

Pseudo-code for MDLINFER

Here, we assume the existence of the calibrate procedure for epidemiological model O_M and term it as $CALIBRATE(O_M, \{\text{list of datasets}\})$. The list of datasets have at least reported infections time sequence D_{reported} , and can also have others. This $CALIBRATE$ uses loss function to learn parameterizations, usually root mean squared error (RMSE) or log likelihood [3, 2] loss functions. We then give the pseudo-code for MDLINFER:

Algorithm 1 MDLINFER

Input: Epidemiological model O_M , calibration procedure $CALIBRATE$, reported infections time sequence D_{reported} .

- 1: Calibrate baseline parameterization $\mathbf{p} = CALIBRATE(O_M, \{D_{\text{reported}}, \text{others}\})$
- 2: Step 1: Find $\alpha_{\text{reported}}^* = GETALPHA(O_M, CALIBRATE, D_{\text{reported}}, \mathbf{p})$
- 3: Step 2: Find $D^* = GETTOTALINFECTIONS(O_M, CALIBRATE, D_{\text{reported}}, \alpha_{\text{reported}}^*, \mathbf{p})$

Output: Total infections D^*

Next, we give the pseudo-code for $GETALPHA$ and $GETTOTALINFECTIONS$:

Algorithm 2 GETALPHA (Step 1: Find the $\alpha_{\text{reported}}^*$)

Input: O_M , $CALIBRATE$, D_{reported} , \mathbf{p} .

- 1: The array to save the MDL cost: $CostArray = []$
- 2: **for** α_{reported} in the grid search space from 0.01 to 1 with step 0.01 **do**
- 3: $D = \frac{D_{\text{reported}}}{\alpha_{\text{reported}}}$
- 4: $D_{\text{unreported}} = D - D_{\text{reported}}$
- 5: Calibrate candidate parameterization $\mathbf{p}' = CALIBRATE(O_M, \{D_{\text{reported}}, D_{\text{unreported}}, \text{others}\})$
- 6: Save the MDL cost for α_{reported} in $CostArray[\alpha_{\text{reported}}] = L(D_{\text{reported}}, D, \mathbf{p}', \mathbf{p})$
- 7: **end for**
- 8: Find the $\alpha_{\text{reported}}^* = \arg \min_{\alpha_{\text{reported}}} CostArray[\alpha_{\text{reported}}]$

Output: Reported rate $\alpha_{\text{reported}}^*$

Algorithm 3 GETTOTALINFECTIONS (Step 2: Find the D^* given $\alpha_{\text{reported}}^*$)

Input: O_M , $CALIBRATE$, D_{reported} , $\alpha_{\text{reported}}^*$, \mathbf{p} .

- 1: Find the $D^* = \arg \min_D L(D_{\text{reported}}, D, \mathbf{p}', \mathbf{p})$. (using the Nelder-Mead algorithm*).

Output: Total infections D^*

*Note: Nelder-Mead is a commonly used numerical algorithm used for optimiation in multidimensional space [1].

References

- [1] GAO, F., AND HAN, L. Implementing the nelder-mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications* 51, 1 (2012), 259–277.
- [2] HAO, X., CHENG, S., WU, D., WU, T., LIN, X., AND WANG, C. Reconstruction of the full transmission dynamics of covid-19 in wuhan. *Nature* 584, 7821 (2020), 420–424.
- [3] KAIN, M. P., CHILDS, M. L., BECKER, A. D., AND MORDECAI, E. A. Chopping the tail: How preventing superspreading can help to maintain covid-19 control. *Epidemics* 34 (2021), 100430.