This repository contains training data for an ML-based parameter optimization for modeling morphogen gradient formation by a source-diffusion-sink (SDS) mechanism [1], as described by the following one-dimensional (1D) reaction-diffusion partial differential equation:

$$\frac{\partial c(x,t)}{\partial t} = f_{\text{source}}(x) + D\frac{\partial^2}{\partial x^2}c(x,t) - k_{\text{sink}}c(x,t)\,. \tag{1}$$

Here, $c(x,t)$ is the scalar concentration field of the morphogen in space x at time t, D is the constant homogeneous diffusion coefficient, k_{sink} is the sink rate scaling with $c(x,t)$, describing morphogen degradation by the cells and proteases in the extracellular space, and f_{source} is the source term describing morphogen secretion by the source cells.

f_{source} depends on the location as only a group of cells produce the morphogen, i.e.,

$$f_{\text{source}}(x) = \begin{cases} k_{\text{source}} & \text{for} \quad x|d_{\text{margin}}(x) \leq w_{\text{source}} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

in a 1D diffusion domain $\Omega = \{x| -\frac{L}{2} \leq x \leq \frac{L}{2}\}$ of length $L = 2.0$. The source is a region of width $w_{\text{source}} = 0.3L$ restricted to a distance $d_{\text{margin}}(x) \leq w_{\text{source}}$ from the left margin. We consider the scenario of a zero morphogen concentration in Ω as initial condition

$$c(x,0) = 0, \quad x \in \Omega \tag{3}$$

and homogeneous zero-flux Neumann boundary conditions at all boundaries $\partial\Omega$, i.e.,

$$\left.\frac{\partial c(x,t)}{\partial \boldsymbol{n}}\right|_{x\in\partial\Omega} = 0\,, \tag{4}$$

with $\boldsymbol{n}$ as the normal vector on $\partial\Omega$. We discretize Ω by 64 grid points, resulting in a grid spacing of $\Delta x = 0.031746$. The timestep size depends on Δx and the diffusion coefficient as

$$\Delta t = \frac{1}{8D}\Delta x^2\,. \tag{5}$$

We solve Eq. (1)–(5) using first-order forward-time central-space finite differences until Eq. (1) reaches a steady-state. This steady-state morphogen concentration profile can be flat, exponential, or step-wise depending on the parameters k_{source}, k_{sink}, and D. The goal is to generate a model that predicts a parameter set given an input gradient. To train this model, we

produce simulated training data by solving Eq. (1) until steady-state for different parameter sets.

The output folders contain a parameters.csv file, where the first row defines the parameter type of the respective column, and each row below contains a set of parameters with index i (see Fig. 1). The corresponding gradients folder contains gradient_i.csv files, each corresponding to one row i of the parameters.csv file. The columns of gradient_i.csv contain the concentration value at each grid point and each row contains one iteration, as shown in Fig. 1.

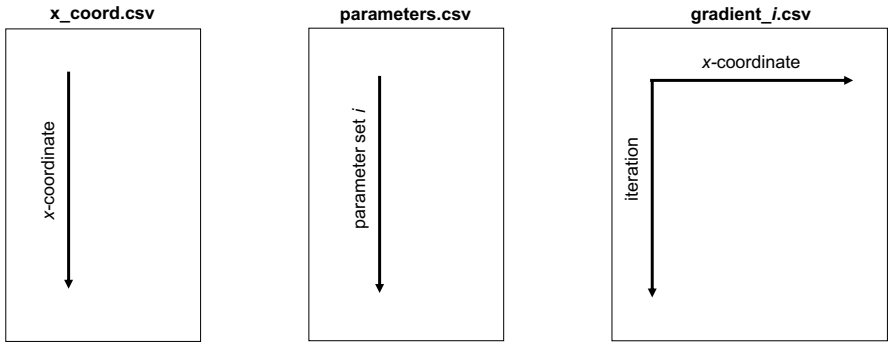

Figure 1

Each gradient_i.csv file contains two columns: the first containing x, the second containing the steady-state concentration field $c(x, t_{\max})$.

References

[1] S. R. Yu, M. Burkhardt, M. Nowak, J. Ries, Z. Petrášek, S. Scholpp, P. Schwille and M. Brand. Fgf8 morphogen gradient forms by a source-sink mechanism with freely diffusing molecules. *Nature* **volume**(461), 533–536, 2009.