# JOHN DAY RIVER BASIN TMDL

# APPENDIX D:
# DISSOLVED OXYGEN ASSESSMENT

**Final**
**November 2010**

DEQ

State of Oregon
Department of Environmental Quality

**Prepared by:**

**Kevin Brannan, Oregon Department of Environmental Quality**

# Table of Contents



# Figures



# Tables

# *1. OVERVIEW AND SCOPE*

The water quality parameter controlling DO in the John Day River Basin was determined to be stream temperature (TEMP). This result applies only for the cool-water criterion of 6.5 mg/l during the non-spawning season. DO is not a pollutant itself: it represents a condition that must be maintained. Pollutants, along with other water column conditions, affect the DO levels in a stream. Regression analysis was performed among DO and other water quality parameters to determine which water quality parameters have the greatest influence on DO during time of year that digressions of the cool-water criteria occur. The equations from the regression analysis were used to determine the DO TMDL for the John Day River.

## 1.1 Impaired Segments for DO Water Quality Standard

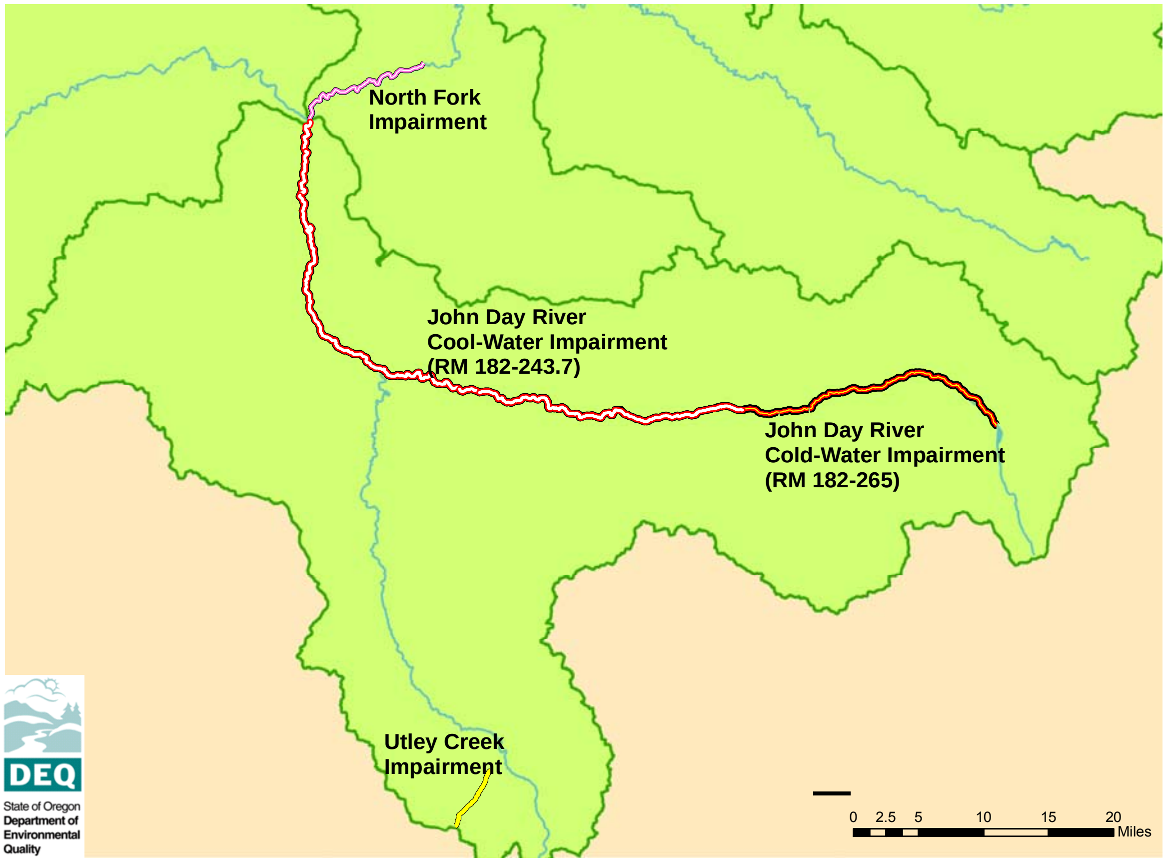
In the 2004-2006 303(d) report, there are four impaired segments for violations of the DO standard identified in the John Day River Basin. The impairments and associated information are presented in Table D-1. The Department questions the designation of three of the listed segments in the upper portions of the John Day River Basin: one segment on the John Day River above the confluence with the North Fork John Day River, one segment on a tributary stream (Utley Creek), and one segment on the North Fork (**Figure D-1**). The two segments covering river miles 182 to 265 on the John Day River have two different beneficial use designations. The John Day River listing from 1998 (**Table D-1**) for river miles 182 to 265 was for a station identified by a STORET number (404158), which corresponds to LASAR station 11479. The ‘cold water’ listing from 1998 for river miles 182 to 265 was carried over to the 2004-2006 303(d) report (Personal Communication, Karla Urbanowicz, Oct. 2008). For the 2004-2006 303(d) report, the use designation of this segment was split into river miles 182 to 243.7 designated as cool water and the river miles 243.8 to 265 being cold water. The designated use from the 2004-2006 303(d) report is used in this TMDL and is the current length and fish use designation for this segment. The Department considers the earlier listing no longer valid. The upper section of the 1998 listing (river mile 243.8 to 265) is a decisional artifact of locating the upper boundary of the segment. The current preference for cool water designation is based on discussions with the ODFW District Fish Biologist indicating an historic cool water aquatic community in the reach (Tim Unterwegner, personal communication with Don Butcher, DEQ) and discussions with EPA regarding delineation of cool/cold water criteria applicability (DEQ Memo to Randy Smith, EPA Region 10, February 2004). The North Fork and Utley Creek 303(d) listings are discussed in the next section.

**Table D-1. DO water quality standard impaired segments from the 2004/2006 303(d) report for the John Day River Basin**

| Name | River Mile | LLID | Criteria | Status | Supporting Date |
|---|---|---|---|---|---|
| John Day River | 182 to 265 | 1206499457318 | Cold Water: Not less than 8.0 mg/l or 90% saturation | 303(d)/1998 | DEQ Data (Site 404158; RM 215.4)[1]; 32% (9 or 28( May – August values exceeded dissolved oxygen standard (8.0 mg/l or 90% saturation) with a minimum of 3.9 mg/l between WY 1986 – 1995 (cold water rearing approximately May – August). |
| John Day River | 182 to 243.7 | 1206499457318 | Cool water: Not less than 6.5 mg/l | 303(d)/2004 | 2004 Data: [DEQ/ODA – Salem] LASAR 11479 River Mile 212.3: From 6/15/1994 to 12/31/2003, 3 out of 40 samples (8%) < 6.5 mg/l and applicable % saturation. Adequate information exists to show that dissolved oxygen falls below 6.5 mg/l as a 30-day mean minimum, 5.0 mg/l as a 7-day minimum mean, and 4.0 mg/l as absolute minimum. More information is needed to determine if cool or cold water aquatic life is present. |
| North Fork John Day River | 0 to 13.1 | 119639447553 | Spawning: Not less than 11.0 mg/l or 95% of saturation | 303(d)/2004 | 2004 Data: [DEQ/ODA – Salem] LASAR 11017 River Mile 0.1: From 2/9/1994 to 4/17/2003, 2 out of 19 samples (10%) < 11 mg/l and applicable % saturations. |
| Utley Creek | 0 to 5.5 | 1193701440183 | Spawning: Not less than 11.0 mg/l 0r 95% of saturation | 303(d)/2002 | LASAR 11644 RM 0.9: 50% 3/6 of samples LASAR 11645 RM 2.2: 80% 4/5 samples |

[1] Corresponds to LASAR Station 11479

**Figure D-1. Impaired segments for DO listed on 2004-2006 303(d) in John Day Basin.**




## 1.2 Impairments Addressed in this TMDL

The DO TMDL developed in this document only addresses the cool water impairment on the John Day River (RM 182-243.7). For North Fork and Utley Creek, the Department will begin the process to change the status of these segments from Category 5 (impaired). The status of North Fork John Day River will be changed to Category 2 (Attaining) and Utley Creek will be changed to Category 3 (Insufficient Data).The impaired segment on the North Fork John Day River was for the digressions of the spawning season DO criteria using data from Station 11017 (RM 0.1). There were two digressions one in 2001 and another in 2002. However, the DO spawning-season criteria have been met since the last digression in 2002. For this reason, DEQ believes the status of North Fork John Day River should be changed to Category 2 (Attaining). The necessary information will be assembled to make the case for the change in status to Category 2. In addition, a review of historic data seems to indicate that Corral Creek, a tributary to the South Fork John Day River in the Upper John Day Sub-basin, may be impaired for the cool-water DO criteria. This data is from the early 1990's. A review of the 1996 and 1998 303(d) report background will be conducted to determine if there is a reason why impairment is not designated for DO on Corral Creek. The status of the cold-water impairment on the mainstem of the John Day River (RM 182 to 265) will be changed from Category 5 to Category 2. The status of this segment should not be impaired based on the current designation and available data. DEQ will work with their Assessment Section to change the status to attaining (Category 2) for this segment due to the administrative change in beneficial use designation. For Utley Creek, Station 11645 RM 2.2 on Utley Creek the impairment for the spawning season will be changed from Category 5 to Category 3 based on evidence presented in the following section.

## Analysis of Utley Creek Dissolved Oxygen Data

The status of Utley Creek, in the John Day Basin, is impaired (Category 5) for the 2002-2004 303(d) report. This impairment was based on small number of data points. A recording error of one measurement that caused the digression may have occurred. A statistical analysis of the dataset for outliers was under taken.

The impairment was based on a single sample at one station (11646) on Utley Creek, though all the other samples from this and other stations met water quality criteria. The locations of the station along Utley Creek are shown in **Figure D-2**. This impairment was based on data collected over the period of 1991 to 1993. The data from all the stations on Utley Creek are presented in **Table D-2**.

**Figure D-2. Station locations on Utley Creek.**

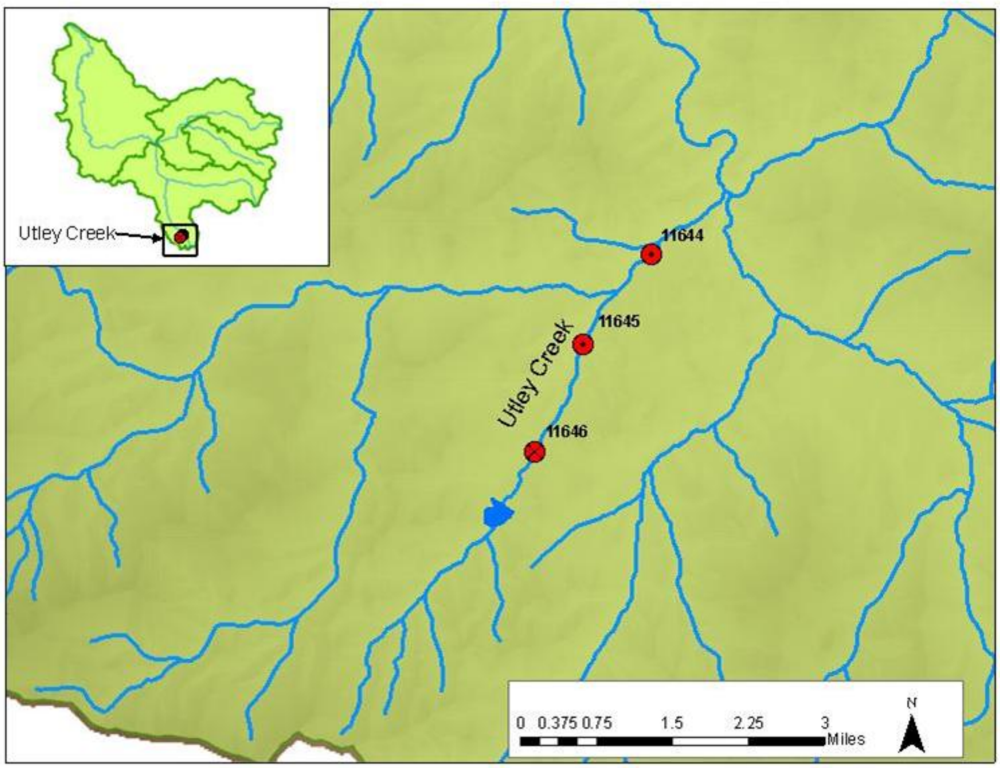


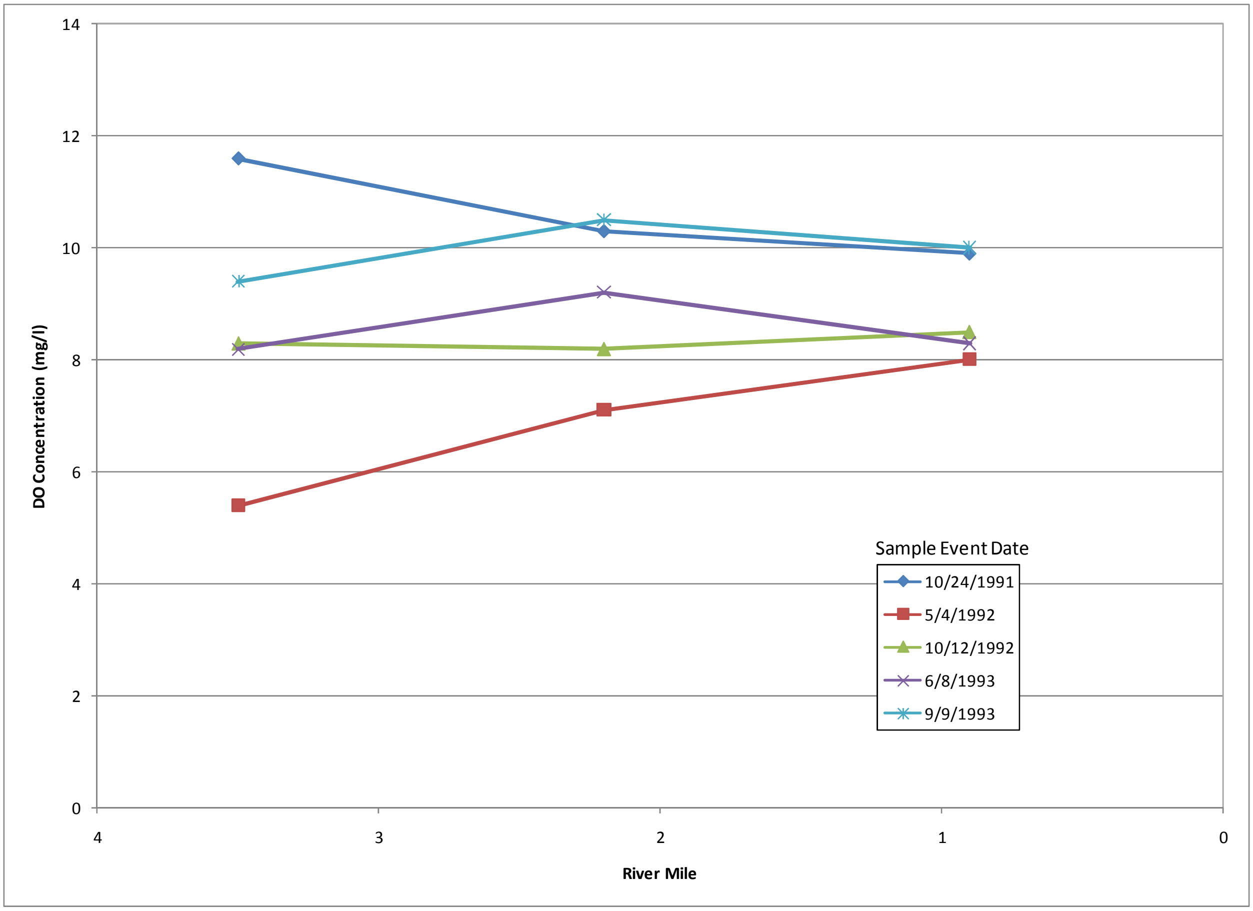
**Table D-2. Dissolved oxygen data collected at Utley Creek.**

| Station | Date Collected | DO (mg/l) | % Saturation | River Mile |
|---|---|---|---|---|
| 11644 | 10/24/1991 | 9.9 | 88 | 0.9 |
| 11644 | 5/4/1992 | 8 | 106 | 0.9 |
| 11644 | 10/12/1992 | 8.5 | 75 | 0.9 |
| 11644 | 6/8/1993 | 8.3 | 100 | 0.9 |
| 11644 | 9/9/1993 | 9.8 | 105 | 0.9 |
| 11644 | 9/9/1993 | 10 | 105 | 0.9 |
| 11645 | 10/24/1991 | 10.3 | 88 | 2.2 |
| 11645 | 5/4/1992 | 7.1 | 106 | 2.2 |
| 11645 | 10/12/1992 | 8.2 | 75 | 2.2 |
| 11645 | 6/8/1993 | 9.2 | 100 | 2.2 |
| 11645 | 9/9/1993 | 10.5 | 105 | 2.2 |
| 11646 | 10/24/1991 | 11.6 | 88 | 3.5 |
| 11646 | 10/12/1992 | 8.3 | 75 | 3.5 |
| 11646* | 5/4/1992 | 5.4 | 106 | 3.5 |
| 11646 | 6/8/1993 | 8.2 | 100 | 3.5 |
| 11646 | 9/9/1993 | 9.4 | 105 | 3.5 |

* Sample that resulted in the impairment.

The data point that caused the impairment is highlighted (in gray) in **Table D-2**. There are a few characteristics of the data set and this point (collected on 5/4/1992) in particular that suggest the impairment status does not accurately reflect the DO conditions for Utley Creek. The first characteristic of the data set is there are a small number of observations. There were only 5 samples collected at stations 11645 and 11646 (station with the digression) and 6 at station 11644. No data were collected after September 1993. During TMDL monitoring in 2004, DEQ sought permission to access the property for further data collection, but the landowner was unavailable. The next data characteristic is the data that caused the digression has a value of 5.4 mg/l, which is the same as the month/day that the sample was collected. The lab report for the sample event (19920394) was checked and 5.4 mg/l is what was recorded. However, this is probably a recording error by the field technician. The value of 5.4 mg/l is far lower than all the other values from each of the sample events from all of the stations (**Figure D-3**).
The last piece of evidence that the observation of 5.4 mg/l may be erroneous is from the results of a statistical analysis for outliers. Two tests were used for this analysis: the Grubbs' Test (Grubbs, 1969) and Dixon's Q Test (Dean & Dixon, 1951). The Grubbs' Test compares the maximum deviation of each observation from the mean to the standard deviation of the dataset to derive the ratio of the deviation from the mean to the standard deviation. The Grubs' Test assumes the data follows a Gaussian distribution (Grubbs, 1969) and if the estimated test statistic is greater than the test statistic for a given significance level, the observation is identified as an outlier. Dixon's Q Test is a nonparametric test that compares the difference (or gap) between each ordered pair of observations to the range (max – min) of the dataset (Dean & Dixon, 1951). The ratio of the gap to the range is the test statistic. In both tests, an observation is identified as an outlier if the deviation of an observation from the other data is greater than the variation of the entire dataset.

**Figure D-3. DO concentration versus river mile for each sample event.**




Both tests identified the suspect sample collected at station 11646 on 5/4/1992 as an outlier. The null hypothesis of the tests was that all the data come from the same distribution and are error free. The significance level chosen was 90% (α = 0.10) and a two-sided test was applied. The test results are presented in **Table D-3**. The tests were applied first to data from just station 11646 and then to data from all the stations. All of the tests for both data sets, except for the Grubbs' test on the data from station 11646 only, identified the suspect sample collected at Station 11646 on 5/4/1992 with the value of 5.4 mg/l as an outlier at an α-level of 0.10. Given the small number of data points, it was decided that visual analysis of the data using box-plots or histograms would not be appropriate.

**Table D-3. Outlier test results.**

| Station* | Test | Number of Observations | Test Statistic | p-value |
|---|---|---|---|---|
| 11646 | Grubbs' | 5 | 1.422 | 0.1043 |
| 11646 | Dixon's | 5 | 0.700 | 0.0861 |
| All | Grubbs' | 17 | 2.257 | 0.0834 |
| All | Dixon's | 17 | 0.572 | 0.0341 |

* "11646" – only data from station 11646 were used in the application of the tests and "All" – data from all of the stations (11644, 11645, and 11646) were used in the applications of the tests.

Based on the outlier analysis and date coincidence, DEQ believes the data point for 11646 on 5/4/1992 was misreported or otherwise compromised and should not be used in determining the DO water quality standard status of Utley Creek. All other data from this and the nearby stations indicate Utley Creek was meeting water quality criteria. More recent data would be needed to effectively evaluate ongoing attainment status and DEQ recommends that this data collection be included in the WQMP. Based on the results of outlier analysis and due to an update in the designated use, the DO water quality standard status of Utley Creek will be changed from Category 5 (Impaired) to Category 3 (Insufficient Data).

# 2. METHODOLOGY

Regression analysis was performed to develop a DO TMDL for the John Day River Basin. The specific analysis procedure used was Quantile Regression (QR). The QR analysis was performed on data grouped together based on when digressions occurred along with the season of the year and management activities in the basin. The equations from the QR were then used to determine the DO TMDL.

## 2.1 Regression Analysis

In order to determine pollutant reductions to achieve the DO criteria, regression analysis was conducted among the DO data and several water quality parameters. Since DO is not a pollutant, but rather a water quality condition needing improvement, the TMDL was not developed for DO loads. Rather, the TMDL was developed for any pollutant (and the relevant processes) causing poor DO conditions. Various approaches of different levels of complexity can be used for investigating the causes of DO conditions. For the John Day River Basin, regression analysis was used to determine equations that relate controllable water quality/quantity parameters to DO conditions. The load reductions for the TMDL can be determined from water quality targets that are based on these regression equations. The regression was only performed for data from station 11479. Station 11479 is one of DEQ's long-term 'ambient' water quality stations, where a full suite of supporting water quality parameters has been measured through the period of record. As discussed previously, the 303(d) listing of concern is based on this station. The other stations located on the mainstem of the John Day River around station 11479 did not have many water quality parameters available and if data were available the sample size was very small (less than 10). This is true from Station 11479 to the headwaters. The equations developed through this regression analysis apply to the segment listed from RM 182 to 243.7 on the John Day River mainstem.

### Parameter Selection

Twenty-one parameters that related physical, chemical and biological state to DO conditions along with water quantity were used in the regression analysis. The specific parameters are presented in **Table D-4**. The data was downloaded from DEQ LASAR. The time-period that the data used for the regression analysis was 1993 to 2008. Data quality levels used were A+, A, and B. There were some cases where continuous data was available. The maximums (TEMP), minimums (DO and DOSAT), or averages (CONDUCT and pH) were used to aggregate this data to a daily interval. When there was not an observation of a parameter available on the date the DO/DOSAT was measured, the parameter observation was assigned "NA" to signify a missing observation.

**Table D-4. List of parameters considered for regressions.**

| Parameter Name | Description |
|---|---|
| BOD | Biochemical Oxygen Demand Stream mg/l |
| CHLORA | Chlorophyll-α in µg/l |
| COD | Chemical Oxygen Demand in mg/l |
| CONDUCT | Field Conductivity in µmhos/cm |
| DIN | Dissolved inorganic nitrogen in mg/l calculated as NH3N + NO23N |
| ECOLI | E. coli concentration in cfu/100 ml |
| FLOW | From USGS 14040500 and using regression of USGS 14040500 = 0.2577* USGS 14046500 for missing values |
| NH3N | Ammonia as N in mg/l |
| NO23N | Nitrate/nitrite as N in mg/l |
| NPRATIO | Nitrogen/Phosphorus ration calculated from TOTN/TOTP |
| pH | Field pH in SU |
| PHEOPA | Pheophytin a in µg/l |
| PO4P | Dissolved Orthophosphate as P in mg/l |
| TEMP | Field Temperature in °C |
| TKN | Total Kjeldahl Nitrogen in mg/l |
| TOC | Total Organic Carbon in mg/l |
| TOTN | Total Nitrogen in mg/l calculated from DIN + TKN |
| TOTP | Total Phosphorus in mg/l |
| TOTS | Total Solids in mg/l |
| TSS | Total Suspended Solids in mg/l |
| TURB | Field Turbidity in NTU |

For some of the samples, percent saturation and concentration were not reported together. There were samples where % saturation was reported and DO concentration was not and vice versa. For samples where % saturation (or DO concentration) was missing, a value was calculated from the measured % saturation (or DO concentration). For the missing values, % saturation was calculated using the following equations:

$$\%\,saturation = \frac{DO}{DO_{sat}} \times 100$$

where,
% *saturation* is percent saturation of dissolved oxygen,
$DO_{sat}$ is saturated dissolved oxygen concentration in mg/L, and
*DO* is dissolved oxygen concentration in mg/L.

For estimating DO concentration, the % saturation equation was solved for DO concentration. $DO_{sat}$ was calculated considering the elevation of the station and the temperature of the water in the stream. The following equation (Duke & Masch, 1973) was used to calculate $DO_{sat}$:

$$DO_{sat} = (14.62 - 0.3898 \times T + 0.006969 \times T^2 - 0.00005897 \times T^3) \times (1 - 0.00000697 \times Elev)^{5.167}$$

where,
$DO_{sat}$ is saturated dissolved oxygen concentration in mg/L,
*T* is water temperature in °C, and
*Elev* is elevation above mean sea level in feet.

The elevations of all the stations were not available in LASAR, so they were determined from the DEM for USGS 30-m digital elevation model file clipped to the John Day River Basin. Temperature data were downloaded from LASAR. Temperature data was available for all the date-times that the DO concentrations and % saturations were measured. Continuous temperature data was aggregated to daily data using the maximum temperature for a given day.

## Flow Data

River flow was used in the regression analysis because it influences DO conditions in the water column. The flow data from USGS station 14040500 was used in the analysis of the DO data at station 11479. As described in **Appendix E** (*Derived Flow Data* section) the flow record from this station was filled-in where missing, via regression with the Service Creek gage station (USGS 14046500). The flow data were then adjusted in proportion to drainage area to approximate the flow at station 11479, approximately ten miles upstream.

## Regression Time Period for Analysis of DO Conditions

The seasons, pertinent management actions, and occurrence of criteria digressions were considered when selecting the time period of the data used in the regression analysis. The regression analysis was done only for time periods within the non-spawning season (May 16-December 31 when the cool water criterion of DO > 6.5 mg/l applies), but time periods were also included in the selection process to understand larger patterns in the data. The first approach was to use the data that fell within either spawning or non-spawning seasons. The dates relating to the spawning season are:

- Spawning Season (January 1 through May 15)
- Non-Spawning Season (May 16 through December 31)

Compared to smaller time frames, using the entire non-spawning season maximizes the number of observations used in the regression analysis. The second approach used the seasons of the year (standard 3-month definition). The dates of the seasons of the year are:

- Fall (September through November)
- Winter (December through February)
- Spring (March through May)
- Summer (June through August)

These time periods account for the seasonal influence on water quality parameters, such as temperature and flow. Next, the data was grouped by the 303(d) seasons which are:

- Summer (June through September)
- Fall/Winter/Summer (October through May)

These are the time periods used to develop the 303(d) list and were used to identify the existing DO impairment on the John Day River (RM 182-243.7). The next time periods considered the different season during the water year (October through September) and irrigation schedule. The period when irrigation occurs was set as April through June. This period was based on input from the water master (Personal Communication with Eric W. Julsrud, October 27, 2008) and stakeholders. Irrigation did not continue throughout the summer because many of the water rights excluded the low flow periods (after June) from times water could be withdrawn from the John Day River. The stakeholders indicated that there was little to no irrigation after July because there was little water left for most of the diversions downstream after the highest seniority water rights used their allocations. The highest seniority water rights could withdraw water all summer and were in the upper parts of the Upper John Day River sub-basin. The highest seniority water right diversions used up all the water that could be diverted in mid to late summer. However, much of the water diverted after July was going on a small portion of the land surface. Based on the feedback from the stakeholders, irrigation started around April and ended around early July for almost all of the fields in the Upper John Day River sub-basin. There will be still withdrawals after July, but the water will be going to a smaller and smaller area in the watershed. Based on this information the irrigation season was set to April through July. The water seasons were combined with the spawning season period to focus on the period when digressions were occurring. The water-spawn seasons were defined as:

- Irrigation-Spawn (April through May 15)
- Irrigation-Non-Spawn (May 16 through June)
- Dry (July through September)
- Wet Non-Spawn (October through December)
- Wet Spawn (January through March)

The time period ultimately selected is based on visual inspection and the descriptive statistics of the data. Boxplots were used to inspect the distribution of the data visually. The distribution of the DO data was also compared among seasons by plotting the boxplots for all of the seasons of a grouping (e.g. water-spawn season). This allowed for the assessment of seasonal patterns of the DO data. The descriptive statistics listed in **Table D-5** were also estimated for the time periods of the different groupings. These statistics provided additional information about the characteristics of the data sets that is more specific than the boxplots. All of this information was used to determine if the data set for a time period had favorable characteristics for the regression analysis. Once a time-period was selected, the data only from that time-period was used in the regression analysis. For the John Day dissolved oxygen TMDL, the selected time period is specified in the subsequent section ***Time Period Selection*** in the regression results chapter of this appendix.

**Table D-5. Descriptive statistics used for time period selection.**

| Statistic | Description |
|---|---|
| N | Number of observations |
| $\mu$ | Arithmetic average |
| $\sigma$ | Standard deviation |
| CV | Coefficient of variation ($\sigma/\mu$) |
| Min | Minimum |
| Median | Median (50$^{th}$ percentile) |
| Max | Maximum |
| IQR | Inter-quartile range is the 3/4$^{th}$ quartile minus the 1/4$^{th}$ quartile or 75$^{th}$ percentile minus the 25$^{th}$ percentile |

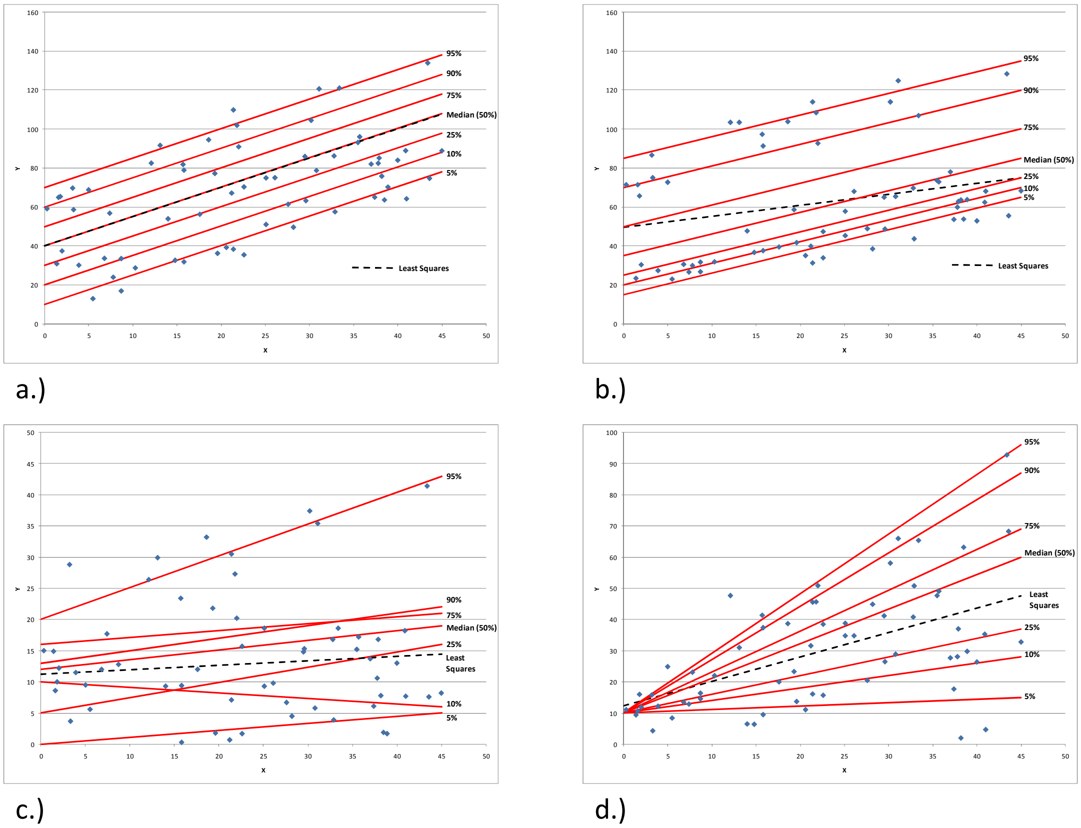
## Quantile Regression – Method Background

Quantile regression (QR) was used to develop equations that relate DO concentrations to other water quality parameters. QR is basically the estimation of a family of equations for the quantiles (e.g. 10%, 25%, and so on) of a data set. This allows for the estimation of relationships among variables "for all portions of a probability distribution" (Cade & Noon, A Gentle Introduction to Quantile Regression for Ecologists, 2003). QR was initially developed for use in economics (Koenker, 2005), but has been used extensively in ecological and environmental research(Cade, errell, & Schroeder, Estimating Effects of Limiting Factors with Regression Quantiles, 1999; Dunham, Cade, & Terrell, 2001; McClain & Rex, 2001; Cade & Noon, A Gentle Introduction to Quantile Regression for Ecologists, 2003; Bryce & Lomnicky, 2008). In the DO TMDL for the John Day River Basin, QR was used rather than least squares for several reasons. First, processes that influence DO concentrations are often constrained by conditions and processes that are measured indirectly through the collection of water quality data. This is especially true for the biological processes (Dunham, Cade, & Terrell, 2001). For example, the consumption of DO by algae in a stream is not measured directly, but is inferred through the measurement of plant nutrients or chlorophyll-α concentrations in the water column. Second, unmeasured factors (Cade & Noon, A Gentle Introduction to Quantile Regression for Ecologists, 2003) may be limiting DO concentrations. There are many processes that control DO concentration and not all of them are captured by the water quality monitoring data. Furthermore, processes operate on multiple temporal scales and this factor may not be represented in the water quality data, especially the data collected monthly. None of these reasons could be addressed using least-square methods for the regression analysis of the water quality and DO data.

QR is related to least squares with respect to the information provided about the conditional probability function of the data. The conditional probability function is simply the probability that Y happens given that X occurred. A simple example is the conditional probability of getting sunburn (Y) at night (X) is very low. For the topic of DO TMDL, we are concerned with the probability of a DO concentration occurring given values of other water quality parameters. For instance, what is the probability of measuring a DO

concentration of 10 mg/l, if the stream temperature was measure to be 20°C? The information about the conditional probability function is more specific when using regression. Regression analysis (QR or least-squares) estimates the conditional expected value for the conditional probability function. Some examples of expected values are the arithmetic average (mean), median (50%-quantile), 95%-quantile, and so on. Besides the specific computation methods, the main difference between least-squares and QR is the expected values that each method estimates. Least-squares regression estimates the expected value for the mean conditional probability function. Conversely, QR estimates the expected values for the quantiles of the conditional probability function. If the conditional probability function is symmetric, then the mean equals the median and the least-squares equation and QR equation for the median would be the same. The power of QR is that the equations for multiple quantiles can be estimated for the conditional probability function, which provides more information about the relationship among the dependent and independent variables. One thing that should be made clear is that the quantiles of the conditional probability function are not the same as the quantiles of the independent variable. For instance, the QR equation that relates DO to flow for the 75% quantile is not the same as the 75% quantile of the DO data. The correct interpretation is that the quantile for the equation where an amount of the data pairs are above and below. For instance, the equation for the DO-Flow 75% quantile means that 75%of the flow-DO data pairs are below the line that crosses the conditional probability function. It does not mean that 75% of the DO measurements will be less that the DO estimated from the equation. The QR equations provide a statistical model that can be used to investigate a system, such as stream DO conditions.
In order to better understand how QR is used for the analysis of the water quality and DO data, a few general conditions are presented. Four different data distributions along with the QR results are shown in **Figure D-4**. For case a.), the least-squares estimate would be as good as the QR for the median (Koenker, 2005). However, if the portion of the conditional-probability distribution of x-y was away from the center of the data, where the mean and the median are, QR would need to be used. For case b.) in **Figure D-4**, the conditional probability function for x and y is skewed. The slopes of the equations for the different quantiles are the same, but the difference among the intercepts increases for increasing quantiles. As can be seen for case b.), the least-squares estimate is not a good representation of the relationship between x-y. The least-squares equation over-estimates low values and under-estimates high values of y. There are both limiting and unmeasured factors shown in case c.). There is much uncertainty for the quantiles of 90% and below signified by the crossing of the lines. This uncertainty could be evident in the lack of statistical significance of the slope and intercept estimates. For case c.) in **Figure D-4**, there are unmeasured factors other than x that may influence y, but there is a strong relationship between x and y for the 95% quantile. It is also evident that x acts as a limiting factor for maxima of y. For this case, the least-squares regression does not capture any of this information and would likely result in the slope and intercept estimates not being statistically significant. Finally, the QR for case d.) reveals the heteroskedacity of the relationship between x and y (**Figure D-4**). Heteroskedacity occurs when the dispersion of data depends on the magnitude of the data. As can be seen in case d.), the slopes of the regressions increase as the quantile increases. The effect of this heteroskedacity on the least-squares regression is basically under-prediction of y throughout the range of x (except for very small values of x where y is over-predicted). The four cases shown in **Figure D-4** are distinct examples of conditions where QR would be more useful than least-squares regression in understanding the relationship between parameters. There are other data characteristics that are better represented using QR. However, the biggest advantage of QR over least-squares is the ability to focus on specific ranges of the dataset.

**Figure D-4. Examples of different conditions highlighted with QR: a.) Identically independently distributed, b.) skewed, c.) limited and unmeasured factors, and d.) heteroskedastic.**




## Quantile Regression John Day – TMDL Application

The set of 22 water quality parameters was reduced to a sub-set using QR. For each of the 22 parameters, QR was performed in relation to DO for the quantiles of 5%, 10%, 25%, 50%, 75%, 90%, and 95% along with the quantile for the cool-water criteria (6.5 mg of DO/ml). Only water quality parameters that had statistically significant QR slope and intercept estimates for any quantiles considered was selected for the next step of the analysis. The slope and intercept estimates were considered statistically significant if the p-value for the estimate was less the 0.10 (the α-level selected for the QR analysis). The null hypothesizes are the intercept and slope estimates are equal to zero and the alternatives are that the slope and intercept estimates are not equal to zero.

The next step further reduced the number of water quality parameters by considering the controllability and whether the QR relationship between the water quality parameter-DO made physical sense. An example of non-controllability is decreasing DO concentrations with respect to increasing air temperatures. The amount of DO water can hold is strongly related to water temperature, which turn is related to air temperature. However, the air temperature is not very controllable. Another example of a non-controllable relationship is the need for extreme changes in a water quality parameter in order to improve DO conditions. It may not be possible to change the level of a water quality levels to improve DO conditions, say if nitrate levels of zero are required. An example of QR relationships not making physical sense would be if the QR results indicate that as water temperature increase DO conditions improve. This is opposite to what is commonly understood about the physical relationship between DO and temperature. After all the screening is complete, the performance of the statistical model that is the QR equations will be investigated in order to select a single equation for use in the DO TMDL.

## DO Model Selection from Quantile Regression Equations

As with other models, the performance of statistical models needs to be assessed for accuracy. For Quantile Regression (QR) results, the different equations for the different quantiles were considered competing models once the independent variable(s) are selected. Several methods were used to assess the QR equations performance. The first was to compare the estimated DO concentration from the equations to observed DO concentration by calculating the relative difference. Small absolute magnitudes of relative difference are better with zero being the best. Next, statistical properties of the estimated and observed DO were compared. The properties were represented using the performance statistics for the minimum, quantiles (5%, 10%, 25%, 50%, 75%, 90%, and 95%), and maximum of the estimated and observed DO data. Remember the quantiles of the estimated and observed DO concentrations are not the same as the quantiles of the QR equations. The QR equations were thought of as separate models that are being compared. Two other performance statistics were used to assess the statistical models (QR equations). The first is the Nash-Sutcliffe model efficiency coefficient (Nash & Sutcliffe, 1970). The equation for the Nash-Sutcliffe model efficiency coefficient (*E*) is:

$$E = 1 - \frac{\sum_{i=1}^{N}\left(O_i - \hat{O}_i\right)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2}$$

where,
$O_i$ is the $i^{th}$ observed value,
$\hat{O}_i$ is the $i^{th}$ modeled value,
$\bar{O}$ is the arithmetic average of the observed values, and
*N* is the number of observed-modeled pairs.

The values for *E* range from −∞ to 1. An *E* of 1 corresponds to the observed and modeled data being a perfect match. When *E* is 0, the model predictions are as accurate as the mean of the observed data. Finally, the mean of the observed data is a better predictor than the model when *E* is less than zero. Basically, the closer *E* is to one the better the model. The data used to calculate *E* was the minimums, quantiles, and maximums calculated for the observed and modeled data. The final performance statistic calculated was the duration (*D*) expressed as a percent that the modeled and observed DO data digressed from the cool-water criterion (6.5 mg/l). This duration was calculated using the following equation:

$$D = \frac{N_D}{N_T} \times 100\%$$

where,
$N_D$ is number of the digressions of the cool-water criterion and
$N_T$ is the total number of values.

All of these statistics and associated graphs were used to select the statistical model (QR equation) used in the DO TMDL. These performance statistics were tabulated and compared visually (**Figure D-5**). The data displayed in **Figure D-5** is a facsimile of the cumulative distribution function of the observed and modeled DO concentrations. If the model is under-predicting the DO concentrations, the line in **Figure D-5** will be below the line for the observed data and vice versa for the case where the model over-predicts the DO concentrations. Also, the performance statistics were compared quantitatively using the percent difference between the value for the performance statistic of an equation and the statistic for the observed data using the following formula.

$$\%\,Differnce = \frac{\left(Stat_{Eq} - Stat_{Obs}\right)}{Stat_{Obs}}$$

An example of the percent difference would be comparing the maximum DO concentration estimated using the QR-50% equation to the maximum of the observed DO concentrations. The acceptable ranges for the performance statistics are listed in **Table D-6**. The observed DO concentrations used to calculate the performance statistics were observations that were collected during July-August for any year data was available.

**Figure D-5. Comparison of Observed and Modeled Statistics.**

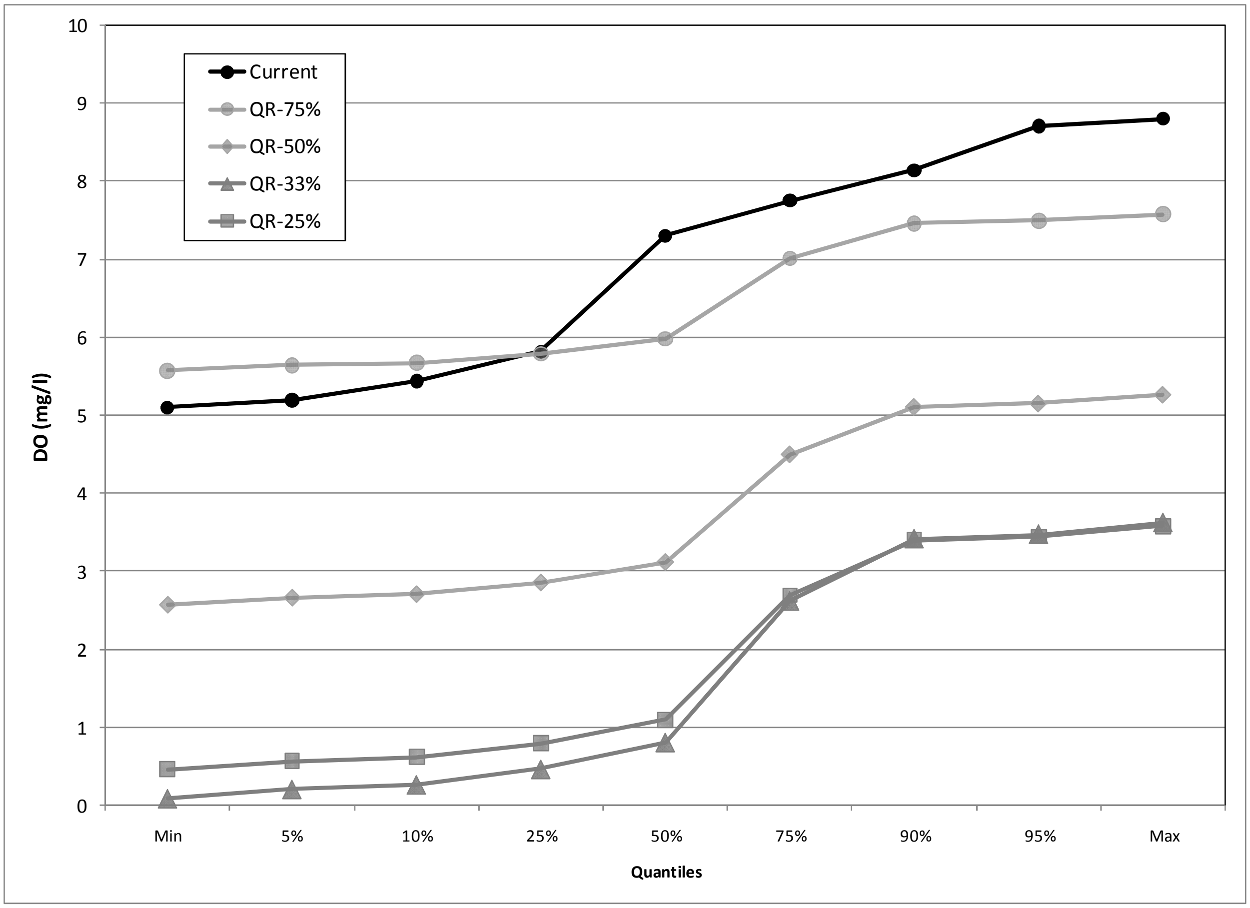



**Table D-6. Acceptable Ranges for performance statistics.**

| Statistic | Min | 5% | 10% | 25% | 50% | 75% | 90% | 95% | Max | *E*[2] | *D*[3] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Range[1] | ±25% | ±25% | ±25% | ±25% | ±25% | ±25% | ±25% | ±25% | ±25% | >0, closer to 1 better | ±50% |

[1]All ranges are relative difference to observed except *E*
[2]Nash-Sutcliffe Model Efficiency Coefficient
[3]Duration that DO concentrations were below the cool-water criterion of 6.5 mg/l

# 3. DO REGRESSION RESULTS

After going through the selection process above, the target water quality parameter selected for the DO TMDL is stream temperature (TEMP). This selection was based on Quantile Regression (QR) results and a family of equations for the quantiles (e.g. 10%, 25%, and so on) of a dataset estimated for the TEMP and DO observed data. The analysis of the water quality data was done for several seasonal groupings to consider spawning/non-spawning seasons, different climatic conditions, and management activities in the basin. The first step was to select the season to use for the QR analysis.

## 3.1 Time-Period Selection

The time-period selected for the QR analysis focused on the summer part of the non-spawning season when there is little to no irrigation occurring and little precipitation. The time-period selection criteria considered the occurrence of digressions, season, and pertinent management actions. The focus of the time-period selection was for the non-spawning season (May 16-December 31).

After preliminary analysis, the spawning/non-spawning season grouping of the DO data was not considered suitable for the regressions. The distribution of the DO data is shown in **Figure D-6**. For the non-spawning season, there appears to be two distinct groupings of the DO data. The digressions (shown in red) are separated from the rest of the data by a region (6.5 mg/l to 8 mg/l DO) where there is a sparse amount of data. The summary statistics for the DO data grouped by spawning season are in **Table D-7**. This possibly bi-modality of the DO data during the non-spawning season is also evident by the high value of the Coefficient of Variation (CV) for the non-spawning season (Table D-7). Furthermore, the non-spawning season (May 16-December 31) extends from spring to winter. During this period, there are a lot of changes in the climate that influence DO concentrations and may not be controllable, like air temperature. For these reasons, the full non-spawning season was not used for the QR analysis of the DO data.

**Figure D-6. Spawn/non-spawn season boxplots.**
**(Seasons are defined starting on page 10)**

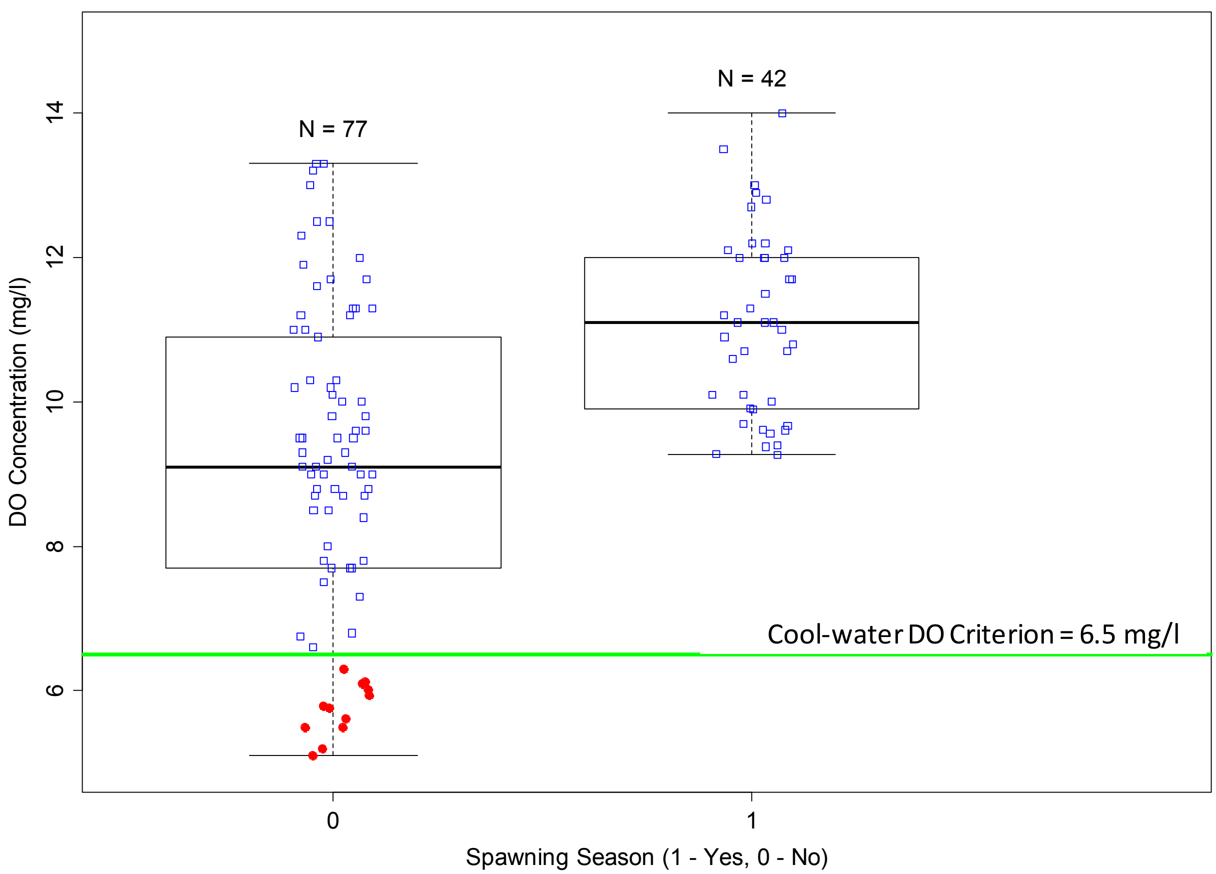


**Table D-7. Summary statistics for DO grouped by spawning season.**

| Season | N | *μ* | *σ* | CV[1] | Min | Median | Max | IQR[2] |
|---|---|---|---|---|---|---|---|---|
| **Spawning** | 42 | 11.1 | 1.26 | 0.11 | 9.3 | 11.1 | 14.0 | 2.1 |
| **Non-Spawning** | 77 | 9.2 | 2.15 | 0.23 | 5.1 | 9.1 | 13.3 | 3.2 |

[1] Coefficient of Variation is $\sigma/\mu$
[2] Inter-quartile range is the 3/4$^{th}$ quartile minus the 1/4$^{th}$ quartile or 75$^{th}$ percentile minus the 25$^{th}$ percentile

Following preliminary data screening, the 3-month season grouping was not used because the digressions were split into two periods. The distribution of the DO data for the 3-month seasons is shown in **Figure D-7**. The digressions are split between summer (June through August) and fall (September through November). Also, the DO concentrations are split between two sub-groups for the fall season, which is most likely due to the increase in precipitation in November (**Figure D-7**) and a steady decrease in air temperature from September to November. The summary statics for the 3-month seasons are presented in **Table D-8**. The high IQR and σ for the fall data is also unsatisfactory. The 3-month season grouping was not used for the QR regression based on these deficiencies.

**Figure D-7. Standard 3-month season boxplots.**
**(Seasons are defined starting on page 10)**

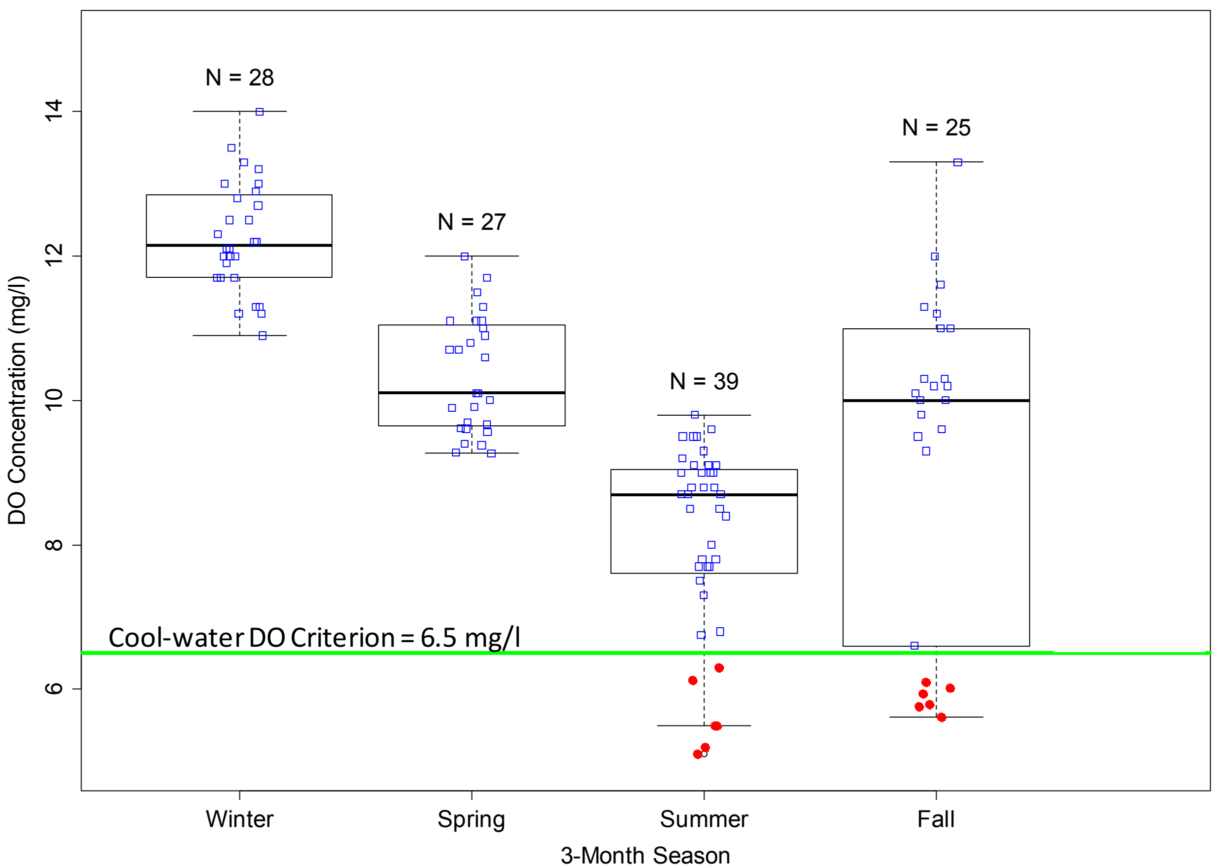


**Table D-8. Summary of DO data for standard 3-month seasons.**

| Season | N | $\mu$ | $\sigma$ | CV[1] | Min | Median | Max | IQR[2] |
|---|---|---|---|---|---|---|---|---|
| Winter | 28 | 12.3 | 0.77 | 0.06 | 10.9 | 12.1 | 14.0 | 1.1 |
| Spring | 27 | 10.4 | 0.81 | 0.08 | 9.3 | 10.1 | 12.0 | 1.4 |
| Summer | 39 | 8.1 | 1.32 | 0.16 | 5.1 | 8.7 | 9.8 | 1.5 |
| Fall | 25 | 9.3 | 2.29 | 0.25 | 5.6 | 10.0 | 13.3 | 4.4 |

[1] Coefficient of Variation is $\sigma/\mu$

[2] Inter-quartile range is the 3/4$^{th}$ quartile minus the 1/4$^{th}$ quartile or 75$^{th}$ percentile minus the 25$^{th}$ percentile

The 303(d) summer season was from June 1 through September 30. All of the digressions occurred during this period. The distribution of the data is shown in **Figure D-8**. As with the previous seasonal groupings, the summer season appears to have two groups. This could result from the irrigation in the beginning of the summer and tapering off to zero before the end of the season. The summary statistics for the 303(d) seasons are presented in **Table D-9**. The high σ of the summer season is similar to the previous seasons where the digressions occurred. The distribution of the summer DO data may also be skewed demonstrated by the difference between the mean and the median. The skewness may be the result of the different process affecting DO concentrations as the water management practices change during the summer season. The change in the water management during the 303(d) summer season , potential skew and high standard deviation of the data were the reasons that 303(d) seasonal grouping was not used for the QR analysis.

**Figure D-8. CWA 303(d) list season boxplots.**
**(Seasons are defined starting on page 10)**

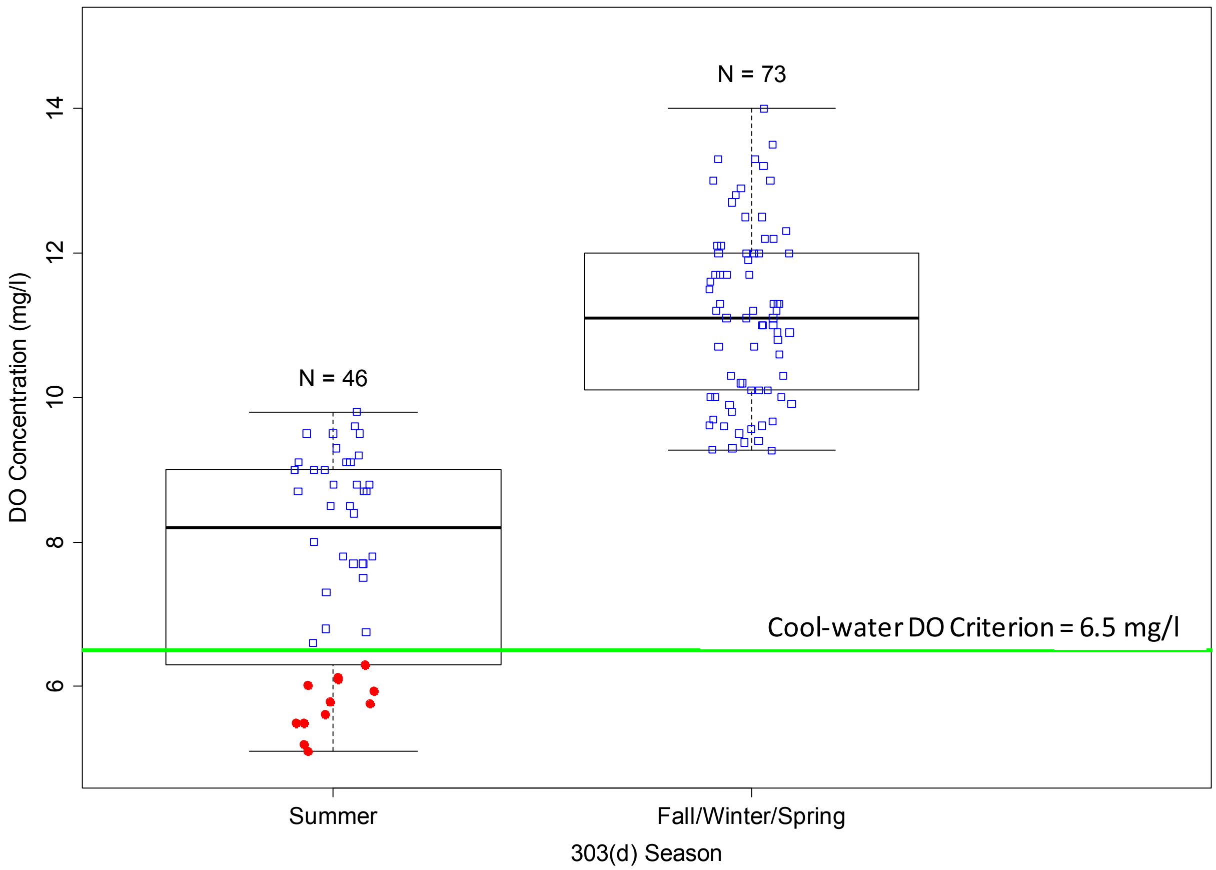



**Table D-9. Summary of DO data for CWA 303(d) list seasons.**

| Season | N | $\mu$ | $\sigma$ | CV[1] | Min | Median | Max | IQR[2] |
|---|---|---|---|---|---|---|---|---|
| **Summer** | 46 | 7.8 | 1.44 | 0.18 | 5.1 | 8.2 | 9.8 | 2.6 |
| **Fall/Winter/Spring** | 73 | 11.1 | 1.22 | 0.11 | 9.3 | 11.1 | 14 | 1.9 |

[1] Coefficient of Variation is $\sigma/\mu$
[2] Inter-quartile range is the 3/4th quartile minus the 1/4th quartile or 75th percentile minus the 25th percentile

The dry season of the water-spawn season was selected for the QR analysis because this grouping isolated important water management activities and had favorable data characteristics. The dry water-spawn season is from July 1 to September 30. The distribution of the DO data for the water-spawn seasons is shown in **Figure D-9**. For the Dry season (July through September), the data tends to be well distributed across the range of values with a slight gap possibly occurring around 6.5 ml/l to 6.7 ml/l DO. The summary statistics for the water-spawn season are presented in **Table D-10**. The data for the dry season appear to be symmetrically distributed demonstrated by the agreement between the mean and median. Also, the σ is smaller than for the other seasonal groupings, but sufficient to provide enough variation for the regression. The dry water-spawn season (July to September) was used in the QR analysis because the grouped data had the relevant conditions and processes contributing to the digressions of the non-spawning season criteria represented and favorable data characteristics.

**Figure D-9. Water-spawn season boxplots. (Seasons are defined starting on page 10)**

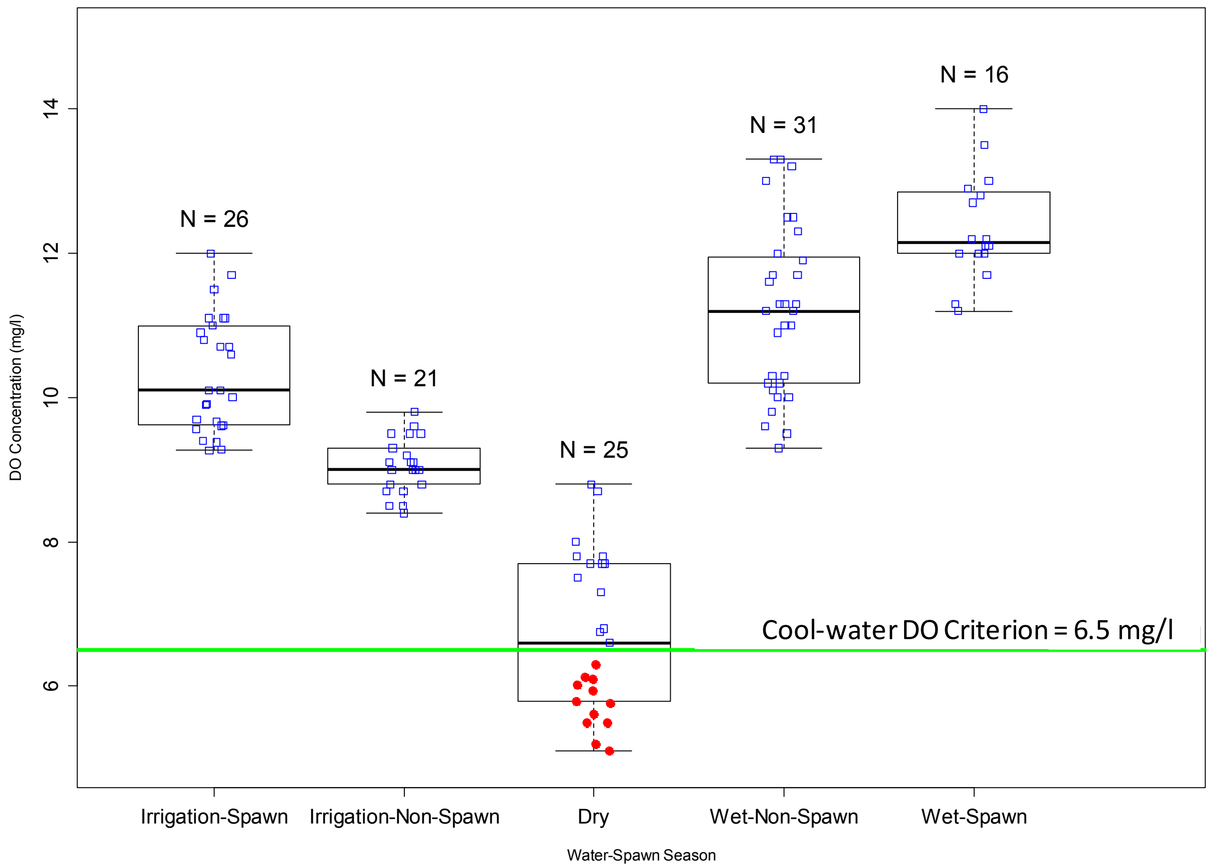



**Table D-10. Summary of DO data for water-spawn seasons.**

| Season | N | μ | σ | CV[1] | Min | Median | Max | IQR[2] |
|---|---|---|---|---|---|---|---|---|
| **Irrigation-Spawn** | 27 | 10.4 | 0.81 | 0.08 | 9.3 | 10.1 | 12.0 | 1.4 |
| **Irrigation-Non-Spawn** | 21 | 9.1 | 0.39 | 0.04 | 8.4 | 9.0 | 9.8 | 0.5 |
| **Dry** | 25 | 6.7 | 1.10 | 0.16 | 5.1 | 6.6 | 8.8 | 1.9 |
| **Wet Non-Spawn** | 31 | 11.21 | 1.17 | 0.1 | 9.3 | 11.2 | 13.3 | 1.75 |
| **Wet Spawn** | 16 | 12.36 | 0.75 | 0.06 | 11.2 | 12.15 | 14 | 0.83 |

[1] Coefficient of Variation is *σ/μ*
[2] Inter-quartile range is the 3/4th quartile minus the 1/4th quartile or 75th percentile minus the 25th percentile

## 3.2 Parameter Selection and Regression Results

A set of equations relating DO concentration to stream temperature (TEMP) were selected for the dry season based on the QR results. The selection of the water quality parameters (and associated equations) proceeded in a step-wise fashion. First, QR was performed for all water quality parameters. The equations with statistically significant slopes ($\alpha \leq 0.10$) were analyzed further (Step 1 in **Table D-11**). Next, the influence of individual observations on the regressions was investigated. Potential outliers or high-leverage points were identified based on visual inspection and excluded from the regressions (Step 2 in **Table D-11**). The water quality parameters that still had significant slopes were used in the final step of the analysis (Step 3 in **Table D-11**). The controllability and physical consistency of the relationship described by the regressions were assessed in this final step. The progression of the screening of water quality parameters for the dry season is presented in **Table D-11**. After the final step in parameter selection, a set of regression equations was produced for the remaining parameter.

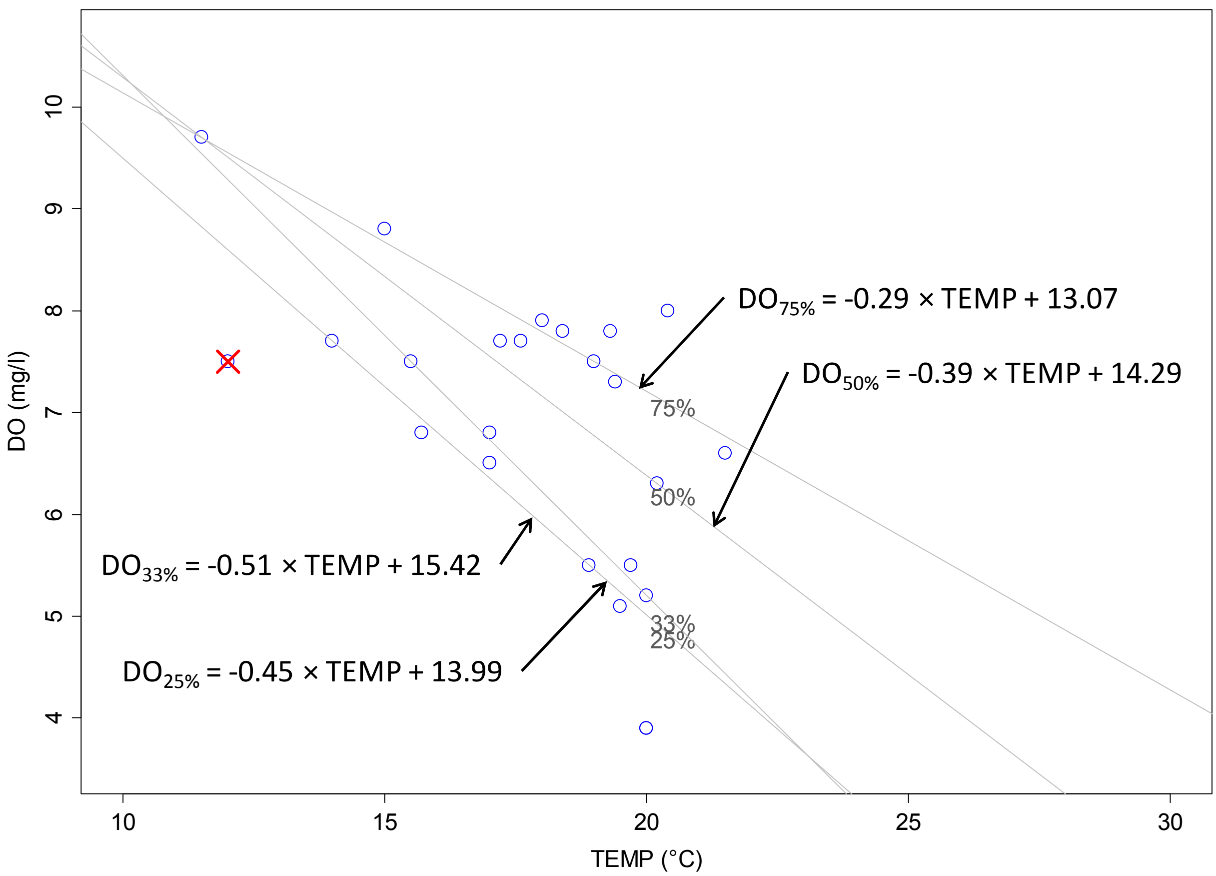
**Table D-11. Progression of water quality parameter screening for DO during Dry season. (Parameter abbreviations are explained in Table 4)**

| All | Step 1 | Step 2 | Step 3 |
|---|---|---|---|
| AIRTEMP | AIRTEMP | AIRTEMP | AIRTEMP |
| BOD | BOD | BOD | BOD |
| CHLORA | CHLORA | CHLORA | CHLORA |
| COD | COD | COD | COD |
| CONDUCT | CONDUCT | CONDUCT | CONDUCT |
| DIN | DIN | DIN | DIN |
| ECOLI | ECOLI | ECOLI | ECOLI |
| FLOW | FLOW | FLOW | FLOW |
| NH3N | NH3N | NH3N | NH3N |
| NO23N | NO23N | NO23N | NO23N |
| NPRATIO | NPRATIO | NPRATIO | NPRATIO |
| pH | pH | pH | pH |
| PHEOPA | PHEOPA | PHEOPA | PHEOPA |
| PO4P | PO4P | PO4P | PO4P |
| TEMP | TEMP | TEMP[1] | TEMP |
| TKN | TKN | TKN | TKN |
| TOC | TOC | TOC | TOC |
| TOTN | TOTN | TOTN | TOTN |
| TOTP | TOTP | TOTP | TOTP |
| TOTS | TOTS | TOTS | TOTS |
| TSS | TSS | TSS | TSS |
| TURB | TURB | TURB | TURB |

[1] TEMP was used in Step 2 and had significant slopes after the omission of a low TEMP value from regression.

There is only one component to the DO criteria during the dry season of the water-spawn seasons: the cool water criterion of DO ≥ 6.5 mg/l (for example, accounting for saturation is not included in the criterion). After reducing the number of parameters in Step 2, only the DO data with DOSAT ≤ 100% were used for these regressions. The rational for this exclusion is based on the definition of "Daily Mean" in OAR 340-041-0002(15). Exclusion of DO data with DOSAT > 100% reduced the number of quantiles that parameters had with significant slopes, but did not eliminate parameters. Even though TEMP was not identified as having significant slopes in Step 1, it was further analyzed in Step 2. After the omission of a low value (TEMP = 12°C and DO = 7.5 mg/l), there were significant slopes for the 25%, 33% (the quantile of DO criterion of 6.5 mg/l), 50%, and 75% quantiles. AIRTEMP, CONDUCT, ECOLI, pH, and TOTS are considered to be indicator parameters for DO rather than parameters that could be used to control in-stream DO. FLOW could be used as a back-up parameter if more information is needed or an alternate approach rather than controlling TEMP. TEMP (stream water temperature) was selected because it is controllable through restoration of shade, channel morphology and flow, which are addressed in temperature TMDL. The equations for the QR of DO vs. TEMP are shown in **Figure D-10**.

**Figure D-10. QR for DO vs. TEMP during Dry season.**

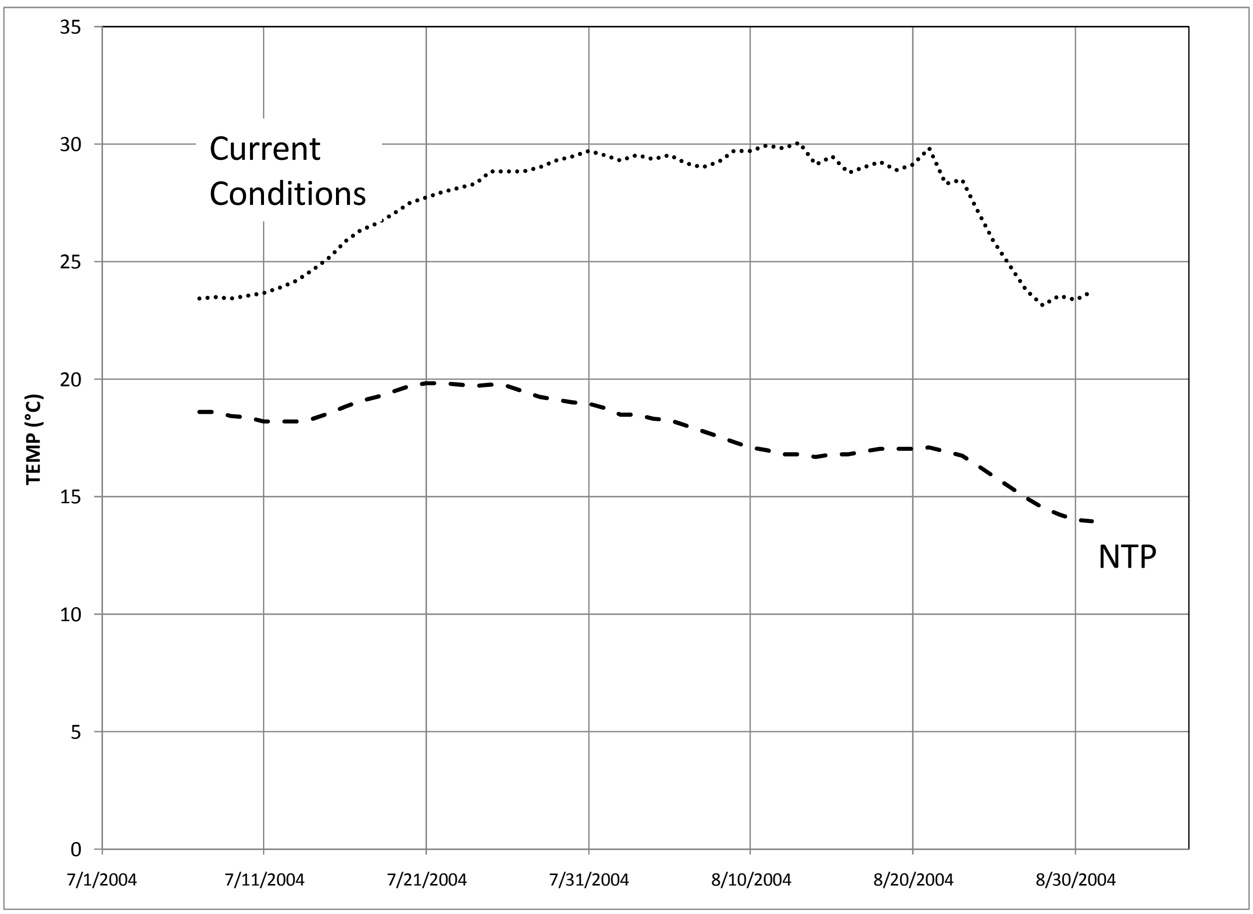


## 3.3 Selection of QR Equation to Model DO

The QR equation used to model DO was selected by comparing the simulated and observed DO concentrations. The simulated temperature from the temperature TMDL was used as input to the QR equation. There were four QR equations (statistical models) available to simulate DO using temperature. The candidate models are shown in **Figure D-10**. The model performance statistics were used to select the QR-75% equation for use in the DO TMDL.

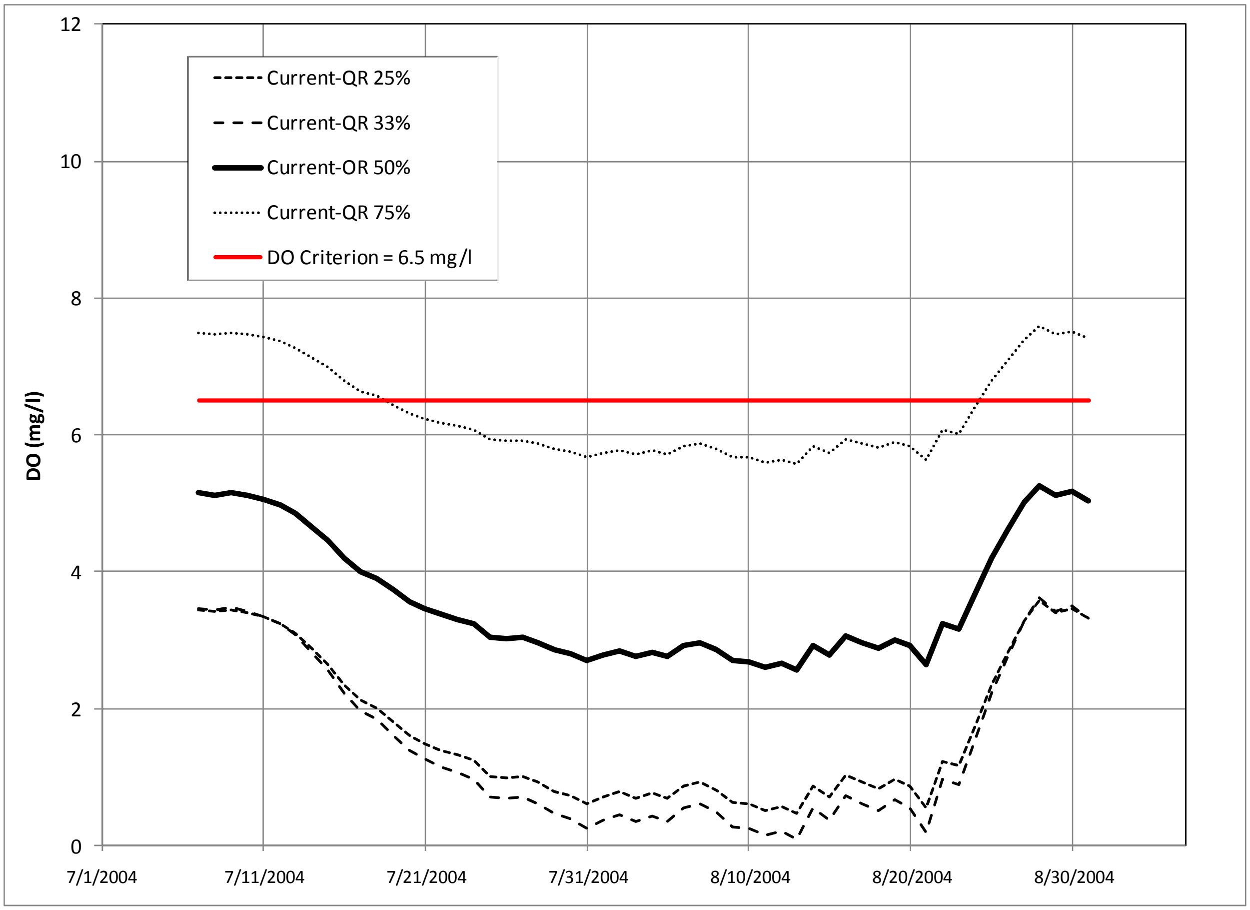
The simulated temperature from the temperature TMDL for river kilometer 333 was used as input to the DO-TEMP statistical model. Station 11479, which is the station that the data used in the QR analysis was collected at, is located at river kilometer 333. The simulated temperature for current conditions and the Natural Thermal Potential (NTP) scenarios are shown in **Figure D-11**. The simulation period was from July to August 2004. This is the simulation period used in the temperature TMDL. As described in Chapter 2.1, the NTP condition is the general target of the temperature TMDL in the warm season.

**Figure D-11. Simulated temperature for rKM 333 at station 11479.**



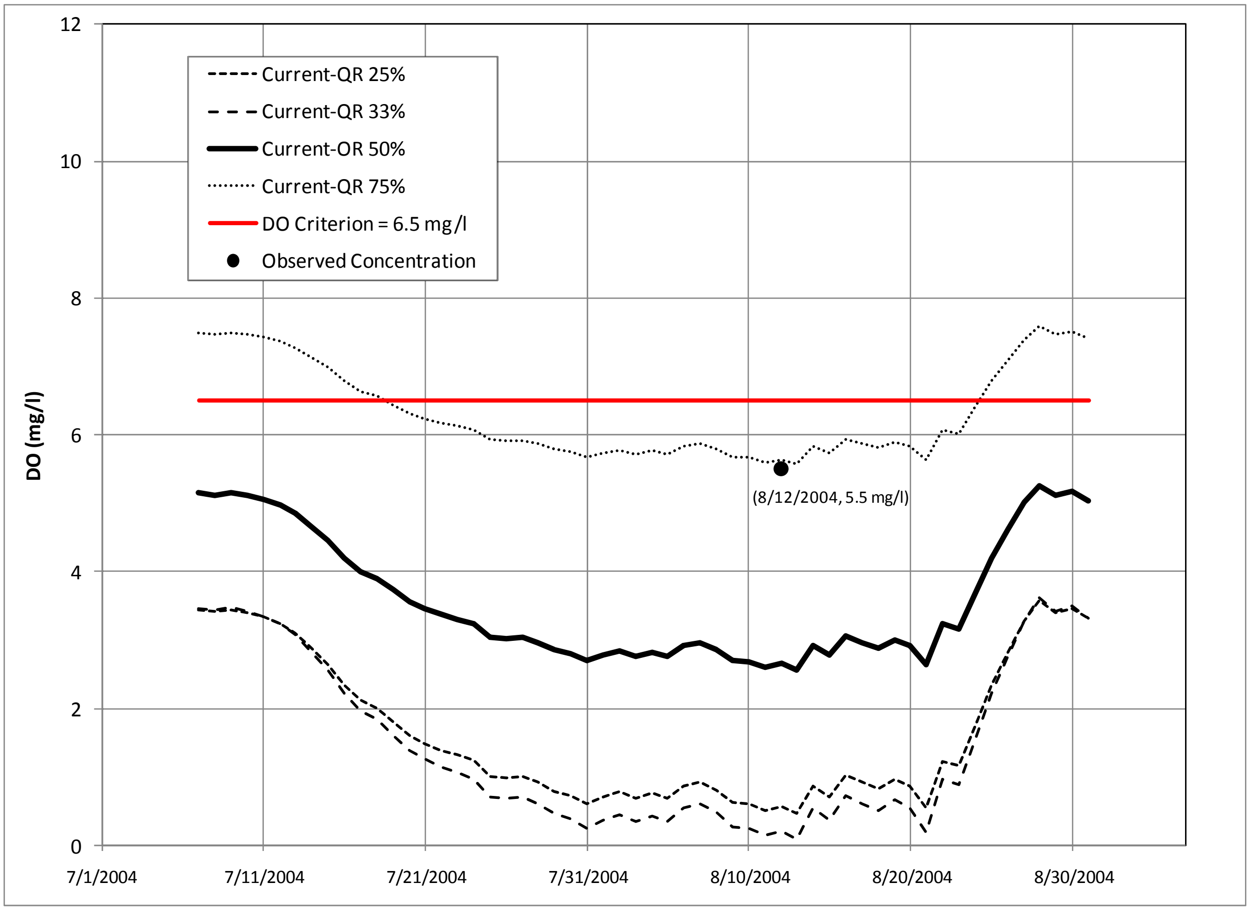

The simulated temperature was used with the DO-TEMP equations to simulate DO for existing and natural potential conditions. The simulated DO concentration time-series are shown in **Figure D-12**. The simulated DO for QR-33% drops below the DO for QR-25%. This is the result of the lines for the QR-25% and QR-33% crossing around DO of 4 mg/l (**Figure D-10**). The crossing of the lines indicates that the processes influencing DO at these higher values for TEMP are not captured very well using one or both of the QR-25% and QR-33% equations. For the QR-50% equations, the simulated DO concentrations are below the cool-water criteria throughout the simulation period. This implies that there are always digressions during this time of the year, which is not the case. The QR-75% equation has simulated DO concentrations both above and below the cool-water criteria implying that there are some digressions during this season of the year.

**Figure D-12. Simulated TEMP-DO for all QR equations.**




The simulated DO concentrations for the QR-75% equation were very close to an observed value. The simulated and observed DO data are shown in **Figure D-13**. There was only one DO observation for this simulation period. However, it is worth noting that the QR-75% simulated DO is very close (within 3%) of the observed DO. The simulated DO for the other equations are much lower than the observed DO concentration. The simulated QR-25% DO was 90% less than the observed, QR-33% was 96% less than, and QR-50% was 52% less than the observed DO concentration. Since there was only one observed DO concentration available within the simulation period, the extent that these differences among the simulated and observed DO could be used to draw conclusions about the different models is limited. Yet, the fact that the independently calibrated simulated temperature was used to simulate DO and the QR-75% DO concentration was so close to the observed DO provides strong qualitative evidence for the selecting the QR-75% equation to simulate DO – as the best representation of relevant processes. Additional quantitative evidence was investigated next in the model selection process. The observed DO concentrations used to calculate the performance statistics were observations that were collected during July-August for any year data was available.

**Figure D-13. Simulated TEMP-DO for all QR equations with observed DO concentration.**




The QR-75% TEMP-DO model performed better than all of the other models for all of the performance statistics. The values for performance statistics are in **Table D-12**. The minimum, quantiles, and maximum values for the all the TEMP-DO models are much lower than the observed except for the simulated DO for the QR-75% model, which is visually evident in **Figure D-14**. The statistics for QR-75% are very similar to the observed. It should be noted that the maximum for the QR-25% and QR-33% simulated DO are slightly more than the minimum observed DO (**Table D-12**). The minimum observed DO is equal to the value for the 90% quantile of the data simulated DO from the QR-50% model. These under predictions indicate that the QR-25%, QR-33%, and QR-50% models do not represent the current DO conditions very well.

The poor performance of the QR-25%, QR-33%, and QR-50% models and good performance of the QR-75% model are evident in the value of *E* (**Table D-12**). The negative for *E* indicate that the average of the observed data simulates DO concentrations better than the QR-25%, QR-33%, and QR-50% models. The relative differences for the performance statistics are listed in **Table D-13**. All of the QR models under-predict the DO concentrations except of the QR-75% model. Also, the *E* value of 0.7 indicates that QR-75% model simulates DO concentrations accurately (*E* of one indicates perfect accuracy of simulated data).

The final statistic considered was the duration that DO concentrations were below the cool-water criterion (*D*). The observed data are less than the cool-water criterion 37% of the time (**Table D-12**). All of the QR models had *D* outside the acceptance range of ±50%. The simulated DO concentration for QR-25%, QR-33%, and QR-50% models are always below the criterion (*D* = 100%).In contrast, *D* for QR-75% model is closer to the *D* of the observed (**Table D-13**). Although *D* for QR-75% model is greater than *D* for the observed, simulated DO for the QR-75% model does have times when the cool-water criterion is met, where the other three models do not. This absence of meeting the cool-water criterion and the under-prediction of the QR-25%, QR-33%, and QR-50% models (**Table D-13**) lead to the conclusion that these models do not represent the current DO conditions well and would also be excessively restrictive if used.

**Table D-12. Performance statistics for the TEMP-DO models.**

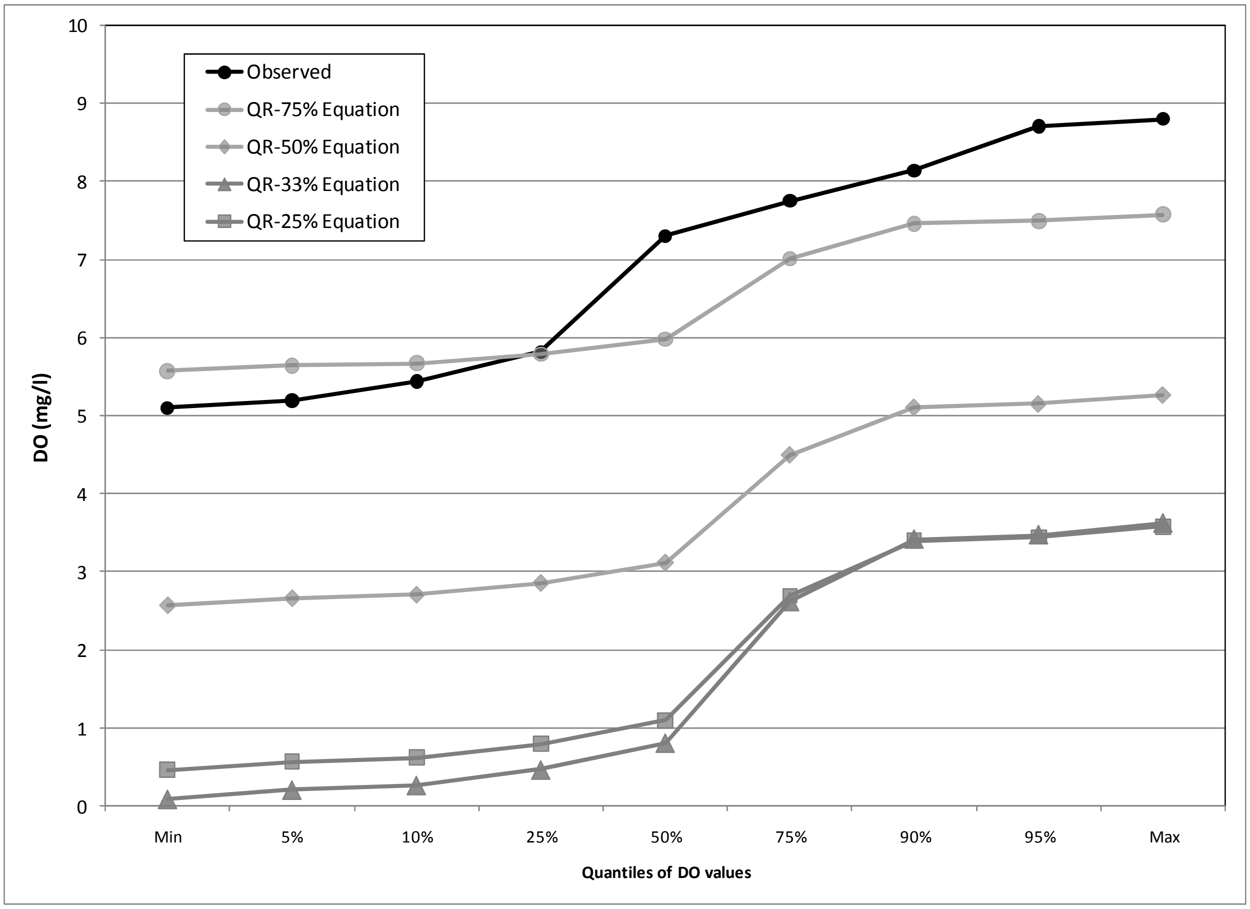
| | **Min** | **5%** | **10%** | **25%** | **50%** | **75%** | **90%** | **95%** | **Max** | ***E*[1]** | ***D*[2]** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| QR-25% | 0.5 | 0.6 | 0.6 | 0.8 | 1.1 | 2.7 | 3.4 | 3.4 | 5.1 | -11.4 | 100% |
| QR-33% | 0.1 | 0.2 | 0.3 | 0.5 | 0.8 | 2.6 | 3.4 | 3.5 | 5.3 | -12.3 | 100% |
| QR-50% | 2.6 | 2.7 | 2.7 | 2.9 | 3.1 | 4.5 | 5.1 | 5.2 | 6.6 | -3.9 | 100% |
| QR-75% | 5.6 | 5.6 | 5.7 | 5.8 | 6.0 | 7.0 | 7.5 | 7.5 | 7.6 | 0.7 | 66% |
| **Observed[3]** | **5.1** | **5.2** | **5.4** | **5.8** | **7.3** | **7.8** | **8.1** | **8.7** | **8.8** | **---** | **37%** |

[1] Nast-Sutcliffe Model Efficiency Coefficient
[2] Duration that DO concentrations were below the cool-water criterion of 6.5 mg/l
[3] Observed data is only for August for multiple years. August was the only month data was collected for the months of the simulation period

**Table D-13. Relative difference of performance statistics for the TEMP-DO models.**

| | Min | 5% | 10% | 25% | 50% | 75% | 90% | 95% | Max | D |
|---|---|---|---|---|---|---|---|---|---|---|
| QR-25% | -91% | -89% | -89% | -86% | -85% | -65% | -58% | -60% | -59% | 170% |
| QR-33% | -98% | -96% | -95% | -92% | -89% | -66% | -58% | -60% | -59% | 170% |
| QR-50% | -50% | -49% | -50% | -51% | -57% | -42% | -37% | -41% | -40% | 170% |
| QR-75% | 9% | 9% | 4% | -1% | -18% | -10% | -8% | -14% | -14% | 79% |

**Figure D-14. Min, quantiles, and max performance statistics for TEMP-DO models compared to observed data for current conditions.**




In summary, the statistical model of the QR-75% equation for DO-TEMP was used based on the evidence provided in the model selection process and the performance statistics, and comparisons of simulated to actual data. The QR-75% model provided simulated DO concentrations closest to the observed DO (**Figure D-13**). Even though there was only one observed DO concentration available, the simulated temperature was independently calibrated and when used to simulate DO with the QR-75% equation the DO concentration was close to the observed DO, thus providing strong qualitative evidence that QR-75% is the optimal choice. The QR-75% model performed best of the four models for all of the statistics considered in the selection process (**Table D-12**). This indicates that this model accurately simulates the DO conditions in the John Day River at station 11479. Almost all of the performance statistics for the QR-75% model were within the acceptable ranges (**Table D-13**). The duration that DO was less than the cool-water criterion (*D*) was the exception. The *D* for the QR-75% equation was 79% greater than the observed, which was outside the ±50% range. Finally, the QR-75% model only slightly under and over-predicted the observed data (**Figure D-14**). Based on this evidence, the DO-TEMP QR-75% model was used for the DO TMDL.

## 3.4 DO TMDL Target

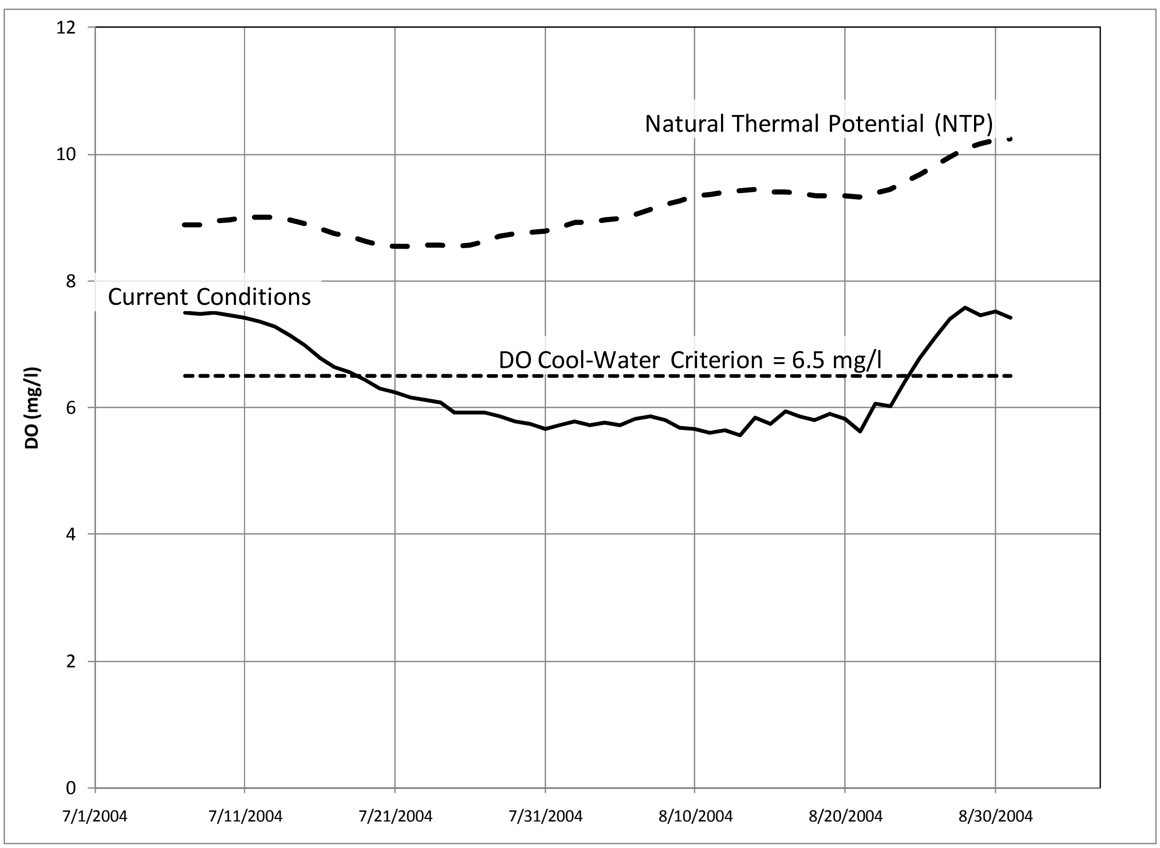
The target stream temperature (TEMP) relationship that controls DO conditions in the John Day River Basin is calculated using the QR equation. The target TEMP QR equation is applied to existing and natural thermal potential (NTP) temperatures simulated for the temperature TMDL. The simulated TEMP data are for the same location of the data used in the QR analysis (station 11479, rKM 333). The DO concentration output from the simulated TEMP and QR equation was then compared to the DO cool-water criterion of 6.5 mg/l.

The simulated TEMP time-series from the temperature TMDL were converted to DO concentrations using the QR equation. The DO time-series from the QR equation and simulated TEMP was compared to the cool-water criterion (6.5 mg/l) to simulate current DO conditions and to evaluate if the DO criterion would be met using the load allocations of the temperature TMDL. The simulated TEMP from the temperature TMDL for rKM 333 is shown in **Figure D-11**. The natural thermal potential (NTP) is the temperature TMDL target during the time frame shown. The temperature time-series given in **Figure D-11** was converted to DO using the QR equation selected to simulate DO conditions. As long as the DO time-series for the NTP temperatures does not go below the cool-water criterion (6.5 mg/l), the DO TMDL Target is met through the load reductions of the temperature TMDL, which is the case (**Figure D-15**). Load reductions were not directly calculated for the John Day River DO TMDL. This is because TEMP was used to address DO, and thermal loads have been established for TEMP in the temperature TMDL.

Accordingly, and as stated in **Chapter 2.3**, the load allocations, waste load allocations and reserve capacity for temperature comprise the DO TMDL as well. Additional margin of safety is provided for the dissolved oxygen TMDL through the following assumptions:

- DO and TEMP continuous data used daily minimum and maximum to aggregate data
- Focused on the driest and warmest part of year
- Integrated the margin of safety of the temperature TMDL into the DO TMDL

**Figure D-15. Simulated DO from temperature TMDL time-series and QR equation.**



# 4. REFERENCES

Bryce, S. A., & Lomnicky, G. A. (2008). Development of Biologically Based Sediment Criteria in Mountain Streams of the Western United States. *28*, 1717-1724.

Cade, B. S., & Noon, B. R. (2003). A Gentle Introduction to Quantile Regression for Ecologists. *Frontiers in Ecology , 1* (8), 412-420.

Cade, B. S., errell, J. W., & Schroeder, R. L. (1999). Estimating Effects of Limiting Factors with Regression Quantiles. *80* (1), 311-323.

Cleland, B. (2002, August 15). TMDL development from the “bottom up” – Part II: using duration curves to connect the pieces. *America’s Clean Water Foundation* .

Dean, R. B., & Dixon, W. J. (1951). implified Statistics for Small Numbers of Observations. *Analytical Chemistry , 23* (4), 636-638.

Duke, J. H., & Masch, F. D. (1973). *Computer Program Documentation for the Stream Water Quality Model DOSAG3, Vol. 1.* Wasshington D. C.: Prepared for the USEPA, System Development Branch.

Dunham, J. B., Cade, B. S., & Terrell, J. W. (2001). Influences of Spatial and Temporal Variation on Fish-Habitat Relationships Defined by Regression Quantiles. *Transactions of the Amrerican Fisheries Society , 131*, 86-98.

Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Tecnometrics , 11* (1), 1-21.

Koenker, R. (2005). *Quantile Regression* (Vol. 38). New York: Cambridge University Press.

McClain, C. R., & Rex, M. A. (2001). The Relationship Between Dissolved Oxygen Concentratio and Maximum Size of Deep-Sea Turid Gastropods: An Application of Quantile Regression. *139*, 681-685.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology , 10* (3), 282-290.

R Development Core Team. (2009). *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing) Retrieved 2009, from The R Project for Statistical Computing: http://www.R-project.org

R Development Core Team. (2009). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

USGS. (2009). *USGS Surface-Water Daily Data for the Nation*. Retrieved 2009, from National Water Information System: Web Interface: http://waterdata.usgs.gov/nwis/dv/?referred_module=sw