Environmental Science & Technology

pubs.acs.org/est

Article

Predictive Modeling of Virus Inactivation by UV

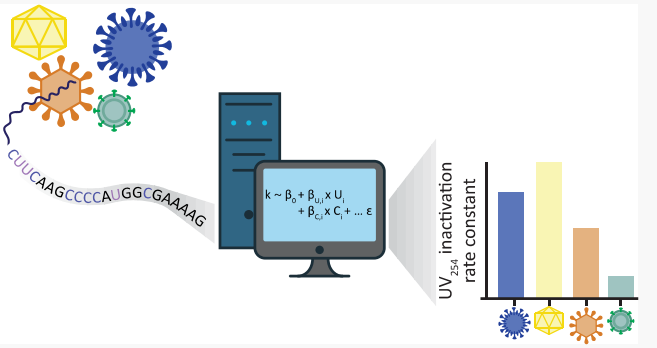
Nicole C. Rockey, James B. Henderson, Kaitlyn Chin, Lutgarde Raskin, and Krista R. Wigginton*

Cite This: https://dx.doi.org/10.1021/acs.est.0c07814

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: UV_{254} disinfection strategies are commonly applied to inactivate pathogenic viruses in water, food, air, and on surfaces. There is a need for methods that rapidly predict the kinetics of virus inactivation by UV_{254}, particularly for emerging and difficult-to-culture viruses. We conducted a systematic literature review of inactivation rate constants for a wide range of viruses. Using these data and virus characteristics, we developed and evaluated linear and nonlinear models for predicting inactivation rate constants. Multiple linear regressions performed best for predicting the inactivation kinetics of (+) ssRNA and dsDNA viruses, with cross-validated root mean squared relative prediction errors similar to those associated with experimental rate constants. We tested the models by predicting and measuring inactivation rate constants of a (+) ssRNA mouse coronavirus and a dsDNA marine bacteriophage; the predicted rate constants were within 7% and 71% of the experimental rate constants, respectively, indicating that the prediction was more accurate for the (+) ssRNA virus than the dsDNA virus. Finally, we applied our models to predict the UV_{254} rate constants of several viruses for which high-quality UV_{254} inactivation data are not available. Our models will be valuable for predicting inactivation kinetics of emerging or difficult-to-culture viruses.



INTRODUCTION

Viruses can cause diverse and costly illnesses in humans and other animals.[1] A variety of approaches have therefore been developed to decontaminate food, water, air, and surfaces that may contain infectious viruses.[2−7] UV_{254} treatment, in particular, is gaining popularity as an alternative to more traditional chemical disinfection strategies.[8−10] Viruses can have highly variable UV_{254} susceptibilities.[11,12] For example, two dsDNA viruses, adenovirus type 40 and bacteriophage T6, are inactivated by UV_{254} at widely varying rates, with rate constants of ∼0.06 cm^2 mJ^{-1},[13−18] and ∼5.4 cm^2 mJ^{-1},[19] respectively.

Viruses have diverse genome types, including double-stranded RNA (dsRNA), single-stranded RNA (ssRNA), double-stranded DNA (dsDNA), and single-stranded DNA (ssDNA). UV_{254} inactivates by primarily targeting viral genetic material, and the different biochemical structures associated with these viral genome types result in distinct sensitivities to UV_{254}.[20] Nucleic acid primary structure, or nucleotide base sequence, also affects UV_{254} genome reactivity; pyrimidine bases, for instance, are about an order of magnitude more reactive with UV_{254} than with purine bases.[21,22] The replication modes of viruses differ, and an enzyme of one virus may stall at a UV_{254} lesion that does not affect the replication enzyme of another virus. For example, the reverse transcriptase enzymes involved in the generation of retrovirus mRNA may have different sensitivities to photochemical modifications in nucleic acids compared to the RNA dependent RNA polymerase enzymes used by other RNA viruses to synthesize mRNA.[23] Additional differences in viral infection cycles impact virus sensitivity to UV_{254}.[24] dsDNA virus genomes, for instance, can undergo nucleic acid repair once inside host cells.[24−26] This means that a virus may be inactivated by UV_{254} treatment through base modification, only to be repaired and thus rendered infectious again when such repair mechanisms are available. We note these differences in virus genome type and mode of mRNA generation are utilized in the Baltimore virus classification system (e.g., Group 1: dsDNA viruses, Group IV: (+) ssRNA viruses).[1,27]

Virus disinfection methods are evaluated by enumerating infectious viruses before and after treatment, typically with virus culture systems. Reliance on culture-based approaches to evaluate inactivation kinetics is often challenging. Most notably, many human viruses that are spread through the environment are not readily culturable. For highly pathogenic viruses that are culturable, disinfection experiments are complicated by biosafety restrictions. Disinfection experiments with severe acute respiratory syndrome (SARS) coronaviruses (SARS-CoV-1 and SARS-CoV-2), for example, are limited to

Received: November 18, 2020
Revised: January 23, 2021
Accepted: January 26, 2021

biosafety level 3 laboratories, and work with ebola viruses require biosafety level 4 facilities. Alternative approaches for determining virus inactivation kinetics would be valuable, especially for difficult-to-culture and emerging viruses. Earlier studies have worked toward a predictive manner of evaluating UV_{254} virus inactivation based on virus attributes.[28,29] Recently developed modeling strategies, an improved understanding of virus UV_{254} inactivation mechanisms, and additional high-quality inactivation data published in recent years provide the necessary tools and information to expand upon these initial predictive approaches.

In this study, we develop models to predict rate constants for virus inactivation with UV_{254} treatment in aqueous suspension using variables that are expected to play a role in inactivation, such as genome sequence composition and genome repair information. We conducted a systematic review to gather high quality virus inactivation data from the literature and used the resulting data set to train and validate the predictive performance of four different models (i.e., multiple linear regression, elastic net regularization, boosted trees, and random forests). The models developed in this research will facilitate rapid evaluation of UV_{254} inactivation rate constants for a broad class of virus types based solely on virus genome sequence and genome repair information.

MATERIALS AND METHODS

Systematic Review of UV_{254} Virus Inactivation Data. We conducted a systematic literature review to capture high quality UV_{254} virus inactivation data using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.[30] Data were extracted from studies if they adhered to all of the following criteria: the UV_{254} lamp fluences were measured and reported; sources emitted UV irradiation principally at wavelengths of 253, 253.7, 254, or 255 nm; viruses were irradiated in a liquid suspension; infectious viruses were enumerated with quantitative culture-based approaches (e.g., plaque assay); attenuation through the sample solution was taken into account, or negligible UV_{254} attenuation was reported (transmittance >95%) or could be assumed based on the reported viral stock purification techniques and matrix solution composition; stirring was reported when attenuation was significant (transmittance <95%); first-order kinetics were reported or could be confirmed with reported data points for at least two UV_{254} fluences; the first-order inactivation rate constant or log-removal dose (e.g., D_{99}) was provided or could be determined with data presented in a plot or table. For publications that contained valuable data, but for which not all criteria could be evaluated, corresponding authors were contacted when possible to inquire about the criteria. For studies that reported multiple UV_{254} inactivation experiments for the same virus (e.g., in different solutions, with multiple UV_{254} sources), we combined all data to determine a single inactivation rate constant with linear regression analysis. All data were re-extracted by a second reviewer and discrepancies were addressed. Additional details of our systematic review process are included in the Supporting Information (SI).

Final Data Set Used in Modeling. An inactivation rate constant collected in the systematic review was included in the modeling work if the virus' genome sequence was available through NCBI and if the error associated with the inactivation rate constant was available. Information on NCBI sequence selection is provided in the SI. For viruses with three or more inactivation rate constants obtained from the systematic review, outlier rate constants (i.e., values lying >1.5 times the interquartile range above the third quartile or below the first quartile) were not included in model development. We calculated the inverse variance weighted mean inactivation rate constant for each virus using the following equation

$$\bar{k}_v = \frac{\sum_{i=1}^{n} k_i \cdot w_i}{\sum_{i=1}^{n} w_i} \quad (1)$$

where $\bar{k}_v$ is the inverse variance weighted mean for the virus, n is the number of experimental rate constants for the virus, k_i is the inactivation rate constant for experiment i, and w_i is the weight for experiment i, defined as

$$w_i = \frac{1}{SE_i^2} \quad (2)$$

where $SE_{,i}$ is the standard error of the inactivation rate constant for experiment i. The standard error of the inverse variance weighted mean, SE_v, was evaluated for each virus as

$$SE_v = \sqrt{\frac{1}{\sum_{i=1}^{n} w_i}}. \quad (3)$$

We estimated the interexperimental error for viruses with more than one experimental rate constant in the literature by determining the residual standard deviation from a weighted least-squares regression. Virus was the categorical variable in the regression, and the experimental rate constant was the dependent variable. Weighting was done using the inverse of the squared experimental standard error normalized by the mean rate constant for that virus.

Predictors. For model development, we used predictors related to virus structure and behavior that are known or hypothesized to affect UV_{254} inactivation. The specific predictors incorporated included the structure of nucleic acid strands (i.e., double-stranded or single-stranded), genome length, pyrimidine base content in the genome, sequential pyrimidine bases, genome repair mode, and host cell type. Our reasoning for inclusion of predictors and the methods used to determine values for each predictor are included in the SI. A list of the exact predictors as well as the values used for each virus are available in SI Table S1.

Predictive Model Optimization. We used four model classes, namely, multiple linear regression, elastic net regularization, boosted trees, and random forests to predict virus inactivation during UV_{254} disinfection. For each model class, we developed individual models using only (+) ssRNA viruses and only dsDNA viruses. We also generated a single model developed using all viruses included in the collected data set and thus not separated by virus Baltimore classification groups. We assessed model performance using leave-one-virus-out cross-validation. Further details of model training, validation, and prediction performance evaluation are included in the SI. Data manipulation, statistical analyses, and modeling work were conducted in R software version 4.0.0.[31] The raw data files and the scripts for model development and prediction are available in Github at https://github.com/nrockey/uv-virus-inactivation-prediction.

Multiple Linear Regression. Several of the genomic variables are collinear (e.g., numbers of U and UU). We therefore conducted principal component analysis (PCA) on the genomic variables prior to linear modeling to reduce variable dimensionality and eliminate collinearity. The

https://dx.doi.org/10.1021/acs.est.0c07814
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

predictor nucleic acid type, genome repair mode, and host cell type were not included in the PCA. We then developed linear regression models containing either the first, first and second, or first, second, and third principal components, as well as the other predictors. Only the first through third principal components were assessed for inclusion in the linear regression models, because they cumulatively explained 97% of the variation in genomic variables. Genomic variables were standardized to unit variance prior to PCA to eliminate dissimilarities in the magnitude of variable values. Linear regression can include one or more predictors that can affect model accuracy. We therefore used best subset selection to evaluate a wide range of potential multiple linear regression models.

Elastic Net Regularization. As an alternative to best subset selection, we considered linear regression with parameter regularization using L1 ("Lasso") and L2 ("Ridge") penalties, a technique known as the elastic net. We used the "glmnet" package in R to create models with elastic net regularization. The α and λ hyperparameters, which control the relative contribution and overall scale of the L1 and L2 penalties, respectively, were tuned using a grid search to find the optimal hyperparameters for the data set as determined by leave-one-virus-out cross-validation. Specifically, 11 different values ranging from 0 to 1 with a step of 0.1 were assessed for the hyperparameter α, and 100 different λ values were evaluated for each α.

Random Forests. To accommodate the use of the modified inverse variance weights, the random forests model was developed in R using the "xgboost" package with a single round of boosting, and other hyperparameters were set to match defaults from the "randomForest" package as well as possible.[32]

Boosted Trees. Boosted trees modeling was conducted using the "xgboost" package in R. The number of boosting rounds was selected to minimize the cross-validated error. The hyperparameters for learning rate, tree depth, and minimum terminal node weight were 0.3, 6, and 1, respectively.

Experimental and Predicted UV_{254} Inactivation of Murine Hepatitis Virus (MHV) and Bacteriophage HS2. To consider how well the models may predict inactivation of a virus not already included in the collected data set, we determined the UV_{254} inactivation rate constant of MHV, a virus in the *Coronaviridae* family and *Betacoronavirus* genus, and of HS2, a marine bacteriophage, and compared experimental inactivation to the model's predicted inactivation. Virus propagation and enumeration details are provided in the SI.

UV_{254} Inactivation of Viruses. All UV_{254} inactivation experiments were conducted with a custom-made collimated beam reactor containing 0.16 mW cm^{-2} lamps (model G15T8, Philips). UV_{254} irradiance was determined using chemical actinometry[33,34] and *Escherichia coli* bacteriophage MS2 (ATCC 15597-B1) was included in all experimental solutions as a biodosimeter to further confirm UV_{254} doses. Infective MS2 was assessed using the double agar overlay approach with the host *Escherichia coli* (ATCC 15597).[35] For each UV_{254} exposure, 2 mL of the experimental solution was added to a 10 mL glass beaker, and the solution was continuously stirred. Sample solution depth (0.8 cm) and transmittance (~47% to 53% for MHV experiments, ~79% to 80% for HS2 experiments) were used to determine the average UV_{254} irradiance of the sample according to the Beer−Lambert law.[36] All UV_{254} inactivation experiments were conducted at room temperature (20 to 21 °C). Infectious viruses were assayed immediately following experiments. Dark controls were conducted with each experiment and consisted of the virus suspended in experimental solution but stored in the dark on ice for the duration of experiments. Three independent replicates were conducted for each inactivation experiment.

For MHV experiments, solutions contained MHV and MS2 diluted in 1X PBS to a final concentration of ~10^5 and ~10^{10} pfu/mL, respectively. Samples were exposed to UV_{254} for 0, 5, 15, 25, and 35 s, which corresponded to UV_{254} doses of approximately 0 mJ cm^{-2}, 0.62, 1.2, 1.9, 3.1, and 4.3 mJ cm^{-2}. MS2 infectivity was assayed after larger UV_{254} doses due to its slower inactivation kinetics, namely 37 and 74 mJ cm^{-2}. For HS2 experiments, solutions contained HS2 and MS2 diluted in 1X PBS to a final concentration of ~10^8 and ~10^9 pfu/mL, respectively. Samples were irradiated for 0, 180, 300, 480, 600, and 720 s, which resulted in UV_{254} doses of approximately 0, 26, 44, 70, 88, and 105 mJ cm^{-2}.

The inactivation rate constant, k_{exp} in cm^2 mJ^{-1}, for MHV, HS2, and MS2 was determined by the following equation

$$\ln\left(\frac{C}{C_0}\right) = -k_{exp}\cdot D_{UV254} \quad (4)$$

where C_0 and C are infectious virus concentrations before and after UV_{254} exposure, respectively, in pfu/mL, and D_{UV254} is the average UV_{254} dose in mJ cm^{-2}.

Experimental inactivation rate constants (i.e., k_{exp}) were determined with linear regression analyses conducted in Prism version 8.4.2 (GraphPad) to obtain experimental inactivation rate constants (i.e., k_{exp}). UV_{254} inactivation curves for some viruses exhibited tailing at high doses. In these situations, only the linear portions of the inactivation curves were included in the linear regression analyses.

MHV and HS2 Inactivation Rate Constant Prediction. The UV_{254} inactivation rate constants of MHV and HS2 were predicted using the best-performing inactivation models for (+) ssRNA viruses and dsDNA viruses, respectively. The MHV genome sequence was provided by Dr. Leibowitz (SI Text File S1), and the HS2 genome sequence is available in NCBI (accession no. KF302036).

Predicting UV_{254} Inactivation of Emerging or Difficult-to-Culture Viruses. The inactivation rates of several emerging and difficult-to-culture viruses, including SARS-CoV-2, were predicted using the best-performing inactivation model. Sequence data for these viruses were obtained from NCBI, and all viruses with sequence information are included in SI Table S1.

RESULTS

Numerous UV_{254} Rate Constants Are Available, but Only for a Limited Subset of Viruses. We conducted a systematic review to collect UV_{254} inactivation rate constants and used them for the training and validation of models developed to predict virus inactivation kinetics. Of 2416 initial studies, 531 underwent full text review and 103 studies were included in the final data set (SI Figure S1). Only data from studies passing a set of experimental criteria (SI) were included to ensure collection of high-quality rate constants. These studies produced 224 experimental inactivation rate constants for 59 viruses (Figure 1; SI Table S2). Viruses of different strains and types were considered unique.

https://dx.doi.org/10.1021/acs.est.0c07814
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

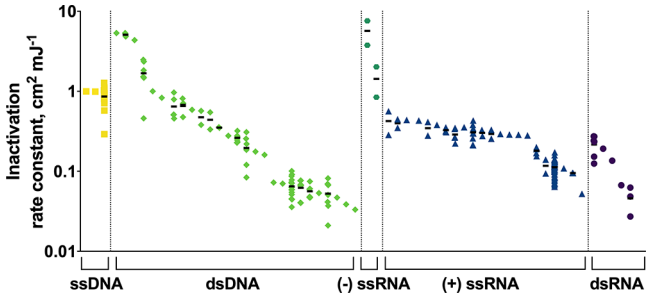


Figure 1. Distribution of UV_{254} inactivation rate constants collected from the systematic literature review. Black bars denote arithmetic means of inactivation rate constants for viruses with more than one experimental rate constant. Outliers are not included. ssDNA viruses: three viruses, 13 rate constants; dsDNA viruses: 26 viruses,* 84 rate constants; (−) ssRNA viruses: two viruses, four rate constants; (+) ssRNA viruses: 22 viruses, 107 rate constants (four outlier rate constants removed); dsRNA viruses: five viruses, 12 rate constants. Viruses within each Baltimore classification are ordered from highest to lowest mean rate constant from left to right. Rate constants are reported in SI Table S2. *Considers two viruses (i.e., adenovirus 5 and adenovirus 41) assayed in host cells with reduced repair abilities as different from the same viruses assayed in wild-type host cells.

More than 350 studies from the full text review that reported conducting UV virus inactivation in aqueous suspension were not included in the final data set. Data were excluded most commonly because the article did not address UV_{254} attenuation in the experimental solution and it could not be ruled out based on details in the materials and methods. Nearly 50% of the extracted rate constants represented only five different viruses. For example, there were 62 different experimental inactivation rates for bacteriophage MS2; in contrast, several viruses, including hepatitis E virus, only had one reported inactivation rate constant, and there were many human viruses with no data that met the review criteria (e.g., influenza viruses, ebolaviruses, coronaviruses, herpesviruses). Ultimately 13, 84, 111, 4, and 12 experimental inactivation rate constants were extracted for ssDNA, dsDNA, (+) ssRNA, (−) ssRNA, and dsRNA viruses, respectively, representing 3, 26, 22, 2, and 5 unique viruses (Figure 1). No rate constants met the inclusion criteria for retroviral (+) ssRNA viruses, referred to as RT-ssRNA viruses. The inactivation rate constants spanned ∼2.5 orders of magnitude (Figure 1) and ranged from 0.021 to 7.6 cm^2 mJ^{-1}. The (−) ssRNA viruses had the largest rate constants on average (k = 3.6 cm^2 mJ^{-1}), while dsRNA viruses had the lowest average rate constants (k = 0.15 cm^2 mJ^{-1}). dsDNA virus constants exhibited the widest range of rate constants, spanning from 0.021 to 5.4 cm^2 mJ^{-1} with a mean of 0.55 cm^2 mJ^{-1}. The rate constants collected were associated with the linear portion of the UV_{254} virus inactivation curve and did not incorporate regions of the curve where tailing occurred. Overall, first-order kinetics were observed up to at least 4-log_{10} virus inactivation. This suggests that our models are applicable up to approximately 4-log_{10} virus inactivation. Beyond that point, our models could overestimate inactivation levels for viruses that exhibit tailing effects during inactivation.

Individual models were developed for the (+) ssRNA and dsDNA virus classes. The limited data sets for viruses in the other Baltimore classifications made it infeasible to develop individual predictive models for the other groups. The data sets used for (+) ssRNA and dsDNA model training and validation included 19 (+) ssRNA viruses with 93 experimental inactivation rate constants and 16 dsDNA viruses with 50 inactivation rate constants, respectively (SI Table S2). The model developed with all viruses from the systematic review included 43 viruses with 168 experimental inactivation rate constants.

Rate Constants Predicted Using Common Modeling Approaches. We used the data collected in the systematic literature review to develop linear regression, elastic net regularization, random forests, and boosted trees models for predicting inactivation rate constants based on several predictors (SI Table S1). These model classes were selected to cover a range of different linear and nonlinear approaches that are commonly applied in the predictive modeling field.[37]

(+) ssRNA Virus Model. The cross-validated root mean squared relative prediction errors (RMSrPEs) for the four optimized models varied from 0.22 to 0.95 (Figure 2 and SI Table S3), with the top performing multiple linear regression resulting in the lowest RMSrPE out of the four optimized

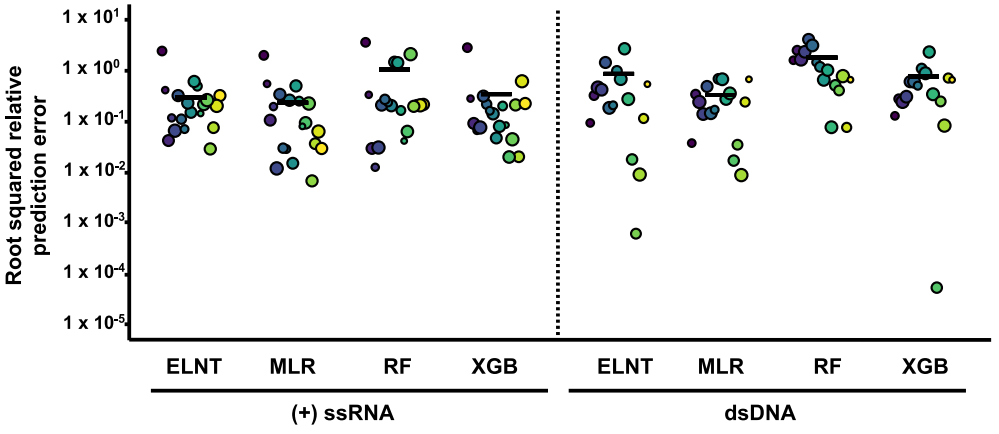


Figure 2. Root squared relative prediction error of virus inactivation rate constants using top performing models from each model class developed with only (+) ssRNA viruses (left) or dsDNA viruses (right) in the training and validation set. Individual symbols indicate the root squared relative prediction error of each virus, and the black bar indicates the model's root mean squared relative prediction error. Distinct colors represent different viruses, and the symbol sizes represent the weight of the experimental inactivation rate constant used for inverse variance weighting, where a larger symbol indicates a greater weight. MLR = multiple linear regression, ELNT = elastic net regularization, XGB = boosted trees, RF = random forests.

https://dx.doi.org/10.1021/acs.est.0c07814
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

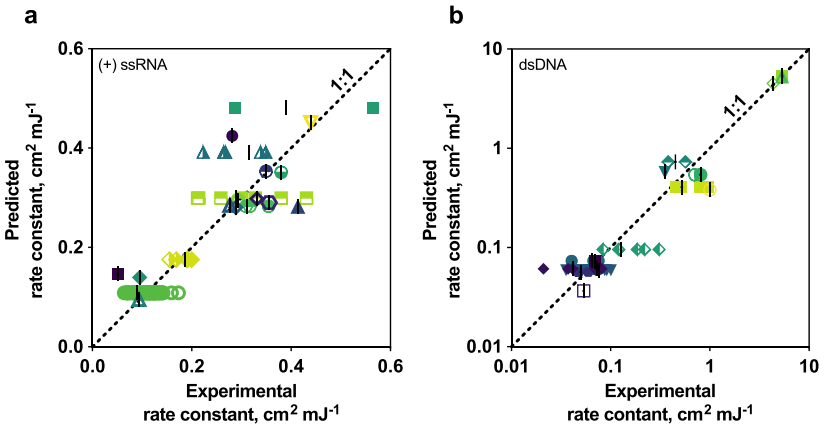


Figure 3. Experimental and predicted cross-validated inactivation rate constants for (+) ssRNA viruses (a) and dsDNA viruses (b) present in the training and validation set. Different colors and symbols represent different viruses. Black lines represent the estimated experimental rate constant for each virus. Data included in the models were obtained from the literature with a systematic review, and all predicted and experimental inactivation rate constants are provided in SI Tables S1 and S6.

model classes. Various subsets of genomic variables were included in multiple linear regression development. Because these genomic variables are highly collinear, we used principal components that incorporated various genomic variable subsets as predictors in the regression models. Ultimately, the multiple linear regression model with one principal component that incorporated the numbers of cytosines (Cs), uracils (Us), uracil doublets (UUs), and uracil triplets (UUUs) resulted in the lowest RMSrPE (0.22 ± 0.23; RMSrPE ± standard error; SI Table S3). Other multiple linear regressions performed similarly (SI Table S4). The optimized elastic net regularization and boosted trees models resulted in slightly higher RMSrPEs than the top performing multiple linear regression model ($RMSrPE_{elastic\ net}$ = 0.28 ± 0.26, $RMSrPE_{boosted\ trees}$ = 0.32 ± 0.28; SI Table S3), and the random forests model had the largest RMSrPE of the (+) ssRNA virus models ($RMSrPE_{random\ forests}$ = 0.95 ± 0.48; SI Table S3). Model performance was significantly reduced in the elastic net and random forests models as compared to the multiple linear regression model (SI Table S5).

Predicted (+) ssRNA virus rate constants from the top performing model were within 51% of the mean experimental virus inactivation rate constants obtained from the systematic review, with the exception of the rate constant for Atlantic Halibut Nodavirus (percent error = 182%; SI Figure S2a). The RMSrPE from the top performing linear regression model was lower than the estimated relative interexperimental error of viruses with multiple rate constants in the literature (RMSrPE = 0.22 ± 0.23; relative interexperimental error = 0.33; Figure 3a). In other words, the predicted rate constants for new (+) ssRNA viruses would be at least as accurate as the rate constants determined through experimental studies.

dsDNA Virus Model. The genomic variables used in dsDNA model development were equivalent to the (+) ssRNA models, with the exception that thymines (Ts) were substituted for Us (SI Table S1). A major distinction of dsDNA viruses is that their genomes can undergo repair in host cells and this impacts their susceptibility to UV_{254}.[24,38−40] Genome repair can be mediated by the host cell or by viral genes,[24] and the varied efficacy of host-mediated dsDNA repair[41−44] impacts virus UV_{254} sensitivity. We included categorical predictors for genome repair mode (i.e., host cell mediated, virus-gene controlled using one repair system, or virus-gene controlled using multiple repair systems) and host cell type (i.e., prokaryotic host, eukaryotic host with wild type repair, or eukaryotic host with reduced repair) in the dsDNA virus inactivation rate constant models. Genome repair mode and host cell type were assigned based on available information and are described in the SI.

The RMSrPE of the four optimized dsDNA model classes ranged from 0.31 to 1.6 (SI Table S3), and the optimized multiple linear regression model outperformed the three other optimized model classes (RMSrPE = 0.31 ± 0.28; Figure 2 and SI Table S3). The optimized elastic net and boosted trees RMSrPEs were slightly higher ($RMSrPE_{elastic\ net}$ = 0.79 ± 0.46, $RMSrPE_{boosted\ trees}$ = 0.70 ± 0.43), though the difference in model performance was not significant (SI Table S5), and the random forests model performed significantly worse ($RMSrPE_{random\ forests}$ = 1.6 ± 0.66). The top linear regression model included the genome repair mode and host cell type predictors, as well as one principal component comprised of thymine doublets (TT), thymine quintuplets (TTTTT), and Cs. As with the top-performing (+) ssRNA model, many of the regressions tested with different genomic variable subsets had similar prediction performance, making it difficult to identify which genomic variables were critical for predicting dsDNA virus rate constants (SI Table S4). A point estimate comparison of the regression coefficients for the standardized principal component (β_{PC1} = 0.46), genome repair mode ($\beta_{genome\ repair\ mode}$ = 2.7), and host cell type ($\beta_{host\ cell\ type}$ = −0.37) predictors indicates that the genome repair mode predictor is approximately 5.9 times more important than the principal component predictor ($\beta_{genome\ repair\ mode}/\beta_{PC1}$ = 2.7/0.46). Host cell type was comparable in importance to the genomic variable contribution, collectively represented by the principal component. Prediction performance dropped significantly without genome repair mode as a predictor ($RMSrPE_{opt}$ = 0.31 ± 0.28, $RMSrPE_{no\ repair}$ = 1.0 ± 0.52; SI Table S5), further highlighting the importance of genome repair in UV_{254} inactivation.

The multiple linear regression model accurately predicted inactivation rate constants across the wide range of dsDNA virus susceptibilities to UV_{254} (Figure 3b). As with the top performing (+) ssRNA model, the predicted error for the top performing dsDNA model was lower than the estimated interexperimental error for viruses with more than one

https://dx.doi.org/10.1021/acs.est.0c07814
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

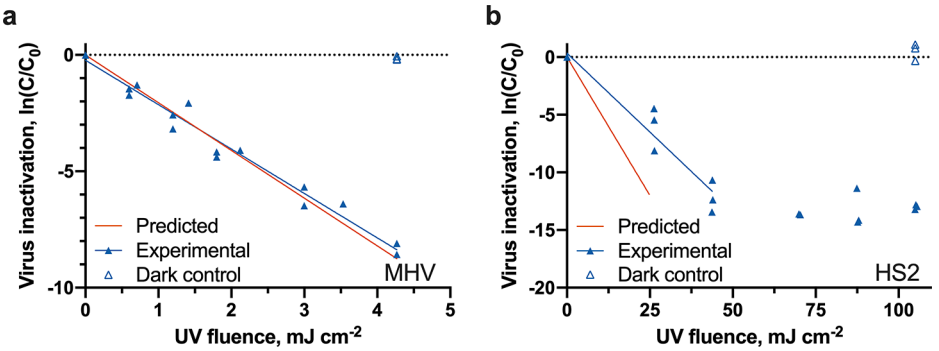


Figure 4. Experimental and predicted UV_{254} inactivation of MHV A59 (a) and HS2 bacteriophage (b). All independent replicates (N = 3) from experiments are shown as individual points. The experimental HS2 inactivation rate constant was determined using the first two UV_{254} fluences due to significant tailing beyond UV_{254} fluences of 50 mJ cm^{-2}.

experimental rate constant (RMSrPE = 0.31 ± 0.28; interexperimental error of k_{virus} = 0.45). Predictions were poorest for T7M, B40−8, and λ (percent error = 62%, 63%, and 62%, respectively; SI Figure S2b), which are bacteriophages with the same form of genome repair mode. The poor prediction of viruses from this group indicates that some of the rate constants in the training data for viruses with these attributes may be inaccurate, leading to worse performance for bacteriophages with host mediated repair.

All-Virus Model. Larger data sets generally add predictive power to models, though the increased signal from additional data can be attenuated or negated by increased heterogeneity. We therefore compared the performance of the separate (+) ssRNA and dsDNA virus models with a model that incorporated data from all Baltimore classes. In addition to the genomic variables and repair-related predictors (i.e., genome repair mode and host cell type) included for (+) ssRNA and dsDNA viruses, a categorical predictor for nucleic acid type (i.e., double-stranded or single-stranded) was included. Boosted trees models were the top performing models using all viruses (SI Table S3); these performed significantly worse than the models trained using only (+) ssRNA viruses ($RMSrPE_{(+)\ ssRNA}$ = 0.22 ± 0.23, $RMSrPE_{all}$ = 0.45 ± 0.33; SI Table S5) or only dsDNA viruses ($RMSrPE_{dsDNA}$ = 0.31 ± 0.28 vs $RMSrPE_{all}$ = 0.45 ± 0.35; SI Tables S3 and S5). This suggests that using our modeling approach and combining viruses with diverse genome types and infection cycles into one model can negatively impact performance of virus predictions, possibly owing to insufficient data from less studied classes. On the basis of these results, we used the separate (+) ssRNA and dsDNA models for subsequent analyses.

Predicted Rate Constants Align with New Experimental Rate Constants. We applied the optimized (+) ssRNA and dsDNA models to predict the rate constants of one (+) ssRNA virus and one dsDNA virus for which experimental data were not available and then measured the rate constants experimentally. Specifically, we predicted and measured the rate constants for MHV, a (+) ssRNA mouse coronavirus, and HS2, a dsDNA marine bacteriophage. On the basis of its large genome size (i.e., ∼270% longer than the largest (+) ssRNA virus genome included in the training and validation set) MHV provided an opportunity to assess the (+) ssRNA model's predictive power using a virus with attributes outside those in the training and validation set (SI Figure S3). HS2 bacteriophage has similar genomic attributes to many of the other viruses in the data set (SI Figure S3), and genome repair-related predictors are the same as those for most of the phages. Bacteriophage MS2 was included in each experimental solution to confirm UV_{254} doses; the measured MS2 rate constants were in line with those in the literature (0.12 to 0.14 cm^2 mJ^{-1}; SI Figure S4 and SI Table S2).

The predicted inactivation rate constant for MHV (k_{pred} = 2.05 ± 0.88 cm^2 mJ^{-1}; mean ± 95% margin of error) was not significantly different than the experimental rate constant (k_{exp} = 1.92 ± 0.17 cm^{-2} mJ^{-1}), with a percent error of only 7% (Figure 4a). The prediction accuracy the model achieved despite MHV's elevated UV_{254} sensitivity compared with other (+) ssRNA viruses in the data set highlights how linear regression approaches are capable of extrapolating predictions to values distinct from those used in training and validation. In comparison, the MHV inactivation rate constant predicted with the top performing nonlinear approach, boosted trees, was 79% different than the experimental value, with a rate constant of 0.40 ± 0.25 cm^2 mJ^{-1}. The accuracy of the MHV rate constant prediction and the relatively low RMSrPE obtained for the top performing (+) ssRNA virus model provide some confidence that the (+) ssRNA model can effectively predict UV_{254} rate constants for emerging or difficult-to-culture (+) ssRNA viruses. Additional out-of-sample validation will be needed, however, to better understand how well the models generalize to new viruses.

The experimental HS2 inactivation kinetics exhibited significant tailing beyond UV_{254} fluences of 50 mJ cm^{-2}; we therefore modeled the first ∼5-log_{10} of inactivation to obtain a rate constant from the first-order portion of the curve. The resulting dsDNA HS2 bacteriophage experimental rate constant of k_{exp} = 0.28 ± 0.08 cm^2 mJ^{-1} was 71% lower than the predicted rate constant of k_{pred} = 0.48 ± 0.29 cm^2 mJ^{-1} (Figure 4b). Although the error of this dsDNA estimate was larger than that of the (+) ssRNA estimate, the HS2 predicted and experimental constants are not significantly different. This result, in combination with the cross-validation results, suggest that the dsDNA model can effectively predict if a dsDNA virus is particularly resistant to UV_{254} treatment.

Predictive Models Estimate Inactivation of Several Emerging and Difficult-to-Culture Viruses. Our systematic review identified a number of important human viruses that lack published high quality UV_{254} inactivation rate constants in the literature. We therefore applied the (+) ssRNA and dsDNA predictive models to estimate the inactivation rates constants for several viruses, including human norovirus, dengue virus, SARS-CoV-2, and several herpesviruses (Table 1). These predictions resulted in a range

https://dx.doi.org/10.1021/acs.est.0c07814
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Table 1. Predicted UV_{254} Inactivation Rate Constants for Several Viruses without High-Quality Experimental Inactivation Rate Constants

Virus	NCBI accession number	Predicted inactivation rate constant, k (cm^2 mJ^{-1})[a]
(+) ssRNA viruses		
SARS-CoV-1	NC_004718	1.9 ± 0.82
SARS-CoV-2	MN908947	2.0 ± 0.86
Middle eastern respiratory syndrome coronavirus (MERS-CoV)	JX869059	2.1 ± 0.91
Dengue virus	NC_001477	0.38 ± 0.16
Zika virus	NC_035889	0.39 ± 0.17
Human rhinovirus (B14)	K02121	0.34 ± 0.15
Human norovirus (GII.4 Sydney)	JX459908	0.28 ± 0.12
dsDNA viruses		
Herpes simplex virus 1 (strain 17)	NC_001806	1.8 ± 1.1
Epstein-Barr virus	NC_007605	1.9 ± 1.2
Human cytomegalovirus	NC_006273	3.0 ± 1.8
Variola virus (major)	L22579	2.5 ± 1.5

[a]Error shown represents the 95% margin of error of predicted rate constant, as determined by the model's 95% margin of error, estimated as 1.96 times the standard error, where standard error = RMSrPE x virus rate constant.

of inactivation rate constants, from 0.28 for human norovirus to 3.0 cm^2 mJ^{-1} for human cytomegalovirus. Although these virus rate constants have not been validated with experiments, the performance of our models gives us confidence that the predicted values are good estimates of the actual inactivation rate constants.

DISCUSSION

Through the evaluation of a large set of models from four distinct model classes developed with the best currently available data, we identified effective models for predicting UV_{254} inactivation rate constants of (+) ssRNA and dsDNA viruses using simple virus attributes as model predictors. UV_{254} primarily targets viral nucleic acid during irradiation. Pyrimidine bases are more photoreactive than purines,[45] and pyrimidine dimers, in particular, cause a large portion of the UV-induced damage to DNA.[45−51] Limited research centered on ssRNA photolysis suggests pyrimidine hydrates are a common form of UV damage.[52] Photochemical damage to nucleic acids can stall or inhibit enzymes required for productive viral infection of host cells.[53−55] On the basis of this *a priori* knowledge, we included several combinations of pyrimidine bases as predictors in our (+) ssRNA and dsDNA models, namely, the numbers of U, UU, UUU, UUUU, UUUUU, C, UC, and CU in (+) ssRNA models and the numbers of T, TT, TTT, TTTT, TTTTT, C, TC, and CT in dsDNA models.

Ultimately, the top performing (+) ssRNA virus model employed one principal component incorporating multiple genomic variables (i.e., numbers of C, U, UU, and UUU), and the top performing dsDNA virus model employed repair mode, host cell type, and one principal component representing three genomic variables (i.e., numbers of C, TT, TTTTT). The relative importance of variables in our top performing predictive models may provide insight into the mechanisms driving UV_{254} inactivation of viruses. Among the (+) ssRNA models, many of the multiple linear regression models that included distinct subsets of genomic variables performed similarly. This is likely because these genomic variables are so highly correlated that different variable combinations resulted in a similar set of principal components as predictors in modeling. This ultimately yielded similar performance among different models. Separating the effects of individual genomic variables was therefore difficult in the (+) ssRNA model. Although the top performing model incorporated multiple genomic variables, several linear regression models using as few as one genomic variable as a predictor resulted in similar model performance. This finding demonstrates that simple aspects of the (+) ssRNA genome provide all the necessary information to accurately predict rate constants for this class of viruses.

In the dsDNA model, performance was significantly improved when genome repair predictors were included in addition to principal components incorporating genomic variables. The importance of genome repair was expected. For example, the two dsDNA bacteriophages T2 and T4 have similar genome sizes and composition (SI Figure S3b and SI Table S1) but dissimilar UV_{254} inactivation rate constants (5.1 cm^{-2} mJ^{-1} for T2 and 1.7 cm^{-2} mJ^{-1} for T4; SI Table S2). T4 phage's UV_{254} resistance is due to an additional virus-controlled repair gene in the T4 genome not present in the T2 genome.[56,57] Interestingly, the relative contribution of genomic variables in the dsDNA model was significantly less than the genome repair predictors, which suggests that genome repair is a more important factor in dsDNA UV_{254} inactivation than genomic variables.

The inclusion of genome repair as a model predictor presented some limitations. First, the mode and extent of genome repair is not known for many viruses and has not been well-studied across virus families. A single predictor encompassing the contribution of genome repair was therefore not possible. We instead applied multiple categorical predictors. With this approach, only viruses that shared a particular genome repair mode or host cell type with at least one other virus in the dsDNA data set could be used in cross-validation. Ultimately, the data set used for dsDNA model

https://dx.doi.org/10.1021/acs.est.0c07814
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

development and validation lacked numerous forms of dsDNA viruses with distinct repair modes and host cell types. This resulted in uncertainty in model performance for certain dsDNA viruses not represented in the training and validation set. To improve future dsDNA virus models, it is critical to have a better understanding of genome repair mechanisms and how they affect UV_{254} inactivation.

Our top performing UV_{254} virus prediction models provide improvements over earlier prediction approaches.[28,29] On average, the (+) ssRNA and dsDNA virus models predicted rate constants to within ∼0.2x and ∼0.3x of experimental constants, respectively. A previous approach using genome length to determine genome size-normalized sensitivity values for a number of virus families expected uncertainties in predicted values of ∼2x.[28] A more recent approach developed predictive models for ssRNA and dsDNA UV_{254} inactivation using genome dimer formation potential, a value that incorporated pyrimidine doublets, genome length, and purines with adjacent pyrimidine doublets.[29] Their reported error as a coefficient of determination (i.e., R^2) was 0.67 for ssRNA viruses compared to 0.74 (adjusted R^2) for our model, and an R^2 value of 0.62 for dsDNA viruses compared to 0.99 (adjusted R^2) for our model. Several factors can be attributed to the improved performance of our models, including extensive curation of data based on quality and the incorporation of genome repair into dsDNA modeling.

In light of the coronavirus disease 2019 (COVID-19) pandemic and the need for effective decontamination strategies, our predictive models provided an opportunity to predict rate constants for a critical group of viruses with very little published inactivation data. Limited data on UV_{254} inactivation for coronaviruses in aqueous suspension are available and the published information did not pass the inclusion criteria of our systematic review.[10,58−60] This paucity of information on the susceptibility of coronaviruses to UV_{254} is of critical importance for developing effective decontamination strategies. Our predicted rate constants for SARS-CoV-1, SARS-CoV-2, and MERS, and our measured rate constant for the mouse coronavirus MHV, suggest that coronaviruses are much more susceptible to UV_{254} inactivation than other (+) ssRNA viruses. A recent estimate of SARS-CoV-2 UV_{254} susceptibility using the previously developed Lytle and Sagripanti approach[28] is ∼1.7x greater than our estimate.[61] Discrepancies in new experimental coronavirus data still persist, likely stemming from a lack of checks on UV_{254} attenuation of suspensions.

More robust models are possible with larger data sets that consist of more diverse viruses. Unfortunately, a large portion of UV_{254} inactivation data found during the systematic review did not pass our inclusion criteria. The most common reason for excluding data from our systematic review was a failure to report solution UV_{254} attenuation. An earlier study of SARS-CoV-1 inactivation by UV_{254},[60] for example, did not account for UV_{254} attenuation in the experimental DMEM suspension. The reported inactivation rate constant of 0.003 cm^2 mJ^{-1} was nearly 3 orders of magnitude lower than our predicted rate constant for SARS-CoV-1 and our measured value for MHV, likely in part due to solution attenuation. We estimate that their rate constant would be closer to 0.35 cm^2 mJ^{-1} after accounting for solution attenuation. This value more closely aligns with our coronavirus values. Similarly, several studies reported UV_{254} inactivation of viruses in blood products without describing how attenuation was considered in their reported doses.[10,62−64] Although these doses are likely representative for these fluids, they cannot be extrapolated to other matrices. More stringent reporting of UV_{254} experimental conditions,[65] including matrix solution transmission at 254 nm, will facilitate future modeling efforts. Our models predict UV_{254} inactivation rate constants for solutions with 100% transmittance (e.g., purified virus in buffer solution). These rate constants can be adjusted to predict virus inactivation in a solution with significant attenuation using the Beer−Lambert law, which takes into account sample absorbance.[36]

The developed models allow us to predict the effectiveness of current UV_{254} treatment strategies on viral pathogens that are difficult or impossible to culture. For example, human norovirus, which causes gastrointestinal disease, is a major target of UV_{254} disinfection processes in water treatment and food processing. Our (+) ssRNA virus model predicts an inactivation rate constant of 0.28 cm^2 mJ^{-1} for human norovirus GII.4. This is similar to our recently reported rate constant of k = 0.27 cm^2 mJ^{-1} for human norovirus GII.4 Sydney using RT-qPCR data coupled with a full-genome extrapolation approach.[66] This finding indicates that current UV_{254} water treatment guidelines that are defined to treat adenovirus 41[67] are more than sufficient to inactivate human norovirus to acceptable levels. In fact, none of the viruses for which we predicted rate constants had UV_{254} resistance greater than viruses in the *Adenoviridae* family.

The limited and unbalanced data set that we obtained from the systematic review and used in modeling efforts created challenges in our modeling work. The primary concern was that we could not take a commonly used approach for evaluating model performance in which a portion of data is held back during model development. A holding back of the typical 10−20% of data would correspond to the holding back of only two to four viruses from the (+) ssRNA or dsDNA classes for testing. This could result in high variance estimates of prediction performance that would also be highly dependent on the viruses withheld during training. We consequently used leave-one-virus-out cross-validation to more efficiently estimate prediction performance on out of sample data. Another limitation of our models is that they were developed and validated for only (+) ssRNA and dsDNA viruses. Although many human viruses are in these two classes, many emerging and noteworthy human viruses belong to other classes. In particular, the (−) ssRNA virus class includes several important human pathogens, such as lassa virus, nipah virus, influenza virus, and ebola virus. Since only two (−) ssRNA viruses were included in our data set, we were unable to assess whether inactivation rate constants for viruses in this group could be accurately predicted with our (+) ssRNA model. More high quality UV_{254} experimental inactivation data for a broader set of viruses would facilitate the holdout approach for validating models and the development of models for other virus Baltimore classification groups.

Additional experimental data could also support an expanded set of predictors beyond the primary genome structure and genome repair parameters included here. Virus attributes, like the secondary structure of single-stranded nucleic acids[22] or nucleic acid interactions with viral proteins,[68] may play a role in virus inactivation by UV_{254}. These structural virus characteristics are not as readily available as genome sequence information and were therefore not considered in the present study. Future research could incorporate these attributes as additional model parameters when more data

https://dx.doi.org/10.1021/acs.est.0c07814
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

become available. Another focus of future work could be the expansion of these models to predict the tailing of virus inactivation that is often observed during UV_{254} treatment. Our models focus on the first order portion of the UV_{254} inactivation curve; with an understanding of the mechanisms that underlie tailing kinetics and including the appropriate predictors, this model could be updated to predict both the first order and tailing regions of the UV_{254} virus inactivation curve.

This research demonstrates the value of predictive models for estimating virus fate in various settings. Using readily available viral genome data, we developed models to predict UV_{254} inactivation of (+) ssRNA and dsDNA viruses. The benefits of predictive models are underlined by the ongoing COVID-19 pandemic: access to the biosafety level 3 laboratories required to work with SARS-CoV-2 has been limited and, as a result, few experimental inactivation studies have been performed. Our approach can rapidly determine virus susceptibility to UV_{254} using available genomes but without reliance on culture systems that are often unavailable or difficult to access. Other potential applications of our models including identification of outlier UV_{254} data that are published and prediction of potential worst-case scenarios for viruses and their susceptibility to UV_{254}. Ultimately, we expect that this predictive modeling approach can be applied to estimate inactivation of microorganisms with other disinfectants and in different settings, such as on surfaces or in air.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.0c07814.

Details about the systematic literature review conducted to collect UV_{254} virus inactivation rate constants, information about predictors included in modeling work, and further details about model training, validation and prediction performance evaluation (PDF)

Inactivation rate constants collected for viruses during the systematic review and predicted rate constants from the top-performing models for viruses included in modeling work (XLSX)

Predictor information for all viruses included in predictive modeling development, validation, and predictions (XLSX)

Model performance metrics obtained during cross-validation (XLSX)

Results of pairwise model comparison (XLSX)

Percent errors for viruses included in predictive models; principal component analyses of genome composition for (+) ssRNA and dsDNA viruses used in model development, validation, and predictions (XLSX)

MS2 inactivation during MHV and HS2 experiments (XLSX)

Text file including the genome sequence for the MHV A59 strain used in this study (TXT)

■ AUTHOR INFORMATION

Corresponding Author

Krista R. Wigginton − *Department of Civil & Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, United States;* orcid.org/0000-0001-6665-5112; Phone: +1 (734) 763-2125; Email: kwigg@umich.edu; Fax: +1 (734) 764-4292

Authors

Nicole C. Rockey − *Department of Civil & Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, United States;* orcid.org/0000-0003-1966-5312

James B. Henderson − *Consulting for Statistics, Computing and Analytics Research, University of Michigan, Ann Arbor, MI 48109, United States*

Kaitlyn Chin − *Department of Civil & Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, United States*

Lutgarde Raskin − *Department of Civil & Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.0c07814

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Water Research Foundation (WRRF #15-07) and by National Science Foundation projects #1545756 and #2023057. Nicole C. Rockey was supported by a Rackham Predoctoral Fellowship and a National Science Foundation Graduate Research Fellowship (award no. 2015205675).

■ REFERENCES

(1) Flint, S. J.; Racaniello, V. R.; Rall, G.; Skalka, A. M.; Enquist, L. W. Foundations. In *Principles of Virology, Vol. 1: Molecular Biology*; ASM Press: Washington, D.C., 2015; pp 2−23.

(2) Hirneisen, K. A.; Black, E. P.; Cascarino, J. L.; Fino, V. R.; Hoover, D. G.; Kniel, K. E. Viral Inactivation in Foods: A Review of Traditional and Novel Food-Processing Technologies. *Compr. Rev. Food Sci. Food Saf.* **2010**, *9* (1), 3−20.

(3) Rutala, W. A.; Weber, D. J.; Healthcare Infection Control Practices Adviosry Committee. *Guideline for Disinfection and Sterilization in Healthcare Facilities*; **2018**.58

(4) Adhikari, A.; Clark, S. Disinfection of Microbial Aerosols. In *Modeling the Transmission and Prevention of Infectious Disease*; Hurst, C. J., Ed.; Springer International Publishing: Cham, 2017; pp 55−71. DOI: 10.1007/978-3-319-60616-3_3.

(5) Benjamin, M. M.; Lawler, D. F. *Water Quality Engineering*; Wiley: Hoboken, N.J, 2013.

(6) Metcalf; Eddy; Tchobanoglous, G.; Stensel, H. D.; Tsuchihashi, R.; Burton, F. L.; Abu-Orf, M.; Bowden, G.; Pfrang, W. *Wastewater Engineering : Treatment and Resource Recovery*, 5th ed.; McGraw-Hill Education: New York, 2014.

(7) Gunter-Ward, D. M.; Patras, A.; Bhullar, M. S.; Kilonzo-Nthenge, A.; Pokharel, B.; Sasges, M. Efficacy of Ultraviolet (UV-C) Light in Reducing Foodborne Pathogens and Model Viruses in Skim Milk. *J. Food Process. Preserv.* **2018**, *42* (2), e13485.

(8) Pavia, M.; Simpser, E.; Becker, M.; Mainquist, W. K.; Velez, K. A. The Effect of Ultraviolet-C Technology on Viral Infection Incidence in a Pediatric Long-Term Care Facility. *Am. J. Infect. Control* **2018**, *46* (6), 720−722.

(9) Ward, D. M.; Patras, A.; Kilonzo-Nthenge, A.; Yannam, S. K.; Pan, C.; Xiao, H.; Sasges, M. UV-C Treatment on the Safety of Skim Milk: Effect on Microbial Inactivation and Cytotoxicity Evaluation. *J. Food Process Eng.* **2019**, *42* (4), e12944.

(10) Eickmann, M.; Gravemann, U.; Handke, W.; Tolksdorf, F.; Reichenberg, S.; Müller, T. H.; Seltsam, A. Inactivation of Ebola Virus and Middle East Respiratory Syndrome Coronavirus in Platelet

Concentrates and Plasma by Ultraviolet C Light and Methylene Blue plus Visible Light, Respectively. *Transfusion* **2018**, *58* (9), 2202−2207.
(11) Hijnen, W. A. M.; Beerendonk, E. F.; Medema, G. J. Inactivation Credit of UV Radiation for Viruses, Bacteria and Protozoan (Oo)Cysts in Water: A Review. *Water Res.* **2006**, *40* (1), 3−22.
(12) Shimizu, A.; Shimizu, N.; Tanaka, A.; Jinno-Oue, A.; Roy, B. B.; Shinagawa, M.; Ishikawa, O.; Hoshino, H. Human T-Cell Leukaemia Virus Type I Is Highly Sensitive to UV-C Light. *J. Gen. Virol.* **2004**, *85* (8), 2397−2406.
(13) Jacangelo, J. G.; Loughran, P.; Petrik, B.; Simpson, D.; McIlroy, C. Removal of Enteric Viruses and Selected Microbial Indicators by UV Irradiation of Secondary Effluent. *Water Sci. Technol.* **2003**, *47* (9), 193−198.
(14) Linden, K. G.; Thurston, J.; Schaefer, R.; Malley, J. P. Enhanced UV Inactivation of Adenoviruses under Polychromatic UV Lamps. *Appl. Environ. Microbiol.* **2007**, *73* (23), 7571−7574.
(15) Guo, H.; Chu, X.; Hu, J. Effect of Host Cells on Low- and Medium-Pressure UV Inactivation of Adenoviruses. *Appl. Environ. Microbiol.* **2010**, *76* (21), 7068−7075.
(16) Thurston-Enriquez, J. A.; Haas, C. N.; Jacangelo, J.; Riley, K.; Gerba, C. P. Inactivation of Feline Calicivirus and Adenovirus Type 40 by UV Radiation. *Appl. Environ. Microbiol.* **2003**, *69* (1), 577−582.
(17) Meng, Q. S.; Gerba, C. P. Comparative Inactivation of Enteric Adenoviruses, Poliovirus and Coliphages by Ultraviolet Irradiation. *Water Res.* **1996**, *30* (11), 2665−2668.
(18) Malley, J.; Linden, K.; Mofidi, A.; Bolton, J.; Crozes, G.; Cushing, B.; Mackey, E.; Laine, J. M.; Janex, M.-L. *Inactivation of Pathogens with Innovative UV Technologies*; American Water Works Association Research Foundation: 2004; Vol. *1*.
(19) Luria, S. E.; Dulbecco, R. Genetic Recombinations Leading to Production of Active Bacteriophage from Ultraviolet Inactivated Bacteriophage Particles. *J. Bacteriol.* **1949**, *34* (2), 93−125.
(20) Qiao, Z.; Ye, Y.; Chang, P. H.; Thirunarayanan, D.; Wigginton, K. R. Nucleic Acid Photolysis by UV254 and the Impact of Virus Encapsidation. *Environ. Sci. Technol.* **2018**, *52*, 10408.
(21) Smith, K. C.; Hanawalt, P. C. Photochemistry of the Nucleic Acids. In *Molecular Photobiology: Inactivation and Recovery*; Horecker, B., Kaplan, N. O., Marmur, J., Eds.; Academic Press: 1969; pp 57−84. DOI: 10.1016/B978-1-4831-9742-5.50009-2.
(22) Pearson, M.; Johns, H. E. Suppression of Hydrate and Dimer Formation in Ultraviolet-Irradiated Poly (A + U) Relative to Poly U. *J. Mol. Biol.* **1966**, *20* (2), 215−229.
(23) Henderson, E. E.; Tudor, G.; Yang, J.-Y. Inactivation of the Human Immunodeficiency Virus Type 1 (HIV-1) by Ultraviolet and X Irradiation. *Radiat. Res.* **1992**, *131* (2), 169−176.
(24) Harm, W. *Biological Effects of Ultraviolet Radiation*; IUPAB biophysics series. 1; Cambridge University Press: Cambridge, Eng.; New York :, 1980.
(25) Harm, W. Gene-Controlled Reactivation of Ultraviolet-Inactivated Bacteriophage. *J. Cell. Comp. Physiol.* **1961**, *58* (S1), 69−77.
(26) DayIII, R. S. Cellular Reactivation of Ultraviolet-Irradiated Human Adenovirus 2 in Normal and Xeroderma Pigmentosum Fibroblasts. *Photochem. Photobiol.* **1974**, *19* (1), 9−13.
(27) Baltimore, D. Expression of Animal Virus Genomes. *Bacteriol. Rev.* **1971**, *35* (3), 235−241.
(28) Lytle, C. D.; Sagripanti, J.-L. Predicted Inactivation of Viruses of Relevance to Biodefense by Solar Radiation. *J. Virol.* **2005**, *79* (22), 14244−14252.
(29) Kowalski, W. J.; Bahnfleth, W. P.; Hernandez, M. T. A Genomic Model for Predicting the Ultraviolet Susceptibility of Viruses. *IUVA News* **2009**, *11* (2), 15−28.
(30) Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D. G.; Group, T. P. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Med.* **2009**, *6* (7), e1000097.
(31) R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria 2020.
(32) xgboost developers. Random Forests in XGBoost https://xgboost.readthedocs.io/en/latest/tutorials/rf.html.
(33) Rahn, R. O.; Bolton, J.; Stefan, M. I. The Lodide/Lodate Actinometer in UV Disinfection: Determination of the Fluence Rate Distribution in UV Reactors. *Photochem. Photobiol.* **2006**, *82* (2), 611−615.
(34) Rahn, R. O. Potassium Iodide as a Chemical Actinometer for 254 Nm Radiation: Use of Lodate as an Electron Scavenger. *Photochem. Photobiol.* **1997**, *66* (4), 450−455.
(35) United States Environmental Protection Agency. Method 1601: Male-Specific (F+) and Somatic Coliphage in Water by Two-Step Enrichment Procedure; Washington, D.C., 2001.
(36) Mozowitz, H. J. Absorption Effects in Volume Irradiation of Microorganisms. *Science (Washington, DC, U. S.)* **1950**, *111* (2879), 229−230.
(37) Kuhn, M.; Johnson, K. *Applied Predictive Modeling*. Springer New York: New York, 2013. DOI: 10.1007/978-1-4614-6849-3.
(38) Harm, W. On the Relationship between Host-Cell Reactivation and UV-Reactivation in UV-Inactivated Phages. *Mol. Genet. Genomics* **1963**, *94* (1), 67−79.
(39) Rupert, C. S.; Harm, W.; AUGENSTEIN, L. G.; MASON, R.; ZELLE, M. A. X. B. T.-A. Reactivation After Photobiological Damage. *Advances in Radiation Biology* **1966**, *2*, 1−81, DOI: 10.1016/B978-1-4832-3121-1.50006-2.
(40) Day, R. S. Studies on Repair of Adenovirus 2 by Human Fibroblasts Using Normal, Xeroderma Pigmentosum, and Xeroderma Pigmentosum Heterozygous Strains. *Cancer Res.* **1974**, *34* (8), 1965 LP−1970.
(41) Rainbow, A. J. Defective Repair of UV-Damaged DNA in Human Tumor and SV40-Transformed Human Cells but Not in Adenovirus-Transformed Human Cells. *Carcinogenesis* **1989**, *10* (6), 1073−1077.
(42) MacRae, S. L.; Croken, M. M.; Calder, R. B.; Aliper, A.; Milholland, B.; White, R. R.; Zhavoronkov, A.; Gladyshev, V. N.; Seluanov, A.; Gorbunova, V.; Zhang, Z. D.; Vijg, J. DNA Repair in Species with Extreme Lifespan Differences. *Aging* **2015**, *7* (12), 1171−1184.
(43) Henderson, E. E. Host Cell Reactivation of Epstein-Barr Virus in Normal and Repairdefective Leukocytes. *Cancer Res.* **1978**, *38* (10), 3256−3263.
(44) Lytle, C. D.; Aaronson, S. A.; Harvey, E. Host-Cell Reactivation in Mammalian Cells. *Int. J. Radiat. Biol. Relat. Stud. Phys., Chem. Med.* **1972**, *22* (2), 159−165.
(45) Smith, K. C. Physical and Chemical Changes Induced in Nucleic Acids by Ultraviolet Light. *Radiat. Res., Suppl.* **1966**, *6*, 54−79.
(46) Schreier, W. J.; Schrader, T. E.; Koller, F. O.; Gilch, P.; Crespo-Hernández, C. E.; Swaminathan, V. N.; Carell, T.; Zinth, W.; Kohler, B. Thymine Dimerization in DNA Is an Ultrafast Photoreaction. *Science (Washington, DC, U. S.)* **2007**, *315* (5812), 625−629.
(47) Meistrich, M. L. Contribution of Thymine Dimers to the Ultraviolet Light Inactivation of Mutants of Bacteriophage T4. *J. Mol. Biol.* **1972**, *66* (1), 97−106.
(48) Law, Y. K.; Forties, R. A.; Liu, X.; Poirier, M. G.; Kohler, B. Sequence-Dependent Thymine Dimer Formation and Photoreversal Rates in Double-Stranded DNA. *Photochem. Photobiol. Sci.* **2013**, *12* (8), 1431−1439.
(49) Becker, M. M.; Wang, Z. Origin of Ultraviolet Damage in DNA. *J. Mol. Biol.* **1989**, *210* (3), 429−438.
(50) Kundu, L. M.; Linne, U.; Marahiel, M.; Carell, T. RNA Is More UV Resistant than DNA: The Formation of UV-Induced DNA Lesions Is Strongly Sequence and Conformation Dependent. *Chem. - Eur. J.* **2004**, *10* (22), 5697−5705.
(51) Setlow, R. B.; Carrier, W. L. The Disappearance of Thymine Dimers from DNA: An Error-Correcting Mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **1964**, *51* (2), 226−231.
(52) Small, G. D.; Tao, M.; Gordon, M. P. Pyrimidine Hydrates and Dimers in Ultraviolet-Irradiated Tobacco Mosaic Virus Ribonucleic Acid. *J. Mol. Biol.* **1968**, *38* (1), 75−87.

https://dx.doi.org/10.1021/acs.est.0c07814
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

(53) Sinha, R. P.; Häder, D.-P. UV-Induced DNA Damage and Repair: A Review. *Photochem. Photobiol. Sci.* **2002**, *1* (4), 225−236.
(54) Eglin, R. P.; Gugerli, P.; Wildy, P. Ultraviolet Irradiation of Herpes Simplex Virus (Type 1): Delayed Transcription and Comparative Sensitivities of Virus Functions. *J. Gen. Virol.* **1980**, *49* (1), 23−31.
(55) Yuan, F.; Zhang, Y.; Rajpal, D. K.; Wu, X.; Guo, D.; Wang, M.; Taylor, J.-S.; Wang, Z. Specificity of DNA Lesion Bypass by the Yeast DNA Polymerase *η*. *J. Biol. Chem.* **2000**, *275* (11), 8233−8239.
(56) Streisinger, G. The Genetic Control of Ultraviolet Sensitivity Levels in Bacteriophages T2 and T4. *Virology* **1956**, *2* (1), 1−12.
(57) Harm, W. Mutants of Phage T4 with Increased Sensitivity to Ultraviolet. *Virology* **1963**, *19* (1), 66−71.
(58) Pratelli, A. Canine Coronavirus Inactivation with Physical and Chemical Agents. *Vet. J.* **2008**, *177* (1), 71−79.
(59) Darnell, M. E. R.; Taylor, D. R. Evaluation of Inactivation Methods for Severe Acute Respiratory Syndrome Coronavirus in Noncellular Blood Products. *Transfusion* **2006**, *46* (10), 1770−1777.
(60) Darnell, M. E. R.; Subbarao, K.; Feinstone, S. M.; Taylor, D. R. Inactivation of the Coronavirus That Induces Severe Acute Respiratory Syndrome, SARS-CoV. *J. Virol. Methods* **2004**, *121* (1), 85−91.
(61) Sagripanti, J.-L.; Lytle, C. D. Estimated Inactivation of Coronaviruses by Solar Radiation With Special Reference to COVID-19. *Photochem. Photobiol.* **2020**, *96* (4), 731−737.
(62) Faddy, H. M.; Fryk, J. J.; Prow, N. A.; Watterson, D.; Young, P. R.; Hall, R. A.; Tolksdorf, F.; Sumian, C.; Gravemann, U.; Seltsam, A.; Marks, D. C. Inactivation of Dengue, Chikungunya, and Ross River Viruses in Platelet Concentrates after Treatment with Ultraviolet C Light. *Transfusion* **2016**, *56* (6pt2), 1548−1555.
(63) Blázquez, E.; Rodríguez, C.; Ródenas, J.; Navarro, N.; Riquelme, C.; Rosell, R.; Campbell, J.; Crenshaw, J.; Segalés, J.; Pujols, J.; Polo, J. Evaluation of the Effectiveness of the SurePure Turbulator Ultraviolet-C Irradiation Equipment on Inactivation of Different Enveloped and Non-Enveloped Viruses Inoculated in Commercially Collected Liquid Animal Plasma. *PLoS One* **2019**, *14* (2), e0212332.
(64) Mohr, H.; Steil, L.; Gravemann, U.; Thiele, T.; Hammer, E.; Greinacher, A.; Mueller, T. H.; Voelker, U. A Novel Approach to Pathogen Reduction in Platelet Concentrates Using Short-Wave Ultraviolet Light. *Transfusion* **2009**, *49* (12), 2612−2624.
(65) Bolton, J. R.; Linden, K. G. Standardization of Methods for Fluence (UV Dose) Determination in Bench-Scale UV Experiments. *J. Environ. Eng.* **2003**, *129* (3), 209−215.
(66) Rockey, N.; Young, S.; Kohn, T.; Pecson, B.; Wobus, C. E.; Raskin, L.; Wigginton, K. R. UV Disinfection of Human Norovirus: Evaluating Infectivity Using a Genome-Wide PCR-Based Approach. *Environ. Sci. Technol.* **2020**, *54* (5), 2851−2858.
(67) U.S. Environmental Protection Agency. *National Primary Drinking Water Regulations: The Long Term 2 Enhanced Surface Water Treatment Rule. EPA-HQ-*; 2006.
(68) Wigginton, K. R.; Menin, L.; Sigstam, T.; Gannon, G.; Cascella, M.; Hamidane, H.; Ben; Tsybin, Y. O.; Waridel, P.; Kohn, T. UV Radiation Induces Genome-Mediated, Site-Specific Cleavage in Viral Proteins. *ChemBioChem* **2012**, *13* (6), 837−845.

https://dx.doi.org/10.1021/acs.est.0c07814
Environ. Sci. Technol. XXXX, XXX, XXX−XXX