**5.A DETAILED COMMENTS FOR MODFLOW**

**5.A.1 Detailed Comments for MODFLOW (Hal Davis)**

**Chapter 3: MODEL CONFIGURATION:**

Model Code Selection page 25

The selection of MODFLOW-NWT is an appropriate choice for the following reasons:

1. The ability to rewet dry cells (since two steady-state models are used).
2. The code is fully 3-dimensional (which fits the appropriateness requirements of ASTM (2010),
3. MODFOLW is widely accepted (a requirement recommenced by ASTM 2010),
4. MODFLOW is credible (a requirement recommenced by ASTM 2010),
5. MODFLOW is well documented (a requirement recommenced by ASTM 2010),
6. MODFLOW has readily available graphical user interfaces (a recommendation by ASTM 2010),
7. The choice of a finite-difference model code over a dual-porosity model code also seemed reasonable, considering that regional groundwater flow modeling was the goal. A dual-porosity model code would only be an advantage if a considerable number of karst dissolution caves were known, and, only a very small percentage of these features have been delineated in the study area.

NFSEG Grid page 27

The choice of square grid cells, 2,500 feet per side, seems appropriate. Given that this resulted in 752 rows and 704 columns (with 7 layers) creates a model that will have long run times and a smaller grid size would probably have resulted in a model that would have had unacceptably long solve times.

Model Layers page 27

Using separate model layers (generally) assigned to each of the major hydrologic divisions of the groundwater flow system was good. Although, having the UFA cross layers 1 and 2 did add some complication to the documentation.

Lateral Boundary Conditions page 32

Extending the lateral model boundaries to the natural groundwater flow system boundaries was appropriate, as recommended by Reilly and Harbaugh (2004) and ASTM (2008). Where this was not possible, the boundaries were placed away from the critical model area and along groundwater flowlines, also recommended by Reilly and Harbaugh (2004). The model boundaries were well discussed as recommended by ASTM (2008).

**Chapter 4: MODEL CALIBRATION:**

Groundwater Levels page 45

Table 4-1. NFSEG PEST Observation Groups: The observation groups labeled Temporal head differences were not described in the text.

Page 45’ last paragraph: Report states that “Statistical methods were used to augment the number and quality of water level observations in areas of limited water level data availability, as detailed in Appendix A.” Appendix A is just a list of values and there is no description of the statistical methods used.

For the Suwannee river in 2001, the simulated cumulative flows between the Ellaville and Wilcox gages (river reaches in contact with the UFA) matches the measured cumulative flows at the gages pretty well (Table 1). This indicates that overall the recharge in the groundwater basin near the Suwannee and lower Withlacoochee rivers is pretty good.

Similarly, in 2009, the simulated cumulative flows between the Ellaville and Wilcox gages also matches the measured cumulative flows pretty well (Table 2). Again, this indicates that the recharge in this part of the model is reasonably close.

The baseflow pickups on the Withlacoochee, Alapaha, Suwannee river (figures 2-38 and 2-41) show an erratic pattern, with large net gains in some reaches and followed by small net gains (or negative gains) in other reaches. This could be a function of the hydrology and accurate; or it may be indicating errors in the estimated baseflow pickups. If it is the latter, then the cumulative flows match better because the errors are averaging out. And this highlights the need for the best estimates of the net baseflows possible.

Near the GA-FL line, the simulated cumulative river baseflows are significantly higher than the measured flows in the Withlacoochee and Alapaha rivers indicating that the recharge rates in this area maybe too high. Groundwater to these river reaches is recharged to the north where there is a significant thickness of surficial aquifer/intermediate confining unit sediments. Thus, recharge occurring in 2001 and 2009 will probably take many years to make it to the UFA. For this situation the HSPF recharge rates could be used as guide but may need to be adjusted because of the movement through surficial aquifer/intermediate confining unit (which will tend to average multiple years of recharge).

For the Santa Fe river the simulated and measured cumulative flows are very close indicating the method of calculating recharge rates are about right.

Table 5A-1. Year 2001 measured and simulated cumulative flows.



Table 5A-2. Year 2009 measured and simulated cumulative flows.



Recharge and Maximum Saturated ET Multipliers page 54

Not varying the recharge manually or using PEST essentially sets the amount and distribution of groundwater across the model (except for some boundary conditions and lake leakage) to the HSPF values. For PEST to match water levels, river baseflows, and spring baseflows during calibration only the hydraulic conductivities can be varied (except for some boundary conditions and lake leakage). During a PEST run, if a baseflow has a high residual, PEST can only vary the hydraulic conductivities in an attempt to lower the residual (when changing the recharge may be more appropriate). This may force PEST to use inappropriate hydraulic conductivities to make up for an inappropriate recharge rate. It is difficult to know how much of the error in the river baseflow matches are due to this, but as the report states, the match to the river baseflows is poor. It is also difficult to know the effect of this on the parameter estimation of the hydraulic conductivities. But as seen in figures 4-76 and the figures in the 0b Additional\_Info\_Request\_20180507\_Final.pdf there is not a strong correlation between measured and simulated hydraulic conductivities and transmissivities.

Horizontal Hydraulic Conductivity of Layer 1 page 65

Figure 4-70: Need to add measured horizonal hydraulic conductivity values where available.

Horizontal Hydraulic Conductivity of Layer 5 page 65

Figure 4-73: Measured transmissivities should be posted on the map where available.

Transmissivity of Layer 3 page 66

The figures below need a linear regression line added (with mean, absolute value mean, standard deviation, and R-squared).

Figure 4-76. Multi-Well-APT-Derived Transmissivity versus Calibration-Derived Transmissivity (Feet Squared per Day), Upper Floridan Aquifer

Figure 2. NFSEG UFA Transmissivity vs. USGS Sim 3204 APT Wells – Confined Region

Figure 3. NFSEG UFA Transmissivity vs. USGS SIM 3204 APT Wells - Unconfined Region

Figure 5. NFSEG UFA Transmissivity vs. NFSEG APT Database – Confined Region

Figure 6. NFSEG UFA Transmissivity vs. NFSEG APT Database - Unconfined Region

**Chapter 6: WATER BUDGET ANALYSIS page 81**

For figures 6-3 to 6-33, many of the figures have arrows showing flow in only one direction for each model layer, indicating flow only goes in one direction. It would probably be more appropriate in many cases to have arrows pointed in both directions indicating flow both into and out of a layer.

Predictive Uncertainty Analysis Results page 95

On page 55, the report states that structural errors are typically the dominant source of errors in groundwater models. There should be some discussion on how to interpretive the uncertainty analysis considering that structural errors are included in model calibration.

**5.A.2 Detailed Comments for MODFLOW (Louis Motz)**

**Chapter 1 Introduction**

**p. 3 and Figure 1-5:** **Long-term average rainfall within the model domain is approximately 50 inches.** Annual rainfall should be expressed in ***inches per year***. Also, **Lake City** and **Live Oak** are not plotted in their correct locations in Figure 1-5. Averaging the rainfall totals in Figure 1-5 yields 45.3 inches for 2001, 53.2 inches for 2009, and 51.8 inches per year for the long-term average. The average value of 51.8 inches per year is somewhat different from 50 inches per year. Does 50 inches per year represent a spatially weighted average or a longer term value? Please explain.

**Chapter 2 Hydrology of the Area**

pp. 12 and 13: In paragraph discussing **hydraulic properties of the intermediate confining unit**, consider adding: **“Higher leakance values on the order of as much as 10-3 day-1 have been determined beneath some of the karstic lakes in Keystone Heights.”**

**Chapter 3 Model Configuration**

**pp. 25-26: The simulation of groundwater flow within the Floridan aquifer in the surrounding [add area?] and including Wakulla Springs and its network of mapped and inferred conduits using the standard MODFLOW approach…was shown by Kuniansky (2016) to compare well to that of an alternative MODFLOW model in which conduit flow to Wakulla Springs was represented more rigorously using the MODFLOW Conduit Flow Package….The results of the study indicated that the presence of conduits…should not necessarily preclude application of the standard Darcian flow approach…for simulation of flows averaged over a month or longer (Kuniansky 2016)….Based on these results,...the standard MODFLOW approach is assumed to be applicable throughout the NFSEG model domain.**

This obviously is a major assumption for the applicability of the NFSEG model. Are Kuniansky’s (2016) results based only on the temporal, i.e., monthly, requirement for averaging flows, or is the scale of discretization (Δx by Δy) and/or other factors considered as well by Kuniansky (2016)? Are there other USGS or other published studies where the issue of simulating conduit flow in regional groundwater flow models is investigated? If so, what are the results and conclusions of these studies?

**pp. 27-29: “…minimum thickness approach…”** Has the minimum thickness approach been used in other comparable regional groundwater flow models or is this approach unique to this study? If it has been used before, what are the results and conclusions, i.e., was the application of this approach successful?

**Chapter 4 Model Calibration**

**p. 46, Table 4-1:** What are “***temporal* head differences”** in layers 1-7?

**p. 48:** Estimating lake leakance rates as the **“difference between rainfall and potential evapotranspiration”** assumes that all other water-budget components including the change in storage (dS/dt) are negligible or that the sum of all of the other inflows and outflows and dS/dt = 0. Was this assumption verified for any of the lakes? Is this a potential source of error?

**pp. 58-59:** The **residual statistics** in Table 4-4 should be compared with residual statistics that have been obtained for other comparable regional groundwater flow models, e.g., steady-state results in Sepúlveda et al. (2012). Please refer to a similar comment pertaining to p. 101 in the **Summary and Conclusions** chapter.

**pp. 60-63, Figures 4-13, 4-14, 4-23, 4-24, 4-37, 4-38, 4-43, 4-44, 4-45, 4-46, 4-53, 4-54, 4-57, 5-58:** The number of data points (n = ) used in each plot should be indicated on each of the plots for the observed versus simulated hydraulic heads, observed versus simulated spring discharges, observed versus spring-group discharges, estimated versus simulated baseflow pickups, and estimated versus simulated cumulative baseflows. Histograms of residuals (simulated minus observed values) for each of the plots should be plotted and evaluated to determine whether the residuals are normally distributed about their means or skewed to the left or right.

**p. 66 and Figure 4-76:** The discussion and plot of model-derived transmissivities and transmissivities derived from aquifer performance tests (APT’s) is a good first step toward validating the model-derived transmissivities for the Upper Floridan Aquifer. In Figure 4-76, the number of APT results should be indicated, and a figure should be added showing the location of the APT’s from which the transmissivity values were obtained. A table listing the APT locations, APT-and corresponding model-derived transmissivity results, and other details such as pumping rates and numbers and depths of pumped and observation wells should be provided in an appendix. Additional discussion should include a comparison of the scale of the APT’s in terms of the affected aquifer area and/or volume compared to the discretization of the model, i.e., are the scales of impact of the pumping tests selected for Figure 4-76 comparable to the discretization of the model? In addition to the line of equality shown in Figure 4-76, a statistical line of best fit should be plotted. These results should indicate (with a weak correlation) whether the model-derived transmissivities are greater or less than the APT-derived transmissivities at the same locations. Finally, have similar model- and APT-derived transmissivity results been determined and reported for other comparable groundwater flow models? If so, these results should be referenced and compared to the results shown in Figure 4-76.

**Chapter 5 Model Simulation**

**p. 68:** Model-wide values for rainfall and ET should be provided for 2001, 2009, and 2010, and for the long-term mean.

**p. 70, Figure 5-10:** Are all of the observation wells shown in Figure 5-10 located in the Upper Floridan aquifer (layer 3)? Please make it clear in the text and on the figure in which aquifer layer(s) the various wells are located.

**p. 72 and Figure 5-14:** The residual groundwater level statistics for model layers 1, 3, and 5 indicate a very good result for all three layers (layers 1, 3, and 5).

**p. 72 and Figure 5-16:**  The residual spring discharge statistics indicate reasonably good results for spring flows for 2001, 2009, and 2010. Can spring conductances be adjusted to improve the results?

**p. 73 and Figure 5-18:**  The residual baseflow pickup statistics indicate reasonably good results for baseflows for 2001, 2009, and 2010, given the difficulties in estimating “observed” values.

**p. 74 and Figure 5-20:**  Similar to the results for the residual baseflow pickup statistics, the residual cumulative baseflow statistics indicate reasonably good results for cumulative baseflows for 2001, 2009, and 2010, given the difficulties in estimating “observed” values.

**p. 75 and Figures 5-23 and 5-24:** The simulated 2010 UFA potentiometric surface compares very favorably with the observed 2010 UFA potentiometric surface.

**pp. 75-76, Table 5-5:** Table is titled **“Comparison of simulated net fluxes into the model in 2010, compared to 2001 and 2009”**, but Table 5-5 apparently contains the distribution of water-level residuals for Layer 3 by GWB. There appears to be a missing Table 5-6 that compares net fluxes in 2001, 2009, and 2010, which is an important result. If Table 5-6 is missing and needs to be added, then the references to **“Table 5-6”** on p. 78 and the table title on p. 79 need to be changed to **Table 5-7.**

**pp. 77 and 78 and Figures 5-26 and 5-27:** The USGS pre-development UFA potentiometric surface map and the simulated no-pumping layer 3 potentiometric surface appear to match reasonably well. However, would a simulated no-pumping ***UFA*** potentiometric surface map be significantly different from the simulated no-pumping ***layer 3*** potentiometric map? Also, a plot of the simulated layer 3 (or UFA) map should follow Figure 5-26 before overlaying the two maps as shown in Figure 5-27.

**pp. 78 and 79 and Table 5-6 (should this table be re-numbered Table 5-7?):** As noted, the no-pumping simulated spring discharges compare favorably with the ranges and means of the discharges observed by Stingfield (1936) except for Juniper and White springs. Can the results for these springs, particularly White Springs, be improved?

**Chapter 6 Water Budget Analysis**

**Tables 6-1 to 6-32:** These tables need to be referred to in the text. Also, in addition to the water-budget information provided in **Tables 6-1 to 6-4**, tables that show the inflow and outflow water-budget components in inches per year for all layers in total should be provided for the 2001, 2009, 2010, and the 2009 no-pumping simulations.

**Figures 6-2 to 6-33:** In addition to these figures, figures that show the inflow and outflow water-budget components in inches per year for all layers in total should be plotted for the 2001, 2009, 2010, and the 2009 no-pumping simulations similar to the following figures:



Figure 5A-1. Model-Wide Water Budget for 2001



Figure 5A-2. Model-Wide Water Budget for 2009



Figure 5A-3. Model-Wide Water Budget for 2010



Figure 5A-4. Model-Wide Water Budget for 2009 Pumps Off

**Chapter 7 Sensitivity and Uncertainty Analysis**

**p. 91, Figures 7-1 to 7-6:** The results of the sensitivity analysis demonstrate very clearly that the standard deviations of the residuals for simulated groundwater levels, baseflows, and spring flows have been minimized by the values for recharge, evapotranspiration, aquifer parameters, and boundary conditions in the calibrated model.

**pp. 93-95: Uncertainty Analysis and Hypothetical 2035 Pumping Scenario….**How was this scenario derived? How were values derived for pumping, recharge, evapotranspiration, boundary heads, and other water-budget components derived? A better explanation needs to be provided.

**Chapter 8 Model Limitations**

**p. 97: “The results…indicate that the model can be used…with the same level of accuracy as existing regional-scale models….”** A table identifying other regional-scale models and providing the calibration statistics for comparison with this model should be included.

**p. 97: “…NFSEG individual grid cells are relatively small…and their size is comparable to or better than other existing groundwater models.”** A table identifying other regional-scale models and providing the details about discretization size and area of model domain for comparison with this model should be included.

**p. 98: The effects of lateral boundaries may limit the accuracy of model results near lateral boundaries.** How restrictive is this result? Is it possible to estimate the area of the model in which these limitations apply compared to the total model area, i.e., does this limitation significantly affect the overall accuracy and usefulness of the model or is it limited to a very small proportional area of the model?

**Chapter 9 Summary and Conclusions**

**p. 101: “The 2001 and 2009 steady state simulations yielded reasonable head and springflow residuals. Although the calibration goals were not fully met…it is important to note that the calibration goals were not intended as absolute requirements but as ambitious goals…”** The residual statistics in Table 4-4 and in Figures 5-14, 5-16, 5-18, and 5-20 should be compared with residual statistics that have been obtained for other comparable regional groundwater flow models, e.g., steady-state results in Sepúlveda et al. (2012).

**5.A.3 Detailed Comments for MODFLOW (Dann Yobbi)**

**Chapter 3. Model Configuration**

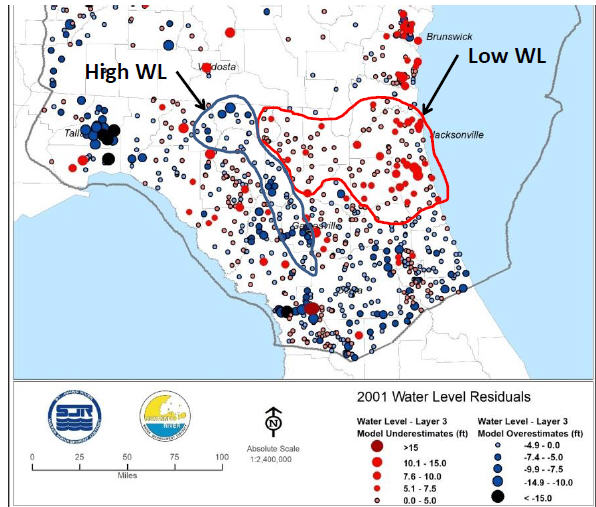
1. The discretization used in this model is appropriate at annual time steps and at regional scales but exercise extreme caution when simulating at local scales and shorter time steps. Although this report states that the NFSEG model can be used for “local-scale evaluations with the same level of accuracy as existing regional-scale models” (p. 97), this model should cautiously be used to interpret local-scale conditions. The regional model scale (grid size) likely is too coarse to accurately simulate the hydrologic behavior of individual spring discharge, focused river discharge, fluxes to streams, or discharge from wells. If details of these hydrologic features are of interest in the future, a smaller grid size (finer resolution) should be implemented. Finer resolution can be easily achieved using the MODFLOW LGR package because it allows significant spatial variations, or local grid refinement.

**2.** It should be noted that the model does not actually simulate flow from individual springs and river reaches, rather the model provides volumetric fluxes from potentially multiple sources (wells, seepage, spring flow, etc.) from 0.22 mi2-sized cells. The PEST code allows conductance values to vary without any apparent constraints to achieve (match) the desired volumes. Individual spring and river flows were accounted for through use multiple conductance terms associated with GHB and River assignments. The model can be teased into portioning and assigning fluxes by “source” by multiple occurrences and types of assignments (GeoHydros, 2014). Separate assignments provide a method for parsing the flows making comparisons of simulated fluxes to individual spring flows and river reach fluxes. These manipulations should be explicitly stated and transparent to stakeholders.

**Chapter 4. Model Calibration**

**3.** An important characteristic of an accurate model (well calibrated) is the spatially random distribution of weighted residuals (Poeter and Hill, 1997). However, model-wide trends in the spatial distribution of water-level residuals are apparent. The NFSEG report states “*the residual map of 2001 shows a relatively high concentration of underestimated and overestimated water levels along the coast of northeast Florida and southeast Georgia. For both calibration periods, clusters of relatively large residuals occur in Leon, Wakulla, and Citrus Counties, Florida.*” A contour map provided by the stakeholders verified other spatial trends in the water-level residuals as well (see Figure 5A-5). A non-random, spatial distribution in residuals often indicates model bias and possible model error. What is the hydrologic/hydrogeologic significance of these trends?

**2001**



**2009**

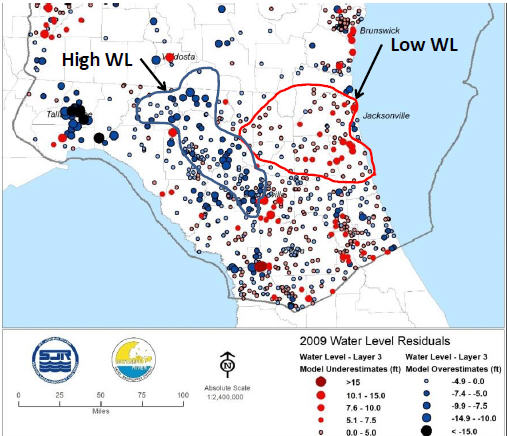


Figure 5A-5. Examples showing selected model-wide trends in spatial groupings of positive and negative water-level residuals in layer 3 (written com. Liquid Solutions Group, LLC, 2018).

**4.** An important calibration target omitted from evaluation is the head distribution along the rivers and at discharging river nodes. However, the report fails to demonstrate the degree to which the model was able to match river elevations along with river and spring flows. In terms of calibration, an acceptable match to spring and river flow requires an appropriate value of stage and aquifer head. If an accurate head difference is not simulated, the resulting calibrated conductance values may become too low/high, reducing/increasing the predicted effects of induced stresses on streamflow/spring flow.

**5.** Some of the report figures are misleading by presenting flows as negative values in the scatter plots. This is counter intuitive and not consistent with published modeling reports. The report statistics, figures, and table should be revised to provide positive values of flow; residuals are “difference” calculations and may be negative. All residuals should be computed as the difference between observed minus simulated and stated on each figure and table.

**Chapter 5. Model Simulations**

**6**. **No Pumping Simulation:** Realistic results from the no pumping scenario are unreliable using this application of the steady-state calibrated NFSEG model because return flow is included in the no pumping simulation (see Figure 5A-6). Conceptually, return flows are how excess water is recharged to the groundwater system from pumping; there can be no return flow in the absence of pumping.

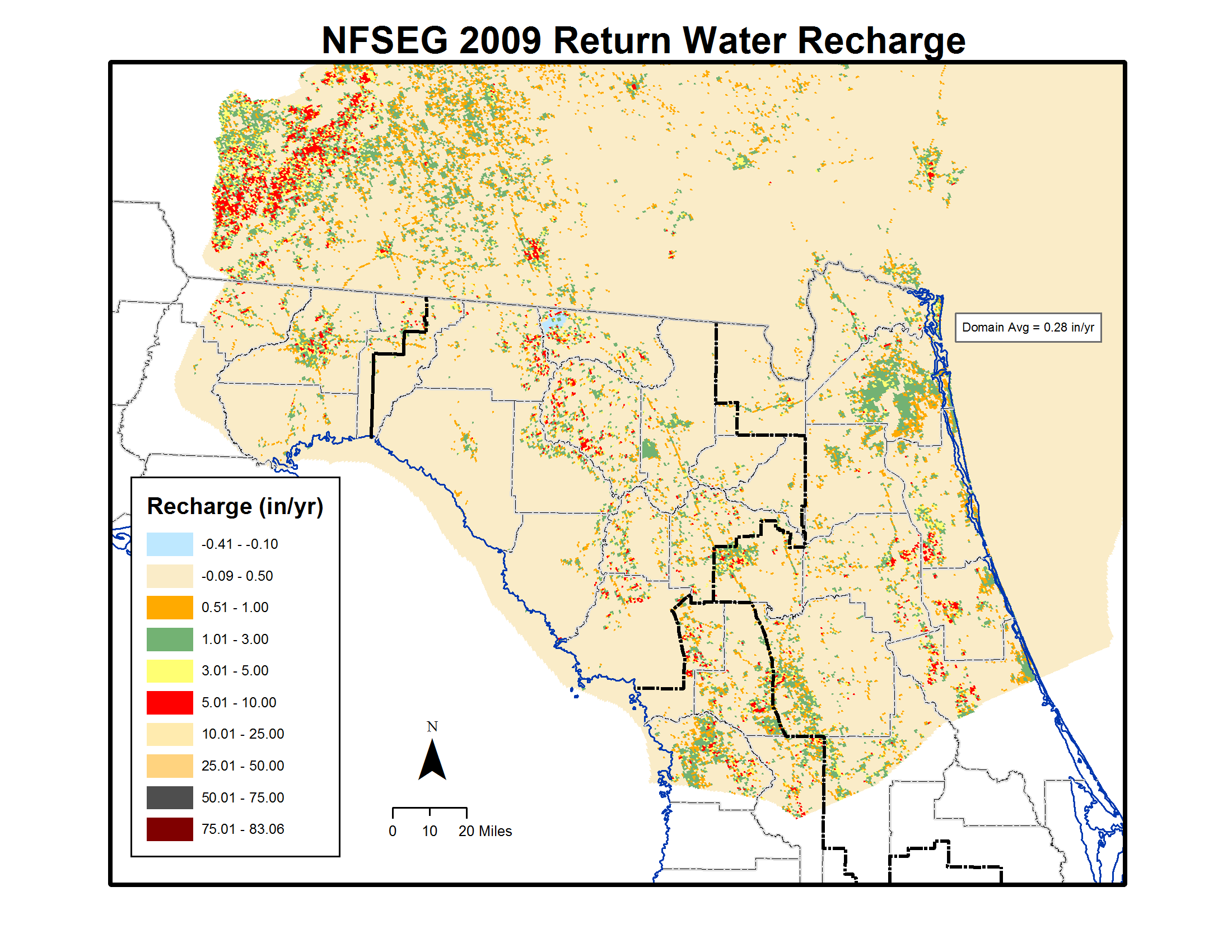


Figure 5A-6. NFSEG V1.1 2009 Return Water Recharge (written communications SWFWMD, 2018)

If the no pumping scenario is desired the following requirements should be adopted: (1) return flow must not be included in the simulation; (2) evaluate and provide quantitative determinations including magnitude and consistency of spring and baseflow, and **key** **water budget components**; (3) provide quantitative comparison between pumping and no pumping heads and fluxes to evaluate reasonability of simulation results; and (4) running transient model simulations to determine if transient conditions improve model results for evaluating the effect of ground water withdrawals (or lack thereof) on heads and flows. Andersen and Stewart (2016) discuss other scenarios that may be worthwhile for determining baseline (pre-development) conditions for comparison to future pumping scenarios: (1a) vary pumping rates by a specified percentage of plus or minus 50 percent or (1b) vary by actual pumping rates rather than percentages--then plot predicted flows against total pumpage and determine intercept; and (2) setting model transient run time (simulation time) under a “pumps off” condition to one or more years.

**Chapter 6. Water Budget Analysis**

**7.** The ability to evaluate the water balance by groundwater basin is a very useful metric to confirm the model’s accuracy. A weakness of the report, however, is the omission of an independent water budget compiled for the model domain and a comparison between the independent and simulated water budgets. This comparison is needed because no statements are made in the report about what would constitute an acceptable mass balance in any part of the model. Annual quantities that can be obtained from the independent water budget and that can are directly comparable to the simulated values are natural recharge at the water table, net recharge to the UFA, baseflow, and spring flow.

**Chapter 8. Model Limitations:**

**8.** Model limitations section should include the following: 1) Grid size limitation on lakes, wetlands, and streams. 2) Results should be interpreted at scales larger than the represented grid cells and should cautiously be used to interpret local-scaled conditions. 3) The model is not unique and many combinations of aquifer properties and recharge-discharge distributions can produce the same results. 4. The selection of a 0.22mi2 uniform orthogonal grid limits the model’s ability to simulate complicated geometries characteristic of the rivers and springs in the model domain (the absence of a discussion of this misleads readers and model users into believing that the chosen approach and software represent either the only or best available option).