Appendix B: Methods to Develop Statistical Models to Explain and Predict Methylmercury Concentrations in Predatory Fish in California Reservoirs

Contents

Introduction

Water Board staff conducted an extensive review of the peer-reviewed literature to determine the factors that have been found important elsewhere in controlling fish methylmercury concentrations in lakes and reservoirs. Physical and chemical reasons why these factors may be correlated with fish methylmercury concentrations are summarized in the conceptual model (Chapter 4). Chapter 5 then summarizes the linkage analysis that establishes the quantitative relationships between predatory fish methylmercury concentrations and the environmental factors that control methylmercury production, bioaccumulation, and biomagnification in California reservoirs. This appendix provides a more detailed description of the statistical analyses and model development methods that form the basis of the linkage analysis.

The purpose of the statistical analyses and model development is to determine the environmental factors that have the strongest correlation with fish methylmercury concentrations. This information is used to develop quantitative links between fish concentrations and environmental factors, specifically aqueous methylmercury, chlorophyll, and sediment total mercury concentrations. These quantitative links are used to determine aqueous methylmercury and sediment total mercury concentration goals necessary to attain the proposed sport fish target of 0.2 milligrams per kilogram (mg/kg). These goals are evaluated in Chapter 7 for TMDL allocations and to predict fish methylmercury reductions associated with potential implementation actions.

In other words, the results of the linkage analysis guide the development of an implementation program to reduce fish methylmercury concentrations in California reservoirs (Chapter 9). Potential implementation actions could include reservoir management practices to reduce aqueous methylmercury, source control actions to reduce reservoir sediment mercury concentration where there is contamination, and fisheries management practices to increase primary production in oligotrophic reservoirs and reduce methylmercury bioaccumulation in the food web (see Chapters 4, 7, and 9 and Appendix A).

This appendix has six parts. The first four parts describe the environmental data used in the linkage analysis to characterize factors identified in Chapter 4, the computer software and macros used for statistical calculations, the correlations and quantitative relationships between individual factors and reservoir fish methylmercury, and the development of multiple linear regression models to explain and predict reservoir fish methylmercury concentrations. Part 5 describes the best model for predicting fish methylmercury concentrations. Part 6 describes the models used to determine goals and predictions for fish methylmercury impairment reductions associated with potential implementation actions. Water Board staff consulted with a University of California statistician to ensure that the analyses and conclusions presented in Chapter 5 and this appendix are robust.

Part 1. Data Preparation

Water Board staff collected water and sediment samples and compiled data from readily available reports and databases for statistical model development:

- Reservoir data such as fish, water, sediment, soil total mercury and methylmercury, chlorophyll a,[1] organic carbon, sulfate, and suspended sediment compiled from reports and databases including California Environmental Data Exchange Network (CEDEN), Federal Energy Regulatory Commission (FERC) re-licensing studies, and other environmental studies (Appendix Z).
- GIS analyses for reservoir spatial, morphological, and land use data.
- USEPA REMSAD model (Chapter 6) for atmospheric mercury deposition to reservoirs.
- NPDES permit information and California Integrated Water Quality System (CIWQS) for NPDES facility data (e.g., discharge volumes, mercury loading rates, number of discharges from each facility).

The linkage analysis assessed the influence of about 70 factors on fish methylmercury concentrations in California reservoirs (Table 5.1 in Chapter 5 and Table B.1 and Figure B.1 in this appendix). One hundred and twelve out of about 350 reservoirs with fish methylmercury data had sufficient information for use in the linkage analysis. Reservoirs assessed in the

[1] Chlorophyll *a* is referred to as simply “chlorophyll” concentrations for the remainder of this appendix.

linkage analysis were not chosen using a random sampling design but instead were selected based upon the availability of data. In addition, data were not available for all of the factors identified in the conceptual model. Insufficient information was available to assess the importance of dissolved organic carbon, pH, degree of anoxia, and food chain length. Evaluation of the linkage analysis results must consider these data limitations (see section 5.5 in Chapter 5).

Fish Data

As noted in Chapter 1, mercury is typically analyzed as "total mercury" in fish because of the additional cost required for methylmercury analysis. However, mercury exists almost entirely in the methylated form in fish. Consequently, even though the fish mercury data presented in this chapter were generated by laboratory analyses for total mercury, the data are described as "methylmercury concentrations in fish."

Fish methylmercury concentration data are available for 345 reservoirs and lakes throughout California (see Table 1.3 in Chapter 1 for a summary). More than half of these have at least 10 samples; however, nearly one third have only 1 or 2 samples (Table Z.1 in Appendix Z). About 80% of the fish methylmercury concentration data were collected since 2000, and more than half were collected in a three-year period, 2006 to 2008. Of the 345 reservoirs and lakes, 267 (77%) have at least 2 samples collected during the 2006–2008 period.

Reservoir-specific fish methylmercury concentrations standardized for length and species were used in the linkage analysis, rather than average methylmercury concentrations. Standardization reduces variance in fish methylmercury concentrations caused by species- and site-specific bioaccumulation rates, and differences in distribution of sampled fish sizes resulting from differences in sample design and fish present at the time of sampling.

Methods similar to Tremblay and others (1998) and Davis and others (2010) were used to determine length-standardized fish methylmercury concentrations. The standardized fish for most reservoirs was 350 mm length largemouth bass (LMB). The 350 mm length was selected because it represents the middle of the typical size distribution caught for mercury analysis at most California reservoirs. A length of 350 mm also is above the legal size limit of 305 mm for LMB in California. If no LMB were available for a given reservoir, then other predatory fish data were used to calculate standardized size fish methylmercury concentrations. Preference was given to smallmouth bass and spotted bass. If no predatory fish data were available, lower trophic level (TL) fish such as rainbow trout were used. Figure 5.1 identifies the reservoir-specific method for calculating length-standardized tissue concentrations and the species employed. Largemouth bass data were available for 78% of the 112 reservoirs, and other predatory fish were used in 19% of the reservoirs. Predatory fish data were not available for 3% of the 112 reservoirs, so rainbow trout, Sacramento sucker, or a combination of these two species was used.

In the final step of the linkage analysis, statistical models were used to determine aqueous methylmercury and sediment total mercury concentration goals necessary to attain the proposed sport fish target. The proposed sport fish target is an average of 0.2 mg/kg in a legal-sized mixture of TL4 fish ranging from 200 to 500 mm and TL3 fish ranging from 150-500 mm (see Chapter 3). However, the linkage analysis compares predicted standardized fish methylmercury concentrations to the proposed target. An additional analysis was conducted to determine the difference between standardized fish methylmercury concentrations and average methylmercury concentrations in legal-sized top trophic level fish TL4 (200-500) or TL3 (150-500 mm).

Figure 5.2 shows the correlation between standardized fish methylmercury concentrations and the average concentrations in top trophic level legal-sized 200-500 mm TL4 fish. The relationship is:

LN[Standardized fish MeHg] = -0.2321 + 0.8248 * LN(Top TL fish MeHg)

$R^2 = 0.82$, *n = 107 reservoirs*

Solving for standardized fish MeHg equivalent to 0.2 mg/kg in 200-500 mm top trophic level fish:

LN[Standardized fish MeHg] = -0.2321 + 0.8248 * LN(0.2 mg/kg)

Standardized fish MeHg = exp[-0.2321 + 0.8248 * LN(0.2 mg/kg)]

= 0.21 mg/kg

An average methylmercury concentration of 0.2 mg/kg in legal-sized top trophic level fish equates to virtually the same concentration in standardized fish methylmercury. Consequently, later in this chapter the sport fish target of 0.2 mg/kg is compared directly to model-predicted standardized fish methylmercury concentrations without adjustment.

This is a conservative comparison and incorporates an implicit margin of safety because:

- The sport fish target applies to all top trophic level species, while the fish data used to calculate average fish methylmercury concentrations for the linkage analysis are dominated by black bass (i.e., largemouth, smallmouth, and spotted bass). More than 70% of the 112 reservoirs included in the linkage analysis have only black bass data or the number of black bass samples exceeds the number of other TL4 fish samples by more than twofold.
- Largemouth bass and other black bass methylmercury concentrations are typically higher than concentrations in more commonly consumed TL4 species such as catfish. There are 19 reservoirs that have methylmercury data for at least three samples each of (a) 305-500 mm (legal catch) LMB or other black bass and (b) 150-500 catfish. In 12 of these reservoirs, average black bass methylmercury concentrations were about equal to or slightly higher than (up to two times) average catfish methylmercury concentrations. In six of these reservoirs, average black bass methylmercury concentrations were more than 2- to 32-times higher than average catfish methylmercury concentrations.

A linkage analysis based primarily on standardized fish methylmercury is a valid approach for TMDL calculations such as assimilative capacity and allocations. This approach has been used for mercury TMDLs developed elsewhere in the state (e.g., the Sacramento-San Joaquin River Delta and reservoirs in the Los Angeles region; Wood et al. 2010 and 2010b, USEPA 2012e). In addition, this is a conservative approach for reservoirs that do not have LMB because LMB are typically one of the species that bioaccumulate the highest levels of methylmercury. Thus, these TMDL calculations should also be protective of reservoirs that do not have LMB.

Water and Other Reservoir Data

The linkage analysis makes use of several types of concentration data: total mercury in reservoir sediment; methylmercury, total mercury, and chlorophyll *a* in reservoir water; and total mercury in upland watershed soils. Some of these data sets varied by many orders of magnitude, in which case geometric mean is a preferred method for estimates of central tendency. Accordingly, concentration data for each reservoir were summarized using geometric means if the data did not contain non-detect values (i.e., concentrations below the analytical method detection limits). However, if data contain non-detects, then it is not possible to calculate the geometric mean.

In the event of non-detects, the median concentration was preferred for estimates of central tendency. The Kaplan-Meier method was used to calculate the median when there were fewer than 50% non-detects, and by the robust regression on order statistics method when there were more than 50% non-detects. One half the detection limit was used as the concentration for samples collected at the few reservoirs where all data were non-detect values, or where there were too few samples with too large of a proportion of non-detect values (e.g., <10 samples in total and >60% non-detect values). All of the estimates of central tendency are labeled “geomean” in Table B.1 and elsewhere in this report.

Appendix Z identifies the sources of concentration data. It should be noted that much of the concentration data used for the analyses were collected by different entities for other purposes. Very little of the data were collected for the sole purpose of developing a statistical model to explain fish tissue methylmercury concentrations in California reservoirs. As a result, the amount of data for each reservoir varies greatly, e.g., from 1 aqueous methylmercury sample for Bridgeport Lake to 284 samples for Clear Lake. Water Board staff assumed that the data used for the summary statistics is the best estimate of typical reservoir conditions. It is understood that additional data may alter the geometric means for individual reservoirs used in these analyses. Even so, the factors identified by the linkage models (see Parts 5 and 6 of this appendix) are known from other studies to influence methylmercury accumulation in fish (see Appendix A and Chapter 4). Consequently, additional data are not expected to change the overall conclusions of the linkage analyses.

In addition, the linkage analysis includes the following information. More information about information sources is provided in Appendix Z.

- Annual water level fluctuation. Annual water level fluctuation is the water year reservoir maximum elevation minus the water year minimum elevation. The average annual fluctuation is the mean of available years of data. The number of years used to calculate the average ranged from 1 to about 25 years.
- Reservoir surface area and watershed area. Reservoir surface area was estimated from DWR databases (DWR 2010a and 2010b) and USGS topographic quadrangles. Watershed areas were estimated using USGS topographic quadrangles, 30-meter digital elevation models, and ArcGIS (Environmental Systems Research Institute, Inc.).
- Elevation. Elevation of the reservoir water surface was estimated from USGS topographic quadrangles and DWR databases (DWR 2010a and 2010b) and represents the normal pool elevation, spillway elevation, or dam crest elevation.

- Latitude and longitude. Latitude (north–south) and longitude (east–west) coordinates represent the reservoir center points.
- Dam height, year constructed, number of upstream dams, and maximum reservoir capacity. Dam height is a proxy for reservoir depth. Information about these features was compiled from DWR databases (DWR 2010a and 2010b).
- Land use. Land use information includes reservoir surface areas, number and density of historic gold, mercury, and silver mines in reservoir watersheds, and watershed land covers (e.g., percent forested and developed [urbanized] lands, wetlands, and open water). See Chapter 6 for a description of the sources of this information.
- Atmospheric mercury deposition. Area-weighted average atmospheric deposition rates for the reservoir water surface areas and watershed areas were determined using the USEPA's REMSAD model output for 2001 (see Chapter 6). Deposition rates were available for total mercury deposition, wet and dry deposition, and deposition attributed to California industrial emissions. Deposition loads to reservoir water surfaces and watersheds were calculated by multiplying the deposition rates by the reservoir and watershed areas, respectively.
- Facility discharges. Information was compiled for facilities with individual NPDES permits that discharge directly to and upstream of reservoirs. Facility discharge volumes, concentration data, and treatment processes were obtained from permit project files and CIWQS (see Chapter 6). Both annual and dry season discharges were evaluated. During the dry season, facility discharges may comprise a larger proportion of reservoir inflows. Dry season discharges are evaluated as average daily discharges and loads for October and November, typically the low flow period for California watersheds. In addition, facility types and treatment processes that are more likely to have elevated effluent methylmercury concentrations are specifically evaluated. Bosworth and others review of effluent methylmercury concentration data for >100 facilities in the Central Valley indicated municipal WWTPs typically contribute more discharge volume and methylmercury load than non-municipal facilities and tend to have higher and more variable effluent methylmercury concentrations (Bosworth et al. 2010). Bosworth and others (2010) also found that some municipal WWTP treatment processes—nitrification/denitrification, filtration, and ultraviolet disinfection—result in statistically significantly lower effluent methylmercury concentrations.

Table B.1 provides the summary data used in the linkage analysis and Table B.2 provides definitions of the codes used in Table B.1.

Data Transformations

For environmental factors (variables) that contained zero values, a constant equal to the smallest non-negative, non-zero value was added to all values for that variable so that Box-Cox Power transformations could be performed. The ratio of methylmercury to chlorophyll was calculated by dividing reservoir geomean aqueous methylmercury concentration by reservoir geomean chlorophyll concentration, and the Box-Cox Power transformation was performed subsequently. The final ratio represents the mass of water column methylmercury divided by the mass of chlorophyll.

The reservoir data were transformed using the Box-Cox Power transformation to normalize values for parametric statistical analyses. The Lambda transformation values in Table B.2 indicate the power to which the different types of data were raised for transformation. For data with a Lambda value equal to 0.00, the transformation is the natural logarithm. The transformed data were then assessed for normality with the Ryan-Joiner test (similar to Shapiro-Wilk). Unfortunately, a majority of the transformed variables were still not normal. Non-normality can reduce the power or skew the interpretation of some parametric statistical tests, so additional non-parametric tests also were performed. Transformed data are denoted by adding "_1" to the end of the variable code in Table B.3 an Figures B.1, B.2, and B.4, for example Chla_1.

In addition, for use in multivariable model development, variables were z-score standardized (mean centered and divided by the standard deviation) to give variables equal weights. Z-score standardized data are denoted by adding "-z" to the end of the variable code in Figures B.2 and B.4, for example Chla_1-z.

Almaden and Guadalupe Reservoirs have fish methylmercury concentrations that are much higher than elsewhere (Table B.1). These two reservoirs are located in the Guadalupe River watershed, which contains the largest mercury mining district in California; mining waste has been deposited in the reservoir sediment. These reservoirs are included in the overall analyses because they exist in California and represent the possible outcome of reservoir creation and operation coupled with upstream mining influences. However, additional analyses were conducted with these reservoirs excluded from the data set to observe whether their presence altered the overall conclusions. It is noted in the text when inclusion/exclusion of the two reservoirs changed the conclusions.

Part 2. Statistical Software and Macros

All statistical analyses were performed using MINITAB® Release 14.20. Macros developed by Dennis Helsel (practicalstats.com) were used in some of the analyses. Minitab macros used for the analyses include the following:

CROS.MAC v. 4.2. CROS draws a boxplot and normal or lognormal probability plot for left-censored data (i.e., non-detect values on the low end of the range). It estimates summary statistics (mean, standard deviation, median and interquartile range) for left-censored data assuming that non-detect values follow either a lognormal or normal distribution using the robust regression on order statistics (robust ROS, or the MR probability plot method of Helsel and Cohn, 1988). Using the optional PCTL subcommand, any other percentile may be estimated. Robust ROS and lognormal distribution were used for reservoir concentration data where greater than 50% of the data were non-detects.

KMSTATS.MAC v. 1.7. KMSTATS computes descriptive statistics for left- (or optionally, right-) censored data using the Kaplan-Meier method. For left-censored data it first flips the data, then computes the stats and "unflips" the resulting statistics back into their original units. The Efron bias correction is used when the lowest value in the data set is censored. Kaplan-Meier and Efron bias correction were used for reservoir concentration data where less than 50% of the data were non-detects.

CORRALL.MAC. CORRALL computes four correlation coefficients—PEARSONS r, SPEARMAN'S rho, KENDALL'S tau-A, and KENDALL'S tau-B—and their corresponding *p*-values for the 2-sided tests of significance. CORRALL also plots y vs x and fits a lowess smooth to the scatter plot.

PARTPLOT.MAC. PARTPLOT computes all the partial plots for specified explanatory variables in the context of multiple regressions. The output consists of n partial plots, one each of dependent versus one of the n explanatory variables. These plots are adjusted for the remaining explanatory variables, and the slope of the plot equals the slope printed in the multiple regression output.

SQRK.MAC. SQRK computes the squared-ranks test, a non-parametric test for equal variances among groups (Conover 1980). Its major advantage over Bartlett's test is its robustness for non-normal data. Bartlett's test will reject the null hypothesis of equal variance if the data are non-normal but not necessarily heteroscedastic; the squared-ranks test will not.

Part 3. Correlations

Parametric (Pearson's r) and non-parametric (Spearman's rho and Kendall's tau) correlation coefficients and their two-sided test of significance were calculated for each environmental factor (variable) to determine which are correlated with reservoir fish methylmercury concentrations (Table B.3). Pearson's correlation (*r*) results are noted here unless there are discrepancies in results between the parametric and non-parametric methods. The absolute value of some correlation coefficients is displayed in the text because some variables (e.g., aqueous methylmercury and chlorophyll) were inversely transformed, and for these a negative correlation coefficient would represent positive associations (i.e., both variables increase together when not transformed).

Though many of the variables appeared to be statistically significant ($p < 0.05$) based on Pearson's r, many of the predictor variables and fish methylmercury concentrations are not normally distributed. In addition, of the variables with normal distributions, only the predictor variable MeHg/Chl-a had equal variance with fish methylmercury concentration data (Levene's Test, $p > 0.05$). Consequently, caution should be used when interpreting the hypothesis tests for the Pearson's r correlation coefficients for the other environmental variables. However, most of the variables had statistically significant correlation coefficients using the non-parametric analyses. As a result, these associations likely do exist in this data set and the variables are further considered in the linkage analyses.

For the most part, the parametric and nonparametric results agreed on the relative strength of associations (Table B.3). Exceptions were chlorophyll (Chla), aqueous total mercury (ATHg), percent of watershed with low density development (Urban Low), reservoir maximum capacity (MaxCap), and ratio of watershed area to reservoir surface area (WA:SA) (Table B.3). Of these, all but WA:SA were statistically significant with parametric but not nonparametric tests.

Fish methylmercury concentrations were most strongly correlated (r = 0.72, $p < 10^{-6}$) with the ratio of unfiltered aqueous methylmercury to chlorophyll concentration. The ratio represents biodilution, the amount of methylmercury entering the base of the food web and available for biomagnification divided by the amount of carbon available for tissue growth. See Chapter 4 and Appendix A for a detailed literature review about the relationship between primary productivity and fish methylmercury concentrations.

Fish methylmercury concentrations also were correlated with other environmental factors (Table B.3), including the following in order of the strength of their associations: longitude; watershed soil mercury concentration; annual reservoir water level fluctuation;; chlorophyll *a* concentration; aqueous total mercury; and reservoir depth.

Fish methylmercury concentrations were not correlated ($p > 0.05$) with the magnitude of several other mercury source types: municipal and industrial facilities, urban runoff, watershed mine density, and upstream wetlands (Table B.3). The lack of a correlation with fish methylmercury suggests that methylmercury bioaccumulation in the food web is dominated by in-reservoir processes more than by upstream processes.

Fish methylmercury concentrations also were strongly correlated ($p < 0.05$) with concentrations of total mercury and methylmercury in water and sediment (Table B.3), as expected from the literature review. Likewise, the various forms of mercury concentrations in reservoirs were correlated with each other (Table B.4), suggesting internal cycling of mercury within the reservoirs ($|r|$ = 0.38–0.70) (e.g., more inorganic mercury is associated with more methylmercury). This supports the literature review finding that the magnitude of inorganic mercury contamination is important in determining fish methylmercury levels, and that reductions in aqueous and sediment inorganic mercury will result in reductions in fish methylmercury concentrations.

Aqueous methylmercury and sediment total mercury in reservoirs were positively associated with several upstream sources of mercury ($|r|$ = 0.30 – 0.51), e.g., watershed soil mercury concentrations and atmospheric deposition mercury loading rates (Table B.4). Additionally, aqueous total mercury was positively associated with watershed soil mercury concentrations (r = 0.40, $p < 0.05$). To a lesser extent, the magnitude of mercury sources were positively correlated to standardized fish methylmercury concentrations. These include the total mercury atmospheric loads (g/year) to a reservoir's surface from California emission sources (r = 0.2, $p < 0.05$).

These positive associations again reinforce the importance of the magnitude of sources of total mercury to controlling fish methylmercury concentration; however, the weakness of the associations, compared to other environmental factors (variables), suggests there may be other processes with a more direct and important influence on methylmercury bioaccumulation in fish. (See Chapter 7 for a comparison of upstream mine density and number of productive mines in watershed to reservoir bottom sediment mercury concentration.)

Aqueous total mercury concentrations correlation coefficients dropped considerably when the two Guadalupe watershed reservoirs were excluded. This suggests that the high aqueous total mercury and fish methylmercury concentrations in these two reservoirs could be the drivers of the associations in the larger data set. Conversely, the correlation coefficients increased for chlorophyll when these two reservoirs were excluded. It is possible that the high mercury concentrations in these to two reservoirs partially mask the influence that chlorophyll has on fish methylmercury concentrations.

No single environmental factor explained a majority of the variability in fish methylmercury concentrations. This suggests that a combination of factors influence fish methylmercury concentrations, and that a multiple variable analysis is required to explain more of the variability in reservoir fish methylmercury concentrations. In addition, many of the environmental factors are auto-correlated with each other and multiple regression analyses controls for this.

All the factors with significant parametric or non-parametric correlation coefficients were plotted against transformed standardized fish methylmercury concentrations so that the relationships could be visually assessed. All of the variables other than percent watershed developed-low intensity and reservoir maximum capacity are used to develop the multiple variable models. Percent watershed developed-low intensity and reservoir maximum capacity were removed from model development because further evaluations through regressions showed that the associations were not actually statistically significant or the residuals were not normally distributed. No procedures were used to control false discovery rate (e.g., Bonferroni correction) in the correlation analyses, so the possibility of false positives exist. However, the correlation analyses were used only as a screening tool to reduce the number of environmental variables used to develop the multiple linear regression models.

Part 4. Multiple Linear Regression Model Development

Seventeen variables were evaluated in multiple linear regression models to determine their influence on fish methylmercury concentrations (Table B.3). Best subsets regression was used to determine the combination of variables that explained the greatest amount of variability in fish methylmercury. Variables were z-score standardized (x - mean)/standard deviation) to give the coefficients equal weight. Z-scores were calculated for both the pre and post lambda transformed data, so that the validity of the transformations could be evaluated for each model. The results are presented for the regressions that were performed using natural log transformed standardized fish methylmercury concentrations values ($\lambda = 0$) because these models were more consistent with the model assumptions of residual normality and constant variance than using data transformed with Lambda of 0.26 as in previous evaluations. The best subsets regression output displayed the top two models for each set of variables. The overall measures of model quality (Mallows' C_p, prediction error sum of squares (PRESS), predicted R^2, and adjusted R^2) were included and were used to determine the "best" models. The best models explain the greatest amount of variability in fish methylmercury levels.

The best subsets regression limits the number of reservoirs included in the analyses to those with data for the all environmental variables considered in the analyses. The best subsets regression analysis was performed including chlorophyll and methylmercury concentrations as separate variables and together as a ratio. Water Board staff started by analyzing multiple linear regression models to determine the combination of variables resulting in the lowest Mallows' C_p value. Staff proceeded to include additional variables that resulted in the next lowest Mallows' C_p value, and then decrease the number of predictor variables until regression equations resulted in slope coefficients that were all statistically significant.

The best subsets regression analysis produced several statistically significant models that explained differing amounts of variability in fish methylmercury. The number of reservoirs included in the different models ranged from 26 to 50, depending on the availability of data for the suite of variables being evaluated. The amount of variability (adjusted R^2) in reservoir fish methylmercury concentrations explained by the different models ranged from 57 to 84%.

All the statistically significant regression models included some form of reservoir mercury concentration as aqueous methylmercury alone, aqueous methylmercury in a ratio with chlorophyll, aqueous total mercury, sediment total mercury, or watershed soil mercury. The models also included reservoir annual fluctuations as a positively correlated predictor variable, which indicates that the magnitude of reservoir fluctuations is likely an important environmental factor in predicting reservoir fish methylmercury concentrations. In addition, the models included reservoir chlorophyll concentration as a negatively correlated predictor, either independently or as a ratio with aqueous methylmercury, except for the models where sediment total mercury concentration was included. The environmental, biological, and geochemical significance of these variables are described in Chapters 4 and 5 and Appendix A.

Part 5. Best Model for Predicting Fish Methylmercury Concentrations

The regression model that explains the greatest amount of variability (adjusted R^2 and predicted R^2) in fish methylmercury concentrations with statistically significant partial correlation coefficients is displayed in Figure B.2, along with model diagnostics including plots of residuals. This model is referred to hereon as the “best model” (Model 1) because it explains the greatest amount of variability in fish methylmercury. All the variables in Model 1 are known from other studies to influence methylmercury accumulation in fish (see Chapters 4 and 5 and Appendix A). The best model (Model 1) to predict methylmercury concentrations in California reservoir fish is:

LN[Fish MeHg] = - 0.958 + 0.544 (Ratio [MeHg]/[Chl-a]_1-z) + 0.271 [ATHg]_1-z + 0.330 (AnnFLuc)_1-z

Adjusted $R^2 = 0.84$, $p < 0.001$, $n = 26$ reservoirs

Where:

Fish MeHg: Length-standardized methylmercury concentration (mg/kg, wet weight) in highest tropic level fish (typically 350 mm bass)
AMeHg: Aqueous methylmercury concentration (ng/L)
Chl-a: Chlorophyll *a* concentration (µg/L)
ATHg: Aqueous total mercury concentration (ng/L)
AnnFluc: Reservoir annual water level fluctuation (feet)

Model 1 shows minimal multicollinearity with a maximum variance inflation factor (VIF) of 2.2, indicating that there is not a high degree of correlation between predictor variables. Model 1 was cross-validated using PRESS to prevent over-fitting the model, and the predicted R^2 is displayed in Figure B.2. Model 1 meets the assumptions of residual normality (Ryan-Joiner normality test, $p > 0.1$, Figure B.2). Residuals exhibited constant variance in the fitted value versus residual plot (Figure B.2).

Partial residual plots were made using the transformed and untransformed predictor variables for Model 1 to validate the use of the transformed variables (Figure B.3). The partial residual plots developed using the transformed data resulted in more linear relationships between predictor variables and log-normally transformed fish methylmercury than the plots developed using the non-transformed data for all of the predictor variables. This suggests that transformations did assist in linearizing the relationships between the predictor variables and fish methylmercury concentrations when the other predictor variable influences are considered.

Three lines of evidence were assessed to evaluate whether the reservoirs included in Model 1 development were representative of the larger suite of mercury impaired reservoirs in California. First, the reservoirs included in Model 1 had a wide range of fish methylmercury concentrations, sediment mercury concentrations, sizes, locations, chlorophyll concentrations, atmospheric deposition amounts, and watershed mining densities (Table B.1).

Second, Water Board staff compared the median of standardized fish methylmercury concentrations of the reservoirs used to develop Model 1 and the median for the greater population of reservoirs. The median for reservoirs used to develop Model 1 was significantly greater than the median for all reservoirs with fish methylmercury data (Table B.5, Mann-Whitney Test, $p < 0.01$). However, no statistical differences (Mann-Whitney Test, $p>0.05$) were found between the median for reservoirs used to develop Model 1 and the 74 303(d)-listed reservoirs without adopted control programs, all 81 303(d)-listed reservoirs (including ones with adopted control programs), or the 149 reservoirs with standardized fish methylmercury concentrations that exceed the proposed sport fish target of 0.2 mg/kg.

Finally, the range of standardized fish methylmercury concentrations for the reservoirs used to develop Model 1 nearly covers the range for all reservoirs with available fish methylmercury data, 0.06 to 4.2 mg/kg versus 0.01 to 4.2 mg/kg, respectively (Table B.6). In addition, the range of standardized fish methylmercury concentrations for the reservoirs included in Model 1 covers the range of concentrations for 303(d)-listed reservoirs and other reservoirs with standardized fish methylmercury concentrations that exceed the proposed sport fish target.

Consequently, Model 1 may be applicable for describing important factors driving fish methylmercury levels in reservoirs used to develop the model and other reservoirs identified as mercury impaired in the future.

Part 6. Models Used to Determine Goals

One objective of the statistical analyses and model development is to identify the environmental factors that have the strongest correlation with fish methylmercury concentrations. Another objective is to develop quantitative links between fish methylmercury concentrations and environmental factors that can be used to (a) determine aqueous methylmercury and sediment total mercury concentration goals necessary to attain the proposed sport fish target of 0.2 mg/kg and (b) predict fish methylmercury reductions associated with potential implementation actions.

As noted in the previous section, other statistically significant models besides Model 1 were identified. Several of these models are considered in the following sections because they are useful for predicting fish methylmercury reductions associated with reductions in aqueous methylmercury and sediment total mercury, and increases in chlorophyll concentrations.

Aqueous Methylmercury and Chlorophyll

The best model (Model 1) to predict methylmercury concentrations in California reservoir fish described in the previous section confirms the importance of aqueous methylmercury levels and primary production in controlling fish methylmercury levels (see Chapter 4 and Appendix A). Model 1 is statistically significant ($p < 0.001$) and has an R^2 of 86%. Model 1 includes aqueous methylmercury as a ratio with chlorophyll. Model 1 may not be statistically appropriate to evaluate aqueous methylmercury as a variable separate from chlorophyll because Model 1 assessed aqueous methylmercury as a ratio with chlorophyll.

Consequently, to predict fish methylmercury reductions associated with reductions in aqueous methylmercury and chlorophyll concentrations, Water Board staff developed several multiple linear regression equations that include aqueous methylmercury and chlorophyll without the use of a ratio. The best fit model for predicting fish methylmercury concentrations without the use of a ratio (Model 2) is:

LN[Fish MeHg] = - 1.02 + 0.426*[Chl-a]_1-z - 0.721*[AMeHg]_1-z
+ 0.280*(AnnFluc)_1-z

Where:

Fish MeHg: Length-standardized methylmercury concentration (mg/kg, wet weight) in highest tropic level fish (typically 350 mm bass)
Chl-a: Chlorophyll *a* concentration (µg/L)
AMeHg: Aqueous methylmercury concentration (ng/L)
AnnFluc: Reservoir annual water level fluctuation (feet)

To provide a range of predictions, Water Board staff developed a new best fit model (Model 3) using chlorophyll *a* as the independent variable. Best subsets regression analyses were performed with fish methylmercury concentrations as a required predictor variable in the multiple linear regressions. As described in Chapter 4 and Appendix A, associations have been observed between fish methylmercury concentrations and primary productivity rates (or chlorophyll *a* concentrations) in many reservoirs and other aquatic systems. This indicates there is an underlying mechanism that affects both, and therefore their association can be predictive even if changes in fish methylmercury concentrations do not cause changes in chlorophyll concentrations.

Model 3 is based on a much larger sample size than Models 1 and 2 (43 compared to 26 and 35 reservoirs, respectively) and was developed explicitly to minimize the variance around chlorophyll levels. The regression equation for Model 3 is:

[Chl-a]_1 = 0.873 + 0.0682*LN[Fish MeHg]_z + 0.135*[AMeHg]_1-z
+ 0.0949*(%Veg)_1-z

Where:

%Veg: Watershed percent vegetation

Both Models 2 and 3 were developed using best subsets regression and, although they have lower R^2 values than Model 1 (62% and 58%, compared to 86%), they are statistically significant ($p < 0.001$) and have several advantages. Models 2 and 3 assess aqueous methylmercury and chlorophyll without the use of a ratio. In addition, they include more reservoirs in their development: 35 and 43 reservoirs for Models 2 and 3 respectively, compared to 26 reservoirs for Model 1.

Water Board staff used Models 2 and 3 to predict the aqueous methylmercury and chlorophyll concentrations that would be necessary to attain the sport fish target of 0.2 mg/kg. Differences in model predictions for aqueous methylmercury concentrations are discussed in Chapter 5 (section 5.2). Differences in model predictions for chlorophyll concentrations are discussed Appendix A.

The model equations were used to calculate predictions only for reservoirs used to develop each model. The PRESS cross-validation results suggests that the models may provide good estimates for extrapolation; consequently, while Water Board staff limited prediction calculations to the reservoirs used to develop each model, the results can be used to determine goals for the remaining impaired reservoirs.

To further assess if the reservoirs used to develop Models 2 and 3 could be representative of mercury impaired reservoirs not included in model development, Water Board staff compared the median of standardized fish methylmercury concentrations of the reservoirs used to develop Models 2 and 3 and the median for the greater population of reservoirs. The medians for reservoirs used to develop Models 2 and 3 were significantly greater than the median for all reservoirs with fish methylmercury data (Table B.5, Mann-Whitney Test, $p < 0.01$). However, no statistical differences (Mann-Whitney Test, $p > 0.05$) were found between the medians for reservoirs used to develop Models 2 and 3 and the 74 303(d)-listed reservoirs without adopted control programs, all 81 303(d)-listed reservoirs (including ones with adopted control programs), or the 149 reservoirs with standardized fish methylmercury concentrations that exceed the proposed sport fish target of 0.2 mg/kg.

Likewise, the range of standardized fish methylmercury concentrations for the reservoirs used to develop Models 2 and 3 nearly cover the range for all reservoirs with available fish methylmercury data, 0.06 to 4.2 mg/kg versus 0.01 to 4.2 mg/kg, respectively (Table B.6). In addition, the range of standardized fish methylmercury concentrations for the reservoirs included in Models 2 and 3 cover the range of concentrations for 303(d)-listed reservoirs and other reservoirs with standardized fish methylmercury concentrations that exceed the proposed sport fish target. The reservoirs used to develop the statistical models to predict aqueous methylmercury and chlorophyll concentrations necessary to attain the proposed sport fish target have similar fish methylmercury concentrations and distributions as the reservoirs that are of concern for this control program.

Sediment Mercury

Reservoir sediment mercury had the second strongest positive correlation of any single factor with reservoir fish methylmercury concentrations. However, sediment total mercury was not a variable that automatically occurred in any of the statistically-significant multiple linear regression models described in previous sections. To create an equation for predicting fish methylmercury reductions associated with reductions in sediment mercury concentrations, best subsets regression analyses were performed with sediment total mercury as a required predictor variable in the multiple linear regressions. Model A in Figure B.4 is the multiple linear regression model with sediment total mercury that had the highest predicted R^2 (54%) and most reservoirs included (50 reservoirs). The regression equation for Model A is:

$$LN[Fish\ MeHg] = -0.883 + 0.475*[STHg]\_1\text{-}z + 0.358*(AnnFluc)\_1\text{-}z - 0.321*(Elevation)\_1\text{-}z + 0.210*(WA{:}SA)\_1\text{-}z$$

Where:

Fish MeHg:	Length-standardized methylmercury concentration in highest tropic level fish (typically 350 mm bass)
STHg:	Reservoir sediment total mercury concentration
AnnFluc:	Reservoir annual water level fluctuation
Elevation:	Reservoir water surface elevation

To provide a range of predictions, a regression based only on reservoir sediment mercury and standardized fish methylmercury (Figure 5.3) also was used for predictions and is labelled Model B in Table 5.3. Model B has an adjusted R^2 of 23% and includes 62 reservoirs. The regression equation for Model B is:

$$LN[Fish\ MeHg] = -0.3158 + 0.3134*LN[STHg]$$

Although Models A and B have lower R^2 values than Models 1, 2, and 3 (54% and 23% compared to 62%, 58%, and 86%, respectively), Models A and B are statistically significant ($p < 0.001$). Models A and B were used to calculate predictions only for reservoirs used to develop each model. The PRESS cross-validation results suggest that Model A may provide good estimates for extrapolation. In addition, reservoir sediment mercury had the second strongest positive correlation of any single factor with reservoir fish methylmercury concentrations. Consequently, while Water Board staff limited prediction calculations to the reservoirs used to develop each model, the results can be used to determine goals for the remaining impaired reservoirs.

To further assess if the reservoirs used to develop Models A and B could be representative of mercury impaired reservoirs not included in model development, Water Board staff compared the median of standardized fish methylmercury concentrations of the reservoirs used to develop Models A and B and the median for the greater population of reservoirs. The medians for reservoirs used to develop Models A and B were significantly greater than the median for all reservoirs with fish methylmercury data (Table B.5, Mann-Whitney Test, $p < 0.01$). However, no statistical differences (Mann-Whitney Test, $p>0.05$) were found between the medians for reservoirs used to develop Models A and B and the 74 303(d)-listed reservoirs without adopted control programs, all 81 303(d)-listed reservoirs (including ones with adopted control programs), or the 149 reservoirs with standardized fish methylmercury concentrations that exceed the proposed sport fish target of 0.2 mg/kg.

Likewise, the range of standardized fish methylmercury concentrations for the reservoirs used to develop Models A and B nearly cover the range for all reservoirs with available fish methylmercury data, 0.02 to 4.2 mg/kg versus 0.01 to 4.2 mg/kg, respectively (Table B.6). In addition, the range of standardized fish methylmercury concentrations for the reservoirs included in Models A and B cover the range of concentrations for 303(d)-listed reservoirs and other reservoirs with standardized fish methylmercury concentrations that exceed the proposed sport fish target. The reservoirs used to develop the statistical models to predict sediment mercury concentrations necessary to attain the proposed sport fish target have similar fish methylmercury concentrations and distributions as the reservoirs that are of concern for this control program.

Predictions

Models 2, 3, A, and B are used to (a) determine aqueous methylmercury and sediment total mercury concentration goals in Chapter 5 (and TMDL allocations in Chapters 7 and 8) necessary to attain the proposed sport fish target of 0.2 mg/kg and (b) predict fish methylmercury reductions associated with potential implementation actions. Chapter 5 describes how Water Board staff assessed the model predictions to develop goals. In addition, Table B.7 in this appendix provides reservoir-specific predictions for the amount of impairment reduction per different implementation action scenarios for each reservoir included in the models.

The amount of reduction for each reservoir is defined by the model-predicted difference in fish methylmercury concentrations before and after aqueous methylmercury reduction (or sediment total mercury reduction or aqueous chlorophyll increase) divided by the amount of reduction needed to meet the proposed sport fish target.

Model predictions for fish methylmercury impairment reduction were calculated for each reservoir using six scenarios. The first three scenarios (1a, 1b, and 1c) are based on reducing ambient annual geomean reservoir aqueous methylmercury concentration [AMeHg] to 0.02, 0.009, and 0.005 ng/L, respectively, which are the current and experimental analytical method detection limits. (See section 5.2.1 in Chapter 5 for more discussion about method detection limits.) The next two scenarios (2a and 2b) are based on reducing ambient geomean reservoir sediment total mercury concentration [STHg] to modern and natural background sediment and soil mercury levels. Natural background ranges 0.05–0.1 mg/kg in the Coast Ranges and 0.005–0.05 mg/kg elsewhere in California. Modern background typically ranges up to 0.3 mg/kg in the Coast Ranges and 0.1 mg/kg elsewhere in California. The upper extents of the background ranges are used for the prediction scenarios for each reservoir based on reservoir location. The last scenario (3) is based on increasing annual geomean reservoir chlorophyll concentration by twofold in oligotrophic reservoirs where chlorophyll ≤3 µg/L. (See Appendix A for more discussion about the selection of a chlorophyll threshold of 3 µg/L.)

Based on the amount of fish methylmercury reduction predicted by each model, Water Board staff assessed where reducing aqueous methylmercury to 0.009 ng/L, reducing sediment total mercury to background levels, or a combination of the two methods, are expected to resolve each reservoir impairment. This assessment is detailed in Table B.7 and illustrated on Figure B.5. Staff assumed that if a given method was predicted to reduce a reservoir impairment by ≥75%, that method likely could be used to resolve the impairment completely.

For some reservoirs there were not enough water data or sediment data to determine which methods may resolve the fish methylmercury impairment. Even so, for all reservoirs where both aqueous methylmercury data <u>and</u> sediment total mercury data were available, the models predicted that one or a combination of both of the methods could be used to resolve the impairments completely.

The model predictions for each reservoir fall into one of five general categories for methods of reducing fish methylmercury (MeHg) concentrations to attain the proposed sport fish target:

1. Either implement source controls to reduce reservoir geomean sediment mercury concentration (STHg) to natural background (NBG) levels **OR** implement reservoir water chemistry management practices to reduce geomean aqueous methylmercury concentrations to 0.009 ng/L. Model predictions indicate that both of these methods could completely resolve the fish methylmercury impairment in about 6% of the reservoirs evaluated.
2. Implement source controls to reduce reservoir geomean STHg to NBG or modern background (MBG) levels. Model predictions indicate that, at a minimum, source control may be useful for completely resolving the fish methylmercury impairment in another 14% of the reservoirs evaluated. Implementing reservoir water chemistry management practices to reduce geomean aqueous methylmercury concentrations may also be effective at reducing fish MeHg levels in these reservoirs, but aqueous MeHg and chlorophyll data were not available to enable evaluation of these approaches.
3. Implement reservoir water chemistry management practices to reduce geomean aqueous methylmercury concentrations to 0.009 ng/L. Model predictions indicate that, at a minimum, water chemistry management may be useful for completely resolving the fish methylmercury impairment in another 54% of the reservoirs evaluated. Source control is expected to reduce fish MeHg in many of these reservoirs, but not completely resolve the impairments.
4. A combination of both source controls to reduce reservoir STHg to NBG levels **AND** reservoir water chemistry management practices to reduce geomean aqueous methylmercury concentrations to 0.009 ng/L likely will be needed to attain the proposed sport fish target in about 2% of impaired reservoirs.
5. Additional water or sediment data are needed for about 24% of the impaired reservoirs to identify methods to completely resolve the fish methylmercury impairments.

Model predictions indicate a reduction of sediment total mercury concentrations to near zero anthropogenic inputs, in conjunction with attaining an aqueous methylmercury concentration of 0.009 ng/L, will achieve the proposed sport fish target in California’s impaired reservoirs. There are many reservoirs where reducing sediment mercury is not expected to make substantial fish methylmercury reductions because sediment mercury levels are already at background levels. Attaining the proposed sport fish target at these reservoirs likely will require implementing reservoir water chemistry management practices to reduce aqueous methylmercury concentrations and/or fisheries management practices to increase primary production in oligotrophic reservoirs and reduce methylmercury bioaccumulation in the food web. Collection of more data is expected to enable identification of additional methods for reducing fish methylmercury levels in many of the reservoirs.