

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:

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

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,  $k_{\text{sink}}$  is the sink rate scaling with  $c(x, t)$ , describing morphogen degradation by the cells and proteases in the extracellular space, and  $f_{\text{source}}$  is the source term describing morphogen secretion by the source cells.

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

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

with source width  $w_{\text{source}} = 0.3L$  in a 1D diffusion domain  $\Omega = \{x | 0 \leq x \leq L\}$  of length  $L = 2.0$ . We consider the scenario of a zero morphogen concentration in  $\Omega$  as initial condition

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

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

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

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

$$\Delta t = \frac{1}{8D} \Delta x^2. \quad (5)$$

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  $k_{\text{source}}$ ,  $k_{\text{sink}}$ , 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  $i$ . The corresponding gradients folder contains `gradient_` $i$ `.csv` files, each corresponding to one row  $i$  of the `parameters.csv` file.

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.