Fuzzy-Probabilistic Calculations of Water-Balance Uncertainty

Boris Faybishenko
Lawrence Berkeley National Laboratory
1 Cyclotron Road, MS 90-1116
Berkeley, CA 94720

bafaybishenko@lbl.gov
Tel: 510-486-4852, Fax 510-486-5686

Fuzzy-Probabilistic Calculations of Water-Balance Uncertainty

Boris Faybishenko
Lawrence Berkeley National Laboratory

Abstract

Hydrogeological systems are often characterized by imprecise, vague, inconsistent, incomplete, or subjective information, which may limit the application of conventional stochastic methods in predicting hydrogeologic conditions and associated uncertainty. Instead, predictions and uncertainty analysis can be made using uncertain input parameters expressed as probability boxes, intervals, and fuzzy numbers. The objective of this paper is to present the theory for, and a case study as an application of, the fuzzy-probabilistic approach, combining probability and possibility theory for simulating soil water balance and assessing associated uncertainty in the components of a simple water-balance equation. The application of this approach is demonstrated using calculations with the RAMAS Risk Calc code, to assess the propagation of uncertainty in calculating potential evapotranspiration, actual evapotranspiration, and infiltration—in a case study at the Hanford site, Washington, USA. Propagation of uncertainty into the results of water-balance calculations was evaluated by changing the types of models of uncertainty incorporated into various input parameters. The results of these fuzzy-probabilistic calculations are compared to the conventional Monte Carlo simulation approach and estimates from field observations at the Hanford site.

Keywords: Water balance, uncertainty, fuzzy-probabilistic approach, fuzzy calculations.

1. Introduction

Hydrogeological predictions are subject to numerous uncertainties, including the uncertainty involved in developing conceptual, mathematical, and numerical models, and determining their parameters. The uncertainty quantification of hydrological processes has recently become a subject of intense research in stochastic simulation (Meyer et al. 1997; Neuman and Wierenga 2003; Neuman 2003; Winter 2004; Orr and Meystel 2005; Wagener and Gupta 2005; Ye et al. 2004; 2005). Stochastic simulation of hydrogeological systems is usually based on the assumption that spatial and temporal variations in input and output hydrogeological parameters are random, and described by known probability distributions—so that their uncertainties are also quantified using probability distributions. However, some hydrogeological parameters are characterized by imprecise, vague, inconsistent, incomplete, or subjective information, insufficient for constructing reliable probability distributions and limiting the application of conventional stochastic methods.

Still, during the last 30 to 40 years, several alternative approaches for modeling complex systems with uncertain parameters have been developed, based on fuzzy set theory, possibility theory (Zadeh 1978; 1986; Dubois and Prade 1994; Yager and Kelman 1996), rough sets, imprecise probability, belief functions, the Dempster-Shafer theory of evidence (Dempster 1967; Shafer 1976; Smets 1990; Zadeh 1986), and fuzzy random variables. Some of these approaches include the blending of interval or fuzzy-interval analysis with probabilistic methods (Ferson and Ginzburg 1995; Ferson 2002; Ferson et al. 2003). Applying methods of fuzzy calculus and fuzzy-probabilistic (hybrid) modeling to hydrological problems is a relatively new direction for hydrological research, risk assessment, and sustainable water resources management under uncertainty (Chang 2005). The motivation of this paper is to demonstrate the potential of the application of this approach for water-balance calculations, given the general inadequacy of meteorological information and field data collected at regional scale.

The objective of this paper is to present the theory and application (through a case study) of the fuzzy-probabilistic approach for assessing the uncertainty involved in hydrogeological modeling, based on the integration of probability and possibility theories. The paper will present the main concepts involved in combining statistical and fuzzy-calculus analyses. A case study, using the data from the Hanford, Washington, USA, site, is developed to evaluate the uncertainty involved in assessing soil water balance—potential evapotranspiration, evapotranspiration, and infiltration. Since the water-balance calculations are conducted in several steps, the author will also analyze how uncertainty could be propagated across multiple submodels of the water-balance model.

The structure of the paper is as follows: In Section 2, the general approaches involved in a hydrogeological uncertainty analysis, as well as the concepts of fuzzy modeling and the integration of stochastic and fuzzy modeling approaches, will be discussed. In Section 3, the notion of fuzziness within a hydrogeological system will be introduced, including the description of (1) a general form of the water-balance equation, (2) the Penman equation for evaluating potential evapotranspiration (E_o), (3) a modified Budyko's model for evaluating evapotranspiration (ET), and (4) an approach for evaluating net infiltration (i.e., drainage below the bottom of the evapotranspiration zone within the soil profile). In Section 4, input parameters and calculation results for the Hanford site will be presented. In Section 5, conclusions will be provided.

2. Types of and Approaches to Uncertainty Analysis

2.1. Types of Uncertainties

The uncertainties involved in hydrogeological predictions can generally be categorized into two groups—*aleatory* and *epistemic* uncertainties. Aleatory uncertainty arises because of the natural, inherent variability of subsurface properties or input and/or model parameters, such as meteorological parameters or subsurface heterogeneity. If a sufficient amount of information is available, stochastic

simulations with parameters generated by probability density functions (PDFs) can be used to assess aleatory uncertainty. This type of uncertainty is also referred to as irreducible, stochastic, or random uncertainty.

Epistemic uncertainty arises because of a lack of knowledge about processes and/or models, insufficient experimental data to characterize subsurface flow and transport processes, or poor understanding of coupled-physics phenomena and events. One of the reasons for epistemic uncertainty is the lack of reliable experimental data, which can be ambiguous, conflicting, insufficient, or not in agreement with existing conceptual models used for predictions and quantifying uncertainty. This type of uncertainty is also referred to as subjective or reducible uncertainty, because it can be reduced as new information becomes available, and by using various models for uncertainty evaluation. Epistemic uncertainty decreases as knowledge about a system or model (and their parameters) increases.

Note that *uncertainty* (both aleatory and epistemic) caused by a lack of knowledge regarding either the model or subsurface properties is different from *error*, which may arise from a human mistake or faulty sensor. Moreover, positive and negative errors may cancel each other out, while uncertainty is always additive (just as variances are always positive). Generally, variability, imprecise measurements, and errors are distinct features of uncertainty; however, they are very difficult, if not impossible, to distinguish (Ferson and Ginzburg 1996). In this paper, the author will consider the effect of aleatory uncertainty on water-balance calculations by assigning the extent of input parameters, and the effect of epistemic uncertainty is considered by assigning different models for input parameters, using probability distributions and fuzzy numbers.

2.2. Approaches to Assessing Uncertainty

One classical approach to assessing uncertainty is based on Monte Carlo simulations (i.e., random sampling) of PDFs describing system parameters. Other approaches, used for assessing uncertainty in the event of insufficient information for constructing reliable PDFs, are based on the specification of uncertain parameters using probability boxes (p-box), interval numbers, and fuzzy numbers, which are schematically shown in Figure 1. For example, an uncertain number expressed with a probability distribution, as shown in Figure 1a, can be represented as an interval number, as shown in Figure 1b. The application of interval analysis for modeling is based on the assumption that model inputs are within their realistic (usually, the largest) intervals. Interval analysis is often used for assessing extreme values and ranges of uncertainty—for example, worst-case scenarios—when the input information is poor and cannot be used to develop reliable PDFs.

The probability box (p-box) approach combines both interval and probability methods, by imposing bounds on a cumulative distribution function (CDF) used to express different sources of uncertainty. This method provides an envelope of distribution functions, which bounds all possible dependencies (Ferson 2002; Ferson et al. 2003). Figure 1c illustrates

a p-box [$\bar{F}, \underline{F}$] representation of a random variable *X*, for which the exact distribution *F* is unknown, except that it is located within the p-box. In Figure 1c, the right curve $\underline{F}(x)$ is a lower bound on an imprecisely known probability, showing that the random variable *X* is smaller than *x*, and the left curve $\bar{F}(x)$ is an upper bound on probability and a lower bound on the *x* quantity. With better empirical information, when input information is abundant, the p-box bounds are narrower, and the results of predictions come close to those from traditional probability theory.

Using probability for evaluating uncertainty is a way of predicting an object's (or event's) occurrence. In other words, probability theory can be used to determine the likelihood of (and the uncertainty over) an event's occurrence. Because probability is the theory of random events, it is not capable of capturing uncertainty resulting from vagueness of input parameters and system behavior. Possibility theory, on the other hand, using fuzzy numbers, is concerned with event ambiguity, or the extent to which an event occurs, given incomplete information expressed in terms of fuzzy numbers. Fuzzy modeling theory or possibility theory, introduced by Zadeh (1978), can be used for modeling of uncertainty in complex systems—systems that contain imprecise, vague, inconsistent, incomplete, or subjective information.

A fuzzy set is defined as a generalized set to which objects can belong with various degrees of membership over the interval [0,1]. A *fuzzy number* can be defined as a family of intervals representing an estimate of an uncertain number. Based on this definition of a fuzzy number, fuzzy arithmetic can be considered a generalization of the interval analysis for treating uncertainty (Dubois and Prade 1981; Ferson 2002). Because fuzzy arithmetic theory is generally based on using less stringent axioms than probability theory, it can be applied to a wider spectrum of uncertainty problems. For instance, it may be appropriate for handling nonstatistical uncertainty, e.g., resulting from inaccurate measurements or subjective evaluation, including the consensus of opinions of observers.

The fuzzy-probabilistic approach (which is also called a hybrid approach) is used when some quantities can be represented by fuzzy numbers and other quantities by probability distributions and interval numbers (Kaufmann and Gupta 1985; Ferson 2002; Guyonnet et al. 2003; Cooper et al. 2006). This approach is presented in Section 2.3.2.

2.3. Fuzzy System and Fuzzy-Probabilistic Approach

2.3.1. Fuzziness and a Fuzzy Number

For the past 30 to 40 years, the theory of fuzziness has successfully been applied to describe such systems as complex, large-scale engineering systems, social systems, economic systems, management systems, medical diagnostic processes, human perception, and others. The term *fuzziness* is, in general, used to describe objects or processes that cannot be given precise definition or precisely measured. *Fuzziness*

identifies a class (set) of objects with nonsharp (i.e., fuzzy) boundaries, which may result from imprecision in the meaning of a concept or measurements used to characterize and model the system. Fuzzy processes can be defined as vague processes with some inherent uncertainty in their description, so that a fuzzy system may not have precise boundaries or parameters. Fuzzification implies replacing a set of crisp (i.e., precise) numbers with a set of fuzzy numbers, using fuzzy membership functions based on the results of measurements and perception-based information (Zadeh 1978).

The central idea of fuzzy theory and fuzzy arithmetics is the notion of a fuzzy number. A fuzzy number is a quantity whose value is imprecise, rather than exact (as is the case with "ordinary" single-valued numbers). Any fuzzy number can be thought of as a function whose domain is a specified set of real numbers. Each numerical value in the domain is assigned a specific "grade of membership," with 0 representing the smallest possible grade (full nonmembership), and 1 the largest possible grade (full membership). The grade of membership is also called the degree of possibility. In other words, a fuzzy number is a fuzzy subset of the domain of real numbers, which is an alternative approach to expressing the uncertainty. A fuzzy variable has a fuzzy membership function (FMF), which is defined on a *universe of discourse* that ranges over a set of possible values for the fuzzy variables. Several types of membership functions are commonly used for fuzzy-systems modeling: triangular, trapezoidal, Gaussian, sigmoid, bell-curve, Pi-, *S*-, and *Z*-shaped curves. As an illustration, Figure 1d shows a triangular fuzzy membership function given by

$$f(x) = \begin{Bmatrix} 0, x \le a \\ \dfrac{x-a}{b-a}, a \le x \le b \\ \dfrac{c-x}{c-b}, b \le x \le c \\ 0, c \le x \end{Bmatrix}, \tag{1}$$

where a=6, b=10, and c=14,

a trapezoidal fuzzy membership function given by

$$f(x) = \begin{Bmatrix} 0, x \le a \\ \dfrac{x-a}{b-a}, a \le x \le b \\ 1, b \le x \le c \\ \dfrac{d-x}{d-c}, c \le x \le d \\ 0, d \le x \end{Bmatrix}, \tag{2}$$

where a=6, b=9, c=11, and d=14,

and a Guassian fuzzy membership function given by

$$f(x) = \exp\left(\frac{-(x-c)^2}{2\sigma^2}\right), \tag{3}$$

where c=10, and σ^2=1.

These membership functions are convex, starting at the grade of zero, monotonically increasing to a maximum of 1, and then monotonically declining to zero as the x-domain increases. (Note that an ordinary set is a special case of a fuzzy set, with the degree of membership either 0 or 1.)

One of the most important attributes of fuzzy numbers is the notion of an α-cut, illustrated in Figure 1d. The range of the fuzzy variable at an α-cut of near 0 may represent a pessimistically wide uncertainty of the variable, and the range of the fuzzy variable at an α-cut value of near 1 represents the optimistically narrowest uncertainty of the variable. The α-cut interval is a crisp interval, limited by a pair of real numbers. In this case, FMFs are themselves precise/deterministic, although they are used to represent uncertain parameters. (This type of fuzzy numbers is called a Type-1 fuzzy number.) In this paper, the following characteristic parameters are used to express the uncertainties of calculations using fuzzy numbers and p-boxes: mean, core, iqrange, and breadth of uncertainty. The definitions of these parameters (given as intervals) are described in Table 1.

Arithmetic operations on fuzzy numbers can be based on the idea of using the interval calculus, using several α-cuts of the fuzzy numbers, and then constructing corresponding output fuzzy numbers, as illustrated in Figures 2a and 2b. An interval is defined as a bounded subset of real numbers

$$X = [a, b] \text{ is an interval} \Longleftrightarrow (X = \{x \in \mathbf{R} \mid a \le x \le b\}). \tag{4}$$

Arithmetic operations on intervals always contain every possible outcomes of the corresponding arithmetic operation on real numbers, so that the result of $X \diamond Y$ is an interval, Z, given by

$$X \diamond Y = Z = \{z = x \diamond y \mid x \in X, y \in Y\}, \tag{5}$$

where $\diamond$ denotes one of the operations $\{+, -, \times, /\}$, which can be expressed in terms of ordinary arithmetic on the interval endpoints:

$$\boldsymbol{X} + \boldsymbol{Y} = [\underline{\boldsymbol{X}} + \underline{\boldsymbol{Y}}, \bar{\boldsymbol{X}} + \bar{\boldsymbol{Y}}] \qquad \boldsymbol{X} - \boldsymbol{Y} = [\underline{\boldsymbol{X}} - \bar{\boldsymbol{Y}}, \bar{\boldsymbol{X}} - \underline{\boldsymbol{Y}}] \tag{6}$$

$$\boldsymbol{X} \times \boldsymbol{Y} = [\min(\underline{\boldsymbol{X}}\cdot\underline{\boldsymbol{Y}}, \underline{\boldsymbol{X}}\cdot\overline{\boldsymbol{Y}}, \overline{\boldsymbol{X}}\cdot\underline{\boldsymbol{Y}}, \overline{\boldsymbol{X}}\cdot\overline{\boldsymbol{Y}}), \max(\underline{\boldsymbol{X}}\cdot\underline{\boldsymbol{Y}}, \underline{\boldsymbol{X}}\cdot\overline{\boldsymbol{Y}}, \overline{\boldsymbol{X}}\cdot\underline{\boldsymbol{Y}}, \overline{\boldsymbol{X}}\cdot\overline{\boldsymbol{Y}})] \quad (7)$$

$$\boldsymbol{X} / \boldsymbol{Y} = [\min(\underline{\boldsymbol{X}}/\underline{\boldsymbol{Y}}, \underline{\boldsymbol{X}}/\overline{\boldsymbol{Y}}, \overline{\boldsymbol{X}}/\underline{\boldsymbol{Y}}, \overline{\boldsymbol{X}}/\overline{\boldsymbol{Y}}), \max(\underline{\boldsymbol{X}}/\underline{\boldsymbol{Y}}, \underline{\boldsymbol{X}}/\overline{\boldsymbol{Y}}, \overline{\boldsymbol{X}}/\underline{\boldsymbol{Y}}, \overline{\boldsymbol{X}}/\overline{\boldsymbol{Y}})], \quad (8)$$

with the condition for division that $0 \notin \boldsymbol{Y}$, and that the underbar expresses the smallest value and the overbar expresses the largest value of the interval.

Fuzzy arithmetic is generally applicable for evaluating all kinds of uncertainty, regardless of its source or nature. It is based on the application of both hard data and the subjective (perception-based) interpretation of data. Fuzzy arithmetic approach provides a distribution characterizing the results of all possible magnitudes, rather than just specifying upper or lower bounds. Fuzzy methods can be combined with calculations with PDFs, interval numbers, or p-boxes, using the RAMAS Risk Calc code (Ferson 2002), described in Section 2.3.2.

When fuzzy measures serve as upper bounds on probability measures (Figure 2a), one could expect to obtain a conservative (bounding) prediction of system behavior (Figure 2b). Although calculations using fuzzy arithmetic may enclose the true answer, the results may overestimate uncertainty. For example, the application of fuzzy arithmetic is not optimal (i.e., it overestimates uncertainty) when sufficient data are available to construct reliable PDFs needed to perform a Monte Carlo analysis. Moreover, fuzzy arithmetic may lead to erroneous results if the α-levels used for the fuzzy calculus are not comparable for different variables. Fuzzy arithmetic does not yield conservative results when the dependencies among variables are unknown. Also, fuzzy arithmetic cannot make use of knowledge about correlations among input variables to reduce bounds on estimates. As with using interval analysis and probability methods, repeated variables may make calculations cumbersome (Ferson 2002).

2.3.2. Fuzzy-Probabilistic Calculations using RAMAS Risk Calc Code

A pair (*F*, *P*), or a combination of a probability distribution, *P*, with a fuzzy number, *F*, is called a "hybrid number," and can be obtained by convolving the respective *P* and *F* functions, according to fuzzy arithmetic and probability theory rules (Ferson 2002). Hybrid numbers can be represented by nesting the interval numbers with uncertain bounds, which are defined using probability distributions at several discrete α-levels of fuzzy uncertainty (Figure 2c). Arithmetic operations using fuzzy numbers and probability distributions include applying fuzzy-interval calculus to a selected number (called hybrid levels) of the fuzzy-interval α-levels. Figure 2d shows an example of hybrid multiplication, using numbers given by PDFs and fuzzy numbers for four α-levels. Calculations are performed using the RAMAS Risk Calc software (Ferson 2002), which supports calculations using scalars, probability bounds, fuzzy arithmetic, and interval analysis. This code automatically carries all the uncertainties through calculations, using

standard arithmetic operations, mathematical functions, comparison operations, basic logical connectives, and string manipulation.

3. Fuzzy Water-Balance Equations

3.1. Fuzziness of Hydrogeologic Systems and Fuzzy-Probabilistic Water Balance Equation

Fuzziness in the description and prediction of hydrogeologic systems arises from the potential deficiency in assessing both epistemic and aleatory uncertainties. Fuzziness in epistemic uncertainty arises from vagueness in selecting adequate conceptual, mathematical, and numerical models, as well as types of input parameters for hydrogeological modeling. In hydrogeology, fuzziness in aleatory uncertainty results from vagueness in assessing the hydrogeological parameters used to describe the flow-field heterogeneity of subsurface fractured-porous media. For example, fuzziness arises from insufficient measurements needed to quantify the spatial and temporal randomness of hydrogeological parameters such as hydraulic conductivity, transmissivity, porosity, and moisture content. Vague, sparse, or insufficient data collection in individual wells or weather stations, insufficient observations for assigning parameters of models, multiple models describing subsurface water flow and chemical transport—all lead to fuzzy knowledge about the subsurface, causing fuzziness in describing the components of the water-balance equation.

The general form of the fuzzy-probabilistic water-balance equation is the same as that for conventional water-balance calculations, which is given by

$$P = ET + I_n + S + (R_{off} - R_{on}) \tag{9}$$

where P is total precipitation, including snowmelt, ET is evapotranspiration, S is the soil water change in storage, R_{off} is runoff, R_{on} is runon, and I_n is the water flux below the root zone (if $I_n > 0$, the flux is downward, which is the source of net infiltration, and if $I_n<0$, it is upward flow, which is the source of evapotranspiration). For fuzzy-probabilistic calculations, all terms of Equation (9) are to be assigned using probability distributions or fuzzy numbers. For example, if time-series data from meteorological stations are known and contain sufficient information to construct PDFs, precipitation and evapotranspiration can be assigned using PDFs. In the event of a deficit of data, fuzzy numbers are used.

For first-order, long-term water-balance calculations within arid and semi-arid areas, we can reasonably assume that (a) soil water storage does not change (S=0), (b) lateral water motion within the shallow subsurface is negligible, (c) the terms of the surface-water runoff and runon for regional-scale calculations simply cancel each other out, and (d) ET is determined as a function of the aridity index, ϕ: $ET=f(\phi)$, where $\phi = E_o/P$, which is the ratio of potential evapotranspiration, E_o, to precipitation, P (Arora 2002). Then, a simplified water balance equation can be given by

$$P= ET + I = f(\phi) + I \tag{10}$$

The steps involved in evaluating net infiltration, using the input probability distributions and/or fuzzy parameters, are as follows: (1) calculate E_o from the Penman equation, (2) evaluate *ET* from a modified Budyko equation, given in Section 3.2, and (3) calculate net infiltration from Equation (10).

3.2. Modified Budyko's Equation for Evaluating Evapotranspiration

Using the data from a number of catchments around the world, Budyko (1974) found that the empirical relationship between the ratio of *ET/P* and the aridity index could be described using:

$$ET/P = \{\phi \tanh (1/\phi) [1 - \exp (-\phi)]\}^{0.5} \quad (11)$$

Equation (11) can also be given as a simple exponential expression:

$$ET/P = a[1-\exp(-b\phi)] \quad (12)$$

with coefficients a =0.9946 and b =1.1493. The correlation coefficient between the calculations using (11) and (12) is 0.999. Application of the modified Budyko's equation, given by an exponential function (12) with the ϕ value in single term, will simplify calculations using the RAMAS Risk Calc code.

Although the original Budyko's model was developed to determine surface runoff, a Budyko-like approach can be used to assess an infiltration-runoff component of the water balance and the catchment-scale soil moisture capacity (Potter et al. 2005; Sankarasubramanian and Vogel 2003).

The input parameters used in these calculations are given in Section 4.1. For the sake of simplicity of simulations, these input parameters are assumed to be independent variables.

4. Input Parameters and Results of Calculations

4.1. Input Parameters for the Hanford Site

4.1.1. Site Description and Field Estimates of E_o, *ET* and Infiltration

The Hanford Site in Southeastern Washington State is the largest environmental cleanup site in the USA, comprising 1,450 square km (560 square miles) of semiarid desert. Located north of Richland, Washington, the Hanford Site is bordered on the east by the Columbia River and on the south by the Yakima River, which joins the

Columbia River near Richland, in the Pasco Basin. The Pasco Basin is one of the structural and topographic basins of the Columbia Plateau. The areal topography is gently rolling and covered with unconsolidated materials, which are sufficiently thick to mask the surface irregularities of the underlying material. Sedimentary material consisting of silts, sands, and gravel of the Hanford Formation and Ringold Formation cover thick basalt flows. Partially saturated sediments have a high absorption capacity (Neitzel 1996). Areas adjacent to the Hanford Site are primarily agricultural lands.

Meteorological information (temperature and precipitation time-series data) used to assign model input parameters were taken from the Hanford Meteorological Station (HMS--see http://hms.pnl.gov/), located at the center of the Hanford Site just outside the northeast corner of the 200 West Area. Wind data were taken from the DOE report (DOE 1996).

At the Hanford Site, the *ET* is estimated to be 160 mm/yr, and the groundwater recharge (or infiltration) ranges from <0.1 mm/yr to ~100 mm/yr (Figure 3). The natural average recharge rate at the 200 Area is 1 mm/yr, through fine-textured soil with deep-rooted vegetation (Gee et al. 1992; 2007). This value is approximately one-tenth of the recharge volume from artificial sources, such as Hanford cribs. Based on the results of lysimeter studies, groundwater recharge at the 200 Area is from 0 to 2 mm/yr (Routson and Johnson 1990). Some episodic groundwater recharge may occur following periods of high precipitation in topographic depressions, highly permeable (such as gravel or coarse sand) surface deposits, and no-vegetation (bare-ground) land, as well as during the snow melting in the spring. According to independent geochemical estimates, using $^{87}Sr/^{86}Sr$ isotope ratios in the pore water, acid extracts, and sediments of a 70 m vadose zone at the Hanford site, the long-term (centuries to millennia) natural infiltration flux is 7 ± 3 mm/yr (Maher et al. 2003). The lower values of the recent estimates of infiltration and groundwater recharge by Routson and Johnson (1990) and Gee et al. (1992; 2007) are likely caused by the overall present-day increase in evapotranspiration, which has, in turn, caused a decrease in infiltration and groundwater recharge. A comparison of field estimates with the results of calculations performed in this paper is given in Section 4.2.

4.1.2. Input Parameters for Different Calculation Scenarios

For water-balance calculations, we used the temperature and precipitation time-series data representing a period of active infiltration from the surface (i.e., no freezing) from March through October for the years 1990–2007. These time-series data were used to determine a set of meteorological parameters summarized in Table 2, which were then used to develop the input PDFs and fuzzy numbers shown in Figure 4. Several modeling scenarios were developed (Table 3) to assess how the application of different models for input parameters affects the uncertainty of water-balance calculations. Scenario 0 was modeled using input PDFs by means of Monte Carlo simulations, using RiskAMP Monte Carlo Add-In Library version 2.10 for Excel. Scenarios 1 through 8 were simulated by means of the RAMAS Risk Calc code. Scenario 1 was simulated using input PDFs, and

the results are given as p-box numbers. Scenarios 2 through 6 were simulated applying both PDFs and fuzzy number inputs, using basic (corresponding to α=0) and an additional 10 hybrid levels (corresponding to α-cuts from 0.1 to 1). Scenarios 7 and 8 were simulated using only fuzzy numbers.

4.2. Results and Comparison with Field Data

The results of simulations are shown in Figures 5, 6, and 7. Figure 5 illustrates the calculated E_o, *ET*, and *I* values for all eight scenarios, including the mean values from Monte-Carlo simulations (Scenario 0), the mean value ranges from p-box calculations (Scenario 1) and hybrid calculations (Scenarios 2 through 6), and the core values, corresponding to α=1, from fuzzy calculations (Scenarios 7 and 8). This figure also demonstrates the interquartile ranges of the calculated distributions (with endpoints at the 25th and 75th percentiles). Figure 6 presents the results of calculations of the breadth of uncertainty, and Figure 7—a comparison of the p-boxes, hybrid numbers, and fuzzy numbers for calculated E_o, *ET*, and infiltration rates.

4.2.1. Potential Evapotranspiration

Figure 5a demonstrates that the E_o mean from Monte Carlo simulations is within the mean ranges from the p-box (Scenario 1) and fuzzy-probabilistic scenarios (Scenarios 2-6). It also corresponds to a midcore of the fuzzy scenario with trapezoidal FMFs (Scenario 7) and the core of the fuzzy scenario with triangular FMFs (Scenario 8). The range of means from the p-box and fuzzy-probabilistic calculations for α=1 is practically the same, indicating that including fuzziness within the input parameters does not change the range of most possible E_o values.

Figure 5a shows that the core uncertainty of the trapezoidal FMFs (Scenario 7) is the same as the uncertainty of means for fuzzy-probabilistic calculations for α=1. Obviously, the output uncertainty decreases for the input triangular FMFs (Scenario 8), because these FMFs resemble more tightly the PDFs used in other scenarios. Figure 5a also illustrates that a relatively narrow range of field estimates of E_o (from 1,400 to 1,611 mm/yr, which was reported for the Hanford site [Ward 2005]), is well within the calculated uncertainty of E_o values. One can see from Figures 5a, 5b, and 5c that the uncertainty ranges from p-box, hybrid, and fuzzy calculations significantly exceed those from Monte Carlo simulations.

Figure 6a demonstrates that while the range of means for α=1 remains practically the same, the overall breadth of E_o uncertainty increases for the basic hybrid scenario and decreases for the hybrid level 10 calculations, as the number of input fuzzy parameters increases from Scenario 2 to Scenarios 5 and 6. The overall increase in uncertainty is caused by the fact that (generally) a fuzzy number bounds a corresponding PDF. Because precipitation is not used in the Penman equation, the same E_o ranges are calculated for Scenarios 5 and 6. The breadth of E_o uncertainty decreases for fuzzy calculations in Scenarios 7 and 8. Figures 7a,b,c also illustrate that the breadth of uncertainty from fuzzy

probabilistic and fuzzy calculations of Eo are significantly larger than that from Monte Carlo calculation. Figures 7a,b demonstrate that the CDFs for E_o in the p-box and fuzzy-probabilistic calculations are practically of the same shape as that from Monte Carlo calculations.

4.2.2. Evapotranspiration

Figure 5b shows that the mean *ET* of ~184 mm/yr from Monte Carlo simulations (Scenario 0) is practically the same as the *ET* means for Scenarios 1 through 5 and the core value for Scenario 8. A greater *ET* uncertainty for Scenario 6 (precipitation is simulated using a fuzzy number) can be explained by the relatively large precipitation range for α=0—from 46 to 324 mm/yr. At the same time, the means of *ET* values for α=1 range within relatively narrow limits, as the precipitation for α=1 changes only from 157.2 to 212.8 mm/yr (see Table 2). The breadth of uncertainty of *ET* (Figure 6b) is practically the same for Scenarios 1 through 5, and increase for Scenarios 6, 7, and 8, in the account of calculations using a fuzzy precipitation.

For Scenarios 1 through 5, the *ETs* are merely probability distributions, which are close to that from Monte Carlo simulations (illustrated in Figure 7d for Scenario 1). When the fuzzy precipitation is included in Scenarios 6, the E_o uncertainty is largely characterized by fuzziness (note in Figure 7e that CDFs in the hybrid calculations of E_o for Scenario 6 are practically vertical lines). Figure 7f demonstrates that a possibility distribution from fuzzy calculations only slightly exceeds that from Monte Carlo simulations.

The calculated means for Scenarios 0, 1–5, and 8 exceed the field estimates of *ET* [160 mm/yr (Gee et al., 1992; 2007)] by 22 to 24 mm/yr. This difference can be explained by Gee et al.'s calculations being based on using a lower value of annual precipitation (160 mm/yr for the period prior to 1990), while our calculations are based on using a greater mean annual precipitation (185 mm/yr), averaged for the years from 1990 to 2007. The field-based data are within the *ET* uncertainty range for Scenarios 6 and 7, since the precipitation range is wider for these scenarios.

4.2.3. Infiltration

Figure 5c demonstrates that the mean infiltration of 1.18 mm/yr from Monte Carlo simulations (Scenario 0) is within the calculated mean uncertainty range for all other scenarios. The extended mean infiltration range for Scenario 6 can be explained by using a wide precipitation range for α=0 (which is also the case for the *ET* calculations). Note that the *y*-axis (infiltration) in Figure 5c is plotted using the log scale, because the infiltration distribution is lognormal, with a positive skewness and an asymmetric tail, extending toward larger values. The breadth of uncertainty of infiltration (Figure 6c) for a basic hybrid level is slightly increases for Scenarios 1 through 5, and significantly increases for Scenario 6, in the account of calculations using a fuzzy precipitation. Figures 7g,h, and i, showing a comparison of CDFs and possibility distributions for different scenarios, indicate that both statistical and fuzzy components are involved in the calculated infiltration uncertainty. Calculated infiltration for all scenarios ranges within the recharge-rate estimates given in

publications by Routson and Johnson (1990) and Gee et al. (1992; 2007), as well as in independent estimates (7±3 mm/yr) by Maher et al. (2003).

5. Conclusions

The motivation of this paper is to demonstrate the potential of the application of a fuzzy-probabilistic approach for water-balance calculations, given the general inadequacy of meteorological information and field data collected at regional scale. Depending on the information available, some input parameters can be given using probability distributions, and some parameters can be presented using fuzzy numbers when the information for constructing reliable PDFs is limited or vague. The fuzzy-probabilistic approach to modeling and uncertainty quantification of hydrogeological systems is based on a combination of statistical and fuzzy-calculus calculations. This paper provides the theory for, and a case study as an application of, the fuzzy-probabilistic approach, combining probability and possibility theory for assessing the uncertainty in the components of a water-balance equation—potential evapotranspiration, evapotranspiration, and net infiltration.

Simulation results for the Hanford site indicate that the E_o, ET, and infiltration uncertainties increase with the increase in a number of fuzzy inputs, because the uncertainty of fuzzy numbers is generally greater than that of corresponding PDFs (as the input fuzzy numbers bound corresponding PDFs). Application of fuzzy-probabilistic calculations provides more conservative estimates of uncertainty, i.e., wider uncertainty ranges compared to conventional Monte Carlo simulations. The results of field observations of E_o, ET, and infiltration at the Hanford site are within the calculated uncertainty ranges, which also suggest the possibility of increasing the ranges of these parameters over longer time periods.

The evaluation of water-balance uncertainty would be essential for predicting and controlling the deep-percolation rate through the unsaturated zone, groundwater recharge, contaminant transport in the unsaturated-saturated zone, as well as planning and assessing the risk of nuclear waste disposal in geologic media and remediation activities at contaminated and agricultural areas.

Acknowledgement: Critical and constructive comments of two anonymous reviewers, Stefan Finsterle of LBNL and Ming Ye of Florida State University are very much appreciated. Scott Ferson of Applied Biomathematics, the developer of the RAMAS Risk Calc code, provided a number of publications related to this software. This work was partially supported by the LBNL Laboratory Directed Research and Development Program, the Nuclear Energy Advanced Modeling and Simulation (NEAMS) program within the DOE Advanced Fuel Cycle Initiative, and the Director, Office of Science, Office of Biological and Environmental Remediation Sciences of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 to Lawrence Berkeley National Laboratory.

References

Arora VK (2002). The use of the aridity index to assess climate change effect on annual runoff, J. of Hydrology, 265: 164–177.

Batchelor CH (1984). The accuracy of evapotranspiration estimated with the FAO modified penman equation, Irrigation Science, 5(4): 223-233.

Burmaster DE, Lloyd KJ, and Thompson KM (1995). The need for new methods to backcalculate soil cleanup targets in interval and probabilistic cancer risk assessments. Human and Ecologial Risk Assessment 1:89–100.

Chang N-B (2005). Sustainable water resources management under uncertainty, Stochast Environ Res and Risk Assess, 19: 97–98.

Cooper JA, Ferson S, Ginzburg L (2006). Hybrid processing of stochastic and subjective uncertainty data, Risk Analysis, 16(6): 785–791.

Dempster AP (1967). Upper and lower probabilities induced by a multivalued mapping." The Annals of Statistics 28: 325-339.

DOE (1996). Final Environmental Impact Statement for the Tank Waste Remediation System, Hanford Site, Richland, Washington, DOE/EIS-0189. See also http://www.globalsecurity.org/wmd/library/report/enviro/eis-0189/app_i_3.htm

Dubois D, Prade H (1981). Additions of interactive fuzzy numbers, IEEE Transactions on Automatic Control 26: 926-936.

Dubois D, Prade H (1994). Possibility theory and data fusion in poorly informed environments. Control Engineering Practice 2(5): 811-823.

Ferson S, Ginzburg L (1995) Hybrid arithmetic. Proceedings of the 1995 Joint ISUMA/ NAFIPS Conference, IEEE Computer Society Press, Los Alamitos, California, 619-623.

Ferson S, Kreinovich V, Ginzburg L, Myers DS, Sentz K (2003). Constructing probability boxes and Dempster-Shafer structures, SAND REPORT, SAND2002-4015.

Ferson S (2002). RAMAS Risk Calc 4.0 Software: Risk assessment with uncertain numbers, CRC Press.

Gee GW, Oostrom M., Freshley MD, Rockhold ML, Zachara JM (2007). Hanford site vadose zone studies: An overview, Vadose Zone Journal 6: 899-905.

Gee GW, Fayer MJ, Rockhold ML, Campbell MD (1992). Variations in recharge at the Hanford Site. Northwest Sci. 66: 237–250.

Guyonne D, Dubois D, Bourgine B, Fargier H, Côme B., Chilès J-P (2003). Hybrid method for addressing uncertainty in risk assessments. Journal of Environmental Engineering 129: 68-78.

Kaufmann A, Gupta MM (1985). Introduction to Fuzzy Arithmetic, New York: Van Nostrand Reinhold.

Maher K, DePaolo DJ, Conrad MS, Serne RJ (2003). Vadose zone infiltration rate at Hanford, Washington, inferred from Sr isotope measurements. Water Resources Research 39(8):1204-1217.

Meyer PD, Rockhold ML, Gee GW (1997) Uncertainty Analysis of infiltration and subsurface flow and transport for SDMP sites, NUREG/CR-6565, US Nuclear Regulatory Commission.

Möller B., Beer M. (2005) Fuzzy Randomness. Uncertainty in Civil Engineering and Computational Mechanics, Computational Mechanics 36(1).

Neitzel DA (1996) Hanford Site National Environmental Policy Act (NEPA) Characterization. PNL-6415, Rev. 8. Pacific Northwest National Laboratory. Richland, Washington.

Neuman SP (2003). Maximum likelihood Bayesian averaging of alternative conceptual-mathematical models. Stochast Environ Res and Risk Assessment 17(5):291–305.

Neuman SP, Wierenga PJ (2003). A comprehensive strategy of hydrogeologic modeling and uncertainty analysis for nuclear facilities and sites, NUREG/CR-6805. US Nuclear Regulatory Commission, Washington, DC.

Orr S, Meystel AM (2005). Approaches to optimal aquifer management and intelligent control in a multiresolutional decision support system, Hydrogeol J 13: 223–246

Penman HL (1963). *Vegetation and hydrology*. Tech. Comm. No. 53, Commonwealth Bureau of Soils, Harpenden, England. 125 pp.

Routson RC and Johnson VG (1990). Recharge estimates for the Hanford Site 200 Areas Plateau, Northwest Science, 64(3): 150-158.

Sankarasubramanian A. and Vogel RM (2003). Hydroclimatology of the continental United States, *Geophysical Res. Letters*, 30(7): 1363.

Shafer G (1976). *A mathematical theory of evidence*. Princeton University Press, Princeton, New Jersey, USA.

Smets P (1990). The combination of evidence in the transferable belief model. *IEEE Pattern Analysis and Machine Intelligence*, 12: 447-458.

Winter CL (2004). Stochastic hydrology: practical alternatives exist, Stochast Environ Res and Risk Assessment 18: 271 – 273.

Wagener T, Gupta HV,(2005). Model identification for hydrological forecasting under uncertainty Stoch Environ Res and Risk Assessement (2005) 19: 378–387

Ward AL, Freeman EJ, White MD, Zhang ZF (2005). STOMP: Subsurface Transport Over Multiple Phases, Version 1.0, Addendum: Sparse Vegetation Evapotranspiration Model for the Water-Air-Energy Operational Mode, PNNL-15465.

Yager R and Kelman A (1996). Fusion of fuzzy information with considerations for compatibility, partial aggregation, and reinforcement. International Journal of Approximate Reasoning, 15(2): 93-122.

Ye M., Neuman SP, Meyer PD (2004). Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff. Water Resour. Res. 40(5): W05113.

Ye M, Neuman SP, Meyer PD, Pohlmann K (2005). Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff. Water Resour. Res. 41:W12429.

Zadeh L (1978). Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1:3-28.

Zadeh LA (1986). A Simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. The AI Magazine 7: 85-90.

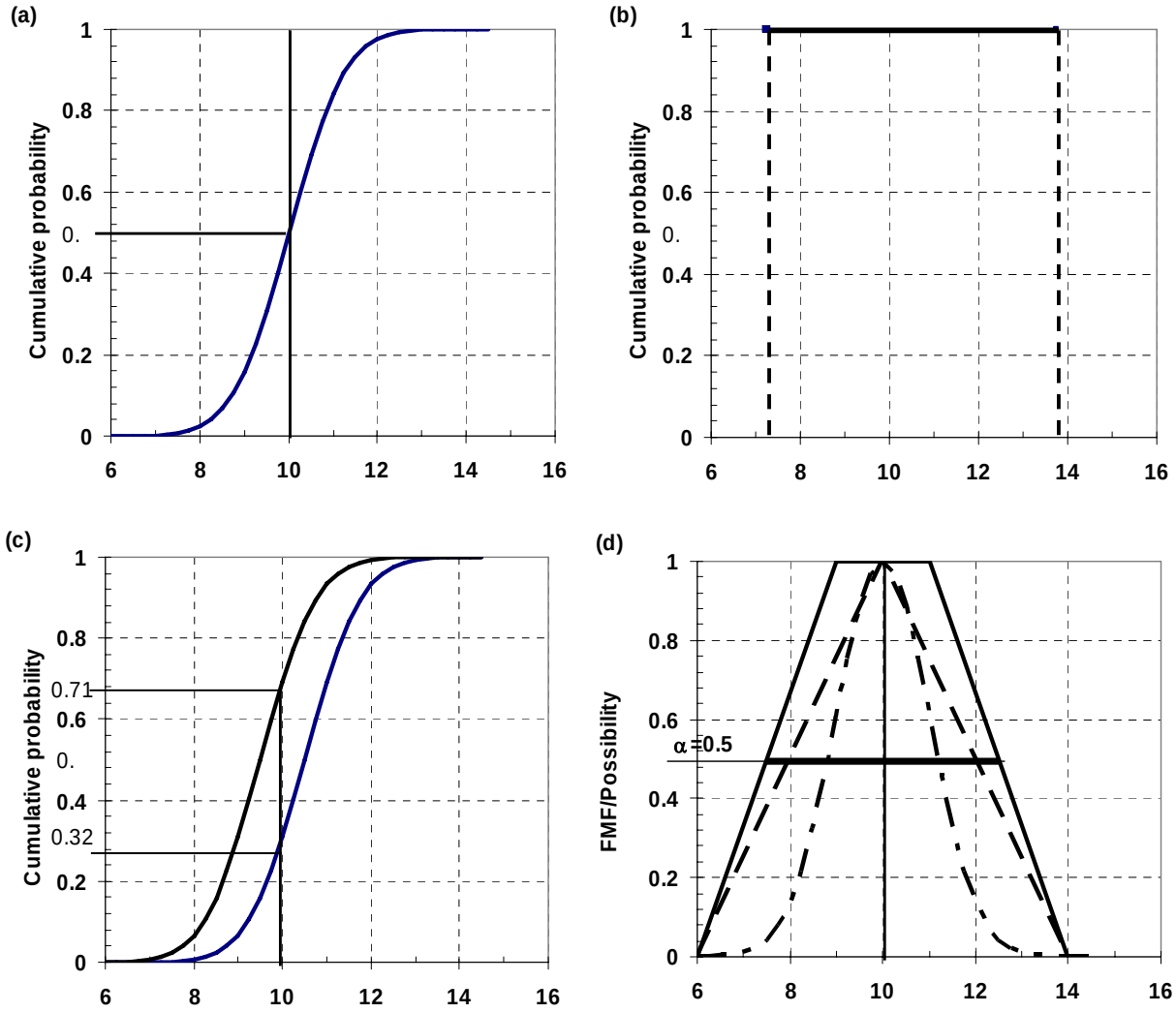

Figure 1. Graphical illustration of uncertain numbers: (a) Cumulative normal distribution function, with mean M=10 and standard deviation σ=1, (b) interval number, corresponding to the interval from 7.25 to 13.75, (c) p-box—left bound with M=9.5 and σ=0.9, and right bound with M=10.5 and σ=1.1, and (d) fuzzy triangular (dashed line) number, with FMF=1 corresponding to M=10, and trapezoidal (solid line) number, with FMF=1 corresponding to the interval M=9-11, and a Gaussian FMF function with M=10, and σ=1. Note that the Gaussian FMF is obtained by normalization of the Gaussian probability density function to the largest value of the probability density. Figure (d) also shows an α-cut (thick horizontal line) through the trapezoidal fuzzy number (for α=0.5).

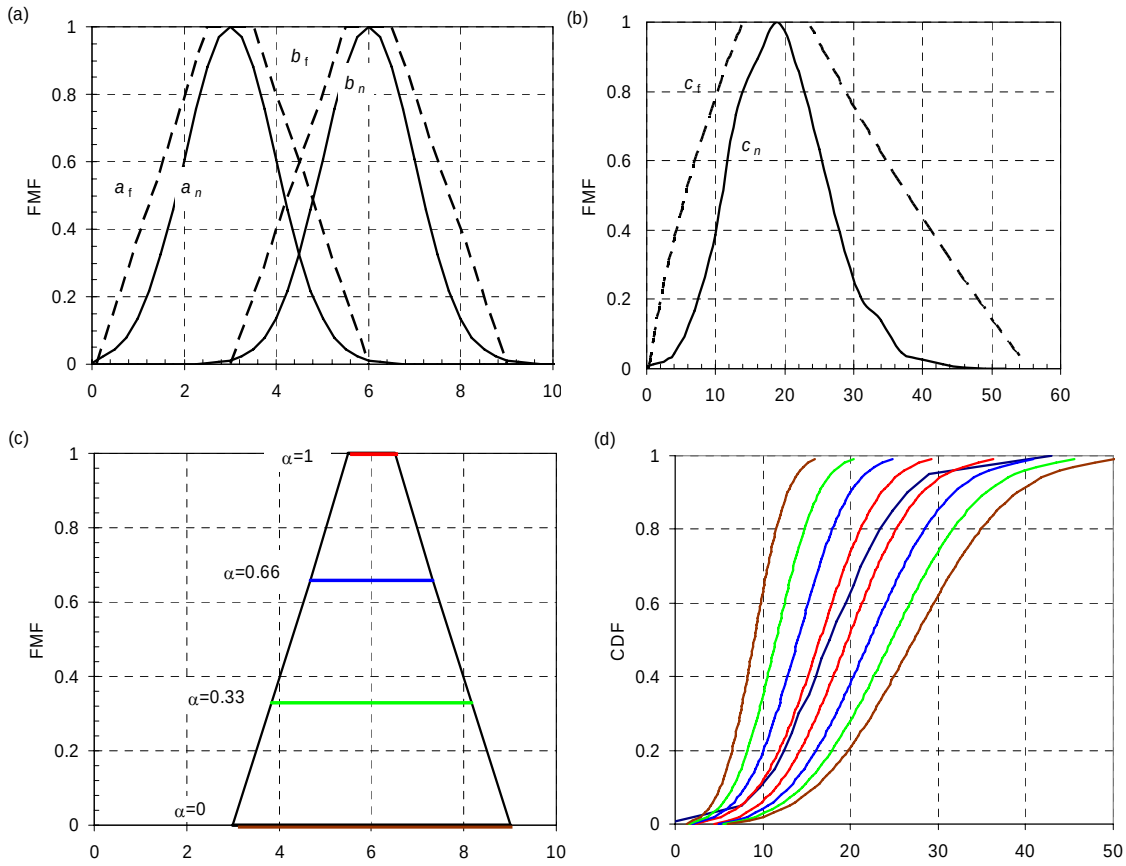

Figure 2. Illustration of the concepts of fuzzy interval-arithmetic and hybrid calculations:

(a) fuzzy numbers a_f=[0.1, 2.5, 3.5, 6] and b_f =[3, 5.5, 6.5, 9], shown by dashed lines, and normal probability density functions a_n=[mean=3, standard deviation=1], and b_n=[mean=6, and standard deviation=1], shown by solid lines;

(b) results of multiplication of fuzzy numbers $c_f=a_f \cdot b_f$ are shown by a dashed line, and Monte Carlo multiplications $c_n=a_n \cdot b_n$ are shown by a solid line;

(c) fuzzy number b_f used for hybrid calculations, showing the intervals α=0, 0.33, 0.66, and 1; and

(d) results of hybrid multiplication $c_h=a_n \cdot b_{f,}$ which are shown as a family of p-boxes: bounding black lines correspond to a multiplication of a_n by an interval for α=0, green lines—a_n by an interval for α=0.33, blue lines—a_n by an interval for α=0.66, and red lines—a_n by an interval for α=1; the middle black line is c_n from Monte Carlo simulations, which is shown for comparison.

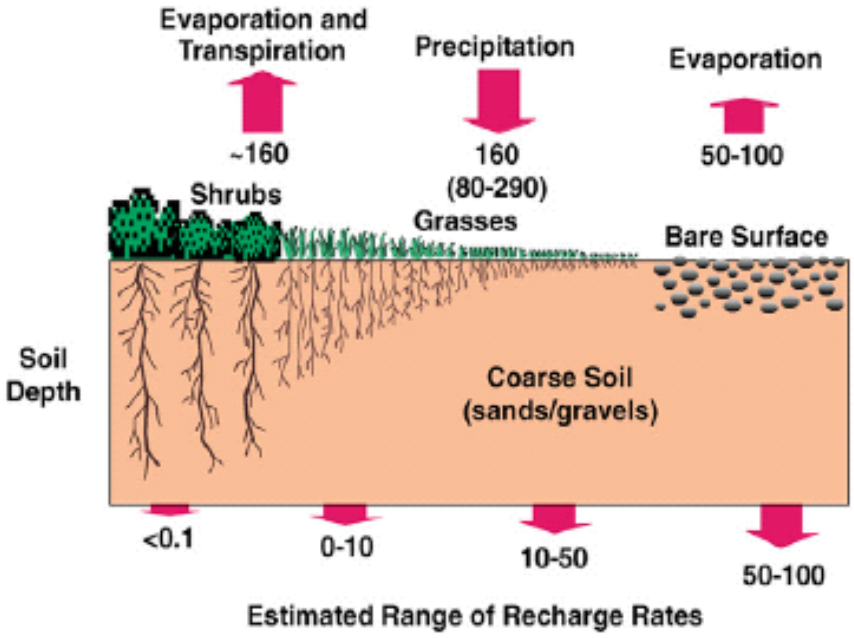

Figure 3. Estimated *ET* and recharge/infiltration at the Hanford site (Gee et al., 2007).

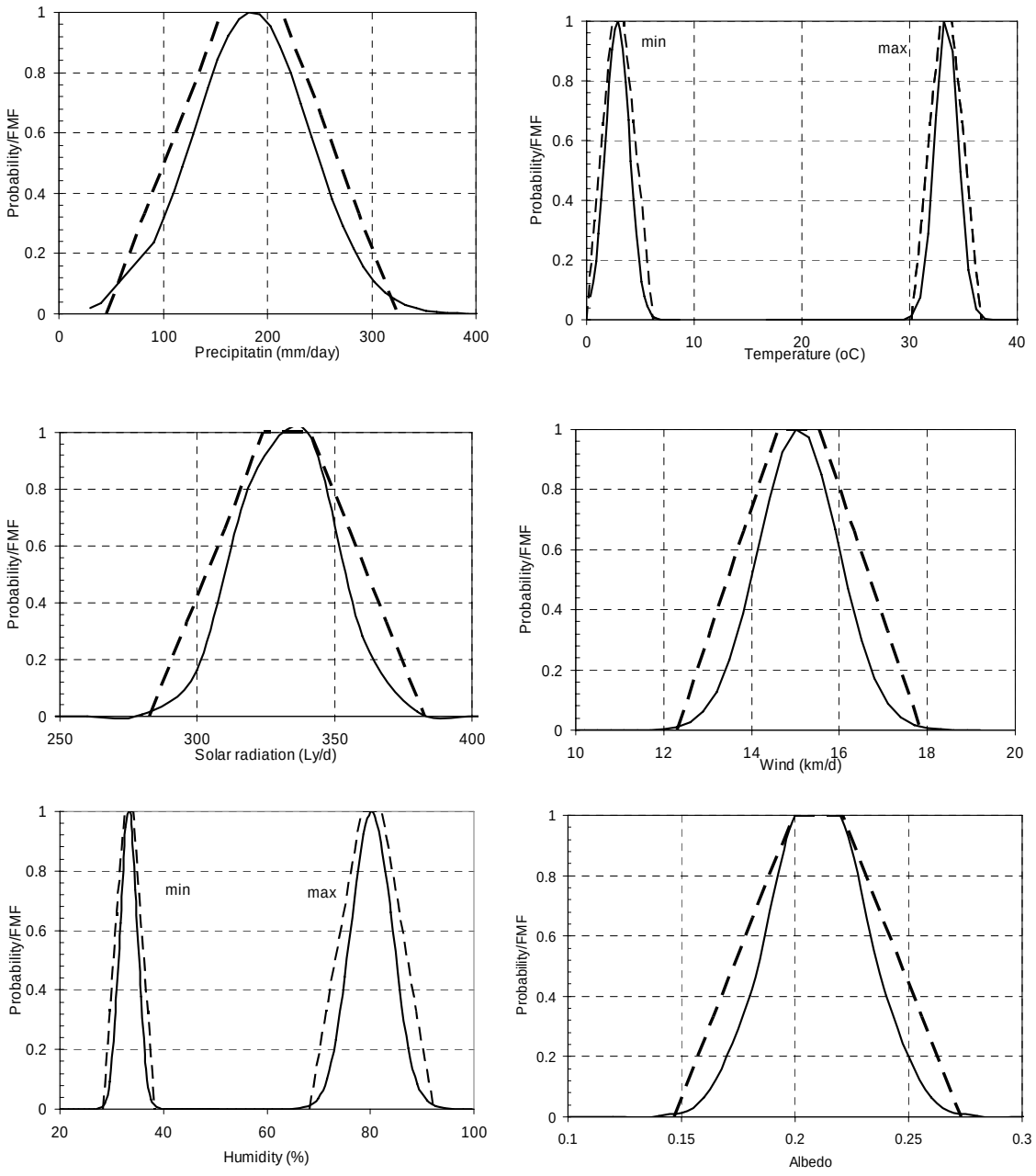

Figure 4. Input PDFs (solid lines) and fuzzy numbers (dashed lines) used for calculations.

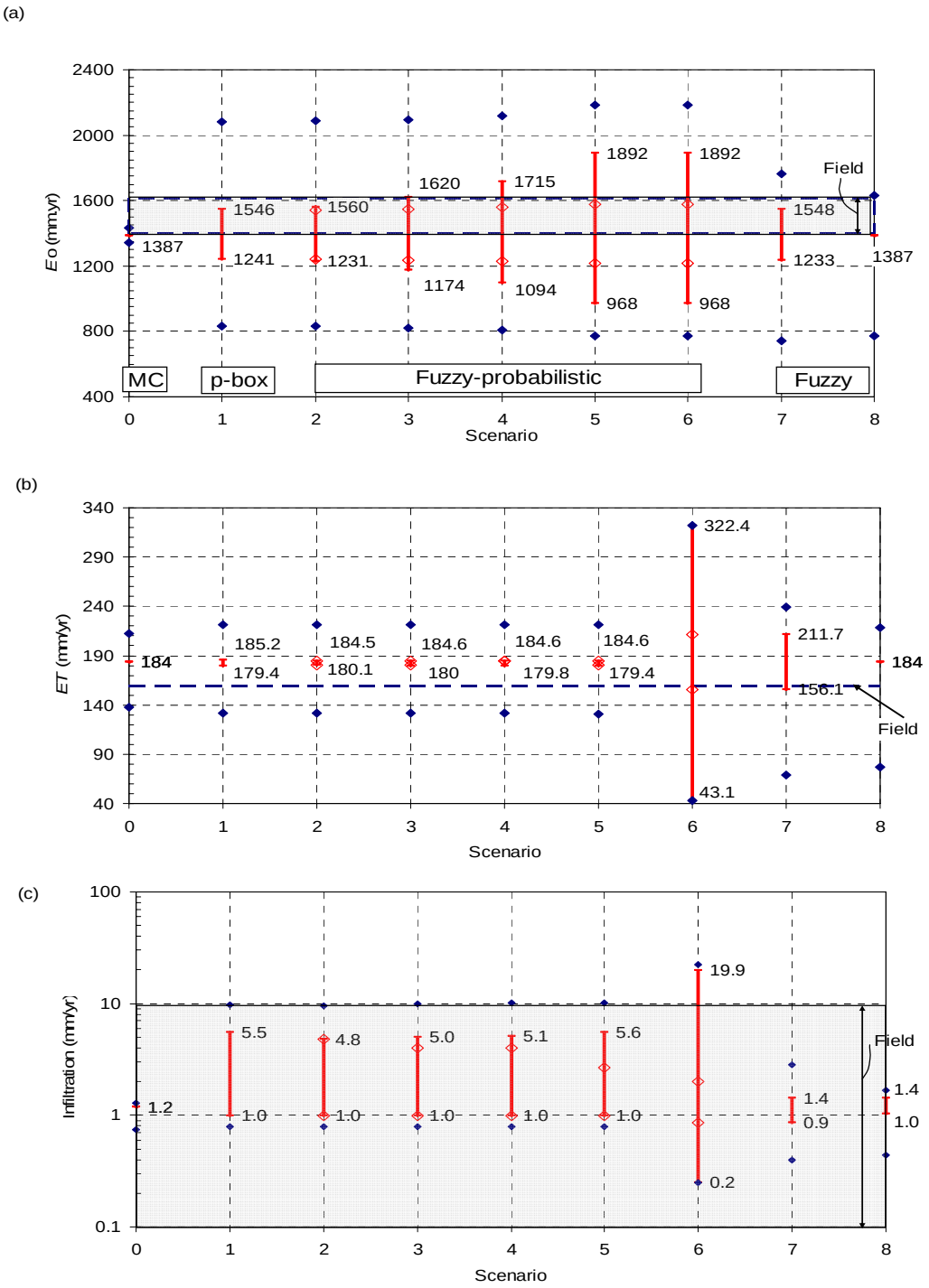

Figure 5. Results of calculations of E_o (a), *ET* (b), and infiltration (c) for all scenarios. Red lines are the mean intervals (Scenarios 1-6) and core intervals (Scenarios 7 and 8), the blue diamonds indicate the interquartile ranges with endpoints at the 25th and 75th percentiles of the underlying distribution. Red diamonds for Scenarios 2-6 indicate the mean intervals for the hybrid level=10. The height of shaded areas in figures *a* and *b* indicate the range of E_o and infiltration from field measurements.

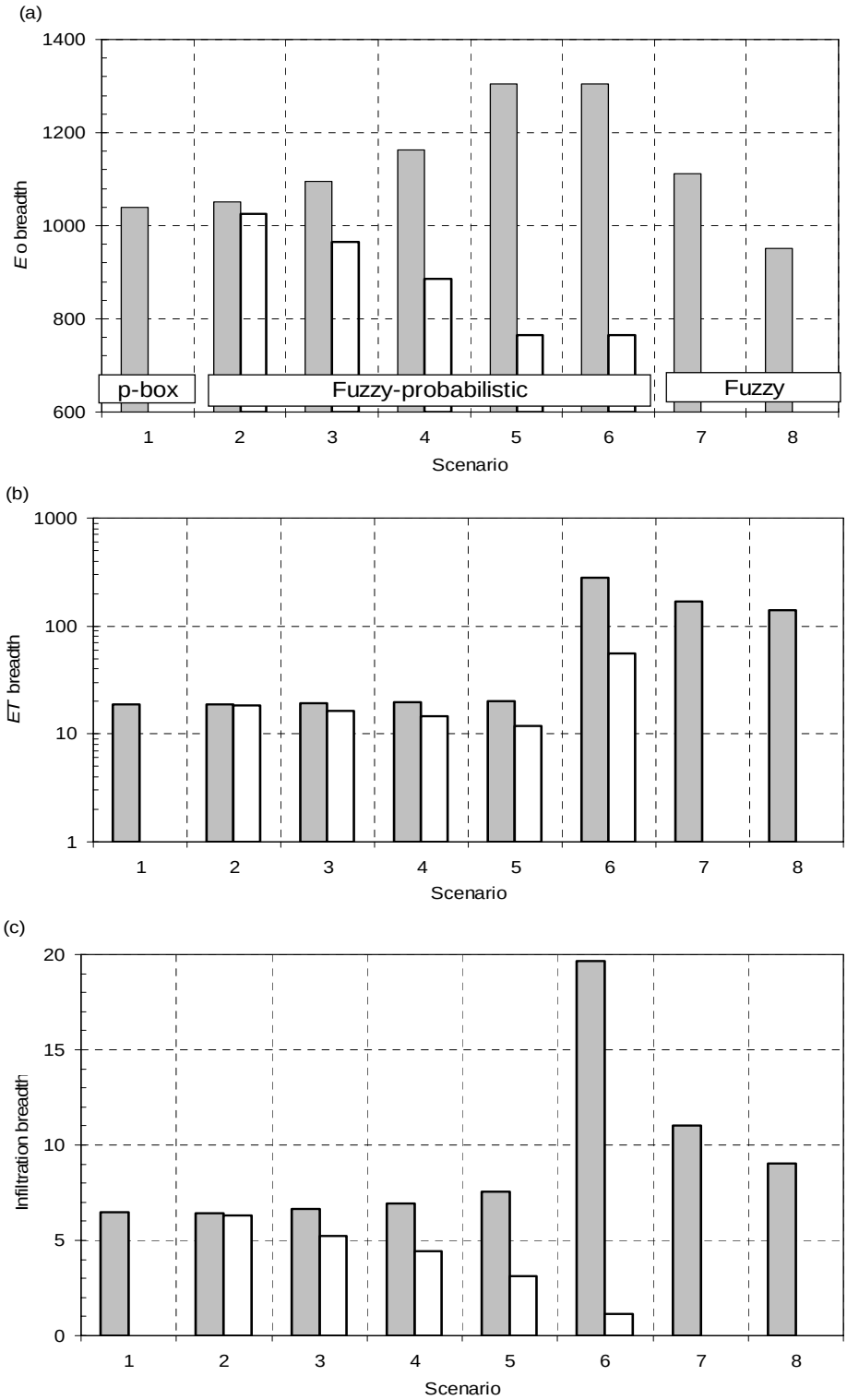

Figure 6. Breadth of uncertainty of E_o, *ET*, and infiltration. For Scenarios 2-6, grey bars are for the basic hybrid level (indicating the maximum uncertainty), and open bars are for the hybrid level=10 (indicating the minimum uncertainty).

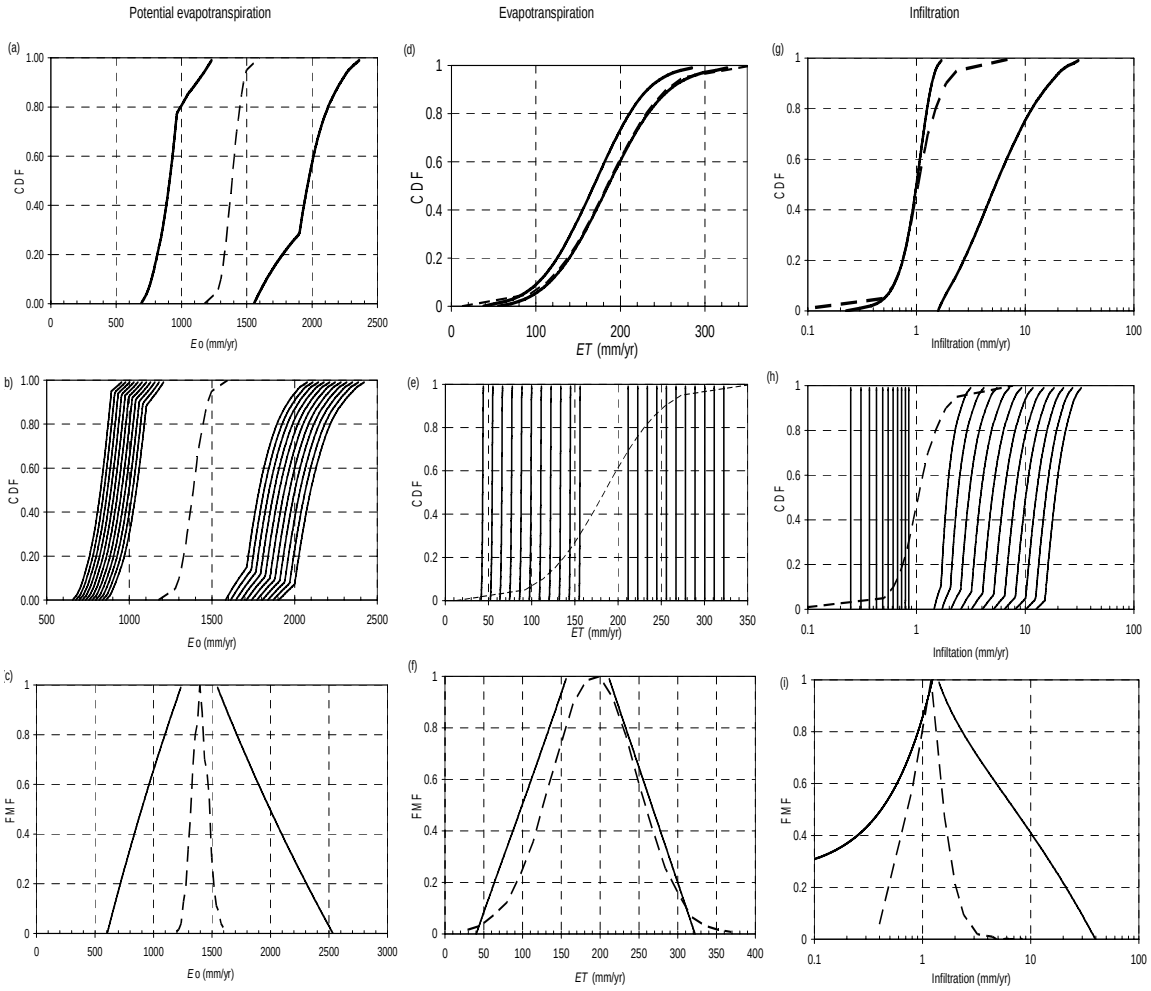

Figure 7. Comparison of the p-boxes, hybrid numbers, and fuzzy numbers for calculated E_o (left column—figures a, b, and c), *ET* (middle column—figures d, e, and f), and infiltration (right column—figures g, h, and i) with corresponding Monte Carlo simulations (shown by dashed lines). The upper row presents the results of p-box calculations (Scenario 1), the middle row—hybrid calculations (Scenario 6), and the lower row—fuzzy calculations (Scenario 7). Note the results of p-box and hybrid calculations are given using CDFs, and fuzzy calculations are given using fuzzy numbers and compared with normalized probability distributions.

Table 1. Definitions for the main characteristics of fuzzy numbers and p-boxes representation of uncertain numbers (calculated using Risk Calc code)

Characteristic	Definitions
core	The most possible value(s) of the uncertain number *n*, i.e., value(s) with a possibility of one, or for which the probability can be any value between zero and one.
mean	An interval between the means of the lower (left) and upper (right) bounds of the uncertain number *n*.
iqrange	An interval guaranteed to enclose the interquartile range (with endpoints are the 25th and 75th percentiles) of the underlying distribution.
breadth of uncertainty	For fuzzy numbers, it is given by the area under the membership function; and for *p*-boxes, it is given by the area between the upper and lower bounds. The uncertainty decreases as the breadth of uncertainty decreases.

Table 2. Meteorological parameters used for calculations

Type of data	Parameters		Wind speed (km/hr)	Relative humidity (%)		Albedo	Solar radiation Ly/day[1)]	Annual precipi-tation (mm/yr)	Temperature (^{o}C)	
				Max	Min				Max	Min
PDFs	**Mean**		15.07	80.2	33.3	0.21	332.55	185	33.41	2.87
	Standard Deviation		0.92	4.01	1.66	0.021	16.63	55.62	1.08	1.11
Trapezoidal FMFs	**α=0**	**Min**	12.31	68.17	28.29	0.15	282.66	46.0	30.17	0.0
		Max	17.84	92.23	38.31	0.27	382.44	324.1	36.65	6.17
	α=1	**Min**	14.61	78.2	32.47	0.22	324.24	157.2	32.87	2.32
		Max	15.53	82.2	34.14	0.27	382.44	212.8	33.95	3.42

Note: [1)] The abbreviation Ly stands for a Langley, and 1 Ly=41,8440 joules per square meter or 1 cal/cm^2.

Table 3. Scenarios of input and output parameters used for water-balance calculations

Scenarios	**Input parameters**						**Output parameters**
	Wind speed	**Humidity**	**Albedo**	**Solar radiation**	**Precipitation**	**Temperature**	
0	PDF	PDF	PDF	PDF	PDF	PDF	PDF
1	PDF	PDF	PDF	PDF	PDF	PDF	p-box
2	Fuzzy	PDF	PDF	PDF	PDF	PDF	Hybrid
3	Fuzzy	Fuzzy	PDF	PDF	PDF	PDF	Hybrid
4	Fuzzy	Fuzzy	Fuzzy	PDF	PDF	PDF	Hybrid
5	Fuzzy	Fuzzy	Fuzzy	Fuzzy	PDF	PDF	Hybrid
6	Fuzzy	Fuzzy	Fuzzy	Fuzzy	Fuzzy	PDF	Hybrid
7[1)]	Fuzzy	Fuzzy	Fuzzy	Fuzzy	Fuzzy	Fuzzy	Fuzzy
8[2)]	Fuzzy	Fuzzy	Fuzzy	Fuzzy	Fuzzy	Fuzzy	Fuzzy

Notes: [1)] In Scenario 7, all FMFs are trapezoidal. [2)] In Scenario 8, all FMFs are triangular: the mean values of parameters, which are given in Table 2, are used for $\alpha=1$; and the minimum and maximum values of parameters, given in Table 2 for trapezoidal FMFs (Scenario 7), are also used for $\alpha=0$ of triangular FMFs in Scenario 8.