INDEPENDENT TECHNICAL PEER REVIEW
OF
NORTH FLORIDA SOUTHEAST GEORGIA GROUNDWATER MODEL
(NFSEG v1.1)

Prepared by

NFSEGv1.1 Technical Peer Review Panel

Louis H. Motz, Ph.D., P.E., D.WRE, Chair
Brian R. Bicknell
J. Hal Davis, P.G.
James Rumbaugh, P.G.
Dann Yobbi, P.G.

Prepared for

St. Johns River Water Management District
Palatka, Florida
and
Suwannee River Water Management District
Live Oak, Florida

August 22, 2018

CONTENTS

1. INTRODUCTION

The North Florida Southeast Georgia (NFSEG) groundwater model is being developed by the St. Johns River Water Management District (SJRWMD) and the Suwannee River Water Management District (SRWMD) to provide a shared tool that can be used by both water management districts to assess the impacts of current and future groundwater withdrawals on water resources in north Florida. The model encompasses parts of Florida, Georgia, and South Carolina covering an area of approximately 60,000 square miles. The model is fully three-dimensional and utilizes seven layers to represent the surficial aquifer system, the intermediate confining unit, the Upper Floridan aquifer, the middle semiconfining unit, the upper zone of the Lower Floridan aquifer, the lower semiconfining unit, and the Fernandina Permeable zone of the Lower Floridan aquifer where these hydrogeologic units are present. In its present form, the model has been calibrated to steady-state hydrologic conditions representing 2001 and 2009. To improve initial estimates of recharge and maximum saturated evapotranspiration for input to the NFSEG groundwater model, surface-water models have been developed for all surface-water basins within the groundwater model boundaries using the Hydrological Simulation Program-FORTRAN (HSPF) software. Version 1.0 of the NFSEG groundwater model and the HSPF-derived surface-water models was completed in 2016 and distributed in August 2016 to stakeholder groups that consisted of government organizations, water utilities, private industry, and environmental organizations and other interested parties throughout north Florida and south Georgia for their use and review. Version 1.1 of the NFSEG groundwater model and the HSPF-derived surface-water models has been developed to address changes and improvements recommended for Version 1.0. Preliminary calibration results for Version 1.1 of the NFSEG groundwater model and the HSPF-derived surface-water models were completed in May 2017, and documentation and model files of Version 1.1 of the NFSEG and HSPF models were completed for final peer review in April 2018.

A panel of modeling experts was convened by SJRWMD and SRWMD in March 2017 to provide independent technical peer review of the NFSEG groundwater model and the HSPF models as the final phase of Version 1.1 of the model was being developed. This was intended to provide opportunities for the SJRWMD and SRWMD modeling team to incorporate peer review suggested changes into the model as it was being completed. Responsibilities of the Peer Review Panel included conducting a thorough review of the groundwater and surface-water models and model documentation reports and assessing the following topics:

- Model objectives, conceptualization, and design;
- Assumptions and limitations of input data;
- Model calibration and sensitivity;
- Model documentation (explanation of model, data sources, and assumptions);
- Suitability of MODFLOW and related HSPF models for the intended applications;
- Appropriateness, defensibility, and validity of the model/relationships;
- Validity and appropriateness of all assumptions used in the development of the model/relationships; and

- Deficiencies, errors, or sources of uncertainty in model/relationship development, calibration, and application.

Also, the Peer Review Panel has provided answers to a set of questions concerning model documentation, implementation, calibration, and application including questions listed in Appendix A of the Charter of the SJRWMD – SRWMD Cooperative Groundwater Model Development Project.

To date, the Peer Review Panel has completed the first three tasks (Tasks A, B, and C) and Task D.1 of the final Task D. The SJRWMD posted all interim deliverables to the NFSEG website at https://northfloridawater.com/groundwaterflowmodel.html. Below is a summary of each task.

Task A consisted of reviewing applicable documents and background materials prepared for Version 1.0 of the NFSEG model and proposed improvements for Version 1.1 (Task A.1), attending a kick-off meeting at the SJRWMD in Palatka on March 29, 2017 (Task A.2), preparing draft initial recommendations that were presented at a teleconference to SJRWMD, SRWMD, and stakeholders on April 13, 2017 (Task A.3), and preparing and submitting a technical memorandum on May 1, 2017 (Task A.4), which contained the panel’s final initial recommendations for changes and modifications to the MODFLOW and HSPF models. The panel’s recommendations were grouped into recommendations for changes to Version 1.1 of the NFSEG model that could be completed by July 1, 2017 (Phase 1) and changes that could be considered later for Phase 2 or for future updates.

Task B consisted of reviewing the Phase 1 results for Version 1.1 of the NFSEG model. This included reviewing preliminary model calibration results presented by SJRWMD and SRWMD at a teleconference on May 5, 2017 and making suggestions to facilitate the model improvements proposed by SJRWMD and SRWMD (Task B.1) and reviewing Phase 1 model files, draft figures and tables, and calibration statistics and attending a technical review meeting in Palatka on June 21, 2017 (Task B.2). A technical memorandum was prepared to present a summary of key findings as well as specific suggestions from each Peer Review Panel member for completing outstanding tasks during the remainder of the NFSEG Version 1.1 development period (Phase 2 period) so that the NFSEG Version 1.1 model could be finalized. The specific suggestions include consideration of a no-pumping/pre-development scenario, an uncertainty analysis, and a verification run for the model, and editorial suggestions for figures and tables. Preliminary answers to Task D.2 questions regarding questions #2A-F Model Implementation and #3A-D, G, and H Model Calibration and Application also were included in the Task B.2 technical memorandum.

Task C consisted of reviewing the Phase 2 results for Version 1.1 of the NFSEG model. Task C.1 consisted of reviewing the interim Phase 2 model calibration results presented by SJRWMD and SRWMD at a Preliminary Phase 2 Results Meeting on July 26, 2017 and making suggestions to facilitate model improvements. Task C.2 consisted of reviewing an update of the Phase 2 calibration results presented by SJRWMD and SRWMD at a Phase 2 Review Meeting on December 7, 2017.

Task D consisted of reviewing the final NFSEG v1.1 model and documentation. Task D.1 consisted of reviewing the draft NFSEG v1.1 and supporting HSPF documentation in preparation for a Task D.1 Draft NFSEGv1.1 meeting on April 18, 2018. Task D.2 consisted of reviewing an updated (5/7/2018) draft NFSEG v1.1 and supporting HSPF documentation, preparing peer review documentation, and preparing the Draft Peer Review Report dated August 10, 2018. Task D.3 consisted of preparing the Final Peer Review Report dated August 22, 2018 contained in this document, which is based on the Draft Peer Review Report, new information received during meetings, and other information received from SJRWMD and/or SRWMD.

2. Response to Critical Questions

The response of the Peer Review Panel to the critical questions listed in Appendix A of the Charter of the SJRWMD – SRWMD Cooperative Groundwater Model Development Project is as follows:

2.1 MODFLOW

1. **Model Documentation Provided by WMD's:**

   A. **Does the documentation provide a clear and appropriate description of the NFSEG groundwater flow model and supporting HSPF surface-water models?**
   Yes, the MODFLOW documentation is quite extensive and well written. There are some additional items discussed in the chapter-by-chapter review that need clarification and a number of editorial comments and suggestions that are included in this report. However, this should not take away from the overall assessment of the Peer Review Panel that this documentation provides a clear and appropriate description of the NFSEG groundwater flow model and supporting HSPF surface-water models**.**

   B. **Are the purposes and scope of the documentation clearly stated and sufficient to document the models? Is the content of the documentation consistent with the stated purpose and scope of the document?**
   Yes, the purposes and scope of the NFSEG MODFLOW model are clearly stated in the report, and the documentation provides analyses and results consistent with the stated purpose. However, the Purpose and Scope section contains additional information that is background material. It would be clearer if the background information were placed into a Background section immediately prior to the Purpose and Scope section.

   C. **Is the documentation readable? Are the figures clear? Does the format of the documentation need to be modified or expanded?**
   Yes, overall, the MODFLOW documentation is quite readable, the figures reflect what they are intended to show, and the format is consistent with that of a well written modeling report. However, there are some specific issues that need to be addressed. Some figures need additional technical information added (the specific items are listed in the chapter-by-chapter review), and some geographic features and places that are named in the text do not appear on any figure (a list of these is included in the editorial corrections and suggestions section of this report).

   D. **After reading the documentation, are the purposes, scope, strengths/weaknesses, intended use, and limitations of the NFSEG model understandable?**
   Yes, the purposes and scope of the MODFLOW model are well described and understandable in the report. The strengths and weaknesses of the model and its intended use and limitations are discussed in Section 8 Model Limitations. The report is very well

done, and it should be easily understood by other groundwater modelers and stakeholders.

2. **Model Implementation:**

A. Is the conceptual model appropriate for the intended use of the model? For example, are critical physical and hydrologic processes represented appropriately?
Yes, the conceptual model for MODFLOW, which is described by Durden et al. (2013) in a separate report, does a good job of describing the physical structure of the aquifers. However, neither that report, nor this report, sufficiently discusses the springs and the baseflows in the rivers. The baseflow discussion is too short, and for the intended use of this model, a thorough documentation and understanding of the baseflows is very important. In addition, ASTM (2018) (also see response to Question 3.J.5) recommends that the error range associated with each calibration target be identified, in addition to the value to be used for calibration. This was not done for the baseflows. Overall, the conceptual model is appropriate and consistent with other models of the Floridan Aquifer System such as the USGS East Central Florida groundwater model (Sepúlveda et al. 2012) and SWFWMD'S District Wide Regulation Model (DWRM) version 3 and Northern District Model (NDM) version 5. The most significant physical process not simulated is flow through conduit systems, but this is addressed in Section 8 Model Limitations.

B. Is the [MODFLOW] model code appropriate, given the intended use of the model?
Yes, MODFLOW-NWT is a recent version of MODFLOW from the USGS and is appropriate given the stated objectives of the model.

C. Was the numerical [MODFLOW] model constructed in a manner that is consistent with the underlying conceptual model, using appropriate data and methods of analysis?
Yes, layered model construction where layers represent aquifer and aquitard units and with a rectangular uniform grid is a standard method of simulating the UFA. In addition, the underlying data analysis methods are appropriate. Calibration with PEST and pilot points is quite sophisticated and appropriate given the complex nature of the aquifer system. However, there are some major items that need further attention. The description of baseflows requires further discussion in this report. Also, it appears that spring flows were given much larger weights than river baseflows in the calibration, causing PEST to produce closer matches to the springs and poorer matches to the rivers; this point needs to be discussed in this report as well. Not having recharge or evapotranspiration as PEST parameters requires further discussion in this report and further consideration as possible calibration parameters in any future revision of version 1.1 of the NFSEG model. Further discussion in this report should include an estimate of the accuracy of the recharge and evapotranspiration values calculated in HSPF, an explanation of why springs were simulated in layer 3 and rivers were simulated in layer 1, and whether manually adjusting recharge and evapotranspiration would result in better

matches for the river baseflows. Other lessor items are listed in the chapter-by-chapter discussion in the detailed comments section of this report.

D. Was the hydrologic model code selected appropriate for its intended use?
Yes. [see answer to question #2B above.]

E. Was the use of HSPF as a method to develop recharge and maximum saturated ET that is assigned to the MODFLOW groundwater flow model a valid and defensible method?
Yes. [See answer to question 2.2 HSPF 2. E Model Implementation below.]

F. Questions specific to HSPF Models:
These questions are addressed in Section 2.2 HSPF.

3. <u>Model Calibration and Application:</u>

A. Is the parameterization scheme used in the PEST calibration appropriate?
Yes, the parameterization scheme used in MODFLOW is quite complex but well thought out and appropriate. One criticism, as stated in previous comments and presentations, is not making the evapotranspiration rate an adjustable parameter. Up to 10,000 cells in layer 1 of the model have heads significantly above the top of the layer (e.g., land surface). The extent of layer 1 "flooding" is up to 359 feet above the top of the layer. Some of these areas were addressed by adding drains to simulate additional surface drainage. However, the number of flooded cells and the maximum extent of flooding increased from run 004b to 007h (current model). The suggestion is repeated that consideration should be given to allowing evapotranspiration to be adjusted during the PEST run, at least in these areas of extreme flooding.

Also, not making recharge a PEST parameter needs an explanation (or inclusion as a parameter). By not making recharge a PEST parameter, the only way PEST can significantly modify groundwater flow is by varying conductivities. This may be a contributing factor to the poor match between river baseflows and simulated baseflows and a contributing factor to the poor match between measured and simulated conductivities.

The justifications for treating evapotranspiration and recharge as constants in the PEST calibration in NFSEG Version 1.1 need to be discussed further in this report. Allowing evapotranspiration and recharge to be adjusted during PEST runs should be evaluated further in any future revision of Version 1.1 of the NFSEG model.

B. Were the types of observations and their implementation in the PEST calibration appropriate, given the intended use of the model?
Generally yes. The PEST calibration in MODFLOW uses data in many different ways. For example, both head and change in head (vertically and horizontally) make maximum

use of existing data in the calibration. A better description of the reasoning for the weights assigned to each observation group should be provided, however. The report does a good job of documenting the weights that were used but does not really get into the logic behind the choice of weights. Also, the river baseflow determination needs more discussion and documentation (as mentioned earlier).

C. **Have the differences between observations and their simulated equivalents (model residuals) been described sufficiently? For example, have an appropriate set of summary statistics, plots, and maps been presented that allow for evaluation of model limitations, (such as model bias and uncertainty) in a manner that meets or exceeds existing professional practices?**
Mostly yes, but providing some additional information related to simulated differences (residuals) in the MODFLOW calibration is recommended. First, since the report goes into considerable detail on parameter and observation groups, it would be consistent to add a table of the contributions of each observation group to the objective function. The objective function is described in general in the report, but the actual results from the PEST run are not documented. A table is provided of head statistics but not for spring flows and base flows. Spring data and baseflow pick-up estimates in Appendices E and F should also show the percent error in spring flow and base flow values to give the reader a better indication of the degree of fit with the flow observations. In addition, the match for important springs is provided in table form for the 2010 verification simulation (Table 5-2), but spring flow matches should also be tabulated and evaluated for the 2001 and 2009 calibration periods. Also, in the no-pumping simulation, estimates for historical heads and spring flows were used to evaluate the no-pumping simulation results, but estimates for baseflows were not made. A number of rivers in the model domain have gages that date back to the 1930s; if possible, these data should be used to estimate historical baseflows, which could also be used to evaluate the no-pumping simulation. In addition, however, return flow should not be included in the no-pumping simulation. Conceptually, return flows are how excess water is recharged to the groundwater system from pumping, and, thus, there should be no return flow in the absence of pumping.

D. **Have the values of calibrated parameters been described appropriately, using (for example) maps illustrating the range and spatial distribution of parameter values?**
In general, yes. The maps in the MODFLOW report are very useful and can be compared to the actual MODFLOW input files by a knowledgeable reader. The text description is a bit limited; however, most modeling reports tend to give cursory descriptions of complex parameter fields, preferring instead to rely on the figures.

E. **Does the final version of the model appear to be adequately calibrated given the available data for calibration and the state of knowledge (and lack thereof) of the hydrologic system prior to development of the model?**
Yes, the degree of calibration of the MODFLOW model is quite good. No model is perfect and there are always outliers and areas that cannot be explained, and this model has a several such areas. However, given the regional nature of the model and grid cell

sizes, the calibration compares favorably with results obtained for similar regional models (e.g., SWFWMD'S District Wide Regulation Model (DWRM) version 3 and Northern District Model (NDM) version 5, see Table 1). Also, the NFSEG model will certainly be adequately calibrated if the concerns of the reviewers that recommend additional discussion and explanation are addressed.

Table 1. Summary of Statistics for Heads for NFSEG and Three Comparable Regional Groundwater Flow Models in Florida

Model	Residual Mean (ft)	Absolute Residual Mean (ft)	RMS Error (ft)
NFSEG	0.17	4.45	8.02
DWRM3 (1995)	1.54	4.00	5.50
DWRM3 (2005)	-1.56	5.63	8.10
NDM (2010)	0.86	5.23	6.74

F. **Is the final version of the model appropriate for the intended planning and regulatory uses in the SRWMD and SJRWMD areas of the model domain? Is the NFSEGv1.1 groundwater flow model a sufficient tool for evaluating individual CUP's and compliance with individual spring MFL's?**
Yes, the NFSEG model is well suited for its intended planning and regulatory uses. The model will be a sufficient tool for evaluating individual CUPs and compliance with individual spring MFLs if the concerns of the reviewers recommending additional discussion and explanation are addressed. At some point, models need to be used to be effective. Experience with SWFWMD's District Wide Regulation Model revealed that it was only after version 1 was released that areas were found that needed to be improved. Weaknesses in any model will invariably reveal themselves through application to real world problems; as the NFSEG model evolves through time, it should get better and better.

G. **Has the complete model water balance, accounting for all water sources and sinks, been assessed and found reasonable?**
Yes, the water budgets described in Section 6 Water Budget Analysis for the model domain and for seven groundwater basins that make up the model domain have been assessed and appear to be reasonable.

H. **Have the uncertainty of key model parameters and predictions been assessed using methods that are appropriate and that meet or exceed typical practice for developing groundwater flow models? Has a detailed statistical assessment of uncertainty in modeled groundwater level and spring flow estimates been provided?**
Yes, the uncertainty analysis documented for the NFSEG model in Section 7 Sensitivity and Uncertainty Analysis is very detailed and comprehensive, including evaluating predictive uncertainties at selected locations for heads, baseflows, and spring flows (Table 4-1, Appendix L). The modeling team should be commended for undertaking

such a significant effort, particularly since very few models ever undergo such a detailed and complete uncertainty analysis.

I. Have the limitations of the final version of the NFSEG groundwater flow model been adequately described in the model documentation?

Yes, Section 8 Model Limitations concisely explains the limitations of the model with the following exception: in the calibration section (Section 4 Model Calibration, p. 55), it is stated that structural errors typically are the largest source of errors in a model. This should be repeated in Section 8.

J. Have the Measures of Success for NFSEG Charter Objectives 2, 5, and 6 been met?

Objective	**Measure of Success**
2. The model output helps to answer all regional-scale model questions in Appendix A of the NFSEG Charter.	A reasonable groundwater modeling technical expert would judge the model output useful in answering the questions in Appendix A.
5. The model is calibrated to industry standards.	The model calibration statistics meet industry standards in ASTM Standard Guide for Calibrating a Ground-Water Flow Model Application, Designation D 5981-96 (2008).
6. The model is accepted as a useful tool.	Success would be (1) a reasonable, independent groundwater modeling technical expert judging the model developed by this project to be acceptable by the standards of the profession for helping to answer the modeling questions that have been asked; and (2) a clear understanding by all involved parties of the uncertainties and limitations of the model for answering the modeling questions in Appendix A.

Objective 2. The model output helps to answer all regional-scale model questions in Appendix A of the NFSEG Charter. A reasonable groundwater modeling technical expert would judge the model output useful in answering the questions in Appendix A.

Yes, the model output will be useful in answering the regional-scale model questions in Appendix A if the suggestions made by the peer reviewers recommending additional discussion and explanation are addressed.

Objective 5. The model is calibrated to industry standards. The model calibration statistics meet industry standards in ASTM Standard Guide for Calibrating a Ground-Water Flow Model Application, Designation D 5981-96 (2008).

Mostly the calibration statistics meet industry (ASTM) standards. The exception is that the range of errors in the river and spring baseflows is not listed. ASTM (2018, updated from 2008) states: for a medium- to high-fidelity model application, calibration targets should be established by first identifying all relevant available data regarding groundwater heads (including measured water levels, bottom elevations of dry wells, and top of casing elevations of flowing wells) and flow rates (including records of pumping well or wellfield discharges, estimates of baseflow to gaining streams or rivers or recharge from losing streams, discharges from flowing wells, spring flow measurements, and/or contaminant plume velocities). For each such datum, error bars associated with the measurement or estimate should be included. In the MODFLOW simulation, calibration targets for heads were established prior to the calibration process but not for spring flows and baseflows, which should be established. Based on ASTM (2018), one criterion for accepting a calibration is that the residual for heads should be a small fraction of the difference between the highest and lowest heads across the model area. This criterion should be checked in addition to the calibration results for heads and residuals described in Section 4 Model Calibration of the draft model report. In addition, targets for spring flows and baseflows should be established based on the accuracy of the observed (or estimated) values for these parameters. ASTM (2018) recognizes that errors in the estimates for groundwater flow rates will usually be larger than errors in the estimates of heads and, in particular, that baseflow estimates are generally accurate only to within an order of magnitude. In such cases, the upper and lower bounds on the acceptable modeled value of baseflow can be equal to the upper and lower bounds on the estimate. This limit should be recognized when establishing calibration targets and evaluating the calibration for baseflows in the NFSEG groundwater model.

Objective 6. The model is accepted as a useful tool. Success would be (1) a reasonable, independent groundwater modeling technical expert judging the model developed by this project to be acceptable by the standards of the profession for helping to answer the modeling questions that have been asked; and (2) a clear understanding by all involved parties of the uncertainties and limitations of the model for answering the modeling questions in Appendix A.
Yes, the model should be able to help answer the modeling questions that are asked in Appendix A if the suggestions made by the peer reviewers recommending additional discussion and explanation are addressed. Also, model limitations are discussed in Section 8 Model Limitations, making it possible for involved parties to understand clearly the uncertainties and limitations of the model.

2.2 HSPF

1. <u>**Model Documentation Provided by WMD's:**</u>

   A. **Does the documentation provide a clear and appropriate description of the NFSEG groundwater flow model and supporting HSPF surface-water models?**
   Yes, the HSPF model documentation provides a clear and appropriate description of the model's approach, conceptual model, and development, i.e., the input data development and construction of the model inputs. It also includes sufficient documentation of the model calibration results. The areas that are lacking are: 1) the presentation of the calibrated model parameters and 2) calibrated model water balances. HSPF models

should include documentation of the key hydrologic parameters that are used to calibrate a model. The NFSEG HSPF model is quite complex due to the large geographic area and the large number of unique HSPF models that are included. In order for a reviewer to determine whether the various parameters are within reasonable/valid ranges, the documentation should include an appendix that summarizes the parameter values with tables and maps.

B. Are the purposes and scope of the documentation clearly stated and sufficient to document the models? Is the content of the documentation consistent with the stated purpose and scope of the document?

Yes, the purposes and scope of the HSPF documentation are to present the model conceptualization, input datasets, implementation/construction of the model, calibration approach, and calibration results for the HSPF models. The documentation is consistent with these purposes and scope with a couple of caveats that are described in the responses to some of the other questions. The conceptualization, input data, implementation, and calibration approach sections are mostly complete and clear. The calibration approach should include discussion of the effects of calibration of flows affected by tides and significant man-made influences on the predicted recharge. The calibration results (shown in the 55 watershed-specific appendix sections) should include a brief discussion of man-made influences and other causes of poor calibration for poorly calibrated gauges. The calibration section should also include documentation of the key hydrologic parameter values obtained or reproduced from a nearby watershed during calibration (see question #1D), and also the simulated water balance summaries described under question #3G.

C. Is the documentation readable? Are the figures clear? Does the format of the documentation need to be modified or expanded?

Yes, the HSPF documentation is readable, and the figures are clear. The format of the documentation is good, with the exceptions noted above in **Question 1A.**

D. After reading the documentation, are the purposes, scope, strengths/weaknesses, intended use, and limitations of the NFSEG model understandable?

Yes, the purpose, scope, strengths/weaknesses, intended use, and limitations of the NFSEG HSPF model are generally understandable. The calibration of watersheds with tidal and man-made influences on measured flows should be discussed in the calibration approach, and the possible effects on computed recharge should be evaluated. Since PEST is used for the automated calibration, the effects of specific objective function components on calibration should be discussed in this section.

2. <u>Model Implementation:</u>

A. Is the conceptual model appropriate for the intended use of the model? For example, are critical physical and hydrologic processes represented appropriately?

Yes, the conceptual model is appropriate for the intended use of the model. The HSPF model provides recharge and maximum saturated ET for use in the MODFLOW model that includes the critical physical and hydrologic processes that are required for estimating those quantities. HSPF models the hydrologic cycle for the unsaturated zone plus the groundwater that contributes to surface water bodies. The key processes resulting from rainfall, i.e., evapotranspiration, surface runoff, infiltration, recharge to the saturated zone, and streamflow are well represented, and the model can be segmented adequately to provide recharge results at the resolution of the groundwater model. There are limitations to use of HSPF or any similar model in flat areas such as Florida, where groundwater levels are often above the land surface. Most aspects of the HSPF approach used for the NFSEG model are appropriate, and in several cases innovative. These aspects include:

- the use of NLDAS rainfall data;
- use of NLDAS PET, adjusted by USGS PET-derived monthly factors to represent the appropriate PET quantity;
- use of Special Actions to model closed basins by collecting the watershed runoff and directing it to a virtual sink with an appropriate flow rate;
- use of a subsurface reach to collect recharge in springsheds and calibrate the recharge to the observed spring flows;
- use of PEST in an automated environment to consistently calibrate more than 50 watersheds to USGS stream flow using a comprehensive set of metrics in the PEST objective function; and
- inclusion of calibration of total actual ET (by land cover) to literature-derived expected annual amounts to constrain this key component in the water balance to a consistent and appropriate value.

B. **Is the [MODFLOW] model code appropriate, given the intended use of the model?**
This is addressed in Section 2.1 MODFLOW.

C. **Was the numerical [MODFLOW] model constructed in a manner that is consistent with the underlying conceptual model, using appropriate data and methods of analysis?**
This is addressed in Section 2.1 MODFLOW.

D. **Was the hydrologic model code selected appropriate for its intended use?**
Yes. [see answer to Question #2E below.]

E. **Was the use of HSPF as a method to develop recharge and maximum saturated ET that is assigned to the MODFLOW groundwater flow model a valid and defensible method?**
Yes, the use of HSPF to develop recharge (and maximum saturated ET) that is input to groundwater flow models is definitely a common and defensible method. The most prominent example is the Integrated Hydrologic Model (IHM), which was developed initially for Tampa Bay Water and the Southwest Florida Water Management District (Geurink and Basso 2013). It should be noted that the IHM contains a dynamic coupling of HSPF with the MODFLOW groundwater model, which is more complex than the one-way coupling used for NFSEG.

F. Questions specific to HSPF Models:

a. The version of HSPF utilized for the hydrologic models is a non-standard version of HSPF that is not publicly available. Is the version of HSPF utilized appropriate and defensible?

Yes, the version of HSPF used is defensible and appropriate, based primarily on personal communications regarding this issue with the SJRWMD staff in past years. However, this could be backed up more clearly in the documentation, including a description of the feature(s) that are non-standard, and citation of a document that confirms the District's prior validation of the non-standard version. The primary feature that is not in the publicly-available version is an optional method for computing surface runoff from a standard pervious land area (PERLND). This feature is utilized to improve the simulation of surface runoff from the land areas categorized as wetlands and water in the NFSEG model.

b. Was the best available information utilized to develop the HSPF hydrologic models?

Yes, in summary, the data and other information used to develop the hydrologic models is the best available, given the limitations imposed by the scale of the model (i.e., the large area and number of sub-watersheds modeled). Input rainfall and potential evapotranspiration data, which are the primary driving forces of the model, utilize the best available, consistent data source (NLDAS) that covers the entire model area. The analysis and decisions related to the various rainfall data sources (gauges, radar, NLDAS) is impressive, and the source that exhibited the best combination of accuracy, consistency and geographical resolution was selected. The NLDAS PET data were adjusted (tensioned) using a consistent PET dataset for Florida developed by USGS to represent the correct quantity required by HSPF, i.e., lake evaporation.

The primary data used for calibration, i.e., the measured streamflows, are the best/only available data for this purpose. As documented, the quality of these data is quite variable, and many of the streamflow gauges are rated as poor quality. This is one of the primary sources of error in the NFSEG HSPF model.

Another data source consists of literature-derived estimates of the expected total ET from the various pervious land cover categories. These data are used to calibrate the actual/simulated ET so that this major component of the water balance is reasonable and consistent across the model domain. It appears that the literature data are reasonable, and most of the data are obtained from Florida.

The model area was segmented (delineated) into watersheds based on the USGS HUC8 watersheds, with sub-watersheds based on elevation (DEM) data. The hydrology of separate land covers/uses is appropriately represented by segmenting the sub-watersheds into individual hydrology computational units (pervious and impervious land segments) using primarily NLCD 2001 coverages.

The effects of irrigation are represented using standard methodologies. The agricultural quantities are estimated using an appropriate model of specific plant needs, soil moisture and available rainfall, and the amounts are consistent with known records of local pumping and withdrawal data. The urban irrigation is based on utility records of usage, where available, and golf course irrigation is based on permitted/measured data, where available. The implementation of the irrigation water input to the soil, and removals from surface water bodies appear to be correct.

c. Unique aspects of these systems were represented with Special Actions or with other features of HSPF and are these conceptually sound and implemented appropriately:

1. **RCHRES representation of Inactive Groundwater Storage to represent spring discharges?**
2. **Closed basins?**
3. **Drainage wells and swallets?**
4. **Implementation of water use:**
    a. **Agricultural irrigation?**
    b. **Urban:**
        i. **Septic?**
        ii. **Irrigation?**
    c. **Golf courses?**
    d. **Reuse spray fields?**

Yes, items listed above in 1-4d above are conceptually sound and implemented appropriately. The HSPF model includes an innovative spring representation that uses a RCHRES to collect the inactive groundwater inflow (IGWI) within a designated springshed, and then routes this "reach" to the surface reach where the actual spring is located. This simulated spring outflow was calibrated to measured spring flows, which is very innovative. However, this aspect of the spring feature does not seem to be included in the documentation.

The closed basin flow into sinks/drainage wells is represented using Special Actions. Review of a HSPF model input file with a closed basin indicates that it is implemented correctly; however, the values of the reach-specific parameters used to represent the invert, the maximum flow and depth above invert where maximum flow begins could not be verified as part of this peer review. The use of this feature is impressive, since it allows more reasonable parameter values to be used from an adjacent calibrated watershed. This avoids the extreme parameter values that would sometimes result from forcing the runoff flows to near zero.

The description of the water use components implemented in the model appears to be complete. These include computation of irrigation for agricultural and urban areas based on plant needs and rainfall/ET (agriculture), urban water use (urban), and typical usage (golf courses). The incorporation of the water time-series in the model as additions to the various soil compartments in HSPF is correct based on review of the document.

3. **Model Calibration and Application:**

   A. **Is the parameterization scheme used in the PEST calibration appropriate?**
   Generally yes, based on experience with PEST that is limited to an understanding of its general usage and theory with HSPF calibration and not a full calibration. The main complaint that some have with using automated calibration with HSPF is the tendency to arrive at a calibration endpoint with parameter values that vary from the ranges that are valid with respect to the specific hydrologic process algorithms in HSPF. If the automated process can be constrained to vary parameters within those limits, and if the calibration can be monitored and adjusted by modelers who are familiar with these limitations, then that issue can be mitigated. Based on the documentation, the HSPF parameters optimized in the PEST calibration are the appropriate set. Furthermore, the establishment of relative values of four key parameters (LZSN, UZSN, INFILT, LZETP) to land use/cover categories is appropriate, and the assignment of AGWRC and DEEPFR on a watershed basis rather than by land cover is definitely standard usage for many modelers.

   B. **Were the types of observations and their implementation in the PEST calibration appropriate, given the intended use of the model?**
   Yes, in general, the observations used in the PEST calibration are appropriate, even with using such a large number of observations, including a series of baseflow-related measures, minimum flows, and flow reversal measures. The calibration of total actual ET measures by land cover is critical to constraining the ET to reasonable and consistent values, and thereby achieving consistent recharge results over the domain. Because PEST is not yet in common usage by HSPF modelers, it is recommended that the objective function components be more completely described, especially the effects of adjusting the relative weights.

   C. **Have the differences between observations and their simulated equivalents (model residuals) been described sufficiently? For example, have an appropriate set of summary statistics, plots, and maps been presented that allow for evaluation of model limitations, (such as model bias and uncertainty) in a manner that meets or exceeds existing professional practices?**
   Yes. The appendix (appendices) described at the end of the Calibration section for HSPF contain separate documents for each of the 51 HUC8 watersheds that were calibrated. Each document contains detailed statistics and three graphs for each calibrated gauge location. They also provide maps showing the gauge locations, land cover, and subwatershed delineation. There is also a graphic that depicts the flow gauges in the HUC 8 watershed, including 1) location/USGS ID number, 2) period of record, 3) whether the gauge was used for calibration, and 4) mean flow in cfs. These appendices provide sufficient information to evaluate calibration errors. The main recommendation is to include a very brief discussion of the modelers' conclusions and evaluation of the reasons for poor agreement in the calibration results at gauges that are poorly calibrated. These

reasons can be a combination of poor observed data, tidal effects, man-made influences in the watershed, unmodeled groundwater gains/losses, and uncertainty in a key input.

D. Have the values of calibrated parameters been described appropriately, using (for example) maps illustrating the range and spatial distribution of parameter values?
No, the HSPF documentation does not include appropriate description of the primary hydrologic parameter values obtained during calibration. (Refer to Question #1A.) The minimum set of calibrated parameters that should be documented in an appendix (tables and maps) are listed below.

AGWRC - Base groundwater recession
BASETP - Fraction of remaining ET from baseflow
CEPSC - Interception storage capacity
DEEPFR - Fraction of groundwater inflow to deep recharge
INFILT - Index to infiltration capacity
INTFW - Interflow inflow parameter (omit due to low value)
IRC - Interflow recession parameter (omit due to low value of INTFW)
KVARY - Variable groundwater recession
LZETP - Lower zone ET parameter
LZSN - Lower zone nominal soil moisture storage
UZSN - Upper zone nominal soil moisture

The main purposes of this recommendation are to: 1) ensure that the parameters have reasonable values, i.e., they are within valid ranges for the respective process formulations and for the specific land cover and climate; and 2) ensure that the variation over the model domain and within specific watersheds is reasonable. The standard requirement for any HSPF model documentation includes summaries of the key calibrated (and assumed) hydrologic parameters listed above.

E. Does the final version of the model appear to be adequately calibrated given the available data for calibration and the state of knowledge (and lack thereof) of the hydrologic system prior to development of the model?
Yes, based on reviewing the Calibration Results section and Table 17 in the main HSPF document and the more detailed calibration statistics/graphics for all gauges in the Appendix. The automated calibration appears to use appropriate criteria, based on the Parameter Estimation with PEST section of the document and the use of percent bias and Nash-Sutcliffe coefficients as key criteria for determination of calibration performance.

Overall, the calibration of large areas of the model domain is very good for an automated procedure, considering that the measured flows at many gauges are: 1) affected by man-made influences that are not included in the models, 2) subject to tidal effects, and 3) poor quality due to difficulty with measuring flows in flat terrain and areas where groundwater effects are large. It is noted that at several gauges, there are large, virtually constant differences between the simulated and observed flows that are caused by either an error in the model or a significant man-made influence. These should have been investigated and either documented, if it is man-made; or corrected, if a model error was the cause. Examples are gauges 02197500 and 02198500, both in the Savannah River. It is assumed in these cases that the calibration criteria used by PEST were affected by objective function components other than the total flow, (e.g., total actual ET). Otherwise, the automated calibration might have improved the total flow agreement. The discussion and graphics (Figures 27 and 28) correlating poor calibration performance with poor flow measurement are useful. In reviewing many individual gauge results, it was observed that this correlation is quite apparent.

The main questions or concerns with the calibration are related to the effects on recharge of calibration to observed flows that are affected by tidal and (especially) man-made influences. The discussion should include an analysis of this impact. Possibly, the effect is small for the same reason that the calibration did not adjust the simulated flow to match observed in the examples of large, constant differences in the two Savannah River gauges noted above. It is assumed that other criteria in the objective function prevented the large changes that would be needed to bring the flows into better agreement.

F. Is the final version of the model appropriate for the intended planning and regulatory uses in the SRWMD and SJRWMD areas of the model domain? Is the NFSEGv1.1 groundwater flow model a sufficient tool for evaluating individual CUP's and compliance with individual spring MFL's?

Yes, While the HSPF model results are not intended to be used directly for regulatory decisions, the use of the NFSEG HSPF results as recharge input to the groundwater model is appropriate based on the model conceptualization, implementation, and calibration.

G. Has the complete model water balance, accounting for all water sources and sinks, been assessed and found reasonable?

Not completely. This question seems to be addressed primarily to the MODFLOW model. However, it is also applicable to the HSPF model. The Districts should generate and document (in an appendix) summaries of the average annual HSPF water balance results for the individual land areas (PERLND and IMPLND). This water balance provides a summary of the: 1) inputs (rainfall, irrigation), 2) evapotranspiration losses, 3) runoff losses to streams (by soil layer), and 4) groundwater recharge. Weighted average summaries can be generated for each land cover in a watershed in addition to averages over all land covers. The primary purpose for this output is to determine the reasonableness of the amounts. It allows the modeler to identify errors in the input data such as rainfall, PET, and irrigation; and unreasonable water balance quantities caused by the automated calibration. In addition, the calibration of total actual ET to expected annual amounts can be verified.

Based on a review of preliminary water balance data that the District recently produced for individual years (2001, 2009, and 2010), it is recommended that the water balance should be computed for the full period of calibration instead of individual years, and it should be included in an appendix so that model reviewers can compare the results with input data (rainfall, irrigation, etc.) and the calibrated of total ET, in addition to verifying that the other components are reasonable.

H. Have the uncertainty of key model parameters and predictions been assessed using methods that are appropriate and that meet or exceed typical practice for developing groundwater flow models? Has a detailed statistical assessment of uncertainty in modeled groundwater level and spring flow estimates been provided?
This is addressed in Section 2.1 MODFLOW.

I. Have the limitations of the final version of the NFSEG groundwater flow model been adequately described in the model documentation?
This is addressed in Section 2.1 MODFLOW.

J. Have the Measures of Success for NFSEG Charter Objectives 2, 5, and 6 been met?
This is addressed in Section 2.1 MODFLOW.

3. CONCLUSIONS AND RECOMMENDATIONS

As part of its responsibilities to conduct a thorough review of the groundwater and surface-water models and model documentation reports, the Peer Review Panel has assessed the topics listed below. The conclusions and recommendations of the Peer Review Panel are as follows:

3.1 MODFLOW

1. Model Objectives, Conceptualization, and Design

The model objectives, as stated in the Purpose and Scope (p. 1), i.e., the "…primary purpose of the NFSEG model is to enable improved evaluations of inter-District (e.g., SJRWMD/SRWMD) and inter-state (e.g., Florida/Georgia) water-level changes in the surficial and Floridan aquifer systems from groundwater use over the model domain.", are addressed in the report. The previously prepared conceptualization report (Durden et al. 2013) details the plan for construction of the NFSEG groundwater model, including model extent, configuration, and lateral and internal boundary conditions; an analysis and interpretation of data needed for determination of the model calibration years; a plan for determination of groundwater recharge and maximum saturated evapotranspiration rates; and proposed NFSEG model calibration objectives.

The design structure of the MODFLOW model is considered excellent for the following reasons:

- Model boundaries extend to natural flow boundaries;
- Model grid cell size was as small as possible;
- Boundary conditions are no-flow where possible;
- Surficial aquifer was modeled as an active layer;
- ICU was modeled as an active layer;
- Geology was well researched;
- Aquifers were well researched and described;
- Representation of the aquifer layering in the model was generally good (having the UFA cross layers 1 and 2 did add some complexity); and
- Recharge was heavily researched using HSPF.

2. Assumptions and Limitations of Input Data

The hydrology of the area (Section 2. Hydrology of the Area) is described in terms of the surface-water and groundwater systems including rivers, lakes, swamps and wetlands, the Atlantic Ocean and Gulf of Mexico, the Surficial and Floridan Aquifer Systems, and the Intermediate Confining Unit (ICU). Sources of data for the groundwater systems include previously published reports, e.g., Miller (1986) and Bush and Johnston (1988), published aquifer test results, measured groundwater levels from water management district files and reports, and reported spring flows. Baseflows, which are an important calibration metric along with groundwater levels and spring flows, were estimated by averaging the results of four different hydrograph separation techniques and a fifth approach that utilizes flow duration curves. In addition, an alternative approach was utilized for determining baseflow pickups between adjacent gages that bound river reaches with contributing basin areas in which the ICU is absent completely, and which lack a well-developed, channelized, surface drainage network. In this approach, baseflow pickup was determined by taking the difference in the total observed flows of the upstream and downstream bounding gages, based on the assumption that overland runoff is negligible in areas in which the Upper Floridan aquifer is unconfined. Cumulative baseflow estimates and baseflow pickup estimates were mostly derived using the same averaging technique used to estimate baseflows. Concentrated groundwater inflows, e.g., rapid infiltration basins (RIBs) and drainage wells, and injection wells, are briefly described, along with groundwater withdrawals for public/commercial/industrial/institutional supply; agricultural irrigation supply; recreational irrigational supply (e.g., golf courses); and domestic self-supply. Input for recharge and maximum saturated ET are obtained from the results of the HSPF model simulations.

The description of the surface-water system is acknowledged to be brief (p. 5), and expanding the discussion of baseflows should be considered. The relative accuracy of the available data for groundwater heads and groundwater flows needs to be acknowledged, i.e., groundwater heads would be expected to be accurate to within a few tenths of a foot, but errors in estimates of groundwater flows (spring discharges and baseflows) would likely be much larger, e.g., the baseflow estimates may be accurate only to within an order of magnitude (ASTM 2018). Also, the discussion of groundwater inflows and withdrawals (pp. 22-23 and Figures 2-44 – 2-47) and the representation of the inflows and outflows in the MODFLOW well package (p. 41 and Figures 3-41 – 3-44) needs additional explanation and detail that could be provided in an appendix. Such detail would include well locations, pumping rates, and water-use categories for 2001, 2009, and 2010.

The assumption that groundwater flow in the Floridan Aquifer System can be approximated as laminar flow and represented as a porous medium in MODFLOW is applicable at the scale of the NFSEG grid spacing (2,500 feet x 2,500 feet discretization). Within the model domain, localized areas of non-linear laminar or turbulent spring flow may occur, and the accuracy of model results would be affected in such sub-regional areas. However, based on a comparison of the application of the MODFLOW Conduit Flow Package and a standard MODFLOW application at Wakulla Springs by Kuniansky (2016), the assumption that the standard MODFLOW porous medium approach is applicable throughout the NFSEG model domain is reasonable and defensible.

3. Model Calibration and Sensitivity

The NFSEG model was calibrated for 2001 (a relatively dry year) and 2009 (a relatively wet year) using hydrologic data (observations) including water levels, spring discharges, and estimated base flows. The year 2010 was selected for the NFSEG model verification. An initial steady-state manual calibration was performed, and the results from the initial calibration were used to guide the PEST- (Doherty 2010) facilitated process to achieve the calibration criteria of minimizing differences between observed and simulated water levels, head gradients, baseflows, and stream flow gains and losses. Numerous realizations (model runs) were performed in PEST by automatically varying various model parameters. Pilot points were used to estimate the spatial distribution of particular hydraulic properties within a model layer. Hydraulic parameters (model input), listed below, were varied spatially to achieve acceptable water levels and flows at hydrologic features of interest:

1. K_h and K_v multipliers of each layer;
2. Anisotropic ratio of each layer;
3. GHB conductance for springs;
4. River-bed conductance multipliers for stream baseflows;
5. Drain conductance multipliers for ephemeral streams;
6. River-bed conductance multipliers for lakes; and
7. Lake-zone multipliers for ICU (layer 2) K_v beneath lakes.

Recharge and ET, obtained from the HSPF results, were treated as constant inputs and not adjusted during the PEST runs. The calibration results were evaluated for heads, flows, and input parameters. The results between observed and simulated values for heads and flows (spring discharges and baseflows) were evaluated using standard statistical comparisons that include mean error (ME), mean of absolute error (MAE), standard deviation of error (SD), and correlation coefficient (R^2). The calibration results for heads are compared to calibration targets for which 80% of the groundwater heads residuals should be within ±5 feet and for which 50% of the groundwater head residuals should be within ±2.5 feet. In the simulation results for groundwater heads, 72% of the residuals are within ±5 feet and 42% of the residuals are within ±2.5 feet for 2001, and 74% of the residuals for heads are within ±5 feet and 48% of the residuals are within ±2.5 feet for 2009. Thus, the simulation results do not achieve the calibration targets, but it is concluded that "…the percentage of the groundwater level residuals within 2.5 feet and 5 feet, generally indicate a very good match between observed and corresponding simulated values." (pp. 58-58). Calibration targets were not established for spring discharges and baseflows. As described in Section 7 (Sensitivity and Uncertainty Analysis), the sensitivity of NFSEG model outputs to individual parameters was evaluated thoroughly to understand the importance of the various model input parameters to the behavior of simulated flows and levels using two methods, i.e., by calculating "traditional" parameter sensitivities as well as calculating composite-scaled sensitivities.

The justifications for treating evapotranspiration and recharge as constants in the PEST calibration in NFSEG Version 1.1 need to be discussed further in this report. Allowing evapotranspiration and recharge to be adjusted during PEST runs should be evaluated further in any future revision of Version 1.1 of the NFSEG model. Also, it is recommended that the calibration targets for groundwater heads be re-examined to determine if a broader range of statistical analyses such as criteria for mean error (ME), mean absolute error (MAE), and root mean square error (RMSE) (e.g., Anderson and Woessner 1992) would provide a better set of metrics to judge the results for 2001, 2009, and 2010. Similarly, calibration targets should be established for spring discharges and baseflows, keeping in mind that the observed (or estimated) values may not be nearly as accurate as measured groundwater heads. Also, the residual statistics in Sections 4 and 5 (Model Calibration and Model Simulations) and results of other statistical analyses should be compared to residual statistics that have been obtained for other comparable regional groundwater flow models, e.g., SWFWMD'S District Wide Regulation Model (DWRM) version 3 and Northern District Model (NDM) version 5 and steady-state results in the USGS East-Central Florida transient model (Sepúlveda et al. 2012).

4. Model Documentation (explanation of model, data sources, and assumptions)

In general, supporting documentation for the NFSEG model is adequate to assess the model results. However, additional statistical metrics and tests of random and normal distribution of residuals on heads, spring flows, and base flow residuals are needed to strengthen technical assessment of the calibration. Also, the "brief description of the surface-water system" (p. 5) needs to be expanded to include more descriptive material and details about baseflows. A weakness of the report is the use of qualitative statements such as "*good match, good agreement, generally good match overall, very good agreement, generally poor to fair comparison, generally poor comparison, and aspirational values*" to assess the goodness of fit between simulated and observed groundwater heads, spring flows, and base flows. Such qualitative descriptors are not easily evaluated because one's person view of what represents "good" agreement between the model and observations can vary from another, and, thus, the use of these descriptors should be avoided.

5. Suitability of MODFLOW and Related HSPF Models for Intended Applications

MODFLOW-NWT is a recent version of MODFLOW from the USGS and is appropriate given the stated objectives of the model. In this application, the use of MODFLOW is suitable for water-resource assessment, determining minimum flows and levels, and evaluating water-use permit applications on a regional scale.

Note: items 6, 7, and 8 are combined into one response below.

6. Appropriateness, Defensibility, and Validity of Model/Relationships

7. Validity and Appropriateness of All Assumptions Used in Development of Model/Relationships

8. Deficiencies, Errors, or Sources of Uncertainty in Model/Relationship Development, Calibration, and Application

The model objectives, as stated in the purpose and scope, are addressed in this report, and the model conceptualization is described in the report by Durden et al. (2013), which presents the geology, physical structure of the aquifer, and other background information needed to begin assembling the model. In the report reviewed here, the range of errors in the determination of the baseflows is not reported as recommended by ASTM (2018), and additional documentation and discussion of spring flows and baseflows is needed. Calibration targets for spring flows and baseflows also need to be established, and consideration needs to be given to adjusting recharge and/or ET during the PEST calibration in any subsequent revision of version 1.1 of the NFSEG model. Additionally, there is some indication that head, spring flow, and base flow residuals are not randomly distributed in the model domain. A non-random, spatial distribution in residuals often indicates model bias and possible model error. To determine the validity of spatial randomness, the “run statistics” (Hill 1998) calculated by the MODFLOW Observation Process or similar code should be used as an independent measure of randomness.

These additional considerations potentially will have important impacts on the applicability of the NFSEG groundwater model. The chapter-by-chapter comments in the Detailed Comments section of this draft peer review report should also be considered. Most aspects of the NFSEG model are valid and appropriate; if the concerns of the reviewers recommending additional discussion and explanation are addressed, then the NFSEG model should be quite appropriate and defensible.

3.2 HSPF

1. Model Objectives, Conceptualization, and Design

The NFSEG HSPF model’s objectives are clearly described, and the model is very well conceptualized and designed. All major hydrologic inputs and processes are accounted for at a reasonable level, and the closed basin representation and spring simulation features are innovative. The model is appropriately delineated, i.e., segmented into watersheds and individual land cover-based hydrologic units. The calibration procedures and PEST objective function conceptualization appear to be valid. The correct output quantities are used to provide the recharge and maximum saturated ET to the MODFLOW model.

2. Assumptions and Limitations of Input Data

The input data for the NFSEG HSPF model include rainfall, PET, water use/irrigation, land elevation and slope, land use/cover, contributing watershed areas, waterbody characterization, and observed streamflow for calibration, i.e., comparison with simulated flow. The model does a good job of recognizing and mitigating the assumptions and limitations of the most important inputs, including rainfall and PET, observed flows, and watershed delineation-related data. The preparation and limitations of somewhat less critical data, such as water use and irrigation, are addressed in less detail.

3. Model Calibration and Sensitivity

Based on the statistical and graphical results presented, the model is generally well-calibrated to observed flow data at most stream gauges where the measured flow is reliable and watersheds that don’t have major man-made or tidal influences. Some stream gauges where the data are uncertain (i.e., poor quality as judged by USGS) have unsatisfactory calibration statistics. These are generally locations that are influenced by tidal flows, man-made structures and flow modifications, and unusually flat or areas of strong groundwater interaction with surface flows. These poorly-calibrated gauges should be discussed briefly in the calibration summaries for each HUC8 watershed.

Sensitivity is not addressed in the documentation for the NFSEG HSPF model. If reasonable care and appropriate assumptions are used in constructing the model and input data, then HSPF models are generally most sensitive to the major driving force inputs (rainfall and PET), and the major parameters for affecting ET and infiltration, such as the upper and lower soil storage parameters, infiltration rate, and ET from interception storage and the plant root zone. A possible future enhancement would include sensitivity analysis of these parameters in selected watersheds.

4. Model Documentation (explanation of model, data sources, and assumptions)

The HSPF documentation is very good at explaining the model, data sources, most assumptions, and the calibration results. In the calibration approach section, watersheds with tidal and man-made influences on measured flows should be discussed, and the possible effects on computed recharge should be evaluated. Since PEST is used for the automated calibration, the effects of specific objective function components on calibration should be discussed in the section on PEST. In the calibration section, the final parameter values of selected HSPF parameters should be compiled and summarized, and HSPF water balance summaries should be compiled and summarized to verify their reasonableness and verify that the total actual ET calibration to expected/literature values is adequate.

5. Suitability of MODFLOW and Related HSPF Models for Intended Applications

The use of HSPF to develop recharge (and maximum saturated ET) for input to MODFLOW-based groundwater flow models is suitable.

Note: items 6, 7, and 8 are combined into one response below.

6. Appropriateness, Defensibility, and Validity of Model/Relationships

7. Validity and Appropriateness of All Assumptions Used in Development of Model/Relationships

8. Deficiencies, Errors, or Sources of Uncertainty in Model/Relationship Development, Calibration, and Application

The NFSEG HSPF model is conceptualized, constructed, and calibrated appropriately over the majority of the model domain, and it is defensible for its intended purpose over the entire domain. The primary sources of possible errors derive from developing the model input rainfall and the implicit assumption that the measured streamflow is representative of the constructed watershed. Some of the watersheds are affected by processes that are not included in these models due to the limitations imposed by the large area and large number of models. These include man-made modifications, tidal effects, and large groundwater influence on surface water flows. The modelers made a decision to not include man-made changes in the models, and HSPF is generally not capable of representing significant groundwater or tidal effects without additional conceptualization and use of special features. Therefore, it is fair to say that the underlying HSPF process relationships are somewhat limited for accurately calibrating watersheds with these conditions unless they are explicitly included by the modeler. This is illustrated in many of the poorly calibrated gauges in the model. However, some of the poorly calibrated watersheds are likely resulting in reasonable and appropriate recharge, since many of the objective function criteria are being satisfied. In those watersheds where the percent bias is extremely high (and therefore the recharge is more likely to be invalid), it is recommended (in future calibrations of the model) that the model be modified to represent the man-made influences, or alternatively those watersheds should be assigned parameter values from a nearby watershed that is well calibrated. This recommendation of using calibrated parameters from another watershed should also be applied to gauges that have strong tidal influences.

4. LITERATURE CITED

Andersen, P., and M. Stewart. 2016. *Peer Review of the Northern District Model Version 5 and Predictive Simulations,* Final report to St. Johns River Water Management District and Southwest Florida Water Management District, Palatka and Brooksville, Fla.

Anderson, M., and W. Woessner. ed 1992. *Applied groundwater modeling: Simulation of Flow and Advective Transport*: San Diego: Academic Press Inc.

[ASTM] ASTM International. 2008. *Standard guide for defining boundary conditions in groundwater flow modeling.* D5609-16, vol. 04.08. West Conshohocken, Pa.

——. 2010., *Standard guide for selecting a groundwater modeling code*. D6170-17, vol. 04.09. West Conshohocken, Pa.

——. 2018. D5981, *Standard guide for calibrating a groundwater flow model application.* D5981 / D5981M-18, vol. 04.09. West Conshohocken, Pa.

Doherty, J. 2010. *PEST, Model-independent parameter estimation: User manual (5th ed., with minor revisions).* Brisbane, Australia: Watermark Numerical Computing.

Doherty, J. E. and R. J. Hunt. 2010. *Approaches to highly parameterized inversion: A guide to using PEST for groundwater-model calibration,* Scientific Investigations Report 2010–5169. Reston, Va.: U.S. Geological Survey

Kincaid, T. and B. A. Meyers. 2014. *North Florida numerical groundwater flow model: Limitations and Implication for groundwater resource management*. Live Oak, Fla.: GeoHydros LLC.

Geurink, J.S., and R. Basso. 2013. *Development, calibration, and evaluation of the Integrated Northern Tampa Bay Hydrologic Model*. Prepared for Tampa Bay Water and the Southwest Florida Water Management District, Clearwater and Brooksville, Fla.

Hill, M. C. 1994. *Five computer programs for testing weighted residuals and calculating linear confidence and predictive intervals on results from the ground-water parameter-estimation computer program MODFLOWP.* Open-File Report 93-481. Denver, Colo.: U.S. Geological Survey.

——. 1998, *Methods and guidelines for effective model calibration.* Open-File Report 98-4005. Denver Colo.: U.S. Geological Survey

Knowles, L., Jr., O'Reilly, A.M., and J.C. Adamski. 2002. *Hydrogeology and simulated effects of ground-water withdrawals from the Floridan aquifer system in Lake County and in the Ocala National Forest and vicinity, north-central Florida.* Water-Resources Investigations Report 02–4207. U.S. Geological Survey

Kuniansky, E. L. 2016. *Simulating Groundwater Flow in Karst Aquifers with Distributed Parameter Models: Comparison of Porous-Equivalent Media and Hybrid Flow Approaches.* Scientific Investigations Report 2016–5116. Reston, Va.: U.S. Geological Survey.

Miller, J.A., 1986. *Hydrologic framework of the Floridan aquifer system in Florida and in parts of Georgia, Alabama, and South Carolina*. Professional Paper 1403-B, U.S. Geological Survey.

Poeter, E., and M. C. Hill. 1997. Inverse models: A necessary next step in ground-water flow modeling. *Ground Water* 35 (2): 250-60.

Reilly, T.E., and A.W. Harbaugh. 2004. *Guidelines for evaluating ground-water flow models.* Scientific Investigations Report 2004-5038, U. S. Geological Survey.

Sepúlveda, N., C. R. Tiedeman, A. M. O'Reilly. J.B. Davis. and P. Burger. 2012. *Groundwater flow and water budget in the surficial and Floridan aquifer systems in east-central Florida.* Scientific Investigations Report 2012–5161. Reston: Va.: U.S. Geological Survey.

Smith, M., D. Laurine. V. Koren. S. Reed, and Z. Zhang. 2003. Hydrologic model calibration strategy accounting for model structure: Calibration of Watershed Models, In *Water Science and Application*, vol. 6. Wash., D.C.: American Geophysical Union.

APPENDIX A – ADDITIONAL CHAPTER COMMENTS

A.1 HAL DAVIS – MODFLOW DOCUMENTATION

Chapter 2: HYDROLOGY OF THE AREA

Surficial Aquifer System page 9:
Table 2-1: The superscript * needs to be replaced with a 1 in the legend.

Intermediate Confining Unit page 11:
Figure 2-9: Need to indicate if -/+ are upward or downward gradients.

Floridan Aquifer System page 13:
Page 15, 4th paragraph: A figure should be added showing the locations of the confining units described in the statement: “Another complication is related to the discontinuous nature of the middle confining unit. Miller (1986) mapped four different middle confining units in the model domain (numbers 1, 2, 3, and 7).”

Page 15, 5th paragraph: A figure, similar to the one below, should be added to help illustrate the configuration of the zones described in the statement: “It also allowed for continuity of model layering with areas where Miller (1986) had not mapped the presence of a middle confining unit, thus resulting in the definition of three continuous layers representing the Floridan aquifer system throughout the model domain, referred to hereafter as zones 1, 2, and 3 in the present report (Table 2-2).”

Page 16: Report states: “The thickness of the Upper Floridan aquifer ranges from near 0 ft along the Gulf Trough in Georgia to nearly 1,000 ft in the northwest region of the model domain (Figure 2-14).” In Miller’s report (1986) the thickness of the Upper Floridan is shown as about 300 ft in the Gulf Trough area.

Figure 2-16: Outline of Miller’s confining layers 1,2,3, and 7 would show where this zone 2 was permeable/impermeable.

Figure 2-24: Contours in legend appear to be reversed.

Page 19, last paragraph: Appendix B should be Appendix C.

Figure 2-33: Should Wakulla? Spring Creek? be on figure?

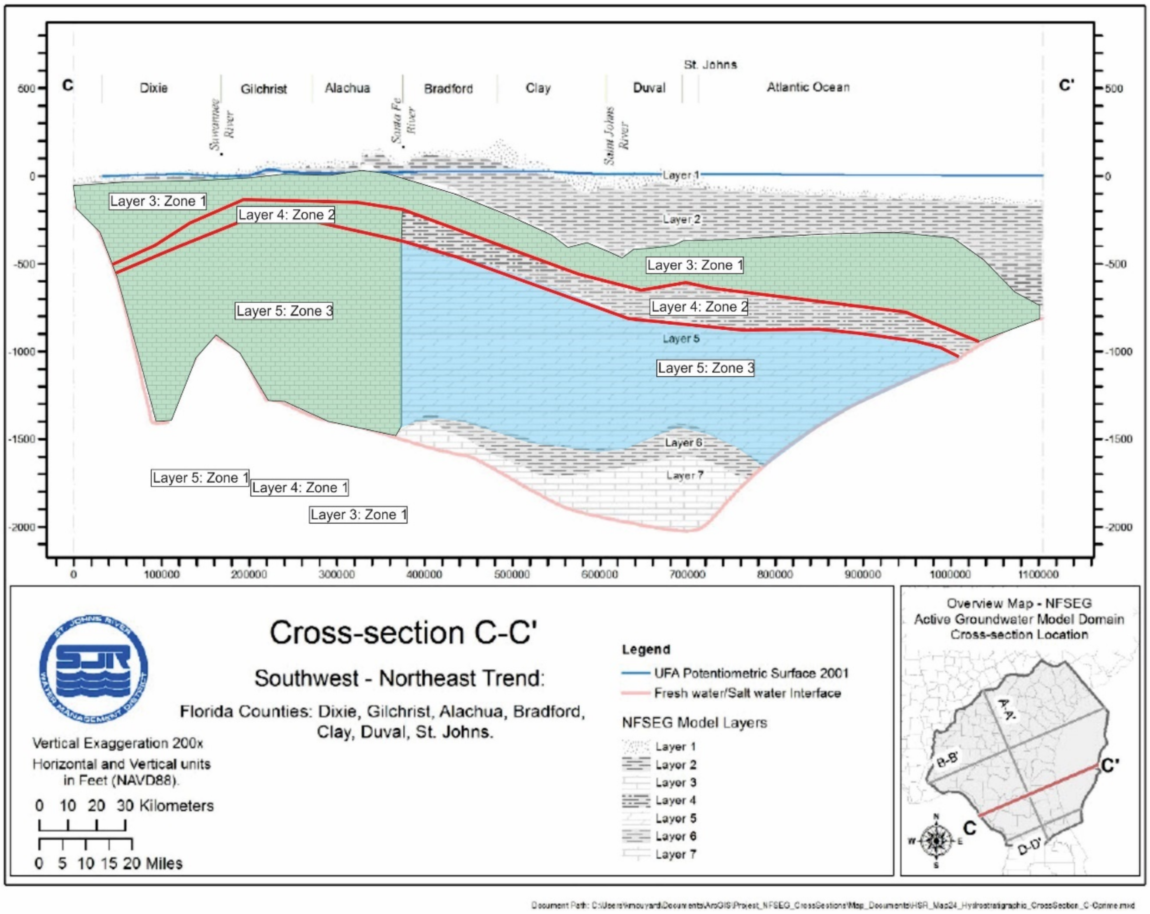


Figure 3-5. Hydrogeologic Cross Section C-C'

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) MODFLOW 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 <u>cumulative</u> 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 <u>cumulative</u> 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 A-1. Year 2001 measured and simulated cumulative flows.

		Year 2001			
Gage	**River**	**Measured, in cfs**	**Simulated, in cfs**	**Diff**	**Diff %**
2319000	WITHLACOOCHEE RIVER NEAR PINETTA, FLA.	445	627	182	34
2319394	WITHLACOOCHEE RIVER NR LEE, FLA	815	918	103	12
Alapaha R.					
2317500	ALAPAHA RIVER AT STATENVILLE, GA	164	403	239	84
2317620	ALAPAHA RIVER NEAR JENNINGS FLA	224	464	240	70
Suwannee R.					
2319500	SUWANNEE RIVER AT ELLAVILLE, FLA	1,918	2,196	278	14
2319800	SUWANNEE RIVER AT DOWLING PARK, FLA	1,976	2,282	306	14
2320000	SUWANNEE RIVER AT LURAVILLE, FLA.	2,290	2,407	117	5
2320500	SUWANNEE RIVER AT BRANFORD, FLA.	2,966	2,922	-44	-1
2323500	SUWANNEE RIVER NEAR WILCOX, FLA.	3,818	4,149	331	8
Santa Fe R.					
2322500	SANTA FE RIVER NEAR FORT WHITE, FLA.	563	535	-28	-5
2322700	ICHETUCKNEE R @ HWY27 NR HILDRETH, FL	202	202	0	0
2322800	SANTA FE RIVER NR HILDRETH FLA	912	874	-38	-4

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

		Year 2009			
Gage	**River**	**Measured, in cfs**	**Simulated, in cfs**	**Diff**	**Diff %**
2319000	WITHLACOOCHEE RIVER NEAR PINETTA, FLA.	482	841	359	54
2319394	WITHLACOOCHEE RIVER NR LEE, FLA	773	1,276	503	49
Alapaha R.					
2317500	ALAPAHA RIVER AT STATENVILLE, GA	247	773	526	103
2317620	ALAPAHA RIVER NEAR JENNINGS FLA	342	810	468	81
Suwannee R.					
2319500	SUWANNEE RIVER AT ELLAVILLE, FLA	2,552	3,013	461	17
2319800	SUWANNEE RIVER AT DOWLING PARK, FLA	2,668	3,130	462	16
2320000	SUWANNEE RIVER AT LURAVILLE, FLA.	2,910	3,308	398	13
2320500	SUWANNEE RIVER AT BRANFORD, FLA.	3,320	3,921	601	17
2323500	SUWANNEE RIVER NEAR WILCOX, FLA.	4,964	5,497	533	10
Santa Fe R.					
2322500	SANTA FE RIVER NEAR FORT WHITE, FLA.	730	727	-3	0
2322700	ICHETUCKNEE R @ HWY27 NR HILDRETH, FL	254	273	19	7
2322800	SANTA FE RIVER NR HILDRETH FLA	1,200	1,147	-53	-5

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.

MISCELLANEOUS ITEMS

Geographic places mentioned in the text that did not appear to be located on a map (and page that it was mentioned) are listed below:

St. Johns, Suwannee, Altamaha, Satilla, and Savannah Rivers 3
Flint, Ochlocknee, Aucilla, Steinhatchee, Wacissa, St. Marks, St. Marys, and Ocklawaha Rivers 3
Clay County, Florida, the Keystone Heights 3
Fernandina Beach 3
Suwannee River and Santa Fe River 5
Alapaha and Withlacoochee 6
Dead River 6
Alapaha Rise 6
St. Johns River 6
Lake George 6
Volusia, Marion, and Putnam counties 6
Silver Glen and Salt springs 6
Green Cove Springs 6

Woodville Karst Plain 65
Silver and Rainbow springs basins 65
Camden County, Georgia 65
Nassau County, Florida, into St. Johns County, Florida 65
Baker County, Florida, and Charlton County, Georgia 65
Leon and Jefferson counties, Florida 65
Marion and Levy counties 68
Gainesville 75

A.2 Louis Motz – MODFLOW Documentation

Chapter 1 Introduction

p. 1: The primary purpose of the NFSEG model is to enable improved evaluations of inter-district…and interstate…water-level changes in the surficial and Floridan aquifer systems resulting from groundwater use over the model domain. Consider adding determination of changes in spring flows and base flows to the description of the primary purpose of the NFSEG model.

p. 3 and Figure 1-1: In the paragraph Municipalities and Other Major Pumping Centers, reference should be made to Figure 1-1. Should **Valdosta** and **Ocala** be included in the list of pumping centers?

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."**

p. 14: …Gulf Trough in Georgia…. Suggest adding **…Gulf Trough in *south* Georgia….**

p. 19 and References, p. 109: There are two references for Kuniansky and Bellino (2012) in the References section on p. 109 Thus, 2012a and 2012b should be indicated on p. 109 in **References**, and the reference on p.19 should be identified as Kuniansky and Bellino (2012a) or (2012b).

p. 21: "**Springs with discharge rates that are greater than or equal to 100 cfs on average are classified as first magnitude springs."** Please provide a reference for this.

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?

p. 31 and Figure 3-1: Please indicate on Figure 3-1 that the northern NFSEG Active Model Boundary is the **"approximate up-dip limit of (the) productive part of the Upper Floridan aquifer…"** and refer to Figure 3-1 again on p. 31.

p. 41: …equivalent freshwater head…" Please provide a reference for the equation given for calculating equivalent freshwater head.

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. 68, Figure 5-1: The bar graphs for annual precipitation should be in the order of 2001, 2009, and 2010.

pp. 68-69, Figures 5-3 and 5-5: The bar graphs for annual ET and recharge should be in the order of 2001, 2009, and 2010.

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 Stringfield (1936) except for Juniper and White springs. Can the results for these springs, particularly White Springs, be improved?

p. 79, Table 5-6: The reference to **Stringfield (1936)** needs to be included in the References section.

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:

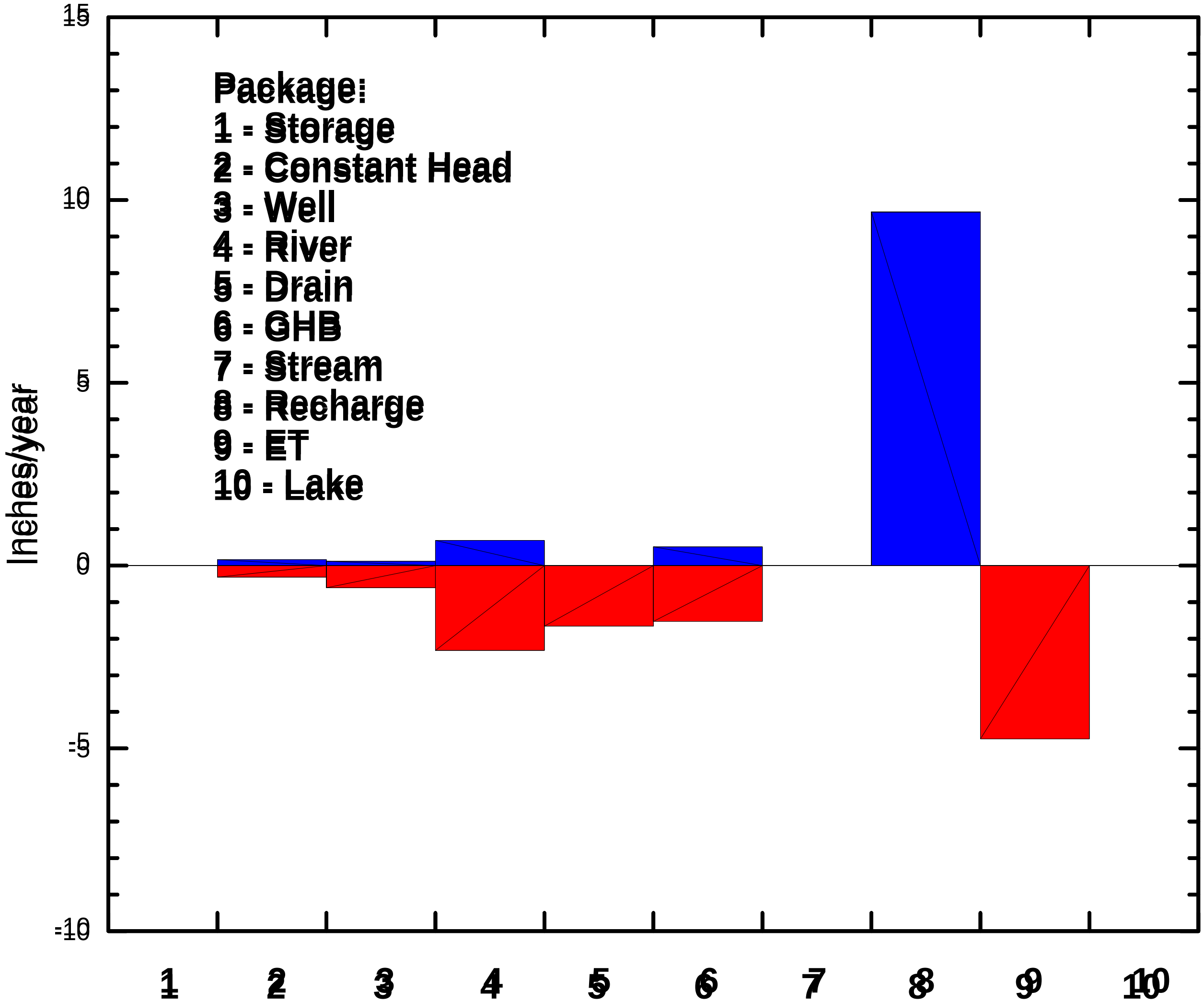


Figure A-1. Model-Wide Water Budget for 2001

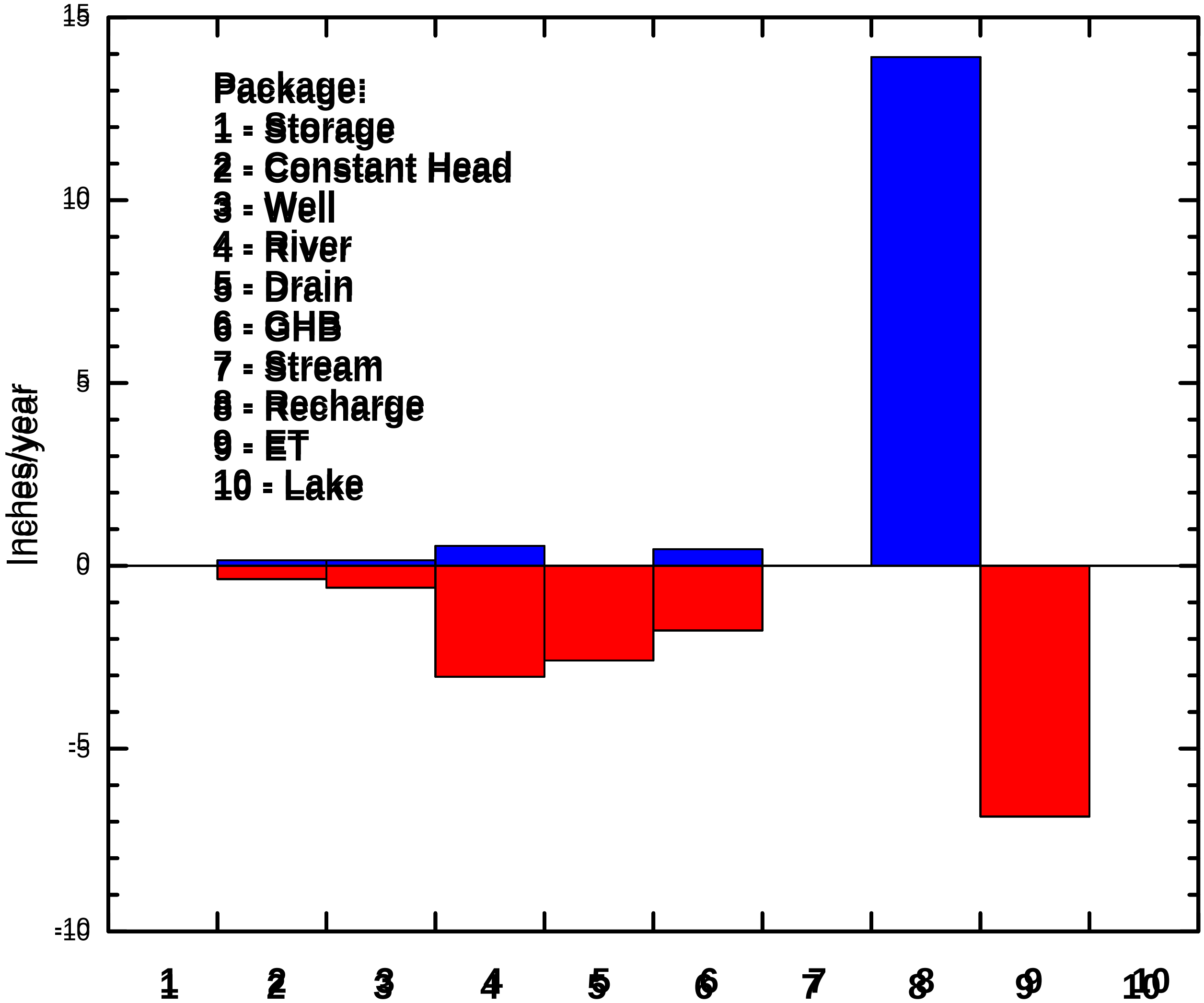


Figure A-2. Model-Wide Water Budget for 2009

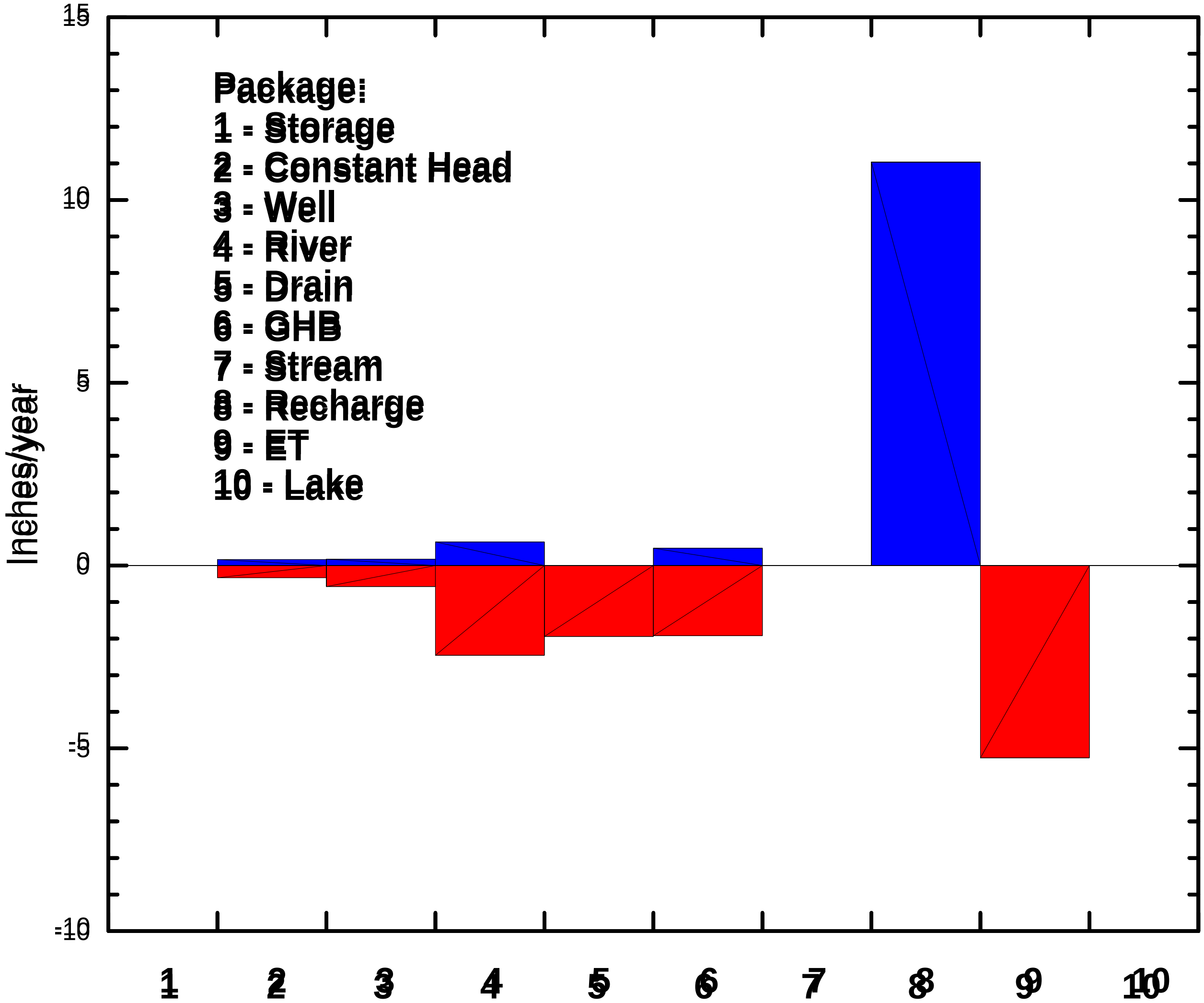


Figure A-3. Model-Wide Water Budget for 2010

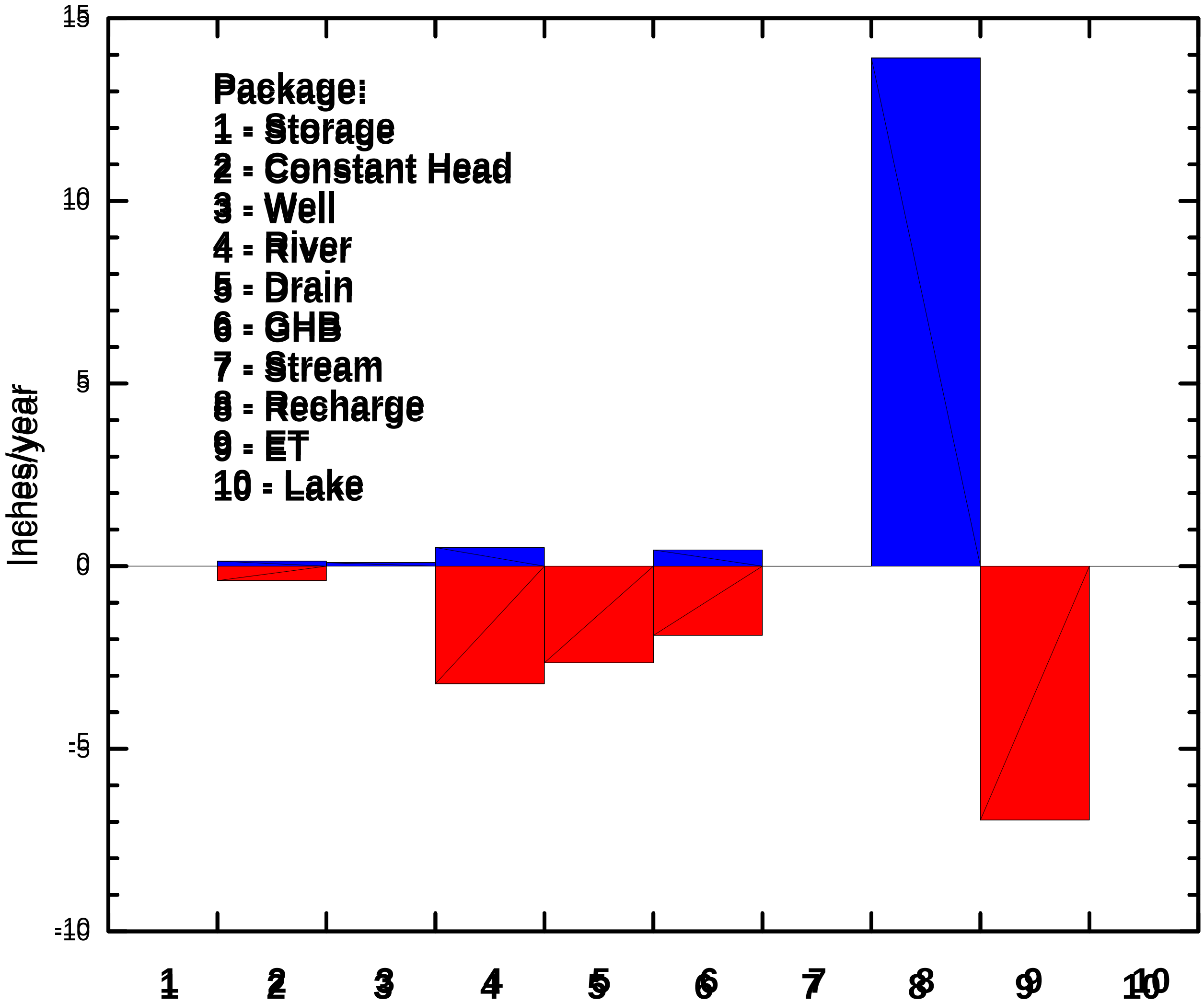


Figure A-4. Model-Wide Water Budget for 2009 Pumps Off

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. 100: Land surface elevations range from sea level to more than 450 feet, NAVD88 in *northern* Georgia. Change *northern* to *south* Georgia.

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).

References

p. 109: The two references to **Kuniansky and Bellino (2012)** should be delineated as **2012a** and **2012b**.

A.3 DANN YOBBI – MODFLOW DOCUMENTATION

Chapter 3. Model Configuration

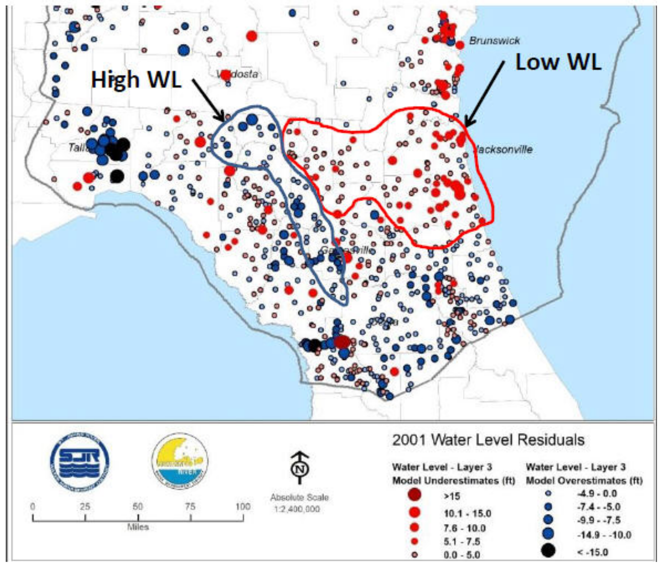
1. The discretization used in this model is appropriate at annual time steps and at regional scales but exercise extreme caution when simulating at local scales and shorter time steps. Although this report states that the NFSEG model can be used for "local-scale evaluations with the same level of accuracy 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 mi^2-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 (Kincaid and Meyers 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

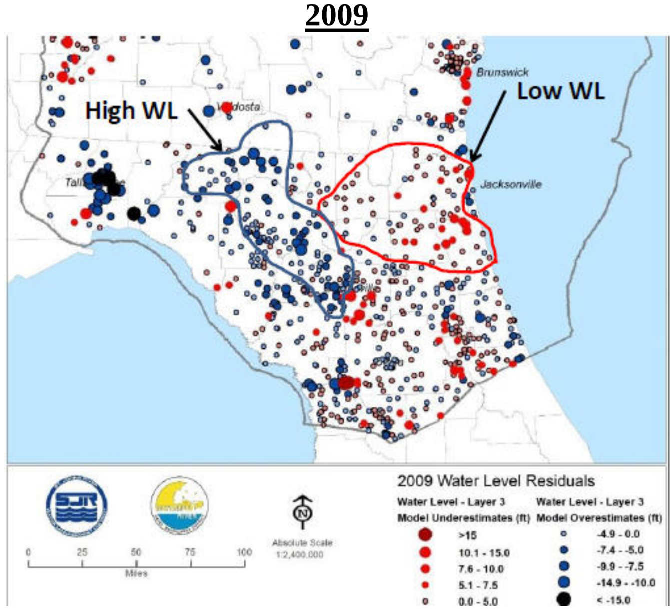


Figure A-5. Examples showing selected model-wide trends in spatial groupings of positive and negative water-level residuals in layer 3 (written com. Liquid Solutions Group, LLC, 2018).

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. In areas of the model domain where model calibration needs improvement, manual ("trial and error") simulations in addition to PEST simulations should be performed to ascertain if model results are in better agreement with field data, aquifer-test values, other published numerical models, and professional experience and judgment.

6. 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

7. 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.

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.

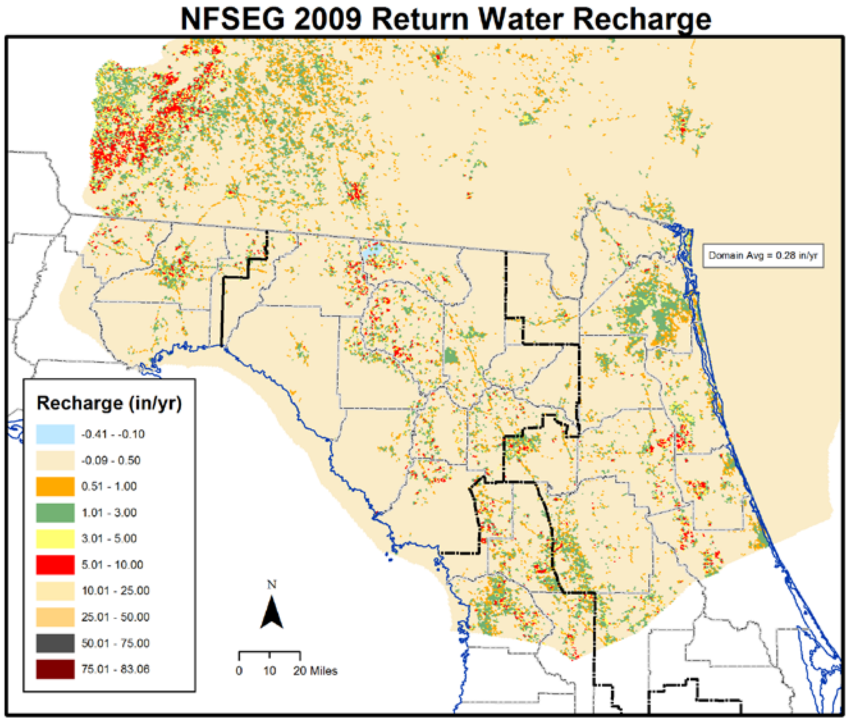


Figure A-6. NFSEG V1.1 2009 Return Water Recharge (written communications SWFWMD, 2018)

Chapter 6. Water Budget Analysis

8. 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

9. 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 0.22mi^2 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).

Editorial Comments for MODFLOW (Errata Sheet)

1. Title page--Change publication year to 2018.
2. Add list of acronyms and abbreviations.
3. Table of Contents is incomplete. 3rd and higher-level order headings are omitted.
4. Label all counties names mentioned in report on one or more figures.
5. Label within the map area the "active model boundary" on each figure where it is shown.
6. Adopt consistent nomenclature for "legends" throughout report. Legend in fig. 3-6 is proper standard.
7. Delete measurement units from report figures.
8. Add fall line to fig 1-2. Text mentions it on p. 2 ¶5
9. Need figure showing major surface water basins. Text mentions them on p. 3 ¶1.
10. Label all hydrographic features named in text to one or more maps—Atlantic Ocean, Gulf of Mexico, rivers, lakes, cities, etc.
11. Potentiometric highs not shown on fig 1.4 but mentioned in text on p.3 ¶2
12. Figure 2-44 is incorrectly labeled.
13. Adopt consistent x axis label on figs 3-4 through 3-7.
14. Check legend scale for figure 3-8 through 3-24. Cannot identify lower intervals on maps?
15. Add histograms to figures 4-13 and 4-14. Define lines and shaded area shown on graphs and revise axis labels. X axis should read *Observed Water Level* and y axis should read *Simulated Water Level.*
16. Add histograms to figures 4-19, 4-20, 4-23, 4-24, 4-27, 4-28. Define lines and shaded area shown on graph. Revise x and y axis titles.
17. Statistics on figure 4-24 do not correspond to values in table 4.4.
18. Figures 4-29 and 4-30--Contour interval of 5 ft is too detailed for the range in values. Suggest a minimum 10 ft interval. Where are the observed 2001 and 2009 maps? On page 61, "*these surfaces represent a good to excellent match to the respective **observed potentiometric surfaces of 2001 and 2009***".
19. Add proper x and y axis titles to figure 4-34.
20. Add histograms to figures 4-37, 4-38. Define lines and shaded area shown on graph. Add proper x and y axis titles.
21. Figures 4-39 and 4-34. Change to a 10 ft contour interval.
22. Why are flow rates on figure 4-41, 4-43 through 4-46 negative? Flow is a positive number (see fig.4-42). Also, report flow rates to a maximum of 3 significant figures.
23. What springs are plotted on figures 4-43 and 4-44? Many more springs are listed in appendix E.
24. Add histogram to figure 4-53 and 5-54. Define lines and shaded area shown on graph. Add proper x and y axis titles.
25. Figures 4-55 and 4-56. Report flow rates to a maximum of 3 significant figures.
26. Define lines and shaded area shown on figure 4-57 and 4-58.
27. Check legend scale on figure 4-61. 3,500 is very high.

28. Check legend scale on figure 4-65 and 4-66. 500 is very high.
29. Check legend scale on figure 4-67 and 4-68. 2,047.7 is very low value.
30. Check legend scale on figure 4-69. 53,902 is very high.
31. Check legend scale on figure 4-71. 77,851,888 is very high.
32. Add histograms to figures 5-11 through 5-13. Define lines and shaded area shown on graphs and correct axis labels. X axis should read *Observed Water Level* and y axis should read *Simulated Water Level.*
33. Add histograms to figures 5-15 and 5-17. Define lines and shaded area shown on graphs and correct axis labels. X axis should read *Observed Water Level* and y axis should read *Simulated Water Level.* Observed discharge is a positive value not a negative value.
34. What springs are plotted on figure 5-15? Number plotted here do not match number in appendix J.
35. Why are observed baseflow pickups negative on figure 5-17 and on table 5-3? Baseflow is a positive number.
36. Why is a different scale used in figure 5-21 and 5-22 than in scale used in 2001/2009 maps?
37. Table 6-4—Why is there 0.10 in/yr model-wide pumpage for no-pumping simulation?
38. Table 6-16-- Why is there 0.24 in/yr pumpage in GWB3 for no-pumping simulation?
39. Table 6-32-- Why is there 0.01 in/yr pumpage in GWB7 for no-pumping simulation?
40. Define symbols used on figures 7-7 through 7-11.
41. Floridian is incorrect spelling on page 2 ¶5.
42. Need to reference Groundwater Vistas on page 51 ¶1.
43. Page 57 ¶2, what "*critical lakes*" were assigned a different weight on page 57 ¶2?
44. Page 58 ¶3, what is the "*large range in groundwater levels*"?
45. Page 58 ¶2, provide proof for the statement "*the scatter plot that is present is to be expected for such a varied and complex range of conditions*".
46. Page 72 ¶1, change appendix x to appendix I.
47. Page 72 ¶2, Why different sets of springs assessed in 2010 for "*important first magnitude springs and spring groups*".
48. Page 72 ¶3, Why no table for 2001 and 2009 for "*important first magnitude springs and spring groups*".
49. Page 74 ¶1, Change Appendix Z to Appendix K.
50. Page 74 table 5-6, once again, why a different set of springs evaluated? Need consistent data sets to evaluate results.
51. Page 97 ¶1, no proof is provided to substantiate your statement "*the model can be used for* ***sub regional and local-scale evaluations*** *with the same accuracy as existing models*". How do you know this?

A.4 JAMES RUMBAUGH – MODFLOW DOCUMENTATION

Editorial Comments for MODFLOW

Page 45 – the text mentions that wetting penalty was only used early in the calibration process, however, run 007h did use this observation type.

Page 48 – Keystone Heights is mentioned here. It would be good to have this on a map for reference.

Page 49 – The text states that flooding issues were due to lack of representation of surface water. That may be, but there are still a lot of flooded cells with some very much above top of layer 1.

Page 50/51 – For the sake of completeness it would be good to document the variogram parameters used in the kriging of pilot points.

Page 55 – the text discusses the problems with water levels in layer 1 being above the top of the model cell. One thing to consider for the future is to calibrate on depth to water in layer 1, rather than elevation.

It would be useful to show a map of the areas covered by the wetting penalty observations.

On Table 4-4, it would be good practice to provide scaled calibration statistics (divide by range in head).

Figure 4-76 – X and Y axes should be the same length.

Figure 4-13/14 – what is the green shaded area on these types of plots? Should discuss in the text.

Figures 4-15/16 – rather than contours, perhaps use color shading as the contours are impossible to read.

A.5 BRIAN BICKNELL – HSPF DOCUMENTATION

HSPF Model Development:

Representation of Springs to Improve HSPF Calibration
I recommend that this section include discussion of the calibration of spring flows that was done.

Calibration Process:

This section discusses the calibration period. I recommend including a discussion of the decision to calibrate with all available data instead of calibration with part of the data and validation/verification with the remaining data. Both approaches are acceptable, and this project with its utilization of many gauges that have partial records, would have been unnecessarily complicated if the gauges had to be subdivided into two parts.

Also, some reviewers have apparently recommended calibrating to individual years to improve recharge for just those years. That is not a good idea; HSPF should utilize all available data (time spans) in a single calibration so that the model more robustly simulates different hydrologic conditions. Calibration to short periods eliminates this advantage of calibration to a long period of record.

Parameter Estimation with PEST:

The discussion of PEST objective function should include more details about the weighting factors that were applied to various components. How sensitive is the model to these weights, and did you try to assign much heavier weights to any components to try to improve the agreement of percent bias and Nash-Sutcliffe coefficient? Also, since PEST calibration is relatively uncommon, I recommend more backup or description of the more obscure components in Table 13.

Calibration Results:

This section includes all calibration results, including the detailed appendices of statistics and graphics for all models. I recommend including description of the calibrated parameter sets and water balance summaries that I have described elsewhere, and refer to those data sets in the respective appendices.

I recommend brief discussion of the apparent reason(s) for poor calibration agreement for those gauges that are poorly calibrated.