Response to Preliminary Review Comments from Liquid Solutions Group, LLC, representing the North Florida Utility Coordinating Group (NFUCG) on the Draft Minimum Flows and Minimum Water Levels Re-Evaluation for the Lower Santa Fe and Ichetucknee Rivers and Priority Springs Draft Report Released on December 20, 2019

INTRODUCTION

This response addresses the 14 specific comments provided by consultants for the North Florida Utility Coordinating Group (NFUCG) dated January 30, 2019. This document augments and replaces a prior response addressing the NFUCG's four summary comments.

It is important to recognize that the review comments addressed below are to the *Minimum Flows and Minimum Water Levels Re-Evaluation for the Lower Santa Fe and Ichetucknee Rivers and Priority Springs* Draft Report (LSFI MFL) released on December 20, 2019. Some of the comments are not applicable to the Final LSFI MFL released on January 4, 2021. Among other changes relevant to the NFUCG comments a primary change is that only the estimated median flow at the US441 gage was used in the Final LSFI MFL Re-Evaluation, i.e., issues with estimated flows at the extremes (flows at or near zero cfs and greater than 3200 cfs) are not relevant for this MFL determination.

It should be noted that the NFUCG's comments have been inserted into this response document to provide context and improve readability. However, the tables and graphics contained in the NFUCG document were not inserted. Therefore, at appropriate points in this response, a note to see the original NFUCG submittal is provided.

Summary Comment – Topic 1:

Calculation of historical flow for Lower Santa Fe River (LSFR) at US 441 gage. The accuracy of the multiple linear regression (MLR) equation used to estimate flows at US 441 is inadequate for use in this process. The required statistical conditions for use of MLR are not met.

Summary Response – Topic 1:
The District disagrees with the above statement. Additional information has been provided in the Peer Review Panel (PRP) Resolution document. Briefly, most of the statistical conditions allegedly required for using MLR results are not required for using the predictive equation(s) for predictions (gap filling). Numerous models and modeling approaches were tested, statistical assumptions were evaluated, and marginal improvements were obtained. All the models and modeling approaches tested produced very similar estimates of the median flow at US441– the objective of the modeling used in the final LSFI MFL report application. Specific comments are repeated below with responses following:

Specific Comments – Topic 1:
The synthetic historical discharge time series for the LSFR at the US 441 gage does not meet the technical standards required for use in development of an MFL. The following comments relate to

the use of a statistical method for synthesizing historical LSFR discharge estimates at the US 441 gage:

1. The discharge series for the LSFR at the US 441 gage was derived through multiple linear regression (MLR), using measured discharge at the LSFR Ft. White and Worthington gages as explanatory variables. Quantitatively, the accuracy of the MLR equation can be measured in several ways.

   The adjusted coefficient of determination as documented in the report is 0.86 which is generally considered acceptable. However, as documented in Helsel and Hirsch (USGS, 2002), “Values of R2 close to 1 are often incorrectly deemed an indicator of a good model. This is a dangerous, blind reliance on the computer software. An R2 near 1 can result from a poor regression model.” In the case of the MLR for the US 441 gage, there are other metrics which show significant issues with the MLR equation.

   The Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) are 102 cubic feet per second (cfs) and 186 cfs, respectively. These values represent a deviation of about 20% and 37% relative to the average observed (filtered) discharge of 509 cfs at the US 441 gage. Furthermore, the Mean Absolute Percentage Error (MAPE) of the MLR is 293%. The MAPE is an estimation of the average size of the error in percentage terms at individual observations. Together, the MAE, the RMSE and the MAPE indicate a substantial deviation between the predicted and the observed discharges at the US 441.

   A comparison of the Frequency Distribution Curves (FDC) for the observed and predicted discharge series at the US 441 gage, resulting from the MLR, indicate that the predicted discharge series underestimates peak flows. Furthermore, the MLR equation produces negative discharges about 6% of the time despite filtering observed values with zero discharge for the MLR (i.e., no zero-discharge included as observed values in the creation of the MLR). Also, at the median value (p50) the predicted discharge series is 16 cfs less than the corresponding observed value.

2. Chapter 7 of the Draft MFL Report includes the use of “Rating Reliability Relative to Median Reference Timeframe (RTF)” as a statistical measure to define the acceptable or unacceptable level of uncertainty associated with the long-term MLR prediction of springflows. Though not presented in the Draft MFL Report, the Rating Reliability Relative to Median RTF for the US 441 gage is 32% (see Table 1 – *see original NFUCG submittal*). This level of error is at the high end of the Rating Reliability Relative to Median RTF computed for various springs in the Draft MFL Report.

   Based on these values presented, the Draft MFL Report concludes that development of spring-specific MFLs using MLR was “impractical” because “the accuracy and precision warranted to establish spring-specific MFLs” was not being met. We would also conclude that the MLR developed for the US 441 gage is impractical for use in setting an MFL at this gage. With due consideration of the best data available and the uncertainties, an appropriate way to afford protection from significant harm to the LSFR upstream of US 441 is to utilize the Ft. White gage downstream.

3. Chapter 7 of the Draft MFL Report documents an attempt to develop a MLR equation for the Santa Fe River Rise. The Santa Fe River Rise, which is just upstream of US 441, provides over 75 percent of the flow at the US 441 gage. As expected, the Santa Fe River Rise and LSFR at US 441 have statistically similar flow frequency distributions. Also as expected, both

the MLR for LSFR and US 441 share similar characteristics as presented in Table 2 (*see original NFUCG submittal*).

The Draft MFL Report concludes that development of an MFL for the Santa Fe River Rise using MLR is “impractical” because “the accuracy and precision warranted to establish spring-specific MFLs” was not being met. With due consideration of the best data available and the uncertainties, an appropriate way to afford protection from significant harm to the LSFR upstream of US 441 is to utilize the Ft. White gage downstream.

4. According to Helsel and Hirsch (USGS, 2002), there are five inherent assumptions associated with the use of linear regression that must be met. These assumptions and their corresponding association to the MLR analysis for the US 441 gage are listed below:

a. <u>The model form must be correct (i.e., Y is linearly related to X))</u>. There must be sufficient evidence that the relationship between discharges at the US 441 and those at Ft. White and Worthington is linear. In this case, a plot of regression residuals (expressed as predicted – observed discharge) vs. observed discharge in Figure 1 (*see original NFUCG submittal*) shows a curved departure from the zero-residual line. This curvature evidences that a linear relationship assumption is not valid, especially at higher flows. Therefore, this required assumption for use of MLR is not met.

This issue might result from the hydrologic character of the LSFR in this area. Specifically, the discharge time series associated with the three Santa Fe River gages used in the MLR analysis exhibit statistically different flow distributions for the POR 1992 through 2018 (see Figure 2) (*see original NFUCG submittal*). This difference increases the difficulty of finding linear relationships among the gaged flows.

b. <u>Data used to fit the model must be representative of data of interest.</u> The LSFR discharges at the Ft. White and Worthington gages within the POR used for the MLR analysis (1992-2018) must be representative of the discharges from 1932 through 1992. However, as shown in Table 2 (*see original NFUCG submittal*), the discharges at Ft. White, post-1992, are between 6 and 35 percent lower than flows pre-1992. Similarly, Worthington exhibits discharges post-1992 that are between 16 and 100 percent lower than flows pre- 1992. As a result, the MLR relationships have been developed for a drier period of lower river flows and applied to a wetter period of higher river flows. Thus, this required assumption for use of MLR is not met.

c. <u>Variance of the residuals must be constant.</u> The variance of the regression residuals changes with the predicted values (see Figure 3) (*see original NFUCG submittal*). Lower variance occurs at lower discharges, and higher variance occurs at higher discharges. This observed heteroscedasticity was confirmed through statistical analysis and is statistically significant. This required assumption for use of MLR is not met.

d. <u>The residuals must be independent.</u> A plot of residuals vs. time shows autocorrelation in the residuals (see Figure 4) (*see original NFUCG submittal*) which means that the estimates and statistical significance of the linear regression coefficients and variance of the errors may be wrong (USGS, 2002). The autocorrelation was confirmed statistically and is statistically significant. This required assumption for use of MLR is not met.

e. <u>The residuals must be normally distributed.</u> A probability plot of residuals shows nonnormality (see Figure 5) (*see original NFUCG submittal*). Non-normality was confirmed through statistical testing and is statistically significant. Normality of the regression residuals is required to validate the statistical significance of the slope coefficients. That is, without normality of residuals the assumption of linear correlation, as well as the predicted magnitude of the proportionality (slope coefficient) between an explanatory and a response variable may be inaccurate. As such, this is a significant issue affecting the validity of the use of MLR and this assumption is not met.

5. The MLR analysis used to synthesize discharges at the US 441 gage, excluded flows at the Ft. White gage above 3,500 cfs, and zero flows at the US 441 gage. Nevertheless, the resulting MLR equation was ultimately applied to Ft. White flows above 3,500 cfs. As stated above, the data used to fit the model must be representative of data of interest. Exclusion of these data violates this core assumption. Furthermore, no explanation as to the reason for this filtering of observed data is provided.

6. For some days, the MLR equation predicts negative flows associated with the US 441 gage. When a negative flow was calculated, the daily flow was increased to 0 cfs as part of the development of the RTF series. Then, as a subsequent step in the development of the RTF, the estimated groundwater impact was added to this 0 cfs flow. In essence, this assumes that the LSFR at US 441 would never experience 0 cfs flow under a pumps-off condition and that the only reason that the LSFR at US 441 would experience no flow is due to pumping. Unfortunately, this assumption has not been validated and is likely to be incorrect. The use of the MLR-predicted flows (adjustment of negative values) for this US 441 gage in this way for this application is inappropriate.

Specific Responses – Topic 1:

R1. The peer reviewers raised some concerns regarding the regression equation used to estimate (infill and extend) flow at the US441 gage that are addressed in detail in the District's response to General Comment 6 and Attachment A (including Exhibit C) of the December 2020 Peer Review Resolution Document. In the Final LSFI MFL, a piecewise regression (Exhibit C) was used that offers the benefit of including flow values of 0.0 and estimating a zero-flow segment, but the regression can also be applied in a way that allows negative values to be estimated by ignoring the zero-flow segment. Numerous modeling approaches were tested with good results (difference from observed median flow < 5%) and one or more may be appropriate for estimating flow depending on future assessments.

R2. The District does not concur with the conclusion that the MLR developed for US441 is impractical to use in setting an MFL. The District proposed an MFL for the US441 gage to facilitate water management, and the R-square and RMSE of the piecewise prediction equation (and others provided in the Peer Review Resolution Document) are more than acceptable for this purpose. With respect to the "Rating Reliability Relative to Median" metric, the number of spring-flow measurements range between 12 and 137, while there are 9,572 daily streamflow values at US441 used in the regression. The objective of the regression model is to estimate (i.e., predict) the mean expected response, the confidence of which is greatly influenced by the number of samples (see standard error in Helsel, et.al 2020). The benefit of many sample values is a high degree of confidence in the mean response.

The table below is from the District's Peer Review Resolution Document Appendix A (Table AC 2) and the Final LSFI MFL Re-Evaluation report Attachment A (Table 2). The estimate of the median is within about 1% of the measured value.

Table AC 2. Estimated flow exceedances calculated using a 2-knot piecewise linear regression model for US441 compared to gaged flow for Water Years 1993-2019 (May 1, 2019)

Model/Exceedance	1	5	10	25	50	75	90	95	99
US441 gaged flow data	3,701	1,648	1,212	696	356	133	33	0	0
Piecewise 2 knots	3,347	1,693	1,191	695	352	148	33	7	0

R3. As mentioned in the preceding response to comment 2, the issue with Santa Fe River Rise, and other springs, is not the location or the statistical parameters, it is the lack of data to have confidence in the mean response. If there are 20 flow measurements available that range between 100 and 200 cfs, there is little confidence in the actual mean or median flow estimate; however, if there are 10,000 flow measurements available within the same range, there is a high degree of confidence in the mean and median, regardless of the distribution of the data.

Also, it was recommended in the Upper Santa Fe River MFL document that an MFL be developed at the US441 gage (Water Resource Associates, Inc. 2007).

R4. The assumptions listed do not apply to all uses of regression as outlined in Table 9.1 of the reference (Helsel, et.al 2020). Only assumptions 1 and 2 are applicable to predict Y as a function of Xs. Assumption 1 is that the model form is correct, and 2 is that data used are representative of the data of interest.

a. Numerous model forms were investigated after the draft report was provided and are described in Attachment A of the District's Peer Review Resolution Document. The greatest improvement regarding the distribution of residuals and R-square was found by using lagged Worthington Springs flow. In general, the resulting distributions of residuals are symmetrical (see associated histograms) and do not exhibit the curvature depicted in the comment Figure 1 (see original NFUCG submittal).

   Regarding comment Figure 2 (*see original NFUCG submittal*), the graphic depicts skewed distributions of similar shape typical of river gage data. Fewer bins would result in even better concordance of the distributions. The US441 and Fort White flow distributions are nearly identical and the association between flow at these two gages accounts for 96% of the model fit, even given the large number of bins. Note that Helsel, et. al (2020) discourage the use of overlapping histograms.

   Approximate normality of residuals was achieved by removing statistical outliers using Cooks Distance statistic (117 values out of 9,957 – i.e., remove the peaks and valleys in the commentor's Figure 4) (*see original NFUCG submittal*), but it is noteworthy that the median flow estimate was within a few cfs of that produced using the selected piecewise model. As a matter of practice, statistical outliers of hydrologic data are not removed unless measurement error is suspected or the data are outside a range of interest (e.g., flow near 0.0 or greater than 3,500 cfs).

Additional improvements might be gained by varying the lag as a function of flow, but the District is satisfied with the current models given the objective – predicting the mean response at US441 given flow at the Fort White and Worthington Springs gages.

b. The District contends that there are sufficient data over a broad range of flows to be representative of the POR. The response at each gage to rainfall could be expected to vary with longer term climatic conditions (i.e., periods of relatively wet and dry conditions); however, the association between the explanatory variables (Fort White and Worthington springs flow) and the dependent variable (US441 flow) is not expected to vary substantially as flow at these gages are influenced by the same climatic stresses.

As a simple check, the median flows were determined for the 2014 to 2019 period as the average flow at Fort White is comparable to the period of record (POR of 1933 to 2015) and it is a continuous record of six years. The ratio of reference timeframe (RTF) flow medians used for setting the MFL at US441 is 0.435. The ratio of the medians for the measured and infilled data for the POR is 0.425, and the same statistic for measured data in the recent six-year period is 0.478. The recent 6-year record has much less variability (based on the range of average annual flow) than the POR. Nonetheless the ratios of medians are comparable.

Statistic	US441	Fort White
Median flow (2014 to 2019)	622	1,300
Ratio of medians		0.478
Average flow (2014 to 2019)	833	1,467
	US441 estimated	Fort White
Median flow (1933 to 2015)	518	1,220
Ratio of medians		0.425
Average flow (1933 to 2015)	751	1,471

c – e. Other arguments are presented in the District's Peer Review Resolution Document Attachment A, mostly related to the quantity of data and application of central limit theorem principles, but the purpose of the model is only to predict, and only the median is of interest at this time (i.e., subject to rule development). The District would note, however, that the "Variance Change" lines in Figure 3 (*see original NFUCG submittal*) depict a change in data range and exaggerate the actual change in variance.

R5. Only the median is being used in the final MFL. Flows greater than 3,500 cfs were not intended to be evaluated for this MFL so it was appropriate to exclude these data with the goal of improving model predictions over the range of interest. These high flows are exceeded less than 5% of the time, often are part of an extreme high flow event, and not evaluated in this MFL re-evaluation.

R6. For developing the RTF flow series, predicted flows could be (and were) negative. RTF adjustments were added to the infilled and extended flow series, including the negative flows. If the resulting RTF flow was still negative, then the RTF flow was set to zero.

Zero flows are a special case; and if used as an MFL metric, would be evaluated by other methods.

Summary Comment – Topic 2:

Estimation of Water Resource Values Upstream for the LSFR at the US 441 Gage. The WRVs used for the US 441 gage are mathematically derived from work related to the Ft. White gage without adequate field work or modeling. The mathematical method used for derivation of a WRV at the US 441 gage is not technically based and inappropriate for application.

Summary Response – Topic 2:
Please see response to Summary Comment No. 1; the gap filling approach is appropriate and standard hydrological practice. In the final report, only the estimate of the median flow at US441 is used in an apportionment procedure at the median flow.

Specific Comments – Topic 2:
The following comments and questions relate to the establishment of Water Resource Value (WRV) metrics upstream of the US 441 gage:

7. The Draft MFL Report indicates that the LSFR at US 441 gage has been incorporated based on a recommendation from 2007 (Water Resources Associates, 2007). However, the Draft MFL Report also states that "There is a paucity of hydrologic data available for the Santa Fe River near the parks during the RTF period on which to base a WRV-specific assessment of flow metrics similar to those evaluated for the reach between the US441 and Ft. White gages". It appears that not much data is available because collection of these data was not undertaken as is typically performed for development of an MFL.

   As a result, the WRVs presented involve translation, apportionment, or other analytical methods to utilize previous work without confirmation of their applicability at the US 441 gage. Sound science is a critical component of any public policy decision, including MFL setting, and the SRWMD should initiate the collection of adequate data and perform additional modeling, if required, in the reach upstream of US 441 if an MFL is desired.

8. Downstream of the US 441 gage, the "Instream Habitat" WRV was determined using site specific data and evaluated, in part, through SEFA Modeling at four sites on the LSFR. These sites included the US 441 gage, and three other locations downstream from the US 441 gage. The resulting Instream Habitat WRV calculated for the LSFR at Ft. White was then adjusted by median flow apportionment to the US 441 gage citing the work of Jacobs and Romesser (2006).

As stated by Jacobs (2006), in response to peer reviewer query, “The exact allocation scheme upstream of a gage is ultimately a policy decision.” Furthermore, it is clear in Jacobs and Romesser (2006) and Jacobs (2006), that the method cited was intended to support data that had been collected and not act to replace such data collection. As stated above, for an issue as important as the development of MFLs, sound science is a critical component and the SRWMD should initiate the collection of adequate data and perform additional modeling, if required, in the reach upstream of US 441 if an MFL is desired.

9. Notwithstanding the previous comment, the LSFR Ft. White and US 441 gages are subject to substantially different hydrologic conditions which result in substantially different distribution of flows as shown in Figure 6 (*see original NFUCG submittal*). As a result, some of the assumptions included in the Jacobs and Romesser method are not valid for application at this location.

10. Section 5 of the Draft MFL report states that daily Area Weighted Suitability (AWS) values for flows greater than 3,200 cfs at the Ft. White gage were not used for the flow reduction assessment associated with the evaluation of the Instream Habitat WRV. Additional information on the rationale for excluding these data should be provided.

Specific Responses – Topic 2:

R7. Additional data are being collected as part of the District’s ongoing efforts. However, the absence of sufficient data to set more robust MFLs does not negate the District’s charge of protecting the resource. The District contends that an apportionment approach is reasonable and necessary to prevent significant harm to the section of the river upstream of US441.

Following is an excerpt from the Final LSFI MFL Re-Evaluation report (p. 106):

> “As a back-check on the reasonableness of this approach, the method was applied to the Worthington Springs gage on the upper Santa Fe River for which an MFL has been determined previously (Water Resource Associates, Inc. 2007). The median RTF flow at the Worthington Springs gage is 123 cfs. The MFL adopted by rule for the median flow condition is 111 cfs, which implies that 12 cfs would be available at the median flow condition. The WRVs that provided the best opportunity to establish an MFL protective of all the applicable water resource values identified for the upper SFR are fish and wildlife habitats and the passage of fish.
>
> The ratio of the median RTF flows at the Fort White and Worthington Springs gages is 0.0969 (i.e., 123/1,270). Applying this adjustment factor to the most applicable hydrologic shift determined for the Fort White gage (115 cfs), yields an adjusted shift of 11 cfs for instream habitat. The adjusted shift is close to the available water associated with the MFL established for the Worthington Springs gage (i.e., 12 cfs). The results support the conclusion that in the absence of site-specific resource information at the O’Leno/River Rise Parks, the apportionment of hydrologic shifts is a reasonable approach for determining a hydrologic shift at the US441 gage that is protective of water resources upstream from the gage such that significant harm, due to withdrawals, is avoided. Additional work efforts are recommended in Section 6.2.3.”

R8. The District agrees that additional data upstream of US441 are needed and is currently collecting and evaluating recently collected data. Also see response to Comment 9 below.

R9. The District contends the apportionment approach is appropriate and the best available information for establishing a MFL at the median given the absence of sufficient data to do otherwise. Apportionment is used in almost all MFLs being set. For example, when evaluating wetted perimeter and fish passage under varying flows for the Middle Peace River in Florida, "all flows for all cross-sections associated with a breakpoint were converted to flows at the Arcadia gage." The same was done for fish passage (Locke, et. al, 2008). Instream habitat model results, for example, are translated from sites on the river to gages miles away as a percent change in flow not as an absolute value of flow. What is inferred using the US441 gage as a reference gage is that if available water at Fort White is 10% of flow at the median, for example, then the available water at the US441 gage should be about 10%. To assess an upstream section of a river, the downstream needs of the river must be considered, and apportionment is a logical way of doing this. Even if the WRV analysis upstream of US441 allowed for more water, a more restrictive MFL at Fort White, or the Lower Suwannee MFL, might prevent or influence allowable withdrawals upstream of US441.

R10. A flow of 3,200 cfs at Fort White is the upper limit of the HEC-RAS simulated flow scenarios and associated rating curves used in the application of the SEFA model to evaluate suitable instream habitat. This flow encompasses all needed flows for analysis of WRVs in the Lower Santa Fe; a flow of 2,826 cfs was the highest flow used for evaluating other WRVs.

Summary Comment – Topic 3:

Estimates of historical pumping impacts on LSFR. The North Florida Southeast Georgia (NFSEG) groundwater model is poorly calibrated at the US 441 gage leading to inaccurate impact calculations. The NFSEG model shows anomalous impacts at springs along the LSFR which affect the calculation of historical pumping impacts.

Summary Response – Topic 3: Impact calculations for the US 441 gage are less than those at the Ft. White gage and exhibit consistent trends. While individual spring flows can be modeled with NFSEG, individual spring impacts were not utilized in the 2019 draft MFL report. The NFSEG model has been peer reviewed and represents the best information available for this system. Subsequent review of the calibration in the US 441 region indicated that the final model calibration in that area appropriately minimized model error at provided observation points in 2001 and 2009.

Specific Comments – Topic 3:

The following comments and questions relate to the use of the NFSEG Model for estimation of historical pumping impact on the LSFR:

11. NFSEG model calibration of baseflow at the US 441 gage ranges from being 35 to 84 percent overpredicted as shown in Table 3 (*see original NFUCG submittal*). This high level of error indicates that one or more significant issues are affecting the accuracy of the model in this area. The source of this error(s) should be identified and these errors reduced prior to using the NFSEG model for these evaluations.

12. One potential source of error is the conductance value of the NFSEG model general head boundary (GHB) representing the Santa Fe Spring upstream of US 441. As shown in Table 4 (*see original NFUCG submittal*), this conductance value is significantly greater than other GHB cells along the LSFR. The high conductance for the Santa Fe Spring GHB has significant effect on the predicted pumping impacts at Santa Fe Spring from pumps off to 2001/2009 conditions. In fact, the impacts from pumps off at Santa Fe Springs represent approximately

90 percent of the total simulated impact at the US 441 gage, for both 2001 and 2009 stress periods of the NFSEG Model. And the simulated impact at Santa Fe Springs at is an outlier compared to all other springs on the LSFR which are included in the NFSEG model as presented in Table 4 (see *original NFUCG submittal*). This level of impact does not appear to be reasonable. Reducing the conductance value of the GHB representing the Santa Fe Spring to a value more representative of other values along the LSFR, results in better calibration of discharge at the US 441 gage. For example, as presented in Table 5 (*see original NFUCG submittal*), when the conductance value is reduced to approximately 600,000 ft2/day (which is greater than the average of LSFR spring conductance value), then the error in the LSFR baseflow at US 441 is reduced dramatically. In order to address this issue, we strongly recommend adjustment of this conductance value.

Specific Responses – Topic 3:

R11. In addition to the above summary response, it should be recognized that the primary use of the model by the District is production of change estimates, not estimates of absolute flows. The model was developed with this use in mind and the District believes the model is appropriate for this purpose as calibrated.

R12. The conductance values for Santa Fe Spring upstream of US 441 are based on the overall fit of the model resulting from the set of calibration targets developed from the best available data, and through calibration with the PEST software package. A more detailed discussion of this topic is provided in Attachment A to this response document.

Summary Comment – Topic 4:

Proposed allowable change in flow on the Ichetucknee River (IR). The proposed allowable change of 2.8% is out of the typical range for spring MFLs throughout the state of Florida and requires justification. The 15% change used as a standard to prevent significant harm for WRVs is based on precedent alone and not justified by the data presented in the report.

Summary Response – Topic 4: Both the percent time and SEFA results support the allowable change from the RTF/baseline condition, as does the 2013 MFL report and the current MFL Rule. The 15% threshold is discussed in detail in the PRP Resolution document and language is added to the final LSFI MFL report. The 15% time and area thresholds were used because they were utilized and acknowledged as part of the 2013 report and rule, are accepted as preventing significant harm, and are the most widely used threshold metric for setting MFLs for Florida rivers.

Specific Comments – Topic 4:

13. We note that the proposed allowable reduction of flow in the Ichetucknee River is 2.8 percent from the RTF. This appears to be among the lowest (or lowest) allowable reduction among spring systems that we can document (see Table 6) (see *original NFUCG submittal*). Generally, we believe that additional justification is warranted when an analysis results in an outlier value.

14. The most restrictive WRVs for the Ichetucknee River are based on a 15 percent change in overall temporal availability. This is the same methodology applied in the setting of the current MFL which was peer reviewed by the University of Florida (UF) Water Institute (Graham, et. al., 2013). The previous peer review noted the following:

> *"The justification for the proposed threshold of a 15% habitat loss in the establishment of the MFLs is based on precedent alone and cannot be justified on the basis of the data presented in the report. Although there is a precedent for the adoption of a 15% value, its general applicability is unproven." (Page 9)*

> *"To prevent significant harm MFLs should include considerations of duration and return interval of both low-flow and high-flow events in addition to cumulative frequency" (Page 10)*

> *"The panel recommends that the 15% threshold of change be more fully justified as it applies specifically to the Santa Fe and Ichetucknee Rivers." (Page 11)*

The UF Peer Review goes on to illustrate significant differences between WRVs which are consider duration and return interval as opposed to WRVs developed based on a 15 percent change in overall frequency. We agree with the UF Peer Review and believe that WRVs based on a scientifically-determined flowrate, duration, and return interval may be more appropriate and justified in the absence of information to the contrary. We are unable to find such information or even a discussion of this matter in the Draft MFL Report.

Specific Responses – Topic 4:

R13.The allowable flow reduction for the Ichetucknee River and Springs is very similar to the previous MFLs adopted by the District after a substantial peer review and administrative hearing. The previous allowable flow reduction at median flow conditions is 3 percent. The proposed net water available at median conditions under the proposed MFL is 10 cfs. Under the previous MFL it was 11 cfs.

The Ichetucknee River is classified as a first-magnitude spring system and an Outstanding Florida Spring group. The upper portion of the Ichetucknee River is a National Natural Landmark and is one of the most pristine spring runs in Florida. The river provides habitat to many species of plants, fish, and wildlife, including federally designated threatened and endangered species such as the Florida manatee and oval pigtoe mussel.

As compared to other MFLs for spring systems, the allowable flow reduction of 2.8 percent for Ichetucknee River is less restrictive than the 2.5 percent reduction for Silver Glen Spring. Additionally, three WRV metrics for Ichetucknee River would stand a higher chance of being significantly harmed with a flow reduction of 2.8 to 3 percent (10 - 11 cfs): woody habitat – exposed roots, floodplain habitat, and submerged aquatic vegetation (see Table 32 of the Final LSFI MFL re-evaluation report).

R14. The subjects of a 15% threshold and the Event Approach are addressed in detail in the District's responses to General Comments 10 and 11 and Attachment C of the Peer Review Resolution Document. The cumulative frequency approach (i.e., 15% change in exceedance), the 15% metric for Area Weighted Suitability (Jowett, et al. 2014), and wetted area in general are the most widely used criteria for setting MFLs and undergo regular verification as additional data are collected. And the criteria are not used without verification, as asserted by reviewer. The available water that results from the MFL is compared to default metrics developed by others, and the Indicators of Hydrologic Alteration (IHA) software is used to identify flow changes that might be problematic.

References

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., and Gilroy, E.J., 2020, Statistical methods in water resources: U.S. Geological Survey Techniques and Methods, book 4, chapter A3, 458 p., https://doi.org/10.3133/tm4a3 [Supersedes USGS Techniques of Water-Resources Investigations, book 4, chapter A3, version 1.1.]

Jowett, I., T. Payne, and R. Milhouse. 2014. SEFA: System for environmental flow analysis software manual version 1.21, New Zealand: Aquatic Habitat Analysts, Inc.

Locke, A., C. Stalnaker, S. Zellmer, K. Williams, H. Beecher, T. Richards, C. Robertson, A. Wald, A. Paul and T. Annear. 2008. Integrated Approaches to Riverine Resource Management: Case Studies, Science, Law, People, and Policy, Instream Flow Council, Cheyenne, WY. 430 pp.

Water Resource Associates, Inc. 2007. MFL establishment for the Upper Santa Fe River. Live Oak, FL: Suwannee River Water Management District.

ATTACHMENT A

NFSEG V1.1 Santa Fe Spring Review

I. LSFI NFUCG Comments (January 30th, 2019)

12. One potential source of error is the conductance value of the NFSEG model general head boundary (GHB) representing the Santa Fe Spring upstream of US 441. As shown in Table 4 (*see original NFUCG submittal*), this conductance value is significantly greater than other GHB cells along the LSFR. The high conductance for the Santa Fe Spring GHB has significant effect on the predicted pumping impacts at Santa Fe Spring from pumps off to 2001/2009 conditions. In fact, the impacts from pumps off at Santa Fe Springs represent approximately 90 percent of the total simulated impact at the US 441 gage, for both 2001 and 2009 stress periods of the NFSEG Model. And the simulated impact at Santa Fe Springs at is an outlier compared to all other springs on the LSFR which are included in the NFSEG model as presented in Table 4 (*see original NFUCG submittal*). This level of impact does not appear to be reasonable. Reducing the conductance value of the GHB representing the Santa Fe Spring to a value more representative of other values along the LSFR, results in better calibration of discharge at the US 441 gage. For example, as presented in Table 5 (*see original NFUCG submittal*), when the conductance value is reduced to approximately 600,000 ft2/day (which is greater than the average of LSFR spring conductance value), then the error in the LSFR baseflow at US 441 is reduced dramatically. In order to address this issue, we strongly recommend adjustment of this conductance value.

II. Representation of Springs in NFSEG V1.1

Springs in the NFSEG V1.1 model were represented using the General Head Boundary (GHB) package. Each NFSEG spring representation was assigned its own pool elevation in addition to its own conductance to enable the independent determination of discharge at each spring location. Spring pool elevations were obtained from observation data, stage estimates, estimated from topographic maps, or the USGS 3DEP 10m DEM representation. GHB conductance is scale dependent and is not readily observable and was therefore determined through automated calibration using PEST. The primary objective of the calibration process with respect to the determination of spring conductance was to enable the model to simulate spring flow rates adequately. In general, a greater emphasis was placed on matching the flows of larger springs and the total flows of spring groups, as the estimates of are generally more reliable due to greater data availability. They are more significant in terms of their effects on the flow system as well and therefore more critical to the overall representation of the groundwater system.

A summary of the spring information associated with Santa Fe Spring for the steady state 2001 and 2009 calibration years and 2010 verification year are included in Table 1 through Table 3. The conductance value estimated by PEST during model calibration for Santa Fe River Spring was 60,408,727 ft^2/day. It should be noted that, although this conductance value is higher than what was estimated for surrounding Lower Santa Fe River (LSFR) springs, the upper and lower bounds assigned to the Santa Fe Spring conductance parameter (sc4030) during model calibration were identical to those assigned to surrounding LSFR springs, 1E-05 ft^2/day and 1E+15 ft^2/day, respectively. Furthermore, conductance is not an observable property, so there are no estimates available with which to evaluate the conductance estimated through model calibration, except to compare how the model matches observed flows and groundwater levels. Additional information describing the implementation of PEST and representation of springs in the model can be found in Chapter 4 of the NFSEG V1.1 report (Durden et. al, 2019). The model underestimated spring discharge in 2001 (Table 1) and overestimated spring discharge in 2009 (Table 2) and 2010 (Table 3), when compared to observed flows for those periods. As stated above, an emphasis was placed on matching the flows of first magnitude springs and total flows of spring groups during model calibration due to greater data reliability and overall effect on groundwater flow.

Table 1. Spring information for Santa Fe Spring for the 2001 model calibration year.

Spring Name	GHB conductance (ft^2/day)	PEST target weight 2001	Pool Elevation 2001 (feet, NAVD88)	2001 Estimated Discharge (cfs)	2001 Simulated Discharge (cfs)	2001 Residual Discharge (cfs)
Santa Fe Spring	60,408,727	1.45	36.53	34.44	11.62	-22.82

Table 2. Spring information for Santa Fe Spring for the 2009 model calibration year.

Spring Name	GHB conductance (ft^2/day)	PEST target weight 2009	Pool Elevation 2009 (feet, NAVD88)	2009 Estimated Discharge (cfs)	2009 Simulated Discharge (cfs)	2009 Residual Discharge (cfs)
Santa Fe Spring	60,408,727	1.01	37.94	49.64	64.28	14.64

Table 3. Spring information for Santa Fe Spring for the 2010 model verification year.

Spring Name	GHB conductance (ft^2/day)	Pool Elevation 2010 (feet, NAVD88)	2010 Estimated Discharge (cfs)	2010 Simulated Discharge (cfs)	2010 Residual Discharge (cfs)
Santa Fe Spring	60,408,727	37.54	70.86	87.29	16.42

III. Review of Available Observation Data at Santa Fe Spring

Uncertainty in the values of spring flow targets is a function of a variety factors. These include uncertainties in field measurements, rating curve estimation error (for spring measurements that are based on ‘rating curves’ that are fit to field measurements), insufficient numbers of measurements (time sampling errors), and departures from steady state assumptions (in which spring flows used to calibrate a model are generated by significant changes in storage during the calibration period, in addition to average re-charge rates during the calibration period). Some studies indicated that the errors in individual flow measurements could range from 2 to 20 percent (Sauer and Meyer 1992). The spring flow targets for selected for Santa Fe Springs were reviewed relative to available measured and modeled data for 2001 and 2009. The selected targets were reasonable based on the available measured and modeled data for the spring.

IV. Santa Fe Spring GHB Conductance Sensitivity Simulations

To evaluate the model sensitivity to the GHB conductance at Santa Fe Spring, the calibrated conductance value of 60,408,727 ft^2/day was multiplied by 0.01, 0.1, 0.5 and 1.5. For each sensitivity simulation, the 2001 and 2009 calibration and pumps off simulations were run with the modified GHB conductance and post-processed to extract flows at springs and gages. Simulated 2001 and 2009 flows at Santa Fe Spring and selected gages along the LSFR are compared in Table 4 and Table 5, respectively. The change in flow between 2009 and the 2009 pumps off simulations is shown in Table 6. This analysis is limited to the evaluating impacts at important locations along the LSFR that were highlighted by the NFUCG comments and most likely to be impacted by changes in spring conductance. To more thoroughly evaluate the model implications of decreasing the conductance at Santa Fe Spring, model calibration and verification statistics for observation groups, including groundwater levels, should be calculated.

The results summarized in the Table 4 through Table 6 show that the model is sensitive to the GHB conductance at Santa Fe Spring. This suggests that the spring conductance value estimated through automated calibration at this spring was systematically adjusted, in combination with other model parameters, to minimize model error at provided observation points in 2001 and 2009 simultaneously. A similar conclusion was determined by parameter sensitivity analyses performed using the model (see NFSEG V1.1 report Chapter 7) that indicated simulated spring flows were highly sensitive to spring conductance while simulated baseflows were moderately sensitive (Durden et al., 2019).

Although it is suggested that reducing the conductance by two orders of magnitude improves calibration at the US441 gage, the effects on Santa Fe Spring also need to be considered since this spring was utilized as calibration target by PEST. Simulated spring flow in 2001 at Santa Fe Spring was increasingly underestimated as the GHB conductance was decreased when compared to the flow estimate of 34 cfs (Table 4). Adjusting the conductance to approximately 600,000 ft^2/day (CONDX0.01 in Table 4) reduced simulated flow at Santa Fe Spring to 2.31 cfs, approximately 32 cfs below the estimated spring flow of 34 cfs in 2001. This scenario resulted in a percent error of 93% at Santa Fe Spring, up from 66% in the 2001 calibrated simulation (CONDX1.0 in Table 4). Similarly, the percent error at Santa Fe Spring in 2009 increased from 29% to 75% by reducing the conductance to 600,000 ft^2/day (CONDX0.01 in Table 5).

Decreasing the conductance by two orders of magnitude at Santa Fe Spring reduced simulated flows at the Fort White gage from 532 cfs to 528 cfs in 2001, slightly increasing the percent error at this gage based on an estimated flow of 563 cfs (Table 4). Similarly, a conductance of 600,000 ft^2/day at Santa Fe Spring reduced simulated flows at the Fort White gage from 724 cfs to 699 cfs in 2009, increasing the model percent error from 1% to 4% based on an estimated flow of 730 cfs (Table 5).

It should be noted that cumulative baseflow at US441 gage was not used as a calibration target by PEST in the NFSEG V1.1 model. The implication of this is that modelled or observed total baseflow at the US441 gage was not used to adjust parameter values, and therefore did not contribute to the objective function used to be PEST to estimate spring conductance values. Based on the estimated flows provided in Table 5 of the attached comments at the US441 gage, the model is overestimating flow in 2001 by 16 cfs (CONDX1.0 in Table 4) and in 2009 by 38 cfs (CONDX1.0 in Table 5). Reducing the conductance at Santa Fe Spring by two orders of magnitude does improve simulation of flow at this gage in 2001 (Table 4) and 2009 (Table 5), as well as reduce the impact from pumping (Table 6), however, it results in an increase in residual flow at Santa Fe Spring and the Fort White gage. There is no information to support that the conclusion that PEST overestimated conductance at this spring, or that reducing the conductance at Santa Fe Spring would improve the regional NFSEG V1.1 model calibration in either calibration year. Future versions of the model may include additional calibration targets, such as cumulative baseflow at the US441 gage, to improve calibration in this region of the model.

Table 4. Simulated flow at selected springs and gages in 2001 for each adjusted GHB conductance simulation. Each column 'COND' represents a Santa Fe Spring conductance multiplier simulation and the simulated discharge results (cfs).

Spring/ Gage	**2001 Estimated Discharge (cfs)**	**CONDX1.0 60,408,727 ft²/day**	**CONDX0.5 30,204,364 ft²/day**	**CONDX1.5 90,613,091 ft²/day**	**CONDX0.1 6,040,873 ft²/day**	**CONDX0.01 604,087 ft²/day**
Santa Fe Spring	34.44	11.62	11.17	11.78	8.50	2.31
Fort White	562.69	532.30	532.08	532.37	530.83	527.92
US 441	20*	35.81	35.42	35.94	33.19	27.99

*Cumulative baseflow at the US 441 gage was not used as a calibration target in NFSEG V1.1, although a baseflow target estimate was produced for during the NFSEG development process.

Table 5. Simulated flow at selected springs and gages in 2009 for each adjusted GHB conductance simulation. Each column 'COND' represents a Santa Fe Spring conductance multiplier simulation and the simulated discharge results (cfs).

Spring/ Gage	**2009 Estimated Discharge (cfs)**	**CONDX1.0 60,408,727 ft²/day**	**CONDX0.5 30,204,364 ft²/day**	**CONDX1.5 90,613,091 ft²/day**	**CONDX0.1 6,040,873 ft²/day**	**CONDX0.01 604,087 ft²/day**
Santa Fe Spring	49.64	64.28	61.73	65.18	46.85	12.62
Fort White	730.42	723.58	722.36	724.01	715.26	698.91
US 441	108*	146.39	144.25	147.14	131.78	103.09

*Cumulative baseflow at the US 441 gage was not used as a calibration target in NFSEG V1.1, although a baseflow target estimate was produced for during the NFSEG development process.

Table 6. Simulated change in flow at selected springs and gages between 2009 and 2009 pumps off for each adjusted GHB conductance simulation. Each column 'COND' represents a Santa Fe Spring conductance multiplier simulation and the simulated discharge results (cfs).

Spring/Gage	2009 Estimated Discharge (cfs)	CONDX1.0 60,408,727 ft^2/day	CONDX0.5 30,204,364 ft^2/day	CONDX1.5 90,613,091 ft^2/day	CONDX0.1 6,040,873 ft^2/day	CONDX0.01 604,087 ft^2/day
Santa Fe Spring	49.64	43.12	41.41	43.72	31.43	8.46
Fort White	730.42	68.76	67.94	69.04	63.19	52.19
US 441	108*	49.58	48.15	50.09	39.79	20.54

*Cumulative baseflow at the US 441 gage was not used as a calibration target in NFSEG V1.1, although a baseflow target estimate was produced for during the NFSEG development process.

V. Considerations for Evaluation of Baseflow Simulations

There is uncertainty relevant to the development of baseflow targets for river gages. Uncertainty regarding baseflow estimates is widely acknowledged. ASTM states that "…baseflow estimates are generally accurate only to within an order of magnitude" (ASTM 2018), as noted in the NFSEG v1.1 final report. With ratios of simulated baseflow to estimated baseflow of 1.79 and 1.36, respectively, the NFSEG v1.1 2001 and 2009 corresponding estimated and simulated baseflows at the US 441 gage are easily within an order magnitude of one another. This lends credence to the general magnitudes of the estimates and corresponding simulated values, suggesting both sets of values are reasonable. Regarding this topic, the NFSEG v1.1 final report states that "Even in the absence of extreme conditions or conditions that pose challenges for estimating flow records, baseflow estimation is commonly subject to high levels of uncertainty."

Other comparisons indicate reasonableness in the 2001 and 2009 baseflow estimates. For instance, comparing the average total flows at the US 441 gage in 2001 and 2009 to the respective baseflow estimates, we see that the respective baseflow percentages of total flow are 44.7 and 50.1, which generally fall within a range of such values as derived for the other gages represented in the model. The percentages imply that the baseflow estimates are less than the corresponding values of total flow but of the same general magnitude (44.7 cfs total flow vs. a 20 cfs baseflow estimate in 2001 and 215.6 cfs total flow vs. a 108 cfs baseflow estimate in 2009). However, the simulated values of baseflow meet the same general criteria (44.7 cfs total flow vs. a 35.8 cfs simulated baseflow in 2001 and 215.6 cfs total flow vs. a 146.4 cfs simulated baseflow in 2009).

VI. Conclusion

District staff performed a detailed review based on the concerns raised by the NFUCG. This review included a review of the data for Santa Fe Spring, the regional river gages as well as estimates of baseflow. Model simulations were run to evaluate the impact of changing spring conductance as proposed by the NFUCG. The Districts have concluded that the calibrated model is performing appropriately in this region based on the available data.