MEMORANDUM

TO: Pete Loucks, Cornell University, Chair
David Chin, Consulting Engineer
Robert Prucha, Integrated Hydro Systems, LLC

FROM: Ken Konyha, Section Lead, Hydrologic and Environmental Systems Modeling
Christine Carlson, Sr. Supervising Geographer, Kissimmee Division

DATE: August 14, 2008

SUBJECT: SFWMD Responses to the Peer Review Panel Task 3 Report for the Kissimmee Basin Modeling and Operations Study (KBMOS)

The Model Peer Review Panel Task 3 Report for the KBMOS (June 14, 2008) found no critical defects in the modeling and operation study. The panel did identify remaining issues and opportunities. This memo responds to the Panel's Findings and Recommendations. District responses are shown in *italics.*

Comments relative to Peer Review Scope Objective 3A: Assess whether the selected models have been appropriately formulated and calibrated to evaluate existing and proposed Kissimmee Basin structure operating criteria

<u>Effect of operating policy deviations</u>

The MIKE 11 model was appropriately calibrated to match the hydrology and hydraulics of the system; however, since the actual gate operations deviated from the operating rules during the calibration period, it was necessary to artificially specify the target headwaters at the structures as being equal to the observed headwaters. If the structures were operated in accordance with their rules during calibration, there would be substantial additional deviation between model results and observations. The model is calibrated appropriately for comparing alternatives using fixed sets of operating rules, with the understanding that, in reality, the rules will not be followed exactly and actual stages and flow rates will be different than those simulated. The comparative methodology being used in this study assumes that these deviations from operating rules do not affect the relative rankings of alternatives.

District Response: *none*

<u>Use of Performance Measures in model calibrations</u>

PM values must be accurate enough to distinguish between AES scores for different alternatives. Instead of using just surface water stage, flow and ground-water elevations, other information more relevant to obtaining more accurate PM values could have been used to also guide model calibration. For example, surface-water flow reversals, peak flows, or ascension/recession curve information could have provided additional constraints on model calibration. Calculating accurate performance measures is essential for intended purpose of the model.

***District Response**: none*

Use of Performance Measures in model comparisons

The validity of the MIKE 11 and OKISS models in matching the hydrology and hydraulics of the system was measured in terms of errors in stages and flow rates; however, since these models are to be used in formulating and ranking alternatives based on performance measures (PMs), it would have been preferable to assess the validity of the models by their ability to accurately simulate the PMs derived from observations. Apparent good agreement in stages and flow rates does not necessarily translate into good agreement with respect to PMs. Unfortunately, the short length of record used for validation did not allow for an evaluation of the models relative to their ability to accurately simulate PMs. Based on these results, it must be assumed that the relative rankings of alternatives derived from these models (using PMs) is accurate, even if the absolute accuracy of the individual PMs cannot be assured.

***District Response**: none*

Different assumptions when calibrating and validating models

The MIKE 11 model is fairly robust in accounting for physical changes in the system, since it was calibrated under post-Phase I conditions (2001-2004) and validated under pre-Phase I conditions (1994-1998).

***District Response:** Agreed.*

Lateral inflow assumptions

MIKE 11 and OKISS use a constant set of lateral inflows in formulating and ranking alternatives, even though these lateral inflows are in reality affected by the different alternatives. Since the system is driven by lateral inflows, there is a concern that using a constant set of lateral inflows will compromise the accuracy of the alternative rankings. It is important that the insensitivity of the alternative rankings to the lateral inflows be demonstrated. To support the assumption of insensitivity to lateral inflows, the panel concurs with the District's plan that when the final three alternatives are run using the MIKE SHE/MIKE 11 model, then these lateral inflows will not be fixed and will vary in response to precipitation and operating policy driven river flows. The resulting PM values will be compared to those resulting from fixed lateral flow simulations to see just how sensitive this fixed lateral flow assumption is and to see if this does not affect the alternative rankings of the top three alternatives.

***District Response:** none*

Conceptual flow model component interactions

More effort could have been made to validate the conceptual flow model performance. This is often an overlooked step in model development, but it can substantially add credibility in the model. For example, numerous watershed water-budget tables were provided throughout the KUB and KLB surface drainage area. Yet, there was little discussion of what processes dominate the lateral inflows to the main surface drainage. It would have been helpful to illustrate how the various key components interact with one another, and which are most important for the intended use of the model. The following are key lateral inflow components:

- baseflow (model only defines surface water-ground water flows between two cells. It doesn't account for baseflow beneath wider channels, or lakes).
- saturated zone drainage (used to simulate overland flow in man-made canals)
- overland flow
- lake-groundwater (flood cells in MIKE SHE/MIKE 11). This wasn't included in the water-budget table but should be, because it is effectively the same as baseflow. Including this might actually show that 'baseflow' is a more important lateral inflow component than suggested in the water-budget tables.

It is often useful to produce spatial distributions of water components, on a cell-by-cell basis. This typically gives good insight where, for example, the most significant flows from drain-cells occur. It would also show where overland flow is derived compared to the saturated zone drainage areas. Available field data, even more qualitative data like known areas of surface saturation versus dry areas could be used to justify model performance and use of the saturated zone drainage feature in MIKE SHE to simulate overland runoff. This can increase overall model credibility. This information can also point out where data is insufficient and where there is greater uncertainty in the model.

> ***District Response:*** *Agreed. A recalibration effort of AFET is underway and this provides a good opportunity to examine the water budget components in light of the calibration expectations and targets.*

<u>Surface drainage</u>

The saturated zone drain feature in MIKE SHE could have been described better. This is important because it represents the largest component of lateral inflows calculated in the calibrated model. In the MIKE SHE model (AFET), is the procedure used to simulate the surface water flow from the large number of local-scale man-made canals to the main channel in MIKE 11 assumes the SAS aquifer groundwater levels exceed specified drain depths (i.e., between 0 and 2.5 feet below ground surface). This procedure may not account for flow in the man-made canals when groundwater levels are below the drain depths. It is also worth investigating whether specifying the same drain constants in cells several miles from the main surface water drainage channel as those immediately adjacent to the drainage channel is appropriate. It was difficult to assess whether this assumption significantly impacts surface water flow in the main surface water drainage.

> ***District Response:*** *The surface drainage module in MIKE SHE is limited and requires a simplified modeling of the tertiary (i.e. non MIKE 11) drainage network. An additional limitation not noted by the Panel is the model's inability to conceptualize detention storage. This is particularly important in Florida where Regulations require that all developed lands provide one inch of detention storage. In contrast, MIKE SHE assumes that urban lands have directly connected impervious areas that drain directly to the MIKE 11 system. The recalibration effort should consider this limitation and consider reformulating urban land components in MIKE SHE.*

<u>Surface – ground water interactions</u>

Considerable effort has been made to show why an integrated surface-groundwater flow code was necessary for this project. Yet, it was difficult to see where the surface-groundwater coupling was most important. The water-budget tables in the calibrated AFET model report indicate that direct baseflow represents only a small percentage of lateral inflow to each catchment compared to the saturated zone drainage inflow. To simulate actual drain flow to the MIKE 11 drainage network, groundwater levels in the SAS aquifer must exceed specified drain elevations. Therefore the SAS aquifer groundwater elevation (or depth below ground) in non-MIKE 11 drainage areas (large area in model) is a critical simulated output that could be discussed at greater length. If data used to define this surface drainage are sparse (especially related to drain depths), increased uncertainty is produced in these areas.

> ***District Response:*** *Agreed. The ongoing model recalibration effort should consider the surface water – groundwater interactions.*

<u>Predictive uncertainty</u>

Calibration error is an important factor that affects the ultimate objective of this modeling effort. It translates into predictive uncertainty and therefore can affect the ability to distinguish between proposed alternatives. To address this predictive uncertainty, the District assessed base condition sensitivity to adjustments in key parameter values and found that changes of 2 to 3 inches per year had significant impacts on PM values. A high and low estimate of lateral runoff (i.e., current versus future base conditions) will be used to modify the AES scoring. Additional recommendations for addressing uncertainty include the following:

- Uncertainty is a complex and challenging issue in all hydrologic modeling. Despite this, efforts should be made to define the various components of uncertainty, and possibly show how these relate to model predictions in a simple illustrative diagram. Summarizing the potential sources of uncertainty (model structure, parameter values, data) in a table could also be prepared. This can help identify key sources of uncertainty, and also promotes transparency to readers. This can be useful even if levels of uncertainty for each model input (i.e., topography, or precipitation, or aquifer thickness) are assigned simple qualitative rankings of low, medium and high. Identifying these up front can show where future model updates, improvements etc can focus efforts, and it can also be useful in describing model limitations.
- The District has addressed model uncertainty in past workshops. An effort should be made to build on this information when developing complex models such as this.

> ***District Response****:* *Agreed. The District's Hydrologic and Environmental Systems Modeling Department is aware of these issues and has initiated a model uncertainty project to begin addressing them.*

<u>Predictive uncertainty effects on AES</u>

The ability of the KBMOS models as formulated and calibrated to evaluate proposed structure operating criteria could be better assessed by considering the effects of predictive uncertainty on Alternative Evaluation Scoring (AES). The uncertainty in AES scores might be significant and different for each alternative operating criteria simulation, and if so could limit the ability to distinguish between alternatives. Because simulated output from proposed alternative structure operating criteria have not been provided, it is difficult to fully assess whether the calibrated models are actually capable of distinguishing among different alternatives. If they were found not to distinguish given the uncertainty, the models would require further calibration to reduce model error to the point that alternatives could be distinguished. In future modeling efforts such as this, where a complex fully integrated flow model like MIKE SHE/MIKE 11 is developed, it might make sense to consider the calibration phase of the project as an iterative step that is not really complete until fully testing the models with all alternatives. An analogy would be developing a very complex race car that you've calibrated in the lab, but haven't really tested fully on various race courses, for different weather conditions.

> ***District Response:*** *This should be considered as part of the review of the KBMOS results. Even though so far, the models have been able to differentiate between alternatives using the AES.*

Comments relative to Objective 3B: Do the results from the base conditions simulations provide an appropriate benchmark for use in the Alternative Plan Selection Process?

Evapotranspiration values

The potential/reference evapotranspiration values appear to be suspect and these values should be revised prior to using the models to rank alternatives.

> ***District Response:*** *The District concurs. A District-wide reference evaporation data set has recently been developed and this improved data set is the basis for the AFET recalibration effort.*

Use of Performance Measures for model comparisons and alternative rankings

It has been adequately demonstrated that all models are capable of implementing specified operating rules and generating comparable stages and flows for a given set of operating rules under current and future base conditions. However, when compared using PMs, significant discrepancies can occur between models using the same base conditions and it must be demonstrated that these discrepancies do not affect the ranking of alternatives. This could be demonstrated by comparing the ranking of the final three alternatives using the MIKE 11 model with the ranking of these same alternatives using the MIKE SHE/MIKE 11 model.

There are some significant discrepancies in the ability of the OKISS, MIKE 11, and MIKE SHE/MIKE 11 to simulate the same PMs under the same base conditions, apparently due to the high sensitivity of some of the PMs to stage-duration conditions. These highly sensitive PMs should be revised to more stable measures and the ability of the models to estimate comparable values of these PMs revisited.

> ***District Response:*** *The validation exercise will be updated using a revised version of the PMs. The final version of the Evaluation of Base Conditions Report will include the updated validation.*

Impact of uncertainty on comparing alternatives

If proposed changes associated with different alternatives do not significantly change flow conditions compared to base conditions, it is unlikely that differences in uncertainty between alternatives will be significant. In this case, AES scores will likely be able to distinguish between alternatives. If, on the other hand, proposed alternatives produce significantly different flow conditions compared to base conditions, greater levels of uncertainty may be introduced in the alternative AES scores. This could make it difficult to distinguish between alternatives.

District Response: *Agreed.*

<u>*Other comments*</u>

Portions of the OKISS model documentation report could be clarified. In particular, descriptions of the flow relationships could be revised, and diagrams could be added to aid in the explanations.

District Response: *OKISS model documentation report will be updated to address panel suggestions.*

The Uncertainty Analysis report describes a sensitivity analysis and the title and wording within the report should be revised appropriately. Parameter value variations of ± 50% are used. It would be more useful to use the actual variations of individual parameters. It should be made clear in the report that the accuracy of the PMs cannot be established using a sensitivity analysis. If experts could identify the likely probability distributions of significant (sensitive) parameter values, it would go a long way towards completing at least a partial uncertainty analysis. A further step would be to estimate the uncertainty in the PMs themselves.

District Response: *The Uncertainty Analysis and references within will be renamed and reworded per the panel's suggestions.*

It is not clear to the panel how conflicts among stakeholders will be addressed should not all stakeholders concur with the results of the AES process.

District Response*: The Computer-Aided Participation process being used to screen alternative plans and refine performance measures and AES parameters is an open process that allows interested stakeholders a seat at the table. This process is intended to allow concerns with the study objectives, performance measures, and AES parameters to percolate to the surface prior to moving forward with the promotion of alternatives to formulation and evaluation. Over the last 6 months, additional time has been added to the Study to allow stakeholders the time and opportunity to participate in alternative plan development process and the refinement of performance measures and AES parameters. This open process has reduced and avoided conflicts and has provided interested stakeholders a seat at the table.*

M E M O R A N D U M

TO: Pete Loucks, Cornell University, Chair
David Chin, Consulting Engineer
Robert Prucha, Integrated Hydro Systems, LLC

FROM: Ken Konyha, Section Lead, Hydrologic and Environmental Systems Modeling
Christine Carlson, Sr. Supervising Geographer, Kissimmee Division

DATE: August 14, 2008

SUBJECT: Action required to address the Modeling Peer Review Panel Review of AFET Model Documentation / Calibration Report

The following memo summarizes the status of Peer Review Panel comments received on the AFET documentation. The purpose of this memo is to identify specific processes which address Panel Recommendations. District annotations are shown in Blue

The status of each comment is provided following each item below. RESPONSES HAVE NOT BEEN INCLUDED. Only status and follow-up activities are identified. The summary at the end of the memo includes comments that require modification to the project documentation.

Comments:

1. (DAC) The calibration period is post Phase I (11/1/01 – 12/31/04) and validation period is pre-Phase I (1/1/94 – 10/15/98). This creates an added challenge for the model since both the hydrologic loading and internal structure of the (MIKE 11) model are both changing. The performance of the model seems to demonstrate that it can adapt to such structural changes as Kissimmee construction moves along.

Addressed in teleconference, no action required

2. (DAC) The procedure for operating the gates (control structures) during model calibration is questionable. During calibration, the gates are operated based on the difference between observed and simulated headwater stages, thereby using data that would not be available when the model is used in the prediction mode. As a consequence, the model parameters derived during calibration might not be optimal in reproducing observed stages and flows when structure operating rules based on simulated headwater elevations are used. Furthermore, this method of gate operation during calibration will pretty much guarantee that the simulated and observed headwaters are close. Using this approach, the calibration results appear to be adequate, especially at the gates. However, much of the stage (and flow) agreement was forced by operating the control structures using observed data.

The Peer review panel included or considered this comment in subsequent documentation and refined comment(s). The SFWMD has prepared and distributed a formal response which includes the actions necessary to address this comment.

3. (DAC) The planned uncertainty analysis will need to account for both parameter uncertainty and structural (model) uncertainty. It is entirely plausible that errors associated with the model structure are much greater than errors caused by parameter uncertainty. The report (page 6-4) seems to indicate that only parameter uncertainty will be considered. As a prelude to assessing the impact of model uncertainty, it would be useful to establish whether the errors are homoscedastic and, if they are not, what transformation of the model output would be appropriate (e.g. log transform) for characterizing model error with a constant variance.

The Peer review panel included or considered this comment in subsequent documentation and refined comment(s). The SFWMD has prepared and distributed a formal response which includes the actions necessary to address this comment.

4. (DAC) Overall, the calibration effort to date meets professional standards. To further apply of the model to meet the KBMOS objectives it is recommended that the points raised here be taken into consideration.

Goal of this document is to summarize how panels comment will be "…taken into consideration."

5. (DAC) The methodology used in the report to establish parameter sensitivity relates the percentage change in model output to the percentage change in parameter values. Presumably, the most sensitive parameters are those that produce the greatest percentage change in output per percentage change in parameter value. The limitation of this approach is that it does not recognize that different parameters change over different orders of magnitude, and hence a parameter that can change over three orders of magnitude might have a more significant impact on model output than a parameter that varies over one order of magnitude, even though the parameter that varies over one order of magnitude might create a larger percentage change in output per percentage change in input.

Addressed in teleconference, no action required

6. (RHP) The PowerPoint overview of calibration was much more helpful in assessing the calibrated model performance than the report. The location maps showing calibration targets and their performance in the PowerPoint file would have been much more helpful/convincing in the report.

Addressed in teleconference, no action required

7. (RHP) There are two types of sensitivity analyses typically performed with hydrologic models;

   a. as performed in this study, before calibration, to identify key parameters, and
   b. performed after calibration as per ASTM standards like D5611 such as for a groundwater flow model. The second approach is used to examine sensitivity of calibration residuals and model conclusions to model inputs to assess adequacy of the model with respect to its objectives.

The Peer review panel included or considered this comment in subsequent documentation and refined comment(s). The SFWMD has prepared and distributed a formal response which includes the actions necessary to address this comment.

Clarifications:

1. (DAC) Page 2-39 states that "Design drawings were used to represent Weir 1, 2, and 3 in the Pre-Phase 1 model as a broad-crested weir". Aren't these sharp-crested weirs?

Addressed in teleconference, no action required

2. (DAC) It is not clear how the results of the sensitivity analysis were used in the calibration of the model. Specifically, were the calibration variables a subset of those variables used in the sensitivity analysis or was the sensitivity analysis simply used to establish a priority of adjusting variables? Also, it is not clear what is gained by normalizing the relative sensitivity by the annual rainfall (page 3-11).

Addressed in teleconference, no action required

3. (DAC) I might have missed it, but the acceptance criteria for R and RMSE should be stated in the report near the tables with performance metrics (e.g. Table 4-4). It is a bit curious that a RMSE of 2.88 ft would be acceptable for PC52 in Pool BC (Table 4-4). These are others like this, for example a RMSE of 2.97 ft in the SAS water table at PINEISL is acceptable. What are the acceptance criteria?

Requires update to AFET Documentation Report, see Memorandum Summary at the end of this document.

4. (DAC) The calibration results have produced many areas of good agreement and some areas of bad agreement (e.g. errors in UFA potentiometric surfaces as much as 20 ft). However, it is not possible at this stage to determine whether the model is providing acceptable results since the ultimate test of the model is whether it can discriminate between specified hydrologic performance outcomes for varying structure operating rules. The impact of the model errors on discriminating between performance outcomes has not been established and so it cannot be determined whether the model errors and their spatial distribution are small enough for the purposes of the model. Can the District comment on this?

The Peer review panel included or considered this comment in subsequent documentation and refined comment(s). The SFWMD has prepared and distributed a formal response which includes the actions necessary to address this comment.

5. (RHP) Performance of the calibrated model in reproducing observed surface water stage and flow seemed much better than for the groundwater aquifers. Given that surface water calibration is much more important in meeting objectives (i.e., nearly all performance measures are related to surface water), I didn't get a sense of how important, or sensitive the surface water flow response is to the groundwater flow

system. Did the modeling team develop a sense of this, and is it still considered important for intended use of the AFET model?

Addressed in teleconference, no action required

6. (RHP) Given that this fully integrated model is more sophisticated than previous groundwater flow models (i.e., driven by continuous PET and distributed precipitation and better SW-GW interaction), it is surprising that more groundwater parameters, and their distributions were not considered for adjustment (i.e., 6.3.1) during calibration. It is also surprising that spatially distributed recharge calculated with the fully integrated model is compared to recharge distributions from a 1980s-era study that probably uses less data. It is unclear whether recharge from the 1980s study guided calibration (i.e., Figure 4.2)?

Addressed in teleconference, no action required

7. (RHP) Calibration doesn't consider conceptual model adjustments, or zonation of many of the parameters. Although, it is infeasible to adjust all model parameters, it is unclear how sensitive the model performance is to non-adjusted parameter values. Is there a sense of the level of uncertainty associated with spatial distributions of all parameters in addition to their values, especially for groundwater aquifer system?

The Peer review panel included or considered this comment in subsequent documentation and refined comment(s). The SFWMD has prepared and distributed a formal response which includes the actions necessary to address this comment.

8. (RHP) I think the calibration approach outlined on Figure 4.2 is reasonable, but pretty confusing. More explanation seems warranted. For example, I'm not sure what the following really means:
   c. ET error <15%. Does this mean model calculated error for ET component?
   d. ET too high/low. Are model simulated values compared against AET estimates? At points, lakes, ground surface, other?
   e. Close to correct SAS recharge, or SAS recharge too high or too low. Does this mean recharge is compared against other model-calculated values? How accurate are the former recharge values if so?
   f. Decrease or Increase ICU K. What were the ranges of K value adjustments (more/less than sensitivity analysis?)?
   g. Is Kinf the same as unsaturated zone Kv (or Keff in Table A-1)?

   It is standard/typical to describe the range over which calibration parameters were adjusted from initial values (presumably based on prior information). Were the ranges reasonable?

Requires update to AFET Documentation Report, see Memorandum Summary at the end of this document.

9. (DPL) The last column in Table 2.2 on page 2-2 is not clear to me. In fact much of the table isn't as clear as it should be I think. Can you clarify this?

Table 2-2 Relationship of statistical and additional criteria applied to the KBMOS MIKE SHE/MIKE 11 model during the calibration and verification periods to issues in the Kissimmee Basin.

KB Issues	Primary Location	Primary Model Capabilities Needed	Temporal Scale (A)	Level First Simulated	Variable	Calibration Criteria Column
Seasonal Flooding	KB	Operations	Daily	Screening	OCSE*	1(B)
	KB	Water surface profiles	Daily	Screening	Stage	
	KB	Channel and structure hydraulics	Daily	Formulation	Flow	
	KB	Lake levels	Daily	Screening	Stage	
Hurricane or Storm Event Flooding	KB	Operations	Hourly	Formulation	Stage/Flow	2(C)
	KB	Flow Routing	Hourly	Formulation	Flow	
	KB	Lake levels	Hourly	Formulation	Stage	
	KB	Channel and structure hydraulics	Hourly	Formulation	Flow	

Addressed in teleconference, no action required

10. (DPL) I didn't find any discussion on the relationship between model accuracy and spatial and temporal resolution. Page 2-4. What determined the selected resolutions for different parts of the model?

2.2 Model Domain and Grid

The model domain and grid have been defined based on the defined extent of the Kissimmee Basin. This was defined during Phase I of the KBMOS project. This boundary is defined primarily by topography. The domain of the KBMOS model encompasses parts of five South Florida Counties (Orange, Polk, Highlands, Osceola, and Okeechobee) and three water management districts (South Florida Water Management District, South West Florida Water Management District, and St Johns River Water Management District). The domain covers an area approximately 1.65 million acres. This area is divided into square grids that are 1,000 × 1,000 ft or approximately 23 acres in area. The entire domain therefore consists of nearly 72,000 square grid cells. An alternate grid has also been developed that consists of approximately 7,520 cells with a cell size equal to 3,000 × 3,000 ft (~205 acre cells) and will be used for the MIKE SHE model that includes the Surficial Aquifer System (SAS), Intermediate Aquifer System (IAS), and Upper Floridan Aquifer System (UFAS), this coarser grid model is also referred as the Regional KB Model.

Addressed in teleconference, no action required

11. (DPL) Wouldn't extreme climatic conditions be an unusually dry dry or unusually wet wet seasons, or unusually dry dry and wet seasons, etc? P. 2-6.

A seasonal analysis of the fully distributed daily precipitation dataset for the Kissimmee basin is shown in Figure 2.4. For the KBMOS model, the wet and dry seasons have been defined as the period from May 1 to October 31 and November 1 to April 30, respectively. This analysis was used identify extreme climatic conditions (*i.e.,* a wet dry season, a dry wet season, *etc.*) during the calibration and verification periods. Extreme periods include the 1996 wet season, the 1996 to 1997 dry season, the 1997 wet season, the 1997 to 1998 dry season, the 2001 to 2002 dry season, the 2002 to 2003 dry season, the 2003 wet season, and the 2004 wet season. Upper Floridan Aquifer potentiometric surface, maximum overland depths, and average overland depths are presented for these extreme periods to evaluate the ability of the model to simulate conditions during these periods.

Addressed in teleconference, no action required

12. (DPL) In Figure 2.21, where are the flood codes defined? p. 2-44.

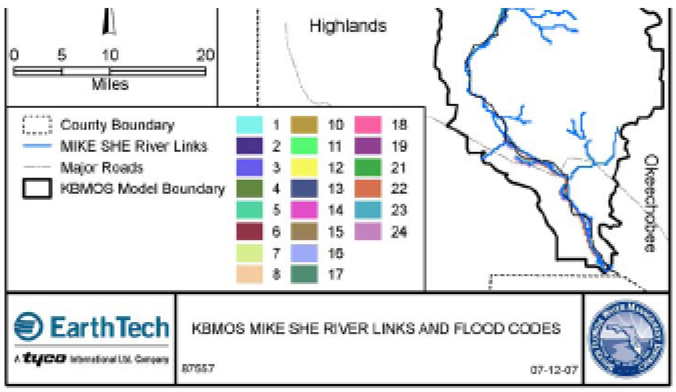

Figure 2.21 MIKE SHE River Links and Floodcodes used in the KBMOS model.

Requires update to AFET Documentation Report, see Memorandum Summary at the end of this document.

13. (DPL) P. 2-45. Are Manning's M values for overland flow constant with depth?

- Manning's *M* values – Manning's *M* values lump the friction effects due to bedding and vegetation and directly impact the velocity of overland flow. These values are constant over the entire flow regime and spatially distributed based on the specified land-use types. Manning's *M* values are equal to the inverse of Manning's *n* values and are equivalent to the Stickler roughness coefficient.

Addressed in teleconference, no action required

14. (DPL) During model calibration, were any parameter values outside the 'usual' or 'expected' range of those values? Could multiple sets of parameter values result in the same model output? P. 2-73,74

Land-use based drainage parameters were adjusted during model calibration to improve model calibration. Drainage levels were modified from land-use based values during calibration in the Lake Conlin (Watershed 11) and Lake Gentry (Watershed 16) to improve model performance in Alligator Lake. Drainage was added to this area after reviewing aerial photographs which suggested wetland features in these watersheds drained to Alligator Lake through Lake Gentry.

Addressed in teleconference, no action required

15. (DPL) Did the uncertain groundwater pumping and withdrawals cause a problem? Was the model sensitive to this? p. 2-78

Potable groundwater well data was developed using data provided by District staff for the ECFT MODFLOW (A.K.A. KBECF model) model currently being developed by the District. Screen depths were developed for UFA wells in the KBMOS model area from hydrostratigraphy data used in the model and have been constrained to the UFA. The ECFT pumping well dataset had withdrawal data from 1995 through 1999. Based on discussions with District staff monthly withdrawal rates for 1995 were used in 1994 and pumping rates for 1999 were used in 2001 through 2004. The locations of the potable water supply groundwater wells in the KBMOS model domain are shown Figure 2.44. Potable groundwater wells are not included in the 1,000 × 1,000 SAS model.

Addressed in teleconference, no action required

16. (DPL) What are the units used in Table 3.8 on pages 3-12?

Table 3.8 Average, minimum and maximum water budget changes resulting from parameter perturbations.

Simulation	Simulation Number	Average	Minimum	Maximum
Base Sensitivity Model		0.0	0.0	0.0
Surficial AQ (Decrease)	1	0.1	-0.5	0.5
Surficial AQ (Increase)	2	-0.1	-0.5	0.0
Confining Unit (Decrease)	3	0.0	-0.3	0.3
Confining Unit (Increase)	4	0.1	-0.1	0.5
Floridan AQ (Decrease)	5	-0.5	-4.6	0.5
Floridan AQ (Increase)	6	0.5	-0.5	4.0
Kc (Decrease)	7	-1.6	-57.2	23.0
Kc (Increase)	8	1.8	-19.8	52.4
OL Manning (Decrease)	9	0.0	-0.5	1.0
OL Manning (Increase)	10	0.0	-0.7	0.5
River Manning (Decrease)	11	0.0	-0.5	0.6
River Manning (Increase)	12	0.0	-0.6	0.4
UZ inf. (Decrease)	13	0.0	-0.1	0.1
UZ inf. (Increase)	14	0.0	-0.5	0.4
Soil Moisture Content$_{sat}$ (Decrease)	15	0.4	-5.5	4.4
Soil Moisture Content$_{sat}$ (Increase)	16	-0.2	-2.4	3.5
Soil Moisture Content$_{Fc}$ (Decrease)	17	-0.3	-1.0	0.1
Soil Moisture Content$_{Fc}$ (Increase)	18	0.5	-0.6	2.1
Drain constant (Decrease)	19	-0.2	-5.5	3.1
Drain constant (Increase)	20	0.2	-1.3	4.4

Addressed in teleconference, no action required

17. (DPL) p. 4-1. My congratulations on accomplishing this very challenging job! 17. Are the models now considered ready for operating policy simulations?

> Calibration of the model was difficult because of the complexity of the Kissimmee Basin and the size of the KBMOS model. The calibrated model meets the criteria defined for the project and adequately representing hydrologic processes in the Kissimmee Basin and can be used to evaluate alternative operational criteria and restoration of the Kissimmee River. Surface water, groundwater, overland water, water budget, and surface water-groundwater interaction results are presented below.

Addressed in teleconference, no action required

18. (DPL) Why were the ground water boundary conditions based on a 3000x3000 ft. grid when the model is based on a 1000x1000 ft. grid? p. 4-1.

> The results of the sensitivity analysis and the KBMOS AFET Acceptance Test Plan (Earth Tech, 2006b) were used to develop a diagram that details the model calibration process that was used during calibration of the KBMOS MIKE SHE/MIKE 11 model. A regional model (RM) that uses a 3,000 × 3,000 ft grid spacing and includes the ICU and the UFA will be used to develop dynamic groundwater boundary conditions of the UFA for a higher resolution MIKE SHE/MIKE 11 model of the SAS. The higher resolution SAS model (SM) uses a 1,000 × 1,000 ft grid spacing and be used to optimize structure operations in the Kissimmee Basin in a later phase of the project.

Addressed in teleconference, no action required

19. (DPL) Tables 4.4, 4.5 (p. 4-39+)and similar tables later in the text show both unsatisfactory as well as satisfactory calibration results for the same modeled area. Am I reading this correctly? I'm confused.

Table 4-4 Stage statistics for the calibration period. Shading is used to indicate locations that do not meet specified criteria.

MODEL AREA	Station	RMSE	R(correlation)	RMSE	R(correlation)	RMSE	R(correlation)
Stages in Upper Basin Lake Management Units							
LMU K-H-C	S65H	H	H	0.10	1.00		
LMU K-H-C	S61T	H	H	0.72	0.97		
LMU K-H-C	S63AT	M	M	0.84	0.96		
LMU Toho	S61H	H	H	0.37	0.99		
LMU Toho	S59T	H	H	0.83	0.91		
LMU Etoho	S59H	H	H	0.29	0.96		
LMU Etoho	S62T	M	M	0.19	0.98		
LMU Hart	S62H	M	M	0.27	0.88		
LMU Hart	S57T	M	M	0.29	0.89		
LMU Myrtle	S57H	M	M	0.18	0.98		
LMU Myrtle	S58T	M	M	0.18	0.99		
LMU Alligator	S58H	M	M	0.36	0.91		
LMU Alligator	S60H	M	M	0.38	0.89		
LMU Gentry	S60T	M	M	0.27	0.92		
LMU Gentry	S63H	M	M	0.22	0.94		
LMU s63a	S63T	M	M	0.16	0.94		
LMU s63a	S63AH	M	M	0.12	0.97		
Stages in Upper Basin's unmanaged watersheds							
ws_UpperReedy	REEDYLOU	M	M	1.80	0.76		
Stages in Lower Basin Lake Management Units							
Pool A	S65T	H	H	0.22	0.98		
Pool A	S65AH	H	H	0.15	0.99		
Pool BC	S65AT	H	H	1.57	0.86		
Pool BC	PC52	M	M	2.88	0.80		
Pool BC	PC45	M	M	3.30	0.68		
Pool BC	PC33	H	H	0.45	0.90		
Pool BC	PC21	M	M	1.29	0.86		
Pool BC	S65CH	H	H	0.11	1.00		
Pool D	S65CT	H	H	0.12	0.88		
Pool D	S65DH	H	H	0.11	0.94		
Pool E	S65DT	L	H		0.79		
Pool E	S65EH	H	H	0.19	0.82		
Stages in Lower Basin's unmanaged watersheds							
D_Chandler	CYPRS	H	H	0.59	0.81		
D_Chandler	CHAND1	H	H	0.62	0.86		
Lake O	S65ET	H	H	0.03	1.00		

"H" not meeting criteria
Does meet criteria
"M" not meeting criteria
NOTE: Calibration points with Low (L) utility have been omitted

Addressed in teleconference, no action required

20. (DPL) What calibration/verification procedures are planned if any to improve model results for storms?

Addressed in teleconference, no action required

21. (RHP) The section 5.1.3 and 5.1.4 seem overly qualitative – can't they be compared against observed data, aerial photos etc.? What is the benefit other than they look reasonable?

Addressed in teleconference, no action required

****

Objective 3.A: Model formulation and calibration:

Possible critical defects

Calibration and verification issues:

1. (DAC) Since a primary objective of KBMOS is to look at the impact of different gate operating rules, it would seem important to calibrate the model using historical gate operating rules such that the flows and stages are adequately reproduced. It does not seem that this was done. Was it?

The Peer review panel included or considered this comment in subsequent documentation and refined comment(s). The SFWMD has prepared and distributed a formal response which includes the actions necessary to address this comment.

2. (RHP) The calibrated model appears to reproduce much of the observed calibration and verification period surface water stage and discharge well, though at some locations, some time periods show greater deviation. Is it possible that uncertainty in rainfall and non-distributed PET time series could be the cause of this? How much of the calibrated model response might be due to uncertainty in these inputs? Knowing climate input is subject to some level of uncertainty, wouldn't final parameter values be different because only these were adjusted during calibration to improve performance against observation data? How much of the discrepancy between simulated/observed response for 2004 hurricane verification is due to error in rainfall/PET input?

The Peer review panel included or considered this comment in subsequent documentation and refined comment(s). The SFWMD has prepared and distributed a formal response which includes the actions necessary to address this comment.

3. (RHP) Appendix A – Table A.1. Is the equation for harmonic mean Ke on page A-4 right? It seems like the effective K values based on soil horizons within root zone are too low in Table A.1 (i.e., Ke is significantly lower than any of the individual horizon K values). If the Ke values were in fact set too low, would other parameter values have been adjusted outside of reasonable ranges to improve model performance against observation data? It seems like SC, FC and WP values in same table are OK. Is Ke the same parameter as Kinf in Table 4.2?

Addressed in teleconference, no action required

Model development issues:

4. (RHP) There are two possible issues with the use of the saturated zone drainage option to simulate surface water drainage to M11 channels. The time flow travels in the saturated zone is much slower than surface flow. Groundwater discharge will be delayed by the rate of infiltration to the saturated zone and saturated zone dynamics – which don't seem consistent with dynamics of overland flow to channels. In addition, it

doesn't seem like overland surface water elevations are known well – so this approach can't really be validated. Another issue is that the drainage algorithm appears to remove groundwater from a cell based on the simulated groundwater level above specified cell's drainage level at a specified drainage rate (1/t). But it doesn't seem to account for the apparent large distances (i.e., > 5 to 10 miles) between drainage cells and M11 channels. Wouldn't this distance need to be considered in the drain constants? The maps are too small to clearly see drainage routes (Figures 2.40 to 2.42). While it seems possible to define decreasing time constants with distance to the intercepting M11 channel, time constants don't seem to reflect this (Figure 2.42). If this is true, what is the implication to the drainage response?

Addressed in teleconference, no action required

5. (DPL) If run times (page 2-5) are from 3 to 4 hours per year of simulation, will the screening model be able to effectively screen out inferior policy alternatives to save time?

> Run-times for the calibrated KBMOS model range from 2.31 hours per year for the three-layer regional KBMOS model to 3.79 hours per year for the one-layer, higher-resolution surficial aquifer KBMOS model. These run-times meet the run-time goals (4 to 5 hours per year) defined in the TDD (Earth Tech, 2006a).

Addressed in teleconference, no action required

6. (DPL) P. 2-38. If the control structures are operated differently than what is assumed in the model, might the model results be in error?

> For calibration purposes only, the control structures in the Pre- and Post-Phase 1 models are dynamically operated using the observed headwater stage data to open and close the gates and dynamically simulate structure discharges. The gates begin to open when the simulated headwater stage is at least 0.1 feet above the observed stage and begin to close when simulated stages are within 0.1 feet of the observed stage. During the course of the project the modeling team evaluated using reported gate opening data provided by the SFWMD. It was decided to not use the reported gate opening data because 1) it varied on a 15-minute frequency which is contrary to the modeling team's understanding that all structures except S63A are operated on a daily basis and 2) it was unclear how the 15-minute gate opening data was developed and the original observed gate opening information was not available.

Addressed in teleconference, no action required

<u>Suggested remedies</u> or <u>appropriate caveats</u>

1. (DPL) Wouldn't the fraction of rainfall on impervious areas reaching a water body vary depending on the amount of rainfall? P. 2.7 If so model predictions of surface runoff could be improved if the non-linear response of rainfall/runoff were defined.

> The paved area option allows runoff from urban areas to be represented in a conceptual sense and requires that a fractional portion of rainfall that is directly routed by urban surface water management features be defined.

Addressed in teleconference, no action required

Recommendations

Sensitivity and uncertainty issues:

1. (RHP) There are two types of sensitivity analyses typically performed with hydrologic models. One is as performed in this study, before calibration, to identify key parameters, and the other is performed after calibration as per ASTM standards like D5611 such as for a groundwater flow model.

   The second approach is used to examine sensitivity of calibration residuals and model conclusions to model inputs to assess adequacy of the model with respect to its objectives.

   Why were parameters adjusted only 10 to 20 percent of initial values? This seems to disregard the possible range that these parameters may vary across the model domain, or that parameter values can vary. For example the range of possible crop coefficient values is notably different than the range of possible hydraulic conductivity values for an aquifer. In the later case, the horizontal conductivities range orders of magnitude, while crop coefficients typically vary less than a factor. It would have been more informative to adjust parameters the same percent, but normalized by their possible ranges (either over model domain, or even published ranges). As reported, the purpose of the sensitivity analyses seems vague and can be misleading. For example, this approach suggests Kc, Mannings n and M and moisture saturation were the most sensitive parameters, though aquifer hydraulic conductivities, or vertical K values for soils might be the most sensitive parameters, even for surface flows.

   It would have been useful to assess sensitivity after initial calibration, to assess calibration residual sensitivity to model parameters. After model parameter adjustments were made in the calibrated model, it would have been nice to see if the model performance was most sensitive to the same parameters, over their possible ranges, as the pre-calibration sensitivity analysis.

   Finally, the sensitivity analyses could have been extended to evaluate 'process sensitivity' instead of just parameter sensitivity. For example, it would have been informative during and after calibration to develop a sense of what processes are most important at different calibration targets. In other words, is the stage or discharge response at for example S65C more sensitive to lateral inflows, upstream or downstream controls, baseflow, leakage, or ET?

The Peer review panel included or considered this comment in subsequent documentation and refined comment(s). The SFWMD has prepared and distributed a formal response which includes the actions necessary to address this comment.

Calibration and verification issues:

2. (RHP) Calibration statistics in terms of R and RMSE seem over-emphasized, rather than how well (accurately) the calibrated model generates valid performance measures eventually used to evaluate various alternatives using the AES. One suggestion might be to evaluate the calibrated model performance by comparing its AES score for the calibration period against actual observation data performance. This may help identify deficient areas of the calibrated model performance which could be improved, or the AES altered. For example, in E0-3 Kissimmee River Stage Hydrograph, simulated levels are compared to very specific elevations and depending on model calibration, valid PM values may be difficult to achieve. Therefore it seems essential to assess performance of the calibrated model in simulating valid PM values, seeing where issues may be, modifying the calibrated model and rechecking until the calibrated model meets appropriate level of accuracy for later AES steps.

The Peer review panel included or considered this comment in subsequent documentation and refined comment(s). The SFWMD has prepared and distributed a formal response which includes the actions necessary to address this comment.

Objective 3.B: Adequacy of base conditions simulations

Possible critical defects

Calibration and verification issues:

1. (DPL) Calibration results illustrated on pages 4-19 – 38, and the verifications are generally impressive. Perhaps the worst on is shown in Figure 5.80. Any reasons? p.5-26.

Addressed in teleconference, no action required

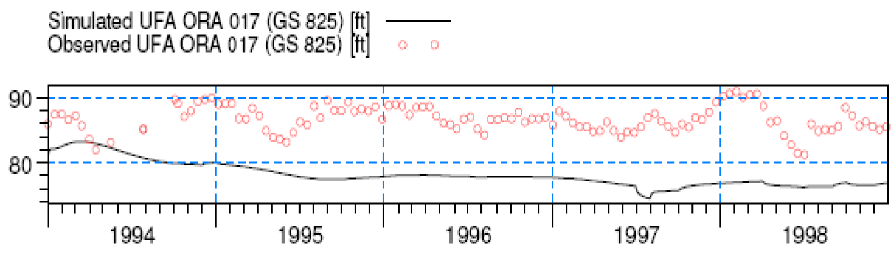

Figure 5.80 **Simulated and observed water level at UFA ORA 017/GS 825 during the verification period.**

2. (DPL) Did I miss seeing overland flow verifications (except hurricane data p. 5-45)?

Addressed in teleconference, no action required

3. (DPL) Figure 5.101 shows a poor fit. p. 5-45. Why?

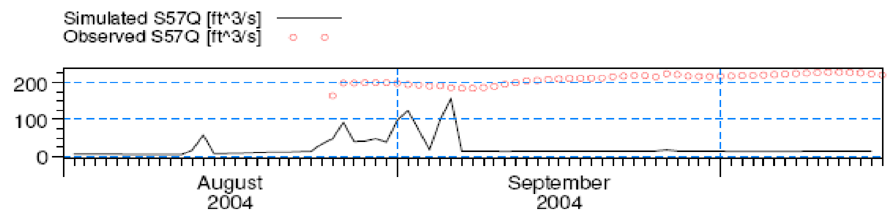

Figure 5.101 Simulated and observed discharge at S57 during the storm verification period.

Addressed in teleconference, no action required

4. (DPL) Figure 5.144 suggests to me there was some inflow to the system in late September 2004 that the model data base doesn't have. Is that true? P. 5-56

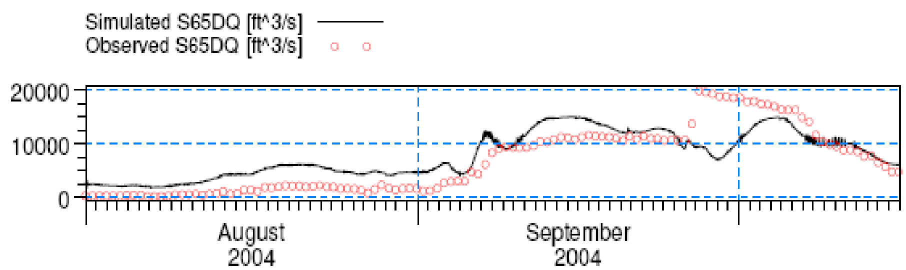

Figure 5.144 Simulated and observed discharge at S65D during the storm verification period.

Addressed in teleconference, no action required

5. (DPL) Figure 4.9. What causes these occasional mismatches? p. 4-16

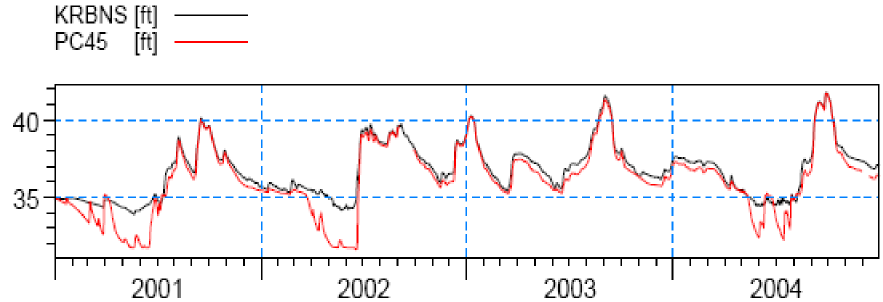

Figure 4.9 Observed stage at KRBNS and PC45 in the restored portion of the Kissimmee River during the calibration period.

Addressed in teleconference, no action required

6. (DPL) Are there datum problems in the following Figures?

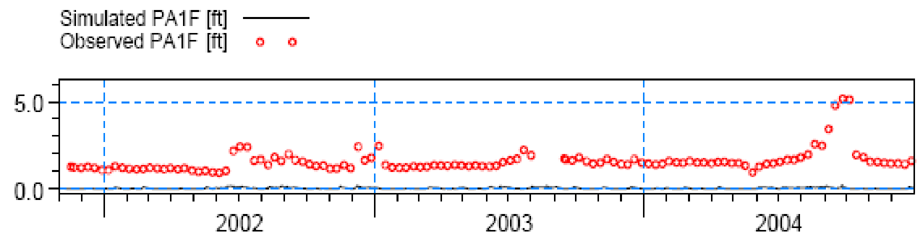

Figure 4.62 Simulated and observed water depth at PA1F during the calibration period.

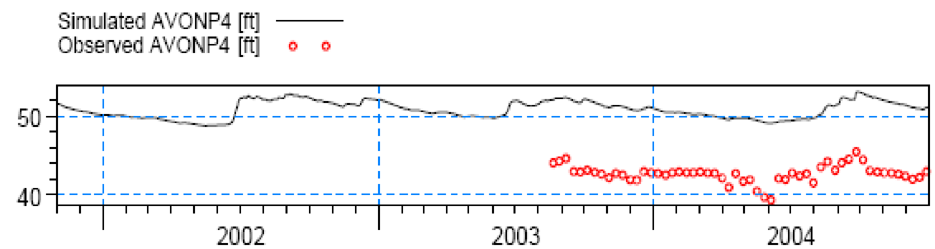

Figure 4.70 Simulated and observed stage at AVONP4 during the calibration period.

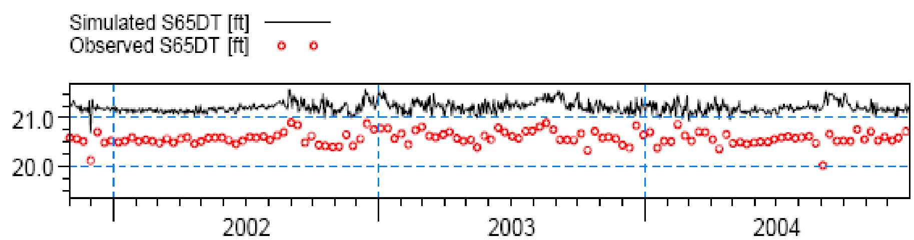

Figure 4.80 Simulated and observed tailwater stage at S65D during the calibration period.

Addressed in teleconference, no action required

7. (DPL) The calibration for aquifer stages seems to be difficult. Why? Might it be related to model or observed spatial and/or temporal resolutions? P. 4-43-49.

Addressed in teleconference, no action required

8. (DPL) Figures 5-106 and 107 suggests to me that perhaps the model doesn't take into account the dampening of variations in stages as it should during hurricanes. Is there another reason?

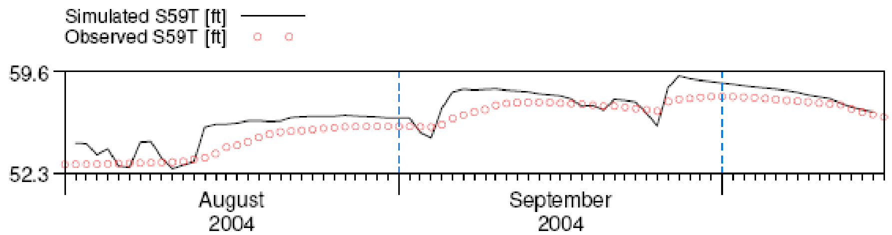

Figure 5.106 Simulated and observed tailwater stage at S59 during the storm verification period.

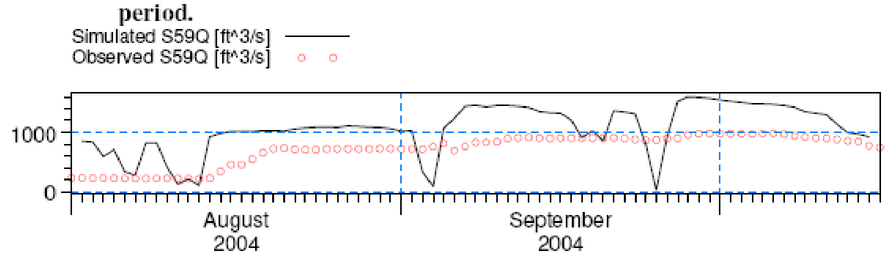

Figure 5.107 Simulated and observed discharge at S59 during the storm verification period.

Addressed in teleconference, no action required

<u>Suggested remedies</u> or <u>appropriate caveats</u>

None

<u>Recommendations</u>

None

Memorandum Summary:

The following summarizes the comments that require modification/update to the project documentation.

Clarifications:

3. (DAC) I might have missed it, but the acceptance criteria for R and RMSE should be stated in the report near the tables with performance metrics (e.g. Table 4-4). It is a bit curious that a RMSE of 2.88 ft would be acceptable for PC52 in Pool BC (Table 4-4). These are others like this, for example a RMSE of 2.97 ft in the SAS water table at PINEISL is acceptable. What are the acceptance criteria?

Action: The acceptance criteria were provided in earlier tables. Tables should be modified to better link the acceptance criteria to model performance reported in the tables.

8. (RHP) I think the calibration approach outlined on Figure 4.2 is reasonable, but pretty confusing. More explanation seems warranted. For example, I'm not sure what the following really means:

a. ET error <15%. Does this mean model calculated error for ET component?
b. ET too high/low. Are model simulated values compared against AET estimates? At points, lakes, ground surface, other?
c. Close to correct SAS recharge, or SAS recharge too high or too low. Does this mean recharge is compared against other model-calculated values? How accurate are the former recharge values if so?
d. Decrease or Increase ICU K. What were the ranges of K value adjustments (more/less than sensitivity analysis?)?
e. Is Kinf the same as unsaturated zone Kv (or Keff in Table A-1)?

It is standard/typical to describe the range over which calibration parameters were adjusted from initial values (presumably based on prior information). Were the ranges reasonable?

Action: Text needs to be added to address comments.

12. (DPL) In Figure 2.21, where are the flood codes defined? p. 2-44.

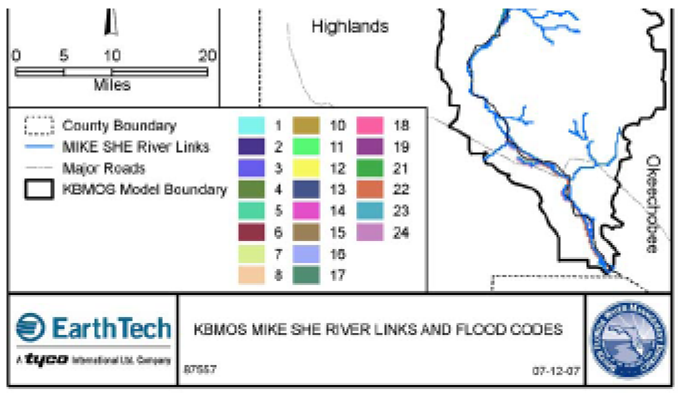

Figure 2.21 MIKE SHE River Links and Floodcodes used in the KBMOS model.

Action: clarification added to the text that the referenced flood codes identify geography, not a numerical model input parameter.

M E M O R A N D U M

TO: Pete Loucks, Cornell University, Chair
David Chin, Consulting Engineer
Robert Prucha, Integrated Hydro Systems, LLC

FROM: Ken Konyha, Section Lead, Hydrologic and Environmental Systems Modeling
Christine Carlson, Sr. Supervising Geographer, Kissimmee Division

DATE: July 7, 2008

SUBJECT: Action Required to address the Modeling Peer Review Panel Comments on OKISS Model Development and History Matching Report

The following memo summarizes the status of Peer Review Panel comments received on the OKISS Model and History Matching Report. The purpose of this memo is to identify comments that may require an update/modification to project documents or preparation of additional project documentation.

The status of each comment is provided following each item below. RESPONSES HAVE NOT BEEN INCLUDED. Only status and follow-up activities are identified. The summary at the end of the memo includes comments that require modification to the project documentation.

KBMOS Model Peer Review Panel Comments on
OKISS Model Development and History Matching Report (Dec 17, 2007)

Report Clarity and Editing

- (RHP) Many references are made to Pool B/C, Pool A, B, C, D, E, yet I couldn't find one map that shows where these are referenced exactly. It would be very helpful to have such a map. (A map was sent by the district – a similar map will be included in the final report).

- (RHP) Figure 2.1. Some clarification would be useful to reviewers not familiar with OASIS. OKISS Report text will be clarified.

    - Are the red triangles versus yellow trapezoid at 169 all Nodes?

    - Which of the nodes shown are Junction, Reservoir, or Demand nodes (page 1-2)?

    - Are the arcs simply the lines between red-triangles?

- (RHP) Given the importance of this Screening Tool in assessing alternatives as per the Alternative Plan Selection Process (Figure 2.1 in the Base Conditions Summary Report), why aren't Performance Measures that can be assessed with this Tool, or those Nodes and Arcs most important to PM scores discussed, or even identified up front in this report? It seems like this should be emphasized. OKISS Report text will be modified

- (DAC) Based on the results in Table 7.1, the report concludes that the set of lateral inflows is reproducing the exchange between MIKE SHE and MIKE 11 well. Elaboration on the rationale for this statement would be useful. OKISS Report text will be modified.

- (RHP) Why aren't quantitative values provided for differences between the AFET and OKISS Tool results at specific locations (other than Table 7.1)? It seems warranted, given the importance of the Screening Tool to the Alternative Selection Process, and to be consistent with the AFET performance assessment. This comparison was provided in the Evaluations of Base Conditions Report (3rd package) to reviewers..

- (RHP) What are the simulated differences in flows, or stages for different time periods (other than calibration period results in Table 7.1), or events (i.e., wet-periods, dry-periods etc) between the AFET and OKISS models? This might be useful to emphasize time-periods, or locations where the Screening Tool performs well, and areas where it doesn't perform so well? District indicated this effort was more qualitative. More quantitative comparison included in validation step (3rd package).

- (DPL) Is it really true that OASIS always finds a system operation that maximizes the number of points, where weights given to operating goals are the points per unit of those goals? I think what happens is that OASIS maximizes the sum of products of weight times goal value. For example consider the following LP model showing the tradeoff between goal A and goal B, both of which are to be maximized.

$$\text{maximize } 1A + 5B$$
$$A = 100 - 10B$$

  Even though the weight on B is five times that on A, the optimal solution is A = 100, and B = 0. This should be easy to clarify in the document.
  Text will be clarified.

- (DPL) While it is true that OASIS does not optimize operating criteria, it does select one of many possible feasible solutions that define flows and storage volumes. It is not clear in the documentation just how this relates to the solution of AFET, let alone how such solutions would be identified and implemented in practice without some change in the control structures (e.g., gate openings, etc.). If the weights are to be set so that the OKISS produces a result similar to a previous solution of AFET, how will OKISS be used for screening of alternatives prior to running AFET? This comment will be addressed later in this document.

- (DPL) The index notation used for the variables in Equation 2 is a bit strange. Why not just use the subscript t to represent the beginning of the day t, so that $Z_{L,t+1}$ would be the elevation at lake L at the end of day t, etc. Consider revising text. Text will be revised.

- (DPL) Section 4.3.1.1 page 4-3: I believe the units of '*g*' are ft/s^2 To be corrected.

- (DPL) Section 5.1, pages 5-1,2: I don't believe it is necessary to have all components of the objective function in the same (i.e., volume) units in order to assign priorities to each operating objective. Text to be clarified.

Model Formulation

- (DPL) Section 1.2. I note that nodes do not include lakes. Are lakes included under reservoir nodes? I distinguish a lake from a reservoir based on whether or not the release or discharge of water is natural based on stage or controlled based on gate openings, for example. Text to be clarified

- (DPL) It appears to me that some arcs are modeled as storage nodes, and others not. Is it realistic in the Kissimmee Basin to assume non-storage arcs that transfer flow volumes from one node to another are empty at the end of each time period? Text to be clarified

- (DAC) The sections describing the derivation of HW-TW-Flow relationships are confusing. Perhaps the origin of this confusion is the different definitions of HW and TW as it relates to reservoirs and structures. For example, consider the following paragraph in Section 3.1.2.1.1:

  > At the same time, flow is independent of the HW in restored Pool B/C, within the historic range of HW, TW and flows. In fact, the HW would have to exceed the maximum for structural stability before the pool would extend to the upstream structure (S65A), causing HW to affect flow. Instead, flow is simply a function of TW at the upstream structure.

  A diagram would be very helpful in understanding this paragraph. OKISS Report text will be modified.

- (DAC) The derivation of HW-TW-Flow relationship in Section 3.1.2.1.2 is also confusing and would also benefit from supporting diagrams. For example, it is my (perhaps erroneous) understanding that HW and TW in this section is referring to reservoirs, then why is the flow equal to 100 cfs when HW = TW? OKISS Report text will be modified.

- (DPL) Section 3.1.2.1.2 on page 3.3: Functions f_1 and f_2 are segmented for LP. I'm always curious when making non-linear constraints piecewise linear whether they are the right shape for the optimization being performed, or are integer 0,1 variables needed to make sure the value of the piecewise linear function closely approximates the value of the original non-linear function. The shape and order of the piecewise linearization have a strong influence on the run times in OKISS. So other than making sure that the linearization process respects the shape of the original function, it is also required The OKISS Report text will be modified to clarify this.

- (DAC) In Equation 7, shouldn't the headwater be H and the tailwater be h? District agrees.

- (DPL) Section 4.1, page 4-1: Isn't the most accurate approach to computing flows from functions of heads in lakes is to take the average of the function values for flows based on the two heads rather than solving the function using the average of the two heads? For non-linear functions, there will be a difference. Text to be clarified.

- (DAC) In using the culvert flow equations on page 4-6: (a) if HW < TOP, how is the flow through the culvert calculated for a given HW and TW? (i.e., how is y calculated?); and (b) for orifice flow, why does the flowrate depend on the tailwater elevation? Text to be clarified.

- (DPL) Section 5.1, pages 5-1,2: Even though the weights are set based on operating criteria, there are still many ways water can be allocated over the region. That is why LP is used, namely to avoid having to write if-then-else rules that would dictate just how water is best allocated given all possible initial conditions and estimated inflows. Again, I'm confused as to how this will be used for screening. I can understand how it could be used to screen various operating policies, but if the physical infrastructure is fixed, what in reality would allow the flow variables to take on alternative values even within the constraints imposed – as they can in the LP model? Text to be clarified, as it was during our discussions.

- (RHP) Page 6-14. Why can't the Screening Tool permit reverse flows at S-60 and S-58 structures, but can at C-36 and C-37 Canal
The OKISS tool could have been set up to allow reverse flows at S-60 and S-58 but the significance of the reverse flow situations were considered too minor to model. This constraint can be easily removed if an alternative plan requires to.

- (RHP) What determines the 'maximum allowable flow' at structures (i.e., Section 6)?

  Maximum allowable flow is defined the gate flow equations and are functions of gate opening, headwater and tailwater.

- (RHP) For the base condition simulation, what information is used to drive the OKISS model simulation? It wasn't clear (page 6-13) why stages, in addition to lateral inflows, need to be specified at each time-step, instead of just at the start? For evaluation of alternative operating criteria, will OKISS only use lateral inflows from AFET to calculate stage and flows at nodes/arcs (and it calculates daily stages, flows and storage)? The history matching used two methodologies, one of them used a fixed set of stages as a target. However, this methodology was used only for the purpose of that particular history matching exercise. For the Base Conditions runs only the lateral inflows need to be specified for each time step. he OKISS Report text will be modified to clarify this

- (DPL) Section 5.1, pages 5-1,2: Is there any situation in which the ordering of priorities among operating objectives are dependent on the values of the operating objectives? For example if the priority of a municipality is greater than that of a farm, suppose the municipality got 80 % of what its ideal target was, it could be reasonable that the farm would have priority over the next drop of water available. Can such conditions be incorporated into OKISS? Yes, OASIS can assign priorities using conditional statements. he OKISS Report text will be modified to clarify this.

- (DPL) Section 5.1, pages 5-1,2: The number of time steps (days) in each OASIS run is not clear to me. Is indeed each day a separate optimization run, so that there is no consideration of the future in each run? Is this the same assumption when running AFET, i.e., the operating rules only considers what will happen in that current time period? OASIS does not use information from a future time step. The optimization routine used in OASIS can be seen as a means to solve a complex set of simultaneous equations.

- (DPL) Section 5.1, pages 5-1,2: Is it really OKISS that defines weights dependent on actual values of objective function components, or are the weights dependent on operating criteria, and therefore set by people, as previously specified? Both, some weights are defined during the model development stages, these weights are related to the physical constraints of the lakes and canal systems. .Other weights are related to the operating criteria and can be modified during the development of an alternative plan.

- (DPL) Section 5.3.5 page 5-7: I'm curious why small negative weights assigned to (I'm assuming not "added to") storage in the LKB reduces run times. No need to address this.

Model Calibration

- (DAC) Section 6.1 demonstrates that OKISS can sometimes be reasonably successful in modeling the regulation schedule for reservoirs in the KB; however, there can also be significant differences between the observed and predicted stages and flow rates in the channels and pools. These differences should be quantifies in terms of an error, since they reflect the difference between what is predicted by the model and what would actually occur in practice for given operating rules. Quantification of this difference would serve as a benchmark for determining whether differing operating schedules would lead to significantly different outcomes in reality.
  District states that errors associated with deviations from operating rules are not significant in differentiating between alternatives. This explanation with justification will be added to the OKISS Report.

- (DAC) The significance of the discrepancies between modeled arcflows and observed flows should be assessed relative to using the modeled arcflows to estimate the performance measures.
  The Base Conditions report has a section comparing model flows.

- (RHP) It seems like we can't really assess whether the OKISS Tool has been appropriately formulated and calibrated to evaluate existing and proposed Kissimmee Basin structure operating criteria (i.e., Objective 3A) until results of the Validation of Model Tools by comparing PME Tool Results (page 1-4) has been completed. Discrepancies between simulated stage and flow by the AFET and OKISS models (Section 6.1 and 6.2) may qualitatively appear acceptable, but may produce unacceptable differences in PME scores. This would depend on the qualitative calibration, or history matching presented in this report. The validation of model tools was included in the Evaluation of Base Conditions Report (3rd package)

- (RHP) What parameters are adjusted in the 'tuning' of OKISS to AFETS calibration model results?

Model results comparisons:

- (DAC) Section 6.2 compares the OKISS predictions with the AFET predictions for the condition where the headwaters and tail waters at each structure (in OKISS) were set to the end-of-day elevations calculated using the AFET model. The resulting flows generated by the two models were then compared.

  It would be useful to estimate the magnitude of the errors introduced into the OKISS model by using the structure equations at daily time steps to estimate daily-averaged flows.

  It seems that significant flow discrepancies between the OKISS and AFET predictions can be produced even when there is exact agreement in the heads at the structures. It would be useful to quantify these flow discrepancies in terms of their effects on performance measures.

  It is noted in the report that the flow in both the AFET and OKISS models occasionally exceeds the maximum flow for structural stability (p.6-31). The report also indicates that such flows are observed. It seems as if the maximum is not a realistic maximum.
  This issue was examined visually by modifying the final calibration run so that historic gate operations are replaced by Operating Rules. Two distinct types of differences were noted: minor variances in operations caused by operator discretion and major variances caused by recorded Operational Deviations. The minor variances do not appear significant enough to impact PM. [This was presented at the Peer Review Workshop. The results of Run 100 will be shown in the AFET Model Calibration Report]

- (DAC) Verification of OKISS consisted of setting the HW stages at all structures to the AFET calibration values and comparing the OKISS and AFET flows through the structures. This is a conditional verification that does not give a true measure of the errors that result when OKISS is used as a screening tool. For a true measure of the errors (i.e. a true validation) associated with using OKISS instead of AFET, structure operating rules in AFET would need to be the same in both AFET and OKISS and then stages, flows, and performance measures compared. (I recognize the difficulty in comparing performance measures over a short period of record) This is done in the revised Base Conditions report.

  The report concludes that the screening tool (OKISS) was able to follow the Hydraulic Model (AFET) stages closely within its calibration period. This is not surprising given the fact that the HW stages at the structure locations were input as being the same in both models.

- (DAC) Table 7.1 shows the mass balance errors in AFET for various lakes in the KB; these errors range from 0% - 3.7%. Column 4 of this table indicates the difference in net outflow through structure; should this be the net outflow from the lake? (It would seem that mass balance is always preserved at the structures) Was the difference in lake stage between the beginning and ending of the validation take into account in calculating the net outflow? It would seem reasonable to me to normalize the mass imbalances by the lake volumes; this would give a better measure of the impact of the imbalances on the lake stages. The District agreed to revise the report to make the above changes in the OKISS Report.

  The differences in net outflow are attributed to using the KUB SAE/LKB equations, however, to support this assertion, it would need to be demonstrated that the errors introduced by applying the structure equations at daily time steps (OKISS) versus 30-minute time steps (AFET) does not contribute significantly to mass-balance errors within the lakes. Covered in previous comment.

- (DAC) The report notes that a limitation in the current OKISS model is that it does not restrict gate openings (and associated flows) to keep hydraulic jumps on the apron of the structures. However, since such flows would never be (intentionally) allowed in reality and at least some consideration should be given to flagging when/if such flows occurs in the screening scenarios. Issue has been adequately addressed by the District.

- (DAC) Appendix F addresses the effects of lateral inflows on OKISS results by quantifying the impacts of ±10% variations in lateral inflows on OKISS-predicted stages and flows. In support of this analysis, it should be demonstrated that: (1) uncertainties in lateral inflows are in fact within the ±10% range; and (2) the impact of these uncertainties on the performance measures as predicted by OKISS is relatively small . (This problem arises because lateral inflows are assumed to be constant and independent of the operating rules in the OKISS model). The AFET model estimates Future Base lateral inflows for the Upper Kissimmee Basin at 15.4 inches per year compared to Current Base lateral inflows of 12.9 inches per year – an increase that is similar to the 10% range assumed. The impact on PM will be assessed at the end of the Alternative Screening process when all Alternatives are reranked using Current Base lateral inflows.

- (RHP) Page 6-14. If the Screening Tool will not permit reverse flows at S-60 and S-58 structures, what are the steps if Base Conditions Validation shows not simulating negative flows impacts Performance Measures significantly? This is considered to be a low risk scenario, based on the small volume of reverse flows in the historic record. Even if this scenario significantly affects the PMs, the impact is small if the ranking of the Operational Alternatives are unaffected. If they are impacted, additional AFET model scenarios may be required.

- (RHP) Even if validation of base conditions shows acceptable comparisons between AFET and OKISS PME scores, this doesn't seem to guarantee that the scores will remain similar once operating criteria are changed in OKISS. Will there be any additional steps taken to guarantee that PME scores will remain comparable between the two Tools? No additional steps are anticipated as long as the relative rankings of the Operational Alternatives remain the same for both models. If they are impacted, additional AFET model scenarios may be required.

- (RHP) The plan for no analysis of uncertainty related to the OKISS model is based on the assumption that most of the error comes from the AFET calculations (page 1-4). Yet, in Section 6.2, comparison of simulated flows between the OKISS and AFET Tools in several cases doesn't seem that close, at least qualitatively. This suggests that, the OKISS Tool adds to the errors of the AFET simulations for the calibration time-period, at least at some locations (i.e., Figure 6.18, Figure 6.23, Figure 6.24, Figure 6.26, Figure 6.27 etc.) even if 7-day moving averages are considered. Is this true? Except for gate operations, OKISS is designed to echo AFET. OKISS model simplifications of hydrology and hydraulics add to model error. However, it is unclear which model OKISS or AFET adds greater operational error since they have radically different methods of implementing Operating Rules. At this time, it is unclear which type of model errors has greater impact on PMs.

M E M O R A N D U M

TO: Pete Loucks, Cornell University, Chair
David Chin, Consulting Engineer
Robert Prucha, Integrated Hydro Systems, LLC

FROM: Ken Konyha, Section Lead, Hydrologic and Environmental Systems Modeling
Christine Carlson, Sr. Supervising Geographer, Kissimmee Division

DATE: August 14, 2008

SUBJECT: Action required to address the Modeling Peer Review Panel Comments on Base Conditions Report

The following memo summarizes the status of Peer Review Panel comments received on the Base Conditions Report. The purpose of this memo is to identify comments that may require an update/modification to project documents or preparation of additional project documentation.

The status of each comment is provided following each item below. RESPONSES HAVE NOT BEEN INCLUDED. Only status and follow-up activities are identified. The summary at the end of the memo includes comments that require modification to the project documentation.

KBMOS Model Peer Review Panel Comments on
Evaluation of Base Conditions Report (May 28, 2008)

Overall Comments (Response welcome but not necessary.)

(DAC)

- The primary objectives of the Base Conditions Report are to: (1) develop the current and future base conditions models; and (2) validate the base conditions models by comparing the performance measures generated by each model (OKISS, MIKE11, and MIKE SHE/MIKE11) for the same base conditions.

- Development of the base conditions models appears to follow a logical path and there does not appear to be any major shortcomings in this process. It is somewhat suspicious that the annual PET could vary between 45 inches (in 1978) and 63 inches (in 1981) as shown in Figure 2-9. A possible reason for such an unusual climatic variation should be offered. The report mixes the terms “potential” ET and “reference” ET which are fundamentally different. Consistent terminology should be used.

- Minor editorial changes are needed on p.1-14 (“but” to “by”), p.2-17 (y axis on Figure 2-7), p3-2 (“be” to “but”), p.3-9 (duplicate paragraph), p3-9 (“reference” to “actual”), p3-27 (define M_{max}), p3-78 (“abotained” to “obtained”).

- On an editorial note, Appendix B does not seem to differentiate between current and future conditions. For current conditions, the table for S65E is mislabeled as S65.

(DPL) This report represents a massive amount of modeling work on a very complex problem. My compliments to all for the work that have been accomplished.

(DPL) Page 4-8, section 4.1.1.1.5: Again, I wonder why assigning a small weight (0.01) to storage volumes decreases model run times. And how did anyone discover this?

(DPL) Section 5: The comparisons of OKISS, AFT and AET output plotted in Figures 5-3 to 5-55 are to me amazingly similar. I wonder what this implies regarding model structural and algorithm uncertainty? And what does this imply regarding the accuracy of model results – could they all be equally wrong?

(RHP) Page 1-4 “Purpose and Scope” could more clearly define ‘base condition models’ referenced. It is unclear what “both” means in the sentence “The purpose of this effort is to demonstrate that they both provide comparable….” Does this mean current vs. future?

(RHP) The report is well organized and systematically presents development and comparisons of both the current and future base condition models. The true test of whether the differences between the 3 modeling tools for each evaluation component and performance measure (i.e. Appendix B) is small, and permits distinguishing between alternatives, will ultimately be determined by conducting similar comparisons for key alternatives.

(RHP) The graphical plots in Appendix B are good and easy to read.

Specific Comments
(These are requests for clarification and/or responses during our telephone calls.)

(DAC) There are several concerns with respect to validating the screening tool (OKISS). It is convenient to define the error in the screening tool as the relative difference between the performance measure (PM) predicted by OKISS and the same PM predicted by MIKE SHE/MIKE11 for the same scenario. Based on this definition of error, concerns with the validation process are as follows:

- Lake performance measure E is always zero, are the models capable of simulating non-zero values?
  [see response below]

- River performance measure H is always zero, are the models capable of simulating non-zero values?
  [see response below]

- For future conditions, lake performance measure G is zero at all locations except for L01, where the error is 400%. Can this PM be accurately estimated under future conditions?
  [Presented at 2nd peer review workshop] Most of the components used in the validation referred to events that were to be maintained during certain number of days. The evaluation protocol used to calculate the component values discarded events as soon as the stages went above or beyond a threshold value. Since the models have different computation engines and work under different time steps, they produce results that look similar when evaluating the result timeseries but they may reflect artificial oscillations due to the

operating criteria of the structures and the time step used in the operation logic. Those oscillations may trigger an event to be discarded by the evaluation protocol. To avoid these conditions the performance measures components and their targets were revised after the reviewed document was submitted to the panel. Less strict thresholds were defined to allow room for the fluctuations caused by structure operations. An activity was added to the future phases of the study to update this evaluation.

- Is there any particular reason why performance measure J has exactly the same value in OKISS and exactly the same value in MIKE SHE/MIKE11 for L01 (current conditions), L03 (future conditions), and L07 (future conditions)?
  There is no particular reason. Both values are correct.

- Analysis of the validation results obscures the magnitudes of the discrepancies in the performance measures, and does not take advantage of the opportunity to use the discrepancies as a measure of the uncertainty in the performance measures. To illustrate this point, the mean percentage errors and one standard deviation error bars in the lake and river performance measures are shown in Figures 1 and 2 respectively. These results are derived from Appendix B, and give a direct indication of the error ranges to be expected from each of the performance measures, where the errors are due only to model structure. For example, lake performance measure B can be expected to have a one-standard deviation error range between B-23%B and B+23%B. This could be particularly useful in discriminating between alternatives in the screening process. Such an uncertainty component is essential in the ranking and screening of alternatives. Failure to adequately account for this uncertainty could potentially be a critical flaw.

  KMBOS Performance Measures continue to evolve and it is sometimes difficult to separate true differences in model results and false differences caused by the current formulation of the PM. To minimize risk, the project team examines both the PM scores of all Alternatives and their hydrologic behavior as well. Once the PM tools mature, the discrimination tests proposed above can be developed.

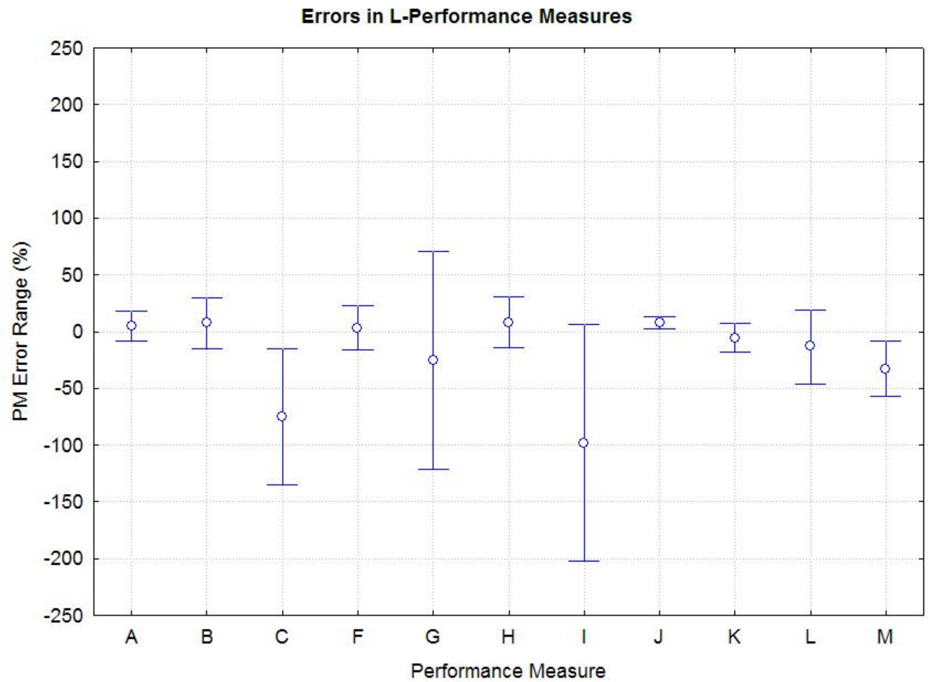

Figure 1: L Performance Measures

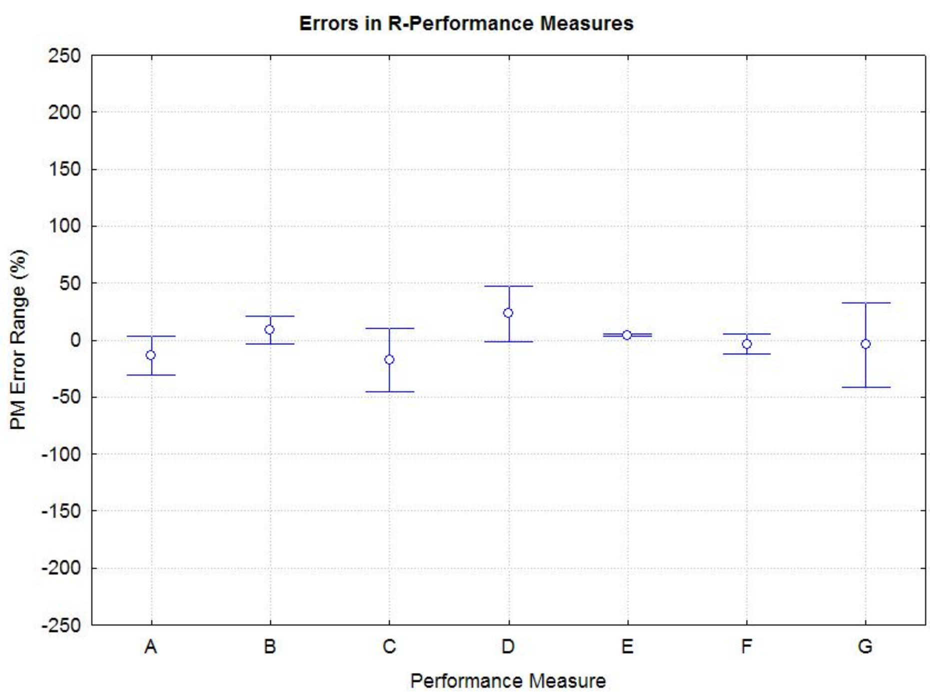

Figure 2: R-Performance Measures

- The validation process used a consistent set of lateral inflows, and this is a primary driver in the level of agreement between the OKISS, AFT, and AET models. However, the impact of using fixed (base) lateral inflows for evaluating all of the various alternatives at the screening and formulation level remains a significant area of concern. The extent to which using the actual lateral inflows for each alternative would affect the model-generated performance measures relative to using baseline lateral inflows must be demonstrated for the effective screening of alternatives. Failure to demonstrate this (in-sensitivity) could potentially be a critical flaw.
  The risk of using a constant set of lateral inflows in screening can be reduced by rerunning Alternatives using a different set of lateral inflow. For plan formulation the Future Base runoff is used while testing will be made using Current Base runoff. The last stage in the Alternative Plan selection process includes the use of the fully integrated model. The effects of the operating criteria in the lateral inflows will be evident in that stage. A check point will be added to the study where it will be decided whether these effects are large enough to cause a revision of the alternative plan screening and formulation process.

(DPL) Page 1-1, Section 1.1:

- For each of the multiple operating objectives (flood control, water supply, aquatic plant management, natural resource management) have operating policies been defined that maximize the benefits to each objective by itself, ignoring the other objectives? Given that conflicts exist among some of them, how is an ‘acceptable balance’ defined, especially in the minds of stakeholders having different priorities?
  Establishing an acceptable balance among stakeholders is achieved through a combination of science, the process of plan formulation and policy. The process being for KBMOS is described in detail in the Alternative Plan Selection Document.

- Will not the modified interim and long-term operating criteria for Kissimmee Basin water control structures be a part of the definition of a plan, and if so, how will the Governing Board be able to pick the preferred alternative plan without knowing the operating criteria and their impact on the performance measures and indicators associated with that plan?
  The Governing Board will require a clear description of the preferred operating criteria and there impact on the PM and Performance Indicators.

(DPL) Page 1-7, Section 1.2.2.6: The revitalization project addresses some but not all hydrologic requirements needed to achieve the ecological health goal. What are missing and why?
The KBMOS project is limited to structure operations. Other aspects of restoration are addressed through the Kissimmee Long Term Management Plan which addressed monitoring needs, water quality issues and Adaptive Management.

(DPL) Page 1-8, Section 1.2.3: How is each alternative plan created? Is it just someone’s idea, or is it partly defined by the optimization included in OKISS?
The development of an Alternative is described in the “Alternative Plan Selection Process.”

(DPL) Page 1-9, Section 1.2.3:

- Would it be useful to list or define the performance measures and indicators associated with each of the management objectives just discussed in the previous section?

- Performance measures are hydrological targets. How does KBMOS define these targets and where is the link between hydrology and ecology (needed for aquatic plant and natural resource (fish, wildlife) management)?

- How does the scoring work in the APS process? If it is subjective and if it involves stakeholders, how stable is it over time?

  The development of the KBMOS Performance Measures is described in the Alternative Plan Selection Document. This document has been reviewed by a separate Scientific Peer Review Panel.

(DPL) Page 1-12: How does the alternative plan screening process produce a 'complete' set of alternative plans when in theory there are an infinity of efficient plans that meet various conflicting objectives to different extents? How are stakeholder workshops going to result in further screening of alternatives? This is really a question about the process of screening and finally coming up with a relatively small number of alternatives to further analyze and eventually submit to the Governing Board.
The development of an Alternative is described in the "Alternative Plan Selection Document."

(DPL) Page 1-13: How can one guarantee flood control constraint compliance?
Flood control is handled as a series of constraints. In screening, the flood performance indicator identifies Alternatives that have a potential compliance issue. Later, the MIKE11 tool will be used by the USACE using short duration, short time step simulations of design storms to assess flooding issues.

(DPL) Page 2-15, Section 2.3.2.1: How will OKISS obtain lateral flow inputs that differ from the base flow conditions, or won't it? Are these lateral flows constant for all alternatives, and are all future scenarios assuming constant land use/cover conditions and hence constant runoff conditions?
OKISS lateral inflows are constant for all alternatives. During the screening project the Future Base inflows are used. At the end of the screening project, the Future Base inflows are replaced by Current Base inflows and the relative rankings of the Alternatives are reassessed. The final promotion of Alternatives will consider only current condition predicted runoff.

(DPL) Page 2-18, Section 2.4.2: Is it assumed that no operating policy changes will be made for Lake Okeechobee, and hence no changes in the boundary conditions?
Yes

(DPL) Page 2-18, Section 2.4.3: Are groundwater boundary conditions constant, and if so is the reasonable?
Lateral boundary head conditions vary throughout the year but are repeated from year to year.

(DPL) Page 3-27, Section 3.2.3.6.1.1: Could you help me understand Equation 3?
Equation 3 is a description of the methodology used to account for the meandering nature of the restored river channel. This correction is necessary because the distance between cross sections in MIKE11 represents the valley length and not the length of the meandering channel.
(DPL) Page 3-73, Section 3.2.2.4: Do you want to label the horizontal axes of all the probability of exceedance plots?

Noted

(DPL) Page 4-14, Section 4.1.2.1.2: What is the MAGO curve? I can't find where it is defined.
"Maximum Allowed Gate Opening" limits gate opening when tail water conditions are too low.

(DPL) Page 5-36, 37: Why wasn't OKISS used to compare its output with AFT for all the performance measures shown in Table 5-1?
Performance Measures included in Table 5-1 were only those that were flagged as critical in Figure 5-56 to 5-61
(DPL) Page 6-6, Section 6.3: Are the performance measures containing values of consecutive days being changed to non-consecutive days to get an improved performance measure for some ecologic indicator, or is it just to get a better model comparison?
The change in the calculation protocol was not promoted to get a better model comparison. This change was proposed by the Interagency Study Team in their review of the first model results. The proposed change was intended to increase the reliability of the PM by eliminating the irregularity (noise) introduced by the some of the structure operations.

(RHP) Appendix A. Is this really necessary to include as a table? Graphical output would be more useful.
Noted.

(RHP) Page 2-4. "Simulation No. 1, the current base condition, will establish a bridge between the real-world context….". It is not really clear how the current base condition simulation will be used "as a means of demonstrating the current effect toward meeting performance targets"?
Text will be expanded for clarity. The Current Base model has significant hydrologic, hydraulic and operational differences compared to the Future Base model. Because current conditions are much better understood than Future Base conditions, the Current Base can be used to validate the model, the operations, and the Performance Measures. This validation is the bridge discussed in the text.

(RHP) Page 2-6. Irrigation Command Areas (ICAs) are concentrated on the western model boundary (mostly in the KUB based on Figure 2-1. Figure 2-2 on the next page shows many associated wells in this area. Won't this affect the lateral boundary conditions for the SAS (and FAS) if the boundary conditions are derived from "available Potentiometric Maps" as stated in Table 2-1, page 2-2 to 2-3? The KBECF model (AFET TDD, 2006) domain seems different along this boundary than modeled with the current AFET model. Was this model used to define the groundwater aquifer boundary conditions?
There are several limitations in the groundwater modeling in AFET. AFET recalibration to address the revised PET datasets is underway in a related project. Groundwater conditions are being given additional attention as part of this effort. This effort is relying on information obtained from the calibrated version of the ECFT (a.k.a. KBECF) model mentioned in the comment.

(RHP) Page 3-11. Figure 3-12. Why is the Reference ET for the later part of period shown during 2000-2009 so different than previous years 1965 to 2000? What are the really high occasional values?
The Reference ET was generated from a collection of sources. Periods after 2000 are based on complete weather station data whereas earlier data are based on adjusted PanET data. This

limitation with reference ET is being addressed by recalibration using newly generated District-wide reference ET data.

(RHP) Page 5-4. Is there a plan to make the location where the OKISS and AFET tools extract results so they compare better?
Modifications to the PM equations are being considered to eliminate artificial sensitivity caused by minor day-to-day differences in gate operations caused by the models. No plans have been made to modify the location. It is considered that the locations selected for AFET provide a more accurate representation of the average stages in the Water Control Units. Due to its nature, OKISS cannot provide stages at those intermediate points.

(RHP) Table 6-1. Are the labels “Location” and “Performance Measure” reversed?
Corrected.

(RHP) Table 6-1. Don’t the statistics for RMSE, R and CE reflect additional error over calibration error (i.e., error between models is added to the AFET calibration error). Is there anyway to use this information in AES (figure 1-2) to distinguish between alternatives?
Although calibration statistics could be used to distinguish between alternatives, there is no known correlation between these statistics and with restoration objectives.

Recommendations
(These could be clarified and/or responded to during our telephone calls.)

(DPL) I wonder if some condensed document showing the model comparisons would be useful to those in the governing board or outside the district that may not want all this detail but just some sense of how well these three modeling approaches work together?
The alternative Plan Selection Document will include an executive summary will be added to this report.

M E M O R A N D U M

TO:	Pete Loucks, Cornell University, Chair David Chin, Consulting Engineer Robert Prucha, Integrated Hydro Systems, LLC
FROM:	Ken Konyha, Section Lead, Hydrologic and Environmental Systems Modeling Christine Carlson, Sr. Supervising Geographer, Kissimmee Division
DATE:	July 7, 2008
SUBJECT:	Action required to address the Modeling Peer Review Panel Comments on KBMOS Sensitivity Analysis

The following memo summarizes the status of Peer Review Panel comments received on the Sensitivity Analysis (previously titled as Uncertainty Analysis). The purpose of this memo is to identify comments that may require an update/modification to project documents or preparation of additional project documentation.

The status of each comment is provided following each item below. RESPONSES HAVE NOT BEEN INCLUDED. Only status and follow-up activities are identified. The summary at the end of the memo includes comments that require modification to the project documentation.

KBMOS Model Peer Review Panel Comments on
Uncertainty Analysis: Kissimmee Basin Modeling and Operations Study (May 7, 2008)

Overall Comments (No response necessary.)

(DAC) The primary objective of the uncertainty analysis (UA) is to quantify the probability distribution of the performance measures that will be used to assess the effectiveness, and discriminate between, alternative operating rules in the Kissimmee Basin.

(DAC) The UA of the AFET model is particularly challenging because of the size and complexity of the model and the computational time required for individual runs. As a consequence, using the benchmark GLUE approach is unrealistic and less-desirable alternatives must be considered. Any alternative UA approach that is chosen will necessarily have limitations, and it is the challenge of the project team to take these limitations into account in applying the results of the UA.

(DPL) It seems to me this report is more a sensitivity analysis of key model parameters than an uncertainty analysis. For example there is no indication that I could find of the probabilities of having any values other than the calibrated base conditions. The report states the analysis will be limited to model uncertainty. In addition to model parameter uncertainty, model uncertainty includes model structure as well, and this was not evaluated, even via a sensitivity analysis. As common in most sensitivity analyses, only one parameter was varied while all others remained at their base condition values. Are there any synergistic or compensating effects? Performing a more elaborate Monte Carlo analyses is indeed time consuming, but if it were carried out it would apparently be based on assumed distributions of parameter values (e.g., normal distributions in which one standard deviation is equivalent to + or - 50% of the parameter value) which themselves are uncertain. This is not a criticism, just an observation that maybe the report title is a little ambitious. We do not really have an uncertainty analysis based on what is contained this report.

(DAC) Overall, the UA [report] appears to be prepared by a well-qualified modeler and engineer, and is a good first step in attacking a very challenging problem.

((RHP) Given the complexities of the integrated flow system, the computational demands of the numerical MSHE/M11 flow model, and challenges of conducting a detailed uncertainty analysis that fully considers the various types of uncertainties (model structure, parameter, measurement error, numerical algorithm etc), the effort presented here represents a good step towards assessing uncertainty in performance measures (PM).

(DPL) The pairs of Figures 4.1 through 4.8 appear to be the same, only the ranges associated with each color have different values.

(RHP) I believe it is still in the best interest of the district to clearly identify and describe all potential sources of uncertainty that may affect estimates of PM values and their possible impacts on PM value uncertainty. I was hoping to see that discussion in this report, and if not here, then it should be added to the Base Conditions report to qualify any predictions, even if it is more qualitative. The integrated model is clearly complex and I believe it is better to acknowledge and discuss the various types of uncertainty (other than parameter) and their potential effects on overall uncertainty of any prediction you make than to leave one with the impression that a) parameter uncertainty as performed in this study is the most significant of all possible types, or b) that only parameter uncertainty warranted further evaluation. I think a qualitative discussion of model (structure) uncertainty and data uncertainty is as important as conducting a quantitative assessment as done here. In my experience, the most significant uncertainty is typically from model uncertainty, where processes, or the spatial and temporal variability of things like aquifer thickness, or effective unsaturated zone hydraulic conductivities are poorly constrained with data, and hence simplified. I think this report, or at a minimum the base conditions report should have appropriate caveats and disclaimers related to uncertainty.=

Specific Comments
(These are requests for clarification and/or responses during our telephone calls.)

(DAC) The following comments and suggestions on the UA are offered:

- Uncertainty in model predictions are conventionally grouped into: model uncertainty (which reflects inadequacies in model structure), parameter uncertainty (which reflects spatial and temporal variability of parameters as well as random variability of parameters), and data uncertainty (which reflects measurement errors). Using these conventional categories, the UA Report considers only parameter uncertainty, without demonstrating that this component of uncertainty is dominant relative to model uncertainty. This is a critical flaw.

A full uncertainty analysis could not be conducted because quantification of model uncertainty and data uncertainty was not feasible within the project scope. The analysis has been re-titled as a sensitivity analysis to reflect this. Based on this, some of the comments included below are not longer applicable because they would apply to a conventional uncertainty analysis and not to a sensitivity analysis.

- The UA does not really use a moment-based approach as stated on Page 2-2; since such an approach relates the moments of the input to the moments of the output. The approach used is closer to a point-estimation method (e.g. Tsai and Franceschni, Evaluation of probabilistic point estimate methods in uncertainty analysis for environmental engineering applications." *J. Environ. Eng.*, 131(3), 387–395, 2005).

The text will be modified to reflect this.

- The variances of five input parameters are estimated based on assumed coefficients of variation (COVs) of 50%. These COV assumptions must be supported by a demonstration that differences between model predictions and observations can be accounted for by these uncertainties in input parameters. This is a critical flaw.

Agreed; the inability to quantify COV for critical parameters limits this analysis. The 50% range was based on best professional judgment but confirmation is needed to demonstrate that the resulting model results remained within the limits of an acceptable calibration. Based on our response to the first comment we believe that if the analysis is seen as sensitivity analysis, this issue is not longer a critical flaw.

- The UA asserts that uncertainty in performance measures (PMs) are indicated by the variability in PMs that occurs when parameter values are varied between their upper and lower limits. This is not the case. For example, by the asserted logic, PM A for L01 (Table 4-4) would have zero uncertainty in model simulations! This is a critical flaw.

Performance Measures evolved throughout this project and the creation of a well-behaved and comprehensive Objective Function was not possible within the limits of the study. Continued work in this area is needed. Based on our response to the first comment we believe that if the analysis is seen as sensitivity analysis, this issue is not longer a critical flaw.

- Clarify the third paragraph on Page 4-23.

(DPL) If one accepts the assumptions made in this report, and also assumes normal distributions of parameter values about their mean base values, one could show on the plots similar to those shown in Appendix B the say 68% confidence ranges of performance indicator values associated with a particular parameter value ranging from 50% less than the base calibrated value to 150% more (i.e., the assumed one standard deviation variation). This would obviously be somewhat misleading without some better idea of just what one standard deviation is. Are there any data that would give one confidence of such likely ranges in parameter values rather than just basing everything on the 50% deviations? If so this might provide a more firm basis for an 'uncertainty analysis.'
The inability to quantify COV for parameters limits this analysis. The 50% range was based on best professional judgment but confirmation is needed to demonstrate that the resulting model results remained within the limits of an acceptable calibration.

(RHP) The following are specific comments/questions:

1) Question: Can I assume that the base conditions model was used to conduct this uncertainty analysis? Was this for existing (2001) or future (2025) conditions (i.e., Scenarios 1, 3 and 4 in Draft Base Condition Summary Report we reviewed in Task 2)? It wasn't clear from the report. If this analysis was only for Scenario 1, would you expect uncertainty in PM values to be much different for Scenarios 3 and 4?

The uncertainty analysis was based on an early version of the current base condition model. A reexamination of the rankings of the top ranked Alternatives will be made using the runoff derived from the current base condition land use using the future condition infrastructure to address this issue.

2) Comment: The first paragraph of Section 5 – Conclusions section sounds like an admission that this uncertainty analysis does not address prediction uncertainty.

3) Question: When one of the five parameters is adjusted to the high or low value, how would it:
    a. affect the overall model calibration?
    b. affect specific areas of model calibration (tables 4-4 through 4-6 in the AFET model documentation/calibration report – Aug 2007)?
    c. affect specific areas of model verification (Tables 5-1 to 5-5)?

    This seems like useful information, even if these parameter adjustments weren't simulated in the calibrated AFET model. If the effects of changing the single parameter, without changing other parameters, markedly changes the calibration to the point that the model no longer meets calibration criteria (as per the AFETS Model Documentation/Calibration Report Aug 2007) is the adjustment too much? Isn't the objective to find the range of parameter values (typically different combinations of parameters) that are capable of producing equally valid models based on specified calibration criteria? By adjusting single parameter values, over a range which likely results in an uncalibrated model, aren't you in effect, producing unrealistic extremes in system response? The adjustment in a parameter's value should be constrained by requiring the model to still meet its calibration criteria.

    Another way to look at this is that the single-parameter adjustments probably produce a greater range of PM values than if the adjustments had been constrained such that each simulation met specified calibration criteria – or were at least comparable to the calibrated AFET model. Therefore, one could argue that the work done here is reasonably conservative in estimating the uncertainty – but you would have to show that the high and low parameter value adjustments caused inferior calibration.
    Parameter variations that cause the model to fail to meet accepted calibration criteria should be rejected. The variations in runoff were adjusted to stay within plus/minus 2 inches per year since this was considered to be the accuracy of the measured data. A more formal process is needed for future uncertainty work.

4) An underlying concept that seems well established in relevant literature is that the specific set of model input parameter values/structure has uncertainty and that alternative sets of values can equally meet pre-defined calibration criteria (i.e., Tables 4-1 through 4-3 in the AFET Model Documentation and Calibration report Aug 2007). The current calibrated AFET model generally meets most of the calibration criteria/targets throughout the model domain. In reality, don't the specified calibration targets dictate the acceptable level of uncertainty? If calibration criteria/targets had been specified to a much higher level, wouldn't this have by design reduced uncertainty – if you were able to produce a model that could be calibrated to this level? As such, I wonder if the targets couldn't be somehow converted to an equivalent range of uncertainty in PM values.

Calibration targets consider data quality and, as such quantify data uncertainty. PM values are more complicated and contain both data uncertainty and (ecologic) model uncertainty. The PM targets consider both a subset of hydrologic variables (for example water depth in the riverine flood plain) and ecologic response to those variables (for example, hydrologic suitability for broadleaf marsh). Using PM targets in lieu of calibration targets would complicate the uncertainty problem unless both data and model uncertainty are both understood. At this point in time, the PM targets of the KBMOS project are not well enough understood to be used in place of the traditional calibration criteria.

5) Is there a way to show that this uncertainty analysis is consistent with other district studies, or by consultants for district? I liked the report "Model Uncertainty Workshop Report, May 2002 by Lall, Phillips, Reckhow and Loucks". It might be nice to show that the district has led the way in assessing uncertainty and has a systematic way of dealing with uncertainty in modeling studies to qualify results. Adding such text to this document would help support this effort, especially if this issue has come up in other studies and considered in more detail. It would also be useful to define up front the various sources of uncertainty. I've noticed a lot of inconsistent terminology in various publications.

   Model Uncertainty Workshop Report, May 2002 by Lall, Phillips, Reckhow and Loucks":

   - Model output uncertainty comes from input variability and measurement errors,
   - parameter uncertainty,
   - model structure uncertainty and
   - algorithmic (numerical) uncertainty.

   David's list:
   - Model uncertainty,
   - Parameter uncertainty, and
   - Data uncertainty.

   US EPA, 1999, Maged M. Hamed, Philip B. Bedient, "Reliability-Based Uncertainty Analysis of Groundwater Contaminant Transport and Remediation":
   - Modeling uncertainty,
   - Prediction uncertainty,
   - information uncertainty, and
   - inherent, intrinsic, or physical uncertainty.

   The discussion of the Sensitivity Report will reference these reports.

6) Will predictive modeling produce hydrologic/hydraulic conditions that are outside the range of that observed during the calibration period? In other words, with specified land-use changes, or changes in use of groundwater etc, will this uncertainty analysis reflect this uncertainty if new extremes in H&H are produced. How does the district plan to handle this?

   The calibration period encompasses both 1-in-10 year droughts and 1-in-25 flood events and should cover the range of conditions experienced in the longer 41 year period of simulation.

Recommendations

(These could be clarified and/or responded to during our telephone calls.)

(DAC) A possible way to begin to address the concerns described in this review is to use the model-validation (and possibly calibration) output with synoptic measurements to compare the PMs estimated from measurements with the PMs estimated by the model. (Admittedly, some statistical creativity will be necessary because of the limited duration of the calibration/validation period.) This discrepancy in PMs can then be compared with the uncertainty in PMs derived from parameter uncertainty (using the UA Report). If the former discrepancy is dominant, this discrepancy can be used directly to estimate PM uncertainty, if the latter is dominant, the results in the UA Report can be used to estimate PM uncertainty.
The Base Conditions report now contains a validation section that compares AET, AFT, and OKISS: calibration statistics are similar; Tier 2 metrics are similar; and differences in PM metrics appear to be related to their formulation rather than to underlying differences in model behavior.

(DPL) Given the tables and plots showing the impact of varying parameter values, one at a time, on performance measures, how will this additional information be used in making decisions regarding infrastructure development and operating policies? What do those making such decisions want to know? Maybe they don't know what they want to know, yet, but if they do it should guide what next steps might be taken to provide such information, even if such information itself is uncertain and based on just professional judgment. In any case I would think some assessment of the probability of actually observing such ranges would be useful.