

Re-examining the relationship between virus and microbial cell abundances in the global oceans

Charles H. Wigington,¹ Derek Sonderegger,² Corina P.D. Brussaard,^{3,4} Alison Buchan,⁵ Jan F. Finke,⁶ Jed Fuhrman,⁷ Jay T. Lennon,⁸ Mathias Middelboe,⁹ Curtis A. Suttle,¹⁰ Charles Stock,¹¹ William H. Wilson,¹² K. Eric Wommack,¹³ Steven W. Wilhelm,^{5,*} and Joshua S. Weitz^{1,14,†}

¹ School of Biology, Georgia Institute of Technology, Atlanta, GA 30332

² Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, AZ

³ Department of Biological Oceanography, Royal Netherlands Institute for Sea Research (NIOZ), Texel, The Netherlands.

⁴ Department of Aquatic Microbiology, Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, The Netherlands.

⁵ Department of Microbiology, University of Tennessee-Knoxville, Knoxville, TN

⁶ Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver BC V6T 1Z4

⁷ Dept of Biological Sciences, University of Southern California, Los Angeles 90089

⁸ Department of Biology, Indiana University, Bloomington, IN 47405

⁹ Marine Biological Section, Department of Biology, University of Copenhagen, Helsingør, Denmark

¹⁰ Departments of Earth, Ocean and Atmospheric Sciences, Microbiology and Immunology, and Botany, University of British Columbia, Vancouver BC V6T 1Z4; Program in Integrated Microbial Diversity, Canadian Institute for Advanced Research, Toronto ON M5G 1Z8

¹¹ Geophysical Fluid Dynamics Laboratory Princeton, NJ 08540-6649

¹² Sir Alister Hardy Foundation for Ocean Science, The Laboratory, Citadel Hill, Plymouth, UK

¹³ Plant and Soil Sciences Delaware Biotechnology Institute Newark, DE

¹⁴ School of Physics, Georgia Institute of Technology, Atlanta, GA 30332

(Dated: August 25, 2015)

Marine viruses are critical drivers of ocean biogeochemistry and their abundances vary spatiotemporally in the global oceans, with upper estimates exceeding 10^8 per ml. Over many years, a consensus has emerged that virus abundances are typically 10-fold higher than prokaryote abundances. The use of a fixed-ratio suggests that the relationship between virus and prokaryote abundances is both predictable and linear. However, the true explanatory power of a linear relationship and its robustness across diverse ocean environments is unclear. Here, we compile 5671 prokaryote and virus abundance estimates from 25 distinct marine surveys to characterize the relationship between virus and prokaryote abundances. We find that the median virus-to-prokaryote ratio (VPR) is 10:1 and 16:1 in the near- and sub-surface oceans, respectively. Nonetheless, we observe substantial variation in the VPR and find either no or limited explanatory power using fixed-ratio models. Instead, virus abundances are better described as nonlinear, power-law functions of prokaryote abundances - particularly when considering relationships within distinct marine surveys. Estimated power-laws have scaling exponents that are typically less than 1, signifying that the VPR decreases with prokaryote density, rather than remaining fixed. The emergence of power-law scaling presents a challenge for mechanistic models seeking to understand the ecological causes and consequences of marine virus-microbe interactions. Such power-law scaling also implies that efforts to average viral effects on microbial mortality and biogeochemical cycles using “representative” abundances or abundance-ratios need to be refined if they are to be utilized to make quantitative predictions at regional or global ocean scales.

Viruses of microbes have been linked to central processes across the global oceans, including biogeochemical cycling [9, 29, 39, 43, 47, 53] and the maintenance and generation of microbial diversity [3, 36, 39, 47, 52]. Virus propagation requires contacting and infecting cells. The per cell rate at which microbial cells – including bacteria, archaea, and microeukaryotes – are contacted by viruses is assumed to be proportional to the product of virus and microbial abundances [33]. If virus and microbe

abundances were related in a predictable way it would be possible to infer the rate of contact, and potentially the relative importance of virus-induced cell lysis, from estimates of microbial abundance alone.

Virus ecology underwent a transformation in the late 1980s with the recognition that virus abundances, as estimated using culture-independent methods, were orders of magnitude higher than estimates via culture-based methods [4]. Soon thereafter, researchers began to report the “virus to bacterium ratio” (VBR) as a statistical proxy for the strength of the relationship between viruses and their potential hosts in both freshwater and marine systems [31]. This ratio is more appropriately termed the “virus-to-prokaryote ratio” (VPR) – a convention which

*Electronic address: wilhelm@utk.edu

†Electronic address: jsweitz@gatech.edu

Year	Observation	Reference
1894	Marine bacteria are first discussed by Certes, Fischer and Russell	Certes [14], Fischer [22, 23], Russell [40]
1915,1917	Bacteriophage are discovered	d’Herelle [21], Twort [49]
1925	The presence of bacteriophage in seawater is noted	Arloing and Chavanne [2]
1946	ZoBell reports that bacteriophage occur only sporadically and in the littoral zone and concludes there is insufficient evidence for viruses to be considered as key to limiting open ocean bacteria	Carlucci and Pramer [12], ZoBell [59]
1947	The presence of bacteriophage described in the oceans	Kriss A..E [30]
1979	Using transmission electronic microscopy, up to 10^4 ml ⁻¹ bacteriophage particles are observed in coastal water, an observations that sparked the rebirth of virus ecology a decade later.	Torrella and Morita [48]
1989	“Rebirth” of virus ecology across a series off papers begins with a report of virus and bacteria abundances for which VBRs range from 0.2 (Raunefjorden) to 50 (North Atlantic)	Bergh et al. [4]
1990	Report of virus particles ranging from 10^6 - 10^{11} per liter, infecting up to 7% of heterotrophic bacteria and each infected cell containing 10-100 mature virions	Proctor and Fuhrman [37]
1991-1993	Estimates of virus abundance exceeding bacteria abundance by 5-10 fold from a series of papers (this observation noted in Fuhrman and Suttle [25])	Cochlan et al. [17], Hara et al. [26], Paul et al. [34, 35], Wommack et al. [58]
1995	Maranger and Bird [31] survey 22 Quebec lakes and collect literature from 14 studies [4, 5, 7, 17, 26–28, 34, 41, 51, 58] and report VBR higher in freshwater (20-25) than marine systems (1-5).	Maranger and Bird [31]
2000	Wommack and Colwell suggest that VBR typically ranges between 3 and 10, and note that VBR decreases as prokaryote abundance increases.	Wommack and Colwell [57]
2000	A VBR “roughly equal to 10” (attributed to Maranger and Bird [31] is designated as a target for parameterizing the Kill-the-Winner theory of virus-microbe interactions.	Thingstad [47]
2004	Consistency in VBR is attributed to the idea that most viruses are phage that infect bacteria. Notes a VBR of 10 in marine systems and attributes to Maranger and Bird [31].	Weinbauer [50]
2004	Chibani-Chennoufi and colleagues advance the notion that VBR is 10:1 in the ocean and that this is justified by the claim that each prokaryote species can be infected by 10 different phage.	Chibani-Chennoufi et al. [15]
2008	VBR ratios reviewed in several publications that collate information from multiple studies, with a 10:1 consensus despite noted variation.	Clasen et al. [16], Wilhelm and Matteson [56]
2011	VBR reviewed across several regimes, with evidence for a linear relationship between viruses and prokaryote in the water column and a nonlinear relationship in sediment.	Danovaro et al. [18]
2014	The BioNumbers database, intended to facilitate quantitative analysis in the biosciences, lists VBR as 10	Milo et al. [32]

TABLE I: Origins and emerging consensus of the 10:1 ratio of virus abundance to bacteria abundance in aquatic systems - from freshwater lakes to the global oceans. We use the convention VPR in this manuscript rather than VBR.

we use here [1].

Observations accumulating over the past 25 years have observed wide variation in VPR, yet there is an emergent census that a suitable first-approximation is that VPR is 10 (see Table I). This ratio also reflects a consensus that typical prokaryote abundances are approximately 10^6 per ml and typical virus abundances are approximately 10^7 per ml [50, 57]. Yet, the use of a fixed ratio carries with it another assumption: that of linearity, i.e., if prokaryote abundance were to double, then viruses are expected

to double as well. An alternative is that the relationship between virus and prokaryote abundance is better described in terms of a nonlinear relationship, e.g., a power-law.

In practice, efforts to predict the regional or global-scale effects of viruses on marine microbial mortality, turnover and even biogeochemical cycles, depend critically on the predictability of the relative density of viruses and microbial cells. Here, we directly query the nature of the relationship between viruses and prokaryotes via

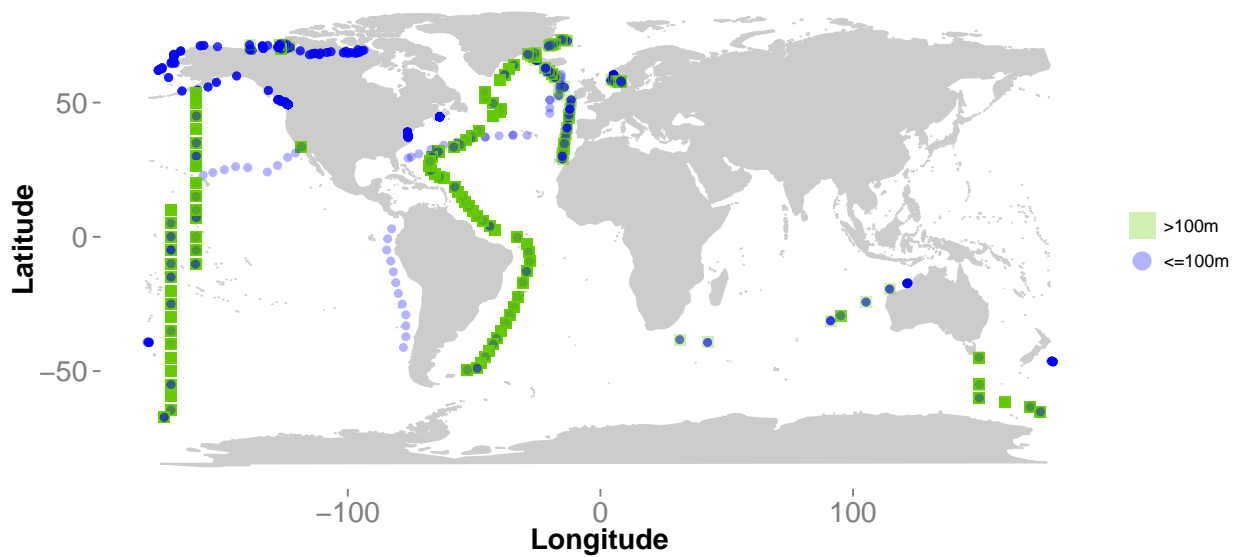


FIG. 1: Global distribution of sample sites. Each point denotes a location from which one or more samples were taken. Samples range from the surface to up to 5,500 meters below sea level, with 2,921 taken near the surface ($\leq 100\text{m}$), noted as squares, and 2,750 taken below the surface ($>100\text{m}$), noted as circles. The number of points for each study – the “Frequency” – is found in Table S2. Full access to data is provided as Supplementary File 1.

a large-scale compilation and re-analysis of abundance data across marine environments.

I. RESULTS

A. VPR exhibits substantial variation in the global oceans

In the compiled marine survey data (see Figure 1, Table S1 and the Materials and Methods), 95% of prokaryote abundances range from 5.0×10^3 to 4.1×10^6 per ml and 95% of virus abundances range from roughly 3.7×10^5 to 6.4×10^7 per ml (Figure 2A). Both prokaryote and virus concentrations generally decrease with depth as reported previously (e.g., see [18]). The median VPR for the near-surface samples ($\leq 100\text{m}$) is 10.5 and the median VPR for the sub-surface samples ($> 100\text{m}$) is 16.0. In that sense, the consensus 10:1 ratio does accurately represent the median VPR for the surface data. We also observe substantial variation in VPR, as has been noted in prior surveys and reviews (see Table 1). Figure 2B shows that 95% of the variation in VPR in the near-surface ocean lies between 1.4 and 160 and between 3.9 and 74 in the sub-surface ocean. For the near-surface

ocean, 44% of the VPR values are between 5 and 15, 16% are less than 5 and 40% exceed 15. This wide distribution, both near- and sub-surface demonstrates potential limitations in utilizing the 10:1 VPR, or any fixed ratio, as the basis for a *predictive* model of virus abundance derived from estimates of prokaryote abundance.

B. Virus abundance does not vary linearly with prokaryote abundance

Figure 3 shows two alternative, predictive models of the relationship between logarithmically scaled virus and prokaryote abundances for water column samples. The models correspond to a fixed-ratio model and a power-law model. To clarify the interpretation of fitting in log-log space consider a fixed-ratio model with a 12:1 ratio between virus and prokaryote abundance, $V = 12 \times B$. Then, in log-log space the relationship is

$$\log_{10}(V) = \log_{10} 12 + \log_{10} B \quad (1)$$

which we interpret as a line with y-intercept of $\log_{10} 12 = 1.08$ and a slope (change in $\log_{10} V$ for a 1-unit change in $\log_{10} B$) of 1. By the same logic, any fixed-ratio model will result in a line with slope 1 in the log-log plot and

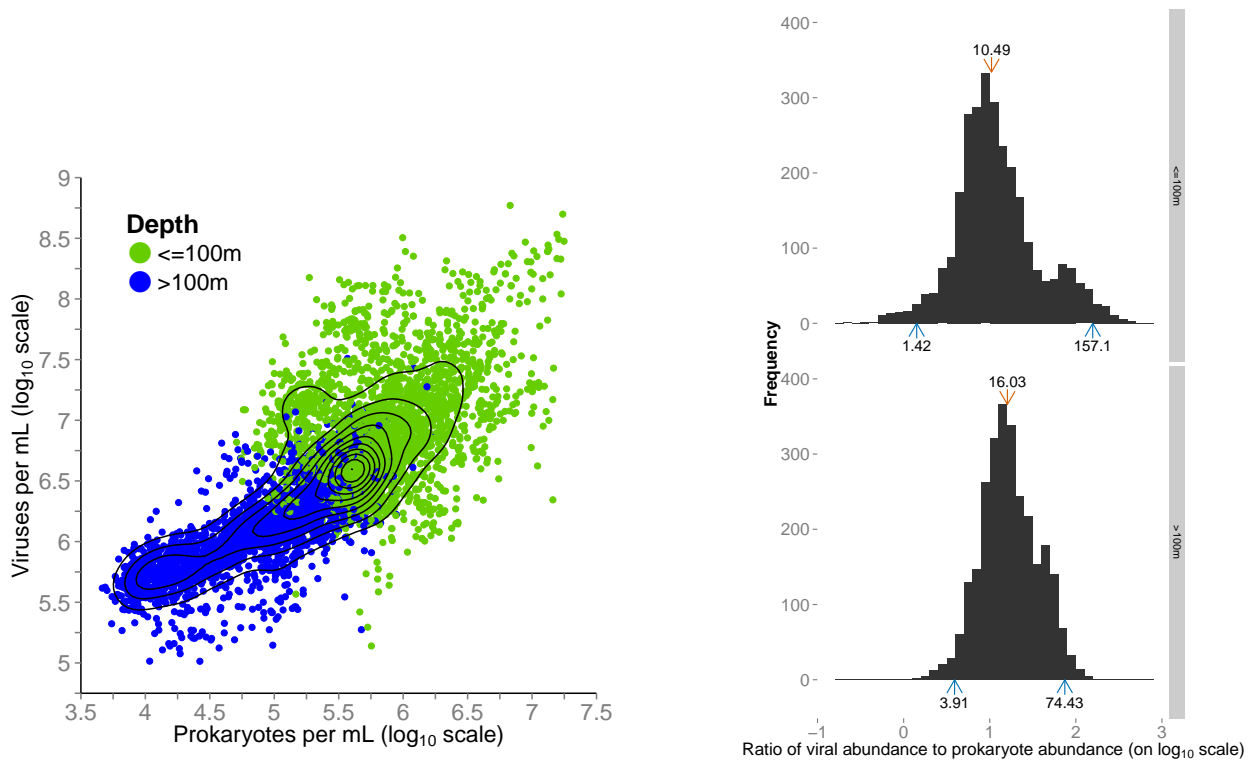


FIG. 2: Variation in virus and prokaryote abundances and the VPR. (A) Prokaryote abundance vs. virus abundance colored by depth. Each point represents a biological sample. (B) Histogram of the logarithm of VPR. The top and bottom panels correspond to near- and sub-surface water column samples, respectively. The red arrow denotes the median value and the blue arrows denote the central 95% range of values - where the numbers associated with each arrow denote the non-transformed value of VPR.

Model	≤ 100m		>100m	
	R^2	AIC	R^2	AIC
10:1	-0.16	-15305.83	-0.25	-14492.09
Power Law	0.15	-16301.81	0.64	-18313.82
Constrained Power Law	0.39	-17292.11	0.66	-18513.48
Power Law by Study	0.79	-20293.10	0.72	-18972.81

TABLE II: Information theoretic comparison of alternative models of the relationship between virus and prokaryote abundance. The values of the Aikake Information Criteria (AIC) are defined in the Materials and Materials and Methods. The value of R^2 for each model denotes the relative amount of variance explained. Negative values of R^2 mean that a model explains less variance than does the overall mean.

the y-intercept will vary logarithmically with VPR. The alternative predictive model is that of a power-law: $V = cB^{\alpha_1}$. In log-log space, the relationship is:

$$\log_{10} V = \log_{10} c + \alpha_1 \log_{10} B, \quad (2)$$

$$\log_{10} V = \alpha_0 + \alpha_1 \log_{10} B. \quad (3)$$

The slope, α_1 , of a fitted line on log-transformed data denotes the *power-law exponent* that best describes the relationship between the variables. The intercept, α_0 , of a fitted line on log-transformed data denotes the logarithmically transformed pre factor.

The 10:1 line has a residual squared error of -16% and -25% in the surface and deep samples, respectively (Table 2). In both cases, this result means that a 10:1 line explains *less* of the variation in virus abundance compared to a model in which virus abundance is predicted by its mean value across the data. In order to evaluate the generality of this result, we considered an ensemble of fixed-ratio models each with a different VPR. In the near-surface samples, we find that all fixed-ratio models explain less of the variation (i.e., have *negative* values of R^2) than does a “model” in which virus abundance is predicted to be the global mean in

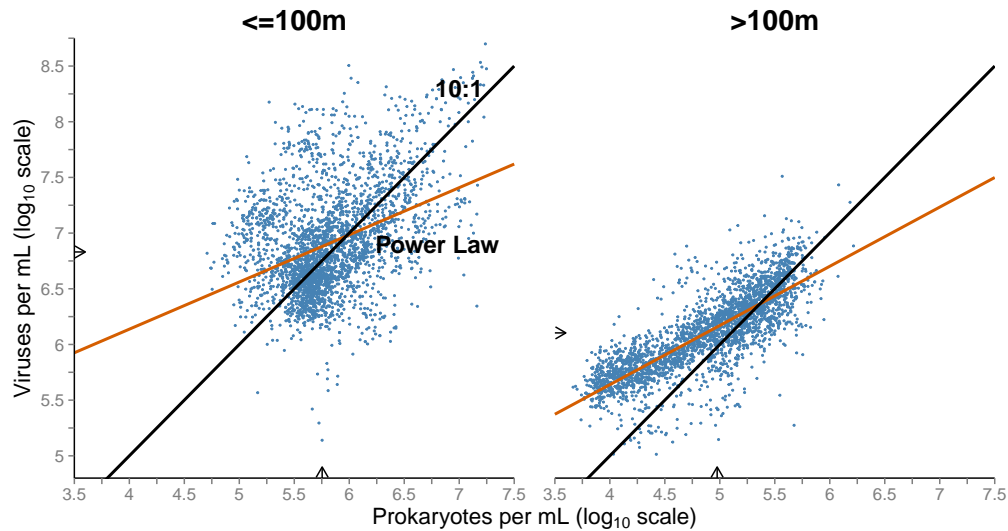


FIG. 3: Virus abundance is poorly fit by a model of 10-fold increase relative to prokaryote abundance. (Left) Surface ocean – the red line denotes the best fit power-law with an exponent of 0.42 while the black line denotes the 10:1 curve. The best-fit power law explains 15% of the variation and the 10:1 line explains -16% of the variation. See text for interpretation of negative R^2 and the importance of outliers in these fits. (Right) Deeper water column – the red line denotes the best fit power-law with an exponent of 0.53 while the black line denotes the 10:1 curve. The best-fit power law explains 64% of the variation and the 10:1 line explains -26% of the variation. In both cases the arrows on the axes denote the median of the respective abundances.

the dataset (Figure S1). This reflects the failure of constant ratio (i.e., linear) models to capture the cluster of high VPRs at low prokaryote density apparent in the density contours of Figure 2A and the shoulder of elevated high VPR frequency in Figure 2B. The largest contributor to this cluster of points is the Arctic SBI study (see Figure S1). Whereas, in the sub-surface samples, fixed-ratio models in which VPR varies between 12 and 22 do have positive explanatory power, but all perform worse than does the power-law model (Figure S1). In contrast, the best fitting power-law model explains 15% and 64% of the variation in the data, for near- and sub-surface samples respectively (Table 2). The best-fit power-law scaling exponent is 0.42 with 95% confidence intervals (CIs) of (0.39, 0.46) for near-surface samples and 0.53 with 95% CIs of (0.52, 0.55) for sub-surface samples. In other words, doubling prokaryote abundance along either regression line is not expected to lead to a doubling in virus abundance, but rather a $2^{0.42} = 1.33$ and $2^{0.53} = 1.44$ fold increase, respectively. The power-law model is an improvement over the fixed ratio model in both cases, even when accounting for the increase in parameters (Table 2). In the near-surface, refitting surface data without outliers improves explanatory power to approximately $R^2 = 0.3$ in contrast to an $R^2 = 0.65$ for the sub-surface (see Methods and Figure S2). In summary, the predictive value of a power-law model is much stronger in the sub-surface than in the near-surface, where confidence in the interpretation of power-law expo-

nents is limited.

C. Study-to-study measurement variation is unlikely to explain the intrinsic variability of virus abundances in the surface ocean

Next, we explored the possibility that variation in methodologies affected the baseline offset of virus abundance measurements and thereby decreased the explanatory power of predicting virus abundances based on prokaryote abundances. That is, if V^* is the true and unknown abundance of viruses, then it is possible that two studies would estimate $\hat{V}_1 = V^*(1 + \epsilon_1)$ and $\hat{V}_2 = V^*(1 + \epsilon_2)$ where $|\epsilon_1|$ and $|\epsilon_2|$ denote the relative magnitude of study-specific shifts. We constrain the relative variation in measurement, such that the measurement uncertainty is 50% or less (see Materials and Methods). The constrained regression model improves the explanatory power of the model (see Table 2), but in doing so, the model forces 18 of the 25 studies to the maximum level of measurement variation permitted (Figure S3). We do not expect that differences in measurement protocols to explain nearly 2 orders of magnitude variation in estimating virus abundance given the same true virus abundance in a sample. Note that when sub-surface samples were analyzed through the constrained power-law model, there was only a marginal increase of 2% in R^2 and, moreover, 9 of the 12 studies were fit given the

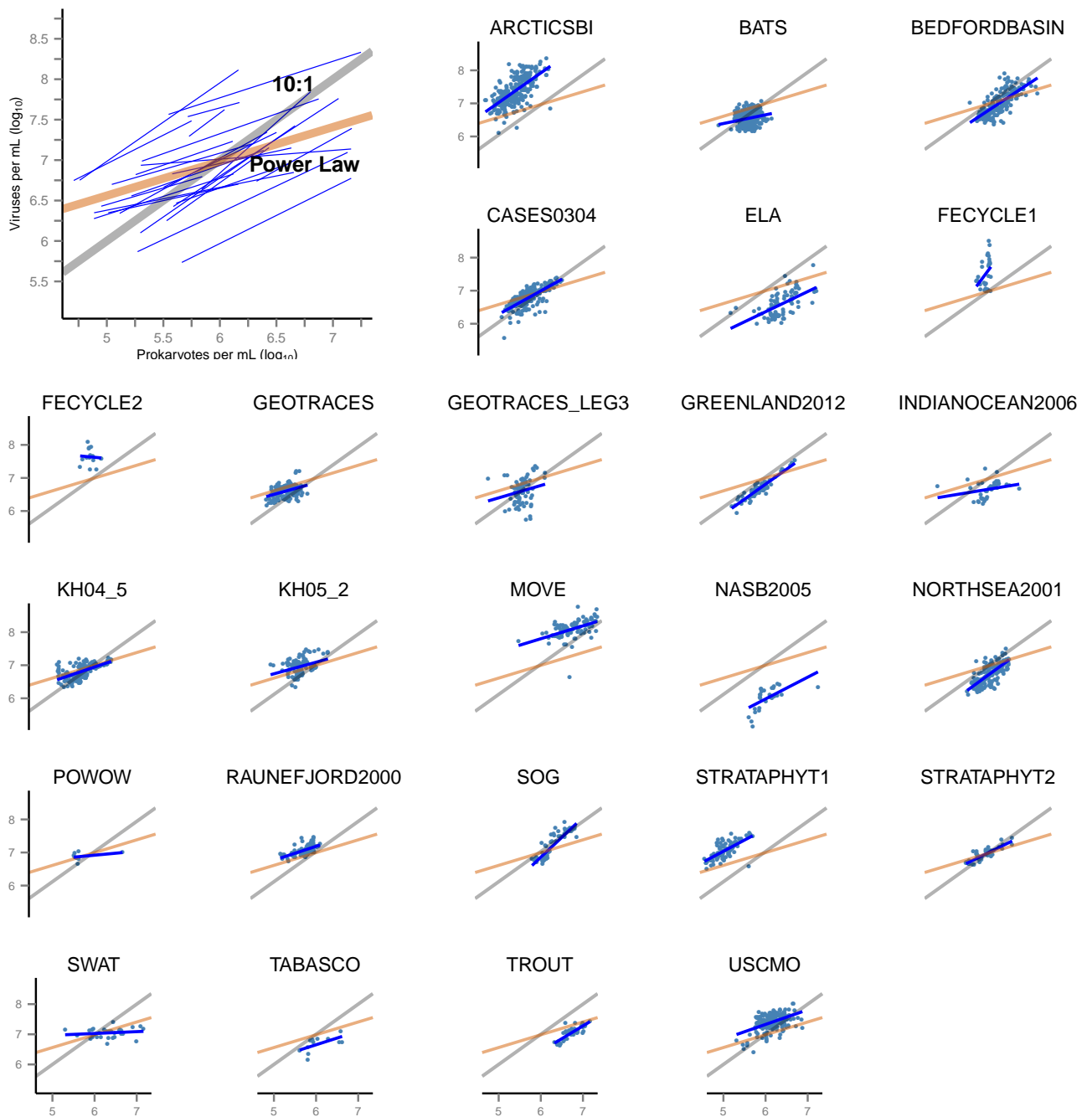


FIG. 4: Virus-prokaryote relationships given the variable slope and intercept mixed-effects model. (Upper-left) Best-fit power-law for each study (blue lines) plotted along with the best-fit power-law of the entire dataset (red line) and the 10:1 line (grey line). (Individual panels) Best-fit power-law model (blue line) on log-transformed data (blue points) for each study, with the power-law model regression (red) and 10:1 line (black) as reference. The power-law exponents and associated confidence intervals are shown in Figure 5,

maximum level of measurement variation permitted (Figure S3)[h!]. The constrained intercept model results suggest that the observed variation in virus abundance in the surface oceans is not well explained strictly by variation in measurement protocol between studies.

D. VPR decreases with increasing prokaryote abundance – a hallmark of power-law relationships

We next evaluate an ensemble of power-law models: $V_i = c_i N^{\alpha_i}$ where the index i denotes the use of distinct intercepts and power-law exponents for each sur-

vey. The interpretation of this model is that the nonlinear nature of the virus to prokaryote relationship may differ in distinct oceanic realms or due to underlying differences in sites or systems, rather than due to measurement differences. Figure 4 shows the results of fitting using the study-specific power-law model in the surface ocean samples. Study-specific power-law fits are significant in 18 of 25 cases in the surface ocean. The median power-law exponent for studies in the surface ocean is 0.50. Furthermore, of those significant power-law fits, the 95% distribution of the power-law exponent excludes a slope of one and is entirely less than one in 11 of 18 cases (see Figure 5). This model in which the power-law exponent varies with study is a significant improvement in terms of R^2 (Table 2). For sub-surface samples, study-specific power-law fits are significant in 10 of 12 cases in the sub-surface (Figure S4). The median power-law exponent for studies in the sub-surface is 0.67. Of those significant power-law fits, the 95% distribution of the power-law exponent is entirely less than one in 6 of 10 cases (see Figure S5). A power-law exponent of less than one means that virus abundance increases less than proportionately given increases in prokaryote abundance. This study-specific analysis extends the findings that nonlinear, rather than linear, models are more suitable to describe the relationship between virus and prokaryote abundances. We find that the dominant trend in both near-surface and sub-surface samples is that VPR decreases as prokaryote abundance increases. The increased explanatory power by study is stronger for near-surface than for sub-surface samples. This increase in R^2 comes with a caveat: study-specific models do not enable *a priori* predictions of virus abundance given a new environment or sample site, without further effort to disentangle biotic and abiotic factors underlying the different scaling relationships.

II. DISCUSSION

Viruses are increasingly considered as part of efforts to understand the factors controlling marine microbial mortality, productivity, and biogeochemical cycles [8, 9, 19, 39, 45, 55]. Quantitative estimates of these viral-induced effects can be measured directly, but are often inferred indirectly, using the relative abundance of viruses to prokaryotes. To do so, there is a consensus that assuming the virus-to-prokaryote ratio is 10 in the global oceans - despite observed variation - is a reasonable starting point. Here, we have re-analyzed the relationship of virus to prokaryote abundance in 25 marine survey datasets. We find that 95% of the variation in VPR ranges from 1.4 to 160 in the near-surface ocean and from 3.9 to 74 in the sub-surface. Although the 10:1 ratio accurately describes the median of the VPR in the surface ocean, the broad distribution of VPR implies that prokaryote abundance is a poor quantitative predictor of virus abundance. Moreover, increases in prokary-

ote abundance do not lead to proportionate increases in virus abundance. Instead, we propose that the virus to prokaryote abundance relationship is nonlinear, and that the degree of nonlinearity - as quantified via a power-law exponent - is typically less than 1. This sublinear relationship can be interpreted to mean that VPR decreases as an increasing function of prokaryote abundance, and generalizes earlier observations [57].

Power-law relationships between virus and prokaryote abundance emerge from complex feedbacks involving both exogenous and endogenous factors. The question of exogenous factors could be addressed, in part, by examining environmental covariates at survey sites. For example, if prokaryote and virus abundances varied systematically with another environmental co-factor during a transect, then this would potentially influence the inferred relationship between virus and prokaryote abundances. In that same way, variation in environmental correlates, including temperature and incident radiation, may directly modify virus life history traits [20, 44]. Likewise, some of the marine survey datasets examined here constitute repeated measurements at the same location (e.g., at the Bermuda Atlantic Time-series Study (BATS)). Time-varying environmental factors could influence the relative abundance of microbes and viruses. It is also interesting to note that viruses-induced mortality is considered to be more important at eutrophic sites [57], where microbial abundance is higher - yet the observed decline in VPR with prokaryote abundance would suggest the opposite.

It could also be the case that variation in endogenous factors, e.g., the life history traits of viruses and microbes, including variability in the structure of cross-infection [54], could drive changes in emergent relationships between virus and prokaryote abundances. Virus-microbe interactions can lead to intrinsic oscillatory dynamics. Indeed, previous observations of a declining relationship between VPR and prokaryote abundance have been attributed to changing ratios across phytoplankton bloom events, including possible virus-induced termination of blooms [57]. Similar arguments were proposed in the analysis of tidal sediments [13]. Alternatively, observations of declining VPR with prokaryote density have been attributed to variation in underlying diversity [6]. Whatsoever the mechanism(s), it is striking that virus abundances in some surveys can be strongly predicted via alternative power-law functions of prokaryote abundances. Mechanistic models are needed to further elucidate these emergent macroecological patterns and relationships.

The present analysis separated the abundance data first according to depth and then according to survey as a means to identify different relationships between virus and prokaryote abundances in the global oceans. The predictive value of total prokaryote abundance is strong when considering sub-surface samples. In contrast, prokaryote abundance is not a strong predictor of virus abundance in the near-surface samples, when utilizing linear or nonlinear models. The predictive power

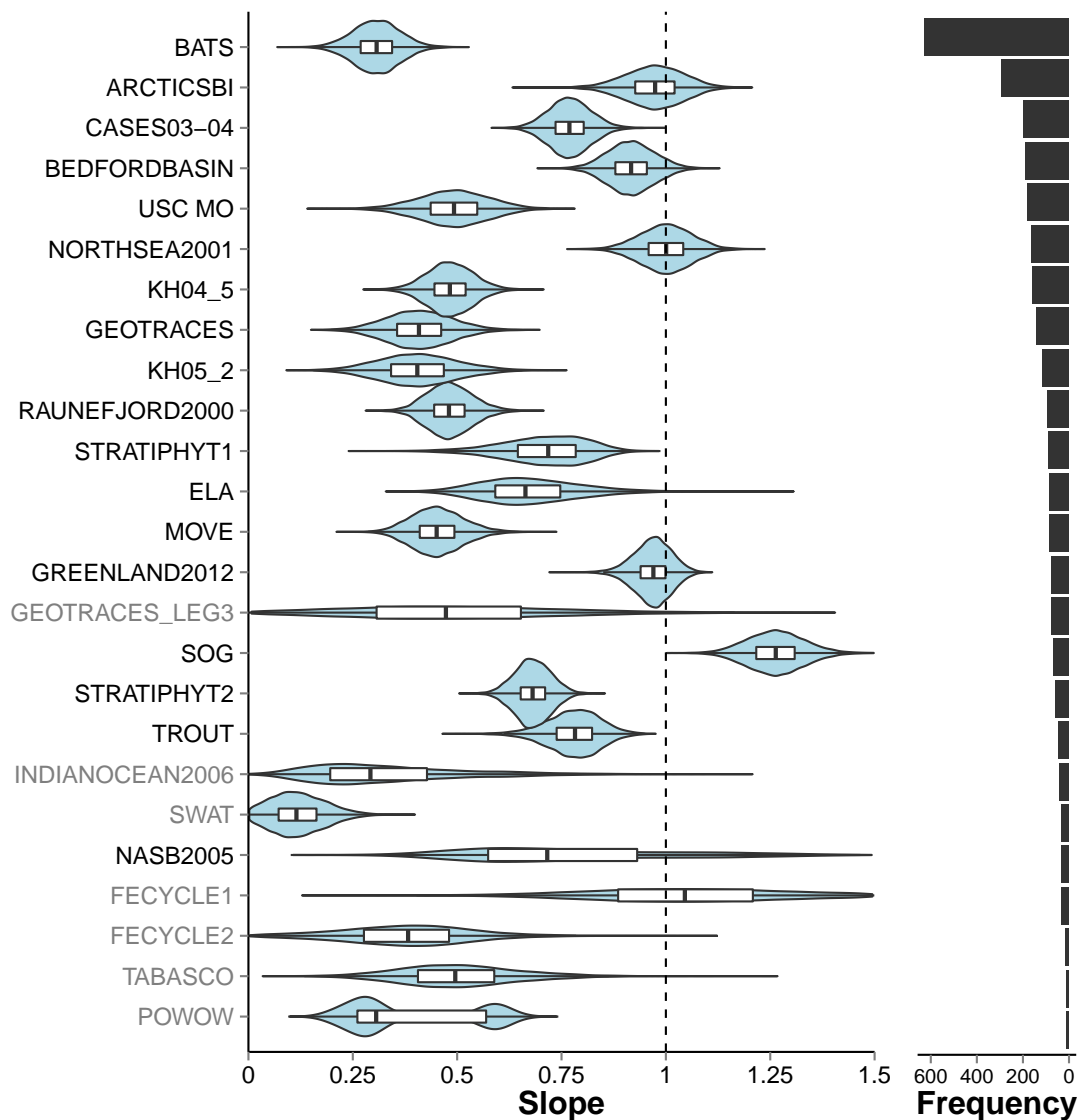


FIG. 5: Study-specific 95% confidence intervals of power-law exponents for relationships between virus and prokaryote abundance in the surface. The confidence intervals are plotting using “violin” plots including the median (center black line), 75% distribution (white bars) and 95% distribution (black line), with the distribution overlaid (blue shaded area). The number of points included as part of each study is displayed on the right-most bar plots. Study labels in black indicate those studies for which the regression fit had a p-value less than $0.002=0.05/25$ (accounting for a multiple comparison correction given the analysis of 25 studies). Study labels in gray indicate a p-value above this threshold.

of nonlinear models improved substantially in the near-surface when evaluating each marine survey separately. The minimal predictive value of prokaryote abundances in the near-surface when aggregating across all surveys is problematic given that virus-microbe interactions have significant roles in driving prokaryote mortality and ecosystem functioning [9, 45, 53]. Indeed the aggregation of abundance measurements in terms of total prokaryote abundances may represent part of the problem. At a given site and time of sampling, each microbial cell in the community is potentially targeted by a subset

of the total viral pool. In moving forward, understanding variation in virus abundance and its relationship to prokaryote abundance requires a critical examination of correlations at functionally relevant temporal and spatial scales, i.e., at the scale of interacting pairs of viruses and microbes. We encourage the research community to prioritize examination of these scales of interaction as part of efforts to understand mechanisms underlying nonlinear virus-microbe abundance relationships in the global oceans.

III. MATERIALS AND METHODS

A. Data source

Marine virus abundance data was aggregated from 25 studies (Table S1). A total of 5671 data points were aggregated. The data collection dates range from 1996 to 2012. Data primarily comes from coastal waters in the northern hemisphere and were collected predominately during the summer months, with the notable exceptions of long-term coastal monthly monitoring sites, i.e., the studies USC MO, BATS, and MOVE.

B. Data processing

Analyses of the data were performed using R version 3.1.1. Scripts and original data are provided at https://github.com/cwington3/VIRBAC_analysis.

C. Power-law model

A power-law regression model used the \log_{10} of the predictor variable, prokaryote abundance per mL N , and the \log_{10} of the outcome variable, virus abundance per mL V . The power-law regression was calculated using the equation $\log_{10} V = \alpha_0 + \alpha_1 \log_{10} N$. The α_0 and α_1 parameters were fit via OLS regression to minimize the sum of square error.

D. Constrained variable-intercept model

The constrained model is a “mixed-effects” regression model using the same predictor and outcome variables, \log_{10} of prokaryote abundance per mL and the \log_{10} virus abundance per mL, respectively. This model includes study-specific intercepts which were constrained such that the value for any of the intercepts were restricted to one standard error above or below the intercept value taken from the power-law model. The standard error value for this model came from the power-law model. The equation for this model is $V = \alpha_0^{(i)} + \alpha_1 N$, where $\alpha_0^{(i)}$ is the study-specific intercept and α_1 is the slope common to all studies, N is the predictor variable, and V is the outcome variable.

E. Variable slope and variable intercept model

A power-law model where the exponent and intercept varied with each study was evaluated using the same predictor variable, \log_{10} prokaryote abundance per mL, and the same outcome variable, \log_{10} virus abundance per mL. In this model, there was a study-specific α_0 and α_1 and a OLS regression calculated using the equation $V = \alpha_0^{(i)} + \alpha_1^{(i)} N$.

F. Bootstrapping model confidence intervals

Bootstrap analyses of the power-law model and mixed effects models were conducted to derive 95% confidence intervals surrounding the parameters estimated by the models. For all models the original dataset was sampled with replacement, by study, to arrive at a bootstrap sample dataset, this process was repeated 10,000 times. Distributions for all parameters were generated and the 2.5%, 50%, and 97.5% points were identified from among the 10,000 parameter estimates.

G. Outlier identification

Outliers in the data were identified by calculating the top and bottom 2% of estimated VPR amongst the entire 5,671 samples. The outliers corresponded to ratios below 1.81 and above 128. Those samples with virus to prokaryote ratios which fell outside of these bounds were considered outliers. There were 218 outlier samples taken at depths $\leq 100\text{m}$ and 10 outlier samples taken at depths $> 100\text{m}$.

This work was supported by NSF grants OCE-1233760 (JSW) and OCE-1061352 (AB and SWW), a Career Award at the Scientific Interface from the Burroughs Wellcome Fund (JSW) and a Simons Foundation SCOPE grant (JSW). This work arose from discussions in the Ocean Viral Dynamics working group at the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation, the U.S. Department of Homeland Security, and the U.S. Department of Agriculture through NSF Award EF-0832858, with additional support from The University of Tennessee, Knoxville.

[1] We use the term virus abundance throughout this manuscript to denote estimates derived from culture-independent methods, including epifluorescence microscopy [46] or flow cytometry [10]. Viruses measured in these methods are generally thought to represent bacteriophage, consistent with the numerical dominance of bacteria in seawater [50]. Yet, current-

ly available methods have potential limitations. For example, ssDNA and RNA viruses [11, 42] and giant viruses [38] are under-counted when estimates are made via epifluorescence microscopy with standard DNA based stains. Likewise, virus-like particles may not always necessarily represent viruses, and total abundance estimates could include contributions from gene-transfer agents

- or vesicles [24]. Similarly, we use the term prokaryote abundance to denote estimates from culture-independent methods, including epifluorescence microscopy or flow cytometry. These abundance estimates are based on counts of cellular-like particles, predominantly bacteria but possibly archaea. Our use of the terms virus and prokaryote abundance represent operational definitions as measured in natural environments.
- [2] F. Arloing and Chavanne. On the influence of the environment on the bacteriophage, electrolytes and the concentration of H ions. *Comptes Rendus Des Sances De La Socit De Biologie Et De Ses Filiales*, 93:531–532, 1925.
- [3] S Avrani, D. A. Schwartz, and D. Lindell. Virus-host swinging party in the oceans: Incorporating biological complexity into paradigms of antagonistic coexistence. *Mobile Genetic Elements*, 2:88–95, 2012.
- [4] O. Bergh, K. Y. Borhseim, G. Bratbak, and M. Heldal. High abundance of viruses found in aquatic environments. *Nature*, 340:467–468, 1989.
- [5] J. Boehme, M. E. Frischer, S. C. Jiang, C. A. Kellogg, S. Pichard, J. B. Rose, C. Steinway, and J. H. Paul. Viruses, bacterioplankton, and phytoplankton in the southeastern Gulf-of-Mexico - distribution and contribution to oceanic DNA pools. *Marine Ecology Progress Series*, 97(1):1–10, 1993.
- [6] G. Bratbak and M. Heldal. Viruses-the new players in the game: their ecological role and could they mediate genetic exchange by transduction? In I. Joint, editor, *Molecular Ecology of Aquatic Microbes*, volume 238, pages 249–264, Berlin, Germany, 1995. Springer-Verlag KG.
- [7] G. Bratbak, M. Heldal, S. Norland, and T. F. Thingstad. Viruses as partners in spring bloom microbial trophodynamics. *Applied and Environmental Microbiology*, 56(5):1400–1405, 1990.
- [8] Jennifer R. Brum and Matthew B. Sullivan. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat Rev Micro*, 13(3):147–159, 03 2015. URL <http://dx.doi.org/10.1038/nrmicro3404>.
- [9] Corina P D Brussaard, Steven W Wilhelm, Frede Thingstad, Markus G Weinbauer, Gunnar Bratbak, Mikal Heldal, Susan A Kimman, Mathias Middelboe, Keizo Nagasaki, John H Paul, Declan C Schroeder, Curtis A Suttle, Dolores Vaqué, and K Eric Wommack. Global-scale processes with a nanoscale drive: the role of marine viruses. *ISME Journal*, 2:575–8, June 2008. ISSN 1751-7370. doi: 10.1038/ismej.2008.31. URL <http://www.ncbi.nlm.nih.gov/pubmed/18385772>.
- [10] Corina P D Brussaard, Jérôme P Payet, Christian Winter, and Markus G Weinbauer. Quantification of aquatic viruses by flow cytometry. In Steve W. Wilhelm and Markus G Weinbauer, editors, *Manual of Aquatic Viral Ecology*, pages 102–109, 2010. ISBN 9780984559107. doi: 10.4319/mave.2010.978-0-9845591-0-7.102.
- [11] Corina P.D. Brussaard, Dominique Marie, and Gunnar Bratbak. Flow cytometric detection of viruses. *Journal of Virological Methods*, 85:175 – 182, 2000. ISSN 0166-0934. doi: [http://dx.doi.org/10.1016/S0166-0934\(99\)00167-6](http://dx.doi.org/10.1016/S0166-0934(99)00167-6). URL <http://www.sciencedirect.com/science/article/pii/S0166093499001676>.
- [12] A. F. Carlucci and D. Pramer. An evaluation of factors affecting the survival of Escherichia-coli in sea water. *Applied Microbiology*, 8(4):251–254, 1960.
- [13] C. Carreira, M. Larsen, R. N. Glud, Brussaard C. P. D., and Middelboe M. Heterogeneous distribution of prokaryotes and viruses at the microscale in a tidal sediment. *Aquatic Microbial Ecology*, 69:183–192, 2013.
- [14] A Certes. Sur la culture, labri des germes atmosphriques, des eaux et des sdiments rapports par les expeditions du travailleur et du talisman 1882 - 1883. *Comptes Rendus Acad Sci*, (98):4, 1884.
- [15] S. Chibani-Chennoufi, A. Bruttin, M. L. Dillmann, and H. Brussow. Phage-host interaction: an ecological perspective. *Journal of Bacteriology*, 186(12):3677–3686, 2004.
- [16] Jessica L. Clasen, Sean M. Brigden, Jerome P. Payet, and Curtis A. Suttle. Evidence that viral abundance across oceans and lakes is driven by different biological factors. *Freshwater Biology*, 53:1090–1100, JUN 2008. ISSN 0046-5070. doi: {10.1111/j.1365-2427.2008.01992.x}.
- [17] W. P. Cochlan, J. Wikner, G. F. Steward, D. C. Smith, and F. Azam. Spatial-distribution of viruses, bacteria and chlorophyll-a in neritic, oceanic and estuarine environments. *Marine Ecology Progress Series*, 92(1-2):77–87, 1993.
- [18] Roberto Danovaro, Cinzia Corinaldesi, Antonio Dell’anno, Jed a Fuhrman, Jack J Middelburg, Rachel T Noble, and Curtis A Suttle. Marine viruses and global climate change. *FEMS Microbiology Reviews*, 35:993–1034, November 2011. ISSN 1574-6976. doi: 10.1111/j.1574-6976.2010.00258.x. URL <http://www.ncbi.nlm.nih.gov/pubmed/21204862>.
- [19] Roberto Danovaro, Antonio Dell’Anno, Cinzia Corinaldesi, Mirko Magagnini, Rachel Noble, Christian Tamburini, and Markus Weinbauer. Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature*, 454:1084–U27, AUG 28 2008. ISSN 0028-0836. doi: {10.1038/nature07268}.
- [20] Marianne De Paepe and François Taddei. Viruses’ life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. *PLoS Biology*, 4:e193, 2006.
- [21] F. d’Herelle. Sur un microbe invisible antagoniste des bacilles dysentériques. *Cr. R. Acad. Sci. Paris*, 165, 1917.
- [22] B Fischer. Bakteriologische untersuchungen auf einer reise nach west- indien. *Zeitschrift fr Hygiene und Infektionskrankheiten*, (1):44, 1886.
- [23] B Fischer. *Die Bakterien des Meeres nach den Untersuchungen der Plankton-Expedition unter gleichzeitiger Bercksichtigung einiger lterer und neuerer Untersuchungen*. Lipsius & Tischer, Keil, 1894.
- [24] P. Forterre, N. Soler, M. Krupovic, E. Marguet, and H.W. Ackermann. Fake virus particles generated by fluorescence microscopy. *Trends in Microbiology*, 21:1–5, January 2013.
- [25] JA Fuhrman and CA Suttle. Viruses in marine planktonic systems. *Oceanography*, 6(2):51–63, 1993.
- [26] S. Hara, K. Terauchi, and I. Koike. Abundance of viruses in marine waters - assessment by epifluorescence and transmission electron-microscopy. *Applied and Environmental Microbiology*, 57(9):2731–2734, 1991.
- [27] M Heldal and G Bratbak. Production and decay of viruses in aquatic environments. *Marine Ecology Progress Series*, 72:205–212, 1991.
- [28] S. C. Jiang and J. H. Paul. Seasonal and diel abundance of viruses and occurrence of lysogeny/bacteriocinogeny in the marine-environment. *Marine Ecology Progress Series*, 104(1-2):163–172, 1994.
- [29] L.F. Jover, T.C. Effer, A. Buchan, S.W. Wilhelm, and

- J.S. Weitz. The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nature Reviews Microbiology*, 12:519–528, 2014.
- [30] Rukina E.A. Kriss A.E. *Bacteriophage in the sea. (In Russian)*, volume 57. Dok Akad Nauk SSSR, 1947.
- [31] R. Maranger and D. F. Bird. Viral abundance in aquatic systems - a comparison between marine and fresh-waters. *Marine Ecology Progress Series*, 121(1-3):217–226, 1995.
- [32] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer. BioNumbers-the database of key numbers in molecular and cell biology. *Nucleic Acids Research*, 38: D750–D753, 2010.
- [33] A. G. Murray and G. A. Jackson. Viral dynamics: a model of the effects of size, shape, motion and abundance of single-celled planktonic organisms and other particles. *Marine Ecology Progress Series*, 89:103–116, 1992.
- [34] J. H. Paul, J. B. Rose, S. C. Jiang, C. A. Kellogg, and L. Dickson. Distribution of viral abundance in the reef environment of key largo, florida. *Applied and Environmental Microbiology*, 59(3):718–724, 1993.
- [35] John H Paul, Sunny C Jiang, and Joan B Rose. Concentration of Viruses and Dissolved DNA from Aquatic Environments by Vortex Flow Filtration. 57:2197–2204, 1991.
- [36] J.P. Payet and C.A. Suttle. To kill or not to kill : The balance between lytic and lysogenic viral infection is driven by trophic status. *Limnology and Oceanography*, 58: 465–474, 2013.
- [37] L. M. Proctor and J. A. Fuhrman. Viral mortality of marine-bacteria and cyanobacteria. *Nature*, 343(6253): 60–62, 1990.
- [38] D. Raoult, S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan, and J. Claverie. The 1.2-megabase genome sequence of mimivirus. *Science*, 306:1344–1350, 2004.
- [39] F. Rohwer and R.V. Thurber. Viruses manipulate the marine environment. *Nature*, 459:207–212, 2009.
- [40] H. L. Russell. Untersuchungen ber im golf von neapel lebende bakterien. *Zeitschrift fr Hygiene und Infektionskrankheiten*, 11:42, 1891.
- [41] D.C. Smith, G.F. Steward, F. Azam, and J.T. Hollibaugh. Virus and bacteria abundance in the drake passage during january and august 1991. *Antarctic Journal of the United States*, 27:125–127., 1992.
- [42] G.F. Steward, A.I. Culley, J.A. Mueller, E.M. Woodcharlson, M. Belcaid, and G. Poisson. Are we missing half of the viruses in the ocean ? *ISME J.*, 7:672–679, 2012.
- [43] C. A. Suttle. Viruses in the sea. *Nature*, 437:356–361, 2005.
- [44] C A Suttle and A M Chan. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. *Marine Ecology Progress Series*, 92:99–109, 1993.
- [45] Curtis A. Suttle. Marine viruses - major players in the global ecosystem. *Nature Reviews Microbiology*, 5: 801–812, OCT 2007. ISSN 1740-1526. doi: {10.1038/nrmicro1750}.
- [46] Curtis A Suttle and Jed A Fuhrman. Enumeration of virus particles in aquatic or sediment samples by epifluorescence microscopy. In S. W. Wilhelm, M. G. Weinbauer, and C. A. Suttle, editors, *Manual of Aquatic Viral Ecology*, pages 145–153, Waco, TX, 2010. American Society of Limnology and Oceanography. ISBN 9780984559107. doi: 10.4319/mave.2010.978-0-9845591-0-7.145.
- [47] T. F. Thingstad. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography*, 45:1320–1328, 2000.
- [48] F. Torrella and R. Y. Morita. Evidence by electron-micrographs for a high-incidence of bacteriophage particles in the waters of yaquina bay, oregon - ecological and taxonomical implications. *Applied and Environmental Microbiology*, 37(4):774–778, 1979.
- [49] T. W. Twort. An investigation on the nature of ultramicroscopic viruses. *Lancet*, 186:1241–1243, 1915.
- [50] MG Weinbauer. Ecology of prokaryotic viruses. *FEMS Microbiology Reviews*, 28:127–181, MAY 2004. ISSN 0168-6445. doi: {10.1016/j.femsre.2003.08.001}.
- [51] M.G. Weinbauer, D. Fuks, and P. Peduzzi. Distribution of viruses and dissolved dna along a coastal trophic gradient in the northern adriatic sea. *Applied and Environmental Microbiology*, 59(12):4074–4082, 1993.
- [52] J. S. Weitz, H. Hartman, and S. A. Levin. Coevolutionary arms races between bacteria and bacteriophage. *Proceedings of the National Academy of Sciences, USA*, 102:9535–9540, 2005.
- [53] Joshua S Weitz and Steven W Wilhelm. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 biology reports*, 4:17, 2012.
- [54] Joshua S Weitz, Timothe Poisot, Justin R. Meyer, Cesar O. Flores, Sergi Valverde, Matthew B. Sullivan, and Michael E. Hochberg. Phage-bacteria infection networks. *Trends in Microbiology*, 21:82–91, 2013.
- [55] Joshua S Weitz, Charles A Stock, Steven W Wilhelm, Lydia Bourouiba, Maureen L Coleman, Alison Buchan, Michael J Follows, Jed A Fuhrman, Luis F Jover, Jay T Lennon, Mathias Middelboe, Derek L Sonderegger, Curtis A Suttle, Bradford P Taylor, T Frede Thingstad, William H Wilson, and K Eric Wommack. A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME J*, 9(6): 1352–1364, 06 2015. URL <http://dx.doi.org/10.1038/ismej.2014.220>.
- [56] Steven W. Wilhelm and Audrey R. Matteson. Freshwater and marine viroplankton: a brief overview of commonalities and differences. *Freshwater Biology*, 53:1076–1089, JUN 2008.
- [57] K.E. Wommack and R.R. Colwell. Viroplankton: Viruses in aquatic ecosystems. *Microbiology and Molecular Biology Reviews*, 64:69–114, MAR 2000.
- [58] K.E. Wommack, R.T. Hill, M. Kessel, E. Russekcohen, and R.R. Colwell. Distribution of viruses in the chesapeake bay. *Applied and Environmental Microbiology*, 58(9):2965–2970, 1992.
- [59] Claude Ephraim ZoBell. *Marine microbiology, a monograph on hydrobacteriology*. Marine microbiology. A monograph on hydrobacteriology. 1946.

Study Name	Lab	Study Type	Location	Regime	Citation
NORTHSEA2001	Bratback	Spatial	North Sea	Coastal	Bratback (unpublished)
RAUNEFJORD2000	Bratback	Temporal	North Sea	Coastal	Bratback (unpublished)
BATS	Breitbart	Temporal	Sargasso Sea	nonCoastal	Parson et al. 2011
STRATIPHYT1	Brussaard	Spatial	N-Atlantic Transect	nonCoastal	Mojica et al., 2015
STRATIPHYT2	Brussaard	Spatial	N-Atlantic Transect	nonCoastal	Brussaard (unpublished)
USC MO	Fuhrman	Temporal	Santa Barbara Channel	nonCoastal	Fuhrman et al. 2006
GEOTRACES	Herndl	Spatial	Atlantic Transect	nonCoastal	De Corte et al. 2012
GEOTRACES.LEG3	Herndl	Spatial	Atlantic Transect	nonCoastal	Herndl (unpublished)
BEDFORDBASIN	Li	Temporal	North Atlantic Ocean	Coastal	Li and Dickie 2001
GREENLAND 2012	Middleboe	Spatial	Greenland Sea	nonCoastal	Middleboe (unpublished)
INDIANOCEAN2006	Middleboe	Spatial	Indian Ocean	nonCoastal	Middleboe (unpublished)
KH04-5	Nagata	Spatial	Southern Pacific Ocean	nonCoastal	Yang et al. 2014
KH05-2	Nagata	Spatial	Northern Pacific Ocean	nonCoastal	Yang et al. 2014
CASES03-04	Suttle	Spatial	Arctic Ocean	nonCoastal	Payet and Suttle, 2013
SOG/ELA/TROUT/SWAT	Suttle	Temporal	Pacific Ocean - Strait of Georgia	Coastal	Clasen et al., 2008
ARCTICSBI	Wilhelm	Spatial	Gulf of Alaska	Coastal	Balsom, MS thesis 2003
FECYCLE2	Wilhelm	Spatial	South Pacific Ocean	nonCoastal	Matteson et al. 2012
FECYCLE1	Wilhelm	Spatial	South Pacific Ocean	nonCoastal	Stzepek et al. 2005
MOVE	Wommack	Temporal	Atlantic - Chesapeake	Coastal	Wang et al. 2011
NASB2005	Wilhelm	Spatial	North Atlantic Ocean	nonCoastal	Rowe et al. 2008
POWOW	Wilhelm	Spatial	Pacific Ocean	nonCoastal	Wilhelm (unpublished)
TABASCO	Wilhelm	Spatial	South Pacific Ocean	nonCoastal	Wilhelm et al. 2003

TABLE S1: Virus and microbial abundance data from 25 different marine virus abundance studies from 11 different lab groups. A total of 5671 data points were aggregated. The data collection dates range from 2000 to 2011. Due to sampling convenience, data primarily comes from coastal waters in the northern hemisphere and were collected predominately during the summer months, with the notable exceptions of long-term coastal monthly monitoring sites (USC MO, BATS, Chesapeake Bay).

Study	≤ 100m	> 100m	Total
ARCTICSBI	292	0	292
BATS	626	756	1382
BEDFORDBASIN	188	0	188
CASES03-04	199	46	245
ELA	85	0	85
FECYCLE1	31	0	31
FECYCLE2	15	0	15
GEOTRACES	141	631	772
GEOTRACES_LEG3	78	351	429
GREENLAND2012	78	46	124
INDIANOCEAN2006	42	10	52
KH04.5	159	383	542
KH05.2	117	238	355
MOVE	84	0	84
NASB2005	31	0	31
NORTHSEA2001	164	27	191
POWOW	9	0	9
RAUNEFJORD2000	95	0	95
SOG	67	0	67
STRATIPHYT1	89	24	113
STRATIPHYT2	59	34	93
SWAT	31	0	31
TABASCO	12	0	12
TROUT	47	0	47
USC MO	182	204	386
Total	2921	2750	5671

TABLE S2: Number of data points per study.

Study	Intercept	Std. Error	Group
ARCTICSBI	4.552594	0.1580157	A
FECYCLE1	4.552594	0.1837236	A
FECYCLE2	4.552594	0.1961546	A
MOVE	4.552594	0.1982541	A
Raunefjord	4.552594	0.1716625	A
StratiphytI	4.552594	0.1543207	A
USCMO	4.552594	0.1806831	A
KH05_2	4.513041	0.1683358	A
SOG	4.480697	0.1869220	A
POWOW	4.408948	0.2060002	B
StratiphytII	4.389658	0.1765001	B
KHO4	4.339907	0.1688285	B
CASES0304	4.339902	0.1669256	B
BEDFORDBASIN	4.336256	0.1825271	B
BATS	4.332784	0.1647363	B
ELA	4.332784	0.1885707	B
GEOTRACES	4.332784	0.1580225	B
GEOTRACES_LEG3	4.332784	0.1679548	B
GREENLAND2012	4.332784	0.1759093	B
INDIANOCEAN2006	4.332784	0.1820002	B
NASB2005	4.332784	0.1876671	B
NORTHSEA2001	4.332784	0.1770218	B
SWAT	4.332784	0.1945882	B
TABASCO	4.332784	0.2047635	B
TROUT	4.332784	0.2019569	B

TABLE S3: Variation in the estimate of the intercept, $\alpha_0^{(i)}$, for each study and associated standard error for the constrained power-law model as applied to surface ocean data. The common intercept in this model is $\alpha_0 = 4.44$ and the common slope is 0.42. The group column denotes whether the study-specific intercept exceeds that of the common intercept (denoted as group A) or is below that of the common intercept (denoted as group B). The table is sorted according to the lab-specific intercept estimates.

Study	≤ 100 m		> 100 m	
	R^2	p -value	R^2	p -value
ARCTICSBI	0.441	<1e-05		
BATS	0.045	<1e-05	0.504	<1e-05
BEDFORDBASIN	0.537	<1e-05		
CASES03-04	0.541	<1e-055	0.072	0.0718
ELA	0.343	<1e-05		
FECYCLE	0.146	0.0341		
FECYCLE2	0.004	0.813		
GEOTRACES	0.163	<1e-05	0.706	<1e-05
GEOTRACES_LEG3	0.043	0.0695	0.396	<1e-05
GREENLAND2012	0.868	<1e-05	0.333	2.7e-05
INDIANOCEAN2006	0.068	0.0955	0.288	0.11
KH04.5	0.325	<1e-05	0.703	<1e-05
KH05.2	0.122	0.000112	0.836	<1e-05
MOVE	0.24	<1e-05		
NASB2005	0.382	0.00021		
NORTHSEA2001	0.542	<1e-05	0.51	2.85e-05
POWOW	0.136	0.329		
RAUNEFJORD2000	0.349	<1e-05		
SOG	0.788	<1e-05		
STRATIPHYT1	0.448	<1e-05	0.471	0.000214
STRATIPHYT2	0.768	<1e-05	0.731	<1e-05
SWAT	0.026	0.389		
TABASCO	0.371	0.0354		
TROUT	0.687	<1e-05		
USC MO	0.229	<1e-05	0.462	<1e-05

TABLE S4: Explanatory power and significance of power-law fits for the model in which the power-law exponent is allowed to vary between studies. Empty cells in a row denote the absence of samples collected at depths > 100 m for the study denoted in the left-most column.

Study	≤ 100 m		> 100 m	
	α_0	α_1	α_0	α_1
ARCTICSBI	2.13	0.97		
BATS	4.81	0.31	2.49	0.72
BEDFORDBASIN	1.32	0.91		
CASES03-04	2.40	0.77	2.80	0.65
ELA	2.38	0.66		
FECYCLE	1.29	1.05		
FECYCLE2	5.36	0.38		
GEOTRACES	4.40	0.41	3.63	0.52
GEOTRACES_LEG3	4.09	0.45	3.43	0.53
GREENLAND2012	0.98	0.97	2.05	0.76
INDIANOCEAN2006	4.97	0.28	2.75	0.66
KH04.5	4.04	0.48	3.00	0.64
KH05.2	4.66	0.40	2.48	0.76
MOVE	5.06	0.45		
NASB2005	1.80	0.69		
NORTHSEA2001	0.74	1.00	1.59	0.84
POWOW	5.14	0.30		
RAUNEFJORD2000	4.30	0.48		
SOG	-0.68	1.25		
STRATIPHYT1	3.42	0.71	4.53	0.45
STRATIPHYT2	2.93	0.68	2.96	0.67
SWAT	6.36	0.11		
TABASCO	3.73	0.49		
TROUT	1.83	0.78		
USC MO	4.37	0.49	2.18	0.79

TABLE S5: Power-law exponents, α_1 , and intercepts, α_0 , for each study from the mixed model allowing study-specific slopes and intercepts. Empty cells in a row denote the absence of samples collected at depths > 100 m for the study denoted in the left-most column.

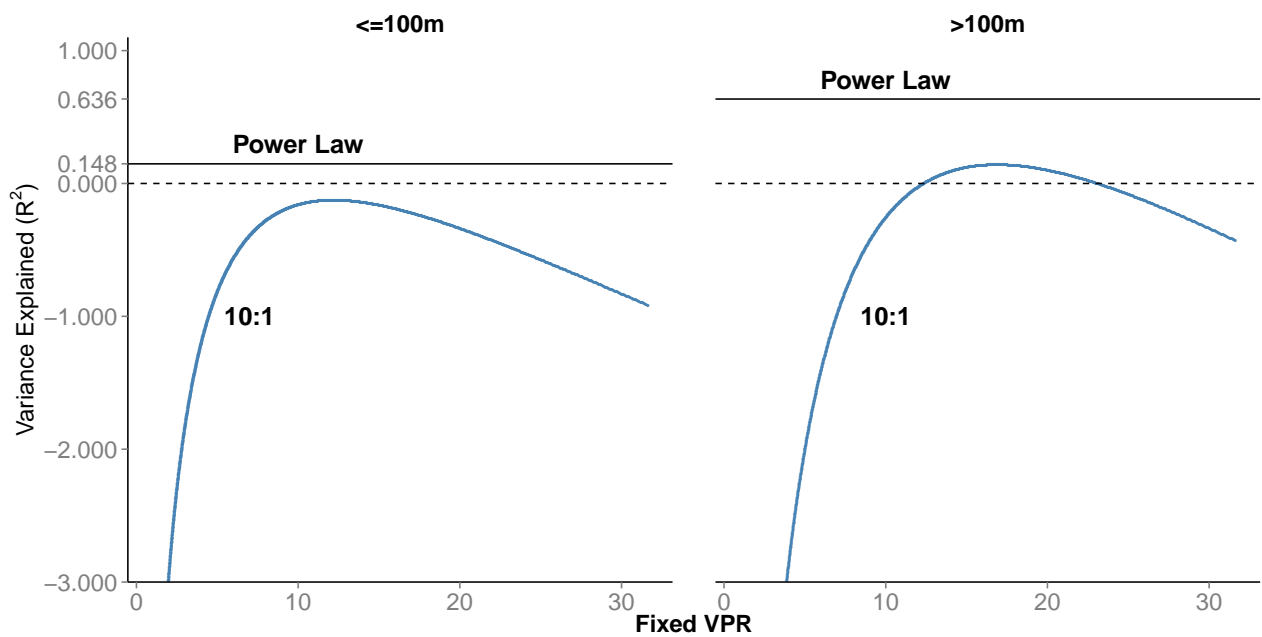


FIG. S1: Explanatory power of fixed VPR models in the surface ocean (left) and deeper water column (right). The x-axis denotes the value r in the model $V = rP$ where V denotes virus abundance and P denotes prokaryote abundance. The y-axis denotes the fraction of variance explained, R^2 . Here, $R^2 = 1 - \text{SSE}_{\text{model}}/\text{SSE}_{\text{total}}$ where $\text{SSE}_{\text{model}}$ is the sum of squared errors for the model and $\text{SSE}_{\text{total}}$ is the sum of total squared errors.

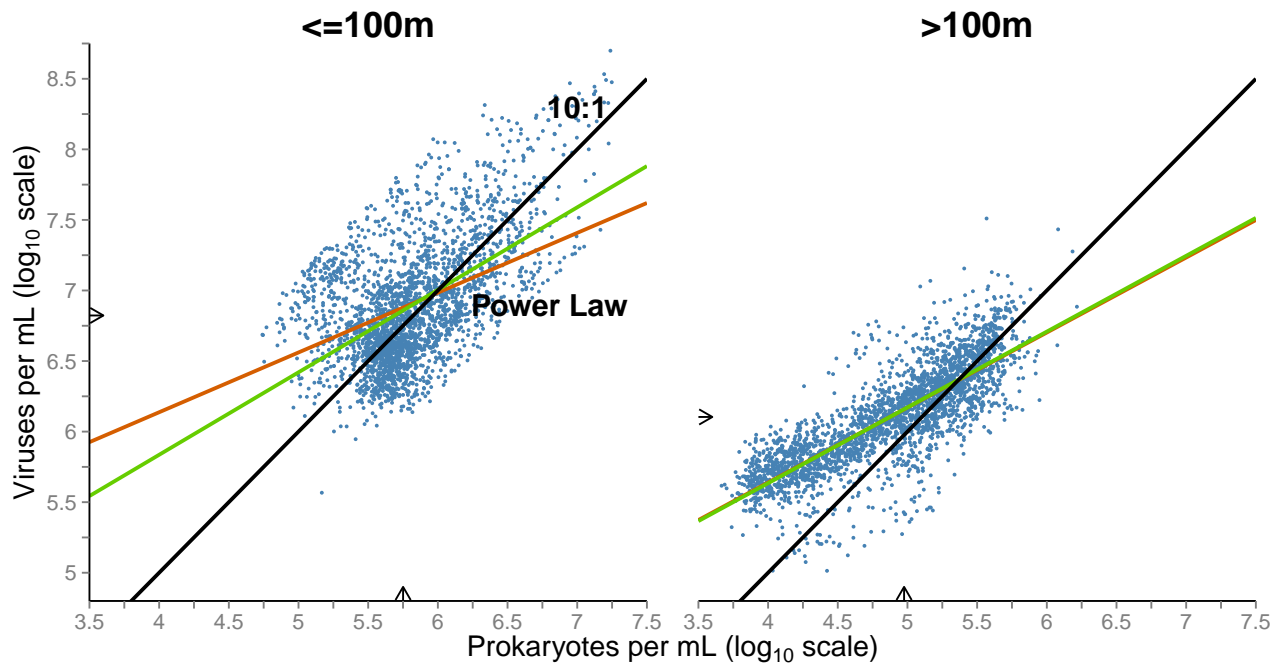


FIG. S2: Explanatory power of fixed VPR models in the near-surface and sub-surface with and without outliers. The three lines in each panel denote the 10:1 line (black), power-law fit (red) and power-law fit when removing outliers (green). The R^2 value for the power law fit for surface data excluding outliers is 0.30, has a slope of 0.58 and an intercept of 3.50. The R^2 value for the power law fit for sub-surface data excluding outliers is 0.65, has a slope of 0.54 and an intercept of 3.49.

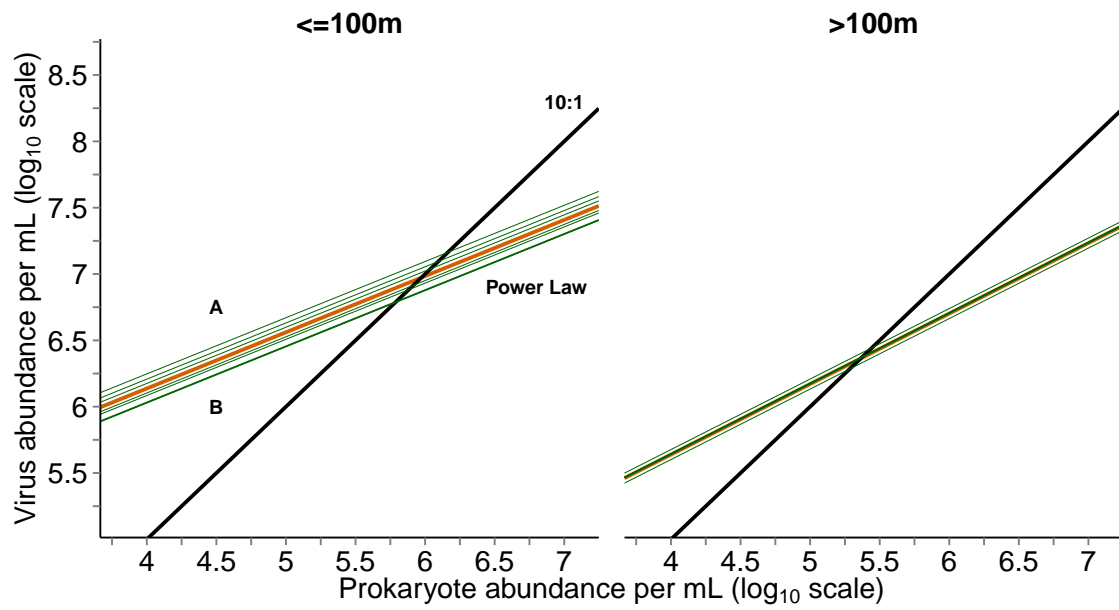


FIG. S3: Constrained regression model for samples taken at depths $\leq 100\text{m}$ (left) and $> 100\text{m}$ (right) where the intercept for each study was permitted to vary (see Materials and Methods). Blue line denotes the 10:1 relationships, the red line denotes the best-fitting power-law model, and the remainder of lines denote the variable intercept model with intercept values reported in Table S3.

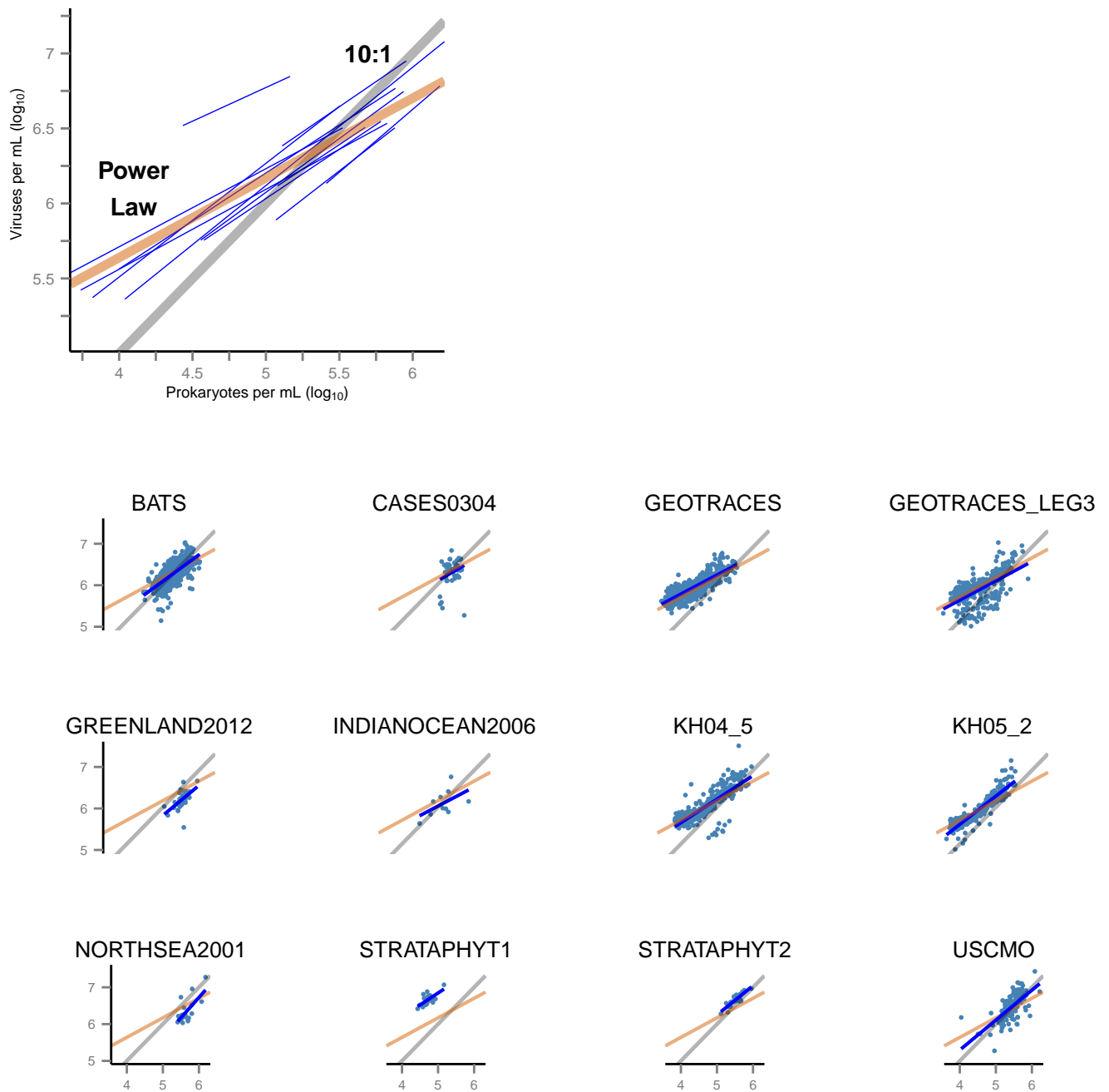


FIG. S4: Virus-prokaryote relationships given the variable slope and intercept mixed-effects model for samples taken at depths greater than 100m. (Upper-left) Best-fit power-law for each study (blue lines) plotted along with the best-fit power-law of the entire dataset (red line) and the 10:1 line (grey line). (Individual panels) Best-fit power-law model (blue line) on log-transformed data (blue points) for each study, with the power-law model regression (red) and 10:1 line (black) as reference. The power-law exponents and associated confidence intervals are shown in Figure S5,

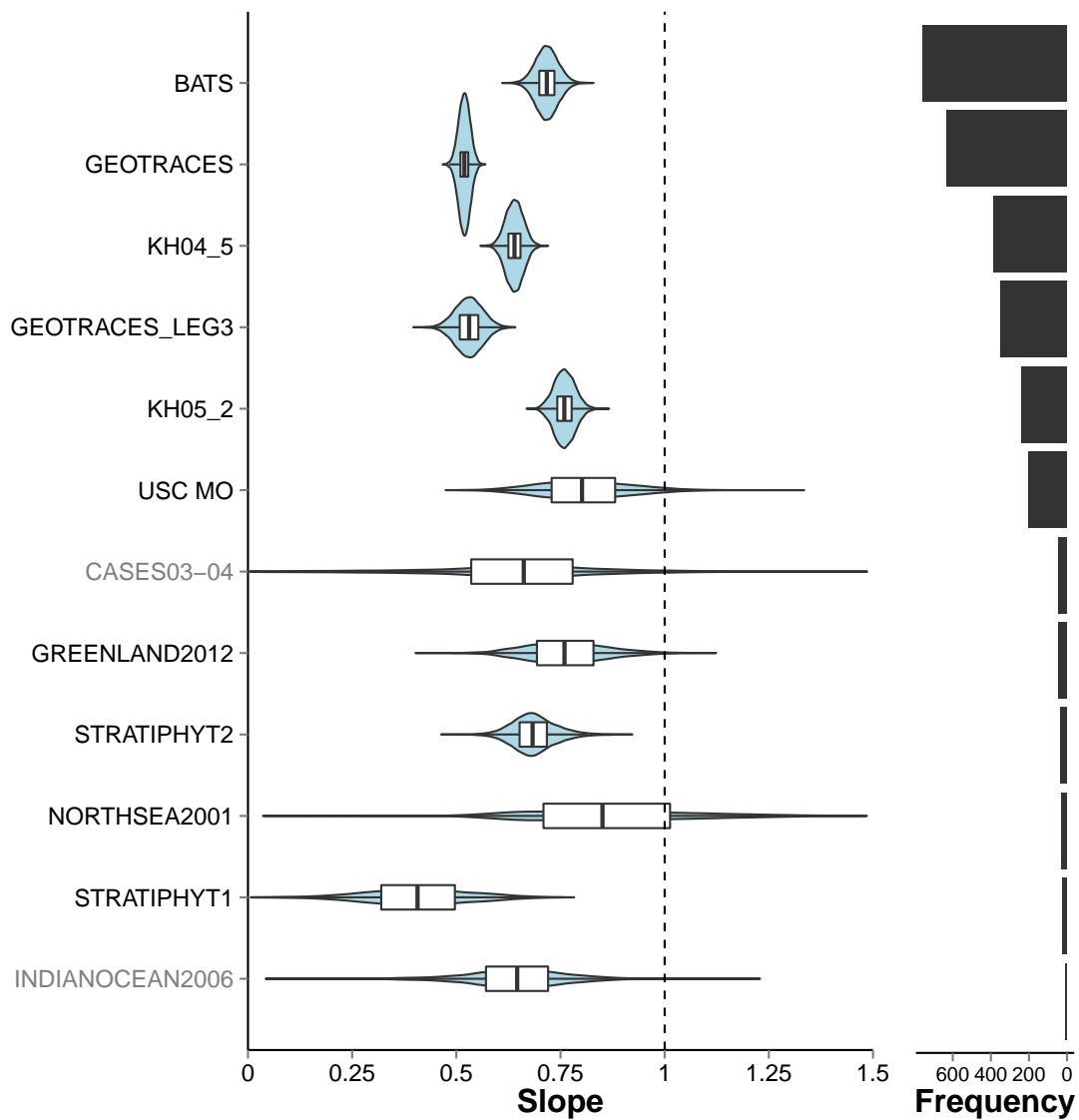


FIG. S5: Study-specific 95% confidence intervals of power-law exponents for relationships between virus and prokaryote abundance from samples taken at depths greater than 100m. The confidence intervals are plotting using “violin” plots including the median (center black line), 75% distribution (white bars) and 95% distribution (black line), with the distribution overlaid (blue shaded area). The number of points included as part of each study is displayed on the right-most bar plots. Study labels in black indicate those studies whose linear regression had a p-value less than .05/12 while labels in gray indicate a p-value above this threshold.