TCU SCHOOL of MEDICINE

Variation of Best Fit Distributions in Single Cell Virus Dynamics Models

Madison Doty[1], Hana Dobrovolny[1,2]

[1]Texas Christian University School of Medicine, Fort Worth, TX

[2]Department of Astronomy and Physics, Texas Christian University, Fort Worth, TX

ABSTRACT

Mathematical modeling of viral kinetics can be used to gain further insight into the viral replication cycle and virus-host interactions. However, many of the virus dynamics models do not incorporate the cell-to-cell heterogeneity of virus yield or the time-dependent factor of virus replication. A recent study of vesicular stomatitis virus (VSV) kinetics in single BHK cells determined that both virus production rate and yield of virus particles varies widely between individual cells of the same cell population. Here we use the results of the previously mentioned study to determine the distribution that best describes the time course of viral production within the single cells. We determined a list of eight potential distributions that are commonly used in viral kinetics models to fit to each data set by minimizing the sum of squared residuals. The model of best fit for each individual cell was determined using Akaike's Information Criterion (AIC_C). Results of this study show that the distribution that best describes viral production varies from cell to cell. This finding could have further reaching implications for incorporating time-dependent viral production into a standard model of virus kinetics in order to better reproduce the diversity of viral replication that occurs over time within a population of cells.

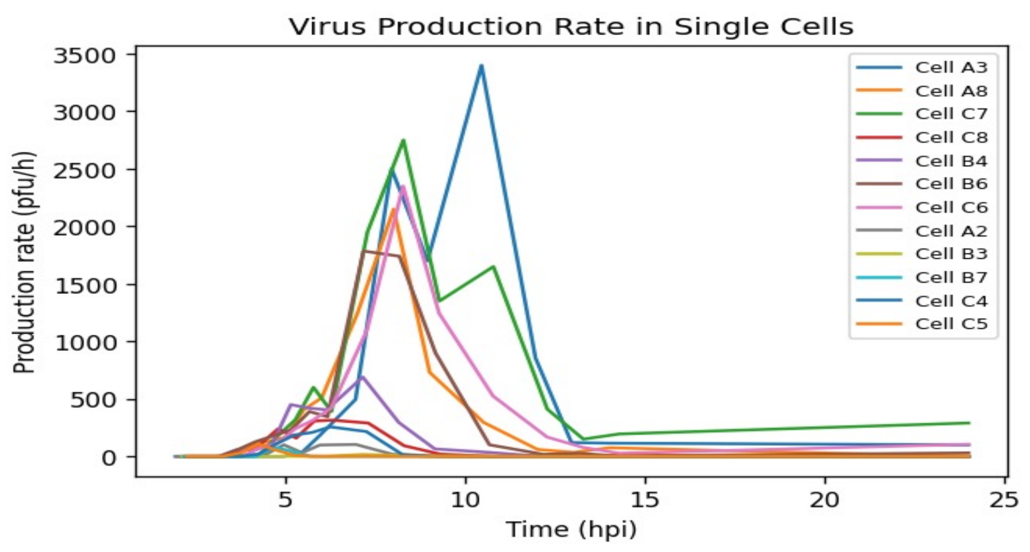

Figure 1. Production rate profiles for 12 cells drawn from the same initial population of infected cells.

METHODS

- The eight candidate distributions (Table 1) were fit to data from a previous study on VSV production (Timm and Yin, 2012) (Figure 1) by minimizing the sum of squared residuals (SSR).
- The best fit distribution was chosen based on Akaike's information criterion (AICc).
- The model with the lowest AIC_C was considered to be the better model given the experimental data it is approximating.

Distribution	Equation
Exponential	$1 - e^{-\lambda t}$
Normal	$\frac{1}{2}\left(1 + \text{erf}\left(\frac{t-\rho}{\sqrt{2\sigma^2}}\right)\right)$
Weibull	$1 - e^{-\left(\frac{t}{\eta}\right)^{\gamma}}$
Gamma	$\frac{1}{\Gamma(\gamma)\eta^{\gamma}} \int_0^t s^{\gamma-1} e^{-\frac{s}{\eta}} \mathrm{d}s$
Poisson	$e^{-\lambda} \sum_{i=0}^{t} \frac{\lambda^i}{i!}$
Gompertz	$1 - e^{-\eta\left(e^{bt}-1\right)}$
Log-normal	$\frac{1}{2}\left(1 + \text{erf}\left(\frac{\ln t - \mu}{\sqrt{2\sigma^2}}\right)\right)$
Logistic	$\frac{a}{1+e^{-b(t-c))}}$

Table 1. List of candidate distributions.

RESULTS

Six of the eight models provided a relatively good fit for the twelve sets of virus production data as seen in the best fit graphs (Figure 2). Two of the models, exponential and log-normal, did not provide a good fit for any of the data sets. The normal distribution provided the best fit according to the AIC_C scores for seven of the data sets. For the remaining data sets, the Weibull distribution was a best fit for three cells and the logistic and Poisson distributions were best fit for one cell each. The AIC_C scores for each candidate distribution are reported below (Table 2).

BEST FIT CURVES

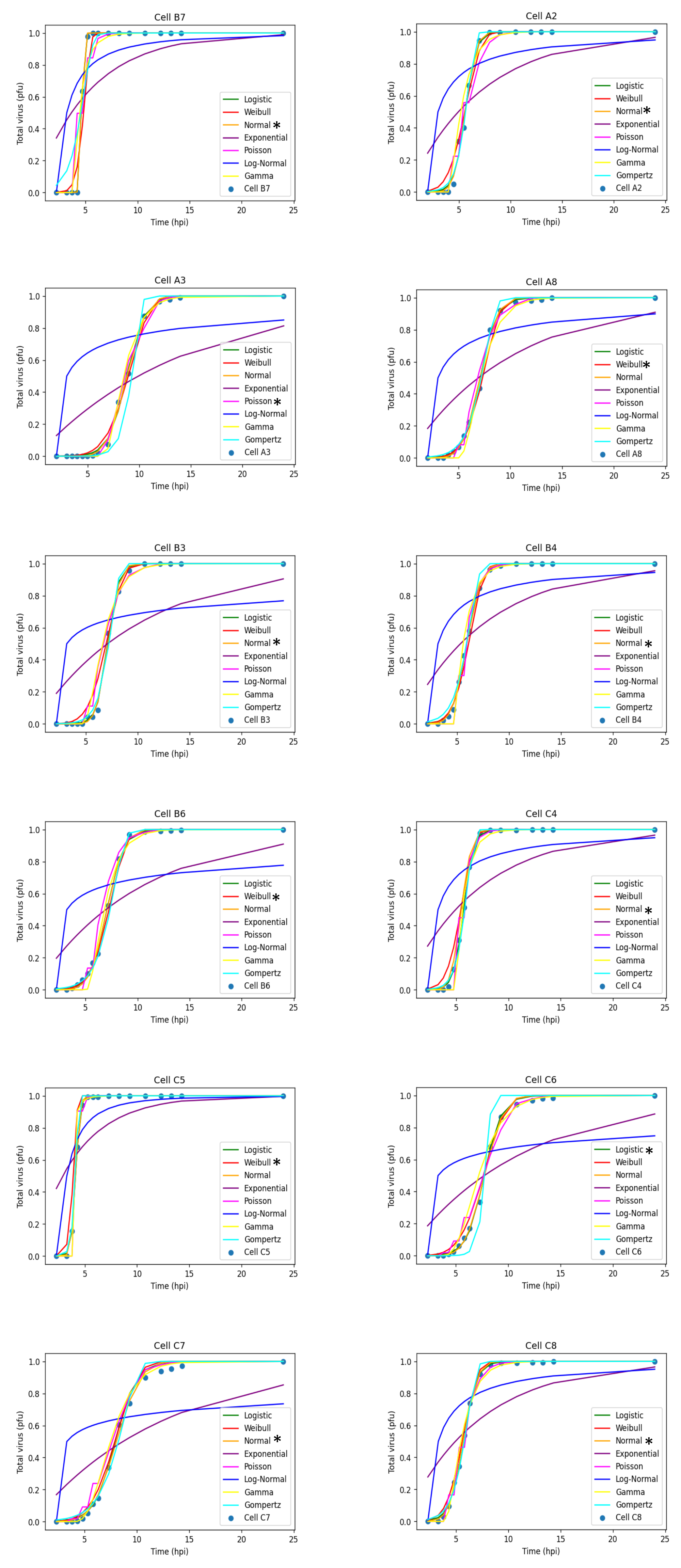

Figure 2. One-step production of virus from single cells with best fit curves of the possible models.

* Denotes the distribution that has the lowest AIC_C score for each cell.

AIC_C SCORES

	A2	A3	A8	B3	B4	B6	B7	C4	C5	C6	C7	C8
Logistic	102.014	192.438	158.727	9.031	157.357	161.373	58.76	132.53	73.141	**165.707**	194.8	113.772
Exponential	152.508	256.824	240.375	79.519	212.658	240.75	120.973	184.809	130.321	244.716	256.151	195.667
Normal	**82.083**	173.279	158.014	**2.825**	**122.645**	152.719	**-1.339**	**75.808**	**25.847**	166.377	**187.182**	**107.294**
Weibull	85.826	180.291	**155.27**	12.528	136.392	**137.335**	-1.18	89.798	44.431	166.979	193.428	115.742
Gamma	101.457	215.285	183.747	36.784	147.087	200.711	71.062	126.567	75.14	191.277	192.491	130.548
Poisson	119.002	**166.33**	181.27	32.884	175.501	193.457	101.634	144.621	100.704	193.806	199.46	156.074
Gompertz	90.446	189.408	161.819	17.688	149.241	145.217	95.893	101.641	46.057	172.827	224.266	127.364
Log-Normal	162.39	275.886	256	92.577	223.703	254.295	122.426	193.229	124.736	259.398	272.664	204.247

Table 2. AIC_C scores for each candidate distribution. Bolded scores denote the distribution of best fit.

CONCLUSIONS

It is clear from previous studies that virus production varies over time from cell to cell. Our results show that the variable viral production rates in this study are best fit different distributions. Overall, it is noted that the normal distribution provides the best fit for most of the one-step virus production curves.

FUTURE DIRECTIONS

- The results of this study will be incorporated with time-dependent production into a standard viral kinetics model.
- We will simulate this model and compare the results to simulations of the model with constant production.

ACKNOWLEDGEMENTS

I would like to thank my mentor for this project, Dr. Hana Dobrovolny, for her help and support. I would also like to sincerely thank Dr. John Yin for generously providing us the viral production data for this project.