Supplemental Information

C.J.Oates and S.Mukherjee

November 15, 2011

Abstract

This supplement accompanies Oates, C.J., Mukherjee, S. (2011) Network Inference and Biological Dynamics.

This document is constructed as follows: Section 1 recalls the dynamical systems and parameters used to simulate single cell trajectories. Section 2 provides derivations relating to the variance function. Section 3 contains additional Figures and results.

1 Dynamical Systems

1.1 Model 1: Cantone *et al.*

Ref: Cantone, I., Marucci, L., Iorio, F., *et al.* (2009) A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches, *Cell*, **137**(1), 172-81.

$$x_1 = [Cbf1],\ x_2 = [Gal4],\ x_3 = [Swi5],\ x_4 = [Gal80],\ x_5 = [Ash1]$$

$$\begin{aligned}
\frac{dx_1}{dt} &= \alpha_1 + v_1 \left(\frac{x_3^{h_1}(t-\tau)}{(k_1^{h_1} + x_3^{h_1}(t-\tau)) \left(1 + \frac{x_5^{h_2}}{k_2^{h_2}}\right)} \right) - d_1 x_1 \\
\frac{dx_2}{dt} &= \alpha_2 + v_2 \left(\frac{x_1^{h_3}}{k_3^{h_3} + x_1^{h_3}} \right) - d_2 x_2 \\
\frac{dx_3}{dt} &= \alpha_3 + v_3 \left(\frac{x_2^{h_4}}{k_4^{h_4} + x_2^{h_4}(1 + \frac{x_4^4}{\gamma^4})} \right) - d_3 x_3 \\
\frac{dx_4}{dt} &= \alpha_4 + v_4 \left(\frac{x_3^{h_5}}{k_5^{h_5} + x_3^{h_5}} \right) - d_4 x_4 \\
\frac{dx_5}{dt} &= \alpha_5 + v_5 \left(\frac{x_3^{h_6}}{k_6^{h_6} + x_3^{h_6}} \right) - d_5 x_5
\end{aligned}$$

Parameter values were taken from the "switch-on" experimental conditions.

Parameter	Description	Value used for simulation
k_1	M. M. const. (SWI5 $\rightarrow$ HO pr)	1 [a.u.]
k_2	M. M. const. (ASH1 $\rightarrow$ HO pr)	0.0356 [a.u.]
k_3	M. M. const. (CBF1 $\rightarrow$ MET16 pr)	0.0372 [a.u.]
k_4	M. M. const. (GAL4 $\rightarrow$ GAL10 pr)	0.01 [a.u.]
k_5	M. M. const.(SWI5 $\rightarrow$ ASH1 pr (GAL80))	1.814 [a.u.]
k_6	M. M. const. (SWI5 $\rightarrow$ ASH1 pr (ASH1))	1.814 [a.u.]
α_1	Basal act. of the HO pr	0 [a.u. min^{-1}]
α_2	Basal act. of the MET16 pr	0.000149 [a.u. min^{-1}]
α_3	Basal act. of the GAL10 pr	0.003 [a.u. min^{-1}]
α_4	Basal act. of the ASH1 pr (GAL80)	0.00074 [a.u. min^{-1}]
α_5	Basal act. of the ASH1 pr (ASH1)	0.00061 [a.u. min^{-1}]
v_1	Max. HO pr trans.	0.04 [a.u. min^{-1}]
v_2	Max. MET16 pr trans.	0.000882 [a.u. min^{-1}]
v_3	Max. GAL10 pr trans.	0.0201 [a.u. min^{-1}]
v_4	Max. ASH1 pr (GAL80) trans.	0.0147 [a.u. min^{-1}]
v_5	Max. ASH1 pr (ASH1) trans.	0.0182 [a.u. min^{-1}]
d_1	Deg. rate of CBF1	0.0222 [min^{-1}]
d_2	Deg. rate of GAL4	0.0478 [min^{-1}]
d_3	Deg. rate of SWI5	0.4217 [min^{-1}]
d_4	Deg. rate of GAL80	0.0980 [min^{-1}]
d_5	Deg. rate of ASH1	0.05 [min^{-1}]
γ	Affinity in the law of SWI5	0.6 [a.u.]
h_1	Hill coef. (SWI5 $\rightarrow$ HO pr)	1
h_2	Hill coef. (ASH1 $\rightarrow$ HO pr)	1
h_3	Hill coef. (CBF1 $\rightarrow$ MET16 pr)	1
h_4	Hill coef. (GAL4 $\rightarrow$ GAL10 pr)	4
h_5	Hill coef. (SWI5 $\rightarrow$ ASH1 pr (GAL80))	1
h_6	Hill coef. (SWI5 $\rightarrow$ ASH1 pr (ASH1))	1
τ	Time delay	100 [min]

1.2 Model 2: Swat

Ref: Swat, M., Kel, A., Herzel, H. (2004) Bifurcation analysis of the regulatory modules of the mammalian G1/S transition, *Bioinformatics*, **20**(10), 1506-11.

$$x_1 = [pRB],\ x_2 = [E2F1],\ x_3 = [CycD_i],\ x_4 = [CycD_a],\ x_5 = [AP-1]$$
$$x_6 = [pRB_p],\ x_7 = [pRB_{pp}],\ x_8 = [CycE_i],\ x_9 = [CycE_a]$$

$$\begin{aligned}
\frac{dx_1}{dt} &= k_1 \frac{x_2}{k_{m1}+x_2}\frac{J_{11}}{J_{11}+x_1}\frac{J_{61}}{J_{61}+x_6} - k_{16}x_1x_4 + k_{61}x_6 - \phi_1 x_1 \\
\frac{dx_2}{dt} &= k_p + k_2 \frac{a^2+x_2^2}{k_{m2}^2+x_2^2}\frac{J_{12}}{J_{12}+x_1}\frac{J_{62}}{J_{62}+x_6} - \phi_2 x_2 \\
\frac{dx_3}{dt} &= k_3x_5 + k_{23}x_2\frac{J_{13}}{J_{13}+x_1}\frac{J_{63}}{J_{63}+x_6} + k_{43}x_4 - k_{34}x_3\frac{x_4}{k_{m4}+x_4} - \phi_3 x_3 \\
\frac{dx_4}{dt} &= k_{34}x_3\frac{x_4}{k_{m4}+x_4} - k_{43}x_4 - \phi_4 x_4 \\
\frac{dx_5}{dt} &= F_m + k_{25}x_2\frac{J_{15}}{J_{15}+x_1}\frac{J_{65}}{J_{65}+x_6} - \phi_5 x_5 \\
\frac{dx_6}{dt} &= k_{16}x_1x_4 - k_{61}x_6 - k_{67}x_6x_9 + k_{76}x_7 - \phi_6 x_6 \\
\frac{dx_7}{dt} &= k_{67}x_6x_9 - k_{76}x_7 - \phi_7 x_7 \\
\frac{dx_8}{dt} &= k_{28}x_2\frac{J_{18}}{J_{18}+x_6}\frac{J_{68}}{J_{68}+x_7} + k_{98}x_9 - k_{89}x_8\frac{x_9}{k_m9+x_9} - \phi_8 x_8 \\
\frac{dx_9}{dt} &= k_{89}x_8\frac{x_9}{k_{m9}+x_9} - k_{98}x_9 - \phi_9 x_9
\end{aligned}$$

Parameter	Value used for simulation
k_1	1
k_2	1.6
k_3	0.05
k_{16}	0.4
k_{34}	0.04
k_{43}	0.01
k_{61}	0.3
k_{67}	0.7
k_{76}	0.1
k_{23}	0.3
k_{25}	0.9
k_{28}	0.06
k_{89}	0.07
k_{98}	0.01
a	0.04
J_{11}	0.5
J_{12}	5
J_{15}	0.001
J_{18}	0.6
J_{61}	5
J_{62}	8
J_{65}	6
J_{68}	7
J_{13}	0.002
J_{63}	2
k_{m1}	0.5
k_{m2}	4
k_{m4}	0.3
k_{m9}	0.005
k_p	0.05
ϕ_1	0.005
ϕ_2	0.1
ϕ_3	0.023
ϕ_4	0.03
ϕ_5	0.01
ϕ_6	0.06
ϕ_7	0.04
ϕ_8	0.06
ϕ_9	0.05
F_m	0.0065

2 Derivations

2.1 Deriving a model in the large sample limit

Suppose the true large sample process obeys $d\mathbf{X}^\infty/dt = \mathbf{F}(\mathbf{X}^\infty)$. Then a Taylor expansion of $\mathbf{X}^\infty$ about t gives (dropping the superscript ∞)

$$\mathbf{X}(t+\Delta) = \mathbf{X}(t) + \Delta\mathbf{F}(\mathbf{X}(t)) + \dots \tag{1}$$

so that when we account for measurement error $\mathbf{Y} = \mathbf{X} + \mathbf{w}$ we have

$$\mathbf{Y}(t+\Delta) - \mathbf{w}(t+\Delta) = \mathbf{Y}(t) - \mathbf{w}(t) + \Delta\mathbf{F}(\mathbf{Y}(t) - \mathbf{w}(t)) + \dots. \tag{2}$$

A Taylor expansion of $\mathbf{F}$ about $\mathbf{Y}(t)$ and a rearrangement gives

$$\frac{\mathbf{Y}(t+\Delta) - \mathbf{Y}(t)}{\Delta} - \mathbf{F}(\mathbf{Y}(t)) = \frac{\mathbf{w}(t+\Delta)}{\Delta} - \left[\frac{\mathbf{I}}{\Delta} + (D\mathbf{F})(\mathbf{Y}(t))\right]\mathbf{w}(t) + \dots \tag{3}$$

so that the variance $\mathbb{V}\left(\frac{\mathbf{Y}(t+\Delta)-\mathbf{Y}(t)}{\Delta} - \mathbf{F}(\mathbf{Y}(t))\right)$

$$= \frac{\mathbf{M}}{\Delta^2} + \left(\frac{\mathbf{I}}{\Delta} + D\mathbf{F}\right)\mathbf{M}\left(\frac{\mathbf{I}}{\Delta} + D\mathbf{F}\right)' + \dots. \tag{4}$$

2.2 Deriving a model for longitudinal single cell measurements

An Euler-Maruyama approximation for single cell expression $\mathbf{X}$ gives

$$\mathbf{X}(t+\Delta) = \mathbf{X}(t) + \Delta\mathbf{f}(\mathbf{X}(t)) + \mathbf{g}(\mathbf{X}(t))\Delta\mathbf{B} + \dots \tag{5}$$

so that when we account for measurement error $\mathbf{Y} = \mathbf{X} + \mathbf{w}$ we have

$$\mathbf{Y}(t+\Delta) - \mathbf{w}(t+\Delta) = \mathbf{Y}(t) - \mathbf{w}(t) + \Delta\mathbf{f}(\mathbf{Y}(t) - \mathbf{w}(t)) + \mathbf{g}(\mathbf{Y}(t) - \mathbf{w}(t))\Delta\mathbf{B} + \dots. \tag{6}$$

Taking a diffusion $\mathbf{g}(\mathbf{X}) = \sigma\mathcal{D}(\mathbf{X})$, a Taylor expansion of $\mathbf{f}$ about $\mathbf{Y}(t)$ and a rearrangement gives $\frac{\mathbf{Y}(t+\Delta)-\mathbf{Y}(t)}{\Delta} - \mathbf{f}(\mathbf{Y}(t))$

$$= \frac{\mathbf{w}(t+\Delta)}{\Delta} - \left[\frac{\mathbf{I}}{\Delta} + (D\mathbf{f})(\mathbf{Y}(t))\right]\mathbf{w}(t) + \sigma[\mathcal{D}(\mathbf{Y}(t)) - \mathcal{D}(\mathbf{w}(t))]\Delta\mathbf{B} + \dots \tag{7}$$

so that $\mathbb{V}\left(\frac{\mathbf{Y}(t+\Delta)-\mathbf{Y}(t)}{\Delta} - \mathbf{f}(\mathbf{Y}(t))\right)$

$$= \frac{\mathbf{M}}{\Delta^2} + \left[\frac{\mathbf{I}}{\Delta} + D\mathbf{f}\right]\mathbf{M}\left[\frac{\mathbf{I}}{\Delta} + D\mathbf{f}\right]' + \frac{\mathbf{g}\mathbf{g}'}{\Delta} + \dots. \tag{8}$$

Notice that this variance is larger than the corresponding variance in Eqn. 4, showing that a single cell data set contains less information than the corresponding dataset for an averaged process.

2.3 Approximating $h_{\mathbf{true}}$ for Cantone

From Eqn. 4 a natural choice of variance function h_{true} is

$$h_{\text{true}}(\Delta)\mathcal{D}(\sigma_1^2,\ldots,\sigma_P^2) \approx \mathbf{M}\Delta^{-2} + (\mathbf{I}\Delta^{-1} + D\mathbf{F})\mathbf{M}(\mathbf{I}\Delta^{-1} + D\mathbf{F})' \tag{9}$$

where the large sample process obeys $d\mathbf{X}^{\infty}/dt = \mathbf{F}$. This can be made precise under a given matrix norm:

$$h_{\text{true}}(\Delta) \approx \frac{\left\|\mathbf{M}\Delta^{-2} + (\mathbf{I}\Delta^{-1} + D\mathbf{F})\mathbf{M}(\mathbf{I}\Delta^{-1} + D\mathbf{F})'\right\|}{\|\mathcal{D}(\sigma_1^2,\ldots,\sigma_P^2)\|} \tag{10}$$

Under the (strong) assumption that $\mathbf{F} \equiv \mathbf{F}(\mathbf{X}^{\infty})$ we have that $D\mathbf{F}|_{\mathbf{x}^{\infty}=\mathbf{0}} = D\mathbf{f}|_{\mathbf{x}^{k}=\mathbf{0}}$ since $\mathbf{X}^{\infty} = \mathbf{0}$ if and only if almost all the $\mathbf{X}^{k} = \mathbf{0}$. So it suffices to find the Jacobian of the single cell drift $\mathbf{f}$.

We seek an approximation to h_{true} for Cantone *et al*, so for simplicity ignore the delay term in the regulation of Cbf1 by Swi5. Then

$$D\mathbf{f}|_{\mathbf{x}=\mathbf{0}} = \begin{bmatrix} -d_1 & 0 & v_1/k_1 & 0 & 0 \\ v_2/k_3 & -d_2 & 0 & 0 & 0 \\ 0 & 0 & -d_3 & 0 & 0 \\ 0 & 0 & v_k/k_5 & -d_4 & 0 \\ 0 & 0 & v_5/k_6 & 0 & -d_5 \end{bmatrix}. \tag{11}$$

Substituting Eqn. 11 for $D\mathbf{F}$ in Eqn. 10 provides an approximation to h_{true}.

3 Additional Figures

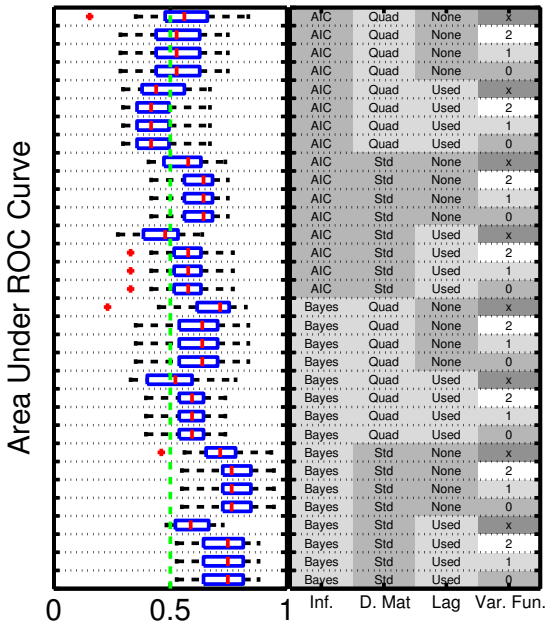
In this section we simply reproduce AUR scores. However the MATLAB R2010a code included in this supplement will automatically display typical datasets, ROC and PR curves in addition to these AUR scores. (For an introduction to receiver operating characteristic (ROC) and precision-recall (PR) analysis, the reader is referred to [Davis, J., Goadrich, M. (2006) The relationship between Precision-Recall and ROC curves, *ICML '06 Proceedings of the 23rd international conference on Machine learning*].)

In order to reproduce these experiments, simply run the .m-file Experiment_k where k = 1,2,3,4,5.

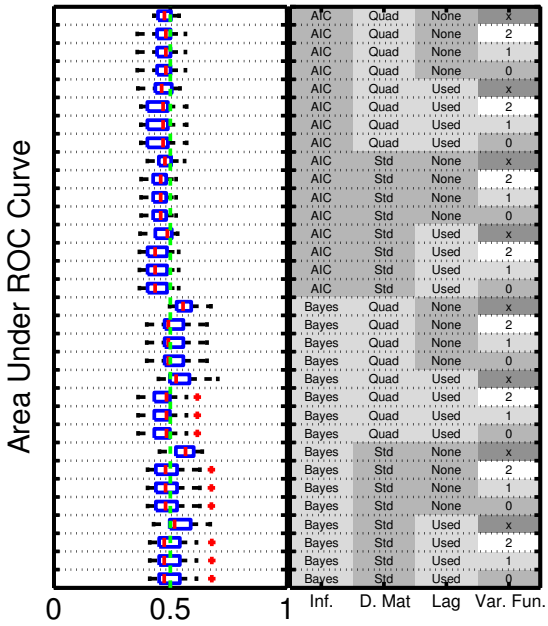
These plots employ the same key as the main text.

3.1 Experiment 1

$N = 10,000$ ($N^* = 30$) simulated microarray experiment. $n = 20$ data points are obtained at regular intervals from independent realizations of the process, mimicking destructive sampling. Additive Gaussian measurement error is applied. The procedure is repeated 20 times over to obtain a distribution of AUR scores for each inference scheme.

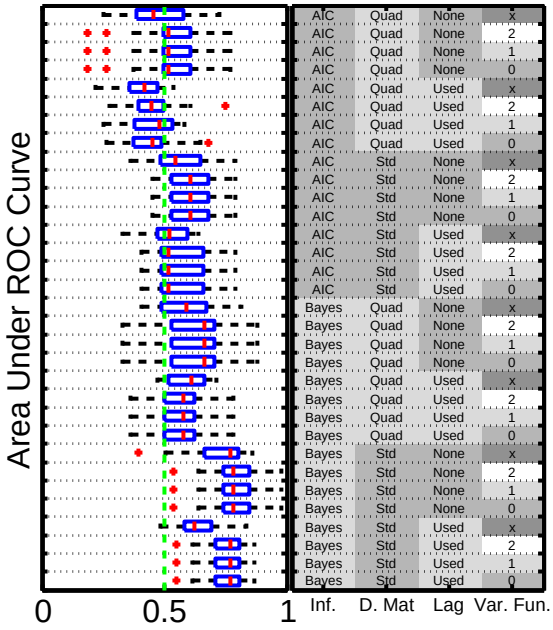


SFig. 1: AUR for Cantone *et al*, even sampling times.

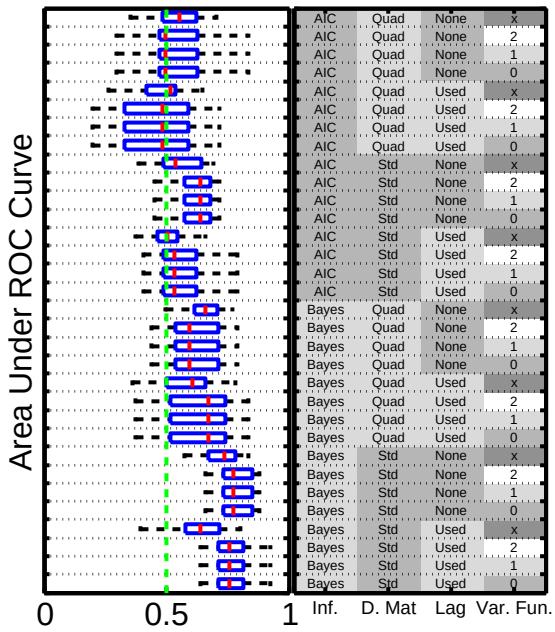


SFig. 2: AUR for Swat *et al*, even sampling times.

Inference was repeated using more nonlinear basis functions (SFig. 3) and longer delay times (SFig. 4).

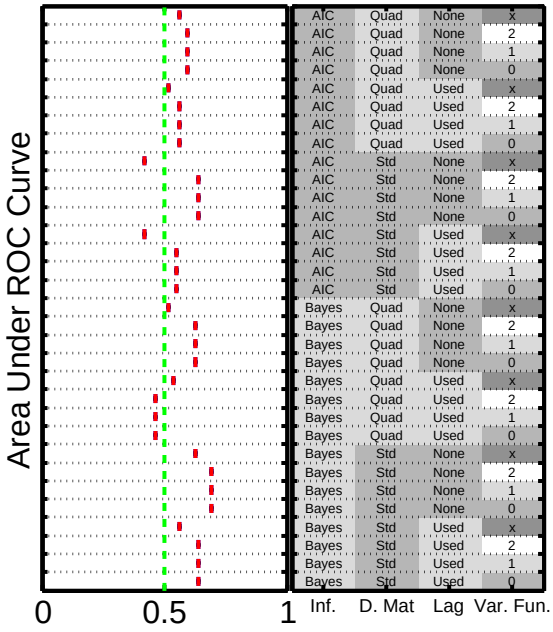


SFig. 3: AUR for Cantone *et al*, even sampling times, with cubic predictors.



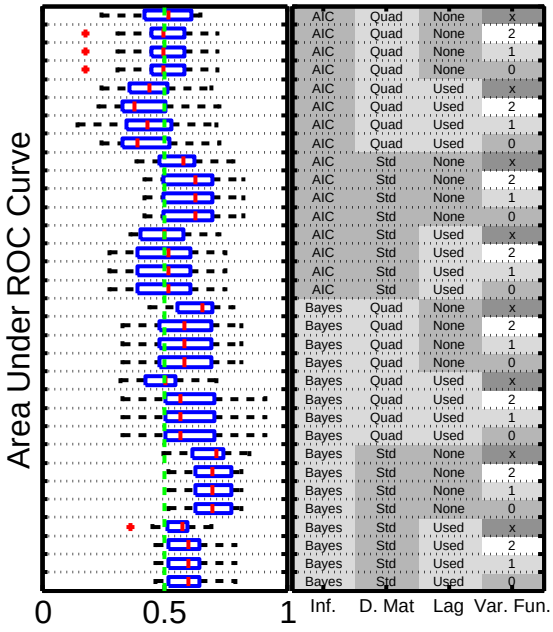
SFig. 4: AUR for Cantone *et al*, even sampling times, with a delay of double duration.

Inference was also performed on the *in vivo* averaged "switch-on" dataset published by Cantone *et al*. Results in SFig. 5 demonstrate reasonable agreement with the *in silico* experiment SFig. 1.



SFig. 5: AUR for Cantone *et al*, even sampling times, based on *in vivo* data.

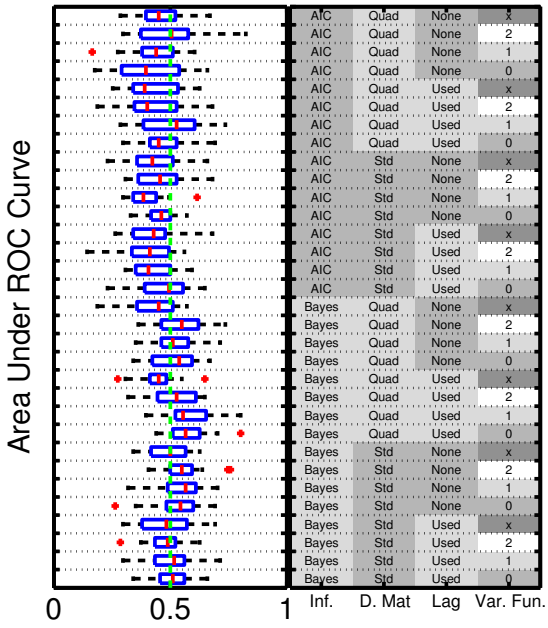
Robustness to $d_{\max}$ was verified by repeating the analyses using $d_{\max} = 3$. One such example is presented in SFig. 6.



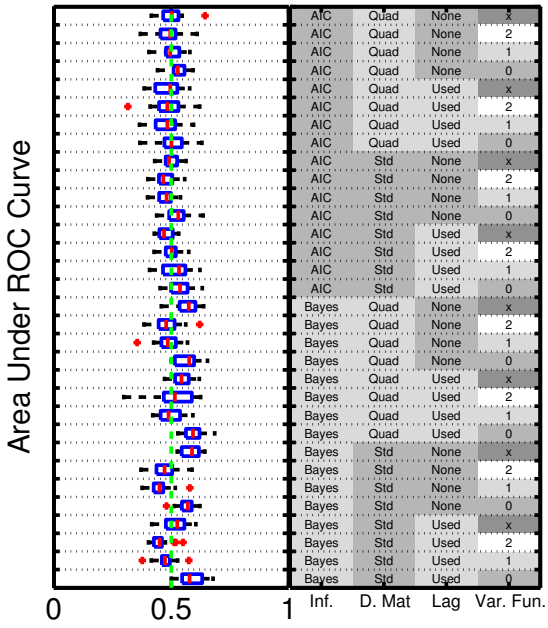
SFig. 6: AUR for Cantone *et al*, even sampling times, using $d_{\max} = 3$.

3.2 Experiment 2

The conditions of Experiment 1 are maintained, with the exception that data are obtained at irregular intervals. For Cantone *et al* these measurements occur at [0, 1, 5, 10, 15, 20, 30, 40, 50, 60, 75, 90, 105, 120, 140, 160, 280, 210, 240, 280] minutes and for Swat *et al* (for consistency) we also use these intervals, normalised to lie in [0,100] minutes.



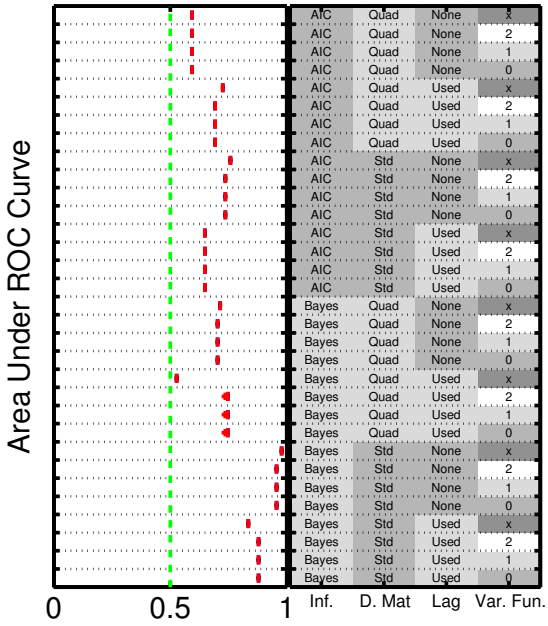
SFig. 7: AUR for Cantone *et al*, uneven sampling times.



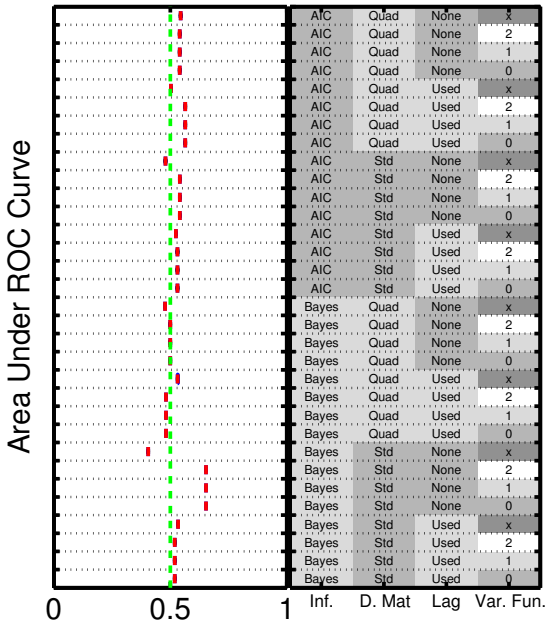
SFig. 8: AUR for Swat *et al*, uneven sampling times.

3.3 Experiment 3

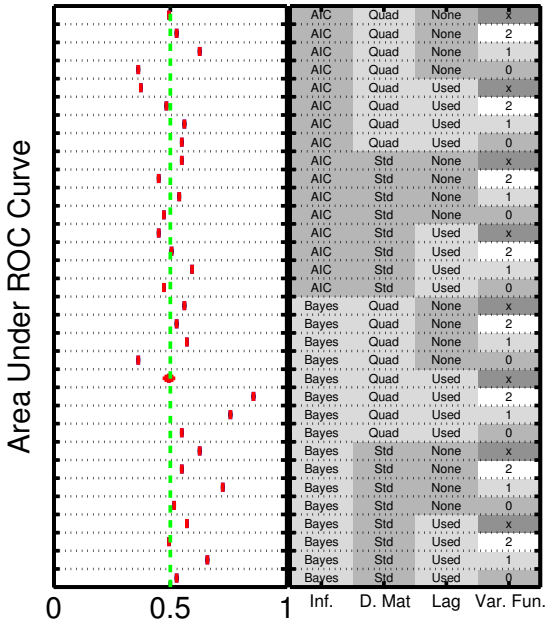
For this experiment we obtain simulated microarray data in the limits of zero measurement noise ($\sigma_{\text{meas}} = 0$), zero cellular heterogeneity ($\sigma_{\text{cell}} = 0$) and large samples ($N \to \infty$). Otherwise we follow Experiments 1 and 2, using both even and uneven sampling times. Since the data are deterministic, the corresponding AUR scores nonrandom.



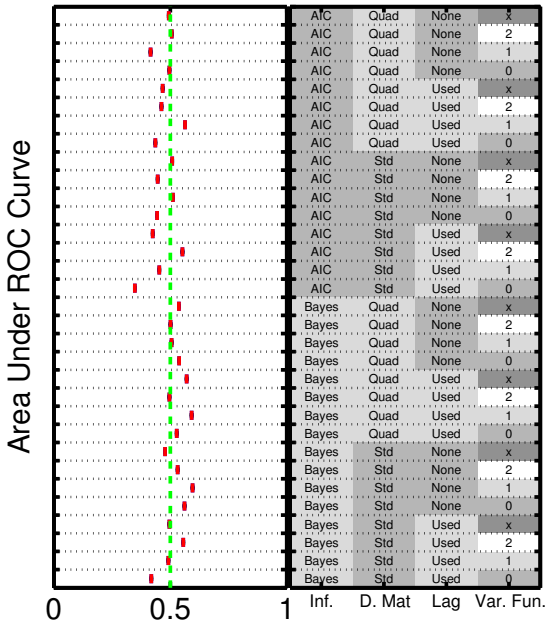
SFig. 9: AUR for Cantone *et al*, even sampling times.



SFig. 10: AUR for Swat *et al*, even sampling times.



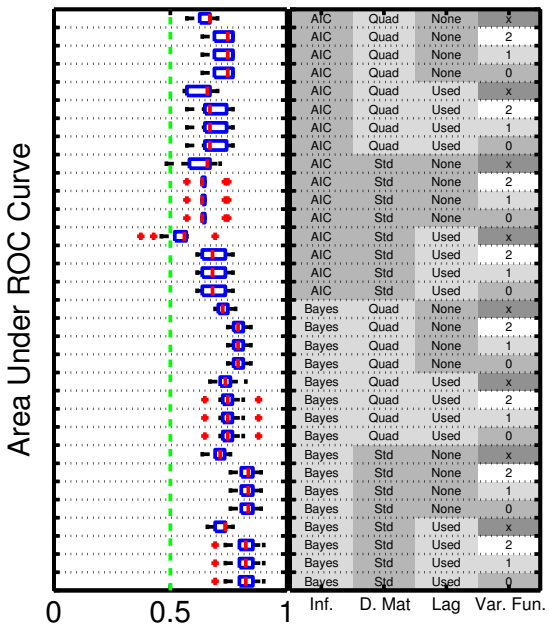
SFig. 11: AUR for Cantone *et al*, uneven sampling times.



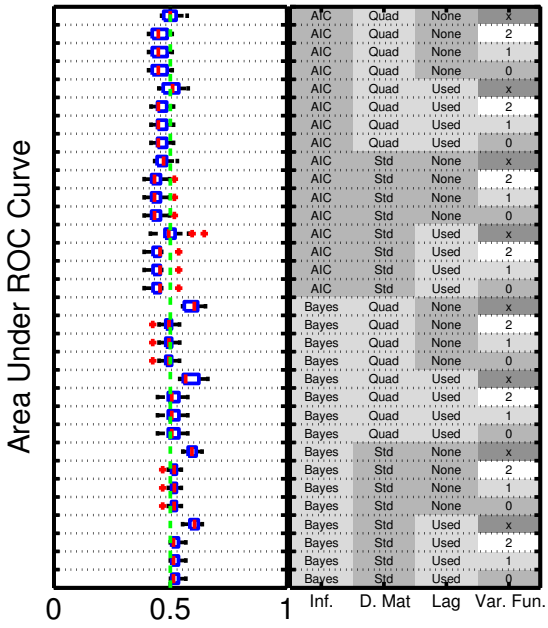
SFig. 12: AUR for Swat *et al*, uneven sampling times.

3.4 Experiment 4

This Experiment investigates the benefit of longitudinal single cell data. Under regular sampling, $C = 10$ single cell trajectories are recorded with additive Gaussian measurement error. The design matrices and predictors are concatenated to produce a total of $n \times C$ time points for inference.



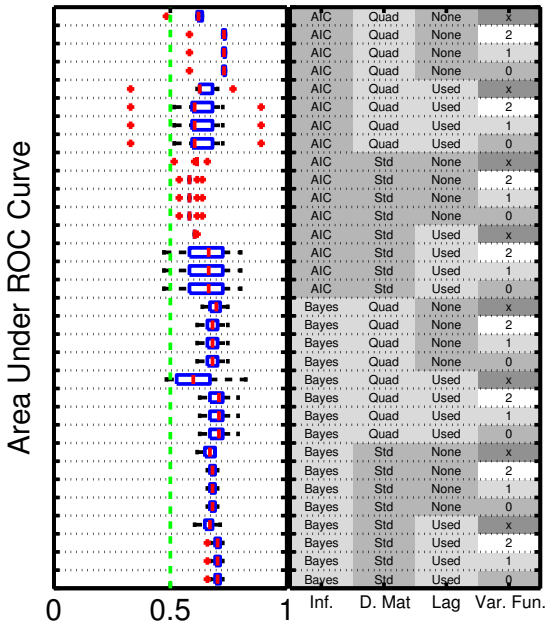
SFig. 13: AUR for Cantone *et al*, even sampling times.



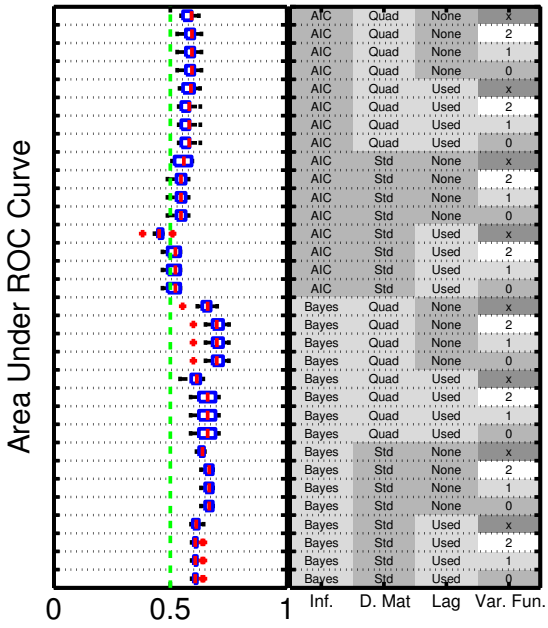
SFig. 14: AUR for Swat *et al*, even sampling times.

3.5 Experiment 5

This Experiment investigates the benefit of inhibition data. ($N = 10,000$, $N^* = 30$.) Under regular sampling, averaged trajectories of the process with variable k inhibited are recorded with additive Gaussian measurement error. The design matrices and predictors are concatenated to produce a total of $n \times P$ time points for inference.



SFig. 15: AUR for Cantone *et al*, even sampling times.



SFig. 16: AUR for Swat *et al*, even sampling times.