Supplementary Information to

Speech-generated Aerosol Settling Times and Viral Viability Predict COVID-19 Transmission Efficiency

Alan Y. Gu[#], Yanzhe Zhu[#], Jing Li and Michael R. Hoffmann*

Linde Laboratories

California Institute of Technology

Pasadena, California 91125, USA

[#]Equal contribution

*Correspondence and requests for materials should be addressed to MRH

Tel: 626-395-4391 Email: mrh@caltech.edu

This file includes:

Supplementary Notes S1

Supplementary Table S1

Supplementary Figures S1 to S2

February 2021

Supplementary Notes

Supplementary Notes 1

Code for this work is available at: https://github.com/zhuyanzhe98/evaptransmission

Supplementary Tables and Figures

a. R^2 for Fitting

	LR	VAR	Simple RNN	LSTM
Harris County, TX	0.362	0.616	0.374	0.99998
King County, WA	0.540	0.702	0.979	0.99998
LA County, CA	0.805	0.882	0.772	0.99997
Maricopa County, AZ	0.530	0.873	0.712	0.99995
Santa Clara County, CA	0.719	0.721	0.745	0.99998

b. Sum of Squares of Residuals for Prediction

	LR	VAR	Simple RNN	LSTM
Harris County, TX	2.02E-03	3.87E-02	9.54E-04	4.37E-03
King County, WA	1.76E-03	3.85E-03	9.97E-04	1.30E-03
LA County, CA	2.32E-02	8.79E-03	1.20E-02	2.47E-03
Maricopa County, AZ	1.27E-03	1.56E-02	6.60E-04	4.19E-04
Santa Clara County, CA	1.02E-02	1.87E-02	9.29E-03	8.47E-03

c. Fitting and prediction for Maricopa County, May-August

	LR	Simple RNN	LSTM
R^2 for Fitting	0.172	0.579	0.99996
Sum of Squares of Residuals for Prediction	1.10E-02	1.56E-02	1.60E-02

Table S1. (a) R-squared of model fitting and (b) sum of squares of residuals of model prediction for each county and model during April 2020. (c) R-squared of model fitting and sum of squares of residuals of model prediction for Maricopa County during extended time frame from May to August 2020. LR: linear regression; VAR: vector autoregression; Simple RNN: simple recurrent neural network; LSTM: long short term memory recurrent neural network

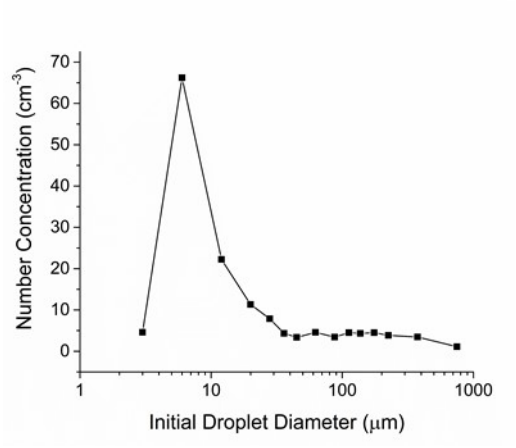
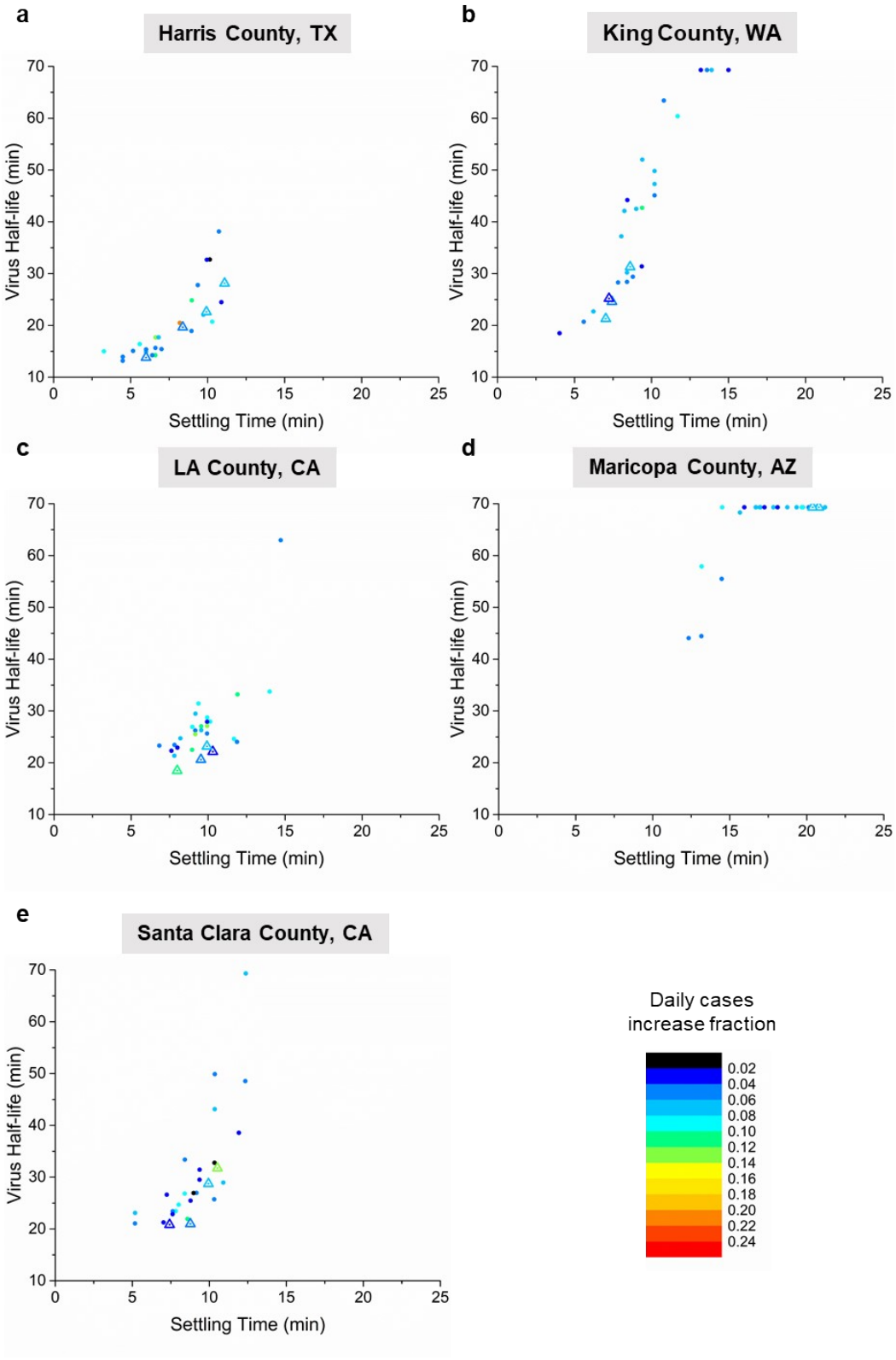


Figure S1. Size distribution of speech-generated droplets before evaporation measured by Morawska et al.[31] The peak number concentration is at 6 μm and is used as the input to the Köhler equation in this work.


a
Harris County, TX
b
King County, WA
c
LA County, CA
d
Maricopa County, AZ
e
Santa Clara County, CA
Virus Half-life (min)
Settling Time (min)
Daily cases increase fraction
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.18
0.20
0.22
0.24

34

Figure S2. Data points available to generate the contour plot in Fig. 5 for each county, with the daily case percentage increase as a function of settling time and viral viability (represented by half-life). Colour shows the daily case percentage increase in decimal. The daily cases percentage increase is shown as the colour of each data point. The dots represent data points for training, and the triangles represent data for prediction. LR: linear regression; VAR: vector autoregression; Simple RNN: simple recurrent neural network; LSTM: long short term memory recurrent neural network