EDITORIAL

STOCHASTIC AND DETERMINISTIC WORLD VIEWS

Introduction

When one reads the water resources literature, it appears that engineers belong to either a stochastic or a deterministic school of thought. It is rare to find an engineer who has a balanced view of the world, one that simultaneously embraces both stochastic and deterministic elements of a modeling problem. This difference in world views stems from our educational system, which tends to propagate such distinctions and tends to emphasize deterministic elements.

High school graduates have had 12 years of mathematics, often without a single course in either probability or statistics. A single course in probability or statistics is the most that one can expect in most undergraduate engineering college curricula. How can one expect engineers to understand statistics, with only a single course? Entire departments are built around the subject of statistics—how can a single course suffice? Would we feel comfortable with our understanding of calculus if we had only a single course? Obviously, our educational system places much greater weight on the deterministic elements than on the stochastic ones.

Are Stochastic and Deterministic Views Really Different?

Consider an example of a watershed model that models streamflow Q based on the climatic inputs precipitation P, and potential evapotranspiration PE. In mathematical terms:

$$Q = f_d(P,PE|\Omega) + \varepsilon \qquad (1)$$

where Ω represents the model parameters and ε represents model error. Streamflow is made up of a deterministic element $f_d(P,\mathrm{PE}|\Omega)$ and a stochastic element ε. Even though a watershed model contains both stochastic and deterministic elements, it is usually classified as a deterministic model, presumably because once P and PE are specified, the deterministic components dominate the process. Deterministic models are not necessarily physically based; they often contain empirical components. A model is classified as deterministic here if the internal structure of the model at least attempts to capture some physical processes. Therefore, a ''black box'' model of a catchment would not be a deterministic model, whereas the various quasi-empirical, hydrologic-design-oriented rainfall-runoff models, such as U.S. Army Corps of Engineers HEC-1 and HMS models or the Natural Resource Conservation Service TR-55 and TR-20 models, are deterministic models, because they attempt to represent watershed processes such as infiltration and channel routing in addition to predicting watershed discharge.

Usually one's mathematical objective is to minimize the error term ε, and one model is often preferred over another if it can be shown that its error term is smaller. So the calibration of a deterministic watershed model reduces to minimizing its stochastic elements. Sometimes validation efforts associated with deterministic models focus on stochastic aspects such as preservation of bias and variance of the streamflows. Sometimes validation efforts focus on model residuals alone, as if the deterministic model were a statistical regression model. Engineers and scientists have not yet learned how to calibrate or validate a deterministic model without resorting to an analysis of its stochastic component.

Now consider a stochastic streamflow model for comparison. Consider the model

$$Q = f_s(P,PE|\Omega) + v \qquad (2)$$

where Ω again represents the model parameters; v represents model error; and P and PE are inputs. Streamflow is still made up of a deterministic element $f_s(P,PE|\Omega)$ and a stochastic element v. What makes this a stochastic streamflow model is that its structure is derived to assure that certain statistical properties of the generated streamflows are preserved. In other words, the deterministic component of the model $f_s(P,PE|\Omega)$ is derived so as to insure reproduction of certain characteristics of the streamflows and to insure reproduction of the covariance matrix between the inputs P and PE and the outputs Q. The deterministic component of a stochastic model is usually not derived from physical processes, as is the deterministic component of a deterministic model, meaning that stochastic models cannot be used for the same purposes intended for most deterministic models. Nevertheless, stochastic and deterministic models have the same general structures because they are both made up of stochastic and deterministic components. The purpose of this editorial is to dramatize the similarities between stochastic and deterministic models in the hopes that one day there will be no distinction between the two. There are advantages to both types of models and, ultimately, improvements in modeling will arise from a unified stochastic/deterministic view of modeling.

Unified Approach to Modeling

Stochastic models are derived to assure that certain properties of the output (streamflow) are reproduced, such as its probability distribution or perhaps its mean, variance, skewness, and serial correlation. Such models are also derived to insure that basic relationships between inputs and outputs are reproduced, such as the cross correlation between P and Q and between PE and Q in (2). Since these properties are often the basis of the model derivation, stochastic models are certain to reproduce these properties, unless the model is either theoretically flawed or improperly verified. See Stedinger and Taylor (1982) for a discussion on model verification and model validation. These important properties of stochastic models come at the expense of a model that cannot reproduce important physical processes, and so stochastic models are appropriately referred to as ''black box'' models. Wouldn't stochastic models benefit if they could also represent certain important internal physical processes?

Deterministic models are derived, on the other hand, to reproduce certain physical processes inherent to the process being modeled. For example, a watershed model of streamflow may contain physical models of evapotranspiration, infiltration, ground water, snowmelt, etc. Deterministic models are designed to represent internal physical processes, enabling a wide range of model applications that stochastic models are unable to accomplish. Yet deterministic modelers often lament that modeled streamflows tend to have lower variance than observed streamflow. Deterministic modelers rarely examine such properties as the skewness or serial correlation of streamflows. Even more rare are evaluations of the ability of deterministic models to reproduce the cross correlations among the inputs and outputs to the model. Wouldn't deterministic models benefit from the types of statistical evaluations that are routinely performed on stochastic models?

Deterministic models will never be able to reproduce the variance of observed streamflows, because as long as the model residuals are independent of the model inputs

$$\mathrm{var}[Q] = \mathrm{var}[f_d(P,PE|\Omega)] + \mathrm{var}[\varepsilon] \qquad (3)$$

so that it will always be true that

$$\mathrm{var}[\, f_d(P,PE|\Omega)] < \mathrm{var}[Q] \tag{4}$$

unless there is no stochastic component, in which case var[ε] = 0. Therefore, the stochastic component will always play a central role, even for deterministic models.

Numerous advantages arise from combining stochastic and deterministic world views when deriving, testing, and applying models. A stochastic world view can help us develop deterministic models that can faithfully represent observed relationships between model inputs and model outputs. A deterministic world view can help us develop stochastic models that can faithfully represent certain observed internal physical processes. Ultimately, if both initiatives are undertaken, models will be neither stochastic nor deterministic. Instead models will balance both perspectives, leading eventually to more believable and useful models. Such developments will occur only after our educational institutions provide a more balanced view of stochastic and deterministic elements.

Recommendations

Statistical Methods Can Lead to Improved Deterministic Models

There are now many examples of researchers who have attempted to apply statistical methods to classical problems in deterministic hydrology. Research has addressed numerous statistical issues in the calibration of deterministic watershed models, including studies that deal with heteroscedasticity of model residuals, model parameter uncertainty, measurement error, and spatial and temporal scaling. One observes, from the watershed modeling literature, a continuing evolution in terms of the statistical rigor incorporated into the analyses. Yet deterministic hydrologic models are still usually calibrated using streamflow alone, as if reproduction of observed streamflow were the holy grail. So even though very sophisticated statistical methods are now in common use, deterministic hydrologists still seem to have much more faith in a single realization of observed streamflow than is warranted. When a stochastic world view is integrated into deterministic modeling, only then will hydrologists fully understand the information content associated with deterministic models, model parameter estimates, model error, and the streamflow sequences such models are designed to mimic.

Deterministic Methods Can Lead to Improved Statistical Models

Although statistical methods are widely used for solving many problems in water resource engineering, such models are not always reliable. For example, statistical hydrologic models of low flow design events are known to perform poorly. To improve much models, Vogel and Kroll (1992) showed that a physically based watershed model of groundwater outflow can suggest variables and the functional form for regional statistical models of design low flow statistics. Perhaps the best example is provided by Wallis (1965), who showed how difficult it is to uncover basic physical relationships using multivariate statistical procedures but without prior knowledge of the physical relationships involved. The development of flood and low flow frequency methods derived from basic deterministic watershed models provides a good example of how deterministic methods can lead to improvements in statistical models. Raines and Valdes (1993) review recent developments in derived distributions of flood frequency.

On the Value of Cross Disciplinary Research

There are many cross-disciplinary lessons to be learned from previous research on what appears to be an unrelated problem. It is often fruitful to exploit developments in related fields. This is commonplace in hydrology, which has benefited from thousands of developments in the fields of mathematics, geomorphology, and meteorology, just to name a few. There are also lessons within one subfield of hydrology that could benefit other subfields. For example, flood frequency analysis has evolved to the point where it is now common practice to exploit regional information, in addition to at-site data, for the estimation of design streamflows for various water resource engineering applications. This idea goes against the natural intuition of most deterministic hydrologists, who tend to trust only ''at-site'' data. The deterministic hydrologists argument is that neighboring basins exhibit fundamentally different behavior due to natural regional heterogeneity. Statistical hydrologists have shown over and over that use of regional information in flood frequency analysis can lead to dramatic improvements over the use of at-site information alone. Such regional hydrologic methods as the ''index flood'' method and regional hydrologic regression methods are now preferred over at-site methods for flood frequency analysis, unless one is fortunate to have well in excess of about 50 years of streamflow. A natural extension would be to use regional information when calibrating deterministic hydrologic models. Hopefully, improvements in deterministic hydrologic modeling will result when regional hydrologic structure and response is integrated into the structure of, and the calibration of, physically based models.

On the Value of Education in Statistics

This editorial has argued for the development of a more balanced view of stochastic and deterministic elements within the framework of water resources engineering and hydrologic modeling. The explosion in the amount of hydrologic, climatic, geologic, geomorphic, topographic, economic, and other data available to engineers and scientists creates an even greater need for expertise in the field of statistics. Yet high school and undergraduate programs have not created corresponding growth in their curriculum relating to statistical issues (Higgins 1999). The challenge of increasing the educational background of water resource engineers and hydrologists in the area of statistics is significant. Only 5% of all colleges and universities have departments of statistics, whereas computer science has achieved departmental status at approximately 50% of all U.S. academic institutions (Higgins 1999). Even within the field of mathematics, the subfield of statistics has its challenges. Higgins argues that the current culture in most math departments makes it difficult for statistics to prosper. Increased attention to statistical issues within high school, undergraduate, and graduate university curricula is required to enable hydrologists and water resource engineers to achieve the type of balance between statistical and deterministic issues described in this editorial.

APPENDIX. REFERENCES

Higgins, J. J. (1999). ''Nonmathematical statistics: A new direction for the undergraduate discipline.'' *The American Statistician*, 53(1), 1–6.

Raines, T. H., and Valdes, J. B. (1993). ''Estimation of flood frequencies for ungauged catchments.'' *J. Hydr. Engrg.*, ASCE, 119(10), 1138–1154.

Stedinger, J. R., and Taylor, M. R. (1982).''Synthetic streamflow generation, 1. Model verification and validation.'' Water Resour. Res., 18(4), 909–918.

Vogel, R. M., and Kroll, C. N. (1992). ''Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics.'' *Water Resour. Res.*, 28(9), 2451–2458.

Wallis, J. R. (1965). ''Multivariate statistical models in hydrology: A comparison using data of known functional relationship.'' *Water Resour. Res.*, 1(4), 447–461.

Richard M. Vogel
Dept. of Civil and Environmental Engineering
Tufts University
Medford, MA 02155
E-mail: rvogel@tufts.edu