CHAPTER 5

Evaluation of NPS Selenium Loading to the River Using a Mass Balance Model

5.1. Mass Balance Model Applied to the LARV

The purpose of the mass balance model in this thesis is to determine the mass of unaccounted for dissolved selenium being transported into and out of the the study reaches in the LARV. The mass transport is called the mass loading, where mass loading refers to mass entering the river channel and mass unloading refers to mass leaving the river channel. The basic concept of the mass balance lies in Equation 1. This equation is generalized for any dissolved or suspended constituent. All references to loadings (L) in this thesis is specific to selenium.

$$L = Q \cdot C \tag{1}$$

Where:

$L =$ Mass loading (when positive) or mass unloading (when negative).

$Q =$ Water flow rate.

$C =$ Concentration of the dissolved constituent under investigation.

Using this equation as a basis of understanding and applying it to Equation **??** in Chapter **??**, we arrive at a very basic equation mass balance model (Equation 2).

$$\frac{\Delta S_M}{\Delta t} = \tag{2}$$
$$Q_{in,US} \cdot C_{in,US} + \sum (Q_{in} \cdot C_{in}) + P \cdot C_P + R \cdot C_R + B \cdot C_B$$
$$-Q_{out,DS} \cdot Cout, DS - \sum (Q_{out} \cdot C_{out}) - E \cdot C_E - T \cdot C_T$$
$$-F \cdot C_F + X \cdot C_X$$

The C terms are the concentration for the dissolved constituent at each of the model design points. Simplifying by using Equation 1 results in Equation 3.

$$\frac{\Delta S_M}{\Delta t} = L_{in,US} + \sum L_{in} + L_P + L_R + L_B - L_{out,DS} - \sum L_{out} - L_E - L_T - L_F + L_X \tag{3}$$

Where:

$\frac{\Delta S_M}{\Delta t}$ = Mass storage change in the study reach.

$L_{in,US}$ = Mass loading to the river along the main stem of the river at the upstream end of the study reach.

$\sum L_{in}$ = Sum of the mass loadings to the river from tributaries and other gauged sources.

L_P = Mass loading to the river from precipitation.

L_R = Mass loading to the river from precipitation runoff off of adjacent land.

L_B = Mass loading to the river from subsurface flow.

$L_{out,DS}$ = Mass unloading along the main stem of the river at the downstream end of the study reach.

$\sum L_{out}$ = Sum of the mass unloadings from the river to canals and other gauged sinks.

L_E = Mass unloading from the river due to evaporation.

L_T = Mass unloading from the river due to plant transpiration.

L_F = Mass unloading from the river due to infiltration into subsurface flow.

L_X = Mass loading/unloading to/from unknown sources and sinks.

Equation 3 is a direct application of the two equations. Some of the terms need to be re-defined to be more appropriate for the situation. L_P is a possible process for many dissolved constituents, but it is unknown whether or not it occurs in any significant magnitude when the constituent is dissolved selenium. Selenium does not naturally occur in the atmosphere. It is transferred via biomethylation processes into the atmosphere. This is the L_T term. While the transport directions are the same, the L_E term is the direct volatilization of dissolved selenium species into the atmosphere. At this time, volatilization has been found to be an insignificant factor in the transport of dissolved selenium species into the atmosphere.

Using the same justifications as used in the previous chapter, Equation 3 is simplified to define the mass loading from and mass loading to unaccounted for non-point source (L_{UNPS}).

$$L_{UNPS} = \left(L_{out,DS} + \sum L_{out} - L_{in,US} - \sum L_{in}\right) + \frac{\Delta S_M}{\Delta t} \tag{4}$$

Where:

L_{UNPS} = The sum of mass gains from non-point sources and losses to non-point sinks

= $L_R + L_B - L_T - L_F + L_{U,in} - L_{U,out} + L_P - L_E L_X$.

This equation includes two terms not previously used. $L_{U,in}$ and $L_{U,out}$ are subsets of $\sum L_{in}$ and $\sum L_{out}$, respectively. They are the ungauged and non-point source river reach gains and losses.

5.2. Mass Storage Change

River water storage change calculations are developed and presented in Section **??**. Water storage change values between consecutive time steps are the basis for the calculations developed in this section. Stored selenium mass changes between two consecutive time steps is dependent on the change in river stored water volume and the concentration of selenium in the water. The stored selenium mass change in a given study region section is the sum of the stored selenium mass changes of the reaches within a study region river section. Individual reach selenium storage changes are calculated using equation 5.

$$\frac{\Delta S_{M,i}}{\Delta t} = \frac{C_{in,i} + C_{out,i}}{2} \cdot \frac{\Delta S_i}{\Delta t} \tag{5}$$

Where:

$\frac{\Delta S_{M,i}}{\Delta t}$ = Stored volume change in study reach segment i between time steps.

$C_{in,i}$ = Calculated dissolved selenium concentration at the upstream end of segment i.

$C_{out,i}$ = Calculated dissolved selenium concentration at the downstream end of segment i.

$\frac{\Delta S_i}{\Delta t}$ = Stored water volume change for segment i as calculated in Section **??**.

Ideally, the average selenate concentration should be calculated as the difference between the current calculation day stored mass and the prior calculation day stored mass. This requires that sufficient data is available to calculate the concentrations in the two consecutive calculation days. The actual stored water volume was not calculated and therefore the stored selenium mass could not be calculated. The methodology shown in equation 5

was assumed to perform as an approximation of the stored selenium mass change between two consecutive days.

The stored water volume change is significantly larger than the average concentration. Therefore it was assumed that small but significant changes in selenium concentrations between consecutive days would not significantly impact the stored selenium mass change.

5.2.1. *Solute Concentration Models*

One of the purposes of this thesis is to determine the selenium loading and unloading rate from unaccounted for sources/sinks. Elemental selenium does not exist in an aqueous form. The two dominant aqueous species are selenate and selnite. Of these two, selenate is the most dominant to the extent that in most cases, selenite is immeasureable. Therefore, throughout this thesis, selenate is used as the surrogate for all aqueous selenium species. References in this thesis to dissolved selenium and aqueous selenium are in fact discussing aqueous selenate.

All concentrations discussed in this thesis refer only to dissolved selenium concentration. Therefore, the typical method of denoting concentration with the aqueous species in the subscript (i.e. C_{Se}) is not going to be used in this thesis. Instead, subscripts are being reserved to designate the location within a study region for which the concentration is used. As such, Tables 5.2 and 5.3 shows the symbolic relationship between the various gauged flow and river segment storage changes and their associated dissolved selenium concentration for the USR and DSR, respectively. The concentration symbols use subscripts that designate the sample location as noted in Figures **??** and **??**.

The concentration associated with the $Q_{RFDRETCO}$ gauged flow is the same as that used for the $Q_{RFDMANCO}$ gauged flow. It was assumed that the two gauges would have the same dissolved selenium concentration due to their close proximity to each other. The concentrations in the USR with the designation $C_{ARK,d=x}$ are based on the concentrations for all locations in the main stem of the Arkansas River. Since it is known that the concentration is not constant along the entire reach, a variable was needed to differentiate the concentration for the various irrigation canal diversions. It was found that there was a slight correlation between the distance between the sample point and the upstream end of the study region

Table 5.2. USR gauged flow and river segment storage change symbolic relationship with dissolved selenium concentrations.

USR gauged flow and aqueous selenium concentration relationships.

Gauged Flow Symbol	Concentration Symbol
$Q_{ARKCATCO}$	C_{U163}
$Q_{ARKLASCO}$	C_{U201}
$Q_{CANSWKCO}$	C_{U74}
$Q_{CONDITCO}$	$C_{ARK,d=85.0}$
$Q_{FLSCANCO}$	$C_{ARK,d=16.4}$
$Q_{FLYCANCO}$	$C_{ARK,d=47.2}$
$Q_{HOLCANCO}$	$C_{ARK,d=12.5}$
$Q_{HRC194CO}$	C_{U207}
$Q_{RFDMANCO}$	C_{U167}
$Q_{RFDRETCO}$	C_{U167}
$Q_{TIMSWICO}$	C_{U60}
$Q_{LAJWWTP}$	$C_{LAJWWTP}$

USR river segment water volume change and aqueous selenium concentration relationship.

River Segment Symbol	US Concentration Symbol	DS Concentration Symbol
$\frac{\Delta S_A}{\Delta t}$	C_{U163}	$C_{ARK,d=12.5}$
$\frac{\Delta S_B}{\Delta t}$	$C_{ARK,d=12.5}$	$C_{ARK,d=16.4}$
$\frac{\Delta S_C}{\Delta t}$	$C_{ARK,d=16.4}$	$C_{ARK,d=47.2}$
$\frac{\Delta S_D}{\Delta t}$	$C_{ARK,d=47.2}$	$C_{ARK,d=85.0}$
$\frac{\Delta S_E}{\Delta t}$	$C_{ARK,d=85.0}$	C_{U201}

and the dissolved selenium concentration. Therefore, this distance was used and is noted in the subscript as x, where x denotes the distance between the irrigation canal diversion and the upstream end of the study reach. The La Junta WWTP does not have a stream gauge nor were selenium analyses performed by the university for this location. Plant operators

were kind enough to provide us with the total daily discharge from the plant in units of million gallons per day (mgd). This value was converted to appropriate units as used in the previous chapter. They also provided the monthly selenium analyses results. These analyses were performed in accordance with State directives. University researchers did not question the validity of the discharge or selenium concentration results.

Table 5.3. DSR gauged flow and river segment storage change symbolic relationship with dissolved selenium concentrations.

DSR gauged flow and aqueous selenium concentration relationships.

Gauged Flow Symbol	Concentration Symbol
$Q_{ARKLAMCO}$	C_{D101C}
$Q_{ARKCOOKS}$	C_{D106C}
$Q_{BIGLAMCO}$	C_{D23}
$Q_{BUFDITCO}$	$C_{ARK,d=37.7}$
$Q_{FRODITKS}$	C_{D106C}
$Q_{WILDHOCO}$	C_{D57}

DSR river segment water volume change and aqueous selenium concentration relationship.

River Segment Symbol	US Concentration Symbol	DS Concentration Symbol
$\frac{\Delta S_F}{\Delta t}$	C_{D101C}	C_{D104C}
$\frac{\Delta S_G}{\Delta t}$	C_{D104C}	C_{D106C}

As can be observed in Figure **??**, the flow gauge for Buffalo Ditch ($BUFDITCO$) has a sample location within close proximity on the channel. The concentrations associated with this location ($D36$), were found to be inconsistent with concentrations at $D104C$. A chance encounter with the individual who owns the land immediately adjacent to the gauge location informed us that a small spring discharged from his property into Buffalo Ditch between the gauge location and the sample location. The landowner informed us that the

spring discharged at approximately $0.056\,\mathrm{m^3\,s^{-1}}$ (2 cfs). He also stated that the flow rate was fairly constant throughout the year. Since there was a significant variance between the measured concentrations at sample points $D36$ and $D104C$ and since there is a known discharge into the channel with an unknown concentration, it was determined that the gauged flow $Q_{BUFDITCO}$ would be associated with the concentration $C_{ARK,d=37.7}$.

Dissolved selenium samples were taken as discussed in Chapter **??**. Samples were not taken for every time step in the study time frame. Concentration estimations for all sample locations except one were performed using linear regression. Linear regression models are defined as models where the functions of the predictor variables are not themselves variable. This is shown in equation 6. A regression model is considered linear if f_i does not contain any fitting parameters (β_i).

$$\hat{y} = \beta_0 + \beta_1 f_1 + \beta_2 f_2 + \ldots + \beta_n f_n \tag{6}$$

Where:

$\hat{y}$ = fitted or predicted value

β_i = fitting parameters

f_i = functions of the predictor variables x_i

Whenever possible, ordinary least squares regression was used to generate best fit equations with a given set of independent variables. Pearson's r-squared value is used as an initial goodness-of-fit value so that individual linear regression models can be evaluated both independently and comparatively with regards to how well they fit the data. R-squared values for linear models are positive, non-negative values less than one (1) with one (1) indicating a perfect fit. R-squared values account for the percent of the dependent variable that can be

accounted for in the independent variables. The adjusted r-squared value is calculated and compared to the r-squared value. This allows for some accounting for uncertainty when using multiple independent variables. If the two are considerably different, then the estimating model is missing an explanatory independent variable (Johnson and Wichern 2007).

The f-statistic was also calculated and compared to the critical f-statistic for each estimating model. An f-statistic greater than the critical f-statistic indicates that at least one of the explanatory independent variables is linearly associated with the calculated dependent variable. When comparing estimating models for suitability, the model with the greater f-statistic is more suitable (Johnson and Wichern 2007). The model f-statistic is returned as part of the statistics software linear regression summary. The critical f-statistic is calculated using the f-distribution and the desired significance level and the degrees of freedom.

The significance level, α, is closely tied to the desired confidence interval for this study. Considering the number and source of the input variables, it was considered desirable to have all models calculated to account for 95% of the variability. This means that 5% of the variability in any model can be attributed to chance. With all models being two-tailed, the central inter-percentile range (CIR) was calculated for the range between 2.5% and 97.5% as a means to comprehend the daily change in variability. This takes the 5% unaccounted variability and distributes it to the two tails. The α is not changed on account of the two-tailed nature of the models.

The two-sided p-value was used to determine if the estimating model independent variables were statistically significant. The p-values greater than α indicated that the independent variable was not significant and did not contribute significantly to describe the variability of the dependent variable. These variables were considered for removal during linear regression model optimization.

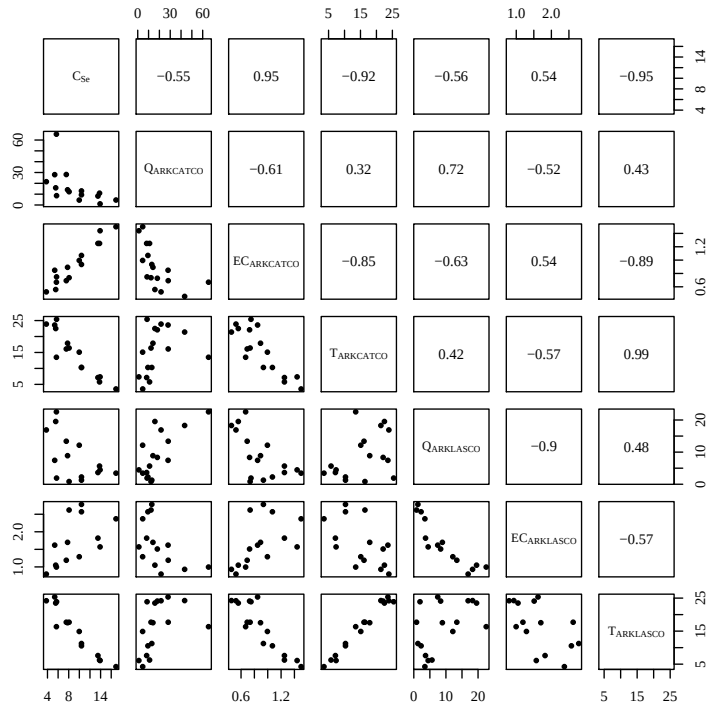
The selenium concentration field data collection effort provided excellent data for specific locations at specific times. This data needed to be expanded to include as many intermediate locations and times as possible to provide a more complete set of results. The methodology for determining the dissolved selenium concentration at various locations in both study region river sections is the same with only one exception. The set of starting independent variables, or starting terms, changes depending on the specific location, but the method of reducing the equation to the final equation is the same.

Estimating equations were calculated using multivariate linear ordinary least squares regression. The 'lm' function in the statistical software 'stats' package was used to fit linear models. Measured selenium concentration values were initially fitted to equations that included the average daily stream flow, EC, and water temperature values for the same day as the selenium concentration sample was collected. Equations were then refined through a process that will be discussed.

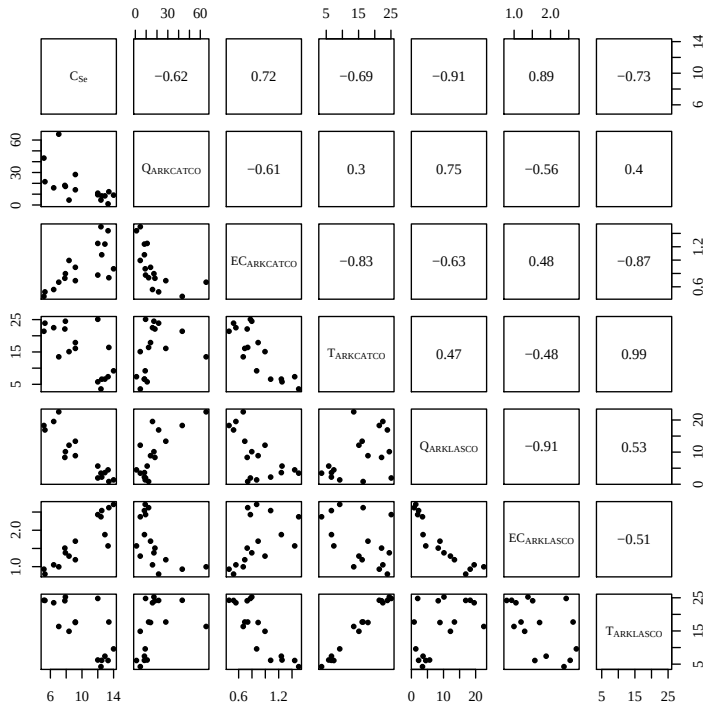
All data points were included in the linear regression even if some terms were missing. The 'lm' function has an argument that allows the user to determine what should happen if missing data is encountered. All analyses were performed such that data points with missing values were excluded from the analysis. As the number of terms was reduced during the equation refinement process, these excluded data points may or may not be included in the analysis. This allowed for the maximum number of data points without reforming the data set with each equation refinement iteration.

Determining which terms to include in each initial regression analysis began with visual analysis of an enhanced scatter-plot matrix of the selenium concentration response vector and the independent variable terms. Figures 2.1 and 2.1 contain the scatter-plot

matricies for all concentration points in the USR and DSR, respectively. The diagonal contains the variable names for the row and column. The lower triangle shows the individual variables when plotted against each other. Individual graphs are in appropriate unites for the investigated variables. Flow is in units of $m^3 s^{-1}$, EC is in units of $dS m^{-1}$, and temperature is in units of °C. The upper triangle presents the Pearson correlation value for the respective variable pair. Similar figures for the other regression analyses are included in the appendix.

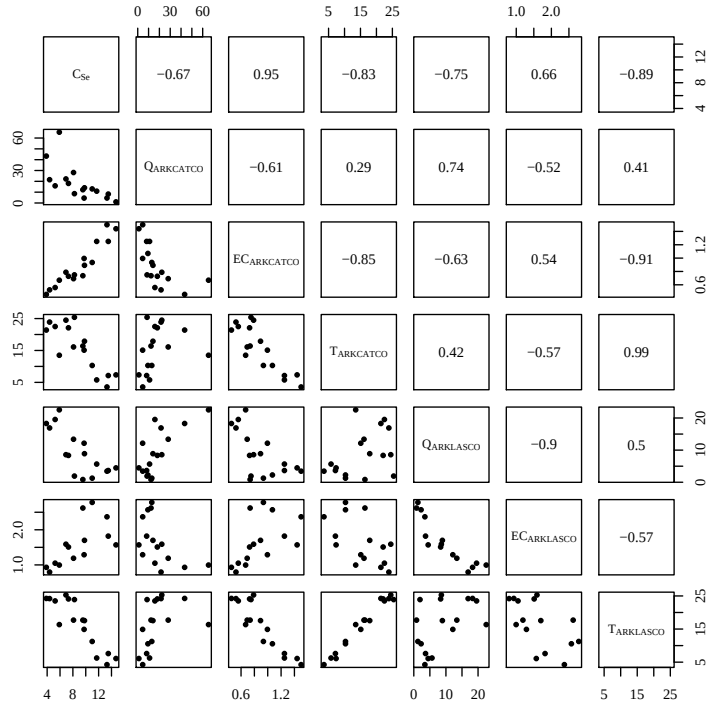


C_{U163}

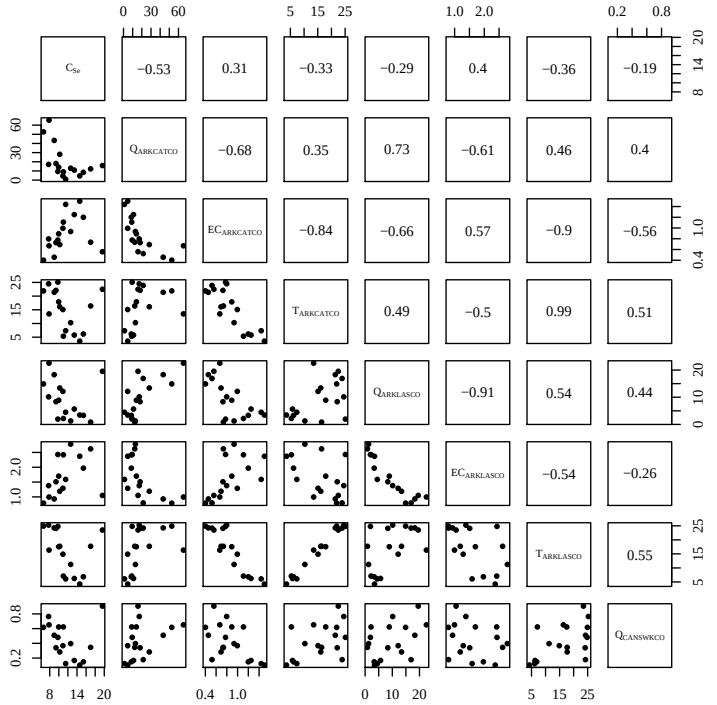


C_{U201}

Figure 2.1. Scatter-plot matricies of the input variables used to estimate dissolved selenium concentrations in the USR. Variable names are plotted down the diagonal. Values in the upper triangle are Pearson correlation values for the intersecting variables. Scatter-plots for the intersecting variables are plotted in the lower triangle. Scales are in the units for the given variable. C_{Se} is in $\mu g\,L^{-1}$. Q values are in $m^3\,s^{-1}$. EC values are in $dS\,m^{-1}$. T values are in °C.

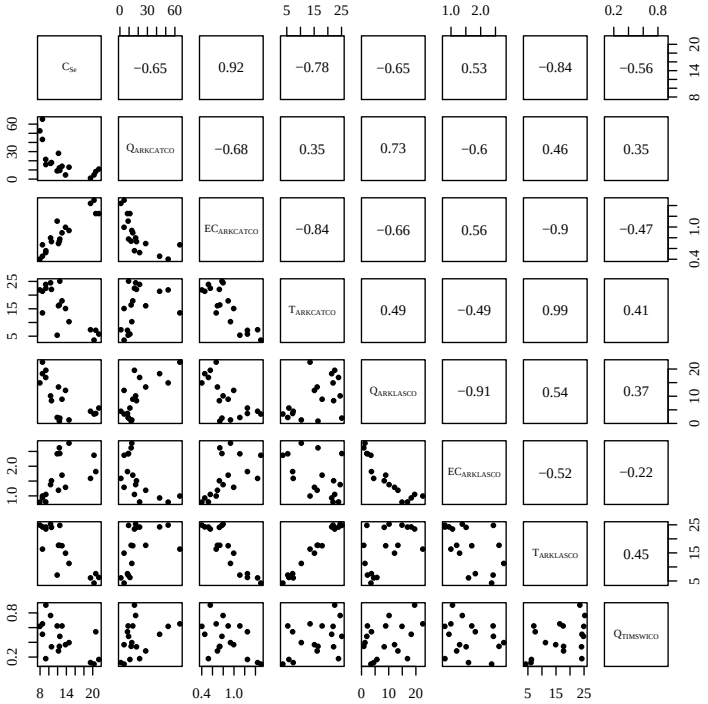


C_{U167}

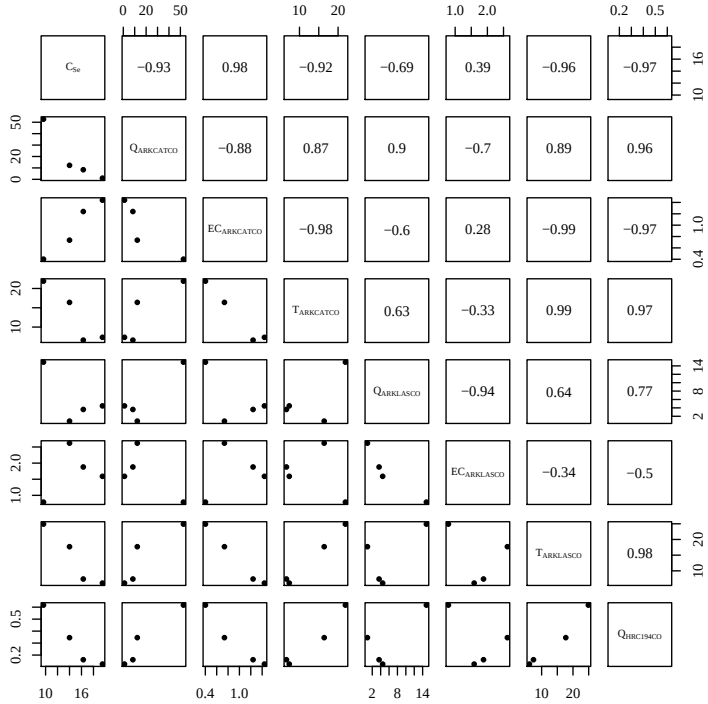


C_{U74}

Figure 2.1 (Cont). Scatter-plot matricies of the input variables used to estimate dissolved selenium concentrations in the USR.

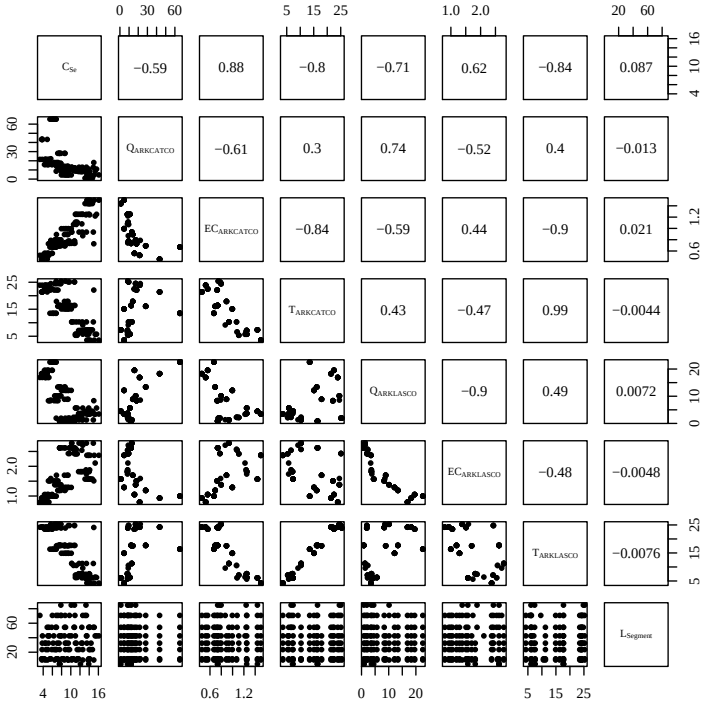


C_{60}

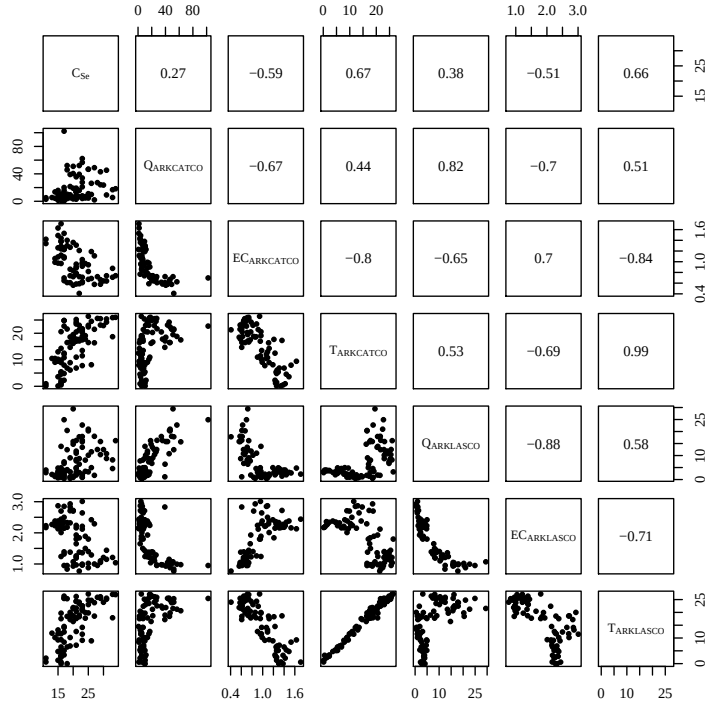


C_{U209}

Figure 2.1 (Cont). Scatter-plot matricies of the input variables used to estimate dissolved selenium concentrations in the USR.

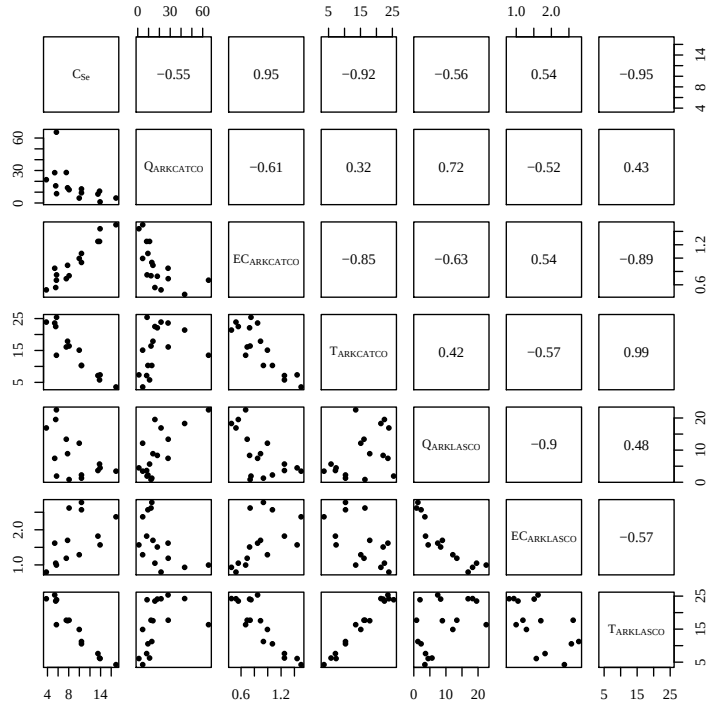


$C_{ARK,d=x}$

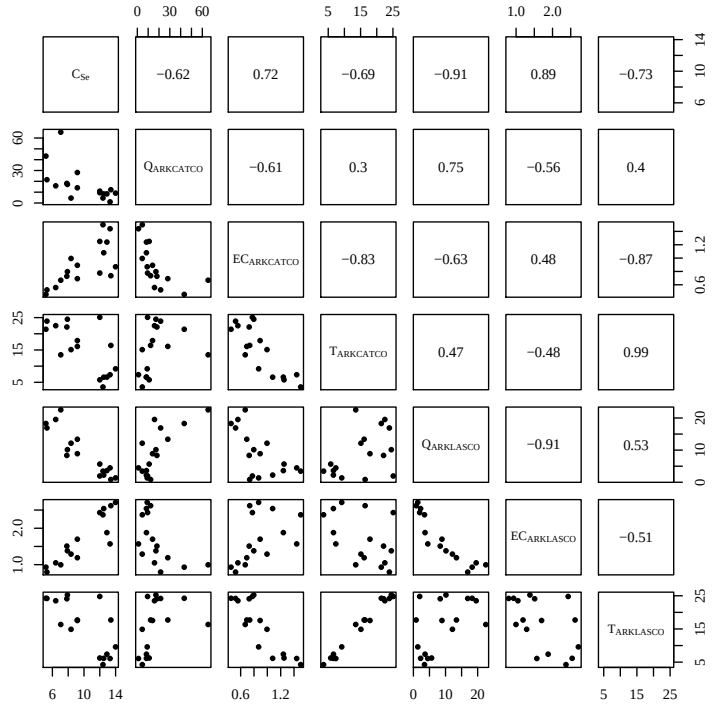


$C_{LAJWWTP}$

Figure 2.1 (Cont). Scatter-plot matricies of the input variables used to estimate dissolved selenium concentrations in the USR.

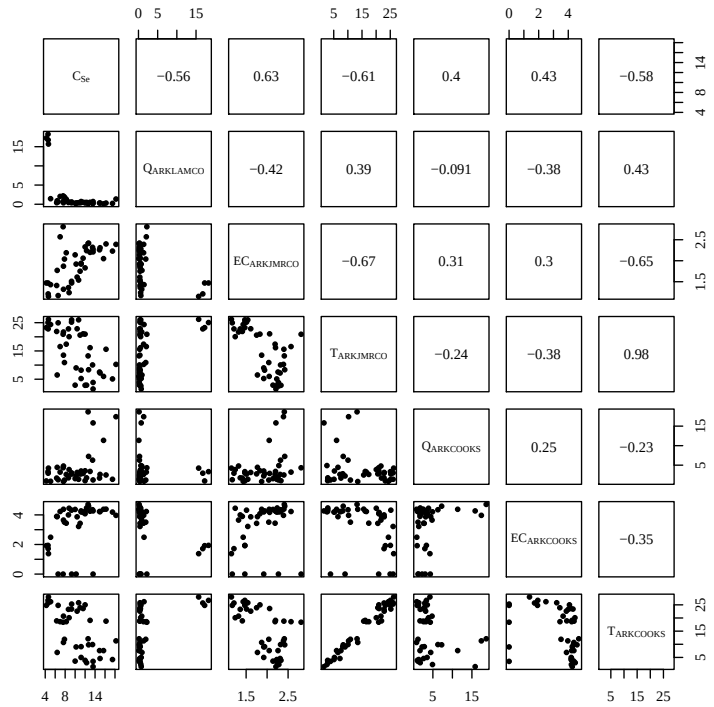


C_{U163}

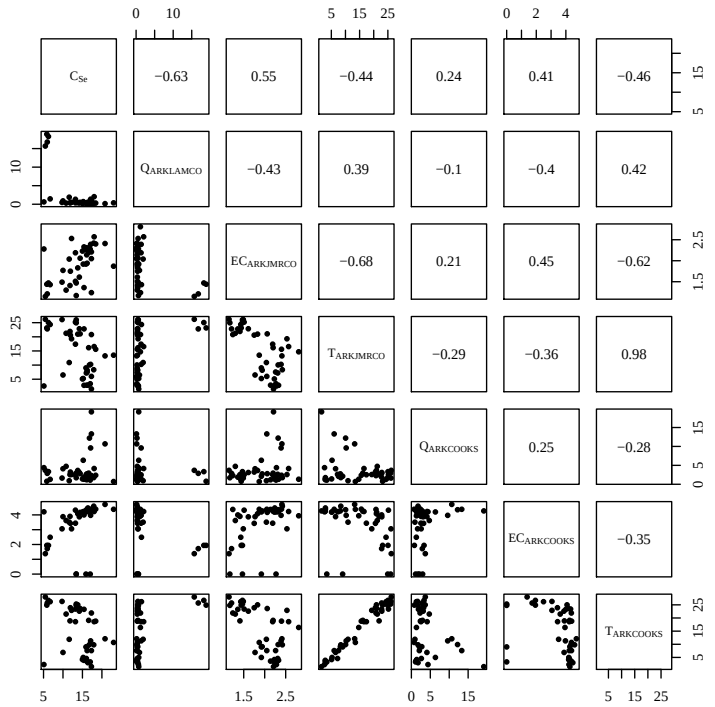


C_{U201}

Figure 2.2. Scatter-plot matricies of the input variables used to estimate dissolved selenium concentrations in the DSR. Variable names are plotted down the diagonal. Values in the upper triangle are Pearson correlation values for the intersecting variables. Scatter-plots for the intersecting variables are plotted in the lower triangle. Scales are in the units for the given variable. C_{Se} is in $\mu g\,L^{-1}$. Q values are in $m^3\,s^{-1}$. EC values are in $dS\,m^{-1}$. T values are in °C.

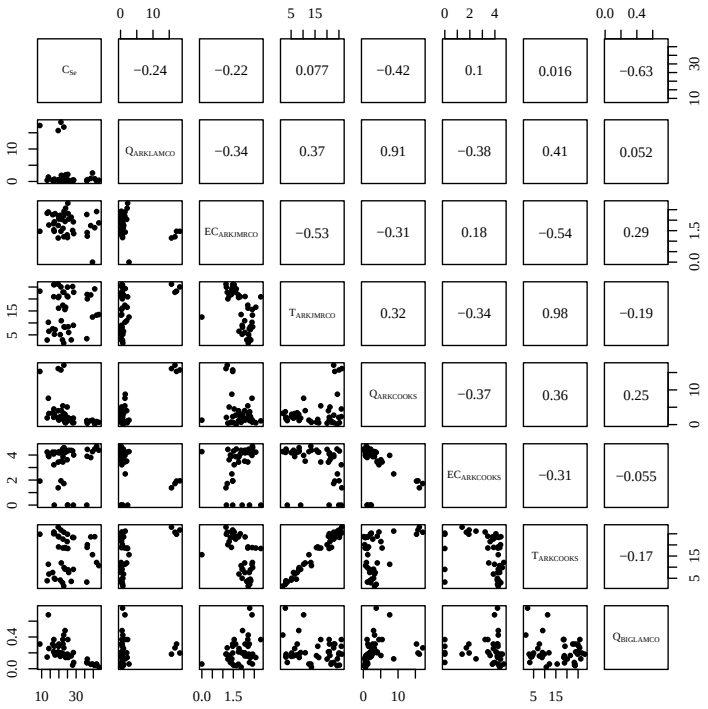


C_{D101C}

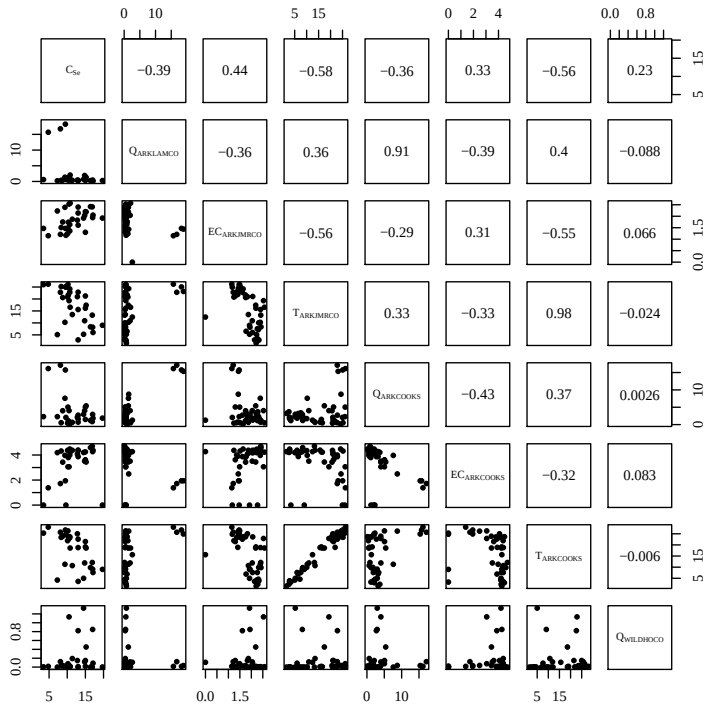


C_{D106C}

Figure 2.2 (Cont). Scatter-plot matricies of the input variables used to estimate dissolved selenium concentrations in the DSR.

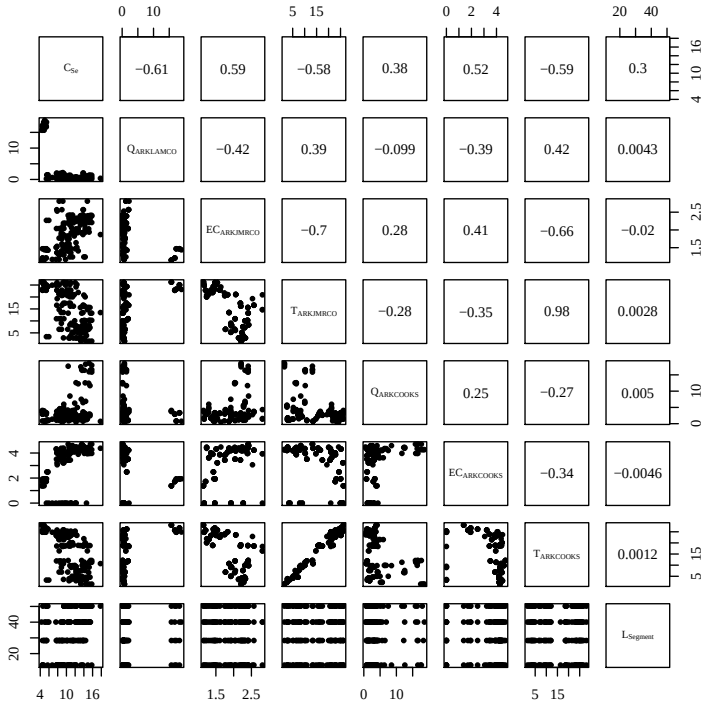


C_{D23}



C_{D57}

Figure 2.2 (Cont). Scatter-plot matricies of the input variables used to estimate dissolved selenium concentrations in the DSR.



C_{D23}

Figure 2.2 (Cont). Scatter-plot matricies of the input variables used to estimate dissolved selenium concentrations in the DSR.

In this case and as with most other regression analyses, there are strong correlations between terms which would indicate that removing all but the EC value would give an adequate estimation. The EC, which is related to the total salt concentration, is proportional to C_{Se}. However, temperature and flow are negatively correlated, compared with EC's positive correlation with C_{Se}. An increase in flow volume dilutes the total salt load and therefore C_{Se}. Temperature affects the solubility of all salts. Including temperature allows the difference in solubility constants to be expressed. Also, flow has opposing correlations with EC and temperature. These trends show that although EC alone can be used to estimate C_{Se}, it may not be not sufficient to completely describe how C_{Se} reacts to the environment.

An initial regression form was used and individual terms were removed based on their individual p value. Only one variable was removed at a time and the new, reduced form was re-analyzed. This step-wise method was performed until the p value for the individual variables was less than an α of 0.05. Residual standard error values, r-squared values, adjusted r-squared values, F-statistics, and p-values were collected for each analysis. If the statistics did not appear to support the conclusion that the final, reduced form linear mode was representative of the data, then the process was re-started with a different, more complex, initial regression form.

Tables 5.4 and 5.5 present the initial regression forms for the river sections in the USR and DSR respectively. These initial regression forms presented do not include the resulting coefficients for the respective variables as they are insignificant to the results of this particular analysis and the complete model results.

Selenium concentration at the La Junta WWTP was not able to be sufficiently estimated using linear models. The La Junta water distribution system receives most of its

Table 5.4. USR Selenium Concentration Initial Regression Models.

Concentration	Initial regression model.
C_{U163}	$Q_{ARKCATCO} + EC_{ARKCATCO} + t_{ARKCATCO}$
C_{U201}	$Q_{ARKLASCO} + EC_{ARKLASCO} + t_{ARKLASCO}$
C_{U167}	$Q_{ARKCATCO} + EC_{ARKCATCO} + Q_{ARKLASCO}$ $+EC_{ARKLASCO} + t_{ARKLASCO}$
C_{U74}	$Q_{ARKCATCO} + EC_{ARKCATCO} + Q_{ARKLASCO}$ $+EC_{ARKLASCO} + t_{ARKLASCO} + Q_{CANSWKCO}$ $+Q_{ARKCATCO}Q_{CANSWKCO} + EC_{ARKCATCO}Q_{CANSWKCO}$ $+Q_{ARKLASCO}Q_{CANSWKCO} + EC_{ARKLASCO}Q_{CANSWKCO}$ $+t_{ARKLASCO}Q_{CANSWKCO}$
C_{U60}	$Q_{ARKCATCO} + EC_{ARKCATCO} + Q_{ARKLASCO}$ $+EC_{ARKLASCO} + t_{ARKLASCO} + Q_{TIMSWICO}$
C_{U207}	$Q_{ARKCATCO} + EC_{ARKCATCO} + Q_{ARKLASCO}$ $+EC_{ARKLASCO} + t_{ARKLASCO} + Q_{HRC194CO}$
$C_{ARK,d=x}$	$Q_{ARKCATCO} + EC_{ARKCATCO} + Q_{ARKLASCO}$ $+EC_{ARKLASCO} + t_{ARKLASCO} + d$
$C_{LAJWWTP}$	$\beta_1 \cdot Q_{WTP}^{\beta_2}$

water from wells. It was not known which aquifer supplied the city. It was assumed that the shallow river riparian aquifer was the city's water source as there are no known deep aquifers in the area. It was also assumed that the water treatment plant processes and waste water treatment plant processes could change the selenium concentration. The average daily flow produced by the plant is the only variable available to estimate selenium concentrations. Visual analysis of the scatter plot similar to Figure **??** showed that the relationship between the plant discharge rate and selenium concentration could not be easily defined. A non-linear model was used with the power function described in Table 5.4. This model produced fairly

Table 5.5. USR Selenium Concentration Initial Regression Models.

Concentration	Initial regression model.
C_{D101C}	$Q_{ARKLAMCO} + EC_{ARKJMRCO} + t_{ARKJMRCO}$ $+Q^2_{ARKLAMCO} + Q_{ARKLAMCO}EC_{ARKJMRCO}$ $+Q_{ARKLAMCO}t_{ARKJMRCO} + EC^2_{ARKJMRCO}$ $+EC_{ARKJMRCO}t_{ARKJMRCO} + t^2_{ARKJMRCO}$
C_{D106C}	$Q_{ARKCOOKS} + EC_{ARKCOOKS} + t_{ARKCOOKS}$ $+Q^2_{ARKCOOKS} + Q_{ARKCOOKS}EC_{ARKCOOKS}$ $+Q_{ARKCOOKS}t_{ARKCOOKS} + EC^2_{ARKCOOKS}$ $+EC_{ARKCOOKS}t_{ARKCOOKS} + t^2_{ARKCOOKS}$
C_{D23}	$Q_{ARKLAMCO} + EC_{ARKJMRCO} + Q_{ARKCOOKS}$ $+EC_{ARKCOOKS} + t_{ARKCOOKS} + Q_{BIGLAMCO}$ $+Q_{ARKLAMCO}Q_{BIGLAMCO} + EC_{ARKJMRCO}Q_{BIGLAMCO}$ $+Q_{ARKCOOKS}Q_{BIGLAMCO} + EC_{ARKCOOKS}Q_{BIGLAMCO}$ $+t_{ARKCOOKS}Q_{BIGLAMCO}$
C_{D57}	$Q_{ARKLAMCO} + EC_{ARKJMRCO} + t_{ARKJMRCO} + Q_{ARKCOOKS}$ $+EC_{ARKCOOKS} + Q_{WILDHOCO} + Q^2_{ARKLAMCO}$ $+Q_{ARKLAMCO}EC_{ARKJMRCO} + Q_{ARKLAMCO}t_{ARKJMRCO}$ $+Q_{ARKLAMCO}Q_{ARKCOOKS} + Q_{ARKLAMCO}EC_{ARKCOOKS}$ $+Q_{ARKLAMCO}Q_{WILDHOCO} + EC^2_{ARKJMRCO}$ $+EC_{ARKJMRCO}t_{ARKJMRCO} + EC_{ARKJMRCO}Q_{ARKCOOKS}$ $+EC_{ARKJMRCO}EC_{ARKCOOKS} + EC_{ARKJMRCO}Q_{WILDHOCO}$ $+t^2_{ARKJMRCO} + t_{ARKJMRCO}Q_{ARKCOOKS}$ $+t_{ARKJMRCO}EC_{ARKCOOKS} + t_{ARKJMRCO}Q_{WILDHOCO}$ $+Q^2_{ARKCOOKS} + Q_{ARKCOOKS}EC_{ARKCOOKS}$ $+Q_{ARKCOOKS}Q_{WILDHOCO} + EC^2_{ARKCOOKS}$ $+EC_{ARKCOOKS}Q_{WILDHOCO} + Q^2_{WILDHOCO}$
$C_{ARK,d=x}$	$Q_{ARKLAMCO} + EC_{ARKJMRCO} + Q_{ARKCOOKS}$ $+EC_{ARKCOOKS} + t_{ARKCOOKS} + d$

reasonable results. Better results could be obtained if average daily EC values of the plant discharge were available.

Tables 5.6 and 5.7 present the final, reduced regression equations with coefficients resulting from the regression analysis.

Table 5.6. USR Selenium Concentration Final Regression Models.

Concentration	Final regression model.
C_{U163}	$-0.05106 \cdot Q_{ARKCATCO} + 4.69 \cdot EC_{ARKCATCO}$ $-0.3063 \cdot t_{ARKCATCO} + 10.12$
C_{U201}	$-0.3138 \cdot Q_{ARKLASCO} - 0.1348 \cdot t_{ARKLASCO} + 14.91$
C_{U167}	$-0.3538 \cdot Q_{ARKLASCO} - 2.021 \cdot EC_{ARKLASCO}$ $-0.3306 \cdot t_{ARKLASCO} + 21.15$
C_{U74}	$-13.01 \cdot EC_{ARKCATCO} - 1.022 \cdot Q_{ARKLASCO} - 0.4132 \cdot t_{ARKLASCO}$ $-16.40 \cdot Q_{CANSWKCO} - 0.3138 \cdot Q_{ARKCATCO}Q_{CANSWKCO}$ $+2.0730Q_{ARKLASCO}Q_{CANSWKCO} + 39.40$
C_{U60}	$11.42 \cdot EC_{ARKCATCO} - 3.160 \cdot Q_{TIMSWICO} + 4.605$
C_{U207}	$-17.78 \cdot Q_{HRC194CO} + 20.41$
$C_{ARK,d=x}$	$4.710 \cdot EC_{ARKCATCO} - 0.1568 \cdot Q_{ARKLASCO} - 0.1491 \cdot t_{ARKLASCO}$ $+0.0203 \cdot d + 8.0831$
$C_{LAJWWTP}$	$19.18 \cdot Q_{WTP}^{0.07664}$

Tables 5.8 and 5.9 present the summary statistics from the regression analysis. The relative standard error with the degrees of freedom (RSE, DoF), multiple R-squared, adjusted R-squared, the F-statistic with the degrees of freedom (F-statistic, DoF), and the p-value generated by the statistical software after regression analyses are presented in these tables. All fitted linear regression equations show statistics that indicate that they fit the measured data with a fairly high degree of accuracy.

Individual regression models were analyzed to determine if they were representative of the data. Figures 2.3 and 2.4 graphs used for each concentration point to make this analysis. For each sub-figure, the top left panel shows the residuals plotted against the fitted values. All points should be evenly distributed in both directions. If the points are not symmetrical along the fitted values axis, then heteroscedasticity should be suspected. If the points are

Table 5.7. DSR Selenium Concentration Final Regression Models.

Concentration	Final regression model.
C_{D101C}	$3.429 \cdot EC_{ARKJMRCO} - 0.005623 \cdot Q_{ARKLAMCO}EC_{ARKJMRCO}$ $-0.07581 \cdot EC_{ARKJMRCO}t_{ARKJMRCO} + 6.355$
C_{D106C}	$-0.37 \cdot t_{ARKCOOKS} + 0.0627 \cdot EC_{ARKCOOKS}t_{ARKCOOKS} + 16.56$
C_{D23}	$0.08951 \cdot Q_{ARKLAMCO} - 0.9925 \cdot Q_{ARKCOOKS}$ $-1.376 \cdot Q_{BIGLAMCO} - 0.007387 \cdot Q_{ARKLAMCO}Q_{BIGLAMCO}$ $+0.006882 \cdot Q_{ARKCOOKS}Q_{BIGLAMCO} + 36.96$
C_{D57}	$-42.64 \cdot EC_{ARKJMRCO} - 3.309 \cdot t_{ARKJMRCO}$ $-14.18 \cdot EC_{ARKCOOKS} - 0.006969 \cdot Q_{ARKLAMCO}Q_{WILDHOCO}$ $+0.005895 \cdot Q_{ARKCOOKS}Q_{WILDHOCO}$ $-0.2522 \cdot Q_{WILDHOCO}EC_{ARKJMRCO}$ $+5.779 \cdot EC_{ARKJMRCO}EC_{ARKCOOKS}$ $+1.029 \cdot EC_{ARKJMRCO}t_{ARKJMRCO}$ $+0.2643 \cdot EC_{ARKCOOKS}t_{ARKJMRCO} - 112.7$
$C_{ARK,d=x}$	$-0.006576 \cdot Q_{ARKLAMCO} + 1.322 \cdot EC_{ARKJMRCO}$ $+0.003794 \cdot Q_{ARKCOOKS} + 0.4512 \cdot EC_{ARKCOOKS}$ $-0.07653 \cdot t_{ARKCOOKS} + 0.1066 \cdot d + 5.709$

Table 5.8. USR Selenium Concentration Regression Models Summary Statistics.

Concentration	RSE, DoF	Multiple R-squared	Adjusted R-squared	F-statistic, DoF	p-value
C_{U163}	0.5923, 11	0.9818	0.9769	198.3, 3 and 11	7.4×10^{-10}
C_{U201}	0.9467, 14	0.9145	0.9022	74.84, 2 and 14	3.3×10^{-8}
C_{U167}	0.7016, 10	0.9684	0.9589	102.1, 3 and 10	8.4×10^{-8}
C_{U74}	0.9868, 10	0.9501	0.9202	31.74, 6 and 10	6.0×10^{-6}
C_{U60}	1.749, 15	0.8633	0.845	47.35, 2 and 15	3.3×10^{-7}
$C_{ARK,d=x}$	1.196, 130	0.8808	0.8772	240.2, 4 and 130	$<2.2\times10^{-16}$
$C_{LAJWWTP}$	4.153, 85	NA	NA	NA	NA

RSE = Residual standard error

DoF = Degrees of Freedom

not symmetrical about either axis, then the regression model is missing an estimating term. The bottom left panel shows a scale-location plot of the data. This panel shows the same information as the first panel with the exception that the square-root of the standardized

Table 5.9. DSR Selenium Concentration Regression Models Summary Statistics.

Concentration	RSE, DoF	Multiple R-squared	Adjusted R-squared	F-statistic, DoF	p-value
C_{D101C}	2.553, 38	0.5472	0.5114	15.31, 3 and 38	1.1×10^{-6}
C_{D106C}	3.534, 39	0.3404	0.3065	10.06, 2 and 36	3.0×10^{-4}
C_{D23}	5.701, 37	0.5819	0.5254	10.3, 5 and 37	3.5×10^{-6}
C_{D57}	2.337, 21	0.7278	0.6111	3.238, 9 and 21	2.7×10^{-4}
$C_{ARK,d=x}$	1.724, 156	0.7104	0.6993	63.78, 6 and 156	¡2.2×10^{-16}

RSE = Residual standard error
DoF = Degrees of Freedom

residuals is used. This reduces the effect of skewness on the analysis. Like the first panel, the points should be symmetrical along both axes. The top right panel shows a normal quantile-quantile (Q-Q) plot of the residuals. It is a reasonable assumption that residuals of a linear model are normally distributed. If the residuals are normally distributed, then the points should fit closely to the line at $y = x$. The bottom right panel shows the residuals plotted against the leverage of the residuals. The dashed red lines show the Cook's distance which is a measure of the influence a particular data point has on the regression model. Points with a higher leverage have a higher influence on the model. Points with leverage values higher than the majority of the data may be outliers.

In all panels, potential outliers are indicated by having the index number of the data printed next to the point. If the same data points are indicated as outliers on most of the panels, then it is most likely true that they are outliers. None of the models had the outliers removed to improve the model. In fact, outliers were expected. Most of the dissolved selenium samples were collected during similar times in the year. Bias towards certain flow regimes and other factors are most likely present in the data. The outliers represent the samples taken outside of the normal sampling season and are more representative of the extremes. These outliers are necessary to the complete analysis and were not removed from the linear regression analyses.

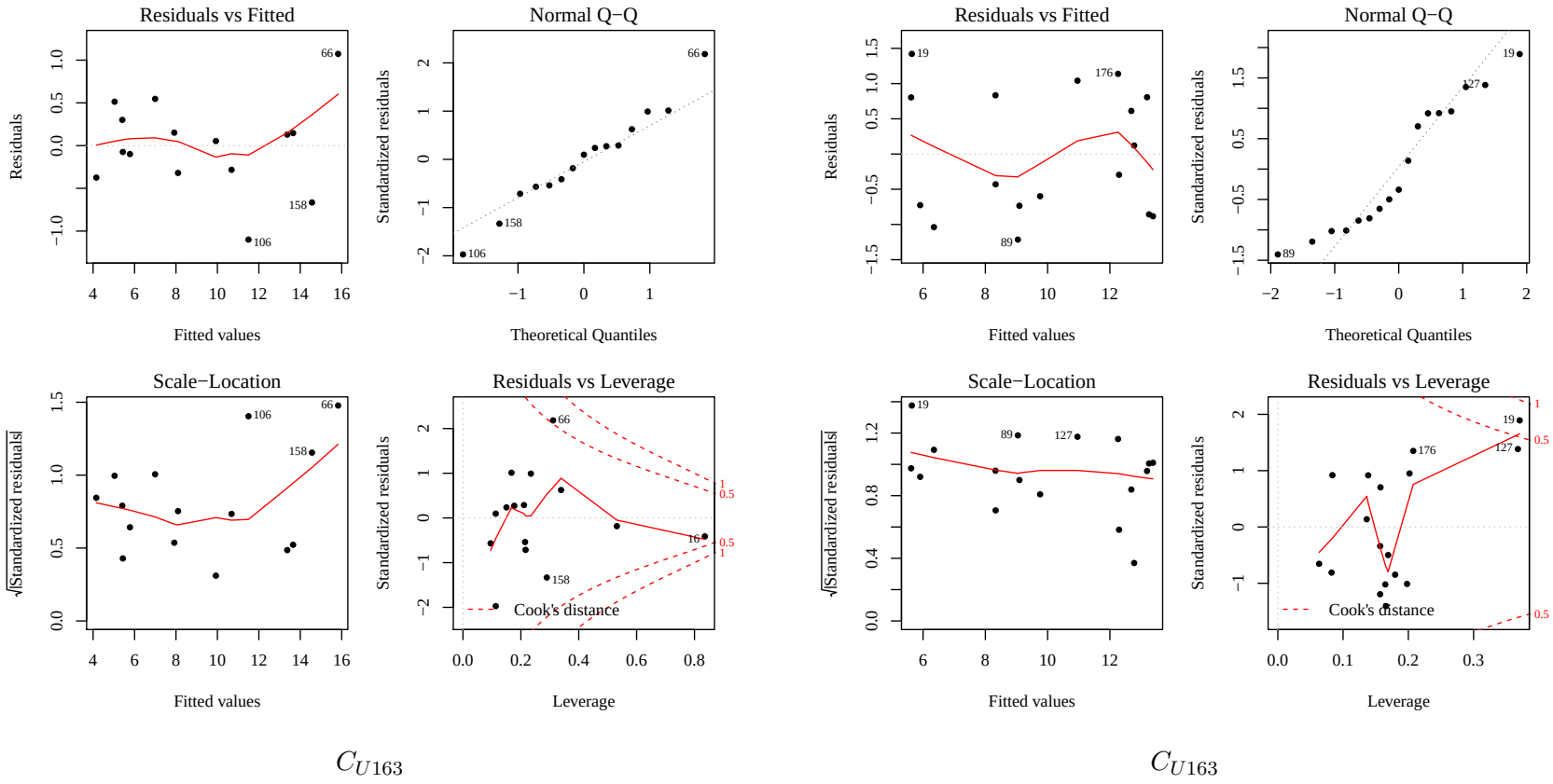
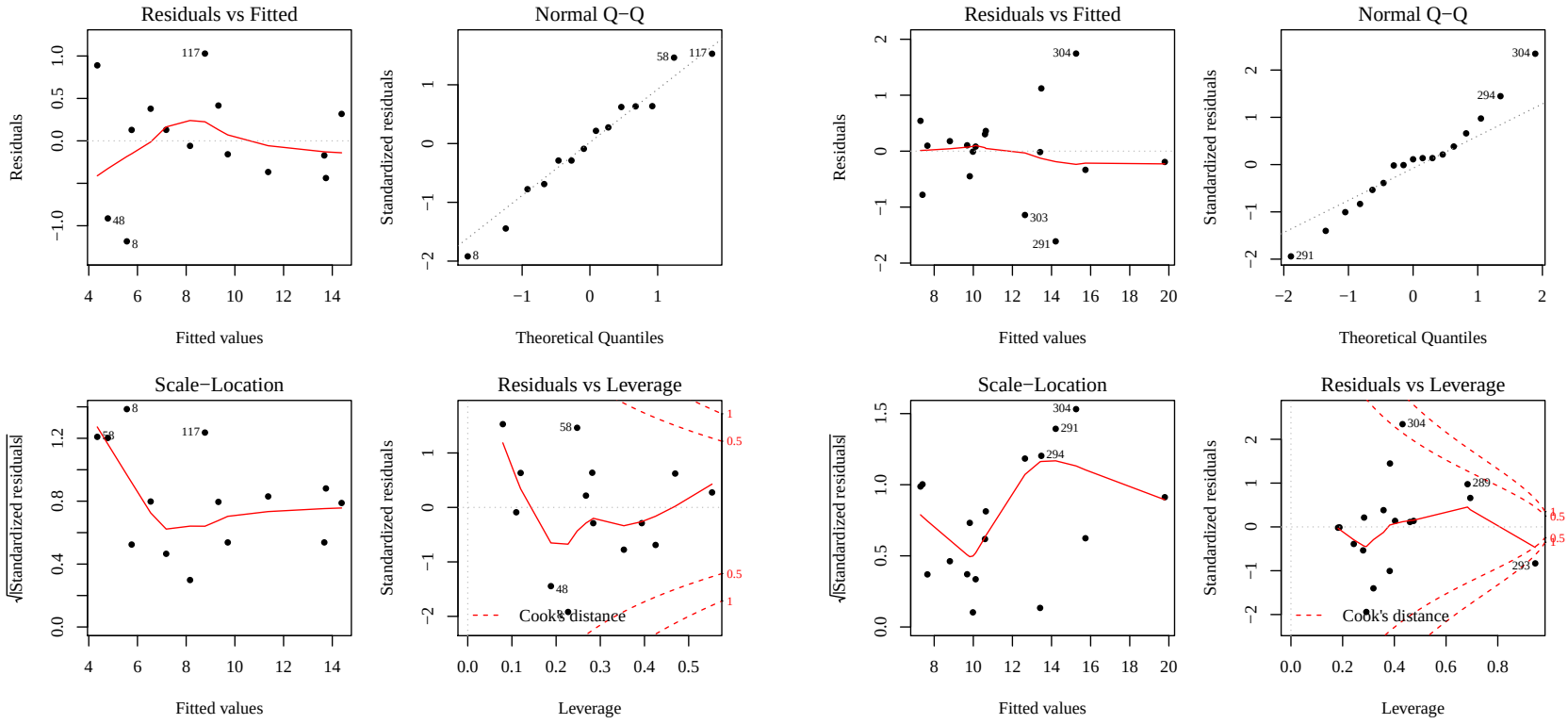
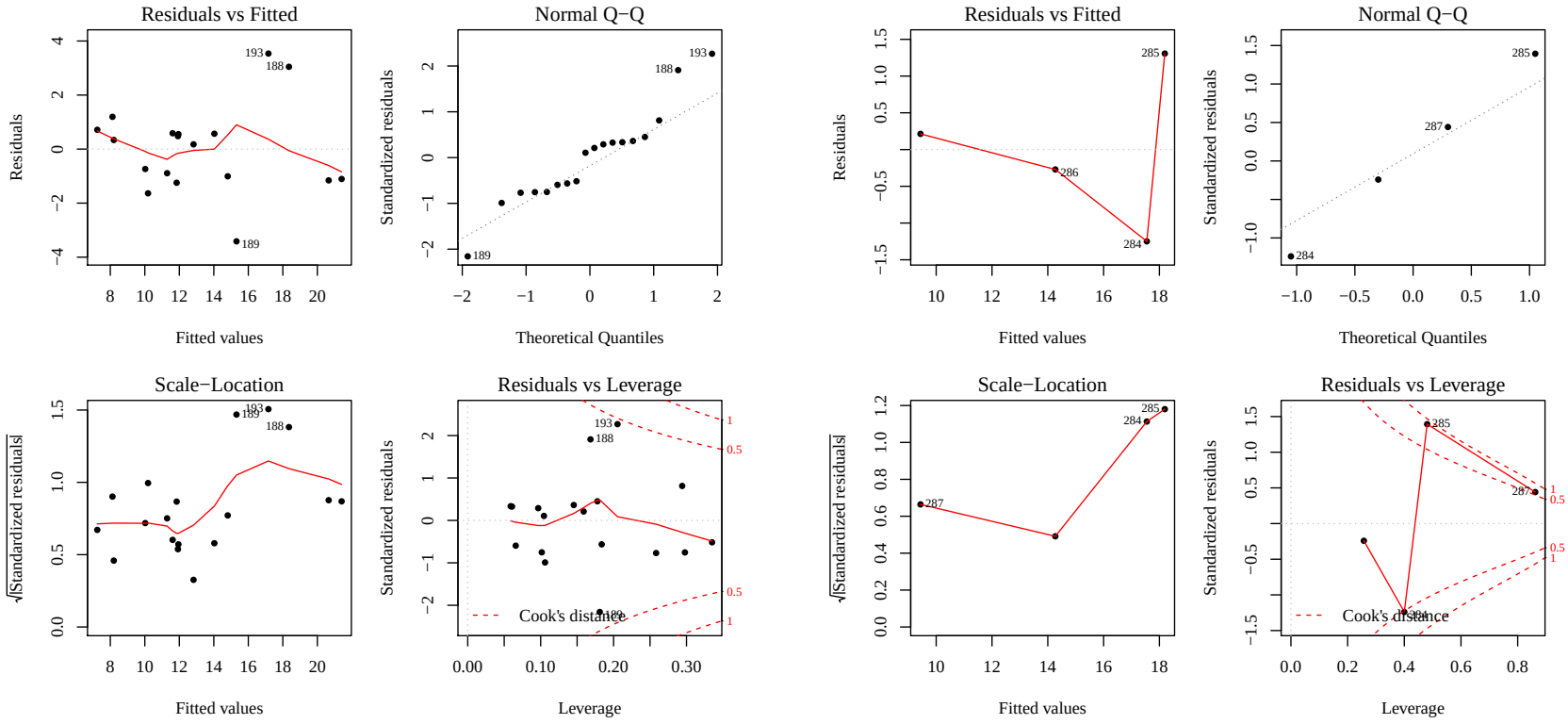


Figure 2.3. USR selenium concentration linear model analysis graphs. The black lines indicate the stochastic model mean time-series. The blue band indicates the 95% central inter-percentile range (CIR) of the stochastic values. The red dashed line in the flow depth portion of the figure indicates the reported flow depth values.



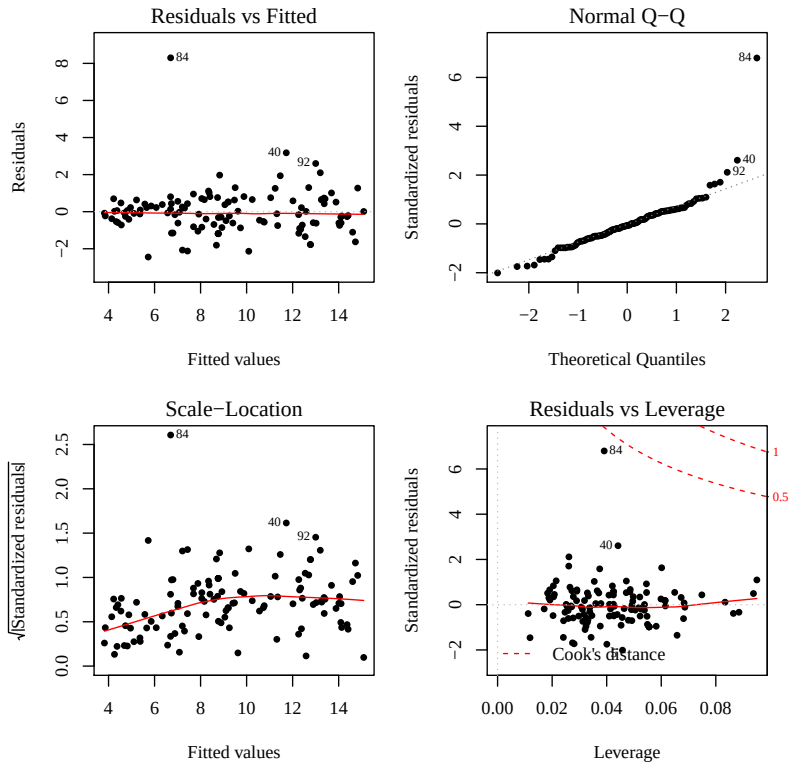
C_{U167} C_{U74}

Figure 2.3 (Cont). USR selenium concentration linear model analysis graphs.



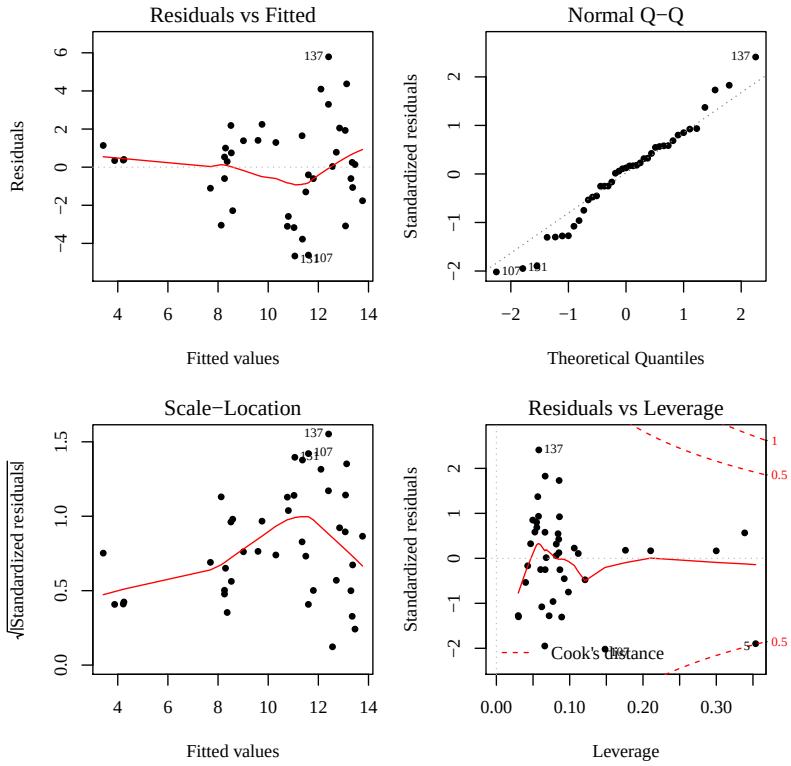
C_{U60} C_{U207}

Figure 2.3 (Cont). USR selenium concentration linear model analysis graphs.

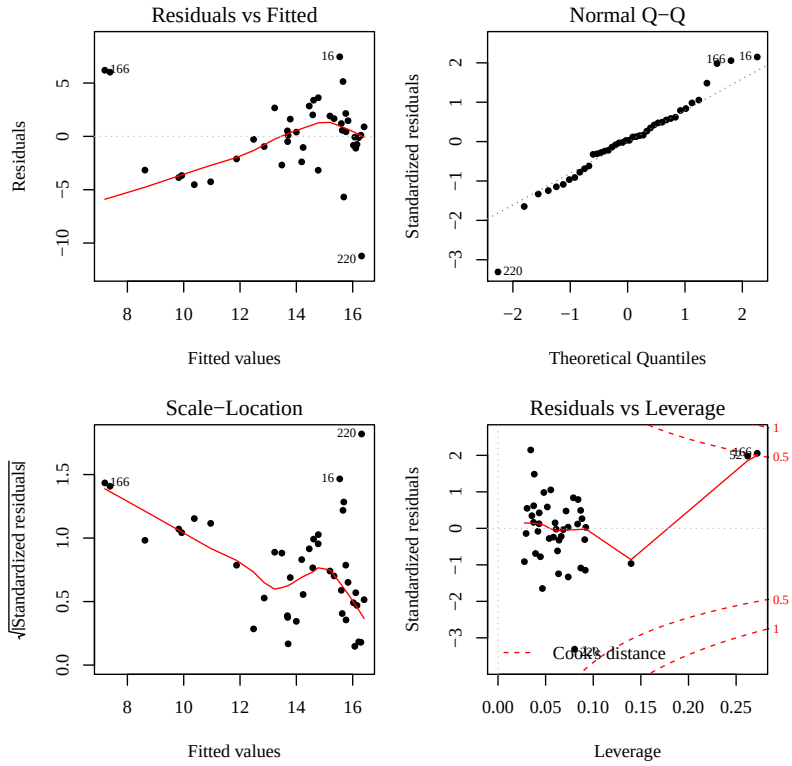


$C_{ARK,d=x}$

Figure 2.3 (Cont). USR selenium concentration linear model analysis graphs.

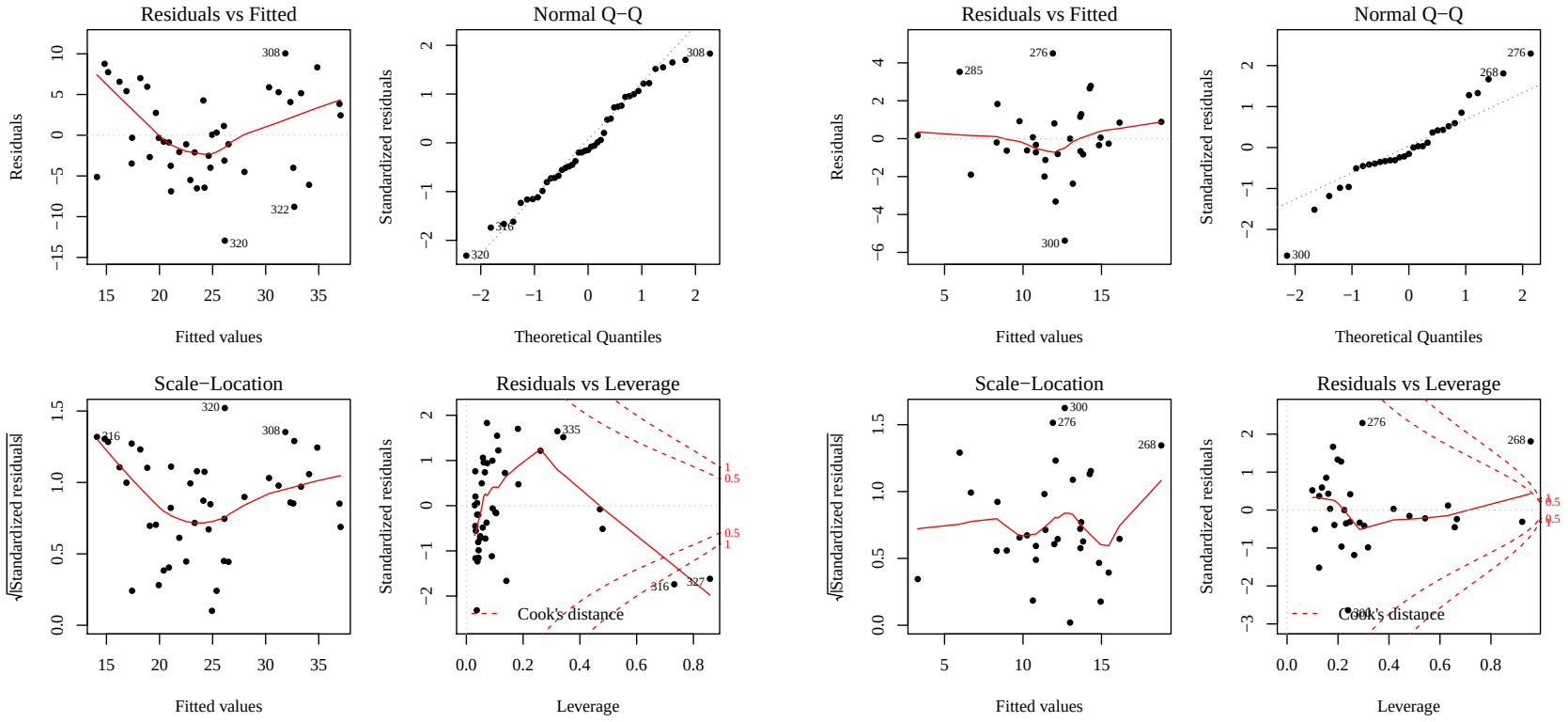


C_{D101C}



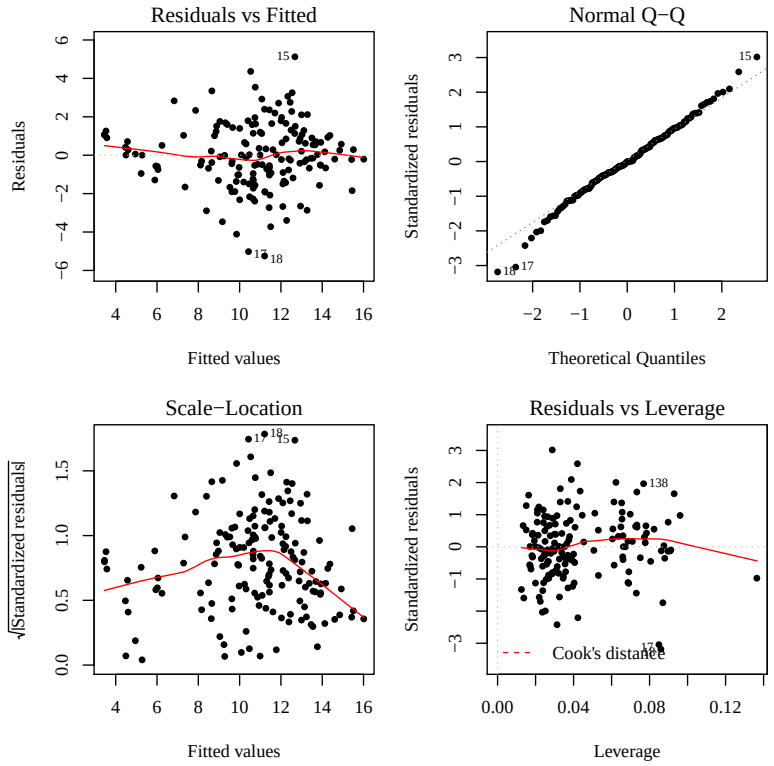
C_{D106C}

Figure 2.4. DSR selenium concentration linear model analysis graphs. The black lines indicate the stochastic model mean time-series. The blue band indicates the 95% central inter-percentile range (CIR) of the stochastic values. The red dashed line in the flow depth portion of the figure indicates the reported flow depth values.



C_{D23} C_{D57}

Figure 2.4 (Cont). DSR selenium concentration linear model analysis graphs.



$C_{ARK,d=x}$

Figure 2.4 (Cont). DSR selenium concentration linear model analysis graphs.

Since the dissolved selenium concentration at the La Junta WWTP is derived using a non-linear equation, the graphs used for linear models are not suitable. In order to prevent a biased comparison between linear and non-linear models and their respective goodness-of-fit, it was determined to judge the model results for the La Junta WWTP selenium concentration on its own merits. The purpose of this model development was to determine which model provided the lowest root mean squared error while simultaneously ensuring that heteroskedasticity was not present. The final $C_{LAJWWTP}$ model was analyzed with the graphs shown in Figure 2.5. The upper sub-figures shows a plot of the fitted values versus the residuals. This plot would show if there was a correlation between these two values, which is heteroskedasticity. The lower sub-figures shows an autocorrelation analysis of the residuals. In a well formed model, the residuals should not be correlated with themselves. Like heteroskedasticity, this indicates that there is a factor missing in the model. Figure 2.5 shows that both heteroskedasticity and autocorrelation are not significant factors in the final $C_{LAJWWTP}$ analysis.

The measured concentration values were compared to the predicted concentration values for all regression analyses. Figures 2.6 and 2.6 show the graphs used in this comparison. An $y = x$ line was plotted . Points below the line show that the estimating equation under-estimated the selenium concentration. The vertical distance between the point and the line corresponds to the estimate error, or residual. Similar figures for the other regression analyses are presented in the appendix.

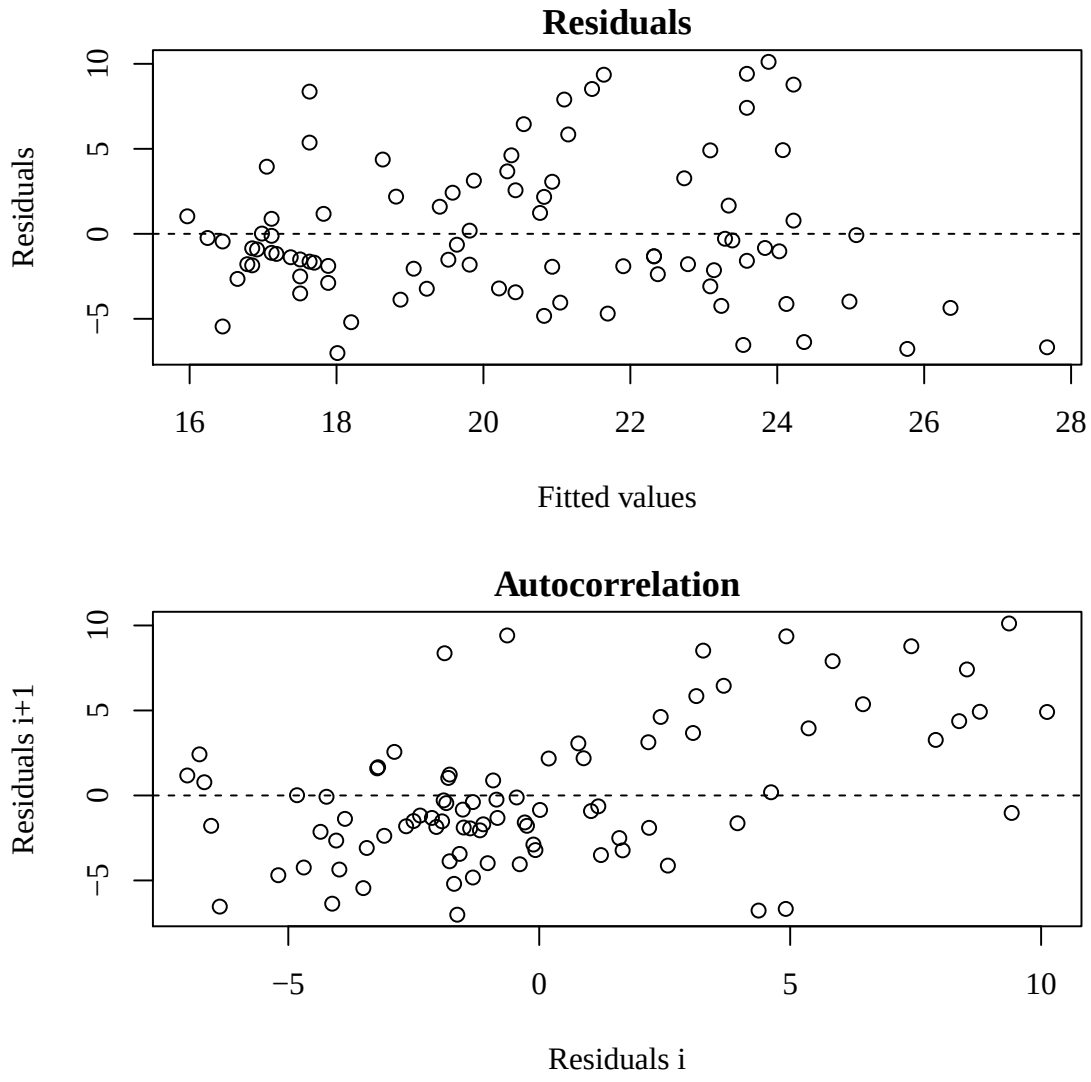
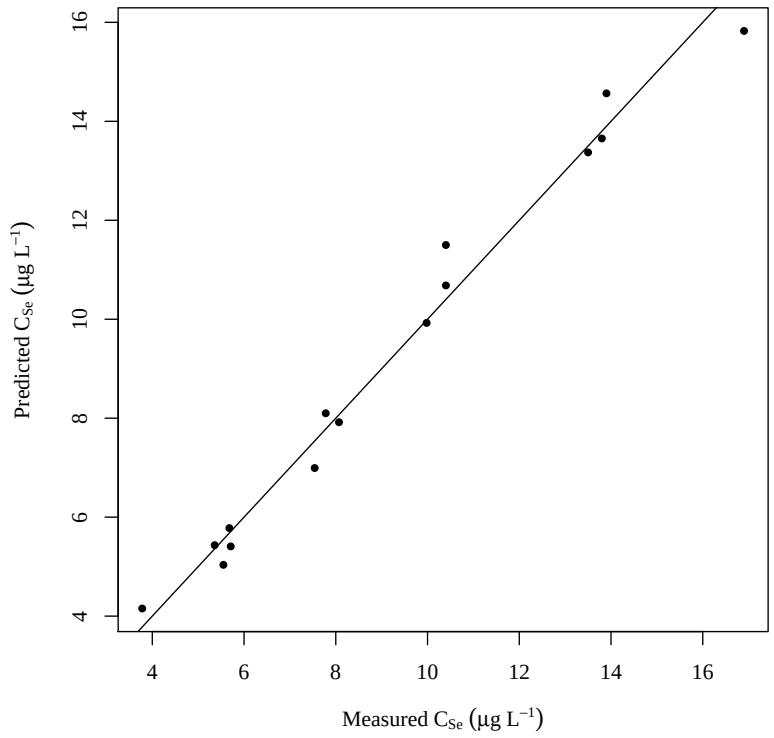
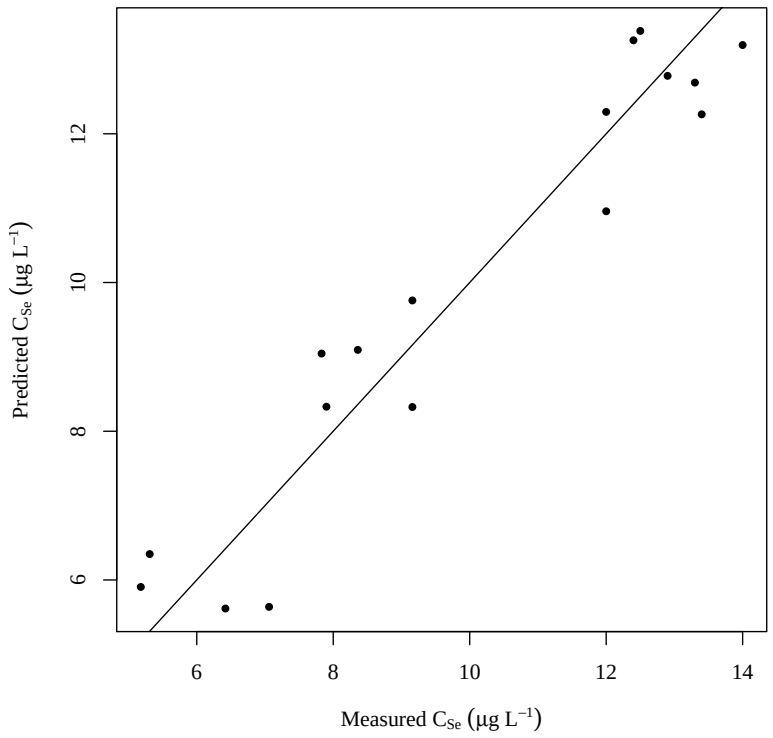


Figure 2.5. La Junta WWTP selenium concentration residuals analysis.

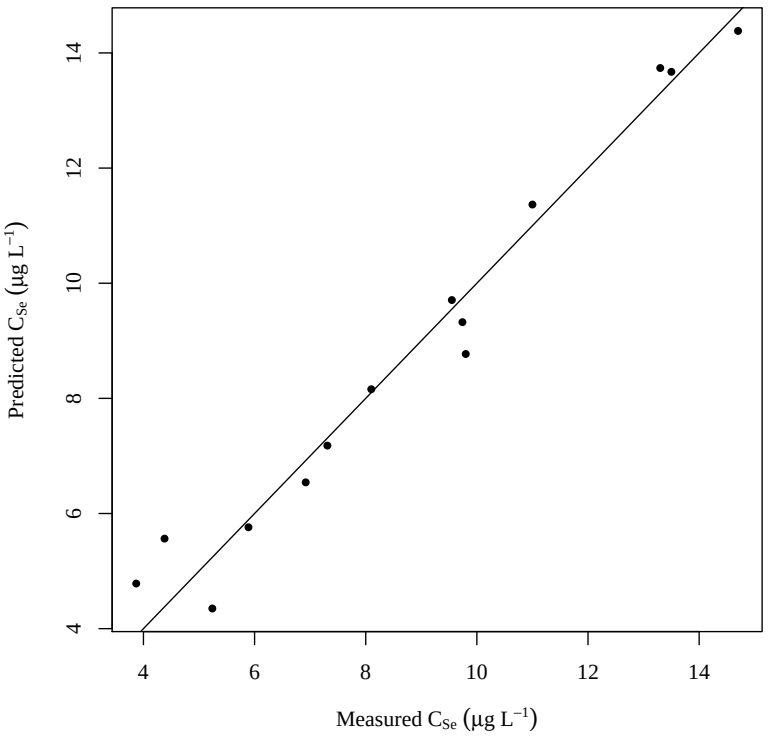


C_{U163}

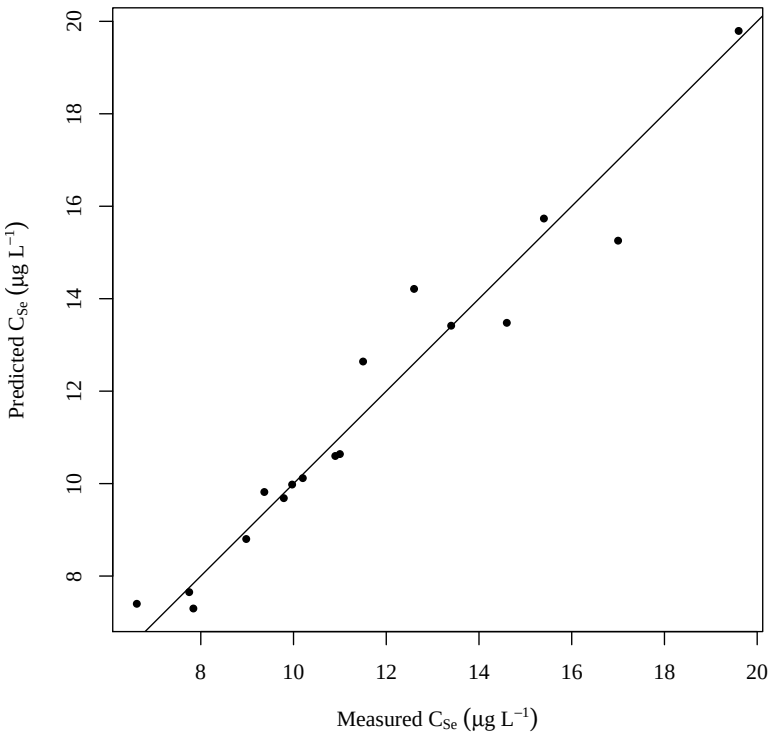


C_{U201}

Figure 2.6. USR measured vs. estimated selenium concentration comparison. The predicted values are calculated using the respective final regression equation. The diagonal line has a slope of 1 passing through the origin to show the over or under estimation. All concentrations are presented in $\mu g\, L^{-1}$.

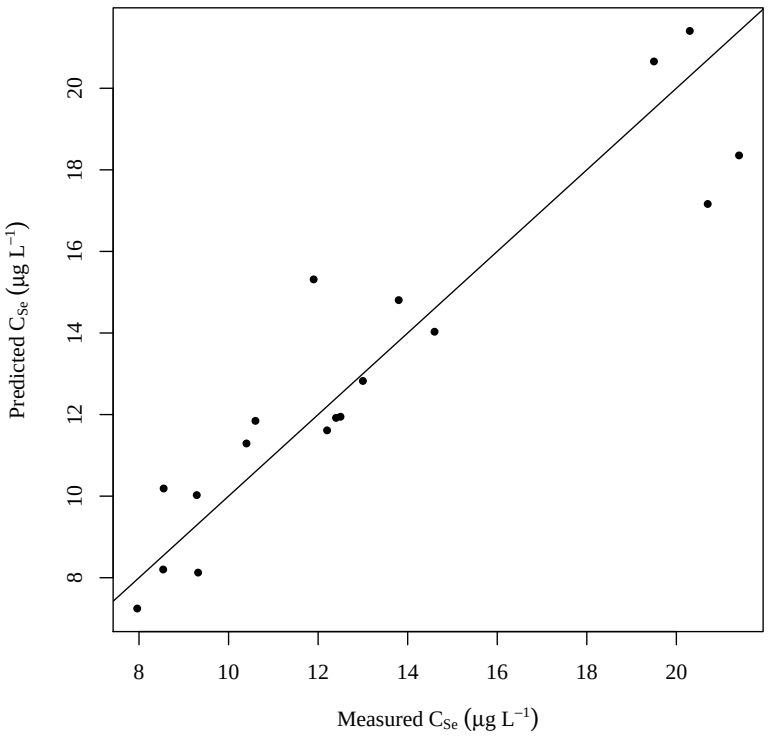


C_{U167}

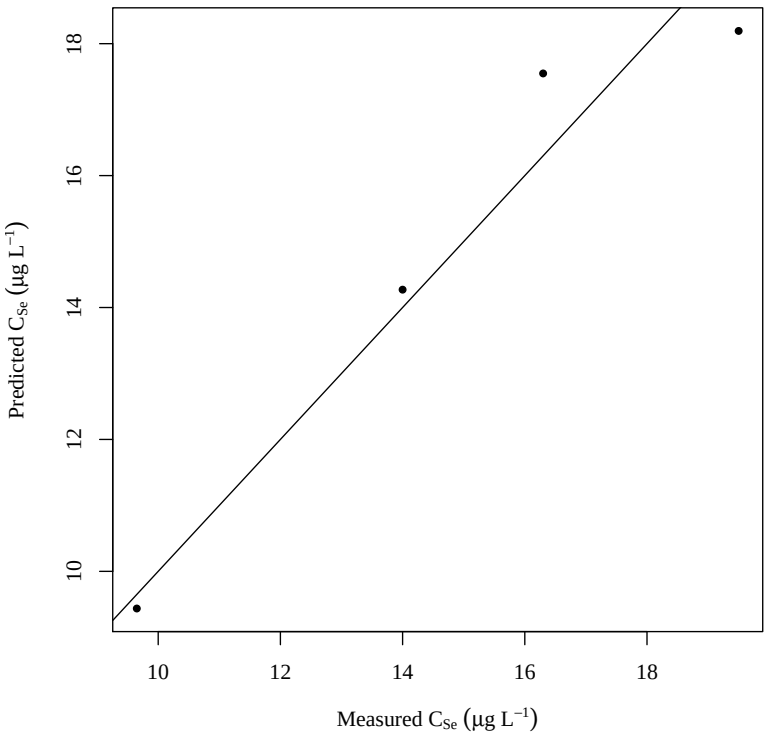


C_{U74}

Figure 2.6 (Cont). USR selenium concentration linear model analysis graphs.

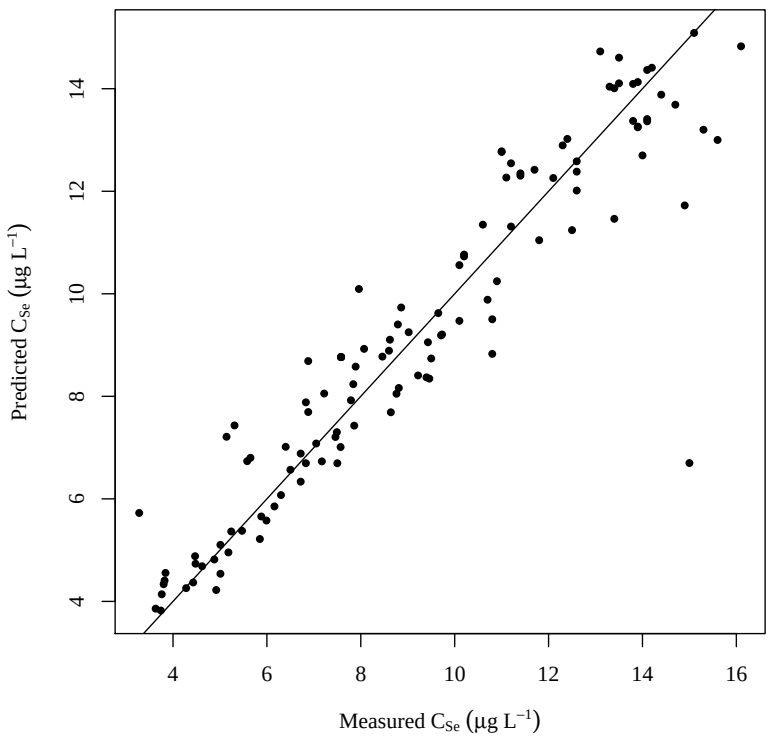


C_{U60}

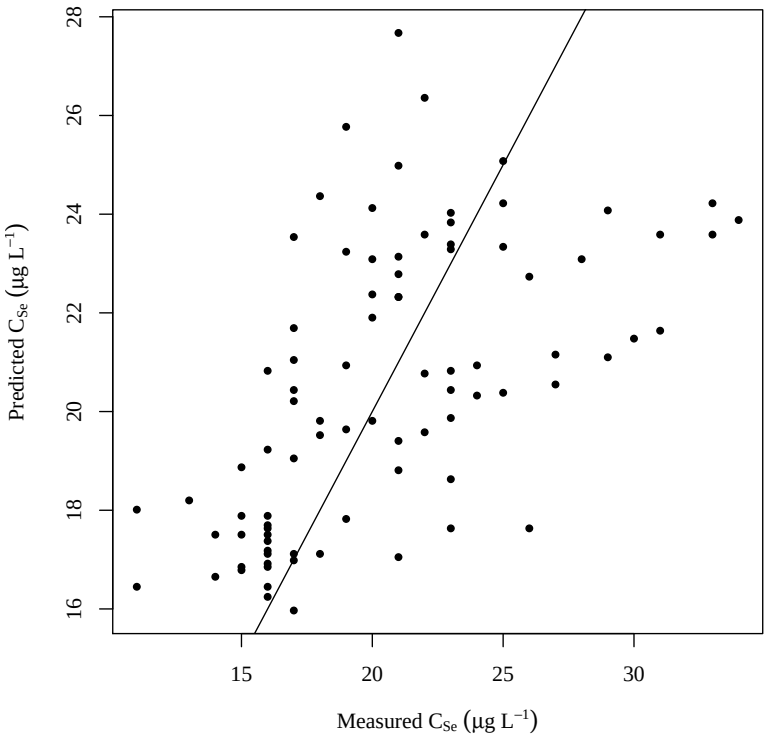


C_{U207}

Figure 2.6 (Cont). USR selenium concentration linear model analysis graphs.

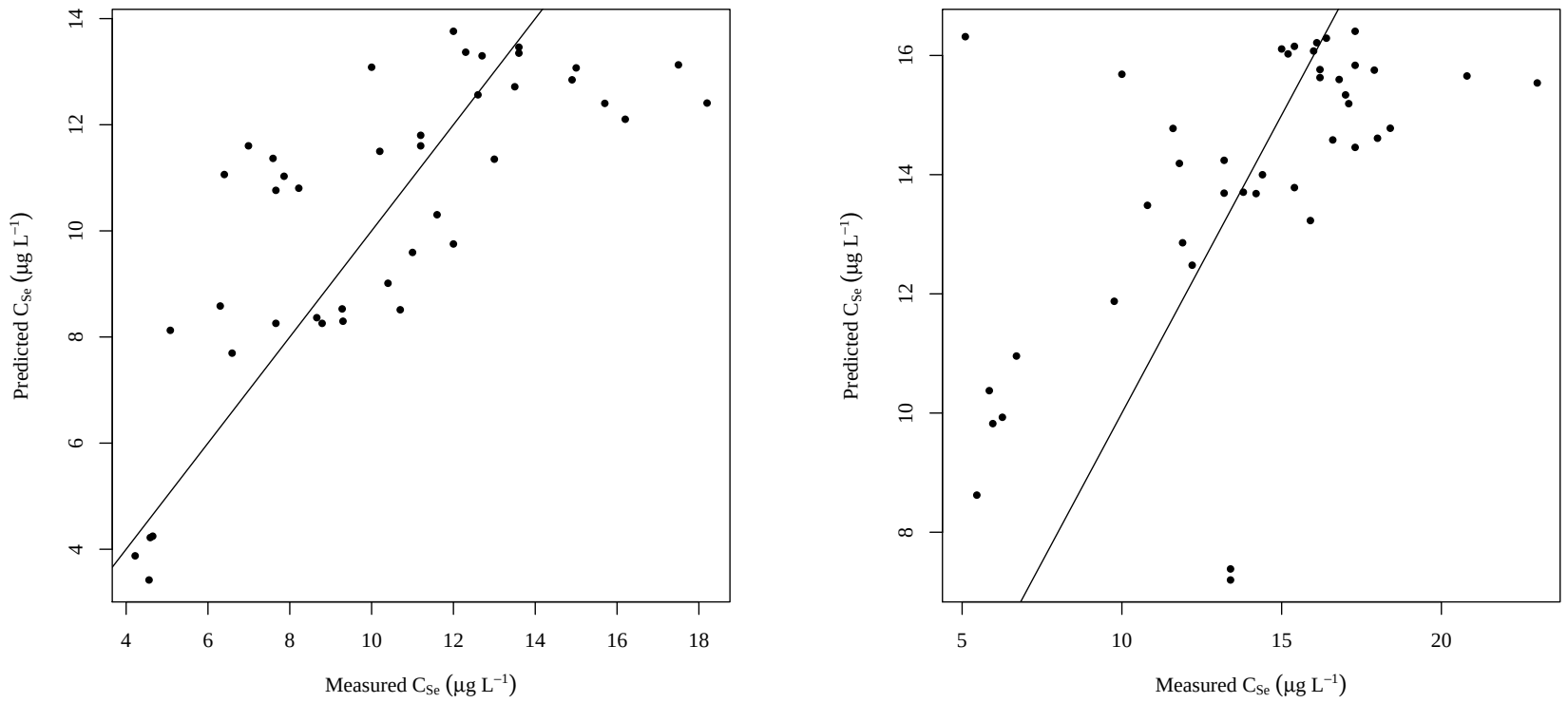


$C_{ARK,d=x}$



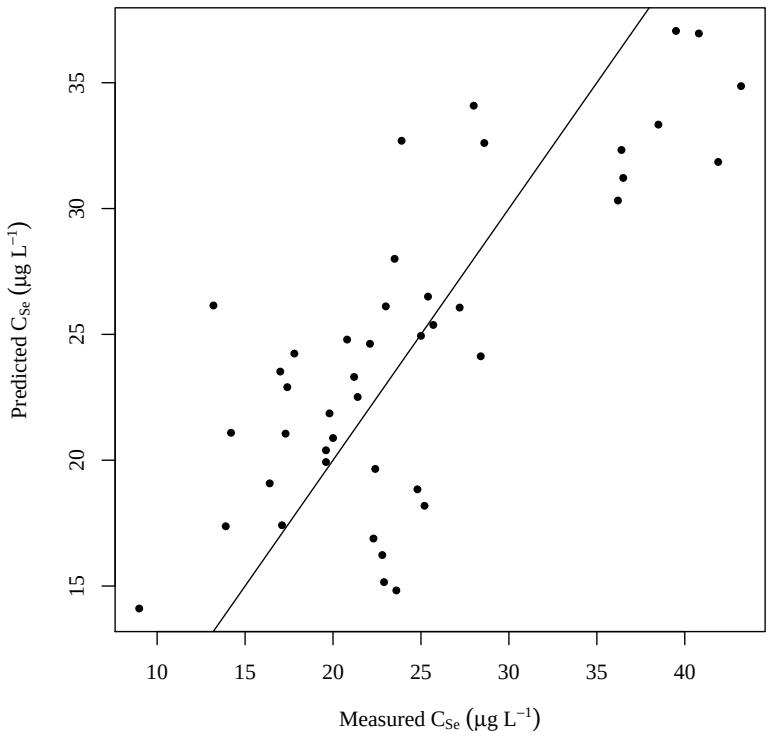
$C_{LAJWWTP}$

Figure 2.6 (Cont). USR selenium concentration linear model analysis graphs.

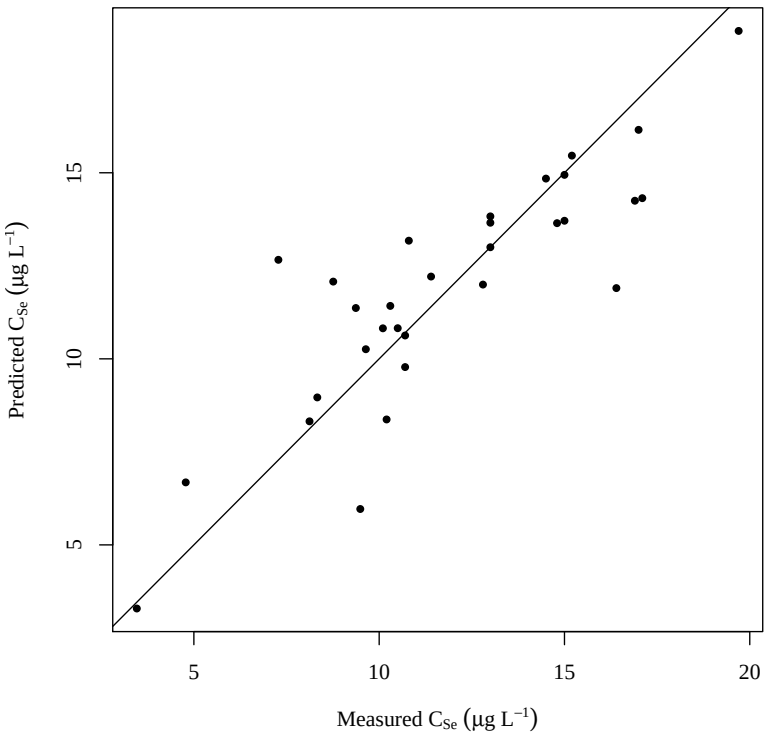


C_{D101C} C_{D106C}

Figure 2.7. DSR measured vs. estimated selenium concentration comparison. The predicted values are calculated using the respective final regression equation. The diagonal line has a slope of 1 passing through the origin to show the over or under estimation. All concentrations are presented in $\mu g\, L^{-1}$.

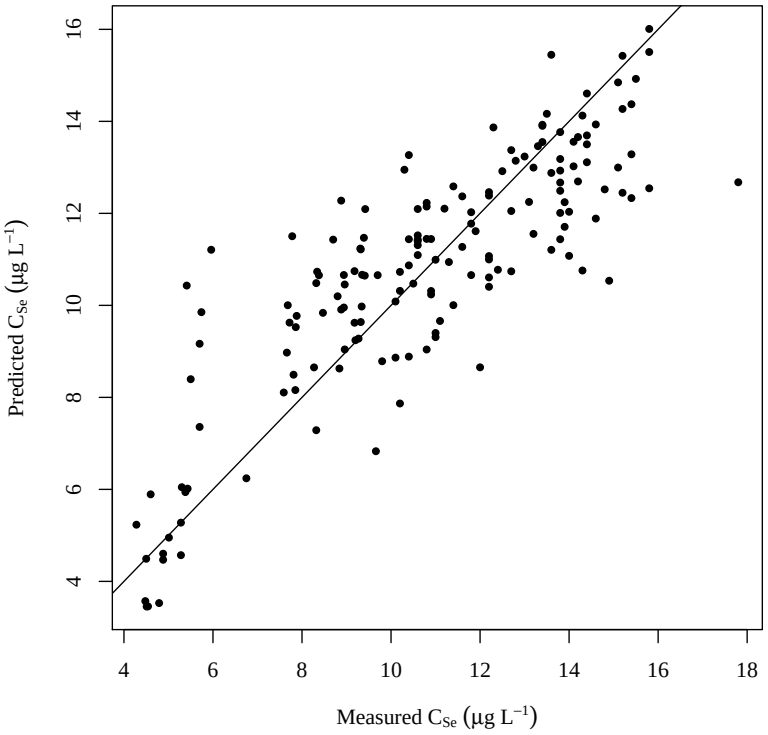


C_{D23}



C_{D57}

Figure 2.7 (Cont). DSR selenium concentration linear model analysis graphs.



$C_{ARK,d=x}$

Figure 2.7 (Cont). DSR selenium concentration linear model analysis graphs.

Each of the dissolved selenium estimating equations is accompanied by two uncertainty values as show in Equation 7. The uncertainty found in the regression equation ε_1 is due solely to the inability of the a regression equation to accurately describe the measured dissolved selenium concentration values.

$$C_x + \varepsilon_1 + \varepsilon_2 \tag{7}$$

Where:

$C_x =$ A calculated concentration value. X denotes any one of the concentration locations.

$\varepsilon_1 =$ Uncertainty found in the regression equation.

$\varepsilon_2 =$ Uncertainty in the measured dissolved selenium concentration lab values.

Selenium estimation calculation error, ε_1 was analyzed to determine the best fit distribution for each selenium estimating equation. Normal and logistic distributions were fit to the regression model residuals. Both of these distributions are unbounded and are simple to apply. They also fit the assumption that the linear model residuals are normally distributed. Logistic distributions were included because they are very similar to normal distributions, but with heavier tails. The best-fit normal and logistic distributions were compared to the regression model residuals by using Kolmogorov-Smirnov, Cramer von Mises, and Anderson-Darling goodness-of-fit tests. The results of these test determined which of either the best fit normal distribution or best fit logistic distribution described the regression model residuals. Results from the goodness-of-fit tests are presented in Tables 5.10 and 5.11 for the USR and DSR, respectively

Table 5.10. USR selenium concentration residuals goodness-of-fit test results. Kolmogorov-Smirnov (K-S), Cramer von Mieses (CvM), and Anderson-Darling (AD) test statistics are presented for each regression model.

Concentration	Tested Distribution	Test Statistics		
		K-S	CvM	A-D
C_{U163}	Logistic*	0.107	0.022	0.151
	Normal	0.116	0.028	0.197
C_{U201}	Logistic*	0.180	0.129	0.766
	Normal	0.178	0.140	0.792
C_{U167}	Logistic*	0.091	0.020	0.165
	Normal	0.100	0.025	0.192
C_{U74}	Logistic*	0.136	0.053	0.315
	Normal	0.140	0.076	0.419
C_{U60}	Logistic*	0.128	0.076	0.476
	Normal	0.161	0.086	0.548
$C_{ARK,d=x}$	Logistic*	0.041	0.032	0.352
	Normal	0.109	0.354	2.47
$C_{LAJWWTP}$	Logistic*	0.090	0.163	1.22
	Normal	0.130	0.281	1.57

* = best fit distribution

Table 5.11. DSR selenium concentration residuals goodness-of-fit test results. Kolmogorov-Smirnov (K-S), Cramer von Mieses (CvM), and Anderson-Darling (AD) test statistics are presented for each regression model.

Location	Tested Distribution	Test Statistics		
		K-S	CvM	A-D
C_{D101C}	Logistic	0.0923	0.038	0.263
	Normal*	0.090	0.378	0.259
C_{D106C}	Logistic*	0.082	0.289	0.199
	Normal	0.110	0.060	0.384
C_{D23}	Logistic	0.103	0.082	0.508
	Normal*	0.092	0.073	0.426
C_{D57}	Logistic*	0.117	0.058	0.340
	Normal	0.140	0.093	0.518
$C_{ARK,d=x}$	Logistic*	0.041	0.020	0.112
	Normal	0.043	0.050	0.319

* = best fit distribution

Table 5.12. USR selenium concentration residual distribution summary statistics.

Concentration	Best Fit Distribution	n	Parameter Estimate	
			Param. 1[1]	Param. 2[2]
C_{U163}	Logistic	15	5.7×10^{-3}	0.2810
C_{U201}	Logistic	17	-3.0×10^{-2}	0.5365
C_{U167}	Logistic	14	1.5×10^{-2}	0.3383
C_{U74}	Logistic	17	1.7×10^{-3}	0.4049
C_{U60}	Logistic	18	-6.3×10^{-2}	0.8743
$C_{ARK,d=x}$	Logistic	135	-5.2×10^{-2}	0.5615
$C_{LAJWWTP}$	Logistic	87	-3.8×10^{-1}	2.313

[1] For normal distributions, mean. For logistic distributions, location.

[2] For normal distributions, standard deviation. For logistic distributions, scale.

The summaries of the fitted distributions are presented in Tables 5.12 and 5.13, for the USR and DSR respectively. In all but one case, the location parameters are near zero. The location parameter for logistic distributions and the mean parameter for normal distributions provide the same information; they describe the central tendency of the distribution. For distributions that describe model error, the goal is to have this value near zero. The location parameter for the La Junta WWTP selenium concentration error distribution is a significant distance from zero. The selenium concentration lab results minimum detection level is $0.4\,\mu g\,L^{-1}$ and the La Junta WWTP location parameter approaches this value. This indicates that the selenium concentration estimating model for the La Junta WWTP is missing an estimating parameter. Unfortunately, no other parameters were available for the collected data.

Statistical plots of the residual distributions, shown in Figures 2.8 and 2.8, were created to visually analyze the data distribution and determine if the chosen distribution represented the data for the USR and DSR, respectively. These figures are diagnostic plots that are produced by statistical software. They were not altered or customized. The top left panel shows a histogram of the residuals. The red curve is the chosen best-fit distribution.

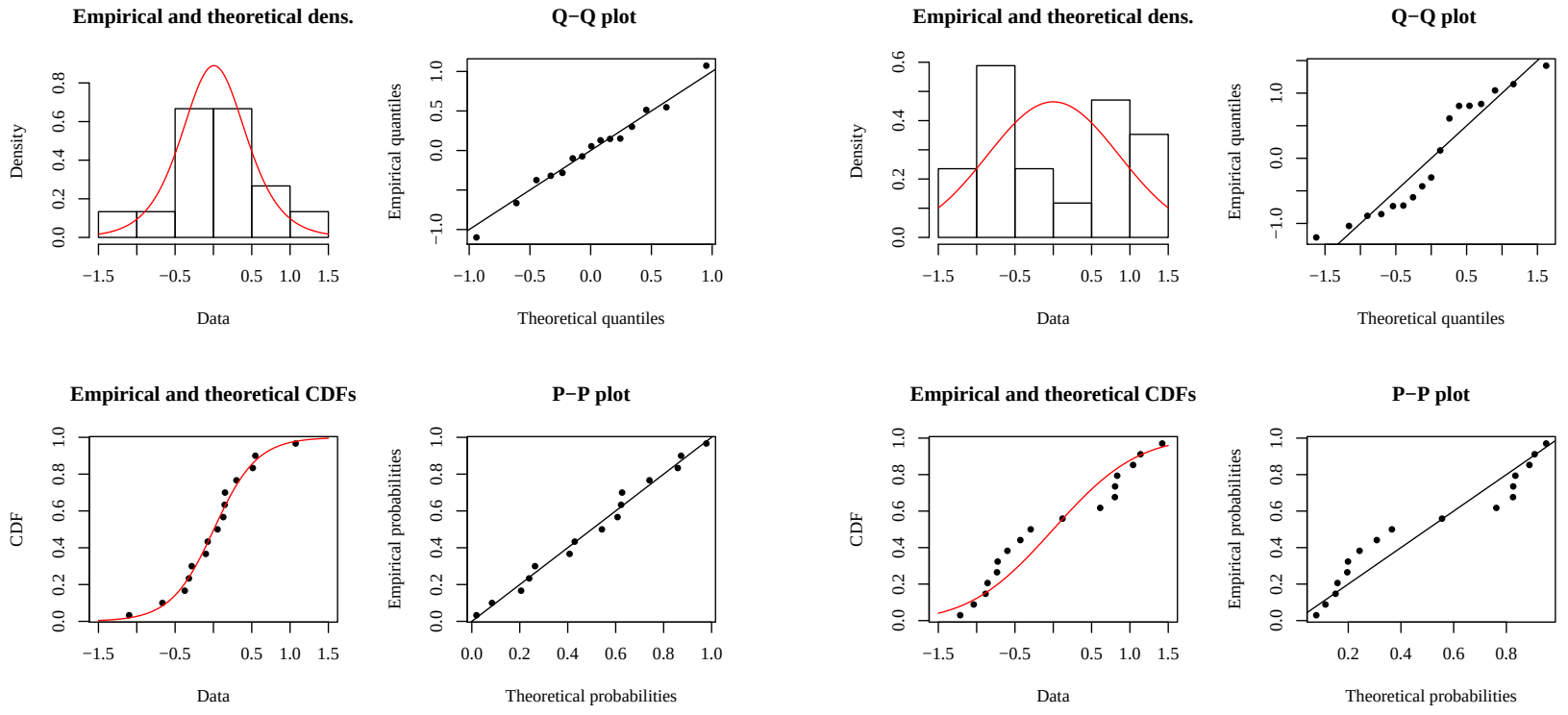
Table 5.13. DSR selenium concentration residual distribution summary statistics.

Concentration	Best Fit Distribution	n	Parameter Estimate	
			Param. 1[1]	Param. 2[2]
C_{D101C}	Normal	42	-1.7×10^{-17}	2.429
C_{D106C}	Logistic	42	7.5×10^{-2}	1.834
C_{D23}	Normal	43	-1.2×10^{-16}	5.289
C_{D57}	Logistic	31	-8.4×10^{-3}	1.026
$C_{ARK,d=x}$	Logistic	163	1.2×10^{-2}	0.9364

[1] = For normal distributions, mean. For logistic distributions, location.

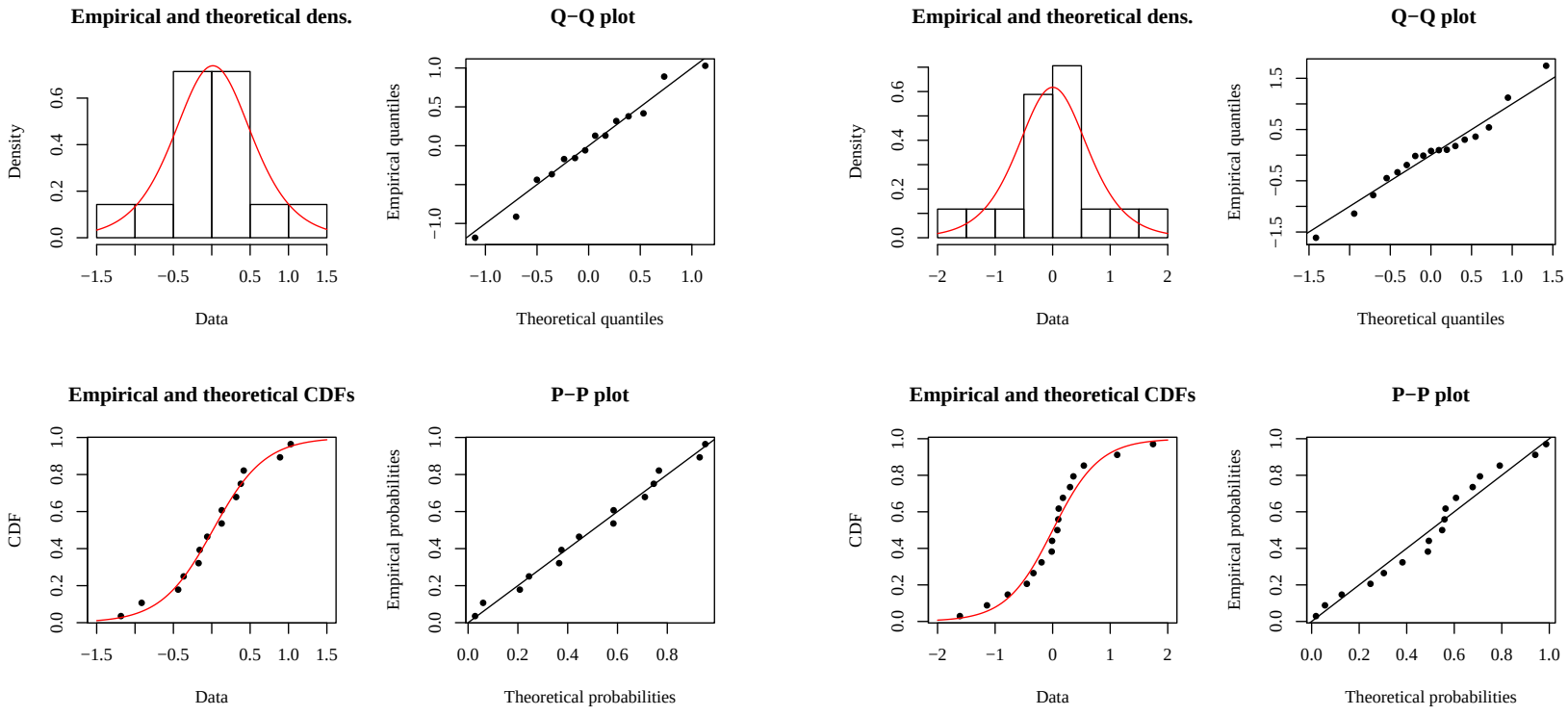
[2] = For normal distributions, standard deviation. For logistic distributions, scale.

The bottom left panel shows the cumulative distribution function of the chosen distribution in red. The points represent the actual collected data. In a well fit distribution, the points should lie on or near the fitted distribution cumulative distribution function line. The top right panel shows the well known quantile-quantile plot. This is a visual test for normalcy. The points should lie on or near the line $y = x$ if the distribution is normal. The focus on this panel is to see how well the tails fit a normal distribution. The bottom right panel is a probability-probability plot. This is also a visual test for normalcy and the points should lie on or near the line $y = x$ if the distribution is normal. The focus on this panel is to see how well the center of the data fits a normal distribution.



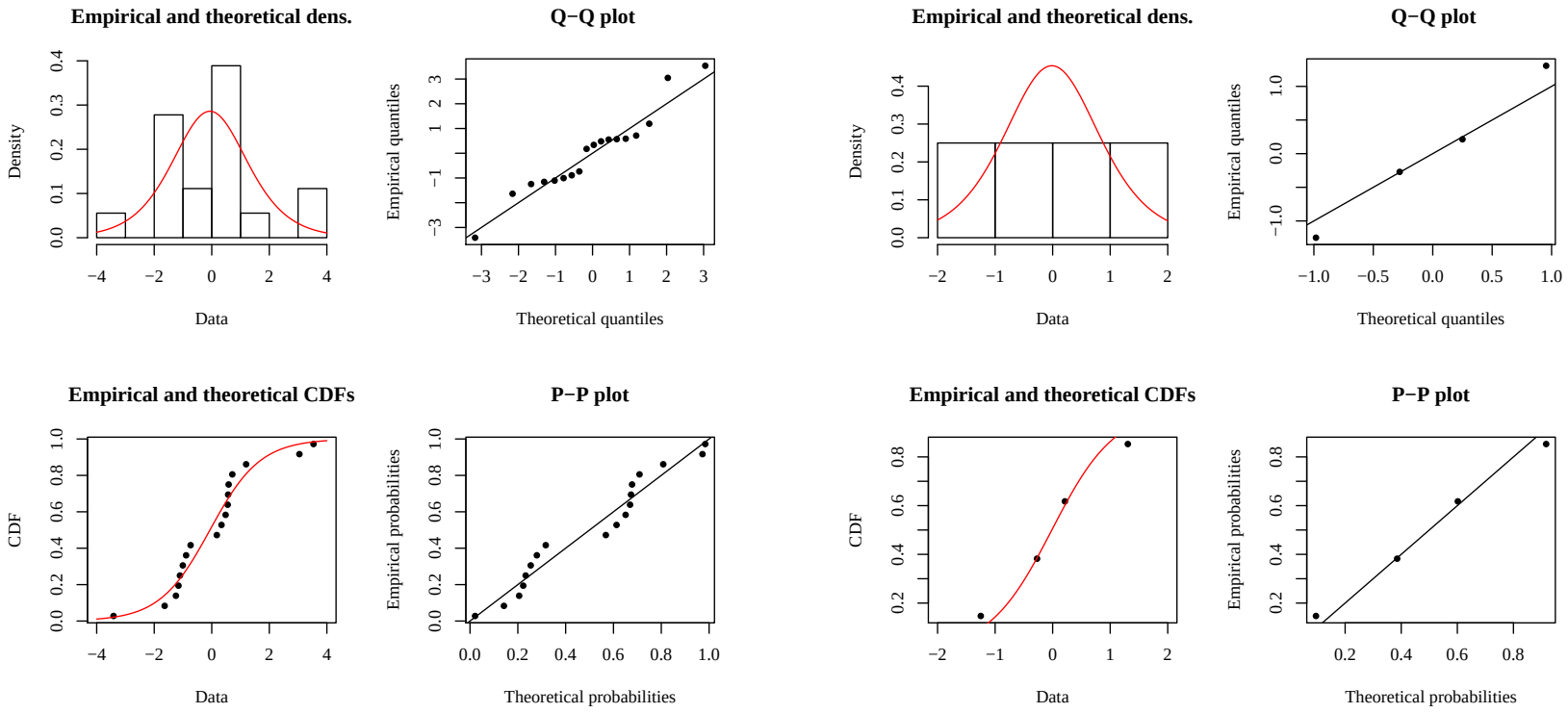
C_{U163} C_{U201}

Figure 2.8. USR upstream boundary selenium estimate residual distribution analysis. The top left plot is a histogram of the residuals with the estimated logistic distribution plotted over top. The top right plot is a quantile-quantile (Q-Q). The bottom left is a plot of the theoretical cumulative distribution function (CDF) against the empirical CDF. The bottom right is a probability-probability plot.



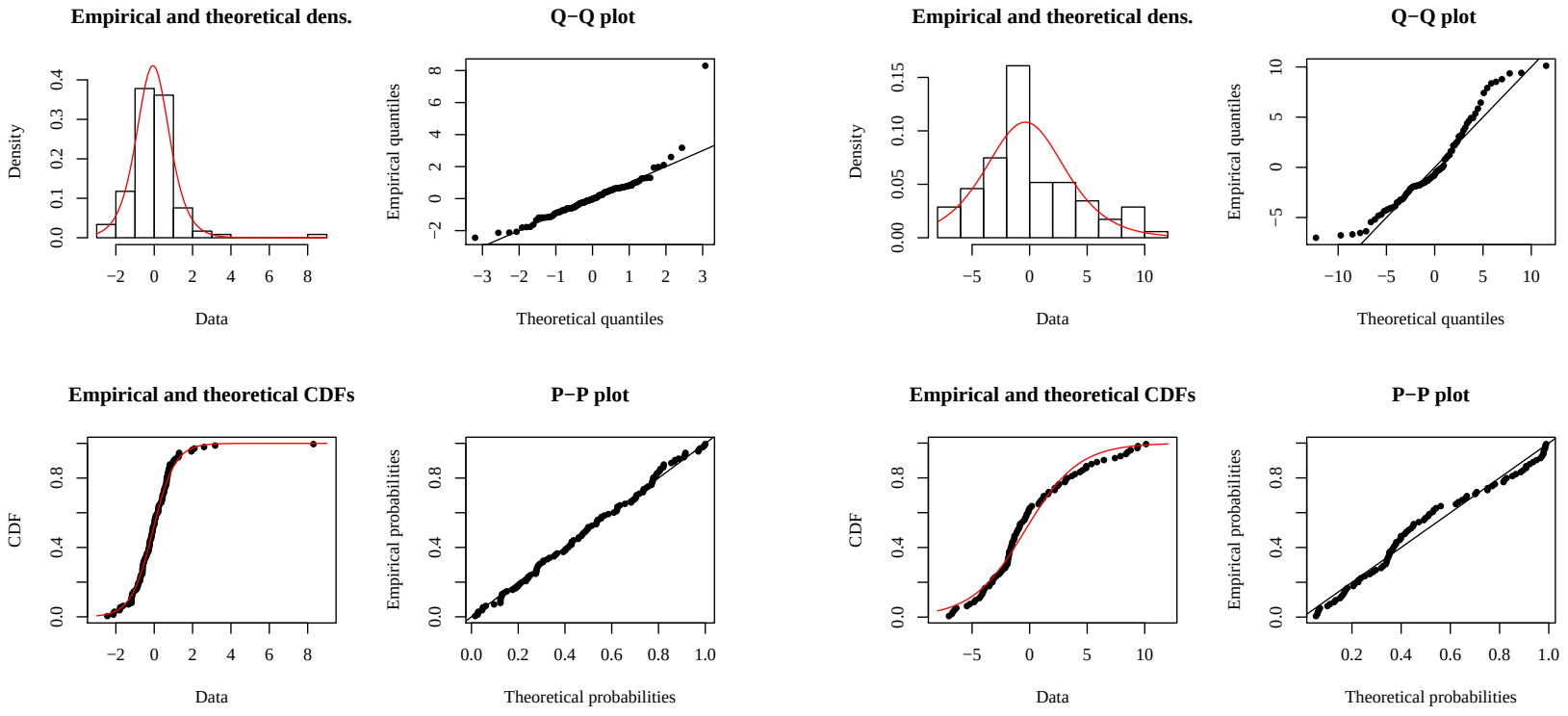
C_{U167} C_{U74}

Figure 2.8 (Cont). USR upstream boundary selenium estimate residual distribution analysis.



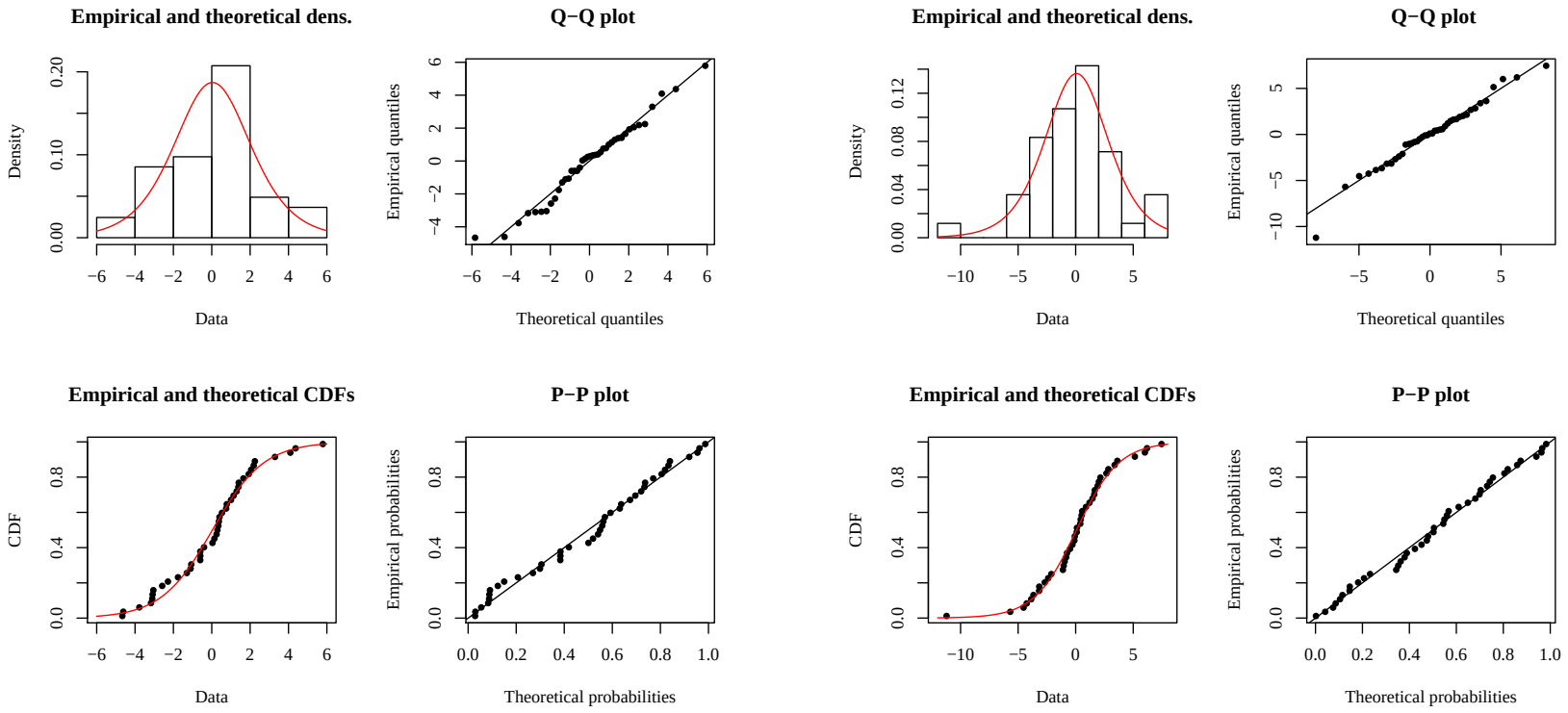
C_{U60} C_{U207}

Figure 2.8 (Cont). USR upstream boundary selenium estimate residual distribution analysis.



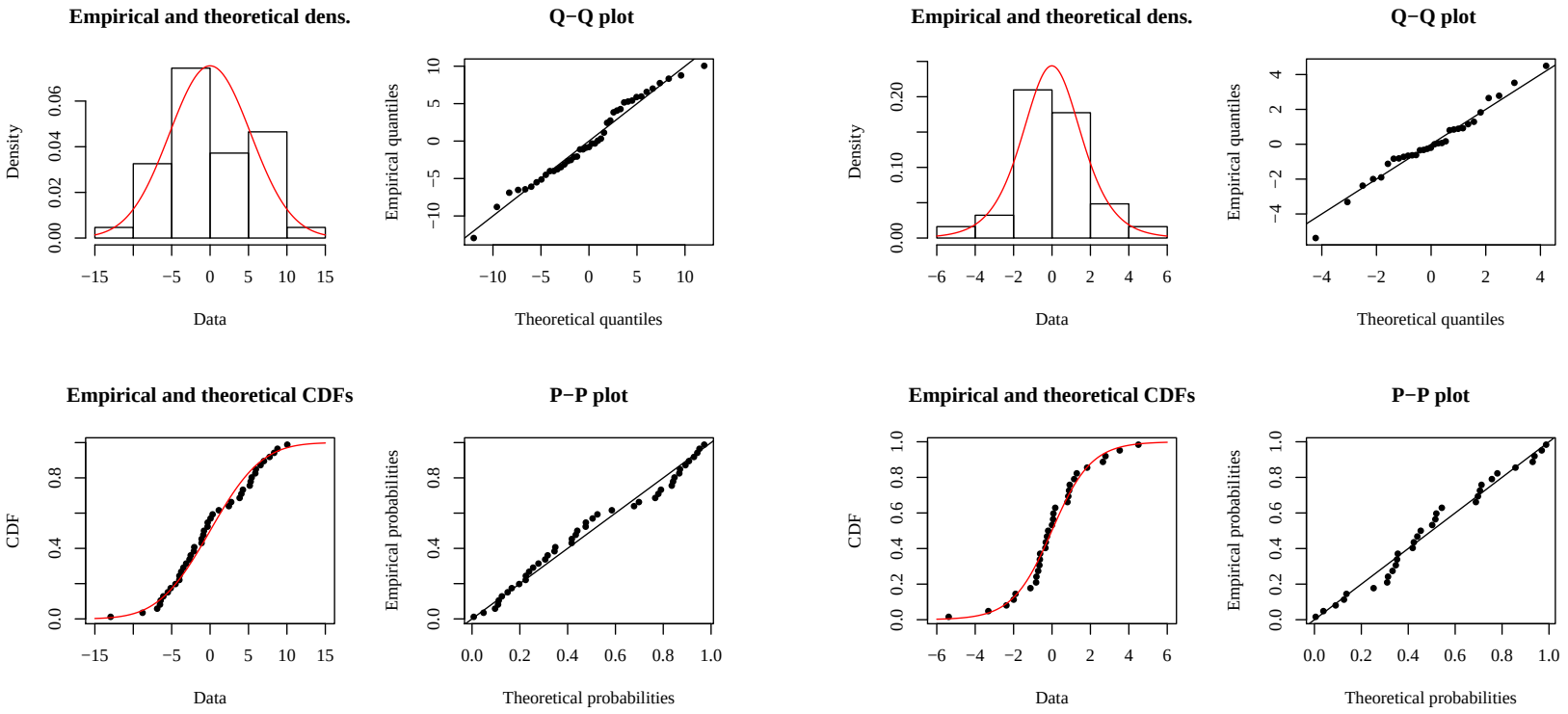
$C_{ARK,d=x}$ $C_{LAJWWTP}$

Figure 2.8 (Cont). USR upstream boundary selenium estimate residual distribution analysis.



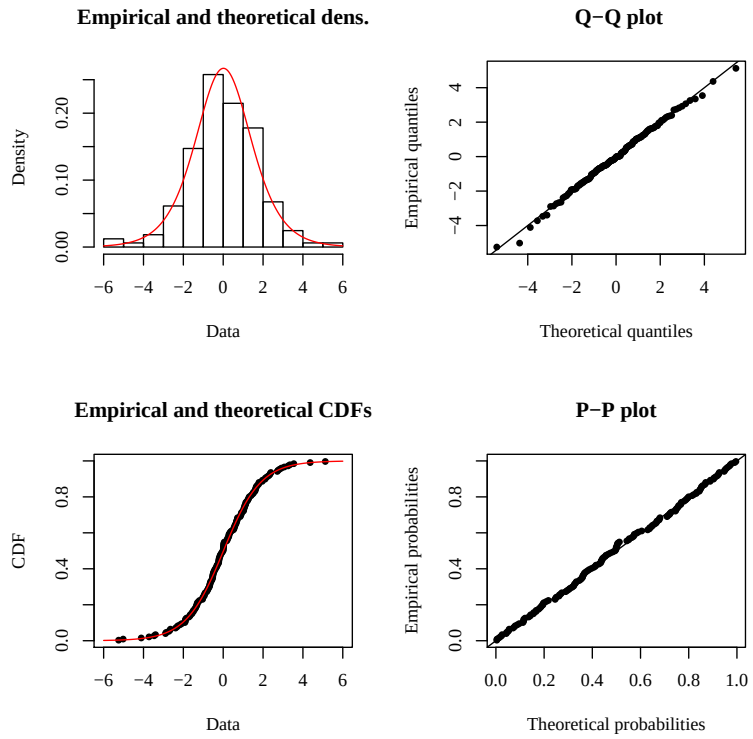
C_{D101C} C_{D106C}

Figure 2.9. DSR upstream boundary selenium estimate residual distribution analysis. The top left plot is a histogram of the residuals with the estimated logistic distribution plotted over top. The top right plot is a quantile-quantile (Q-Q). The bottom left is a plot of the theoretical cumulative distribution function (CDF) against the empirical CDF. The bottom right is a probability-probability plot.



C_{D23} C_{D57}

Figure 2.9 (Cont). DSR upstream boundary selenium estimate residual distribution analysis.



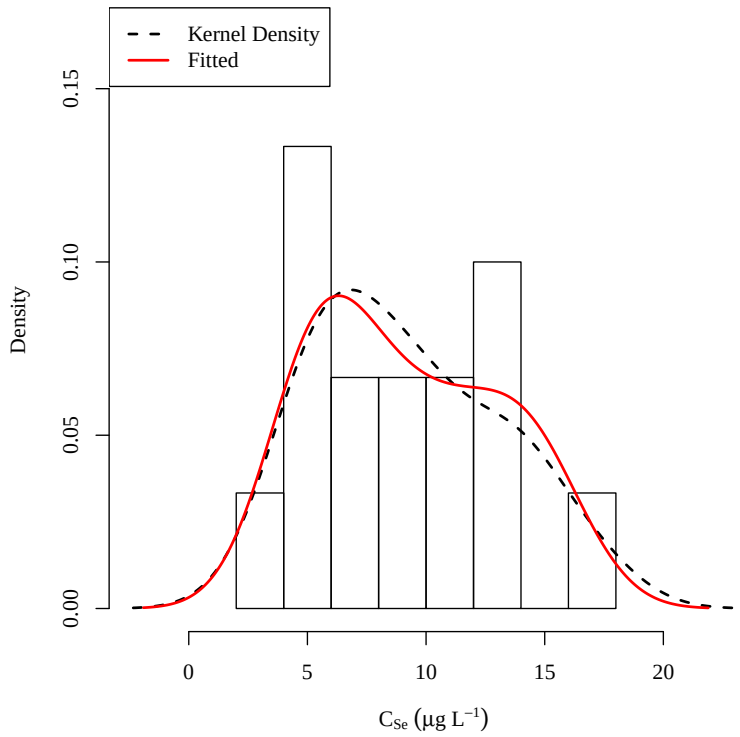
$C_{ARK,d=x}$

Figure 2.9 (Cont). DSR upstream boundary selenium estimate residual distribution analysis.

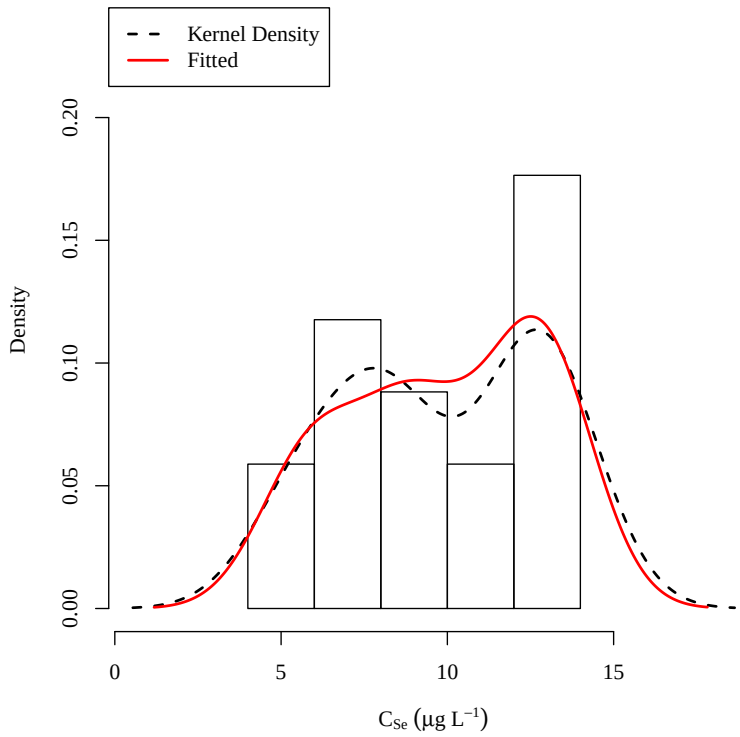
A number of the chosen residual distributions do not appear to fit the data. The residuals for the USR outlet appear to not be normally distributed. The distributions for CANSWKCO and TIMSWICO do not appear to be good fits. The data and distribution for HRC194CO does not have enough data to allow for a conclusive analysis. For the most part, the lack of a good fit can be traced back to data collection. First, for the poorly fit distributions, there was insufficient data collected. Second, there was a tendency for the samples to be taken during the same time frame each year. If selenium concentration is seasonal, then the samples should have been taken relatively equally spaced throughout the year to capture the seasonal variation.

These selenium concentration error distributions were not discarded due to these findings. It was assumed that the error could be best described by the best fit distributions determined in this analysis. Future data should be included in future analyses to develop more accurate estimation models and error distribution.

To test statistical software's ability to generate the required distribution, residuals were plotted as a histogram overlain with a kernel density estimate as shown in Figures 2.10 and 2.11 for the USR and DSR, respectively. The black line is the kernel density estimate of the residuals. The red line is the kernel density estimate of 5000 draws from the best fit regression model. In spite of the earlier findings that some of the best fit selenium concentration error distributions are not good fits, all graphs visually indicated that the fitted distributions are not as poor as expected.

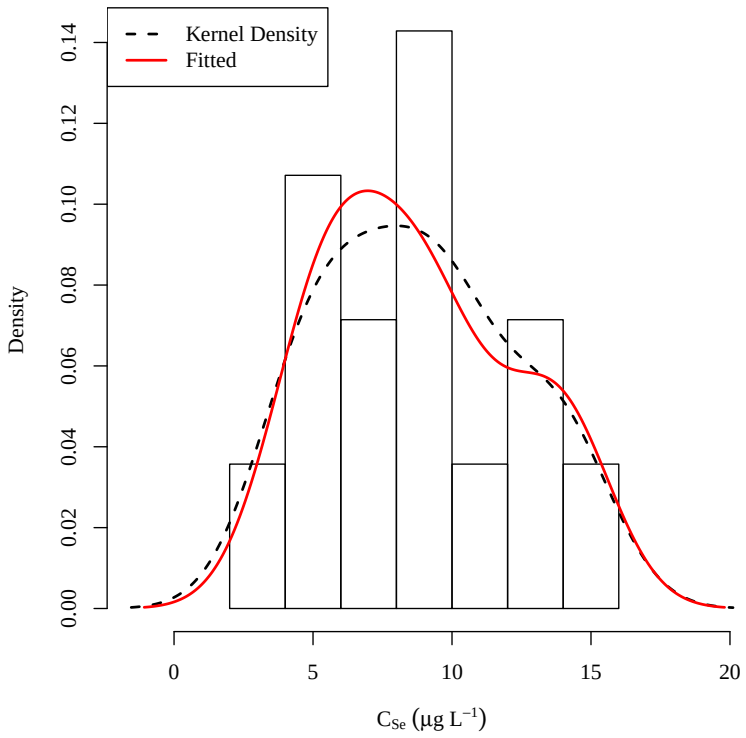


C_{U163}

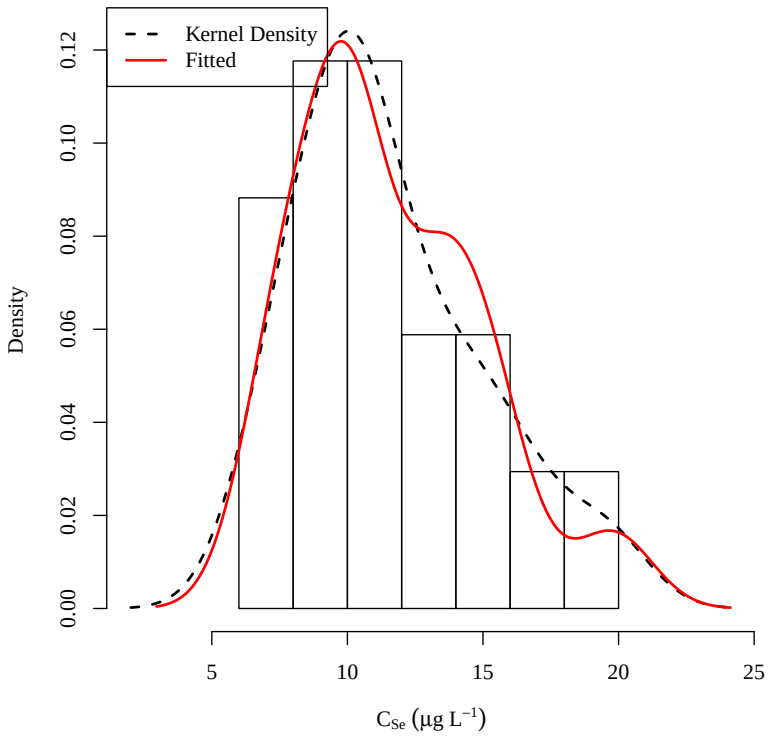


C_{U201}

Figure 2.10. USR selenium estimate residual histogram. The kernel density of the residuals is plotted over the histogram. Similar plots for the rest of the linear model analyses are included in the appendix.

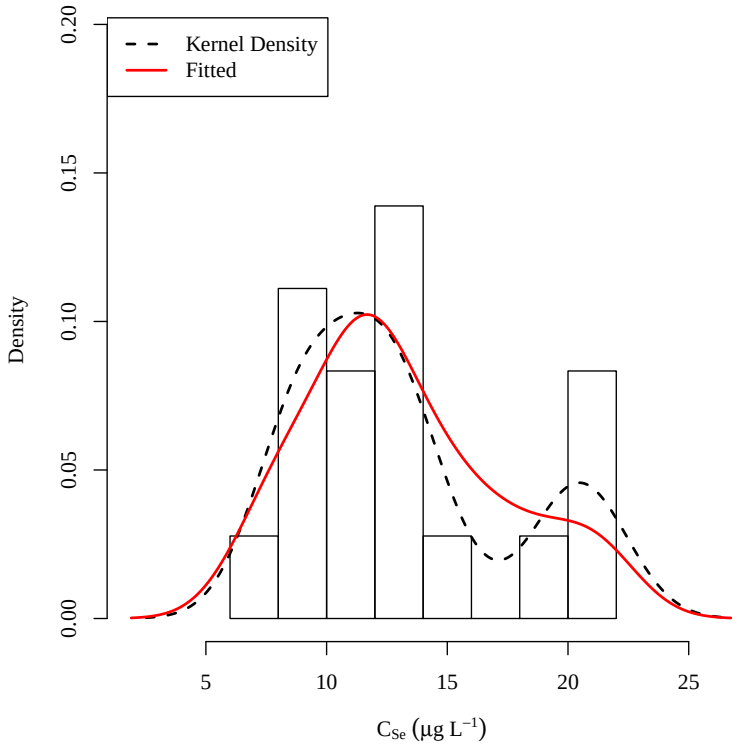


C_{U167}

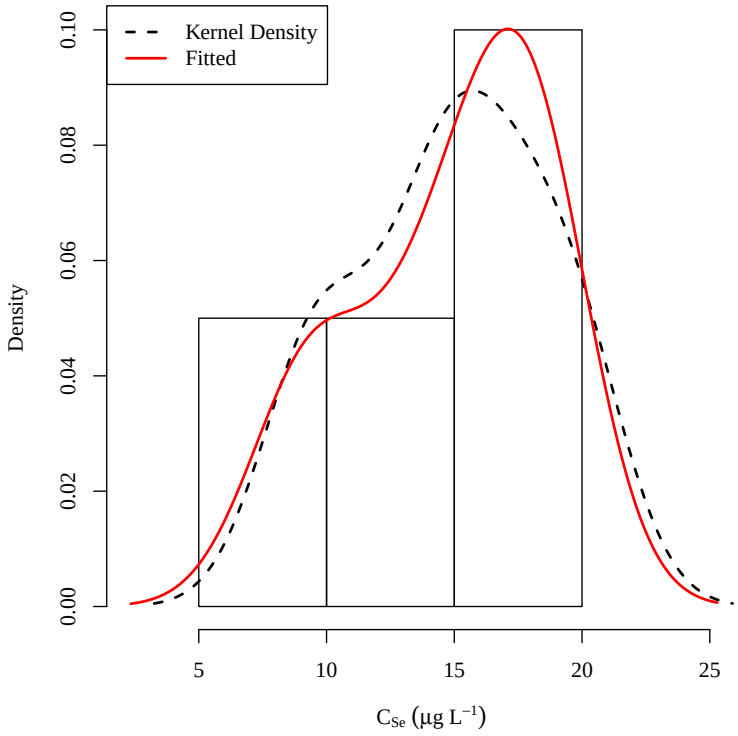


C_{U74}

Figure 2.10 (Cont). USR selenium estimate residual histogram.

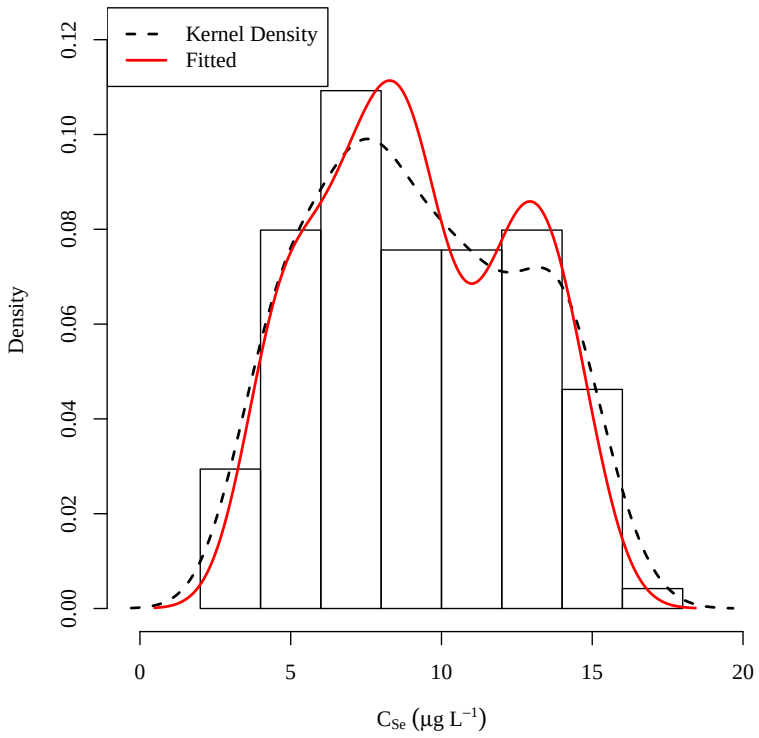


C_{U60}

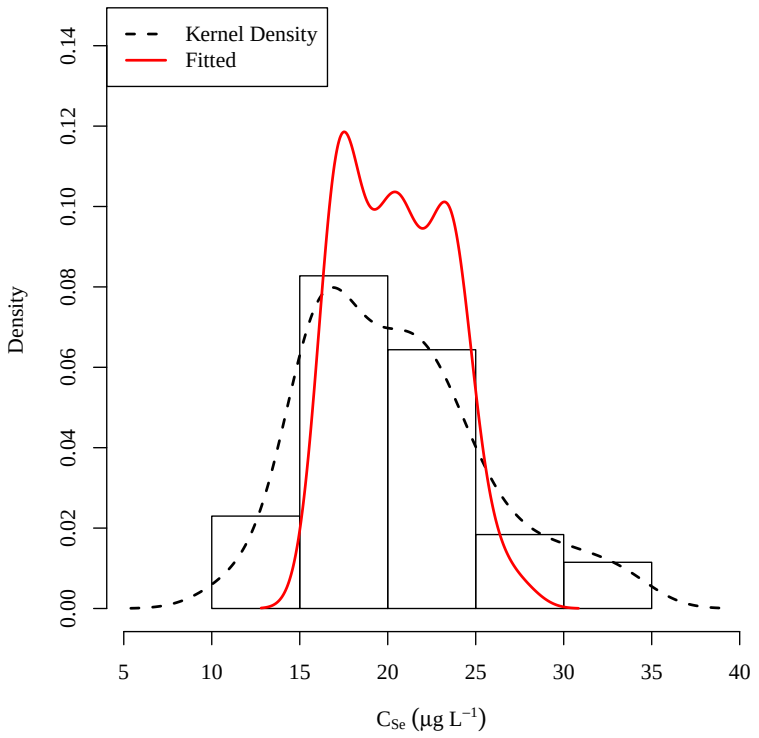


C_{U207}

Figure 2.10 (Cont). USR selenium estimate residual histogram.

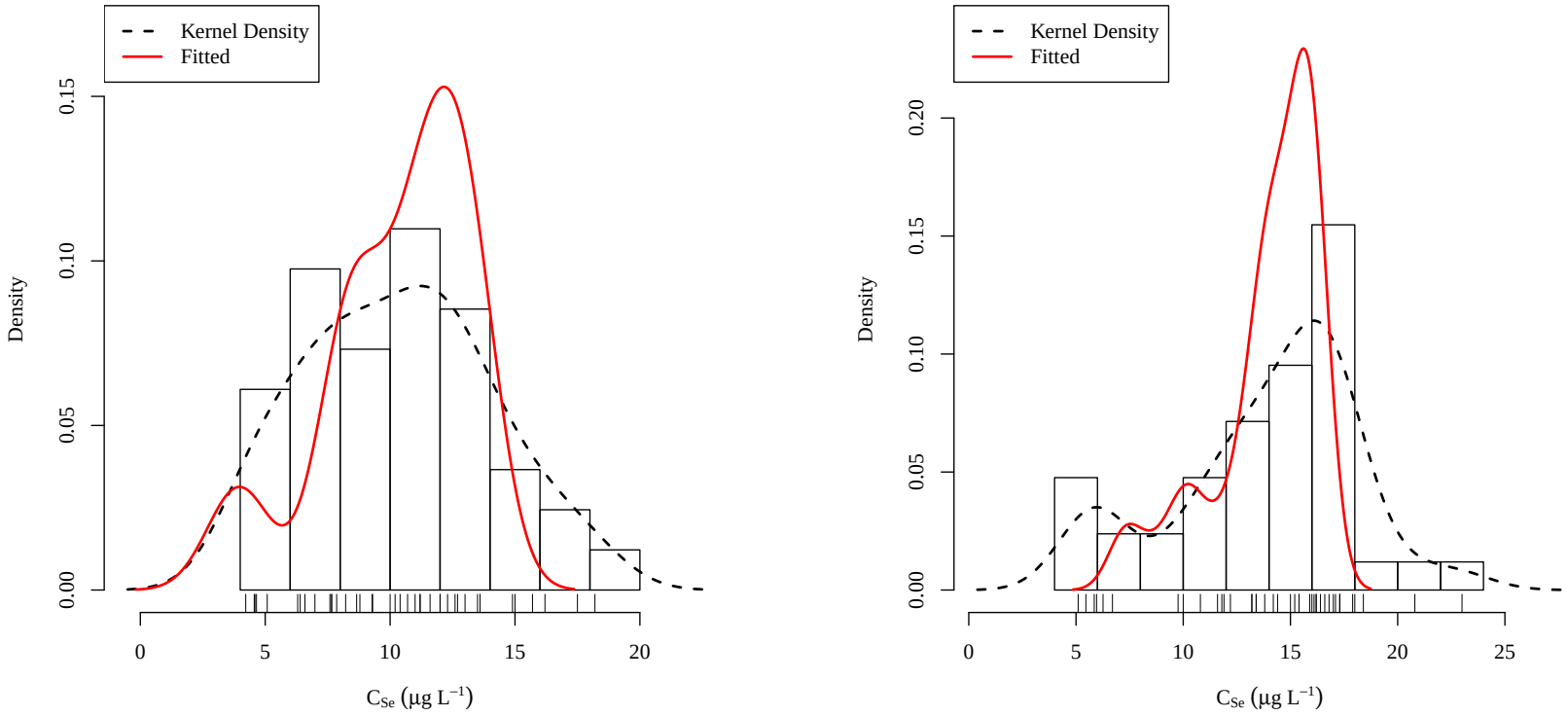


$C_{ARK,d=x}$



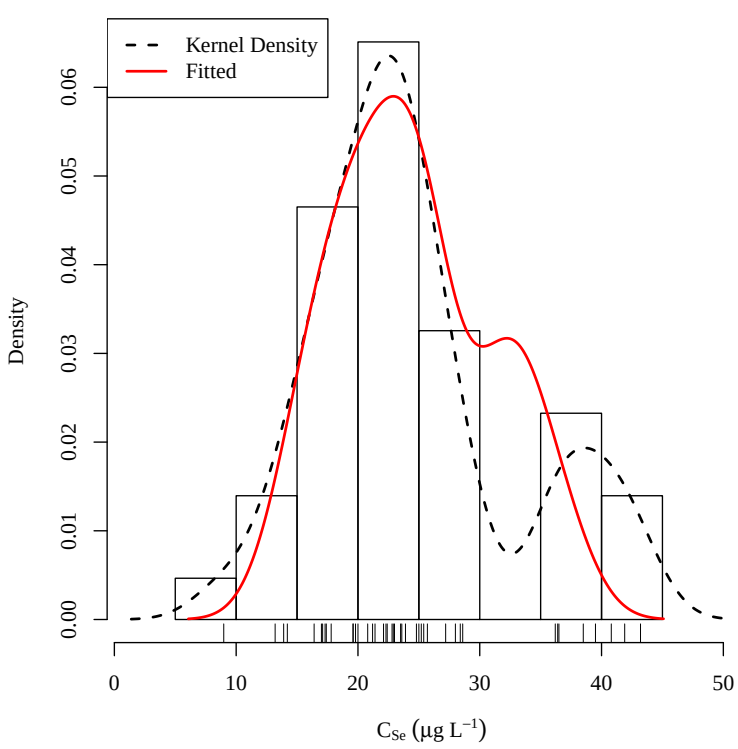
$C_{LAJWWTP}$

Figure 2.10 (Cont). USR selenium estimate residual histogram.

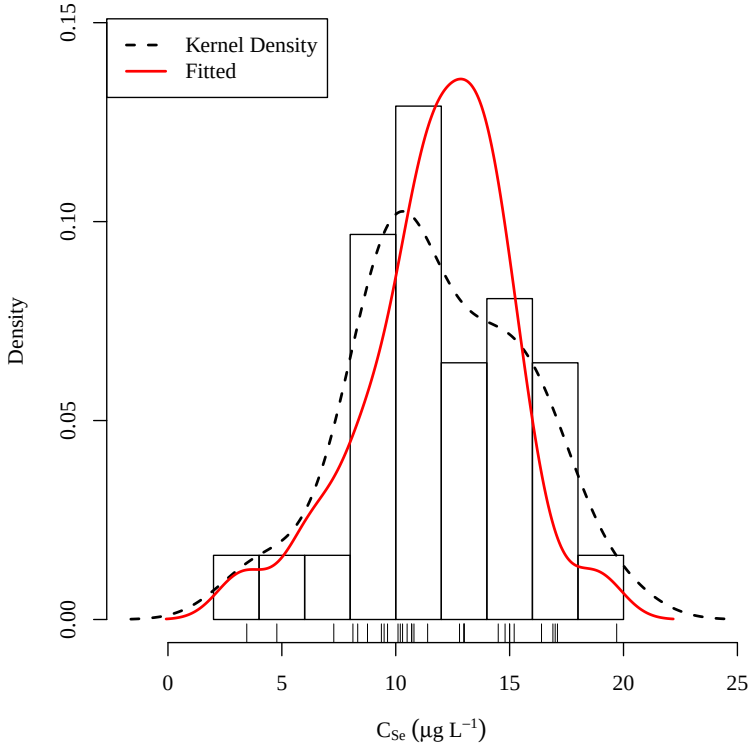


C_{D101C} C_{D106C}

Figure 2.11. DSR selenium estimate residual histogram. The kernel density of the residuals is plotted over the histogram. Similar plots for the rest of the linear model analyses are included in the appendix.

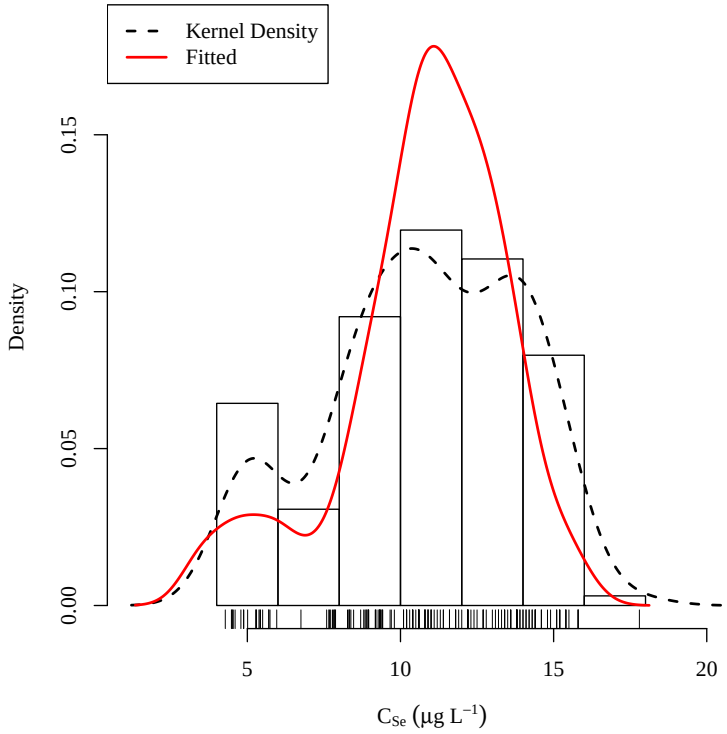


C_{D23}



C_{D57}

Figure 2.11 (Cont). DSR selenium estimate residual histogram.



$C_{ARK,d=x}$

Figure 2.11 (Cont). DSR selenium estimate residual histogram.

The primary purpose of the next set of analyses was to determine if the computational code and assumptions used to generate the stochastic distributions of dissolved selenium concentrations were performed correctly. The first analysis was to compare the calculated dissolved selenium concentration values to the measured results from the collected field samples. Figure 2.12 is a box plot of the sampled selenium concentrations at the various sampling locations along the main stem of the river and the tributaries in the USR. The sample locations are arranged with the upstream on the left and the downstream on the right, in order. The value "n" above each sample location is the number of samples collected at each site. Concentrations are measured and reported in $\mu g\,L^{-1}$ of dissolved selenium. The boxes encompass the first to the third quartile. The whiskers extend to 1.5 times the inter quartile range. Blue tinted boxes indicate dissolved selenium concentrations within tributaries all other boxes are from samples collected within the main stem of the Ark R.

Concentrations for the Rocky Ford Return Ditch in the USR and Frontier Ditch in the DSR are not included. Both ditches are assumed to have the same dissolved selenium concentration as a nearby calculated location. The Rocky Ford Return Ditch (RFDRETCO) uses the same concentration as the Rocky Ford Ditch (RFDCANCO) as it returns water from the main ditch to the Arkansas R. less than 1 km downstream of the main ditch head gate. The Frontier Ditch (FRODITKS) diverts water near the downstream end of the DSR and uses the concentrations calculated for this point.

Figure 2.13 is a box plot of the calculated estimated selenium concentration at the various key points in the USR mass balance model. The value "n" indicates the number of steps in the time series. The values used in the box plot are from the 1-D mean stochastic model. The blue tinted boxes indicate calculated dissolved selenium concentrations within

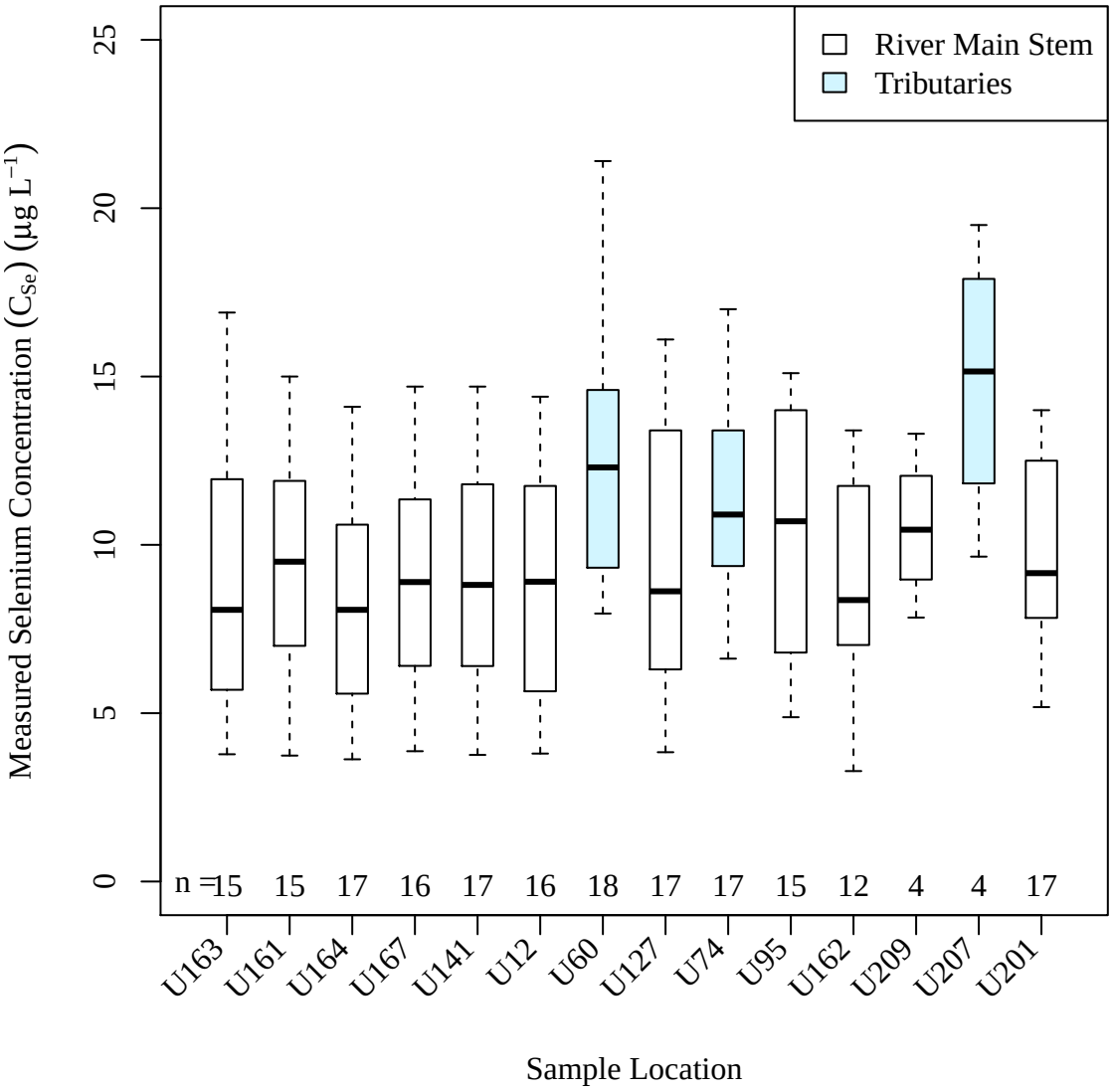


Figure 2.12. Measured Dissolved Selenium Concentrations in the USR.

tributaries and the tan tinted boxes indicate calculated dissolved selenium concentrations at the irrigation canal head gates.

Figure 2.14 is a box plot of the measured dissolved selenium concentrations at sample points in the main stem of the river and its main tributaries in the DSR. This plot is similar in fashion to figure 2.12.

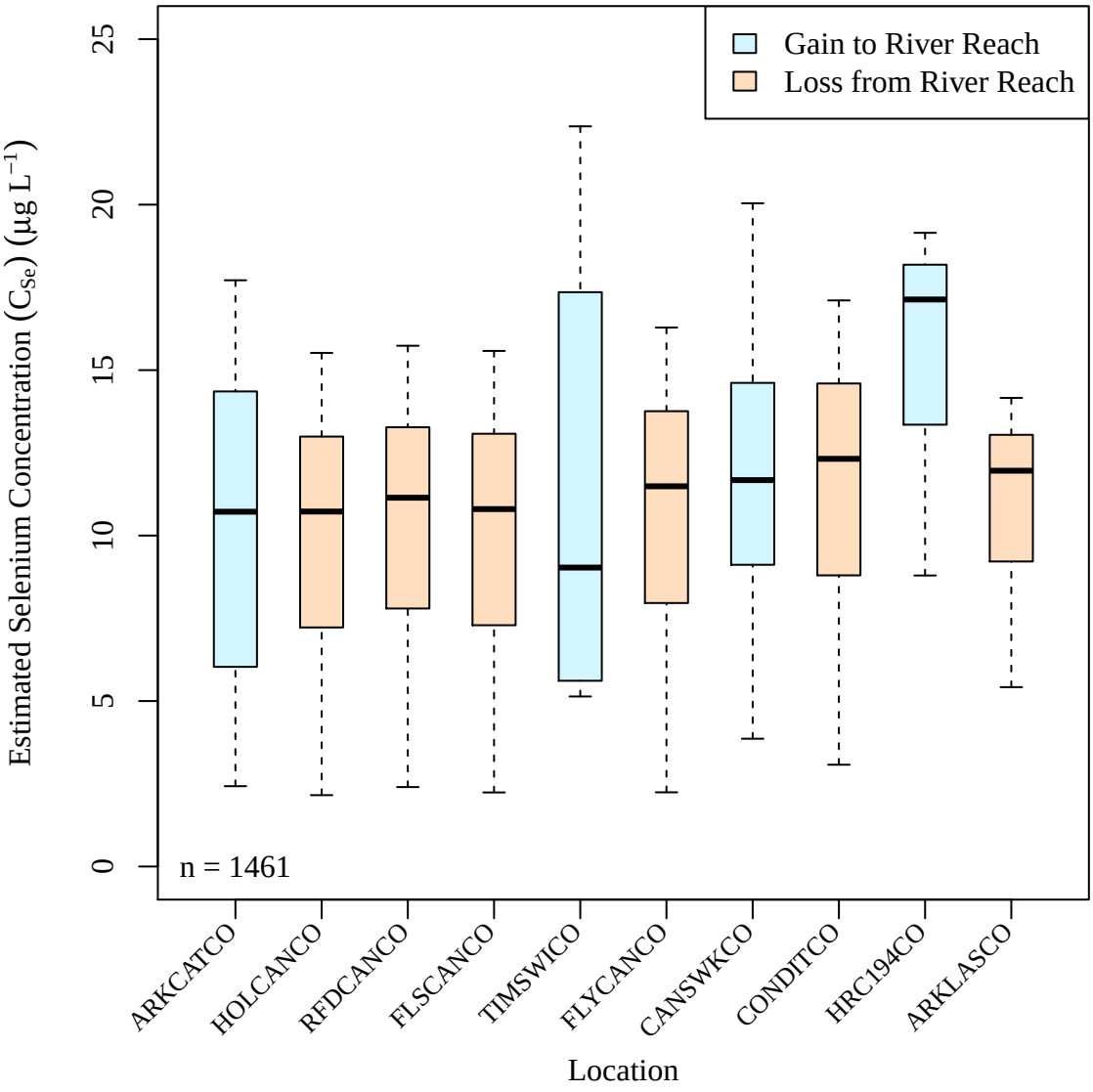


Figure 2.13. Calculated Dissolved Selenium Concentrations in the USR.

Figure 2.15 is a box plot of the measured dissolved selenium concentrations at sample points in the main stem of the river and its main tributaries in the DSR. This plot is similar in fashion to figure 2.13.

These four figures (2.12 to 2.15) compare the measured dissolved selenium concentration values with the estimated values. These figures are used along with tables **??** and **??** in chapter **??** to make this comparison. Study region sample locations along the Arkansas R.

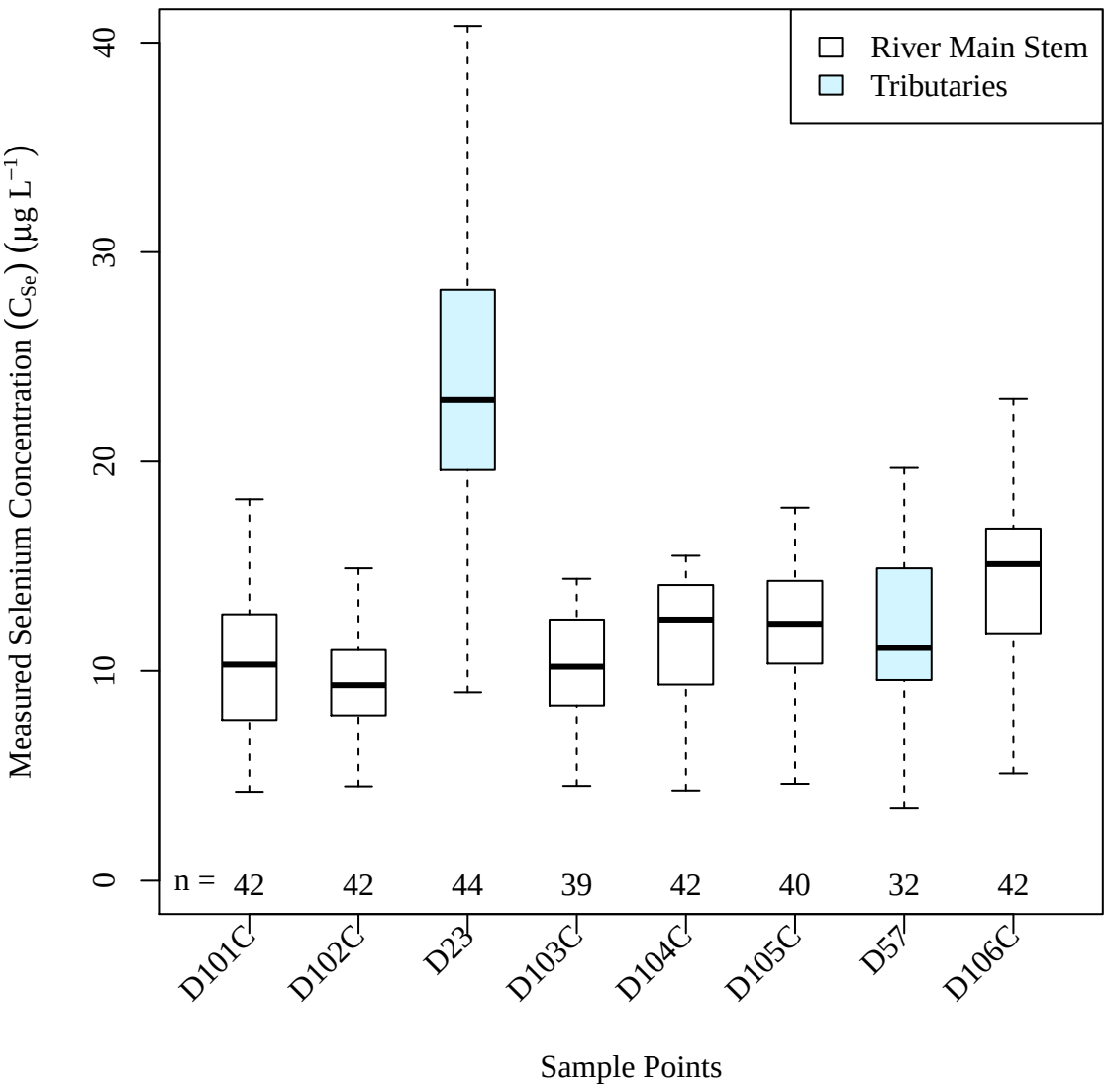


Figure 2.14. Measured Dissolved Selenium Concentrations in the DSR.

are not necessarily located at the same places where calculated dissolved selenium concentrations are required.

In all cases but one, the graphs support the statement that the calculated selenium concentration values are representative of the actual recorded values. Timpas Creek (TIMSWICO) in the USR appears to be the only exception. Here it appears that the calculated dissolved selenium concentrations are far lower than the measured values.

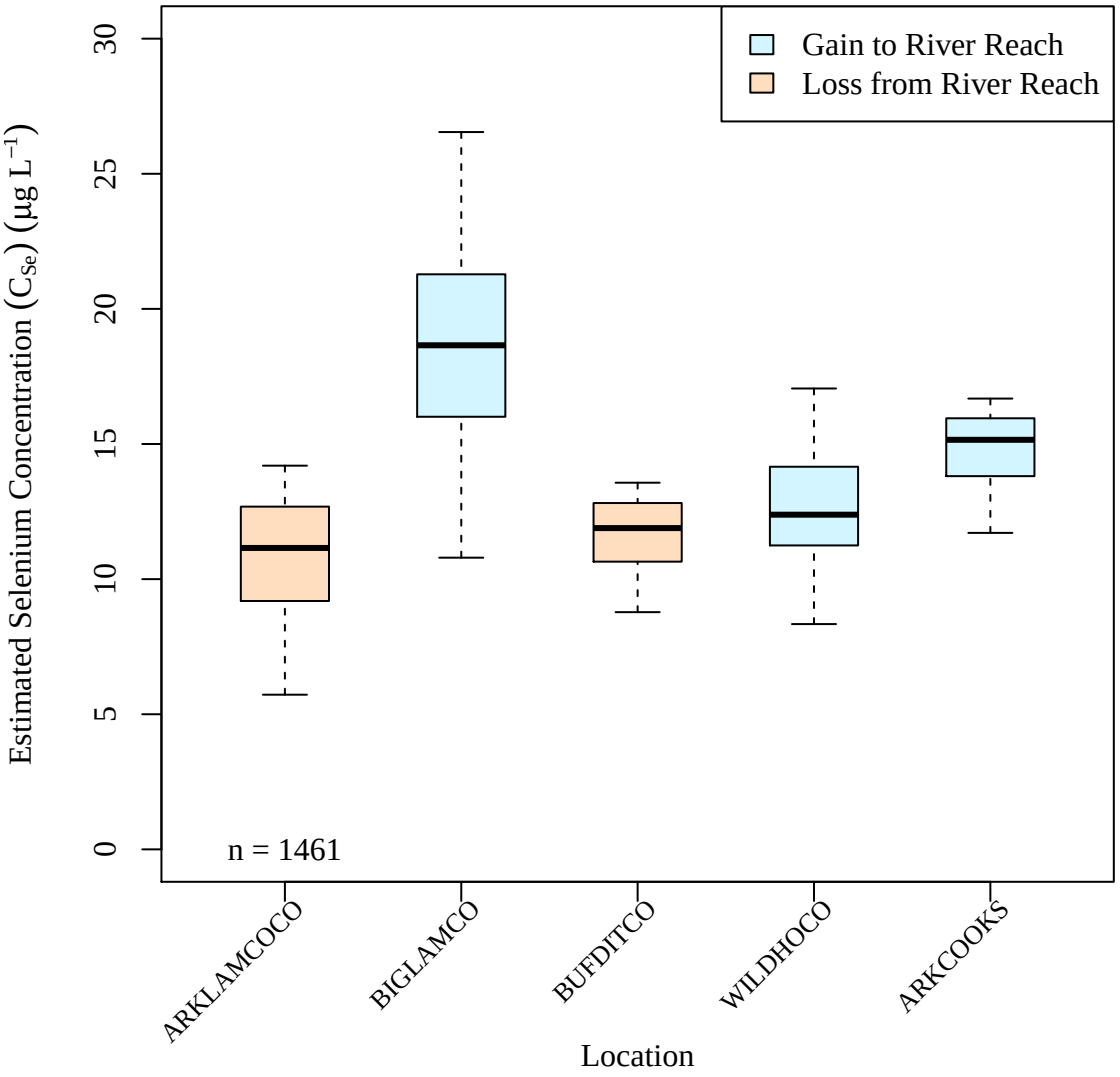
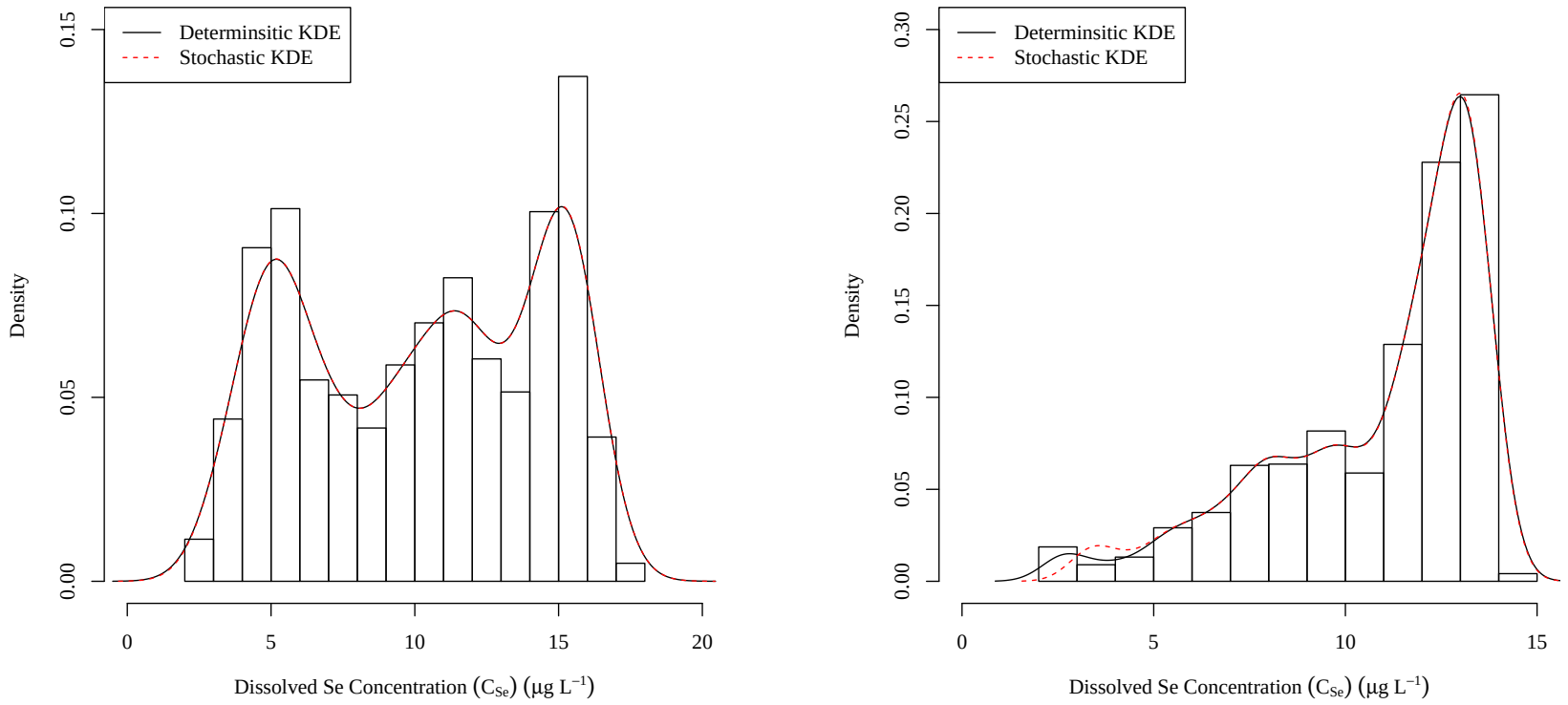


Figure 2.15. Calculated Dissolved Selenium Concentrations in the DSR.

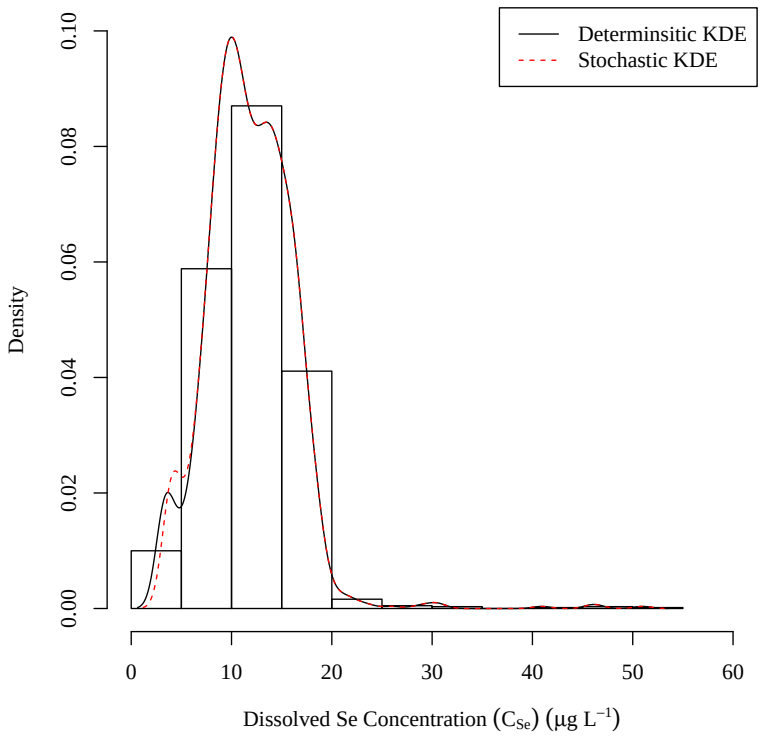
This is possibly caused by three factors. The first is the sampling frequency. The sample results represented in figure 2.12 are not a uniform representation of the possible concentration values throughout a calendar year. The sample values more heavily consider three months, March, May, and July, and either minimal consider or ignore all other months. The second is the nature of flows within Timpas Creek. The lower portion of the creek serves as a return flow channel for field irrigation runoff. The selenium concentration of the runoff and the effects of other water constituents are not known.

The second analysis was to compare the results between the deterministic model time-series and the stochastic model mean time-series results to determine if there is any unacceptable variance between the models. The distributions of the dissolved selenium concentrations calculated for both the deterministic and stochastic models were graphically compared in Figures 2.16 and 2.17 for the USR and DSR, respectively. The histogram and the black KDE are of the calculated deterministic model values. The red dashed KDE represents the distribution of the stochastic model mean time-series. Similar figures for all calculated concentrations are provided in the appendix. The third factor is the uncertainty with which dissolved selenium concentrations were calculated for Timpas Creek.

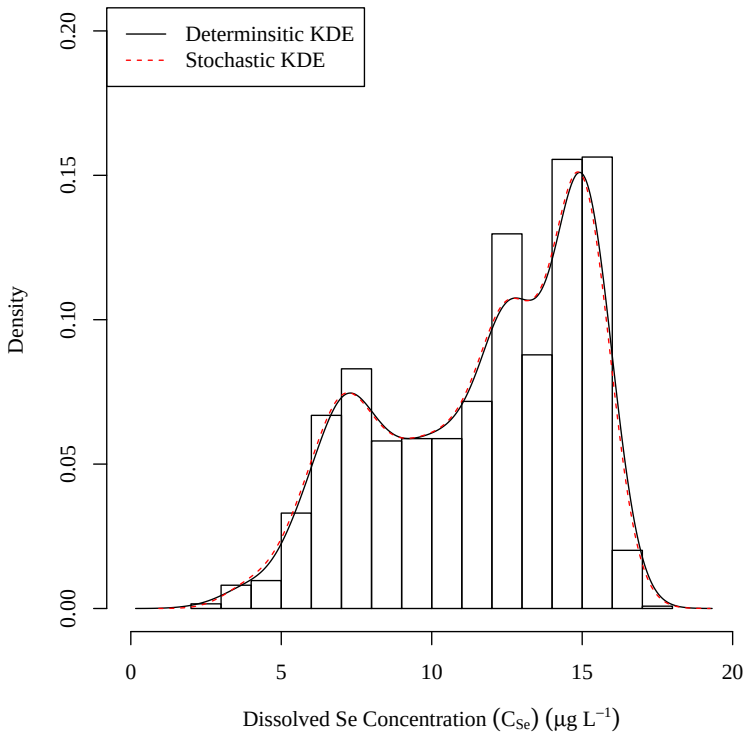


C_{U163} C_{U201}

Figure 2.16. USR dissolved selenium concentration distribution analysis. The histogram and the black KDE are of the calculated deterministic model values. The red dashed KDE represents the distribution of the stochastic model mean time series.

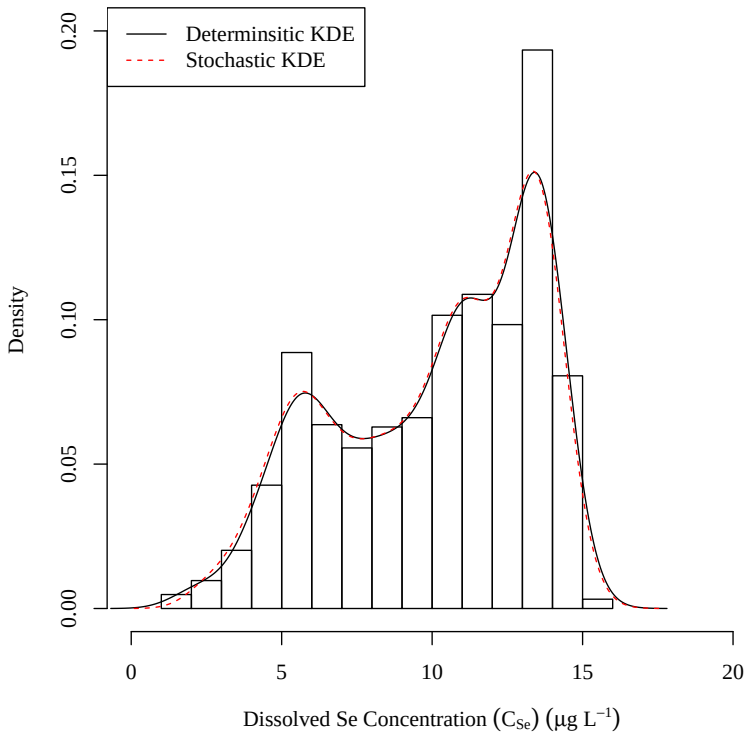


C_{U74}

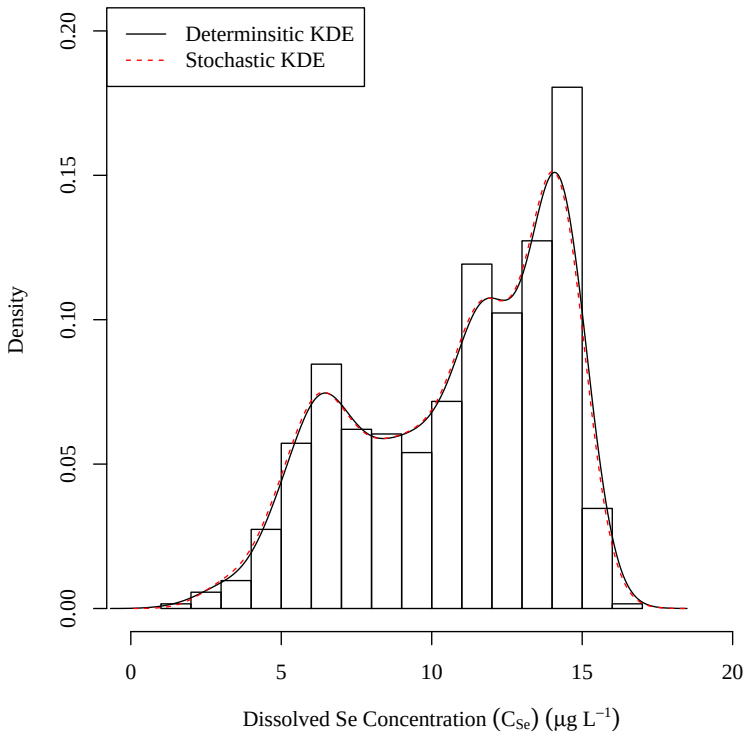


$C_{ARK,d=85.0}$

Figure 2.16 (Cont). USR selenium concentration linear model analysis graphs.

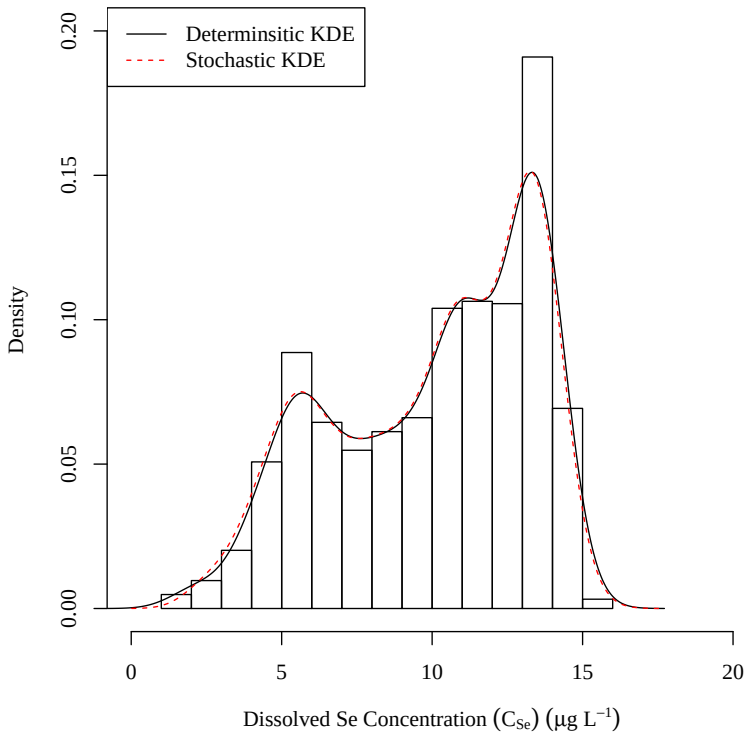


$C_{ARK,d=16.4}$

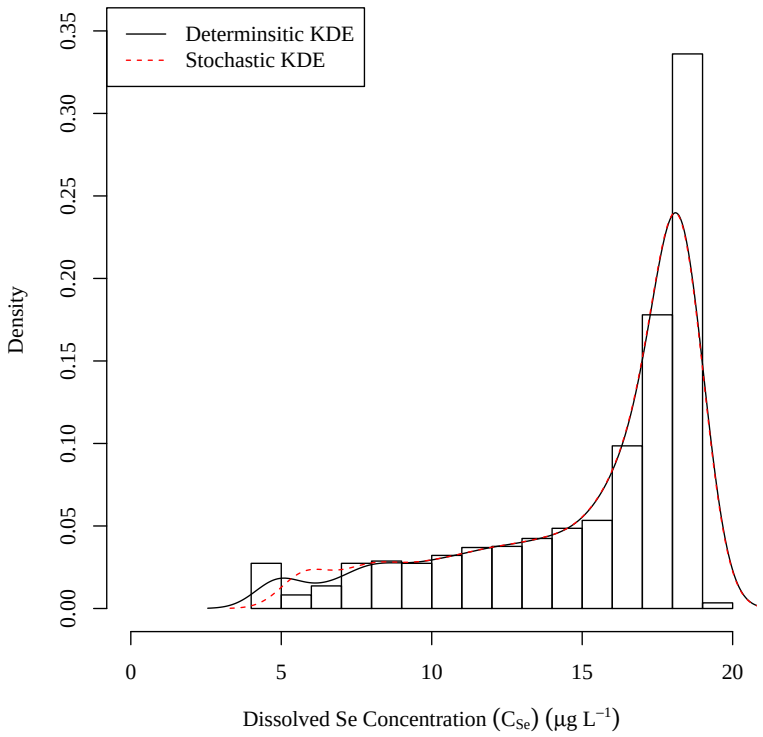


$C_{ARK,d=47.2}$

Figure 2.16 (Cont). USR selenium concentration linear model analysis graphs.

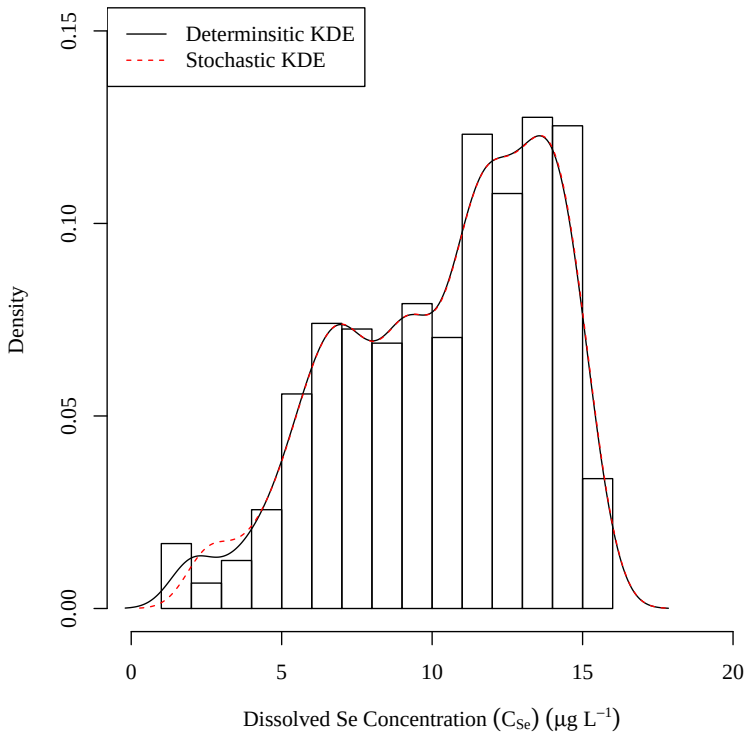


$C_{ARK,d=12.5}$

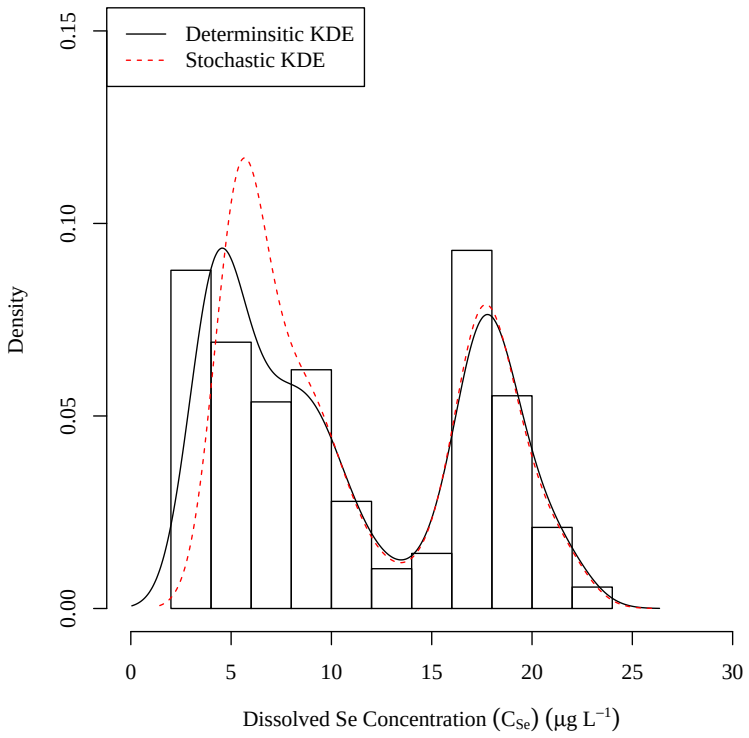


C_{207}

Figure 2.16 (Cont). USR selenium concentration linear model analysis graphs.



C_{U167}



C_{60}

Figure 2.16 (Cont). USR selenium concentration linear model analysis graphs.

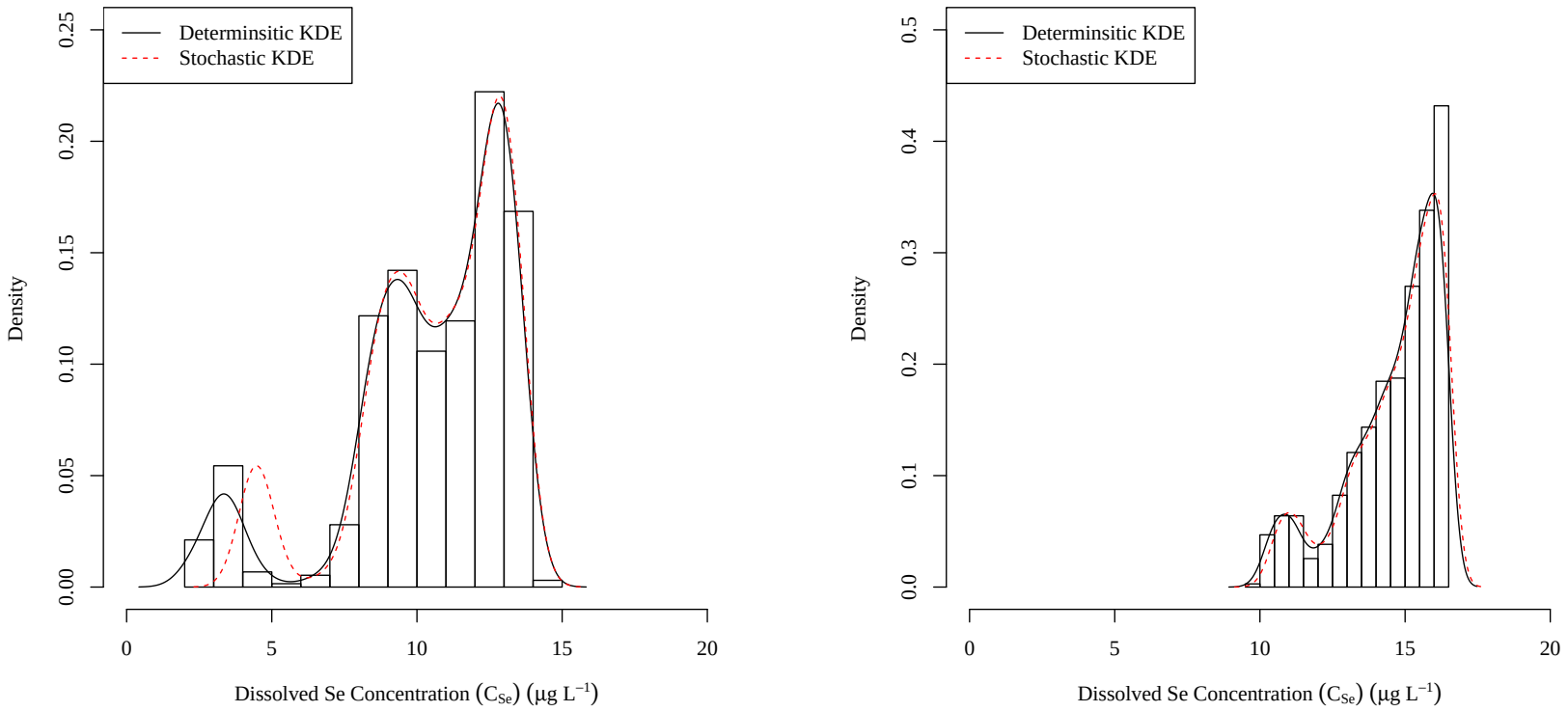
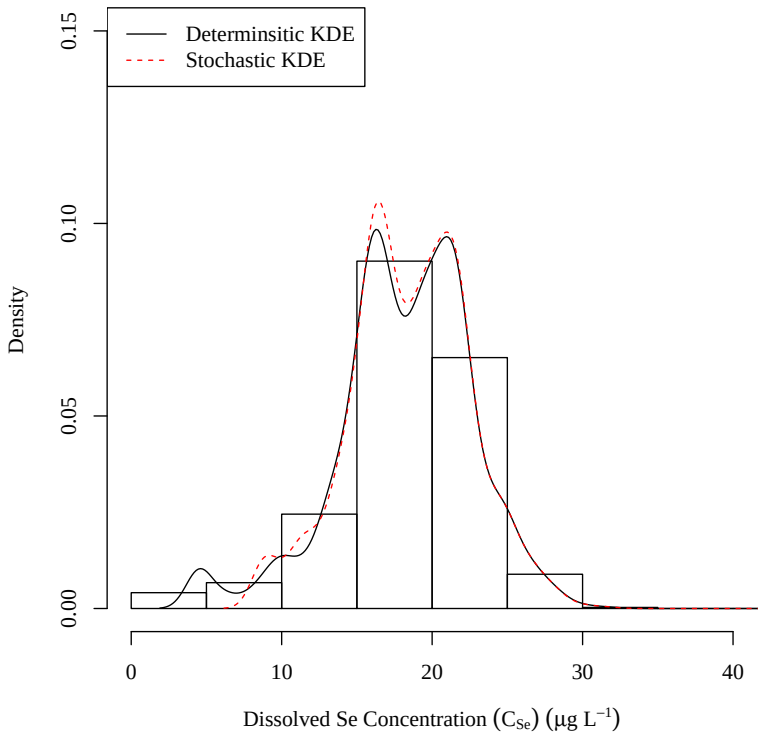
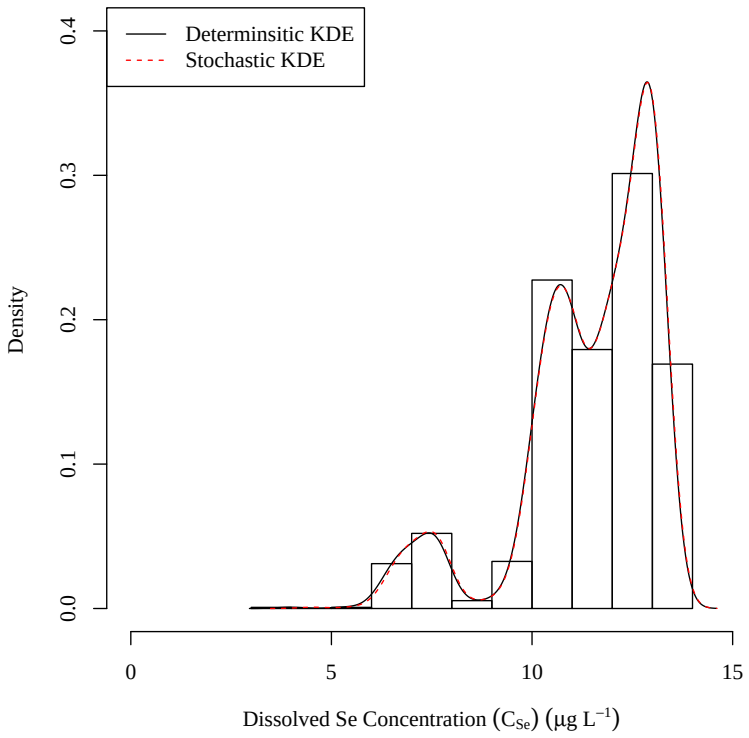


Figure 2.17. DSR dissolved selenium concentration distribution analysis. The histogram and the black KDE are of the calculated deterministic model values. The red dashed KDE represents the distribution of the stochastic model mean time series.

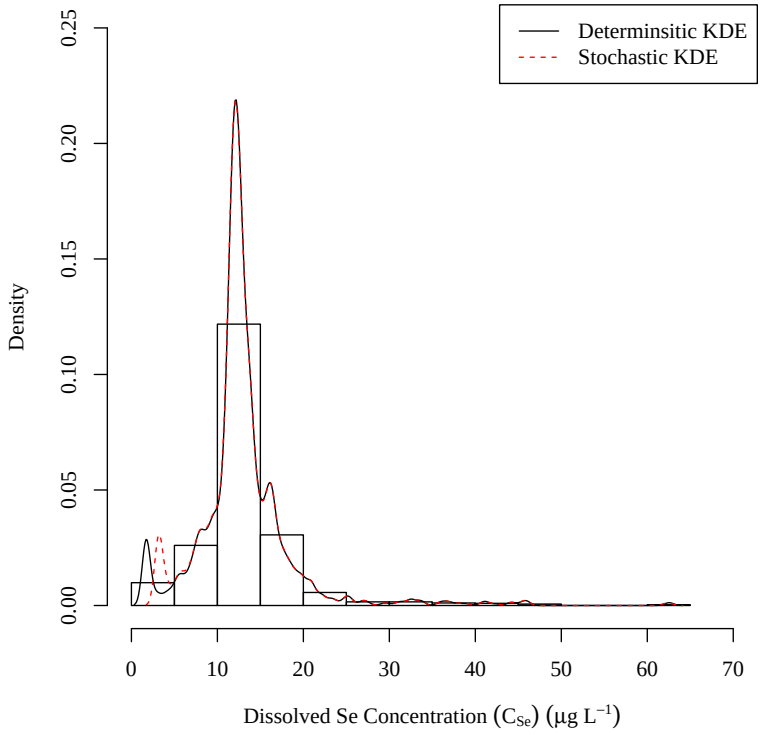


C_{D23}



$C_{ARK,d=37.7}$

Figure 2.17 (Cont). DSR dissolved selenium concentration distribution analysis.

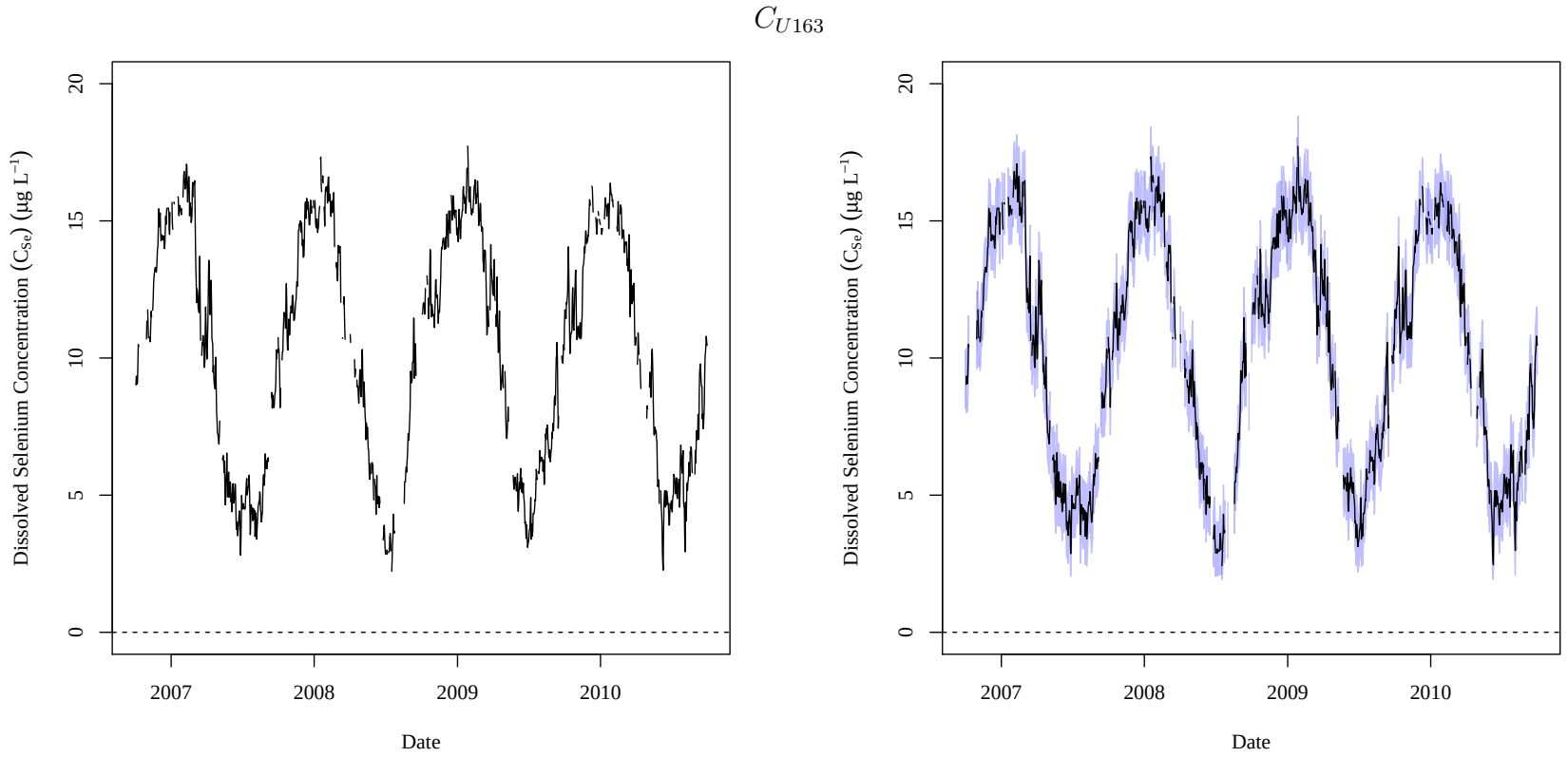


C_{D57}

Figure 2.17 (Cont). DSR dissolved selenium concentration distribution analysis.

These figures show that the values used for the stochastic model and the deterministic model have very similar distributions. In some cases there are slight deviations between the two distributions at lower concentration values. This is due to the uncertainty assigned to the stochastic concentration estimates being more noticeable at lower calculated concentrations.

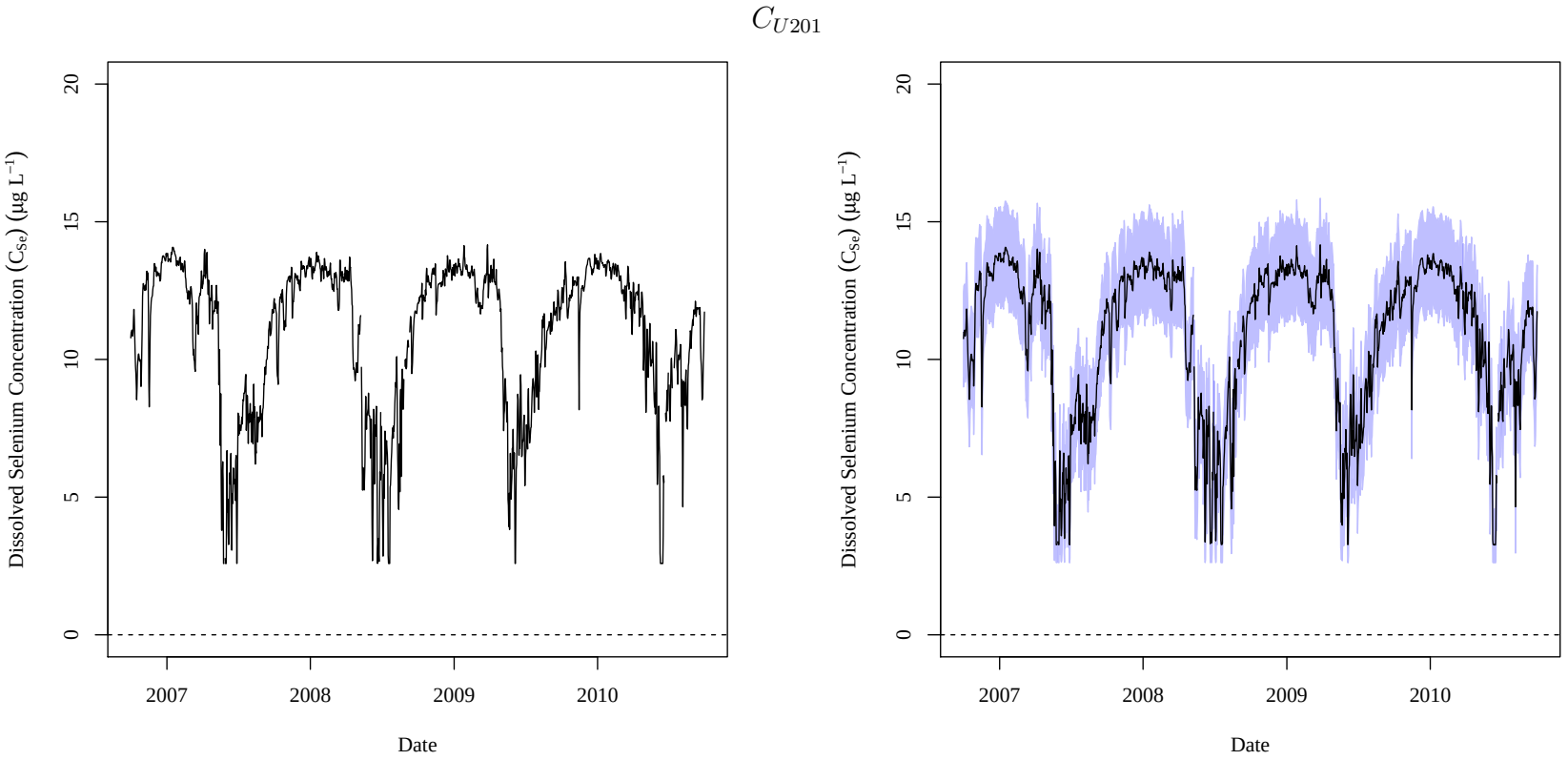
Time series plots of the concentration results from both the deterministic and stochastic models were prepared to visually analyze the relationship between the dissolved selenium concentration and the calendar date as shown in Figures 2.18 and 2.19 for the USR and DSR, respectively. Two sub-figures are provided. Sub-figure a is the deterministic model time series and sub-figure b is the stochastic model time series. The line is the mean of the realizations and the blue band is the 97.5th CIR.



(a) Deterministic Model.

(b) Stochastic Model.

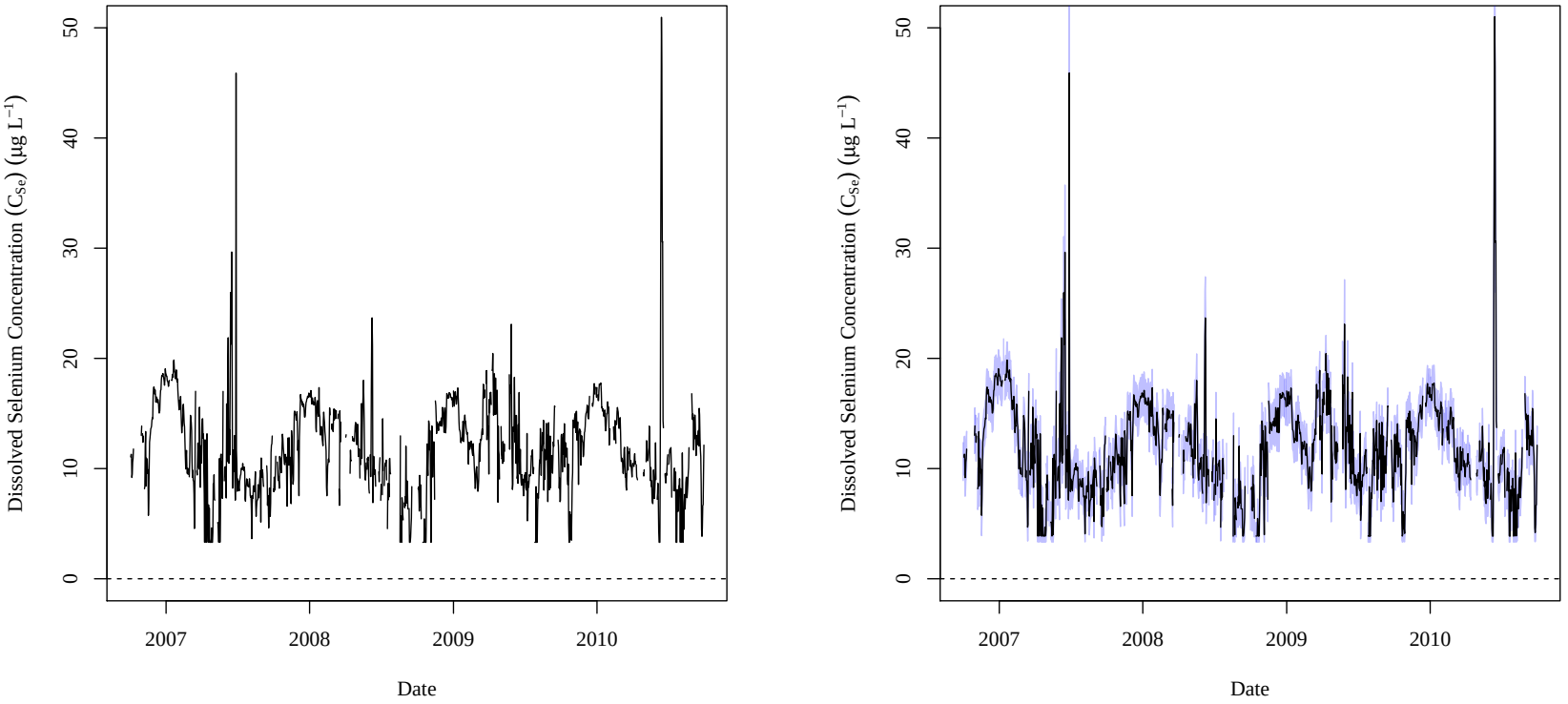
Figure 2.18. USR Deterministic and stochastic model time series of dissolved selenium concentration. For sub-figure b, the black line is the mean of the realizations and the blue band is the 97.5th CIR.



(a) Deterministic Model.

(b) Stochastic Model.

Figure 2.18 (Cont). USR Deterministic and stochastic model time series of dissolved selenium concentration.

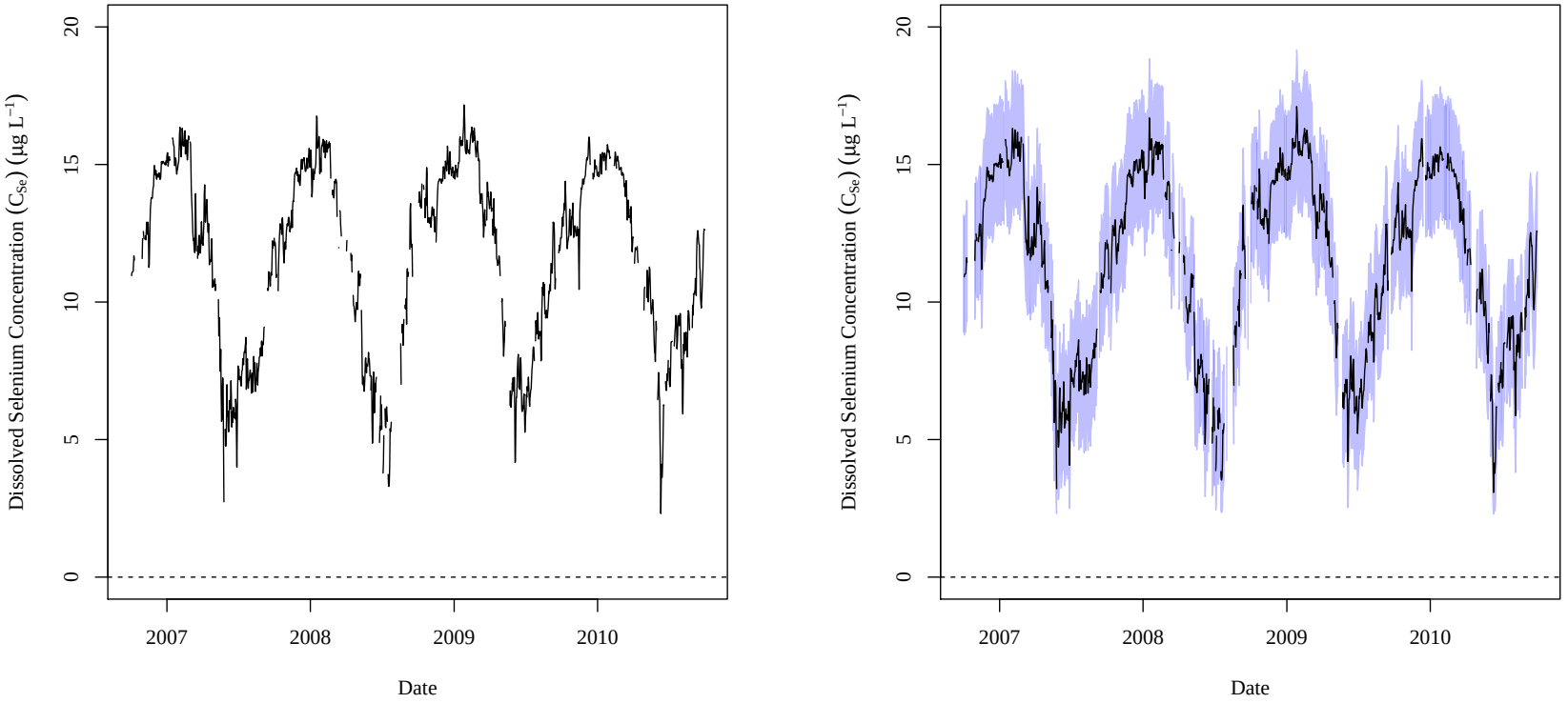


(a) Deterministic Model.

(b) Stochastic Model.

Figure 2.18 (Cont). USR Deterministic and stochastic model time series of dissolved selenium concentration.

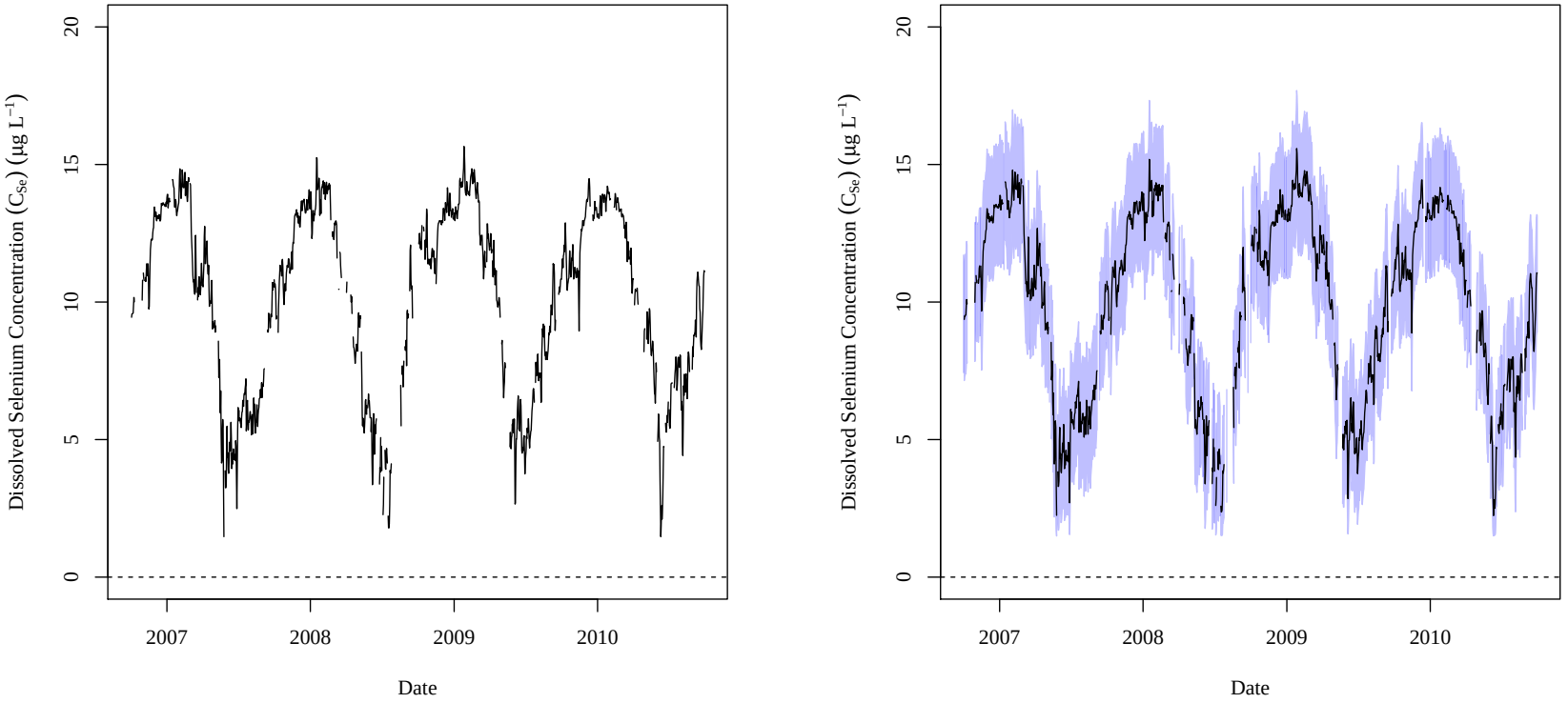
$C_{ARK,d=85.0}$



(a) Deterministic Model.

(b) Stochastic Model.

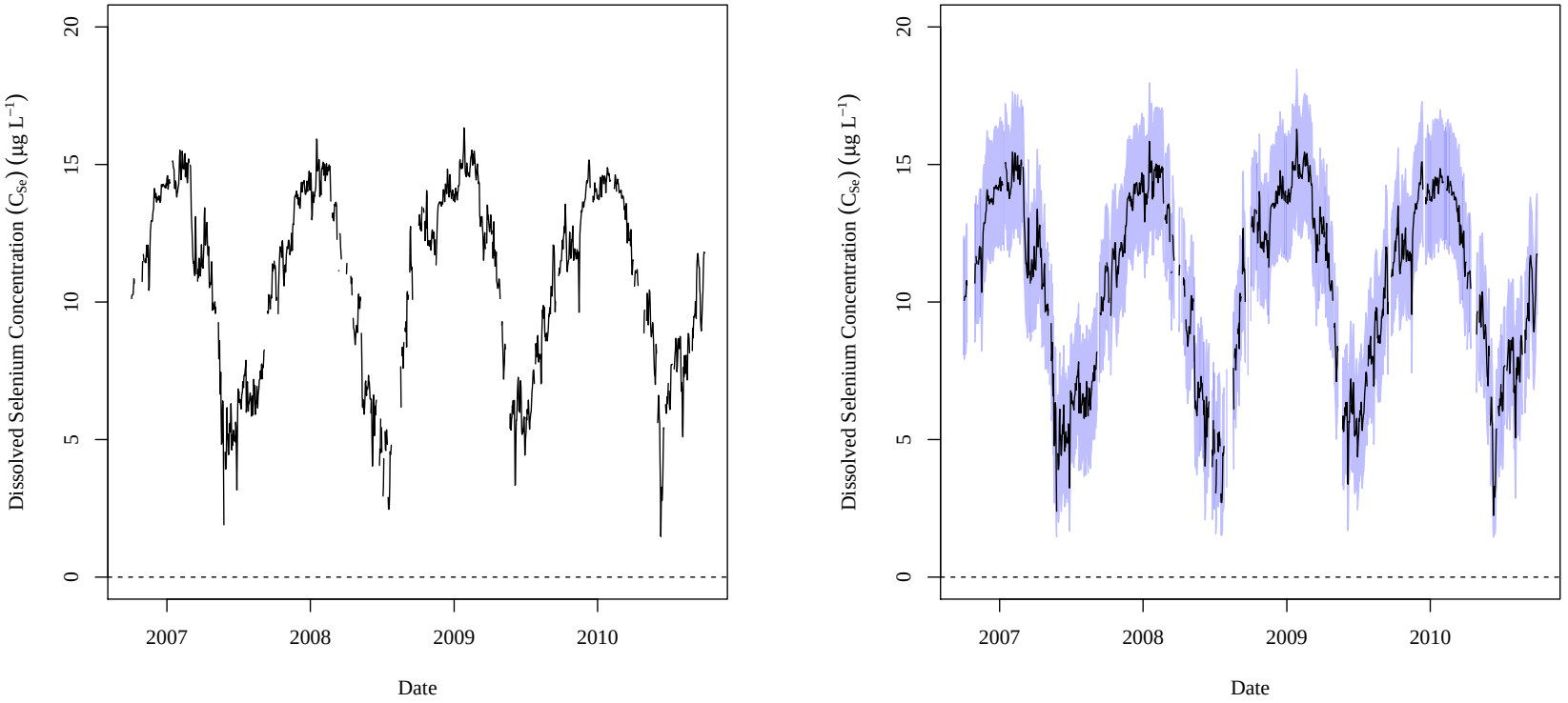
Figure 2.18 (Cont). USR Deterministic and stochastic model time series of dissolved selenium concentration.



(a) Deterministic Model.

(b) Stochastic Model.

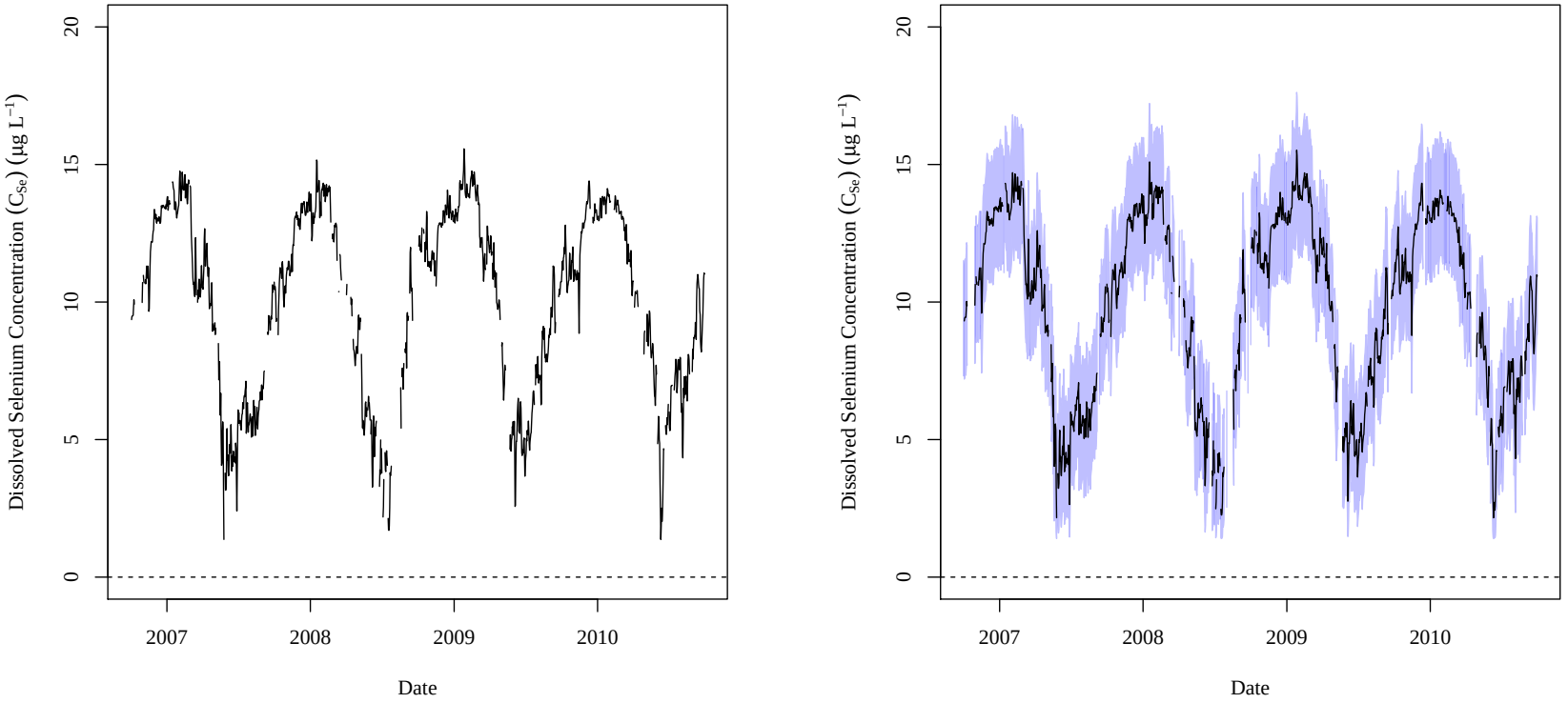
Figure 2.18 (Cont). USR Deterministic and stochastic model time series of dissolved selenium concentration.



(a) Deterministic Model.

(b) Stochastic Model.

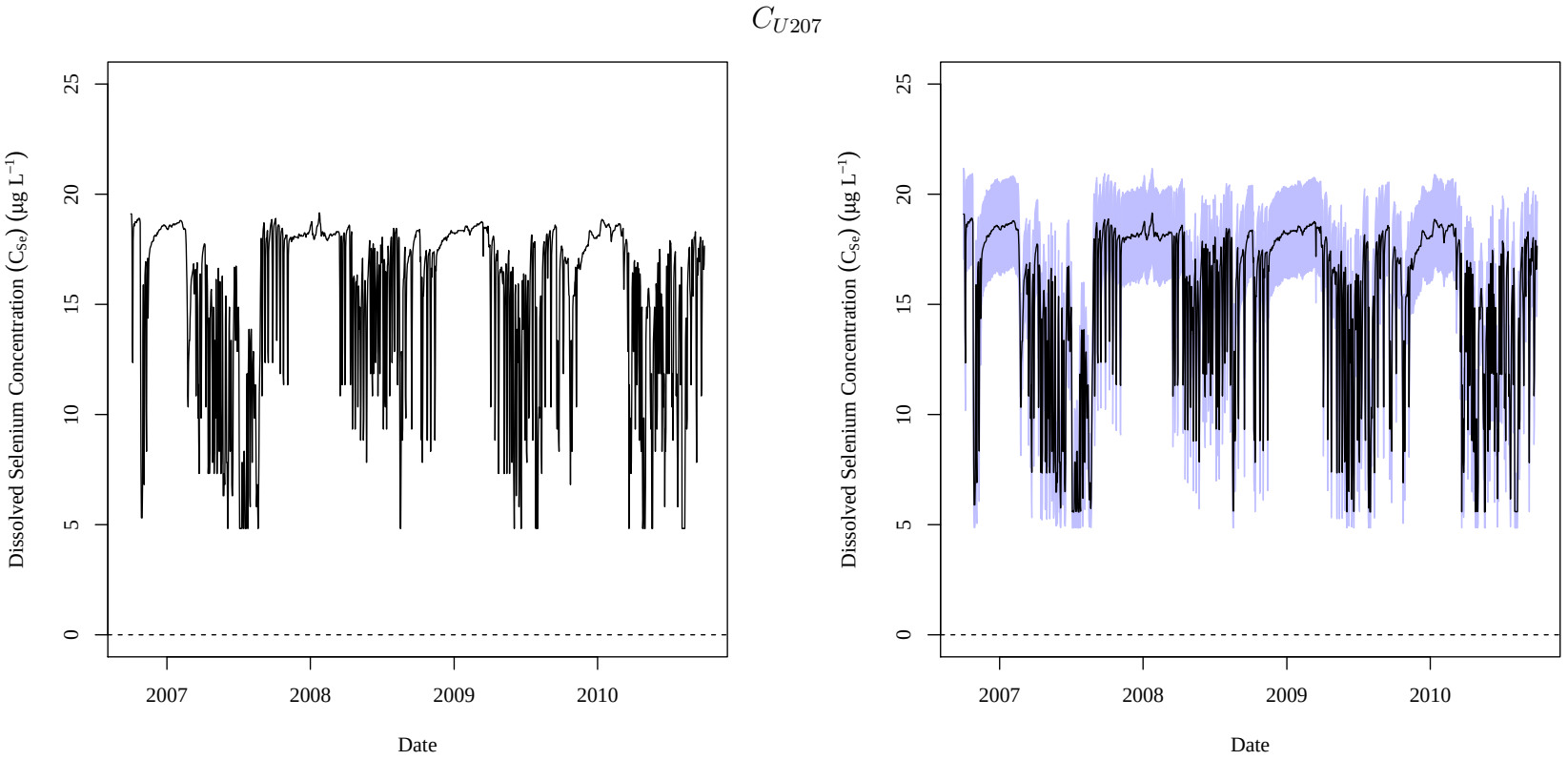
Figure 2.18 (Cont). USR Deterministic and stochastic model time series of dissolved selenium concentration.



(a) Deterministic Model.

(b) Stochastic Model.

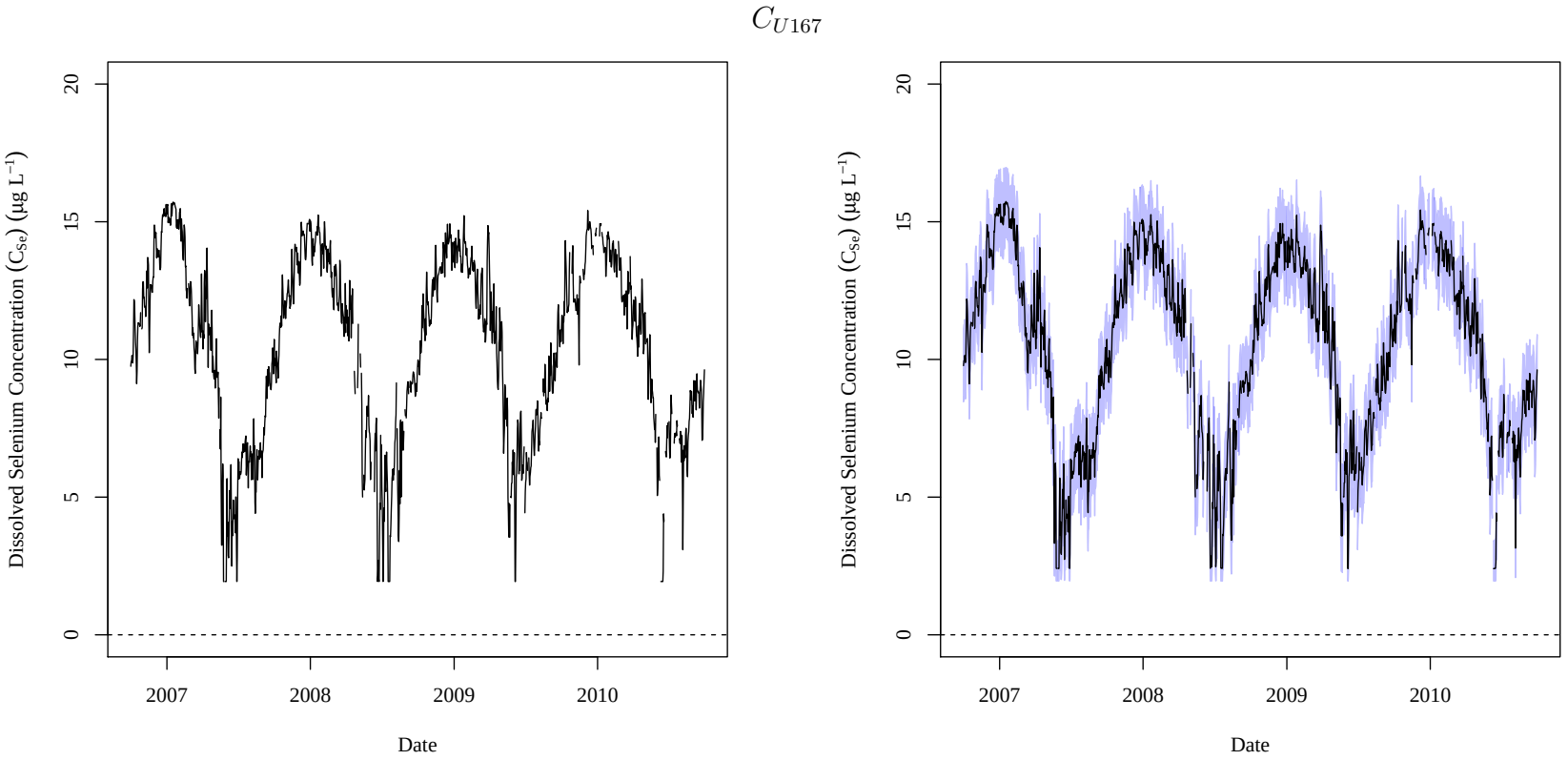
Figure 2.18 (Cont). USR Deterministic and stochastic model time series of dissolved selenium concentration.



(a) Deterministic Model.

(b) Stochastic Model.

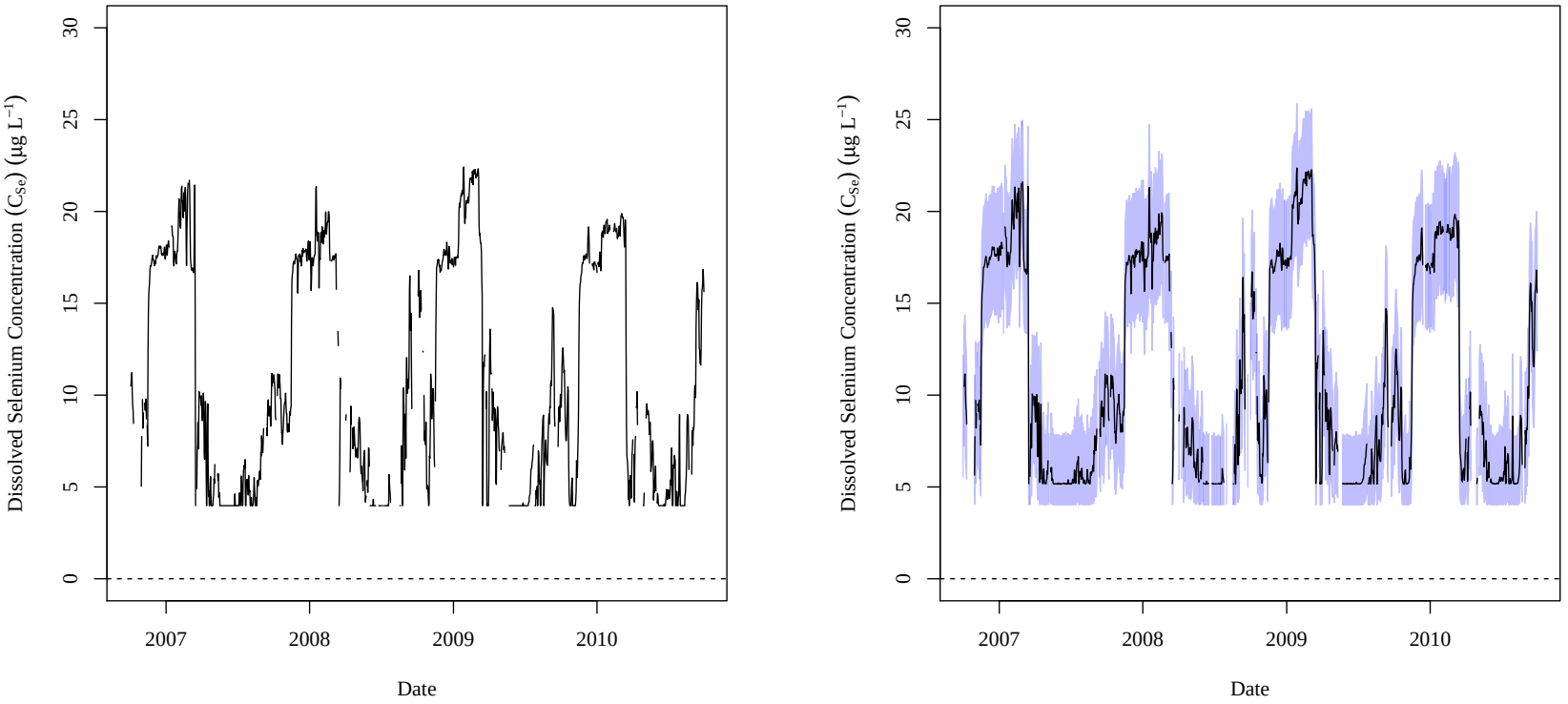
Figure 2.18 (Cont). USR Deterministic and stochastic model time series of dissolved selenium concentration.



(a) Deterministic Model.

(b) Stochastic Model.

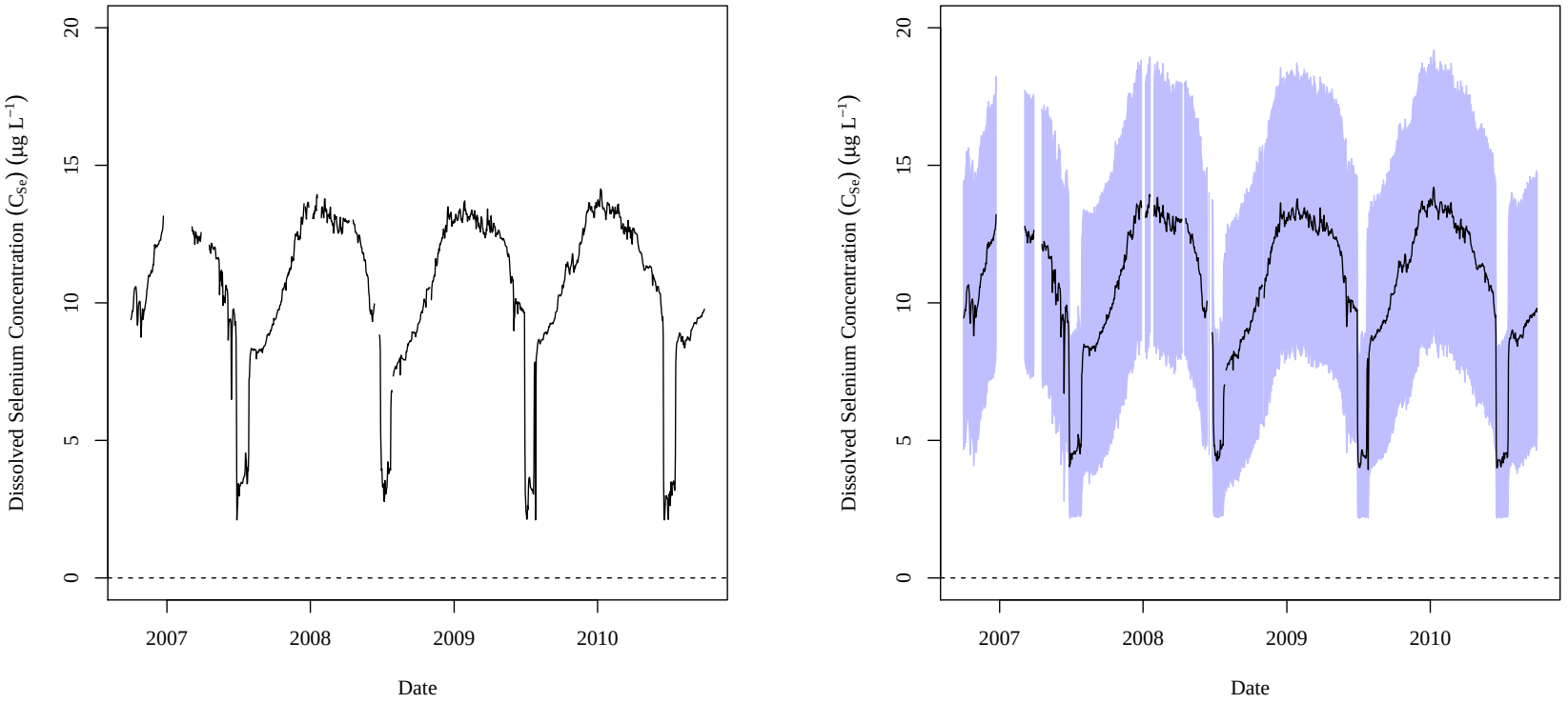
Figure 2.18 (Cont). USR Deterministic and stochastic model time series of dissolved selenium concentration.



(a) Deterministic Model.

(b) Stochastic Model.

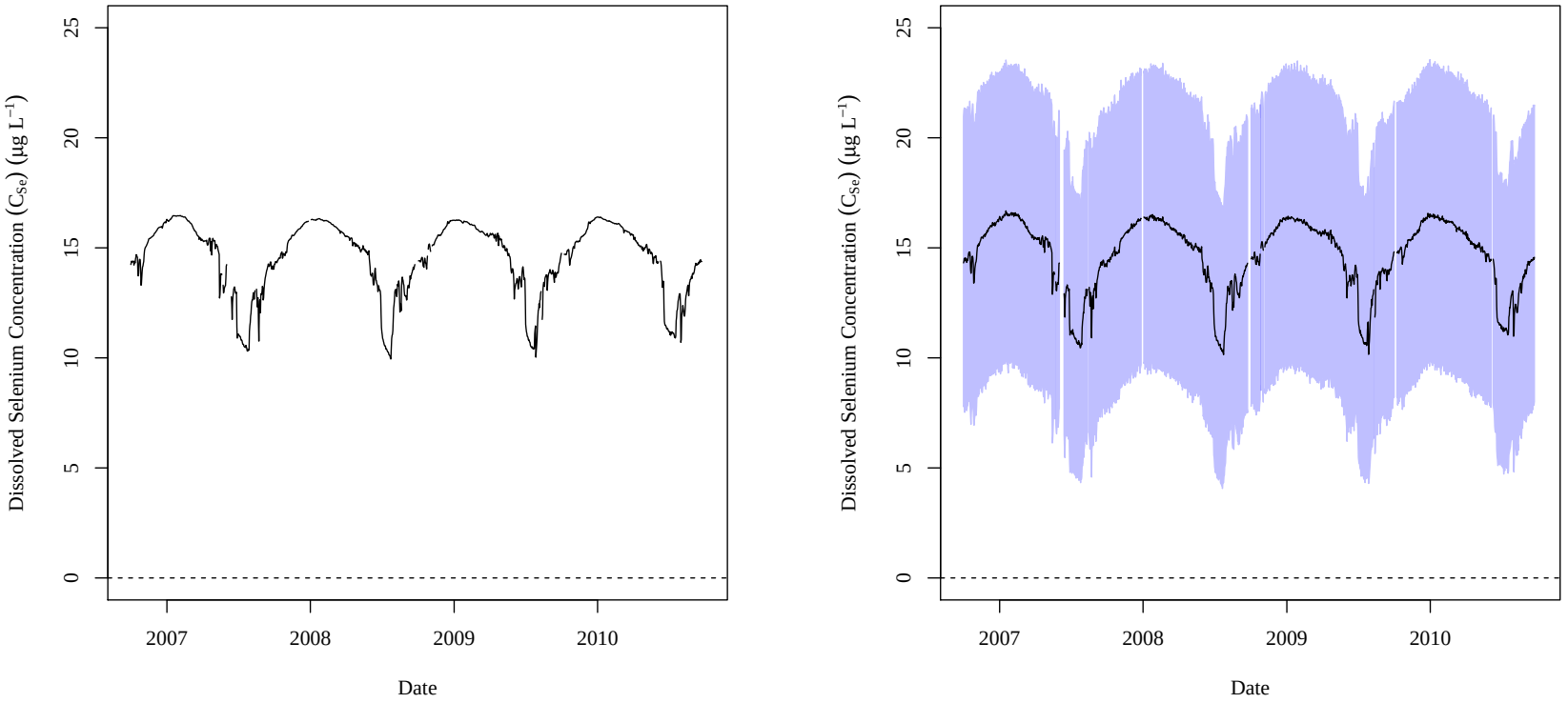
Figure 2.18 (Cont). USR Deterministic and stochastic model time series of dissolved selenium concentration.



(a) Deterministic Model.

(b) Stochastic Model.

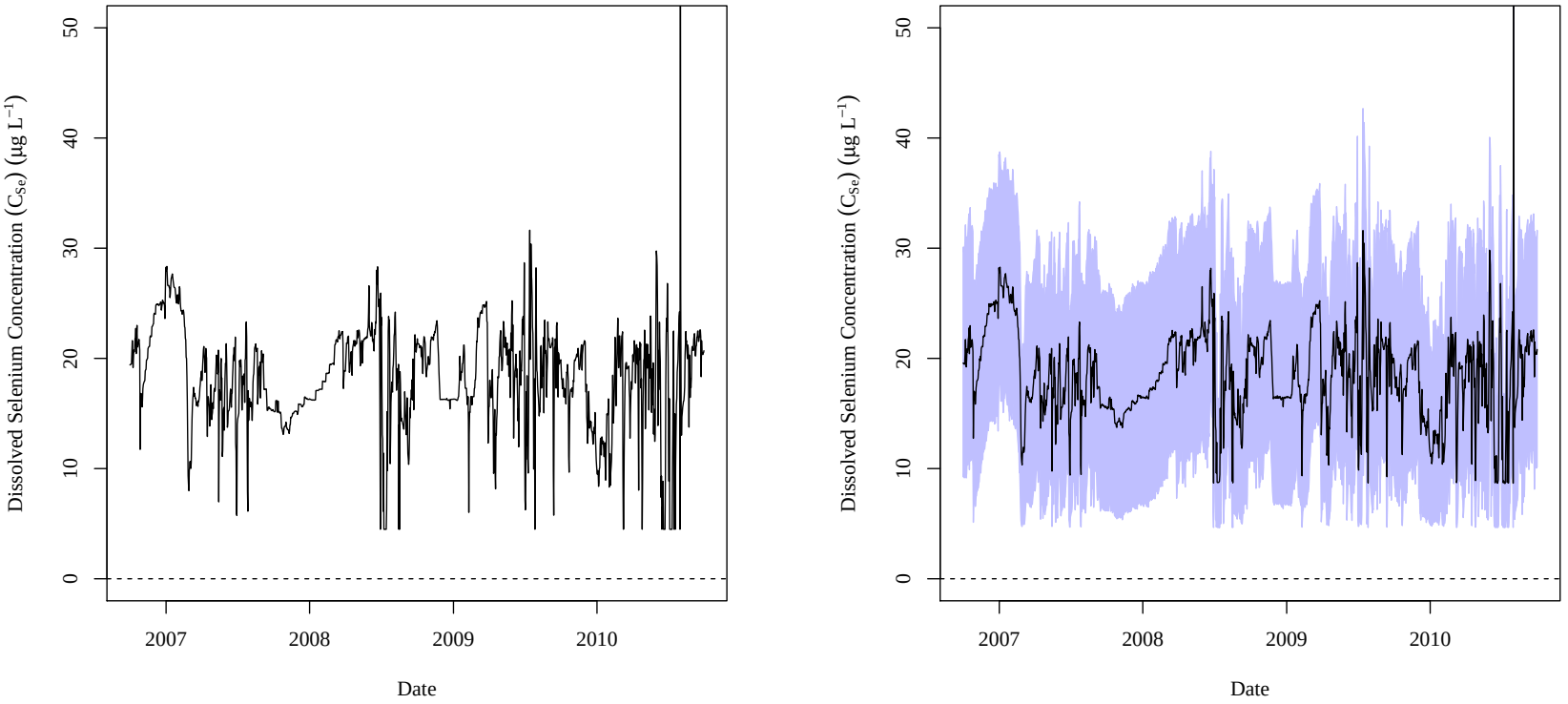
Figure 2.19. DSR Deterministic and stochastic model time series of dissolved selenium concentration. For sub-figure b, the black line is the mean of the realizations and the blue band is the 97.5th CIR.



(a) Deterministic Model.

(b) Stochastic Model.

Figure 2.19 (Cont). DSR Deterministic and stochastic model time series of dissolved selenium concentration.

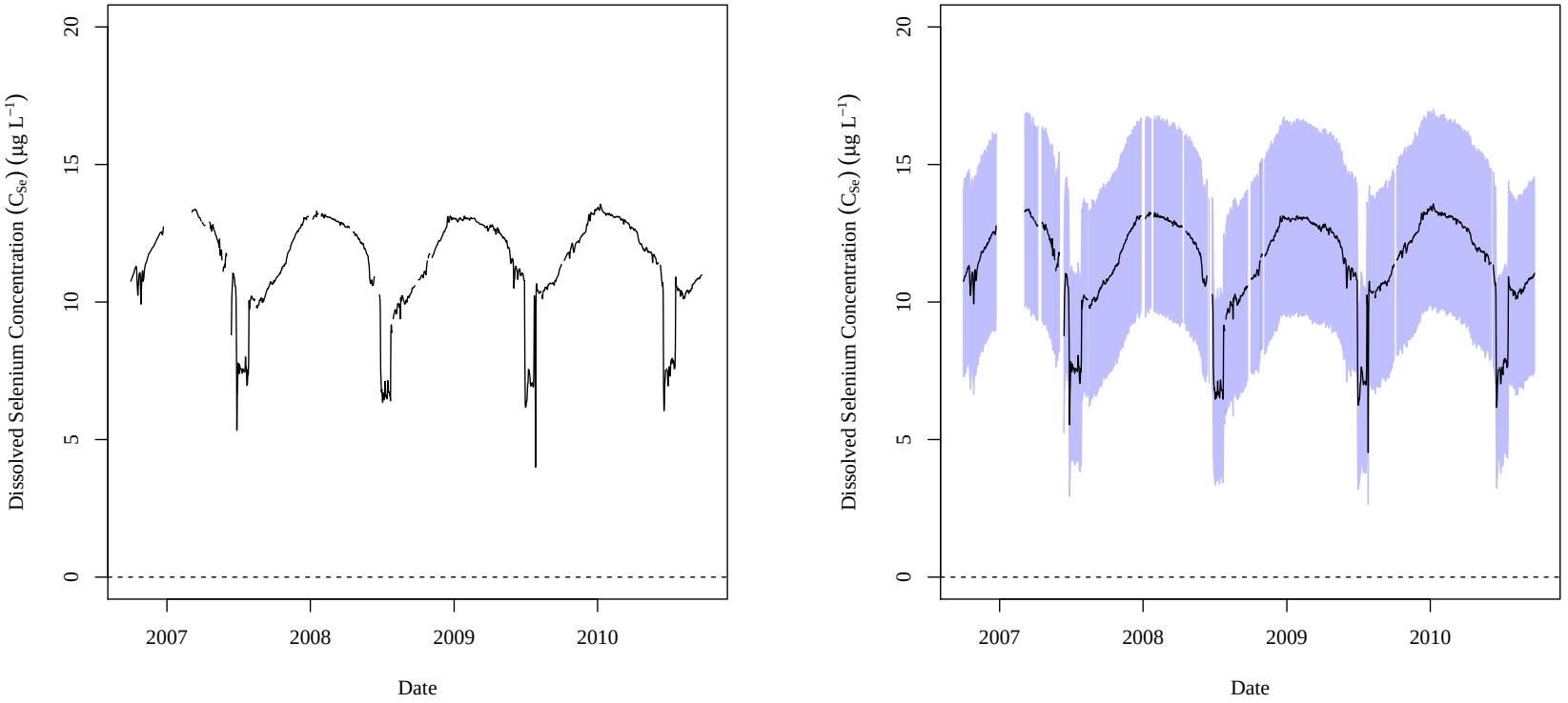


(a) Deterministic Model.

(b) Stochastic Model.

Figure 2.19 (Cont). DSR Deterministic and stochastic model time series of dissolved selenium concentration.

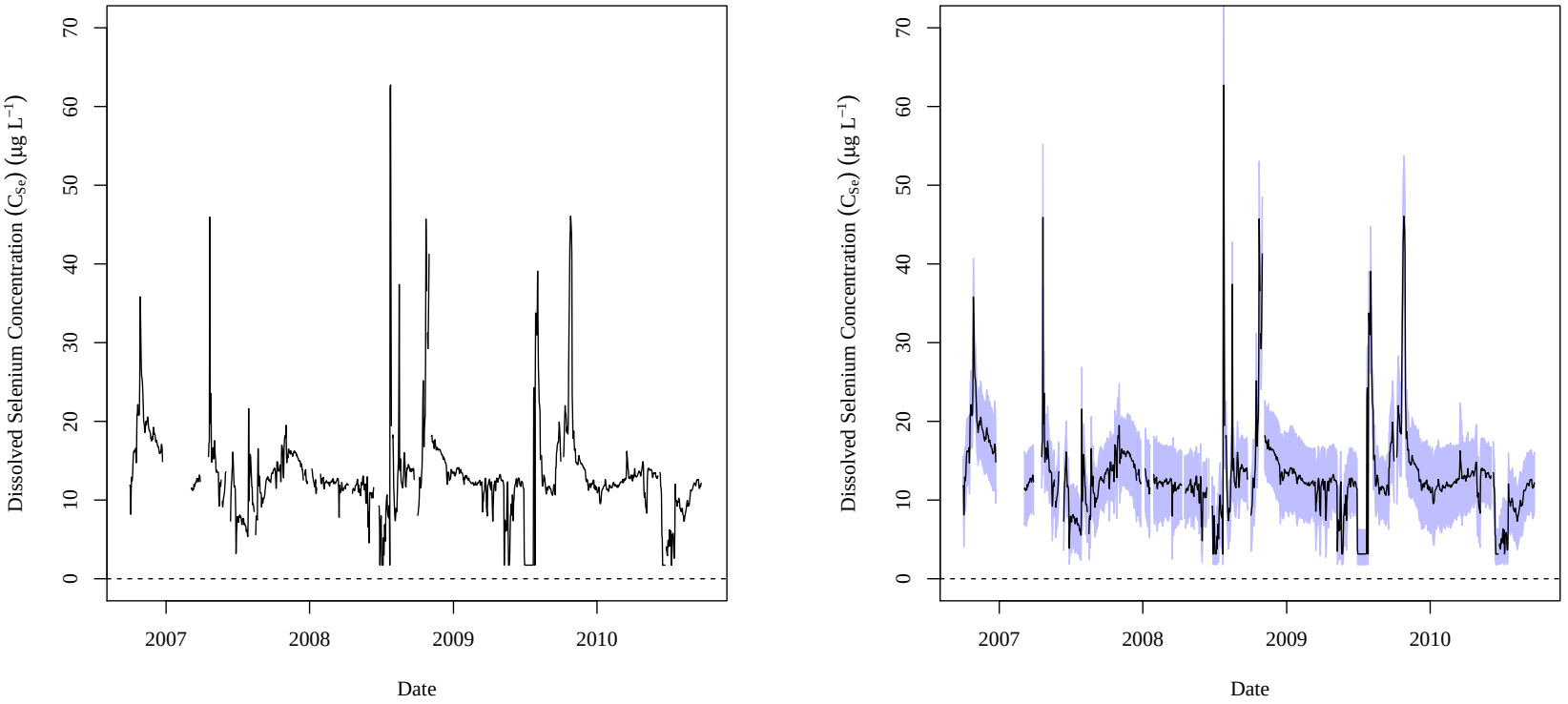
$C_{ARK,d=7.7}$



(a) Deterministic Model.

(b) Stochastic Model.

Figure 2.19 (Cont). DSR Deterministic and stochastic model time series of dissolved selenium concentration.



(a) Deterministic Model.

(b) Stochastic Model.

Figure 2.19 (Cont). DSR Deterministic and stochastic model time series of dissolved selenium concentration.

These figures show a definite cyclical pattern for concentrations within the main stem of the Arkansas R. Although the pattern varies, it is interesting to note that higher concentrations are calculated during the colder months in all cases.

There were not any significant discrepancies between the calculated data and the measured data nor between the deterministic and stochastic models. The results and comparison of the results are presented in tables 2.14 and 2.15 for the USR and DSR, respectively. These tables present the mean, 2.5th, and 97.5th percentile of the deterministic time series results. These tables also present the mean, 2.5th, and 97.5th percentile of the 1-D stochastic mean time series results. The last column provides the percent difference between the two calculated mean values. Again, the stochastic and deterministic calculated dissolved selenium concentration resluts are not significantly different.

Table 2.14. USR dissolved selenium concentration results table. Values are in units of $\mu g\, L^{-1}$.

C_{U163}

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
3.49	10.3	16.3

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	4.53	11.3	17.4
Mean	3.5	10.3	16.3
2.5th Percentile	2.46	9.2	15.2
Pearson Correlation = 1			

C_{U201}

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
3.81	10.9	13.7

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	5.67	12.6	15.4
Mean	3.96	10.9	13.7
2.5th Percentile	2.7	9.25	12.1
Pearson Correlation = 0.9995			

Table 2.14 (Cont). USR dissolved selenium concentration results table.

C_{U74}

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
3.31	11.9	18.6

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	5.15	13.7	20.6
Mean	3.89	11.9	18.6
2.5th Percentile	3.33	10.2	16.9

Pearson Correlation = 0.9997

$C_{ARK,D=85.0}$

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
5.35	11.7	15.9

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	7.39	13.7	18
Mean	5.3	11.6	15.8
2.5th Percentile	3.26	9.49	13.7

Pearson Correlation = 0.9999

Table 2.14 (Cont). USR dissolved selenium concentration results table.

$C_{ARK,D=16.4}$

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
3.83	10.1	14.4

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	5.92	12.2	16.5
Mean	3.83	10.1	14.3
2.5th Percentile	1.98	8	12.2

Pearson Correlation = 0.9999

$C_{ARK,D=47.2}$

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
4.51	10.8	15.1

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	6.56	12.9	17.2
Mean	4.47	10.8	15
2.5th Percentile	2.44	8.66	12.9

Pearson Correlation = 0.9999

Table 2.14 (Cont). USR dissolved selenium concentration results table.

$C_{ARK,D=12.5}$

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
3.75	10.1	14.3

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	5.84	12.1	16.4
Mean	3.74	10	14.2
2.5th Percentile	1.87	7.91	12.1

Pearson Correlation = 0.9999

C_{U207}

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
4.82	15.4	18.8

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	7.58	17.4	20.8
Mean	5.73	15.4	18.7
2.5th Percentile	4.86	13.3	16.7

Pearson Correlation = 0.9995

Table 2.14 (Cont). USR dissolved selenium concentration results table.

C_{U167}

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
3.25	10.4	15.1

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	4.74	11.8	16.4
Mean	3.33	10.5	15.1
2.5th Percentile	2.13	9.19	13.9

Pearson Correlation = 0.9998

C_{U60}

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
3.98	11	21.3

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	7.69	14.3	24.7
Mean	5.16	11.3	21.3
2.5th Percentile	4.02	8.68	17.9

Pearson Correlation = 0.9983

Table 2.14 (Cont). USR dissolved selenium concentration results table.

$C_{LAJWWTP}$

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
16.1	19.7	25

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	24.1	27.9	33.3
Mean	15.8	19.4	24.7
2.5th Percentile	8.09	11.2	16.2

Pearson Correlation = 0.9997

Table 2.15. DSR dissolved selenium concentration results table. Values are in units of $\mu g\, L^{-1}$.

C_{D101C}

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
3.07	10.5	13.6

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	8.53	15.5	18.6
Mean	4.35	10.6	13.6
2.5th Percentile	2.21	6.02	8.69

Pearson Correlation = 0.9971

C_{D106C}

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
10.5	14.6	16.4

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	17.3	21.4	23.3
Mean	10.7	14.7	16.5
2.5th Percentile	4.62	8.09	9.87

Pearson Correlation = 0.9995

Table 2.15 (Cont). DSR dissolved selenium concentration results table.

C_{D23}

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
6.19	18.2	26.4

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	18.1	28.8	36.9
Mean	9.57	18.5	26.4
2.5th Percentile	4.7	8.94	16

Pearson Correlation = 0.9927

$C_{ARK,d=37.7}$

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
6.91	11.4	13.3

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	10.3	14.9	16.7
Mean	6.97	11.5	13.3
2.5th Percentile	3.77	8.05	9.88

Pearson Correlation = 0.9998

Table 2.15 (Cont). DSR dissolved selenium concentration results table.

C_{D57}

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
1.73	13.1	27.2

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	6.27	17.5	33.4
Mean	3.16	13.1	27.2
2.5th Percentile	1.79	8.81	22.1

Pearson Correlation = 0.9991

5.2.2. *Uncertainty of Lab Selenium Concentration Values*

Any sampling methodology is subject to error from a multitude of sources. The additional combined selenium concentration estimating error, ε_2, includes error due to variations in field sampling technique, environmental variations, and lab analysis variations. The samples collected in this study were also subject to an additional unknown error due to environmental conditions during transport from the field to the lab. In some cases, samples reached the lab seven days after being taken in the field. Field technicians took great efforts to keep the samples chilled throughout the field collection process. At the end of a field sampling trip, samples were sent in a chilled insulated container by mail to the lab. The environmental conditions during this transport phase were not and could not be monitored.

Upon receipt at the lab, samples were stored in a refrigerator until they were analyzed. The temperature upon receipt was not recorded by any of the labs. The labs stored the samples in a refrigerator for a maximum of four days. It is not known what, if any, chemistry changes within the samples from the time the samples are taken to the time they are analyzed at the lab. It is also unknown if there is any difference due to minor variations in sampling technique.

Preferably, these error sources could be accounted for on an individual basis. There were a number of factors that determined that this methodology would not work. Error analysis would have to be performed for each field technician. The total project data collection time frame spanned 10 years and included an unknown number of field technicians. The data collection methodology previously described was not entirely adhered for the entire data collection time frame. Not all field technicians recorded what was later considered valuable information such as date and time of sample collection, field technician name, and sampling variances.

The travel distance between the field locations and the lab is also a factor that cannot be overcome. Preferably, the lab would be located fairly close to either the university or the study region. At the start of the sampling time frame, there was no lab in Colorado capable of handling the required analysis with the additional physical, schedule, and fiscal requirements imposed by the project supervisor. The additional distance made determining error due to transport time difficult to determine.

All dissolved selenium samples were treated with nitric acid to stabilize the sample for transport. The stabilization method and acceptable sample storage duration was discussed at length with the lab before any sampling was undertaken. The samples were additionally preserved by storing and transporting them on ice. We were assured by the sampling lab that this additional preservation step would only serve to lengthen the time the sample would be considered viable.

The temperature variations experienced by the samples was not considered a factor during the sample collect time frame. On hind-sight, this could have easily been performed by adding a temperature transponder to the sample container before shipping. This technique might have brought transport temperature control issues to the technician's or the project supervisor's attention if the existed. Unfortunately, this information is not available and we are left to assume that even though significant temperature variations may have existed, those variations did not significantly change the sample chemistry due to the applied sample preservation.

Field and lab blanks were used to determine if the samples were subjected to contamination from the environment or cross contamination from other samples. No lab or field blanks exhibited any evidence of contamination. Since the blanks contained only de-ionized

water, they did not have any chemical or physical markers to show whether they experienced unacceptable environmental conditions.

Lab analysis errors are known to exist and the lab states these ranges. Since the lab was USEPA certified and subscribed to USEPA proficiency testing, we can assume that the lab errors are as stated by the labs. Verification through a different lab was not performed at any time. Although the lab error range was known, it was not known if that error range could be influence by variances in the sampling technique or transport environment.

Combining all individual unaccountable errors into a single error term seemed to be the most pragmatic means to estimate the total error. The only data available to analyze was the set of duplicate samples. As previously discussed at least two samples per sample trip were taken as duplicates. Duplicate samples were taken near the beginning and the end of the sampling trip. Only the 'A' sample was used for concentration estimation. The 'B' samples were taken to monitor for equipment malfunction, significant deviation in sampling methodology, and significant lab error. The 'A' and 'B' samples were taken using the same equipment and transported in the same container from the sample location to the lab. Since they experienced the same environmental conditions, it was unreasonable to assume that this method could be used to estimate the error due to extreme transport environmental conditions. The samples were well preserved and it was assumed that temperature variations did not significantly affect the samples.

Lab results for the 'A' and 'B' samples from both the USR and DSR were complied into a single data set. Date, location, and all other identifying markers were removed from the data set to reduce potential bias due to prior knowledge of the individual samples. 'A' samples were assumed to be the expected value for the following samples.

Figure 2.20 shows the comparisons analyzed. The top graph plots the 'A' and 'B' sample concentrations against the difference from the mean of the 'A' and 'B' samples. The bottom graph plots the same data, but with respect to the percent difference from the mean of the 'A' and 'B' samples.

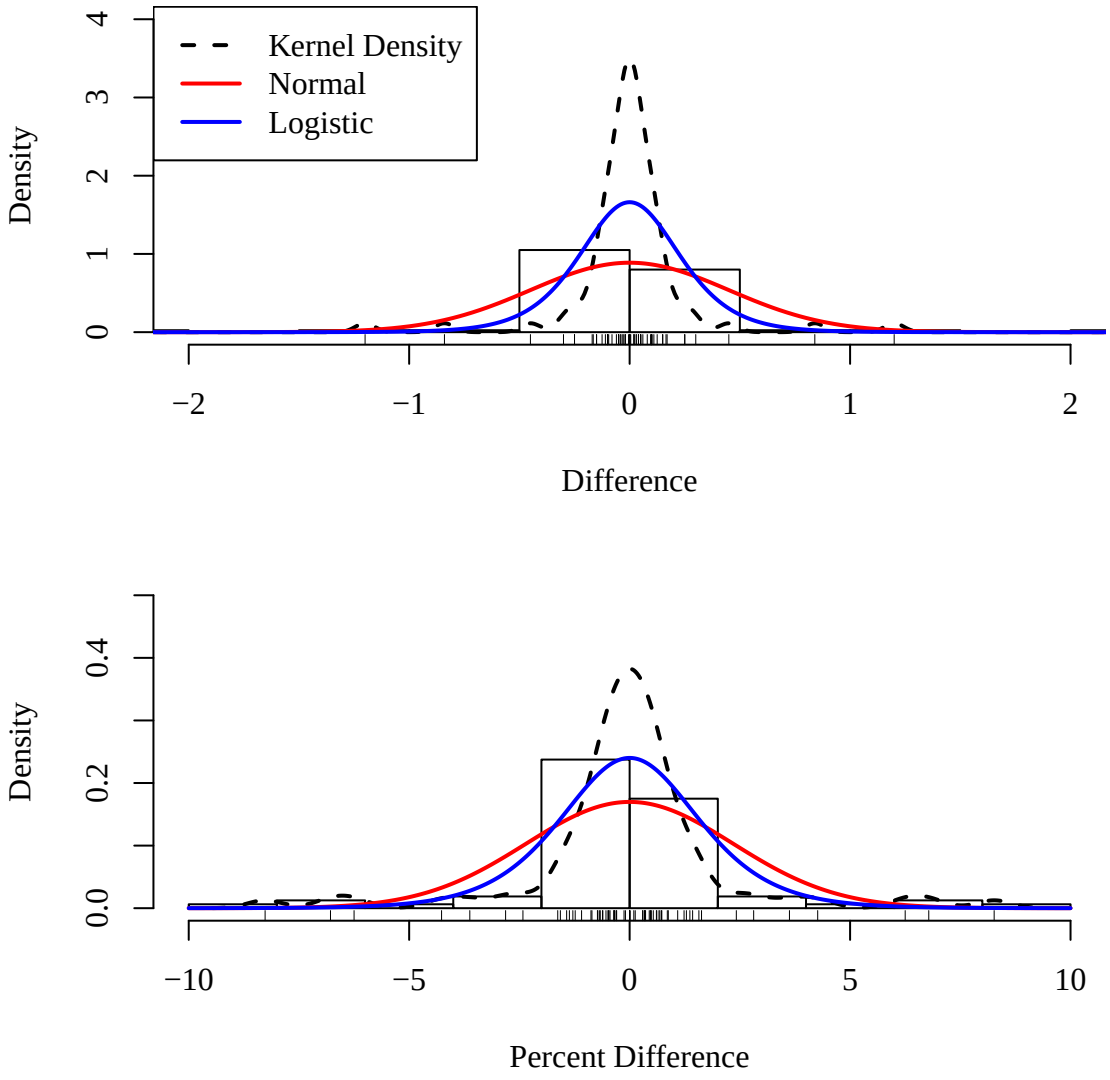


Figure 2.20. Dissolved Selenium Concentration Uncertainty. Each sub-figure shows the comparison of the lab reported dissolved selenium concentration and the mean for each pair of duplicate samples.

Figure 2.21 shows the 'A' and 'B' samples plotted against the difference and percent difference. This analysis was performed to determine if there was a correlation between the magnitude of the concentration and the magnitude of the difference. No such correlation was found.

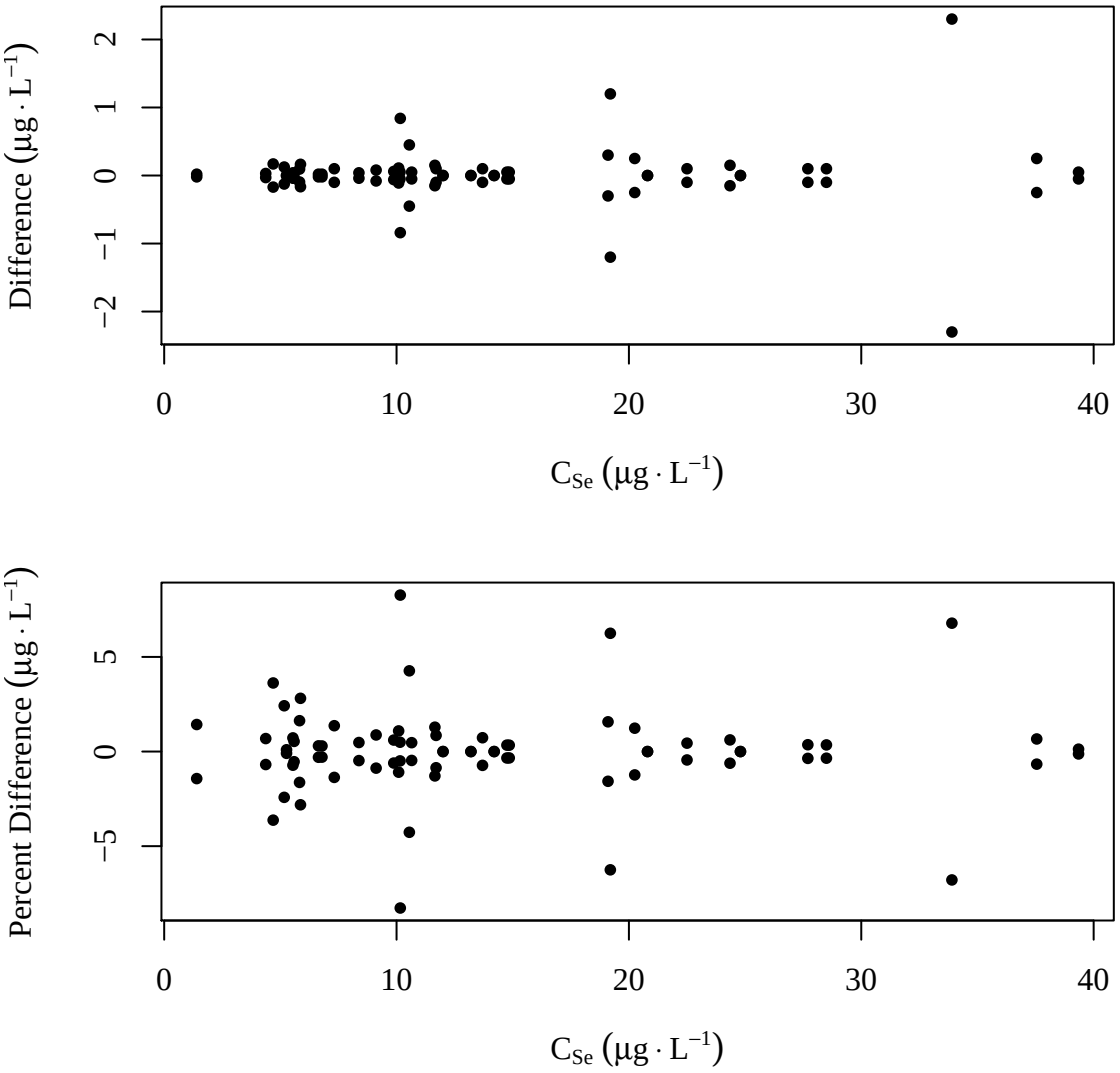


Figure 2.21. Dissolved Selenium Concentration Variation.

The percent difference between the reported lab values and the mean duplicates was best fit with a logistic distribution. The logistic distribution had a location parameter of

Table 2.16 (Cont). Selenium combined error analysis goodness-of-fit test results.

Calculation	Distribution	Fit Test Result		
		K-S	CvM	A-D
Difference	normal	0.2876	2.598	13.48
	logistic	0.1644	0.8538	5.225
Percent Difference	normal	0.173	1.150	6.365
	logistic	0.1195	0.4139	2.729

-0.067 and a scale parameter of 1.807. This comparison and distribution combination was chosen because it had the best fit when compared to others using Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling goodness-of-fit test statistics. The duplicate samples had a mean difference from the mean of their respective 'A' and 'B' samples of 0% with a standard deviation of 4.16%. This corresponds to 95% of the data within ±10% of the reported value

Calculated dissolved selenium concentrations at specific sites were constrained to fit between one-half of the minimum historically reported value and 1-1/2 of the maximum historically reported value. This range should allow for dissolved selenium concentration variations that are comparable to the values reported from the field samples. The range allowed for variation beyond the reported concentration range to allow for the possibility that concentrations beyond the range may possibly have existed at some time.

Visual analysis seems to indicate that using the percent difference distribution would lead to the best characterization of the selenium sample errors. This hypothesis was tested by using Kolmogorov-Smirnov, Cramer von Mises, and Anderson-Darling goodness-of-fit tests. Results from these tests are presented in Table 2.16 (Cont). The logistic distribution of the percent difference calculation was shown to have the best fit calculation.

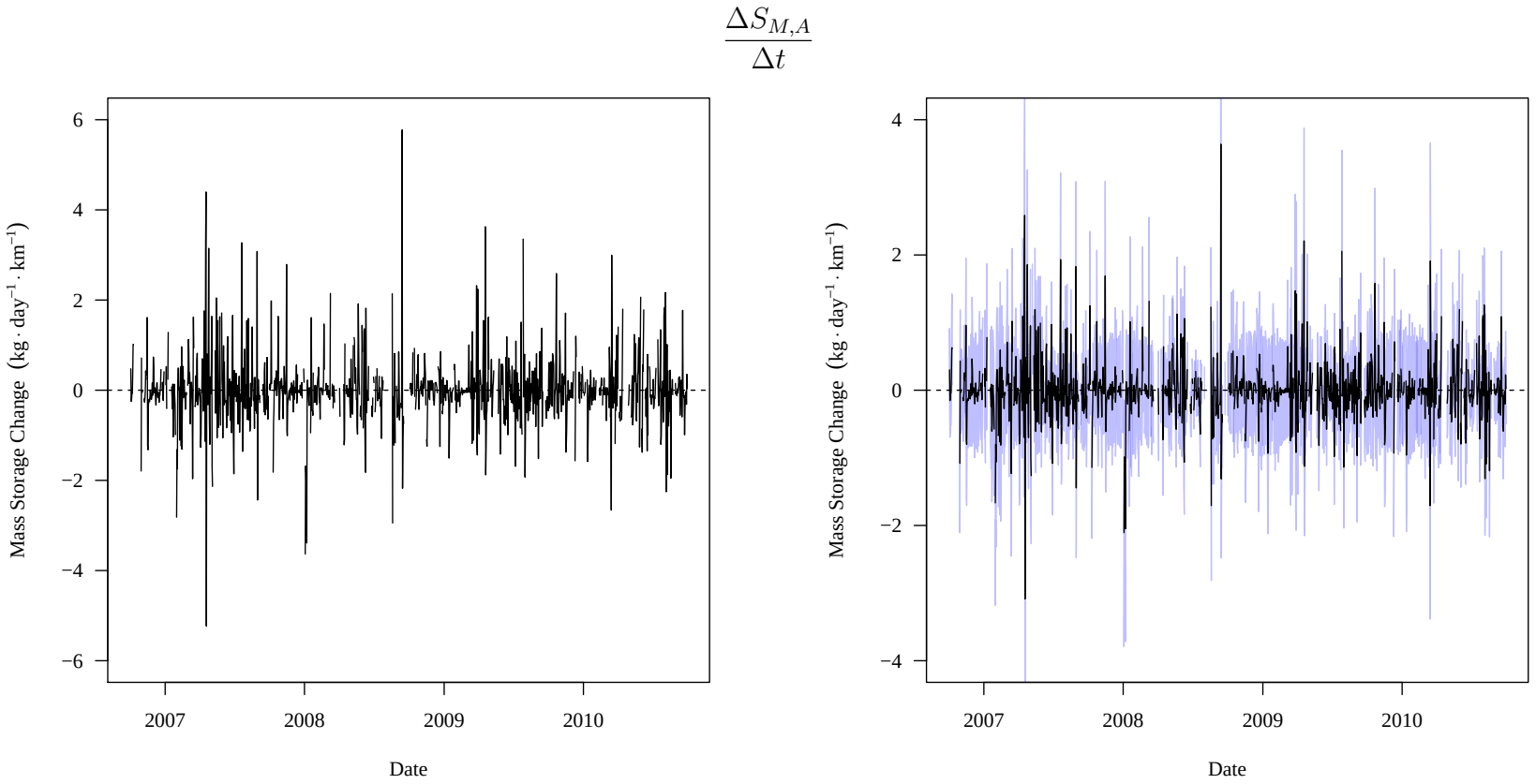
Given these results, the total field sampling and lab error distribution is described by a logistic distribution with a location parameter of -0.06693 and a scale parameter of 1.807.

The combined field sampling and lab error is bounded such that 95% of the distribution lies in the range of approximately ±6.6% of the expected value.

The combined field sampling and lab error is calculated independently from the selenium concentration estimation error previously described. The estimated selenium concentration, without the estimation estimation, is taken as the expected value for the combined field sampling and lab error.

5.2.3. *Mass Storage Change Results*

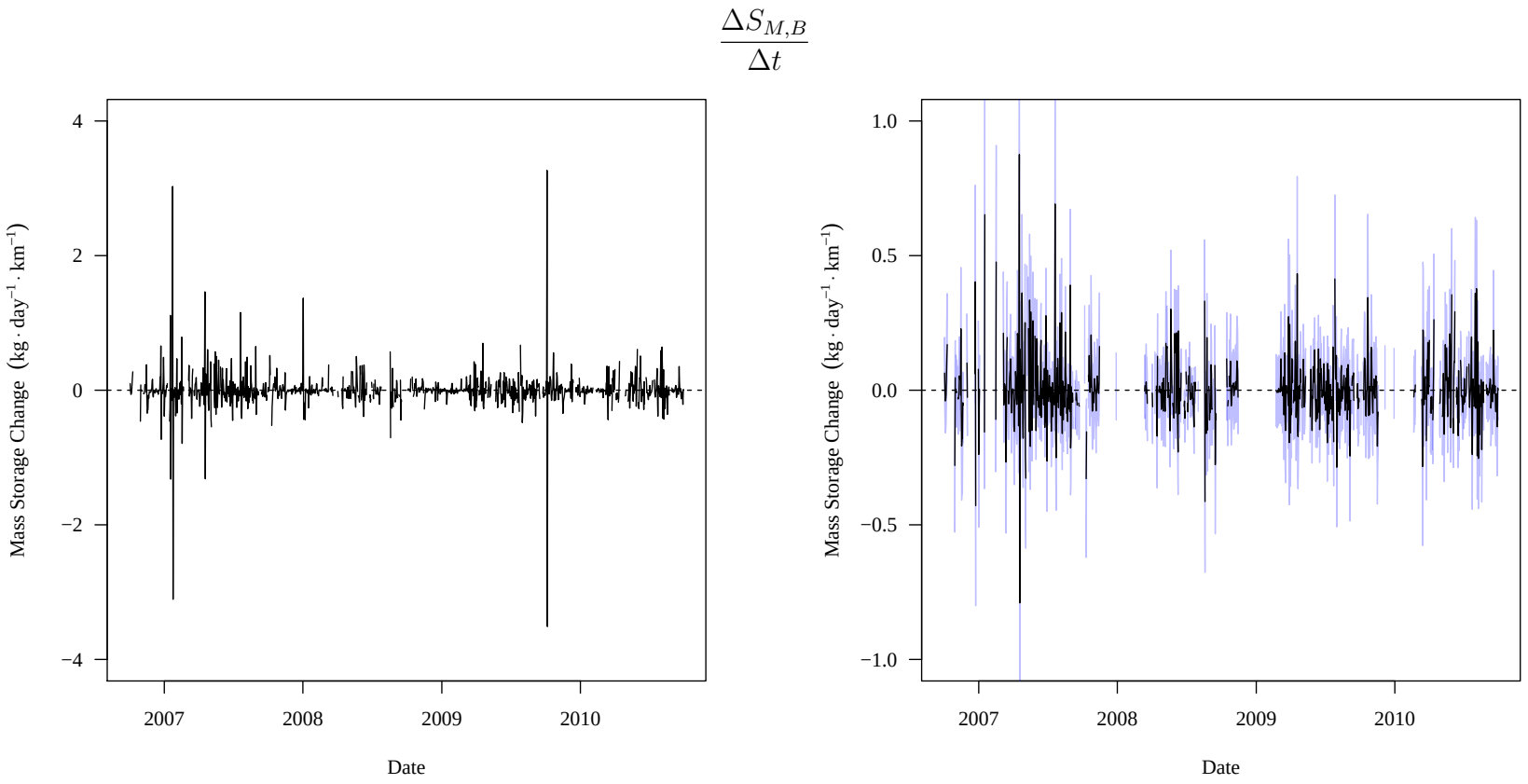
Figures 2.22 and 2.23 show the deterministic and stochastic time series of the mass storage change within each of the river segments in the USR and DSR, respectively. The black line in the stochastic model sub-figure is the mean of the realizations and the blue band is the 97.5th CIR. Results in these figures and tables 2.17 and 2.18 are presented in units of $\mathrm{kg\,d^{-1}\,km^{-1}}$. Standardizing values to mass storage per unit length allows for comparison between all river segments in both study reaches. This also allows for comparison between the mass storage change components and the mass transport components of the mass balance models.



(a) Deterministic Model.

(b) Stochastic Model.

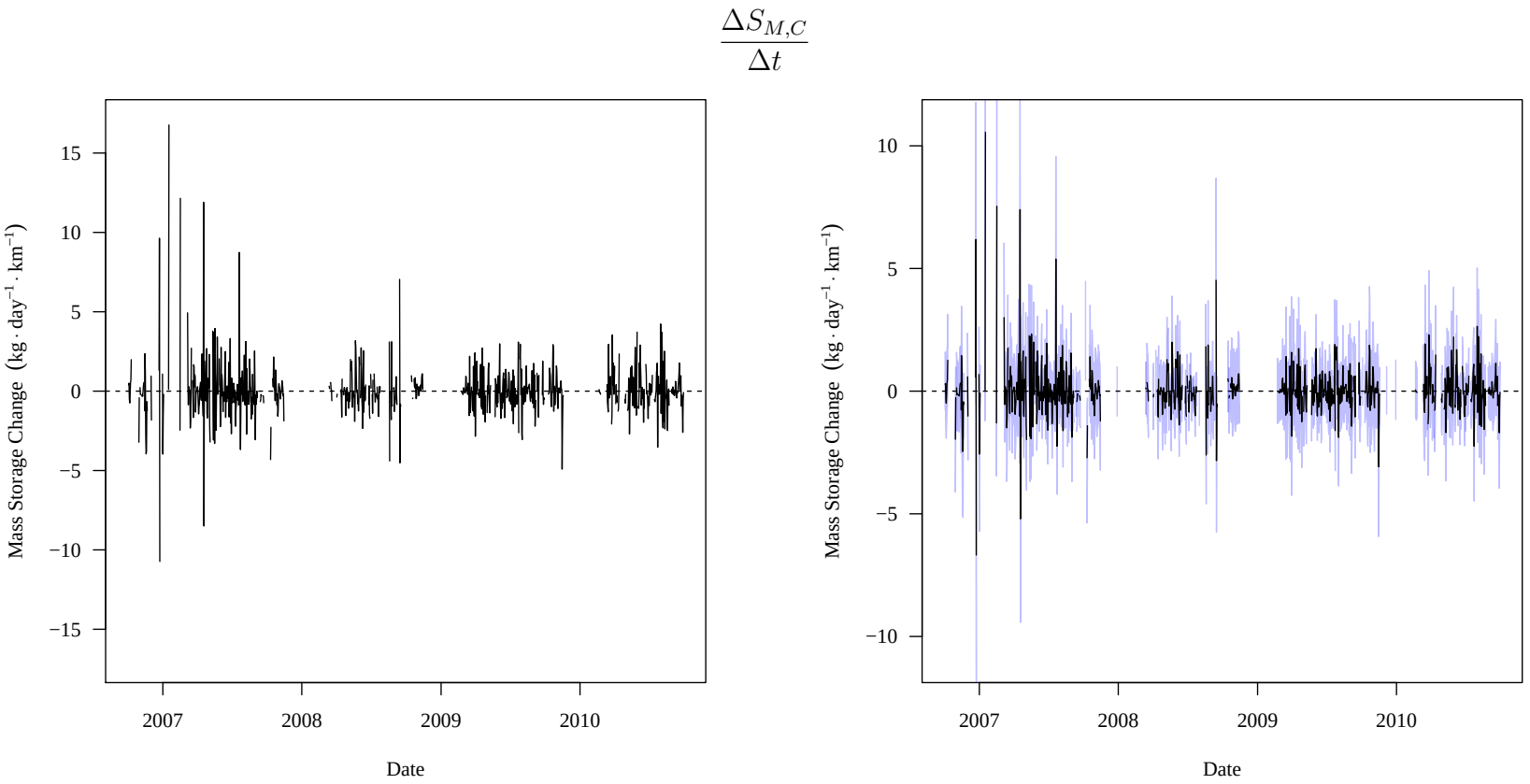
Figure 2.22. USR river segment deterministic and stochastic dissolved selenium mass storage change time series. For the stochastic model, the line is the mean of the realizations and the blue band is the 97.5th CIR.



(a) Deterministic Model.

(b) Stochastic Model.

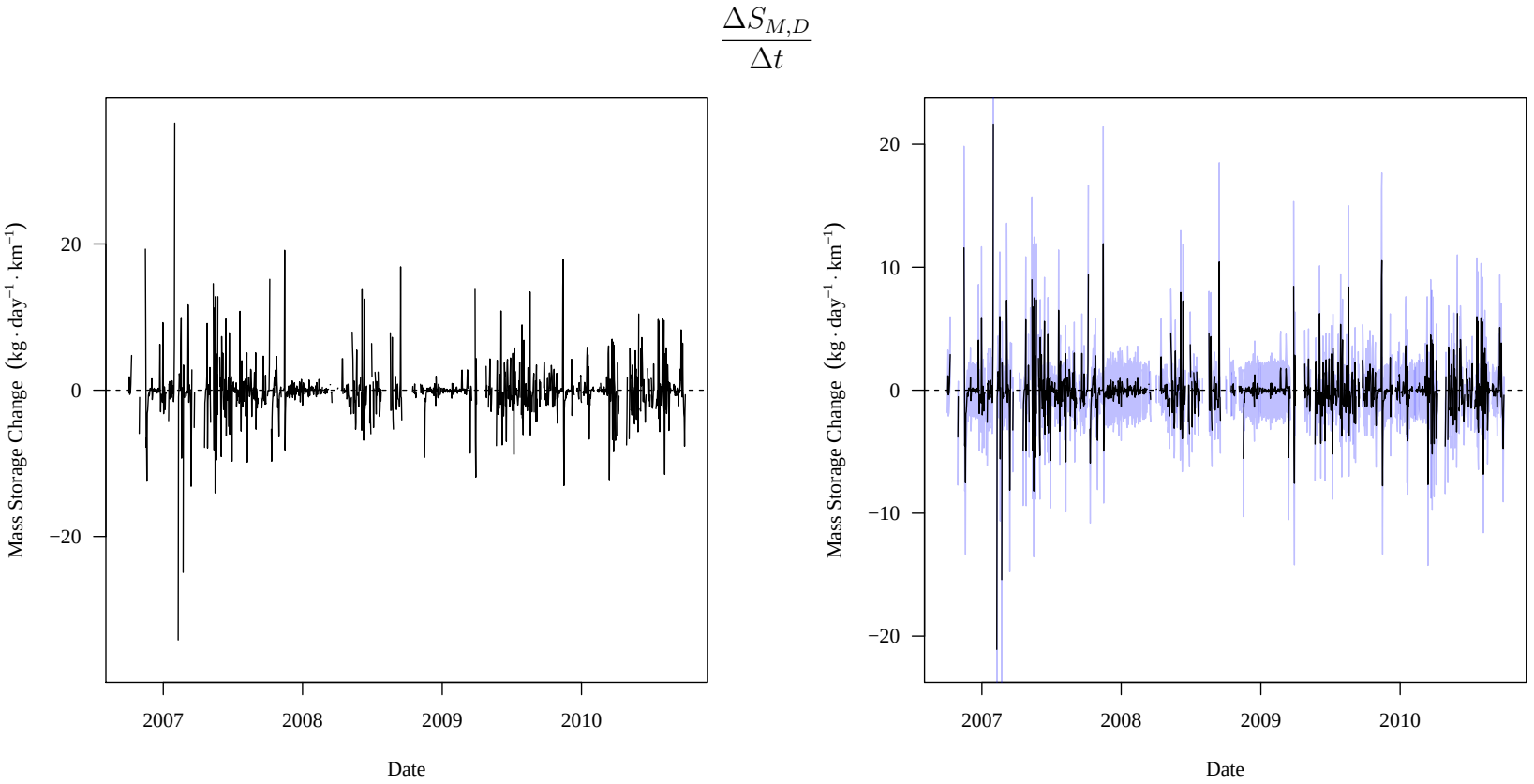
Figure 2.22 (Cont). USR river segment deterministic and stochastic dissolved selenium mass storage change time series.



(a) Deterministic Model.

(b) Stochastic Model.

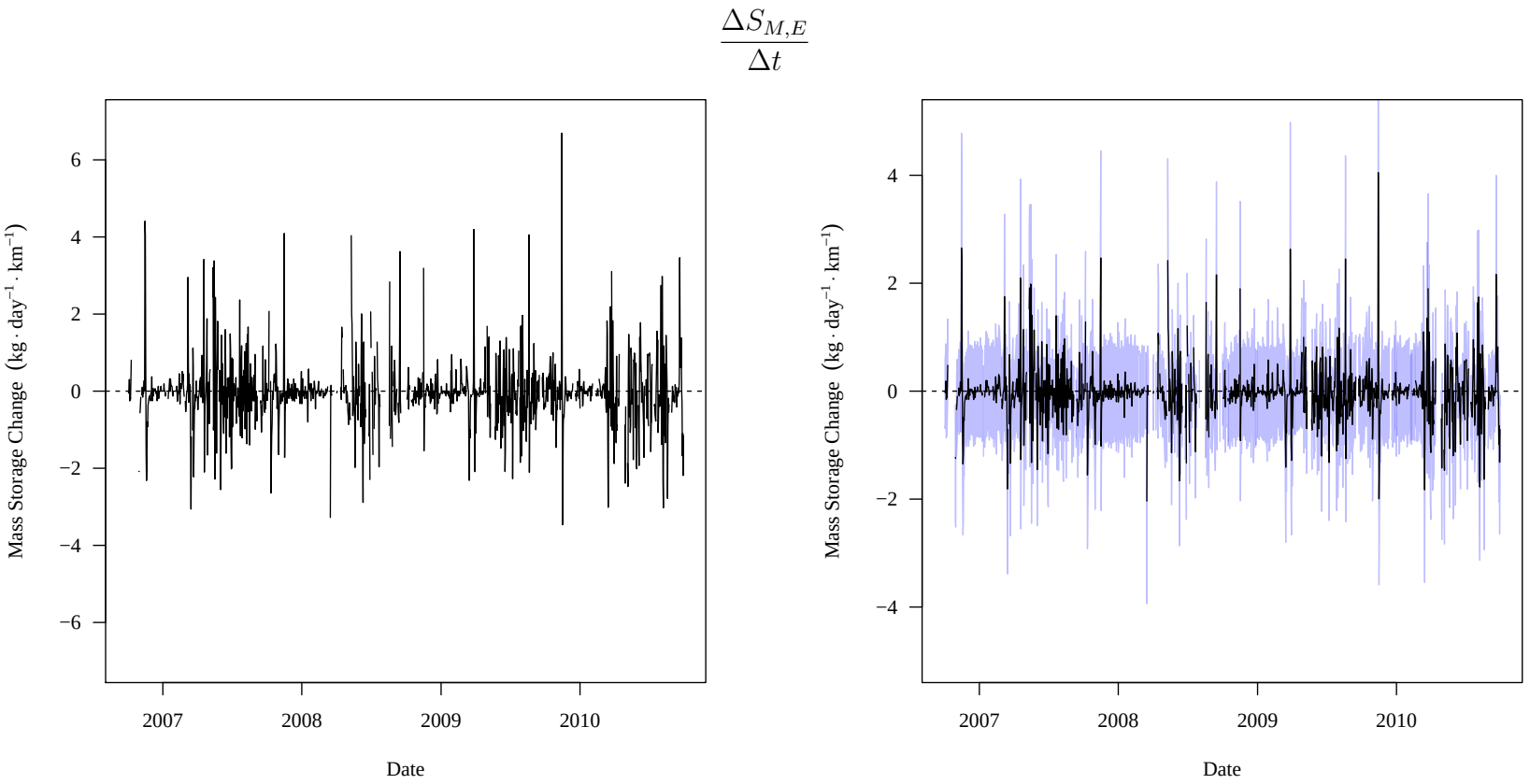
Figure 2.22 (Cont). USR river segment deterministic and stochastic dissolved selenium mass storage change time series.



(a) Deterministic Model.

(b) Stochastic Model.

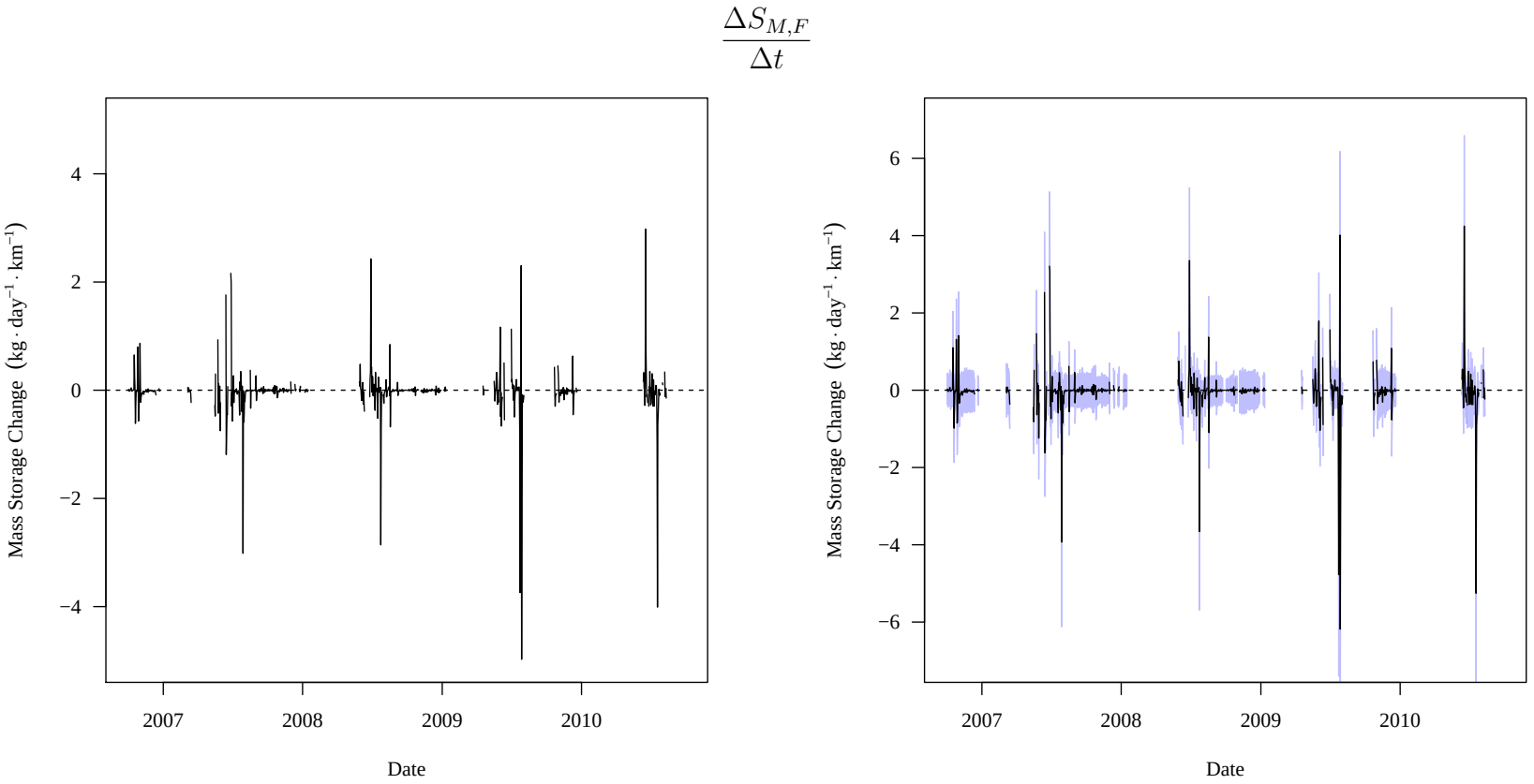
Figure 2.22 (Cont). USR river segment deterministic and stochastic dissolved selenium mass storage change time series.



(a) Deterministic Model.

(b) Stochastic Model.

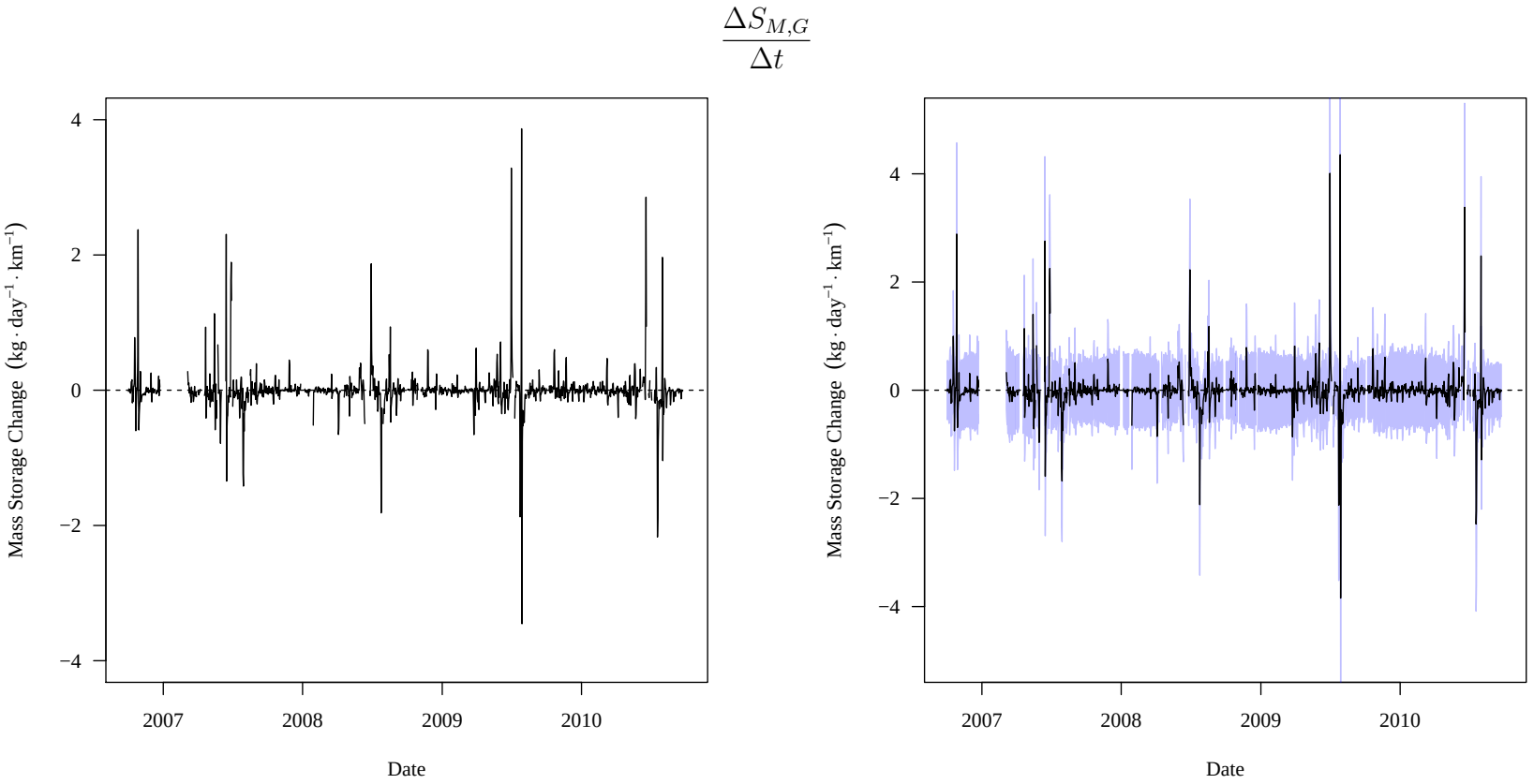
Figure 2.22 (Cont). USR river segment deterministic and stochastic dissolved selenium mass storage change time series.



(a) Deterministic Model.

(b) Stochastic Model.

Figure 2.23. USR river segment deterministic and stochastic dissolved selenium mass storage change time series. For the stochastic model, the line is the mean of the realizations and the blue band is the 97.5th CIR.



(a) Deterministic Model.

(b) Stochastic Model.

Figure 2.23 (Cont). USR river segment deterministic and stochastic dissolved selenium mass storage change time series.

Comparing the sub-figures provides for a visual goodness-of-fit analysis between the deterministic and stochastic models. All USR river segment stochastic models agree with the deterministic models. It should be noted that there is quite a fair amount of uncertainty associated with the values calculated for the stochastic model. This is the compounding of uncertainties from the multiple input variables. This is as expected for a complex multi-variate model.

The calculated selenium storage change values for each reach were compared between the deterministic and stochastic models and are reported in tables 2.17 and 2.18 for the USR and DSR, respectively. All selenium storage change values in the figures and tables are presented in units of $\mathrm{kg\,d^{-1}\,km^{-1}}$. This table is presented in a similar fashion to other comparison tables in this chapter.

These tables show that the deviation between the deterministic model time-series and the stochastic model mean time-series is low, but higher than the values calculated for the individual input variables. This is most likely indicative of the compounding of input variable uncertainties. The percent deviation between the models for segment B in the USR is deceptive. Both the deterministic model time-series mean value and the stochastic model mean time-series mean value are near zero. This causes any deviation to appear large.

Table 2.17. USR river segment deterministic and stochastic dissolved selenium mass storage change results tables. Values are in units of $\mathrm{kg\,d^{-1}\,km^{-1}}$.

$$\frac{\Delta S_{M,A}}{\Delta t}$$

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
-1.44 (-3.17)	-0.0094 (-0.0207)	1.66 (3.66)

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	-0.173 (-0.381)	0.576 (1.27)	1.97 (4.34)
Mean	-0.858 (-1.89)	-0.00545 (-0.012)	1.01 (2.23)
2.5th Percentile	-1.84 (-4.06)	-0.596 (-1.31)	0.229 (0.505)
Pearson Correlation = 0.9994			

$$\frac{\Delta S_{M,B}}{\Delta t}$$

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
-0.344 (-0.758)	-3.08e-05 (-6.79e-05)	0.423 (0.933)

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	-0.0492 (-0.108)	0.115 (0.254)	0.489 (1.08)
Mean	-0.214 (-0.472)	0.00229 (0.00505)	0.272 (0.6)
2.5th Percentile	-0.408 (-0.899)	-0.113 (-0.249)	0.0617 (0.136)
Pearson Correlation = 0.9994			

Table 2.17 (Cont). USR river segment deterministic and stochastic dissolved selenium mass storage change results tables.

$$\frac{\Delta S_{M,C}}{\Delta t}$$

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
-2.7 (-5.95)	0.0261 (0.0575)	3.1 (6.83)

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	-0.35 (-0.772)	1.18 (2.6)	3.83 (8.44)
Mean	-1.64 (-3.62)	0.0155 (0.0342)	1.88 (4.14)
2.5th Percentile	-3.64 (-8.02)	-1.17 (-2.58)	0.435 (0.959)

Pearson Correlation = 0.9993

$$\frac{\Delta S_{M,D}}{\Delta t}$$

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
-7 (-15.4)	-0.0873 (-0.192)	8.82 (19.4)

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	-1.02 (-2.25)	1.93 (4.25)	9.18 (20.2)
Mean	-4.34 (-9.57)	-0.0546 (-0.12)	5.31 (11.7)
2.5th Percentile	-8.31 (-18.3)	-2.11 (-4.65)	1.28 (2.82)

Pearson Correlation = 0.9994

Table 2.17 (Cont). USR river segment deterministic and stochastic dissolved selenium mass storage change results tables.

$$\frac{\Delta S_{M,E}}{\Delta t}$$

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
-1.96 (-4.32)	-0.0439 (-0.0968)	1.83 (4.03)

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	-0.248 (-0.547)	0.695 (1.53)	2.22 (4.89)
Mean	-1.14 (-2.51)	-0.0252 (-0.0556)	1.13 (2.49)
2.5th Percentile	-2.21 (-4.87)	-0.768 (-1.69)	0.245 (0.54)
Pearson Correlation = 0.9996			

Table 2.18. DSR river segment deterministic and stochastic dissolved selenium mass storage change results tables. Values are in units of $\mathrm{kg\,d^{-1}\,km^{-1}}$.

$$\frac{\Delta S_{M,F}}{\Delta t}$$

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
-0.593 (-1.31)	-0.0288 (-0.0635)	0.491 (1.08)

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	-0.324 (-0.714)	0.453 (0.999)	1.6 (3.53)
Mean	-0.894 (-1.97)	-0.0356 (-0.0785)	0.786 (1.73)
2.5th Percentile	-1.7 (-3.75)	-0.535 (-1.18)	0.262 (0.578)

Pearson Correlation = 0.9947

$$\frac{\Delta S_{M,G}}{\Delta t}$$

Deterministic Model Time Series

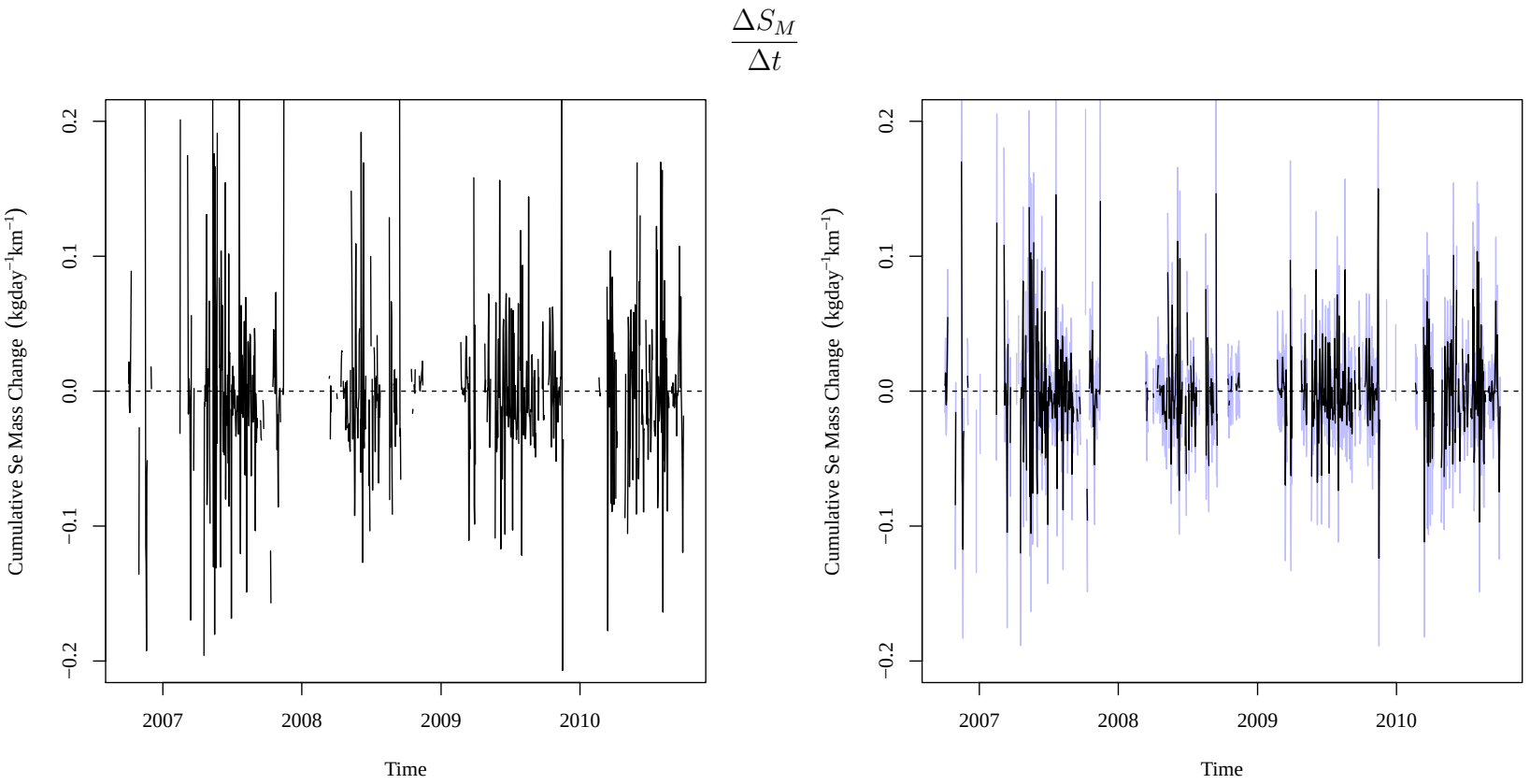
2.5th Percentile	Mean	97.5th Percentile
-0.415 (-0.915)	-0.00606 (-0.0134)	0.486 (1.07)

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	-0.00124 (-0.00273)	0.593 (1.31)	1.33 (2.93)
Mean	-0.518 (-1.14)	-0.00729 (-0.0161)	0.609 (1.34)
2.5th Percentile	-1.25 (-2.76)	-0.613 (-1.35)	0.0606 (0.134)

Pearson Correlation = 0.9986

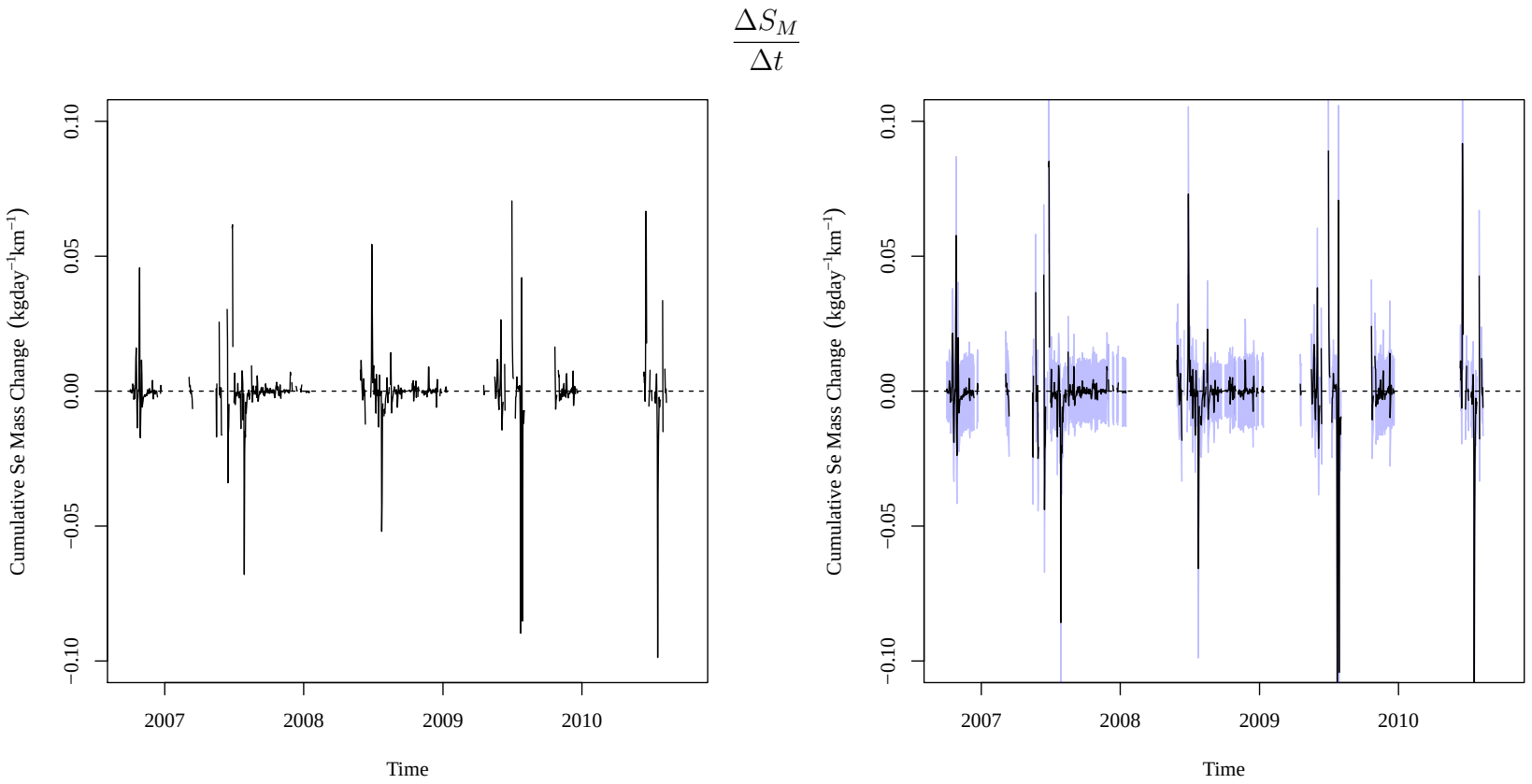
Figures 2.24 and 2.25 are the time-series of the deterministic and stochastic river reach selenium mass storage changes for the USR and DSR, respectively. The black line in the stochastic sub-figure is the mean for each time step. The blue band encompasses the 95% CIR for each time step. Values in these figures and associated tables are in units of $\mathrm{kg\,d^{-1}\,km^{-1}}$. Standardizing the results allows for comparisons to be made between the two study reaches and between mass balance model components. Positive values indicate that the reach gained selenium during the given time step.



(a) Deterministic Model.

(b) Stochastic Model.

Figure 2.24. USR river reach deterministic and stochastic dissolved selenium mass storage change time series. For the stochastic model, the line is the mean of the realizations and the blue band is the 97.5th CIR.



(a) Deterministic Model.

(b) Stochastic Model.

Figure 2.25. DSR river reach deterministic and stochastic dissolved selenium mass storage change time series. For the stochastic model, the line is the mean of the realizations and the blue band is the 97.5th CIR.

Figures 2.24 and 2.25 show that there is a definite seasonal variation in the selenium storage component. This temporal relationship follows the same pattern identified with the water balance model storage component. There is a very strong visual relationship between the water balance model flow component and the mass balance model selenium transport component. This is to be expected since the water balance storage component is the prime contributor to the mass balance selenium storage component. These figures also show that uncertainty with the selenium storage component is very large. This is to be expected since the mass balance models contain all of the uncertainty from the water balance model and the selenium concentration estimation linear models.

The distribution of all realizations within each time step was analyzed to determine a distribution type. This analysis was performed to determine if the assumption that the deterministic model results were representative of the stochastic model. Testing was performed by comparing K-S statistics for the best fit normal, log-normal, logistic, exponential, gamma, and Weibull distributions. In the USR In the DSR . This indicates that for both the USR and DSR, the distributions across the realizations are normal with some slight skewness.

do calcs

do calcs

Tables 2.19 and 2.20 present the summary statistics for the deterministic model time-series and the three stochastic model summary statistics time-series for the USR and DSR, respectively. All values in these tables are in units of $\mathrm{kg\,d^{-1}\,km^{-1}}$

The mean of the percent difference between the deterministic and 1-D stochastic mean models is low, but not insignificant. This indicates that the deterministic model is fairly representative of the stochastic model expected value. The high percent differences at the 2.5th and 97.5th percentile indicate that there is still a large range of uncertainty

Table 2.19. USR river reach deterministic and stochastic dissolved selenium mass storage change time series summary statistics. Values are in units of $\mathrm{kg\,d^{-1}\,km^{-1}}$.

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
-0.12 (-0.265)	-0.00131 (-0.00289)	0.16 (0.353)

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	-0.0326 (-0.0719)	0.0238 (0.0525)	0.151 (0.333)
Mean	-0.073 (-0.161)	-0.000772 (-0.0017)	0.0961 (0.212)
2.5th Percentile	-0.115 (-0.254)	-0.0258 (-0.0569)	0.041 (0.0904)
Pearson Correlation = 0.9995			

Table 2.20. DSR river reach deterministic and stochastic dissolved selenium mass storage change time series summary statistics. Values are in units of $\mathrm{kg\,d^{-1}\,km^{-1}}$.

Deterministic Model Time Series

2.5th Percentile	Mean	97.5th Percentile
-0.0168 (-0.037)	-0.000302 (-0.000666)	0.0178 (0.0392)

Stochastic Model Summary Statistics Time Series

Time Series	2.5th Percentile	Mean	97.5th Percentile
97.5th Percentile	-0.00843 (-0.0186)	0.0121 (0.0267)	0.0415 (0.0915)
Mean	-0.0227 (-0.05)	-0.000346 (-0.000763)	0.0237 (0.0522)
2.5th Percentile	-0.039 (-0.086)	-0.0129 (-0.0284)	0.00974 (0.0215)
Pearson Correlation = 0.9964			

contained within the stochastic model that the deterministic model cannot replicate. The deterministic model can be used to determine how changes can affect a reach over a span of

time, but using it to estimate values for specific time steps is unwise as the differences noted at individual time steps is too large to account for.

5.3. Mass Transport in Gauged Streams and Diverted Canals

The combined river section selenium surface transport rate ($L_{Surface}$) is calculated as shown in Equation 8 which contains all portions of the mass balance model that transport selenium in and out of the river channel through a flow gauge. This equation is a part of Equation 4. Selenium surface transport rates for individual stream gauges are calculated as positive values regardless of whether they discharge to or receive water from the main stem of the river.

$$\sum L_{Surface} = \left(L_{out,DS} + \sum L_{out} - L_{in,US} - \sum L_{in} \right) \tag{8}$$

$\sum L_{Surface} =$ Sum of all loadings transported in gauged surface waters.

$L_{out,DS} =$ Loading being discharged at the downstream end of the reach.

$\sum L_{out} =$ Sum of oadings being discharged through the reach irrigation canals.

$L_{in,US} =$ Loading being received at the upstream end of the reach.

$\sum L_{in} =$ Sum of loadings being received through the reach tributaries.

References

Johnson, Richard A. and Dean W Wichern (2007). *Applied Multivariate Statistical Analysis*. 6th. Pearson.