

DETERMINING THE PRECISION OF MORPHOGEN GRADIENT FORMATION AND READOUT BY NUMERICAL SIMULATIONS

Jan Adelman

Supervision: Dr. Roman Vetter and Prof. Dagmar Iber

Computational Biology Group
ETH Zürich
Zürich, Switzerland

ABSTRACT

In the early stages of embryonic development, morphogen gradients provide the positional information necessary for patterning. Using computational methods, the accuracy of morphogen gradients was established. Introducing variable cell size as a source of noise did not significantly increase the variability in the morphogen gradients. Further, simulations imply that cells with small diameters lead to more robust gradients. For example, cells in the developing chick neural tube have small cell diameters, which leads to more robust gradients. Different spatial averaging methods reveal significantly different results in the reported concentration variability. Calculating the readout position of noisy gradients showed that the deviation from the mean position is as little as 1-2 cell diameters, over the whole domain considered. Therefore, morphogen gradients are accurate enough to provide the spatial information needed for precise patterning.

1. INTRODUCTION

During the early stages of development, cells acquire different cell fates. A cell 'knows' its spatial location in a tissue through the spatial information mediated by morphogen gradients. According to the French flag model, a morphogen diffuses from a source into a patterning domain [1]. Cells in the patterning domain

readout the concentration and change their cell identity according to concentration thresholds C_θ . In the domain itself the concentration declines over many orders of magnitude [2]. It is still not known if cells can sense such low concentration and how the noisy gradients can lead to the accurate spatial patterning found in animals. The morphogen gradients can be approximated well by exponential functions,

$$C(x) = C_0 \exp\left(-\frac{x}{\lambda}\right). \quad (1)$$

The gradient defined in equation 1 is characterised by an amplitude C_0 and a decay length λ . The readout position, defined by the concentration thresholds C_θ , is obtained by rearranging equation 1. Which results in the following readout position x_θ for a specified threshold concentration C_θ ,

$$x_\theta = \lambda \ln\left(\frac{C_0}{C_\theta}\right).$$

Computational- and theoretical work by Vetter and Iber [3] assessed whether the noisy gradients are accurate enough to define the extremely precise patterning domains. According to their model, most uncertainty in the gradients comes from noise in the measurements of the gradients. The study showed that morphogen gradients are accurate enough to define the spatial resolution needed for patterning. This would mean that the

inaccuracy reported so far comes from the experimental methods. In this study their findings were extended and validated.

2. METHODS

2.1. Simulating noisy gradients and variability quantification

Noisy gradients were obtained by solving a one dimensional reaction-diffusion system, equivalent to the system set up by Vetter and Iber [3]. The reaction-diffusion equation is described by the following second order differential equation,

$$pH(-x) - dC(x) = -D \frac{\partial^2 C(x)}{\partial x^2}. \quad (2)$$

Equation 2 was solved on a continuous domain consisting of two subdomains. A source domain L_s on the interval $(-L_s \leq x \leq 0)$ and a patterning domain L_p on the interval $(0 \leq x \leq L_p)$, cf. figure 1. The reaction-diffusion equation contains a production, term with production rate p , a linear degradation term with degradation rate d and a Fickian diffusion term with diffusion constant D . The term $H(-x)$ ensures that production only occurs in the source domain, where H is the Heaviside step function. In comparison, degradation takes place over the whole domain. Imposing zero flux boundary conditions,

$$\frac{\partial C(-L_s)}{\partial x} = 0 = \frac{\partial C(L_p)}{\partial x},$$

allows for an analytical solution of the equation. The solution is characterised by hyperbolic cosine concentration profiles, as is evident from equation 3.

$$C(x) = \frac{p}{d} \left(H(-x) \left(1 - \cosh \left(\frac{x}{\lambda} \right) \right) + \frac{\sinh(L_s/\lambda)}{\sinh((L_s + L_p)/\lambda)} \cosh \left(\frac{L_p - x}{\lambda} \right) \right). \quad (3)$$

Analogously to the paper by Vetter and Iber [3] the kinetic parameters $k = p, d$ and D were drawn from log-normal distributions with prescribed mean μ_k and respective coefficients of variation CV_k ,

$$k \sim \ln \mathcal{N}(\hat{\mu}, \hat{\sigma}^2).$$

Where $\hat{\mu}$ and $\hat{\sigma}^2$ are the estimated mean and variance parameters. To extend their work, this study introduces cell size variability as an additional source of noise. The cell areas were drawn from log-normal distributions with prescribed mean μ_{area} and coefficient of variation CV_{area} . The mean cell diameter of $4.9 \mu\text{m}$ corresponds to the average cell diameter of cells in the neural tube (NT) as described in the supplementary materials by Zagorski et al. [4]. In a study performed by Guerrero et al. [5], they measured cell areas in the NT. The distribution of cell areas obtained in this study resembles a log-normal distribution. As the distribution is only known for areas, the sampled areas had to be converted to diameters. The simulations were repeated for fixed $CV_k = 0.3$ and varying CV_{area} . For each CV_{area} thousand noisy gradients were computed. The two gradient parameters λ and C_0 were extracted by log-fitting hyperbolic cosines to the numerical solutions. To quantify the variability in the gradient parameters λ and C_0 the coefficient of variation (CV) was used. For example the CV for the gradient decay length λ is calculated as $CV_\lambda = \sigma_\lambda / \mu_\lambda$.

2.2. Impact of spatial averaging

Spatial averaging methods can potentially reduce noise. The methods considered in this study reduce the different concentrations a cell 'experiences' along its cell surface to a single scalar value. The concentration in each cell is selected according to three different principles:

1. Average cellular concentration.
2. Random cellular concentration.
3. Subset of adjacent grid points.

To calculate the unique concentration in each cell, the numerical solutions of equation 2 were matched with cell boundaries, extracting the solutions that lie within each cell. To find the average concentration of a cell, the solutions were numerically integrated using the trapezoid method. The result was normalised by the cell area. Another strategy for a cell to readout a concentration is to select a random coordinate in the cell and the concentration at that location is the readout concentration of the cell. As the numerical solutions of the reaction-diffusion equation only consisted of discrete

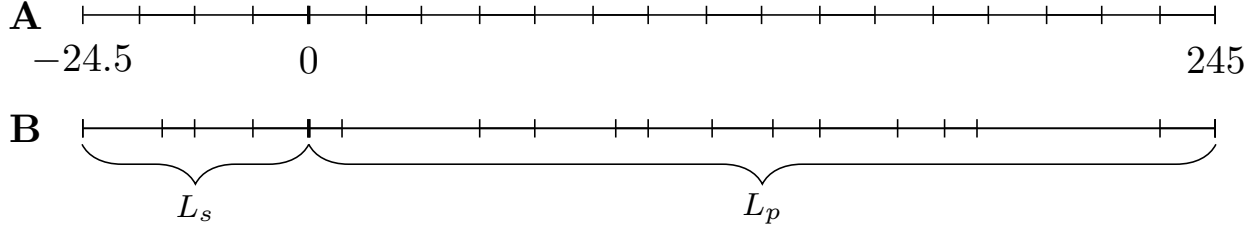


Fig. 1. Schematic depictions of the one dimensional source domain L_s and patterning domain L_p . The overall length of the two subdomains was kept constant. A Source and patterning domain with evenly spaced cells, each with a diameter of $4.90 \mu\text{m}$. **B** Introducing variable cell diameters, with a mean cell diameter of $4.9 \mu\text{m}$.

values, the concentration at the selected location was obtained by interpolation. The third procedure considers a small subset of the cell. Where concentrations are always readout at that particular location in the cell. This can be thought of as a readout at a small compartment such as a cilium [6]. This method is called 'cilium method' for the rest of this report. The readout concentration was determined by interpolation of the numerical solutions.

2.3. Defining a readout threshold C_0

Hill functions are often used when modelling the reaction kinetics of a small diffusible molecule which binds to its receptor. These interactions are often characterised by sigmoidal responses. During development the small molecules are the morphogens which bind to the respective receptors on the cell. T. Gregor et al. modelled the interaction between Hunchback and Bicoid with a Hill equation [7]. The Hill equation has the following functional form [8],

$$\omega = \frac{C(x)^n}{C(x)^n + K^n}. \quad (4)$$

$C(x)$ are the concentrations after the respective averaging readout method was applied, n is the hill coefficient. The parameter K defines the half occupation concentration. The output of the Hill equation ω displays the fraction of receptor bound by the ligand (morphogens). The sigmoidal shape for varying hill coefficients is shown in figure 2. In the following, the readout threshold is fixed at $\omega = 0.5$. This threshold is to a certain degree arbitrary, as it is not clear at which threshold cells readout concentrations. Setting a threshold of $\omega = 0.5$ simplifies equation 4. As concen-

trations and K are both positive and non-zero numbers in the domain considered, equation 4 can be rewritten as,

$$\omega = \frac{1}{1 + \left(\frac{K}{C(x)}\right)^n}.$$

Setting $\omega = 0.5$, the relationship between K and $C(x)$ simplifies to $K = C(x)$, independently of n . In this deterministic case, the half-occupation concentration K defines the readout concentration and thus ultimately also the readout position x . By varying K , the readout position x is shifted in the domain. In biological systems, K itself is noisy. Because K can take only positive values, noise can be introduced by drawing K from a log-normal distribution with mean K and coefficient of variation CV_K . Two approaches of retrieving the readout position x from the concentrations were considered. firstly, a series of K values that span the whole patterning domain was chosen. K is equivalent to the readout concentration C_θ . To find the readout position x at which K is attained, the concentrations obtained by solving the reaction diffusion equation were interpolated. The second approach uses Hill functions. Concentrations were plugged into the Hill equation, and through interpolation, the readout position x was obtained at the above mentioned threshold of $\omega = 0.5$. To assess if the interpolation has an effect on the result, both linear and cubic interpolation was applied. Performing these measurements for repeated simulations, lead to a mean readout position μ_x and a positional error,

$$\sigma_x = \text{SD}\{x_{\theta,i}\}. \quad (5)$$

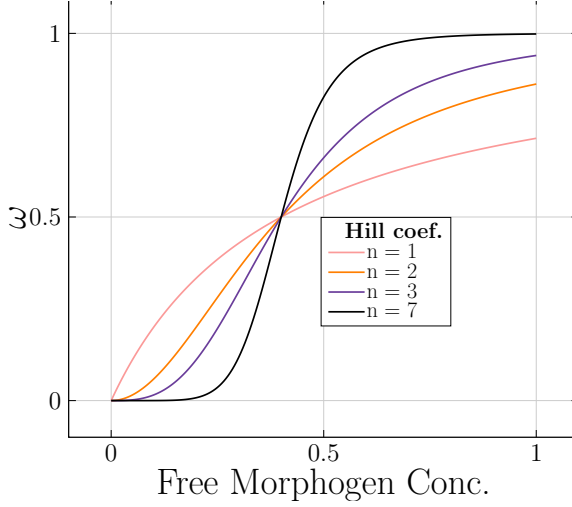


Fig. 2. Hill Curves plotted for different Hill coefficients and a half occupation $K = 0.4$. A higher hill coefficient leads to faster saturation of the receptor. In the limit $n \rightarrow \infty$ the response approaches a step function. The fraction of receptor bound by morphogens is denoted by ω .

3. RESULTS

3.1. Effect of variability in the cell size

For the following simulations the variability in the kinetic parameters $k = p, d$ and D was fixed at a coefficient of variation $CV_k = 0.3$. The mean cell diameter was fixed at $4.9 \mu\text{m}$, but the coefficient of variation CV_{area} was increased from 0.01 to 10. Results are shown in figure 3. Considering the biological relevant scale of $0 \leq CV_{\text{area}} \leq 1$, the coefficient of variation for both gradient parameters remains almost constant. Which implies that introducing a certain degree of cell size variability does not affect gradient parameter variability. Nevertheless, there is a moderate linear increase in CV_λ after $CV_{\text{area}} \gtrsim 1$. It is interesting to note that CV_0 stays roughly constant over the whole CV_{area} range considered.

3.2. Effect of varying cell diameters

The average cell diameter was fixed at $4.9 \mu\text{m}$ for the simulations presented in subsection 3.1. In order to study the impact of varying cell diameters the variability in the kinetic parameters was fixed at $CV_k = 0.3$ and the area variability was fixed at $CV_{\text{area}} = 0.5$. Fur-

thermore, the mean diameter of cells was varied and gradients were simulated. Whereas the overall domain size was kept constant. Figure 4 displays a linear increase in both parameters λ and C_0 as the cell diameter increases. Consequently, lining a domain with a lot of small cells leads to smaller variability in the gradient parameters. On the other hand, having only a few big cells in the domain leads to bigger variability in the gradients. To check the diameter of cells in the early stages of development, data from the chick neural tube (cNT) by Iber and Kokic [9] was used. The measured mean cell area of the cNT is $6.61 \mu\text{m}^2$. Assuming circular cell geometry leads to an average cell diameter of $2.90 \mu\text{m}$. Normalising by the mean gradient decay length $\mu_\lambda = 19.26 \mu\text{m}$, yielded a value of diameter/ $\mu_\lambda = 0.15$. Comparing this value to figure 4, results in a value for CV_λ of ≈ 0.011 and a value for CV_0 of ≈ 0.105 . Compared to other diameters from different tissues, collected by Iber and Kokic [9], the cNT has one of the smallest cell diameters. Hence, to minimise variability in the gradients, it can be a strategy to place many small cells in the patterning domain.

3.3. Spatial averaging of cellular concentration

The coefficient of variation for the three methods is plotted at various positions in the domain, cf. figure 5. The overall variability for all methods is increasing along the domain. At the end of the domain a saturation occurs. To test if the methods provide significantly different results, the one sample t-test was used. If two methods provide the same results, the mean difference between them should be zero. The null hypothesis therefore assumes that the mean difference of two methods is zero. If the null hypothesis is rejected the mean difference is not equal to zero and therefore the two methods give rise to different variabilities $CV_{C(x)}$. The null hypothesis is not rejected for the cilium and average method with a p-value = 0.266, cf. table 1. Variability in the concentration for both methods is not significantly different. In contrast, the random method is significantly more variable than the other two methods.

3.4. Accuracy of readout position

Patterning in the developing embryo is an incredibly precise process. Placing wrong differentiation bound-

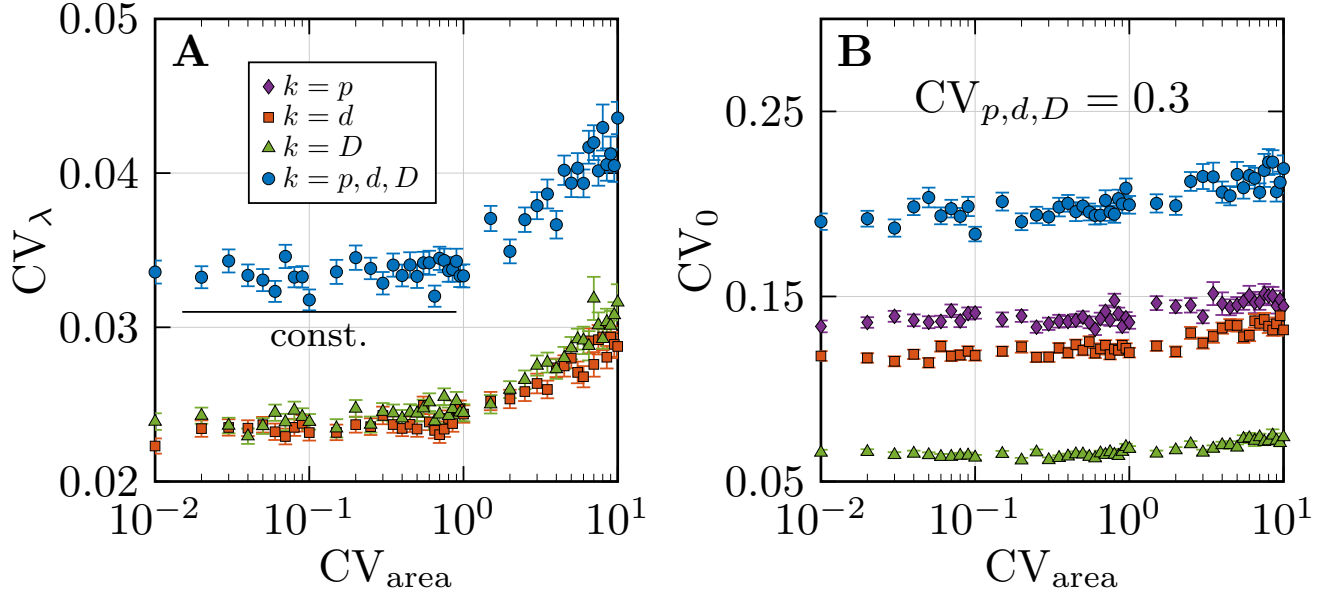


Fig. 3. Predicted variability in gradients from numerical model, allowing for variable cell sizes and fixing $CV_k = 0.3$. Where $k = p, d, D$. For each cell the parameters were drawn independently from a log-normal distribution. A The variability in λ is constant in the region $0.01 \leq CV_{area} \leq 1$. For higher CV_{area} the variability in λ increases linearly. **B** The variability in C_0 is almost constant over the whole CV_{area} spectrum considered. In the region $CV_{area} \geq 1$ a small increase in variability occurs.

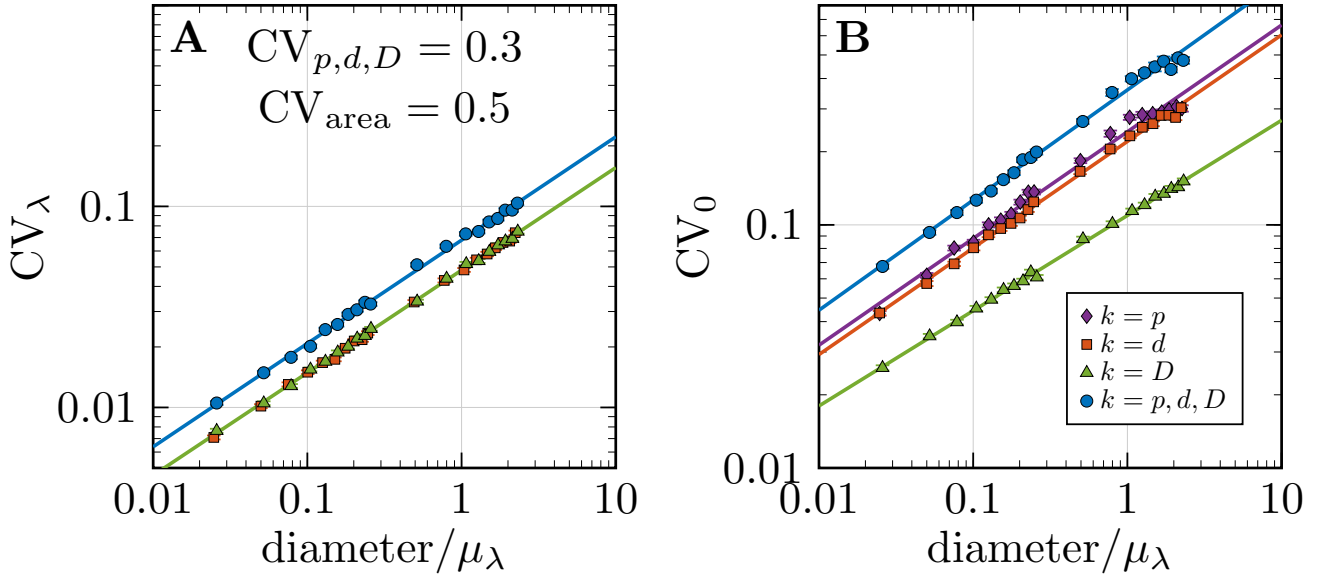


Fig. 4. Effect of varying mean cell size on the gradient variability. The coefficient of variation CV_k was fixed at 0.3 and the coefficient of variation CV_{area} was fixed at 0.5. The diameter of each cell was divided by the mean decay length of the gradient leading to a dimensionless variable. The mean decay length was calculated for fixed mean cell diameter configuration. The parameters were drawn for each cell independently from a log-normal distribution. A Smaller average cell diameters lead to less variability in the parameter λ , the variability increases linearly on the log-log plot. **B** The variability in the amplitude C_0 also increases linearly over the course of the domain.

Comparison	Average difference	Standard error (SE)	P-value (one sample t-test, two sided)
Cilium:Average	3.193×10^{-5}	2.856×10^{-5}	0.266
Random:Average	0.01039	0.000466	9.100×10^{-41}
Random:Cilium	0.01040	0.000456	1.873×10^{-41}

Table 1. Summary statistics comparing the three readout methods.

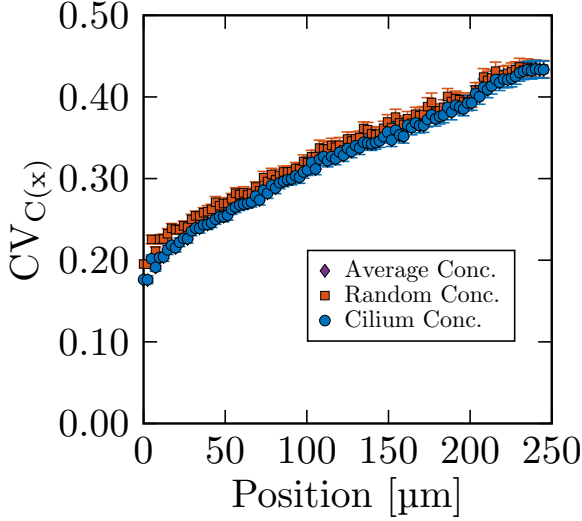


Fig. 5. Variability in the concentration readout at different locations in the domain. Saturation of the increasing variability occurs after approximately 230 μm for all of the three methods. The averaging method and the cilium method show almost identical variability. The random concentration method has a higher variability compared to the other two. Standard Errors were obtained by bootstrapping.

aries could potentially lead to abnormal limb formation and/or organ growth. In section 3.3, different spatial averaging methods were presented and the variability of the concentrations at different positions in the domain were quantified. Now, the accuracy of the readout position is of interest. Noisy gradients with $\text{CV}_{\text{area}} = 0.5$ and $\text{CV}_K = 0.3$ were simulated. The accuracy of the readout position x was quantified as the standard deviation of the mean readout position, described by equation 5. Results presented in this section are with respect to the average concentration method. The other two methods provide very similar results. Using K , the half occupation constant, to define the readout position results in figure 6. Increasing CV_K leads to a higher positional error σ_x . For example if

$\text{CV}_K = 1$ the positional error is approximately three times bigger than in the noise free setting, cf. figure 9. This is expected, as an increase in noise leads on average to more variability. At the beginning of the domain the variability decreases fast. In contrast, at the end of the domain a sharp increase in σ_x is occurring. The behaviour can be explained by looking at the interpolation method chosen to find the readout position x . The first readout position is chosen such that the first K value is equal to the solution of equation 3 at the point $x = 0$. If the gradients are noisy it is possible that the position where this concentration is reached lies in the source domain. Therefore, the increase in variability is an artefact of the simulations, as at these points no interpolation but extrapolation takes place. The extrapolation is associated with more uncertainty, thus the increased σ_x . Regions in between the two black lines denote the subset of the domain where only interpolation took place. The black lines were plotted with respect to the data obtained by setting $\text{CV}_K = 0$. In the numerically stable part of the domain 12 – 208 μm , the positional error increases moderately. The positional error increases above one cell diameter after approximately 75 μm for the simulations with low variability in K ($\text{CV}_K = 0, 0.01$ and 0.1). At 200 μm , σ_x increases to around 1.5 cell diameters. Concluding that the concentration based threshold model can lead to high spatial readout accuracy, even with noisy gradients. The cNT cell diameter (2.9 μm) is lower than the average diameter considered above (4.9 μm). But even when taking a smaller diameter, the positional error in the middle of the domain is not higher than 2 cell diameters. This result implies that gradients are accurate enough to allow for the high spatially resolved patterning in the NT. So far, the concentration K defined the readout position x . Another family of threshold functions to consider are Hill functions. A natural threshold to choose is the half-occupation $\omega = 0.5$. In section 2.3 the relationship $C(x) = K$ was established for this particular threshold. In a noise free system

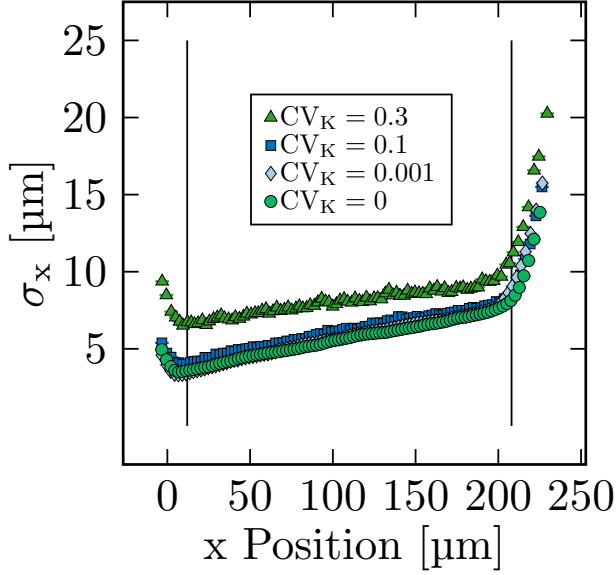


Fig. 6. Accuracy of the readout positions for varying CV_K values with the threshold concentration being defined by K . Increasing CV_K leads to an increase of variability (σ_x) in the readout position. In-between the two black lines only interpolation and no extrapolation takes place. Standard errors for σ_x were obtained by bootstrapping.

the two approaches presented above should be equivalent. Even in a noisy system the two approaches give a similar result, cf. figure 6. The overlap of between the K based readout and the Hill function readout is almost perfect. Interestingly the Hill function threshold readout seems to have a slightly higher variability. The same qualitative behaviour is observed for varying CV_K as in the K based approach, cf. figure 7. Increasing CV_K increases the positional error σ_x .

4. CONCLUSIONS

The work conducted in this project extended the study by Vetter and Iber [3]. Their findings were reproduced and validated. The impact of cell size variability in the relevant biological region is negligible. Cell area variability does not increase the variability in the gradient parameters λ and C_0 exceedingly. A domain lined with small cells leads to less variability in the gradient parameters and thus to more robust gradients. To assess if this behavior is observed in nature, more data needs to

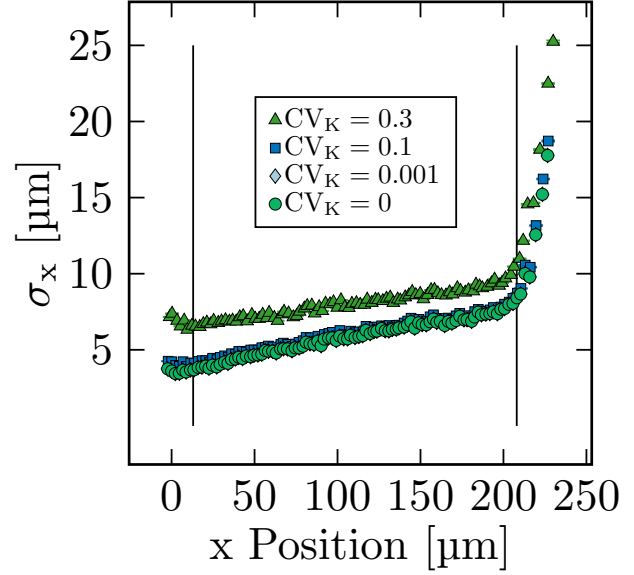


Fig. 7. Positional accuracy for varying CV_K , with the threshold being defined via a Hill function with threshold $\omega = 0.5$. Increasing CV_K increases the variability in the readout position. Standard errors for σ_x were obtained by bootstrapping.

be analysed. Taking the average cellular concentration or the concentration at a small subset of a cell showed identical variability in the concentrations $CV_{C(x)}$. As expected, selecting a random concentration in the cell lead to the highest variability in the concentrations at a certain location. All three methods provide similar readout precision. The readout precision obtained by simply using a concentration or by first applying a Hill equation with hill coefficient $n = 1$ revealed similar results when choosing a threshold of $\omega = 0.5$ for the Hill function. In a next step, higher order Hill coefficients could be considered and varying readout thresholds. Overall the positional error σ_x increased from approximately half a cell diameter to two cell diameters over the course of the domain, assuming an average cell diameter of $4.9 \mu m$. Increasing the noise level of the readout concentration increased the variability of the positional readout. Moderate noise levels, $CV_K = 0.1$, only had a small impact on readout precision. Even with noise, positional accuracy throughout the whole domain is very high. We conclude that morphogen gradients are accurate enough to pattern a domain with high precision.

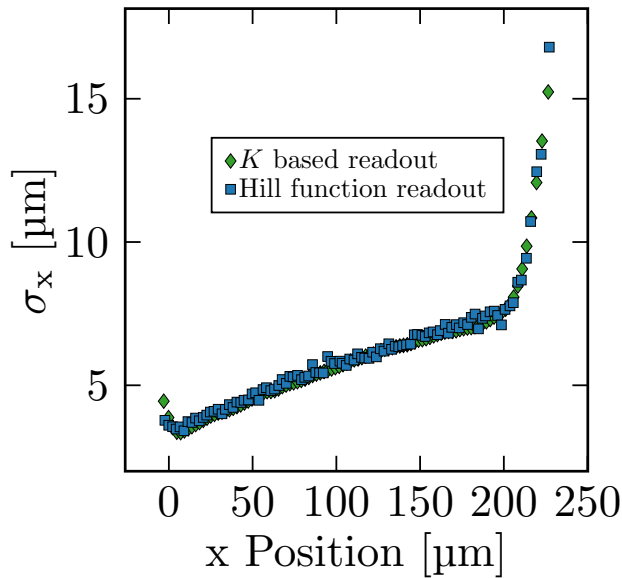


Fig. 8. Comparison of the two readout methods. The green points were obtained by defining a series of K values that specify the readout concentration. Readout positions were found by interpolation. The blue points were obtained using the Hill function defined in equation 4 as a threshold function. The threshold was set to $\omega = 0.5$. The two approaches show the same qualitative behaviour and overlap well. The Hill function based readout is noisier.

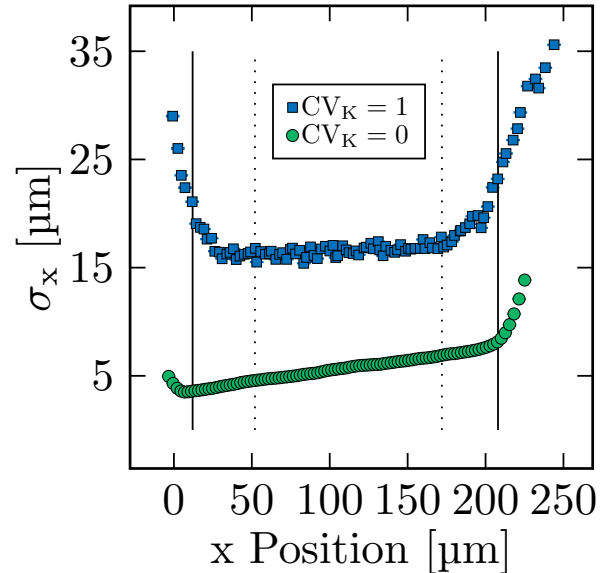


Fig. 9. Error in the readout position x increases with higher CV_K . For large CV_K values, $CV_K \approx 1$, the positional error σ_x stays almost constant in the numerically stable regions. $CV_K = 1$, approximately triples the positional error compared to the noise free setting, $CV_K = 0$. In between the solid and dotted black lines numerical solutions were interpolated. Solid lines correspond to $CV_K = 0$, dotted lines to $CV_K = 1$.

5. REFERENCES

- [1] Wolpert L. 1969 Positional information and the spatial pattern of cellular differentiation. *Journal of Theoretical Biology* **25**, 1, 1–47. (doi:[https://doi.org/10.1016/S0022-5193\(69\)80016-0](https://doi.org/10.1016/S0022-5193(69)80016-0)).
- [2] Gurdon JB, Bourillot PY. 2001 Morphogen gradient interpretation. *Nature* **413**, 6858, 797–803. (doi:[10.1038/35101500](https://doi.org/10.1038/35101500)).
- [3] Vetter R, Iber D. 2021 Inference of morphogen gradient precision from molecular noise data. *bioRxiv* (doi:[10.1101/2021.04.20.440728](https://doi.org/10.1101/2021.04.20.440728)).
- [4] Zagorski M, Tabata Y, Brandenberg N, Lutolf MP, Tkačik G, Bollenbach T, Briscoe J, Kicheva A. 2017 Decoding of position in the developing neural tube from antiparallel morphogen gradients. *Science* **356**, 6345, 1379–1383. (doi:[10.1126/science.aam5887](https://doi.org/10.1126/science.aam5887)).
- [5] Guerrero P, Perez-Carrasco R, Zagorski M, Page D, Kicheva A, Briscoe J, Page KM. 2019 Neuronal differentiation influences progenitor arrangement in the vertebrate neuroepithelium. *Development* **146**, 23.
- [6] Flaherty J, Feng Z, Peng Z, Young YN, Resnick A. 2020 Primary cilia have a length-dependent persistence length. *Biomechanics and Modeling in Mechanobiology* **19**, 2, 445–460. (doi:[10.1007/s10237-019-01220-7](https://doi.org/10.1007/s10237-019-01220-7)).
- [7] Gregor T, Tank DW, Wieschaus EF, Bialek W. 2007 Probing the limits to positional information. *Cell* **130**, 1, 153–164. (doi:<https://doi.org/10.1016/j.cell.2007.05.025>).
- [8] Somvanshi PR, Venkatesh KV. 2013 *Hill Equa-*

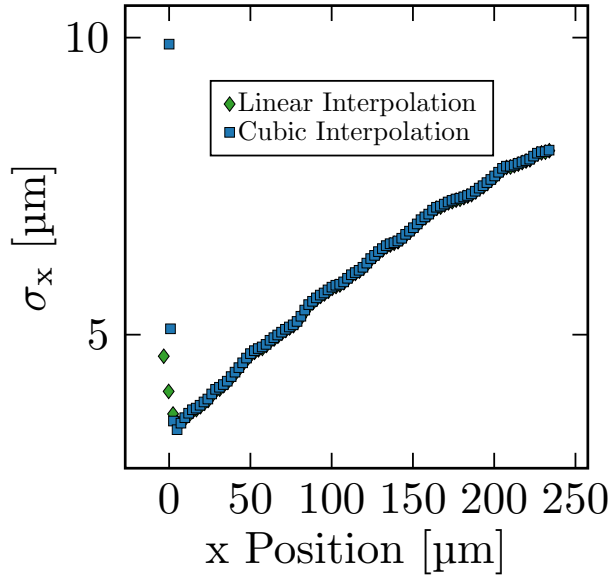


Fig. 10. Increased variability at the start end end of domain is due to extrapolation. Comparison of two interpolation methods.

tion, pp. 892–895. New York, NY: Springer New York. (doi:10.1007/978-1-4419-9863-7_946).

- [9] Kokic M, Iannini A, Villa-Fombuena G, Casares F, Iber D. 2019 Minimisation of surface energy drives apical epithelial organisation and gives rise to lewis’ law. *bioRxiv* (doi:10.1101/590729).

Appendix

Validation of the results found by Vetter and Iber [3]

Introducing area variability ($CV_{\text{area}} = 0.5$) did not significantly change the results. The findings are presented in figure 11. All trends observed without area variability are also observed when including area variability. For example, the saturation taking place when increasing the variability of the kinetic parameters is still observable, cf. figure 11.a. Similarly, no saturation is observed for the variability in the gradient parameter C_0 , cf. figure 11.c.

Extrapolation as a source of increased positional variability σ_x

Interpolations methods used in this study don’t give reliable results the methods switch from interpolation to extrapolation. This can be visualised by solving the reaction-diffusion equation on a bigger domain. For this purpose the size of the patterning domain was doubled Readout points close to 250 μm now have numerical solutions in their neighbourhood. Therefore no extrapolation takes place. Accordingly, the positional error σ_x does not increase rapidly anymore, cf. figure 10. Note that the source domain was not changed, for this reason, the variability is still high at the start of the patterning domain. In the region where interpolation takes place, cubic- and linear interpolation give the same result, as expected. This does not hold for the region where extrapolation happens.

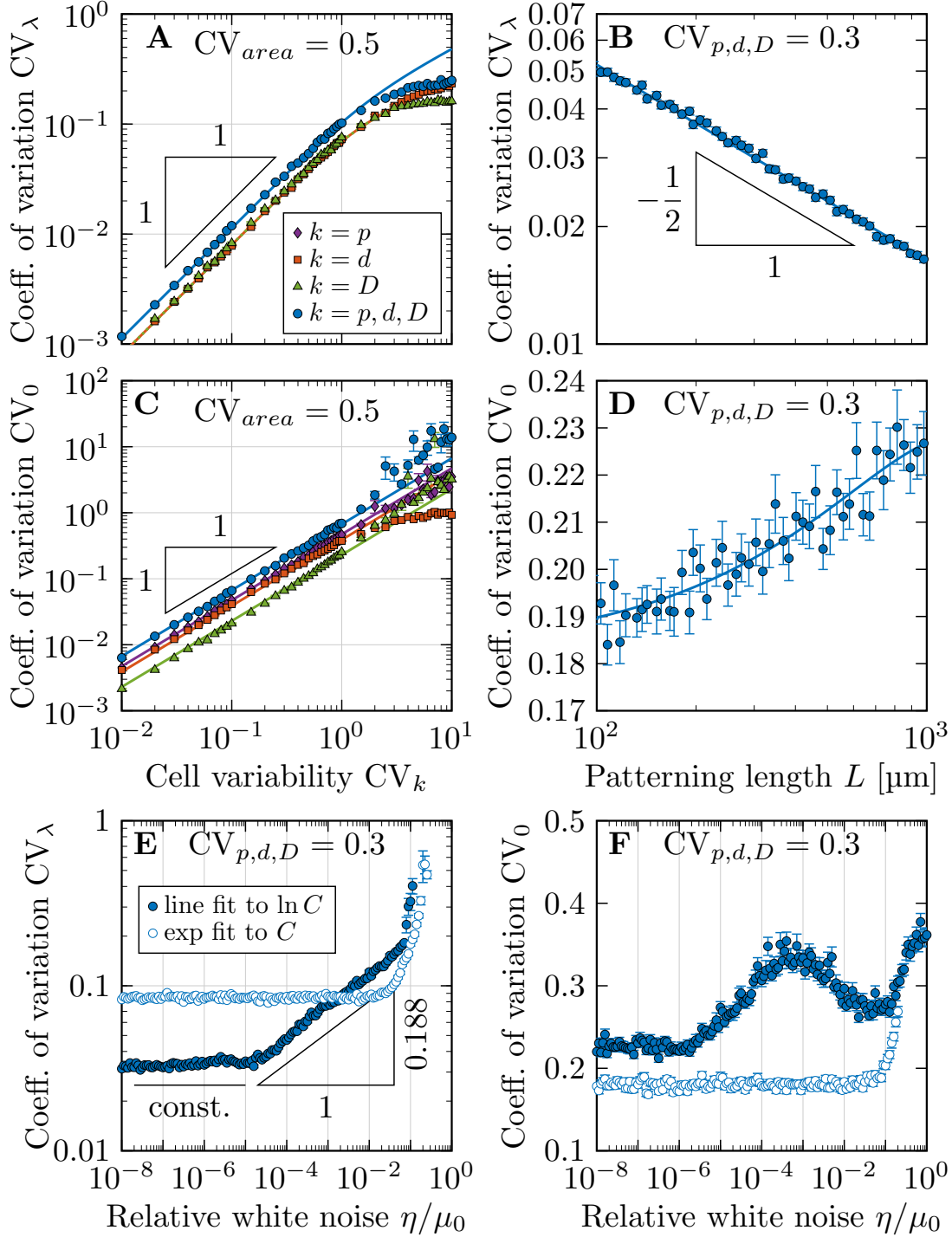


Fig. 11. Extending the simulations by Vetter and Iber [3]. In every simulation a CV_{area} value of 0.5 was added. **A** The kinetic parameters were drawn from a log-normal distribution with fixed mean and varying CV_k , where $k = p, d, D$. The variability in λ increases linearly until a saturation takes place for higher CV_k values. **B** Kinetic parameters are fixed at $CV_k = 0.3$. As the length of the domain is increased, the variability in λ drops. **C** The variability in the amplitude C_0 increases linearly as in **A**, but for $CV_k \gtrsim 1$ no saturation takes place. **D** An increase in the variability of the amplitude C_0 occurs as CV_k increases. **(E, F)** To account for random noise, Gaussian white noise $\mathcal{N}(0, \eta^2)$ was added to every cell. The additional noise lowered the range where a line could be fitted to (for small concentration values, noise dominates). Therefore also log-fitting was performed. **E** linear fitting leads to lower variability. **F** Log-fitting leads to an increased variability compared to linear fitting. Linear fitting remains constant until $\eta \gtrsim 10^{-1} \mu_0$.