TECHNICAL REPORT

Technical Review: FDEP DRAFT Evaluation of Waters for Dissolved Oxygen Site Specific Alternative Criteria (SSAC) Development

Paul Julian[1]

[1]Sanibel-Captiva Conservation Foundation, PO Box 839, Sanibel, FL 33957

Document History
Compiled May 14, 2021

Summary
The Florida Department of Environment Protection (FDEP or Department) evaluated several waterbodies that have been placed in the assessment category 4c for the potential development of site-specific alternative criteria (SSAC). Waterbodies in this assessment category are impaired for one or more criteria or designated uses but do not require TMDL development because the impairments are not caused by a pollutant. Based on the Department's assessment they recommend the adoption of Type II dissolved oxygen (DO) SSAC for eight waterbodies. For purposes of this review, three of the eight proposed DO SSACs were evaluated. The three waterbodies evaluated in this report include Daughtrey Creek (WBID 3240F), Popash Creek (WBID 3240Q), and Cypress Creek (WBID 3235C) and were selected due to the proximity to the Caloosahatchee River. The Impaired Waters Rule (IWR) database was retrieved from the FDEP webpage and evaluated based on methods outlined in the Department's report. The Department evaluated trends in DO saturation levels using the parametric linear model approach. In this report, additional analysis was performed to evaluate the suitability of applying parametric linear models to evaluate trends and Kendall trend analysis from an overall waterbody and site-specific perspective. Overall, the resulting linear models presented by the Department did not fit the assumptions of the statistical test. Moreover, when proper trend analysis was performed notable declining trends in DO saturation levels were detected for Daughtrey and Popash Creeks at both the overall waterbody and individual station level. In the derivation of the DO SSACs for these waterbodies, the Department used a parametric percentile approach with minimal documentation, justification of using the 10th percentile, or verification that the data conform to the assumptions of the statistical test. Based on the assessment of the data, the data is not normally distributed therefore a parametric approach should not be taken, doing so could increase the chances of Type II error. Given significantly declining trends at most monitoring locations, data not fitting the assumptions of the parametric percentile approach and some discrepancies between available and presented data additional and/or alternative analyses are recommended to ensure any proposed criteria account for the natural variability within the waterbodies while reducing both Type I and II error rates.

Open Source Information
https://github.com/SwampThingPaul/SFL_DOSSAC/

CONTACT Paul Julian. Email: pjulian@sccf.org

1. Purpose of this Report

The purpose of this report is to evaluate proposed Florida Department of Environmental Protection (FDEP) Dissolved Oxygen (DO) Site Specific Alternative Criteria (SSACs) for select streams listed in the assessment category 4c. In February 2021, FDEP through the triennial review process, proposed DO SSAC for eight waterbodies (FDEP 2021) This analysis focuses on three of the eight waterbodies located in Charlotte, Lee, Glades and Hendry counties and includes Daughtrey (*WBID 3240F*), Popash (*WBID 3240Q*) and Cypress (*WBID 3235C*). In addition to evaluating the proposed SSACs for the waterbodies listed above, the data was also evaluated for an alternative SSAC.

Below are specific comments regarding FDEPs technical support document (FDEP 2021):

1. While significant attention was dedicated to background information regarding surface water quality standards, SSACs, and generally applicable DO criteria (Section 1 of the report) very little was dedicated as to why the 10th percentile approach was selected. The report states *"To account for the natural variability within the waterbodies, the final proposed SSACs were calculated as the 10th percentiles of the DO saturation levels."*. This ultimately sets the criterion at a very low value well below the central tendency of the data (mean or median), muting the natural variability. In most cases, average DO saturation levels across the period of record are above that of the current DO criterion with only a small portion of the samples dropping below the current criterion during the selected period of record.

2. Additional information regarding how the 10th percentile was calculated (i.e. spatially averaged versus raw daily DO values) would be helpful in understanding the applicability of the proposed SSAC. Moreover, in the methods section, the analysis assumes a normal distribution of the data but does not verify the data distribution for any of the waterbodies in question. To reduce Type II error it is recommended to include an evaluation of data distribution with the appropriate statistical analysis to verify the data fits the assumptions of the test.

3. Presented for each waterbody a time-series of DO was presented with a linear model (and R^2) as evidence of a trend. Typically the period or record presented is different from the assessment period of record. Furthermore, least-squared (parametric) linear models are not suitable for trend analysis due to the statistical method typically violates the assumptions of parametric linear models. Trend analysis typically violates the assumption of independence of model residuals, where the residuals of the model should lack autocorrelation (Helsel and Hirsch 2002). Analysis presented below suggests that trends reported are different from the trend in the data using suitable statistical analyses.

4. More transparency is needed in what monitoring locations were used beyond just points on a map. Furthermore, some clarity of screening protocols and/or data inclusion criteria is needed as some monitoring locations have very limited data (i.e. one sample in the period of record).

5. Prior to development of the proposed DO SSACs it is recommend that an evaluation of the numeric nutrient criteria (NNC) be assessed during the period of record to ensure that nutrient conditions are not contributing the potential degredation of DO conditions within the waterbody.

2. Methodology

2.1. *Data Source*

Water quality data collected for each waterbody of interest during the period from 2006 through 2020 were retrieved from the FDEP Impaired Water Rule (IWR) database (version Run 60; `http://publicfiles.dep.state.fl.us/DEAR/IWR/`).The IWR database and assessment only includes ambient water quality of surface water within a given waterbody. A station will not be included if it samples wastewater effluent or if it is a rain gauge, or it samples groundwater.

2.2. *Data Screening and Handling*

Water quality data were screened based on laboratory qualifier codes, consistent with the FDEP Quality Assurance Rule (Chapter 62-160, F.A.C.). Any datum associated with a fatal qualifier (e.g., H, J, K, N, O, V, Q, Y, G, or ?) indicating a potential data quality problem was removed from the analysis. Values that exceeded possible physical or chemical measurement constraints (e.g., if resulting pH is greater than 14), had temperatures well outside seasonal norms (e.g., 6 degrees Celsius [° C] in July), or represented data transcription errors were excluded. For field parameters, measurements collected at multiple depths at the same location on the same day were considered one sample, with the arithmetic mean used to represent the vertical profile.

Additional considerations in the handling of water quality data are the accuracy and sensitivity of the laboratory method used. For purposes of summary statistics presented in this chapter, data reported as less than the MDL were assigned a value of one-half the MDL unless otherwise noted. All data presented in this chapter, including historical results, were handled consistently with regard to screening and MDL replacement.

2.3. *Statistical Analysis*

To verify FDEP's original analysis, trend analysis was performed consistent with FDEP's method of using a linear model to evaluate trend on the spatially averaged DO time-series across the each WBID during the analysis period of record. Model residuals were evaluated for autocorrelation using the Breucsch-Godfrey test (`bgtest` in the `lmtest` R-package; Zeileis and Hothorn 2002) and compared to a normal distribution using the Shapiro-Wilk Normality Test. Global Validation of Linear Models Assumptions (`gvlma` R-package; Peña and Slate 2006) was also used to evaluate if the model fit the assumptions of linear models. Additional trend analysis was performed using Kendall's trend test on the spatially averaged DO time-series and individual stations with greater than 20 samples per time-series using sampling date as the time index. Kendall's seasonal trend test (`kendallSeasonalTrendTest` in the `EnvStats` R-package; Millard 2013) was also applied to spatially averaged monthly data across each waterbody using month as season.

To verify that the data fit the assumptions of the parametric percentile approach, the distribution of the data was compared to a normal distribution using the Shapiro-Wilk, and Anderson-Darling (`ad.test` in the `nortest` R-package; Gross and Ligges 2015) tests. Other tests of normality (i.e. Kolmogorv-Smirnov, Lilliefors, etc.) were initially considered but due to the statistical power of the Shapiro-Wilk and Anderson-Darling tests (Mohd Razali and Yap 2011) and to avoid confusion only these tests were used

to evaluate the data. Additionally, empirical and theoretical data distribution plots (`plotdist` in the `fitdistrplus` R-package; Delignette-Muller and Dutang 2015) were generated to qualitatively evaluate the data.

All statistical operations were performed using R (Ver 4.0.4, R Foundation for Statistical Computing, Vienna Austria). Unless otherwise stated, all statistical operations were performed using the base R library. The critical level of significance was set at $\alpha = 0.05$. All R code for the analyses outline above can be found in **Appendix A** and source files for the analysis and this report have been posted to https://github.com/SwampThingPaul/SFL_DOSSAC/.

3. Results

3.1. *Daughtrey Creek (WBID 3240F)*

Within the IWR database, WBID 2340F has a total of 12 sites with DO data during the period of record (Fig 1; **Appendix B** Table S1). In the FDEP report, site 112WRD 264140081494600 was excluded from the analysis, presumably due to the proximity of the site to the Caloosahatchee River. In evaluating the current data, slight differences in summary statistics between the FDEP's report (FDEP 2021) and this summary is noted below for DO in Table 1. Over the analysis period of record, three of the 12 monitoring locations consistently exceed the established DO stream water quality criterion (Table 2).

Table 1. Summary statisic of dissolved oxygen (DO) saturation for Daughtrey Creek (WBID 3240F) during the period of record (Calendar Year: 2006 - 2020) as reported by FDEP (2021) and summaried data (this report)

Parameter	Source	Count	Average	Std Dev	10th %tile	25th %tile	Median	75th %tile
DO Saturation, %	FDEP 2021	822	45.0	19.1	18.2[1] / 20.5[2]	32.2	45.4	58.9
	This Analysis	819	45.1	19.0	18.5[1] /20.7[2]	32.5	45.4	59.0

Note:

Std Dev=Standard Deviation; %tile=Percentile

[1] Percentiles based on ranking of data

[2] 10th percentile based on normal distribution using mean and standard deviation

Table 2. Number of years where average dissolved oxygen (DO) saturation values exceed the stream time-of-day DO water quality criteria more than 10% of the time during the analysis period of record for monitoring locations within wBID 3240F.

Station ID	Number of Years WQS Exceeded	Number of Years with Data
21FLEECO20-29GR	12	15
21FLEECO20-9GR	7	15
21FLEECO20A-11GR	8	15
21FLEECO20A-19GR	15	15
21FLEECOAB96009	0	1
21FLEECOGATRGR91	11	15
21FLFTM 28020037	1	1
21FLFTM 28020231	0	2
21FLFTM CALUSA0030FTM	1	1
21FLFTM G3SD0164	0	1
21FLGW 42393	0	1
21FLWQSPLEE634US	0	1

3.1.1. Trend Analysis

In FDEP's report, the Department notes that *"A statistically significant decreasing trend in DO levels was not found..."*. No statistical results were reported to accompany this statement and the only evidence of a trend was reported as a linear regression and R^2 value in Figure 18 of the report. Furthermore, the caption of this figure includes a statement of *"The regression equation and r2 value provided in the inset are from simple linear regression analysis and demonstrate no significant decreasing trend in DO levels."* to justify this trend. Moreover this figure includes data outside the reported period of record. Trend analysis by a linear model typically violates the assumption of independence of model residuals, where the residuals of the model should lack autocorrelation (Helsel and Hirsch 2002).

Using the linear model approach of trend analysis (Fig 2), the model had a low degree of fit (R^2 = 0.01), significantly auto-correlated residuals (Statistic = 27.4, df = 3, ρ <0.01; Fig 3) and the residuals of the model was significantly different from a normal distribution (W=0.99, ρ <0.01). The global validation of linear models assumptions also indicate that the link function, skewness and overall assumptions were not satisfied (Global Stat = 20.37, ρ <0.001). When this same data is evaluated using the Kendall's correlation test it results in a significant monotonic decline in DO across the WBID (τ = -0.07, ρ = <0.05). At individual monitoring locations, of the five monitoring locations with sufficient data, two locations had significantly declining trends while three had no significant trend (Table 3). From a seasonal perspective the DO saturation levels also significantly declined (τ = -0.14, ρ <0.05) at a rate of -0.31 % DO Yr^{-1} over the period of record. The significantly declining trends within and across this waterbody should be taken into considered when establishing any water quality criterion. Given these results, setting a DO SSAC for this waterbody could potentially perpetuating a declining existing condition.
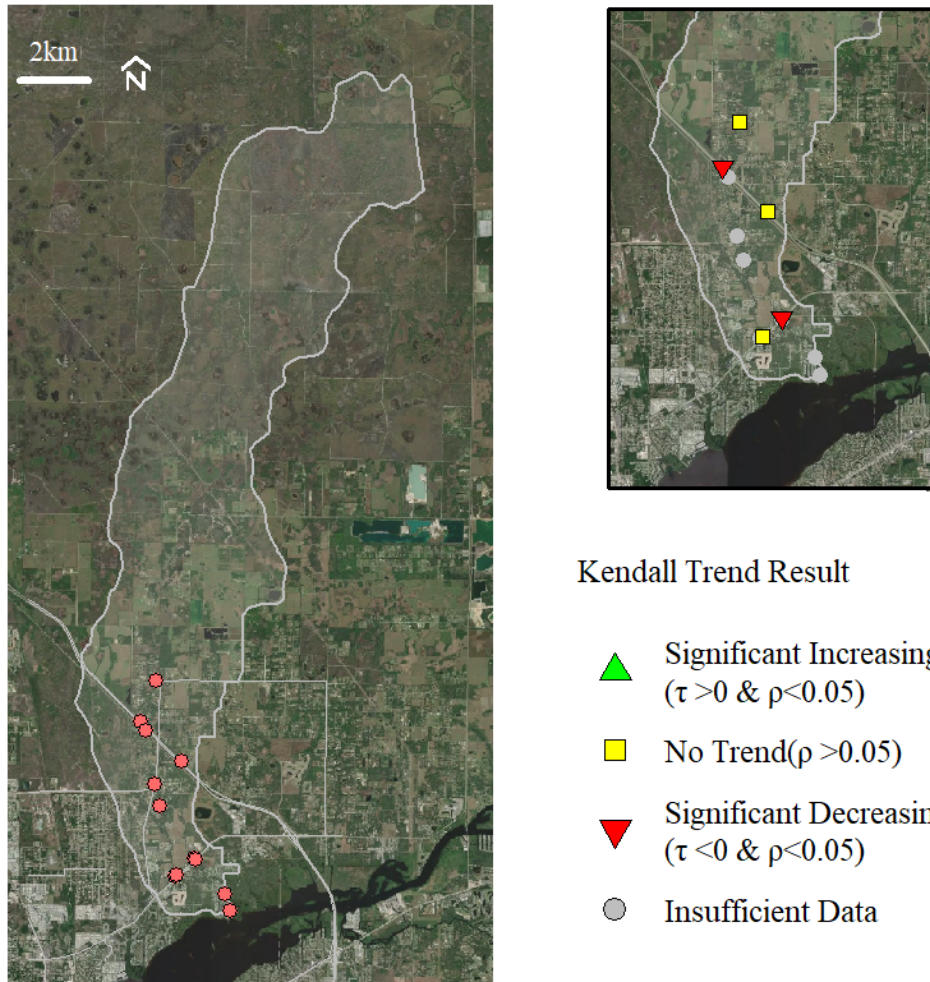
5

Figure 1. Monitoring locations and individual station trend results within WBID 3240F - Daughtrey Creek.
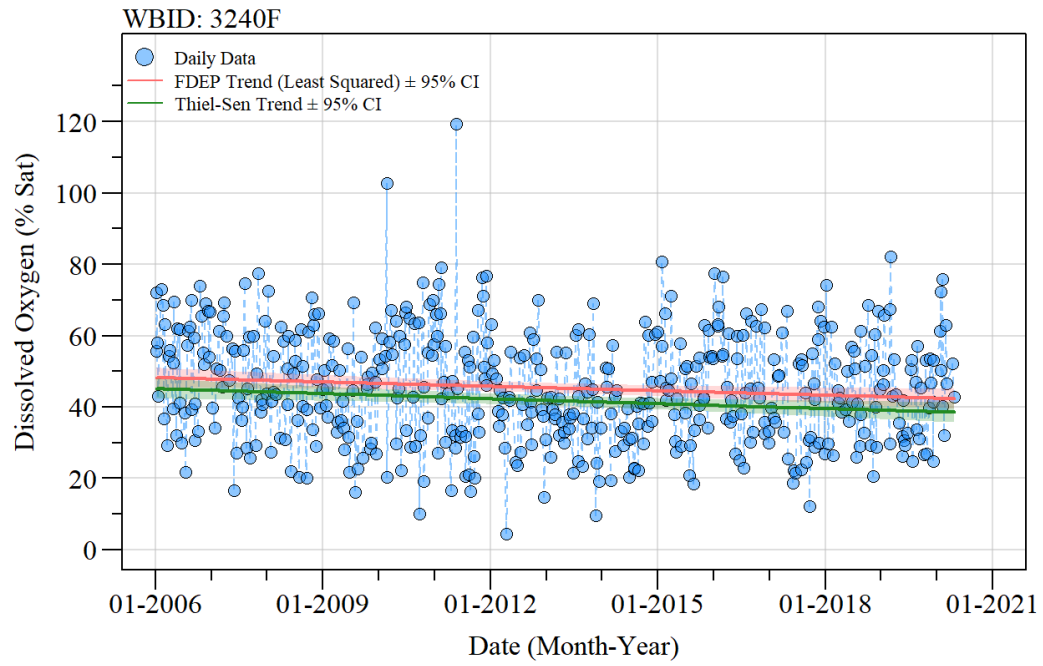
Figure 2. Spatially average daily dissolved oxygen percent saturation from Janurary 2006 till December 2020 for WBID 3240F - Daughtrey Creek. Both least squared (FDEP trend analysis) and Thiel-Sen trends are displayed.
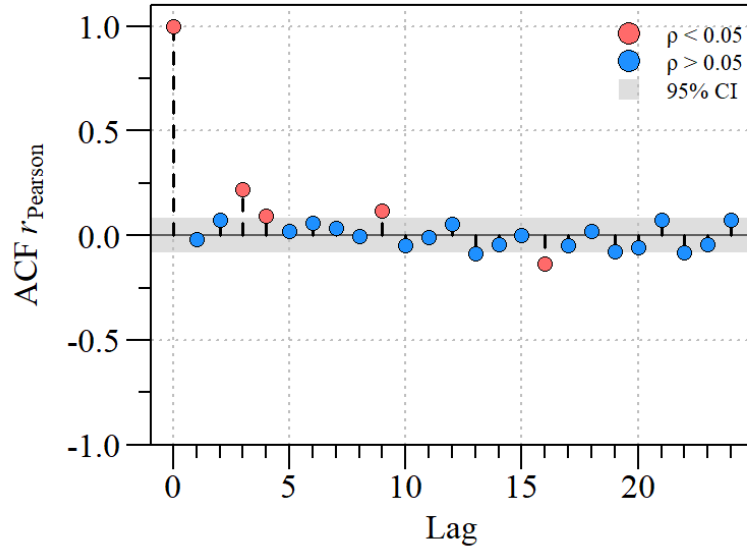

Figure 3. Autocorrelation function (ACF) of least squared trend model residual for the spatially averaged dissolved oxygen time-series. WBID 3240F - Daughtrey Creek.

Table 3. Kendall's trend test results for each individual moinitoring locations with suitable data in WBID 3240F - Daughtrey Creek

Station ID	Sample Size	Kendall's τ	ρ -value
21FLEECO20-29GR	158	-0.09	0.10
21FLEECO20-9GR	170	-0.02	0.64
21FLEECO20A-11GR	159	-0.14	<0.01
21FLEECO20A-19GR	148	-0.03	0.56
21FLEECOGATRGR91	162	-0.19	<0.01

3.1.2. Data Distribution

The assumption of using a parametric percentile approach, as proposed by FDEP (2021) is that the data must be normally distributed. Both Shapiro-Wilk (W=0.99, $\rho<0.001$) and Anderson-Darling (A=1.22, $\rho<0.01$) test indicated that the data was significantly different from a normal distribution. Moreover, quantile-quantile and density histogram (Fig 4) indicate a slight divergence from the normal distribution. Given that the data does not fit a normal distribution, application of the parametric 10th percentile could result in increased Type-II error and an alternative approach considered.
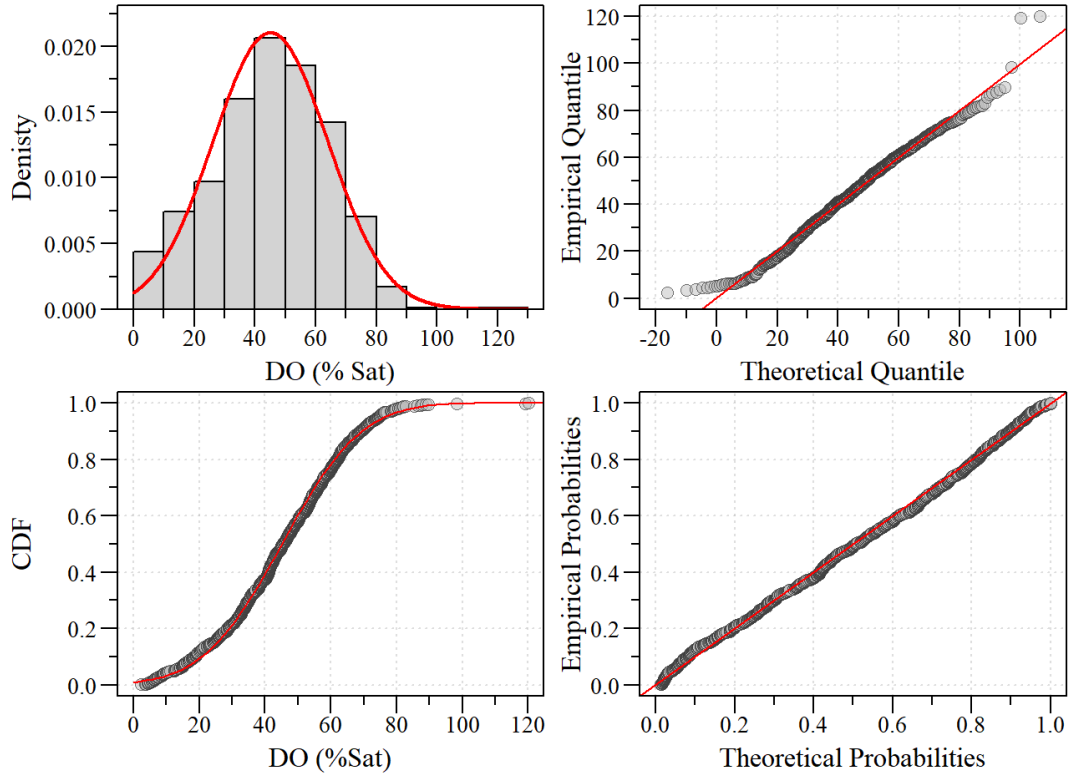

Figure 4. Evaluation of dissolved oxygen empirical and theoretical normal distribution for Daughtrey Creek (WBID 3240F) during the period of record. Top left: Density histogram of the data compared to the theoretical distribution (red line); Top right: quantile-quantile (Q-Q) plot; Bottom left: cumulative distribution function (CDF) relative to theoretical normal distribution CDF (red line); Bottom right: probability-probability (P-P) plot.

3.2. *Popash Creek (WBID 3240Q)*

Within the IWR database, WBID 2340Q has a total of 12 sites with DO data during the period of record (Fig 5; **Appendix B** Table S2). In evaluating the current data, slight differences in summary statistics between the FDEP's report (FDEP 2021) and this summary is noted below for DO in Table 4. The difference between FDEP's report and this summary is the inclusion of four additional samples which slightly change the summary statistics over the period of record. Over the analysis period of record, two of the 12 monitoring locations consistently exceed the established DO stream water quality criterion (Table 5).

Table 4. Summary statisic of dissolved oxygen (DO) saturation for Popash Creek (WBID 3240Q) during the period of record (Calendar Year: 2006 - 2020) as reported by FDEP (2021) and summaried data (this report)

Parameter	Source	Count	Average	Std Dev	10th %tile	25th %tile	Median	75th %tile
DO	FDEP 2021	345	45.6	18.7	20.2^1 / 21.6^2	31.6	46.3	58.5
Saturation, %	This Analysis	349	45.4	17.7	21.0^1 / 22.7^2	31.8	46.7	58.0

Note:

Std Dev=Standard Deviation; %tile=Percentile

[1] Percentiles based on ranking of data

[2] 10th percentile based on normal distribution using mean and standard deviation

Table 5. Number of years where average dissolved oxygen (DO) saturation values exceed the stream time-of-day DO water quality criteria more than 10% of the time during the analysis period of record for monitoring locations within wBID 3240Q.

Station ID	Number of Years WQS Exceeded	Number of Years with Data
21FLEECO23-27GR	15	15
21FLEECO23-5GR	10	15
21FLEECOAC26956	0	1
21FLEECOAD03858	0	1
21FLFTM 28020038	0	1
21FLFTM 28020232	0	2
21FLFTM CALUSA0020FTM	NA	1
21FLFTM G3SD0079	0	3
21FLGW 45794	1	1
21FLGW 51902	0	1
21FLGW 53944	0	1
21FLGW 56327	0	1

3.2.1. *Trend Analysis*

Similar to Daughtrey creek, the department justified no declining trend in DO saturation levels using the linear model approach. Using the linear model approach of trend analysis (Fig 6), the model had a low degree of fit ($R^2 = 0.06$), significantly auto-correlated residuals (Statistic = 48.3, df = 3, $\rho < 0.001$; Fig 7) and the residuals of the model was significantly different from a normal distribution (W=0.99, $\rho < 0.01$). The global validation of linear models assumptions analysis also indicate that the link function, residual kurtosis and overall assumptions were not satisfied (Global Stat = 13.32, $\rho < 0.01$). When this same data is evaluated using the Kendall's correlation test

it results in a significant monotonic decline in DO across the WBID (τ = -0.14, ρ <0.001). At individual monitoring locations, of the two monitoring locations with sufficient data, both had significantly declining trends while (Table 6). From a seasonal perspective the DO saturation levels also significantly declined (τ = -0.21, ρ <0.001) at a rate of -0.90 % DO Yr^{-1} over the period of record. From a visual perspective, DO saturation levels were much more variable and elevated early in the analysis period of record when comparing later in the period of record(Fig \ref{fig:fig6}). Furthermore, Figure 21 presented by FDEP 2021, with an extended period of record accentuates this change in DO condition. The significantly declining trends within and across this waterbody and the number of sites with a large enough sample size should be taken into considered when establishing any water quality criterion. Given these results, setting a DO SSAC for this waterbody could potentially perpetuating a declining existing condition.
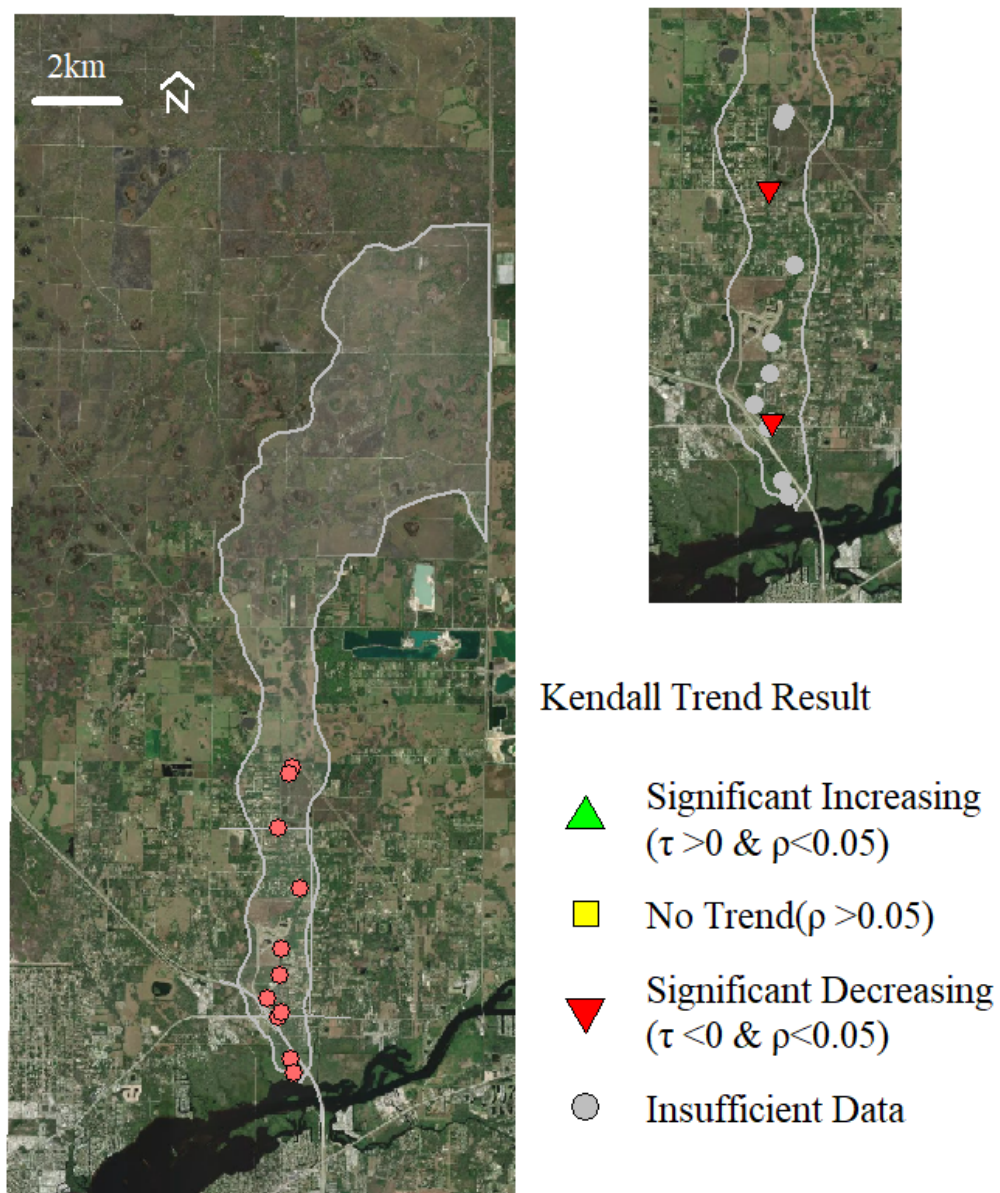
Figure 5. Monitoring locations and individual station trend results within WBID 3240Q - Popash Creek.
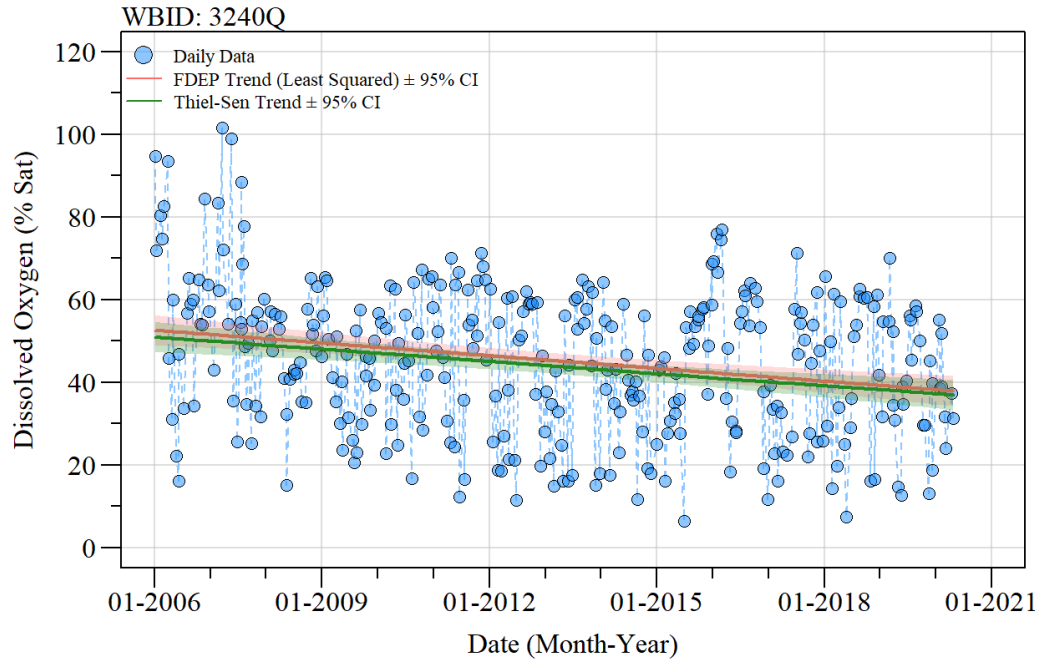
Figure 6. Spatially average daily dissolved oxygen percent saturation from Janurary 2006 till December 2020 for WBID 3240Q - Popash Creek. Both least squared (FDEP trend analysis) and Thiel-Sen trends are displayed.
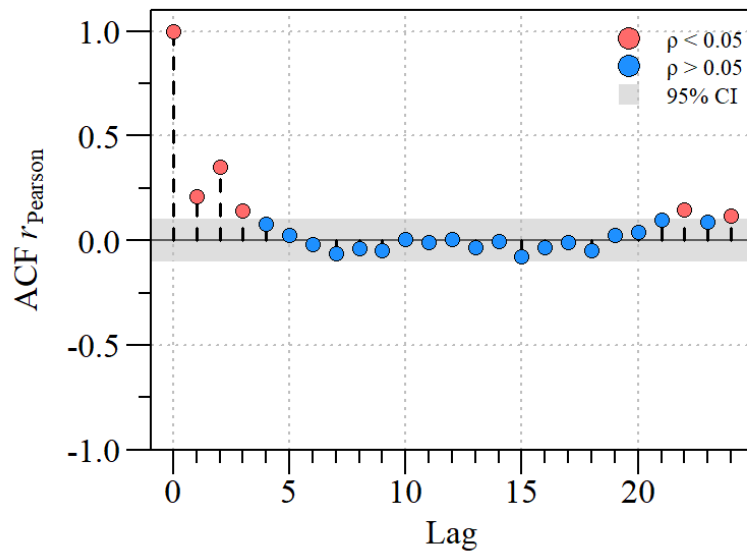

Figure 7. Autocorrelation function (ACF) of least squared trend model residual for the spatially averaged dissolved oxygen time-series. WBID 3240Q - Popash Creek.

Table 6. Kendall's trend test results for each individual monitoring locations with suitable data in WBID 3240Q - Popash Creek

Station ID	Sample Size	Kendall's τ	ρ -value
21FLEECO23-27GR	165	-0.17	<0.01
21FLEECO23-5GR	170	-0.14	<0.01

3.2.2. Data Distribution

Both Shapiro-Wilk (W=0.99, ρ<0.001) and Anderson-Darling (A=1.74, ρ<0.001) test indicated that the data was significantly different from a normal distribution. Moreover, quantile-quantile, cumulative distribution function and density histogram (Fig 8) indicate a divergence from a normal distribution. Given that the data does not fit a normal distribution, application of the parametric 10th percentile could result in increased Type-II error and an alternative approach considered.
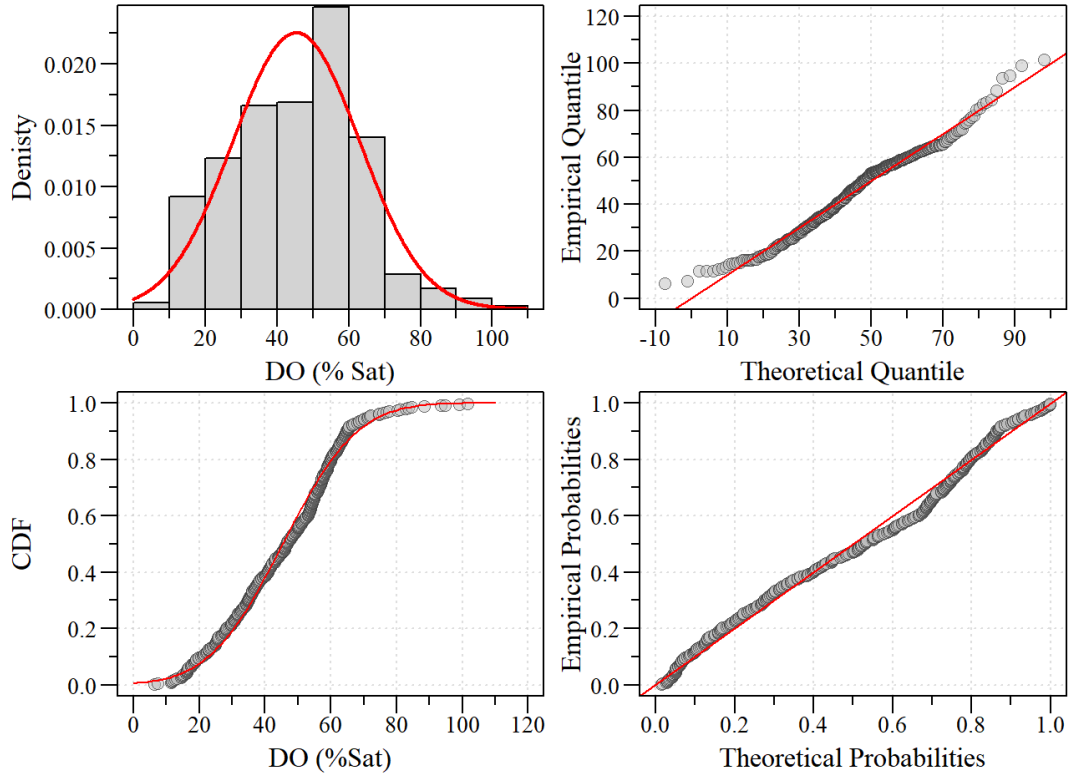

Figure 8. Evaluation of dissolved oxygen empirical and theoretical normal distribution for Popash Creek (WBID 3240Q) during the period of record. Top left: Density histogram of the data compared to the theoretical distribution (red line); Top right: quantile-quantile (Q-Q) plot; Bottom left: cumulative distribution function (CDF) relative to theoretical normal distribution CDF (red line); Bottom right: probability-probability (P-P) plot.

3.3. *Cypress Creek (WBID 3235C)*

Within the IWR database, WBID 3235C has a total of 11 sites with DO data during the period of record, seven of which are located within Cypress Creek (Fig 9; **Appendix B** Table S3). However, Figure 23 in the FDEP report only indicates potentially four monitoring locations. Given these data discrepancies Table 7 includes all available data within the waterbody compared to values reported by FDEP (FDEP 2021). Given the disparity between what is presented in the report and what is available in the IWR database, it is unclear as to what data was used to derive the proposed DO SSAC for Cypress Creek (WBID 3235C). Over the analysis period of record given the limited data, two of the 11 monitoring locations consistently exceed the established DO stream water quality criterion. The remaining monitoring locations within Cypress Creek proper and locations across the waterbody had limited annual data with the exception of 21FLEECOSPANISHGR (Table 8).

Table 7. Summary statisic of dissolved oxygen (DO) saturation for Cypress Creek (WBID 3235C) during the period of record (Calendar Year: 2006 - 2020) as reported by FDEP (2021) and summaried data (this report)

Parameter	Source	Count	Average	Std Dev	10th %tile	25th %tile	Median	75th %tile
DO Saturation, %	FDEP 2021	379	48.6	15.8	28.6[1] / 28.4[2]	37.0	49.4	59.5
	This Analysis	530	53.4	17.8	31.9[1] /30.7[2]	40.3	53.3	65.0

Note:

Std Dev=Standard Deviation; %tile=Percentile

[1] Percentiles based on ranking of data

[2] 10th percentile based on normal distribution using mean and standard deviation

Table 8. Number of years where average dissolved oxygen (DO) saturation values exceed the stream time-of-day DO water quality criteria more than 10% of the time during the analysis period of record for monitoring locations within wBID 3235C.

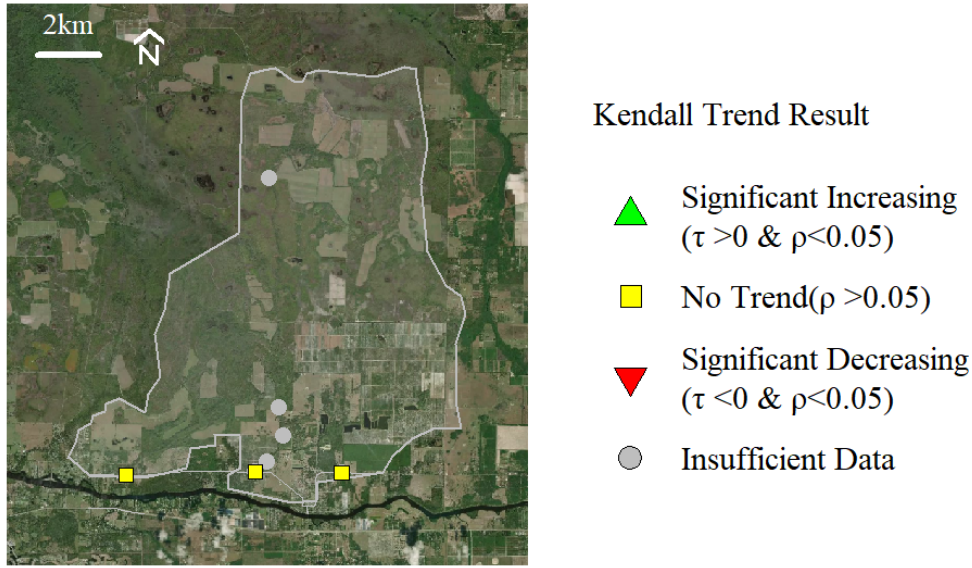| Station ID	Number of Years WQS Exceeded	Number of Years with Data
21FLBABRCYPRESS_HEAD	3	3
21FLBABRCYPRESS_OUTFLOW	3	4
21FLEECOCYPRESSGR	10	15
21FLEECOFICHTERSGR	11	15
21FLEECOSPANISHGR	1	15
21FLFTM 28020043	0	1
21FLFTM 28020237	0	2
21FLFTM CYPRESSGR	1	1
21FLFTM G3SD0084	0	2
21FLGW 53948	0	1
21FLGW 56335	0	1

Figure 9. Monitoring locations and individual station trend results within WBID 3235C - Cypress Creek.

3.3.1. Trend Analysis

Using the linear model approach of trend analysis (Fig 10), the model had a low degree of fit ($R^2 = 0.02$) and significantly auto-correlated residuals (Statistic = 17.6, df = 3, $\rho < 0.001$; Fig 11). While the residuals of the trend model were normally distribution (W=0.99, $\rho = 0.68$), the model did not meet the other assumptions of linear models based on the global validation of linear models assumptions analysis (Global Stat = 14.80, $\rho < 0.01$). When evaluated using the Kendall's correlation test no monotonic trend in DO across the WBID was detected ($\tau = 0.09$, $\rho = 0.05$) or individual monitoring locations (Table 9 and Fig 9). From a seasonal perspective the DO saturation levels had a borderline significantly increasing trend ($\tau = 0.12$, $\rho = 0.05$). These results are for monitoring locations across the waterbody and not Cypress Creek specifically. However, one monitoring location within Cypress Creek (21FLEECOCYPRESSGR) did not have a statistically significant monotonic trend (Table 9). Given these results, setting a DO SSAC for this waterbody could be feasible given the potentially improving seasonal or a lack of a monotonic trend in DO saturation levels across the waterbody.

Table 9. Kendall's trend test results for each individual mointoring locations with suitable data in WBID 3235C - Cypress Creek

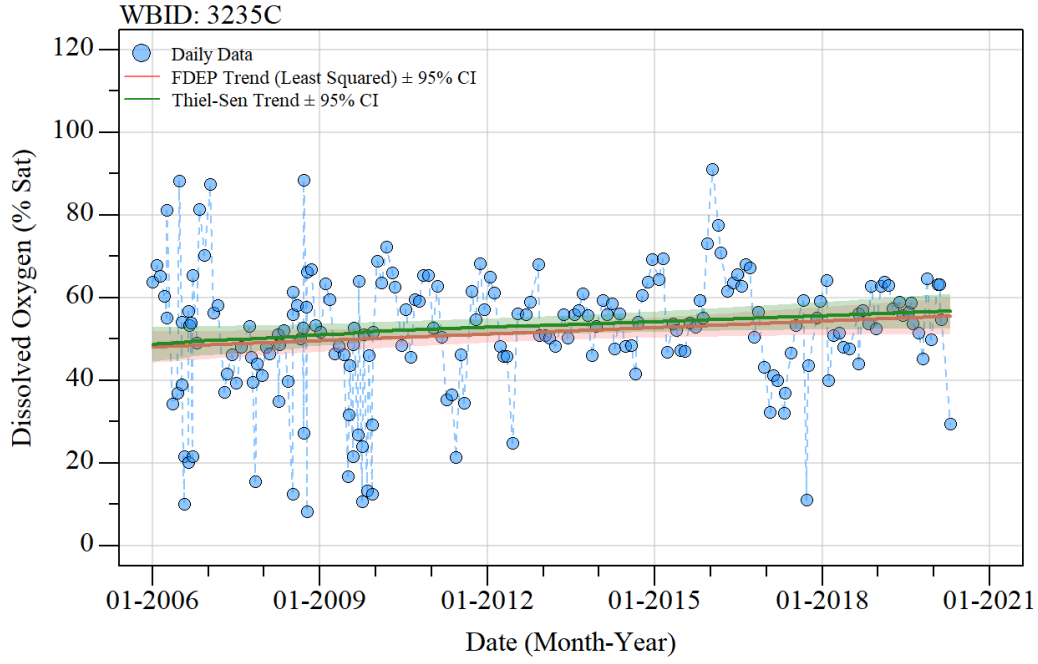| Station ID	Sample Size	Kendall's τ	ρ -value
21FLEECOCYPRESSGR	165	0.03	0.53
21FLEECOFICHTERSGR	166	0.01	0.85
21FLEECOSPANISHGR	152	0.06	0.29

Figure 10. Spatially average daily dissolved oxygen percent saturation from Janurary 2006 till December 2020 for WBID 3235C - Cypress Creek. Both least squared (FDEP trend analysis) and Thiel-Sen trends are displayed.
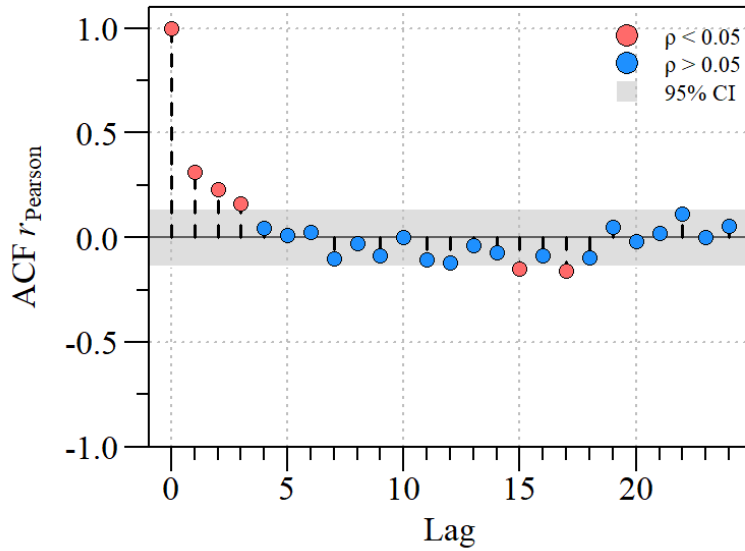
Figure 11. Autocorrelation function (ACF) of least squared trend model residual for the spatially averaged dissolved oxygen time-series. WBID 3235C - Cypress Creek.

3.3.2. Data Distribution

Including all the data across the waterbody, DO saturation levels are not significantly different from a normal distribution based on the Shapiro-Wilk (W=1.00, ρ=0.47) and Anderson-Darling (A=0.37, ρ=0.43) normality tests (Fig 12). However, if only

16

monitoring locations within Cypress Creek (See **Appendix B**, Table S3) were used, DO saturation levels are significantly different from a normal distribution (Shapiro-Wilk: W=0.99, ρ<0.05; Anderson-Darling: A=1.00, ρ<0.05; Fig 13). If monitoring locations within Cypress Creek proper were considered to develop the DO SSAC, the data do not fit a normal distribution, application of the parametric 10th percentile could result in increased Type-II error and an alternative approach considered.

Figure 12. Evaluation of dissolved oxygen empirical and theoretical normal distribution for all monitoring locations across the Cypress Creek waterbody (WBID 3235C) during the period of record. Top left: Density histogram of the data compared to the theoretical distribution (red line); Top right: quantile-quantile (Q-Q) plot; Bottom left: cumulative distribution function (CDF) relative to theoretical normal distribution CDF (red line); Bottom right: probability-probability (P-P) plot.
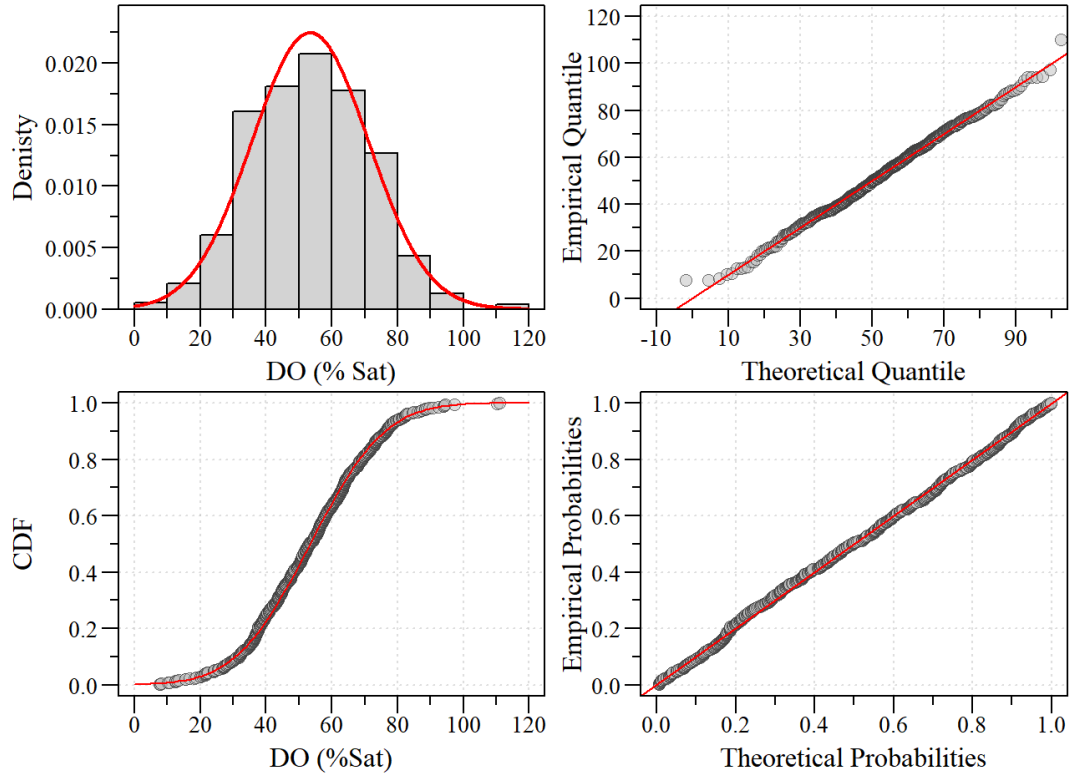
Figure 13. Evaluation of dissolved oxygen empirical and theoretical normal distribution for for only moin-toring locations within Cypress Creek (WBID 3235C; see **Appendix B**) during the period of record. Top left: Density histogram of the data compared to the theoretical distribution (red line); Top right: quantile-quantile (Q-Q) plot; Bottom left: cumulative distribution function (CDF) relative to theoretical normal distribution CDF (red line); Bottom right: probability-probability (P-P) plot.
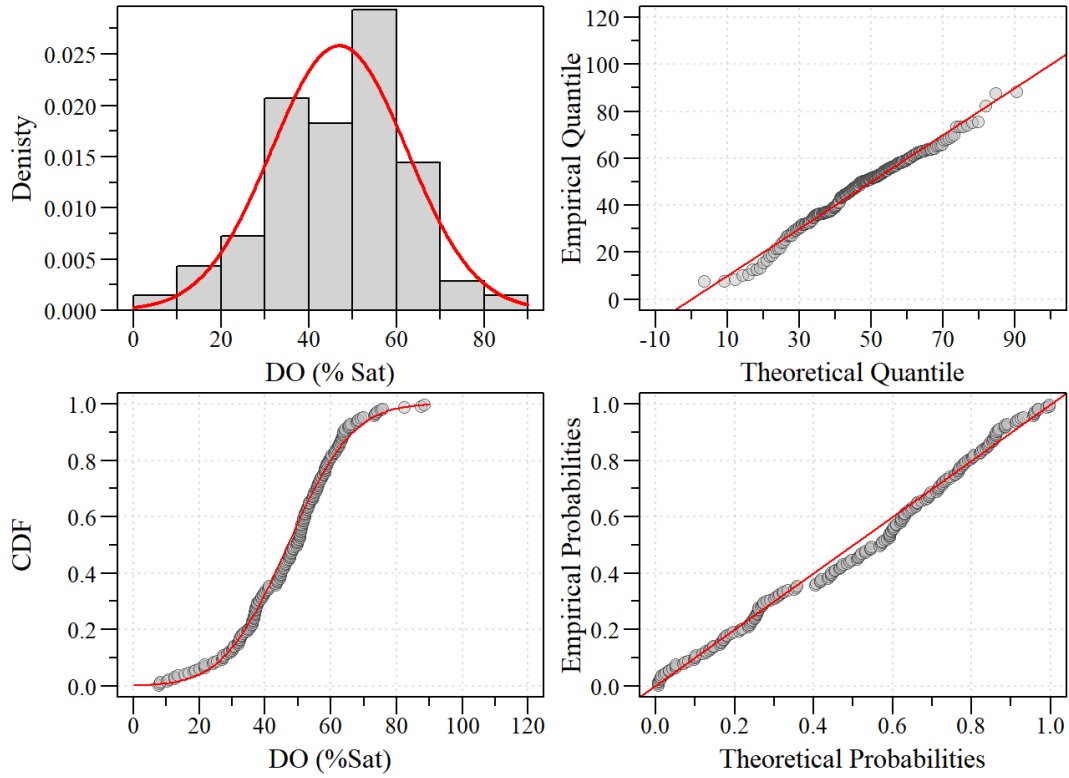
4. Conclusions

- This report only evaluated three of the eight waterbodies proposed DO SSACs by the Department (FDEP 2021).
- Based on the analysis presented here, Daughtrey and Popash creek have significantly declining trends in DO saturation levels averaged across the waterbodies and at individual stations.
- In evaluating the data used to propose the DO SSACs, the data do not fit the assumptions of the parametric percentile statistic therefore alternative methods should be considered to reduce both Type I and II errors.
- Given that significantly declining trends were detected, the central tendency of the data across the period of record is above the current DO criterion and the data does not fit the assumptions of the parametric percentile approach, it is recommended that an alternative approach be considered for Daughtrey and Popash creek.
- Based on discrepancies between what was reported by FDEP and the available data in the IWR database additional analysis and documentation is recommended for Cypress Creek (WBID 3235C). Moreover, additional clarity is needed in what data was used in the derivation of the proposed Cypress Creek

DO SSAC. If data outside of Cypress Creek was used in the derivation of the criterion this could result in increased Type I error.

- While not evaluated in this report, FDEP recommends the adoption of the proposed DO SSACs for these waterbodies as *"each waterbody supports a healthy biological community as indicated by passing Stream Condition Index (SCI) scores, and each waterbody has a high percentage of natural lands, such as wetlands and forests, and are not highly influenced by anthropogenic inputs."* However, the SCI data presented in Appendix A of the FDEP report (FDEP 2021) is very limited and is temporally and spatially limited.

- Not presented in this report, it would appear that several of the monitoring locations within each waterbody exceeds either the total phosphorus or total nitrogen NNC (62-302.531(2)(c)2 FAC) during the evaluation period of record. It is recommended that an assessment be conducted to ensure that nutrient conditions are not contributing to degraded DO conditions within these waterbodies prior proposing alternative DO criterion.

5.   References

- Delignette-Muller M, Dutang C (2015) Fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software 64:1–34. doi: 10.18637/jss.v064.i04
- FDEP (2021) DRAFT Evaluation of Waters for Dissolved Oxygen Site Specific Alternative Criteria (SSAC) Development. Florida Department of Environmental Protection, Tallahassee, FL 85.
- Gross J, Ligges U (2015) nortest: Tests for Normality.
- Helsel DR, Hirsch RM (2002) Statistical methods in water resources. United States Geological Survey, Washington DC.
- Millard SP (2013) EnvStats: An R Package for Environmental Statistics, 2nd edn. doi: 10.1007/978-1-4614-8456-1
- Mohd Razali N, Yap B (2011) Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. Journal of Statistical Modeling and Analytics 2:21–33.
- Peña EA, Slate EH (2006) Global Validation of Linear Model Assumptions. Journal of the American Statistical Association 101:341–354. doi: 10.1198/016214505000000637
- Zeileis A, Hothorn T (2002) Diagnostic Checking in Regression Relationships. R News 2:7–10.

6. Appendix A

- The `AnalystHelper` R-package documentation can be found on Github (`https://github.com/SwampThingPaul/AnalystHelper`). All other packages are available on CRAN.
- Data used in this assessment was retrieved from FDEP's IWR database found here: `http://publicfiles.dep.state.fl.us/DEAR/IWR/`.

`R` code used in this analysis.

```r
# Libraries
library(AnalystHelper)
library(plyr)
library(reshape2)
library(openxlsx)
library(zoo)
library(EnvStats)
library(lmtest)
library(nortest)
library(gvlma)

data.path="./iwr2020_run60_2020-10-05/"

# -----------------------------------------------------------------------
dat.qual=data.frame(QUALIFIER=c(NA,"!","A","D","E","F","I","R","T","U"
                                ,"*","?","B","H","J","K","L","M","N",
                                "O","Q","V","Y","Z"),
                    FATALYN=c("N",rep("N",9),rep("Y",14)))

# WBID3240F - Daughtrey ---------------------------------------------
# Data exported from the IWR database as an xlsx file
dat3240F=read.xlsx(paste0(data.path,"periodOfRecordData3240F.xlsx"))
dat3240F$date=date.fun(convertToDate(dat3240F$date))
dat3240F=subset(dat3240F,year%in%seq(2006,2020,1))
# site not included in original analysis
dat3240F=subset(dat3240F,station.id!="112WRD 264140081494600");

# QA/QC qualifiers
unique(dat3240F$rcode)
unique(dat3240F$xcode)
quals=as.character(unique(dat3240F$xcode))
spl=strsplit(quals,split="")
quals=data.frame(xcode=quals,
                 q1=sapply(spl,"[",1),
                 q2=sapply(spl,"[",2),
                 q3=sapply(spl,"[",3))
quals$Fatal=with(quals,
                 ifelse(q1%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER|
```

```r
                              q2%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER|
                              q3%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER,
                              "Y","N"))
dat3240F=merge(dat3240F,quals[,c("xcode","Fatal")],"xcode",all.x=T)
dat3240F=subset(dat3240F,Fatal=="N")
dat3240F$HalfMDL=with(dat3240F,
                    ifelse(is.na(mdl)==T,result,
                          ifelse(result<=mdl,result/2,result)))


params.keep=c("CHLAC","COLOR","COND",
              "DO","DOSAT","TN","TP","TOC","TEMP")
dat3240F.xtab=dcast(subset(dat3240F,master.code%in%params.keep),
                              wbid+station.id+lat+long+date~master.code,
                              value.var="HalfMDL",mean)
dat3240F.xtab$Sal=with(dat3240F.xtab,SalinityCalc(COND,TEMP))
dat3240F.xtab$DOsat.calc=with(dat3240F.xtab,DO_PerSat(TEMP,DO,Sal))
dat3240F.xtab$DoY=as.numeric(format(dat3240F.xtab$date,"%j"))
dat3240F.xtab$CY=as.numeric(format(dat3240F.xtab$date,"%Y"))
dat3240F.xtab$month=as.numeric(format(dat3240F.xtab$date,'%m'))


## Spatially Average Data
dat3240F.xtab.mean=ddply(dat3240F.xtab,"date",summarise,
                          mean.DOSat=mean(DOsat.calc,na.rm=T))
dat3240F.xtab.mean$DoY=as.numeric(format(dat3240F.xtab.mean$date,"%j"))
dat3240F.xtab.mean$CY=as.numeric(format(dat3240F.xtab.mean$date,'%Y'))
dat3240F.xtab.mean=subset(dat3240F.xtab.mean,is.na(mean.DOSat)==F)
dat3240F.xtab.mean$dum.val=1:nrow(dat3240F.xtab.mean) #time index


# WQS Eval
dat3240F$time=with(dat3240F,ifelse(nchar(time)==3,
                                  paste0(0,time),time))
dat3240F$datetime=with(dat3240F,
                        date.fun(paste(date,time),form="%F %H%M"))
dat3240F.xtab2=dcast(subset(dat3240F,
                            master.code%in%params.keep&
                              station.id!="112WRD  264140081494600"),
                    wbid+station.id+lat+long+datetime~master.code,
                    value.var="HalfMDL",mean)
dat3240F.xtab2$CY=as.numeric(format(dat3240F.xtab2$date,"%Y"))
dat3240F.xtab2$Sal=with(dat3240F.xtab2,SalinityCalc(COND,TEMP))
dat3240F.xtab2$DOsat.calc=with(dat3240F.xtab2,DO_PerSat(TEMP,DO,Sal))
dat3240F.xtab2$DO.TOD.WQS=with(dat3240F.xtab2,
                                DO.TOD.WQS.stream(datetime))
dat3240F.xtab2$exceed=with(dat3240F.xtab2,
                          ifelse(DOsat.calc<DO.TOD.WQS,1,0))


rslt.3240F=ddply(dat3240F.xtab2,c("station.id","CY"),
                summarise,N.exceed=sum(exceed,na.rm=T),
                N.val=N.obs(DOsat.calc))
```

```r
rslt.3240F$PerExceed=with(rslt.3240F,N.exceed/N.val)*100
rslt.3240F$status=with(rslt.3240F,ifelse(PerExceed>10,1,0))
rslt.3240F=ddply(rslt.3240F,"station.id",summarise,
                  sum.status=sum(status),n.val=N.obs(status))

# Kendall Trend
ken.rslt=with(dat3240F.xtab.mean,
              cor.test(mean.DOSat,as.numeric(date),method="kendall"))

# Seasonal Kendall
dat3240F.xtab.mean.sea=ddply(dat3240F.xtab,c("month","CY"),summarise,
                              mean.DOSat=mean(DOsat.calc,na.rm=T))
sea.rslt=kendallSeasonalTrendTest(mean.DOSat~month+CY,
                                  data=dat3240F.xtab.mean.sea)
print(sea.rslt)

# Individual station Kendall Trend
samp.screen=ddply(dat3240F.xtab,"station.id",summarise,
                  N.val=N.obs(DOsat.calc))
samp.screen=subset(samp.screen,N.val>20)
dat3240F.trend=subset(dat3240F.xtab,
                      station.id%in%samp.screen$station.id)
dat3240F.trend$date.num=as.numeric(dat3240F.trend$date)
site.trend=ddply(dat3240F.trend,"station.id",summarise,
                  N.val=N.obs(DOsat.calc,"NaN"),
                  est=cor.test(DOsat.calc,date.num,
                               method="kendall")$estimate,
                  pval=cor.test(DOsat.calc,date.num,
                               method="kendall")$p.value)

# lm trend
dep.trend=lm(mean.DOSat~date,dat3240F.xtab.mean)
summary(dep.trend)
gvlma(dep.trend)
layout(matrix(1:4,2,2));plot(dep.trend)

bgtest(dep.trend,order=3)
shapiro.test(residuals(dep.trend))

## Data distribution
tmp=subset(dat3240F.xtab,is.na(DOsat.calc)==F)$DOsat.calc
shapiro.test(tmp)
ad.test(tmp)

plotdist(tmp,"norm",para=list(mean=mean(tmp),sd=sd(tmp)))

# WBID3240Q - Popash  -----------------------------------------
dat3240Q=read.xlsx(paste0(data.path,"periodOfRecordData3240Q.xlsx"))
dat3240Q$date=date.fun(convertToDate(dat3240Q$date))
```

```r
dat3240Q=subset(dat3240Q,year%in%seq(2006,2020,1))

# QA/QC qualifiers
unique(dat3240Q$rcode)
unique(dat3240Q$xcode)
quals=as.character(unique(dat3240Q$xcode))
spl=strsplit(quals,split="")
quals=data.frame(xcode=quals,
                 q1=sapply(spl,"[",1),
                 q2=sapply(spl,"[",2),
                 q3=sapply(spl,"[",3))
quals$Fatal=with(quals,
                 ifelse(q1%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER|
                         q2%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER|
                         q3%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER,
                         "Y","N"))
dat3240Q=merge(dat3240Q,quals[,c("xcode","Fatal")],"xcode",all.x=T)
dat3240Q=subset(dat3240Q,Fatal=="N")
dat3240Q$HalfMDL=with(dat3240Q,
                      ifelse(is.na(mdl)==T,result,
                             ifelse(result<=mdl,result/2,result)))

params.keep=c("CHLAC","COLOR","COND",
              "DO","DOSAT","TN","TP","TOC","TEMP")
dat3240Q.xtab=dcast(subset(dat3240Q,master.code%in%params.keep),
                    wbid+station.id+lat+long+date~master.code,
                    value.var="HalfMDL",mean)
dat3240Q.xtab$Sal=with(dat3240Q.xtab,SalinityCalc(COND,TEMP))
dat3240Q.xtab$DOsat.calc=with(dat3240Q.xtab,DO_PerSat(TEMP,DO,Sal))
dat3240Q.xtab$DoY=as.numeric(format(dat3240Q.xtab$date,"%j"))
dat3240Q.xtab$CY=as.numeric(format(dat3240Q.xtab$date,"%Y"))
dat3240Q.xtab$month=as.numeric(format(dat3240Q.xtab$date,'%m'))

## Spatially Average Data
dat3240Q.xtab.mean=ddply(dat3240Q.xtab,"date",summarise,
                         mean.DOSat=mean(DOsat.calc,na.rm=T))
dat3240Q.xtab.mean$DoY=as.numeric(format(dat3240Q.xtab.mean$date,"%j"))
dat3240Q.xtab.mean$CY=as.numeric(format(dat3240Q.xtab.mean$date,'%Y'))
dat3240Q.xtab.mean=subset(dat3240Q.xtab.mean,is.na(mean.DOSat)==F)
dat3240Q.xtab.mean$dum.val=1:nrow(dat3240Q.xtab.mean) #time index

# WQS Eval
dat3240Q$time=with(dat3240Q,ifelse(nchar(time)==3,
                                   paste0(0,time),time))
dat3240Q$datetime=with(dat3240Q,
                       date.fun(paste(date,time),form="%F %H%M"))
dat3240Q.xtab2=dcast(subset(dat3240Q,
                            master.code%in%params.keep),
                     wbid+station.id+lat+long+datetime~master.code,
```

```r
                             value.var="HalfMDL",mean)
dat3240Q.xtab2$CY=as.numeric(format(dat3240Q.xtab2$date,"%Y"))
dat3240Q.xtab2$Sal=with(dat3240Q.xtab2,SalinityCalc(COND,TEMP))
dat3240Q.xtab2$DOsat.calc=with(dat3240Q.xtab2,DO_PerSat(TEMP,DO,Sal))
dat3240Q.xtab2$DO.TOD.WQS=with(dat3240Q.xtab2,
                               DO.TOD.WQS.stream(datetime))
dat3240Q.xtab2$exceed=with(dat3240Q.xtab2,
                           ifelse(DOsat.calc<DO.TOD.WQS,1,0))

rslt.3240Q=ddply(dat3240Q.xtab2,c("station.id","CY"),
                 summarise,N.exceed=sum(exceed,na.rm=T),
                 N.val=N.obs(DOsat.calc))
rslt.3240Q$PerExceed=with(rslt.3240Q,N.exceed/N.val)*100
rslt.3240Q$status=with(rslt.3240Q,ifelse(PerExceed>10,1,0))
rslt.3240Q=ddply(rslt.3240Q,"station.id",summarise,
                 sum.status=sum(status),n.val=N.obs(status))


# Kendall Trend
ken.rslt=with(dat3240Q.xtab.mean,
              cor.test(mean.DOSat,as.numeric(date),method="kendall"))

# Seasonal Kendall
dat3240Q.xtab.mean.sea=ddply(dat3240Q.xtab,c("month","CY"),summarise,
                             mean.DOSat=mean(DOsat.calc,na.rm=T))
sea.rslt=kendallSeasonalTrendTest(mean.DOSat~month+CY,
                                  data=dat3240Q.xtab.mean.sea)
print(sea.rslt)

# Individual station Kendall Trend
samp.screen=ddply(dat3240Q.xtab,"station.id",summarise,
                  N.val=N.obs(DOsat.calc))
samp.screen=subset(samp.screen,N.val>20)
dat3240Q.trend=subset(dat3240Q.xtab,
                      station.id%in%samp.screen$station.id)
dat3240Q.trend$date.num=as.numeric(dat3240Q.trend$date)
site.trend=ddply(dat3240Q.trend,"station.id",summarise,
                 N.val=N.obs(DOsat.calc,"NaN"),
                 est=cor.test(DOsat.calc,date.num,
                              method="kendall")$estimate,
                 pval=cor.test(DOsat.calc,date.num,
                               method="kendall")$p.value)


# lm trend
dep.trend=lm(mean.DOSat~date,dat3240Q.xtab.mean)
summary(dep.trend)
gvlma(dep.trend)
layout(matrix(1:4,2,2));plot(dep.trend)

bgtest(dep.trend,order=3)
```

```r
shapiro.test(residuals(dep.trend))

## Data distribution
tmp=subset(dat3240Q.xtab,is.na(DOsat.calc)==F)$DOsat.calc
shapiro.test(tmp)
ad.test(tmp)

plotdist(tmp,"norm",para=list(mean=mean(tmp),sd=sd(tmp)))

# WBID3235C - Cypress --------------------------------------------
dat3235C=read.xlsx(paste0(data.path,"periodOfRecordData3235C.xlsx"))
dat3235C$date=date.fun(convertToDate(dat3235C$date))
dat3235C=subset(dat3235C,year%in%seq(2006,2020,1))

# Sites in IWR databased within Cypress Creek
cypresscreeksites=c("21FLBABRCYPRESS_HEAD",
                    "21FLBABRCYPRESS_OUTFLOW",
                    "21FLFTM 28020237",
                    "21FLGW  56335",
                    "21FLFTM G3SD0084")
# data discrepancies
# dat3235C=subset(dat3235C,station.id%in%cypresscreeksites)

# QA/QC qualifiers
unique(dat3235C$rcode)
unique(dat3235C$xcode)
quals=as.character(unique(dat3235C$xcode))
spl=strsplit(quals,split="")
quals=data.frame(xcode=quals,
                 q1=sapply(spl,"[",1),
                 q2=sapply(spl,"[",2),
                 q3=sapply(spl,"[",3))
quals$Fatal=with(quals,
                 ifelse(q1%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER|
                        q2%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER|
                        q3%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER,
                        "Y","N"))
dat3235C=merge(dat3235C,quals[,c("xcode","Fatal")],"xcode",all.x=T)
dat3235C=subset(dat3235C,Fatal=="N")
dat3235C$HalfMDL=with(dat3235C,
                      ifelse(is.na(mdl)==T,result,
                             ifelse(result<=mdl,result/2,result)))

params.keep=c("CHLAC","COLOR","COND",
              "DO","DOSAT","TN","TP","TOC","TEMP")
dat3235C.xtab=dcast(subset(dat3235C,master.code%in%params.keep),
                    wbid+station.id+lat+long+date~master.code,
                    value.var="HalfMDL",mean)
dat3235C.xtab$Sal=with(dat3235C.xtab,SalinityCalc(COND,TEMP))
```

```r
dat3235C.xtab$DOsat.calc=with(dat3235C.xtab,DO_PerSat(TEMP,DO,Sal))
dat3235C.xtab$DoY=as.numeric(format(dat3235C.xtab$date,"%j"))
dat3235C.xtab$CY=as.numeric(format(dat3235C.xtab$date,"%Y"))
dat3235C.xtab$month=as.numeric(format(dat3235C.xtab$date,'%m'))

## Spatially Average Data
dat3235C.xtab.mean=ddply(dat3235C.xtab,"date",summarise,
                         mean.DOSat=mean(DOsat.calc,na.rm=T))
dat3235C.xtab.mean$DoY=as.numeric(format(dat3235C.xtab.mean$date,"%j"))
dat3235C.xtab.mean$CY=as.numeric(format(dat3235C.xtab.mean$date,'%Y'))
dat3235C.xtab.mean=subset(dat3235C.xtab.mean,is.na(mean.DOSat)==F)
dat3235C.xtab.mean$dum.val=1:nrow(dat3235C.xtab.mean) #time index

# WQS Eval
dat3235C$time=with(dat3235C,ifelse(nchar(time)==3,
                                   paste0(0,time),time))
dat3235C$datetime=with(dat3235C,
                       date.fun(paste(date,time),form="%F %H%M"))
dat3235C.xtab2=dcast(subset(dat3235C,
                            master.code%in%params.keep),
                     wbid+station.id+lat+long+datetime~master.code,
                     value.var="HalfMDL",mean)
dat3235C.xtab2$CY=as.numeric(format(dat3235C.xtab2$date,"%Y"))
dat3235C.xtab2$Sal=with(dat3235C.xtab2,SalinityCalc(COND,TEMP))
dat3235C.xtab2$DOsat.calc=with(dat3235C.xtab2,DO_PerSat(TEMP,DO,Sal))
dat3235C.xtab2$DO.TOD.WQS=with(dat3235C.xtab2,
                               DO.TOD.WQS.stream(datetime))
dat3235C.xtab2$exceed=with(dat3235C.xtab2,
                           ifelse(DOsat.calc<DO.TOD.WQS,1,0))

rslt.3235C=ddply(dat3235C.xtab2,c("station.id","CY"),
                 summarise,N.exceed=sum(exceed,na.rm=T),
                 N.val=N.obs(DOsat.calc))
rslt.3235C$PerExceed=with(rslt.3235C,N.exceed/N.val)*100
rslt.3235C$status=with(rslt.3235C,ifelse(PerExceed>10,1,0))
rslt.3235C=ddply(rslt.3235C,"station.id",summarise,
                 sum.status=sum(status),n.val=N.obs(status))

# Kendall Trend
ken.rslt=with(dat3235C.xtab.mean,
              cor.test(mean.DOSat,as.numeric(date),method="kendall"))

# Seasonal Kendall
dat3235C.xtab.mean.sea=ddply(dat3235C.xtab,c("month","CY"),summarise,
                             mean.DOSat=mean(DOsat.calc,na.rm=T))
sea.rslt=kendallSeasonalTrendTest(mean.DOSat~month+CY,
                                  data=dat3235C.xtab.mean.sea)
print(sea.rslt)
```

```r
# Individual station Kendall Trend
samp.screen=ddply(dat3235C.xtab,"station.id",summarise,
                  N.val=N.obs(DOsat.calc))
samp.screen=subset(samp.screen,N.val>20)
dat3235C.trend=subset(dat3235C.xtab,
                      station.id%in%samp.screen$station.id)
dat3235C.trend$date.num=as.numeric(dat3235C.trend$date)
site.trend=ddply(dat3235C.trend,"station.id",summarise,
                 N.val=N.obs(DOsat.calc,"NaN"),
                 est=cor.test(DOsat.calc,date.num,
                              method="kendall")$estimate,
                 pval=cor.test(DOsat.calc,date.num,
                              method="kendall")$p.value)


# lm trend
dep.trend=lm(mean.DOSat~date,dat3235C.xtab.mean)
summary(dep.trend)
gvlma(dep.trend)
layout(matrix(1:4,2,2));plot(dep.trend)

bgtest(dep.trend,order=3)
shapiro.test(residuals(dep.trend))

## Data distribution
tmp=subset(dat3235C.xtab,is.na(DOsat.calc)==F)$DOsat.calc
shapiro.test(tmp)
ad.test(tmp)

plotdist(tmp,"norm",para=list(mean=mean(tmp),sd=sd(tmp)))
```

7.   Appendix B

Table S1.   Monitoring locations, coordinates and dissolved oxygen sample size for monitoring locations within Daughtery Creek (WBID 3240F).

Station ID	Longitude	Latitude	DO Sample Size
21FLEECO20-29GR	-81.85140	26.75538	158
21FLEECO20-9GR	-81.84479	26.70417	170
21FLEECO20A-11GR	-81.83972	26.70870	159
21FLEECO20A-19GR	-81.84360	26.73402	148
21FLEECOAB96009	-81.83070	26.69922	1
21FLEECOGATRGR91	-81.85569	26.74459	162
21FLFTM 28020037	-81.84528	26.70389	6
21FLFTM 28020231	-81.84991	26.72227	6
21FLFTM CALUSA0030FTM	-81.83950	26.70847	5
21FLFTM G3SD0164	-81.84477	26.70420	1
21FLGW 42393	-81.85417	26.74220	1
21FLWQSPLEE634US	-81.85161	26.72822	2

Note:

Longitude and Latitude are in decimal degrees and WGS84 (EPSG:4326) coordinate system.

Table S2.   Monitoring locations, coordinates and dissolved oxygen sample size for monitoring locations within Popash Creek (WBID 3240Q).

Station ID	Longitude	Latitude	DO Sample Size
21FLEECO23-27GR	-81.80949	26.75546	165
21FLEECO23-5GR	-81.80862	26.71575	170
21FLEECOAC26956	-81.80634	26.70566	1
21FLEECOAD03858	-81.80528	26.70290	1
21FLFTM 28020038	-81.80917	26.71472	2
21FLFTM 28020232	-81.81172	26.71862	2
21FLFTM CALUSA0020FTM	-81.80428	26.74247	1
21FLFTM G3SD0079	-81.80863	26.71581	3
21FLGW 45794	-81.80628	26.76847	1
21FLGW 51902	-81.80694	26.76700	1
21FLGW 53944	-81.80855	26.72923	1
21FLGW 56327	-81.80897	26.72387	1

Note:

Longitude and Latitude are in decimal degrees and WGS84 (EPSG:4326) coordinate system.

Table S3. Monitoring locations, coordinates and dissolved oxygen sample size for monitoring locations within Cypress Creek (WBID 3235C).

Station ID	Longitude	Latitude	DO Sample Size	Cypress Creek Stations
21FLBABRCYPRESS_HEAD	-81.62303	26.81074	13	Y
21FLBABRCYPRESS_OUTFLOW	-81.61950	26.74482	17	Y
21FLEECOCYPRESSGR	-81.62700	26.72601	165	Y
21FLEECOFICHTERSGR	-81.66846	26.72482	166	N
21FLEECOSPANISHGR	-81.59914	26.72579	152	N
21FLFTM 28020043	-81.59916	26.72581	3	N
21FLFTM 28020237	-81.61800	26.73641	4	Y
21FLFTM CYPRESSGR	-81.62702	26.72603	6	Y
21FLFTM G3SD0084	-81.62705	26.72599	2	Y
21FLGW 53948	-81.59924	26.72586	1	N
21FLGW 56335	-81.62318	26.72902	1	Y

Note:

Longitude and Latitude are in decimal degrees and WGS84 (EPSG:4326) coordinate system.