Reconstruction of ensemble of single-cell time trajectories from discrete-time fluorescence data: Oscillatory MAPK dynamics

Girija S. Kalantre, Ganesh A. Viswanathan*

Department of Chemical Engineering, Indian Institute of Technology Bombay, Powai, Mumbai – 400076
*Email: ganeshav@iitb.ac.in

Abstract: Continuous time profiles of intracellular protein levels in a collection of isogenic cells is needed to achieve quantitative prediction of heterogeneity in cellular systems. However, intracellular staining based quantitative single-cell detection of protein levels, reported by emitted fluorescence, using confocal microscopy or flow cytometry can only result in discrete time series due to arresting of cell state to enable entry of reporters – antibody – into the cell. We propose a method to reconstruct the time-series of oscillatory dynamics of phosphorylated ERK (pERK), the terminal protein in the ubiquitously found MAPK cascades, based on the discrete time series data consisting of a distribution of fluorescence emitted by different ensemble of cells at different time points post-stimulation. This method employs a model autocorrelation function to predict the fluorescence from the experimental data that will correspond to a specific time point in a randomly reconstructed trajectory. We validate the method using the single-cell pERK oscillatory dynamics data consisting of 12100 data points measured in transfected cells across 121 time points by pairing reconstructed trajectories with those from original based on the constraint that the pair satisfied a certain cut-off for both mutual information score and Euclidean distance between them. Out of the 100 trajectories in the original data, our algorithm was able to reconstruct ~30 of them capturing a reasonable fraction of the amplitudes of the Fast fourier transform modes present in the original trajectory. Using the developed method, we reconstructed 2471 trajectories from pERK discrete dynamics data set consisting of distribution of fluorescence data across 16 time points obtained from single-cell flow cytometry. The dynamics of the standard deviation of the reconstructed trajectories is comparable to that of the original fluorescence data.

Keywords: Time-series reconstruction, Single-cell, Cell-to-cell variability, Discrete data, Flow cytometry, Autocorrelation

1. INTRODUCTION

Quantitative dynamical analysis of cellular systems hinges on the availability of continuous time-series experimental data (Spiller et al., 2010). Cell-to-cell variability due to various intrinsic and extrinsic sources causing distribution of intra-cellular protein levels in isogenic population of cells is commonly observed in both prokaryotic and eukaryotic cells (Raser and O'shea, 2005; Raj and van Oudenaarden, 2008). This cell-to-cell variability is captured experimentally by measuring the intra-cellular protein levels across a population of cells using either confocal microscopy or flow cytometry technique (Bendall and Nolan, 2012).

Measurement of intra-cellular protein levels using single-cell experimentation techniques requires the presence of a reporter in the cell that emits fluorescence when the protein of interest is active. Such reporters can be incorporated into cells either via genetic fusion by a plasmid, that is, transfection (Meyer and Dworkin, 2007; Miyawaki et al., 2003), which is performed pre-stimulation during cell-culture or by intracellular staining (Krutzik and Nolan, 2003), which is incorporated post-stimulation. While the former method is amenable to continuous time *in situ* detection of levels of proteins of interest in an ensemble of live cells using single-cell tracking methods integrated with confocal microscopy (Lippinott-Schwartz et al., 2003), latter method is *not* amenable to *in situ* measurements altogether and thus only discrete time measurements is possible from cells in which metabolism and signalling are arrested. This is due to the fact that the intracellular staining which inserts monoclonal antibody (conjugated with dyes that emit fluorescence upon activation) that is specific to the protein of interest requires fixation and permeabilization of cells (Krutzik and Nolan, 2013). While fixation arrests the metabolic activity and freezes the state of the cells, permeabilization creates pores in the cell wall to help antibodies enter the cell and bind to the protein of interest. Thus, *in situ* continuous-time live-cell measurement at single-cell level using intra-cellular staining method is infeasible.

Since transfection of reporters into cells is a tedious and difficult task, such plasmid reporter-based detection cannot be achieved for all proteins of interest. However, intracellular staining can be done for many proteins as antibodies for several proteins are now readily available. While single-cell experimentation to detect intracellular proteins is feasible, it can at best only offer discrete time-series information for an ensemble of cells and the individual cell time-continuous trajectories cannot be obtained. Given that continuous time-series is needed for meaningful analyses to understand behaviour of a population of cells, to construct and validate systems biology based stochastic models, we ask the question is it possible to reconstruct the time-continuous single-cell trajectories using the discrete time-series data. This study

proposes a method to address this question by back-calculating the fluorescence values at different time points in a trajectory by assuming appropriate model autocorrelation information. In section 2, we briefly present the nature of discrete time-series experimental data on phosphorylated ERK (pERK) dynamics and describe the method for reconstruction of the trajectories. In section 3, we use a plasmid-reporter based data measured at single-cell level using confocal microscopy to validate the predictions of the algorithm. In section 4, we use the algorithm to reconstruct a discrete time-series single-cell data of pERK levels in an ensemble of cells measured using flow cytometry.

2. RECONSTRUCTION OF SINGLE-CELL TRAJECTORIES

2.1 *Information reflecting protein levels in cells*

Intracellular protein levels are usually measured by detecting the intensity of fluorescence emitted by an active conjugated dye excited at a certain pre-specified, precise wavelength laser beam. These dyes become active only when the monoclonal antibody to which it is attached is bound to the protein whose level is to be measured. Monoclonal antibodies are designed to have an exclusive binding affinity towards an epitope, a sequence of amino acids, that is a unique signature of the protein of interest. (Note that monoclonal antibody specific to an active form of protein does not bind to the inactive form of the same protein.) Thus the intensity of fluorescence emitted by the conjugated dye from a cell quantifies the number of active protein molecules present in the cell to which the antibody is bound.

2.2 *Reconstruction of intracellular single-cell phosphor-protein time-series*

We propose a 4-step algorithm for the reconstruction of intracellular single-cell phosphor-protein time trajectories based on the discrete-time fluorescence information in ensemble of cells. Flow chart outlining the steps involved in the reconstruction algorithm is in Figure 1. In this study, we assume that (a) fluorescence intensities across an ensemble of cells at certain discrete time points of interest are available, (b) protein level in a cell (reflected by the fluorescence emitted) at a certain time point is independent of that in any other cell in the same population, and (c) average behaviour is oscillatory, that is, the time series constructed by joining the average fluorescence at different discrete time points display/indicate oscillations. As the intracellular antibody staining requires arresting cell's state at a certain measurement time, distribution obtained at different time points are from different ensemble of isogenic cells grown under identical conditions and thus discrete.

We next briefly explain the four steps involved in the proposed novel reconstruction algorithm (Fig. 1) implemented in Matlab®.

I: Experimental data (Fig. 1, Step I): The fluorescence data is assembled as an unordered set

$$\{\boldsymbol{E}\}_{(\geq N)\times T} = (\boldsymbol{E}_1, \cdots, \boldsymbol{E}_T) \quad [1]$$

where, $\boldsymbol{E}_1, \cdots, \boldsymbol{E}_T$, respectively are the fluorescence vectors at time $t = 1, \cdots, T$, where T is the total number of time points for which measurements are available and N representing the number of trajectories to be reconstructed. Note that we set N

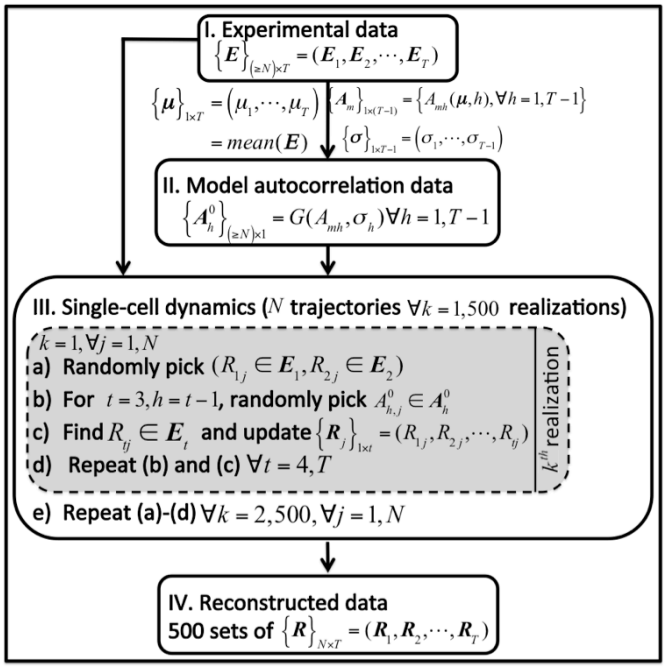

Fig. 1: Reconstruction of an ensemble of single-cell time trajectories from discrete-time fluorescence data

as the number of raw data points in the distribution at that time point which contains the minimum number of experimentally detected cells across all datasets. We next estimate the mean fluorescence vector $\{\boldsymbol{\mu}\}_{1\times T} = (\mu_1, \cdots, \mu_T)$ which constitute the elements of the mean fluorescence time series and the vector of lag-h autocorrelation $\{\boldsymbol{A}_m\}_{1\times(T-1)} = \{A_{mh}(\boldsymbol{\mu}, h), \forall h = 1, T-1\}$ of the mean time series, which will henceforth be referred to as mean autocorrelation.

II. Model autocorrelation data (Fig. 1, Step II): Every single-cell trajectory must have its own lag-h autocorrelation leading to a distribution of autocorrelations $\boldsymbol{A}_h^0, \forall h = 1, T-1$ corresponding to an ensemble of such N trajectories. However, as the trajectories are not available, it is not possible to *a priori* estimate the distribution of autocorrelations from the experimental data set $\boldsymbol{E}$, which is discrete in time. Therefore, we assume a certain model form G such as Gaussian distribution for $\boldsymbol{A}_h^0$. (We show in the next section that Gaussian distribution is a good model for the MAPK system whose time trajectories reconstruction is demonstrated in this study. For cases where prior information is unavailable, Gamma distribution, which captures protein level distribution in resting cells could be used.) In order to achieve this, based on the literature time-series data for similar systems, we assume a certain standard deviation vector $\{\boldsymbol{\sigma}\}_{1\times(T-1)} = (\sigma_1, \cdots, \sigma_{T-1})$ for the autocorrelation distributions $\boldsymbol{A}_h^0, \forall h = 1, T-1$ to be constructed. Next, we use the mean autocorrelation vector $\boldsymbol{A}_m$ and the literature data-based standard deviation vector $\boldsymbol{\sigma}$ to arrive at a distribution of autocorrelations $\boldsymbol{A}_h^0, \forall h = 1, T-1$ lags using the probability density function $G(A_{mh}, \sigma_h)$ corresponding to the Gaussian distribution as the base model. (Note that use of any other

probability density function as model for $G(A_{mh},\sigma_h)$ does not alter the proposed algorithm.)

III. Single-cell trajectories: Using the autocorrelation distributions and the original experimental data, we next construct the single-cell trajectories, one-by-one, *N* of them. We refer to one set of *N* trajectories as one reconstruction realization and construct $k=1,\cdots,500$ such realizations altogether. We next outline the procedure for obtaining a reconstruction realization.

In order to construct the j^{th} trajectory $\{\boldsymbol{R}_j\}_{1\times T}$ ($\forall j=1,N$), we first randomly pick one cell each from the experimental fluorescence distributions for the first and second time points (Fig. 1, Step III(a)). Assume the corresponding fluorescence values are $R_{1j}\in\boldsymbol{E}_1$ and $R_{2j}\in\boldsymbol{E}_2$, respectively. Next, to predict the fluorescence value in the j^{th} trajectory for the next time point, say t=3, we randomly pick an autocorrelation value $A^0_{2,j}$ from the model autocorrelation distribution corresponding to lag-2, that is, from $\boldsymbol{A}^0_2$ (Fig. 1, Step III(c)). Using a re-arranged expression for the autocorrelation $A^0_{2,j}$ of the form

$$A^0_{2j}=\frac{\left(R_{1j}-\left(2\bar{R}_{2j}+R_{3j}\right)/3\right)\left(R_{3j}-\left(2\bar{R}_{2j}+R_{3j}\right)/3\right)}{\frac{1}{3}\left(2\bar{\sigma}^2_{2j}+\left(R_{3j}-\left(2\bar{R}_{2j}+R_{3j}\right)/3\right)^2\right)} \quad [2]$$

where, $\bar{R}_{2j}=\frac{1}{2}\sum_{k=1}^{2}R_{kj}$ and $\bar{\sigma}^2_{2j}=\frac{1}{2}\sum_{l=1}^{2}\left(R_{lj}-\bar{R}_{2j}\right)^2$ which, upon expanding takes the form of a quadratic equation in R^e_{3j}. (Detailed derivation of Eq. (2) is presented in Appendix I.) We estimate R^e_{3j} by solving this underlying quadratic equation. After rejecting the unfeasible solution – negative – for the estimate, we identify the cell having a fluorescence of say $R_{3j}\in\boldsymbol{E}_3$ in the original data that is closest or equal to the estimate R^e_{3j}. We choose this fluorescence value $R_{3j}\in\boldsymbol{E}_3$ to be that corresponding to the t=3 time point in j^{th} trajectory. We then update, upto the current time point (t=3), the reconstructed time-series vector $\{\boldsymbol{R}_j\}_{1\times3}=(R_{1j},R_{2j},R_{3j})$. Next, in Step III(d) (Fig.1), we repeat the estimation of the fluorescences R_{tj}, $\forall t=4,T$ by following the same procedure of randomly picking autocorrelation value $A^0_{h,j}$ with $h=t-1$ from the model autocorrelation distribution data $\boldsymbol{A}^0_{h,j}$ corresponding to lag-*h* and by solving the underlying quadratic equation for the autocorrelation $A^0_{h,j}$ of the form

$$A^0_{hj}=\frac{\left(R_{1j}-\left(h\bar{R}_{hj}+R_{tj}\right)/t\right)\left(R_{tj}-\left(h\bar{R}_{hj}+R_{tj}\right)/t\right)}{\frac{1}{t}\left(h\bar{\sigma}^2_{hj}+\left(R_{tj}-\left(h\bar{R}_{hj}+R_{tj}\right)/t\right)^2\right)} \quad [3]$$

where, $\bar{R}_{hj}=\frac{1}{h}\sum_{l=1}^{h}R_{lj}$ and $\bar{\sigma}^2_{hj}=\frac{1}{h}\sum_{l=1}^{h}\left(R_{lj}-\bar{R}_{hj}\right)^2$. (Detailed derivation is in Appendix I.) Note that while solving the quadratic equation for the t^{th} timepoint $\forall t=4,T$, to estimate R^e_{tj}, we assume all elements of $\{\boldsymbol{R}_j\}_{1\times(t-1)}=(R_{1j},\cdots,R_{(t-1)j})$ are already estimated. We then update $\{\boldsymbol{R}_j\}_{1\times T}=(R_{1j},\cdots,R_{Tj})$ and thereby obtain the trajectory.

Steps III(b) and (c) are then repeated to obtain trajectories $\forall j=1,N$ constituting trajectory data set for 1 realization. Steps III (a-d) are then repeated *k*=500 times to obtain as many reconstruction realizations.

IV. Reconstructed data sets: The data sets corresponding to all the realizations are then placed in an appropriate database for further use of the reconstructed trajectories.

3. RECONSTRUCTION OF CONTINOUS TIME SAMPLE: VALIDATION

In this section, we validate the proposed method by implementing the algorithm on a test continuous time-series dataset from literature where the fluorescence-tagging is achieved by inserting an appropriate plasmid into the cells. This validation is carried out by first assuming that the distribution of fluorescence measured is not continuous time, that is, assume that the fluorescence distribution across different time points are from different cell population. We then reconstruct the trajectories, pair those reconstructed with the original using appropriate metrics and quantify the extent of prediction.

3.1 *Sample fluorescence data*

Using a lentiviral transfected PC12 cells, Ryu et al. (Ryu et al., 2015) quantitatively measured the levels of pERK continuously in live-cells stimulated with various concentrations of growth factors NGF/EGF. (Note that the transfection via a plasmid enabled genetic fusion of a biosensor capable of fluorescing upon activation of a certain interested protein.) We used the dataset consisting of 12100 data points (121 (=60 minutes x 2 measurement per minute + 0 minute) x 100 cells) corresponding to sustained stimulation of transfected PC12 cells with 25 ng/ml of EGF, source data for which is available as supplementary information. This continuous time trajectories measurement was obtained by immobilising the cells through an appropriate highly sophisticated, micro-fluidic device and by detecting the fluorescence emitted by individual cells using confocal microscopy.

3.2 *Reconstruction of the trajectories of pERK levels in transfected PC12 cells stimulated with EGF*

Assuming the fluorescence distribution at 121 time points to be discrete, we estimated the model autocorrelation data $\boldsymbol{A}^0_h$ with $\forall h=1,120$ from the original data. The distribution of the autocorrelation displayed a gaussian distribution behaviour at all time points. The mean and the standard deviation of the autocorrelation distributions across 121 time points along with sample histograms (at two different time lags) fit with a Gaussian probability density function is presented in Fig. 2.

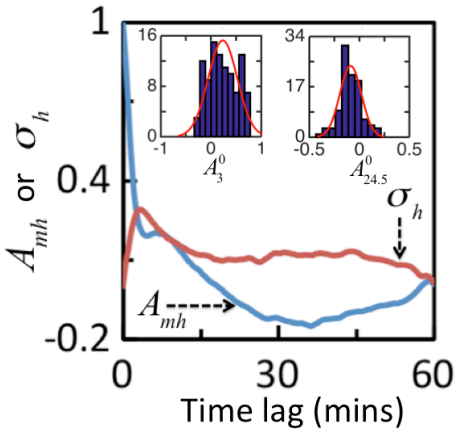

Fig. 2: Mean and standard deviation of the autocorrelation distributions estimated with different time lags. Inset: Autocorrelation distribution with time lag of 3 mins and 24.5 mins fit with Gaussian model (red line).

After constructing the autocorrelation distributions, we employed the algorithm to reconstruct the time trajectories: 500 realizations each consisting 100 reconstructed trajectories. Demonstration of the ability of the algorithm to reliably reconstruct, at least a fraction of the maximum possible trajectories hinges strongly on the pairing of the reconstructed ones with the original trajectories. Identification of the right pair is a complex problem. However, in the next sub-section, we present a possible quantitative approach to address the pairing problem.

3.3 *Pairing of reconstructed and original time trajectories*

Rational pairing of the two time profiles requires quantitative pattern matching techniques. A good match will typically have same number of occurrences of fluorescence values in both time-series and will also have them ordered in the same sequence. In order to account for these two primary features while pairing, we employ as metrics

(a) mutual information score

$$M_{jn}(x,y) = \sum_x \sum_y p(x,y)\log_2\left[p(x,y)/(p_1(x)p_2(y))\right] \quad [4]$$

and

(b) the Euclidean distance P_{jn}

between the jth reconstructed trajectory and nth original trajectory. In Eq. (4), $p(x,y), p_1(x)$, and $p_2(y)$, respectively are the joint probability of finding values x and y in the vector of fluorescences in reconstructed $(\boldsymbol{R}_j)$ and original $(\boldsymbol{E}_n)$ trajectories, probability of finding value x in $\boldsymbol{R}_j$, and that of finding value y in $\boldsymbol{E}_n$. We estimated the mutual information score and the Euclidean distance (pdist function in Matlab®) between all reconstructed and all original trajectories for all 500 realizations. Note that while larger M_{jn} indicate that the fraction of time points containing same fluorescence values is similar in both $j^{th} \in \boldsymbol{R}$ and $n^{th} \in \boldsymbol{E}$ trajectories, smaller P_{jn} indicate good similarity between the $j^{th} \in \boldsymbol{R}$ and $n^{th} \in \boldsymbol{E}$ trajectories. In order to account for both these factors, an appropriate cut-off on the M_{jn} and P_{jn} is needed. As there is no independent way to set a cut-off for these and the reconstructed data is not expected to capture 100% information present in the original data, we selected pairs by considering different cut-offs for M_{jn} in $[0.1,1]$ and for P_{jn} in $[0.25,2.5]$. In particular, for every realization and for every M_{jn} and P_{jn} cut-off, we choose as unique pairs, out of 100×100 possible pairs, those that exhibit (a) M_{jn} above its cut-off and (b) P_{jn} below its cut-off. In Fig. 3, we present a heatmap of the average number of unique pairs that can be selected for a pre-specified combination of cut-offs for M_{jn} and P_{jn}. (Note that the average here is across the 500 realizations.) Standard deviation corresponding to each of the combinations is indicated on the heatmap.

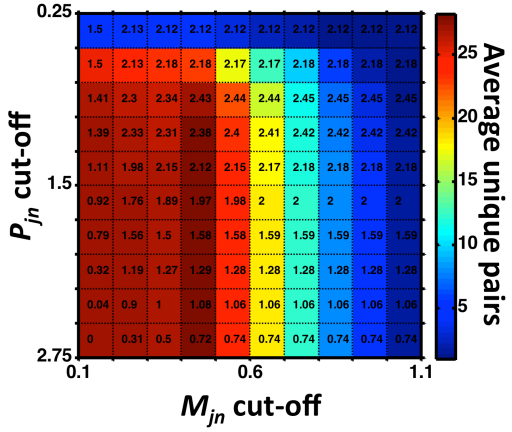

Fig. 3: Heatmap of the average number of unique pairs of reconstructed and original trajectories satisfying the cut-offs on M_{jn} and P_{jn} between them. The average is across 500 realizations and for each realization every selected unique pair satisfied both the cut-offs. The numbers presented in every box represents the standard deviation.

3.4 *Comparison of the two trajectories of the unique pairs*

Extent of representation of information in a reconstructed trajectory can be assessed by comparing features of the reconstructed and original trajectories paired. We consider the case that satisfies the conditions

$$M_{jn} > 0.5, P_{jn} < 1.5 \quad [5]$$

for which the average unique pairs is 28.1 (±2.12). In Fig. 4, for a particular realization, we present a comparison of the trajectories from the reconstructed set and that it is likely to correspond to in the original dataset based on the conditions in Eq. (5). In Fig. 5, we show comparison of top three (out of 28) based on the cut-offs chosen (Eq. 5).

Figures (4) and (5) together suggest that the reconstruction algorithm is able to capture, within the scope of the pairing method chosen, reasonable number of original trajectories. The algorithm is unable to capture all the crests and troughs of the trajectories accurately.

Visual inspection of the profiles is insufficient to exactly assess what features of the original trajectory the algorithm is able to accurately capture in the reconstruction. In order to understand the extent of capture of the features of the original profile, we perform fast fourier transform (FFT) on the paired reconstructed and original trajectories and compare the amplitudes of the leading modes. We conducted this analyses on a selected few realizations and present for one of them.

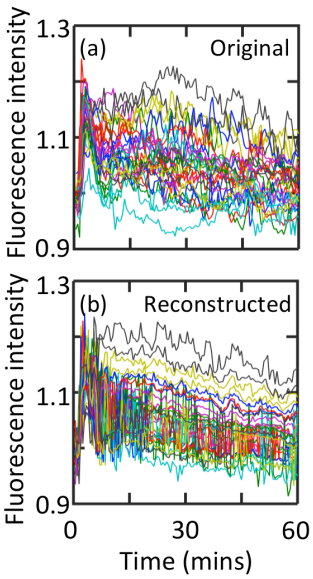

Fig. 4: Trajectories in the (a) original data and (b) reconstructed data that correspond to the 26 unique pairs.

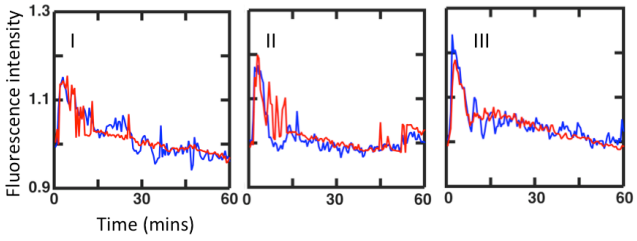

Fig. 5: Comparison of the time profiles of the top three reconstructed (red) and the original (blue) trajectory pairs for the realization corresponding to that in Fig. 4.

Figure 6a shows the amplitudes of the FFT modes of the original and reconstructed trajectory pair in Fig. 5(I). Moreover, in Fig. 6b, we present, for the first 10 FFT modes, box plots showing the quartiles of distribution of the fractional deviation of the amplitude of reconstructed trajectory from that of the corresponding mode of the original trajectory. Figure 6b suggests that the there are pairs where the amplitudes of the first 10 modes (which is expected to capture the dominant dynamics) match very well. However, unfortunately the selection methodology also has led to pairs where the amplitude of the dominant modes in reconstructed trajectory deviate appreciably from that in the original, as reflected by the high standard deviation in the FFT amplitudes (Fig. 6b). Moreover, a comparison of the paired trajectories (Fig. 5a) suggest that the reconstruction does not capture the larger frequencies (higher modes) very well, as is also evident from the deviation in the amplitudes of the higher modes for the matched pair (Fig. 6a).

4. RECONSTRUCTION USING FLOW CYTOMETRY DATA

Measurement of intracellular protein levels using flow cytometry requires cell fixation that arrests the cell's metabolic and signaling state and thus rendering it (cell) unusable any further. Thus, the only dynamical information that is available is discrete distribution of the fluorescence emitted by different (isogenic) cells at different time points.

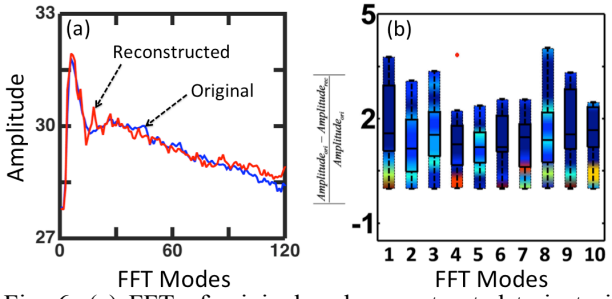

Fig. 6: (a) FFT of original and reconstructed trajectories in Fig. 5(I). (b) Heatmap of and boxplot – overlayed – showing different quartiles of the distribution of the fractional difference in the amplitudes of the first 10 FFT modes.

As demonstrated in the previous section, the algorithm provides a reasonable reconstruction of such discrete data. In this section, we present reconstruction of an ensemble of single-cell trajectories using discrete flow cytometry fluorescence data.

4.1 *pERK experimental data from flow cytometry*

Jurkat-E6.1 cells were constantly stimulated with 10ng/ml of PMA, which is a pan-TLR agonist (Kalantre et al., 2016). pERK levels were estimated by detecting the fluorescence emitted by cells stained with pERK monoclonal antibody conjugated with Alexa Fluor 488 dye (BD Biosciences). Single-cell pERK levels were captured at 16 discrete time points between 0, 90 mins with a uniform interval of 6 mins. The number of cells captured and the mean of the distribution at these time points are in Table 1. (Note that since different time point data is from different population of isogenic cells, the number of dead cells/debris varies and thus capturing fluorescence from same number of cells is impractical.)

Table 1: Number of cells captured at and mean pERK level at discrete experimental time points

t (mins)	No. of cells	A_{mh}	t (mins)	No. of cells	A_{mh}
0	26759	448	48	24460	509
6	6828	493	54	8285	508
12	5743	493	60	3958	509
18	25289	502	66	8236	491
24	16686	500	72	6766	516
30	5408	508	78	7660	506
36	12811	522	84	8001	502
42	2471	515	90	14141	511

4.2 *Reconstruction of the single-cell trajectories*

We used the algorithm (Fig. 1) and the discrete raw fluorescence data and reconstructed 2471 single-cell time-series trajectories. As the distribution at different time points measured in sample continuous time data exhibited Gaussian distribution for the autocorrelation (Fig. 2) for all time points, we employed the same distribution along with the corresponding time-averaged standard deviation of 0.125 as the model input parameter (σ_h). Reconstruction was also performed for various $\sigma_h \in [0.015, 0.28]$ (Fig. 2). Fig. (7a) shows the reconstruction along with the mean and standard deviation at every time point. Comparison of time-series of

standard deviation of original data and those of reconstructed trajectories for various (σ_h) (Fig. 7b) suggests that reconstruction is insensitive to model input parameter.

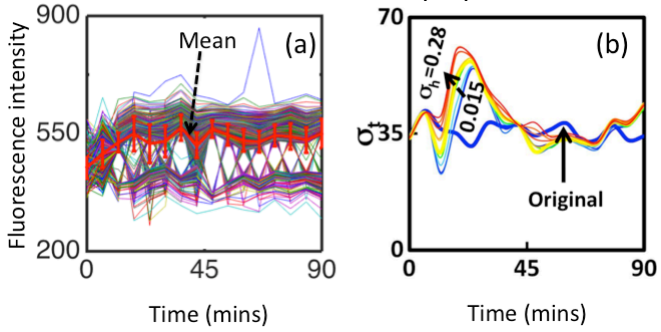

Fig. 7: (a) Time-series of 2471 reconstructed trajectories from flow cytometry data for $\sigma_h = 0.125$. (b) Comparison of the time series of the standard deviation of reconstructed with various σ_h and original data.

5. CONCLUSIONS

Knowledge of continuous time-series is of immense value for quantitative prediction of intracellular dynamics. Such continuous time-series information is unavailable when measuring single-cell intra-cellular protein levels using antibody as reporter. Only time-discrete distribution of time-series information is available. Our study is the first attempt to reconstruct an ensemble of trajectories based on such discrete data set when the mean data across the measurement time points display oscillatory behaviour. Model autocorrelation function was used to predict the fluorescence values at every time point.

We validated the reconstruction algorithm using the continuous time-series pERK dynamics in transfected cell lines in which plasmid was used to genetically fuse the reporters. The algorithm was able to predict ~30 reconstructed trajectories that matched with as many distinct original ones based on the cut-offs on the mutual information score and Euclidean distance as the pairing constraints.

We used the algorithm to predict 2471 trajectories of the pERK dynamics from the intra-cellular measurement of the fluorescence at 16 time points using flow cytometry. Comparison of the dynamics of the standard deviation of the reconstructed trajectories and those of the original data suggest that the reconstruction is reasonable and is insensitive to model input parameter σ_h.

Trajectory predictions made in this work is based on the approximate autocorrelation function for estimating the value of the fluorescence at the time point of interest. Several improvements can be made, for example, use of the complete autocorrelation function or introduction of iterative correction scheme or a different model autocorrelation function itself. If prior data indicating good model is unavailable, Gamma distribution capturing the underlying cell-to-cell variability of inactive protein levels could be used as a first approximation (Cai et al., 2006) for the autocorrelation distribution.

ACKNOWLEDGEMENTS

We thank DST, Govt. of India for support of this work.

REFERENCES

Bendall, S.C., Nolan, G. (2012) From single cells to deep phenotypes in cancer. *Nat. Biotech.* 30, 639-647.

Cai, L. et al. (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature*. 440, 358-362.

Kalantre, G., Viswanathan, G.A., (2016) Unpublished results.

Krutzik, P., Nolan, G. (2003) Intracellular phospho-protein staining techniques for flow cytometry: Monitoring single cell signaling events. *Cyto. Part A* 55A:61–70.

Lippinott-Schwartz, J., et al. (2003). Photobleaching and photoactivation: following protein dynamics in living cells. *Nat. Cell biol.* 5,S7-14.

Meyer, P., Dworkin, J. (2007) Applications of fluorescence microscopy to single bacterial cells. *Res Microbiol*. 158, 187-194.

Miyawaki, A., et al.. (2003) Lighting up cells: labelling proteins with fluorophores. *Nature Cell biology*. 5, S1-S6.

Raj, A., van Oudenaarden, A. (2008) Nature, nurture or chance: Stochastic gene expression and its consequences. *Cell* 135,216-226

Raser, J.M., O'Shea, E.K. (2005) Noise in gene expression: Origins, consequences, and control. *Science* 309, 2010-2013.

Ryu H, et.al.(2015) Frequency modulation of ERK activation dynamics rewires cell fate. *Mol. Syst. Biol*.11:838.

Spiller, et.al.(2010)Measurement of single-cell dynamics, *Nature* 465, 736–745

Appendix A: AUTOCORRELATION FUNCTION

Suppose that $\boldsymbol{R} = (R_1, \cdots, R_{t-1})$ data points in a time-series are already estimated and it is desired to estimate R_t, the data point at *t*. Autocorrelation function of a time-series for a time lag of *h*=*t*-1 of with data points $\boldsymbol{R} = (R_1, \cdots, R_t)$ is given by

$$A_h^0 = (R_1 - \bar{R}_t)(R_t - \bar{R}_t) \big/ \bar{\sigma}_t^2 \qquad \text{[A1]}$$

where, mean

$$\bar{R}_t = \frac{1}{t}\sum_{l=1}^{t} R_l = \frac{1}{t}\left[R_t + \sum_{l=1}^{t-1} R_l\right] = \frac{1}{t}\left[R_t + h\bar{R}_h\right] \qquad \text{[A2]}$$

and variance approximated as

$$\bar{\sigma}_t^2 = \frac{1}{t}\sum_{l=1}^{t}\left(R_l - \bar{R}_t\right)^2 \approx \frac{1}{t}\left[\left(R_t - \frac{1}{t}\left[R_t + h\bar{R}_h\right]\right)^2 + h\bar{\sigma}_h^2\right] \qquad \text{[A3]}$$

Note that here we express the variance $\bar{\sigma}_t^2$ at the current time point *t* as an approximate function of the penultimate variance $\bar{\sigma}_h^2$ at *h*=*t*-1. Using Eqs (A2) and (A3) into Eq. (A1), Eq. (A1) can be re-written into that in Eq. (3) in the main text. Eq. (2) can be obtained by setting *t*=3 in Eq. (3).