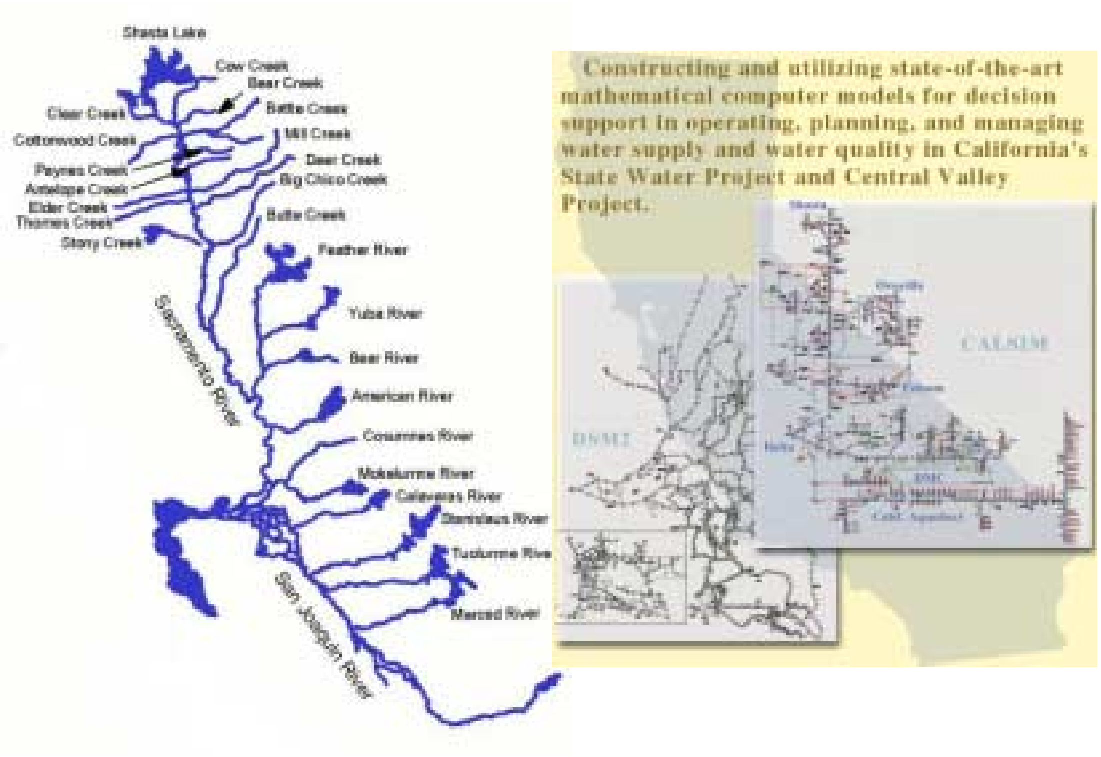
A Strategic Review of CALSIM II and its Use for Water Planning, Management, and Operations in Central California



Constructing and utilizing state-of-the-art mathematical computer models for decision support in operating, planning, and managing water supply and water quality in California's State Water Project and Central Valley Project.

Submitted to the

California Bay Delta Authority Science Program
Association of Bay Governments
Oakland, California

by

A. Close, W.M. Haneman, J.W. Labadie, D.P. Loucks (Chair),
J.R. Lund, D.C. McKinney, and J.R. Stedinger

December 4, 2003

Executive Summary

1. Summary

The central all-encompassing question put to the panel is whether the CALFED program has adopted an appropriate approach to modeling the CVP-SWP-Central Valley system. Is the general CALSIM modeling approach appropriate for predicting the performance of the general facilities and for use in allocation planning, assessing water supply reliabilities and for carrying out operational studies? We believe the use of an optimization engine for simulating the hydrology and for making allocation decisions is an appropriate approach and is in fact the approach many serious efforts of this kind are using. It is a substantial improvement of the previous modeling approaches and provides a basis for consensus among federal and state interests. The modeling approach addresses many of the complexities of the CVP-SWP system and its water management decisions.

There exists a common tension between those who wish for greater detail and those who want less detail from the model. This argues for a more comprehensive, modular and flexible approach than is now available. In this report we suggest some ways this might be accomplished in the future. We also propose some management procedures that could be considered to improve model and model application quality control and documentation. The openness and availability of the model is admirable and very important given the numerous stakeholders who have interests in the management and allocation of water in the state. To increase the public's confidence in the many components and features of CALSIM II, we suggest that these components of CALSIM be subjected to careful technical peer review by appropriate experts and stakeholders.

2. Background

The California Department of Water Resources (DWR) and the U.S. Bureau of Reclamation (USBR) have developed a computer model called CALSIM II that simulates much of the water resources infrastructure in the Central Valley of California and the Delta region. This infrastructure is referred to as the CVP-SWP system. In particular CALSIM II provides quantitative hydrologic-based information to those responsible for planning, managing and operating the State Water Project (SWP) and the federal Central Valley Project (CVP). As the official model of those projects, CALSIM II is the default system model for any inter-regional or statewide analysis of water in the Central Valley of California.

CALSIM II has a central role in the analysis of many CVP-SWP and related issues, some of which require capabilities beyond those included in the model. California needs a large-scale relatively versatile inter-regional operations planning model and CALSIM II currently serves that purpose reasonably well. As the primary State and Federal-sponsored model available for water operations and planning, CALSIM II is critical to the study of many technical and policy issues related to water supply reliability, environmental management and performance, water demands, economics, hydrology and climate, and regulatory compliance.

CALSIM II is a particular application of the California Water Resources Simulation Model called CALSIM. It uses a mixed integer linear programming model solver to route water through a network over time. Currently it uses monthly time steps. Policies and priorities are implemented through the use of user-defined weights applied to the flows in the system (represented by arcs of the network). Simulation cycles at different temporal scales allow for successive implementation of constraints. The model can simulate the operation of relatively complex environmental water accounts and state and federal environmental regulations. In our judgment CALSIM II represents a very impressive modeling effort on the part of all those involved with its development and application.

The CALFED Science Program commissioned this external review panel (Appendix D) to 1) provide an independent analysis and evaluation of the strengths and weaknesses of CALSIM and CALSIM II, and 2) to offer suggestions on the appropriate uses of these modeling tools, on ways their use might complement or be complemented by other models, and on further development, quality assurance, and use in major water systems operations and planning in California.

The panel received background documents (Appendix B), including a survey by the University of California at Davis of stakeholder responses to questions about CALSIM II. We subsequently met for one and a half days in Sacramento for discussions and presentations (Appendix A) by CALFED, DWR and USBR staff. The discussions concluded with a summary presentation by the panel outlining our tentative conclusions.

The information we received and the shortness of our meetings with modeling staff precluded a thorough technical analysis of CALSIM II. We believe such a technical review should be carried out. Only then will users of CALSIM II have some assurance as to the appropriateness of its assumptions and to the quality (accuracy) of its results. By necessity our review is more strategic. It offers some suggestions for establishing a more complete technical peer review, for managing the CALSIM II applications and for ensuring greater quality control over the model and its input data, and for increasing the quality of the model, the precision of its results, and their documentation.

In this review we were asked to address the following questions:

1. Is CALSIM a reasonable modeling approach for current and proposed applications and problems?
2. Do other modeling approaches show similar or greater promise and flexibility for such problems? If so, how?
3. What are the major comparative strengths and weaknesses of the current CALSIM approach and alternative approaches?
4. What are major scientific, technical, and institutional limitations, uncertainties, and impediments for current and proposed applications of CALSIM?
5. What model, software, and data developments, special studies or tests would be beneficial to improve CALSIM for current and proposed uses?
6. How might CALSIM development and applications be managed and overseen to improve the quality assurance of model results for current and proposed applications?

7. What are your suggestions for long-term use, development, or replacement of the current suite of models and data available for the current and proposed uses of CALSIM?

The following sections of this summary present our responses to these questions. The main parts of this report and its appendices provide additional detail.

3. CALSIM Modeling Approach

CALSIM II is a simulation model developed as a joint venture between the California Department of Water Resources (DWR) and the U.S. Bureau of Reclamation (USBR) to (i) provide a significant modernization and upgrading of the DWRSIM and PROSIM models developed and used by these organizations, (ii) develop a comprehensive modeling system that simultaneously addresses the current and future needs of both the SWP and CVP systems; and (iii) develop a generalized modeling system that could be applied in any river basin system, in contrast with the previous models that were less generalized and more specifically designed for the existing SWP and CVP systems. In this respect, CALSIM II represents a state-of-the-art modeling system that is similar in general concept, while differing in specific details, to other data-driven river basin modeling systems such as ARSP, MODSIM, OASIS, REALM, RiverWare and WEAP.

CALSIM uses linear programming to solve sets of equations that simulate water movement through the CVP-SWP system in accordance with various objectives and constraints. This is a modeling approach which has been used successful in California (Johnson et al., 1991). In a complex system such as that being modeled, it is essential to have some mathematical representation of system flows that reflects all of the interconnections and constraints. Use of an optimization algorithm allows good decisions to be identified from among all possible and feasible decisions. To the extent this simulates what actually occurs, it is a good modeling approach. To the extent it optimizes when in reality no such optimization is implemented, it has the potential to produce inaccurate and overly optimistic outputs.

Most successful applications of optimization that attempt to simulate the behavior of a system have calibrated their objective functions (i.e., set the weights that prioritize flows over time and space) so that the model results correspond to what actually happens or would happen under a particular hydrologic and demand scenario. In these cases the model's decisions correspond to those the operators would make, as often prescribed by rules that have been worked out in a legal/political process. It does not appear that such a calibration of the objective function weights in CALSIM has yet been completed.

4. Other Modeling Approaches

There are two aspects of modeling, the model structure and algorithms used, and the model software. The use of linear optimization algorithms to solve simultaneous equations for simulating hydrology is a common way of avoiding a typically long list of procedural rules for simulating regional water systems. Such sets of procedures can be difficult to generate for complex systems, and very different and new rule sets may be needed if structural or significant policy changes are to be investigated. In addition the performance of the system when simulated will be less than that which can be achieved in practice if a good set of rules is not provided. Optimization models are generally easier to reformulate when system changes are to be investigated. However unless the optimization is calibrated in such a way as to actually resemble what takes place in practice it can produce an optimistic description of system performance. This is particularly true if the optimization model is allowed to have perfect foresight of future events that in practice would not be available to system operators.

Large simulation models using optimization and procedural rules both need to have internal checks to ensure to the extent possible that errors in mass balances, for example, do not occur due to errors made when the model is being defined or created. Such internal checking is not apparent to us in our admittedly brief review of CALSIM II. Nor were calibration procedures well defined.

One obvious limitation of using linear optimization procedures is its inability to model accurately and efficiently some of the non-linear hydrologic and decision processes that occur in systems as complex as the CVP-SWP. One approach to addressing this issue of model accuracy, and possibly for decreasing the computational time as well, is to link linear optimization models to non-linear simulation models in a way that permits the simulation to represent the hydrology in any spatial and temporal detail desired. The optimization is used to determine what the decisions should be at every site where a water allocation, reservoir release, or other management decisions must be made. The time steps for simulation could be daily, or weekly or longer, depending on the needs of the user, but would likely be of shorter durations than the optimization time steps. After a predetermined number of simulation time steps, the optimization model would be run. The initial state of the optimization should be set at the beginning of each optimization time step. The optimization component should include multiple future time periods, with imperfect hydrologic and demand forecasts, but once solved only the current period's solutions are implemented – i.e., these decision variable values are sent to the simulation component. The decisions indicated for future periods are ignored. When appropriate, the initial state of the multi-period optimization model is updated and the model is again solved. And so on. Such a modeling approach may prove to be both more realistic, more accurate, and require less time, once developed. We believe such an approach might be worth considering for future development.

CALSIM II currently consists of a combination of software modules developed in several languages, including FORTRAN, Java and C. Several of the modules require proprietary software packages in order to run CALSIM II (Lahey FORTRAN and XA Solver). DWR and USBR staff have said that these components are being replaced by public domain software that can be obtained free of charge. We agree with this decision. Very good public domain software packages of optimization, visualization, file management, and data base support are currently available, and new ones will continually be produced. Periodic updates should be anticipated as part of the business of maintaining the modeling system.

Significant thought should be given to the sustainability of the CALSIM II software. How will future programmers be able to maintain this software? How will future software developments be incorporated into the system? Will the solver currently being developed by LBNL be adequate in terms of accuracy and computation speed? Will other solvers need to be tested? Can the system accommodate these future developments without major modifications? What reasonable modifications could be made now in anticipate of future developments?

5. Comparative Strengths and Weaknesses

Many of the stakeholder perceived strengths and weaknesses of CALSIM and CALSIM II are very well identified in the survey report from the University of California at Davis (Ferreira, et al. 2003). Our background materials and briefings covered various strengths and weaknesses, but without first hand experience, all we can do here is to summarize those that we have heard expressed by others.

Here we provide a brief summary list.

5.1 Some Prominent Strengths

The strengths of CALSIM II are many. Most are expressed in comparison to previous DWRSIM and PROSIM models DWR and USBR were using. Some of these strengths include:

- Consensus model. CALSIM II is the official joint modeling environment of the State DWR and USBR. This includes a common schematic, hydrologic representation of the system, common set of facility capacities, and common representation of system operating policies. This helps all parties improve representations, rather than compete over representations.

- Common effort. The joint development of CALSIM II by USBR and DWR has provided more focused and effective use of resources and expertise than previous development of agency-specific models. CALSIM II development has also involved other agencies and consulting expertise more than pervious models of this system.

- Data-driven model. CALSIM II is a rather data-driven simulation model with an optimization engine. This modeling approach provides:
    a. greater flexibility than its predecessors and traditional water resources simulation approaches.
    b. a promising framework for improving transparency, data, and model documentation, compared to other approaches.

- Public domain. The model and data are substantially in the public domain, facilitating transparency and adaptability for California's decentralized water system.

- Steady improvements. Data improvements have been steadily pursued following the adoption of CALSIM II, although deficiencies remain.

- Improved Delta water quality representation. Although problems appear to remain, the model developers have made substantial gains in representing Delta water quality operating criteria and performance.

- Better groundwater representation. Efforts to better include groundwater and non-CVP-SWP project operations merit continuation and expansion.

- Benchmark Studies. The development of documented benchmark studies have resulted in significant model improvements and aided in the development of comparative model applications. Such exercises should be continued and improved.

- Long-term vision. The vision of a more transparent and publicly available model that can be employed by those outside the major agencies is excellent. This is a major change in direction, and achieving this vision will require adjustments over time. Often, these adjustments will be externally driven. Externally-driven improvements are a price of success and evidence of success for an open, public, modeling policy.

- Important CALSIM II features:
    a. CALSIM II is able to simulate the operation of the complete CVP-SWP system in all areas that contribute flow to the Delta in monthly time-steps.
    b. CALSIM II is being applied to examine a diverse range of options including flood control, water conservation and supply, power generation, recreation, water transfers, groundwater banking, recycling, desalination, conjunctive use, the purchase of options and streamflow and water quality protection.
    c. CALSIM II has successfully been applied by both DWR and USBR to examine both structural and non-structural changes to the CVP-SWP system as well as to ascertain the risks involved with different potential operating scenarios and to quantify the impacts of proposed actions.
    d. CALSIM II can dynamically model operation of environmental water accounts.
    e. Demands may vary according to various levels of development (e.g. 2001, 2020) and to hydrologic conditions.
    f. The regulatory environment under which the projects must operate can be simulated.
    g. CALSIM II can link to external modules as needed, e.g., to estimate the salinity at water quality stations within the Delta.

5.2 Some prominent weaknesses

As its strengths are many, so are its weaknesses. It seems worth saying, however, that no model can perfectly (meaning efficiently and effectively) serve all interests in a system as complex as the CVP-SWP. Tradeoffs need to be made. This can result in what some would call weaknesses. Such weaknesses are often accepted to gain strengths in another ways.

We heard that the CALSIM II model was too complex. We also heard that it did not handle particular components of the system with sufficient detail. And such is the dilemma of any complex model, such as CALSIM II. The model is clearly too complex, and not complex enough. The root of this difficulty is that when such a model is constructed, it is not clear what level of detail is needed, so the model must be made sufficiently complex to ensure it is complex enough. And the complexity needed to address some issues will remain in the model when it is used to address other less complex issues, or the same issues at less complex locations. One approach to addressing this issue is to develop different linkable modules of CALSIM II having different complexities. In this way the level of detail can be varied to be consistent the application or study at hand, and level of sophistication and resources available to the user.

Other weaknesses model users would like addressed include:

- The model provides limited and inadequate coverage of non CVP or SWP water and of the California water system south of the Delta.
- The model assumes that facilities, land-use, water supply contracts and regulatory requirements are constant over this period, representing a fixed level of development rather than one that varies in response to hydrologic conditions or changes over time.
- Groundwater has only limited representation in CALSIM II.
- Groundwater resources are assumed infinite, i.e., there is no upper limit to groundwater pumping.
- The linear programming model considers only the current month, and hence CALSIM II operating rules are required to determine annual water allocations, to establish reservoir carryover storage targets, and to trigger transfers from north of Delta to south of Delta storage.
- Better quality control is needed both for the model and its current version and the input data. Procedures for model calibration and verification are also needed. Currently many users are not sure of the accuracy of the results. A sensitivity and uncertainty prediction capability and analysis is needed.
- Need improved ways of altering the model's geographic scope and resolution and its temporal resolution to better meet the needs of various analyses and studies.
- Need to improve the model's comparative as well as absolute (or predictive) capabilities.
- CALSIM II needs better capabilities for analyzing economic, water quality, and groundwater issues.
- Need improved documentation explaining how the model works, its assumptions, its limitations, and its applicability to various planning and management issues.
- DWR and USBR have not provided a centralized source of support for CALSIM II. More training for CALSIM II is needed. There is a need for more people who can run CALSIM II. There is a need for a well-publicized user group. A more extensive users' guide is needed.
- Improved capabilities are needed for real-time operations especially during droughts, gaming involving stakeholders during a simulation run, handling of evapotranspiration and agriculture demand changes over time, water transfers, Delta storage, carryover contract rights, refuge water demands and more up to date representation of Feather River, Stanislaus River, Upper American River, San Joaquin River and Yuba River operations.

- Need an improved graphical user interface to facilitate input of model data, setting of model constraints and weights, operating the model, and displaying and post analysis of model results.
- Need to be able to change the model time period durations for improved accuracy of model results.

6. Limitations, Uncertainties, and Impediments

6.1 Absolute Values or Comparative Results

Modelers sometimes make a distinction between the use of a model for *absolute* versus *comparative* analyses. In an absolute analysis one runs the model once to predict an outcome. In a comparative analysis, one runs the model twice, once as a baseline and the other with some specific change, in order to assess change in outcome due to the given change in model input configuration. The suggestion is that, while the model might not generate a highly reliable absolute prediction because of errors in model specification and/or estimation, nevertheless it might produce a reasonably reliable estimate of the relative change in outcome. The panel is somewhat skeptical of this notion because it relies on the assumption that the model errors which render an absolute forecast unreliable are sufficiently independent of, or orthogonal to, the change being modeled that they do not similarly affect the forecast of change in outcome; they mostly cancel out. This feature of the model is something that would need to be documented rather than merely assumed.

In our opinion CALSIM II has not yet been calibrated or validated for making absolute predictions values. Yet it is apparent that there has been a distinct need by model users for absolute predictions. In the absence of alternatives, users are adopting CALSIM II results as the best absolute prediction available and they are likely to continue to do so. We recommend that model developers recognize the requirement for CALSIM II to provide absolute predictions. To satisfy this new purpose, additional calibration of the model will be required to ensure that the output it produces is fit for this purpose. Regardless of how possible it is to match the model closely with observed behavior, statistics on the accuracy of the calibration run should be supplied to users to enable them to gauge the likely errors involved with using the model output.

6.2 Sensitivity and Uncertainty Analyses

Sensitivity analyses would be useful to identify which parameters and input data have major impacts on decisions and system performance criteria of concern. Uncertainty analyses would help users of the model understand better the risks of various decisions and the confidence they can have in various predictions.

6.3 Graphical User Interface

Having a graphical user interface would substantially aid those who use the model in managing both input and output data, and in controlling or managing model operations. This model will not likely become as available to and as well understood by the public, to the extent desired by the model developers, until an effective menu-driven GUI has been created that can help create and draw from a database of system parameters and characteristics, and simulation results.

.

6.4 Documentation and Training

When if ever is adequate documentation and training available? Rarely, but we believe there is a serious need to improve the documentation as well as the training available for all those interested in using CALSIM II.

7. Options for Improving CALSIM

7.1 CALSIM Model Software

We encourage the developers of CALSIM to convert their present software to that which is publicly available and to develop a useful graphic based user interface that can facilitate the input, editing, and display of all the data that are input to and output from CALSIM II. There are many options, some of which we have discussed with the model developers.

The CALSIM package should be made more modular and capable of linking to other more complex models of components of the CVP-SWP system. If the changes in code and modeling approach result in a quicker running model, it might be possible to link, when desired, modules that facilitate position analyses and other types of uncertainty analyses. A modular system would allow alternative representations of different components of the system. Thus different levels of spatial detail, or representations of the fundamental processes, would be allowed within the overall system representation and record of California hydrology. This will allow the use of more general and streamlined models for use of preliminary investigation and general planning, as well as a more detailed representation of the system for final analyses and more detailed studies. This would be very useful.

7.2 Sensitivity and uncertainty analyses

Both sensitivity analyses need to be performed, and procedures need to be developed to enable the estimation of measures of uncertainty associated with model output. Perhaps workshops focused on just these needs should be scheduled to better determine how best to meet these needs. There are numerous procedures available that could be applied. Appendix H contains some approaches for performing sensitivity and uncertainty analyses.

7.3 Model calibration

There is a need to develop the model so that it is able to provide absolute estimates of key model outputs rather than limiting the use of the model to comparative studies. One way to do this is to subject the model to a comprehensive calibration process where it is fine-tuned until it is able to reproduce the historical behavior of the system with sufficient accuracy to provide absolute results. The calibration of the model should aim to test all the key outputs of model including water quality in the San Joaquin River and in the Delta. It is necessary to test the monthly values of outputs for those outputs for which the monthly pattern is important.

7.4 Other extensions and improvements

- The opportunity of improving the collection of data on the use of water (preferably broken down by irrigation district and water source) should be investigated. The use of groundwater should be included in this investigation.
- It would be useful to expand the geographic extent of the model so that it includes all the components of the linked water supply system, including both the San Joaquin and Tulare Lake Basins of the Central Valley. The model should also account in some manner for imported supplies of water to users in southern California from the Colorado River.
- The linkage between surface water and groundwater would appear to be of critical importance and output that would enable the impact of surface water use on groundwater extractions would appear to be useful.
- Examination of the report *'CALSIM II Simulation of Historical SWP/CVP Operations',* DWR (2003) indicates that the current formulation of CALSIM II:
  - Overestimates water deliveries to SWP and CVP contractors,
  - Determines carryover storage target values that differ from those the operators have determined in the past, and
  - Operates the San Luis Reservoir at lower levels and fills it later in the season than operators have in the past.

8. Managing CALSIM Development and Applications

The predicted impacts and other information derived from CALSIM II applied to the CVP and SWP can influence major investment decisions. It is thus self evident that those who use the model results need to have some confidence as to their precision. Is the science behind the information derived from CALSIM II been reviewed and judged correct? Is the model software free from errors? Are the assumptions made when performing the modeling the correct ones? Are the model results accurately and fully reported? In other words, just how much credence should decision makers place in the model output? Users of the model results should be assured that they are credible and unbiased. One way to help ensure this is to have the models, their associated software, and their applications under the control of some interagency organization that can oversee and provide quality control over model development, application and documentation. They can also plan and implement needed peer reviews.

One possible means of facilitating the peer review processes and for maintaining control on the particular versions of CALSIM II and accompanying models used for CVP-SWP planning and management decisions is to create an interagency modeling consortium (IMC) consisting of DWR, USBR, and other stakeholder organization (including university) personnel if they are interested and want to participate. This center would be responsible for maintaining a toolbox of 'acceptable' models for use by the agencies and contractors. The models placed in the toolbox should be peer reviewed with respect to their applicability and suitability for use in particular applications. Those that are not peer reviewed should be considered for peer review. New models proposed for use in California should be peer reviewed with respect to their suitability, and for their strengths and limitations, before being placed in the toolbox. The review should be of the theory underlying the model, the model's software, the documentation of the model as well as of its software, the model's functions and capabilities including those pertaining to model data input and output, the input data themselves, model calibration and verification, capabilities for sensitivity and uncertainty analyses, user control of all model operations including pre and post analyses (GUIs), spatial and temporal resolutions, and its limiting assumptions.

9. Future Use, Development, or Replacement of CALSIM

9.1 A coupled optimization simulation approach

Given a system as complex as the SWP/CVP system, it seems to us it might make sense to consider the development of a more detailed simulation 'engine' and couple it to an optimization or management 'engine'. The simulation component can more accurately model hydrologic processes. For example it can include the deterministic non-linear routing of flows and their quality constituents through the system on a smaller time step (e.g., daily) and hence much more realistically or accurately, than can linear optimization using longer time steps, even with all the known tricks for linearizing separable (single variable) non-linear functions and 'if-then-else' statements. The simulation engine itself may require a simultaneous equation solver, especially for the Delta. But the simulation engine needs to know what to do, i.e., what decisions to make. Periodic use of the optimization, say once a week or even less frequently if conditions are relatively constant, for determining the decisions to be simulated, e.g., the water allocation and reservoir release decisions, eliminates much of the maze of rules that otherwise would be required and which developers of CALSIM II are avoiding through the use of optimization. Each time the optimization or management 'engine' is run it is first updated with the current state of the system as determined from the more precise simulation 'engine'. The optimization component would include multiple time periods only to the extent that the current period's solution is not affected by the time horizon in the optimization. The other time period solutions are ignored. This coupled optimization-simulation approach has the potential to be both more accurate as well as quicker to execute. In our opinion it is worth considering for future development.

9.2 Models as hypotheses

CALSIM II is really about the future, not the past. Benchmarking studies can help establish the credibility of the model and provide estimates of its accuracy by comparing its performance to actual historical operations. A concern is how well the model reproduces historical operations, not whether it is valid or invalid on some absolute scale of perfection. But the real issue is how well CALSIM can predict what might happen in the future with sets of hydrological and meteorological conditions that have not yet been experienced, and may be significantly different from the past if climate variability and climate change are considered. In these cases the ability of the model to forecast what will happen depends both upon its ability to describe what would happen should a particular system operating policy, priorities and water demands be adopted. In this sense CALSIM II modeling studies should be thought of as the exploration of a hypothesis that particular policies and priorities have been adopted. Our ability to predict the future has generally been poor, but it is the obligation of agencies such as DWR and USBR to attempt to ensure that should water demands, water supplies, and water policies evolve as one would expect, society is prepared for the consequences. And that would seem to be what CALSIM II is about.

9.3 Future Model Development and Use

From the list of perceived weaknesses above, there are clearly many opportunities for further refinement of CALSIM II. Rather than attempt to meet all needs using only one model, namely CALSIM II, it seems preferable to improve its adaptability to various levels of detail through its ability to link to other models when additional detail in a particular region or for a particular feature is desired. For example, the monthly time step used by CALSIM II is sufficient for many studies. Yet some seasonal (multi-month) decision making is needed in CALSIM II to reflect decisions made by the SWP and CVP as to what Table A and other allocations to honor in full. On the other hand, it is clear that many water quality and ecosystem management decisions would profit from more detailed weekly or daily time steps. However, such shorted time-step models will need the guidance of a longer time-step model. As discussed earlier, models with shorter time scales can require increased spatial resolution, both of which lead to increased model complexity and a strong argument for model modularity.

Additional potential applications of CALSIM II include operational planning using gaming, or the involvement of potential decision makers during the simulation runs via a well developed graphical user interface, and to improve the capability of modeling water quality, energy production, conjunctive groundwater and surface water interactions and use, to mention a few.

There will always be a need to perform alternative ‘what if’ policy analyses where a relatively fast model that also provides some capability for uncertainty analyses is required. Perhaps CALSIM II will never be able to serve this need, and if so another more simplified modeling approach could be developed to fill that need. This simpler screening tool would be calibrated to produce results comparable to those of CALSIM II or observed data. Is this possible? We can not be certain but feel the idea should be seriously considered.

Acknowledgments

We want to acknowledge and thank all those who were involved in preparing the materials we received, in presenting additional information to us, and in arranging our activities and taking care of us during our brief review visit. Our special thanks to Kim Taylor for leading this effort, and to all those who did their best to educate us on the finer points of CALSIM II. All those involved in the development and implementation of CALSIM II are to be congratulated for extending the state-of-the-art in water systems modeling and for making it a critical part of the planning, development and management of California's water resource infrastructure.

Caveat

Just as all models are approximations of reality, so may all advice be an approximation of what it should be. We hope what we have written in this report is correct and useful, but encourage CALSIM model managers and California's water community to take our assessments and suggestions for what they are, arrived at based on our own experiences and some limited exposure to those who know much more about CALSIM and CALSIM II than we do.

Table of Contents

1. CALSIM Compared to Other Modeling Approaches

Management of complex systems such as coordination of the California State Water Project (SWP) and the Federal Central Valley Project (CVP) requires effective decision support tools for simulating and analyzing system components in a fully integrated manner. The classic definition of a decision support system (DSS) provided by Sprague and Carlson (1982) is *"an interactive computer-based support system that helps decision makers utilize data and models to solve unstructured problems."*

A DSS integrates the following interactive subsystems: (i) dialog generation and management subsystem (DGMS) for managing the interface between the user and the system; (ii) data base management subsystem (DBMS); and (iii) model base management subsystem (MBMS).

CALSIM II is a DSS developed as a joint venture between the California Department of Water Resources (DWR) and the U.S. Bureau of Reclamation (Bureau) to (i) provide a significant modernization and upgrading of the previous models DWRSIM and PROSIM employed by these organizations, (ii) develop a comprehensive modeling system that simultaneously addresses the current and future needs of both the SWP and CVP; and (iii) develop a generalized modeling system that could be applied in any river basin system, in contrast with the previous models that were less generalized and more specifically designed for the SWP and CVP. In this respect, CALSIM II represents a state-of-the-art modeling system that is similar in general concept, while differing in specific details, to other river basin modeling systems such as AQUATOOL (Valencia Polytechnic University, Spain), ARSP (Acres Reservoir Simulation Program) (Boss International, 2003), IRAS (Interactive River-Aquifer Simulation) (Loucks, et al. 1996), MIKE BASIN (Danish Hydrologic Institute, 2002), MODSIM (Labadie and Larson, 2000), OASIS (Randall, et al., 1997), RAISON (Young, et al. 2000), ResSim (U.S. Army Corps of Engineers, Hydrologic Engineering Center), Ribasim (River BAsin SIMulation Model) (Delft Hydraulics, Netherlands), REALM (REsource ALlocation Model) (James, 2003), RiverWare (Zagona, et al. 1998), WaterWare (Jamieson and Fedra, 1996), and WEAP (Water Evaluation and Planning System, 2003) (Hansen, 1994). All of these can be categorized as decision support systems since all three subsystems of a DSS are embodied within them.

A distinguishing feature of several of these modeling systems is the use of optimization on a period by period basis (not fully dynamic) to “simulate” the allocation of water under various prioritization schemes, such as water rights, without the presumption of perfect foreknowledge of future hydrology and other uncertain information. This is a valid approach since use of optimization overcomes the disadvantage of employing numerous, unwieldy prescriptive rules governing water allocation. Systems employing optimization in this manner include: ARSP, MODSIM, OASIS, REALM, RiverWare, and WEAP and are therefore more akin to CALSIM II. ARSP, MODSIM, REALM and Ribasim are further distinguished by use of specialized minimum cost network flow optimization algorithms, although of these only MODSIM includes iterative structures using an imbedded scripting language for including non-network “side constraints” in the optimization. The other modeling systems are essentially limited to a pure network structure that does not allow inclusion of all the complex, non-network type constraints necessary to model the complex CVP-SWP system.

It may be useful to compare this use of optimization with some other uses that have appeared in the modeling literature. One use of optimization is purely for computational convenience; in this case optimization is employed as a numerical method for obtaining the solution of a series of simultaneous (often linear) equations. This approach, which was used in the first generation of computational economic models about forty years ago, exploited the fact that some existing computational algorithms for solving optimization problems were faster than those for solving large systems of simultaneous equations. A second use of optimization applies when the solution of the system of equations characterizing a water balance has multiple possible solutions; this is essentially the case described above, where optimization is being used primarily to identify a unique solution for a system of equations. Both of these uses of optimization are primarily descriptive rather than prescriptive (also referred to as positive vs. normative) in intent: the goal is to model how a system, characterized by a set of equations, operates. To the extent that the real–world managers of the system do optimize some objective function, the aim is to mimic their behavior by setting up and solving a similar optimization. But, the goal is to model what they actually do, not to advise them what they ought to do. The third use of optimization adopts an explicitly prescriptive goal and sets out to ascertain what managers ought to do if they wished to optimize some objective function (e.g. maximize economic efficiency). While this is certainly a legitimate analytical exercise, it should be kept conceptually distinct from the use of optimization in a purely descriptive context.

1.1 Advantages of Optimization-Driven Simulation

For large, complex, integrated systems, simulation models that optimize operation and allocation of water within each time-step by operational priorities have become the major simulation approach. Models of similar approach include ACRES (Acres Engineering), AQUATOOL (Spain), MODSIM (Colorado State U.), OASIS (Hydrologics, Inc.), WASP (Australia), and WEAP (Tellus Institute). Priority-based simulation models with optimization engines have become widespread in part because:

- The models are simpler to develop, comprehend, and modify.
- Their software is easier to upgrade, since the data set describing the system and its operating policies is substantially separate from the software code.
- Data are easier to update and modify, since changes require little or no software changes.
- Priority-based operations are a common basis for water rights and operating policies.
- Priority-based operations are relatively easy to explain.

The major exception to this technological trend in simulation modeling is to use more traditional procedural operating rules in simulation models with a graphical user interface for primarily flood control operations (HEC-RESSIM) or for exploratory study of large systems or detailed management of relatively small systems (Stella-type models).

Similar to several of these systems, CALSIM II allows specification of objectives and constraints in strategic planning and operations without the need for reprogramming of complex models. The CALSIM II authors developed the English-like WRESL (Water Resources Engineering Simulation Language) as an intuitive means of defining the objective function and constraints for a mixed-integer linear programming model, similar to the OCL (Operational Control Language) used in OASIS and the Policy Editor employed in RiverWare. In MODSIM, the optimization model is formulated directly through the graphical user interface with no need for a modeling language, but with supplemental features of the optimization defined through the PERL scripting language. WRESL allows planners and operators to specify targets, objectives, guidelines, constraints, and their associated priorities, in ways familiar to them. WRESL provides simple text file output that is converted to FORTRAN 90 code by a parser-interpreter program, whereas PERL is fully embedded in the network optimization code. Both modeling systems are data centered, meaning that model operation is controlled solely by user specification of input data rather than hidden rules or hard-wired data structures.

CALSIM II, OASIS, RiverWare and MODSIM are similar in that all use a high level language with syntax and logical operators; are written to simple text files which are subsequently parsed and interpreted; use rule-based or IF-THEN-ELSE conditional structures; are designed to be easy for planners and operators to use without the need for reprogramming; allow adaptive and conditional rules which are dependent on current system state variable information; include constructs for assigning targets, guidelines and constraints, along with their associated priorities; and include a goal seeking capability. CALSIM employs a mixed integer linear programming solver for repeated period by period solution that is less efficient computationally than the network solver employed in MODSIM, ARSP, REALM and Ribasim.

Unfortunately, unlike these aforementioned modeling systems, CALSIM lacks a comprehensive graphical user interface for constructing and editing the river basin system topology. CALSIM II would be greatly enhanced if, similar to RiverWare, IRAS, and MODSIM, objects representing features of the basin such as reservoirs, canals, and river reaches, could be created on the palette of a graphical user interface by simply clicking and dragging various icons for the objects to the display. The objects are instances of various classes that share certain common characteristics, and each object contains its own physical process methods and associated data. We believe that complaints concerning the complexity of using CALSIM II would be greatly reduced with development of such an object-oriented graphical user interface.

2. Comparative Strengths and Weaknesses

2.1 Some Prominent Strengths

CALSIM II has important strengths as a general inter-regional operations planning model, particularly compared with available alternatives and its predecessors. The primary strengths include:

- Coordination of Federal and State Interests A unique aspect of CALSIM II is the high degree of cooperation between Federal (i.e., U.S. Bureau of Reclamation) and State (i.e., California Department of Water Resources) interests in its development. This kind of cooperation is rare, and in fact this may be the only such example of such coordination for a system of this scale and complexity. Although it is clear that DWR staff have taken the greatest degree of responsibility in the planning, development, coding, testing and application of CALSIM II, it is also clear that USBR staff have also played an important role. CALSIM II can provide a showcase for other states as to what can be accomplished with Federal and State cooperation for river basin management.

- Consensus model. CALSIM II is the official joint modeling environment of the State and USBR. This includes a common schematic, hydrologic representation of the system, common set of facility capacities, and common representation of system operating policies. This saves a lot of unproductive bickering and helps all parties improve representations, rather than compete over representations.

- Common effort. The joint development of CALSIM II by USBR and DWR has provided more focused and effective use of resources and expertise than previous development of agency-specific models. CALSIM II development has also involved other agencies and consulting expertise more than pervious models of this system.

- Data-driven model. CALSIM II is a rather data-driven simulation model with an optimization engine. This modeling approach provides:

    a. much greater flexibility than its predecessors and traditional water resources simulation approaches.
    b. a promising framework for improving transparency, data, and model documentation, compared to other approaches.

- Public domain. The model and data are substantially in the public domain, facilitating transparency and adaptability for California's decentralized water system. Ongoing software development efforts will improve CALSIM in this regard.

- Steady improvements. Data improvements have been steadily pursued following the adoption of CALSIM II, although deficiencies remain widespread.

- Improved Delta water quality representation. Although problems appear to remain, the model developers have made substantial gains in representing Delta water quality operating criteria and performance.

- Better groundwater representation. Efforts to better include groundwater and non-CVP-SWP project operations are good efforts in the right direction, and merit continuation and expansion.

- Benchmark Studies. The development of documented benchmark studies seems to have resulted in significant model improvements and aided in the development of comparative model applications. Such exercises should be continued and improved.

- Long-term vision. The vision of a more transparent and publicly available model that can be employed by those outside the major agencies is excellent. This is a major change in direction, and achieving this vision will require adjustments over time. Often, these adjustments will be externally driven. Externally-driven improvements are a price of success and evidence of success for modeling policy that is open and public.

Few, if any, modeling organizations in the country have consistently done as good a job on model development and application for such a large, complex, and controversial system as the modeling group which developed CALSIM II. They are to be commended for their work to take California water modeling beyond past "closed shop" practices in favor of the development and dissemination of modeling capabilities that are more relevant to California's current water management problems. Most areas and suggestions for improvement noted below are meant to aid the model developers in moving further and faster in the direction they are already heading.

2.2 Some Prominent Weaknesses

The strengths and weaknesses of CALSIM II are not only technical (software, data, and methods), but also are institutional regarding how this model has been developed and employed. The administrative setting and objectives of model development and application are important, and difficult to manage. Alas, the management/policy problems of a system change frequently, while data and modeling capability change more slowly, and effective administrative structures change very slowly, if at all.

- Inadequate data development and management are principal shortcomings of CALSIM II. There has not been a sufficiently systematic, transparent, and accessible approach to the development and use of hydrologic, water demand, capacity, and operational data for CALSIM II. This problem extends beyond inadequate documentation and has led to controversy, confusion, and inefficiency in application of CALSIM II.
    a. Inadequate data management steepens the unavoidably difficult learning curve inherent for a complex system. Data have mostly been considered a "back room" activity of a few experienced experts. Retirement, promotion, or departure of these experts has left many gaps in knowledge and created difficulties for re-developing data for newer policy and planning problems.
    b. The administration of data development is fragmented, disintegrated, and lacks a coherent technical or administrative framework. Data required by CALSIM II are developed by several administrative units, without systematic technical vision or quality control for modeling purposes. Within DWR, different groups develop hydrologic and water demand data under different Deputy Directors, without effective coordination. This division must be overcome for a coherent data and analytical framework to be developed and implemented.
    c. In many cases it appears that water use and other hydrologic data inputs to CALSIM II are based on data collection and analyses that took place during the 1960s when DWRSIM and PROSIM were being constructed. It is important to ensure that data used for CALSIM II are up-to-date and consistent with the best current information

- The expertise and insights of many in local agencies, system operators, and consulting firms have not been prominent in the development of CALSIM II. For such a system with many hundreds of local experts, this is somewhat unavoidable, especially early in model development. Periodic re-examinations of how each area in CALSIM II is represented, in consultation with local agency and consulting experts, might overcome these technical shortcomings, and create and maintain a broader technical, user, and credibility base for CALSIM II. Active involvement of local agencies in CALSIM II development and applications would be much easier with better data management, and would be rewarded with a broader base of CALSIM II expertise and enhanced model credibility.

- Compared to the current CALSIM II, any central operations planning model for California water management should be:
    a. Expanded in geographic scope to include major non-CVP-SWP areas, especially the Tulare Basin, the Colorado River, and southern California. Operations and demands in these regions seem increasingly important for CVP and SWP operations, and are important for the integrated operations of California's major local and regional water management agencies.
    b. Expanded in management scope to include local management options such as water conservation, reuse, water transfers, groundwater and conjunctive use management, etc. These additional water management options are important for local, regional, and statewide water policy, planning, and management efforts and can have significant effects on CVP and SWP water demands.
    c. Made regionally modular, so smaller regional models can be run independently and tested locally, with boundary conditions consistent with the larger model.
    d. Made modular in terms of hydrologic, water management, and water demand processes, allowing better development, comparison, and updating of hydrologic and water demand process models. Agricultural, urban, environmental, and other water demands should be represented more directly, and explicitly. Groundwater should be represented and operated more explicitly. Land use based local hydrology and water demand approaches might be implemented in such standardized modules.
    e. Subject to a systematic model and data testing regime and continuous quality improvement program. As the problems of California water change, different and greater demands will be placed on analytical capability, requiring an essentially continuous testing, re-testing, and improvement of data and models. This might parallel a continuous review of local representations and data involving local agency and consulting experts.
    f. Financed on a broader base, by more than the CVP and SWP projects. Increasing use of CALSIM II is being made by local, regional, State, and Federal agencies interested in developing bilateral or multi-lateral water transfers or projects, which incidentally involve the CVP and SWP. To develop inter-regional modeling capability needed to integrate these activities at local, regional, and inter-regional scales, more sustained funding and involvement from local and regional agencies is needed. In effect, local and regional agencies have been "free riders" on CALSIM

II's analytical capabilities, and it is not necessarily a good bargain for them. Everyone should benefit from broader technical and financial participation.

g. Capable of analyzing a wide range of scenarios. More capability is needed to examine various long-term scenarios with respect to hydrologic, water demand, and operational uncertainties in the future. There also needs to be a better capacity to accommodate other approaches to representing hydrologic uncertainty and variability besides simply simulating 70-plus years of record.

- Input data and its development. Important aspects of CALSIM II rest upon the representations of other models of Delta hydrodynamics and water quality, water demands, and groundwater. The credibility of CALSIM II also rests on testing these models that send important data/representations to CALSIM II, and documenting them adequately. These models include:
  a. CU Model and SIMETAW: The consumptive use model and the newer SIMETAW model, used to develop hydrologic inputs and estimate return flows, also require testing and more explicit documentation. The underlying data for these models also need more systematic, standardized, and transparent treatment.
  b. DSM2: Representation of the Sacramento-San Joaquin Delta will always be important and prone to controversy, given the prominent importance of Delta flows and water quality for the operation and planning of California's water system. The difficulties of representing the Delta in operations and planning models are compounded by the tidal nature of the Delta, which usually implies a need for shorter time-steps. Representation of Delta water quality constraints currently falls heavily on an ANN method within CALSIM II. This ANN is calibrated (trained) based on a hydrodynamics model, DSM2. Thus, controversies regarding Delta representation in CALSIM II are likely to lead to questions of the adequacy of DSM2. The transparency and testing procedures valuable for establishing the credibility and limitations of a Central Valley operations model would also seem to apply to DSM2, or any other Delta hydrodynamics-water quality model. Tests of methods used to represent small-time step phenomena with larger time-steps (e.g., "partial month standards") should be tested in a forum that would give the approach credibility and where its limits could be developed, discussed, and documented.
  c. CVPM/CALAG/LCPSIM/IWR-MAIN: Representations of water demands in CALSIM II rely heavily on other models, particularly CVPM and eventually CALAG for agricultural water demands and LCPSIM and eventually IWR-MAIN for urban water demands. Thus, these models also will attract attention, and will probably require the same types of testing, transparency, and documentation suggested for DSM2 and CALSIM II. Many water contractors of the CVP and SWP also have internal water sources (groundwater, water conservation, and water reuse) and side contracts with other agencies to supply water that can increase or decrease (at different times) their water demands from the CVP and SWP contracts and from the demands estimated from CALAG and IWR-MAIN types of models.
  d. IGSM /CVGSM: Water users in California rely on groundwater as a water source and as the major source of over-year drought storage. Groundwater is also being increasingly used and looked-towards as a source of storage as part of conjunctive use schemes, and water transfer and market schemes. Thus, representation of

groundwater in the system is important, and probably should be expanded considerably. The representation of groundwater quantities, storage, and recharge and pumping capability will also attract attention from interested and critical parties. Thus, the IGSM/CVGSM modeling efforts of DWR and USBR should include the same types of transparency, documentation, and testing suggested for CALSIM II.

e. Agricultural demands: _Agricultural demands in the model are estimated by an external modeling system (CU model). Staff noted that the estimation methods being used are include out of date information on agricultural cropping patterns and irrigation technology, both of which result in inaccurate estimates of agricultural water demands. This estimation process needs to be revised and updated to include current information on an ongoing basis. The methodology needs to be improved to include economic factors in the estimation of cropping decisions and water demands. In many case, the preferred spatial scale for the economic modeling of agricultural water demand is going to be the individual irrigation district rather than very broad areas containing multiple quite heterogeneous districts.

- CALSIM II is currently awkward to apply for broader State and CVP-SWP policy questions. Practically, the time needed to complete analyses is too long and CALSIM II does not explicitly represent many of the management options which policy makers are interested in investigating, evaluating, and orchestrating.

- More CALSIM II modelers are needed. Many water managers and policy makers across California look to CALSIM II for many purposes, and there is near-universal consensus that the application of CALSIM II is currently limited by a dearth of knowledgeable modelers. Current training by DWR and USBR on CALSIM software is useful, but clearly insufficient. To be a functioning and credible CALSIM II modeler one must understand both CALSIM software and the operational complexities of the system (which probably no one can know in its entirety). Improved model and data documentation is also essential here.

- Stakeholders and policy makers are poorly guided in how to interpret CALSIM II results. Not only must CALSIM II become more responsive to current planning and policy concerns and management options, but current policy makers must receive some education in the benefits and limits of such modeling for their purposes. This is a very difficult problem that will often involve the role assigned to modeling and model results within larger politically-driven policy making processes.

- Non-interpretation of model results is not helpful. Several recent DWR reports based on CALSIM II results have been considerable improvements over past practices in terms of presenting model results, discussion of the model, and examination of model performance in a historical context. However, often the studies have not contained the kind of written discussion and interpretation of results that would demonstrate that the authors have thought about the results and drawn conclusions in a realistic and self-critical manner. This detracts from the perceived credibility of the work and makes the study less informative for readers (most of who surely do not have the modeling background of the authors).

- <u>Some needs exist to improve CALSIM software</u>. These are well-known to the model developers and include:
    a. Elimination of the need for the FORTRAN compiler,
    b. A public-domain mixed integer-linear programming (MIP) solver,
    c. A graphical user interface, including ties to databases and GIS display if possible,
    d. Post-processing tools for users to help new users and broader application and scrutiny of CALSIM II results,
    d. Version control software and system (also a problem for model administration),
    e. Better data and database management software and protocols (this has great data management and administration implications),
    f. An ability to more systematically set objective function weights,
    g. More automated input and output data checking is needed to improve productivity in model application and quality control of modeling output. This would also facilitate use of CALSIM II by a broader range of modelers,
    h. Ability to access and employ sensitivity analysis information coming from the MIP solver to identify possible multiple optima and identify binding constraints and slacks,
    i. A debug version of the code where water can be added or subtracted at any location and time (at a great penalty) to quickly identify locations and times of model infeasibilities. (Prof. J. Lund has had great success with this approach to correcting infeasibilities in the CALVIN model of California for a network flow algorithm.),
    j. Time-step issues should be explored and evaluated comparatively. There are major drawbacks to shortening time-steps system-wide (run-time, data development, interpretability of results, etc.), but short time-step components within the model or other approaches might adequately represent short-period aspects of the system for many purposes.

There will be some who argue that CALSIM II is and should remain a model of only the CVP and SWP system. While this would be simpler administratively and financially, it seems technically and politically untenable. California's water system is being asked to operate in an increasingly integrated manner across local and regional scales, with multiple local water demands, supplies, and aquifers being coordinated with the operations of major aqueduct and storage infrastructure. Any model of the CVP and SWP systems must be responsive to this operational integration, either implicitly through better parameterization of local supplies and demands, or explicitly by widening the geographic and functional scope of the model.

3. Limitations, Uncertainties, and Impediments

3.1 Removal of Unnecessary Ties to DWRSIM and PROSIM

Much of the spatial detail employed in CALSIM II is a carryover from the previous DWRSIM model. This is particularly evident in the coarse delineation of watersheds and sub-areas, which may no longer be relevant for future applications of CALSIM II. It is recommended that all unnecessary ties to the previous DWRSIM and PROSIM models be removed in further development of CALSIM II.

3.2 Relative vs. Absolute Predictions

As noted in the Executive Summary, we are skeptical of the usefulness of the distinction between comparative and absolute predictions. To declare that CALSIM II is intended for comparative predictions and should not be used for absolute predictions is not a helpful or desirable strategy. Rather than embracing this limited view of what CALSIM II can be expected to accomplish, we recommend that model developers recognize the requirement for CALSIM II to provide absolute values. To satisfy this purpose, additional calibration of the model will be required to ensure that it provides a reasonably reliable depiction of how the California water system operates. In addition, data on model accuracy and the outcome of the calibration runs should be made available so that users can gauge the likely errors involved in using the model for their own particular purposes. Some methods for doing this and performing sensitivity and uncertainty analyses are contained in Appendix H.

Model uses should realize that model calibration and validation exercises can illustrate only how well the model can reproduce historical decisions and system behavior. Our ability to predict future policy decisions and the emergency responses to water shortages is clearly limited, thus decreasing the absolute precision of any model's predicted values of various system performance measures. Thus it is useful to distinguish between the ability of the model to reproduce correctly the physical operations of the water systems in California (which should be good), its ability to reproduce and anticipate decisions by the agricultural sector that determine the quantities of water the consume, and its ability to mimic historical and current water operation decisions by the CVP, SWP and other water management agencies.

In general, it appears that the developers of CALSIM II do not have a clear idea of how to define the scope of CALSIM II use and many of the applications are evolving in a reactionary manner. Model developers should identify clearly the desired uses for CALSIM II and then determine acceptable approaches for satisfying those desires. Developers should seek to improve data accuracy and overcome unrealistic assumptions to improve confidence in model results.

3.3 Hydropower

CALSIM II is currently greatly lacking in hydropower computations, which is an important of the federal CVP system. This should include risk-based power capacity evaluation, and possible incorporate the ISM (indexed sequential hydrologic modeling) method that the Bureau has used for many years in hydropower capacity analysis. Also, hydropower should not simply be an after-the-fact calculation, but explicitly included in the system objectives.

3.4 Daily operations

A great challenge awaits the developers as they attempt to adapt CALSIM II to daily operations. These challenges are primarily related to the impacts of routing on distribution of flows and scheduling of reservoir releases. Under the current period-by-period optimization structure over daily time increments, without appropriate consideration of routing there is the danger that the model will allow diversion of upstream flows to lower priority users, resulting in injury to higher priority downstream users in the following days where travel times exceed 1 day. The proper inclusion of routing in the daily operations requires some kind of look-ahead capability in CALSIM II, which is currently lacking. In addition, scheduling of reservoir releases on a daily basis creates difficult timing issues in order to minimize unnecessary downstream spills or shortages caused by routing and attenuation of upstream reservoir releases. Another complexity in moving into daily operations is that reservoir discharges now become head-dependent, whereas this can usually be ignored on a monthly time scale. This means that the maximum reservoir release in any day will be dependent on the head, and should be based on the average head over the day, which introduces the potential for time consuming iterative processes to deal with nonlinear relationships in discharge-head curves for any reservoir.

3.5 Groundwater model

Groundwater has only limited representation in CALSIM II. This resource is modeled as a series of inter-connected lumped-parameter basins. Groundwater pumping, recharge from irrigation, stream-aquifer interaction and inter-basin flow are calculated dynamically by the model.

The purpose of the multi-cell groundwater model is to better represent groundwater levels in the vicinity of the streams to better estimate stream gains and losses to aquifers.

In the Sacramento Valley floor, groundwater is explicitly modeled in CALSIM II using a multiple-cell approach based on DSA boundaries. For the Sacramento Valley, there are a total of 14 groundwater cells.

Currently no multi-cell model has been developed for the San Joaquin Valley. Instead stream-aquifer interaction is estimated from historical stream gage data. These flows are fixed and are not dynamically varied according to stream flows or groundwater elevation.

The approach to modeling groundwater in CALSIM II, a lumped-parameter “tank” model seems to be a reasonable approach. However, few details of this implementation were provided to the review panel, that it is not possible to assess its accuracy or reliability. Details of the calibration and verification activities performed to date should be carried out and reported for the groundwater tank model. The effect of using large size tanks should be assessed and the level of uncertainty in computed results reported. In addition, the effect of these uncertainties on CALSIM II calculations should also be assessed. The San Joaquin valley aquifers are not well represented in the tank model, but it is in the CVGSWM. The San Joaquin valley groundwater should also be modeled in CALSIM II.

Groundwater availability from aquifers is poorly represented in the model. This results from the fact that aquifers in the northern part of the state (Sacramento Valley) have not been investigated regarding storage and recharge characteristics. Thus, in the model, upper bounds on potential pumping from aquifers are undefined. This does not represent reality, since, if CALSIM II is used for statewide planning, it would allow pumping of vast quantities of water for export to southern parts of the state, something which agency staff claim is unrealistic.

Realistic upper bounds to pumping from any of the aquifers represented in the model need to be developed and implemented.

In addition, historical groundwater pumping is used to estimate local groundwater sources in the model. However, the information on the historical pumping is very limited, causing these pumping rates to be very uncertain. Better pumping information is needed and an analysis of the effect of this uncertainty on model results needs to be conducted.

In general, the level of representation of groundwater in CALSIM II is not reasonable from the point of view of the reviewers. This is due to several factors, perhaps the most important being the lack of information presented to the reviewers for their assessment. Another factor is the lack of data collected and analyzed by the State of California to properly account for groundwater resources in the Central Valley. These data are critical to an understanding of the availability of water in the state and the operation of the major water systems that supply water to agriculture and small municipalities in the Central Valley. Assumptions of unlimited groundwater resources in the Sacramento Valley are unfounded and unbelievable. Efforts should be taken to make reasonable estimates of these resources.

There are other approaches that provide reasonably accurate estimates of river-aquifer interactions and groundwater basin response, while not sacrificing computer time. The response function approach is a good example, whereby the CVGSM model is used to develop kernel functions describing this response. A similar approach is described in Fredericks, et al. (1998). These kernels may require readjustment as head conditions change in the basin, but they provide a more accurate prediction tool and are easily incorporated in the MIP model since they apply a linear superposition assumption and retain the linearity of the constraints in the model. A dynamically linked CALSIM-CVGSM configuration is not necessary for reasonably accurate solutions. If computer run time for CALSIM II is considered excessive now, it could only considerably worsen if this type of linkage is incorporated.

Soil moisture is not dealt with in a realistic manner and needs to be improved in applications where the model output might be sensitive to these assumptions.

3.6 Dynamic Variation of Priority Weights

A severe restriction in CALSIM II is the inability to dynamically vary the weights used to prioritize flow allocation in the system. It should not only be possible to dynamically vary these weights, but this variation should be conditional on the current system state, however that state (or states) is defined. In addition to dynamic variation of weights, more explanation is needed of the reservoir operating rules and how these rules are incorporated into CALSIM II. The description of operating rules used in the system is not very clear. For example, what kinds of hedging or shortage rules are used to mitigate the effects of drought?

3.7 Expanding Scope of CALSIM II

CALSIM II is a considerable advance on earlier models in that it fully incorporates both the State Water Project run by the Department of Water Resources and the Central Valley Project operated by the Bureau of Reclamation. However to be able to examine the full range of Californian water issues, it would be desirable that all components of the linked system should be incorporated in the model including the Friant system, the larger Tulare Basin, and southern California and its links to the Colorado River. Also because of the very important linkage between surface water and groundwater use, improvements should be made in this area particularly with regard to how that linkage affects demand for surface water and how access to groundwater reduces the economic impact of surface water restrictions.

When expanding the geographical scope of the model to include non CVP-SWP areas, as well as Southern California, a hierarchical, decomposition approach would allow development of separate models for these areas that can then be linked together through iterative processes. Otherwise, the CALSIM II model can become extremely unwieldy. Again, integration can still be achieved through appropriate iterative interaction between the regional models. In the same vein, it is also unnecessary to explicitly integrate water quality and detailed water demand/consumptive use models into the model structure. Iterative schemes involving successive estimation of water quality and other parameters can produce comparable accuracy at reduced computer run times, while reducing the complexity of the model.

The replacement of DSM2 with a neural network is consistent with reinforcement or machine learning methods which are increasingly being used to replace complex, computationally time consuming models employed in decision support systems. The complex models are only used to provide the data sets used for training the neural network. Current research at Colorado State University and elsewhere is using neural networks for groundwater surface water interaction and return flow computation to replace computationally expensive groundwater models.

3.8 Key Model Outputs

In the past, the primary purpose behind the development of CALSIM II and its predecessors has been the examination of the reliability of water supplied to the State Water and the Central Valley Projects. However it is clear that there is now a demand for a model that will provide a wider range of outputs including:

- Water supply reliability for all water users
- Demand for water by existing users
- Outflows to Delta
- Use of groundwater and the rate of depletion of aquifers
- Water quality in the Delta and in the San Joaquin River
- Indicators of ecological health in particular with regard to key fish species
- The value of hydroelectric generation.

Although the modules in the CALSIM II package currently address many of these areas, the recognition that all these outputs are important may necessitate some further model development and a greater degree of testing and calibration of these parameters.

3.9 Modeling Allocation, Accounting and Operating Rules

CALSIM II uses a system of weights and constraints to define the water allocation process and the operating rules for storage reservoirs. Unfortunately these do not accurately reflect how operators of the state and federal water projects behave in managing their complex systems. Ideally, CALSIM should both reflect how the operators behave and be accepted by them as a useful tool when considering their management alternatives. The failure to achieve this limits the usefulness of CALSIM to investigate the specific operating or accounting rules that are of interest to those operators. For example, CALSIM II was not used to test changes to the accounting and allocation rules that have recently been proposed by the Department of Water Resources and the US Bureau of Reclamation because the rules that were changed do not exist in CALSIM II.

4. Options for Improving CALSIM

4.1 Optimization Model and Run Times

Many of the complaints regarding using of CALSIM II relate to long run times, which is not conducive to sensitivity or uncertainty analyses. Since CALSIM II employs a mixed integer linear programming (MIP) solver, the usual sensitivity information available in linear programming solvers, such as dual variables and right-hand-side ranging, are not available. The problem is that small changes in right-hand-side constants or objective coefficients (i.e., weights on water allocation priorities) can produce large abrupt changes in model solutions. In this case, dual variables do not provide useful information for MIP problems. Sensitivity analysis can only be conducted through trial and error processes involving incremental adjustment of important weights, coefficients, and uncertain data inputs with subsequent repetitive execution of the model. In light of this, it is crucial that the MIP solver employed in CALSIM II is upgraded. Significant advances have been made in MIP solvers, as described by Bixby, et al. (2000), which are not reflected in the current XA solver utilized in CALSIM II. There have been many recent improvements to the branch and bound method which should be incorporated, and the LP solver itself can be improved with better sparse matrix analysis. As planned by the CALSIM II developers, removal of the need for use of the FORTRAN 90 compiler will also improve run times when changes in optimization model structure are required.

4.2 Confidence in the model

The usefulness of a computer model in water resource management is only as good as the confidence that the stakeholders have in the accuracy and reliability of the model and the trust that they have in the modelers. There are several factors that affect that confidence and a number of ways that confidence can be improved.

- **Documentation**

Producing documentation of models requires considerable resources to do properly and ongoing resources to maintain especially when model development is continuing. Typically documentation of any water resource model is poorly done. However, where there are external model users, as is the case with CALSIM II, it is important. The survey conducted by Ferreira et al (2003) indicated that many users of the model thought that documentation of CALSIM II was poor.

- **Seminars**

In the Murray-Darling Basin, seminars with key users and interest groups in which the operation of the model is described and discussed have proved to be useful in increasing confidence in models. The practicality of this approach will depend on the number and location of the prospective participants and the resources available to support the process.

- **Data**

A model can only be as good as the data that is used to develop and calibrate it. The agreement over an acceptable set of hydrologic data that occurred during the development of CALSIM II is a considerable advance. However, there appears to be a need to improve the collection and use of data on water diversions and return flows. Because of the close links between the surface water use and groundwater use there also is a need to have better information on the use of groundwater.

The models used to calculate the Local Water Supplies in the Depletion Study Areas depend on estimates of surface water use, crop evapotranspiration rates and water use efficiencies developed using data from the 1970's. Confidence would be improved if more recent data were available to check these estimates.

- **Calibration**

A very good way to improve confidence in a model is to calibrate it against historical data to ensure that the model output is able to reproduce the observed data. Calibration is the process of using the model to reproduce the historical behavior of the system and then fine-tuning the model so that the match between modeled and observed values improves. The calibration of the model assists in detecting errors in the model and the input data. It also enables a comparison to be made between the way that the operators actually manage the system and the way that the model assumes that the system is managed.

A further consequence of the calibration process is that the statistics of the match between modeled and observed values can be used as a reasonable estimate of the absolute accuracy of the model output.

It is legitimate in a calibration/validation run to incorporate changes to infrastructure, institutional or operational rules as they occurred especially if these changes are specified as input parameters to the model. This was done to a limited extent in the CALSIM II validation run with three regulatory periods modeled related to decisions made by the State Water Resources Control Board. It is also legitimate to incorporate growth in demand especially if that growth is described in a manner that is consistent with the way that demand is specified in the production run. Demand north of the Delta was specified in the validation run by inputting the historical crop areas.

A Calibration/Validation report should be very useful in demonstrating the accuracy of the model. However there are a number of elements in the CALSIM II validation run and the validation report which reduce that confidence including:

- State Water Project (SWP) demands south of the Delta were set at historical deliveries in years with no restriction and at the contractor's request level in restricted years. Neither of these pieces of information is available to a production run which calculates demand based on crop areas. Therefore the validation run does not provide reliable information on how well the model can represent these demands.
- The validation run omitted Article 21 deliveries. Although this omission will not affect the delivery of 'Table A' volumes south of the Delta, it will affect flow in the Delta and Delta water quality. Also, in the example model run presented in the paper by Draper A.J. et al (2003) which was supplied as part of the review, changes to Article 21 deliveries constituted the largest impact resulting from a change to the allowable pumping capacity at Banks between March and December. This suggests that the modeling of these demands is important.
- The DWR (2003) report produces estimates of SWP and Central Valley Project (CVP) deliveries south of the Delta but then adjusts them for changes in storage before presenting comparisons of those results with observed deliveries. This process merely checks that the model is preserving a water balance and does not present a legitimate validation of model deliveries.
- The report provides statistics on long term average deliveries and flows but no statistics on the fit for individual years. Additional analysis of the output would assist stakeholders to assess whether the estimate of water supply reliability and in particular the modeled volumes of water available in the most restricted years are accurate.
- In some instances, such as the examination of water quality in the Delta, the ability to accurately model monthly flows and deliveries will be important. The validation report contains no information that would enable the ability to model monthly flows to be assessed.
- A key model output is the water quality in the Delta. It would assist the validation of the model if a comparison of parameters such as the location of the X2 boundary was provided.

The users of CALSIM should recognize that models are a summary of what one believes to be true and important about a system. Validation is then an exercise to test how good that summary and understanding really is.

Appendix I contains brief descriptions of calibration modeling in the Murray-Darling Basin in Australia and in the State of Texas.

4.3 Assessment of the reliability of "delivered" water

An important recent application of CALSIM II which has drawn widespread attention is the "State Water Project Delivery Reliability Report. While this is an important step forward in the use of CALSIM for policy purposes, it highlights a number of issues, both conceptual and empirical, that need to be resolved in order to provide a more adequate assessment of the reliability of water supply in California.

First, it illustrates the need for sound calibration of CALSIM. The question being asked is not a comparative one – What are the consequences of changing some aspect of the system from X to Y? – but rather an absolute one – How does the system function at present? How often can users expect a shortage in deliveries of Z%?

Second, it highlights the fact any water system model such as CALSIM requires a blend of hydrology and behavioral analysis. To conduct a water balance, the model needs to know what deliveries are required by the customers of the given project, and what are the diversions by other user groups who extract water from the same surface or groundwater sources. These are fundamentally questions of economic and institutional behavior, not matters of hydrology. Therefore they cannot be dealt with by hydrologists alone. Like its predecessors, CALSIM tends to treat these as black boxes. The diversions by water users outside the CVP-SWP are taken as exogenously given, based on an assumed "level of development" and simplistic assumptions about the patterns of water use associated with that level of development. The deliveries required by the water users who are served by CVP-SWP are generally taken as given. For reasons explained below, both of these treatments are simplistic and unsatisfactory.

In CALSIM modeling exercises the level of development plays two different roles depending upon the context. In a simulation context, the level of development is used to represent hydrologic variability and uncertainty; in a calibration/validation context, it is used to reflect the actual historical demand for water withdrawals. These are very different purposes and it is important to keep them distinct. In most applications of CALSIM prior to the recent reliability study, the main focus was simulation and the representation of hydrologic variability. The chief purpose served by using 73 years of adjusted streamflow records was to represent the variability and uncertainty in the streamflow that one can expect to observe in any single year. Therefore, the calendar date of the record has no substantive significance, the (adjusted) streamflows for 1952 or 1982 are not being used to represented what happened historically in 1952 or 1982, but rather as an indication of the variation in streamflow that could be expected to occur next year, or any other year. In this context of simulating hydrologic variability, it makes good sense to apply the *same* level of development (i.e. the same pattern of water use) to every year in the sequence, rather than a series of different levels of development that vary with calendar time, because the streamflows represent alternative hydrologies that can occur in any given year.[1] The situation is different when one is conducting a calibration or validation

---

[1] This could be modified to allow for the fact that local weather conditions have a significant impact on irrigation (and urban) demands – e.g., farmers plant fewer acres of crops in a drought year. In that case, one could have different levels of water demand and extraction in different year *types;* but, these would all be keyed to the same overall level of economic development (e.g. the California economy in the 1990s). CALSIM II does not presently

exercise. In that case, one wants to represent the historical demands in 1952 or 1982 in order to compare what the model predicts with what actually happened. Therefore, in a calibration or validation exercise one wants the level of development to change each year in order to reflect the demand that occurred historically.

Both simulation and calibration/validation raise some other important technical issues. In the context of simulation, there are several different ways to generate a hydrologic sequence that is calibrated to a fixed level of development. One can use all 73 years for which data are available. One could use a subset of those years chosen either according to some deterministic rule or randomly. The subset could be oriented, for example, towards the extremes of the 73 sequence of annual records. However, the drawback of any approach based on sampling from the observed historical record is that it *understates* the full variability in streamflow that could be experienced in the future. The 73 years of record are drawings from a probability distribution the extremes of which extend beyond the minimum and maximum flows observed in the historical record. Relying on this record, therefore, understates the true minimum and maximum flows that might be encountered. In a reliability assessment exercise, one might want to take some steps to minimize the potential understatement of streamflow uncertainty. This could be accomplished by fitting a (parametric) probability model to the historical streamflow record and then sampling from the tails of the fitted distribution (Stedinger, 1981). The use of statistical models of streamflow variability could be considered in future applications of CALSIM to assess delivery reliability.

The assessment of delivery reliability requires that particular attention be given to the definition and measurement of the water users' demands. In this context, the user's demands play two roles: they affect the definition of "deliveries" and they influence the assessment of "reliability". With respect to deliveries, CALSIM II considers water to be delivered whenever it has the water irrespective of the ability of a contractor to use the water or to store it; The reality is that, if the contractor does not have a demand for the full quantity of water and is not able to store the excess, that amount will not be delivered. Therefore, the calculation of deliveries would be flawed. Furthermore, reliability cannot be assessed without reference to demand. Stating that a water supply system can deliver 100 acre feet in a wet year but only 70 acre feet in a dry year is useful only if one knows what the demands will be in wet and dry years. The implications are quite different if the user needs 105 acre feet per year than if he or she needs 65 acre feet per year. Thus, the users' demands should serve as the norm against which reliability is assessed. Instead, the recent reliability report uses the so-called 'Table A' water amounts as the norm for assessing deliveries to SWP contractors. This does not seem to be a satisfactory approach because there is no presumption that the Table A amounts, negotiated in 1960, measure the actual demands of SWP contractors in any particular year. The actual demands of the individual contractors will be influenced by how much storage they have, what access they have to other surface water or groundwater, and the demands of the farmers they serve to plant crops and apply water. Without accounting for these factors, it is difficult to generate a meaningful assessment of supply reliability.

---

consider the impact of annual weather conditions on demands. In order to model water demands accurately in a year, the climate conditions would be linked to the flow conditions to provide an input set for a particular year.

The assessment of reliability should ideally go beyond a comparison with quantities demanded to incorporate the notion of a loss function. If a user has a demand for 100 acre feet and can only receive 90 acre feet in one scenario and 80 acre feet in another, while the shortfall is twice as large in the second scenario the actual *consequences* of the shortfall to the user, in terms of lost profit or higher cost, might be more than twice as large. To assess the economic value of reliability, or the economic cost of a lack of reliability, one needs to be able translate shortages into monetary losses. To accomplish this, the warning time provided and the delivery shortfalls from CALSIM would need to be processed through an economic model of the value of water to different SWP contractors.

Because water users face difference demands and have access to different sources of supply, when assessing reliability it is unhelpful to aggregate all contractors and simply present the results in terms of total annual project deliveries, as was done in the report. Precisely because of the potential non-linearity of the loss function, a given aggregate shortfall can have different consequences when distributed differently among the individual contractors. A similar observation applies to the temporal distribution of delivery shortfalls across the year. It is unhelpful to aggregate supply system deliveries into an annual total, as done in the report. For a user to be able to obtain 100% of his or her demands in the period from March to May but only 60% in the next three-month period from June to August has different consequences than being able to obtain 80% in each of the six months. Furthermore, for both agricultural users and many urban users, major decisions affecting water use have to be made in the spring. They are based on the expectation around March about the amount of water that will subsequently be available for delivery during the summer months. What matters to these users when assessing supply reliability is the amount of water they can expect around March to be delivered over the summer, rather than the ultimate total delivery.

For both reliability assessment and also model calibration/validation, it is important to avoid excessive aggregation when describing shortfalls between demand and supply, or deviations between model predictions and actual outcomes. In regression analysis, it is the convention to measure the goodness of fit of a regression equation not by the average deviation but rather by the sum of the squared deviations. In ordinary least squares regression, by definition the average deviation is always zero (that is to say, the average of the predicted values of the dependent variable always equals the average of the actual values) regardless of how well or badly the regression equation fits the data. The average deviation thus provides *no* information regarding the goodness of fit; by contrast, the sum of squared deviations or the sum of the absolute values of the deviations are sensitive measures of goodness of fit. Although the calibration of CALSIM is not an exercise in least squares regression, the same general principle applies. To judge whether the model is doing a good job, the goodness of fit should be measured by reference to the disaggregate results and not simply by the overall average deviation.

Additional comments on the 2003 CALSIM II Validation Report are contained in Appendix F.

5. Managing CALSIM Development and Applications

The costs of not continuously and substantially improving our analytical capabilities are political (in terms of continued controversy and diminished agency credibility), economic (as inferior system performance for agricultural and urban water users), environmental (in terms of inferior environmental system performance), and financial (lawyers and policy consultants are more expensive than engineers and scientists).

CALSIM II is a substantial improvement over its predecessor models, DWRSIM and PROSIM, with a great deal more flexibility, transparency, and potential than these earlier models. The modeling team for CALSIM has identified an exciting and relevant vision of how modeling should be done for this complex and difficult system in the coming years. However, implementation of this vision in a coherent technical manner that leads to both technical and stakeholder credibility will be a difficult process, requiring financial and institutional support if this kind of capability is to be developed and sustained.

To accomplish these objectives CALSIM II developers need to be in an institutional position where they can see the model more as "outsiders" view it. This would allow them to be more responsive in supporting the credibility of their work and the relevancy of their tools and results to the broad range of current water management problems. As such CALSIM II should no longer be solely responsible to CVP-SWP managers, but should be responsible to a broader range of technical managers from additional interests, reflecting its current and prospective uses.

It would be imprudent to manage a state's finances, a business, or a retirement plan without quantification – quantification in such matters is necessarily imperfect, but necessary nonetheless. While shortcomings have been identified in CALSIM II, it would be similarly irresponsible to manage California's water budget without carefully-interpreted quantification. Progressive and continuous improvement in our quantitative understanding of California's water system provides a common basis for improving its performance for all interests.

One possible means of maintaining control of the quality of particular versions of CALSIM II and accompanying models used for SWP-CVP planning and management decisions is to create an interagency modeling consortium (IMC) consisting of DWR, USBR, and persons from other stakeholder organizations if they are interested and want to participate. This consortium would be responsible for maintaining a toolbox of 'acceptable' models for 'official' use by the agencies and contractors.

IMC responsibilities and authority could include:

- Prioritize, coordinate, and provide consistency, technical guidance and oversight for all modeling applications,
- Approve model selection and insure that each requested application is carried out using the most appropriate model(s) and input data,
- Provide or otherwise insure documentation of the modeling process itself as well as the modeling results,

- Insure that the results are expressed and made available in a way such that others can understand and benefit from that modeling application, as applicable.
- Implement peer reviews of models and their applications as deemed appropriate.

To help meet their responsibilities the IMC will need to establish, publish and implement some procedures for insuring the quality of the entire model development and application process. They will need to identify among all the models that might be used, which are the most appropriate to address each of these separate groups of model applications. They must identify various models, i.e., establish a model toolbox, from which clients can choose the one that best meets their needs (or perhaps argue that another model should be added to the toolbox). The IMC will also need to maintain model documentation and provide for peer reviews of any model, its documentation, and/or its use in a project.

Further suggestions and discussion on the creation and operation of a possible IMC for model development and application, as well as for managing peer reviews of both the models and their applications, are contained in Appendix E.

6. Recommendations for Future Use, Development, and Application of CALSIM II

The most concise recommendation we might make would be to fix the shortcomings beginning with what are considered the most serious, and proceeding to those that are less serious, taking into account the time and other resources needed to address each weakness. However, we believe it is more useful to suggest ideas on how to systematically address both present shortcomings and those likely to emerge as stakeholders' quantitative understanding of California's water system and its problems continue to evolve.

6.1 Model development and support consortium

As discussed in the previous section and in Appendix E, it might be useful to explore creation of a broader interagency modeling consortium for developing operations planning models for California. The joint DWR-USBR development strategy used for CALSIM II has shown some notable successes, and should be expanded to include additional parties and sources of expertise. Such a consortium might include staffs from several agencies (DWR and USBR, as well as potential members from MWD, KCWA, CCWD, and other agencies), NGOs, some consultants, and universities. Such a model development forum would:

a. Bring a wider range of expertise to bear on model development problems.
b. Facilitate having more agencies involved in supporting model development with expertise and financial resources.
c. Better enable model developers to see the model as "outsiders" see it.
d. Potentially improve contracting for model development and testing.

e. Take model development and testing outside of the explicit agency framework; a broader consortium should be more conducive to self-critical and transparent technical practices.
f. Provide a common training ground for agency, NGO, and consulting staffs to become effective modelers, broadening the talent base for technical work in California.
g. Reduce impediments to model development and testing arising from current State budgetary and personnel hiring problems.

Many of the questions, concerns, and problems mentioned in the user community interviews could be addressed well in such a distributed model development, testing, and support framework. It would still be necessary for each stakeholder group and agency to maintain its own modeling staff, but these would be partially shared in an interagency modeling consortium.

The governance and finance of such a consortium would be difficult and would probably require a steering committee or governing board, but any resulting model(s) would have broader credibility and a broader and deeper technical base.

In the immediate term, a users' group should be formed and the formal listing of model development activities should be posted on the web, including short descriptions of each development activity and contact information.

6.2 Quality Control Program

The DWR and USBR modeling team (or a broader model development consortium) need an explicit quality control program. Such a program should include a variety of activities:

a. periodic external reviews on the broad modeling program
b. specialized external reviews of model products and applications
c. a standing (or sitting) external technical advisory body
d. software engineering and maintenance
e. a regime of model testing
f. model and data documentation
g. data development and management
h. user group activities
i. local agency and interest involvement
j. model, data, and documentation accessibility (including web site use).
k.

Such a quality control program would benefit from deep consultation with stakeholders and the broad community of water technical people, perhaps via the California Water and Environment Modeling Forum (www.cwemf.org).

6.3 A Training Program

DWR, USBR, and assorted agencies and consultants should establish a more formal common regimen to train new CALSIM II users in both CALSIM software and the complexities of actual system operation. All these groups currently rely on a relatively small pool of perhaps a dozen knowledgeable CALSIM II users and all proclaim a need for many more capable users. A training regimen consisting of current CALSIM II training classes, supplemented by additional training in software application and system operation and apprenticeships or rotations through operations and model development shops would be useful to all concerned. The entire water community would benefit from having such expertise being widespread. Having widespread CALSIM II modeling expertise also makes explaining CALSIM II and its results easier. This might be an appropriate activity for a model development consortium.

6.4 Extend Improvements in Modeling Practice to Supporting Models

CALSIM II is at the center of a web of additional models used by DWR, USBR, and other agencies to prepare inputs for CALSIM II and post-process outputs from CALSIM II.

Delta controversies and difficulties of representation seem endemic to problems of modeling Central Valley operations. The technical basis for representations of Delta operations and water quality performance requires a similar level of transparency and testing to avoid this becoming a “weak link” in the Valley-wide operations planning model. Since so much is based on the DSM2 Delta model, documentation of fairly strenuous tests of the DSM2 model are highly desirable. This would provide a firm foundation for the use of ANN or other approaches for summarizing DSM2 behavior in an operations model. Similar documentation, testing, and development are desirable for the other models mentioned above which provide data for CALSIM II (CVGSM/IGSM, CVPM/CALAG, IWR-MAIN, LCPSIM, CU model, and SIMETAW).

6.5 Hydrologic Data and Data Development

An effort should be made to step back and perhaps re-define a more systematic and solid basis for developing hydrology for water management models of California’s inter-tied water system. Currently, several efforts exist to develop surface or groundwater hydrologies for parts of the Central Valley (sponsored by DWR-USBR, USACE-Sacramento District, USEPA, USGS, CALFED, local agencies, etc.). An effort should be made to broaden the range of hydrologic expertise involved in hydrology data development for management modeling of California’s inter-tied water system, and establish a consistent and high, but reasonable, standard of documentation and testing for developed data and any underlying hydrologic models. Establishing such a standard of documentation and testing would make existing hydrologic studies more accessible and useful for future studies and encourage the comparison and further development of existing representations of the system’s hydrology.

6.7 Performance-Based Optimization

Performance-based optimization should be added to CALSIM’s capabilities; it would not be difficult in terms of software or data, and would add much greater ability to explore and seek improvements in management within a complex system. The multi-period optimization approach being developed (CAM) is an operations-oriented first step in this direction, but could be expanded without great difficulty.

For large-scale water resource systems of great complexity and many options for system management, it is often difficult to find "optimal" operations with simulation modeling. There are simply many myriads of decision options and combinations of options, which theoretically each require a simulation model run – which would be prohibitive in terms of analysis cost and time. In such situations, performance-based optimization models, such as those seeking maximum economic performance, can offer useful insights as to where to look for improving system operations and management. Metropolitan Water District of Southern California (MWD) and San Diego County Water Authority (SDCWA) employ performance-based optimization modeling of parts of California's water system to gain strategic insights for planning and management. An economic-engineering optimization model has been developed for California and, despite significant limitations, shows several insights for California (CALVIN), suitable for identifying promising operational and management strategies worthy of more detailed analysis (Jenkins et al. 2001; Draper et al. 2003; Jenkins et al. 2004). The CALSIM II modeling approach could easily be adapted to provide greater functionality to this type of performance optimization. Having performance-based optimization capability together with a compatible simulation model for more detailed analysis and trade-off evaluation could greatly improve the capability of California's water community to explore and develop promising and creative options for improving operations, facilities, and overall system management.

6.8 Modular and Layered Versions of CALSIM II

Speedier versions of CALSIM II are needed for operations planning and integrated water planning studies. Such versions would be regional modules of CALSIM II (for regional studies) or explicitly aggregated system-wide models from the most detailed CALSIM II schematic for system-wide or statewide studies. Both approaches would simplify the model for particular purposes, yet be tied to a common detailed schematic and detailed hydrologic, operations, and water demand data sets.

Geographically modular or aggregated system-wide versions would allow additional local and regional water management options to be represented for particular operations and policy planning purposes and allow users to more quickly explore and develop operating policies. The final runs from such integrated or exploratory studies could then be evaluated using a more detailed and complete version of CALSIM II.

Modular regional models might represent regions with relatively few inter-ties, such as: Sacramento Valley, Delta and eastside streams, San Joaquin Valley, San Francisco Bay Area, Tulare Basin, and Southern California (DWR's South Coast and Colorado River hydrologic regions). (We have had good success with the CALVIN model of California with 5 modular regional models, which combine to form a system-wide model. These geographic sub-models greatly improved quality control in model development, work flow and data checking, and identification of problems in the model.)

6.9 Model Calibration and Testing

Many approaches exist for model calibration and testing (Modeling Forum 2000). Calibrating a planning model oriented to operations in an uncertain and distant future is always challenging. For a model that serves many uses (including policy-urgent uses unforeseen by developers), use-specific testing will often be impossible within a responsive time frame and budget. Such unavoidable situations call for more thorough, general, and well-documented model calibration and testing than would otherwise be needed.

For the model to have technical credibility, stakeholder credibility, and to serve the kind of training and reference function needed for the water management community, a systematic and coherent means of setting parameter values in the model and documenting these values is needed. Similarly, a systematic self-critical means of testing is needed for a model to establish and retain credibility, and have defined limitations, for a range of applications.

A potentially excellent resource for model testing is comparisons of seasonal operations planning CALSIM II model runs with recent years' seasonal operations, as done by actual operators. Similarly, system operators could scrutinize historical simulations, such as those in the recent November report, for systematic differences from operating practice. Such comparison with operator policies and philosophy could also be performed with SWP or CVP delivery reliability estimates. Such comparative analyses would both help define the likely (and unavoidable) differences between actual and modeled operations and water deliveries and identify potential opportunities to narrow such differences.

Credibility arises, in part, from demonstration that problems and limitations are systematically identified and addressed or considered in model development and in making and interpreting model runs. This can be accomplished by use of documentation, metadata, written guidance, and protocols and logs for identifying model problems and recording model improvements.

Given present and anticipated uses of CALSIM II, the model should be calibrated, tested, and documented for "absolute" or non-comparative uses. This is what many applications require today and will be increasingly desired and required in the future. Maintaining the traditional "comparative-only" use of CALSIM II is undesirable if the model is to be useful for the CVP and SWP systems, the operations of water contractors, or for statewide planning purposes.

6.10 Documentation of Model Improvements

Along with better documentation of model versions, logs of data and model improvements and "bug fixes" should be maintained. Explicit protocols and records for identifying and correcting modeling errors and problems would enhance the credibility of the modeling effort with technical people and policy makers. Such protocols also provide an internal aid to staff and staff development in modeling. I understand that this kind of record-keeping is done, but the precise form of, nature, and extent of this record-keeping is unclear. It would be useful and reassuring to stakeholders and policy makers to know that this kind of record-keeping of the software and data was being done.

6.11 Better Model Integration in Decision-Processes and Stakeholder Education

Greater aid should be given to interested parties and decision-makers who must work with the unavoidable limitations of any model. If possible, a document should be prepared for stakeholders and interested parties outlining the model, summarizing the model's primary limitations, and providing guidelines for interpreting model results. Those developing policy-making forums and processes should thoughtfully incorporate computer models in these processes in ways that do not assume model omniscience, or otherwise place too great or exclusive a reliance on model results.

Models and model results will never be perfect. If models are to be important for planning and policy-making, they be must be presented and used in ways that enlighten policy-makers more than they add confusion and controversy to already difficult situations, if possible.

7. References

Bixby, R., M. Fenelon, E. Rothberg, and R. Wunderling, (2000), "MIP: Theory and Practice – Closing the Bap," in, *System Modelling and Optimization: Methods, Theory and Applications*, M. J. D. Powell and S. Scholtes (eds.), pgs. 19-49, Kluwer Academic Publishers.

Carnegie Mellon University, (1994), The Capability Maturity Model: Guidelines for Improving the Software Process, Addison Wesley, Boston

Draper, A.J., A. Munevar, S.K. Arora, E. Reyes, N.L. Parker, F.I. Chung and L.E. Peterson, (2003), 'CalSim: A Generalized Model for Reservoir System Analysis', Paper supplied to November 2003 CALSIM II Review.

Draper, A.J., M.W. Jenkins, K.W. Kirby, J.R. Lund, and R.E. Howitt, (2003), "Economic-Engineering Optimization for California Water Management," *Journal of Water Resources Planning and Management*, ASCE, Vol. 129, No. 3, pp. 155-164, May

Disco, C. and J. van den Ende (2003), "Strong, Invincible Arguments?: Tidal models as management instruments in twentieth-century Dutch coastal engineering," *Technology and Culture*, Vol. 44, July, pp. 502-535.

DWR (2003), 'CALSIM II Simulation of Historical SWP/CVP Operations', California Department of Water Resources Technical Memorandum Report November 2003.

DWR (2002), 'The State Water Project Delivery Reliability Report 2002', California Department of Water Resources Bay-Delta Office Report 2002.

Ferreira, I.C., S.K. Tanaka, S.P. Hollinshead and J.R. Lund (2003), 'CALSIM II in California's Water Community: Musing on a Model', Draft Report to November 2003 CALSIM II Review, 18 September 2003.

Fredericks, J., J. Labadie, and J. Altenhofen, (1998), "Decision Support System for Conjunctive Stream-Aquifer Management," *Journal of Water Resources Planning and Management*, ASCE, 124(2)

Hansen, E., (1994), "WEAP – A System for Tackling Water Resource Problems," in *Water Management Europe: An Annual Review of the European Water and Wastewater Industry*, Stockholm Environment Institute, Stockholm, Sweden

James, B., (2003), "REALM: Simulation Software for Water Supply Systems," Department of Natural Resources and Environment, Victoria, and Victoria University of Technology, Victoria, Australia,

Jamieson, D. and K. Fedra, (1996), "The WaterWare Decision-Support System for River Basin Planning. 1. Conceptual Design," *Journal of Hydrology*, 177(3-4), 163-175

Jenkins, M.W., A.J. Draper, J.R. Lund, R.E. Howitt, S. Tanaka, R. Ritzema, G. Marques, S.M. Msangi, B.D. Newlin, B.J. Van Lienden, M.D. Davis, and K.B. Ward, (2001), "Improving California Water Management: Optimizing Value and Flexibility," Center for Environmental and Water Resources Engineering Report No. 01-1, Dept. of Civil and Environmental Engineering, University of California, Davis, CA, http://cee.engr.ucdavis.edu/faculty/lund/CALVIN/.

Jenkins, M.W., J.R. Lund, R.E. Howitt, A.J. Draper, S.M. Msangi, S.K. Tanaka, R.S. Ritzema, and G.F. Marques, (2000), "Optimization of California's Water System: Results and Insights," *Journal of Water Resources Planning and Management*, ASCE, in press.

Johnson, S.A., J.R. Stedinger and K. Staschus, (1991), Heuristic Operating Policies for Reservoir System Simulation, Water Resources Research, 27(5), 673-685

Modeling Forum (2000), "Protocols for Water and Environmental Modeling," California Water and Environment Modeling Forum, http://www.cwemf.org

Labadie, J. and R. Larson, (2000), "MODSIM: Decision Support System for River Basin Management: Documentation and User Manual," Colorado State University and U.S. Bureau of Reclamation, Ft. Collins, CO, May

Loucks, D.P., M. Taylor, and P. French, (1996), "IRAS – Interactive River-Aquifer Simulation Model, Program Description and Operating Manual, Cornell University, Ithaca, NY

Randall, D., C. Cleland, C. Kuehne, G. Link, and D. Sheer, (1997), "A Water Supply Planning Simulation Model using a Mixed Integer Linear Programming Engine, Journal of Water Resources Planning and Management, ASCE, 123(2).

Stedinger, J.R., 1981, Chapter 6 in Loucks, Stedinger and Haith, Water Resource Systems Planning and Analysis, Prentice-Hall, Englewood Cliffs, NJ, 560 pp., 1981.

Young, W., D. Lam, V. Ressel, and I. Wong, (2000), "Development of an Environmental Flows Decision Support System," *Environmental Modelling and Software*, 257-267

Zagona, E., T. Fulp, H. M. Goranflo, and R. Shane, (1998), "RiverWare: A General River and Reservoir Modeling Environment," *Proceedings of the First Federal Interagency Hydrologic Modeling Conference*, Las Vegas, NV, pp. 113-120.

A Strategic Review of CALSIM II and its Use for Water Planning, Management, and Operations in Central California

Appendices

Appendix A: CALSIM II Science Review

Dates: Nov 13-14th
Location: Bay-Delta Room, CBDA Offices
650 Capitol Mall, 5th Floor
Sacramento, CA

Day 1: The Management Context, Model and Application Details

9:00 Welcome – Kim Taylor

- Overview of the CALFED Bay Delta Program -
- Introduction of the Panel

9:15 Water issues in California – Francis Chung

- General Hydrology
- SWP/CVP
- Operational challenges
- Sacramento-San Joaquin Delta – Ron Ott (5 min.)

9:35 Panel Q&A

9:45 Planning Models – Andy Draper

- CALSIM software
- CALSIM II application overview
- Interaction with other models

10:10 Panel Q&A

10:20 Break

10:30 Summary of CALSIM Applications

- DPLA/CalFed/US Bureau of Reclamation: Integrated Storage Investigations – Steve Roberts
- Bay Delta Office (DWR): SWP Delivery Reliability Report - Kathy Kelly
- USBR: Multi-layered modeling to simulate CVPIA (b)(2) water and Environmental Water Account Operations – Nancy Parker
- Operations Control Office (DWR): Oroville Relicensing, SWP Allocation decision procedure – Curtis Creel
- Department of Planning and Local Assistance (DWR): California Water Plan Update – Kamyar Guivetchi/Ken Kirby

12:15 Panel Q&A

12:30 Lunch

1:15 Summary of User and Stakeholder Interviews
  1:15 Interview Summary and Findings – UC Davis
  1:35 Panel Q&A
  1:50 Public Comment

2:15 CalSim II Details

- Development philosophy – Francis Chung
- Operation priorities, constraints, common assumptions – Erik Reyes
- Hydrology development – Andy Draper
- Delta water quality constraints – Ryan Wilbur

3:15 CalSim Evaluation

- Historical Operations Study / Sensitivity Analysis – Sushil Arora

3:30 Panel Q&A

3:45 Break

4:00 Future Directions

- Data Structure / Version Control / Multi-Period Prescriptive Optimization – Ryan Wilbur
- Daily Time Step - Dan Easton
- CalSim II – CVGSM Integration – Tariq Kadir
- Water Quality / Upstream Models – Nancy Parker

5:00 Panel organizational meeting (additional information needs, questions of specific staff, discussion plan)

Day 2—Panel Deliberations and Preliminary Report

8:30 Panel Q&A with specific DWR and USBR staff on request

9:30 Panel *in camera* discussions

11:00 Panel presentation of draft main findings—Pete Loucks

12:00 Wrap up and next steps - Kim Taylor i

Appendix B: Briefing Material for CALSIM II Peer Review

California Water
Averting a California Water Crisis (3 pages)
California Water Today, Bulletin 160-0, Chapter 2 (20 pages)
Water Supplies, California Water Plan Update, Bulletin 160-98, Chapter 3 (11pages)
Urban, Agricultural and Environmental Water Use, California Water Plan Update, Bulletin 160-98, Chapter 4 (17 pages)
California's Major Water Projects (map) (1 page)

CVP and SWP
State Water Project Operations (6 pages)
Central Valley Project Operations (16 pages)

CalSim and CalSim II Overview
CalSim: A Generalized Model for Reservoir System Analysis (19 pages)

CalSim Software Details
CalSim water resources simulation model: Users guide (18 pages)
CalSim water resources simulation model: Wresl language reference (11 pages)

CalSim II Details
Network Representation (1 page)
Sacramento-San Joaquin Delta Operations (9 pages)
Coordinated Operating Agreement (3 pages)
Reservoir Rule Curves (2 pages)
CalSim ANN Implementation (8 pages)
*CVPIA (b)(2) Management and Operations (6 pages).*ii
EWA Management and Operations (8 pages)
Multi-Cell Groundwater Model (2 pages)
SWP and CVP Delivery Allocation Logic (3 pages)

Hydrology Development
Surface Water Hydrology Development for CalSim II (8 pages)

Supporting Computer Models
Model Interaction (1 page)
CALAG (2 pages)
CU Model (2 pages)
DSM2 (2 pages)
IGSM2 – CVGSM (4 pages)
LCPSIM (5 pages)

CalSim II Evaluation
Planned Sensitivity Analysis (7 pages)
CalSim II Simulation of Historical SWP-CVP Operations - Extracts (61 pages)

CalSim II Applications

CalSim II Project Applications Summary (not completed)
SWP Delivery Reliability Report – Extracts (25 pages)
North of Delta Offstream Storage Investigations (3 pages)
In-Delta Storage Investigations (3 pages)
California Water Plan Update 2003 (3 pages)
*CalSim II and SWP Operations Control Office (1 page).*iii

Future Model Development

(a) CalSim Software

CalSimMulti-period Prescriptive Optimization (not completed)
CalSim Daily Time Step Model (not completed)
CalSim Water Quality Module (not completed)
Data Structure / Version Control (not completed)
CalSim Graphical User Interface (not completed)

(b) CalSim II Applications

CalSim II – CVGSM Integration (not completed)
CalSim II Geographical Expansion (not completed)
Global Climate Change (not completed)
Refined Spatial Resolution (not completed)
Expansion of Land Use Based Demands (not completed)
CalSim II – CALVIN Integration (not completed)
Revision of Urban Water Demands (not completed)

(c) Supporting Models

Replacement of Consumptive Use Model (not completed)

Appendix C: CALSIM II Review Process and Timeline

Establishing the Peer Review Panel

Dr. Pete Loucks (Cornell University and South Florida Water Management District) has accepted the CALFED Science Program's invitation to chair the panel. Other members are being currently being contacted by the Science Program staff

Organization of Briefing Material

Science Program and key agency staff, in consultation with the review panel chair, are identifying and organizing briefing material for panel members. Target date for completion is Sept 1, 2003. (This was extended to December 8, 2003)

Public Meeting of Review Panel

Target: 2-day session in November, 2003 in Sacramento area

Review workshop structure will include:

- Presentation overviews of California hydrology, water management, current issues, and the development of CALSIM II
- Presentations on the range of different current and potential applications of CALSIM for planning, operations, and supply reliability projects
- A summary of an independent interview project by Dr. Jay Lund of users and stakeholders explaining the major questions people are trying to answer with CALSIM II and other models
- Public comment to the panel
- Detail discussion of the model, including assumptions used in different applications, verification studies, and sensitivity analyses
- Opportunity for panel members to ask follow up questions of CALSIM developers and users
- An in camera session for panelists to discuss and begin compiling review comments
- A public presentation of the panel's draft findings

Panel Chair Provides Final Report to CALFED Lead Scientist

The panelists will be asked to finalize their review comments within 3 weeks of the public meeting and to transmit those directly to the Lead Scientist. The Science Program will transmit the completed review to CBDA and the CALFED community.

Appendix D: Panelists CALSIM II Review, Nov. 13-14, 2003

Name	**Affiliation**	**Position**	**Address/Phone/E-mail**
Andy Close	Murray Darling Basin Commission	Lead Modeler and System Manager	GPO Box 409 Canberra ACT 2601, AUSTRALIA (02)62790102 andy.close@mdbc.gov.au
Michael Haneman	UC Berkeley	"Senior Economist, Professor"	327 Giannini Hall, Berkeley, CA 94720-3310 (510)642-2670 hanemann@are.berkeley.edu
John Labadie	Colorado State University	Professor	B211 Engineering, Fort Collins, CO 80523 (970)491-6898 John.Labadie@colostate.edu
Pete Loucks	Cornell University	Professor	"Civil and Environmental Engineering, 311 Hollister Hall, Ithaca, NY 14853 " (607) 255-4896 DPL3@cornell.edu
Jay Lund	UC Davis	Professor	Civil and Environmental Engineering 3109 Engineering III, Davis, CA 95616" (530)752-5671 jrlund@ucdavis.edu
Daene McKinney	University of Texas at Austin	Professor	Civil and Environmental Engineering Campus Mail Code: C1786, Austin, TX 78712 (512)471-8772 daene_mckinney@mail.utexas.edu
Jery Stedinger	Cornell University	Professor	Civil and Environmental Engineering, Hollister Hall, Ithaca, NY 14853 (607) 255 2351 JRS5@Cornell.edu

Appendix E: Managing Model Development, Application, Documentation and Communication.

One possible means of maintaining control of the quality of particular versions of CALSIM II and accompanying models used for SWP-CVP planning and management decisions is to create an interagency modeling consortium (IMC) consisting of DWR, USBR, and persons from other stakeholder organizations, including NGOs and universities, if they are interested and want to participate. This consortium would be responsible for maintaining a toolbox of ‘acceptable’ models for ‘official’ use by the agencies and contractors.

IMC responsibilities and authority could include:

- Prioritize, coordinate, and provide consistency, technical guidance and oversight for all modeling applications,
- Approve model selection and insure that each requested application is carried out using the most appropriate model(s) and input data,
- Provide or otherwise insure documentation of the modeling process itself as well as the modeling results,
- Insure that the results are expressed and made available in a way such that others can understand and benefit from that modeling application, as applicable.
- Implement peer reviews of models and their applications as deemed appropriate.

To help meet their responsibilities the IMC will need to establish, publish and implement some procedures for insuring the quality of the entire model development and application process. They will need to identify among all the models that might be used, which are the most appropriate to address each of these separate groups of model applications. They must identify various models, i.e., establish a model toolbox, from which clients can choose the one that best meets their needs (or perhaps argue that another model should be added to the toolbox). The IMC will also need to maintain model documentation and provide for peer reviews of any model, its documentation, and/or its use in a project.

CMM Level 3 Performance Expectations

Firms that develop professional software are typically required to meet certain software standards. One such standard is defined in a book from Carnegie Mellon University. These so called Capability Maturity Model (CMM 1994) standards have various levels. For example, the South Florida Water Management District, that develops hydrologic models used as inputs to major investment decisions, strives to meet Level 3 standards. To meet such standards in software development and peer review, one needs to show that

- Modeling related problems are anticipated and prevented

- Model development and application groups work together as an integrated product team.
- Model use training is planned and provided as is needed.
- New modeling methodologies are identified and evaluated for possible implementation on a qualitative basis.
- Data are collected and used in all defined processes.
- Data are systematically shared across various projects.
- Both the models and their applications are evaluated and judged satisfactory by independent reviewers.

It seems to this panel that CALFED could without too much difficulty meet such standards if it chose to. Clearly planning for, conducting, and documenting these activities will require additional time and money. The expectation is that in the long run, such documentation and review will save time and money by redirecting misguided initiatives, identifying alternative approaches, or providing valuable technical support for a potentially controversial decision.

Model Toolbox

The IMC in collaboration with all agencies involved in water resources planning could be responsible for creating and maintaining a collection of models that agencies can use to meet their needs. As shown in Figure 1, this collection of models might be called the model toolbox. The criteria to be used as a basis for deciding whether a proposed model should or should not be included in the toolbox will depend in part on an assessment of the attributes of that model compared to alternative models and the suitability of the model to meet the needs of the project. Associated with the model toolbox is a library of completed model application documents and data bases for use by anyone who could benefit from them.

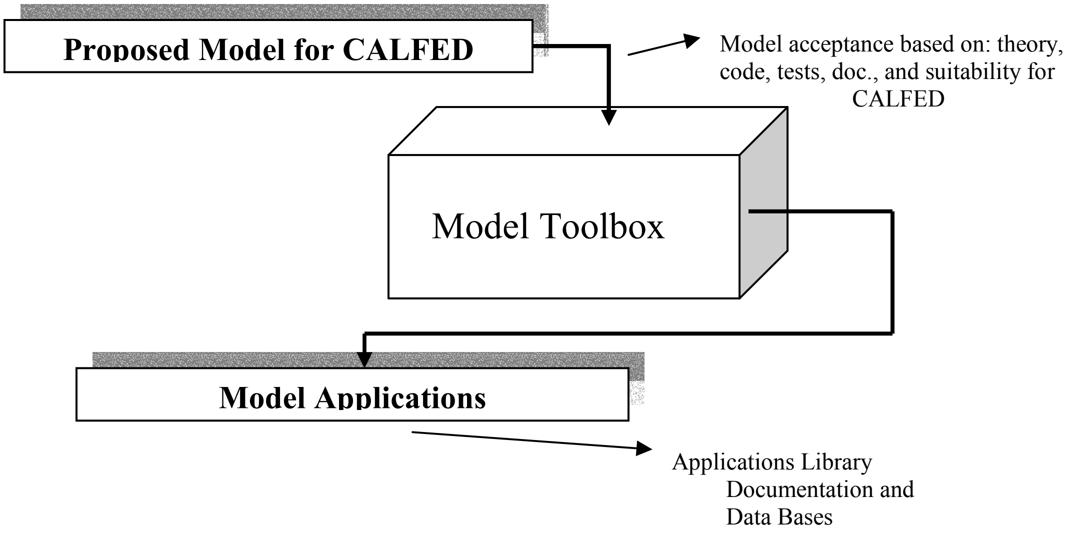


Figure 1. Model Toolbox consisting of approved models for use and Applications Library consisting of documentation and model data bases.

Everyone would agree that all modeling applications should be performed with the 'best' models available. But 'best' does not mean that all models used should be the most detailed, complex, realistic and thus usually the most expensive models available. The decision regarding the 'best' or most appropriate model should be based on the particular issues or questions being addressed, on the quantity and quality of the available input data, and on the time, personnel, and money available to perform the modeling application. The central question to be answered before initiating any modeling application is just what model output information (and precision) is needed to meet the needs of the decision making process. Expressed in other words, just how sensitive will the decision be to the type, amount and precision of the model output?

IMC in consultation with the other agencies could provide guidance on the adequacy of a particular version of CALSIM II or other associated model requested by each client with respect to the theory upon which it is based, its data requirements, its spatial and temporal resolutions, its documentation and status with respect to peer reviews, its capabilities, and its limitations. Similar considerations must be given to the proposed input data. To provide these services to each client requesting services from the IMC would require IMC to be staffed with personnel acquainted with the models in the toolbox, as well as be able to perform or review the simulations requested by various agencies.

There will likely be requests to use models not yet included in the model toolbox. IMC together with others from the DWR and/or USBR will need to judge the merits of such requests and if deemed beneficial, consider including such models in the toolbox. Undoubtedly the extent and quality of the documentation, testing, and peer review of various models in the toolbox will vary. However, a model's inclusion in the toolbox should signify that the model has been judged to be the best available for meeting the goals for which it was designed and is applicable to conditions in California.

Information Flows and Documentation

The IMC will probably be devoting a substantial amount of time giving guidance to clients and, when applicable, to the public. They will need to be working with the clients who are requesting model applications, and in situations where they are not doing this work, they will need to be reviewing and approving the work of the agencies or contractors who are performing the modeling services. IMC would provide technical assistance as well as oversight and coordination among all CALSIM II modeling activities.

Requests for modeling are easy to make, and time and money are required to carry them out. Requests sent to this proposed IMC should reflect some thought by those requesting such model runs as to just why the model application is desired, and just how the results are to be used. We would propose that requests include such items as:

- Reason for modeling,

- Type of modeling (e.g., event based or continuous),
- Particular model preference if any, and why, and possible alternatives,
- Model output information (data) needed and why and when it is needed,
  - What questions are the model results going to answer?
  - What issues are being studied?
  - What decisions are to be made, or at least to be informed, based on these model results?
  - When are the model results needed?
  - What formats are desired for presenting the model results?
- Location or site being modeled and the spatial and temporal scales desired,
- Particular input data assumptions, boundary conditions and other regional assumptions required,
- Source of input data, and format required or desired for the output data,
- Model calibration and verification needs and preferred procedures if any,
- Money and time available for modeling,
- Extent (duration) of the simulations to be performed,
- Desired performance measures, other than variables being simulated, if any,
- Alternative scenarios to be modeled (i.e., number of simulation runs needed),
- Other analyses or model applications that may or will need the output from this model application,
- Sensitivity and uncertainty analyses needed, and for which decision variables and why,
- Client contact person,
- Requirements for intermediate reviews of results or needs for periodic review of modeling application process logs and documents, and
- Other particular requirements or needs.

The use of a model nearly always takes place within a broader context. The model itself can also be part of a larger whole, such as a network of models in which some are using the outputs of other models. These conditions may impose constraints on the simulation modeling project. All these considerations need to be specified in the modeling application request.

Along with the proposal, there should also be a simple order-of-magnitude estimate of the expected values of all relevant decision variables based on simple mass-balance analytical solution methods that can be used without requiring a computer. These estimated values should be used to validate (check the reasonableness of) selected portions of the model runs. If there are any serious discrepancies, it may signify a major problem in the model output.

Is all this paperwork useful? It is to the extent it leads to a more effective and efficient use of personnel, money and time. Preparing a formal modeling application request requires some serious thought as to just why this is necessary and just what information is needed to further the project or analysis. It involves defining the objectives that are to be accomplished. Writing this down in some detail helps reduce the differences in perception that can exist between those who need information and those who are going to provide that information (IMC or a contractor). The problem as stated is often not the problem as understood, by either the client or the model user. In addition, problem perceptions and modeling objectives can change over the duration of a project. One should ask and answer the question of whether or not modeling in general is the right way to obtain the needed information. What are the alternatives to modeling?

The objective of any modeling project should be clearly understood with respect to the domain and the problem area, the reason for using a particular model, the questions to be answered by the model, the model assumptions and limitations, and the scenarios to be modeled. Throughout the project these objective components should be checked to see if any have changed and if they are being met.

If IMC is to serve as a central point to coordinate CALSIM II-related modeling activities, and to provide modeling services, it needs to have the authority to do so. This authority extends to giving advice on issues related to model and input data selection, and for reviewing, approving and prioritizing requests for services. Should contractors be involved in particular model applications, IMC must be authorized to specify the technical terms to be met and oversee the work done by the contractor. Finally IMC will need the financial and human resources needed to do this in a timely manner.

Modeling Application Documentation

One common problem of model studies once they are underway occurs when one wishes to go back over a series of simulation results to see what was changed or why a particular simulation was made or what was learned. It is also commonly difficult if not impossible for third parties to continue from the point at which any previous modeling project was terminated, especially if some time has passed. These problems are caused by a lack of information on how the study was carried out. What was the pattern of thought that took place? Which actions and activities were carried out? Who carried out what work and why? What choices were made? How reliable are the end results? These questions should be answerable if a model journal is kept. Just like computer programming documentation, modeling project documentation is often neglected under the pressure of time and perhaps because writing it is not as interesting as running the models themselves.

The paper trail of what has happened, what assumptions have been made, how calibration and verification were carried out, what results were obtained, why changes, if any, were made, what sensitivity analysis procedures were used and their results, and so on, could be contained in a modeling application documentation (MAD). Once the model application is completed, a copy of the MAD should be given to the requesting agency, as applicable and a copy should remain in IMC. These reports, or at least a summary of them, should be available for downloading from the web. Should further model applications be requested and approved, the requester as well as the IMC can refer to this previously prepared documentation to better understand what was done previously that pertains to the current request.

Model Calibration

Once a model is tested satisfactorily, it can be calibrated. Calibration of models such as CALSIM II are difficult because there are no historical observations of future scenarios to compare with model results. Historical runs, such as have been made, can provide some basis for calibration. In general the smaller the deviation between the calculated model results and the field observations, the better the model. This is true to a certain extent, as the deviations in a perfect model are only due to measurement errors. In practice, however, a good fit is by no means a guarantee of a good model.

The deviations between the model results and the field observations can be due to a number of factors. These factors include possible software errors, inappropriate modeling assumptions such as the (conscious) simplification of complex structures, neglecting certain processes, errors in the mathematical description or in the numerical method applied, inappropriate parameter values, errors in input data and boundary conditions, and measurement errors in the field observations.

To determine whether or not a calibrated model is a 'good' predictor, it should be validated or verified. Calibrated models should be able to reproduce field observations not used in calibration. Validation can be carried out for calibrated models if an independent data set has been kept aside for this purpose. If all available data are used in the calibration process in order to arrive at the best possible results, validation will not be possible. A decision to leave out validation may be a justifiable one especially when data are limited.

Philosophically it is impossible to know if a simulation model of a complex system is 'correct'. There is no way to prove it. Experimenting with a model, such as by carrying out multiple validation tests, can increase confidence in that model. After a sufficient number of successful tests, one might be willing to state that the model is 'good enough', based on the modeling project requirements. The model can then be regarded as having been validated, at least for the ranges of input data and field observations used in the validation.

If model predictions are to be made for situations or conditions for which the model has been validated, there may be some confidence in the reliability of those predictions. Yet one cannot be certain. Much less confidence can be placed on model predictions for conditions outside the range for which the model was validated.

While a model should not be used for extrapolations as commonly applied in predictions and in scenario analyses, this is often exactly the reason for the modeling project. What is likely to happen given events we have not yet experienced? A model's answer to this question should also include the uncertainties attached to these predictions. Depending on the type of model selected and used, one might end up predicting an incorrect future with great accuracy, or predicting the correct future with great uncertainty'. We don't yet know how to predict the correct future with great accuracy – so we do 'what ifs'. One can then argue about what scenarios – the ifs – are the most reasonable or probable, or about the impacts from improbable scenarios that you want to avoid should such scenarios occur.

Use the model

Once the model has been judged 'good enough,' the model may be used to obtain the information desired. Close communication between the client and the modeler during the modeling application process is essential to avoid any unnecessary misunderstandings about what information is wanted and the assumptions on which that information is to be based.

Before the end of this model-use step one should determine whether all the necessary simulations have been performed and whether they have been performed well. Questions to ask include

- did the model fulfill its purpose?
- are the results valid?
- are the quality requirements met?
- was the discretization of space and time chosen well?
- was the choice of the model restrictions correct?
- was the correct model and/or model program chosen?
- was the numerical approach appropriate?
- was the implementation performed correctly?
- what are the sensitive parameters (and other factors)?
- was an uncertainty analysis performed?

If any of the answers to these questions is no, then the situation should be corrected. If it cannot, the reason(s) for why it cannot be corrected should be documented in the model application document (MAD).

Interpret model results

Interpreting the information resulting from models is a crucial step in the modeling application process, especially in situations in which the client may only be interested in those results and not the way they were obtained. The model results can be compared to those of other similar studies. Are the results consistent? IMC must make that judgment. Any unanticipated results should be discussed and explained. The results should be judged with respect to the modeling project objectives.

The results of any modeling project typically include large files of time-series data. Only the most dedicated of clients will want to read those files. Thus these data must be presented in a more concise form. Statistical summaries should explicitly include any restrictions and uncertainties in the results. They should identify any gaps in the domain knowledge, thus generating new research questions or identifying the need for more field observations and measurements.

Report model results

Once the modeling application is completed, the organization doing the modeling will be responsible for preparing a report. The contents of this report should conform to the agreement made between modeling organization and the client prior to the initiation of the modeling application (see above). Although the results of a model are very rarely used as the sole basis for policy decisions, those requesting model applications may have a responsibility to translate their model results into policy recommendations. Policymakers, managers, and indeed the participating stakeholders typically want simple and clear unambiguous answers to complex questions. Much of the scientifically justified discussion, say regarding the uncertainties associated with some of the data, included in the main body of a report are not included in the executive summary of that report. This executive summary is often the only part read by those responsible for making decisions. Therefore, the conclusions of the model study must not only be scientifically correct, but also concisely formulated, without jargon, and fully understandable by managers and policymakers. When preparing or reviewing contractor model results reports, the IMC should consider this need.

These model application and model results reports should include sufficient detail to allow others to reproduce the model study (including its results) and/or to proceed from the point where this study ended. The report therefore requires a clear indication of the validity, usability and any restrictions of the model results.

Data Management

CALSIM II and its associated or linked models will require data. They will also produce data. Many of these data will have spatial and temporal dimensions. This information must be documented (meta data), preserved, and made accessible to IMC customers, coordination agencies and others. IMC should participate in data management strategic development, storage, documentation and dissemination. It should work with data base managers of various agencies to help them satisfy the IMC's data management requirements.

The availability of quality assured data is a critical dependency that must be met to facilitate timely completion of model development, implementation and application. To mitigate the impact of the availability of data on the timeline for the major model completion deadlines, the following issues should be addressed. :

- Updating land use / land cover data at regular and timely intervals.
- Developing and maintaining a common modeling database. This data base should include infrastructure design and operating policy data as well as water quality, ecological, land use, economic and of course hydrological data. Many of these data sets will have spatial as well as temporal dimensions. Each data set should have an associated metadata file.
- Pre-processed and post-processed datasets from previous model runs should be archived along with its metadata file in a central location for ease of access and availability.
- Measures to insure the consistency and quality of the input data.
- Measures to insure adequate communication among model developers, users and stakeholders. This includes measures to assist in developing documentation appropriate for each type of stakeholder.

Support of IMC activities

Common failures of IMC type organizations are typically due to:

- Insufficient staff to enable cross-training. This may lead to the dependency on one person or a very small group of employees for each sub module or the overall effort.
- Inadequate funding to institute good project management discipline.
- Inadequate funding to contract for technical writers and software engineers.
- Inadequate funding to contract for peer reviews.

Risk assessments

A risk assessment of CALSIM II and its associated models and data should be completed. The timely availability of quality assured data for example, is a risk. Project risk management includes the processes concerned with identifying, analyzing, and responding to uncertainties. Risk management attempts to minimize the results of adverse events. As a guide, the template, such as shown at the end of this Appendix, may be used to facilitate the assessment of risks.

Problem Management

Given the high visibility and criticality of the CALSIM II modeling effort an issue or problem management process should be developed within IMC. Issue/problem management includes the process for identifying, communicating, and resolving issues and problems.

The purpose of this procedure is to ensure that:

- Issues are identified, reported, managed, and resolved in a timely and effective manner. Responsibility is assigned to an owner for reporting, managing and resolving each issue
- All affected stakeholders are aware of the status of the issues
- Escalation of unresolved issues take place according to a defined procedure

In order to ensure that project issues and problems are appropriately managed various issue/problem management steps should be identified and followed to track the actions taken to resolve the issue or problem throughout the life of a modeling project.

B. Managing Peer Reviews

One means of quality control involves peer reviews of the models, their associated software, and their applications. One possible means of facilitating the peer review processes and for maintaining control on the particular versions of CALSIM II and accompanying models used for SWP-CVP planning and management decisions is another reason to create an interagency modeling consortium (IMC) consisting of DWR, USBR, and other stakeholder organization personnel if they are interested and want to participate. As suggested above, this consortium could be responsible for maintaining a toolbox of 'acceptable' peer-reviewed models for use by the agencies and contractors. The peer reviews should be of the theory underlying each model, the model’s software, the documentation of that software, the model’s functions and capabilities including those pertaining to model data input and output, model calibration and verification, sensitivity analyses, uncertainty analyses, user control (GUIs), spatial and temporal resolutions, limiting assumptions, and on the model (as opposed to code) documentation.

Just having evidence of published articles about a particular model in peer reviewed journals is not a substitute for a peer review of the model software and its applicability or suitability for certain types of analyses for SWP-CVP. Peer reviews of all models, their software, and their use should be accomplished by experts both within and outside of the originating agencies. ‘Inside’ agency (or internal) reviews may uncover some needed changes and identify other issues or problems that external reviewers could be asked to specifically examine and address. Internal reviews can make the external review process more effective, less costly and less time consuming.

Peer reviews are considered a key process area for Level 3 and higher of the Capability Maturity Model guidelines for improving the software process (Carnegie Mellon University, 1994). The purpose of peer review evaluations is to find defects in the model formulation and software and in its use, i.e., model application. Peer reviewers can also identify possible ways of correcting those defects, if any. If there are no defects, or after all known defects have been corrected, both the developers and users of any model and its software can have a stronger basis for believing that their product and its output are reliable.

Peer reviews serve the same function as accountants. Once a firm’s financial records have been peer reviewed by accountants (assuming they are qualified, objective and honest) the board of directors as well as the stockholders will have more assurance of the liabilities and net worth of their firm, and just how well it is being managed. In this case it is the assurance of the quality of the models, their software, and on their use in project evaluations, that actual and potential users of the model results depend upon.

The types of problems and issues for which a model, its software, and its documentation are designed to address are called the model’s ‘application niche’. Peer review of model development should include the evaluation of the intended application niche along with consideration of other aspects of model performance. Users of any model should be aware of the types of analyses for which the model is best suited and those for which the model is not well suited. This, along with the results of a peer review of any model application, should help the potential model user, or the user of the model results, better understand the limitations of the scientific basis of the model and just how much confidence can be placed on the model output.

Peer review triggers

Clearly judgment will have to be exercised as to just when and in what detail a peer review needs to be implemented. However the triggers on when a decision about a peer review needs to be made can be defined.

As shown in Figure 2, decisions regarding peer review are needed when models are proposed for the tool box and when model applications are completed. Should IMC decide a peer review is warranted when either of those events takes place, they will have to decide on the type of review and its level of detail. They will also need to identify the individuals to be asked to carry out that peer review.

Peer reviews are going to take time and cost money. They will also require IMC time to prepare the documentation needed for the peer reviewers and to read and act on reports prepared by the peer reviewers. This will apply if the peer review is internal or external.

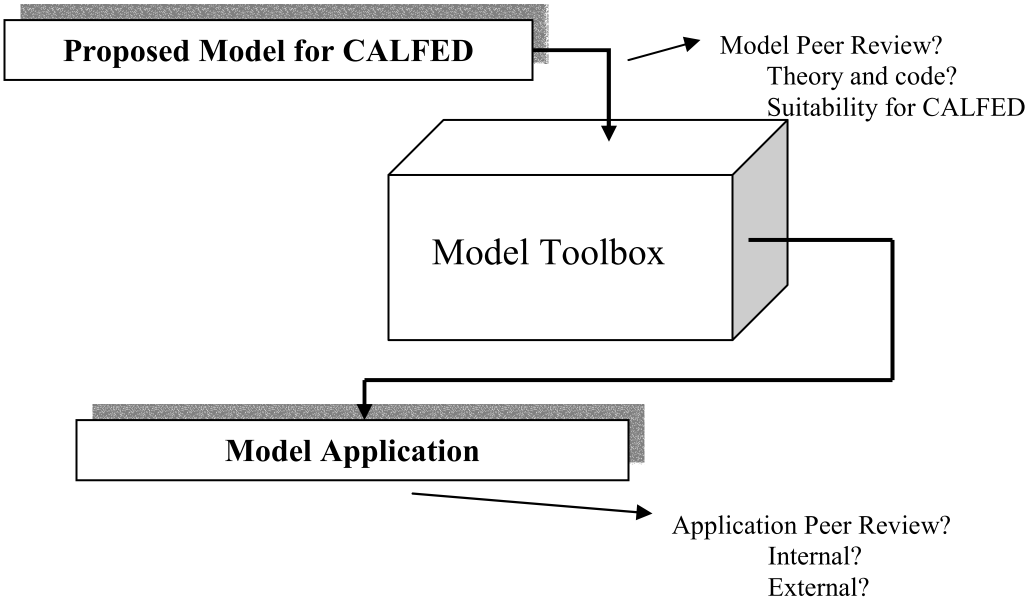


Figure 2. Schematic showing events where a peer review decision can be made.

The particular models and their associated software and documentation to be peer reviewed should be identified by the individuals or departments or agencies. This can include model process descriptions, software source code, documents, test results, and other supporting materials, as needed, for an adequate peer review of the entire model and its software. These products to be reviewed should be identified in writing and a written history of the review of different versions of each item should be maintained.

Events that take place in the progression of model development and use and subsequent modifications that warrant a peer review should be identified and specified in a written document. (This fits in to the model development and use documentation that should be maintained for Level 3 or higher CMM) When these events take place a peer review process should be considered, and if warranted, implemented. Depending on the event, the review can be solely internal, or it can involve an independent external review team as well.

Model application reviews should include an evaluation of the intended model application niche, and its applicability to current needs. Peer review may be appropriate for existing models when new information becomes available that could negate some or all of the conclusions of previous reviews or suggest a change in the currently specified application niche. Peer review of a model's applicability to a particular study should be planned well in advance of when model results are needed. The results of application reviews can influence the decisions made based on the model outputs. Once a peer review has been conducted for a particular model and its input data, peer reviews of subsequent applications of a model with similar inputs might be unnecessary. However, any time the model results may be controversial, or end up in litigation, another peer review may be justified.

Peer Review Process

The extent and process of performing and responding to peer reviews can vary in any organization. The ones discussed in this section attempt to follow the processes recommended by the Capability Maturity Model Level 3 guidelines.

Project peer review process should be specified in writing. A first step in this process should be to identify the particular modeling products and processes that will undergo peer review. This includes the models (i.e. the processes being modeled and the assumptions built into the models for describing these processes), their supporting software, the documentation of the model and its software, as well as all the written guidelines on how the models are to be used.

A second step is to perform an internal peer review prior to a model's use for project evaluation. It should be peer reviewed for accuracy, its suitability for use, and for identifying any possible errors in its logic, its coding, or in its documentation. Following an internal review, an external peer review can be performed.

Following the successful conclusion of internal and external peer reviews of a model and its documentation, the model can be applied to evaluate alternative projects. After the model has been applied to a particular project, the modeling process and its results should be peer reviewed to insure that the model has been applied properly, that the input data were appropriate, and that the conclusions drawn were valid.

Peer review teams should be selected, along with a peer review team leader. The particular personnel on the team will depend on the particular model and its software and documentation being reviewed. CALFED should have a list of qualified peer reviewers representing all applicable disciplines, both internal and external, that it can call upon to perform these reviews. The peer reviews are to be of the models and their use, not of the people who developed or used them. The reviews are to be used to evaluate the quality of modeling products and processes, not of the personnel involved.

Establishing and carrying out ongoing peer review processes costs money. Adequate funding must be made available to

1. identify and recruit a peer review team and team leader
2. prepare and distribute the peer review materials to the peer review team
3. support the time required for the team to review the materials prior to a team meeting
4. support the team meeting and to participate in it as appropriate (e.g., answering questions, conducting model experiments and sensitivity analyses, etc.)
5. reproduce and distribute the team report and to take actions as needed
6. monitor the modifications or changes being made to the model, its software, and its documentation, or redoing the model application, as needed.
7. prepare and distribute to model developers and potential users a report on the results of the peer review and the actions taken.

The particular peer review process may depend on just what is being peer reviewed and the resources and time available to perform the review. In general, however, the steps of a peer review could include the following:

1. DWR or CALFED should identify and establish a pool of possible reviewers representing various disciplines, with sufficient redundancy to allow for scheduling conflicts when ever some subset of those reviewers are needed. This includes both internal as well as external reviewers. What ever administrative work is need to establish this pool should be completed prior to when these reviewers will be needed.

2. At particular milestones in any new model development or in model application an internal peer review process could be initiated, to examine the modeling assumptions, the software that implements those assumptions in the case of model development or the data being used for model inputs in the case of model applications, and the documentation being prepared to describe the processes, to document the software code, and to document the tests that were run to test the code, or to document the results of the model application. If deemed appropriate, an external peer review could also be performed. If an external review is to take place, the particular reviewers need to be selected, notified, sent supporting documents, and be scheduled for one or more meetings, as needed. They should be issued contracts specifying the requirements (the checklist of items to be reviewed) and products expected.

3. Recommendations made by the peer review team need to be addressed and the actions taken along with the rationale for those actions should be documented.

4. The peer review team should review the actions taken and the results obtained from these actions. If not judged acceptable new recommendations should be made and submitted. A final report should be prepared by the peer review team when all recommendations have been successfully implemented or addressed, or if no further actions based on review team’s recommendations will be taken by the model developers or users.

The time and effort required for various levels of review should also be assessed and provided to the review team so that they can carry out the level of review requested of them. Otherwise the reviews may be superficial and while appearing to be peer reviewed, a model and its associated products may in fact be inadequately reviewed. Peer review teams have the responsibility to specify in writing the scope and limitations of their reviews.

As was the case for this peer review panel, the materials to be sent to the review team to allow them to prepare for their meeting should include the statement of review objectives and the level of detail desired, the applicable requirements and standards upon which to judge the adequacy of the products being reviewed, and of course the material that is to be reviewed. There should be a list of questions for the reviewers to address. Each review team member should be assigned and given responsibility for answering specific questions and for completing specific aspects of the overall review. All team members should be given specific review standards or requirements, including the expected completion dates. Checklists should be provided the review team that are applicable to the specific type of product being reviewed and the level of detail to be examined. These checklists will contain the criteria for judging the product, such as compliance with any standards and procedures, completeness, correctness, rules of construction, and maintainability.

Peer Review Issues and Questions

Each model development or application review will dictate its own special set of questions to be addressed. Some of these questions could relate to:

- Model Purpose and Objective
    - Use of model related to decisions being considered.
    - Model application niche, and why.
    - Model strengths and weaknesses –is it the best model?
- Model Processes and Limitations
    - Model processes, spatial and temporal scales, grid resolution.
    - Model variables and level of aggregation.
- Model Theoretical Basis
    - Model algorithms, numerical or analytical methods,
    - Model process formulation
    - Modeling approach in comparison with other models
    - Any shortcomings in relation to application niche
- Model Parameter Estimation
    - Methods used
    - Data available for parameter estimation
    - Parameter estimate reliabilities
    - Boundary conditions and appropriateness.
- Model Input Data Quantity/Quality
    - Data used in design of model
    - Data adequacy (quantity, quality, resolution) for model purpose and application
    - Data necessary for application of model
    - Key data gaps in model application
    - Additional data needs and why
- Model Key Assumptions
    - Basis for major assumptions

    - Sensitivity of model outputs to key assumptions
    - Sensitivity of potential decisions to key assumptions
    - Ease in modifying key assumptions
- Model Performance Measures
    - Criteria for assessing model performance
    - Correspondence of model output with measured observed data
    - Any model bias throughout range of model predictions
    - Variability and uncertainty analyses and representations in model results
    - What determines model's variability and uncertainty.
    - Model performance relative to others in application niche
- Model Documentation and User's Guide
    - Clarity of documentation, comprehensiveness of user's guide
    - Model applicability and limitations
    - Input data requirements for calibration, verification, model runs
    - Post modeling analyses, display and interpretation of results
    - Model code documentation
    - Model application documentation examples for prospective users.
- Review Retrospective
    - How well model and its application meet objectives and needs of project
    - Possible changes in the model to improve model performance
    - Robustness of model solutions to small changes in uncertain parameters, etc.
    - Ease of including uncertainty analyses associated with uncertain input data.
    - Key research needs for model improvement.

Peer Review Completion Reports

Procedures need to be established to track and confirm actions based on suggested changes or modifications in the material being reviewed. Once these actions are taken and completed, and documented, the peer review process for that particular product is completed. Peer review completion reports should contain data on what was reviewed and the results of the review. These data should include a description of the products that were reviewed, the level of detail of the review, any review limitations or qualifications, the number and backgrounds of the reviewers, the time spent preparing for and during review team meetings, the defects found and recommendations made, and the actions taken to address these recommendations.

Overall Peer Review Evaluations

The IMC or initiating agency should document the planning for and scheduling of peer reviews. The products to be reviewed and the level of detail to be examined also need to be specified. The procedures to be followed for selecting peer review team members, and the team leader, should also be determined and documented. Procedures for training potential reviewers, if such training is needed, should be identified and implemented, as required.

Periodically the IMC or applicable agency should assess just how well the plan described in the preceding paragraph is being carried out, and just how beneficial these peer reviews are to the overall modeling effort. Measures should be identified and used to determine the status of the peer review activities.   These measures could include the number of completed peer reviews performed compared to the number expected to be performed, the overall effort expended on peer reviews compared to that expected, and the number and extent of peer review recommendations requiring actions.

At a minimum these periodic reviews should verify that

1. The planned peer reviews and/or audits are conducted.
2. The peer review leaders are adequately trained for their roles.
3. The reviewers are properly trained or experienced in their roles.
4. The processes for preparing for and conducting peer reviews, and for following up on reviewer's recommendations are adequate and are being followed.
5. The reporting of peer review results is complete, accurate, timely and is being made available to model users.

Risk Management Template

Risk Definition Nam	Enter a short name that uniquely defines the risk
Risk #	Enter a unique number assigned to the risk. Range starts with 1 and continues.
Date Risk Identified	Enter the date the risk was identified
Risk Identification Source	Enter the source of the risk identification. In example, meeting name, group, or person.
Risk Owner	Enter the name of the person who will be responsible for ensuring the risk is approved, managed, periodically assessed, communicated, and tracked through closed or transfer.
Risk Detailed Description	Enter a detailed description of the risk so that a reader clearly understands the risk.
Probable Impact of Risk on Project (H, M, L)	Enter the impact on the project. o High = the risk will most likely occur and the impact could prevent the project from achieving its purpose. o Medium = there is a 50/50 change the risk would occur and the impact is serious but the project could still achieve its purpose if appropriately managed. o Low = there is a low probability that the risk would occur and minimal impact to the project's purpose.
Probable Impact of Risk on Project Costs	Enter the impact on the project in dollars. Determine what the potential cost to the project would be if the risk occurs.
Probable Impact of Risk on Project Schedule	Enter the schedule impact on the project. Determine how the schedule would be potentially impacted if the risk occurs.
Probable Impact of Risk on Project Results	Enter the impact on the project. Determine how the overall project purpose and results will be potentially impacted if the risk occurs.
Detailed Plan to Mitigate or Transfer Risk	Enter the detailed plan to mitigate the risk or a statement that the risk will be accepted. Mitigation could include ways to minimize, avoid, or transfer the risk to another party or group. Risk transfer would include evidence of agreement by the accepting party.
Detailed Project Action Items Required to Mitigate or Transfer Risk	Enter the detailed action items required to mitigate the risk. These items will be summarized and assigned within the project Action Log, along with an action item owner, and target completion date.
Detailed Project Plan Tasks Required to Mitigate Risk	Enter the detailed project plan task required to mitigate the risk. These items will be summarized and contained within the MS Project Schedule along with the effort, duration, schedule, and assigned resources.
Comments	Enter any permanent comments that cannot be included in the above items.
Referenced Documents	Enter any documents that a reader should consider in understanding, analyzing, mitigating, or accepting this risk.
Date Risk Closed	Enter the date this risk was closed. This would include when all action items or project tasks were completed, or the risk was transferred to another party or group.

Appendix F: Analysis of the November 2003 CALSIM II Validation Report

The following comments come from an analysis of the model results presented in the validation report *'CALSIM II Simulation of Historical SWP/CVP Operations'*, DWR (2003). The observations relate to the formulation of the model at November 2003. Changes might be made to that formulation which could resolve these issues.

Overestimation of Project Deliveries

The validation run suggests that the modeled demands included in CALSIM II overestimate the actual demands. CVP demands south of the Delta are assumed to be always equal to the contract entitlement whereas the observed deliveries in unrestricted years are consistently less than this amount. The modeled North of Delta deliveries are also consistently higher than observed. The modeled and observed CVP deliveries from the validation report are listed in Table 1.

Table 1. Comparison of modelled and observed CVP deliveries (1975-1998)

Project	Simulated Delivery (Taf/yr)	Historical Delivery (taf/yr)	Difference (taf/yr)	% Difference
CVP North of Delta	1960	1750	210	12
CVP South of Delta	2650	2490	160	6.4

Because the SWP south of delta demands were set to historical deliveries in many years, comparison with the historical deliveries in the validation report is of limited validity. However the fact that the historical SWP deliveries over the last ten years have averaged only 2385 taf/year while the modeled 'year 2001 development' SWP Delta deliveries reported in the 2002 State Water Project Delivery Reliability Report average 3090 taf/year, suggests that modeled SWP deliveries may also be too high.

Allocations to Project Contractors

Seasonal allocations to SWP and CVP contractors are made on the basis of water in storage, forecast inflows, projected carryover storage requirements and in-Basin and Delta regulatory requirements. The allocation processes used by the operators and those used by CALSIM II, are not identical. An examination of the way that CALSIM II has restricted project deliveries during the dry period of 1987-1992 (Figures 10, 16, 17 and 24 of the validation report) suggests that CALSIM II has allocated less water in the early years of the dry sequence than occurred in practice and consequently had more water available in 1991 and 1992 when the most severe restrictions were experienced. The carryover storage rules adopted can have a significant impact on the expected frequency and severity of water supply restrictions. The model rules need to be examined to ensure that the accurately reflect the way the system will be managed in the future.

San Luis Reservoir Operation

The rules used by the system operators for transferring water from headwater storages to the San Luis Reservoir can have a significant impact on:

- the pattern of flow in the Delta,
- the operation of accounting rules between the SWP and the CVP and
- opportunities for SWP wheeling of CVP water and possibly the availability of Article 21 water to SWP contractors.

A comparison of the modeled and observed storage behavior of the SWP component of San Luis (Figure 15) reveals that the model consistently underestimates the volume in storage. A comparison of the CVP component of the storage (Figure 23) indicates that the actual storage is filled earlier in the season and that the actual storage is also slightly higher than the modeled.

Users of CALSIM II output need to be confident that the rules adopted by the model for determining these transfers reflect the way this component of the system will be operated in the future.

Appendix G: Some Principles for Strategic Water Analysis for the California Water Plan Bulletin 160-03 (from the stakeholder review Draft, Sept. 30, 2003)

Strategy:

1) A frequently amended strategic document will lay out DWR's strategic analysis framework and identify the technical objectives, roles, and responsibilities of major DWR data collection efforts and analytical tools and their interactions and their responsible managers.

Transparency:

2) All data and models should be in the public domain and available on the web.
3) All data and models should have significant documentation.
4) Known limitations should be documented.

Longer-term viability:

5) Modularity: Major analytical tools will be designed and implemented to fit modularly and explicitly within the larger strategic analysis framework.
6) Adaptive data management framework: Major data efforts will fall within a larger data management framework, including protocols for data documentation and updating, and documentation of limitations.
7) A frequently-updated document will outline short-term and long-term efforts, budgets, and responsibilities for continuous improvement of analytical tools and data, with policy for continued user, local agency, and stakeholder involvement.

Coverage:

8) Spatial coverage for the basic data and analytical framework will be statewide.
9) Local and regional water management and resources will be explicitly represented.

Accountability and Quality Control:

10) In developing analytical tools, systematic efforts should be made to involve local agencies and stakeholders.
11) Major analytical products will undergo external review by a) external unaffiliated experts and b) local agencies whose systems are included in the model. User groups will exist for all major analytical products.
12) DWR's strategic analysis framework will undergo periodic internal and external review.

Appendix H: Model Sensitivity and Uncertainty Analysis

(This is a draft of a book chapter by DPL/JRS that may be useful for CALSIM II developers)

The usefulness of any model is in part dependent on the accuracy and reliability of its output data. Yet, because all models are imperfect abstractions of reality, and because precise input data are rarely if ever available, all output values are subject to imprecision. The input data and modeling uncertainties are not independent of each other. They can interact in various ways. The end result is imprecision and uncertainty associated with model output. This chapter focuses on ways of identifying, quantifying, and communicating the uncertainties in model outputs.

1. Introduction

Models are the primary way we have to estimate the multiple affects of alternative water resource system design and operating policies. Models predict the values of various system performance indicators. Model outputs are based on model structure, hydrologic and other time-series inputs and a host of parameters whose values describe the system being simulated. Even if these assumptions and input data reflect, or are at least representative of, conditions believed to be true, we know they will be wrong. Our models are always simplifications of the real systems we are studying. Furthermore, we simply cannot forecast the future with precision. So we know the model outputs of future conditions are uncertain estimates, at best.

Some prediction uncertainties can be reduced by additional research and data collection and analysis. Before undertaking expensive studies to gather and analyze additional data it is reasonable to ask what improvement in estimates of system performance or what reduction in the uncertainty associated with those estimates would result if all data and model uncertainties could be reduced. Such information helps determine how much one would be willing to ‘pay’ to reduce prediction uncertainty. If prediction uncertainty on average is costing a lot, it may pay to invest in additional data collection, more studies, or in better models all aimed at reducing that prediction uncertainty. If that uncertainty has no, or only a very modest, impact on the likely decision that is to be made, one should find other issues to worry about.

If it appears that reducing prediction uncertainty is worthwhile, then one should consider how best to do it. If doing this involves obtaining additional information, then it is clear that the value of this additional information, however measured, should exceed the cost of obtaining it. The value of such information will be the increase in system performance, or the reduction in its variance, that one can expect from obtaining such information. If additional information is to be obtained, it should be that information which reduces the uncertainties considered important, not the unimportant ones.

This chapter reviews some methods for identifying and communicating model prediction uncertainty. The discussion begins with a review of the causes of risk and uncertainty in model output. It then examines ways of measuring or quantifying uncertainty and model output sensitivity to model input imprecision, concentrating on methods that seem most relevant or practical for large-scale regional simulation modeling. It builds on some of the statistical methods reviewed in Chapter III and the modeling of risk and uncertainty in Chapter VI.

2. Issues, concerns, and terminology

Outcomes or events that cannot be predicted with certainty are often called risky or uncertain. Some individuals draw a special and interesting distinction between risk and uncertainty. In particular, the term risk is often reserved to describe situations for which probabilities are available to describe the likelihood of various events or outcomes. If probabilities of various events or outcomes cannot be quantified, or if the events themselves are unpredictable, some would say the problem is then one of uncertainty, and not of risk. In this chapter what is not certain is considered uncertain, and uncertainty is often described by a probability distribution. When the ranges of possible events are known and their probabilities are measurable, risk is called objective risk. If the probabilities are based solely on human judgment, the risk is called subjective risk.

Such distinctions between objective and subjective risk, and between risk and uncertainty, rarely serve any useful purpose to those developing and using models. Likewise the distinctions are often unimportant to those who should be aware of the risks or uncertainties associated with system performance indicator values.

Uncertainty in information is inherent in future-oriented planning efforts. Uncertainty stems from inadequate information and incorrect assumptions, as well as from the variability of natural processes. Water managers often need to identify both the uncertainty as well as the sensitivity of, or changes in, system performance indicator values due to the any changes in possible input data and parameter values from what were predicted. They need to reduce this level of uncertainty to the extent practicable. Finally, they need to communicate the residual uncertainties clearly so that decisions can be made with this knowledge and understanding.

Sensitivity analysis can be distinguished from uncertainty analysis. Sensitivity analysis procedures explore and quantify the impact of possible errors in input data on predicted model outputs and system performance indices. Simple sensitivity analysis procedures can be used to illustrate either graphically or numerically the consequences of alternative assumptions about the future. Uncertainty analyses employing probabilistic descriptions of model inputs can be used to derive probability distributions of model outputs and system performance indices. Figure 1 illustrates the impact of both input data sensitivity and input data uncertainty on model output uncertainty.

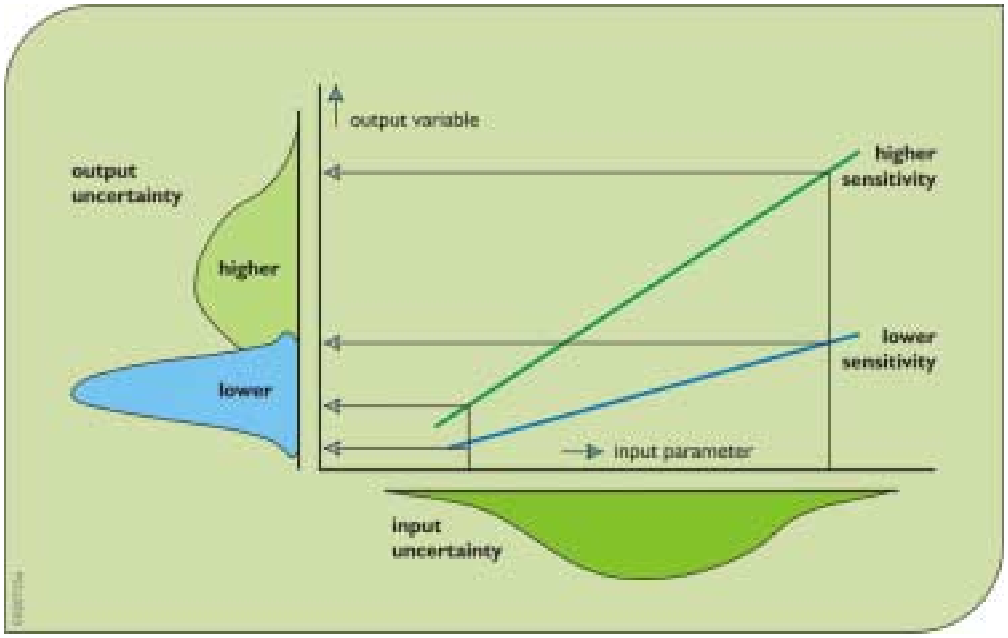


Figure 1. Schematic diagram showing relationship among model input parameter uncertainty and sensitivity to model output variable uncertainty (Lal, 1995).

It is worthwhile to explore the transformation of uncertainties in model inputs and parameters into uncertainty in model outputs when conditions differ from those reflected by the model inputs. Historical records of system characteristics are typically used as a basis for model inputs. Yet conditions in the future may change. There may be changes in the frequency and amounts of precipitation, changes in land cover and topography, and changes in the design and operation of control structures, all resulting in changes of water stages and flows, and their qualities, and consequently changes in the impacted ecosystems.

If asked how the system would operate with inputs similar to those in the historical database, the model should be able to interpolate within the available knowledge base to provide a fairly precise estimate. Still that estimate will not be perfect. This is because our ability to reproduce current and recent operations is not perfect, though it should be fairly good. If asked to predict system performance for situations very different from those in the historical knowledge base, or when the historical data are not considered representative of what might happen in the future, say due to climate change, such predictions become much less precise. There are two reasons why. First, our description of the characteristics of those different situations or conditions may be imprecise. Second, our knowledge base may not be sufficient for calibrating model parameters in ways that would enable us to reliably predict how the system will operate under conditions unlike those that have been experienced historically. The more conditions of interest are unlike those in the historical knowledge base, the less confidence we have that the model is providing a reliable description of systems operation. Figure 2 illustrates this issue.

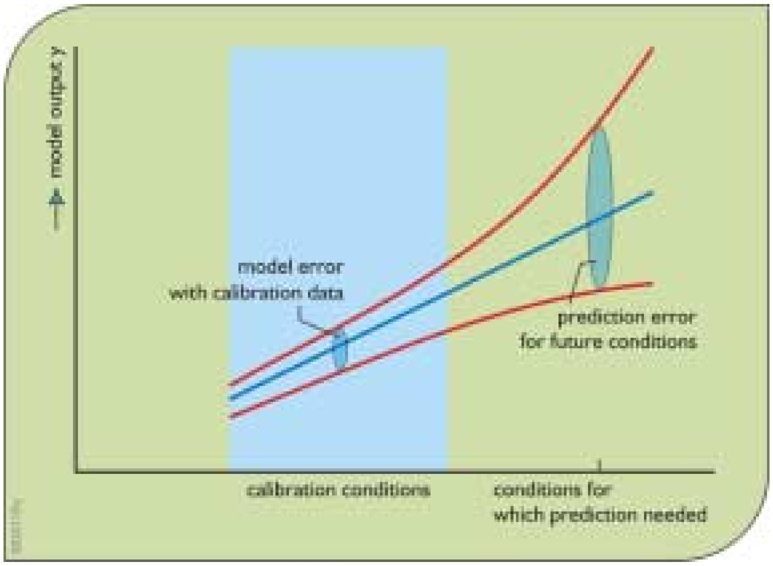


Figure 2. The precision of model predictions is affected by the difference between the conditions or scenarios of interest and the conditions or scenarios for which the model was calibrated.

Clearly a sensitivity analysis needs to consider how well a model can replicate current operations, and how similar the target conditions or scenarios are to those described in the historical record. The greater the required extrapolation from what has been observed, the greater will be the importance of parameter and model uncertainties.

The relative and absolute importance of different parameters will depend on the system performance indicators of interest. Seepage rates may have a very large local effect, but a small global effect. Changes in system-wide evapotranspiration rates will likely impact system-wide flows. The precision of model projections and the relative importance of errors in different parameters will depend upon the:

(1) precision with which the model can reproduce observed conditions,
(2) difference between the conditions predicted and the historical experience included in the knowledge base, and the
(3) system performance characteristics of interest.

Errors and approximations in input data measurement, parameter values, model structure and model solution algorithms, are all sources of uncertainty. While there are reasonable ways of quantifying and reducing these errors and the resulting range of uncertainty of various system performance indicator values they are impossible to eliminate. Decisions will still have to be made in the face of a risky and uncertain future. Decisions can be modified as new data and knowledge are obtained in a process of adaptive management.

There is also uncertainty with respect to human behavior and reaction related to particular outcomes and their likelihoods, i.e., to their risks and uncertainties. As important as risks and uncertainties associated with human reactions are to particular outcomes, they are not usually part of the models themselves. Social uncertainty may often be the most significant component of the total uncertainty associated with just how a water resource system will perform. For this reason we should seek designs and operating policies that are flexible and adaptable.

When uncertainties associated with system operation under a new operating regime are large, one should anticipate the need to make changes and improvements as experience is gained and new information accumulates. When predictions are highly unreliable, responsible managers should favor actions that are robust (e.g., good under a wide range of situations), gain information through research and experimentation, monitor results to provide feedback for the next decision, update assessments and modify policies in the light of new information, and avoid irreversible actions and commitments.

3. Variability and uncertainty in model output

Differences between model output and observed values can result from either natural variability, say caused by unpredictable rainfall, evapotranspiration, water consumption, and the like, and/or by both known and unknown errors in the input data, the model parameters, or the model itself. The later is sometimes called knowledge uncertainty but it isn’t always due to a lack of knowledge. Models are always simplifications of reality and hence ‘imprecision’ can result. Sometimes imprecision occurs because of a lack of knowledge, such as just how a particular species will react to various environmental and other habitat conditions. Other times known errors are introduced simply for practical reasons.

Imperfect representation of processes in a model constitutes model structural uncertainty. Imperfect knowledge of the values of parameters associated with these processes constitutes model parameter uncertainty. Natural variability includes both temporal variability and spatial variability, to which model input values may be subject.

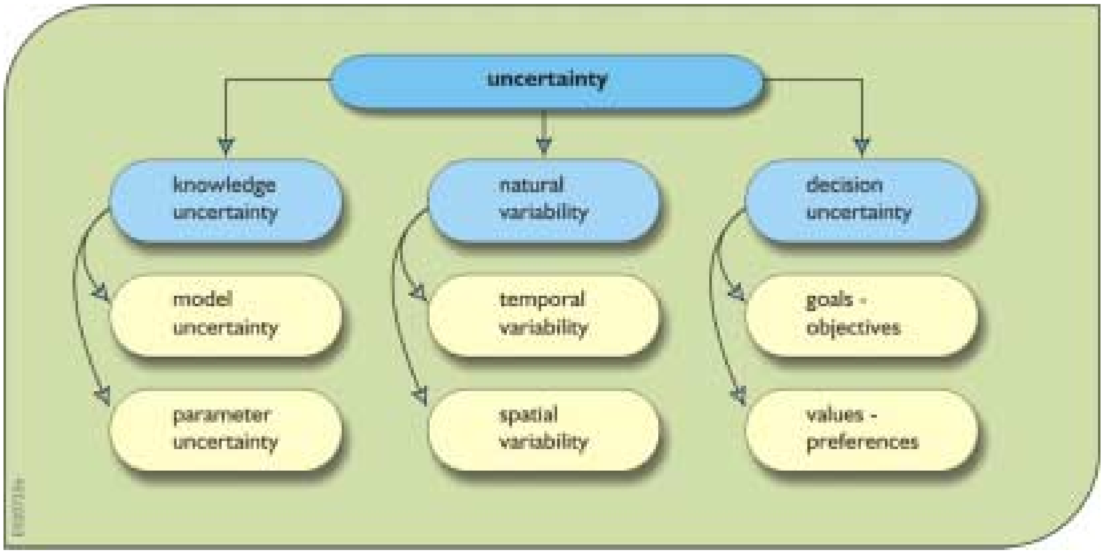


Figure 3. One way of classifying types of uncertainty.

Figure 3 illustrates these different types of uncertainty. For example, the rainfall measured at a weather station within a particular model grid cell may be used as an input value for that cell, but the rainfall may actually vary at different points within that cell and its mean value will vary across the landscape. Knowledge uncertainty can be reduced through further measurement and/or research. Natural variability is a property of the natural system, and is usually not reducible at the scale being used. Decision uncertainty is simply an acknowledgement that we cannot predict ahead of time just what decisions individuals and organizations will make, or even just what particular set of goals or objectives will be considered and the relative importance of each.

Rather than contrasting 'knowledge' uncertainty vs. natural variability vs. decision uncertainty, one can classify uncertainty in another way based on specific sources of uncertainty, such as those listed below, and address ways of identifying and dealing with each source of uncertainty.

Informational Uncertainties:

- imprecision in specifying the boundary and initial conditions that impact the output variable values
- imprecision in measuring observed output variable values

Model Uncertainties:

- uncertain model structure and parameter values
- variability of observed input and output values over a region smaller than the spatial scale of the model
- variability of observed model input and output values within a time smaller than the temporal scale of the model. (e.g., rainfall and depths and flows within a day)
- errors in linking models of different spatial and temporal scales

Numerical Errors:

- errors in the model solution algorithm

3.1 Natural variability

The main source of hydrologic model output value variability is the natural variability in hydrological and meteorological input series. Periods of normal precipitation and temperature can be interrupted by periods of extended drought and intense meteorological events such as hurricanes and tornadoes. There is no reason to think such events will not continue to occur and become even more frequent and extreme. Research has demonstrated that climate has been variable in the past and concerns about anthropogenic activities that may increase that variability increase each year. Sensitivity analysis can help assess the affect of errors in predictions if those predictions are based only on past records of historical time-series data describing precipitation, temperature and other exogenous forces across and on the border of the regions being studied.

Time series input data are often actual, or at least based on, historical data. The time-series values typically describe historical conditions including droughts and wet periods. What is distinctive about natural uncertainty, as opposed to errors and uncertainty due to modeling limitations, is that natural variability in meteorological forces cannot be reduced by improving the model's structure, increasing the resolution of the simulation, or by better calibration of model parameters.

Errors result if meteorological values are not measured or recorded accurately, or if mistakes are made in the generation of computer data files. Furthermore, there is no assurance the statistical properties of historical data will accurately represent the statistical properties of future data. Actual future precipitation and temperature scenarios will be different from those in the past, and this difference in many cases may have a larger affect than the uncertainty due to incorrect parameter values. However, the affects of uncertainties in the parameter values used in stochastic generation models are often much more significant than the affects of using different stochastic generation models (Stedinger and Taylor, 1982).

While variability of model output is a direct result of variability of model input (e.g., hydrologic and meteorological data), the extent of the variability, and the lower and upper limits of that variability, may also be affected by errors in the inputs, the values of parameters, initial boundary conditions, model structure, processes and solution algorithms.

Figure 4 illustrates the distinction between the variability of a system performance indicator due to input data variability, and the extended range of variability due to the total uncertainty associated with any combination of the causes listed in the previous section. This extended range is what is of interest to water resource planners and managers.

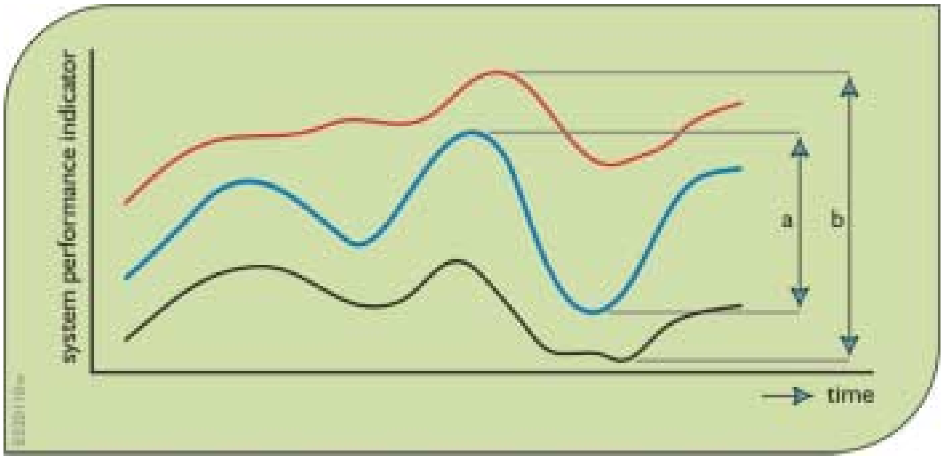


Figure 4. Time-series of model output or system performance showing variability over time. Range "a" results from the natural variability of input data over time. The extended range "b" results from the variability of natural input data as well as from imprecision in input data measurement, parameter value estimation, model structure and errors in model solution algorithms. The extent of this range will depend on the confidence level associated with that range.

What can occur in practice is a time-series of system performance indicator values that can range anywhere within or even outside the extended range, assuming the confidence level of that extended range is less than 100%. The confidence one can have that some future value of a time series will be within a given range is dependent on two factors. The first is the number of measurements used to compute the confidence limits. The second is on the assumption that those measurements are representative of - come from the same statistical or stochastic process yielding - future measurements. Figure 5 illustrates this point. Note that the time series may even contain values outside the range "b" defined in Figure 4 if the confidence level of that range is less than 100%. Confidence intervals associated with less than 100% certainty will not include every possible value that might occur.

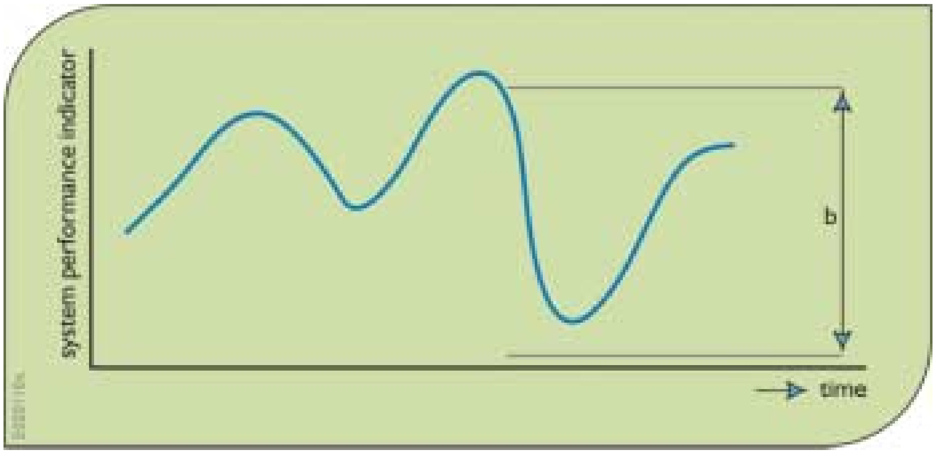


Figure 5. Typical time series of model output or system performance indicator values that are the result of input data variability and possible imprecision in input data measurement, parameter value estimation, model structure and errors in model solution algorithms.

3.2 Knowledge uncertainty

Referring to Figure 3, knowledge uncertainty includes model structure and parameter value uncertainties. First we consider parameter value uncertainty including boundary condition uncertainty, and then model and solution algorithm uncertainty.

3.2.1 Parameter value uncertainty

A possible source of uncertainty in model output results from uncertain estimates of various model parameter values. If the model calibration procedure were repeated using different data sets, different parameter values would result. Those values would yield different simulated system behavior, and thus different predictions. We can call this parameter uncertainty in the predictions because it is caused by imprecise parameter values. If such parameter value imprecision were eliminated, then the prediction would always be the same and so the parameter value uncertainty in the predictions would be zero. But this does not mean that predictions would be perfectly accurate.

In addition to parameter value imprecision, uncertainty in model output can result from imprecise specification of boundary conditions. These boundary conditions can be either fixed or variable. However, because they are not being computed based on the state of the system, their values can be uncertain. These uncertainties can affect the model output, especially in the vicinity of the boundary, in each time step of the simulation.

3.2.2 Model structural and computational errors

Uncertainty in model output can also result from errors in the model structure compared to the real system, and approximations made by numerical methods employed in the simulation. No matter how good our parameter value estimates, our models are not perfect and there is a residual model error. Increasing model complexity to more closely represent the complexity of the real system may not only add to the cost of data collection, but also introduce even more parameters, and thus even more potential sources of error in model output. It is not an easy task to judge the appropriate level of model complexity, and to estimate the resulting levels of uncertainty associated with various assumptions regarding model structure and solution methods. Kuczera (1988) provides an example of a conceptual hydrologic modeling exercise with daily time steps where model uncertainty dominated parameter value uncertainty.

3.3 Decision uncertainty

Uncertainty in model predictions can result from unanticipated changes in what is being modeled. These can include changes in nature, human goals, interests, activities, demands, and impacts. An example of this is the deviation from standard or published operating policies by operators of infrastructure such as canal gates, pumps, and reservoirs in the field, as compared to what is specified in documents and incorporated into the water systems models. Comparing field data with model data for model calibration may yield incorrect calibrations if operating policies actually implemented in the field differ significantly from those built into the models. What do operators do in times of stress? And can anyone identify a place where deviations from published policies do not occur?

What humans will want to achieve in the future may not be the same as what they want today. Predictions of what people will want in the future are clearly sources of uncertainty. A perfect example of this is in the very flat Greater Everglades region of south Florida in the US. Fifty years ago folks wanted the swampy region protected from floods and drained for agricultural and urban development. Today many want just the opposite at least where there are no human settlements. They want to return to a more natural hydrologic system with more wetlands and unobstructed flows, but now for ecological restoration objectives that were not a major concern or much appreciated some half a century ago. Once the mosquitoes return and if the sea level continues to rise, future populations who live there may want more flood control and drainage again. Who knows? Complex changing social and economic processes influence human activities and their demands for water resources and environmental amenities over time. Some of these processes reflect changes in local concerns, interests and activities, but population migration and many economic activities and social attitudes can also reflect changing national and international trends.

Sensitivity scenarios that include human activities can help define the affects of those activities within an area. It is important that careful attention go into the development of these alternative scenarios so that they realistically capture the forces or stresses that the system may face. The history of systems studies are full of examples where the issues studied were rapidly overwhelmed by much larger social forces resulting from, for example, the relocation of major economic activities, an oil embargo, changes in national demand for natural resources, economic recession, sea-level rise, an act of terrorism, or even war. One thing is sure; the future will be different than the past, and no one is certain just how.

3.3.1 Surprises

Water resource managers may also want to consider how vulnerable a system is to undesirable environmental surprises. What havoc might an introduced species like the zebra mussel invading the Great Lakes of North America have in a particular watershed? Might some introduced disease suddenly threaten key plant or animal species? Might management plans have to be restructured to address the survival of some species such as salmon in the Rhine River in Europe or in the Columbia River in North America? Such uncertainties are hard to anticipate when by their nature they are truly surprises. But surprises should be expected. Hence system flexibility and adaptability should be sought to deal with changing management demands, objectives, and constraints.

4. Sensitivity and uncertainty analyses

An uncertainty analysis is not the same as a sensitivity analysis. An uncertainty analysis attempts to describe the entire set of possible outcomes, together with their associated probabilities of occurrence. A sensitivity analysis attempts to determine the relative change in model output values given modest changes in model input values. A sensitivity analysis thus measures the change in the model output in a localized region of the space of inputs. However, one can often use the same set of model runs for both uncertainty analyses and sensitivity analyses. It is possible to carry out a sensitivity analysis of the model around a current solution and then use it as part of a first order uncertainty analysis.

This discussion begins by focusing on some methods of uncertainty analysis. Then various ways of performing and displaying sensitivity analyses are reviewed.

4.1 Uncertainty Analyses

Recall that uncertainty involves the notion of randomness. If a value of a performance indicator or performance measure, or in fact any variable, like the phosphorus concentration or the depth of water at a particular location varies and this variation over space and time cannot be predicted with certainty, it is called a random variable. One cannot say with certainty what the value of a random variable will be but only the likelihood or probability that it will be within some specified range of values. The probabilities of observing particular ranges of values of a random variable are described or defined by a probability distribution. There are many types of distributions and each can be expressed in several ways as presented in Chapter III.

Suppose the random variable is *X*. If the observed values of this random variable can be only discrete values, the probability distribution of *X* can be expressed as a histogram, as shown in Figure 6a. The sum of the probabilities for all possible outcomes must equal 1. If the random variable is a continuous variable that can assume any real value over a range of values, the probability distribution of *X* can be expressed as a continuous distribution as shown in Figure 6b. The shaded area under the density function for the continuous distribution is 1. The area between two values of the continuous random variable, such as between u and v in Figure 6c, represents the probability that the observed value *x* of the random variable value *X* will be within that range of values.

The probability distribution, $P_X(x)$ shown in Figure 6 (a) is called a probability mass function. The probability distributions shown in Figure 6 (b and c) are called a probability density functions (pdf) and are denoted by $f_X(x)$. The subscript *X* of P_X and f_X represents the random variable, and the variable *x* is some value of that random variable *X*.

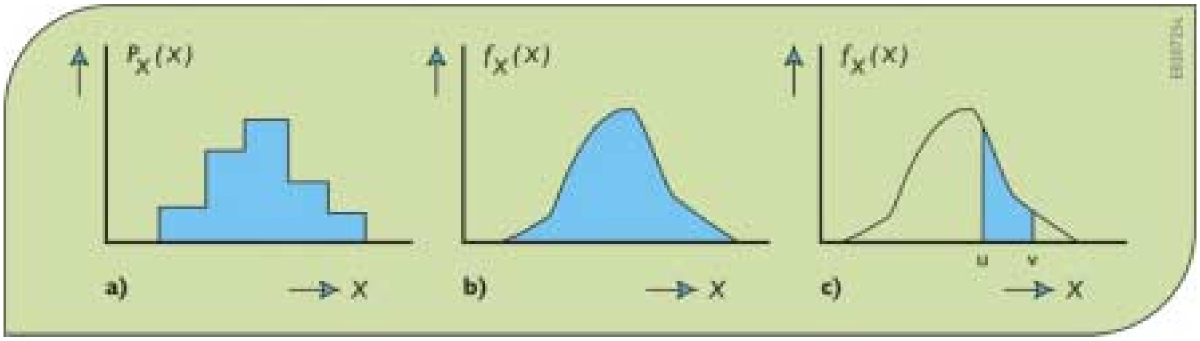


Figure 6. Probability distributions for a discrete or continuous random variable *X*. The area under the distributions (shaded areas in a and b) is 1, and the shaded area in c is the probability that the observed value x of the random variable *X* will be between u and v.

Uncertainty analyses involve identifying characteristics of various probability distributions of model input and output variables, and subsequently functions of those random output variables that are performance indicators or measures. Often targets associated with these indicators or measures are themselves uncertain.

A complete uncertainty analysis would involve a comprehensive identification of all sources of uncertainty that contribute to the joint probability distributions of each input or output variable. Assume such analyses were performed for two alternative project plans, *A* and *B*, and that the resulting probability density distributions for a specified performance measure were as shown in Figure 7. Figure 7 also identifies the costs of these two projects. The introduction of two performance criteria, cost and probability of exceeding a performance measure target (e.g., a pollutant concentration standard) introduces a conflict where a tradeoff must be made.

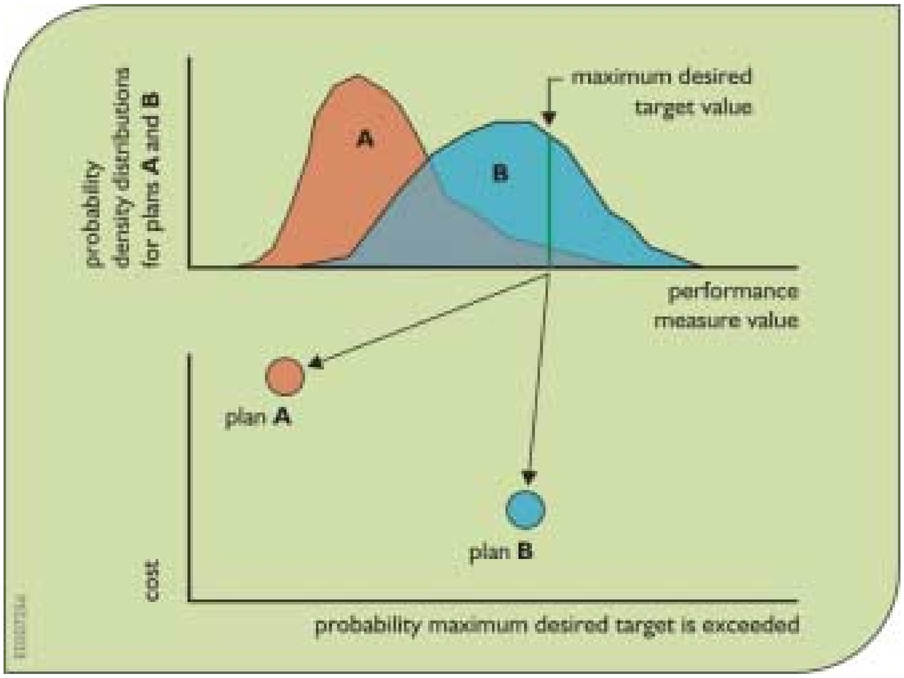


Figure 7. Tradeoffs involving cost and the probability that a maximum desired target value will be exceeded. In this illustration we want the lowest cost (*B* is best) and the lowest probability of exceedance (*A* is best).

4.1.1 Model and model parameter uncertainties

Consider a situation as shown in Figure 8, in which for a specific set of model inputs, the model outputs differ from the observed values, and for those model inputs, the observed values are always the same. Here nothing randomly occurs. The model parameter values or model structure needs to be changed. This is typically done in a model calibration process.

Given specific inputs, the outputs of deterministic models are always going to be the same each time those inputs are simulated. If for specified inputs to any simulation model the predicted output does not agree with the observed value, as shown in Figure 8, this could result from imprecision in the measurement of observed data. It could also result from imprecision in the model parameter values, the model structure, or the algorithm used to solve the model.

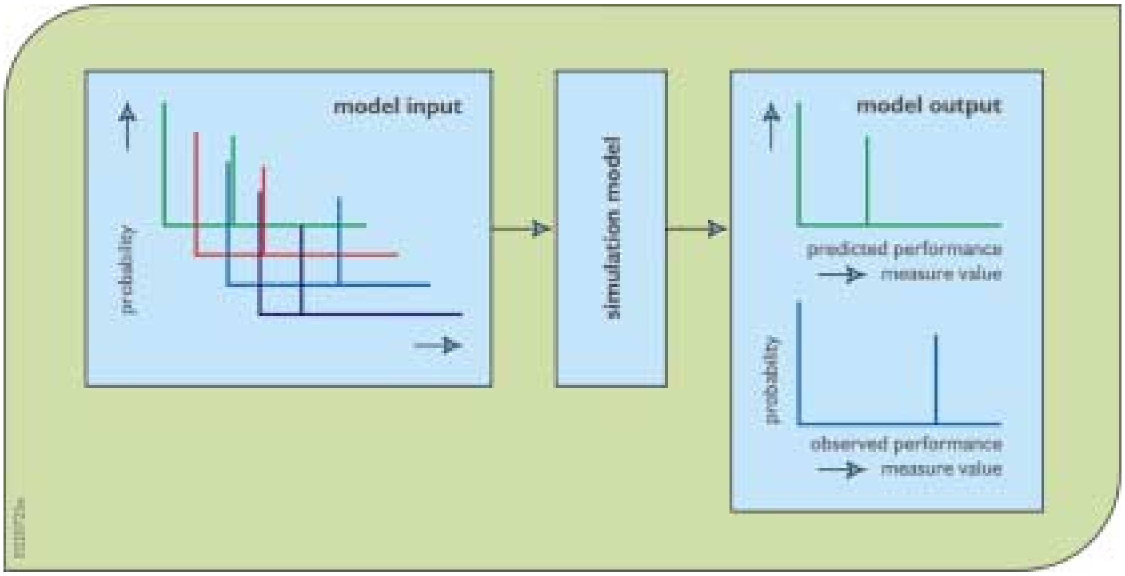


Figure 8. A deterministic system and a simulation model of that system needing calibration or modification in its structure. There is no randomness, only parameter value or model structure errors to be identified and corrected.

Next consider the same deterministic simulation model but now assume at least some of the inputs are random, i.e., not predictable, as may be case when random outputs of one model are used as inputs into another model. Random inputs will yield random outputs. The model input and output values can be described by probability distributions. If the uncertainty in the output is due only to the uncertainty in the input, the situation is similar to that shown in Figure 8. If the distribution of performance measure output values does not fit or is not identical to the distribution of observed performance measure values, then calibration of model parameter values or modification of model structure may be needed.

If a model calibration or 'identification' exercise finds the 'best' values of the parameters to be outside reasonable ranges of values based on scientific knowledge, then the model structure or algorithm might be in error. Assuming the algorithms used to solve the models are correct and observed measurements of system performance vary for the same model inputs, as shown in Figure 9, it can be assumed that the model structure does not capture all the processes that are taking place that impact the value of the performance measures. This is often the case when relatively simple and low-resolution models are used to estimate the hydrological and ecological impacts of water and land management policies. However, even large and complex models can fail to include or adequately describe important phenomena.

In the presence of informational uncertainties there may be considerable uncertainty about the values of the "best" parameters during calibration. This problem becomes even more pronounced with increases in model complexity.

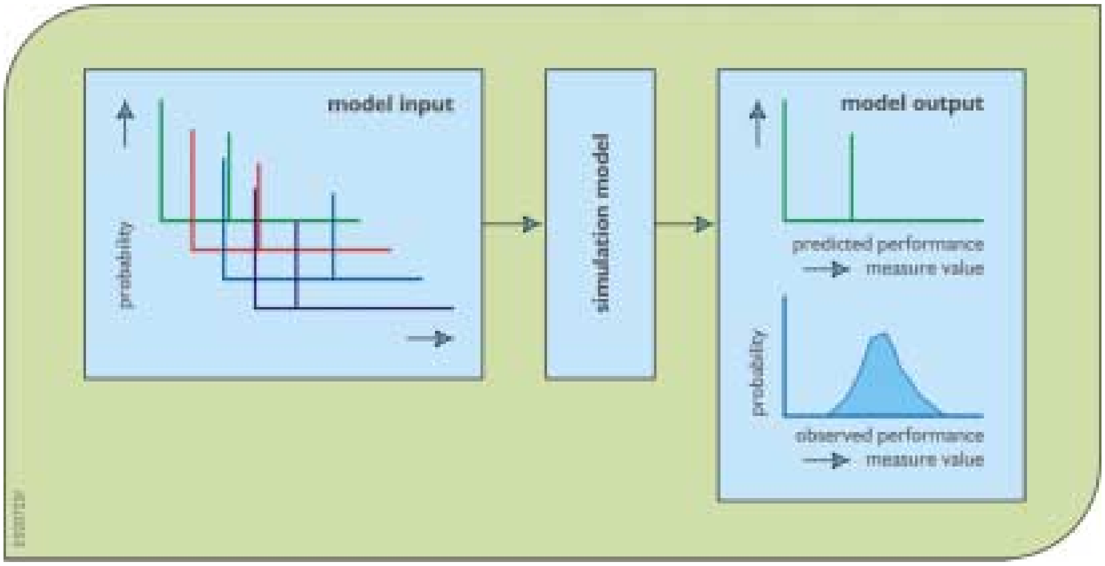


Figure A deterministic simulation model of a 'random or stochastic' system. To produce the variability in the model output that is observed in the real system, even given the same input values, the model's parameter values may need to vary over distributions of values and/or the model structure may need modification along with additional model inputs.

An example: Consider the prediction of a pollutant concentration at some site downstream of a pollutant discharge site. Given a streamflow Q (in units of 1000 m^3/day), the distance between the discharge site and the monitoring site, X (m), the pollutant decay rate constant k (day^{-1}), and the pollutant discharge W (Kg/day), we can use the following simplified model to predict the concentration of the pollutant C (g/m^3 = mg/l) at the downstream monitoring site:

$$C = (W/Q)\exp\{-k(X/U)\}$$

In the above equation assume the velocity U (m/day) is a known function of the streamflow Q.

In this case the observed value of the pollutant concentration C may differ from the computed value of C even for the same inputs of W, Q, k, X, and U. Furthermore, this difference varies in different time periods. This apparent variability, as illustrated in Figure 9, can be simulated using the same model but by assuming a distribution of values for the decay rate constant k. Alternatively the model structure can be modified to include the impact of streamflow temperature T on the prediction of C.

$$C = (W/Q)\exp\{-k\theta^{T-2}(X/U)\}$$

Now there are two model parameters, the decay rate constant k and the dimensionless temperature correction factor □and an additional model input, the streamflow temperature, T. It could be that the variation in streamflow temperature was the sole cause of the first equation's 'uncertainty' and that the assumed parameter distribution of k was simply the result of the distribution of streamflow temperatures on the term $k\theta^{T-20}$.

If the output were still random given constant values of all the inputs, then another source of uncertainty exists. This uncertainty might be due to additional random loadings of the pollutant, possibly from non-point sources. Once again the model could be modified to include these additional loadings if they are knowable. Assuming these additional loadings are not known, a new random parameter could be added to the input variable W or to the right hand side of the equations above that would attempt to capture the impact on C of these additional loadings. A potential problem, however, might be the likely correlation between those additional loadings and the streamflow Q.

While adding model detail removed some 'uncertainty' in the above example, increasing model complexity will not always eliminate or reduce uncertainty in model output. Adding complexity is generally not a good idea when the increased complexity is based on processes whose parameters are difficult to measure, the right equations are not known at the scale of application, or the amount of data for calibration is small compared to the number of parameters.

Even if more detailed models requiring more input data and more parameter values were to be developed, the likelihood of capturing all the processes occurring in a complex system is small. Hence those involved will have to make decisions taking this uncertainty into account. Imprecision will always exist due to less than a complete understanding of the system and the hydrologic processes being modeled. A number of studies have addressed model simplification, but only in some simple cases have statisticians been able to identify just how one might minimize modeling related errors in model output values.

The problem of determining the "optimal" level of modeling detail is particularly important when simulating the hydrologic events at many sites over large areas. Perhaps the best approach for these simulations is to establish confidence levels for alternative sets of models and then statistically compare simulation results. But even this is not a trivial or costless task. Increases in the temporal or spatial resolution typically require considerable data collection and/or processing, model recalibrations, and possibly the solution of stability problems resulting from the numerical methods used in the models. Obtaining and implementing alternative hydrologic simulation models will typically involve considerable investments of money and time for data preparation and model calibration.

What is needed is a way to predict the variability evident in the system shown in Figure 9. Instead of a fixed output vector for each fixed input vector, a distribution of outputs are needed for each performance measure based on fixed inputs (Figure 9) or a distribution of inputs (Figure 10.). Furthermore the model output distribution for each performance measure should 'match' as well as possible the observed distribution of that performance measure.

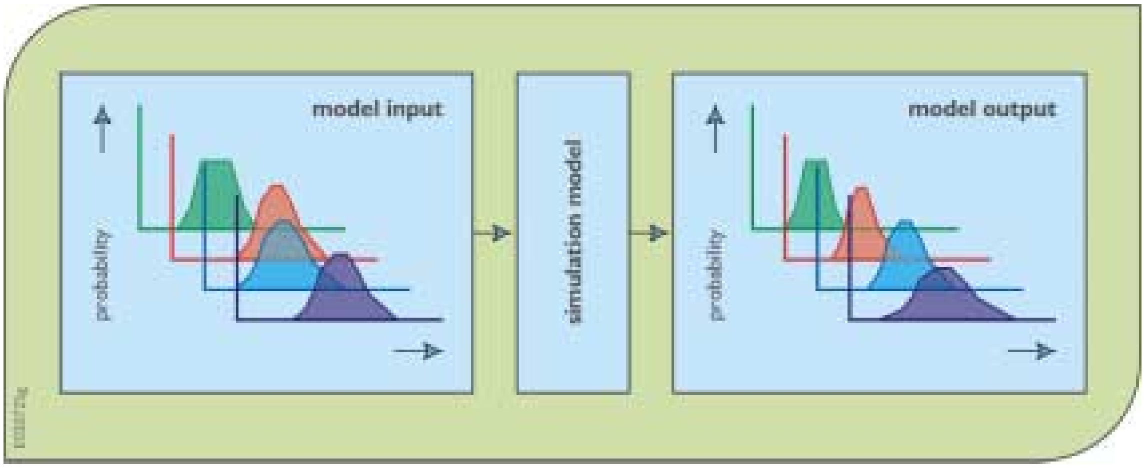


Figure 10. Simulating variable inputs to obtain probability distributions of predicted performance indices that match the probability distributions of observed performance values.

4.1.2 What uncertainty analysis can provide

An uncertainty analysis takes a set of randomly chosen input values (that can include parameter values), passes them through a model (or transfer function) to obtain the distributions (or statistical measures of the distributions) of the resulting outputs. As illustrated in Figure 11, the output distributions can be used to

- Describe the range of potential outputs of the system at some probability level.
- Estimate the probability that the output will exceed a specific threshold or performance measure target value.

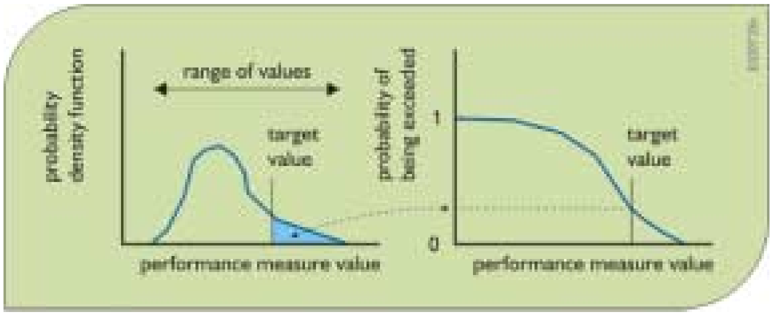


Figure 11. The distribution of performance measures defines range of potential values and the likelihood that a specified target value will be exceeded. The shaded area under the density function on the left represents the probability that the target value will be exceeded. This probability is shown in the probability of exceedance plot on the right.

Common uses for uncertainty analyses are to make general inferences, such as the following:

- Estimating the mean and standard deviation of the outputs.
- Estimating the probability the performance measure will exceed a specific threshold.
- Putting a reliability level on a function of the outputs, e.g., the range of function values that is likely to occur with some probability.
- Describing the likelihood of different potential outputs of the system.

Implicit in any uncertainty analysis are the assumptions that statistical distributions for the input values are correct and that the model is a sufficiently realistic description of the processes taking place in the system. Neither of these assumptions is likely to be entirely correct.

4.2 Sensitivity analyses

"Sensitivity analysis" is aimed at describing how much model output values are affected by changes in model input values. It is the investigation of the importance of imprecision or uncertainty in model inputs in a decision making or modeling process. The exact character of sensitivity analysis depends upon the particular context and the questions of concern. Sensitivity studies can provide a general assessment of model precision when used to assess system performance for alternative scenarios, as well as detailed information addressing the relative significance of errors in various parameters. As a result, sensitivity results should be of interest to the general public, federal and state management agencies, local watershed planners and managers, model users, and model developers.

Clearly, upper level management and the public may be interested in more general statements of model precision, and should be provided such information along with model predictions. On the other hand, detailed studies addressing the significance and interactions among individual parameters would likely be meaningful to model developers and some model users. They can use such data to interpret model results and to identify where efforts to improve models and their input values should be directed.

Initial sensitivity analysis studies could focus on two products:

(1) detailed results to guide research and assist model development efforts, and
(2) calculation of general descriptions of uncertainty associated with model predictions so that policy decisions can reflect both the modeling efforts best prediction of system performance and the precision of such predictions.

In the first case, knowing the relative uncertainty in model projections due to possible errors in different sets of parameters and input data should assist in efforts to improve the precision of model projections. This knowledge should also contribute to a better understanding of the relationships between model assumptions, parameters, data and model predictions.

For the second case, knowing the relative precision associated with model predictions should have a significant effect on policy development. For example, the analysis may show that, given data inadequacies, there are very large error bands associated with some model variables. When such large uncertainties exist, predictions should be used with appropriate skepticism.

Incremental strategies should be explored along with monitoring so that greater experience can accumulate to resolve some of those uncertainties.

Sensitivity analysis features are available in many linear and nonlinear programming (optimization) packages. They identify the changes in the values of the objective function and unknown decision variables given a change in the model input values, and a change in levels set for various constraints (Chapter V). Thus sensitivity analysis addresses the change in "optimal" system performance associated with changes in various parameter values, and also how "optimal" decisions would change with changes in resource constraint levels, or target output requirements. This kind of sensitivity analysis provides estimates of how much another unit of resource would be worth, or what "cost" a proposed change in a constraint places on the optimal solution. This information is of value to those making design decisions.

Various techniques have been developed to determine how sensitive model outputs are to changes in model inputs. Most approaches examine the affects of changes in a single parameter value or input variable assuming no changes in all the other inputs. Sensitivity analyses can be extended to examine the combined effects of multiple sources of error, as well.

Changes in particular model input values can affect model output values in different ways. It is generally true that only a relatively few input variables dominate or substantially influence the values of a particular output variable or performance indicator at a particular location and time. If the range of uncertainty of only some of the output data is of interest, then undoubtedly only those input data that significantly impact on the values of those output data need be included in the sensitivity analysis.

If input data estimates are based on repeated measurements, a frequency distribution can be estimated that characterizes natural variability. The shorter the record of measurements, the greater will be the uncertainty regarding the long-term statistical characteristics of that variability. If obtaining a sufficient number of replicate measurements is not possible, subjective estimates of input data ranges and probability distributions are often made. Using a mixture of subjective estimates and actual measurements does not affect the application of various sensitivity analysis methods that can use these sets or distributions of input values, but it may affect the conclusions that can be drawn from the results of these analyses.

It would be nice to have available accurate and easy-to-use analytical methods for relating errors in input data to errors in model outputs, and to errors in system performance indicator values that are derived from model output. Such analytical methods do not exist for complex simulation models. However methods based on simplifying assumptions and approximations can be used to yield useful sensitivity information. Some of these are reviewed in the remainder of this chapter.

4.2.1 Sensitivity coefficients

One measure of sensitivity is the sensitivity coefficient. This is the derivative of a model output variable with respect to an input variable or parameter. A number of sensitivity analysis methods use these coefficients. First-order and approximate first-order sensitivity analyses are two such methods that will be discussed later. The difficulty of

1. obtaining the derivatives for many models,
2. needing to assume mathematical (usually linear) relationships when obtaining estimates of derivatives by making small changes of input data values near their nominal or most likely values, and
3. having large variances associated with most hydrologic process models have motivated the replacement of analytical methods by numerical and statistical approaches to sensitivity analysis.

Implicit in any sensitivity analysis are the assumptions that statistical distributions for the input values are correct and that the model is a sufficiently realistic description of the processes taking place in the system. Neither of these assumptions is likely to be entirely correct.

The importance of the assumption that the statistical distributions for the input values are correct is easy to check by using different distributions for the input parameters. If the outputs vary significantly, then the output is sensitive to the specification of the input distributions and hence they should be defined with care. A relatively simple deterministic sensitivity analysis can be of value here (Benaman, 2002). A sensitivity coefficient can be used to measure the magnitude of change in an output variable Q per unit change in the magnitude of an input parameter value P from its base value P_o. Let SI_{PQ} be the sensitivity index for an output variable Q with respect to a change ΔP in the value of the input variable P from its base value P_o. Noting that the value of the output $Q(P)$ is a function of P, the sensitivity index could be defined as

$$SI_{PQ} = [\, Q(P_o + \Delta P) - Q(P_o - \Delta P)\,] / 2\,\Delta P \qquad (1)$$

Other sensitivity indices could be defined (McCuen 1973). Letting the index i represent a decrease and j represent an increase in the parameter value from its base value P_o, the sensitivity index SI_{PQ} for parameter P and output variable Q is could be defined as

$$SI_{PQ} = \{\, |(Q_o - Q_i) / (P_o - P_i)| + |(Q_o - Q_j) / (P_o - P_j)|\,\} / 2 \qquad (2)$$

or

$$SI_{PQ} = \max \{\, |(Q_o - Q_i) / (P_o - P_i)|,\ |(Q_o - Q_j) / (P_o - P_j)|\,\} \qquad (3)$$

A dimensionless expression of sensitivity is the elasticity index, EI_{PQ}, that measures the relative change in output Q for a relative change in input P could be defined as

$$EI_{PQ} = [P_o / Q(P_o)]\, SI_{PQ} \qquad (4)$$

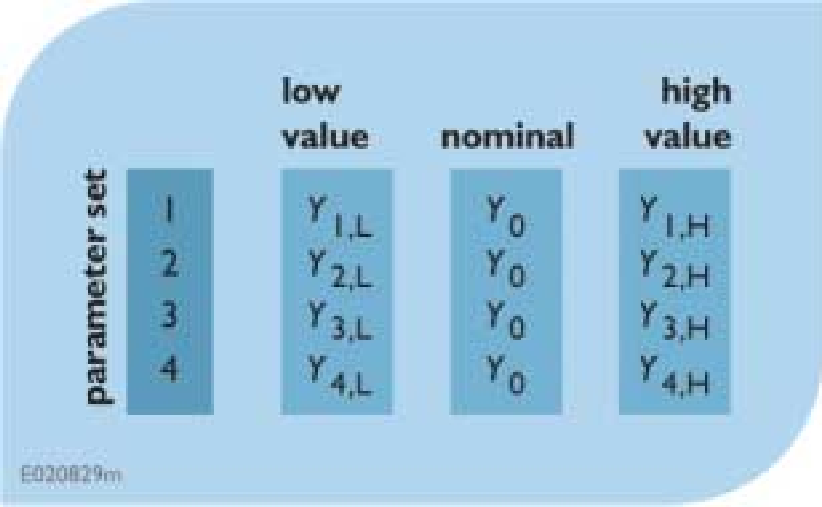
4.2.2 A simple deterministic sensitivity analysis procedure

This deterministic sensitivity analysis approach is very similar those most often employed in the engineering economics literature. It is based on the idea of varying one uncertain parameter value, or set of parameter values, at a time. The ideas are applied to a water quality example to illustrate their use.

The output variable of interest can be any performance measure or indicator. Thus one does not know if more or less of a given variable is better or worse. Perhaps too much and/or too little is undesirable. The key idea is that, whether employing physical measures or economic metrics of performance, various parameters (or sets of associated parameters) are assigned high and low values. Such ranges may reflect either the differences between the minimum and maximum values for each parameter, the 5 and 95 percentiles of a parameters distribution, or points corresponding to some other criteria. The system model is then run with the various alternatives, one at a time, to evaluate the impact of those errors in various sets of parameter values on the output variable.

Table 1 illustrates the character of the results that one would obtain. Here Y_0 is the nominal value of the model output when all parameters assume the estimated best values, and $Y_{i,L}$ and $Y_{i,H}$ are the values obtained by increasing or decreasing the values of the ith set of parameters.

Table 1. Sensitivity of model output Y to possible errors in four parameter sets containing a single parameter or a group of parameters that vary together.



parameter set	low value	nominal	high value
1	$Y_{1,L}$	Y_0	$Y_{1,H}$
2	$Y_{2,L}$	Y_0	$Y_{2,H}$
3	$Y_{3,L}$	Y_0	$Y_{3,H}$
4	$Y_{4,L}$	Y_0	$Y_{4,H}$

A simple water quality example is employed to illustrate this deterministic approach to sensitivity analysis. The analysis techniques illustrated here are just as applicable to complex models. The primary difference is that more work would be required to evaluate the various alternatives with a more complex model, and the model responses might be more complicated.

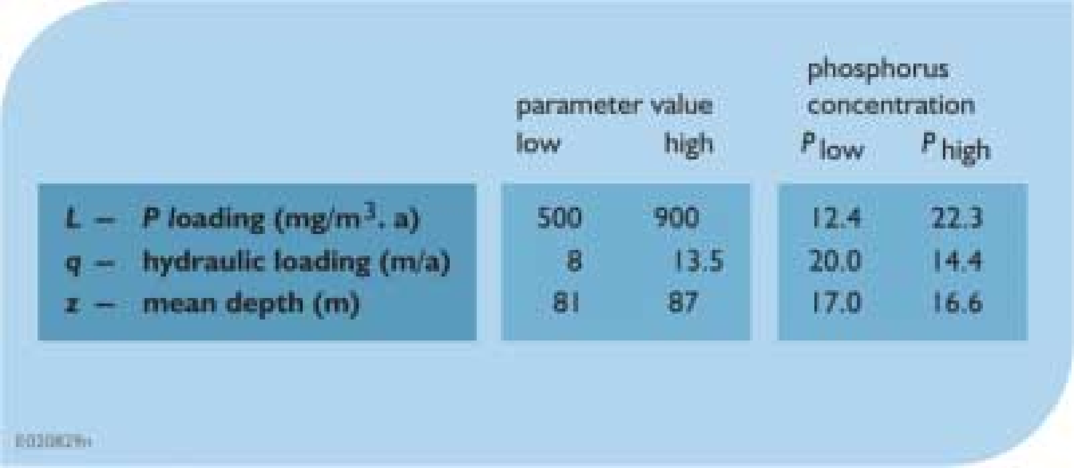
The simple water quality model is provided by Vollenweider's empirical relationship for the average phosphorus concentration in lakes (Vollenweider, 1976). He found that the phosphorus concentration, P (mg/m^3), is a function of the annual phosphorus loading rate, L (mg/m^2•a), the annual hydraulic loading, q (m/a or more exactly m^3/m^2•a), and the mean water depth, z (m).

$$P = (L/q) / [\, 1 + (z/q)^{0.5} \,] \tag{5}$$

L/q and P have the same units; the denominator is an empirical factor that compensates for nutrient recycling and elimination within the aquatic lake environment.

Data for Lake Ontario in North America would suggest that reasonable values of the parameters are $L = 680$ mg/m^3; $q = 10.6$ m/a; and $z = 84$ m, yielding $P = 16.8$ mg/m^3. Values of phosphorus concentrations less than 10 mg/m^3 are considered oligotrophic, whereas values greater than 20 mg/m^3 generally correspond to eutrophic conditions. Reasonable ranges reflecting possible errors in the three parameters yield the values in Table 2.

Table 2. Sensitivity of estimates of phosphorus concentration (mg/m^3) to model parameter values. The two right most values in each row correspond to the Low and High values of the parameter, respectively



	parameter value low	high	phosphorus concentration P_{low}	P_{high}
L – P loading (mg/m^3. a)	500	900	12.4	22.3
q – hydraulic loading (m/a)	8	13.5	20.0	14.4
z – mean depth (m)	81	87	17.0	16.6

One may want to display these results so they can be readily visualized and understood. A tornado diagram (Eschenbach, 1992) would show the lower and upper values of P obtained from variation of each parameter, with the parameter with the widest limits displayed on top, and the parameter having smallest limits on the bottom. Tornado diagrams (Figure 12) are easy to construct and can include a large number of parameters without becoming crowded.

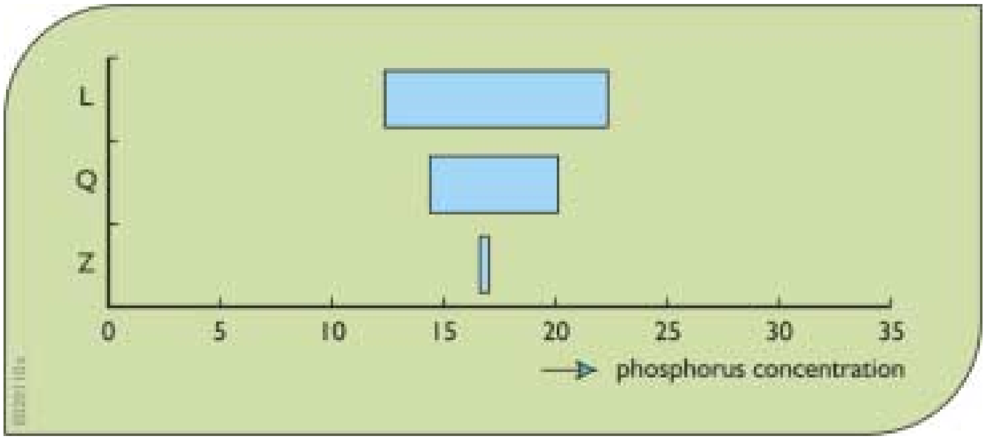


Figure 12. A Tornado diagram showing the range of the output variable representing phosphorus concentrations for high and low values of each of the parameter sets. Parameters are sorted so that the largest range is on top, and the smallest on the bottom.

An alternative to tornado diagrams is a Pareto chart showing the width of the uncertainty range associated with each variable, ordered from largest to smallest. A Pareto chart is illustrated in Figure 13.

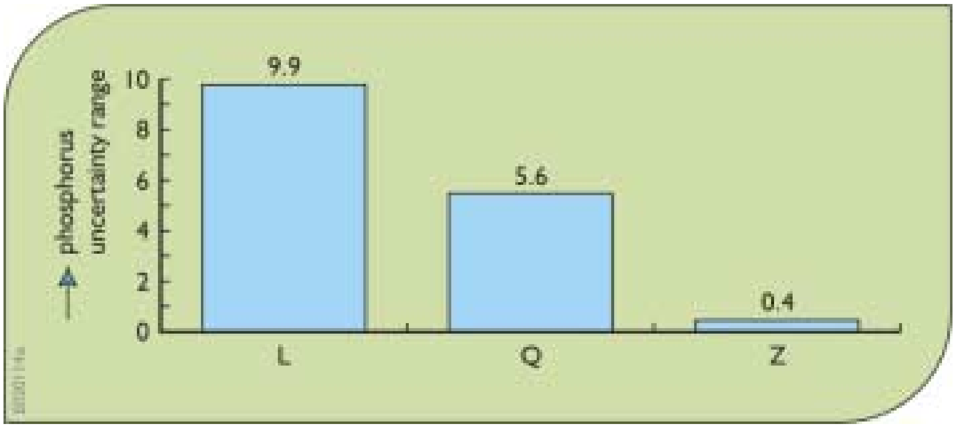


Figure 13. A Pareto Chart showing the range of the output variable representing phosphorus concentrations resulting from high and low values of each parameter set considered.

Another visual presentation is a spider plot showing the impact of uncertainty in each parameter on the variable in question, all on the same graph (Eschenback, 1992; DeGarmo, 1993, p. 401). A spider plot, Figure 14, shows the particular functional response of the output to each parameter on a common scale, so one needs a common metric to represent changes in all of the parameters. Here we use percentage change from the nominal or best values.

Spider plots are a little harder to construct than tornado diagrams, and can generally include only 4 - 5 variables without becoming crowded. However, they provide a more complete view of the relationships between each parameter and the performance measure. In particular, a spider plot reveals nonlinear relationships and the relative sensitivity of the performance measure to (percentage) changes in each variable.

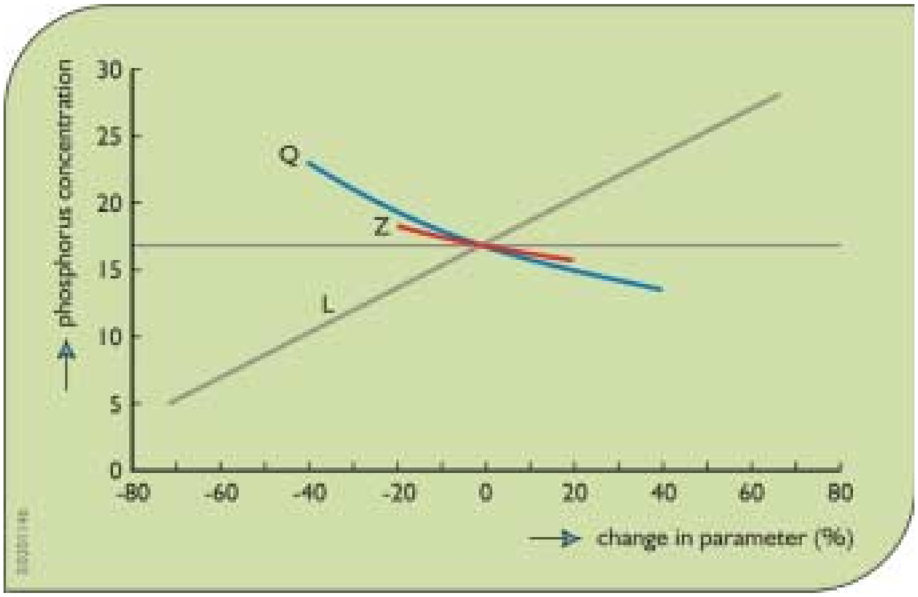


Figure 14. Spider Plot illustrates the relationships between model output describing phosphorus concentrations and variations in each of the parameter sets, expressed as a percentage deviation from their nominal values.

In the spider plot, the linear relationship between *P* and *L* and the gentle nonlinear relationship between *P* and *q* is illustrated. The range for *z* has been kept small given the limited uncertainty associated with that parameter.

4.2.3 Multiple errors and interactions

An important issue that should not be ignored is the impact of simultaneous errors in more than one parameter. Probabilistic methods directly address the occurrence of simultaneous errors, but the correct joint distribution needs to be employed. With simple sensitivity analysis procedures, errors in parameters are generally investigated one at a time, or in groups. The idea of considering pairs or sets of parameters is discussed here.

Groups of factors. It is often the case that reasonable error scenarios would have several parameters changing together. For this reason, the alternatives have been called parameter sets. For example, possible errors in water depth would be accompanied with corresponding variations in aquatic vegetation and chemical parameters. Likewise, alternatives related to changes in model structure might be accompanied with variations in several parameters. In other cases, there may be no causal relationship among possible errors (such as model structure versus inflows at the boundary of the modeled region), but they might still interact to effect the precision of model predictions.

Combinations. If one or more non-grouped parameters interact in significant ways, then combinations of one or more errors should be investigated. However, one immediately runs into a combinatorial problem. If each of m parameters can have 3 values (high, nominal, and low) there are 3^m combinations, as opposed to $2m + 1$ if each parameter is varied separately. [For m = 5, the differences are $3^5 = 243$ versus $2(5)+1 = 11$.] These numbers can be reduced by considering instead only combinations of extremes so that only $2^m + 1$ cases need be considered [$2^5 + 1 = 33$], which is a more manageable number. However, all of the parameters would be at one extreme or the other, and such situations would be very unusual.

Two factors at a time. A compromise is to consider all pairs of two parameters at a time. There are m(m-1)/2 possible pairs of m parameters. Each parameter has a high and low value. Since there are 4 combinations of high and low values for each pair, there are a total of 2m(m-1) combinations. [For m = 5 there are 40 combinations of two parameters each having two values.]

The presentation of these results could be simplified by displaying for each case only the maximum error, which would result in m(m-1)/2 cases that might be displayed in a Pareto diagram. This would allow identification of those combinations of two parameters that might yield the largest errors and thus are of most concern.

For the water quality example, if one plots the absolute value of the error for all four combinations of high (+) and low (-) values for each pair of parameters, they obtain Figure 15.

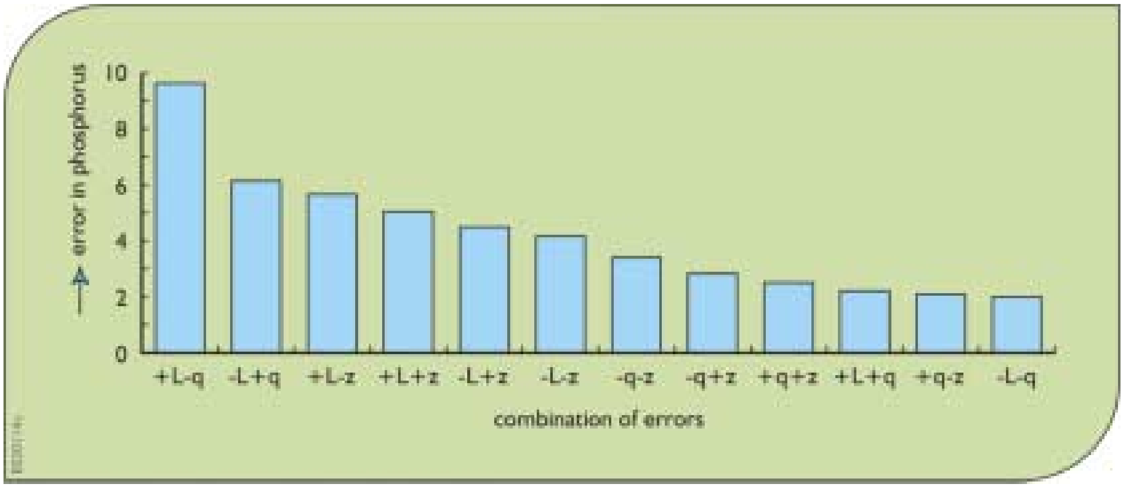


Figure 15. Pareto diagram showing errors in phosphorus concentrations for all combinations of pairs of input parameters errors. A + indicates a high value, and a - indicates a low value for indicated parameter. *L* is the phosphorus loading rate, *q* is the hydraulic loading, and *z* is the mean lake depth.

Considering only the worst error for each pair of variables yields Figure 16.

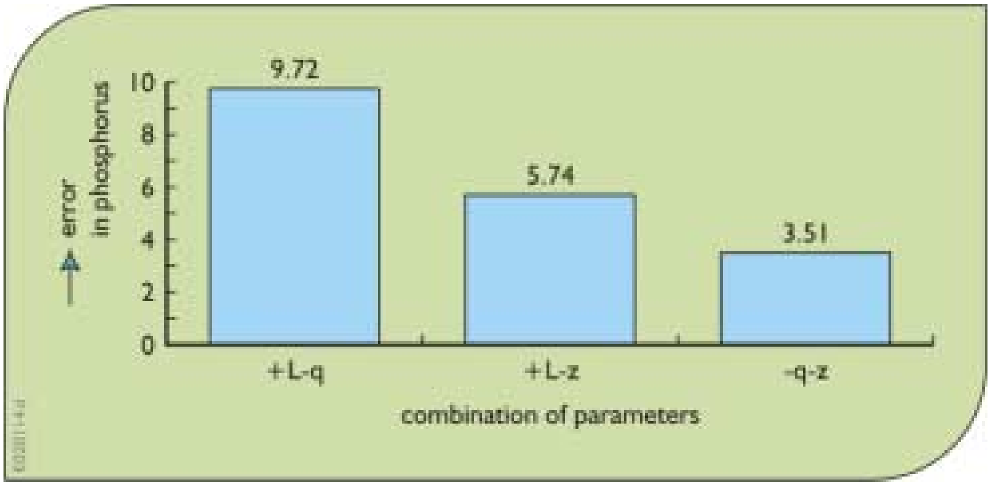


Figure 16. Pareto diagram showing worst error combinations for each pair of input parameters. A '+' indicates a high value, and a '–' indicates a low value for indicated parameter.

Here we see, as is no surprise, that the worst error results from the most unfavorable combination of L and q values. If both parameters have their most unfavorable values, the predicted phosphorus concentration would be 27 mg/m^3.

Looking for non-linearities. One might also display in a Pareto diagram the maximum error for each pair as a percentage of the sum of the absolute values of the maximum error from each parameter separately. The ratio of the joint error to the individual errors would illustrate potentially important nonlinear interactions. If the model of the system and the physical measure or economic metric were strictly linear, then the individual ratios should add to one.

4.2.4 First-order sensitivity analysis

The above deterministic analysis has trouble representing reasonable combinations of errors in several parameter sets. If the errors are independent, it is highly unlikely that any two sets would actually be at their extreme ranges at the same time. By defining probability distributions of the values of the various parameter sets, and specifying their joint distributions, a probabilistic error analysis can be conducted. In particular, for a given performance indicator, one can use multivariate linear analyses to evaluate the approximate impact on the performance indices of uncertainty in various parameters. As shown below, the impact depends upon the square of the sensitivity coefficients (partial derivatives) and the variances and covariances of the parameter sets.

For a performance indicator $I = F(Y)$, which is a function $F(\bullet)$ of model outputs Y, that are in turn a function $g(P)$ of input parameters P, one can use a multivariate Taylor series approximation of F to obtain the expected value and variance of the indicator:

$$\begin{aligned} \mathrm{E}[I] &= F(\text{ based on mean values of input parameters}) \\ &\quad + (1/2)\ \{\Sigma_i \Sigma_j\ [\partial \mathrm{F}^2/\partial P_i \partial P_j]\ \mathrm{Cov}\ [P_i, P_j]\ \} \end{aligned} \qquad (6)$$

and

$$\mathrm{Var}[I] = \Sigma_i \Sigma_j\ (\partial F/\ \partial P_i)(\partial F/\ \partial P_j)\ \mathrm{Cov}\ [P_i, P_j] \qquad (7)$$

where $(\partial F/\partial P_i)$ are the partial derivative of the function F with respect to P_i evaluated at the mean value of the input parameters P_i, and $\partial F^2/\partial P_i \partial P_j$ are the second partial derivatives. The covariance of two random input parameters P_i and P_j is the expected value of the product of differences between the values and their means.

$$\mathrm{Cov}[P_i, P_j] = \mathrm{E}[(P_i - \mathrm{E}[P_i])(\ P_j - \mathrm{E}[P_j])] \qquad (8)$$

If all the parameters are independent of each other, and the second-order terms in the expression for the mean E[I] are neglected, one obtains

$$\mathrm{E}[I] = F(\text{ based on mean values of input parameters}) \qquad (9)$$

and

$$\mathrm{Var}\ [I] = \Sigma_i\ [\partial F/\ \partial P_i]^2\ \mathrm{Var}\ [P_i] \qquad (10)$$

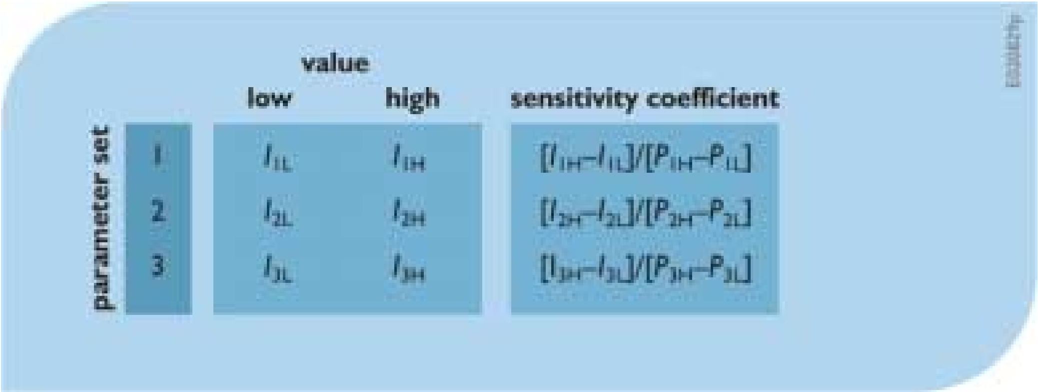
(Benjamin and Cornell, 1970). Equation 6 for E[I] shows that in the presence of substantial uncertainty, the mean of the output from nonlinear systems is not simply the system output corresponding to the mean of the parameters (Gaven and Burges, 1981, p. 1523). This is true for any nonlinear function.

Of interest in the analysis of uncertainty is the approximation for the variance Var[I] of indicator I. In Equation 10 the contribution of P_i to the variance of I equals Var[P_i] times $[\partial F/\ \partial P_i]^2$, which are the squares of the sensitivity coefficients for indicator I with respect to each input parameter value P_i.

4.2.4.1 An example of first-order sensitivity analysis

It may appear that first-order analysis is difficult because the partial derivatives of the performance indicator I are needed with respect to the various parameters. However, reasonable approximations of these sensitivity coefficients can be obtained from the simple sensitivity analysis described in Table 3, as shown below. In that table, three different parameter sets, P_i, are defined in which one parameter of the set is at its high value, P_{iH}, and one is at its low value, P_{iL}, to produce corresponding values (called high, I_{iH}, and low, I_{iL}) of a system performance indicator I.

Table 3. Approximate parameter sensitivity coefficients.



parameter set	value low	value high	sensitivity coefficient
1	I_{1L}	I_{1H}	$[I_{1H}–I_{1L}]/[P_{1H}–P_{1L}]$
2	I_{2L}	I_{2H}	$[I_{2H}–I_{2L}]/[P_{2H}–P_{2L}]$
3	I_{3L}	I_{3H}	$[I_{3H}–I_{3L}]/[P_{3H}–P_{3L}]$

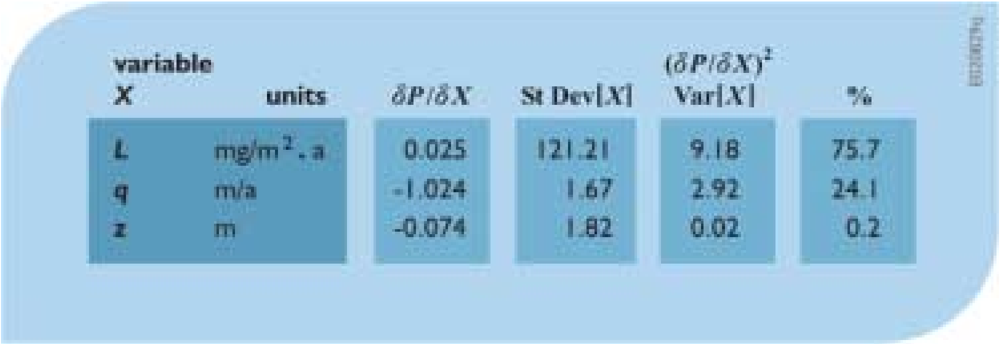
It is then necessary to estimate some representation of the variances of the various parameters with some consistent procedure. For a normal distribution, the distance between the 5 and 95 percentiles is 1.645 standard deviations on each side of the mean, or 2(1.645) = 3.3 standard deviations. Thus, if the high/low range is thought of as approximately a 5-95 percentile range for a normally distributed variate, a reasonable approximation of the variance might be

$$\mathrm{Var}[P_i] = \{ [P_{iH}–P_{iL}]/3.3 \}^2. \tag{11}$$

This is all that is needed. Use of these average sensitivity coefficients is very reasonable for modeling the behavior of the system performance indicator *I* over the indicated ranges.

As an illustration of the method of first-order uncertainty analysis, consider the lake quality problem described above. The "system performance indicator" in this case is the model output, the phosphorus concentration *P*, and the input parameters, now denoted as $X = L$, q, and z. The standard deviation of each parameter is assumed to be the specified range divided by 3.3. Average sensitivity coefficients $\partial P/\partial X$ were calculated. The results are reported in the table below.

Table 4. Calculation of approximate parameter sensitivity coefficients.



variable X	units	$\delta P/\delta X$	St Dev[X]	$(\delta P/\delta X)^2$ Var[X]	%
L	$mg/m^2 \cdot a$	0.025	121.21	9.18	75.7
q	m/a	-1.024	1.67	2.92	24.1
z	m	-0.074	1.82	0.02	0.2

Assuming the parameter errors are independent:

$$\mathrm{Var}[P] = 9.18 + 2.92 + 0.02 = 12.12 \qquad (12)$$

The square root of 12.12 is the standard deviation and equals 3.48. This agrees well with a Monte Carlo analysis reported below.

Note that 100*(9.18/12.12), or about 76% of the total parameter error variance in the phosphorus concentration *P* is associated in the phosphorus loading rate *L* and the remaining 24% is associated with the hydrologic loading *q*. Eliminating the uncertainty in *z* would have a negligible impact on the overall model error. Likewise, reducing the error in *q* would at best have a modest impact on the total error.

Due to these uncertainties, the estimated phosphorus concentration has a standard deviation of 3.48. Assuming the errors are normally distributed, and recalling that ± 1.645 standard deviations around the mean define a 5-95 percentile interval, the 5-95 percentile interval would be about

$$16.8 \pm 1.645\ (3.48)\ \mathrm{mg/m^3} = 16.8 \pm 5.7\ \mathrm{mg/m^3} = 11.1 \text{ to } 22.5\ \mathrm{mg/m^3}. \qquad (13)$$

These error bars indicate there is substantial uncertainty associated with the phosphorus concentration *P*, primarily due to uncertainty in the loading rate *L*.

The upper bound of 22.6 $\mathrm{mg/m^3}$ is considerably less than the 27 $\mathrm{mg/m^3}$ that would be obtained if both *L* and *q* had their most unfavorable values. In a probabilistic analysis with independent errors, such a combination is highly unlikely.

4.2.4.2 Warning on accuracy.

First-order uncertainty analysis is indeed an approximate method based upon a linearization of the response function represented by the full simulation model. It may provide inaccurate estimates of the variance of the response variable for nonlinear systems with large uncertainty in the parameters. In such cases Monte Carlo simulation (discussed below and in Chapter VII) or the use of higher-order approximation may be required. Beck (1987, p. 1426) cites studies that found that Monte Carlo and first-order variances were not appreciably different, and a few studies that found specific differences. Differences are likely to arise when the distributions used for the parameters are bimodal (or otherwise unusual), or some rejection algorithm is used in the Monte Carlo analysis to exclude some parameter combinations. Such errors can result in a distortion in the ranking of predominant sources of uncertainty. However, in most cases very similar results were obtained.

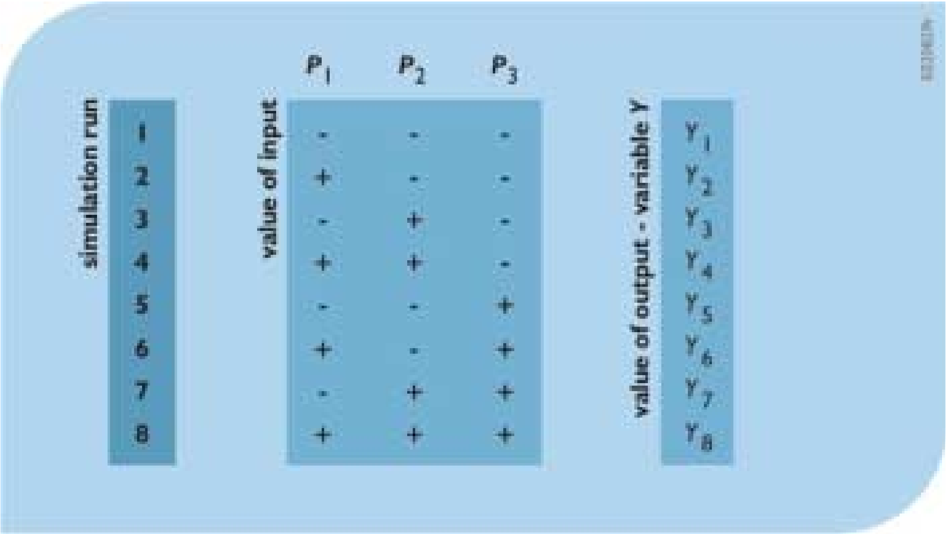
4.2.5 Fractional factorial design method

An extension of first-order sensitivity analysis would be a more complete exploration of the response surface using a careful statistical design. First consider a complete factorial design. Input data are divided into discrete "levels'. The simplest case is two levels. These two levels can be defined as a nominal value, and a high (low) value. Simulation runs are made for all combinations of parameter levels. For *n* different inputs, this would require 2^n simulation runs. Hence for a three-input variable or parameter problem, 8 runs would be required. If 4 discrete levels of each input variable or parameter were allowed to provide a more reasonable description of a continuous variable, the three-input data problem would require 4^3 or 64 simulation runs. Clearly this is not a useful tool for large regional water resources simulation models.

A fractional factorial design involves simulating only a fraction of what is required from a full factorial design method. The loss of information prevents a complete analysis of the impacts of each input variable or parameter on the output.

To illustrate the fractional factorial design method, consider the two-level with three-input variable or parameter problem. Table 5 below shows the 8 simulations required for a full factorial design method. The '+' and the '–' show the upper and lower levels of each input variable or parameter P_i where $i = 1, 2, 3$. If all 8 simulations were performed, seven possible effects could be estimated. These are the individual effects of the three inputs P_1, P_2, and P_3, the three two-input variable or parameter interactions, $(P_1)(P_2)$, $(P_1)(P_3)$, and $(P_2)(P_3)$, and the one three-input variable or parameter interaction $(P_1)(P_2)(P_3)$.

Table 5. A three-input factorial design.



simulation run	value of input P_1	P_2	P_3	value of output - variable Y
1	-	-	-	Y_1
2	+	-	-	Y_2
3	-	+	-	Y_3
4	+	+	-	Y_4
5	-	-	+	Y_5
6	+	-	+	Y_6
7	-	+	+	Y_7
8	+	+	+	Y_8

Consider an output variable Y, where Y_j is the value of Y in the jth simulation run. Then an estimate of the effect, denoted $\delta(Y|P_i)$, that input variable or parameter P_i has on the output variable Y, is the average of the four separate effects of varying P_i:

For i = 1:

$$\delta(Y \mid P_1) = 0.25 \, [\, (Y_2-Y_1)+(Y_4-Y_3)+(Y_6-Y_5)+(Y_8-Y_7) \,] \tag{14}$$

Each difference in parentheses is the difference between a run in which P_1 is at its upper level and a run in which P_1 is at its lower level, but the other two parameter values, P_2 and P_3, are unchanged. If the effect is equal to 0, then, in this case, P_1 has no impact on the output variable Y.

Similarly the effects of P_2 and P_3, on variable Y can be estimated as:

$$\delta(Y \mid P_2) = 0.25 \, \{ \, (Y_3-Y_1)+(Y_4-Y_2)+(Y_7-Y_5)+(Y_8-Y_6) \, \} \tag{15}$$

and

$$\delta(Y \mid P_3) = 0.25 \, \{ \, (Y_5-Y_1)+(Y_6-Y_2)+(Y_7-Y_3)+(Y_8-Y_4) \, \} \tag{16}$$

Consider next the interaction effects between P_1 and P_2. This is estimated as the average of the difference between the average P_1 effect at the upper level of P_2, and the average P_1 effect at the lower level of P_2. This is the same as the difference between the average P_2 effect at the upper level of P_1 and the average P_2 effect at the lower level of P_1:

$$\begin{aligned}\delta(Y \mid P_1, P_2) &= (1/2) \, \{ \, [\, (Y_8-Y_7) + (Y_4-Y_3)]/2 - [\, (Y_2-Y_1) + (Y_6-Y_5)]/2 \, \} \\ &= (1/4) \, \{ \, [\, (Y_8-Y_6)+(Y_4-Y_2)] - [\, (Y_3-Y_1) + (Y_7-Y_5)] \, \} \end{aligned} \tag{17}$$

Similar equations can be derived for looking at the interaction effects between P_1 and P_3, and between P_2 and P_3 and the interaction effects among all three inputs P_1, P_2, and P_3.

Now assume only half of the simulation runs were performed, perhaps runs 2, 3, 5 and 8 in this example. If only outputs Y_2, Y_3, Y_5, and Y_8 are available, for our example:

$$\delta(Y \mid P_3) = \square(Y \mid P_1, P_2) = 0.5 \, \{ \, (Y_8 - Y_3) - (Y_2 - Y_5) \, \} \tag{18}$$

The separate effects of P_3 and of P_1P_2 are not available from the output. This is the loss in information resulting from fractional instead of complete factorial design.

4.2.6 Monte Carlo sampling methods

The Monte Carlo method of performing sensitivity analyses, illustrated in Figure 16, first selects a random set of input data values drawn from their individual probability distributions. These values are then used in the simulation model to obtain some model output variable values. This process is repeated many times, each time making sure the model calibration is valid for the input data values chosen. The end result is a probability distribution of model output variables and system performance indices that results from variations and possible errors in all of the input values.

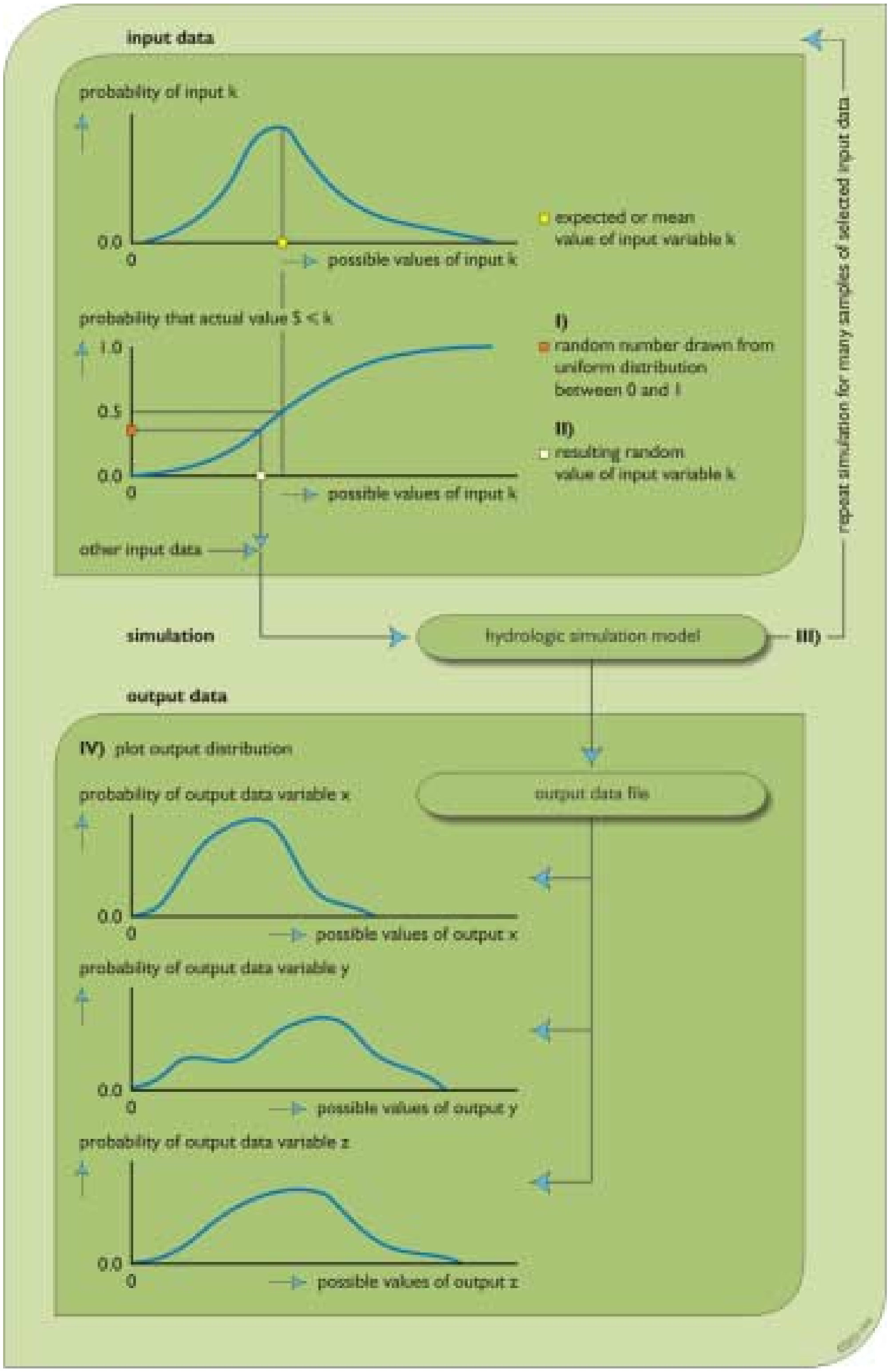


Figure 16. Monte Carlo sampling and simulation procedure for finding distributions of output variable values based on distributions, for specified reliability levels, of input data values. This technique can be applied to one or more uncertain input variables at a time. The output distributions will reflect the combined effects of this input uncertainty over the specified ranges.

Using a simple Monte Carlo analysis, values of all of the parameter sets are selected randomly from distributions describing the individual and joint uncertainty in each, and then the modeled system is simulated to obtain estimates of the selected performance indices. This must be done many times (often well over 100) to obtain a statistical description of system performance variability. The number of replications needed is generally not dependent on the number of parameters whose errors are to be analyzed. One can include in the simulation the uncertainty in parameters as well as natural variability. This method can evaluate the impact of single or multiple uncertain parameters.

A significant problem that arises in such simulations is that some combinations of parameter values result in unreasonable models. For example, model performance with calibration data sets might be inconsistent with available data sets. The calibration process places interesting constraints on different sets of parameter values. Thus, such Monte Carlo experiments often contain checks that exclude combinations of parameter values that are unreasonable. In these cases the generated results are conditioned on this validity check.

Whenever sampling methods are used, one must consider possible correlations among input data values. Sampling methods can handle spatial and temporal correlations that may exist among input data values, but the existence of correlation requires defining appropriate conditional distributions.

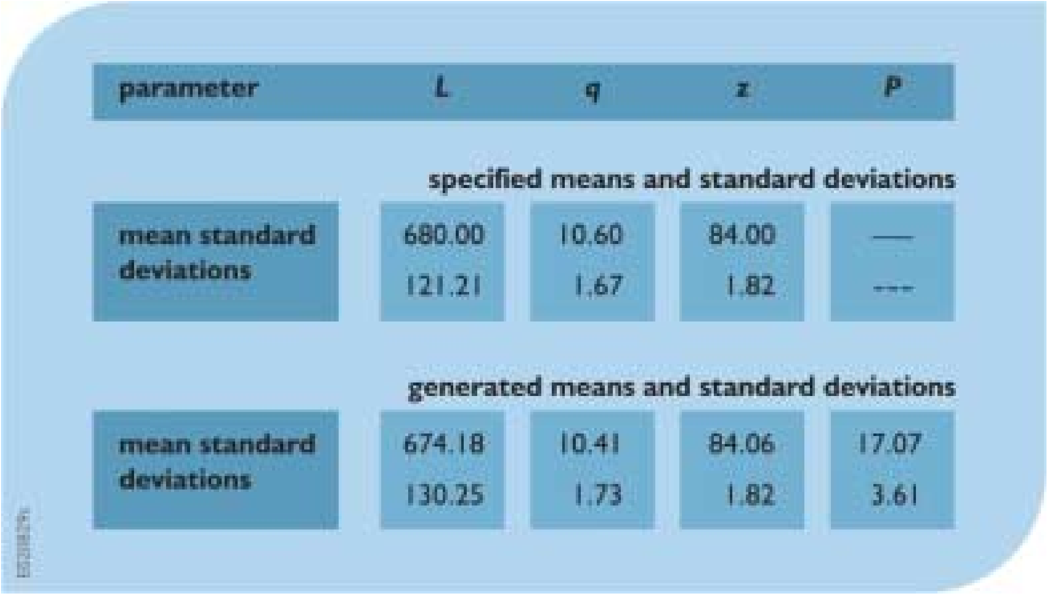
One major limitation of applying Monte Carlo methods to estimate ranges of risk and uncertainty for model output variable values, and system performance indicator values based on these output variable values, is the computing time required. To reduce the computing times needed to perform sensitivity analyses using sampling methods, some tricks and as well as stratified sampling methods are available. The discussion below illustrates the idea of a simple modification (or trick) using a "standardized" Monte Carlo analysis. The more general Latin Hypercube Sampling procedure is also discussed.

4.2.6.1 Simple Monte Carlo sampling

To illustrate the use of Monte Carlo sampling methods consider again Vollenweider's empirical relationship, Equation 5, for the average phosphorus concentration in lakes (Vollenweider, 1976). Two hundred values of each parameter were generated independently from normal distributions with the means and variances as shown in Table 6.

The table contains the specified means and variances for the generated values of L, q and z, and also the actual values of the means and variances of the 200 generated values of L, q, z and also of the 200 corresponding generated output phosphorus concentrations, P. Figure 17 displays the distribution of the generated values of P.

Table 6. Monte Carlo analysis of lake phosphorus levels.



parameter	L	q	z	P
	specified means and standard deviations			
mean standard deviations	680.00 121.21	10.60 1.67	84.00 1.82	— ---
	generated means and standard deviations			
mean standard deviations	674.18 130.25	10.41 1.73	84.06 1.82	17.07 3.61

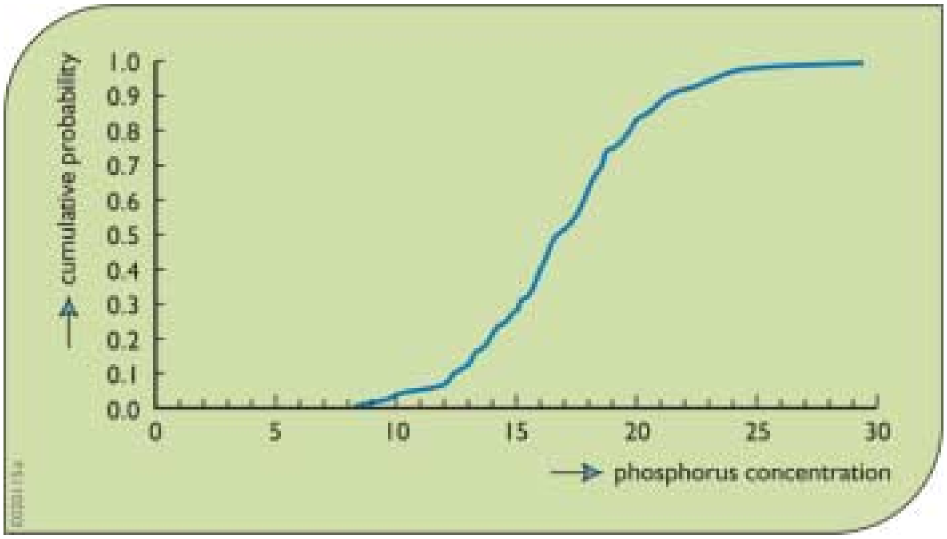


Figure 17. Distribution of lake phosphorus concentrations from Monte Carlo analysis

One can see that given the estimated levels of uncertainty, phosphorus levels could reasonably range from below 10 to above 25. The probability of generating a value greater than 20 mg/m^3 was 12.5%. The 5% to 95 percentile range was 11.1 to 23.4 mg/m^3. In the figure, the cumulative probability curve is rough because only 200 values of the phosphorus concentration were generated, but these are clearly enough to give a good impression of the overall impact of the errors.

4.2.6.2 Sampling uncertainty.

In this example, the mean of the 200 generated values of the phosphorus concentration, *P*, was 17.07. However a different set of random values would have generated a different set of *P* values as well. Thus it is appropriate to estimate the standard error, SE, of this average. The standard error equals the standard deviation σ of the *P* values divided by the square root of the sample size *n*:

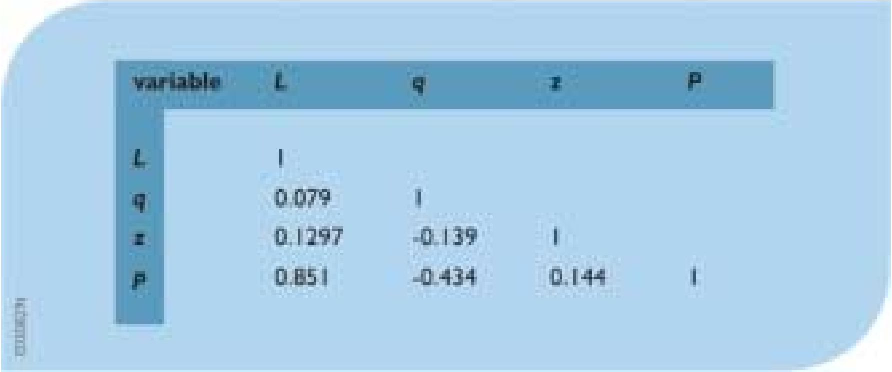
$$\mathrm{SE} = \sigma/(n)^{0.5} = 3.61/(200)^{0.5} = 0.25. \qquad (19)$$

From the central limit theorem of mathematical statistics, the average of a large number of independent values should have very nearly a normal distribution. Thus, 95% of the time, the true mean of *P* should be in the interval 17.1 ± 1.96 (0.25), or 16.6 to 17.6 mg/m^3 . This level of uncertainty reflects the observed variability of *P* and the fact that only 200 values were generated.

4.2.6.3 Making sense of the results.

A significant challenge with complex models is to determine from the Monte Carlo simulation which parameter errors are important. Calculating the correlation between each generated input parameter value and the output variable value is one way of doing this. As Table 7 below shows, based upon the magnitudes of the correlation coefficients, errors in *L* were most important, and those in *q* second in importance.

Table 7. Correlation analysis of Monte Carlo results.

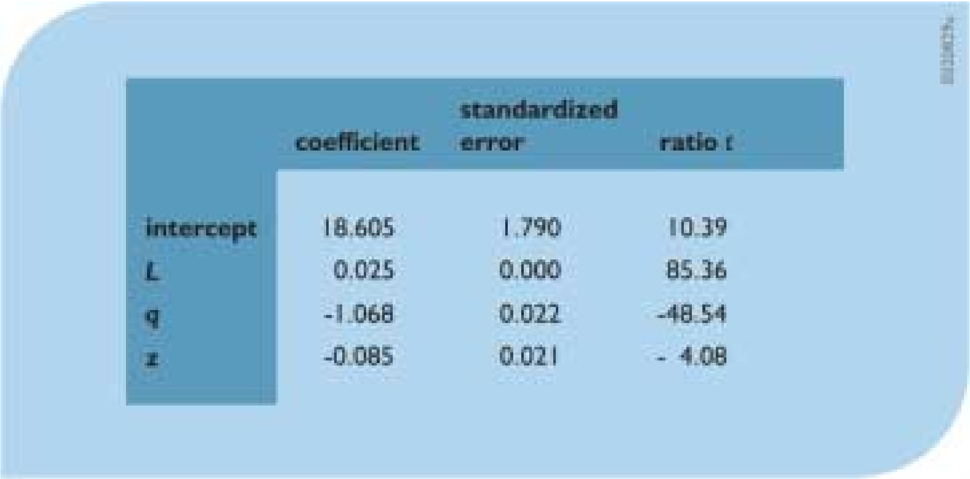


variable	*L*	*q*	*z*	*P*
L	1			
q	0.079	1		
z	0.1297	-0.139	1	
P	0.851	-0.434	0.144	1

One can also use regression to develop a linear model defining variations in the output based on errors in the various parameters. The results are shown in the Table 8. The fit is very good, and R^2 = 98%. If the model for *P* had been linear, a R^2 value of 100% should have resulted. All of the coefficients are significantly different from zero.

Note that the correlation between *P* and *z* was positive in Table 7, but the regression coefficient for *z* is negative. This occurred because there is a modest negative correlation between the generated *z* and *q* values. Use of partial correlation coefficients can also correct for such spurious correlations among input parameters.

Table 8. Results of Regression Analysis on Monte Carlo Results



	coefficient	standardized error	ratio t
intercept	18.605	1.790	10.39
L	0.025	0.000	85.36
q	-1.068	0.022	-48.54
z	-0.085	0.021	- 4.08

Finally we display a plot, Figure 18, based on this regression model illustrating the reduction in the variance of *P* that is due to dropping each variable individually. Clearly *L* has the biggest impact on the uncertainty in *P,* and *z* the least.

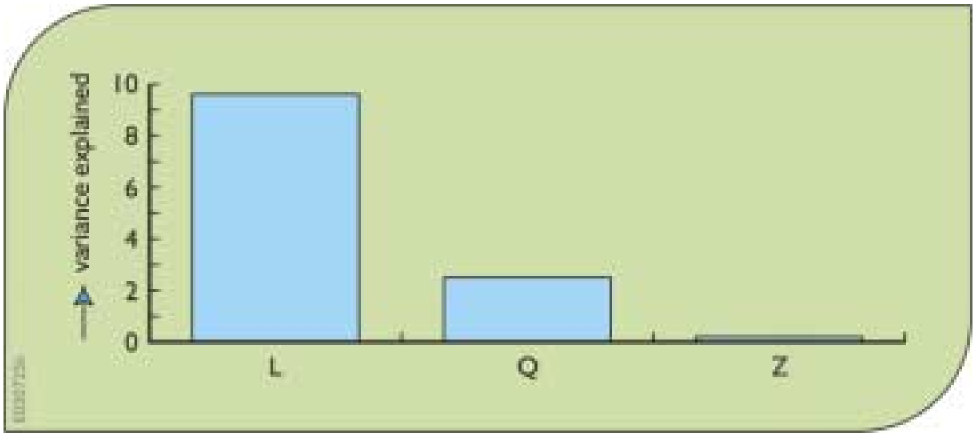
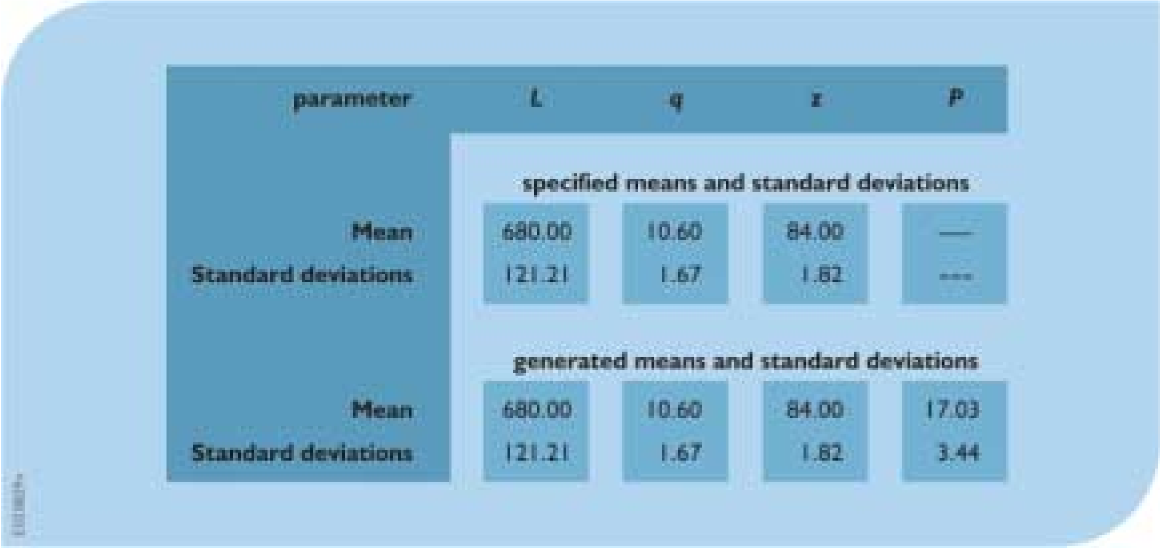


Figure 18. Reduction in the variance of *P* that is due to dropping from the regression model each variable individually. Clearly *L* has the biggest impact on the uncertainty in *P*, and *z* the least.

4.2.6.4 Standardized Monte Carlo analysis

Using a "standardized" Monte Carlo analysis, one could adjust the generated values of *L*, *q* and *z* above so that the generated samples actually have the desired mean and variance. While making that correction, one can also shuffle their values so that the correlations among the generated values for the different parameters are near zero, as is desired. This was done for the 200 generated values to obtain the statistics shown in Table 9.

Table 9. Standardized Monte Carlo analysis of lake phosphorus levels

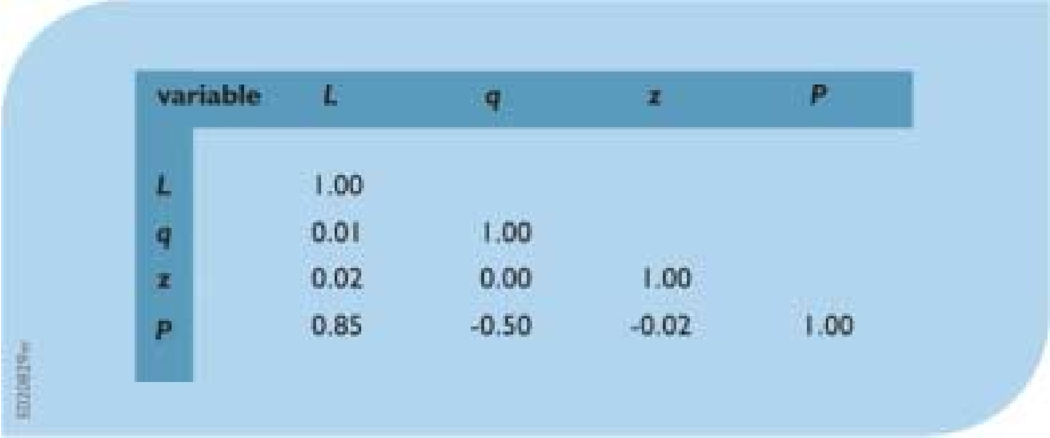


parameter	L	q	z	P
	specified means and standard deviations			
Mean	680.00	10.60	84.00	—
Standard deviations	121.21	1.67	1.82	---
	generated means and standard deviations			
Mean	680.00	10.60	84.00	17.03
Standard deviations	121.21	1.67	1.82	3.44

Repeating the correlation analysis from before (shown in Table 10) now yields much clearer results that are in agreement with the regression analysis. The correlation between *P* and both *q* and *z* are now negative as they should be. Because the generated values of the three parameters have been adjusted to be uncorrelated, the signal from one is not confused with the signal from another.

.

Table 10. Correlation analysis of standardized Monte Carlo results



variable	L	q	z	P
L	1.00			
q	0.01	1.00		
z	0.02	0.00	1.00	
P	0.85	-0.50	-0.02	1.00

The mean phosphorus concentration changed very little. It is now 17.0 instead of 17.1 mg/m^3.

Using control variates with a linear predictive model in conjunction with the standardized Monte Carlo variates, the standard deviation of the errors associated with the 200 observations is only 0.45. Thus the standard error for this estimate of the mean of P is $0.45/(200)^{0.5}$ or just 0.03. Thus this is a highly accurate result. The regressions were also repeated and yielded very similar results. The only real difference was that the parameter estimates had small standard errors and were more significant because of the elimination of correlation between the generated parameters.

4.2.6.5 Generalized likelihood estimation

Beven (1993) and Binley and Beven (1991) suggest a Generalized Likelihood Uncertainty Estimation (GLUE) technique for assessment of parameter error uncertainty using Monte Carlo simulation. It is described as a “formal methodology for some of the subjective elements of model calibration” (Beven, 1989, p. 47). The basic idea is to begin by assigning reasonable ranges for the various parameters and then to draw parameter sets from those ranges using a uniform or some similar (and flat) distribution. These generated parameter sets are then used on a calibration data set so that unreasonable combinations can be rejected, while reasonable values are assigned a posterior probability based upon a likelihood measure which may reflect several dimensions and characteristics of model performance.

Let $L(P_i) > 0$ be the value of the likelihood measure assigned to the i^{th} parameter set’s calibration sequence. Then the model predictions generated with parameter set/combination P_i are assigned posterior probability, $p(P_i)$.

$$p(P_i) = L(P_i) / \Sigma_j L(P_j) \qquad (20)$$

These probabilities reflect the form of Bayes theorem, which is well supported by probability theory (Devore, 1991). This procedure should capture reasonably well the dependence or correlation among parameters, because *reasonable* sequences will all be assigned larger probabilities, whereas sequences that are unable to reproduce the system response over the calibration period will be rejected or assigned small probabilities.

However, in a rigorous probabilistic framework, the L would be the likelihood function for the calibration series for particular error distributions. (This could be checked with available goodness-of-fit procedures; for example, Kuczera, 1988.) When relatively ad hoc measures are adopted for the likelihood measure with little statistical validity, the $p(P_i)$ probabilities are best described as pseudo probabilities or "likelihood" weights.

Another concern with this method is the potential efficiency. If the parameter ranges are too wide, a large number of unreasonable or very unlikely parameter combinations will be generated. These will either be rejected or else will have small probabilities and thus little effect on the analysis. In this case the associated processing would be a waste of effort. A compromise is to use some data to calibrate the model and to generate a prior or initial distribution for the parameters that is at least centered in the best range (Beven 1993, p. 48). Then use of a different calibration period to generate the $p(P_i)$ allows an updating of those initial probabilities to reflect the information provided by the additional calibration period with the adopted likelihood measures.

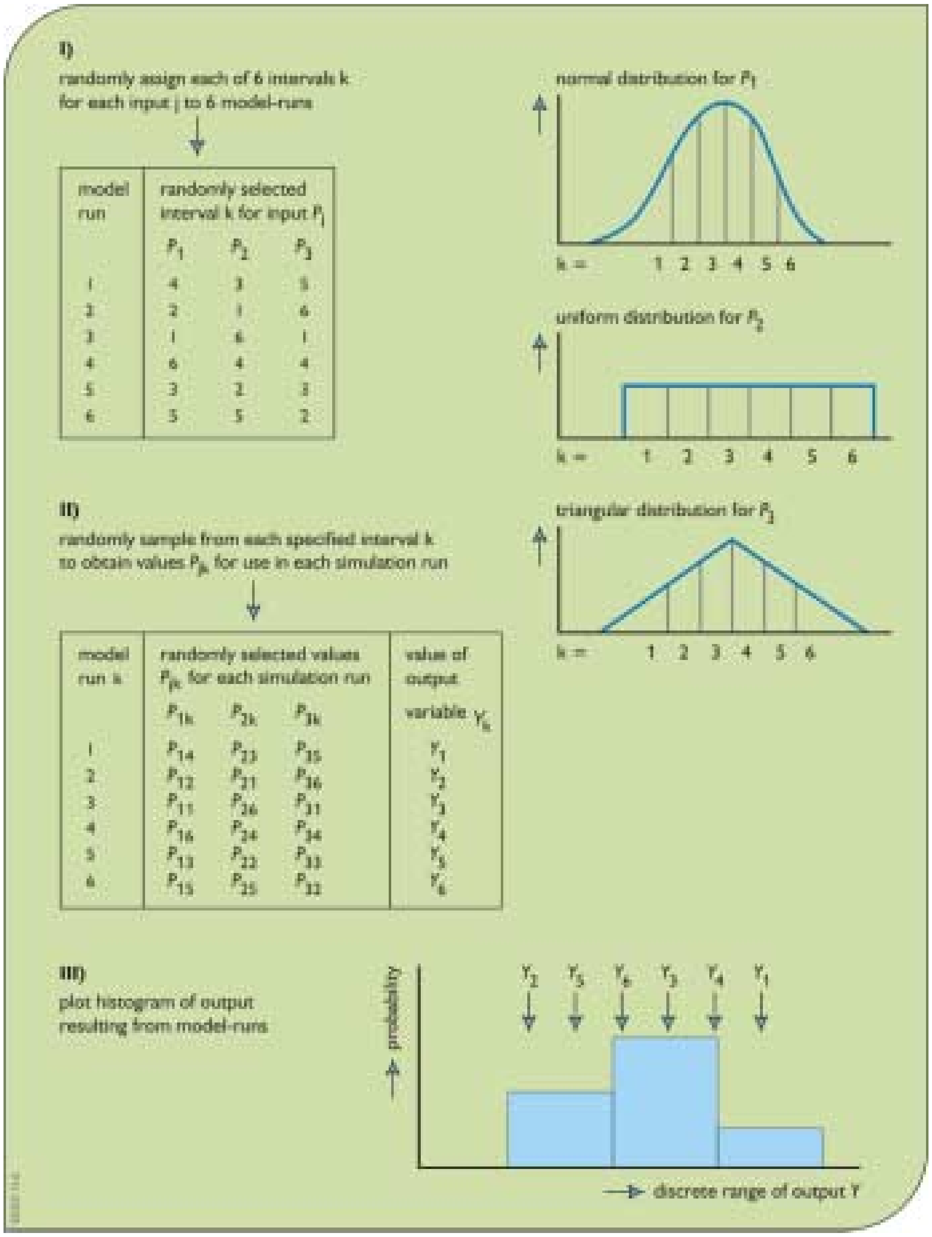
After the accepted sequences are used to generate sets of predictions, the likelihood weights would be used in the calculation of means, variances and quantiles, rather than the customary procedure of giving all the generated realizations equal weight. The resulting conditional distribution of system output reflects the initial probability distributions assigned to parameters, the rejection criteria, and the likelihood measure adopted to assign "likelihood" weights.

4.2.7 Latin hypercube sampling

For the simple Monte Carlo simulations described above, with independent errors, a probability distribution is assumed for each input parameter or variable. In each simulation run, values of all input data are obtained from sampling those individual and independent distributions. The value generated for an input parameter or variable is usually independent of what that value was in any previous run, or what other input parameter or variable values are in the same run. This simple sampling approach can result in a clustering of parameter values and hence both redundancy of information from repeated sampling in the same regions of a distribution and lack of information from no sampling in other regions of the distributions.

A stratified sampling approach ensures more even coverage of the range of input parameter or variable values with the same number of simulation runs. This can be accomplished by dividing the input parameter or variable space into sections and sampling from each section with the appropriate probability.

One such approach, Latin hypercube sampling (LHS), divides each input distribution into sections of equal probability for the specified the probability distribution, and draws one observation randomly from each range. Hence the ranges of input values within each section actually occur with equal frequency in the experiment. These values from each interval for each distribution are randomly assigned to those from other intervals to construct sets of input values for the simulation analysis. Figure 19 shows the steps in constructing a LHS for six simulations involving three inputs P_j (P_1, P_2, and P_3) and six intervals of their respective normal, uniform and triangular probability distributions.



I)

randomly assign each of 6 intervals k for each input j to 6 model-runs

model run	randomly selected interval k for input P_j		
	P_1	P_2	P_3
1	4	3	5
2	2	1	6
3	1	6	1
4	6	4	4
5	3	2	3
6	5	5	2

II)

randomly sample from each specified interval k to obtain values P_{jk} for use in each simulation run

model run k	randomly selected values P_{jk} for each simulation run			value of output variable Y_k
	P_{1k}	P_{2k}	P_{3k}	
1	P_{14}	P_{23}	P_{35}	Y_1
2	P_{12}	P_{21}	P_{36}	Y_2
3	P_{11}	P_{26}	P_{31}	Y_3
4	P_{16}	P_{24}	P_{34}	Y_4
5	P_{13}	P_{22}	P_{33}	Y_5
6	P_{15}	P_{25}	P_{32}	Y_6

III)

plot histogram of output resulting from model-runs

Figure 19. Schematic representation of a Latin hypercube sampling procedure for six simulation runs.

5. Performance indicator uncertainties

5.1 Performance measure target uncertainty

Another possible source of uncertainty is the selection of performance measure target values. For example, consider a target value for a pollutant concentration based on the effect of exceeding it in an ecosystem. Which target value is best or correct? When this is not clear, there are various ways of expressing the uncertainty associated with any target value. One such method is the use of fuzzy sets (Chapter VI). Use of ‘grey’ numbers or intervals instead of ‘white’ or fixed target values is another. When some uncertainty or disagreement exists over the selection of the best target value for a particular performance measure it seems to us the most direct and transparent way to do this is to subjectively assume a distribution over a range of possible target values. Then this subjective probability distribution can be factored into the tradeoff analysis, as outlined in Figure 20.

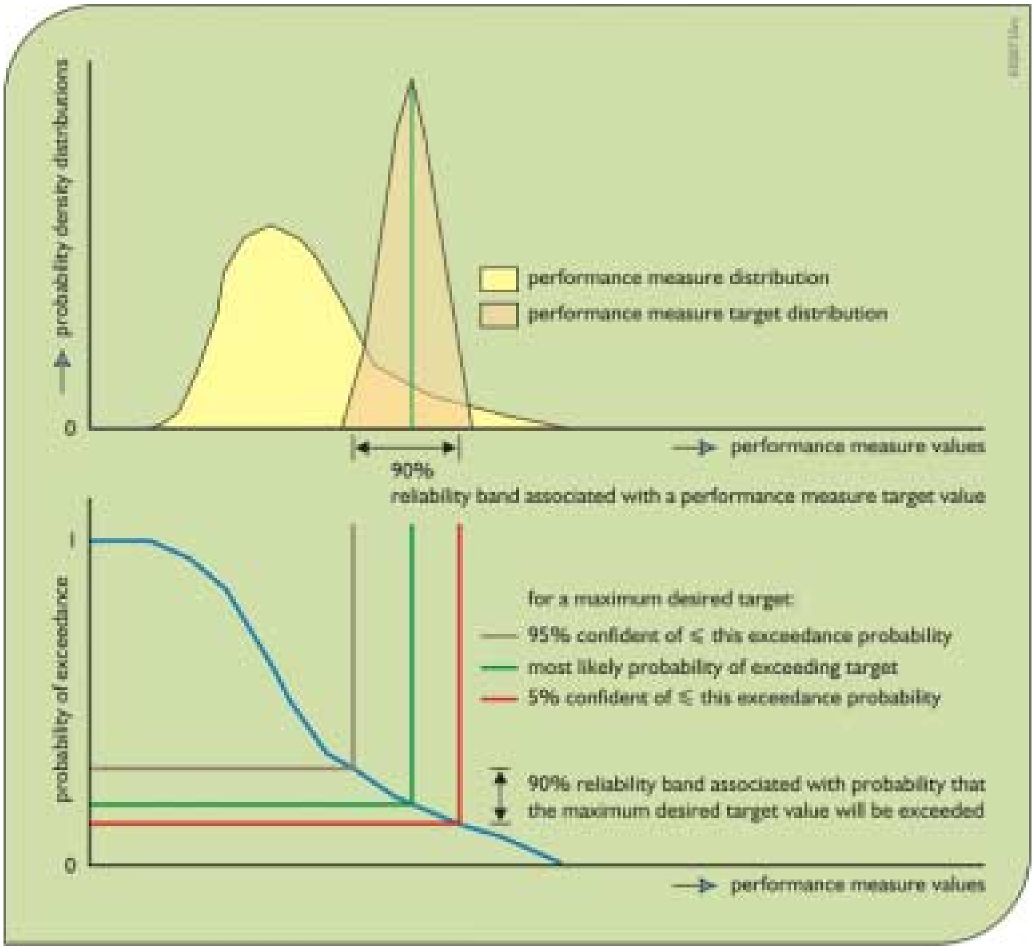


Figure 20. Combining the probability distribution of performance measure values with the probability distribution of performance measure target values to estimate the confidence one has in the probability of exceeding a maximum desired target value.

One of the challenges associated with defining and including in an analysis the uncertainty associated with a target or threshold value for a performance measure is that of communicating just what the result of such an analysis means. Referring to Figure 20, suppose the target value represents some maximum limit of a pollutant, say phosphorus, concentration in the flow during a given period of time at a given site or region, and it is not certain just what that maximum limit should be. Subjectively defining the distribution of that maximum limit, and considering that uncertainty along with the uncertainty (probability of exceedance function) of pollutant concentrations – the performance measure – one can attach a confidence to any probability of exceeding the maximum desired concentration value.

The 95% probability of exceedance shown on Figure 20, say $P_{0.95}$, should be interpreted as "we can be 95% confident that the probability of the maximum desired pollutant concentration being exceeded will be no greater than $P_{0.95}$." We can be only 5% confident that the probability of exceeding the desired maximum concentration will be no greater than the lower $P_{0.05}$ value. Depending on whether the middle line through the subjective distribution of target values in Figure 20 represents the most likely or median target value, the associated probability of exceedance is either the most likely, as indicated in Figure 20, or that for which we are only 50% confident.

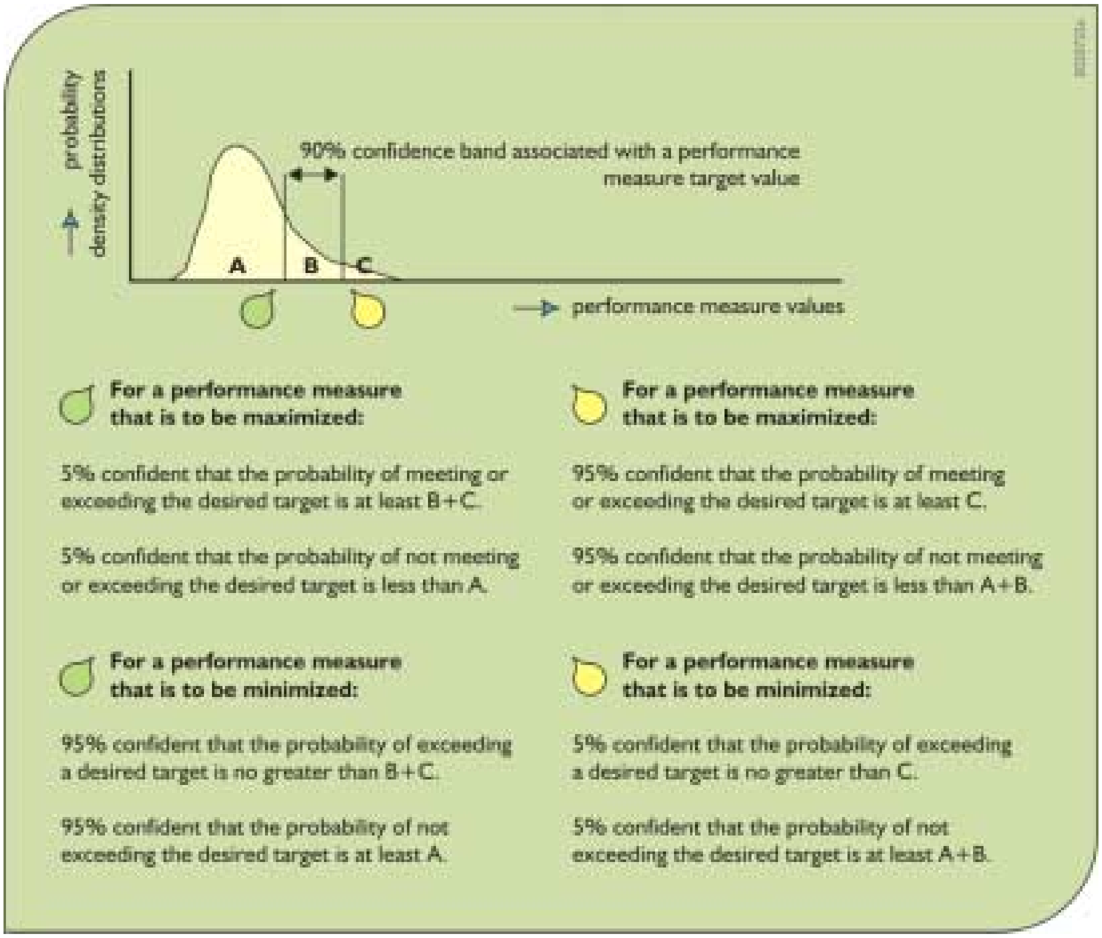
Figure 21 attempts to show how to interpret the reliabilities when the uncertain performance targets are

- minimum acceptable levels that are to be maximized,
- maximum acceptable levels that are to be minimized or
- optimum levels.

An example of a minimum acceptable target level might be the population of wading birds in an area. An example of a maximum acceptable target level might be, again, the phosphorus concentration of the flow in a specific wetland or lake. An example of an optimum target level might be the depth of water most suitable for selected species of aquatic vegetation during a particular period of the year.

For performance measure targets that are not expressed as minimum or maximum limits but that are the 'best' values, referring to Figure 21, one can state that one is 90% confident that the probability of achieving the desired target is no more than B. The 90% confidence level probability of not achieving the desired target is at least A+C. The probability of the performance measure being too low is at least A and the probability of the performance measure being too high is at least C, again at the 90% confidence levels. As the confidence level decreases the bandwidth decreases, and the probability of not meeting the target increases.

Now, clearly there is uncertainty associated with each of these uncertainty estimations, and this raises the question of how valuable is the quantification of the uncertainty of each additional component of the plan in an evaluation process. Will plan evaluators and decision makers benefit from this additional information, and just how much additional uncertainty information is useful?



For a performance measure that is to be maximized:

5% confident that the probability of meeting or exceeding the desired target is at least B+C.

5% confident that the probability of not meeting or exceeding the desired target is less than A.

For a performance measure that is to be maximized:

95% confident that the probability of meeting or exceeding the desired target is at least C.

95% confident that the probability of not meeting or exceeding the desired target is less than A+B.

For a performance measure that is to be minimized:

95% confident that the probability of exceeding a desired target is no greater than B+C.

95% confident that the probability of not exceeding the desired target is at least A.

For a performance measure that is to be minimized:

5% confident that the probability of exceeding a desired target is no greater than C.

5% confident that the probability of not exceeding the desired target is at least A+B.

Figure 21. Interpreting the results of combining performance measure probabilities with performance measure target probabilities depends on the type of performance measure. The letters A, B, and C represent proportions of the probability density function of performance measure values. (Hence probabilities A + B + C = 1.)

Now consider again the tradeoffs that need to be made as illustrated in Figure 7. Instead of considering a single target value as shown on Figure 7, assume there is a 90% confidence range associated with that single performance measure target value. Also assume that the target is a maximum desired upper limit (e.g., of some pollutant concentration).

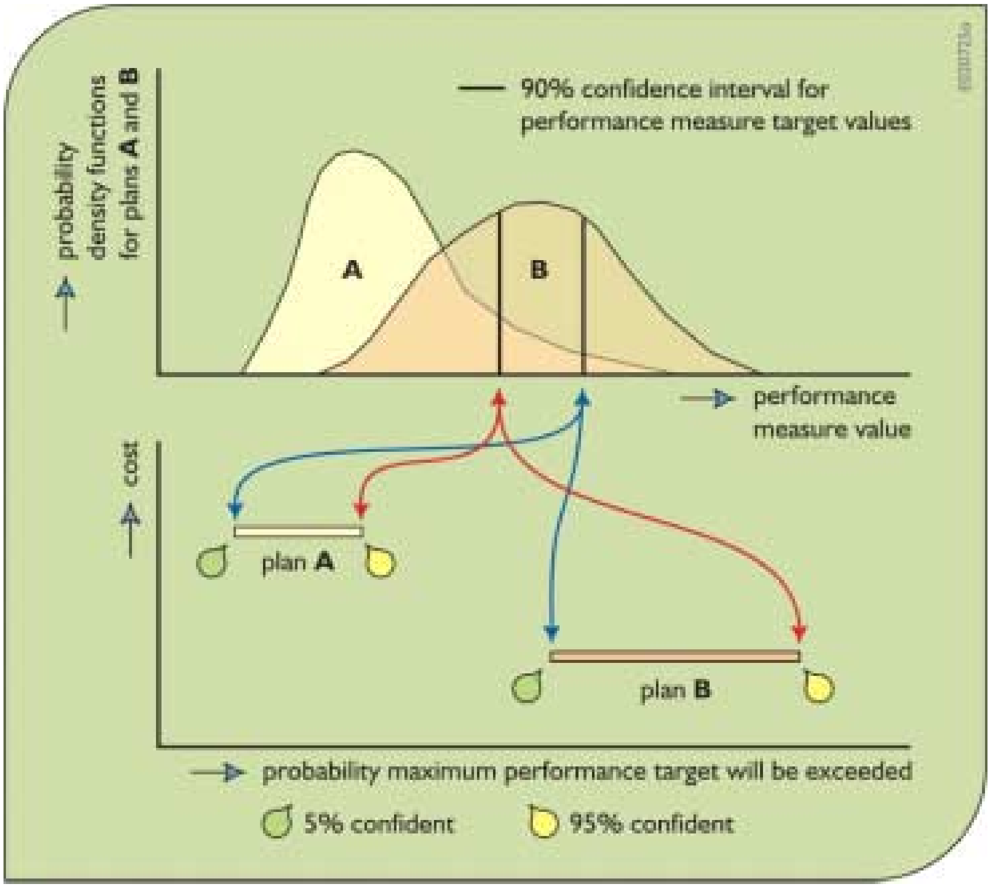


Figure 22. Two plans showing ranges of probabilities, depending on one's confidence, that an uncertain desired maximum (upper limit) performance target value will be exceeded. The 95% confidence levels are associated with the higher probabilities of exceeding the desired maximum target. The 5% confident levels are associated with the more desirable lower probabilities of exceeding the desired maximum target. Plan A with reduced probabilities of exceeding the upper limit costs more than Plan B.

In the case shown in Figure 22, the tradeoff is clearly between cost and reliability. In this example, no matter what confidence one chooses, Plan A is preferred to Plan B with respect to reliability, but Plan B is preferred to Plan A with respect to cost. The tradeoff is only between these two performance indicators or measures.

Consider however a third plan, as shown in Figure 23. This situation adds to the complexity of making appropriate tradeoffs. Now there are three criteria: cost, probability of exceedance (reliability) and the confidence in those reliabilities or probabilities. Add to this the fact that there will be multiple performance measure targets, each expressed in terms of their maximum probabilities of exceedance and the confidence in those probabilities.

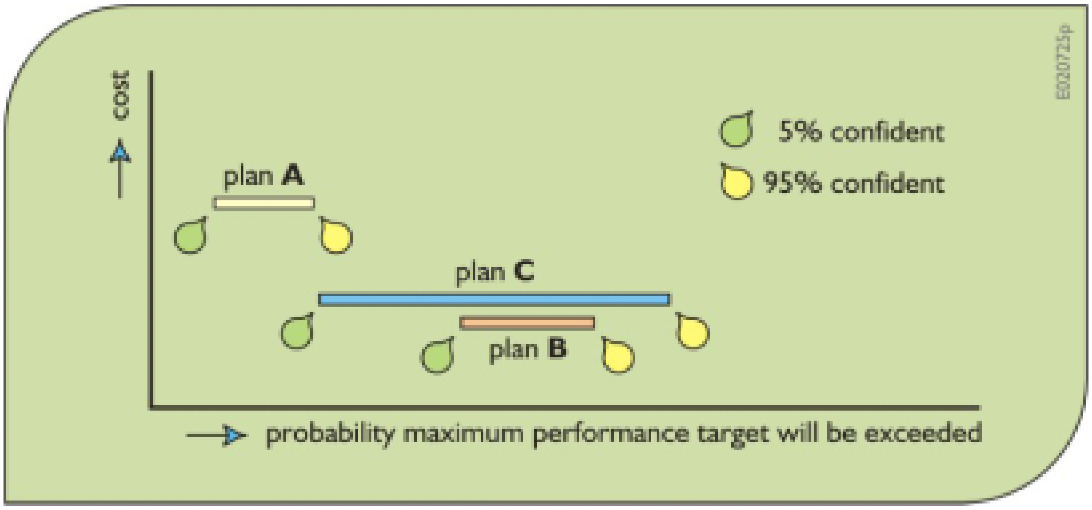


Figure 23. Tradeoffs among cost, reliabilities, and the confidence level of those reliabilities. The relative ranking of plans with respect to the probability of exceeding the desired (maximum limit) target may depend on the confidence given to that probability.

In Figure 23, in terms of cost the plans are ranked, from best to worst, B, C, and A. In terms of reliability at the 90 percent confidence level, they are ranked A, B, and C but at the 50 percent confidence level the ranking is A, C and B.

If the plan evaluation process has difficulty handling all this it may indicate the need to focus the uncertainty analysis effort on just what is deemed important, achievable, and beneficial. Then when the number of alternatives has been narrowed down to only a few that appear to be the better ones, a more complete uncertainty analysis can be performed. There is no need nor benefit in performing sensitivity and uncertainty analyses on all possible management alternatives. Rather one can focus on those alternatives that look the most promising, and then carry out additional uncertainty and sensitivity analyses only when important uncertain performance indicator values demands more scrutiny. Otherwise the work is not likely to affect the decision anyway.

5.2 Distinguishing differences between performance indicator distributions

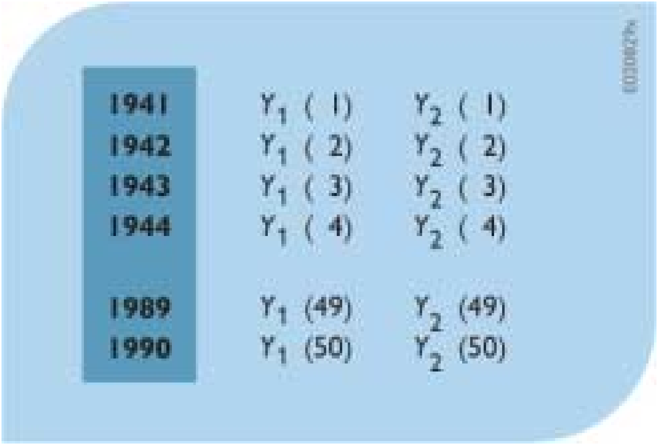
Simulations of alternative water management infrastructure designs and operating policies require a comparison of the simulation outputs – the performance measures or indicators – associated with each alternative. A reasonable question to ask is are the observed differences statistically significant. Can one really tell if one alternative is better than another or are the observed differences explainable by random variations attributable to variations in the inputs and how the system responds?

This is a common statistical issue that is addressed by standard hypothesis tests (Devore, 1991; Benjamin and Cornell, 1970). Selection of an appropriate test requires that one first resolve what type of change one expects in the variables. To illustrate, consider the comparison of two different operating policies. Let Y_1 denote the set of output performance variable values with the first policy, and Y_2 the set of output performance variable values of the second policy. In many cases, one would expect one policy to be better than the other. One measure might be the difference in the mean of the variables; for example is $E[Y_1] < E[Y_2]$?. Alternatively one could check the difference in the median (50 percentile) of the two distributions.

In addition, one could look for a change in the variability or variance, or a shift in both the mean and the variance. Changes described by a difference in the mean or median often make the most sense and many statistical tests are available that are sensitive to such changes. For such investigations parametric and non-parametric tests for paired and unpaired data can be employed.

Consider the differences between "paired" and "unpaired" data. Suppose that the meteorological data for 1941-1990 is used to drive a simulation model generating data as described in Table 11:

Table 11. Possible flow data from a 50-year simulation



1941	$Y_1(1)$	$Y_2(1)$
1942	$Y_1(2)$	$Y_2(2)$
1943	$Y_1(3)$	$Y_2(3)$
1944	$Y_1(4)$	$Y_2(4)$
1989	$Y_1(49)$	$Y_2(49)$
1990	$Y_1(50)$	$Y_2(50)$

Here there is one sample, $Y_1(1)$ through $Y_1(50)$, for policy 1, and another sample, $Y_2(1)$ through $Y_2(50)$, for policy 2. However, the two sets of observations are not independent. For example, if 1943 was a very dry year, then we would expect both $Y_1(3)$ for policy 1 in that year and $Y_2(3)$ for policy 2 to be unusually small. With such paired data, one can use a paired hypothesis test to check for differences. Paired tests are usually easier than the corresponding unpaired tests that are appropriate in other cases. (For example, if one were checking for a difference in average rainfall depth between 1941-1960, and 1961-1990, they would have two sets of independent measurements for the two periods. With such data, one should use a two-sample unpaired test.)

Paired tests are generally based on the differences between the two sets of output, $Y_1(i) - Y_2(i)$. These are viewed as a single independent sample. The question is then are the differences positive (say Y_1 tends to be larger then Y_2), or negative (Y_1 tends to be smaller), or are positive and negative differences are equally likely (there is no difference between Y_1 and Y_2).

Both parametric and non-parametric families of statistical tests are available for paired data. The common parametric test for paired data (a one-sample t test) assumes that the mean of the differences

$$X(i) = Y_1(i) - Y_2(i) \tag{21}$$

are normally distributed. Then the hypothesis of no difference is rejected if the t statistic is sufficiently large, given the sample size n.

Alternatively, one can employ a nonparametric test and avoid the assumption that the differences $X(\mathrm{i})$ are normally distributed. In such a case, one can use the Wilcoxon Signed Rank test. This nonparametric test ranks the absolute values $|X(i)|$ of the differences. If the sum S of the ranks of the positive differences deviates sufficiently from its expected value, $n(n+1)/4$ (were there no difference between the two distributions), one can conclude that there is a statistically significant difference between the $Y_1(i)$ and $Y_2(i)$ series. Standard statistical texts have tables of the distribution of the sum S as a function of the sample size n, and provide a good analytical approximation for $n > 20$ (for example, Devore, 1991). Both the parametric t test and the nonparametric Wilcoxon Signed Rank test require that the differences between the simulated values for each year be computed.

6. Communicating model output uncertainty

Spending money on reducing uncertainty would seem preferable to spending it on ways of calculating and describing it better. Yet attention to uncertainty communication is critically important if uncertainty analyses and characterizations are to be of value in a decision making process. In spite considerable efforts by those involved in risk assessment and management, we know very little about how to ensure effective risk communication to gain the confidence of stakeholders, incorporate their views and knowledge, and influence favorably the acceptability of risk assessments and risk-management decisions.

The best way to communicate concepts of uncertainty may well depend on what the audiences already know about risk and the various types of probability distributions (e.g., density, cumulative, exceedance) based on objective and subjective data, and the distinction between mean or average values and the most likely values. Undoubtedly graphical representations of these ways of describing uncertainty considerably facilitate communication.

The National Research Council (NRC 1994) addressed the extensive uncertainty and variability associated with estimating risk and concluded that risk characterizations should not be reduced to a single number or even to a range of numbers intended to portray uncertainty. Instead, the report recommended managers and the interested public should be given risk characterizations that are both qualitative and quantitative and both verbal and mathematical.

In some cases communicating qualitative information about uncertainty to stakeholders and the public in general may be more effective than quantitative information. There are, of course, situations in which quantitative uncertainty analyses are likely to provide information that is useful in a decision-making process. How else can tradeoffs such as illustrated in Figures 10 and 27 be identified? Quantitative uncertainty analysis often can be used as the basis of qualitative information about uncertainty, even if the quantitative information is not what is communicated to the public.

One should acknowledge to the public the widespread confusion regarding the differences between variability and uncertainty. Variability does not change through further measurement or study, although better sampling can improve our knowledge about variability. Uncertainty reflects gaps in information about scientifically observable phenomena.

While it is important to communicate uncertainties and confidence in predictions, it is equally important to clarify who or what is at risk, possible consequences, and the severity and irreversibility of an adverse effect should a target value, for example, not be met. This qualitative information is often critical to informed decision-making. Risk and uncertainty communication is always complicated by the reliability and amounts of available relevant information as well as how that information is presented. Effective communication between people receiving information about who or what is at risk, or what might happen and just how severe and irreversible an adverse effect might be should a target value not be met, is just as important as the level of uncertainty and the confidence associated with such predictions. A two-way dialog between those receiving such information and those giving it can help identify just what seems best for a particular audience.

Risk and uncertainty communication is a two-way street, It involves learning and teaching. Communicators dealing with uncertainty should learn about the concerns and values of their audience, their relevant knowledge, and their experience with uncertainty issues. Stakeholders' knowledge of the sources and reasons for uncertainty needs to be incorporated into assessment and management and communication decisions. By listening, communicators can craft risk messages that better reflect the perspectives, technical knowledge, and concerns of the audience.

Effective communication should begin before important decisions have been made. It can be facilitated in communities by citizen advisory panels. Citizen advisory panels can give planners and decision makers a better understanding of the questions and concerns of the community and an opportunity to test its effectiveness in communicating concepts and specific issues regarding uncertainty.

One approach to make uncertainty more meaningful is to make risk comparisons. For example, a ten parts per billion target for a particular pollutant concentration is equivalent to 10 seconds in over 31 years. If this is an average daily concentration target that is to be satisfied "99 percent," of the time, this is equivalent to an expected violation of less than one day every three months.

Many perceive the reduction of risk by an order of magnitude as though it were a linear reduction. A better way to illustrate orders of magnitude of risk reduction is shown in Figure 24, in which a bar graph depicts better than words that a reduction in risk from one in a 1,000 (10^{-3}) to one in 10,000 (10^{-4}) is a reduction of 90% and that a further reduction to one in 100,000 (10^{-5}) is a reduction 10-fold less than the first reduction of 90%. The percent of the risk that is reduced by whatever measures is a much easier concept to communicate than reductions expressed in terms of estimated absolute risk levels, such as 10^{-5}.

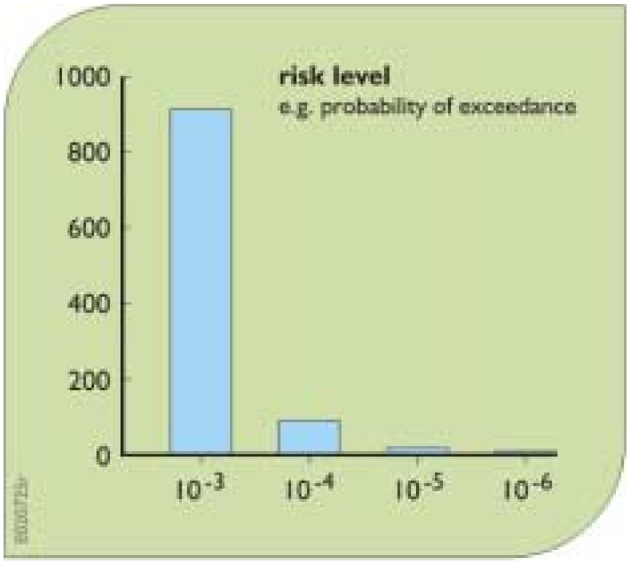


Figure 24. Reducing risk by orders of magnitude is not equivalent to linear reductions.

Risk comparisons can be helpful, but they should be used cautiously and tested if possible. There are dangers in comparing risks of diverse character, especially when the intent of the comparison is seen as minimizing a risk (NRC 1989). One difficulty in using risk comparisons is that it is not always easy to find risks that are sufficiently similar to make a comparison meaningful. How is someone able to compare two alternatives having two different costs and two different risk levels, for example, as is shown in Figure 7? One way is to perform an indifference analysis (Chapter X), but that can lead to different results depending who performs it. Another way is to develop utility functions using weights, where, for example reduced phosphorus load by half is equivalent to a 25 percent shorter hydroperiod in that area, but again each person's utility or tradeoff may differ.

At a minimum, graphical displays of uncertainty can be helpful. Consider the common system performance indicators that include:

- Time-series plots for continuous time-dependent indicators (Figure 25 upper left)
- Probability exceedance distributions for continuous indicators (Figure 25 upper right),
- Histograms for discrete event indicators (Figure 25 lower left), and
- Overlays on maps for space-dependent discrete events (Figure 25 lower right).

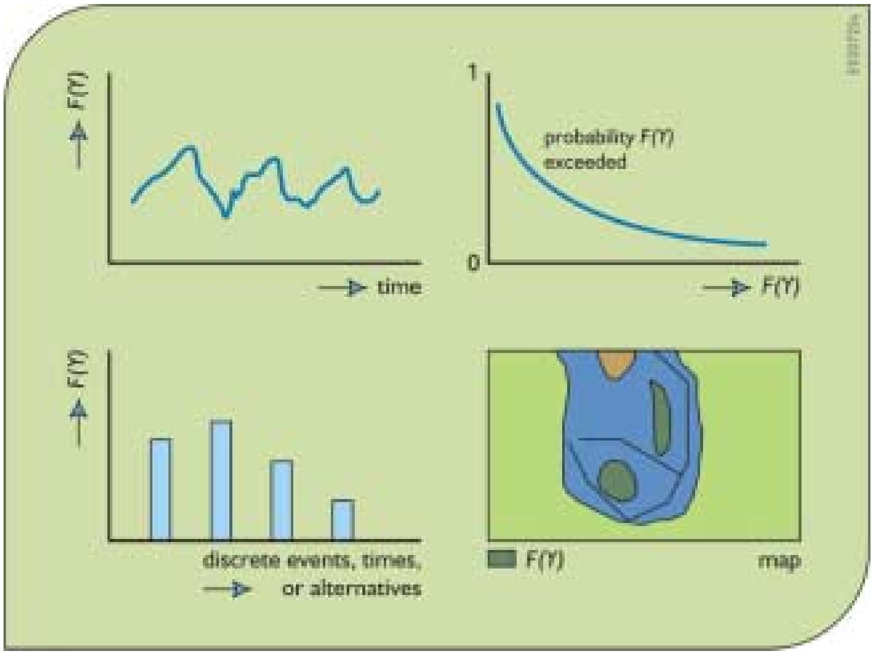


Figure 25. Different types of displays used to show model output *Y* or system performance indicator values *F*(*Y*).

The first three graphs in Figure 25 could show, in addition to the single curve or bar that represents the most likely output, a range of outcomes associated with a given confidence interval. For overlays of information on maps, different colors could represent the spatial extents of events associated with different ranges of risk or uncertainty. Figure 26, corresponding to Figure 25, illustrates these approaches for displaying these ranges.

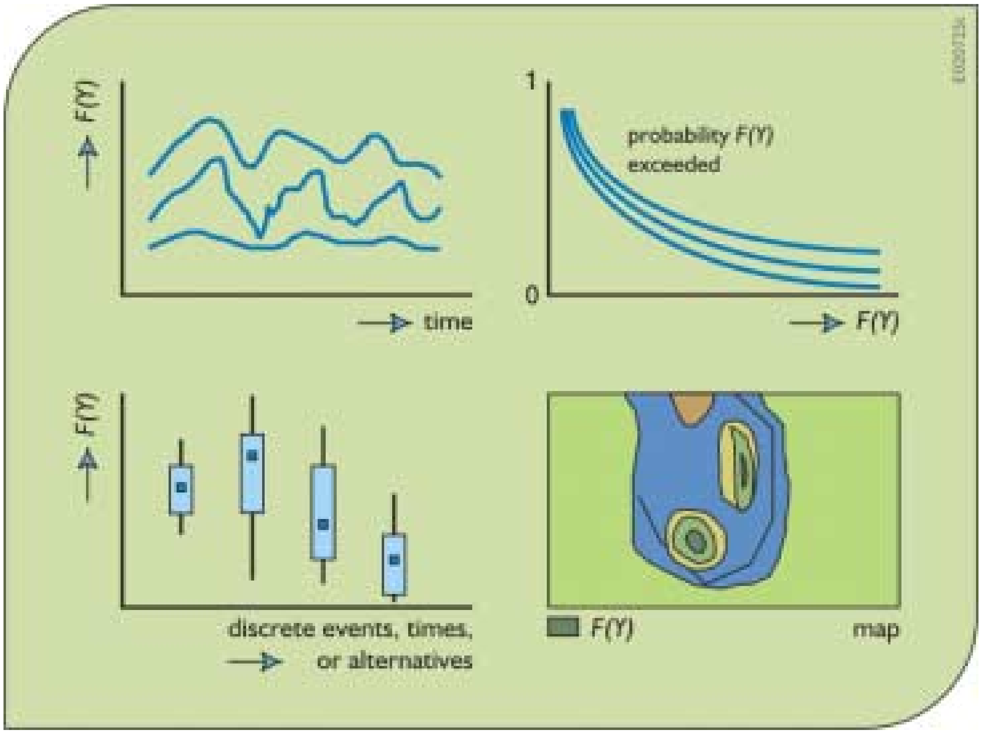


Figure 26. Plots of ranges of possible model output Y or system indicator values $F(Y)$ for different types of displays.

7. Conclusions

This chapter provides an overview of uncertainty and sensitivity analyses in the context of hydrologic or water resources systems simulation modeling. A broad range of tools are available to explore, display, and quantify the sensitivity and uncertainty in predictions of key output variables and system performance indices with respect to imprecise and random model inputs and to assumptions concerning model structure. They range from relatively simple deterministic sensitivity analysis methods to more involved first-order analyses and Monte Carlo sampling methods.

Because of the complexity of many watersheds or river basins, Monte Carlo methods for uncertainty analyses may be a very major and unattractive undertaking. Therefore it is often prudent begin with the relatively simple deterministic procedures. This coupled with a probabilistically based first-order uncertainty analysis method can help quantify the uncertainty in key output variables and system performance indices, and the relative contributions of uncertainty in different input variables to the uncertainty in different output variables and system performance indices. These relative contributions may differ depending upon which output variables and indices are of interest.

A sensitivity analysis can provide a systematic assessment of the impact of parameter value imprecision on output variable values and performance indices, and of the relative contribution of errors in different parameter values to that output uncertainty. Once the key variables are identified, it should be possible to determine the extent to which parameter value uncertainty can be reduced through field investigations, development of better models, and other efforts.

Model calibration procedures can be applied to individual catchments and subsystems, as well as to composite systems. Automated calibration procedures have several advantages including the explicit use of an appropriate statistical objective function, identification of those parameters that best reproduce the calibration data set with the given objective function, and the estimations of the statistical precision of the estimated parameters.

All of these tasks together can represent a formidable effort. However, knowledge of the uncertainty associated with model predictions can be as important to management decision and policy formulation as are the predictions themselves.

No matter how much attention is given to quantifying and reducing uncertainties in model outputs, uncertainties will remain. Professionals who analyze risk, managers and decision makers who must manage risk, and the public who must live with risk and uncertainty, have different information needs and attitudes regarding risk and uncertainty. It is clear that information needs differ among those who model or use models, those who make substantial investment or social decisions, and those who are likely to be impacted by those decisions. Meeting those needs should result in more informed decision making. But it comes at a cost that should be considered along with the benefits of having this sensitivity and uncertainty information.

9. References

Anderson, J.L., 1998, Embracing uncertainty: The interface of Bayesian statistics and cognitive psychology. Conservation Ecology (online) 2(1): 2. Available from the Internet. URL: http://www.consecol.org/vol2/iss1/art2

Beck, M.B., 1987, Water Quality Modeling: A review of the Analysis of Uncertainty, Water Resour. Res., 23(8), 1393-1442

Benaman, J., 2002, A Systematic Approach to Uncertainty Analysis for a Distributed Watershed Model, Ph.D. Thesis. School of Civil and Environmental Engineering. Cornell University, Ithaca, NY

Benjamin J.R., and Cornell, C.A., 1970, Probability, Statistics and Decision for Civil Engineers, McGraw-Hill Book Co., New York

Berry, D.A. 1996, Statistics: A Bayesian Perspective, Belmont, CA: Duxbury.

Beven, K., 1989, Changing Ideas in Hydrology – The Case of Physically-based Models, J. of Hydrol., 105, 157-172

Beven, K., 1993, Prophecy, reality and uncertainty in distributed hydrologic modelling, Adv. in Water Res., 16, 41-51

Binley, A.M., and Beven, K.J., 1991, Physically-based modelling of catchment hydrology: a likelihood approach to reducing predictive uncertainty, in Computer Modelling in the Environmental Sciences, D.G. Farmer and M.J. Rycroft (ed.), Clarendon Press, Oxford

Chapra, S.C. and Reckhow, K.H., 1979, Expressing the phosphorus loading concept in probabilistic terms. J. Fish. Res. Board Can., 36.

Colwell, R.K., 1974, Predictability, consistency, and contingency of periodic phenomena, Ecology Vol. 55, pp 1148-53.

Covello, V.T. and Merkhofer, M.W., 1993, Risk Assessment Methods. Plenum Press. London.

Covello, V. T., 1987, Decision Analysis and Risk Management Decision-Making: Issues and Methods, Risk Analysis, 7, , pp. 131-139

DeGarmo, E.P., Sullivan, W.G. and Bontadelli, J.A., 1993, Engineering Economy, MacMillian, New York

Deutsch C.V. and Journel, A.G., 1992, GS-LIB: Geostatistical Software Library and User's Guide. Oxford University Press

Devore, J., 1991, Probability and Statistics for Engineering and the Sciences, 3rd Edition, Brooks/Cole Publishing Co., London

Devore, J.L. and Peck, R., 1996, Statistics: The exploration and analysis of data, 3rd ed. Brooks/Cole Publishing Co., London

Dilks, D.W., Canale, R.P., and Meier, P.G., 1992, Development of Bayesian Monte Carlo techniques for water quality model uncertainty, Ecological Modelling, Vol. 62, pp 149-162.

Dilks, D. W. and James, R. T., 2002, Application of Bayesian Monte Carlo analysis to determine the uncertainty in the Lake Okeechobee water quality model, Proceedings, Watershed 2002, Water Environment Federation

DOE, 1998, Screening Assessment and Requirements for a Comprehensive Assessment, DOE/RL-96-16, Rev. 1, U.S. Department of Energy, Richland, Washington

Duan, Q., Soroosian, S. and Ibbitt. R.P., 1988, A maximum likelihood criterion for use with data collected at unequal time intervals, Water Resources Research, 24(7), 1163-1173

Eschenback, T.G., 1992, Spider plots versus Tornado Diagrams for Sensitivity Analysis, Interfaces, 22, 40-46

Fedra, K., 1983, A Monte Carlo approach to estimation and prediction, In M. B. Beck and G. Van Straten (Eds.), Uncertainty and Forecasting of Water Quality, Springer-Verlag, Berlin, pp 259-291.

Fitz, H.C., Voinov, A. and Costanza, R., 1995, The Everglades Landscape Model: multiscale sensitivity analysis. South Florida Water Management District, Everglades Systems Research Division, West Palm Beach, FL, 88 pp.

Fontaine, T.A. and Jacomino, V.M.F., 1997, Sensitivity analysis of simulated contaminated sediment transport. Journal of the American Water Resources Association **3**3(2): 313-326.

Frey, H.C. and Patil, S.R., 2002, Identification and review of sensitivity analysis methods, Risk Analysis, Vol. 22, No. 3 pp 553-578

Gardner, R.H., Rojder, B., and Bergstrom, U., 1983, PRISM: A systematic method for determining the effect of parameter uncertainties on model predictions. Technical Report, Studsvik Energiteknik AB report NW-83-/555, Nykoping, Sweden.

Gaven, D.C. and Burges, S.J., 1981, Approximate error bounds for simulated Hydrograph, J. Hydraulic Engineering, 107(11), 1519-1534

Gelman, A., Carlin, J., Stern, H., and Rubin, D.B., 1995, Bayesian Data Analysis. Chapman and Hall, London:

Gleick, J., 1987, Chaos: making a new science, Penguin, NY

Gupta, V.K. and Sorooshian, S., 1985a, The automatic calibration of conceptual watershed models using derivative-based optimization algorithms, Water Resources Research, 21(4), 473-485

Gupta, V.K. and Sorooshian, S., 1985b, The relationship between data and the precision of parameter estimates of hydrologic models, Journal of Hydrology 81, 57-77

Haimes, Y.Y., 1998, Risk Modelling, Assessment and Management, John Wiley & Sons, Inc. New York

Harlin, J., and Kung, C-S., 1992, Parameter Uncertainty in Simulation of Design of Floods in Sweden, J. of Hydrol., 137, 2009-230

Hendrickson, J.D., Sorooshian, S. and L.E. Brazil, 1988, Comparison of Newton-type and direct-search algorithms for calibration of conceptual rainfall-runoff models, Water Research Research, 24(5), 691-700

Holling, C.S., 1978, Adaptive Environmental Assessment and Management, John Wiley & Sons, Chichester, UK

Ibrekk, H. and Morgan, M.G., 1987, Graphical Communication of Uncertain Quantities to nontechnical people, Risk Analysis, Vol. 7, No. 4, pp 519-529.

Isaaks, E.H. and Srivastava, R.M., 1989, An introduction to applied geostatistics. Oxford University Press, New York.

Jaffe, P.R., Paniconi, C., and Wood, E.F., 1988, Model calibration based on random environmental fluctuations, J. Environ. Eng., 114(5), 1136-1145

Jain S, and U Lall, 2001, Floods in a changing climate: Does the past represent the future?, Water Resources Research, 37 (12): 3193-3205 DEC

Jensen, F.V., 2001, Bayesian Networks and Decision Graphs, New York: Springer.

Kann, A. and Weyant, J.P., 1999, A comparison of approaches for performing uncertainty analysis in integrated assessment models, Journal of Environmental Management and Assessment, Vol. 5, No.1, pp 29-46.

Kelly, E. and Campbell, K., 2000, Separating variability and uncertainty-making choices, Human and Ecological Risk Assessment, An International Journal, In Press, February.

Kelly, E. and Roy-Harrison, W., 1998, A mathematical construct for ecological risk: A useful framework for assessments, Human and Ecological Risk Assessment, An International Journal, Vol. 4, No. 2, pp 229-241.

Kelly, E.J., Campbell, K., and Henrion, M., 1997, To Separate or Not to Separate - That is the Question - A Discourse on Separating Variability and Uncertainty in Environmental Risk Assessments, Learned Discourses, Society of Environmental Toxicology and Chemistry (SETAC) News, November.

Kuczera, G., 1988, On the Validity of First-order Prediction Limits for Conceptual Hydrologic Models, J. of Hydrol., 103, 229-247

Lal, W., 1995, Sensitivity and Uncertainty Analysis of a Regional Model for the Natural System of South Florida. South Florida Water Management District, West Palm Beach, FL, Draft report, November

Lal, W., Obeysekera, J. and van Zee, R., 1997, Sensitivity and Uncertainty Analysis of a Regional Simulation Model for the Natural System in South Florida, in Managing Water: Coping with Scarcity and Abundance, Proceedings, 27th Congress of International Association for Hydraulic Research, San Francisco, CA. August.

Lal, A.M.W, 2000, Numerical Errors in Groundwater and Overland Flow Models, Water Resources Research, Vol. 36, No. 5, May, pp 1237-1248

Lemons, J., ed. 1996, Scientific uncertainty and environmental problem solving. Blackwell, Cambridge, MA

Lopez, A. and Loucks, D.P., 1999, Uncertainty representations in water quality prediction. Manuscript, Civil and Environmental Engineering, Cornell University, Ithaca, NY

Loucks, D.P., Stedinger, J.R. and Haith, D.A., 1981, Water Resource Systems Planning and Analysis, Prentice-Hall, Englewood Cliffs, NJ

Ludwig, D., Hilborn, R., and Walters, C., 1993, Uncertainty, resource exploitation, and conservation: lessons from history, Science vol. 260, pp 17-36

Majoni, H. and Quade, E. S., 1980,, Pitfalls of analysis. John Wiley and Sons, New York

McCarthy, James J., Canziani, Osvaldo F., Leary, Neil A., Dokken, David J., White, Kasey S., (eds), 2001, Climate Change 2001: Impacts, Adaptation, and Vulnerability, Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change, Published for the Intergovernmental Panel on Climate Change. http://www.ipcc.ch/pub/tar/wg2/index.htm

McCuen, R., 1973, The role of sensitivity analysis in hydrologic modeling, Journal of Hydrology 18(1): 37-53.

McCuen, R. H. and Snyder, W. M., 1983, Hydrologic Modeling: Statistical Methods and Applications. Englewood Cliffs, NJ, Prentice-Hall.

Meixner, T., Gupta, H. V., Bastidas, L. A., and Bales, R. C., 1999, Sensitivity analysis using mass flux and concentration, Hydrological Processes 13: 2233-2244.

Miser, H J., 1980, “Operations research and systems analysis.” Science., 209, 174 -182.

Morgan, M.G., and Henrion, M., 1990, Uncertainty, A guide to dealing with uncertainty in quantitative risk and policy analysis, Cambridge University Press. Cambridge, UK.

Nearing, M. A., Deer-Ascough, L. and Laflen, J. F., 1990, Sensitivity analysis of the WEPP hillslope profile erosion model, Transactions of the American Society of Agricultural Engineers 33(3): 839-849.

National Research Council (NRC), 1989, Improving Risk Communication, National Academy Press, Washington, DC.

National Research Council (NRC), 1994, Science and Judgment in Risk Assessment, National Academy Press, Washington, DC

National Research Council (NRC), 1996, Committee on Risk Characterization, Understanding Risk: Informing Decision in a Democratic Society," P.S. Stern and H.V. Fineberg, (eds.) National Academy Press, Washington, D.C.

Phillips D.L. and Marks D.G., 1996, Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models. Ecological Modelling 91: 213-229.

Press, S.J., and Tanur, J.M., 2001, The Subjectivity of Scientists and the Bayesian Approach. New York: Wiley.

Reckhow, K.H., 1994, Water quality simulation modeling and uncertainty analysis for risk assessment and decision making. Ecological Modelling. 72:1-20.

Reckhow, K.H., 1999, Water quality prediction and probability network models. Canadian Journal of Fisheries and Aquatic Sciences. 56:1150-1158.

Reckhow K., 2002, Applications of Water Models: Prediction Uncertainty and Decision Making. Presentation at Model Uncertainty Workshop, South Florida Water Management District, West Palm Beach, FL, January

Saltelli, A., Chan, K., and Scott, E.M., (ed.), 2000, Sensitivity Analysis, John Wiley & Sons, Chichester, UK

Schweppe, F.C., 1973, Uncertain Dynamic Systems, Prentice-Hall, Englewood Cliffs, NJ

Shapiro, H.T., 1990, The willingness to risk failure, Science, 250(4981), 609.

.

Simon, I-I. A., 1988, Prediction and prescription in system modeling, 15th Anniversary of IIASA, Int, Inst. for Applied Systems Analysis, Laxenburg, Austria.

Sklar, F. H. and Hunsaker, C.T., 2001, The Use and Uncertainties of Spatial Data for Landscape Models: An Overview with Examples from the Florida Everglades, Chapter 2 in Spatial Uncertainty in Ecology, Implications for Remote Sensing and GIS Applications, Hunsaker et al. eds, Springer

Sorooshian, S., Gupta, V.K., and Fulton, J.L., 1983, Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models--influence of calibration data variability and length on model credibility, Water Resources Research, 19(1), 251-259

Sorooshian, S., Duan, Q. and Gupta, V.K., 1993, Calibration of Rainfall-Runoff Models: Application of global optimization to the Sacramento soil moisture accounting model, Water Resources Research, 29(4), 1185-94

Soutter, M. and Musy, A., 1999, Global sensitivity analyses of three pesticide leaching models using a Monte-Carlo approach, Journal of Environmental Quality 28: 1290-1297.

Spear, R.C. and Hornberger, G.M, 1980, Eutrophication in Peel Inlet – II. Identification of critical uncertainties via generalized sensitivity analysis. Water Research 14: 43-49.

Stedinger, J.R. and Taylor, M.R., 1982, Synthetic Streamflow Generation, Part II. Effect of Parameter Uncertainty, Water Resources Research, 18(4), 919-924

Stokey, E. and Zeckhauser, R., 1977, A primer for policy analysis,. W. W. Norton and Co., Inc., New York, N. Y.

Suter, G.W., II. 1993, Ecological Risk Assessment. Lewis Publishers. Boca Raton, FL

Tattari, S. and Barlund, I., 2001, The concept of sensitivity in sediment yield modelling, Physics and Chemistry of the Earth 26(1): 27-31.

van Griensven, A., Francos, A. and Bauwens, W., 2001. "Sensitivity Analysis and Auto-calibration of an Integral Dynamic Model for River Water Quality." In Progress.

van Harn Adams, B.A., 1998, “Parameter Distributions for Uncertainty Propagation in Water Quality Modeling.” Doctoral Dissertation, Duke University.

van Straten, G., 1983, Maximum likelihood estimation of parameters and uncertainty in phytoplankton models, in Uncertainty and Forecasting of Water Quality, M.B. Beck and G. van Straten (ed.), Springer Verlag, New York

Vollenweider, R.A., 1976, Advances in Defining Critical Loading Levels for Phosphorus in Lake Eutrophication, Memorie dell’Istituto Italiano di Idrobiologia, 33, 53-83

von Winterfeldt, D., and Edwards, W., 1986, Decision Analysis and Behavioral Research, Cambridge University Press, Cambridge, UK.

Warwick, J.J. and Cale, W.G., 1987, Determining likelihood of obtaining a reliable model, J. Env. Eng., 113(5), 1102-1119

Appendix I: Model Calibration Examples

- *Calibration of models in the Murray-Darling Basin*

In the Murray-Darling Basin, in order to preserve water quality, water reliability and the environment, a decision was made in 1995 to restrict water use to the 1993/94 level of development. Computer models of the major tributary streams are now used at the end of each year to determine the annual use target for the previous season based on that level of development. Rules are in place to ensure that long term usage is maintained at the agreed level. Because the models now define the overall water rights of each valley, there are legal requirements to calibrate models and each model is independently audited and certified as being unbiased before being approved as fit for purpose. The key model output of interest is water use but emphasis is also placed on the modeling of downstream flow which impacts the rights of downstream regions. Each model must be calibrated over at least ten years and this often means that changes in infrastructure, operating rules and growth in demand have to be incorporated into the calibration run. Calibration reports contain plots of modeled and observed water use, storage behavior and flow and statistics such as mean error, correlation coefficients and standard errors. The aim of calibration is to ensure that the model is unbiased and to give confidence to stakeholders.

An issue that is sometimes raised with model development is the role of calibration, where the model is fine-tuned to match the observed data, and validation where the model is tested against data that was not used in the calibration process to get an independent assessment of the model's accuracy. For the Murray River, because of the variability of our climate, we like to calibrate our model against a long period of data including the most recent years when the current operating rules were being used and the historical data is generally the most reliable. Validation is considered to be less important and is typically carried out using the two or three years of data available following the completion of model calibration.

- *Use of models for Allocating Water in Texas*

Recent legislation in Texas revised the State Water Planning process and mandated the development of water allocation models for every river basin in the state (http://www.tnrcc.state.tx.us/permitting/waterperm/wrpa/permits.html). Similar to the Murray – Darling situation, these models are used to provide estimates of reliability for all permitted water diversions in the state as well as analysis of the effects of all permit applications. Naturalized, or predevelopment, time series of flows were constructed for the basins, and then the effects of developments were added in to achieve models of the current situation. The process of developing the basin models was an iterative, peer reviewed calibration process subject to stakeholder comment at several critical junctures. The naturalized flows and subsequent development of the basins now form an accepted and legal basis for future water allocations. Currently, similar activities are ongoing to provide calibrated and verified models of the state's groundwater aquifers and usage.