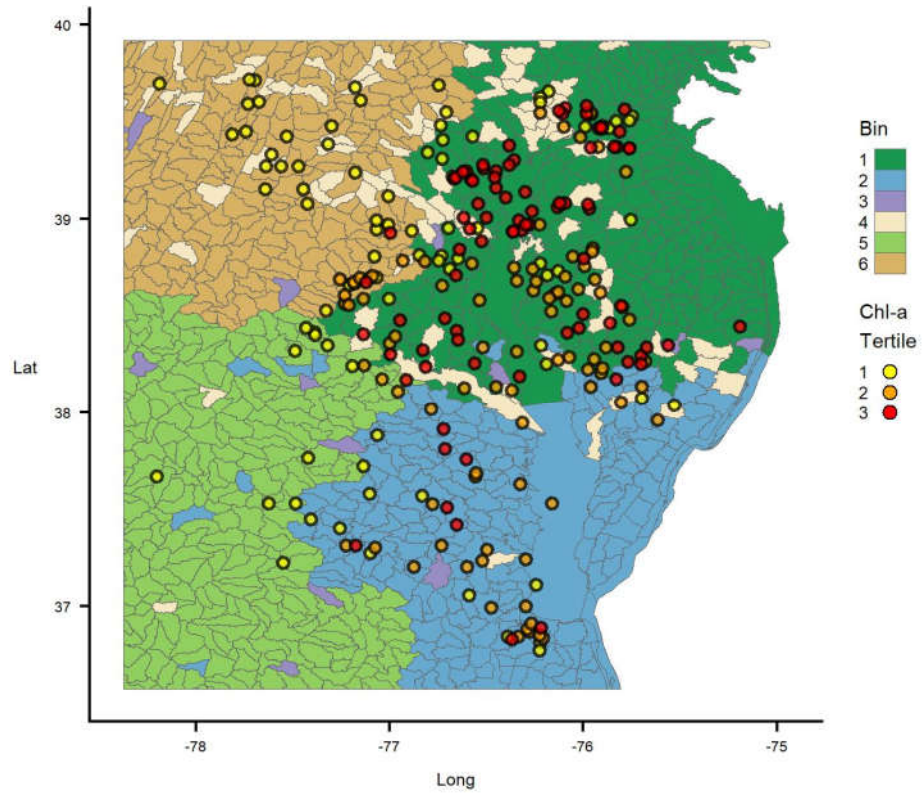
Chesapeake Bayes: Modeling Water Quality with Heterogeneous Data

Pasky Pascual* and Kerry Kuntz±

* U.S. Environmental Protection Agency (Retired)

± U.S. Environmental Protection Agency



ABSTRACT. Historically, total nutrient loads were understood to determine overall autotrophic growth in Chesapeake Bay. However, the speciation and ratio of individual nutrients has since been shown to significantly influence phytoplankton biomass and diversity. We present evidence that the ratio of nitrate to ammonium strongly affects chlorophyll *a* concentrations, more so than concentrations of total nitrogen, or nitrate and ammonium taken alone. Moreover, the ratio's effect may be attributed to the preferences of different phytoplankton taxa and their dynamic competition for particular nitrogen forms.

INTRODUCTION

Chesapeake Bay, the largest estuary in the U.S., is bounded by a watershed spanning 64,000 square miles across six states and the District of Columbia (*1*). Its land to water ratio is 14:1, the most extensive for any coastal watershed in the world (*2*). Home to 18 million people, the watershed is a hub for production, agriculture and development (*3*). With so many people living on a land mass encroaching so significantly on a body of water, it is no wonder the estuary was dubbed by the Washington Post (*4*) as "one of the dirtiest."

The bay's nutrient pollution problems primarily stem from excess nitrogen and phosphorous loading from point and non-point sources (*3*). When excess nutrients from fertilizers, sewage treatment plants and stormwater runoff permeate a watershed, they become available for uptake by algae, which begin to grow excessively (*5*). When excess algae die and decompose, oxygen in the water column is depleted, leading to hypoxia, dead zones, and fish kills (*6*).

Researchers have explained these eutrophic systems with Liebig's law of minimum, stating that plant growth is controlled by the scarcest resource, or limiting factor (*7*). When neither macronutrients nor micronutrients is limiting, algae growth is overstimulated. The law of the minimum has motivated mitigation efforts aimed at reducing the limiting nutrient to decelerate eutrophication and improve water quality.

However, nutrient loading only partially explains the pattern of algae abundance in water columns. While *total* nutrient load prevails over autotrophic growth, the *ratio* of individual nutrients significantly influences phytoplankton biomass (*8*). Nutrient

stoichiometry was first noted in 1933 by the oceanographer Alfred Redfield, who observed that the ratio between nitrogen and phosphorous atoms in plankton cells (**N:P**) was unexpectedly similar to that of dissolved, inorganic nitrate and phosphate in marine waters (*9*). Today, it is generally acknowledged that Redfield's ratio represents N:P averaged over variable environmental conditions and phylogenetic groups (*10*). Further, it is but one manifestation of the theory that stoichiometry exists because a nutrient's concentration in the water column depends not only on its availability in the environment, but also on aquatic organisms' abilities to uptake and assimilate it (*11*). Under varying ecological conditions wherein communities of autotrophs live in waters suffused with various nutrients, different species of phytoplankton respond to each resource differently, according to their own nutritional needs, modulating and being modulated by nutrient stoichiometries in the water column (*8*).

Beyond nutrient ratios, researchers have reported that stoichiometry in different *forms* of nutrients also influence phytoplankton growth. Specifically, the preference of cyanobacteria and diatoms, respectively, for ammonium (NH_4^+) and nitrate (NO_3^-) has been reported in field studies at multiple locations ((*12*), (*13*), (*14*)).

This difference in N-form preferences has been used to explain the counterintuitive observation referred to as the "ammonium paradox" (*15*). It is a broadly held principle that phytoplankton expend less energy taking up NH_4^+ over other N-forms (*16*), and should therefore theoretically thrive in an environment rich in NH_4^+. Despite this supposition, field experiments in the San Francisco Bay-Delta showed higher NH_4^+ linked to diminished algae growth ((*17*), (*18*), (*19*)).

To explain this paradox, Glibert offers several theories (*20*) describing how the

ratio between nitrate and ammonium (**N:A**) affects the composition of a phytoplankton community and therefore, phytoplankton biomass. Essentially, the relative concentrations of reduced (NH_4^+) versus oxidized (NO_3^-) forms of nitrogen in the water column elicit a variety of cellular responses within the individual organism, the cumulative effects of which eventually cascade upwards to the community level. Focusing solely on energetics as the regulatory mechanism for stoichiometry ignores the cellular pathways organisms have opportunistically evolved to maintain homeostasis in the face of variable environments (*21*).

Among cyanobacteria, NH_4^+ is the preferred N-source (*22*), with transportation and assimilation generally unaffected by NO_3^- levels (*20*). For phytoplankton to assimilate N into proteins, it must be in reduced form, where it serves as a substrate inside the cell for glutamine synthetase, which fixes NH_4^+ into glutamine; glutamine is then synthesized into other amino acids (*23*). Researchers previously supposed that cyanobacteria absorbed NH_4^+ through simple diffusion, but it is now apparent that aquaporins—proteins in the cell membrane conducting water and solutes into and out of the cell—must actively maintain intracellular NH_4^+ at sufficient concentrations for glutamine synthetase to operate effectively (*23*).

Dahm reports that, among diatoms, the preferred N-source is NO_3^-, even in the presence of high NH_4^+ levels (*24*), probably because they have evolved cellular transporters with a high affinity for NO_3^- (*25*). Studies show that elevated NH_4^+ inhibits both transport of NO_3^- into phytoplankton cells, as well as assimilation of NO_3^- within cells (*26*). The inhibition of NO_3^- by NH_4^+ appears to occur at both transcriptional and posttranscriptional phases, although phytoplankton seem to be less

sensitive to NH_4^+ alone than to N:A (*26*).

The contrasting preferences for NH_4^+ and NO_3^- by, respectively, cyanobacteria and diatoms, is consistent with their evolutionary development. Cyanobacteria evolved under anaerobic conditions. They are believed to be the only organisms to have organically evolved the ability to photosynthesize, other photoautotrophs having gained this ability through endosymbiosis with cyanobacteria (*27*). They are further believed to have caused the so-called "Great Oxidation Event," that period between 2.45-2.32 billion years ago in which atmospheric oxygen levels rose to their current levels (*28*). Diatoms evolved more recently, about one billion years ago, from the endosymbiosis of red alga and an eukaryotic heterotroph (*29*), in an oxygenated atmosphere that would therefore be more hospitable to N's oxidized forms.

Whatever the specific mechanisms for homeostasis, Sperfeld *et al.* propose it is rooted in the fitness and natural selection of organisms that have evolved metabolic pathways to uptake and assimilate nutrients to maintain their internal composition in a changing environment (*21*). Further, the aggregate, cumulative effects of these subcellular interactions with the environment cascade upwards, until they are ultimately manifest at individual, at community, and at large, ecosystem scales.

The large, ecosystem scale of the Chesapeake Bay is this paper's setting. At this scale, the influence of N:A on phytoplankton growth is further complicated by fluctuations in the relative concentrations of NH_4^+ and NO_3^- within the watershed's network of streams. As opposed to being determined solely by terrestrial input, N:A is dynamic, subject to in-stream transformations, which can sometimes occur within minutes to hours of loading (*30*). In turn, in-stream transformations are facilitated by

dynamic processes—the net result of biological activity, other nutrients, and a stream's physical attributes, such as depth and velocity.

Previous studies have examined nutrient uptake on large spatial and temporal scales, across stream gradients and landscape types ((*31*), (*32*), (*33*)). There have also been numerous studies investigating nutrient dynamics and algae proliferation, such as those examining N:A's impact on the growth of harmful algal species ((*34*), (*20*)). To our knowledge, no one has investigated the nexus between these topics, as we do here.

We derived our data from multiple sources, which introduces heterogeneity and constrains our ability to definitively establish causal relationships. Nonetheless, the data include measurements of water quality indicators, phytoplankton counts, and land attributes over multiple seasons, years, and watersheds. As such, they provide us with the opportunity to holistically investigate biogeochemical patterns underlying eutrophic systems around the Chesapeake.

METHODS

Data. We used three broad types of raw data in this study. The first two, measurements of water nutrients and counts of phytoplankton taxa, were collected at monitoring stations around the Chesapeake Bay. The third are data on elevation, impermeable surfaces, and land use derived from rasters of satellite imagery and from models. The dataset of water nutrients consists of 15,922 observations gathered over 11 years (from 2005-2015) from 260 monitoring stations in 165 watersheds, defined at the 12-digit Hydrologic Unit Code (**HUC12**) level. Spatial distribution of the monitoring stations is shown in Figure 2. The phytoplankton dataset consists of 2,219 observations collected

from 41 monitoring stations in 27 HUC12s over 8 years (from 2005-2012). The data and their provenance are summarized in Table 1.

Table 1. Description and source of data used in this study

Data	Description	Source and Metadata
Water nutrients	Concentrations (µg/L) of chlorophyll *a*, nitrate, ammonium, particulate carbon, dissolved oxygen, total phosphorous, and total nitrogen.	http://datahub.chesapeakebay.net/WaterQuality Metadata: https://www.chesapeakebay.net/documents/3676/wq_data_userguide_10feb12_mod.pdf
Phytoplankton	Counts of individuals in various phytoplankton taxa	http://datahub.chesapeakebay.net/LivingResources
Land Use and Impermeable Surfaces	Raster based on satellite imagery	https://www.sciencebase.gov/catalog/item/5a1c31b9e4b09fc93dd63982
Elevation	Raster based on satellite imagery	https://www.sciencebase.gov/catalog/item/559e96ffe4b0b94a64018fb4
Base Flow Index	Raster based on base flow model	https://water.usgs.gov/GIS/metadata/usgswrd/XML/nhd_bfi.xml

Processing point data. We log-transformed data on nutrient concentrations, replacing zero values with 10^{-4}, filtering data to include only those from monitoring stations with at least three observations per season per year. We binned observations into seasons, defining these as spring (March-May), summer (June-August), fall (September-November), and winter (December-February). We discarded outliers using a Cleveland plot, as prescribed by Zuur *et al.* (*35*).

Raw phytoplankton data were identified at various taxonomic levels, including a group, labelled "blue green sphere", used to refer to miscellaneous cyanobacteria. To organize the data, we first categorized observations by phylum, based on taxonomy in the online source AlgaeBase (*http://www.algaebase.org*). These phyla included *bacillariophyta, chlorophyta, cyanobacteria, dinoflagellata, euglenophyta,*

heterokontophyta, miozoa, ochrophta, and *sarcomastigophora*. Because some phyla, such as *heterokontophyte*, contain a wide range of species, we looked at specific species within these phyla and grouped them by common name, hereafter referred to as functional groups. For example, *bacillariophyta* were reclassified using their common name, diatom. Ultimately, we established four such functional groups: cyanobacteria, diatom, dinoflagellate, and chlorophyte. Phytoplankton types with negligible counts, such as those in the phylum *sarcomastigophora*, were eliminated from the analysis.

For each record, we calculated the proportion of each taxa out of the total phytoplankton count. We then merged the phytoplankton and water nutrients dataset by linking observations taken at the same monitoring site and on the same day, month, and year.

Processing spatial data. We manipulated all spatial data using various R GIS packages ((*36*),(*37*),(*38*),(*39*)). After merging all original GeoTiff rasters into a regional raster covering the Chesapeake Bay area, we generated multiple, smaller rasters by cropping the regional raster around the shapefile defined for each HUC12. We converted these into matrices of pixel counts at the HUC-12 level. For each of these matrices, we counted the number of pixels corresponding to a pixel category or numeric index and estimated the proportion of these counts out of the total number of pixels. These proportions served as an indicator of the proportion of land in a HUC12 associated with a pixel category or index.

For the discrete categories underlying the land use raster, we relied on the Anderson Land Use/Land Cover Classification scheme to categorize the raster pixels (*40*). For the rasters with underlying indices measured on a continuous scale (elevation,

impervious surface, and base flow), we converted these indices into discrete categories by creating tertiles from the continuous scale, and calculated the proportion of each HUC12 in each tertile.

Spatial bins based on land attributes and proximity. To establish spatial bins, we identified the land attributes most closely related to chlorophyll *a* and N:A, performing a canonical coefficient analysis (**CCA**) across all HUC12s (*41*). Using these attributes, we conducted a cluster analysis based on Ward hierarchical clustering, constrained by the geographical proximity of centroids of each HUC12 polygon (*42*). To summarize and visualize the dominant attributes in each spatial bin or cluster, we developed an index ranging from 0-1 for each attribute, using min-max normalization (*43*).

Nutrient level comparisons based on group-level differences. To compare differences in the levels of selected nutrients in different groups, we relied on Bayesian Analyses of Variance (**BANOVA**), a form of hierarchical, Bayesian (**HB**) modeling ((*44*), (*45*)). We conducted a BANOVA in three analyses to evaluate differences: (1) in chlorophyll *a* concentrations among spatial bins; (2) in N:A among spatial bins; and (3) in chlorophyll *a* in bins defined spatially, seasonally, and by N:A groups (see next section). Model details can be found in the Supporting Information online.

N:A groups based on model's inflection points. To investigate the relationship between phytoplankton composition and chlorophyll *a*, we used General Additive Models (**GAMs**) (*46*). We developed GAMs in three analyses. In the first two, we analyzed the phytoplankton dataset. First, we evaluated the relationships between N:A and diatom and cyanobacteria counts; and second, the relationship between N:A and

chlorophyll *a*. In the third application, we used GAM on the water nutrient dataset to analyze the relationship between N:A levels and chlorophyll *a*. In all three applications, we used the GAM inflection points to define three N:A groups of high, medium, and low N:A values.

Nutrient influence on chlorophyll *a* based on group-level differences. We used HB regression to estimate predictor coefficients, allowing these to vary based on groups to which individual observations belonged (*47*). We estimated parameters at both the population level, as well as at each combination of spatial bin and season. Model details are available in the Supplemental Information online.

RESULTS

Spatial bins. The CCA indicates there are two statistically significant canonical functions (Table 2), based on Wilk's lambda test (*41*). These functions show seven attributes correlate with chlorophyll *a* and N:A that have absolute values of structure coefficients greater than 0.40. These are low elevation, elevation variability, elevation range, and proportion in each HUC12 of water (*i.e.* ponds and lakes), wetlands, agriculture, and forests (Table 3). Elevation variability is the variation in elevation within the HUC12; elevation range is the difference between minimum and maximum elevations.

Table 2. Wilk's test for Significance of Canonical functions in Canonical Coefficient Analysis, showing both function are statistically significant with *p*-values < 0.05.

Wilks L	F	df1	df2	p
0.24	11.80	26	300	0.00
0.86	2.10	12	151	0.02

Table 3. Canonical Coefficient Analysis Standardized Weights and Structure Coefficients. Highlighted covariates have structure coefficients > 0.40. For chlorophyll *a*, green and yellow highlights indicate, respectively, positive and negative correlations. For N:A, the reverse is true.

	Standardized Weights		Structure Coefficients	
	Function 1	Function 2	Function 1	Function2
Dependent Variables				
N:A	0.89	0.90	0.99	0.13
Chlorophyll *a*	-0.16	1.26	-0.71	0.70
Covariates				
Low Elevation	-0.31	0.85	-0.84	0.22
Elevation Range	1.40	0.78	0.83	-0.15
Water	-0.40	-0.33	-0.65	-0.12
Elevation Variability	-0.93	-0.72	0.64	-0.15
Wetland	-0.12	-0.44	-0.51	-0.01
Agriculture	-0.09	0.47	0.51	0.43
Forest	-0.30	-0.74	0.30	-0.44
High Elevation	-0.02	0.10	0.28	-0.13
Low Base Flow	0.11	0.39	0.15	0.10
Low Impermeable Surface	0.69	2.60	-0.09	-0.08
Developed	0.50	6.53	0.07	0.12
High Impermeable Surface	-0.17	-4.25	0.06	0.10
High Base Flow	0.02	0.41	-0.01	0.27

The K-means cluster analysis based on these attributes yielded six spatial bins (Figure 1). Because the water and wetland attributes were highly related, we dropped the latter. We then conducted a Ward hierarchical clustering, constrained by geographical proximity. In such a clustering, one has the option of choosing how much weight to allocate values based on land attributes versus geographical proximity. Based on a scree plot, we assigned 90% weight to the former. The geographical distribution of the spatial bins, along with location of the monitoring stations and their average chlorophyll *a* levels broken into tertiles, is shown in Figure 2.

Figure 1. Dominant land attributes in each spatial bin

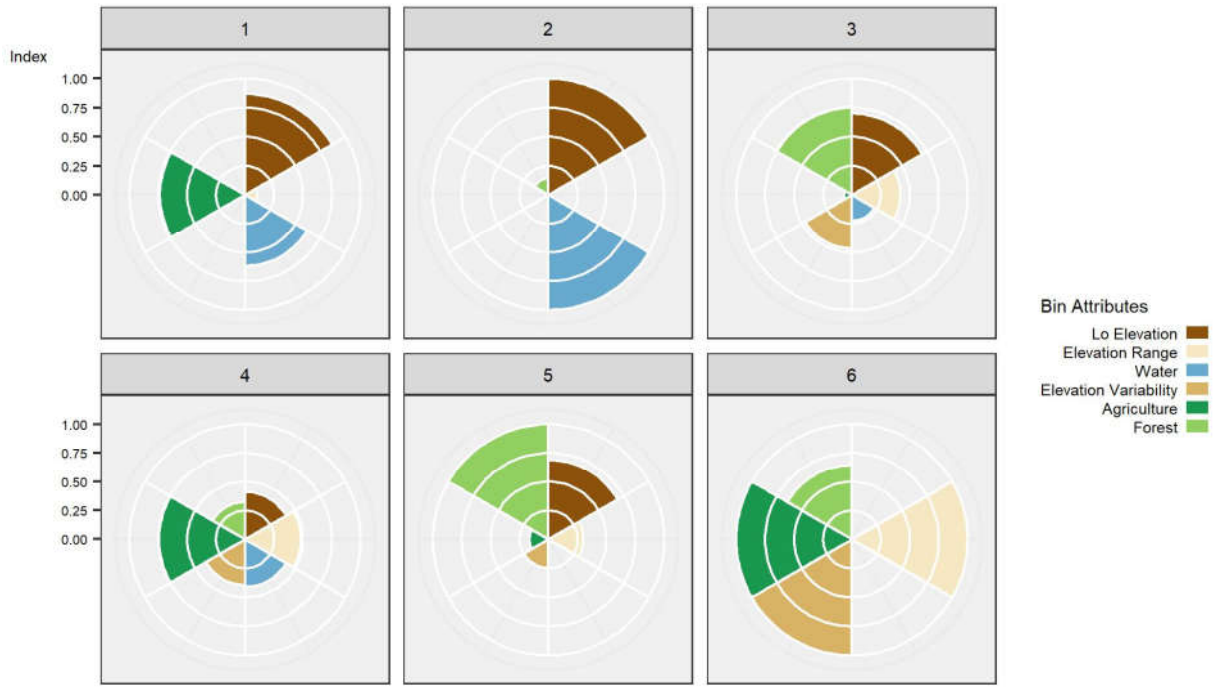
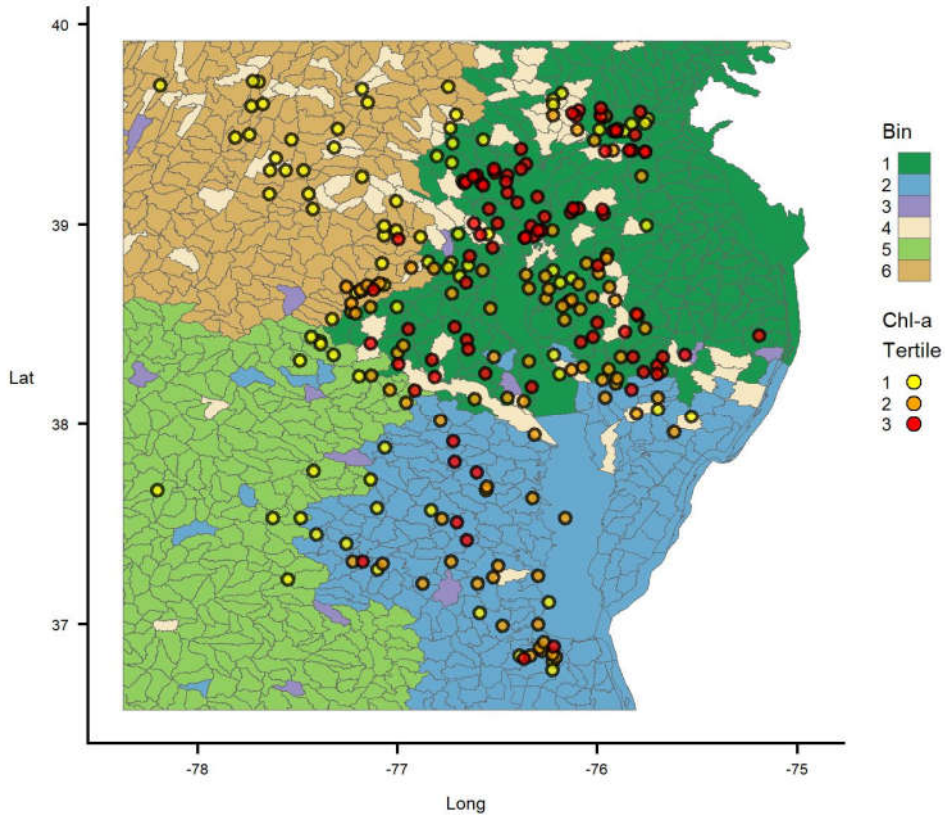
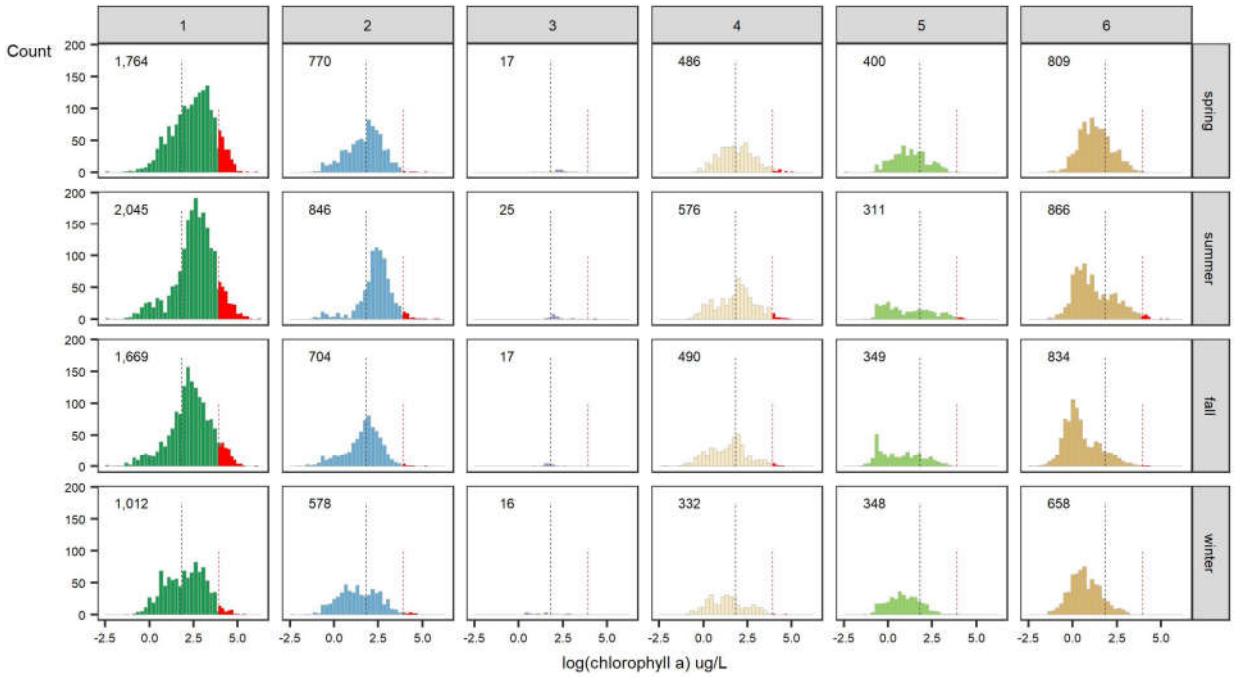


Figure 2. Distribution of chlorophyll *a* levels in monitoring stations around Chesapeake Bay, each station within a spatial bin defined by land attributes.



Nutrient levels. Having clustered the all HUC12s into six spatial bins, we did a frequency count of average chlorophyll *a* concentrations within each bin and season combination (Figure 3).

Figure 3. Histogram showing counts of log(chlorophyll *a*) in each spatial bin and season. Numbers within each cell is the sample size. Taller, dark dashed line represents the grand mean (1.8 µg/L). Shorter, red dashed line represents 3.9 µg/L or log(50), where 50 µg/L is the water quality goal for chlorophyll *a* in Maryland jurisdictions (*48*).



We also conducted three BANOVAs. Kruschke and Liddel provide an overview of different data analytic approaches to scientific inference (*49*). For HB modeling, they recommend using High Density Intervals (**HDI**) to report results.

Our first BANOVA models the mean of average chlorophyll *a* concentrations within each spatial bin across all seasons (Figure 4). The second BANOVA models the mean of N:A within each spatial bin across all seasons (Figure 5). The third is a three-way BANOVA featuring the influence of spatial bins, seasons, and N:A groups on chlorophyll *a* concentrations (Figure 6).

Figure 4. BANOVA showing the distribution of estimated means of log(chlorophyll *a*) in each spatial bin. Tall, red dashed line represents the grand mean, 1.58 µg/L. The group means for bins 1 through 6 are, respectively, 2.43, 1.82, 1.66, 1.64, 0.98, and 0.94 . Numeric details of the HDIs are available in Table S2 in the Supporting Information online.

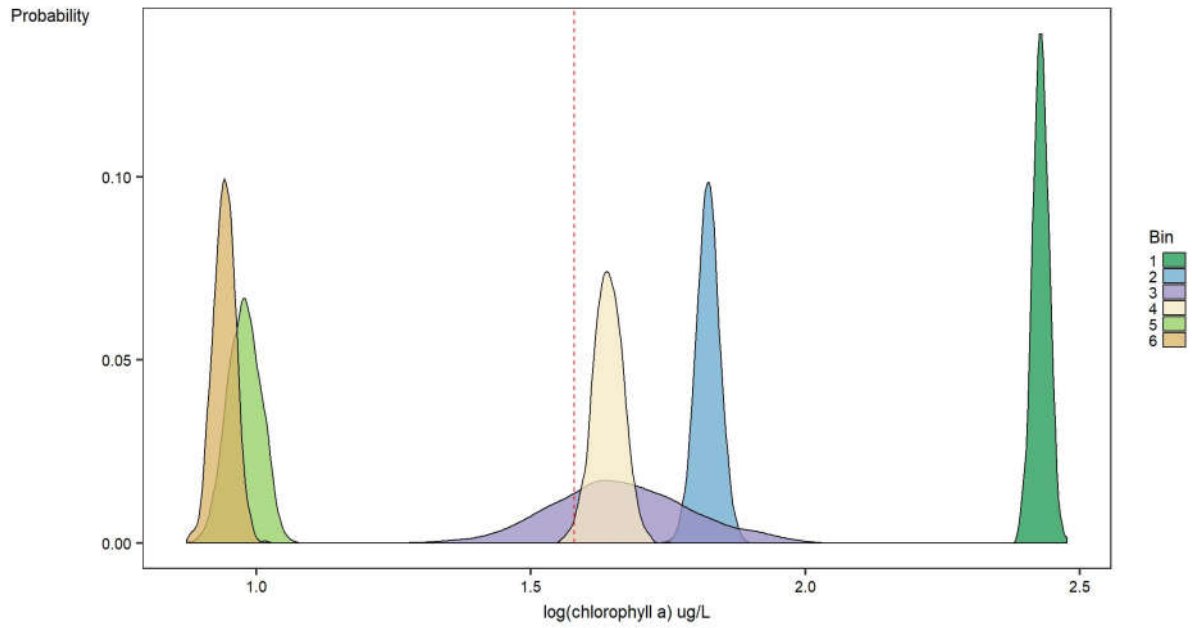


Figure 5. BANOVA showing the distribution of estimated means of N:A in each spatial bin. Tall, red dashed line represents the grand mean, 2.0 µg/L. The group means for bins 1 through 6 are, respectively, 1.46, 0.58, 1.07, 2.58, 2.28, 4.02. Numeric details of the HDIs are available in Table S3 in the Supporting Information online.

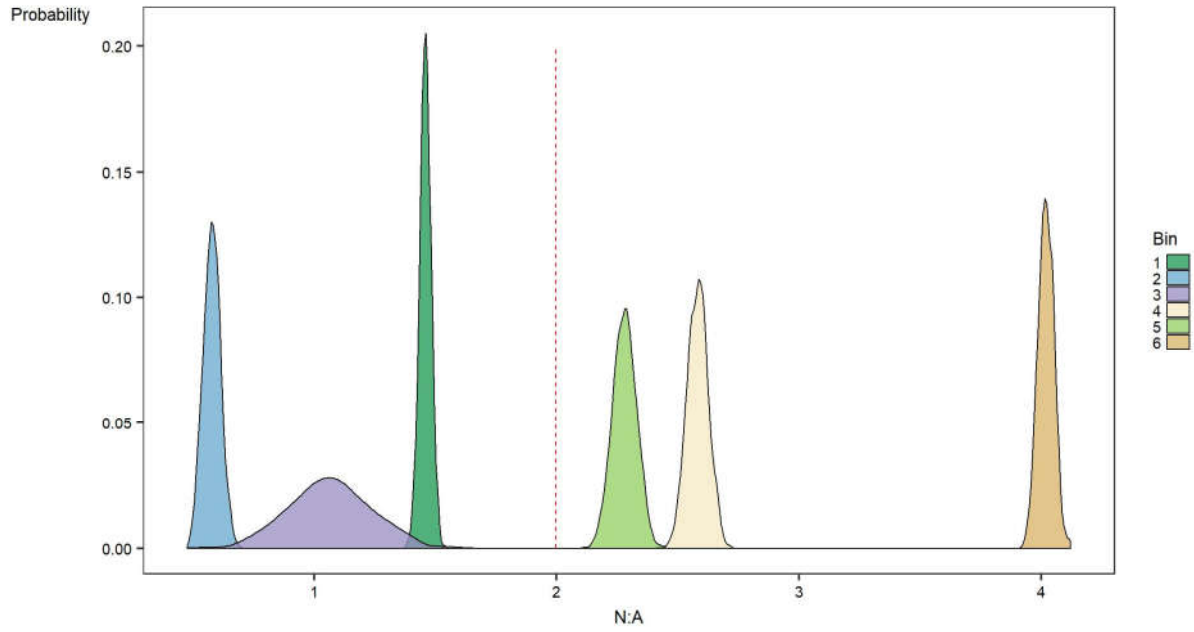
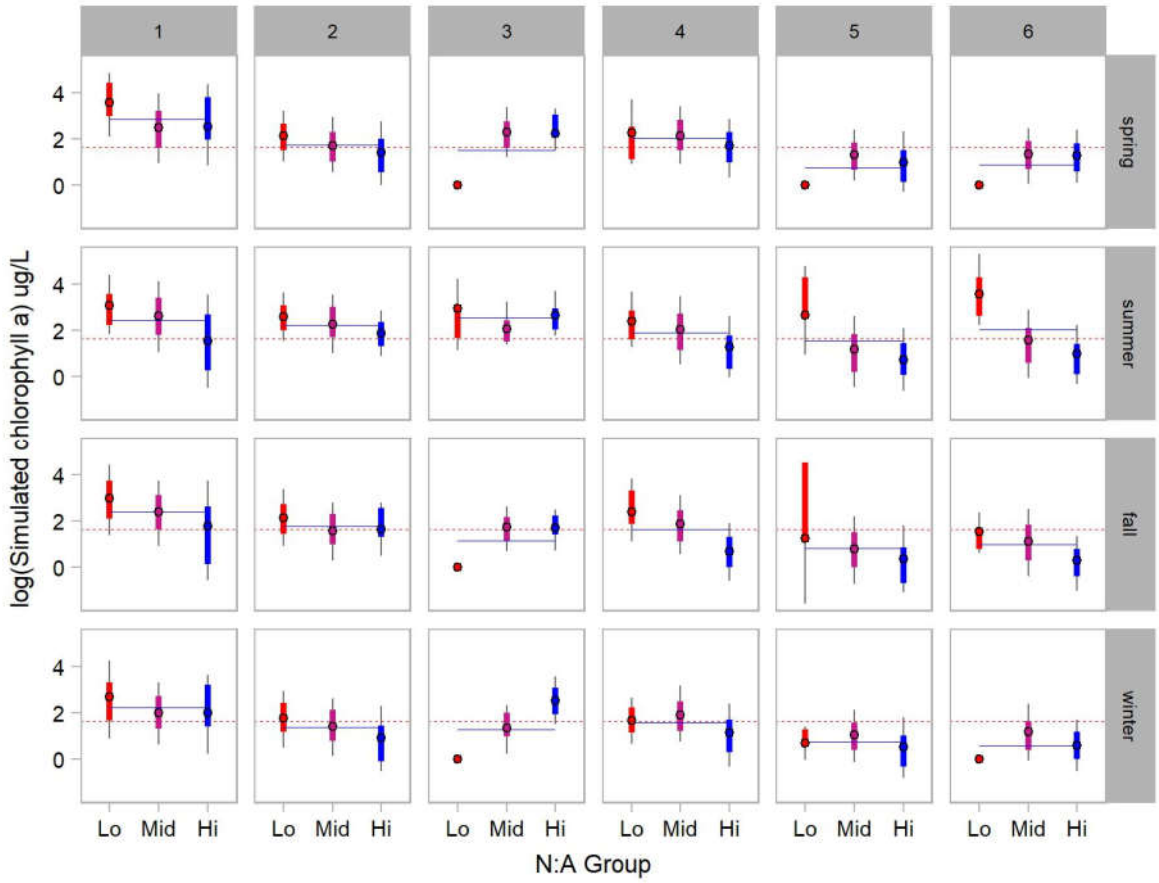


Figure 6. Three-way BANOVA showing distribution of estimated means of the log of simulated chloropyll *a* (µg/L) in each combination of spatial bin, season, and N:A group. (Note: zero values are not logged). Horizontal, red dashed line represents the grand mean, which is 1.62 µg/L . Horizontal, blue solid line is the group mean. For each N:A group, the thin, black vertical line is the 80% HDI. Thicker, colored vertical lines are the 50% HDI. Points in the center are the HDI mean. For example, during the *summer in spatial bin 6*, the group mean for chlorophyll *a* (horizontal, solid, blue line) is higher than the population mean (horizontal, dashed, red line). This is primarily because chlorophyll *a* is high in the low N:A group (the red dot); it is higher than both group and population mean. In contrast, chlorophyll *a* in the high N:A group (the blue dot) is low. Numeric details of the HDIs are available in Table S4 in the Supporting Information online.

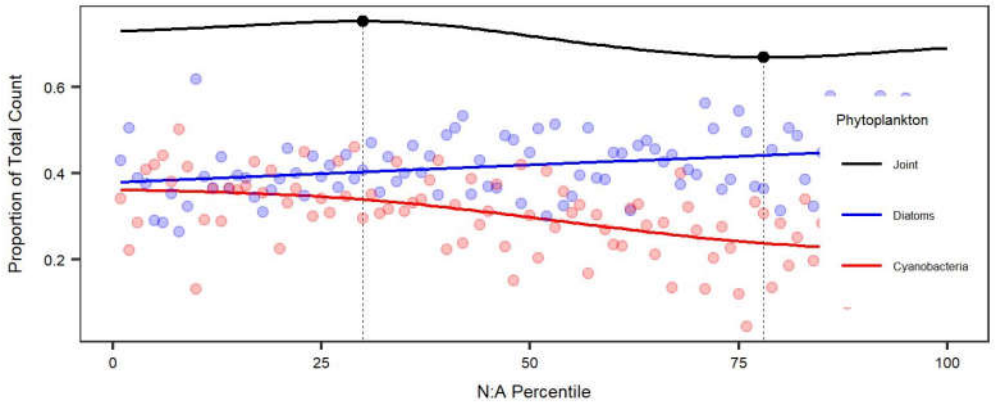


N:A Groups. We developed five GAMs. The first, second, and third GAMs are based on the phytoplankton data set and model the relationship between N:A percentiles and, respectively, the proportion of diatoms and cyanobacteria out of the total count of phytoplankton, as well as their joint proportion (Figure 7*a*). To develop the GAM, we

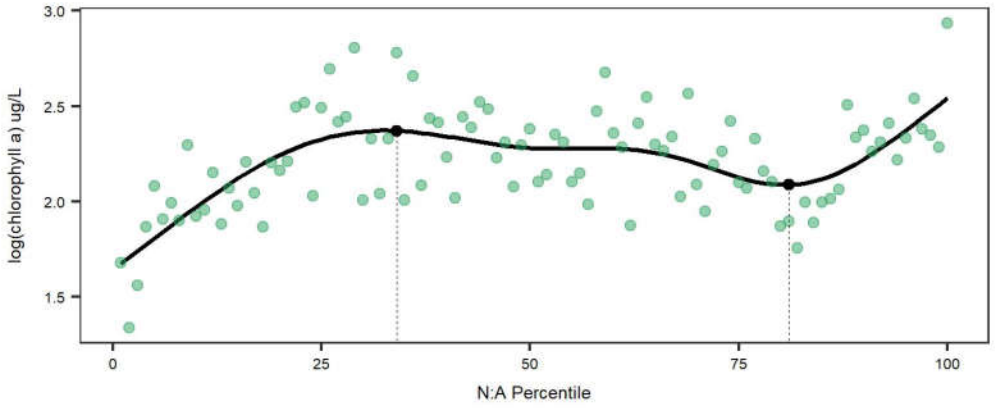
averaged proportions across all observations within each N:A percentile. The fourth GAM is based on the phytoplankton data merged with water nutrients data. For this, we modeled the relationship between N:A percentile and chlorophyll *a* levels (Figure 7*b*). The fifth GAM was developed for the entire water nutrient dataset, modeling N:A percentiles versus chlorophyll *a* levels. The GAM model is not shown but the inflection points occur when N:A is -1.03 and 3.08. For all models, we estimated the inflection points in the GAM by noting shifts in the value and sign of the GAM slopes.

Figure 7. General Additive Models for phytoplankton data. (*a*) N:A percentile vs. proportion of cyanobacteria, diatoms, and their joint proportion. Inflection points occur when N:A percentiles are 30% and 78%, or when N:A is -0.77 and 2.07 . **(*b*)** N:A percentile vs. log(chlorophyll a). Inflection points occur when N:A percentiles are 34% and 81%, or when N:A is -0.53 and 2.28.

(a)



(b)



Nutrient influence on chlorophyll *a*. We developed an HB regression model to analyze the relationship between selected nutrients and chlorophyll *a* at both the population- and group-levels, shown in Figures 8 and 9, respectively.

Figure 8. Parameter estimates for regression coefficients relating nutrients to chlorophyll *a* at the population level. The thin lines, thick lines, and points respectively represent the 80% HDI, 50% HDI, and mean for the estimates. Definitions of the predictors and the HDI values of their estimates are available in Table S5 in the Supporting Information online.

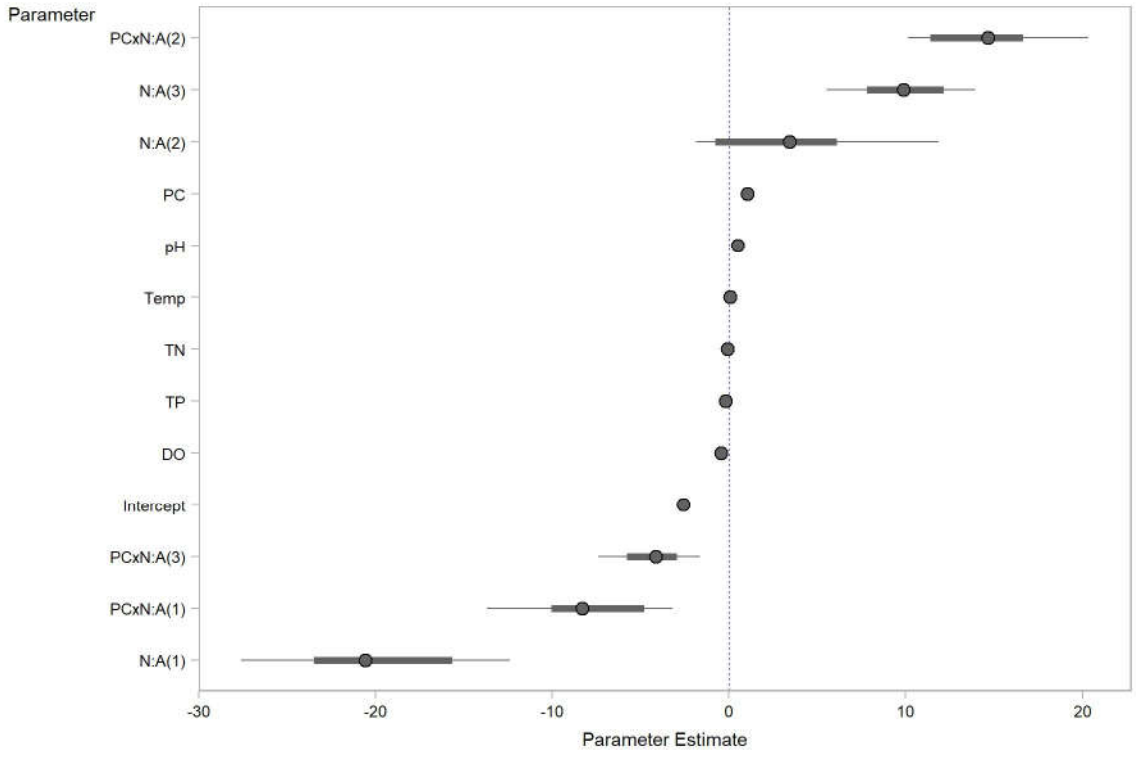
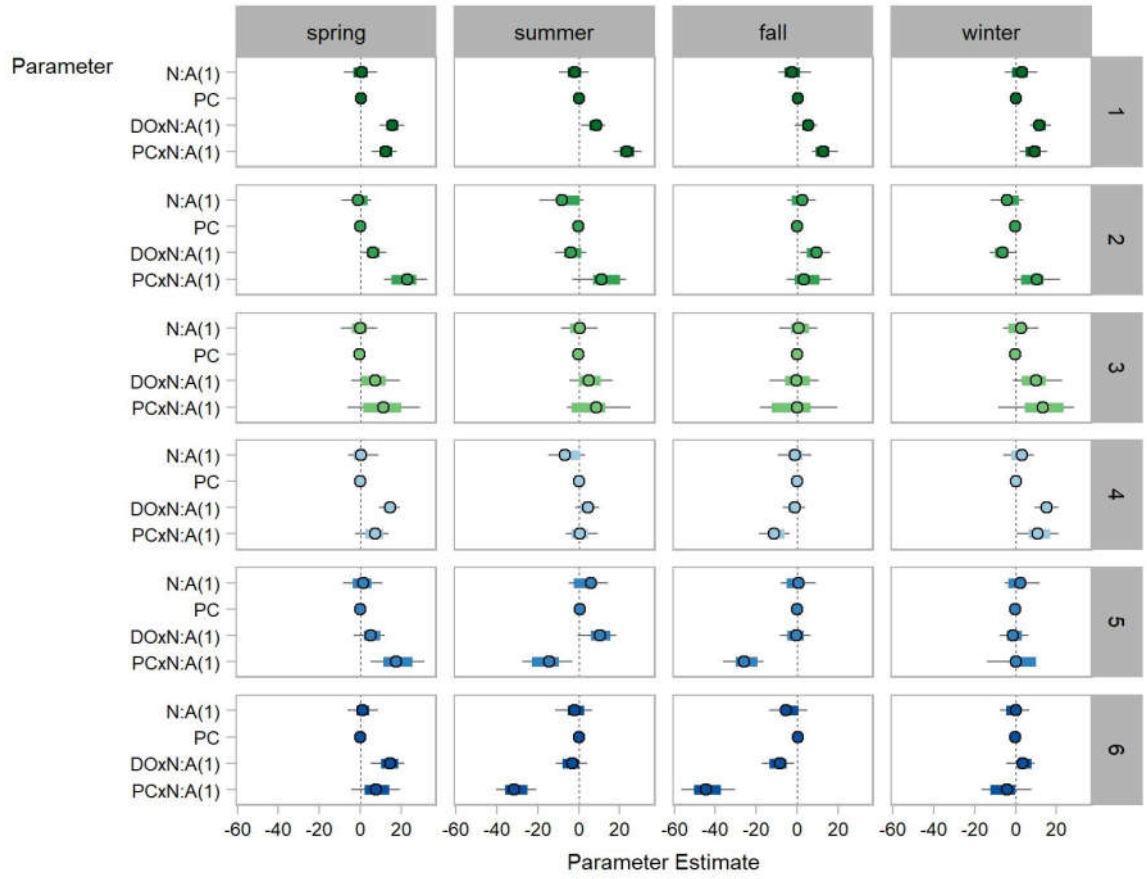


Figure 9. Parameter estimates for regression coefficients relating nutrients to chlorophyll *a* at the group level, defined by season and spatial bin. Only estimates for the first-order polynomial for N:A are shown. HDI values for all parameters are available in Table S6 of the Supporting Information online.



DISCUSSION

Data support the following hypotheses: (1) the ratio of nitrate to ammonium strongly influences chlorophyll *a* concentrations in the water column, more so than concentrations of total nitrogen, or nitrate and ammonium taken alone; (2) N:A's influence on chlorophyll *a* is variable across seasons and across spatial bins defined by land attributes; and (3) N:A's influence is moderated by particulate carbon, dissolved oxygen, and by the composition of the phytoplankton community in the water column, specifically, the relative proportions of diatoms and cyanobacteria. These hypotheses support the argument, made by some researchers, that water quality models based on nutrient loads and rate constants can be improved with information on how nutrients interact with the microbiome (*50*).

Variability in chlorophyll *a* concentrations around Chesapeake Bay. Chlorophyll *a* concentrations are highest in spatial bins 1 and 2. This is apparent in Figure 2's map, and corroborated by Figure 3's histogram and Figure 4's BANOVA. Conversely, observations with the lowest average chlorophyll *a* levels are in bins 5 and 6.

Influence of N:A on chlorophyll *a* concentrations. Considered alongside Figure 4, the BANOVA in Figure 5 highligths the influence of N:A on phytoplankton biomass. Figure 5 is almost an inverse image of Figure 4, indicating that bins with higher N:A tend to have lower chlorophyll *a* concentrations, while lower N:A is associated with higher chlorophyll *a*.

These BANOVA results are corroborated by the parameter estimates generated by the HB regression models featured in Figures 8 and 9. The estimates in Figure 8 suggest

N:A influences chlorophyll *a* more than any other water quality indicator in the dataset. Compared to the parameter estimates for particulate carbon (**PC**), total nitrogen, total phosphate, dissolved oxygen (**DO**), pH, and water temperature, the estimates associated with the polynomial form of N:A, along with their interaction terms, are strongly positive or negative.

The estimate for the first-order polynomial form (**N:A(1)**) provides supporting evidence for the BANOVAs of Figures 4 and 5 showing that N:A is negatively associated with chlorophyll *a*. In Figure 8, chlorophyll *a* relates inversely with N:A(1). While this is the overall trend, Figure 9 shows the dynamic nature of this relationship across seasons and spatial bins.

Variability of N:A's Influence. N:A's complex interactions with chlorophyll *a* is first suggested by the model form itself. Compared to alternative HB models (see Table S1 in the Supporting Information online), this model delivered the best results, suggesting that this form is most capable of computationally capturing the shifting influence of N:A on chlorophyll *a*.

Adding to this dynamism, Figure 9 shows N:A's influence is not only moderated by PC and DO, but also that these interactions vary with season and spatial bin. Despite these variations, some general patterns are distinguishable.

Across all bins, the effects of N:A and PC taken alone are neutral. Their effects on chlorophyll *a* manifest in their interactions with each other and with DO. Across all bins during spring, the moderating influence of PC and DO on N:A is slightly to strongly positive. This pattern is also true across all seasons for bin 1, but differs across other bins and seasons. For example, in bins 5 and 6, PC's moderating influence of PC

becomes negative in summer and fall. We discuss possible reasons for these patterns later.

Influence of N:A and phytoplankton composition on chlorophyll *a* concentrations. The GAMs in Figure 7*a* suggest that as N:A increases, the proportion of diatoms increases at a steady rate, while that of cyanobacteria decreases at varying rates. We also developed a GAM for the taxa's joint proportions. This third GAM has two inflection points, essentially dividing the N:A percentiles into three groups. In the first, the joint proportion increases as the proportion of diatoms increases, while that of cyanobacteria either increases or decreases at a slow rate. In the second, cyanobacteria decrease much faster, resulting in the joint proportion's decrease. In the third, the rate of decreasing cyanobacteria slows and that of diatoms increases steadily, resulting in an increasing joint proportion.

These patterns in Figure 7*a* can be explained by a key difference in the uptake of NH_4^+ and NO_3^-. Generally, NH_4^+ in the water column suppresses pathways responsible for NH_4^+ uptake into the phytoplankton cell, while NO_3^-'s presence enhances its uptake (*20*). Increasing N:A should therefore be strongly associated with increasing levels of diatoms, given their preference for NO_3^-. For cyanobacteria, the effect of increasing N:A is more complex. Increasing N:A implies less available NH_4^+ in the water column relative to NO_3^-, but also less suppression of NH_4^+ uptake into the cell. Cyanobacteria's uptake of NH_4^+ and their subsequent increase in biomass therefore results from the balance between these two opposing factors.

We developed a fourth GAM to investigate the relationship between N:A and chlorophyll *a* (Figure 7*b*), which shows a similar pattern to the joint GAM in Figure 7*a*.

The two GAMS show inflection points at roughly the same N:A levels. We *hypothesize* the pattern in Figure 7*b* is caused primarily by the disparate effects of N:A on diatoms and cyanobacteria, the two dominant taxa in our phytoplankton dataset.

Testing and interpreting the influence of N:A on chlorophyll *a* concentrations. Unfortunately, not all observations in the water nutrient dataset contained phytoplankton counts. We did, however, use all the data to develop a GAM of N:A versus chlorophyll *a*. The resulting model showed two inflection points, at levels where N:A was equal to -1.03 and 3.08. Using this information, we binned the observations into three groups with low (below -1.03), high (above 3.08), and medium values of N:A.

We used these N:A groups to partition simulated data from the HB regression model and conducted a BANOVA on these data (Figure 6). The BANOVA suggests (1) chlorophyll *a* levels differ across combinations of temporal and spatial bins; and (2) within each such combination, chlorophyll *a* is related to N:A in a way that might explain variation of chlorophyll *a* levels in the water nutrients dataset.

In Figure 6, for bin 1 across all seasons, the group means lie above the population mean. This happens because chlorophyll *a* production is high in the low N:A group; across all seasons in bin 1, the HDI mean for the low N:A group exceeds both the group and population means. Contrariwise, in bins 5 and 6 in spring, chlorophyll *a* production falls to 0 within the low N:A group, the net result being that group means fall below the population mean.

If, as we hypothesize, patterns in the water nutrients data can be explained by N-form preferences among different phytoplankton taxa, then Figure 6 suggests that

during spring, bins 5 and 6 do not provide an environment where cyanobacteria thrive. This situation changes in the summer, when chlorophyll *a* production does occur at low N:A, to a point where group means are either above or only slightly below the population mean. On the other hand, given the high levels of chlorophyll *a* production in bin 1 at low N:A across all seasons, watersheds there likely provide hospitable conditions for cyanobacteria.

These hypotheses also help explain the interaction effects among DO, PC, and N:A in Figure 9. PC is a strong correlate of chlorophyll *a* and we assume it is an indicator of phytoplankton that are already in the water column and that moderate N:A's effects. DO is an oxidizing agent influencing N:A levels.

Figure 6 indicates that across all seasons, watersheds in bin 1 contain phytoplankton communities that are sufficiently diverse to have taxa thriving in each N:A group. It follows that in Figure 9, N:A, DO, and PC would interact positively to produce more chlorophyll *a*, because the diverse mix of taxa could uptake and assimilate any N-form along the N:A gradient.

In bins 5 and 6, the situation is more complicated. Figure 6 indicates that *in spring*, these bins are dominated by taxa thriving primarily at high N:A levels. We would therefore expect PC, DO, and N:A to have positive interactions stimulating chlorophyll *a* production. This is corroborated by the interaction effects shown in Figure 9. But Figure 6 also indicates that *in summer and fall*, the phytoplankton communities in these bins diversify to include taxa with N-preferences across all N:A levels. In these situations, Figure 9 shows the over-all interaction effects of PC, DO, and N:A are to inhibit chlorophyll *a* production.

Glibert *et al.* suggest that under non-steady state conditions, as when nutrients are in excess, phytoplankton "chase" rather than "anticipate" environmental changes (*20*). If N:A levels vary from spring to summer and fall, the composition of the phytoplankton community adapted for spring N:A levels may lag behind and not be optimally adapted as N:A shifts to fall and summer levels.

If this is true, then PC, DO, and N:A may interact to ultimately inhibit chlorophyll *a* production. Figure 10 shows that N:A levels in bins 5 and 6 dip in summer and fall, perhaps to levels that are too low for taxa preferring high N:A. But Figure 10 also shows that despite this dip, these levels are still higher than they are in bin 1, perhaps at levels too high for taxa preferring low N:A. The net effect is a mismatch between N:A levels and a phytoplankton community ill-adapted to take advantage of these levels. Similar processes may explain the variability in Figure 9's parameter values.

To summarize, the results from the BANOVAs, HB model, and GAMs provide a speculative, but nonetheless plausible explanation of how variability in N:A influences chlorophyll *a*. Local spatial and temporal conditions determine N:A levels. In turn, N:A levels lead to different levels of chlorophyll *a* production, depending on the N-form preferences of the phytoplankton community within each temporal and spatial gradient.

Influence of temporal and spatial conditions on N:A levels. It is beyond the scope of this paper to provide a full explanation of how temporal and spatial conditions affect N:A levels. But we briefly raise some possible mechanism suggested by Figures 2 and 10 and suggest that this be an area for future research. Figure 10 shows the monthly variation in average log concentrations of selected water nutrients. Across all spatial

bins, N:A levels plunge during the summer months, primarily because of a drop in NO_3^- levels. This dip in N:A is associated with increases in chlorophyll *a*. However, N:A levels in bins 4,5, and 6 remain relatively higher and chlorophyll *a* relatively lower, because NO_3^- levels do not drop as significantly during the summer. Indeed, in bin 6, NO_3^- levels remain relatively high throughout the year. This seasonal pattern of N:A, along with the hypothesis regarding the N-form preferences of diatoms and cyanobacteria, is consistent with the results in Figure 3's histogram and Figure 6's BANOVA.

Figure 10. Average concentrations of selected nutrients per month in a given year.

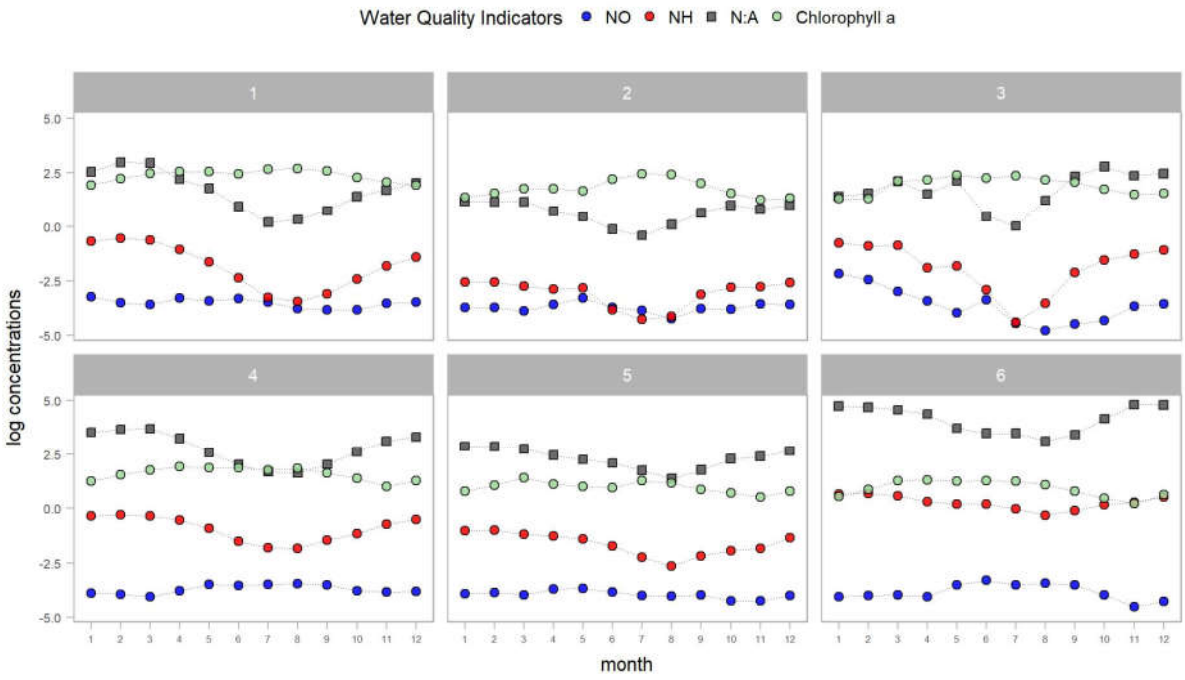


Figure 1 provides a snapshot of the main land attributes of each spatial bin. In bin 6, despite the large proportion of agricultural land, an attribute typically associated with eutrophication, it also has the lowest index for the proportion of water (*i.e,* ponds and lakes), the highest index for elevation range and variability, and mid-levels of forested

land. Bins 1 and 2 have not only the lowest elevation, but also the lowest variability in elevation. Additionally, bin 2 has the highest proportion of water.

When considered alongside Figure 2's map and Figure 5's BANOVA, Figure 1 suggests a spatial gradient wherein N:A decreases as water flows downstream to lower elevations. At these lower elevations, the proportion of water in the HUC12 and the dynamics of the Chesapeake estuary are most influential. Cha *et al.* note a similar pattern in North Carolina's Neuse Estuary and attribute it to the possibility that, downstream, internal processes rather than riverine input exert a more dominant influence on nutrient composition, particularly during the low-flow conditions of the summer months (*51*). Peterson *et al.* report that headwater streams tend to be higher in NO_3^- because riverine input tends to be higher in nitrate than ammonium and because NH_4^+ tends to be rapidly removed by assimilation, sorption, and nitrification (*30*), processes that would be enhanced by the higher stream flows that greater elevation variability would introduce. With regards to the influence of lakes and ponds, Coban *et al.* concluded that wetlands create an ecosystem conducive to N-transformations (*52*). They found that N-transforming bacteria, as well as the functional genes primarily responsible for N-transformations, were mainly distributed along the flow path of water in wetlands.

The foregoing may explain how land attributes influence N:A. At higher elevations, N:A is more reflective of nitrogen loads, which is predominantly NO_3^-, and of the relatively more rapid removal of NH_4^+ from the water column. At lower elevations, particularly in those locations where wetlands provide a haven for N-transformations, N:A is more reflective of in-stream processes favoring NH_4^+. These higher NH_4^+ levels

create conditions where cyanobacteria, based on energetic and metabolic advantages, can compete more effectively for the resources they need, thereby leading to higher chlorophyll *a* concentrations.

REFERENCES

1. C. Moore, D. Guignet, C. Dockins, K. B. Maguire, N. B. Simon, Valuing Ecological Improvements in the Chesapeake Bay and the Importance of Ancillary Benefits 1. *Journal of Benefit-Cost Analysis*, 1-26 (2017).
2. J. D. Wiggert, R. R. Hood, C. W. Brown, in *Modeling Coastal Hypoxia*. (Springer, 2017), pp. 119-147.
3. W. M. Kemp *et al.*, Eutrophication of Chesapeake Bay: historical trends and ecological interactions. *Marine Ecology Progress Series* **303**, 1-29 (2005).
4. A failing bay cleanup. [Editorial]. (2015, July 26). *The Washington Post.*
5. B. E. Lapointe, L. W. Herren, D. D. Debortoli, M. A. Vogel, Evidence of sewage-driven eutrophication and harmful algal blooms in Florida's Indian River Lagoon. *Harmful Algae* **43**, 82-102 (2015).
6. D. M. Anderson, P. M. Glibert, J. M. Burkholder, Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. *Estuaries and Coasts* **25**, 704-726 (2002).
7. P. M. Glibert, Ecological stoichiometry and its implications for aquatic ecosystem sustainability. *Current Opinion in Environmental Sustainability* **4**, 272-277 (2012).
8. D. O. Hessen, J. J. Elser, R. W. Sterner, J. Urabe, Ecological stoichiometry: an elementary approach using basic principles. *Limnol. Oceanogr* **58**, 2219-2236 (2013).
9. A. C. Redfield, *On the proportions of organic derivatives in sea water and their relation to the composition of plankton*. (University Press of Liverpool James Johnstone memorial volume, 1934).
10. C. A. Klausmeier, E. Litchman, T. Daufresne, S. A. Levin, Optimal nitrogen-to-phosphorus stoichiometry of phytoplankton. *Nature* **429**, 171-174 (2004).
11. K. R. Arrigo, Marine microorganisms and global nutrient cycles. *Nature* **437**, 349-355 (2005).
12. R. B. Domingues, A. B. Barbosa, U. Sommer, H. M. Galvão, Ammonium, nitrate and phytoplankton interactions in a freshwater tidal estuarine zone: potential effects of cultural eutrophication. *Aquatic Sciences* **73**, 331-343 (2011).
13. T. D. Harris *et al.*, Combined effects of nitrogen to phosphorus and nitrate to ammonia ratios on cyanobacterial metabolite concentrations in eutrophic Midwestern USA reservoirs. *Inland Waters* **6**, 199-210 (2016).
14. P. M. Glibert *et al.*, Phytoplankton communities from San Francisco Bay Delta respond differently to oxidized and reduced nitrogen substrates—even under conditions that would otherwise suggest nitrogen sufficiency. *Frontiers in Marine Science* **1**, 17 (2014).
15. R. Dugdale, F. Wilkerson, A. Parker, The “ammonium paradox:” a summary of more than a decade of research into phytoplankton processes and nitrogen relationships in the northern San Francisco Estuary. *Suisun Synthesis II. San Francisco (CA): San Francisco Estuary Institute*, (2016).
16. Q. Dortch, The interaction between ammonium and nitrate uptake in phytoplankton. *Marine ecology progress series*, 183-201 (1990).

17. R. Dugdale, F. Wilkerson, A. E. Parker, A. Marchi, K. Taberski, River flow and ammonium discharge determine spring phytoplankton blooms in an urbanized estuary. *Estuarine, Coastal and Shelf Science* **115**, 187-199 (2012).

18. R. C. Dugdale, F. P. Wilkerson, V. E. Hogue, A. Marchi, The role of ammonium and nitrate in spring bloom development in San Francisco Bay. *Estuarine, Coastal and Shelf Science* **73**, 17-29 (2007).

19. F. Wilkerson, R. Dugdale, in *Aquatic Microbial Ecology and Biogeochemistry: A Dual Perspective*. (Springer, 2016), pp. 117-126.

20. P. M. Glibert *et al.*, Pluses and minuses of ammonium and nitrate uptake and assimilation by phytoplankton and implications for productivity and community composition, with emphasis on nitrogen‐enriched conditions. *Limnology and Oceanography* **61**, 165-197 (2016).

21. E. Sperfeld, N. D. Wagner, H. M. Halvorson, M. Malishev, D. Raubenheimer, Bridging ecological stoichiometry and nutritional geometry with homeostasis concepts and integrative models of organism nutrition. *Functional ecology* **31**, 286-296 (2017).

22. M. I. Muro-Pastor, J. C. Reyes, F. J. Florencio, Ammonium assimilation in cyanobacteria. *Photosynthesis research* **83**, 135-150 (2005).

23. R. J. Ritchie, The ammonia transport, retention and futile cycling problem in cyanobacteria. *Microbial ecology* **65**, 180-196 (2013).

24. C. N. Dahm, A. E. Parker, A. E. Adelson, M. A. Christman, B. A. Bergamaschi, Nutrient dynamics of the Delta: effects on primary producers. *San Francisco Estuary and Watershed Science* **14**, (2016).

25. Y. Shangguan, P. M. Glibert, J. Alexander, C. J. Madden, S. Murasko, Phytoplankton assemblage response to changing nutrients in Florida Bay: Results of mesocosm studies. *Journal of Experimental Marine Biology and Ecology* **494**, 38-53 (2017).

26. E. Sanz-Luque, A. Chamizo-Ampudia, A. Llamas, A. Galvan, E. Fernandez, Understanding nitrate assimilation and its regulation in microalgae. *Frontiers in plant science* **6**, (2015).

27. B. E. Schirrmeister, M. Gugger, P. C. Donoghue, Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. *Palaeontology* **58**, 769-785 (2015).

28. T. W. Lyons, C. T. Reinhard, N. J. Planavsky, The rise of oxygen in Earth/'s early ocean and atmosphere. *Nature* **506**, 307-315 (2014).

29. C. Bowler *et al.*, The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239-244 (2008).

30. B. J. Peterson *et al.*, Control of nitrogen export from watersheds by headwater streams. *Science* **292**, 86-90 (2001).

31. A. J. Reisinger, J. L. Tank, E. J. Rosi-Marshall, R. O. Hall, M. A. Baker, The varying role of water column nutrient uptake along river continua in contrasting landscapes. *Biogeochemistry* **125**, 115-131 (2015).

32. J. Tank *et al.*, Partitioning assimilatory nitrogen uptake in streams: an analysis of stable isotope tracer additions across continents. *Ecological Monographs* **88**, 120-138 (2018).

33. R. B. Alexander, R. A. Smith, G. E. Schwarz, Effect of stream channel size on the delivery of nitrogen to the Gulf of Mexico. *Nature* **403**, 758 (2000).

34. D. W. Pritchard, C. L. Hurd, J. Beardall, C. D. Hepburn, Restricted use of nitrate and a strong preference for ammonium reflects the nitrogen ecophysiology of a light‐limited red alga. *Journal of phycology* **51**, 277-287 (2015).

35. A. F. Zuur, E. N. Ieno, C. S. Elphick, A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* **1**, 3-14 (2010).

36. R. J. Hijmans, raster: Geographic Data Analysis and Modeling. *R package version 2.5-8*, (2016).

37. R. B. a. T. K. a. B. Rowlingson, rgdal: Bindings for the Geospatial Data Abstraction Library. *R package version 1.2-11*, (2017).

38. R. B. a. C. Rundel, rgeos: Interface to Geometry Engine - Open Source ('GEOS'). *R package version 0.3-25*, (2017).

39. E. J. Pebesma, R. S. Bivand, Classes and methods for spatial data in R. *R news* **5**, 9-13 (2005).

40. J. R. ANDERSON, E. E. HARDY, J. T. ROACH, R. E. WITMER, D. L. P. Director, A Land Use And Land Cover Classification System For Use With Remote Sensor Data.

41. A. Sherry, R. K. Henson, Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. *Journal of personality assessment* **84**, 37-48 (2005).

42. M. Chavent, V. Kuentz-Simonet, A. Labenne, J. Saracco, ClustGeo: an R package for hierarchical clustering with spatial constraints. *Computational Statistics*, 1-24.

43. L. Al Shalabi, Z. Shaaban, B. Kasasbeh, Data mining: A preprocessing engine. *Journal of Computer Science* **2**, 735-739 (2006).

44. S. M. Lynch, *Introduction to applied Bayesian statistics and estimation for social scientists*. (Springer Science & Business Media, 2007).

45. M. Kéry, *Introduction to WinBUGS for ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses*. (Academic Press, 2010).

46. A. Guisan, T. C. Edwards Jr, T. Hastie, Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling* **157**, 89-100 (2002).

47. H. Singmann, D. Kellen, in *New Methods in Neuroscience and Cognitive Psychology*. (Psychology Press Hove, 2017).

48. M. C. N. TMDL, Total Maximum Daily Loads of Nitrogen and Phosphorus for Mattawoman Creek in Charles County and Prince George's County, Maryland. (2004).

49. J. K. Kruschke, T. M. Liddell, The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic bulletin & review*, 1-29 (2017).

50. P. M. Glibert, T. M. Kana, K. Brown, From limitation to excess: the consequences of substrate excess and stoichiometry for phytoplankton physiology, trophodynamics and biogeochemistry, and the implications for modeling. *Journal of Marine Systems* **125**, 14-28 (2013).

51. Y. Cha, I. Alameddine, S. S. Qian, C. A. Stow, A cross‐scale view of N and P limitation using a Bayesian hierarchical model. *Limnology and Oceanography*, (2016).

52. O. Coban *et al.*, Nitrogen transforming community in a horizontal subsurface-flow constructed wetland. *water research* **74**, 203-212 (2015).