Draft Final Report – Analysis of water quality data from the lower Deschutes River, July-August, 2014, for the Deschutes River Alliance.

John Van Sickle, Oct 23, 2014

Contents

1. **Data processing and analysis methods**
2. **Results**
3. **Summary of patterns seen in Results**

1.0 METHODS

1.1 Data sets and reformatting.

All analyses are based on 2 sets of data delivered to me by R. Hafele, Sept. 19, 2014.

The grab sample data set (*Dwr Deschutes 2014 Field Grab.xlsx)* is a series of separate tables in a spreadsheet. I concatenated these tables into a single table, adding columns to identify the sampling location. The reformatted data set (*all.grab.samples.xlsx*) is suitable for import into statistical analysis software, and an e-copy will be delivered with this report.

Time series readings from 4 data sondes were delivered as a set of 4 spreadsheet files. Again, I concatenated the 4 into a single file (*all.sondes.xlsx*) for importation into statistical software, and an e-copy will be delivered with this report.

1.2 Analysis software

All analysis and graphics were done using the R language (v.3.1.1). The internally-documented R scripts will be delivered with this report.

1.3 Analysis of grab-sample data

Grab samples of river water were collected from 5 sites between River Mile (RM) 4 and RM 96, in midmorning (AM) and again in midafternoon (PM), on 3 successive days (July 21-23). Samples were assessed for 5 analytes: temperature, pH, specific conductance, dissolved oxygen, and turbidity. Grab samples during August at 3 sites were also included in the data, but are not analyzed here.

The 3 AM samples (or PM samples) at one site cannot be considered as independent replicates because of their proximity in time and space. Thus, I calculated the means over the 3 days, separately for the AM and PM measurements of each analyte, giving a single AM (and PM) value for each site. (Four of the 5 sites also had an AM sample on July 24.

However, because one site (RM 4) did not have this sample, the other July 24 AM samples were excluded when calculating the mean).

I then tested for a possible upstream linear trend, for each analyte across the n=5 sites, by estimating the linear (product-moment) correlation coefficient between the site mean and the site river mile. Assuming statistical independence between the sites (adjacent sites are about 25 mi apart), I also tested whether each correlation was significantly different from zero. This analysis was done separately for the AM and PM means.

I plot and tabulate the means for each analyte, versus the site river mile, and the plots report the correlations and their P-values. A significant positive correlation indicates that the analyte was increasing in an upstream direction. No multiple-testing adjustments were made for the correlation significances.

1.4 Analyses of sonde data.

1.4.1 – Data description and screening.

Sondes were deployed at 3 sites (RM 4, 57, and 100) during August 5-8. The fourth sonde was deployed in July at RM96 and is not analyzed in this report. The sondes recorded measurements, every 15 minutes continuously, of 4 analytes: temperature, pH, conductivity, and DO.

Preliminary time-series plots revealed outlying values of temperature and pH at the first 6 time points of the RM100 record, and of conductivity at the last time point of RM4 record. R. Hafele and L. Marxer concurred that the sondes were not fully submerged and acclimatized at these time points, so I deleted all measurements at these time points from my analyses. Data from these times, however, remains in the delivered data file *all.sondes.xlsx.*

After dropping these time points, the final continuous records available for analyses have the following start and end times:

RM	Start time	End time
4	11:00 8/5	11:30 8/8
57	11:00 8/5	07:00 8/8
100	13:30 8/5	13:45 8/7

1.4.2 Summary statistics and models.

As with the grab sample data, the primary analysis goal for the multiple sonde time series was to describe how they differed between sites.

Time series plots (see Results below) of all 4 analytes showed smoothly-changing diurnal variation for all analytes. For temperature and conductivity, these temporal patterns can be easily observed and directly compared between sites, by examining the time series plots. In addition, I tabulate the observed mean, maximum, and minimum of these 2 sets of time series, as well as the time period(s) over which the minimum and maximum were observed.

I did not try to model either the temperature or conductivity time series. The temperature times series show almost total separation between the sites (see below), so modeling seemed unnecessary. The conductivity time series had extended periods of constant conductivity, which would require a complicated and lengthy model with low interpretive value.

The pH and dissolved oxygen (DO) time series (see plots below) show very clear sinusoidal variation over the course of each day, with substantial overlap across sites. For these two analytes, I tried to clarify the sinusoidal pattern and possible time trend, by fitting the following sinusoidal regression model with a linear time-trend, separately at each site:

$$Y(t) = a + b*t + c*\sin(2\pi t) + d*\cos(2\pi t)$$

In this model, time (t) is expressed in elapsed days since the start of 08/01, and sinusoidal terms assume 1 complete cycle per day. Model coefficients (a-d) for each analyte and river mile were estimated with least-squares linear regression applied to the observed time series, Y(t). I report the R^2 of each model as well as the estimated trend coefficient (b). In addition, I report the mean, max and min of observed values for Y(t), as well as the observed time periods over which the max and min occurred.

1.4.3 No statistical inference

Because of their very strong temporal dependence, individual measurements for temperature and conductivity could not be treated as true replicates for constructing such tests or confidence intervals.

For pH and DO, the sinusoidal models accounted for nearly all of the temporal dependence in Y(t) (that is, models had very high R^2). As a result, model residuals might be sufficiently serially independent to justify placing confidence intervals on model coefficients or predictions. However, for these data, the model residuals were so small in magnitude that they were likely due entirely to the round-off significant digits reported for the raw measurements. Thus, they do not express the kind of unexplained natural variation over time and space that would be represented by true multiple replicate samples.

For these reasons, I report all statistical summaries of the sonde time series without confidence intervals, and did not perform any formal hypothesis tests of between-site differences.

1.4.4 Common time period

The sinusoidal models of pH and DO were fit to the full periods of time series data available at each of the 3 sites (Sec. 1.4.1). In addition, the time series plots for all analytes span the full series length at each site..

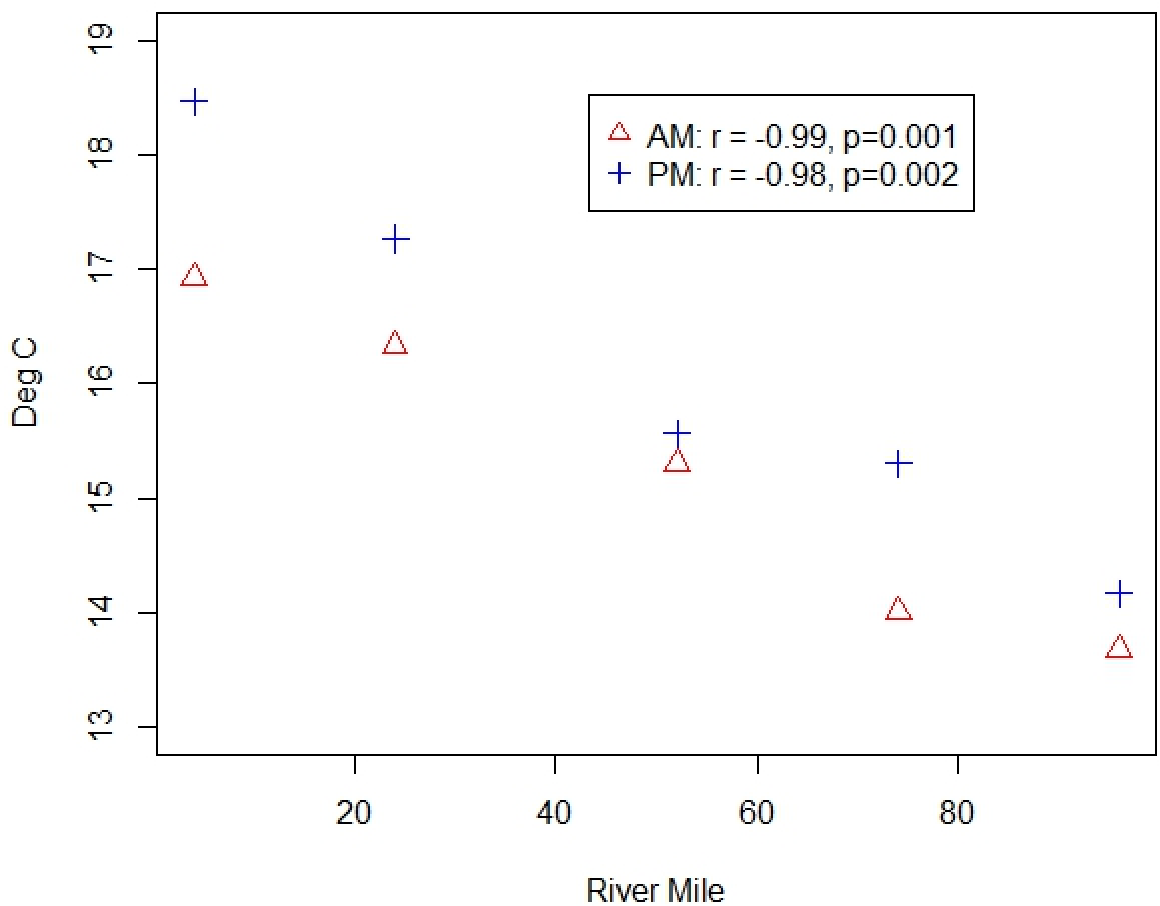
However, estimating the means or other summary statistics over the full periods would generate bias, because of the differing period lengths (see e.g., the temperature time series plots below). To eliminate such bias, I estimated all summary statistics over the exact same time period, defined as longest time period that was covered at all 3 sites. **Thus, all summary statistics from sondes are estimated over the period from 13:30 8/5 to 13:45 8/7** , which is the full sampling period available from the shortest time series (RM 100).

2. Results

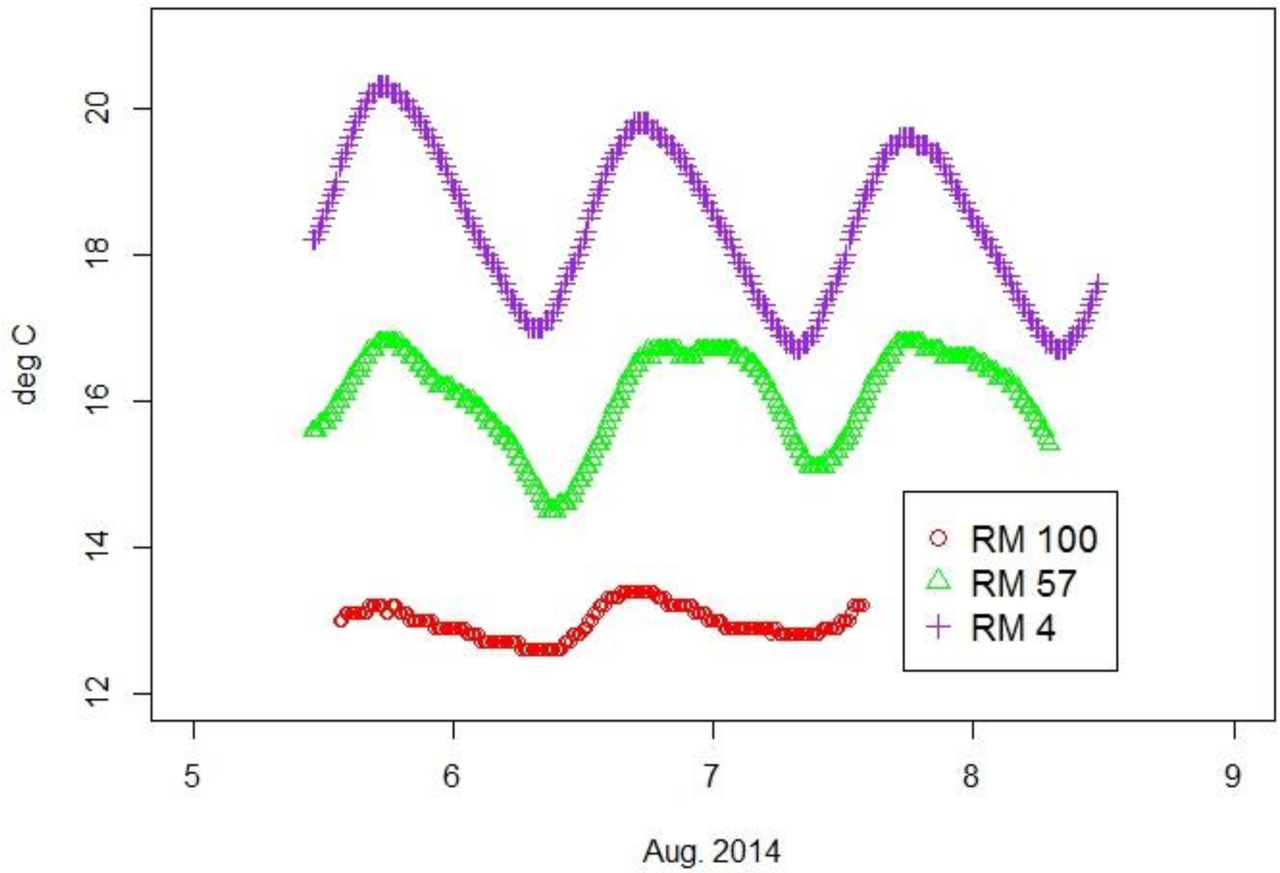
For each analyte, I report its grab sample results, followed by its sonde time series results.

2.1 Temperature

2.1.1 Grab Samples, July 21-23, 2014: Mean temperature versus river mile, separately for AM and PM samples. Correlations and their p-values are for a linear trend. Means are listed in Table 1.



2.1.2 Temperature time series from sonde data at 3 sites (RM 4, 57, 100)

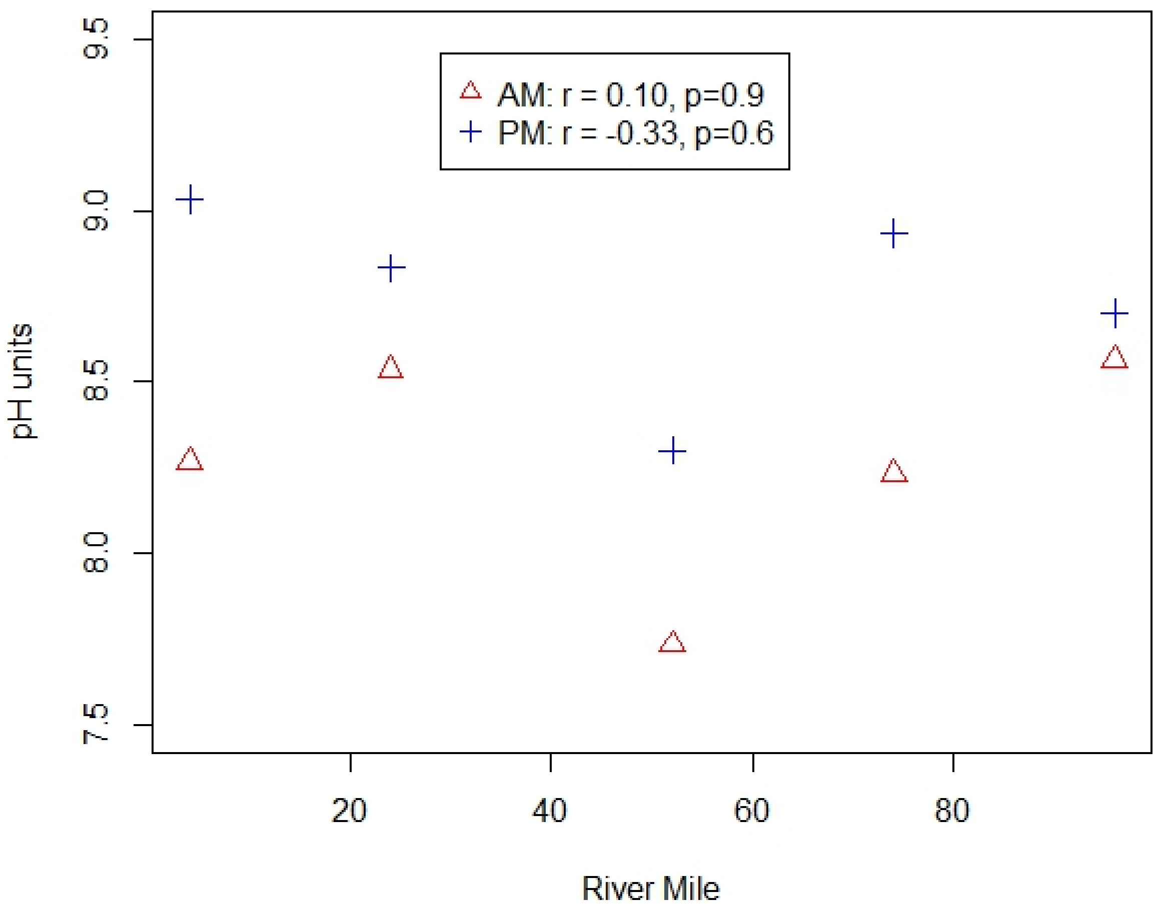


Summary statistics (Deg. C) for observed temperature time series, from 13:30 8/5 to 13:45 8/7

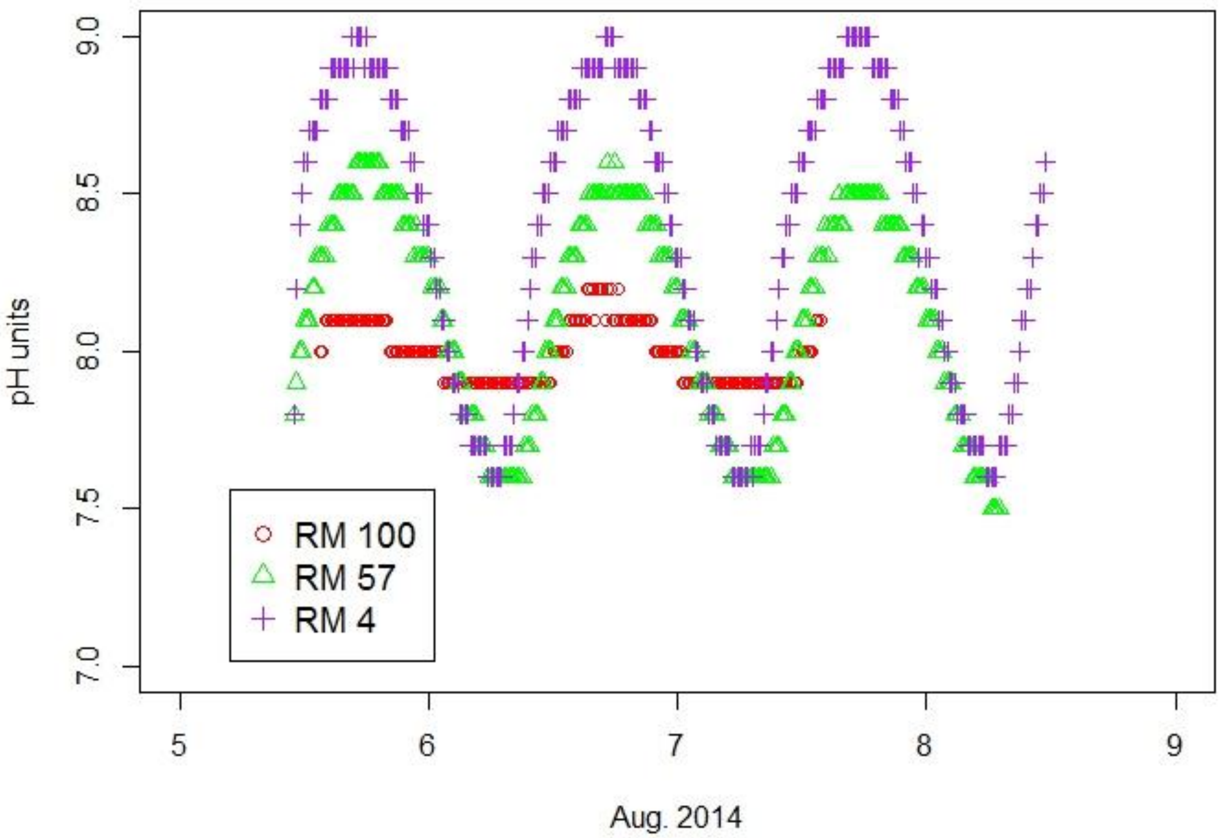
River Mile	Mean	Max	Time of Max	Min	Time of Min
4	18.5	20.3	17:00-18:00 08/05	16.7	7:30-8:00 08/07
57	16.0	16.8	17:00-19:00 08/05	14.5	8:45-9:30 08/06
100	13.0	13.4	15:30-18:30 08/06	13.4	6:15-10:00 08/06

2.2 pH

2.2.1 Grab Samples, July 21-23, 2014: Mean pH versus river mile, separately for AM and PM samples. Correlations and their p-values are for a linear trend. Means are listed in Table 1.



2.2.2 pH time series from sonde data at 3 sites (RM 4, 57, 100)

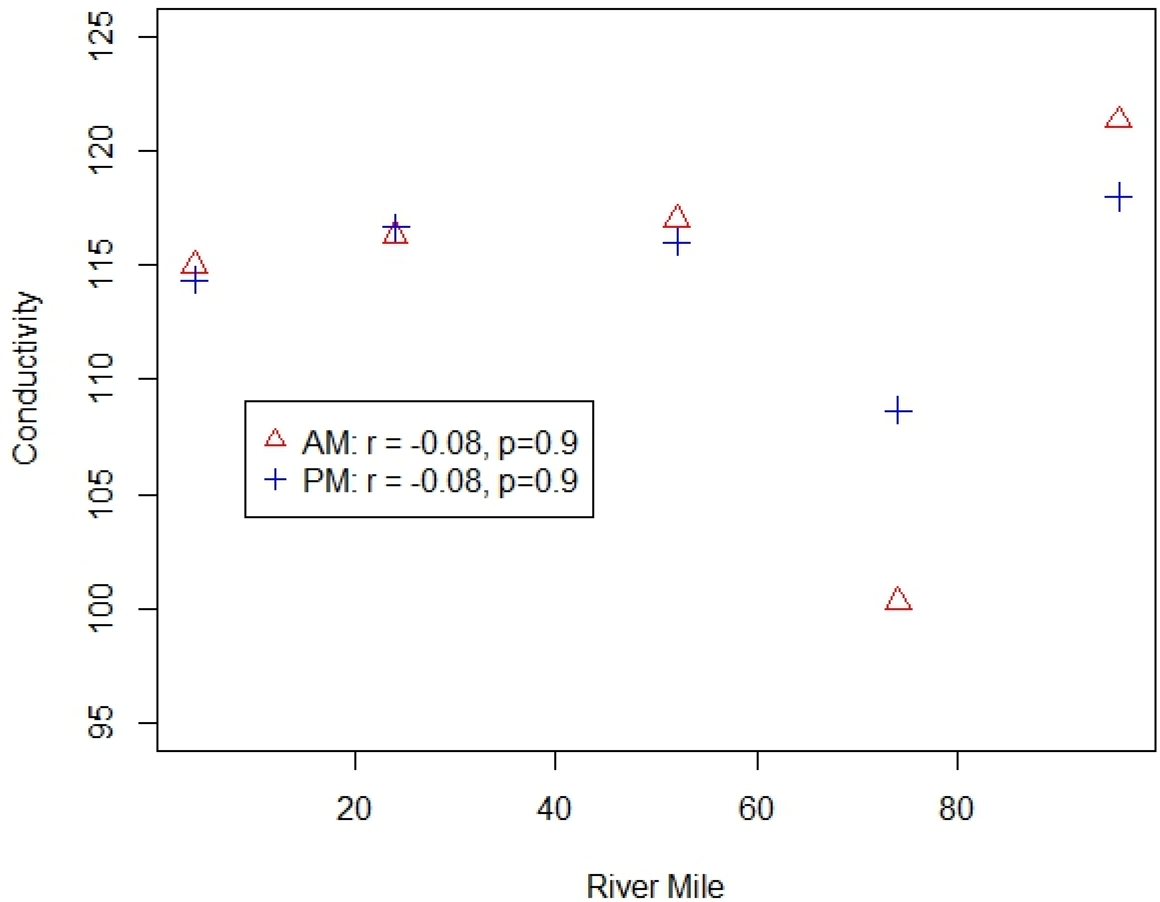


Summary statistics for observed pH time series, and model-estimated trend and R^2 from 13.30 8/5 to 13.45 8/7

River Mile	Mean	Max	Time of Max	Min	Time of Min	Trend (pH units/da)	R^2
4	8.37	9.00	16:30-18:00 08/05, 17:00-17:45 08/06	7.60	5:45-7:00 08/06, 5:15-7:15 08/07	0.0093	0.97
57	8.11	8.60	17:00-19:15 08/05, 17:15-17:45 08/06	7.60	5:45-9:00 08/06, 5:15-9:00 08/07	-0.0361	0.98
100	7.99	8.20	15:15-18:15 08/06,	7.90	1:30-11:45 08/06, 0:30-11:30 08/07	0.0038	0.83

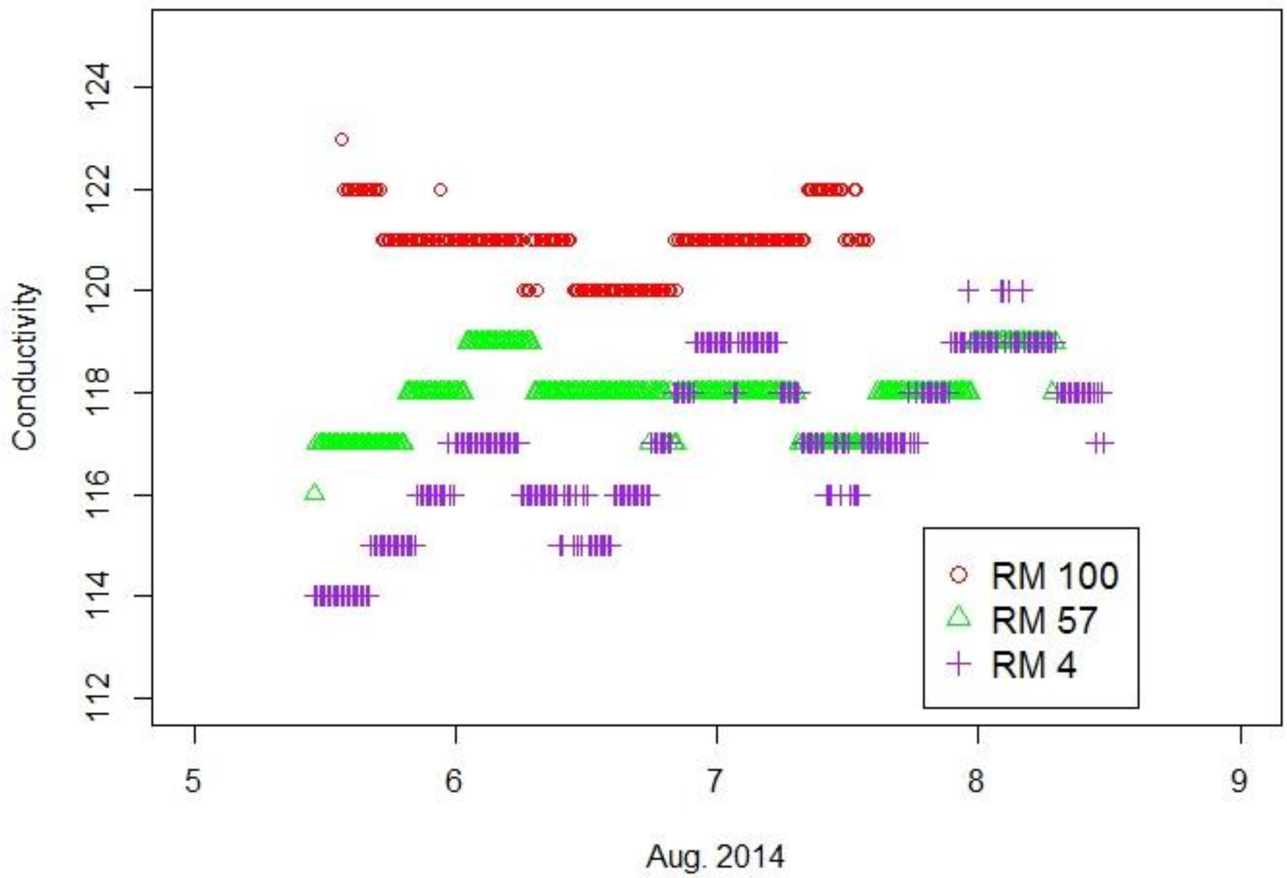
2.3 Conductivity

2.3.1 Grab Samples, July 21-23, 2014: Mean conductivity versus river mile, separately for AM and PM samples. Correlations and their p-values are for a linear trend. Means are listed in Table 1.



Note: There appears to be a slight increasing trend in conductivity with river mile, with RM 74 (Davidson Flat) breaking this trend. This apparent pattern is supported by Spearman rank correlations, which are positive (but nonsignificant) for these data. In contrast, the RM74 conductivities have large influences on linear correlation estimates, resulting in the negative linear correlations reported on the plot.

2.3.2 Conductivity time series from sonde data at 3 sites (RM 4, 57, 100)

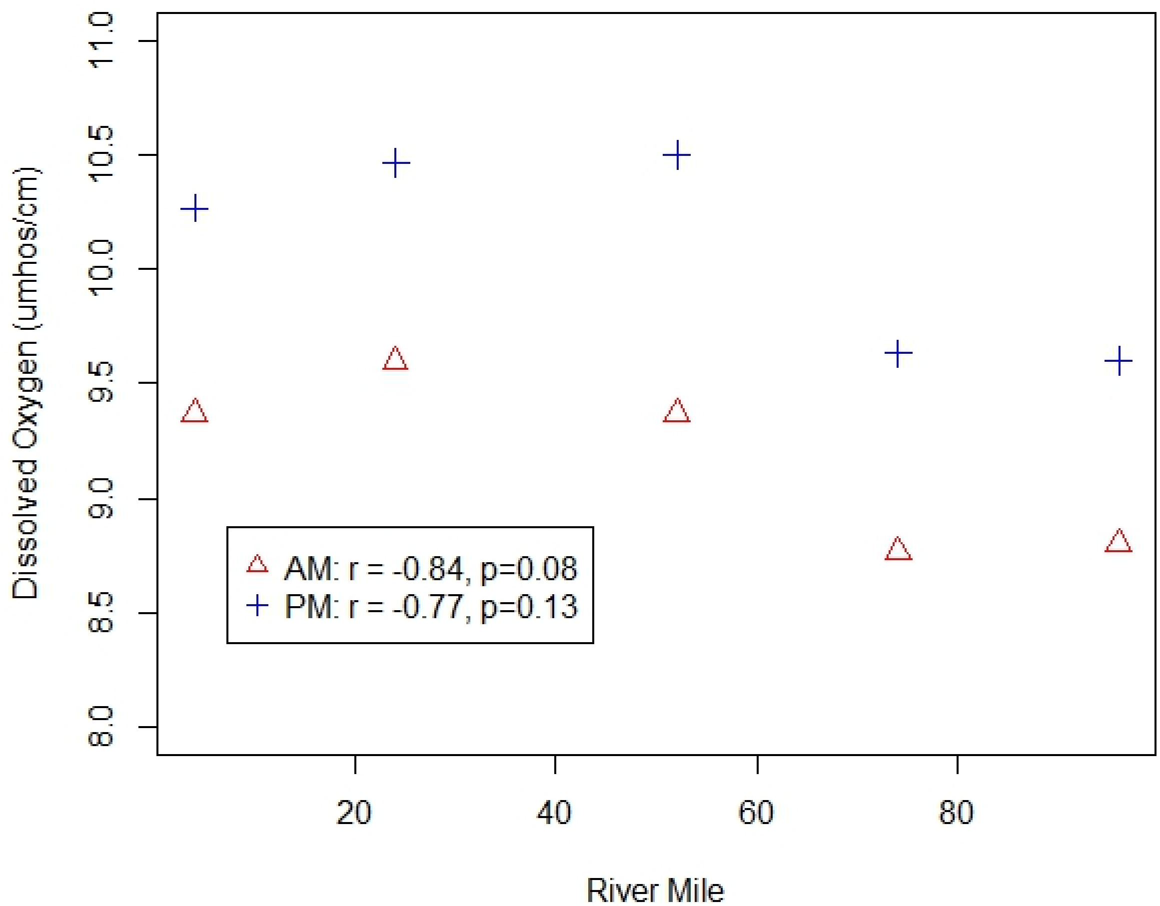


Summary statistics for observed conductivity time series,
from 13:30 8/5 to 13:45 8/7

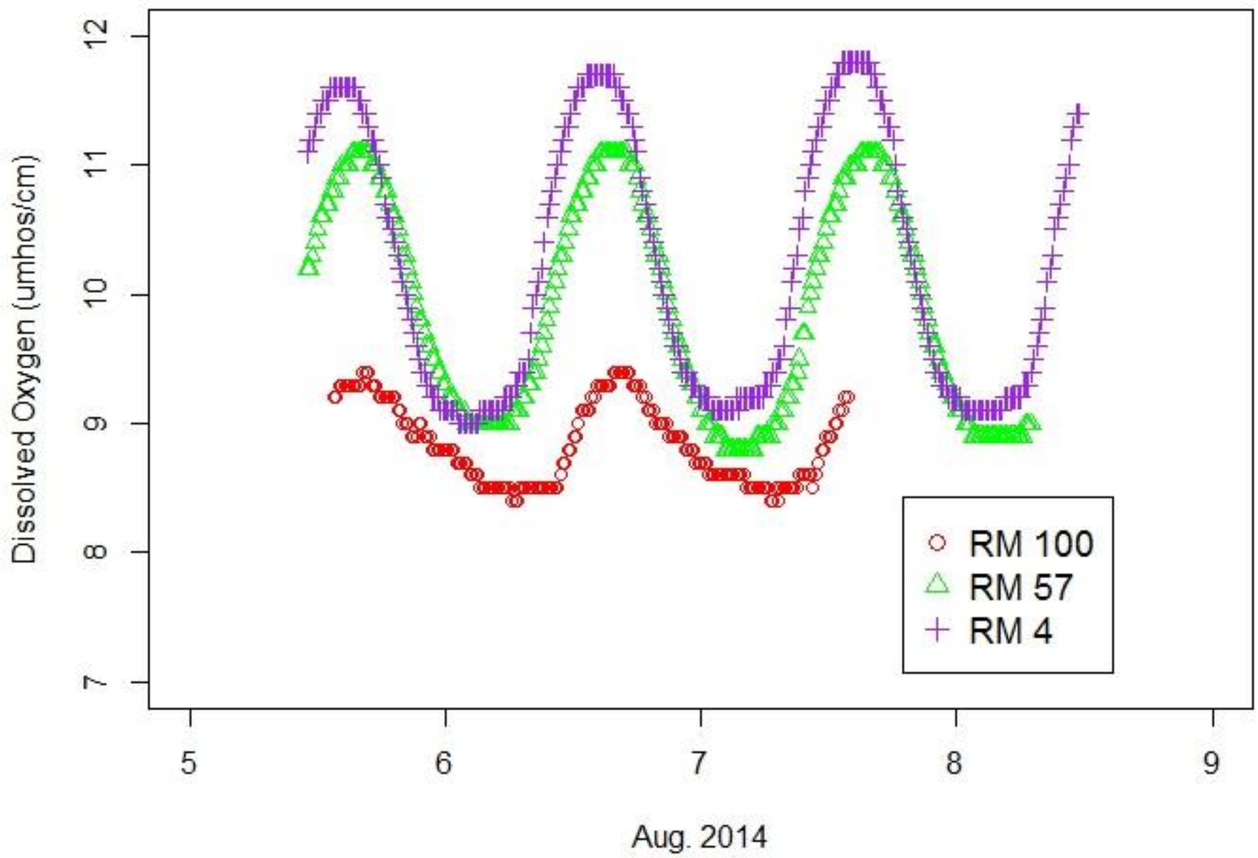
River Mile	Mean	Max	Time of Max	Min	Time of Min
4	116.6	119	22:00-5:30 08/06-07	114	13:30-16:00 08/05
57	117.9	119	1:00-7:00 08/06	117	13:30-22:00 08/06, 7:30-13:45 08/07
100	121	123	1:30 08/05	120	6:15-22:15 08/06

2.4 Dissolved Oxygen (DO)

2.4.1 Grab Samples, July 21-23, 2014: Mean DO versus river mile, separately for AM and PM samples. Correlations and their p-values are for a linear trend. Means are listed in Table 1.



2.4.2 Dissolved oxygen time series from sonde data at 3 sites (RM 4, 57, 100)



Summary statistics for observed DO time series, and model-estimated trend and R^2 from 13:30 8/5 to 13:45 8/7

River Mile	Mean	Max	Time of Max	Min	Time of Min	Trend (umhos/ cm/da)	R^2
4	10.22	11.8	13:15-13:45 08/07	9.0	1:15-3:00 08/06,	0.0381	0.97
57	9.90	11.1	15:15-16:45 08/05, 14:30-17:00 08/06	8.8	2:15-5:00 08/07	-0.0315	0.98
100	8.84	9.4	16:15-16:45 08/05, 15:45-17:15 08/06	8.4	6:15-6:45 08/06, 6:30-7:15 08/07	-0.0312	0.90

2.4 Turbidity

2.4.1 Grab Samples, July 21-23, 2014: Mean turbidity versus river mile, separately for AM and PM samples. Correlations and their p-values are for a linear trend. Means are listed in Table 1.

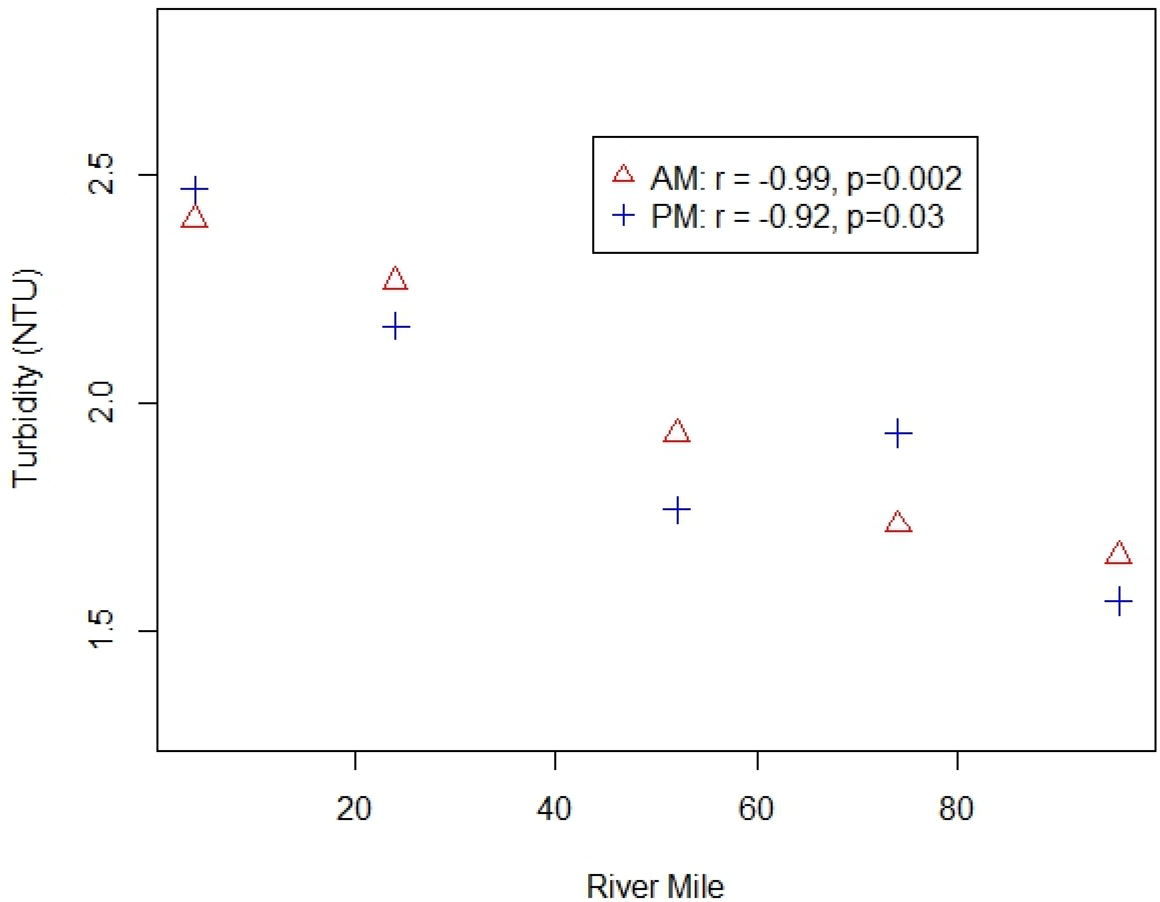


Table 1 – Mean grab-sample concentrations of 5 analytes at 5 sites on the Deschutes River, July 21-23, 2014. Each value is the mean of one sample from each of 3 successive days, taken at approximately the same time (AM or PM) on each day.

Site	Riv_Mile	AM/PM	Temp	pH	SpCond	DO	Turbidity	
WagonBlast	4	AM	16.93	8.27	115.00	9.37	2.40	
WagonBlast	4	PM	18.47	9.03	114.33	10.27	2.47	
Macks Canyon	24	AM	16.33	8.53	116.33	9.60	2.27	
Macks Canyon	24	PM	17.27	8.83	116.67	10.47	2.17	
Oasis	52	AM	15.30	7.73	117.00	9.37	1.93	
Oasis	52	PM	15.57	8.30	116.00	10.50	1.77	
Davidson Flat	74	AM	14.00	8.23	100.33	8.77	1.73	
Davidson Flat	74	PM	15.30	8.93	108.67	9.63	1.93	
Mecca Flats	96	AM	13.67	8.57	121.33	8.80	1.67	
Mecca Flats	96	PM	14.17	8.70	118.00	9.60	1.57	

3. Summary of Results – based on visual inspection of plots and summary statistics.

3.1 Temperature

The mean of July temperature grab samples showed a highly significant, approximately linear decrease moving upstream from RM 4 to RM96, for both AM and PM samples. Temperature time series from August showed clear diurnal variation at all 3 sonde sampling locations, although the diurnal pattern was a clean sinusoid only at RM4.There was also almost no overlap in temperature time series at the 3 locations, consistent with the strong upstream temperature trends seen in the grab samples. Mean temperatures between adjacent sonde locations, over the 2-day period in August, differed by 2.5 to 3 deg. C.

3.2 pH

The grab-sample mean values showed no upstream trends in pH during the July sampling period, either for AM or PM samples. The sonde time series from August at each of the 3 locations were well-fitted by simple sinusoidal models, with very small estimated time trends. The sonde time series had slight upstream decreases in mean pH. However, these differences were quite small relative to the striking diurnal variation in pH at all 3 sites. During each daily cycle, pH ranged over 0.30 units (max-min) at RM100, and this range increased to 1.4 units at RM4. Max and Min pH occurred each day at approximately the same time, at all 3 sampling locations.

3.3 Conductivity

A slight upstream trend of increasing conductivity for the July grab samples is suggested by the plot of AM and PM mean values. However, this pattern was broken by the much-lower grab-sample conductivities sampled at RM 74. The sonde time series from August had higher mean conductivity at RM 57 than at RM4, and at RM100 than at RM57. However, the upward time trend at RM4 suggested a convergence in conductivity with RM57 by the third day of sampling. This example shows that conductivity differences across the sampling locations cannot be safely extrapolated outside the 3 day sampling period. The conductivity time series appeared to vary on a1-day cycle like other analytes. However, the variation was small, especially at RM 57 and 100, relative to the reported measurement precision. As a result, the time series had extended periods of constant values.

3.4 Dissolved Oxygen (DO)

The means of DO grab samples from July 21-23 appeared to generally decline in an upstream direction. However, trend correlation p-values were not convincing. In the August time series, DO values at RM 100 were also consistently lower at than at the 2 downstream locations. Diurnal variation in the 3 time series was sinusoidal, ranging over 2.8 (at RM 4) and 2.3 (at RM 57) umhos/cm. These diurnal ranges were substantially greater than the 0.32 umhos/cm difference between mean DO at RM57 and RM4. Estimated time trends at all three sites were also very small, relative to diurnal variation. The daily max and min DO at RM57 lagged behind the corresponding daily extremes at RM 4 by 1-2 hours.

3.5 Turbidity

Grab-sample mean turbidities from July showed clear and significant linear declines, moving upstream from RM4, for both AM and PM samples. Turbidity measurements were not available in the sonde data.