Charlie Creek Water Quality Analysis Using Generalized Linear Mixed Models
Technical Memo, December 2022

Kristina Deak
Environmental Flows and Levels
Southwest Florida Water Management District

Introduction

Environmental data tend to deviate from a normal distribution and can be impacted by the unique characteristics of a sampling location. Generalized linear mixed models (GLMMs) can be used to predict the probability of an outcome (including a binary response) using inputs of fixed and random effects. Fixed effects are those variables are those assumed to have a constant effect on an outcome. Random effects are characteristics unique to a given sample, such as the influence of unquantified aspects of the sampling location. Input data may be normal or non-parametric and either continuous or categorical (Bolker et al., 2009).

In this analysis, GLMMs were used to predict the probability of exceeding State water quality thresholds (per Chapter 62-302.531, F.A.C.) for Class III waters in Charlie Creek under the proposed minimum flows for the system (Ghile et al., 2023). A similar application of GLMMs was used by Janicki Environmental, Inc. (JEI) through Applied Technology & Management, Inc. (ATM) in their 2018 analysis of water quality in the Chassahowitzka River (ATM and JEI 2018).

The most recently adopted Verified List for Charlie Creek was approved on July 15, 2022. According to this list, Charlie Creek above Peace River (Florida Department of Environmental Protection (DEP) waterbody identification number (WBID) 1763A) is impaired for macrophytes (determined by linear vegetation surveys) and total phosphorus (frequent exceedances of the annual geometric mean threshold) and is a medium priority for TMDL development. The WBID has biological evidence indicating non-attainment of its designated use. Charlie Creek above Old Town Creek (WBID 1763D) is impaired for fecal coliform. Charlie Creek above Oak Creek (WBID 1763B), Little Charlie Bowlegs (WBID 1857), and Fish Branch (WBID 1928) are on the Study List for dissolved oxygen percent saturation exceedances without a known causative pollutant. Apart from a statewide total maximum daily load (TMDL) for mercury (DEP 2013), the DEP has not established a TMDL or basin management plan for any waterbody within Charlie Creek.

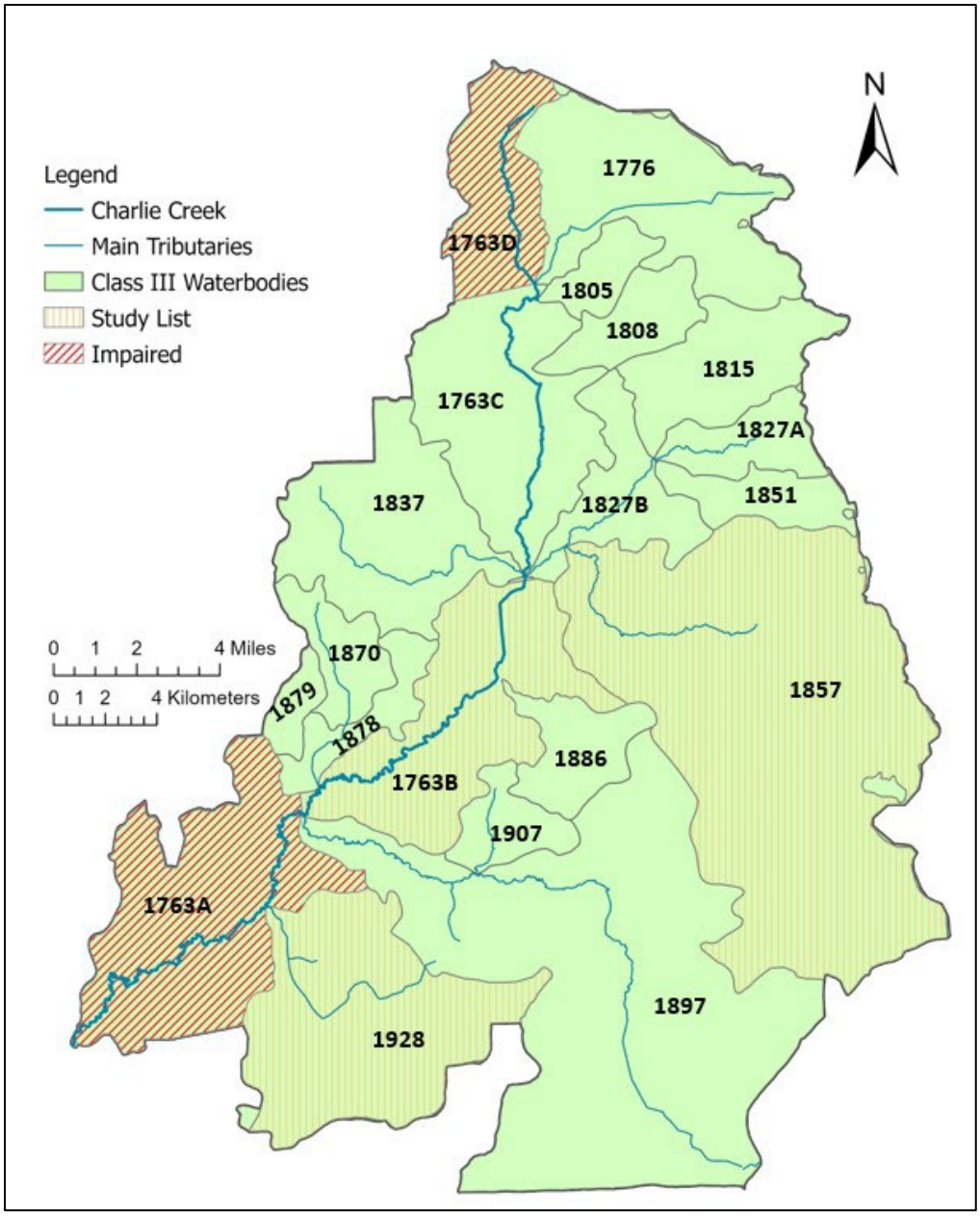

Figure 1. Location of waterbodies by waterbody identification number (WBID) within the Charlie Creek watershed, colored according to designated use classification, impairment status, and inclusion on the study list, according to the DEP's Impaired Waters Rule Run 60 and the Verified List adopted in 2022.

Data

Water quality samples from the DEP, Southwest Florida Water Management District (SWFWMD) and United States Geological Survey (USGS) were used in this analysis (Figure 2). The available period of record and frequency of sampling varied by the sampling agency (Table 1). For consistency with data quality assurance, all water quality data used were pulled from the DEP Impaired Waters Rule Run 59 database, provided by Janicki Environmental, Inc (JEI) through Applied Technology and Management, Inc (ATM and

JEI 2021). Flow data were queried from the USGS website for the USGS Charlie Creek near Gardner, FL (No. 02296500) gage and matched to dates of sample collection. Antecedent flows were not considered in this modeling exercise.

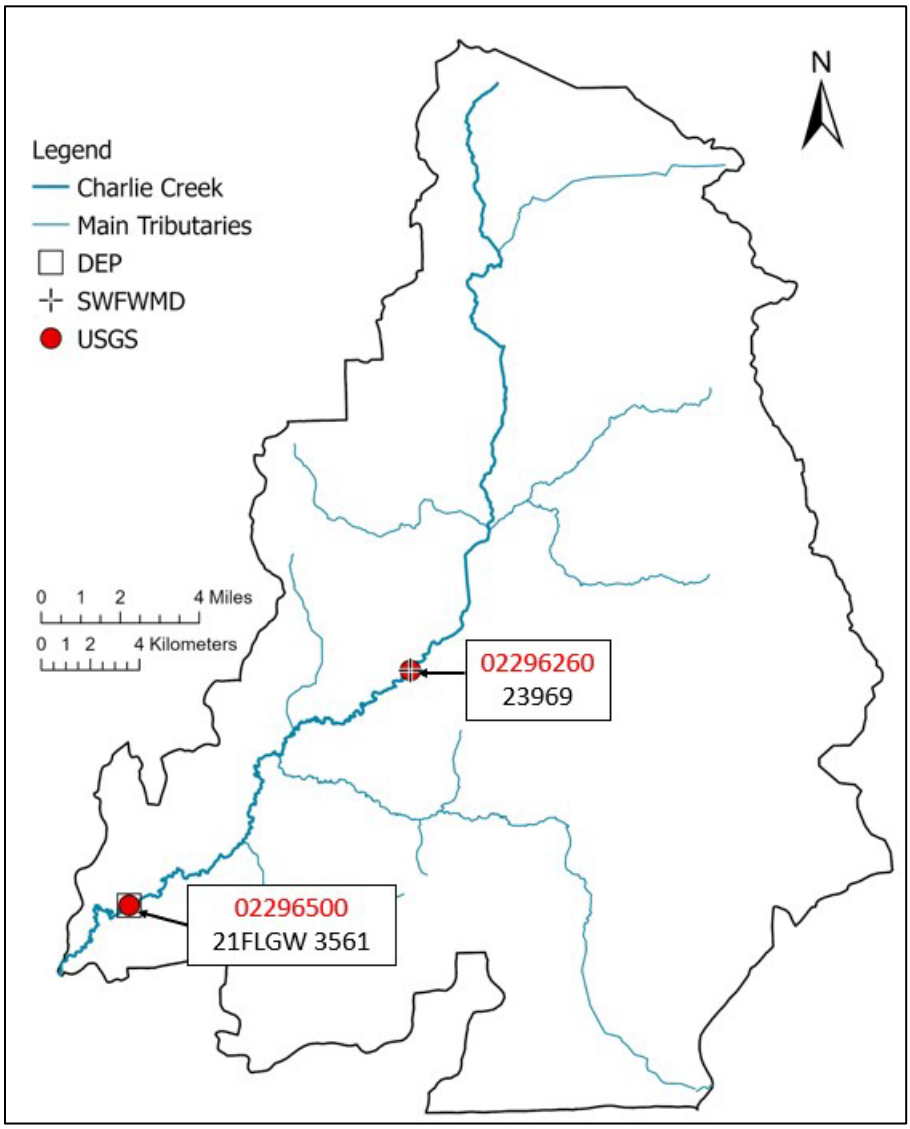

Figure 2. Locations of the water quality sampling sites throughout Charlie Creek considered during analysis. Flow data was obtained from USGS station 02296500 (USGS Charlie Creek near Gardner, FL).

Table 1: Sampling information for each water quality parameter in the GLMM analysis, including station name data source, number (n) of samples, the start and end dates of the period of record (POR), and the approximate sampling frequency over the POR.

Parameter	Station	Source	Samples (n)	POR	Approximate Sampling Frequency
Total Phosphorus	21FLGW 3561	DEP	226	10/1998 - 12/2017	Monthly

	23969	SWFWMD	66	10/2009 - 3/2019	Monthly from 2009 to 2011, bimonthly to 2019
	02296500	USGS	76	5/1970 - 9/1998	Variable; from 1 to 8 samples per year
Dissolved Oxygen (% Saturation)	21FLGW 3561	DEP	221	10/1998 - 12/2017	Monthly
	23969	SWFWMD	104	6/2007 - 3/2019	Variable; from 2 to 15 samples per year
	02296500	USGS	126	5/1967 - 9/1999	Variable; from 1 to 9 samples per year

Methods

To predict the probability of State water quality threshold exceedances with flow reductions (a binomial response), GLMMs were created for total dissolved solids, sulfate, dissolved calcium, total nitrogen, total phosphorus, and dissolved oxygen percent saturation using the glmer function in the lme4 package in R programming language (Bates et al. 2015, R Core Team 2021). Models were run for each analyte, considering combinations of the continuous variables (flow) and categorical variables (season, river kilometer) and the interaction terms among them. "Season" referred to the quarter of the year in which samples were taken, beginning in January. If the model failed to converge with raw flows, the log of flows was taken. The successful model with the lowest Akaiki information criterion (AIC) score was selected for further analysis. The predict function in R was then applied to the selected model to predict the probability of State water quality threshold exceedance at a given flow and location. Flow reduction scenarios were run from 1-20% to determine if such a reduction increased the likelihood of 0.5 probability threshold exceedance of State water quality criteria by more than 15% compared to baseline conditions. The proposed minimum flows for all flow-based blocks were included in the range of flow reduction scenarios.

Results

Total Phosphorus

Data from DEP station 21FLGW 3561, the USGS Charlie Creek near Gardner, FL (No. 02296500) station, and SWFWMD station 23969 were utilized in the creation of GLMMs for total phosphorus. Sample distributions are presented in Figure 3. Exceedance of the State Class III water quality threshold of 0.49 mg/L total phosphorus was utilized as the binary response. Exceedance of the state water quality criteria occurred frequently over the POR for available total phosphorus data.

The GLMM model for total phosphorus with the lowest AIC (362, Table 2) considered the log of flows, season (quarter), river kilometer, and the interaction between river kilometer and flow. When all data were visualized, the probability of exceeding the State water quality criteria for total phosphorus exceeded 0.5 during all flow periods for DEP station 21FLGW 3561 and the USGS 02296500 station (Figure 3). Flow reductions up to 20% in Blocks 2, 3a, and 3b did not significantly impact the probability of total phosphorus exceedance above the 0.5 threshold. The probability of total phosphorus being above 0.5

probability threshold for exceeding the State water quality threshold at SWFWMD station 23969 decreased to 6.9% (from 10.34%) with the recommended 14% flow reduction during Block 2.

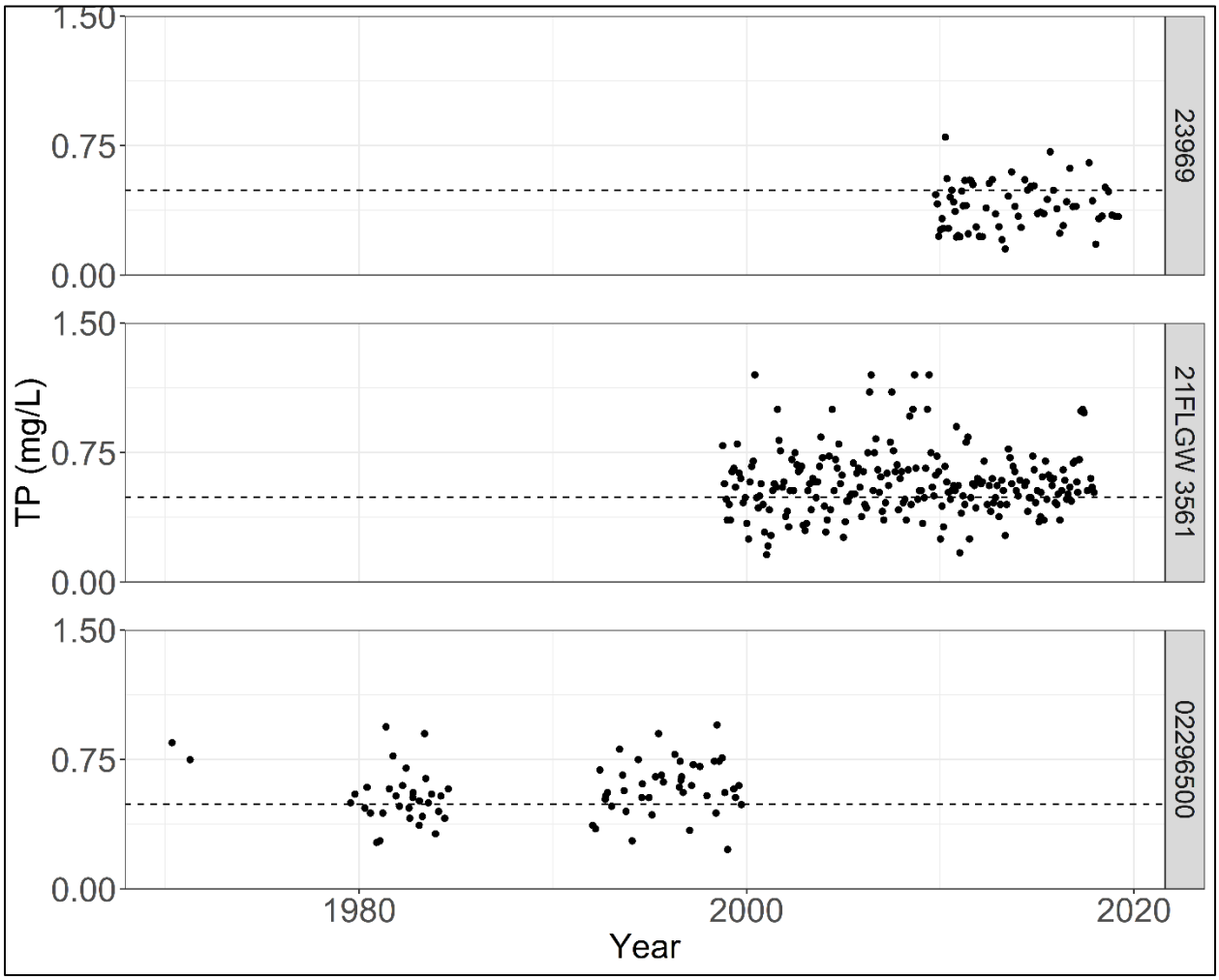

Figure 3. Sample distribution for total phosphorus (TP) at SWFWMD (23969), DEP (21FLGW 3561), and USGS (02296500) stations, listed from upstream to downstream in the watershed. The dashed line indicates the total State water quality threshold for Class III waterbodies (0.49 mg/L). Note this threshold is established for the annual geometric mean rather than for individual samples.

Table 2. Solutions table for fixed effects for the generalized mixed model to predict total phosphorus exceedances.

Fixed effects	Estimate	p-value
Intercept	1.6243	0.009
Log(Flow)	1.0255	0.001
Quarter 1 (Jan-Mar)	0	
Quarter 2 (Apr-Jun)	2.24861	<0.001
Quarter 3 (July-Sep)	3.25961	<0.001
Quarter 4 (Oct-Nov)	2.18496	<0.001

RKM	-0.25109	<0.001
Log(Flow)*RKM	0.06683	0.004

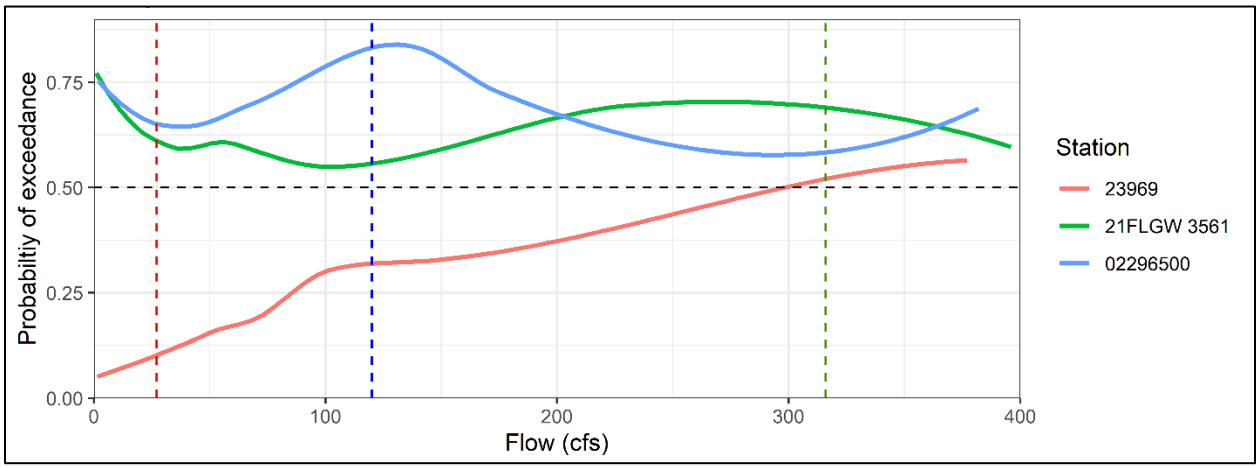

Figure 4: Predicted probability of exceedance of the DEP criteria for total phosphorus in Class III waters (0.49 mg/L) with flow at the USGS Charlie Creek near Gardner, FL gage (No. 02296500). The dashed black line indicates the 0.5 probability of exceedance. The dashed red line indicates the border of Block 1 and Block 2 (27 cfs), the dashed blue line delineates the border between Block 2 and Block 3a (120 cfs) and the green dashed line shows the border between Block 3a and Block 3b (316 cfs). For visualization purposes, the relationship with flow is only shown through 400 cfs. Stations are listed in the key from upstream to downstream.

Dissolved Oxygen

Data from DEP station 21FLGW 3561, the USGS Charlie Creek near Gardner, FL (No. 02296500) station, and SWFWMD station 23969 were utilized in the creation of GLMMs for dissolved oxygen percent saturation. Their sample distribution is presented in Figure 5. Sample values below the DEP Class III water quality threshold of 38% were utilized as the binary response. Over the POR for available dissolved oxygen percent saturation data, exceedance of State water quality criteria for this parameter occurred in approximately 36% of samples at SWFWMD station 23969 (Figure 5).

The GLMM model for dissolved oxygen percent saturation with the lowest AIC (108.4; Table 3) considered the log of flows at time of data collection, river kilometer, and the interaction of these variables. When all data were visualized, the 0.5 probability of exceeding the State water quality threshold increased with increasing flows at station 23969 (Figure 6). Flow reductions up to 20% in Blocks 2, 3a, and 3b did not cause an increase in the 0.5 probability of exceeding the State water quality threshold. The 0.5 probability of threshold exceedance at SWFWMD station 23969 decreased to 16.67% (from 38.89%) with the recommended 12% flow reduction during Block 3a.

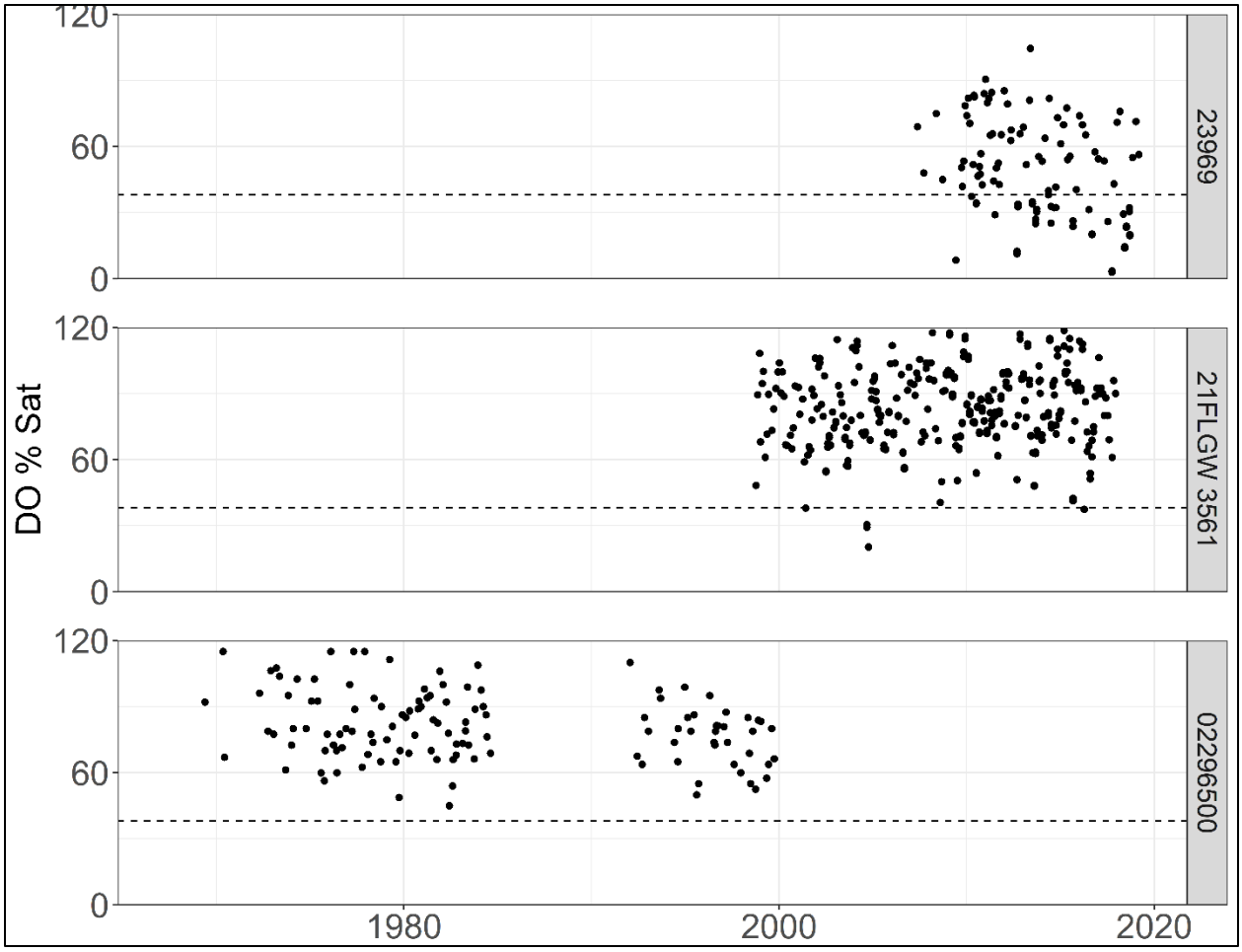

Figure 5. Sample distribution for dissolved oxygen percent saturation (DO % Sat) at SWFWMD (23949), DEP (21FLGW 3561), and USGS (02296500) stations, listed from upstream to downstream in the watershed. The dashed line indicates the total State water quality threshold for Class III waterbodies (38%).

Table 3. Solutions table for fixed effects for the generalized mixed model to predict dissolved oxygen percent saturation exceedances.

Fixed effects	Estimate	p-value
Intercept	-14.98065	0.003
Log(Flow)	3.43896	0.035
RKM	0.15143	0.366
Log(Flow)*RKM	0.01956	0.734

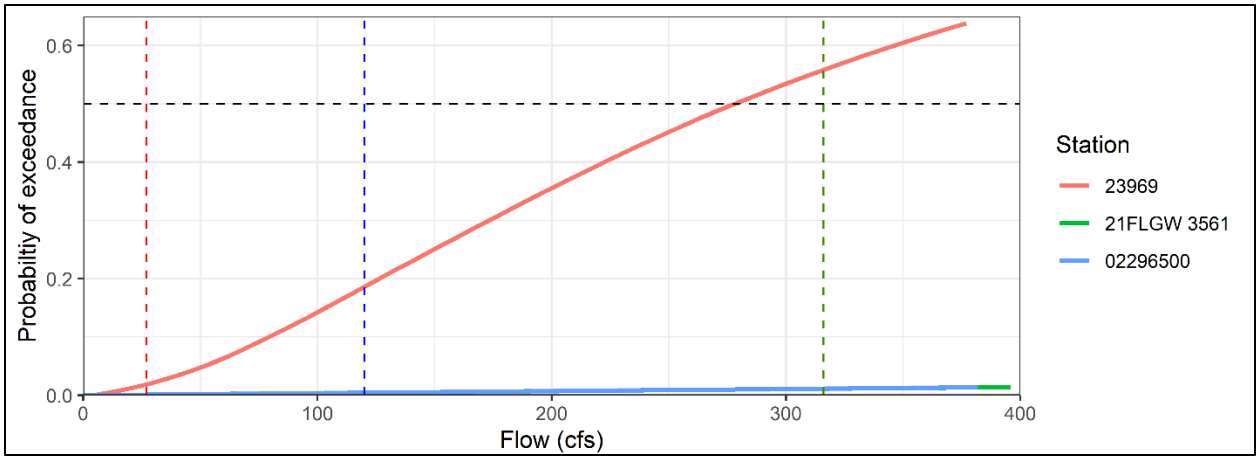

Figure 6: Predicted probability of exceedance of the DEP criteria for dissolved oxygen percent saturation in Class III waters (38%) with flow at the USGS Charlie Creek near Gardner, FL gage (No. 02296500). The dashed line indicates the 0.5 probability of exceedance. The dashed red line indicates the border of Block 1 and Block 2 (27 cfs), the dashed blue line delineates the border between Block 2 and Block 3a (120 cfs) and the green dashed line shows the border between Block 3a and Block 3b (316 cfs). For visualization purposes, the relationship with flow is only shown through 400 cfs. Stations are listed in the key from upstream to downstream.

Summary

The State water quality threshold for total phosphorus is based upon the annual geometric mean of samples frequently exceeding the threshold in a three-year period. The State water quality threshold for dissolved oxygen percent saturation is exceeded if more than 10% of the daily average dissolved oxygen percent saturation values are below 38%. While the statistics calculated below are modeled off available sample data, and therefore, reflect the probability of exceedance on a per sample basis, it is assumed that if the number of samples exceeding the threshold is not substantially increased by flow reduction, the probability of exceeding the State water quality threshold once an annual geometric mean or 10% of daily averages is calculated would also not increase.

Based upon results from this analysis, the proposed minimum flows for Charlie Creek are not anticipated to increase the 0.5 probability of exceeding the State water quality thresholds for total phosphorus or dissolved oxygen percent saturation at the evaluated water quality stations. The likelihood of total phosphorus exceeding the 0.5 probability threshold at SWFWMD station 23969 decreased to 16.67% (from 38.89%) with the recommended 12% flow reduction during Block 3a. The likelihood of 0.5 probability threshold exceedance for dissolved oxygen percent saturation at SWFWMD station 23969 decreased to 6.9% (from 10.34%) with the recommended 14% flow reduction during Block 2.

Of note, water quality data in Charlie Creek are somewhat limited and there was little overlap in the period of record or sampling frequency in the available data. Robust sampling over a longer period of record with additional water quality stations along the length of the river would likely improve confidence in the model outputs.

Literature Cited

Applied Technology and Management, Inc. and Janicki Environmental, Inc. (ATM and JEI) 2018. Exploratory evaluation of water quality and flow relationships for the Chassahowitzka River in support of minimum flows reevaluation. Prepared for the Southwest Florida Water Management District. Brooksville, Florida.

Applied Technology and Management, Inc. and Janicki Environmental, Inc. (ATM and JEI) 2021. Charlie Creek water quality assessment. Prepared for the Southwest Florida Water Management District. Brooksville, Florida.

Bates, D. M. Machler, B. Bolker, and S. Walker. 2015. Fitting linear mixed effects models using lme4. Journal of Statistical Software. 67:1-48.

Bolker, B.M., M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, H.H. Stevens and J-S. S. White. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology and Evolution. 24:119-174.

Deak, K., Y. Ghile, G. Herrick, and D. Leeper. 2023. Proposed minimum flows for Charlie Creek. Southwest Florida Water Management District. Brooksville, Florida.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.