Lund, J.R., "Random Variables Versus Uncertain Values: Stochastic Modeling and Design," *Journal of Water Resources Planning and Management*, ASCE, Vol. 117, No. 2, pp. 179-194, March, 1991.

RANDOM VARIABLES VS. UNCERTAIN VALUES: STOCHASTIC MODELING AND DESIGN

Jay R. Lund, Associate Member, ASCE
Assistant Professor, Department of Civil Engineering
University of California, Davis, CA 95616

Abstract: Recent decades have seen great progress in the use of stochastic methods to model aspects of water resource problems. However, the design implications of stochastic modeling have been relatively overlooked. In particular, there is little explicit realization that the stochastic modeling of single-valued, yet uncertain phenomena yields qualitatively different information than the stochastic modeling of multi-valued and uncertain phenomena in terms of the role of the modeled uncertainty in design decision-making. This paper explores this issue by way of some simple examples.

INTRODUCTION

Stochastic modeling has become a well accepted, but still controversial set of methods in water resources. The water resources engineer now has a plethora of analytical, approximate, and Monte Carlo techniques available for propagating probability distributions for input variables and model parameters through a set of model equations to estimate the probability distribution of model output variables (Burges and Lettenmaier 1975; Loucks et al. 1981). A large number of techniques are also available for incorporating random variables into optimization models (Tung 1986; Wagner 1975). These methods have greatly increased the sophistication of how uncertainty is considered in engineering design.

However, at times applications of stochastic modeling and optimization seem somewhat automatic and ill-considered for a particular design context. This paper argues that stochastic modeling of physically deterministic and single-valued phenomena with imperfectly known characteristics yields a qualitatively different form of information than stochastic modeling of phenomena which are physically time-varying and random. This difference is illustrated for two hypothetical examples: the selection of a water supply source and the design of a wastewater treatment plant. These examples are not intended to necessarily illustrate optimal design methodologies for such facilities. Rather, for each example, decision theoretic formulations reveal the distinction between these two types of uncertainty and the potential importance of this distinction.

UNCERTAINTY: RANDOM VERSUS IMPERFECTLY KNOWN VALUES

Uncertainty generally implies that one is unsure of the particular value a variable will take on. A variable's value may be uncertain both if the variable is single-valued, deterministic, and constant, but has an imperfectly known value, or if the variable's value is constantly fluctuating with a random pattern. In both cases, a variable's uncertainty may be expressed by a probability distribution. In the first case, the single-valued variable's value will often become increasingly well-known through experimentation and experience. In the second case, the value of a multi-valued fluctuating variable, such as streamflow or precipitation, may become no better known as experience with the variable accumulates. This practical difference in types of uncertain variables mirrors the difference between frequentist and Bayesian approaches to probability theory (Jaynes 1986).

Before a reservoir is built, its cost is always an uncertain estimate. But upon completion, the cost, which is a single-valued variable, is well known. However, the streamflow entering the reservoir is likely to remain almost as uncertain after the reservoir's construction as before. This distinction between variables with imperfectly known values and those with randomly fluctuating values can be crucial for engineering design or alternative evaluation problems. This is proven through some simple examples. In each case, important engineering distinctions arise: 1) because some designs allow more flexibility as more is learned about the project's environment while other alternatives are tailored more specifically for a narrow range of circumstances and 2) different designs can be based more or less on variables about which there is more or less ability to learn and narrow the uncertainty with time.

SOURCES OF UNCERTAINTY IN MODEL RESULTS

Uncertainty or error in the results of mathematical models can arise from four sources (Fiering and Kuczera 1982; Jackson 1975):

1) Uncertainties in the values of parameters and constants in a model's equations,
2) Uncertainties in the true values of input data to the model,
3) Errors introduced by the numerical method which solves the model's equations, and
4) Errors or uncertainties in the form of the equations which constitute the structure of the mathematical model.

The first two sources of uncertainty are perhaps most easily handled through different forms of sensitivity analysis and common stochastic modeling techniques. In some lucky cases, where probability distributions of different parameter and input values are known, the resulting uncertainty in model results can be found through analytical or Monte Carlo propagation of these input and parameter distributions through the model's equations, to derive probability distributions which represent the uncertainty in model results.

The third source of error, from numerical solution methods, can be delimited through numerical experiments where the model's numerical solution is compared with the results of other numerical solution methods (including the same method using smaller discretizations) or simple analytical solutions.

Errors in model equations are conceptually fundamental errors in the validity of the model. Their fundamental nature makes errors in equation form the most difficult to delimit and difficult to interpret in terms of their effects on the relation of model results to understanding an engineering problem. In some cases, errors in equation form that result from computationally necessary simplifications in equation form might be delimited somewhat by solving both the complete and simplified model equations for simple problems (where the more complex form is computationally tractable) and comparing results. Commonly, uncertainty in equation form represents a more serious lack of understanding of the modeled phenomenon. The practical seriousness of this lack of understanding for an engineering problem can be studied through model verification and validation studies.

This study is restricted to the relatively simple case where both the model equations and solution methods are certain and errorless. Only input and parameter uncertainties are assumed to exist. This is the most common form of stochastic modeling.

A SIMPLE WATER SUPPLY DESIGN EXAMPLE

Consider a very simple water supply design problem. The engineer has two sources of water to supply a known demand, a stream and an aquifer. The engineer must choose one source for the water supply. There is a single objective, minimizing the expected present value cost of the water supply. The yields of both the stream and the aquifer are uncertain and are presently described by probability distributions. The engineer will use decision theory to make this choice.

The stream's flow is unmodified by storage and its yield is given by the probability distribution of streamflow. This probability distribution may be estimated either by observing actual streamflows over a long period of time to form an empirical description (Jackson 1975; Stedinger 1980) or by propagating stochastic rainfall and hydraulic properties through a watershed model to derive a stochastic description of streamflow (Klemes 1978). After the stream has been observed for some time, the stream's yield will remain stochastic and time-varying, although we are likely to have a better description of the probability distribution of stream yield. In decision theory parlance, the prior probability distribution of stream yield is very much like the posterior probability distribution of stream yield, after the project is implemented and given a sufficiently long prior streamflow record. The probability distribution for stream yield is given by $P_S(Q)$, the probability that a flow of water Q will be available during a particular year.

The aquifer has a relatively constant, single-valued, and physically deterministic yield. Annual variation in yield is very small and arises from accumulated infiltration minus withdrawal. However, because the hydraulic properties of the groundwater basin are relatively unknown, the aquifer's yield cannot presently be estimated with certainty and is described by a probability distribution resulting from the application of a stochastic model of aquifer yield assuming the aquifer's single-valued, but unknown hydraulic properties are represented by random variables (Freeze 1978; Smith and Freeze 1979). Once the aquifer source has been in operation for some time, the aquifer's yield will still be deterministic, relatively constant, and now relatively known, based on operational experience. The aquifer's yield has a prior probability distribution estimated by the stochastic model, but a known single-valued yield some time after the aquifer is developed. The aquifer yield's prior probability distribution is given by $P_G(Q)$.

The demand for the water supply is a constant flow D. The initial capitalized cost of developing either source is given by K if the source always provides a yield greater than D. By coincidence, the costs of developing the stream and aquifer are equal. If the water supply source proves unable to supply demand, there are two supplemental options. The first option if the source fails to meet demand is to purchase water elsewhere. For an annual cost of C, the operation can purchase all the water it requires for a given year of water deficit. The engineer's second option is to develop some third water source with a known and constant yield sufficient to make up the deficit in any year. The development of this additional source costs K_a. The real discount rate is r and the water supply's lifetime is very long. By coincidence again, $P_S(Q) = P_G(Q)$.

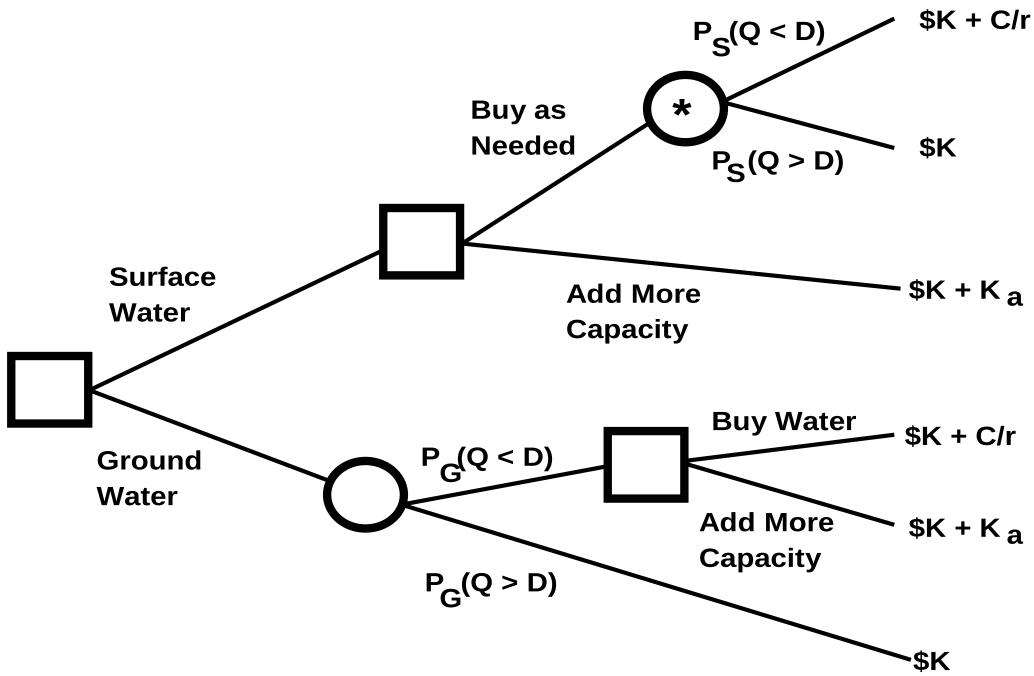

Figure 1: Decision Tree for Water Supply Example

The engineer employs decision theory to select a water source (Raiffa 1968). The decision tree depicting this choice appears in Figure 1, representing the logic of the design problem. In the figure, boxes represent decision or choices which must be made. Circles represent uncertainties, with each link to the right-hand side of a circle representing a different possible outcome. Links between boxes or circles represent outcomes. Where a link emanates from the right-hand side of a box, the designer can choose the outcome. Where the link emanates from a circle, the selection of the link is by chance. The most right-hand entry, which begins with a dollar sign, is the cost of the most specific level of outcome.

The decision tree diagrams in this paper have a slightly different notation, for illustrative purposes. Circles with stars in the center represent uncertainties that recur after the project has been implemented. In resolving the diagram, these recurring uncertainties are handled the same as non-recurring uncertainties, but their placement is juxtaposed with subsidiary design decisions in the logic of the decision tree compared with non-recurring uncertainties.

The two decisions will be compared based on their expected present value costs. For the surface water source:

(1) $EVC_S = K + \text{Min}[P_S(Q < D)\, C/r\, ,\, K_a]$,

where $P_S(Q < D)$ is the cumulative probability that the stream yield Q will be lower than D in a given year. $P_S(Q < D)\, C/r$ is the expected present value cost of annual water purchases over the project's very long lifetime.

The aquifer's expected present value cost before the aquifer's development is:

(2) $EVC_G = K + P_G(Q < D)\, \text{Min}[C/r\, ,\, K_a]$,

where $P_G(Q < D)$ is the cumulative probability that the aquifer's yield will be less than D. Equation 2 assumes that all present uncertainty in the aquifer's yield stems from uncertainty in the aquifer's hydraulic parameters. The more general expression for the aquifer source's cost, where additional temporal uncertainty is introduced from uncertain variation in aquifer inflows, is developed in Appendix II.

Equations 1 and 2 differ because the stream's yield varies each year, whereas the aquifer's yield is physically rather constant, but has an uncertain value prior to the aquifer's development. If the aquifer's yield is less than demand D, the design must suffer from this outcome for all years and one of the supplemental options will always be needed. If the stream's yield is less than D in one year it might be greater than D the next year and the need for supplemental water will be random.

When examined in this design context, the two source options have different total costs despite both sources having the same component costs and ostensibly the same stochastic descriptions of their yields. To further contrast these two forms of uncertainty, the difference between EVC_S and EVC_G is explored. Since $P_S(Q) = P_G(Q)$, $P_S(Q < D) = P_G(Q < D)$, and so let $Pf = P_S(Q < D) = P_G(Q < D)$, the probability that the source will fail in a given year. When will erroneous consideration of yield uncertainty make a difference? The difference in the expected present value costs of the two water sources is given by:

(3) $$\begin{aligned} DEVC &= EVC_S - EVC_G \\ &= K + \text{Min}[P_f\, C/r\, ,\, K_a] - K - P_f\, \text{Min}[C/r\, ,\, K_a] \\ &= \text{Min}[P_f\, C/r\, ,\, K_a] - P_f\, \text{Min}[C/r\, ,\, K_a]. \end{aligned}$$

This difference is examined for several cases.

Case I: $\text{Min}[P_f\, C/r\, ,\, K_a] = K_a$

Since this case implies that $P_f \leq 1$, $\text{Min}[C/r, K_a] = K_a$ as well for this case. This reduces Equation 3 to:

(4) $DEVC = K_a - P_f K_a$

$DEVC \neq 0$, unless $P_f = 1$.

For this case, the qualitative difference in the meanings of yield uncertainty between the stream and the aquifer will always be important unless there is no uncertainty that both sources will always fail, $P_f = 1$.

Case II: $\text{Min}[P_f\, C/r\, ,\, K_a] = P_f\, C/r$

Case II has two sub-cases. The first sub-case (Case II A) is where $\text{Min}[C/r, K_a] = K_a$. For this sub-case,

(5) $DEVC = (P_f\, C/r) - P_f K_a$

$DEVC \neq 0$, unless $K_a = C/r$ or $P_f = 0$.

For this sub-case, the qualitative difference in the meanings of yield uncertainty will be important unless there is no uncertainty that both sources are perfectly reliable, $P_f = 0$ or if the choices of how to react to failure are equivalent in terms of present value cost.

The second sub-case (Case II B) is where $\text{Min}[C/r, K_a] = C/r$. For this sub-case,

(6) $DEVC = -(P_f\, C/r) - P_f - (C/r) = 0$,

always. Here, the qualitative difference in the meaning of yield uncertainty is unimportant, in terms of the design decision.

Two different interpretations in the meaning of yield uncertainty can often have important consequences for the selection of an optimal design, as demonstrated by this simple example. These differences in meaning are derived from the physical aspects of the phenomena and the logical role of uncertainty in the decision-making or design context.

For this example, the streamflow yield can be readily viewed in frequentist term of probability theory (Jaynes 1986), where the yield will vary randomly for each time period and the probability distribution for streamflow yield is readily interpreted as a frequency distribution of yield outcomes for any given future time period. The probability distribution of groundwater yield cannot be interpreted as a frequency distribution, since this yield can only be sampled once. A more general probabilistic approach, the Bayesian approach represented by decision theory, is required to examine problems with such uncertain variables. The difference in the "information" content of the two yield probability distributions is derived in Appendix III. Where a single-valued variable with an uncertain value and a multi-valued (frequentist) random variable have the same probability density function, the density function for the single-valued uncertain variable will contain more "information," i.e., a greater certainty as to the outcome of the uncertain situation over time.

SELECTING A WASTEWATER TREATMENT PLANT DESIGN

A wastewater treatment plant is needed for an industry's effluent. Three designs are to be considered. Two designs are different conventional treatment plants; one would discharge treated effluent to an ocean outfall and the other would discharge effluent into a nearby river. Water quality models are developed for both the river and the ocean outfall to aid in designing the treatment plants. The ocean outfall water quality model has uncertain mixing parameters and uncertain hydrodynamics. However, while currently unknown, these values are assumed to be fairly constant and deterministic in nature. The river water quality model has stochastic inputs, such as streamflow and atmospheric temperature, as well as uncertain parameter values for different reaction and mixing constants. The third design is an evaporation pond, with no discharge into the river. The pond is to be sized based on a long record of evaporation measurements in the area.

When the plant produces more waste than can be treated by a design, two alternatives are assumed to be available:

1) curtail industrial plant's production, and reduce its waste stream for the period when the treatment design is inadequate, or

2) retrofit the industrial and treatment plants to reduce effluent loads permanently.

Curtailing production is a flexible response which can accommodate rare, short-lived events, such as high temperatures and low streamflows which would greatly reduce the assimilative capacity of the river. Curtailment incurs costs only when it is needed. However, curtailing production over a long term might become very expensive.

Retrofitting the industrial and treatment processes incurs a large capital cost. However, this cost and retrofit are assumed to eliminate any impairment of future production due to wastewater problems. There is of course a financial cost for over-designing the wastewater treatment plant in the beginning. The expected present value cost of each design alternative is discussed below. An infinite planning horizon is assumed. These choices are depicted in Figure 2.

Ocean Outfall Plant:

The first treatment plant design has an ocean outfall and a capital cost K1. The plant is evaluated using the ocean outfall water quality model. The model is deterministic, except for its uncertain parameter values. The yearly cost of curtailing production with this design is C1, the cost of retrofitting this plant to greatly reduce wastewater concentrations is R1, and r is the real discount rate. The expected value cost of this design before it is implemented is:

(7) $$EVC_1 = K_1 + \sum_{i=1}^{m} P_i \; F_i \; \min\left[\frac{C_{1i}}{r}, R_{1i}\right]$$

where the first term represents the plant's initial capital cost and the second term represents the antecedent estimate of the expected cost of the plant's failing to meet ocean water quality standards. Pi is the probability that parameter set i represents the real coastal water quality system. $F_i = 1$ if the design fails to maintain ocean water quality above a given standard under a particular set of model parameter values i and $F_i = 0$ if the design succeeds under parameter set i. The limit m is the number of possible sets of parameter values. Actual calculation would require a finite m, which might be done by a Monte Carlo method. If the plant fails to achieve water quality standards, the owners will choose the least expensive means of alleviating the deficiency, either permanently reducing production or retrofitting the plant to improve its treatment efficiency. The subscript i in C_{1i} and R_{1i} represent the curtailment and retrofit costs, respectively, for failure under a specific set of parameter values i. Note that if the original ocean outfall design fails, it fails for all time (unless retrofitted).

The decision to curtail production, retrofit the plant, or accept the plant as initially built, can only be made after the plant has become operational. The ocean monitoring after the plant begins operations is assumed to resolve uncertainties regarding the ocean water quality model and the initial plant's environmental impact. If the initial plant is found, after construction, to unacceptably harm water quality, then its total cost becomes:

(8) $EVC_1 = K_1 + \mathrm{Min}(C_{1i}/r), R_{1i}),$

for the parameter set i, known empirically after the plant begins operations.

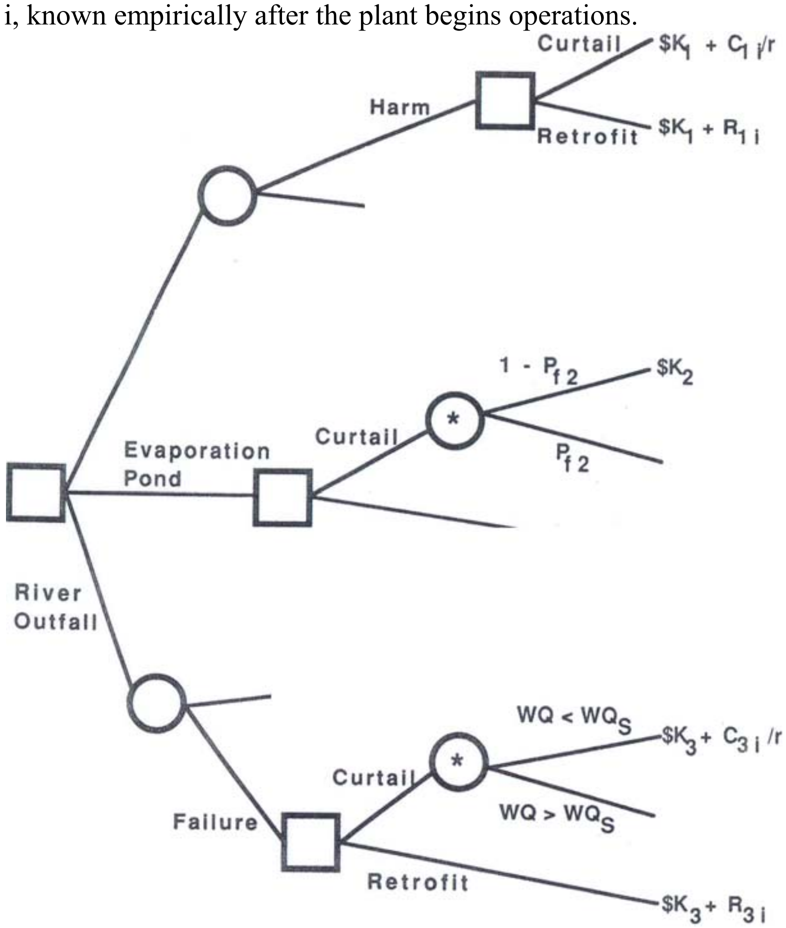

Figure 2: Decision Tree for Wastewater Treatment Example [parts of figure did not scan]

Evaporation Pond:

The evaporation pond is designed probabilistically, on the basis of a long record of local pond evaporation rates. Its initial capital cost is K_2. Since the probability distribution of wastewater treatment capacity is assumed to be known from the long record of evaporation, the design of the plant can be optimized in terms of trading off the costs of pond size and storage capacity versus the cost of curtailing industrial plant operations for extreme events where the pond cannot accommodate the plant's wastes. The expected value cost of the evaporation pond design is:

(9) $$EVC_2 = K_2 + Min(P_{f2}\ C_2/r), R_2],$$

where P_{f2} is the probability that the pond cannot accept the entire normal wastewater stream due to insufficient evaporation, C_2 is the expected loss from curtailing production for that time, and R_2 is the cost of expanding the pond to a size where it incurs a negligible failure rate. For this design, the decision of how to respond to failure can be made before the pond is operational. Little new knowledge about the pond's true operational reliability will be learned from the pond's construction and operation.

River Outfall Plant:

The river outfall plant is evaluated using a river water quality model. The water quality model is assumed to have two sources of uncertainty; the values of parameters and constants are not perfectly known (such as mixing and reaction constants) and some of the inputs into the model are stochastic and vary randomly with time (such as streamflows and atmospheric temperature). The structure of the model, embodied in its equations, is assumed to be perfect. The plant's initial construction cost is K_3. The plant's cost is given by:

(10) $$EVC_3 = K_3 + \sum_{i=1}^{m} (P_i\ F_i)\ Min[P_{3i}(WQ < WQ_s)C_{3i}/r,\ R_{3i}]$$

where Pi is the probability that a set of parameter values i represents the true riverine system, Fi = 1 if failure ever occurs with the ith parameter set ($F_i = 0$ otherwise), and $P_{3i}(WQ < WQs)$ is the probability that the plant cannot meet water quality standards WQ_s at a given time and for a given set of parameter values i. C_{3i} is the expected cost of curtailing production in response to episodic water quality failures for parameter set i, and R_{3i} is the cost of retrofitting the treatment plant and industrial process to accommodate failures under parameter set i.

Once the plant has been operated for some time, monitoring of the river will improve the accuracy of estimates for model parameters and constants. In the extreme case, one of the P_i would become equal to one. Even in this fortunate case where the parameter and constant values become perfectly clear after the plant becomes operational, there is some residual uncertainty resulting from the stochastic inputs to the water quality model and physical process (e.g., streamflow and temperature). It is only after the plant is operational that a decision is made whether to retrofit the plant or curtail production when wastewater exceeds the river's assimilative capacity.

For the riverine treatment plant, the additional information gained from the plant's construction and operation would make the choice of post-construction ameliorative measures clear (curtailment vs. retrofit). However, it is possible that enough may be known about the water quality model's parameter values for a clear-cut ameliorative decision before construction. This may be found before construction using sensitivity analysis. If one ameliorative action is always preferred for any plausible set of parameter values, additional knowledge from monitoring after plant operation is unlikely to affect the decision between ameliorative measures. (Monitoring may still have other practical benefits, however.)

Design Implications:

The design implications of mis-identifying the source of uncertainty for choosing a design can be shown for this case as they were for the previous simpler water supply example. These implications are demonstrated for the simple case where all C1i = C2 = C3i = Cc and all R1i = R2 = R3i = R. When these assumptions are made, the above cost equations can be simplified. Equation 7 becomes,

(11) $$EVC_1 = K_1 + \sum_{i=1}^{m} P_i\ F_i\ \min[Cc/r,\ R]$$

where the summed factor is separable from the minimization operator. This summation factor can be summarized as the prior probability of plant failure arising from parameter error in the model, Pfp. This leads to the further simplification:

(12) $EVC_1 = K_1 + P_{fp}\,\mathrm{Min}(C_c/r), R).$

For expected cost of the evaporation pond, the only type of uncertainty is in the evaporation inputs to the pond operating model. This uncertainty, Pf2 in Equation 9, is represented by PfI in Equation 13 and represents the probability of the plant's failure due to random system inputs.

(13) $EVC_2 = K_2 + \mathrm{Min}[P_{fI} - f(C_c/r)\ , R]$

With Equations 12 and 13, the differences between pure input uncertainty and pure parameter uncertainty can be explored as they were for the water supply case. To summarize these results, where PfI = Pfp and K1 = K2, the costs of the first two designs are only equal when:

$\mathrm{Min}[C_c/r, R] = C_c/r,$

$\mathrm{Min}[P_{fI}\ Cc/r\ , R] = R$ and $P_{fI} = 1$, or

$\mathrm{Min}[P_{fI}\ Cc/r\ , R] = P_{fI}\ C_c/r$ and $P_{fI} = 0$ or $R = C_c/r$.

For many cases, the decisions are not equally valued and the more costly design is likely to be selected if parameter uncertainty and input randomness are confused.

For the plant discharging into the river, the simplifications above do not greatly simplify the cost equation for this alternative. The slightly simpler expression is found in Equation 14.

(14) $$EVC_3 = K_3 + \min\left[\frac{C_c}{r}\sum_{i=1}^{m}(P_i\ P_{fi})\ ,R\sum_{i=1}^{m}(P_i\ F_i)\right]$$

Some additional simplification can be made when $C_c/r\ ^2\ R$. Since $P_{fIi}\ ^2\ F_i$ for any i, then when curtailment costs are less than retrofit costs ($C_c/r\ ^2\ R$), curtailment is always preferred and the minimization term in Equation 14 will have the value of the curtailment term. However, even when this simplification can be made, the choice of treatment designs remains unclear and requires explicit calculation. For this more common case, the cost effects of parameter uncertainty and input randomness are inter-related for the curtailment option.

As with the water supply example, confusing random variables with physically single-valued variables that have uncertain values can lead to a poor design choice. This is remedied by proper decision-theoretic treatment of these two types of stochastic variables. The differences in the ability to learn about these variables' behavior through the course of a series of decisions has important decision-making and decision-modeling consequences.

STOCHASTIC PROGRAMMING WITH RECOURSE

The problem discussed in this paper can arise in design and decision-making problems where a) one or more design alternatives can be modified later in response to new information and b) some random variables in the formulation represent physically single-valued variables with uncertain values whose values will become better known once an alternative is chosen or c) the uncertainty in some multi-valued random variables decreases with time (e.g., because of increased data collection). The above examples illustrate this problem and resolve it for simple cases where the decision space is small and discrete, using decision theory.

For larger problems, stochastic programming methods might be more desirable solution methods. Some such stochastic programming methods have been developed, commonly known as stochastic programming with recourse (Walkup and Wets 1967; Ziemba 1970). These problems can also sometimes be addressed with multi-stage linear programming formulations (Dantzig 1963; Wagner 1975) and stochastic dynamic programming (Yakowitz 1982). Where stochastic optimization methods cannot be reasonably expected to resolve the design problem, stochastic simulation-optimization approaches (including heuristic methods) could be employed. Solution using any of these methods requires correct formulation of the stochastic optimization problem and objective function, such as those described in the decision theory examples above.

OTHER WATER RESOURCES EXAMPLES

The differences between a prior probability distribution of streamflow and aquifer yield and input and parameter uncertainty in water quality models are not the only cases where the physical and decision-making character of an uncertain variable is important in water resources engineering. The interpretation of any stochastic description where a significant proportion of the uncertainty results from unknown values for physical constants or single-valued parameters might be significantly affected by the decision-making context of the problem. Such problems commonly arise in the problem-solving application of stochastic models of groundwater quantity, surface and ground water quality, ecological models, economic, and rainfall-runoff models. Optimization models, which incorporate a decision rule in the form of an objective function, are particularly susceptible to these problems.

CONCLUSIONS

While stochastic modeling has become a common practice, the design and decision-making implications of stochastic models and representations are often not explicitly realized. This paper suggests that the description of uncertainty is not wholly contained within the probability distributions of variables. The "information content," or the long-term ability to predict outcomes implied by two identical prior probability distributions, can be very different (Appendix II). The physical and decision-making origins of the uncertainty are also of great importance in interpreting and using probabilistic descriptions. These problems can arise when some design alternatives could be modified in the future in response to reductions in uncertainty. Such reductions in uncertainty are common results of learning about new technology, increased data collection, and greater experience with water resource and environmental problems. This is demonstrated by two simple design problems resolved by decision theory. As shown in Figs. 1 and 2, the difficulty arises where uncertainty in the real value of physically single-valued variables appears to the left, before subsidiary decisions are made regarding how to respond to any failure in the initial design. Uncertainty in physically multivalued variables appears to the right, after or contemporaneous with subsidiary decisions with regard to failure are made.

This paper does not intend to provide detailed methodologies of how to select water supply or wastewater treatment designs. Rather, it serves to illustrate a potential for mis-application of common stochastic modeling techniques to engineering problems.

Stochastic modeling of an engineering problem is likely to require more than merely developing probabilistic models of variables important to the design. In decision theory terms, the prior versus posterior nature of each probability distribution and its place within the expected value cost equations of each alternative might also be required to complete the description of variables for engineering purposes. Much of this completion of the description of uncertain variables requires physically-based or logistical insights into the nature of the stochastic decision-making problem. Incorporation of these decision theory subtleties into stochastic descriptions will greatly enhance the usefulness of stochastic modeling for engineering design (Marin 1986; Hobbs and Hepenstal 1989). The formulation of the decision-making logic of a design problem can be very important.

Acknowledgements:
The author thanks Miguel Marino, James Englehardt, and two anonymous reviewers for their very useful comments on earlier drafts of this paper.

APPENDIX I - REFERENCES

Amorocho, J. and B. Espildora (1973), "Entropy in the Assessment of Uncertainty in Hydrologic Systems and Models," Water Resources Research, Vol. 9, No. 6, December, pp. 1511-1522.

Brown, C.B. (1980), "Entropy Constructed Probabilities," Journal of the Engineering Mechanics Division, ASCE, Vol. 106, No. EM4, August, pp. 633-640.

Burges, S.J. and D.P. Lettenmaier (1975), "Probabilistic Methods in Stream Quality Management," Water Resources Bulletin, Vol. 11, No. 1, Feb., pp. 115-130.

Dantzig, G.B. (1963), Linear Programming and Extensions, Princeton University Press, Princeton, NJ, 627 pp.

Fiering, M.B. and G. Kuczera (1982), "Robust Estimators in Hydrology," in M.B. Fiering (ed.), Scientific Basis of Water-Resource Management, National Academy Press, Washington, D.C., pp. 84-94.

Freeze, R.A. (1975), "A Stochastic Conceptual Analysis of One-Dimensional Groundwater Flow in Nonuniform Homogeneous Media," Water Resources Research, Vol. 11, No. 5, pp. 725-741.

Hobbs, B.F. and A. Hepenstal (1989), "Is Optimization Optimistically Biased?," Water Resources Research, Vol. 25, No. 2, Feb., pp. 152-160.

Jackson, B.B. (1975), "The Use of Streamflow Models in Planning," Water Resources Research, Vol. 11, No. 1, pp. 54-63.

Jaynes, E.T. (1986), "Bayesian Methods: General Background," in J.H. Justice (ed.), Maximum Entropy and Bayesian Methods in Applied Statistics, Cambridge University Press, N.Y., pp. 1-26.

Kapur, J.N. (1983), "Twenty-five years of maximum-entropy principle," Journal of Mathematical and Physical Sciences, Vol. 17, No. 2, April, pp. 103-156.

Klemes, V. (1978), "Physically Based Stochastic Hydrologic Analysis," Advances in Hydroscience, Vol. 11, pp. 285-356.

Loucks, D.P., J.R. Stedinger, and D.A. Haith (1981), Water Resource Systems Planning and Analysis, Prentice-Hall, Inc., Englewood Cliffs, NJ, 559 pp.

Marin, C.M. (1986), "Parameter Estimation in Water Resources Planning and Management: Optimal Actions or Optimal Parameters," Water Resources Research, Vol. 22, No. 3, March, pp. 353-360.

Raiffa, H. (1968), Decision Analysis, Addison-Wesley, Reading, MA, 309 pp.

Shannon, C.E. (1948), "A Mathematical Theory of Communication," Bell System Technical Journal, Vol 27, pp. 379-423 and pp. 623-659.

Smith, L. and R.A. Freeze (1979), "Stochastic Analysis of Steady Groundwater Flow in a Bounded Domain, 2, Two-Dimensional Simulations," Water Resources Research, Vol. 15, No. 6, pp. 1543-1559.

Stedinger, J.R. (1980), "Fitting Log Normal Distributions to Hydrologic Data," Water Resources Research, Vol. 16, No. 3, June, pp. 481-490.

Tribus, M. (1969), Rational Descriptions, Decisions and Designs, Pergamon Press, N.Y., 478 pp.

Tung, Y-K (1986), "Groundwater Management by Chance-Constrained Model," Journal of Water Resources Planning and Management, ASCE, Vol. 112, No. 1, Jan., pp. 1-19.

Wagner, H.M. (1975), Principles of Operations Research, Prentice-Hall, Inc., Englewood Cliffs, NJ.
Walkup, D.W. and R. J.-B. Wets (1967), "Stochastic Programming With Recourse," SIAM Journal of Applied Mathematics, Vol. 15, No. 5, September, pp. 1299-1314.

Yakowitz, S. (1982), "Dynamic Programming Applications in Water Resources," Water Resources Research, Vol. 18, No. 4, August, pp. 673-696.

Ziemba, W.T. (1970), "Computational Algorithms for Convex Stochastic Programs With Simple Recourse," Operations Research, Vol. 18, No. 3, May-June, pp. 414-431.

Appendix II - Expected Aquifer Source Cost with Uncertain Hydraulic Parameters and Random Inflows

Equation 2 assumed that the only source of uncertainty in the aquifer's yield is the aquifer's hydraulic parameters, its conductivity, storativity, and spatial boundary conditions. However, Equation 2 is inadequate if additional uncertainty is introduced from truly random variables, such as random infiltration resulting from random precipitation over time. Where both forms of uncertainty are present, the expected value cost of choosing the aquifer source is:

(15) $$EVC_G = K + \sum_{i=1}^{m} (P_i \; F_i)\;)\mathrm{Min}\left[\frac{P_{Gi}(Q < D)\; C}{r}, K_a \right]$$

where each i represents a different possible set of aquifer parameter values, Pi is the prior probability of this set of values being the true set, m is the number of sets of parameter values, $F_i = 1$ if the aquifer yield is ever inadequate with this parameter set i ($F_i = 0$ otherwise), and $P_{Gi}(Q < D)$ is the probability that the aquifer with parameter value set i will have inadequate yield in a single year.

In reality the number of possible sets of aquifer parameter values is infinite, $m = \infty$. This makes direct application of Equation 4 difficult. However, EVCG can be estimated using Monte Carlo techniques, randomly choosing possible sets of parameter values for each of a finite number of i. For each of these sets of parameter values, a yield model incorporating random precipitation or inflows would be used to estimate $P_{Gi}(Q < D)$.

Appendix III - A Simple Information-Theoretic Derivation

This appendix contains a simple information-theoretic derivation of the difference in information content between a single-valued variable with an uncertain value and a variable with a constantly fluctuating value over time. This relatively esoteric approach is suggested because the difference between the groundwater and surface water yields in this case appears similar to the dichotomy discussed by Jaynes (1986) between the frequentist and Bayesian approaches to probability theory. Here, the surface water yield distribution has a clear frequentist interpretation, while the groundwater yield distribution is difficult to interpret as a frequency distribution because it will not vary between replicates (over time). This distinction is unlikely to be useful for most practicing engineers. However, it is included to provide a more fundamental distinction between the types of different stochastic situations discussed in the paper.

Shannon (1948) developed a measure of information entropy (S) as a function of event probabilities (P(Q) for event Q). Narrower probability distributions inherently contain less information entropy. At the extreme, a deterministic ($P(Q=q) = 1$) distribution contains the least entropy and typically distributions with greater variances contain progressively more information entropy. This function (Equation 16) is often interpreted as the information content of the probability distribution p(Q):

(16) $$S= -\int_a^b P(Q)\ \ln(P(Q))\ dQ$$

This function has found employment in several areas of engineering and hydrology (Amorocho and Espildora, 1973; Tribus, 1969; Brown, 1980; Kapur, 1983). Note that for two uniform distributions of varying widths (b - a), the wider distribution, which contains less "information," will have a higher entropy value, S.

Examining the water supply design example, the prior probability distribution for yield from the aquifer and stream were assumed to be identical, $P_S(Q_S) = P_G(Q_G)$, even though Q_G is physically single-valued and QS fluctuates in values and is therefore physically many-valued. The information entropy of the prior groundwater yield probability distribution is given by:

(17) $$S_G = -\int_0^\infty P_G(Q)\ \ln(P_G(Q))\ dQ$$

Since Q_G is physically single-valued, only one realization of Q_G will actually occur.

The value of Q_S fluctuates, possessing a single value for each time period, but varying randomly between time periods. The information entropy of the distribution $P_S(Q_S)$ should then be portrayed as:

(18) $$S_s = -\sum_{t=0}^{T} \int_0^\infty P_S(Q)\ \ln(P_S(Q))\ dQ$$

where T is the time horizon of the project. In the case where a large amount of historical surface water data exists, $P_S(Q)$ is unlikely to change greatly over time and Equation 18 becomes:

(19) $$S_s = -T\int_0^\infty P_S(Q)\ \ln(P_S(Q))\ dQ$$

If the prior probability distributions are the same for surface and ground water yields, as assumed above, then the information entropy for the surface water distribution will always be greater than the information entropy for the ground water yield. Even though the probability distributions are the same, the physically single-valued groundwater yield distribution contains considerably greater "information," i.e., less uncertainty and a greater ability to predict the water yield over time, in this case. This might be true even though our abilities to predict groundwater and surface water yields prior to the project are equivalent.

Appendix IV - List of Variables

C - the annual cost of water purchases
C_c - the expected annual cost of production curtailment
C_{ji} - the expected cost of curtailing production given a type j treatment plant and assuming the ith set of parameter values
D - the constant flow demand
EVC_1 - the expected value cost of the treatment plant with an ocean outfall
EVC_2 - the expected value cost of the evaporation pond
EVC_3 - the expected value cost of the riverine treatment plant
EVC_G - the expected value cost of the aquifer source
EVC_S - the expected value cost of the stream source
F_i - the probability that the system can fail (0 or 1) assuming the ith set of parameter values
K - the capital cost of developing either the aquifer or the stream
K_a - the capital cost of developing a supplemental water source
K_1 - the capital cost of developing the treatment plant with an ocean outfall
K_2 - the capital cost of developing the evaporation pond
K_3 - the capital cost of developing the riverine treatment plant
P_i - the probability that parameter value set i represents the system's true set of parameter values
P_f - the probability of a source failing
P_{fI} - the probability of a treatment failure due to random inputs
P_{fp} - the prior probability of a treatment failure due to parameter uncertainty
$P_G(Q)$ - the probability of the aquifer yielding a flow Q
$P_G(Q < D)$ - the probability of the aquifer cannot yield a flow Q
$P_S(Q)$ - the probability of the stream yielding a flow Q in a given year
$P_S(Q < D)$ - the probability of the stream cannot yield a flow Q
r - the real discount rate
R_{ji} - the cost of retrofitting a type j treatment plant assuming the ith set of parameter values
WQ - actual water quality index level
WQs - water quality index standard
DEVC - the difference in expected value costs of the two potential sources