Goodness of fit and bias are

evaluated by regressing on estimates of relative abundances of variants obtained from clinical sequencing data.

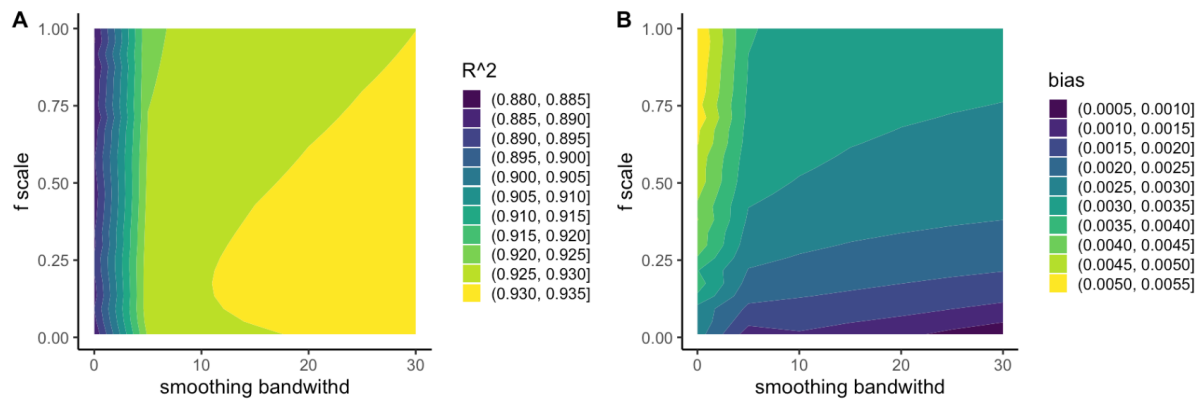


Figure S2: Goodness of fit (A) and bias (B) of the robust kernel deconvolution of the wastewater NGS, as a function of the kernel bandwidth κ and α scale parameter of the robust regression. Goodness of fit and bias are evaluated by regressing on estimates of relative abundances of variants obtained from clinical sequencing data.