Multi-scale trend analysis of water quality using error propagation of generalized additive models

Marcus W. Beck ([mbeck@tbep.org](mailto:mbeck@tbep.org)), Perry de Valpine ([pdevalpine@berkeley.edu](mailto:pdevalpine@berkeley.edu)), Rebecca Murphy ([rmurphy@chesapeakebay.net](mailto:rmurphy@chesapeakebay.net)), Ian Wren ([ianw@sfei.org](mailto:ianw@sfei.org)), Ariella Chelsky ([ariellac@sfei.org](mailto:ariellac@sfei.org)), Melissa Foley ([melissaf@sfei.org](mailto:melissaf@sfei.org)), David B. Senn ([davids@sfei.org](mailto:davids@sfei.org))

Compiled 2020-11-05 17:24:36

# Abstract

# Introduction

Many environmental monitoring programs generate fine-scale but potentially irregular time-series data to assess long-term trends for regulatory, management, or research purposes. The mismatch between the scales of monitoring versus analysis questions or management goals can lead to statistical challenges (Urquhart et al. [1998](#ref-Urquhart98), Cumming et al. [2006](#ref-Cumming06), Forbes and Xie [2018](#ref-Forbes18)) [Marcus will go over refs]. At finer temporal scales, environmental systems may show short-term fluctuations from multiple factors (e.g., weather events, management, or seasonal changes). Such fluctuations may not be of direct interest for longer-term trends or may not be well-suited to multi-scale smoothing methods. However, aggregate features of seasonal patterns that integrate or summarize short-term fluctuations may be of interest. In this paper, we describe methods to estimate across-year trends of within-year features of interest such as a seasonal average, seasonal peak, or seasonal timing of events, while accounting for uncertainties across analysis steps.

Previous methods for water quality trend analysis can be generalized into four basic approaches. Seasonal Kendall tests or related non-parametric approaches have been used for decades in water quality trend assessments to identify monotonic changes that account for seasonal variation between years (Hirsch et al. [1982](#ref-Hirsch82), Helsel et al. [2020](#ref-Helsel20)). A literature survey of Wan et al. ([2017](#ref-Wan17)) revealed non-parametric approaches to be the most commonly used methods in long-term water quality trend analysis, yet they have limited scope. They do not account for changes occurring at different temporal scales, do not adequately evaluate irregularly spaced data [is this true? I would think so.], do not incorporate other explanatory variables, and do not estimate a model that could be useful for other purposes. Thus, while these non-parametric approaches have some degree of robustness, they apply only to narrow goals.

The seasonal trend decomposition using loess (STL) decomposes a time series into additive components of a long-term trend, a seasonal pattern, and residuals (Cleveland et al. [1990](#ref-Cleveland90), Cloern and Jassby [2010](#ref-Cloern10), Stow et al. [2015](#ref-Stow15)). While useful and widely applied, this method also has important limitations. STL decomposition does not incorporate explanatory variables, it is defined more as an algorithm of statistical steps than as a coherent statistical model (e.g., Wan et al. [2017](#ref-Wan17)), and it does not usually include standard errors to allow hypothesis testing (but see Hafen [2010](#ref-Hafen10)). Conventional STL approaches may also over-simplify trends into absolute components that do not change over time, e.g., a seasonal estimate that is constant across years. This limitation presents challenges when addressing questions relevant to long-term water quality data, such as timing of seasonal peaks that can suggest system response to changing environmental conditions (Cloern and Jassby [2010](#ref-Cloern10), Navarro et al. [2012](#ref-Navarro12)).

The more recently developed method of weighted regression on time, discharge, and season (WRTDS) uses a more general local regression scheme than STL (Hirsch et al. [2010](#ref-Hirsch10), Beck and Hagy [2015](#ref-Beck15)). Designed for river data where separating the effect of discharge on constituent concentration is important, WRTDS estimates a moving window regression model with components that allows parameters to vary smoothly in relation to both time and discharge. This yields parameters that are specific to season, year, and flow regime. Conceptually, the approach is similar to localized multi-polynomial smoothing methods, although the application was developed specifically for describing long-term water quality trends. Standard error estimates of predictions are available through a “block bootstrap” approach that uses Monte Carlo estimates of false positive rates from the model results (Hirsch et al. [2015](#ref-Hirsch15)). Although a useful addition to the original method in Hirsch et al. ([2010](#ref-Hirsch10)), the approach requires extensive resampling as a post-hoc application to a previously fitted model.

The final and most recent approach is to use smoothing splines to separate fluctuations on different time scales and do so within the larger framework of generalized additive models (GAMs) (He et al. [2006](#ref-He06), Morton and Henderson [2008](#ref-Morton08), Pearce et al. [2011](#ref-Pearce11), Haraguchi et al. [2015](#ref-Haraguchi15), Murphy et al. [2019](#ref-Murphy19)). These may be seen as generalizing the concepts behind STL and WRTDS. In statistics, the evolution of non-parametric regression methods has largely converged on GAMs rather than more generalized kernel smoothing methods used by both STL and WRTDS. Kernel smoothing and spline smoothing are closely related, and a key challenge for each is to determine the appropriate degree of smoothing. For example, WRTDS can potentially give results similar to the spline-based smoothing methods described next, although at higher computational expense and with the limitation that uncertainty estimates are not readily obtainable from the original method (Beck and Murphy [2017](#ref-Beck17)). User input is also required to specify an acceptable degree of smoothing used by the “windows” that define the localized fit of WRTDS at each point in the time series. These windows are conceptually similar to the kernel (or bandwidth) that used in more conventional smoothing methods. There is no simple rule to guide the choice of defining an appropriate size and a tradeoff between over- and under-smoothing is a hallmark of these approaches.

Compared to kernel smoothing methods, GAMs have various advantages. They are formulated using “basis functions”, and these can be customized for needs such as cyclic splines (e.g., for an annual pattern) and low-dimensional interactions. They can naturally include both parametric (e.g. linear or quadratic) components and non-parametric (spline) components. Importantly, multiple approaches to automatically determine the optimal degree of smoothness have been developed, based on likelihood and/or optimizing out-of-sample prediction error. They have natural frequentist and Bayesian interpretations, are naturally extensible to include random effects (GAMMs), and have computationally efficient implementations that can be optimized more quickly than other approaches (Wood [2017](#ref-Wood17)). For these reasons, GAMs are widely used for non-parametric regression smoothing in many fields.

GAMs have recently been applied to trend analysis of water quality time-series, particularly from long-term monitoring programs (Haraguchi et al. [2015](#ref-Haraguchi15), Murphy et al. [2019](#ref-Murphy19)), but with different formulations and goals than given here. For example, in the US, the Chesapeake Bay Program uses GAMs to decompose time-series into long-term and seasonal trends (Murphy et al. [2019](#ref-Murphy19)) and test trend hypotheses between two points in time. That is related to the methods here, but this paper gives more general methods for analyzing trends of seasonal spline features, describes the relationships among alternative spline formulations when they are used as designed (Wood [2003](#ref-Wood03), [2017](#ref-Wood17)) rather than for ad hoc separation of time scales, and prioritizes full incorporation of uncertainty. Other studies of environmental time-series with GAMs have addressed the use of transformed response data (Yang and Moyer [2020](#ref-Yang20)), serial correlation in high resolution data (Morton and Henderson [2008](#ref-Morton08), Yang and Moyer [2020](#ref-Yang20)), and different time lags in describing relationships between response and predictor variables (Lefcheck et al. [2017](#ref-Lefcheck17), Testa et al. [2018](#ref-Testa18)).

Our motivating problem has several needs that are not satisfied by previous methods, but can use GAMs as a starting point. Our general goal is to understand interannual changes in seasonally averaged water quality metrics, such as chlorophyll. However, the seasonal average within each year must account for different sampling times and intervals, and any trend analysis must incorporate the uncertainties in seasonal averages. STL and/or WRTDS could potentially separate seasonal from long-term trends, but doing so is not necessary to determine seasonal averages. What is needed is an accurate estimate of uncertainty (e.g., a standard error) of seasonal averages, allowing for irregular sampling and the non-independence inherent in time-series. This can be done with GAMs, but we develop an application that is distinct from previous studies. Even if estimates of seasonal averages and their standard errors are available, none of these methods are designed to understand interannual trends in those averages. A Kendall test would not incorporate the standard errors or reveal useful long-term patterns beyond a significance test. Similarly, STL and WRTDS are not designed for this goal.

We illustrate the proposed methods with the motivating example of water quality monitoring in the southern portion of San Francisco Bay, California, USA. For several decades, approximately twice-monthly monitoring has been conducted at fixed locations (stations) of the longitudinal axis of the Bay. Analysis of these data is complicated by irregularities in timing and consistency of data collection, such that simple seasonal averages of raw data may not adequately describe trends. Examples of long-term trend questions include: Are there significant trends in spring mean chlorophyll at four-year (or other) time-scales? At what across-year scale do within-year summer-fall mean chlorophyll levels change? Are there significant across-year trends in within-year timing of the spring bloom in chlorophyll or in baseline levels during periods of low productivity? We also provide an approach for using meta-analysis methods (Gasparrini et al. [2012](#ref-Gasparrini12), Sera et al. [2019](#ref-Sera19)) following signal extraction with GAMs that is new to environmental trend-detection problems. For this step, we give methods for isolating seasonal trends secondarily from GAM results with reasonable certainty and evaluating these trends between years.

# Methods

## Study area and data sources

The San Francisco Estuary (SFE) is the largest estuary on the Pacific Coast of North America and drains an area of approximately 200 thousand km in the US state of California. Major freshwater inputs enter the system through the Sacramento-San Joaquin Delta complex upstream of Suisun Bay, where the combined inflow from both rivers is approximately 28 km per year. The northern subembayments are river-dominated (salinity ranging from 0 to 15 ppt), whereas the southern subembayments are marine-dominated with salinity ranging from 5 to 35 ppt depending on the tidal cycle, effluent discharge from wastewater treatment plants, and stormwater runoff. The South Bay embayment is heavily urbanized and includes thirty-seven wastewater treatments plants that serve 7.2 million people. Secondary treatment occurs at a majority of the treatment plants and the remaining effluent is discharged into the SFE. Agricultural runoff from the upper watershed also contributes to nutrient loading in the SFE with the annual nutrient export estimated as approximately 30 thousand kg dy of nitrogen from the Delta.

Nitrogen and phosphorus levels in SFE generally exceed concentrations that have been observed to promote excess primary production in other large estuarine systems. However, eutrophic conditions have not been regularly observed since routine monitoring began in the 1970s. Historical resistance of SFE to eutrophication has been attributed to several factors, including elevated suspended sediments that reduce light penetration in the water column, regular exchange and mixing with low-nutrient marine waters and export of estuarine nutrients to the Pacific Ocean, and benthic grazing by filter-feeding bivalves that reduce algal concentrations. Renewed interest in the potential for nutrient loading to negatively affect water quality has occurred recently, particularly in South Bay, where harmful algal blooms (HABs), increases in summer-fall chlorophyll concentrations, and low dissolved oxygen concentrations beginning in 1999 (Figure 1) (Cloern et al. [2020](#ref-Cloern20)). Although visual changes in observed data are apparent, statistical analyses to quantify current status and to provide estimates of annual and seasonal trends with appropriate bounds on uncertainty have not been sufficiently developed, particularly on a seasonal basis.

The analysis evaluated near-surface chlorophyll-a data collected biweekly to monthly along the South Bay axis extending from Central Bay (stations 18-23), South Bay (stations 24-32), and Lower South Bay (stations 34-36) (Table 1, Figure 2). Monitoring data were obtained from the SFE Research Program of the US Geological Survey (Cloern and Schraga [2016](#ref-Cloern16), Schraga et al. [2020](#ref-Schraga20)). Discrete chlorophyll concentrations at each station were determined by fluorimetric analysis with 90% acetone pigment extraction on GFF filters. Data collected between 1990-2019 were selected for analysis because it represented a suitable balance among three factors relevant to testing the statistical approaches, including sufficient length of record, consistent biweekly-monthly sampling, and a diverse set of stations covering the salinity gradient across multiple subembayments. While sampling frequency varied somewhat over time or by station, all data were treated as unique time series within the statistical models (i.e., no spatial or temporal binning or averaging was done).

## GAM application

The methods proposed here involve three stages. First, a GAM is used to estimate a smooth pattern of variation in the raw data along with its uncertainty. Second, a feature of interest is calculated from the estimated GAM, along with its propagated uncertainty. For this example, the seasonal averages are extracted, whereas other features could be the timing or magnitude of a seasonal peak, but those are not developed here. Third, a mixed effects meta-analysis is used to estimate trends and test hypotheses about the change in seasonal averages across years. While meta-analysis methods arose from analyses of results from multiple studies, their distinguishing characteristic is propagation of uncertainty (Gasparrini et al. [2012](#ref-Gasparrini12), Sera et al. [2019](#ref-Sera19)). Meta-analysis uses response data that includes standard errors (uncertainties) as needed to address our questions. Two mixed effects meta-analysis approaches are developed, 1) simple comparison of whether seasonal features differ across years, and 2) estimation of short-term linear trends on time-scales chosen by the analyst.

The three-stage approach is motivated in several ways. In the first stage, the GAM is used for signal-extraction at the scale of the raw data and we explain the relationships among various GAM specifications to emphasis concordance between the trend-analysis goal and signal-extraction goal in the formulation of splines. The second stage uses mixed-meta analysis to propagate uncertainty from estimated GAM parameters to features of interest, i.e., seasonal averages. Such features can be practically evaluated as functions of the fitted GAM, rather than as parameters to be estimated within the GAM. On a practical level, the first two stages can be viewed as creating a “data product” that can be used in any subsequent analyses as long as they include the uncertainty associated with each point estimate. This offers an important practical benefit for visualization and management purposes. For the third stage (long-term trend-analysis), we harness existing meta-analysis methods that are well-suited to the task at hand.

### First-stage analysis: GAM estimation

Generalized additive models use sums of parametric terms (e.g., linear) and non-parametric terms (e.g., smoothing splines) to predict response data. In our approach, GAMs are used to estimate signal and uncertainty from the time-series data. Results are subsequently used in the second-stage analysis, described below.

In this section we describe four GAM formulations that can achieve the same or similar fits to data in slightly different ways. We present these to clarify relationships that may otherwise be confusing and to emphasize that results can be similarly achieved if a model unconstrained to have sufficient capability to extract a signal from raw data. Models will be written in the notation of the mgcv R package as formulas for the gam function (Wood [2017](#ref-Wood17)).

The fundamental goal is to smooth the raw data across time to separate a signal of response variable from noise. The simplest GAM for this purpose can be expressed as:

Model S: y ~ s(cont\_year, k = num\_knots\_Y)

Here y is the time-series of interest, such as chlorophyll, cont\_year is “continuous year”, a continuous numerical date (e.g., July 1st 2019 would be 2019.5), y ~ s(...) indicates that y will be explained by a smoothing spline (in this case of cont\_year), and num\_knots\_Y is the number of knots used to create the spline. Following previous authors (Murphy et al. [2019](#ref-Murphy19), Yang and Moyer [2020](#ref-Yang20)), we consider log-transformed (base 10) chlorophyll levels for analysis.

For smoothing in mgcv, appropriate smoothness is achieved not by choice of knots but by a penalty on the net curvature of the spline (Wood [2004](#ref-Wood04)). Smoothing was determined using generalized cross-validation (GCV, as implemented in mgcv), which approximately minimizes out-of-sample prediction error. To allow GCV (or other alternatives) to work as intended, the number of knots must be sufficiently large. Results should not be sensitive to the number of knots; if they are, the number of knots should be increased. Therefore, we choose a sufficiently large number of knots for num\_knots\_Y as 12 times the number of years in the time series that was modeled. This created the potential to have one knot per month as an approach to both prevent under-fitting the observed data and to accurately estimate the seasonal signal within a year. In some cases, this upper limit was too large given the observed data and the multiplier was reduced by one until sufficient degrees of freedom were available (i.e., 12 \* years, 11 \* years, etc.).

The next three spline formulations differ in how spline terms compose a model to smooth the raw data. For convenience, a linear trend in cont\_year can specified explicitly as a separate component. This would be expressed as:

Model SY: y ~ cont\_year + s(cont\_year, k = num\_knots\_Y)

This model is effectively and mathematically equivalent to model S. The spline for cont\_year includes an unpenalized linear trend, so a trend will be estimated in model S. When cont\_year is included explicitly as a linear term in model SY, the mgcv implementation automatically adjusts the basis functions for the spline to exclude the linear term, to avoid over-parameterizing the model. With model S, the estimated trend in cont\_year and its uncertainty can be extracted from the fitted spline. With model SY, it is available more conveniently. Further, package mgcv includes an option select = TRUE to penalize linear trends in splines to provide a method for variable selection, such as when numerous splines are included in the model formulation for variables that may or may not be important. For our approach, this option is not used and all models specify select = FALSE. Details in the supplement explain this justification.

Next, an average within-year cyclic pattern as a separate spline can be separated in the model formulation. This can be expressed as:

Model SYD: y ~ cont\_year + s(cont\_year, k = num\_knots\_Y) + s(doy, bs = 'cc', k = num\_knots\_D)

Here doy is “day-of-year” (i.e., Julian date, a count starting January 1 for each year), bs = 'cc' indicates the spline will be cyclic (constrained to start and end at the same value), and num\_knots\_D is the upper limit for the number of knots for the doy spline. While model SYD is not mathematically equivalent to models S and SY, it should be functionally similar. In this model, the doy spline gives the average pattern of fluctuations within each year. This changes the interpretation of the cont\_year spline to represent smoothed deviations from the average within-year pattern.

Models S, SY, and SYD can all potentially extract a similar signal from the raw data. What differs between the models is the allocation of penalties for curvature used to determine smoothness for each spline. In model SYD, there are separate penalties for the two splines, as compared to S and SY that include penalties only for the cont\_year spline. Caution is needed in interpreting differences in fit between model SYD and models S or SY because the penalties for smoothing splines based on curvature are heuristic (Wood [2017](#ref-Wood17)). For example, if a lower AIC is achieved in one model compared to another, assuming both use sufficient knots, this may just reflect the outcome of alternative penalization heuristics implied by the different formulations. In the examples here, model SYD achieves nearly identical fits to model S or SY, where the latter by definition also achieve identical fits.

Model SYD has the appealing feature that, if some parts of some years have limited data, model SYD will “borrow information” across years. The model will impute an average seasonal pattern where data are limited. However, a conceptual representation of what an average within-year pattern represents is challenging. For example, suppose a spring chlorophyll peak is a notable feature in every year. If the timing is the same in every year whereas the magnitudes vary, then the average within-year pattern will have an average magnitude thta appears as a “typical” year. However, if the magnitude is the same and the timing varies across years, then the average within-year pattern will not appear typical. The imputed values will represent a day-by-day average, appearing more spread out in time than the pattern in any single year.

More importantly, the average within-year pattern (and its uncertainty) can be obtained from estimated models S and SY, as well as from model SYD, but using a different approach. In models S and SY, the average within-year pattern is simply the average across years of the smoothed within-year patterns. From this perspective, model SYD, like model SY, is a convenience, estimating the average within-year pattern directly as a separate estimated spline. The average pattern extracted from model S or SY may not be identical to the doy spline estimated in model SYD, again due to different smoothing formulations, but in practice they should be similar.

Finally, the raw data can be smoothed using a bivariate spline representing an interaction between cont\_year and doy. This can be expressed as:

Model SYDI: y ~ cont\_year + s(cont\_year) + s(doy, bs = "cc") + ti(cont\_year, doy, bs = c("tp", "cc"), k = c(num\_knots\_Y\_ti, num\_knits\_D\_ti))

Here, ti() specifies a tensor-product spline for a surface that varies smoothly as a function of both cont\_year and doy. The number of knots is the product of num\_knots\_Y\_ti in the cont\_year axis and num\_knots\_D\_ti in the doy axis. In SYDI, the need for sufficient knots can be satisfied either by sufficiently large values for num\_knots\_Y\_ti and num\_knits\_D\_ti or a sufficiently large value for knots in s(cont\_year), but not both to avoid over-parameterization.

Following similar rationale above, the relationship of model SYDI to model S is similar to that of model SYD to model S. Model SYDI is more mathematically different, but all of the splines use the same inputs to smooth the same data. The splines in cont\_year and doy will likely not capture as much variation as above because of fewer knots. The ti term represents an interaction by allowing the pattern in cont\_year to vary by doy and vice-versa. The interaction term in model SYDI provides an appearance that this model is fundamentally different from those provided by the other models. However, models S, SY, and SYD all allow within-year fluctuations to vary across years by allowing a spline to be fit through the entire time-series. Although model SYDI is the only model that includes an explicit interaction term, all of the models support the interaction conceptually. By providing this term with sufficient knots, the raw data can be fully smoothed with model SYDI to a similar degree as for the other models. However, this outcome is difficult to achieve for reasons explained below.

The distinct aspect of model SYDI is the anticipation that within-year fluctuations will vary smoothly across years. For example, if one year has an early and large magnitude spring bloom, the years before and after would have similar patterns described by the model, perhaps different but not substantially. For the raw San Francisco Bay data, this may not be the case, as one year might have an early, high spring bloom and the next year a late, low spring bloom, or other such large variation across years. If the interaction spline has sufficient knots, then the degree of smoothness across years can be estimated to be very low, allowing fluctuations within years to vary strongly across years. However, if that is the case, the conceptual motivation for model SYDI may not be justified. In summary, model SDYI may best be suited for systems where within-year fluctuations actually vary smoothly across years, in which case it may out-perform the other models. If that is not the case, it may perform similarly to the others, but the spline structure and penalty structure are more different, making performance harder to compare.

All of the models require a sufficiently large number of knots in one or more of the spline terms to accommodate the fluctuations in the data and use the GAM estimation methods in the manner for which they were designed. In models S, SY, and SYD, the number of knots in the cont\_year term need to be sufficiently large. In model SYDI, a very large number of knots in both the cont\_year spline *and* in both dimensions of the interaction spline is impossible to achieve. Allowing the cont\_year spline to include many knots would somewhat defeat the purpose of the interaction spline. In practice, allocation of knots between the different splines can be chosen by the analyst so long as it is recognized that different choices will arbitrarily lead to more variation being explained between the splines.

Murphy et al. ([2019](#ref-Murphy19)) used a related set of spline formulations but used them differently than we do here. Murphy et al. ([2019](#ref-Murphy19)) have a “gam0” with only a s(doy) term, a “gam1” like our SYD, and a “gam2” like our SYDI. Their gam0 is *a priori* not of interest for our data and was not explored further. Compared to our uses of SYD and SYDI, their uses differ in choice of knots and interpretation of results. For SYD, they set a maximum number of knots in the s(cont\_year) term of 2/3 times the number of years, whereas we use 12 times the number of years, roughly one per month. Thus, they interpret this spline as separating a long-term (or low-frequency) trend from other patterns, whereas we use it to separate signal from noise at a scale informed by the data. In the SYDI model, Murphy et al. ([2019](#ref-Murphy19)) do not explicitly consider number of knots in the interaction spline. In both cases, Murphy et al. ([2019](#ref-Murphy19)) in effect use choice of number of knots in different spline components to serve as ad hoc allocation of variation in the data to different components while not necessarily allowing any component to fully estimate the signal in the data. Murphy et al. ([2019](#ref-Murphy19)) acknowledge that incomplete modeling of fluctuations in the data may lead to inflated Type I error rates, which would apply to their later comparisons of changes across time, but they leave that problem for future work. For the methods presented here, we seek to avoid inflated Type I error rates arising in this way. Finally, Murphy et al. ([2019](#ref-Murphy19)) present large AIC differences between their spline formulations. We instead emphasize that, given sufficient knots, the models represent alternative formulations of conceptually similar explanations for the data and yield similar fits. In our example, such large differences in AIC would only reflect inadequate choice of knots in one or more splines.

### Second-stage analysis: Uncertainty propagation from estimated GAMs to seasonal features

Next we consider the goal of estimating a seasonal average, such as the mean spring chlorophyll concentrations and the estimated uncertainty in each year. Define as the seasonal average in year , as an estimate of , and as the estimated standard error of . The season includes days. For simplicity, the following text omits subscript .

Point estimates of response values for the fitted GAM take the form , where is the vector of estimate parameters and is a model matrix of explanatory variables, including spline basis function values. Here includes both fixed effect parameters and spline parameters, and contains columns corresponding to each. For example, using model SY, if a point estimate for chlorophyll is needed for a single day, given as dec\_year = , then would have a row with with in the first column (for the intercept parameter), (for the linear time trend) in the second column, and an evaluation of each spline basis function at (and possibly its corresponding doy value) in the remaining columns. The number of spline basis functions is related to the number of knots. Note that can be any time, not necessarily the time of an observation.

To obtain a vector, , of fitted point estimates for every day in a season, would have one row for each day. Here, the seasonal averages used in our examples are calculated at the resolution of days. The estimated spline yields both and , an estimate of the covariance matrix of the sampling distribution of . The scalar standard errors of are the square roots of the diagonal elements of . Since parameter estimates are correlated, it is important to use the full to determine the covariance of . This is given by .

Finally, the estimated seasonal average is , where is a row vector with all values equal to . The variance of is and standard error is . Each of these estimates are standard results based on approximate multivariate normality of the sampling distribution of . A similar summary for comparing multi-year averages of month-scale spline values for sets of years at the beginning vs. end of a time-series is provided by Murphy et al. ([2019](#ref-Murphy19)). For example, the above estimates can be used for a test of the null hypothesis that y is the same at two times. To do so, would have two rows for each time to compare. The vector would be [-1, 1], so that gives the point estimate of the difference in chlorophyll between those times, and gives the squared standard error of that difference. This difference can then be used to determine if the two estimates in time are significantly different from 0, e.g., is Chla-a different now than it was before based on GAM estimates.

### Third-stage analysis: Trend analysis of seasonal features with uncertainties

This section explains how the within-year features () from the estimated GAMs, and their standard errors , are used to investigate across-year trends of within-year features. In the examples, we use two seasonal averages of chlorophyll as features, one for January-July and the other for August-December. These periods are chosen for their relevance to phytoplankton bloom phenology and broad seasonal patterns that occur in the San Francisco Bay region. For any feature of interest that includes estimates with associated standard errors, meta-analysis methods can be used to address two kinds of questions:

1. Do two years, or a group of years, differ significantly in a seasonal average?
2. Is there a significant linear trend across a group of years in a seasonal average, where the time-scale of the trend is chosen by the investigator? This question can be posed in a moving-window manner across a time-series.

For all analyses, the response data of interest are , , with their associated standard errors, . is the number of years of the study.

### Differences among years in seasonal averages

To test the null hypothesis that for years and , one can use simple normal distribution theory that has standard error . To test the null hypothesis that all means in a group of (>2) years are equal, one can use a simple case of the meta-analysis given in the next section or a general linear hypothesis test. The latter readily includes consideration of multiple testing, which is provided in the R package multcomp (Hothorn et al. [2008](#ref-Hothorn08)). While this package typically uses the results of a linear or generalized linear model fit as input, alternative values of and could be provided as inputs. Details are given in the Appendix.

### Linear trends across years at a chosen time-scale of interest

The standard meta-analysis mixed effects model to estimate linear trends is given as follows, using notation similar to (Sera et al. [2019](#ref-Sera19)):

where is the intercept, is the year, is the slope, is the random effect for year , and is the residual for year . Here the seasonal average for year is . The “residual”, , represents estimation error in , namely . The residuls is assumed to be normally distributed with mean 0 and variance , where the latter is estimated from the calculations above. The random effect, , is the difference between and and is considered the “residual” in the sense of unexplained variation not due to the estimation error. The random effect is assumed a normal distribution with mean 0 and unknown variance, , to be estimated.

The R package *mixmeta* was used to estimate the model in equation @ref{eqn:mixmet} (Sera et al. [2019](#ref-Sera19)). Results have a similar interpretation as those from regression analysis, but parameter estimates and their standard errors incorporate the known standard errors of the response values. Following meta-analysis theory, evaluates this model as a linear mixed effects model with some variance components fixed, whereas others are estimated (). The default estimation method for meta-analysis models is restricted maximum likelihood (REML).

We apply this method for windows of a chosen width of years. This approach evaluates whether there is a significant linear trend in chlorophyll over a chosen series of years, by providing a more comprehensive assessment of all the data in those years (i.e., the seasonal averages, plus their uncertainty). Some years may have had more sampling, resulting in smaller standard error for , while others may have had less sampling, result in higher standard error. The meta-analysis explicitly includes these differences in uncertainty magnitudes.

As a final analysis, the chosen windows for evaluating a trend for a selected seasonal average can be applied across the time series to assess periods of time within which trends could be significant or not. Although the initial window width and seasonal period to evaluate is chosen by the analyst, applying the method in a moving window approach reduces some of the ambiguity around points in the time series when trends may be changing with reasonable certainty. The moving window approach applies the meta-analysis model from left to right in the time series across the seasonal averages to obtain the slope and significance of the estimated trend. We choose a centered window where the model estimates are based on results at equal half-window widths to the left and right of a given year. Although a left- or right-centered window could also be applied, we limit the analysis to a centered window to demonstrate the concept.

# Results

# Discussion

# Conclusions

# Acknowledgments

# Figures



Figure 1: Observed chlorophyll concentrations for all stations in central and south San Francisco Estuary (18-36, Figure 2), with (a) annual summer/fall concentrations (Aug - Dec) and (b) monthly concentrations by decade.

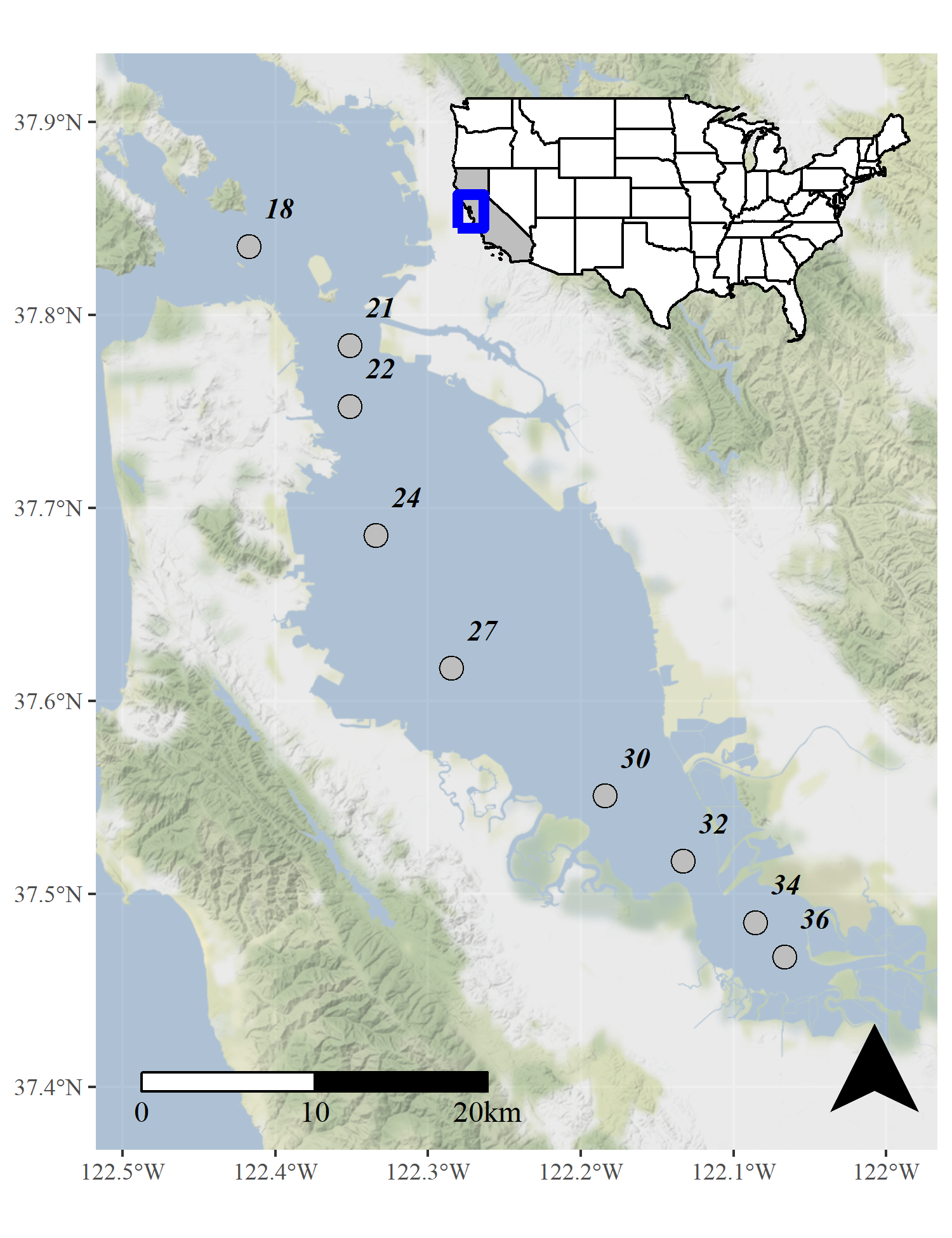


Figure 2: Station locations in the central and south San Francisco Estuary used for analysis. See Table 1 for station descriptions. Full dataset described in Schraga et al. ([2020](#ref-Schraga20)).

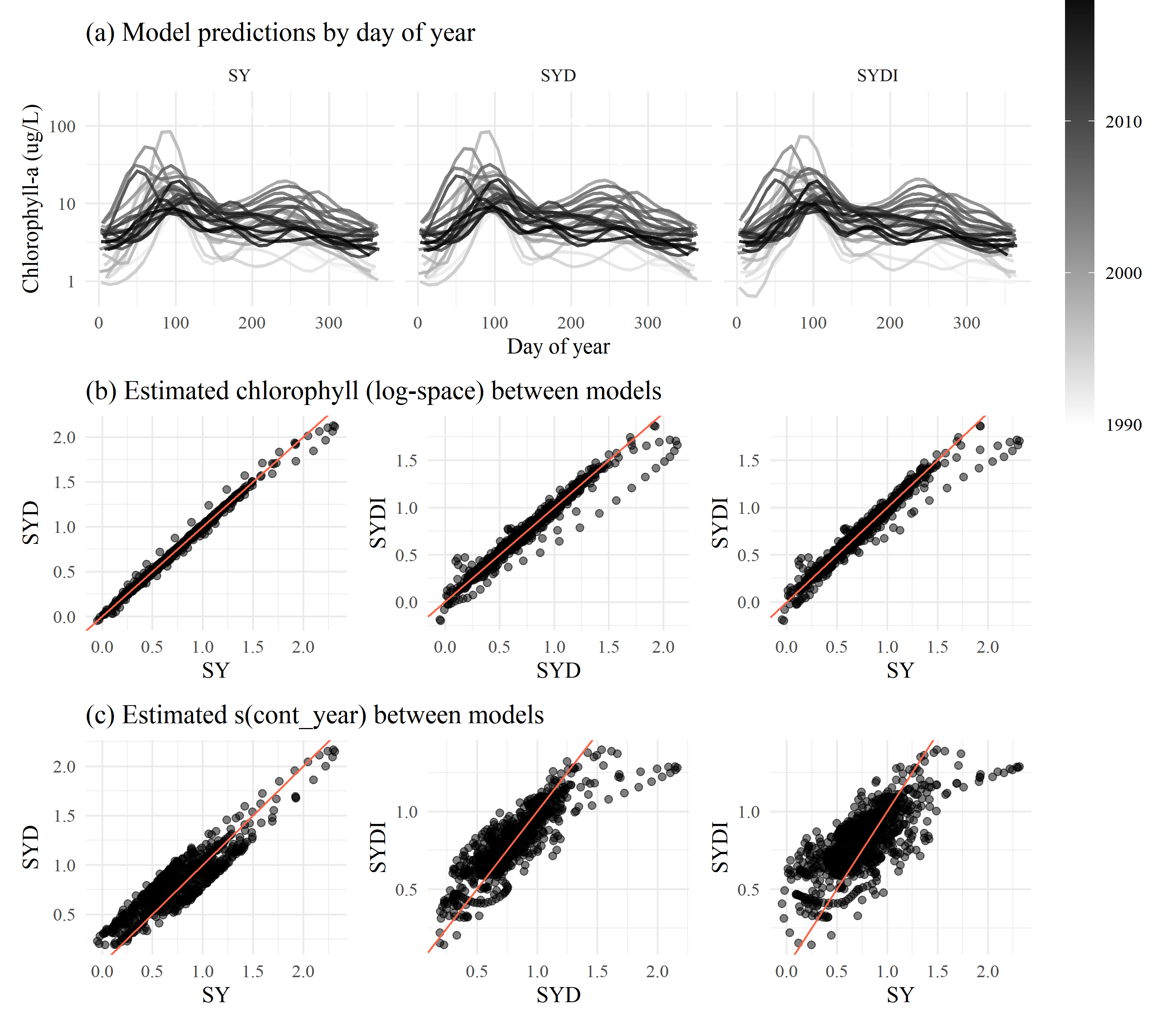


Figure 3: GAM output of estimated chlorophyll for models SY, SYD, and SYDI. Model S is identical to SY and is not shown. Plots in (a) show model predictions by day of year with separate lines for each year. Plots in (b) show pairwise comparisons of predicted chlorophyll between the models and plots in (c) show the same comparisons as in (b) but only for results from the esimated smoother for the cont\_year variable. The plots demonstrate that results between the models are comparable except for a few observations at extreme values(a), but they vary in the penalties applied to the basis functions for any particular smoother depending on which additive components are included in each model (b). The 1:1 lines are in red to faciliate comparisons.

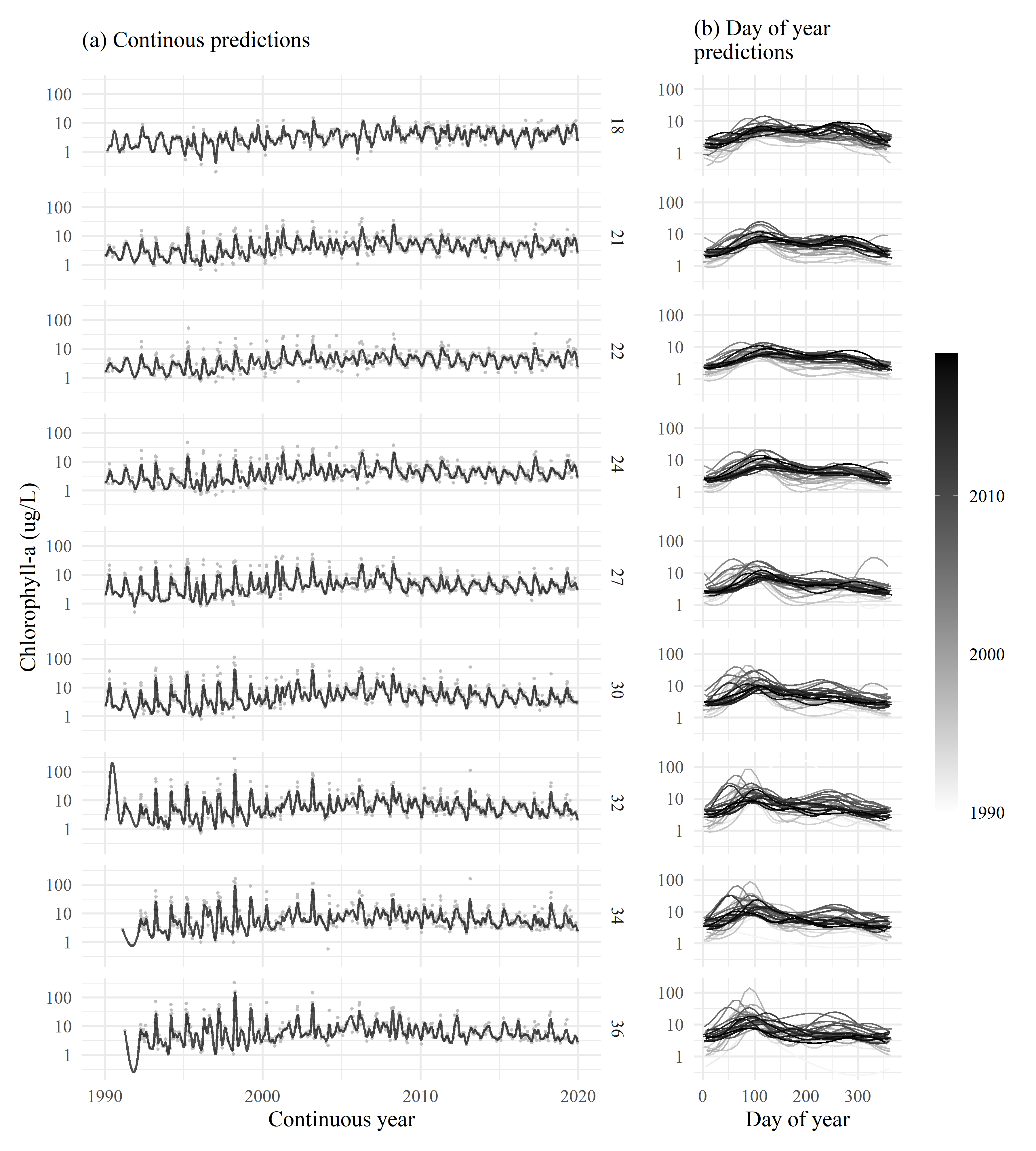


Figure 4: GAM predictions for all stations from north to south for model S. The results show (a) predictions across the time series and (b) predictions by day of year. Observed data in (a) are shown with the gray points. Station locations are in Figure 2.

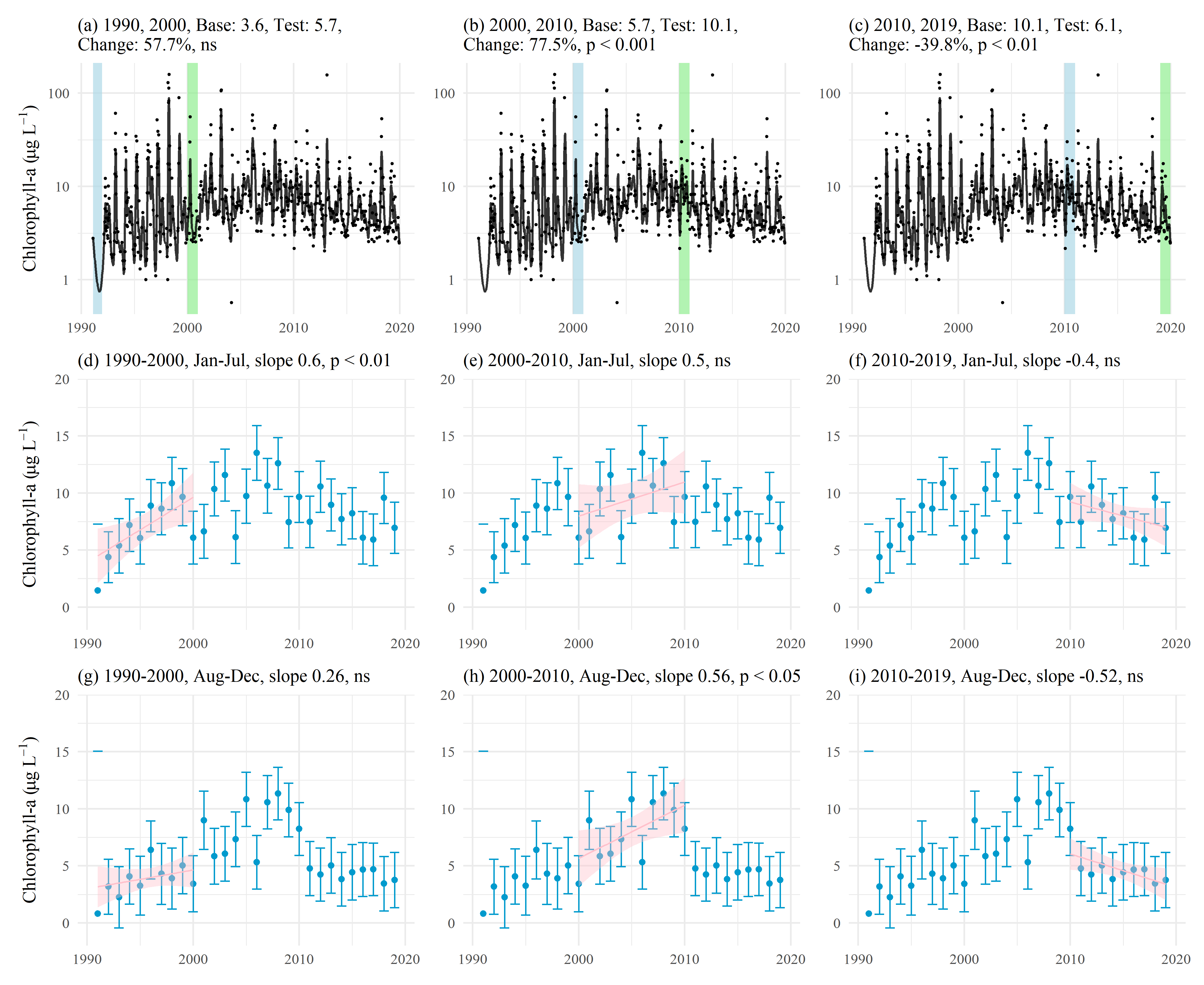


Figure 5: Examples of trend tests using results of GAM predictions for station 34 (Const seas\*). Plots (a), (b), and (c) show estimates of percent change as the difference of means between base and test years for chlorophyll. Plots (d) through (i) show seasonal averages and trend estimates over time. Plot (d), (e), and (f) show trend estimates for January through July and (g), (h), and (i) show trend estimates for August through December. The trend lines in (d) through (i) estimate the rate of change of chlorophyll per year, reported as the slope in the plot title. ns: not significant

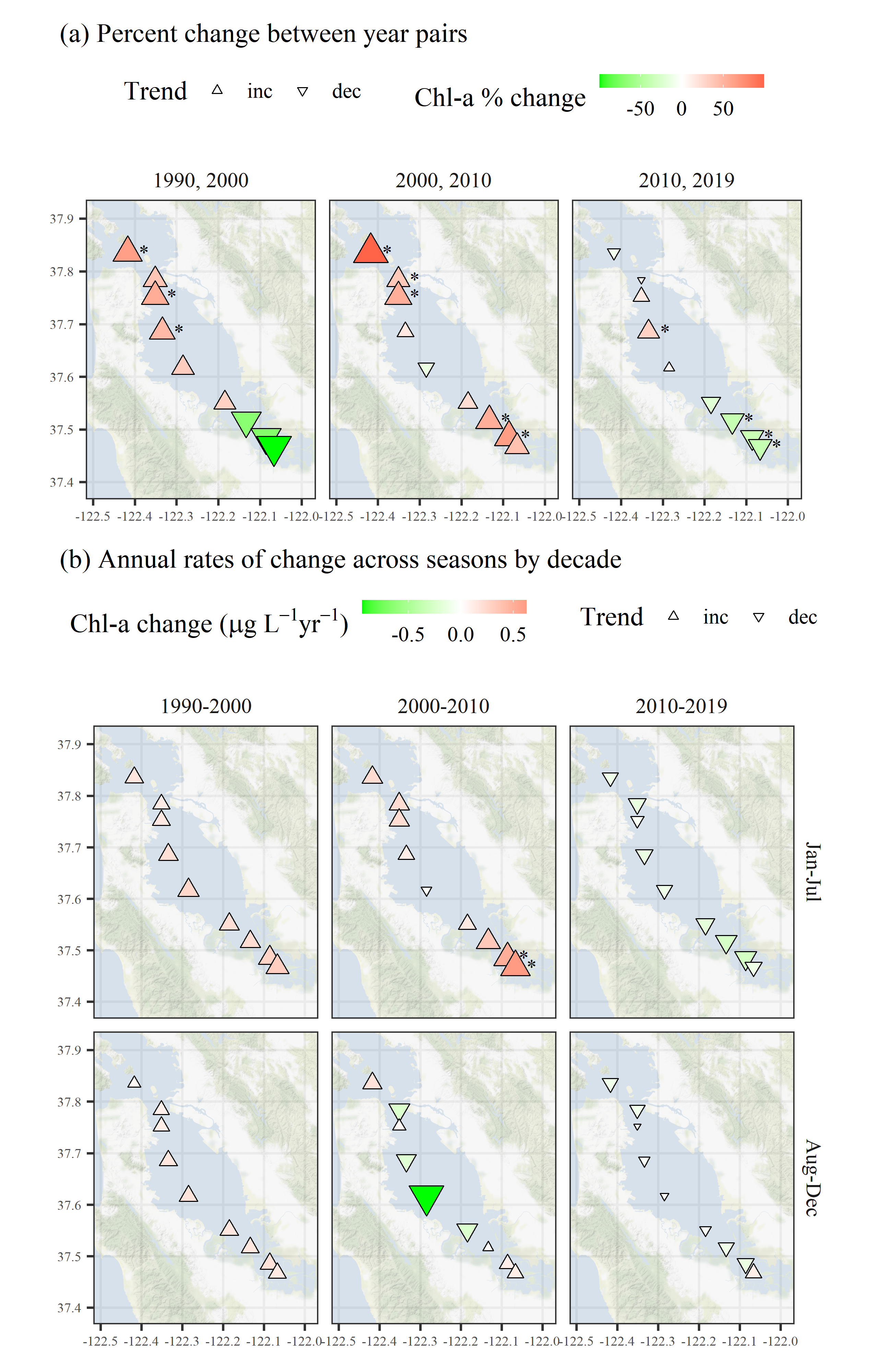


Figure 6: Trend estimates (a) between year pairs and (b) across seasons by decade for chlorophyll at each station. The top plots show percent change comparing the first and last year within a decade and the bottom plots show seasonal estimates of change per year for chlorophyll concentrations for each decade. Point type, shape, and color represent the direction and magnitude of an estimated trend. Trends with are marked with an asterisk. All results are from Model S.

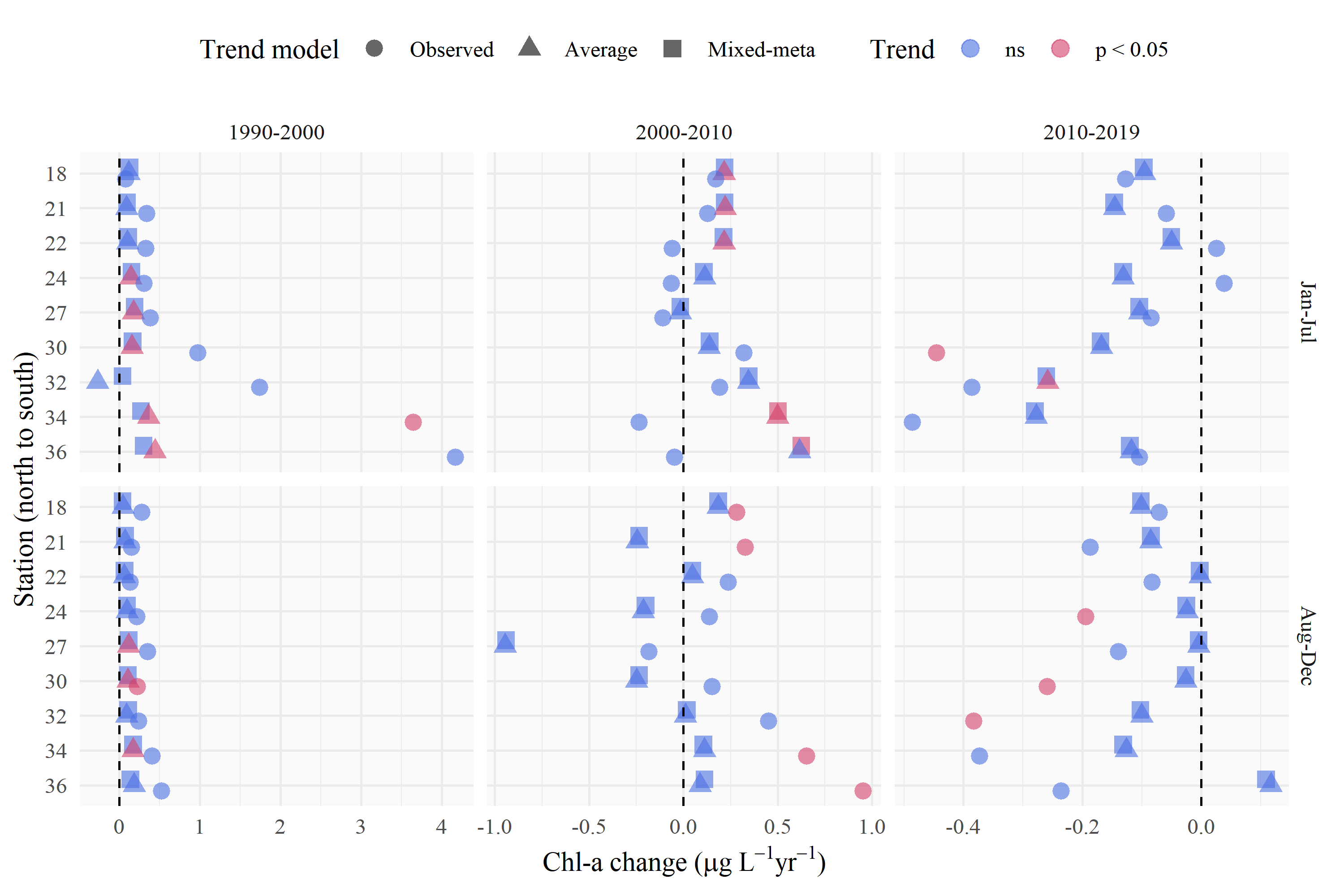


Figure 7: Trend estimate comparisons for three models applied to seasonal averages of chlorophyll in different annual periods at each station. The “observed” trend model is based on a linear fit to the annual averages of chloropyll from the observed data, the “average” trend model is based on a linear fit to the annual averages of chlorophyll from the GAM constant season\* model, and the “mixed-meta” trend model is based on a mixed-meta regression model fit to the annual averages of chlorophyll from the GAM constant season\* model. Values for each model are the slope esimates as annual change per year within each season, with color denoting significant trends.

# Tables

Table 1: Station locations, sample sizes, and summary values (median, minimum, maximum) for chlorophyll. Rows are arranged from north to south.

| Station | Latitude | Longitude | n | Med. | Min. | Max. |
| --- | --- | --- | --- | --- | --- | --- |
| 18 | 37.836 | -122.418 | 414 | 3.6 | 0.2 | 16.6 |
| 21 | 37.784 | -122.351 | 576 | 4.4 | 0.6 | 40.0 |
| 22 | 37.752 | -122.351 | 569 | 4.0 | 0.7 | 53.1 |
| 24 | 37.686 | -122.334 | 595 | 4.2 | 0.7 | 47.3 |
| 27 | 37.617 | -122.285 | 596 | 4.5 | 0.5 | 50.9 |
| 30 | 37.551 | -122.184 | 608 | 5.1 | 0.8 | 112.2 |
| 32 | 37.517 | -122.133 | 591 | 5.9 | 0.7 | 282.1 |
| 34 | 37.485 | -122.086 | 544 | 6.5 | 0.6 | 158.3 |
| 36 | 37.468 | -122.067 | 476 | 6.2 | 1.1 | 328.4 |

Table 2: Summary and details for each of the GAM structures. In practive, a sufficiently large number of knots provided to the additive terms will produce identical or comparable estimates for a response variable. The models differ in the allocation of penalties for the smoothness of each spline (s()).

|  |  |  |
| --- | --- | --- |
| GAM | Additive components | Details |
| S | s(cont\_year) | A single smoother over a continuous year variable |
| SY | cont\_year + s(cont\_year) | A linear continuous year variable and a single smoother over a continuous year variable |
| SYD | cont\_year + s(cont\_year) + s(doy) | A linear continuous year variable, a smoother over a continuous year variable, and smoother over a day of year variable |
| SYDI | cont\_year + s(cont\_year) + s(doy) + ti(cont\_year, doy) | A linear continuous year variable, a smoother over a continuous year variable, smoother over a day of year variable, and an interactin smoother across continuous year and day of year variables |

Table 3: Comparison of the four model structures (S, SY, SYD, SYDI) described in the first stage analysis of GAM estimation. The four models provide either identical or comparable ability to describe chlorophyll trends at an example station (32) in South San Francisco Bay. The models differ in additive smoothers and the amount of effective degrees of freedom (edf) in the smoothers (measure of wiggliness in each component), but the overall model predictions are comparable. GCV: generalized cross-validation score, R2: r-squared values for predictions, edf: effective degrees of freedom, F: F-statistic, p-val: probability value, \*\* p < 0.001

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| model | GCV | R2 | smoother | edf | F | p-val |
| S | 0.06 | 0.74 | s(cont\_year) | 242.23 | 6.27 | \*\* |
| SY | 0.06 | 0.74 | s(cont\_year) | 242.23 | 6.27 | \*\* |
| SYD | 0.06 | 0.74 | s(cont\_year) | 229.33 | 3.88 | \*\* |
|  |  |  | s(doy) | 8.07 | 0.11 | \*\* |
| SYDI | 0.06 | 0.73 | s(cont\_year) | 132.83 | 3.06 | \*\* |
|  |  |  | s(doy) | 9.55 | 0.80 | \*\* |
|  |  |  | ti(cont\_year,doy) | 74.02 | 0.68 | \*\* |

Table 4: Model performance statistics for each station as generalized cross-validation scores (GCV) and r-squared values.

|  |  |  |
| --- | --- | --- |
| station | GCV | R-squared |
| 18 | 0.04 | 0.78 |
| 21 | 0.04 | 0.70 |
| 22 | 0.05 | 0.59 |
| 24 | 0.04 | 0.69 |
| 27 | 0.05 | 0.72 |
| 30 | 0.05 | 0.74 |
| 32 | 0.06 | 0.74 |
| 34 | 0.07 | 0.68 |
| 36 | 0.07 | 0.73 |

Table 5: Seasonal trends from Const seas\* results by station and decade. Results show the annual slope estimate as chlorophyll change per year for each seasonal period across years (JFM: January, February, March; AMJ: April, May, June; JAS: July, August, September, OND: October, November, December). Significant trends based on mixed-meta regression analysis are indicated with asterisks at , \* , and \*\* .

|  |  | Chlorophyll change (ug/l) per year | | | |
| --- | --- | --- | --- | --- | --- |
| Station | Years | JFM | AMJ | JAS | OND |
| 18 | 1990-2000 | 0.08 | 0.09 | 0.11 | 0.03 |
|  | 2000-2010 | **0.25\*** | 0.22 | 0.17 | 0.13 |
|  | 2010-2019 | -0.12 | -0.11 | -0.13 | -0.06 |
| 21 | 1990-2000 | 0.07 | 0.07 | 0.14 | 0.06 |
|  | 2000-2010 | **0.26\*** | **0.25\*** | 0.07 | -0.26 |
|  | 2010-2019 | -0.16 | -0.17 | -0.11 | -0.06 |
| 22 | 1990-2000 | 0.09 | 0.09 | 0.12 | 0.06 |
|  | 2000-2010 | 0.21 | 0.22 | 0.16 | -0.01 |
|  | 2010-2019 | -0.09 | -0.08 | -0.01 | -0.01 |
| 24 | 1990-2000 | 0.1 | 0.11 | 0.19 | 0.08 |
|  | 2000-2010 | 0.21 | 0.15 | -0.01 | -0.21 |
|  | 2010-2019 | -0.14 | -0.16 | -0.07 | -0.02 |
| 27 | 1990-2000 | 0.13 | 0.15 | 0.24 | 0.11 |
|  | 2000-2010 | 0.22 | 0.1 | -0.49 | **-0.58\*** |
|  | 2010-2019 | -0.11 | -0.11 | -0.06 | -0.02 |
| 30 | 1990-2000 | 0.16 | 0.15 | 0.18 | 0.12 |
|  | 2000-2010 | **0.36\*** | 0.19 | -0.05 | -0.19 |
|  | 2010-2019 | -0.23 | -0.19 | -0.08 | -0.04 |
| 32 | 1990-2000 | -1.05 | -0.42 | 0.14 | 0.13 |
|  | 2000-2010 | **0.43\*** | 0.34 | 0.23 | -0.1 |
|  | 2010-2019 | -0.27 | -0.26 | -0.15 | -0.14 |
| 34 | 1990-2000 | **0.43\*** | 0.25 | 0.22 | 0.23 |
|  | 2000-2010 | **0.49\*\*** | **0.46\*** | **0.39\*** | -0.07 |
|  | 2010-2019 | -0.26 | -0.27 | -0.15 | -0.18 |
| 36 | 1990-2000 | **0.38\*** | 0.24 | 0.23 | 0.19 |
|  | 2000-2010 | 0.46 | 0.54 | **0.48\*** | -0.03 |
|  | 2010-2019 | -0.13 | -0.11 | 0 | 0.05 |

# Supplement

When select = TRUE is included, the comparison between models S and SY changes. This option tells mgcv to penalize the linear term’s coefficient in the spline. This would be appropriate if cont\_year was an explanatory variable subject to variable selection – which it is not here – but it would be a strange choice if one is explicitly including a linear term for cont\_year. If select = TRUE is used, models S and SY would still be effectively equivalent, but it would appear by AIC selection that one model is superior. That would be an artifact of the strange choice in model SY to include a linear trend in cont\_year both as a separate term and as part of the spline, with the latter subject to penalization. A link to an RMarkdown file demonstrating this issue is provided.

# References

Beck, M. W., and J. D. Hagy III. 2015. Adaptation of a weighted regression approach to evaluate water quality trends in an estuary. Environmental Modelling and Assessment 20:637–655.

Beck, M. W., and R. R. Murphy. 2017. Numerical and qualitative contrasts of two statistical models for water quality change in tidal waters. Journal of the American Water Resources Association 53:197–219.

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning. 1990. STL: A seasonal-trend decomposition procedure based on Loess. Journal of Official Statistics 6:3–73.

Cloern, J. E., and A. D. Jassby. 2010. Patterns and scales of phytoplankton variability in estuarine-coastal ecosystems. Estuaries and Coasts 33:230–241.

Cloern, J. E., and T. S. Schraga. 2016. USGS measurements of water quality in San Francisco Bay (CA), 1969-2015: U.S. Geological Survey data release. https://doi.org/10.5066/F7TQ5ZPR.

Cloern, J. E., T. S. Shcraga, E. Nejad, and C. Martin. 2020. Nutrient status of San Francisco Bay and its management implications. Estuaries & Coasts 43:1299–1317.

Cumming, G. S., D. H. M. Cumming, and C. L. Redman. 2006. Scale mismatches in social-ecological systems: Causes, consequences, and solutions. Ecology and Society 11:14.

Forbes, D. J., and Z. Xie. 2018. Identifying process scales in the Indian River Lagoon, Florida using wavelet transform analysis of dissolved oxygen. Ecological Complexity 36:149–167.

Gasparrini, A., B. Armstrong, and M. G. Kenward. 2012. Multivariate meta-analysis for non-linear and other multi-parameter associations. Statistics in Medicine 31:3821–3839.

Hafen, R. P. 2010. Local regression models: Advancements, applications, and new methods. PhD thesis, Purdue University, West Lafayette, Indiana.

Haraguchi, L., J. Carstensen, P. C. Abreu, and C. Odebrecht. 2015. Long-term changes of the phytoplankton community and biomass in the subtropical shallow Patos Lagoon Estuary, Brazil. Estuarine, Coastal and Shelf Science 162:76–87.

He, S., S. Mazumdar, and V. C. Arena. 2006. A comparative study of the use of GAM and GLM in air pollution research. Environmetrics 17:81–93.

Helsel, D. R., R. M. Hirsch, K. R. Ryberg, S. A. Archfield, and E. J. Gilroy. 2020. Statistical methods in water resources. Page 458. 2nd editions. U.S. Geological Survey Techniques; Methods, book 4, chapter A3, version 1.1, Reston, Virginia.

Hirsch, R. M., S. A. Archfield, and L. A. De Cicco. 2015. A bootstrap method for estimating uncertainty of water quality trends. Environmental Modelling and Software 73:148–166.

Hirsch, R. M., D. L. Moyer, and S. A. Archfield. 2010. Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay river inputs. Journal of the American Water Resources Association 46:857–880.

Hirsch, R. M., J. R. Slack, and R. A. Smith. 1982. Techniques of trend analysis for monthly water quality data. Water Resources Research 18:107–121.

Hothorn, T., F. Bretz, and P. Westfall. 2008. Simultaneous inference ingeneral parametric models. Biometrical Journal 50:346–363.

Lefcheck, J. S., D. J. Wilcox, R. R. Murphy, S. R. Marion, and R. J. Orth. 2017. Multiple stressors threaten the imperiled coastal foundation species eelgrass (*zostera marina*) in Chesapeake Bay, USA. Global Change Biology 23:3474–3483.

Morton, R., and B. L. Henderson. 2008. Estimation of nonlinear trends in water quality: An improved approach using generalized additive models. Water Resources Research 44:W07420.

Murphy, R. R., E. Perry, J. Harcum, and J. Keisman. 2019. A Generalized Additive Model Approach to evaluating water quality: Chesapeake Bay case study. Environmenal Modelling & Software 118:1–13.

Navarro, G., I. Caballero, L. Prieto, A. Vázquez, S. Flecha, I. E. Huertas, and J. Ruiz. 2012. Seasonal-to-interannual variability of chlorophyll-*a* bloom timing associated with physical forcing in the Gulf of Cádiz. Advances in Space Research 50:1164–1172.

Pearce, J. L., J. Beringer, N. Nicholls, R. J. Hyndman, and N. J. Tapper. 2011. Quantifying the influence of local meteorology on air quality using generalized additive models. Atmospheric Environment 45:1328–1336.

Schraga, T. S., E. S. Nejad, C. A. Martin, and J. E. Cloern. 2020. USGS measurements of water quality in San Francisco (CA), beginning in 2016 (ver. 3.0, March 2020): U.S. Geological Survey data release. https://doi.org/10.5066/F7D21WGF.

Sera, F., B. Armstrong, M. Blangiardo, and A. Gasparrini. 2019. An extended mixed-effects framework for meta-analysis. Statistics in Medicine 38:5429–5444.

Stow, C. A., Y. Cha, L. T. Johnson, R. Confesor, and R. P. Richards. 2015. Long-term and seasonal trend decomposition of Maumee River nutrient inputs to western Lake Erie. Environmental Science and Technology 49:3392–3400.

Testa, J. M., R. R. Murphy, D. C. Brady, and W. M. Kemp. 2018. Nutrient- and climate-induced shifts in the phenology of linked biogeocemical cycles in a temperate estuary. Frontiers in Marine Science 5:1–15.

Urquhart, N. S., S. G. Paulsen, and D. P. Larsen. 1998. Monitoring for policy-relevant regional trends over time. Ecological Applications 8:246–257.

Wan, Y., L. Wan, Y. Li, and P. Doering. 2017. Decadal and seasonal trends of nutrient concentration and export from highly managed coastal catchments. Water Research 115:180–194.

Wood, S. N. 2003. Thin-plate regression splines. Journal of the Royal Statistical Society (B) 65:95–114.

Wood, S. N. 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. Journal of the American Statistical Association 99:673–686.

Wood, S. N. 2017. Generalized additive models: An introduction with r. Page 476. 2nd editions. Chapman; Hall, CRC Press, London, United Kingdom.

Yang, G., and D. L. Moyer. 2020. Estimation of nonlinear water-quality trends in high-frequency monitoring data. Science of The Total Enviornment 715:10.1016/j.scitotenv.2020.136686.