Estimating the average growth rate from turbidostat data, a Bayesian approach

Péter Kómár

June 23, 2014

This is note contains the theory behind the `mufit` algorithm. Section 1 describes the data we are facing. Section 2 defines the problem: fitting the time dependent growth rate from noisy data. Section 3 describes the model, the way the data is filtered and gives the details of fitting. Section 4 contains the basics steps of using the `mufit` application.

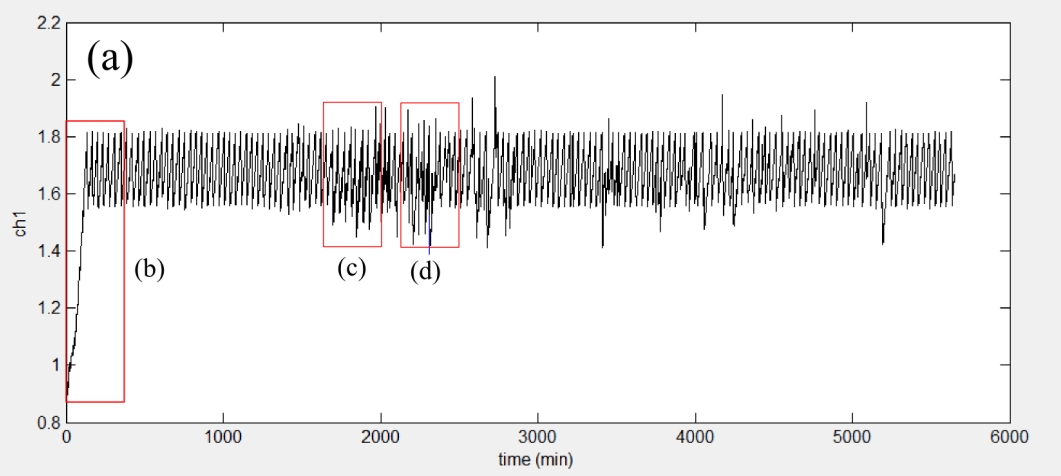
Contents

1 Measurement data

The turbidostat is designed to keep the cell concentration, n, below a certain threshold by measuring the optical absorbance (α) of a representative portion of the culture and diluting it if the absorbance exceeds a certain threshold level. This feedback procedure periodically lowers the cell concentration, allowing the cells to continue growing under unchanged conditions for extended periods of time. Continuous experiments of this kind run typically for multiple days.

Throughout a run, the optical absorbance, $\alpha(t)$, is recorded as a function of time. Typical features (shown on Fig. 1) of such a time series include

- uninterrupted rising at the beginning of the experiment, cells are let to grow, but their density is below the threshold and no dilution is happening,
- sudden drops right after the threshold is reached. The feedback mechanism triggers a dilution step, which quickly lowers n and α,
- short rising segments between dilution steps, corresponding to free growing, (see Fig. 1(b)),
- noise over all parts of the data, present with varying amplitude, (see Fig. 1(c)), and
- sudden jumps (both up and down) of large amplitude comparable to the height of the dilution drop, (see Fig. 1(d)).



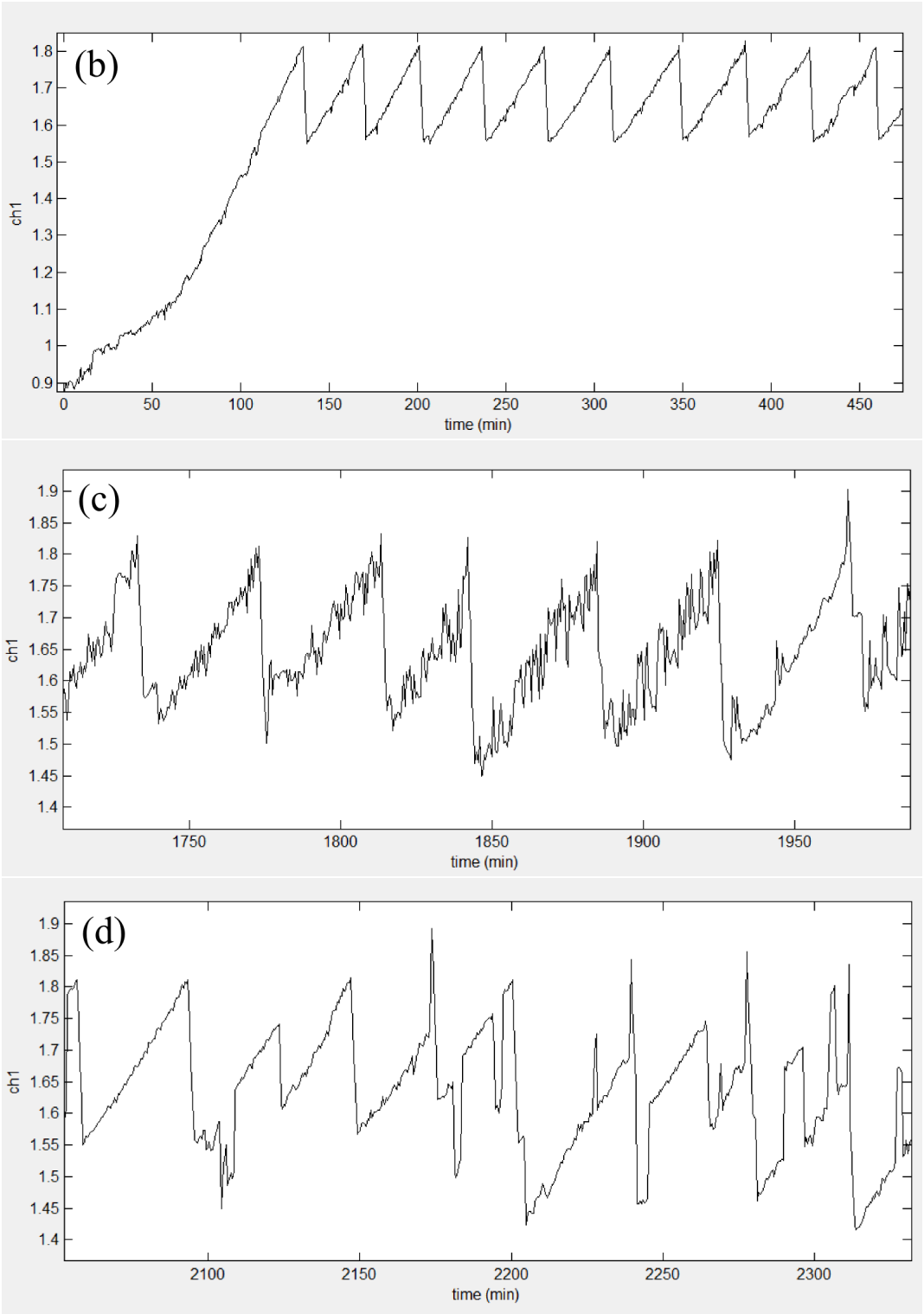


Figure 1: (a) Sample time series of $\alpha(t)$ with typical behaviors: (b) Rising start and smooth oscillations due to alternating diluting and free-growing steps. (c) Noise on top of the same oscillating pattern. (d) Sudden jumps interrupting the smooth growing segments between two diluting steps.

2 Average growth rate

Recordings of the optical absorbance as a function of time, $\alpha(t)$, can be used to determine the average growth rate of the cell population, $\mu(t)$, as a function of time. The differential equation governing the density, n, and its general solution are

$$\dot{n}(t) = \mu(t)n(t) \qquad \Rightarrow \qquad n(t) = n_0 \exp\left(\int_{t_0}^{t} dt\, \mu(t)\right)$$

Even if the proportionality constant c between α and n is unknown, we can find $\mu(t)$ from precise enough data on $\alpha(t)$ as follows

$$\alpha(t) = cn(t) \qquad \rightarrow \qquad \dot{\alpha}(t) = \mu(t)\alpha(t) \qquad \Rightarrow \qquad \mu(t) = \frac{\dot{\alpha}(t)}{\alpha(t)} = \frac{\dot{n}(t)}{n(t)}.$$

The periodically interrupted growth of α, by the diluting steps, and the additional noise in the recorded time series pose the largest obstacle in the determination of μ. A direct estimation of μ, with the naive finite differential formula,

$$\mu_{\text{est}}(t_j) = \frac{n(t_j) - n(t_{j-1})}{(t_j - t_{j-1})n(t_j)},$$

fails due to its inability to account for the additional noise.
Note: Since α and n are assumed to be related by only a proportionality constant c, from now on, we will denote the recorded time series with $n(t)$, but keep in mind that the actual values are not cell concentrations.

3 Estimation from noisy data

3.1 Model

After examining the data from Fig. 1, we model the expected behavior of n by

$$n(t) = n_{\text{smooth}} + n_{\text{noise}} + n_{\text{jumps}}, \tag{1}$$

where $n_{\text{smooth}}(t) = n_0 \exp(\mu(t)t)$, where $\mu(t)$ a slowly changing function of t, $n_{\text{noise}}(t) \sim \mathcal{N}(0, \sigma^2(t))$ is a white noise process with unknown and slowly changing time dependent variance, $\sigma^2(t)$, and n_{jumps} is the sum of appropriately timed offsets corresponding to jumps and drops of varying heights. Note that n_{jumps} tries to account for both the semi-regular drops from diluting steps and the random jumps shown on Fig. 1(d).

3.2 Locating the jumps

Although we separated the white noise part of the noise from the jumps above, in reality neither the noise is exactly white (it does exhibit short-time auto-correlation) and nor the jumps are infinitely narrow in time. In the presence of large noise, it becomes less clear how to categorize certain regions of the time series $n(t)$. (See Fig. 2)

Before tackling this problem we make a short digression and show how we find a rough estimate of μ, μ^*, which will serve as an initial guess of the growth rate. This will be used in our analysis, which will eventually supply a more precise estimate. In order to do so, we determine three important parameters of the time series:

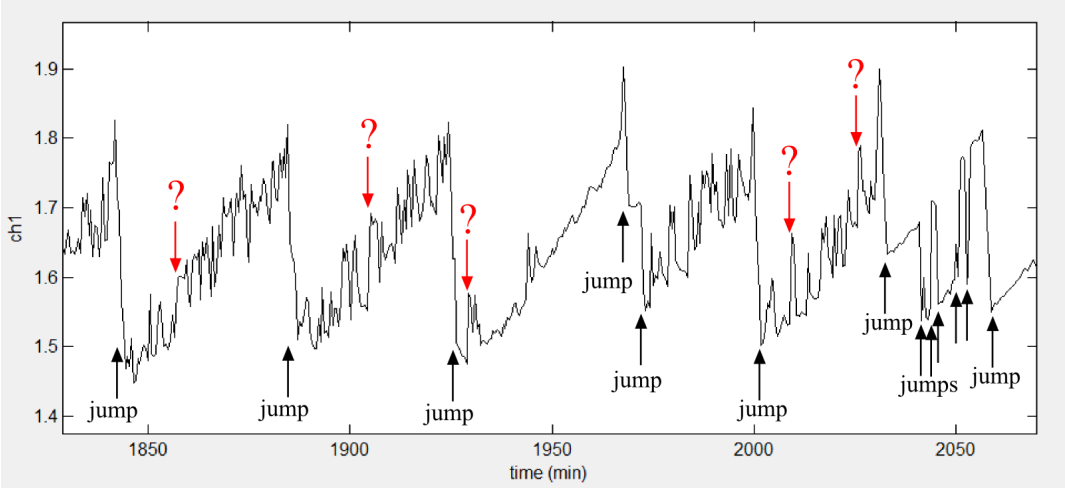


Figure 2: Discriminating between jumps and fluctuations is not an easy task even for the human eye. While we can easily categorize the features marked by black arrows as 'jumps', we are less sure about the features marked by red arrows.

- the average value of n in the oscillating part, $n_{\text{midline}} = \overline{n(t)}$,
- the average amplitude of the oscillations, $n_{\text{width}} = \sqrt{12\,\overline{(n(t) - n_{\text{midline}})^2}}$ (from the assumption of uniform distribution of n between its maximum and minumum), and
- a rough estimate of the period of the oscillations, t_{period} = 90th percentile of the waiting times between the data crossing the midline. (This definition picks out the period of the smooth cycles, which cross the midline only once in one cycle.)

These values are rough estimates, and require human feedback to confirm their approximate correctness. After this, μ^* is calculated as

$$\mu^* = \frac{1}{n_{\text{midline}}} \frac{n_{\text{width}}}{t_{\text{period}}} \qquad \left(\text{from} \quad \mu = \frac{1}{n} \frac{dn}{dt} \right).$$

We employ a simple scheme, designed to work with human feedback, which identify a short region (of length Δt) of the time series as 'jump' if the maximum and minimum values differ by a predefined threshold value, Δn_{limit}. In order to not have a bias towards down-jumps vs up-jumps, we subtract the expected growth slope from the $n(t)$ points in the region of Δt, and discriminate each region by comparing the maximal deviation of $\tilde{n}(t) = n(t) - \mu^* t$ with Δn_{limit}.

$$\max_{t_0 < t < t_0 + \Delta t} \tilde{n}(t) - \min_{t_0 < t < t_0 + \Delta t} \tilde{n}(t) = \begin{cases} < \Delta n_{\text{limit}} & \rightarrow \quad \text{'no jump' at} \quad t_0 + \frac{\Delta t}{2} \\ > \Delta n_{\text{limit}} & \rightarrow \quad \text{'jump' at} \quad t_0 + \frac{\Delta t}{2} \end{cases}$$

The result of this algorithm is shown on Fig. 3 for $\Delta t = 0.1\, t_{\text{period}}$ and $\Delta n = (0.35,\ 0.45,\ 0.6) n_{\text{width}}$. High Δn values allow for larger fluctuations, and therefore pick out only the more pronounced jumps, while lower Δn values discriminate more strictly and categorize regions with large noise as 'jumpy'.

It is clear that human feedback is required once again to choose the most appropriate values of Δt and especially Δn. The goal is to identify continuous segments of the time series which can be faithfully described by only the first two terms $(n_{\text{smooth}} + n_{\text{noise}})$ of the model from Eq. 1.

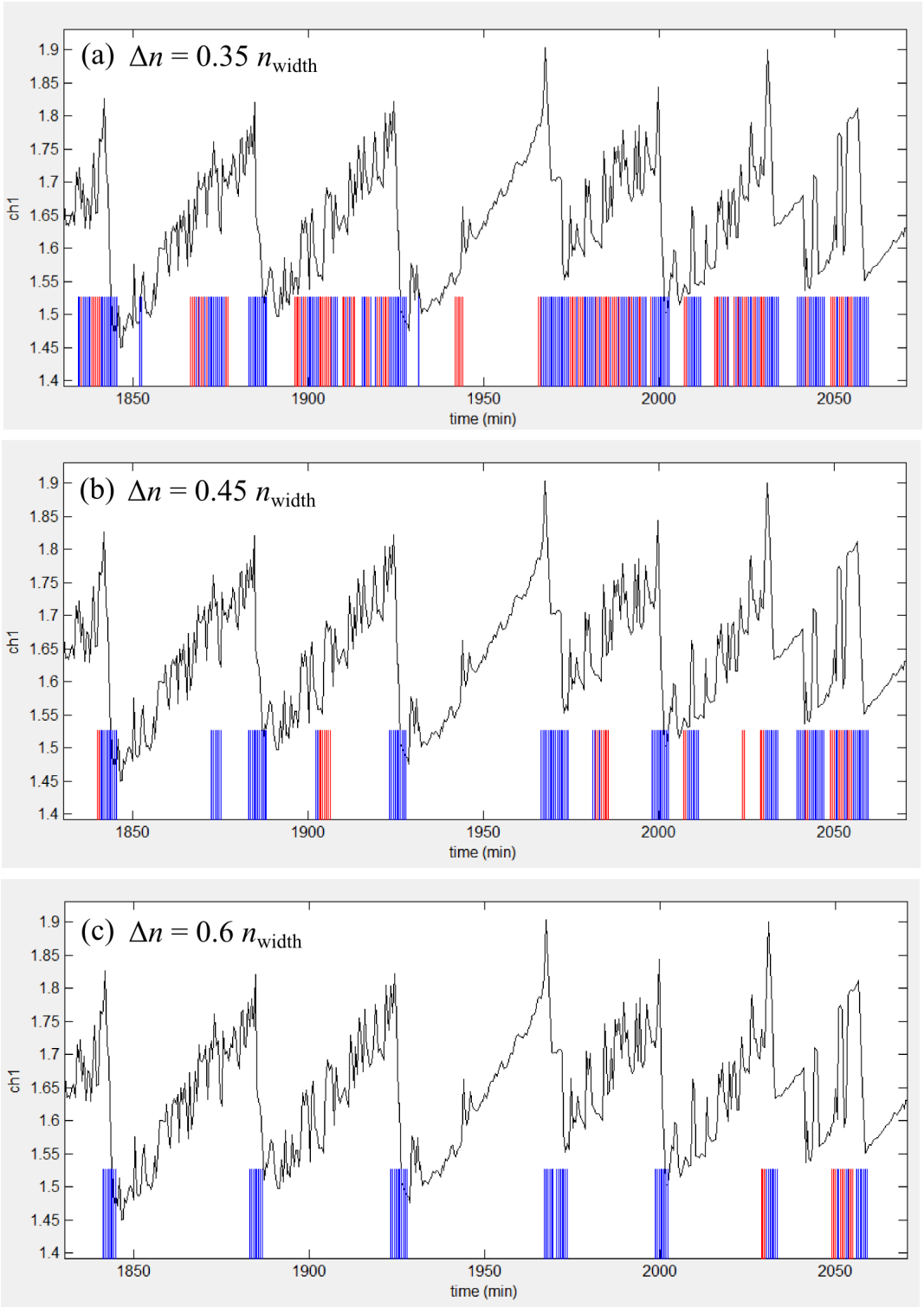


Figure 3: Jumps (blue for down, red for up) identified by our algorithm for $\Delta t = 0.1\, t_{\text{period}} = 3.72\,\text{min}$, $n_{\text{midline}} = 1.67$ and (a) $\Delta n = 0.35\, n_{\text{width}}$, (b) $\Delta n = 0.45\, n_{\text{width}}$ and (c) $\Delta n = 0.6\, n_{\text{width}}$, where $n_{\text{width}} = 0.28$. Compare with the guess on Fig. 2

3.3 Finding the smooth regions

The jumps, identified by the algorithm above, interrupt the smooth evolution of n. In principle, each segment between two jumps could be used to estimate μ, but if the segment is too short the fluctuations from n_{noise} will swamp the smooth signal of n_{smooth}. In order to arrive to an informative collection of regions, we require them to have a minimum length $t_{\text{min length}}$.

It is clear that $t_{\text{min length}} > \Delta t$ should be chosen, but its actual value should be set with human feedback once again. The goal is to keep most of the smooth regions, and filter out too short, noisy regions. By choosing $t_{\text{min length}} = 0.2\, t_{\text{period}} = 7.44\,\text{min}$ for the set of parameters used on Fig. 3(b), we arrive to the set of smooth regions shown in bold on Fig. 4

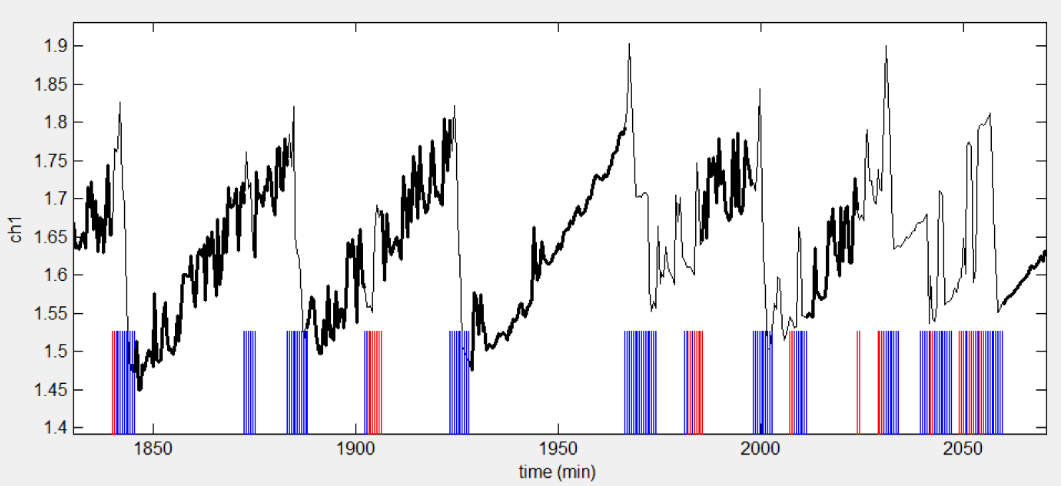


Figure 4: Smooth regions identified by our protocol for $t_{\text{min length}} = 0.2\, t_{\text{period}} = 7.44\,\text{min}$, and $\Delta t = 0.1\, t_{\text{period}} = 3.72\,\text{min}$, $n_{\text{midline}} = 1.67$, $\Delta n = 0.45\, n_{\text{width}}$, where $n_{\text{width}} = 0.28$

Note: One might be worried about the noise still present in the regions categorized as 'smooth' in the above example, however, we will show later that the Bayesien approach we are about to use extracts $\mu(t)$ while correctly taking the uncertainty of each region into account: High-noise regions will contribute with a lower weight to the estimation of the growth rate.

3.4 Fitting μ from a single region

For simplicity, we start by describing the estimation of μ from a single region R. We assume that both $\mu(t)$ and $\sigma^2(t)$ changes slowly enough that we can approximate them with constants throughout this single region. In other words we are fitting the model

$$\begin{aligned} n(t) &= n_0 e^{\mu t} + \text{Gaussian noise of variance } \sigma^2, \\ \text{or, equivalently} & \\ \mathcal{P}\big(t_j, n_j \mid n_0, \mu, \sigma^2\big) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(n_j - n_0 e^{\mu t_j})^2}{2\sigma^2}\right], \end{aligned} \tag{2}$$

where $\mathcal{P}(t_j, n_j \mid n_0, \mu, \sigma^2)$ is the probability density function of measuring the value n_j at time t_j given the values of n_0, μ and σ^2. Performing these steps analytically for our exponential model (Eq. 2), is not feasible due to the need of solving transcendental algebraic equations. We can choose a linear model, however,

$$\begin{aligned} \log n(t) &= \log n_0 + \mu t + \text{Gaussian noise of variance } \sigma^2, \\ \text{or, equivalently} & \\ \mathcal{P}\big(t_j, n_j \mid n_0, \mu, \sigma^2\big) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log n_j - \log n_0 - \mu t_j)^2}{2\sigma^2}\right], \end{aligned} \tag{3}$$

which approximates the exponential model well if the range of n is narrow enough compared to its average value. We can confirm this by looking at the oscillating region of Fig. 1(a).

This problem is identical to linear regression, which we can solve analytically. (See Appendix A for details.) The resulting estimates, and their uncertainties are

$$\begin{aligned} \log n_{0,\text{est}} \pm \Delta(\log n_0) &= \langle \log n \rangle - \mu_{\text{est}} \langle t \rangle \;\pm\; \frac{\sigma_{\text{est}}}{\sqrt{N}} \sqrt{\frac{\langle t^2 \rangle}{\langle t^2 \rangle - \langle t \rangle^2}}, \\ \mu_{\text{est}} \pm \Delta\mu &= \frac{\langle t \log n \rangle - \langle t \rangle \langle \log n \rangle}{\langle t^2 \rangle - \langle t \rangle^2} \;\pm\; \frac{\sigma_{\text{est}}}{\sqrt{N}\sqrt{\langle t^2 \rangle - \langle t \rangle^2}}, \\ \sigma^2_{\text{est}} \pm \Delta(\sigma^2) &= \Big\langle (\log n_{0,\text{est}} + \mu_{\text{est}} t - \log n)^2 \Big\rangle \;\pm\; \sqrt{\frac{2}{N}} \sigma^2_{\text{est}}, \end{aligned}$$

where N is the number of data points in region R and $\langle \cdot \rangle = \frac{1}{N} \sum_j (\cdot)_j$ are numerical values calculated from the data.

By repeating this method for each region R_k and obtaining the μ_k fits, we can get an estimate of the growth rate as a function of time, $\mu(\bar{t}_k) = \mu_k$, where $\bar{t}_k = \langle t \rangle_{R_k}$. This works well for long and smooth regions but it results in abrupt changes of μ_{est} from one region to the next if the noise of n is too large. (See Fig. 5(b).) We need to take our assumption about μ changing slowly into account in order to reduce the noise.

3.5 Fitting all $\{\mu_k\}$ from all regions simultaneously

The linear model described above can be formulated for all regions $\{R_k\}_{k=1}^K$ at once:

$$\mathcal{P}\Big(\{n_j\} \;\Big|\; \{t_j\}, \{R_k\}, \{(n_{0,k}, \mu_k, \sigma_k^2)\}\Big) = \prod_{k=1}^{K} \prod_{j \in R_k} \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left[-\frac{(\log n_j - \log n_{0,k} - \mu_k(t_j - \bar{t}_k))^2}{2\sigma_k^2}\right], \quad (4)$$

where $\bar{t}_k = \langle t \rangle_{R_k} = \frac{1}{N_k} \sum_{j \in R_k} t_j$, and N_k is the number of data point in region R_k. (We define the model with the t values of each region shifted to have zero mean in order to eliminate the correlation between $n_{0,k}$ and μ_k, as we will see below.)

This formulation allows us to define a joint prior for $\{\mu_k\}$ that incorporates our expectation of $\mu(t)$ changing slowly. For now, we use

$$\begin{aligned} &\mu_k = \mu_{k-1} + \text{Brownian motion} \\ &\mathcal{P}_0\Big(\{(n_{0,k}, \mu_k, \sigma_k^2)\} \;\Big|\; \{t_j\}, \{R_k\}\Big) \propto \prod_{k=2}^{K} \exp\left[-\frac{(\mu_k - \mu_{k-1})^2}{2D(\bar{t}_k - \bar{t}_{k-1})}\right], \end{aligned} \quad (5)$$

where D is an appropriately chosen diffusion constant of μ. A detailed argument for this is given in Appendix B.1.

Now, we are ready to use Bayes' theorem to express the posterior distribution of the fitting parameters as

$$\mathcal{P}\Big(\{(\log n_{0,k}, \mu_k, \sigma_k^2)\} \;\Big|\; \{(t_j, n_j)\}, \{R_k\}\Big) \quad \propto \quad \Big[\text{model from Eq. 4}\Big] \times \Big[\text{prior from Eq. 5}\Big].$$

Let us maximize the logarithm of this, $\mathcal{L} = \log \mathcal{P}$, by finding the optimal values of the fit parameters. To facilitate comparing the following calculations with similar results of linear regression (see Appendix A), we switch to the notation $y = \log n$, $x = t$, $b = \log n_0$ and $a = \mu$. At the point where $\mathcal{L}$ takes its maximum,

$$\partial_{b_k}\mathcal{L} = -\frac{1}{\sigma_k^2} \sum_{j \in R_k} [b_k + a_k(x_j - \bar{x}_k) - y_j] = 0, \quad (6)$$

$$\partial_{a_k}\mathcal{L} = -\frac{a_k - a_{k-1}}{D\Delta x_k} - \frac{a_k - a_{k+1}}{D\Delta x_{k+1}} - \frac{1}{\sigma_k^2} \sum_{j \in R_k} [b_k + a_k(x_j - \bar{x}_k) - y_j]\,(x_j - \bar{x}_k) = 0, \quad (7)$$

$$\partial_{\sigma_k^2}\mathcal{L} = -\frac{N_k}{2\sigma_k^2} + \sum_{j \in R_k} \frac{[b_k + a_k(x_j - \bar{x}_k) - y_j]^2}{2\sigma_k^4} = 0, \quad (8)$$

where $\Delta x_k = \bar{x}_k - \bar{x}_{k-1}$, $k = 1, 2 \ldots K$ and $x_0 = -\infty$, $x_{K+1} = +\infty$ by definition. (This avoids coupling a_1 and a_K to the non-existent a_0 and a_{K+1}, respectively.) The solution of this set of coupled equations provide the estimates,

$$\mu_k = a_k \quad \text{and} \qquad n_{0,k} = \exp(b_k),$$

and define the fitted curves

$$n_k(t) = n_{0,k} \exp\Big[\mu_k(t - \bar{t}_k)\Big] \quad \text{in each region } R_k.$$

The values of a_k, b_k and their uncertainties are determined efficiently in a numerical way. See Appendix B.2 and B.3 for details.

3.6 Choosing the optimal D

How much the above protocol takes the neighboring regions into account is governed by the value of D. Its value needs to be chosen as an honest representation of how much we expect $\mu(t)$ to change from region to region. This is one major input information, supplied by the human operator. To facilitate this we express D with another quantity: χ, which stands for the expected maximum of relative change of μ over one doubling time, $t_{\text{doubling}} = \log(2)/\mu$.

The value D tells us the expected maximal change of μ in time dt,

$$d\mu_{\text{typical}} = \sqrt{D\,dt}.$$

The relative change is

$$\frac{d\mu_{\text{typical}}}{\mu^*} = \frac{\sqrt{D\,t_{\text{doubling}}}}{\mu^*} \approx \sqrt{D}\frac{\sqrt{\log(2)}}{(\mu^*)^{3/2}} =: \chi.$$

Now, $\chi = 0.05 = 5\%$ means that we expect μ to change typically not more than 5% of its value over one doubling time, and so $D = \frac{(0.05)^2}{\log(2)}(\mu^*)^3$.

Large values of χ (and D) allow for detecting rapid changes of the fitted growth rate, while small values keep it a slowly changing function. As an example, we show the fitted $\mu(t)$ on Fig. 5(b) and (c) for $\chi = 100\%$ and $\chi = 5\%$, respectively.

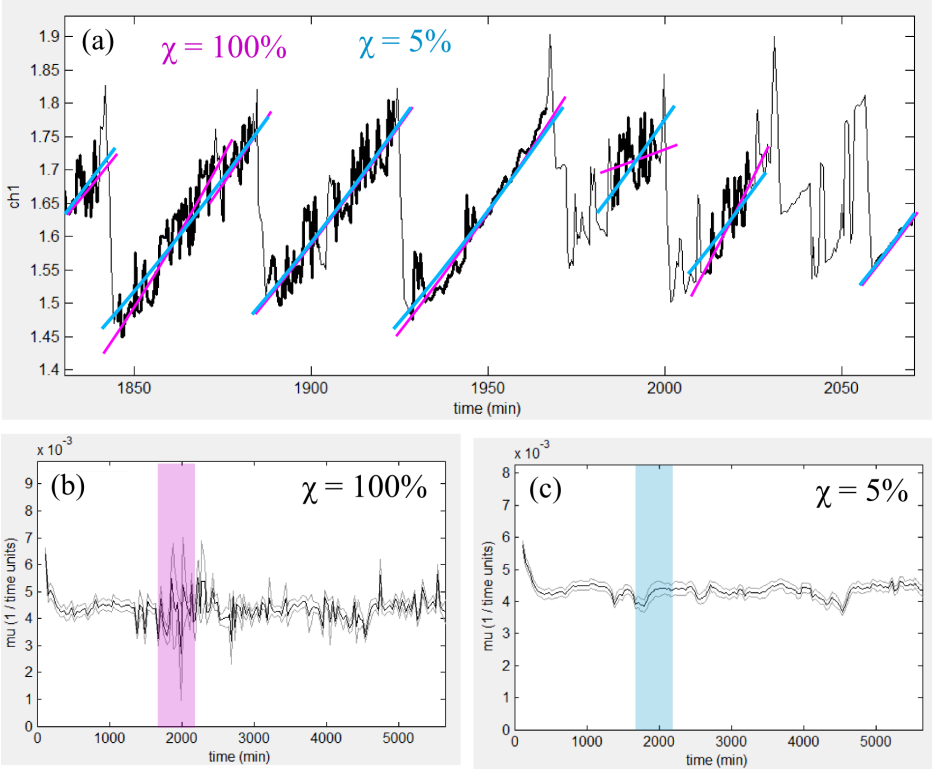


Figure 5: Illustration of the effect of large (magenta) and small (cyan) χ. (a) A notoriously noisy region of a time series with fitted exponentials shown for $\chi = 1$ (practically independent fits) and $\chi = 0.05$. The entire $\mu(t)$ series is shown on (b) and (c) for these values of χ, respectively. Gray lines indicate 95% confidence bounds. Highlighted regions correspond to the part shown on (a).

3.7 A smoother prior for $\{\mu_k\}$

The prior distribution defined by Eq. (5) assumes that $\mu(t)$ behaves as a Brownian motion, (see appendix C.3). This may not be an accurate description of reality. In case of a smooth change of the cell population, μ is also expected to change smoothly under constant environmental conditions (see Appendix C.5).

We can describe this smooth behavior more closely if we model $\mu(t)$ with an integrated Brownian motion, (see appendix C.4). This results in the prior

$$\mu_k = \mu_{k-1} + (\bar{t}_k - \bar{t}_{k-1}) \left(\frac{\mu_{k-1} - \mu_{k-2}}{\bar{t}_{k-1} - \bar{t}_{k-2}} + \text{Brownian motion} \right)$$

$$\mathcal{P}_0\Big(\{\mu_k\} \;\Big|\; \{\bar{t}_k\}\Big) \propto \prod_{k=3}^{K} \exp\left[-\frac{(\nu_k - \nu_{k-1})^2}{\frac{2D}{3}(\bar{t}_k - \bar{t}_{k-2})} \right], \qquad \text{where} \quad \nu_k = \frac{\mu_k - \mu_{k-1}}{\bar{t}_k - \bar{t}_{k-1}}.$$

After combining it with Eq. (4) to get the posterior distribution $\mathcal{P}$, we can maximize its logarithm $\mathcal{L} = \log \mathcal{P}$ by solving the following system of equations. We use $y = \log n, x = t, a = \mu$ and $b = \log n_0$ notation once again.

$$\partial_{b_k} \mathcal{L} = -\frac{1}{\sigma_k^2} \sum_{j \in R_k} [b_k + a_k(x_j - \bar{x}_k) - y_j] = 0 \qquad \text{(identical to Eq. (6))} \tag{9}$$

$$\begin{aligned} \partial_{a_k} \mathcal{L} = {} & -h_{k,k-2} a_{k-2} - h_{k,k-1} a_{k-1} - h_{k,k} a_k - h_{k,k+1} a_{k+1} - h_{k,k+2} a_{k+2} \\ & -\frac{1}{\sigma_k^2} \sum_{j \in R_k} [b_k + a_k(x_j - \bar{x}_k) - y_j] (x_j - \bar{x}_k) = 0 \end{aligned} \tag{10}$$

$$\partial_{\sigma_k^2} \mathcal{L} = -\frac{N_k}{2\sigma_k^2} + \sum_{j \in R_k} \frac{[b_k + a_k(x_j - \bar{x}_k) - y_j]^2}{2\sigma_k^4} = 0 \qquad \text{(identical to Eq. (8))}, \tag{11}$$

where

$$\begin{aligned} h_{k,k-2} &= \frac{3}{D} \frac{1}{\Delta t_k \Delta t_{k-1} (\Delta t_k + \Delta t_{k-1})}, \\ h_{k,k-1} &= -\frac{3}{D} \frac{1}{(\Delta t_k)^2} \left(\frac{1}{\Delta t_{k-1}} + \frac{1}{\Delta t_{k+1}} \right), \\ h_{k,k} &= \frac{3}{D} \left[\frac{1}{(\Delta t_k)^2 (\Delta t_k + \Delta t_{k-1})} + \frac{1}{\Delta t_k \Delta t_{k+1}} \left(\frac{1}{\Delta t_k} + \frac{1}{\Delta t_{k+1}} \right) \right], \\ h_{k,k+1} &= -\frac{3}{D} \frac{1}{(\Delta t_{k+1})^2} \left(\frac{1}{\Delta t_k} + \frac{1}{\Delta t_{k+2}} \right), \\ h_{k,k+2} &= \frac{3}{D} \frac{1}{\Delta t_{k+1} \Delta t_{k+2} (\Delta t_{k+1} + \Delta t_{k+2})}, \end{aligned}$$

where $\Delta t_k = \bar{t}_k - \bar{t}_{k-1}$, and by definition $\Delta t_0 = \Delta t_1 = \Delta t_{K+1} = \Delta t_{K+2} = \infty$ to eliminate the coupling to the non-existent a_{-1}, a_0 and a_{K+1}, a_{K+2} variables.

Solving Eq. (9), (10) and (11) for $\{(b_k, a_k, \sigma_k^2)\}$ and finding their uncertainties can be performed with the procedure described in Appendix B.2 and B.3 with the slight change: now matrix A needs to be

$$A_{k,l} = m_k \delta_{k,l} + h_{k,l}, \qquad \text{where} \quad m_k = \frac{N_k}{\sigma_k^2} \langle (x - \bar{x}_k)^2 \rangle_k,$$

and $h_{k,l} = 0$ if $|k - l| > 2$.

To choose the optimal value of D, we again express it with the expected maximal relative deviation of the growth rate $\Delta\mu$ from the linear course, $\nu\Delta t$ in one doubling time, denoted by χ. The change of μ in time Δt is

$$\Delta\mu = \Delta t\,(\nu + \Delta\nu) = \nu\Delta t + \sqrt{D(\Delta t)^3},$$

where $\nu = \frac{d\mu}{dt}$ is exhibiting Brownian motion with diffusion constant D, (see Appendix C.4 for details). The relative deviation from the linear course in $t_{\text{doubling}} = \frac{\log 2}{\mu^*}$ is

$$\frac{\Delta\mu - \nu\Delta t}{\mu^*} = \frac{\sqrt{D}t_{\text{doubling}}^{3/2}}{\mu^*} = \frac{\sqrt{D}[\log(2)]^{3/2}}{(\mu^*)^{5/2}} =: \chi.$$

This means that if $\chi = 0.05$, then $D = \frac{(0.05)^2}{[\log(2)]^3}(\mu^*)^5$.

On Fig. 6 we compare the fitted $\mu(t)$ for the case of the Brownian and the integrated Brownian priors, for $\chi = 10\%$.

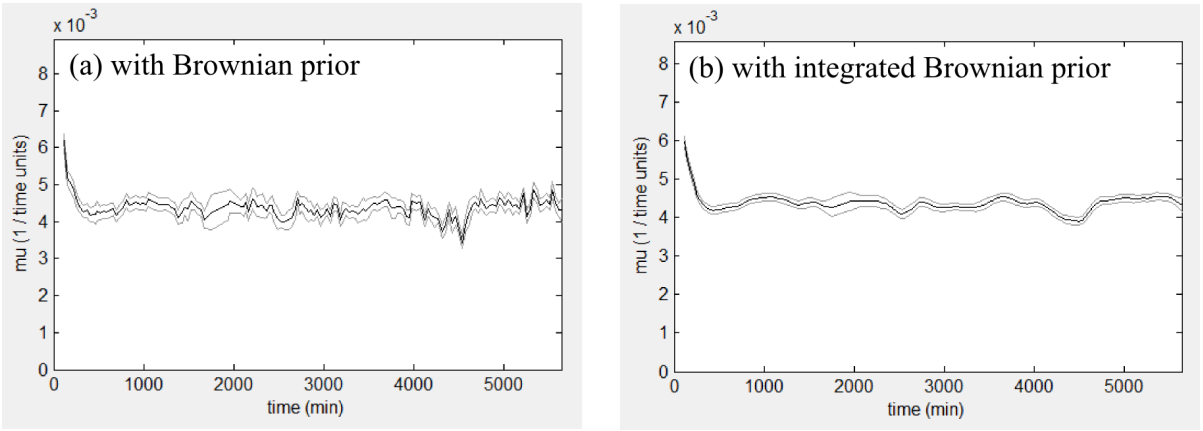


Figure 6: Comparison of the resulting fits from the two stochastic models: (a) The usage of Brownian motion prior results in a zigzag behavior, and larger fluctuations on short timescales. (b) The usage of integrated Brownian motion prior gives a more smoother and slower evolution. Confidence intervals of the 0.95 level are shown with gray lines on both plots.

4 Summary of steps of analysis

After loading the raw data the following steps need to be done:

1. Determine the values of n_{midline}, n_{width} and t_{period}. They should give a rough idea for the algorithm about the scales of the data.

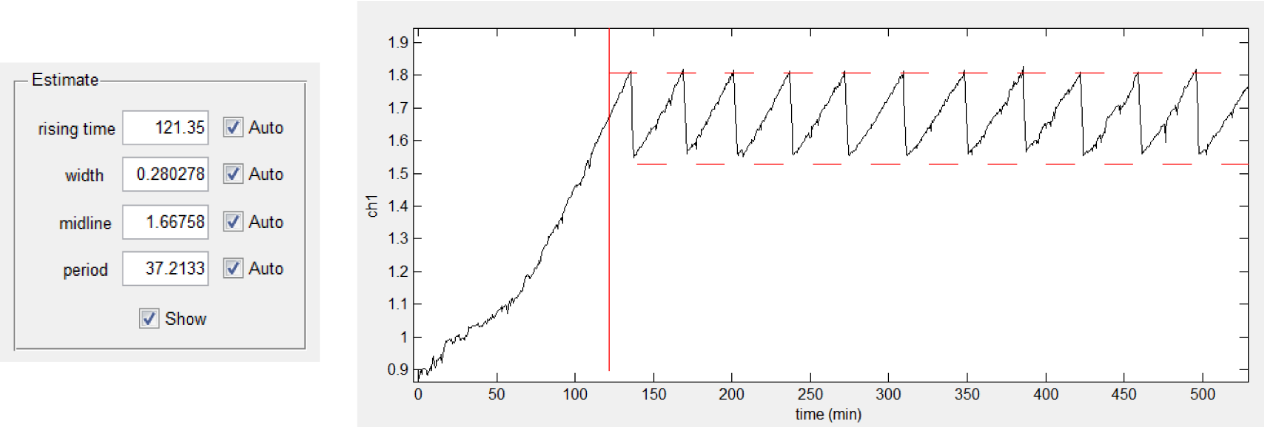


Figure 7: The additional parameter shown above is the length of the initial rising region (the end of which is plotted as a vertical red line). Its value helps the 'Auto' function determine the values of the other three. n_{midline} and n_{width} are jointly depicted as the typical maximum and minimum of the oscillating part of the data, indicated by the level of the two horizontal dashed lines. The period of the dashing indicates t_{period}.

2. Set the time window Δt, the noise limit Δn and the minimum smooth region length $t_{\text{min length}}$, such that the selected regions are not interrupted by jump-like features.

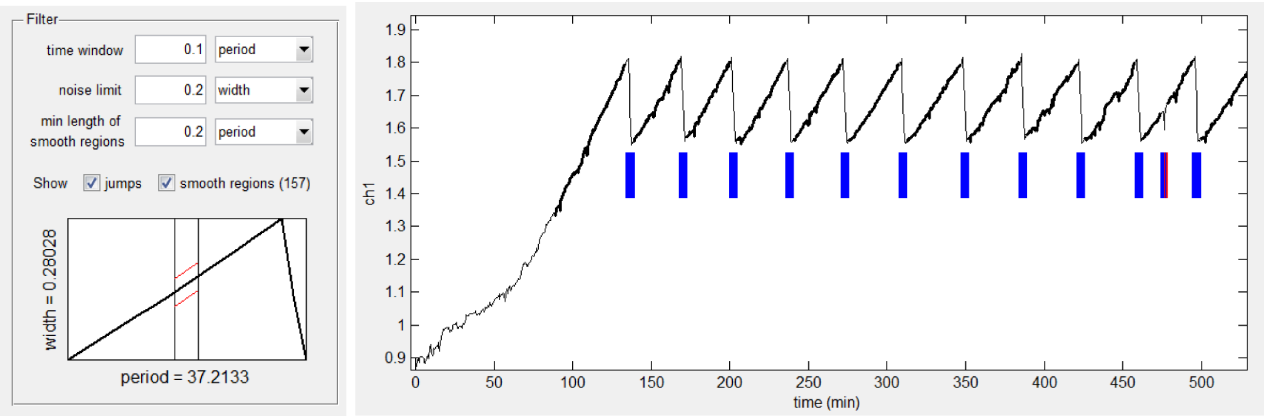


Figure 8: The panel on the left has an additional graphical representation of the filter defined by the numerical values of Δt and Δn. Up and down jumps are marked with red and blue sticks below the data, as determined by the filter. Smooth regions selected by the filter are highlighted by being plotted as a thicker line.

3. Set χ and the type of stochastic model such that they honestly reflect the initial idea about how the growth rate is expected to change.

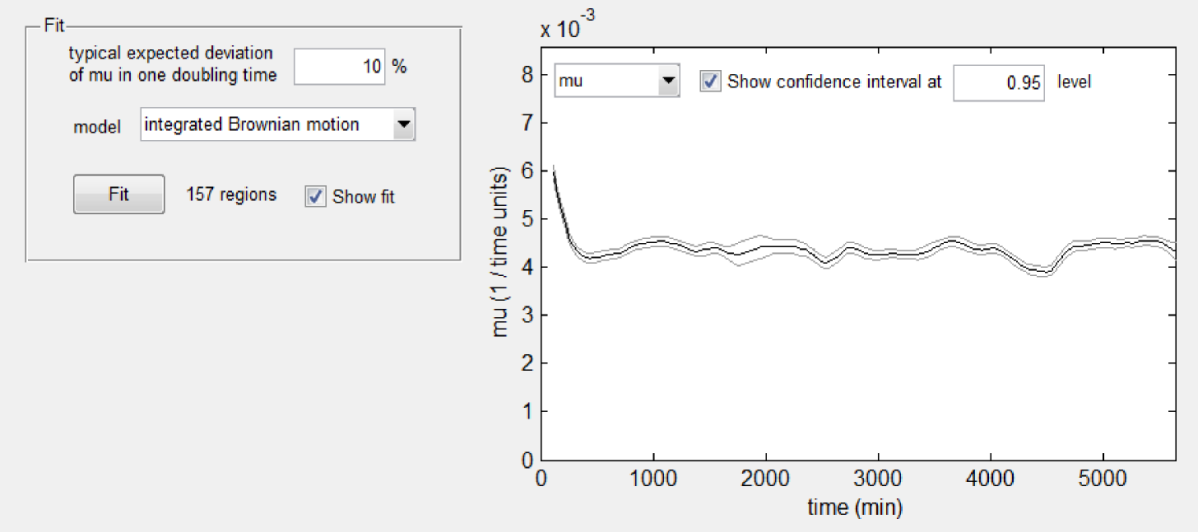


Figure 9: Beside the value of χ, confidence intervals (gray lines) shown on the $\mu(t)$ plot can also be customized.

4. After obtaining the fitted $\mu(t)$ it is worth checking the significance of any outlying values by comparing the fitted curves with the raw data in the respective region.

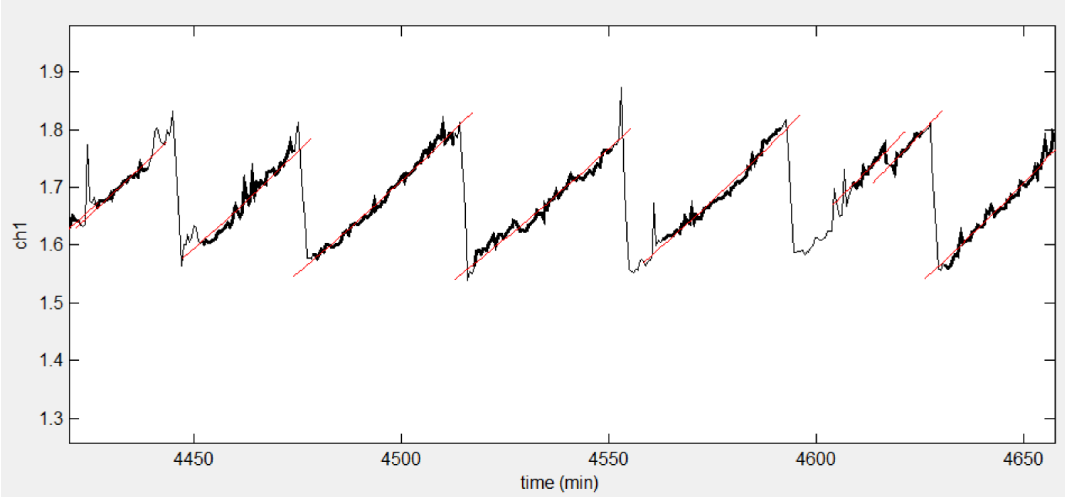


Figure 10: The fitted $\mu(t)$ (on Fig. 9) shows a dip around $t = 4500$, we can check the goodness of the fit near this time point by zooming on its neighborhood. In this example the fit is good enough to conclude this to be a significant feature.

A Linear regression

Given the data $D = \{(x_j, y_j)\}$, we are fitting the model

$$\mathcal{P}(x_j, y_j \mid b, a, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_j - b - ax_j)^2}{2\sigma^2}\right].$$

For simplicity, we are going to use a constant prior $\mathcal{P}(b, a, \sigma^2) = \text{const.}$, which renders the Bayesian scheme identical to maximum likelihood estimation. The likelihood function is

$$\mathcal{L}(b, a, \sigma^2) = \log \mathcal{P}(D \mid b, a, \sigma^2) = \log \prod_j \mathcal{P}(x_j, y_j \mid b, a, \sigma^2) = \sum_j \log \mathcal{P}(x_j, y_j \mid b, a, \sigma^2),$$

which needs to be maximized by finding the optimal values of b, a, σ^2. At the maximum point,

$$\partial_b \mathcal{L} = -\frac{1}{\sigma^2} \sum_j (ax_j + b - y_j) = 0, \tag{12}$$

$$\partial_a \mathcal{L} = -\frac{1}{\sigma^2} \sum_j (ax_j + b - y_j) x_j = 0, \tag{13}$$

$$\partial_{\sigma^2} \mathcal{L} = -\frac{1}{\sigma^2} \sum_i \left[1 - \frac{(ax_j + b_j - y_j)^2}{\sigma^2}\right] = 0. \tag{14}$$

Eq. (12) and (13) form a linear system of equations for (b, a), which can be solved and used to find σ^2. This gives

$$\begin{aligned} b_{\text{est}} &= \langle y \rangle - a_{\text{est}} \langle x \rangle, \\ a_{\text{est}} &= \frac{\langle yx \rangle - \langle y \rangle \langle x \rangle}{\langle x^2 \rangle - \langle x \rangle^2}, \\ \sigma^2_{\text{est}} &= \left\langle (b_{\text{est}} + a_{\text{est}} x_j - y_j)^2 \right\rangle, \end{aligned}$$

where $\langle \cdot \rangle = \frac{1}{N} \sum_j (\cdot)_j$, are numerical values calculated from the data $\{(x_j, y_j)\}$, and N is the total number of data points.

The uncertainties of these estimates need to be determined from the covariance matrix C, the inverse of which is (approximately) given by

$$C^{-1} \approx -[\partial^2 \mathcal{L}]_{\text{est}} = -\begin{bmatrix} \partial_b^2 \mathcal{L} & \partial_b \partial_a \mathcal{L} & \partial_b \partial_{\sigma^2} \mathcal{L} \\ \partial_a \partial_b \mathcal{L} & \partial_a^2 \mathcal{L} & \partial_a \partial_{\sigma^2} \mathcal{L} \\ \partial_{\sigma^2} \partial_b \mathcal{L} & \partial_{\sigma^2} \partial_a \mathcal{L} & \partial_{\sigma^2}^2 \mathcal{L} \end{bmatrix}_{\text{est}} = \frac{N}{\sigma^2_{\text{est}}} \begin{bmatrix} 1 & \langle x \rangle & 0 \\ \langle x \rangle & \langle x^2 \rangle & 0 \\ 0 & 0 & \frac{1}{2\sigma^2_{\text{est}}} \end{bmatrix}.$$

The covariance matrix is then

$$C = \frac{\sigma^2_{\text{est}}}{N} \begin{bmatrix} \frac{1}{\langle x^2 \rangle - \langle x \rangle^2} \begin{bmatrix} \langle x^2 \rangle & -\langle x \rangle \\ -\langle x \rangle & 1 \end{bmatrix} & \\ & 2\sigma^2_{\text{est}} \end{bmatrix}.$$

The variances of each estimated parameter (from its marginal) are given by the diagonal elements. The standard deviations are

$$\Delta b = \frac{\sigma_{\text{est}}}{\sqrt{N}} \sqrt{\frac{\langle x^2 \rangle}{\langle x^2 \rangle - \langle x \rangle^2}}, \qquad \Delta a = \frac{\sigma_{\text{est}}}{\sqrt{N} \sqrt{\langle x^2 \rangle - \langle x \rangle^2}}, \qquad \Delta \sigma^2 = \sqrt{\frac{2}{N}} \sigma^2_{\text{est}}.$$

B Details of simultaneous fitting

B.1 Priors for the growth rate

The formulation given by Eq. 4 allows us to define a prior of the fitting parameters that exhibits correlations between different μ_k values:

$$\mathcal{P}_0\Big(\{(n_{0,k},\mu_k,\sigma_k^2)\}\Big) = \left(\prod_k \mathcal{P}_0(\log n_{0,k})\right)\left(\prod_k \mathcal{P}_0(\sigma_k^2)\right)\mathcal{P}_0\Big(\{\mu_k\}\Big) \propto \mathcal{P}_0\Big(\{\mu_k\}\Big),$$

where we have chosen independent, constant priors for $\log n_{0,k}$ and σ_k^2, but allow for correlations between neighboring μ_k values. Specifically, we formalize our prior knowledge about the growth rate, which reads "consecutive values of μ don't differ much", in the following way:

1. Instead of $\mu_1, \mu_2, \dots \mu_K$, we express the prior probability in terms of $\mu_1, (\mu_2 - \mu_1), (\mu_3 - \mu_2), \dots (\mu_K - \mu_{K-1})$,
2. we prescribe the prior of μ_1 to be $\mathcal{P}_0(\mu_1) = \text{const.}$, and
3. we prescribe the prior of the differences $\mu_k - \mu_{k-1}$ to be Gaussians with width $\sqrt{D(\bar{t}_k - \bar{t}_{k-1})}$,

$$\mathcal{P}_0(\mu_k - \mu_{k-1}) = \frac{1}{\sqrt{2\pi D(\bar{t}_k - \bar{t}_{k-1})}} \exp\left[-\frac{(\mu_k - \mu_{k-1})^2}{2D(\bar{t}_k - \bar{t}_{k-1})}\right],$$

   where D is the diffusion constant of a Brownian motion. It has to be adjusted to faithfully reflect our prior knowledge about the rate with μ itself is changing.

The total prior (given the time coordinates and the selected regions) is therefore

$$\begin{aligned} &\mu_k = \mu_{k-1} + \text{Gaussian noise} \\ &\mathcal{P}_0\Big(\{(n_{0,k},\mu_k,\sigma_k^2)\} \;\Big|\; \{t_j\},\{R_k\}\Big) \propto \prod_{k=2}^{K} \exp\left[-\frac{(\mu_k - \mu_{k-1})^2}{2D(\bar{t}_k - \bar{t}_{k-1})}\right], \end{aligned} \tag{15}$$

the Brownian prior, which we use in Eq. 5.

An even smoother prior can be defined by requiring the continuity the derivative of μ instead of μ itself, and prescribing the Gaussian distribution for its increments:

$$\begin{aligned} &\mu_k = \mu_{k-1} + (\bar{t}_k - \bar{t}_{k-1})\left(\frac{\mu_{k-1} - \mu_{k-2}}{\bar{t}_{k-1} - \bar{t}_{k-2}} + \text{Gaussian noise}\right) \\ &\mathcal{P}_0\Big(\{(n_{0,k},\mu_k,\sigma_k^2)\} \;\Big|\; \{t_j\},\{R_k\}\Big) \propto \prod_{k=3}^{K} \exp\left[-\frac{(\nu_k - \nu_{k-1})^2}{\frac{2D}{3}(\bar{t}_k - \bar{t}_{k-2})}\right], \qquad \text{where} \quad \nu_k = \frac{\mu_k - \mu_{k-1}}{\bar{t}_k - \bar{t}_{k-1}}. \end{aligned} \tag{16}$$

This is the integrated Brownian prior.

Detailed derivation of both Eq. (15) and (16) from the corresponding stochastic processes is given in Appendix C.

B.2 Solving Eq.(6)-(8)

Since $\sum_j (x_j - \bar{x}_k) = 0$, Eq. (6) decouples from the rest, and can be solved separately, giving

$$b_k = \frac{1}{N_k} \sum_{j \in R_k} y_j = \langle y \rangle_k .$$

Unfortunately, a_k remains coupled not just to σ_k^2 but also to a_{k+1} and a_{k-1}. Using $b_k = \langle y \rangle_k$, we write Eq. (7) and (8) as

$$\begin{aligned} v_k &= a_k(g_k + g_{k+1} + m_k) - a_{k-1} g_k - a_{k+1} g_{k+1} =: \sum_l A_{kl} a_l , \\ \sigma_k^2 &= \langle y^2 \rangle_k - \langle y \rangle_k^2 - 2 a_k \langle y \tilde{x} \rangle_k + a_k^2 \langle \tilde{x}^2 \rangle_k , \end{aligned}$$

where $\tilde{x} = x - \bar{x}_k$ is the shifted coordinate in each region R_k, the regional average is defined as $\langle \cdot \rangle_k = \frac{1}{N_k} \sum_{j \in R_k} (\cdot)_j$, and the coefficients are $g_k = \frac{1}{D \Delta x_k}$, $m_k = \frac{N_k}{\sigma_k^2} \langle \tilde{x}^2 \rangle_k$ and $v_k = \frac{N_k}{\sigma_k^2} \langle y \tilde{x} \rangle_k$.

Since m_k and v_k depend on σ_k^2, these equations form a set of $2K$ non-linear equations for $\{a_k\}$ and $\{\sigma_k^2\}$. We solve them with the following iterative procedure:

1. Solve for a_k while setting all $g_k = 0$. This gives $a_k^{(0)} = \frac{v_k}{m_k} = \frac{\langle y \tilde{x} \rangle_k}{\langle \tilde{x}^2 \rangle_k} =: a_k^{(i)}$, for $i = 0$.

2. Calculate $(\sigma_k^2)^{(i)} = \langle y^2 \rangle_k - \langle y \rangle_k^2 - 2 a_k^{(i)} \langle y \tilde{x} \rangle_k + \left(a_k^{(i)} \right)^2 \langle \tilde{x}^2 \rangle_k$.

3. Calculate $m_k^{(i)} = \frac{N_k}{\left(\sigma_k^2\right)^{(i)}} \langle \tilde{x}^2 \rangle_k$ and $v_k^{(i)} = \frac{N_k}{\left(\sigma_k^2\right)^{(i)}} \langle y \tilde{x} \rangle_k$.

4. Solve the now linear set of K equations,

$$\begin{aligned} \vec{v}^{(i)} &= \mathbf{A}^{(i)} \, \vec{a} \\ v_k^{(i)} &= a_k(g_k + g_{k+1} + m_k^{(i)}) - a_{k-1} g_k - a_{k+1} g_{k+1} \end{aligned}$$

   for $\{a_k\}$, where $g_k = \frac{1}{D \Delta x_k}$ for $k = 1, 2 \ldots K$, and $g_0 = g_{K+1} = 0$. This gives the values $\{a_k^{(i+1)}\}$.

5. If the relative difference between $a_k^{(i+1)}$ and $a_k^{(i)}$, is smaller than a predefined limit ϵ for all k, then we can declare $\left\{ a_k^{(i+1)} \right\}, \left\{ (\sigma_k^2)^{(i+1)} \right\}$ to be the solution. Otherwise we repeat from step 2 with $i \mapsto i + 1$.

This procedure finds the solution for arbitrary D in only a few iterations, (typically 3-5, for $\epsilon = 10^{-4}$). Its stability is due to the fact that both $D \to 0$ and $D \to \infty$ limits have a unique and bounded solution:

- $D \to 0$ makes the couplings g_k dominate, and forces all a_k have the same value.
- $D \to \infty$ decouples the neighboring regions, and results in independent fits for $\{a_k\}$.

B.3 Uncertainties

The uncertainties of the estimates from Eq. (6)-(8) need to be determined from the covariance matrix C, the inverse of which is (approximately)

$$C^{-1} \approx -\left[\partial^2 \mathcal{L}\right]_{\rm est} = \begin{bmatrix} B & 0 & 0 \\ 0 & A & G \\ 0 & G & S \end{bmatrix}, \qquad \text{a } 3K \times 3K \text{ matrix,}$$

where B is a diagonal matrix:

$$\begin{aligned} -\partial^2_{b_k}\mathcal{L} &= B_{k,k} = \frac{N_k}{\sigma_k^2} \\ -\partial_{b_k}\partial_{b_l}\mathcal{L} &= B_{k,l} = 0, \quad \text{if } k \neq l \end{aligned}$$

and the $\{b_k\}$ sector does not couple to the others,

$$\begin{aligned} -\partial_{b_k}\partial_{a_l}\mathcal{L} &= 0 \\ -\partial_{b_k}\partial_{\sigma_k^2}\mathcal{L} &= -\frac{N_k}{\sigma_k^4}\Big(b_k - \langle y\rangle_k\Big)\Big|_{\rm est} = 0. \end{aligned}$$

Matrix A is symmetric and 'stripe-diagonal',

$$\begin{aligned} -\partial^2_{a_k}\mathcal{L} &= A_{k,k} = g_k + g_{k+1} + m_k \\ -\partial_{a_k}\partial_{a_{k+1}}\mathcal{L} &= A_{k,k+1} = -g_{k+1} \\ -\partial_{a_k}\partial_{a_l}\mathcal{L} &= A_{k,l} = 0, \quad \text{if } |k-l| > 1, \end{aligned}$$

where the expressions for $g_k = \frac{1}{D\Delta x_k}$ and $m_k = \frac{N_k}{\sigma_k^2}\langle\tilde{x}\rangle_k$. The matrix S is also diagonal

$$\begin{aligned} -\partial^2_{\sigma_k^2}\mathcal{L} &= S_{k,k} = -\frac{N_k}{2\sigma_k^4} + \frac{N_k}{4\sigma_k^6}\left(\langle y^2\rangle_k - \langle y\rangle_k^2 - 2a_k\langle y\tilde{x}\rangle_k + a_k^2\langle\tilde{x}^2\rangle_k\right) \\ -\partial_{\sigma_k^2}\partial_{\sigma_l^2}\mathcal{L} &= S_{k,l} = 0, \quad \text{if } k \neq l, \end{aligned}$$

where $\tilde{x} = x - \bar{x}_k$, and G, the coupling matrix between $\{a_k\}$ and $\{\sigma_k^2\}$ is also a diagonal block,

$$\begin{aligned} -\partial_{a_k}\partial_{\sigma_k^2}\mathcal{L} &= G_{k,k} = \frac{N_k}{\sigma_k^4}\Big(\langle y\tilde{x}\rangle_k - a_k\langle\tilde{x}\rangle_k\Big), \\ -\partial_{a_k}\partial_{\sigma_l^2}\mathcal{L} &= G_{k,l} = 0, \quad \text{if } k \neq l. \end{aligned}$$

The inverse of C^{-1} provides all covariances,

$$C = \begin{bmatrix} B^{-1} & 0 & 0 \\ 0 & \begin{bmatrix} A & G \\ G & S \end{bmatrix}^{-1} \end{bmatrix} = \begin{bmatrix} V_b & 0 & 0 \\ 0 & V_a & C_{a,\sigma^2} \\ 0 & C_{a,\sigma^2} & V_{\sigma^2} \end{bmatrix}.$$

The uncertainties (standard deviations) of the estimates are

$$\Delta\mu_k = \Delta a_k = \sqrt{(V_a)_{k,k}} \quad \text{and} \qquad \Delta(\log n_{0,k}) = \Delta b_k = \sqrt{(V_b)_{k,k}}.$$

C Remarks on the prior of $\{\mu_k\}$

C.1 Time sampling independence

The prior distribution from Eq. (5) defines the deviation, $\Delta\mu_k = \sqrt{\mathrm{Var}(\mu_k - \mu_{k-1})}$ as

$$\Delta\mu_k = \sqrt{D\Delta t_k},$$

which explicitly depends on the times $\{t_k\}$. These times are, however, not known before the measurement is carried out, and therefore we should formulate the prior in a way which is independent of how we actually sample.

More precisely, we need to write the prior distribution $\mathcal{P}_0(\{\mu_k\}\,|\,\{t_k\})$ in a functional form $f(K, \vec{\mu}, \vec{t})$ which is invariant under the elimination of any sampling time point t_k:

$$\int d\mu_k\, f(K, \vec{\mu}, \vec{t}) = f(K-1, \vec{\mu}_{(\backslash k)}, \vec{t}_{(\backslash k)}), \tag{17}$$

where $\vec{\mu} = (\mu_1, \mu_2, \ldots \mu_K)$, and $\mu_{(\backslash k)} = (\mu_1, \mu_2, \ldots \mu_{k-1}, \mu_{k+1}, \ldots \mu_K)$ and the same with $\vec{t}$ and $\vec{t}_{(\backslash k)}$.

This requirement makes sense if we think about a scenario where data from one of the sampling times t_k turn out to be missing due to external reasons. We should arrive to the same prior by averaging over the variable μ_k as if it was not considered in the first place.

C.2 Stochastic processes

One way to make sure that the prior distribution $\mathcal{P}_0(\vec{\mu}\,|\,\vec{t})$ has a functional form which is invariant under removing sampling points is to obtain it from the marginal of a continuous stochastic process M_t at times $\{t_1, t_1, \ldots t_K\}$. A process M_t is defined by the conditional probability

$$\mathcal{P}\Big(M_t = \mu_t \,\Big|\, \{M_s = \mu_s \,:\, s < t\}\Big). \tag{18}$$

By setting the prior distribution of the first value ($\mathcal{P}_0(\mu_1)$), we can use Eq. (18) to construct the full prior $\{\mu_k\}$,

$$\begin{aligned}
\mathcal{P}_0\Big(\{M_{t_k} = \mu_k\}\Big) \;=\;& \mathcal{P}_0\Big(M_{t_1} = \mu_1\Big) \\
&\times\; \mathcal{P}\Big(M_{t_2} = \mu_2 \,\Big|\, M_{t_1} = \mu_1\Big) \\
&\times\; \mathcal{P}\Big(M_{t_3} = \mu_3 \,\Big|\, M_{t_1} = \mu_1, M_{t_2} = \mu_2\Big) \\
&\times\; \mathcal{P}\Big(M_{t_4} = \mu_4 \,\Big|\, M_{t_1} = \mu_1, M_{t_2} = \mu_2, M_{t_3} = \mu_3\Big) \\
&\times\; \ldots \\
=\;& \mathcal{P}_0\Big(M_{t_1} = \mu_1\Big) \prod_{k=2}^{K} \mathcal{P}\Big(M_{t_k} = \mu_k \,\Big|\, \{M_{t_l} = \mu_l \,:\, l = 1, 2, \ldots k-1\}\Big).
\end{aligned} \tag{19}$$

This procedure is invariant under omission of a time point t_k, since it only changes the list of points over which the marginal of the distribution of process M_t needs to be taken.

C.3 Brownian motion

If $\mu(t)$ is moved around by a random white noise, it will exhibit Brownian motion with a diffusion constant D, set by the strength of the noise. This can be modeled with the process

$$M_{t_k} = \mu_1 + W_{\tau_k}, \qquad \text{where} \quad \tau_k = D(t_k - t_1),$$

and W_τ is the Wiener process,

$$\mathcal{P}\Big(W_\tau = x \,\Big|\, \text{history}\Big) = \mathcal{P}\Big(W_\tau = x \,\Big|\, W_{\tau'} = x'\Big) = \frac{1}{\sqrt{2\pi(\tau - \tau')}} \exp\left[-\frac{(x-x')^2}{2(\tau-\tau')}\right], \tag{20}$$

where τ' is the largest time point in the history of τ. In other words, the random increment $\Delta x = x - x'$ on the Wiener process depends only on $\Delta\tau = \tau - \tau'$, and independent of any earlier values of W_τ. This is a Gaussian process with the following properties

$$W_0 = 0, \qquad \langle W_\tau \rangle = 0, \qquad \langle W_\tau^2 \rangle = \tau, \qquad \langle W_\tau W_s \rangle = \min(\tau, s).$$

Using Eq. (20) with Eq. (19) gives the prior

$$\begin{aligned} \mathcal{P}_0\Big(\{M_{t_k} = \mu_k\}\Big) &= \mathcal{P}_0\Big(M_{t_1} = \mu_1\Big) \prod_{k=2}^{K} \mathcal{P}\Big(M_{t_k} = \mu_k \,\Big|\, M_{t_{k-1}} = \mu_{k-1}\Big) \\ &= \mathcal{P}_0(\mu_1) \prod_{k=2}^{K} \frac{1}{\sqrt{2\pi D(t_k - t_{k-1})}} \exp\left[-\frac{(\mu_k - \mu_{k-1})^2}{2D(t_k - t_{k-1})}\right]. \end{aligned} \tag{21}$$

C.4 Integrated Brownian motion

If $\mu(t)$ changes due to a finite value of its derivative $\dot{\mu} = \frac{d\mu}{dt} =: \nu$ at any given time t, then its increment $\Delta\mu$ will depend not only on Δt, but previous values of μ. For the simplest case, we imagine that its derivative $\dot{\mu}$, will move around according to a Brownian motion,

$$\dot{M}_{t_k} = W_{\tau_k} + \nu_1, \qquad \text{where} \quad \tau_k = D(t_k - t_1).$$

The integral of this will follow an integrated Brownian motion,

$$M_{t_k} = \mu_1 + \int_{t_1}^{t_k} dt\, (W_\tau + \nu_1) = \mu_1 + \nu_1(t_k - t_1) + \frac{1}{D}\int_0^{\tau_k} ds\, W_s.$$

Instead of constructing the prior from $\mathcal{P}\Big(M_{t_k} = \mu_k \,\Big|\, \text{history}\Big)$ directly, we introduce a new variable: the average rate of change,

$$V_{t_k} = \frac{M_{t_k} - M_{t_{k-1}}}{t_k - t_{k-1}} = \frac{1}{t_k - t_{k-1}}\left[\nu_1(t_k - t_{k-1}) + \frac{1}{D}\int_0^{\tau_k} ds\, W_s - \frac{1}{D}\int_0^{\tau_{k-1}} ds\, W_s\right] = \nu_1 + \frac{1}{\tau_k - \tau_{k-1}}\int_{\tau_{k-1}}^{\tau_k} ds\, W_s.$$

Here, V_{t_k} is a stochastic process. It is a linear combination (integral) of Gaussian processes, W_s, which makes it also Gaussian. Now, we show that V_{t_k} is also a Brownian motion:

1. We show that $V_{t_k} - V_{t_{k-1}}$ has zero expectation value.

$$\langle V_{t_k} - V_{t_{k-1}} \rangle = \frac{1}{\tau_k - \tau_{k-1}} \int_{\tau_{k-1}}^{\tau_k} ds \, \underbrace{\langle W_s \rangle}_{0} - \frac{1}{\tau_{k-1} - \tau_{k-2}} \int_{\tau_{k-2}}^{\tau_{k-1}} ds \, \underbrace{\langle W_s \rangle}_{0} = 0.$$

2. We show that $V_{t_k} - V_{t_{k-1}}$ is independent of W_r if $r < t_{k-2}$. First,

$$\mathrm{Cov}\,(V, W_r) = \left\langle \frac{1}{\tau_2 - \tau_1} \int_{\tau_1}^{\tau_2} ds \, W_s W_r \right\rangle = \frac{1}{\tau_k - \tau_1} \int_{\tau_1}^{\tau_2} ds \, \underbrace{\langle W_s W_r \rangle}_{\min(s,r)=r} = r,$$

independent of τ_1 and τ_2, if $r < \tau_1, \tau_2$. Therefore

$$\mathrm{Cov}\left(V_{t_k} - V_{t_{k-1}},\ W_r\right) = \mathrm{Cov}\,(V_{t_k},\ W_r) - \mathrm{Cov}\left(V_{t_{k-1}},\ W_r\right) = r - r = 0.$$

Since all processes are Gaussian, zero covariance implies independence.

3. We show that the increment $V_{t_k} - V_{t_{k-1}}$ is independent of all previous values V_{t_l}.

$$\mathrm{Cov}\left(V_{t_k} - V_{t_{k-1}},\ V_{t_l}\right) = \nu_1 \underbrace{\langle V_{t_k} - V_{t_{k-1}} \rangle}_{0} + \frac{1}{\tau_k - \tau_{k-1}} \int_{\tau_{k-1}}^{\tau_k} ds \, \underbrace{\mathrm{Cov}\left(V_{t_k} - V_{t_{k-1}},\ W_s\right)}_{0} = 0$$

4. We find the variance of $\Delta V_k = V_{t_k} - V_{t_{k-1}}$ as a function of $\Delta t_k = t_k - t_{k-1}$ and $\Delta t_{k-1} = t_{k-1} - t_{k-2}$.

$$\begin{aligned}
\left\langle \left(V_{t_k} - V_{t_{k-1}}\right)^2 \right\rangle &= \left\langle \left(\frac{1}{\tau_k - \tau_{k-1}} \int_{\tau_{k-1}}^{\tau_k} ds \, W_s - \frac{1}{\tau_{k-1} - \tau_{k-2}} \int_{\tau_{k-2}}^{\tau_{k-1}} ds \, W_s \right)^2 \right\rangle \\
&= \frac{1}{(\Delta\tau_k)^2} \int_{\tau_{k-1}}^{\tau_k} ds \int_{\tau_{k-1}}^{\tau_k} dr \, \underbrace{\langle W_s W_r \rangle}_{\min(s,r)} + \frac{1}{(\Delta\tau_{k-1})^2} \int_{\tau_{k-2}}^{\tau_{k-1}} ds \int_{\tau_{k-2}}^{\tau_{k-1}} dr \, \underbrace{\langle W_s W_r \rangle}_{\min(s,r)} \\
&\quad - \frac{2}{\Delta\tau_k \Delta\tau_{k-1}} \int_{\tau_{k-1}}^{\tau_k} ds \int_{\tau_{k-2}}^{\tau_{k-1}} dr \, \underbrace{\langle W_s W_r \rangle}_{r} \\
&= \frac{\frac{1}{3}\tau_k^3 - \tau_k \tau_{k-1}^2 + \frac{2}{3}\tau_{k-1}^3}{(\tau_k - \tau_{k-1})^2} + \frac{\frac{1}{3}\tau_{k-1}^3 - \tau_{k-1}\tau_{k-2}^2 + \frac{2}{3}\tau_{k-2}^3}{(\tau_{k-1} - \tau_{k-2})^2} - \frac{2(\tau_k - \tau_{k-1})(\tau_{k-1}^2 - \tau_{k-2}^2)}{(\tau_k - \tau_{k-1})(\tau_{k-1} - \tau_{k-2})} \\
&= \frac{1}{3}(\tau_k - \tau_{k-2}) = \frac{2D}{3} \left(\frac{\Delta t_k + \Delta t_{k-1}}{2} \right).
\end{aligned}$$

Now that we have shown that V_{t_k} is a Brownian motion, we can safely write its conditional distribution

$$\mathcal{P}\Big(V_{t_k} = \nu_k \,\Big|\, \text{history}\Big) = \mathcal{P}\Big(V_{t_k} = \nu_k \,\Big|\, V_{t_{k-1}} = \nu_{k-1}\Big) = \frac{1}{\sqrt{2\pi D(t_k - t_{k-2})/3}} \exp\left[-\frac{(\nu_k - \nu_{k-1})^2}{2D(t_k - t_{k-2})/3} \right],$$

and use it to express the prior for $\{\nu_k\}$ as

$$\mathcal{P}_0\Big(\{V_{t_k} = \nu_k\}\Big) = \mathcal{P}(V_{t_1} = \nu_1) \prod_{k=2}^{K} \mathcal{P}\Big(V_{t_k} = \nu_k \,\Big|\, V_{t_{k-1}} = \nu_{k-1}\Big).$$

The prior for $\{\mu_k\}$ can be obtained by changing variables from $\{V_{t_k}\}$ to $\{M_{t_k}\}$. Since this is a linear transformation, the Jacobian is constant, and we can write

$$\mathcal{P}_0\Big(\{M_{t_k}=\mu_k\}\Big) \;\propto\; \mathcal{P}_0(M_{t_1}=\mu_1)\,\mathcal{P}_0(M_{t_2}=\mu_2)\prod_{k=3}^{K}\mathcal{P}\left(V_{t_k}=\frac{\mu_k-\mu_{k-1}}{t_k-t_{k-1}}\;\middle|\;V_{t_{k-1}}=\frac{\mu_{k-1}-\mu_{k-2}}{t_{k-1}-t_{k-2}}\right).$$

By choosing the priors for μ_1 and μ_2 constant, we get

$$\mathcal{P}_0\Big(\{\mu_k\}\;\Big|\;\{t_k\}\Big)\propto\prod_{k=3}^{K}\exp\left[-\frac{\left(\frac{\mu_k-\mu_{k-1}}{t_k-t_{k-1}}-\frac{\mu_{k-1}-\mu_{k-2}}{t_{k-1}-t_{k-2}}\right)^2}{\frac{2D}{3}(t_k-t_{k-2})}\right].$$

Using this prior instead of Eq. (5) will make the linear equation of $a_k=\mu_k$ couple to not just a_{k-1} and a_{k+1}, but also to a_{k-2} and a_{k+2}, in Eq. (7), but no further complication arises.

We note that this prior satisfies the sampling invariance condition from Eq. (17) for $\{\nu_k\}$ and not for $\{\mu_k\}$.

C.5 Simple population dynamics

At the moment, choosing between the Brownian, the integrated Brownian priors or a prior generated by any other stochastic process seems to be a matter of taste. Here, we argue that, the Brownian and the integrated Brownian priors cover a large set of realistic dynamics of μ.

At any given time t the cell population inside the measurement volume has a distribution of growth rates, $\rho(\mu,t)$. The density of cells with a given growth rate μ is expressed as

$$n_\mu(t)=n(t)\rho(\mu,t),\qquad\text{where}\quad\int d\mu\,\rho(\mu,t)=1,$$

and $n(t)=\int d\mu\,n_\mu(t)$ is the total density, measured by the detector (through the absorbance α).

In a stationary environment and under the assumption that cells produce identical copies of themselves, the density of cells with a certain growth rate is increasing exponentially,

$$n_\mu(t)=n_\mu(0)\exp(\mu t),$$

and the total density is increasing as

$$n(t)=\int d\mu\,n_\mu(0)\exp(\mu t).$$

In a short enough time window $[t,t+dt]$, the change of the total density dn is

$$dn=n(t+dt)-n(t)\approx dt\int d\mu\,n_\mu(t)\mu,$$

and the relative change is

$$\frac{dn}{n}\approx dt\int d\mu\,\frac{n_\mu(t)}{n}\mu=dt\int d\mu\,\rho(\mu,t)\mu=:\langle\mu\rangle\,dt.$$

This means that by fitting a linear function to a short segment of the logarithm of the measurement data, $\log(n(t))$, we can extract $\langle\mu\rangle$ directly. This average value is what we denoted by μ all the way from the start.

Under the stationary assumption, defined above, we can write the time evolution of ρ as

$$\rho(\mu,t) = \frac{n_\mu(t)}{n(t)} = \frac{n(0)}{n(t)}\rho(\mu,0)\exp(\mu t) = \frac{\rho(\mu,0)\exp(\mu t)}{N(t)},$$

where $N(t) = \int d\mu\, \rho(\mu,0)e^{\mu t}$ is a normalization constant. The time evolution of $\langle\mu\rangle$ is entirely determined $\rho(\mu,t)$. Its derivative can be written as

$$\begin{aligned}
\frac{d}{dt}\langle\mu\rangle &= \frac{d}{dt}\int d\mu\, \mu\rho(\mu,t) = \int d\mu\, \mu\left[-\frac{(\int d\mu\, \rho(\mu,0)\mu e^{\mu t})}{[N(t)]^2}\rho(\mu,0)e^{\mu t} + \frac{1}{N(t)}\rho(\mu,0)\mu e^{\mu t}\right] \\
&= -\langle\mu\rangle\int d\mu\, \mu\rho(\mu,t) + \int d\mu\, \mu^2\rho(\mu,t) \\
&= \langle\mu^2\rangle - \langle\mu\rangle^2 = \mathrm{Var}(\mu),
\end{aligned}$$

which tells us that the rate of change of $\langle\mu\rangle$ is entirely determined by the variance of μ. Similar relations can be found for any moment,

$$\frac{d}{dt}\langle\mu^k\rangle = \frac{d}{dt}\int d\mu\, \mu^k\rho(\mu,t) = \langle\mu^{k+1}\rangle - \langle\mu^k\rangle\langle\mu\rangle.$$

This couples all the momentums $\langle\mu^k\rangle$ with each other non-linearly. Since $\rho(\mu,t)$ is an analytic function of t, all momenta are also analytic functions of t. Since the derivatives of $\langle\mu\rangle$ are analytic functions of these momenta, $\langle\mu\rangle$ is infinitely differentiable.

Perturbations change this picture. There are two very different ways this can happen:

1. A global change in the environment results in a change $\delta\mu$ for the entire population. Such an effect will make $\langle\mu\rangle$ follow the noisy dynamics of the relevant environmental variable. If this variable is affected by multiple other variables, then it most likely exhibits Brownian motion, which is then imprinted on $\langle\mu\rangle$. This scenario justifies the choice of the Brownian motion prior. (See Appendix C.3.)

2. In the case of no global environmental fluctuations, noise from internal switching can dominate. If a small fraction of the cells x switches from μ to μ' then we can write the changes of the average and the variance as

$$\begin{aligned}
\langle\mu\rangle' - \langle\mu\rangle &= x(\mu' - \mu) \\
\mathrm{Var}(\mu)' - \mathrm{Var}(\mu) &= 2x(\mu' - \mu)\left(\frac{\mu + \mu'}{2} + \langle\mu\rangle\right) + \mathcal{O}(x^2).
\end{aligned}$$

   With a waiting time Δt, the total change of $\langle\mu\rangle$ is

$$\Delta\langle\mu\rangle \approx x(\mu' - \mu) + \mathrm{Var}(\mu)\Delta t + 2x(\mu' - \mu)\left(\frac{\mu + \mu'}{2} + \langle\mu\rangle\right)\Delta t.$$

   If the waiting time between two switching events is long enough, such that the first term is negligible next to the sum of the second and third, then the integrated Brownian motion (see Appendix C.4) dominates.

Note that noises that affect higher moments of μ would be better described by higher order integrals of the Brownian motion. The conclusion here is that in the absence of environmental changes, the simple Brownian motion overestimates the expected fluctuations of μ, and it is recommended to use the integrated Brownian motion prior.