Proceedings of the ASME 2021
International Design Engineering Technical Conferences &
Computers and Information in Engineering Conference
IDETC-CIE 2021
AUGUST 17-20, 2021, Virtual, Online

IDETC2021-67944

DIFFERENTIAL UTILITY: ACCOUNTING FOR CORRELATION IN PERFORMANCE AMONG DESIGN ALTERNATIVES

Sahar Jolini[*]
Department of Systems Engineering
and Operations Research
George Mason University
Fairfax, Virginia 22030
sjolini@gmu.edu

George A. Hazelrigg[†]
Department of Mechanical Engineering
George Mason University
Fairfax, Virginia 22030
ghazelri@gmu.edu

ABSTRACT

Recognizing expected utility as a valid design criterion, there are cases where uncertainty is such that this criterion fails to distinguish clearly between design alternatives. These cases may be characterized by broad and significantly overlapping utility probability distributions. Not uncommonly in such cases, the utility distributions of the alternatives may be highly correlated as the result of some uncertain variables being shared by the alternatives, because modeling assumptions may be the same across alternatives, or because difference information may be obtained by means of an independent source. Because expected utility is evaluated for alternatives independently, maximization of expected utility typically fails to take these correlations into account, thus failing to make use of all available design information. Correlation in expected utility across design alternatives can be taken into account only by computing the expected utility difference, namely the "differential expected utility," between pairs of design alternatives. However, performing this calculation can present significant difficulties of which excessive computing times may be key. This paper outlines the mathematics of differential utility and presents an example case, showing how a few simplifying assumptions enabled the computations to be completed with approximately 24 hours of desktop computing time. The use of differential utility in design decision making can, in some cases, provide significant additional clarity, assuring better design choices.

NOMENCLATURE

Ω	A possible outcome of a decision
C	Cost
COP	Heat pump coefficient of performance
DD	Degree-days
H	A state-of-information
J	A performance measure, an objective function
P	Probability
R	Insulation rating
T	Temperature
THL	Total heat loss
i	Annual inflation rate
q	Rate of heat exchange
r	Discount rate
t	Time
u	Utility, risk-adjusted value
$E\{u\}$	Expected utility
$<$	Is less than, $x < y$ reads x is less than y
$>$	Is greater than, $x > y$ reads x is greater than y
$\geq$	Is greater than or equal to, $x \geq y$ reads x is greater than or equal to y
$\succ$	Is preferred to, $A \succ B$ reads A is preferred to B

[*]PhD student.
[†]Research faculty.

$\sim$ Is indifferent to, $A \sim B$ reads A is indifferent to B

$\succsim$ Is preferred or indifferent to, $A \succsim B$ reads A is preferred or indifferent to B

$|$ Subject to the condition that, $P\{A|B\}$ reads the probability of A given the value of B

1 Introduction

Consider the following problem. An engineer is asked to design a home heating and air conditioning system. There are two fundamentally different systems from which the engineer may choose: gas heat combined with an air conditioner, and an electric heat pump system. The design criterion is minimum net present value (NPV) of total system cost over a period of 10 years. This design criterion can be reformulated as a maximization of expected utility, accounting for uncertainties including such variables as initial cost, maintenance costs, future energy costs, weather and inflation rates. Relatively good data are available for initial cost and maintenance costs, but future energy costs, weather and inflation rates are quite uncertain. Indeed, these uncertainties dominate the analysis, resulting in rather broad and overlapping distributions of NPV of system costs. Thus, although one system will have the highest expected utility, the engineer will not have a great deal of confidence that the highest expected utility system will indeed be the lowest cost system.

It is interesting to note, however, that while future energy costs, weather and inflation rates are quite uncertain, they are highly correlated across the system alternatives. Obviously, the prices of electricity and gas are strongly correlated, while the weather and inflation rates are actually the same regardless of which system is chosen. These correlations in "exogenous" variables comprise information that is not taken into account by simply choosing the alternative with the highest expected utility.

In the 1970's, sharp increases in the price of oil provided a strong incentive to both the commercial airlines and the U.S. Air Force to reduce aircraft separation on oceanic routes, particularly in the heavily traveled North Atlantic track system. On the one hand, reduction of aircraft separation would significantly reduce fuel consumption while, on the other hand, it would increase the risk of a collision [1]. Despite the fact that there had never been a collision between commercial aircraft on an oceanic route, the concern for collision risk was a paramount consideration. An acceptable level of risk was specified in the form of a Target Level of Safety noting, "the risk of collision due to loss of lateral separation [shall] not exceed 0.2 fatal aircraft accidents in 10^7 flying hours" [2]. Even at current traffic levels, this translates to approximately one collision every 20 years. Clearly, actuarial data could not be used as a basis for separation standards decisions. Thus, decisions would have to be made on the basis of a mathematical model. In the 1960's, a collision risk model was developed by the Royal Aircraft Establishment under the direction of Reich [3–5]. The model incorporated conservative assumptions including, aircraft comprise rectangular cuboids that enclose the entire airframe, and pilots do not see and avoid other aircraft. The navigation errors of aircraft using the North Atlantic track system were estimated based on a sample of roughly 120,000 radar traces obtained on aircraft as they departed the track system, which were parameterized as probability distributions of various types to accommodate the relatively sparse data for extreme navigational deviations. Overall, it was unlikely that the model-estimated collision risk was accurate to an order of magnitude.

Despite the considerable lack of confidence in the accuracy of the Reich model, the model errors across track system design alternatives were very highly correlated, thus enabling confidence in a statement that one design poses less collision risk than another.

The above examples illustrate two cases where there is considerable uncertainty in the estimation of expected utility, one because of uncertainty in exogenous variables such as the weather, and the other because of model assumptions. Nonetheless, correlations between system alternatives enable clear performance distinctions among the system alternatives. In these cases reliance on expected utility decision making would typically neglect the correlation among alternatives which, when taken into account, could significantly improve confidence in system design decision making. This paper addresses this loss of information in the decision making process and offers an approach to its inclusion.

We first address the role of information in engineering decision making, then derive the mathematics for inclusion of correlation among design alternatives in the analysis of these alternatives. Finally, we illustrate the mathematics with an example problem.

2 Information

Decision theory links the concept of *information* to decisions. Roughly, information is the basis upon which a decision is made. However, despite this notion, there are subtle differences in the definitions of *information*. Howard and Abbas [6] note that, "The linking of what we can do and what we want to do is provided by what we know, also known as our *information*." More precisely, they define the concept of a clairvoyant, who is an entity capable of predicting with both precision and certainty the outcome of any measurably observable event such as a coin flip or the outcome of an engineering decision. With this notion, the clairvoyant can provide a decision-maker with *perfect information*. Thus, by this definition, the decision-maker has perfect information when she can predict with precision and certainty the outcomes of all alternatives available in a particular decision situation.

Hazelrigg [7], provides a definition that is slightly different. His definition is that a state of perfect information exists when

the decision-maker can be certain that her preferred choice is the alternative selected from the set of available alternatives that will result in the most preferred outcome. With this definition, Hazelrigg defines a state-of-information as the probability that the decision-maker's preferred choice will indeed result in the most preferred outcome given the set of available alternatives. Computation of the state-of-information proceeds as follows.

We can graph a typical decision as shown in Fig. 1. The

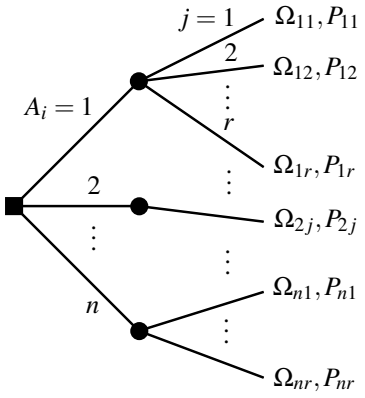

FIGURE 1. A typical decision.

available alternatives may be denoted A_i. Each alternative, A_i, provides a range of possible outcomes, Ω_{ij}, with each outcome occurring with probability P_{ij}. Associated with each outcome is a real scalar measure referred to as *utility*, u_{ij}, such that if

$$\Omega_{ij} \succ \Omega_{IJ} \tag{1}$$

then

$$u_{ij} > u_{IJ} \tag{2}$$

Further, if

$$\Omega_{ij} \sim \Omega_{IJ} \tag{3}$$

then

$$u_{ij} = u_{IJ} \tag{4}$$

Now define the variable

$$\delta_{ij}^{IJ} = \begin{cases} 0 \text{ if } u_{ij} < u_{IJ} \\ 1 \text{ if } u_{ij} \geq u_{IJ} \end{cases} \tag{5}$$

Thus, δ_{ij}^{IJ} is 0 if outcome IJ is preferred to outcome ij, otherwise it is 1. Given this definition of δ_{ij}^{IJ}, the probability that alternative A_i will result in an outcome that is preferred or indifferent to alternative A_I is given by

$$P_i^I = \sum_j P_{ij} \sum_J \delta_{ij}^{IJ} P_{IJ} \tag{6}$$

Let the preferred alternative be A_ι. We now determine the state of information as the probability that the following statement is true,

> *Alternative A_ι will produce an outcome that is preferred or indifferent to any other alternative.*

This probability is

$$H_\iota = \prod_{I \neq \iota} P_\iota^I \tag{7}$$

This definition of the state of information is such that H_ι is unity if the decision maker is certain that alternative A_ι will provide an outcome that is preferred or indifferent to the outcomes that would result from the other available alternatives. The value of H_ι decreases as the state of information is degraded. If the best choice from among n alternatives is completely random, one could expect $H_\iota = 1/n$. For example, this would be the case in calling the flip of a fair coin. It is important to recognize, however, that the case of $H_\iota = 1$ does not imply zero uncertainty. This is a key difference between the definitions of Howard and Hazelrigg.

Given this definition and invoking the preference [8], "I want the best system I can get," it is clear that improving the state-of-information can lead to better design decisions provided that the cost of the improved information is less than its value. This is the motivation for this work.

3 Mathematics

Consider a case where we seek to compare two engineering alternatives, A and B, against a scalar performance criterion J, where $J(A) = f(\mathbf{x}, \mathbf{z})$ and $J(B) = g(\mathbf{y}, \mathbf{z})$. The variables $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ may be uncertain and the functions f and g may be both approximate and uncertain. The variables $\mathbf{z}$ are common to both alternatives. Thus, the performance of alternatives A and B are correlated and, if we were to evaluate A and B independently, for example, using a Monte Carlo method, this correlation would not be a factor in the determination of $J(A)$ and $J(B)$.

In order to take the correlation between $J(A)$ and $J(B)$ into account, we create the difference model, $J(A) - J(B) = h(\mathbf{x}, \mathbf{y}, \mathbf{z})$. This difference model provides an additional condition on the determination of $J(A)$ and $J(B)$, which enables us to better estimate these performances. However, this is not the end of the problem. Simply because the expected performance $\mathrm{E}\{J(A)\}$ may

be greater than the expected performance $\mathrm{E}\{J(B)\}$, we cannot conclude that A is the better alternative. For example, A may allow the possibility of a low probability but highly undesirable consequence, thus making the lower performing but less risky alternative B more desirable. As a result, we must compare the expected utilities $\mathrm{E}\{u[J(A)]\}$ and $\mathrm{E}\{u[J(B)]\}$. In the case where utility equals performance, this evaluation is simple, as the differential utility, $\mathrm{E}\{u[J(A)]\} - \mathrm{E}\{u[J(B)]\}$, does not depend on the values $\mathrm{E}\{u[J(A)]\}$ or $\mathrm{E}\{u[J(B)]\}$. But when this is not the case, such as when the decision-maker is risk averse, then it is necessary to evaluate both $\mathrm{E}\{u[J(A)]\}$ and $\mathrm{E}\{u[J(B)]\}$ given the condition $J(A) - J(B) = h(\mathbf{x},\mathbf{y},\mathbf{z})$. Only this way can we evaluate the *differential utility*, $\Delta E\{u[J(A) - J(B)]\}$. This becomes a rather challenging problem in Monte Carlo analysis.

To begin, we shall invoke Bayes Theorem in an expanded form.

$$P\{J(A)\cdot J(B)|h(\mathbf{x},\mathbf{y},\mathbf{z})\} = \frac{P\{h(\mathbf{x},\mathbf{y},\mathbf{z})|J(A)\cdot J(B)\}P\{J(A)\cdot J(B)\}}{P\{h(\mathbf{x},\mathbf{y},\mathbf{z})\}} \tag{8}$$

where the notation $P\{J(A)\cdot J(B)\}$ means the probability that both $J(A)$ and $J(B)$ occur. Notice that the term $P\{h(\mathbf{x},\mathbf{y},\mathbf{z})|J(A)\cdot J(B)\}$ takes on the value 1 when for the specific values of $J(A)$ and $J(B)$ the condition $J(A) - J(B) = h(\mathbf{x},\mathbf{y},\mathbf{z})$ is met, and it is 0 otherwise. Thus, Eq. 8 reduces to

$$P\{J(A)\cdot J(B)|h(\mathbf{x},\mathbf{y},\mathbf{z})\} = \frac{P\{J(A)\cdot J(B)\}}{P\{h(\mathbf{x},\mathbf{y},\mathbf{z})\}} \tag{9}$$

Recognizing that $P\{J(A)\cdot J(B)\}$ is defined only when the condition on $J(A) - J(B) = h(\mathbf{x},\mathbf{y},\mathbf{z})$ is met. Then,

$$P\{J(A)\cdot J(B)\} = P\{J(A)\cdot J(B)|h(\mathbf{x},\mathbf{y},\mathbf{z})\}P\{h(\mathbf{x},\mathbf{y},\mathbf{z})\} \tag{10}$$

Obtaining analytic solutions to the terms of the right-hand side of Eq. 10 can be very difficult if possible. However, we can eliminate the need for analytic solutions if we solve this equation by means of a Monte Carlo simulation. Also note that the marginal probabilities, $P\{J(A)\}$ and $P\{J(B)\}$, may be found by summing the probabilities, for example, of $J(B)$ for a given value of $J(A)$,

$$P\{J(A)\} = \sum P\{J(B)|J(A)\} \tag{11}$$

In a Monte Carlo simulation, this becomes merely a sum of bin counts divided by the total bin count across all values of $J(A)$.

It is important to note that the equation for the difference, $J(A) - J(B)$, may be simply the difference in the models for $J(A)$ and $J(B)$, but may also derive from an entirely different condition, such as a measurement or estimate of this difference invoking an independent source. For example, suppose you wish to estimate the elevation and azimuth of two proximate stars using a telescope. You could center each star separately in the lens of the telescope and measure the elevation and azimuth of the telescope to obtain the elevation and azimuth for each star. In addition, however, you could measure the angular difference of the stars by observing them directly through the telescope. This difference measure would enable you to refine the independent measurements through the introduction of additional information.

4 Illustrative Example

The problem formulation and subsequent computations imposed by Eq. 10 can be a bit confusing. So, we shall offer a simple engineering example intended only as a guide. An engineer has been asked to recommend a heating system for a house with the goal of minimizing total system cost over a period of 10 years. Two alternative systems have been proposed, a gas-fired furnace and an electric heat pump system. To keep the analysis simple, we shall neglect the possible need for air conditioning and deal only with the heating requirements, but we will adjust the installation costs to account for the additional need of an air conditioner with the gas system. Our example will relate rather loosely to the case of a two-story, mid-size house with a living area of about 2,800 square feet, located in a typical northeastern area of the U.S.

To begin, we shall construct a rudimentary heat-loss model assuming that the house loses heat through its vertical walls and upper ceiling. We shall assume the house to have a rectangular shape of 30 by 48 feet, with a vertical elevation of 18 feet and an upper ceiling area of 1440 square feet. This provides a total heat-loss area, A, of 4248 square feet. We shall further assume an average R-factor for the heat-loss surfaces of 13 feet^2-hour/BTU. This leads to an hourly heat-loss rate of

$$q_L = \frac{A(60-T)}{R} = \frac{4248(60-T)}{13} = 326.8(60-T) \tag{12}$$

Where q_L is the rate of heat loss in BTU/hour, and T is the outdoor temperature, taking 60°F to be the "base" outdoor temperature above which the heat loss is nil. Thus, assuming that heat is required only when the outside temperature is less than 60°F, the total daily heat loss, THL, is

$$\mathrm{THL} = 7843DD \tag{13}$$

where DD are degree-days, given as the integral of $60 - T$ for that time during which the outside temperature is below 60°F. Note that, because of daily fluctuations in temperature, days with

average temperatures above 60°F may still have nonzero degree-days.

The computation of energy consumed by the gas furnace is the total heat loss divided by the efficiency of the gas furnace.

$$q_G = 100,000\, h_G = \frac{\text{THL}}{\eta_G} \tag{14}$$

where q_G is the energy in BTU provided by the gas and h_G is the gas consumption in therms (1 therm equals 100,000 BTU). We can now compute the present values of total cost of gas heat as the sum of the inflated and discounted costs of furnace purchase and installation, annual maintenance and energy.

$$C_G = C_{Ginst} + \sum_{t=0}^{t=n} \left[\frac{1+i}{1+r}\right]^t (C_{Gmaint} + C_{Genergy}) \tag{15}$$

Where r is the annual discount rate and i is the inflation rate, taken here to be a constant. n is the number of days in each case of the simulation.

The energy requirement for a heat pump system is a bit more complex as the relationship between degree-days and total heat loss is not linear. A heat pump system requires less energy input than is transferred into the house. For the heat pump system, we use a *coefficient of performance*, COP, to relate the rate of energy input to the rate of heat-loss,

$$\frac{dq_H}{dt} = \frac{1}{\text{COP}} \frac{dq_L}{dt} \tag{16}$$

where the daily energy input is the integral of dq_H/dt for that time during the day for which the outside temperature is below 60°F. Thus, the daily energy cost for the heat pump system is the daily q_H times the cost per BTU of electrical energy, $0.000293071 C_{elect}$. The coefficient of performance for a typical heat pump is a function of the outside temperature as shown in Fig. 2 [9]. For simulation purposes, it is convenient to obtain an analytical expression for COP. This expression takes the form of a logistic function.

$$\text{COP} = 2.88\{1 - e^{[e^{-[2.410+0.874(60-t)/7.97]}]}\} + 1 \tag{17}$$

We can now compute the present value of total costs for the heat pump system as the sum of the inflated and discounted heat pump purchase and installation cost, annual maintenance costs and energy costs.

$$C_H = C_{Hinst} + \sum_{t=0}^{t=n} \left[\frac{1+i}{1+r}\right]^t (C_{Hmaint} + C_{Henergy}) \tag{18}$$

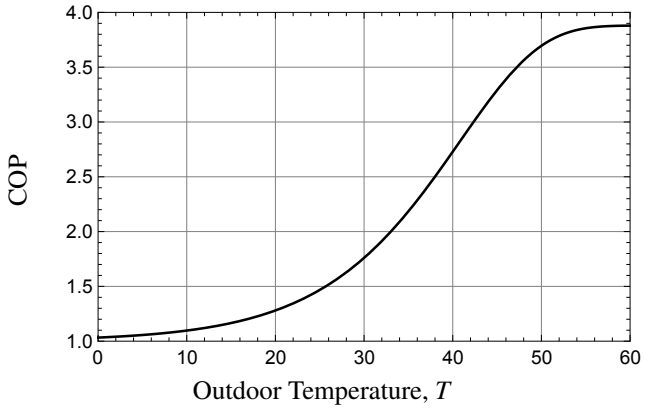

FIGURE 2. COP for a typical heat pump. The COP of the heat pump is the ratio of heat output to electric energy input.

Finally, we convert the present values of total system costs into a measure of utility assuming that the decision maker has a net worth, W, dedicated to home heating of $10,000, with the preference to retain as much of this amount as possible. We then take utility to be $\$10,000 - C$, where C is the total system cost (C_H or C_G), adjusted for the decision maker's risk preference. The risk preference taken for this example is $u = \log(W - C)$. The decision maker would then select the system with the highest expected utility. Via the mechanism of expected utility, we can account for risks that could alter the preferred choice. For example, heat pumps are more prone to rare but expensive failures than gas systems. The possibility, albeit of low probability, of a large expense some years in the future could deter the decision maker from choosing this alternative.

The data for this example are as follows:

Input Variables

Deterministic variables
- $R = 13$
- C_{elect} =$0.10/kWhr
- C_{gas} =10.38 $/1,000 cuft
- C_{Hinst} =$2,000*
- C_{Ginst} =$2,200
- Gas energy, 1027 BTU/cuft
- $\eta = 0.85$
- $i_{gas} - i_{elec} = 1\%$
- kWhr/BTU=0.000293071

Stochastic variables, triangular distributions (min, most likely, max values)
- i_{gas} (0.952%, 1.923%, 3.922%)
- C_{Hmaint} ($450, $700, $1,200)
- C_{Gmaint} ($150, $300, $1,000)

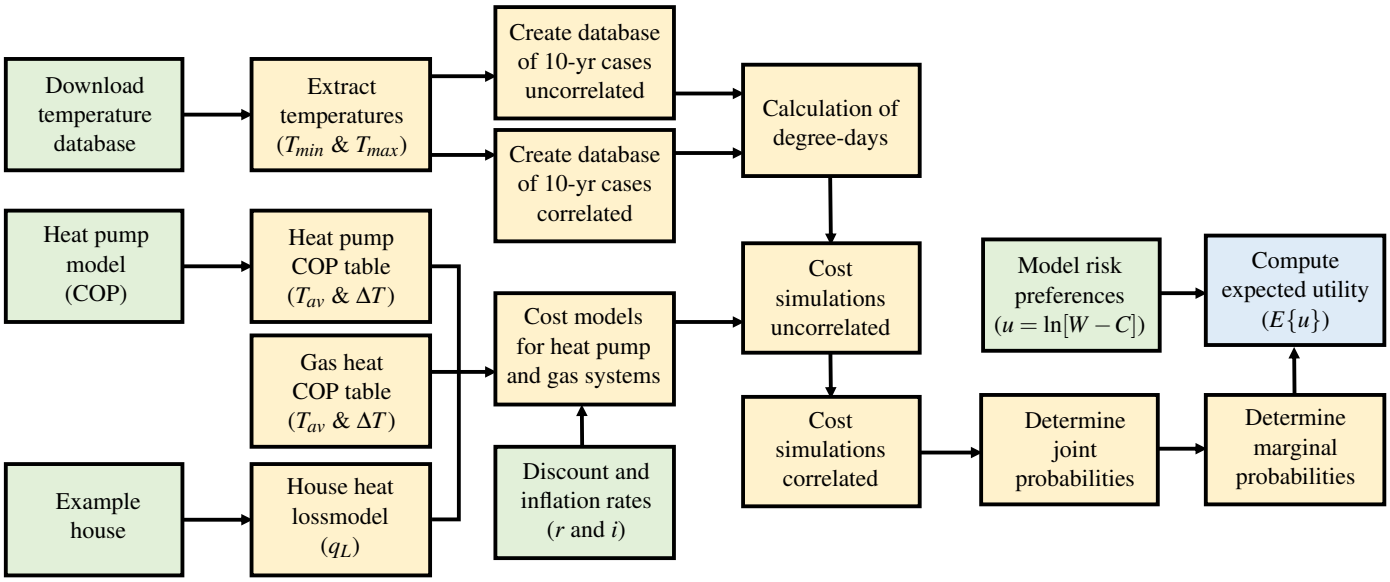

FIGURE 3. A framework for computation of utility difference.

Footnotes from table on previous page

*Price includes credit because a heat pump system does not need air conditioning.

5 Computational Procedure

The illustrative example was solved using a Monte Carlo simulation procedure. The logic flow of this simulation is shown in Fig. 3. The inputs to the analysis are shown in light green, and the result is in light blue. The straw colored boxes denote computations.

Monte Carlo simulations have the distinct advantage that they simplify the mathematics and programming of a solution considerably, thus significantly reducing the likelihood of errors and time spent coding the solution. But this advantage comes at the heavy cost of computational time. Indeed, Monte Carlo simulations can easily require run times that render them useless. The run times even for this simple problem could have easily run into weeks or more of desktop computer time. As a result, we devoted considerable effort to finding ways of reducing the computational operations needed to achieve results.

We began by obtaining a database of daily temperatures, including 85 continuous years of minimum and maximum daily temperatures. The temperature database was obtained from the National Oceanic and Atmospheric Administration (NOAA) [10] covering the time period from August 1, 1935, to July 31, 2020, for a total of 31,047 days. The minimum and maximum daily temperatures, T_{min} and T_{max}, were extracted from this database, and a daily average temperature, $T_{av} = (T_{max} - T_{min})/2$, was computed as the average of these temperatures. We then created a histogram of the daily average temperatures as shown in Fig. 4.

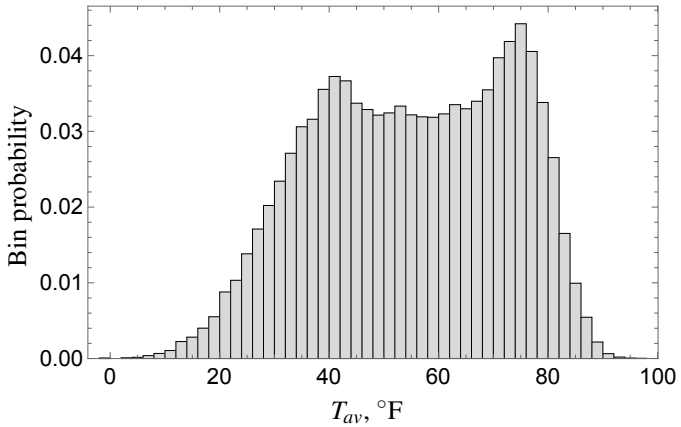

FIGURE 4. Histogram of T_{av} data, bin width 2°F. Data obtained from the NOAA database for Newark, New Jersey, airport (EWR).

Upon first thought, it would seem reasonable that an analysis of heating costs would require the random simulation of a temperature profile over the 10-year period of each simulation. However, on further consideration, it becomes apparent that we need obtain only the total heating requirements, and this computation can be based on a sampling of daily values of T_{av} and the maximum variation of temperature from the average, $\Delta T = T_{max} - T_{av}$, neglecting any day-to-day correlation in these temperatures. We now make one more simplifying assumption, taking daily temperature profiles to be sinusoids with temperatures ranging from $T_{av} - \Delta T$ to $T_{av} + \Delta T$. This greatly simplifies the simulation while

maintaining the essence of the problem.

Given this insight, we computed the daily temperature variations from the average temperature, $\Delta T = T_{max} - T_{av}$, and created histograms of these temperature variations for temperature blocks of 10°F, an example of which is shown in Fig. 5. Com-

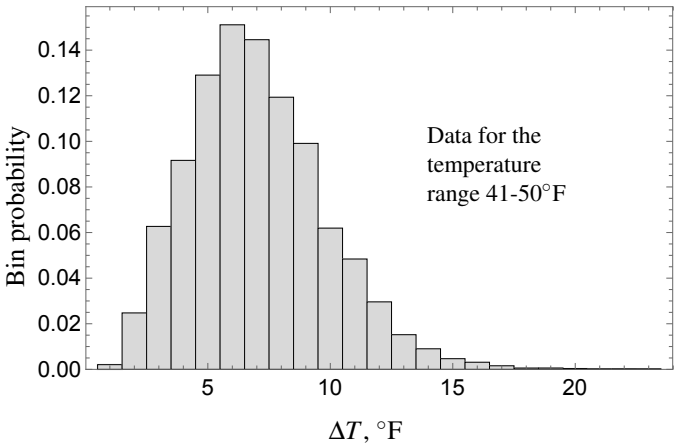

FIGURE 5. Histogram of ΔT data, bin width 1°F. ΔT is the difference between the maximum temperature and the average temperature as recorded on a day-by-day basis.

paring these histograms, it was noted that there is little variation in the distributions of ΔT as a function of T_{av}. Thus, we chose the histogram shown in Fig. 5 as a baseline distribution for ΔT. Then, from these histograms, we created two tables of integer temperature values, one for T_{av} and the other for ΔT. The order in which the temperatures are given in these tables is not relevant. Let n and m be the numbers of entries in each of these tables. Now, we select two uniformly distributed integers, N and M, ranging between 1 and n or m respectively, and by simply looking up in the respective tables the temperature values corresponding to these indices, we can create a random set of daily temperatures with precisely the same distributional forms as shown in the figures. Aside from the computation of the random numbers, which is very quick, generating temperature data for the simulation by this process requires no additional computation.

With the assumptions outlined above, we can now compute heating degree-days for each heating system as a function of T_{av} and ΔT. These computations convert heat loss into required energy input. Results of these computations are shown in Figs. 6 and 7, where degree-days, DD, have been computed for each integer combination of T_{av} and ΔT. These results were stored in a look-up table indexed by T_{av} and ΔT for values of T_{av} from 1 to 100, and ΔT from 1 to 25. This reduced the number of times we would need to evaluate DD from something in excess of 3 billion times to a total of 2,500 times for each heating system.

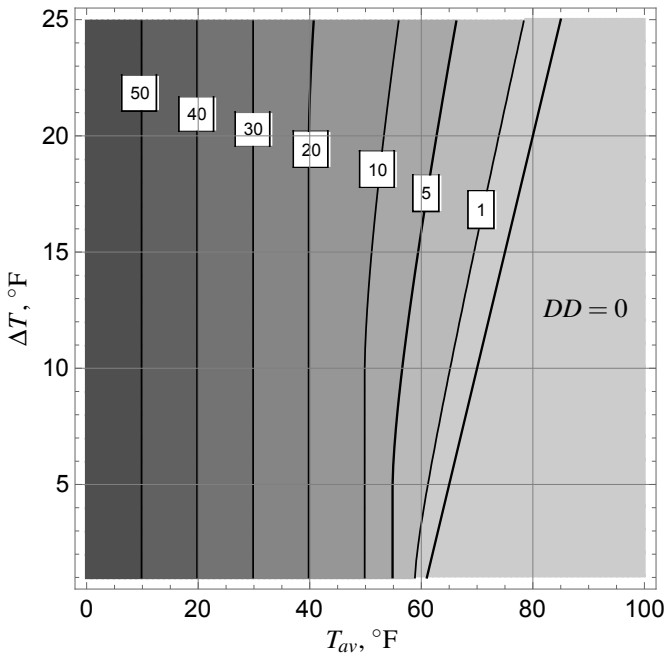

FIGURE 6. Daily gas-heat system degree-days.

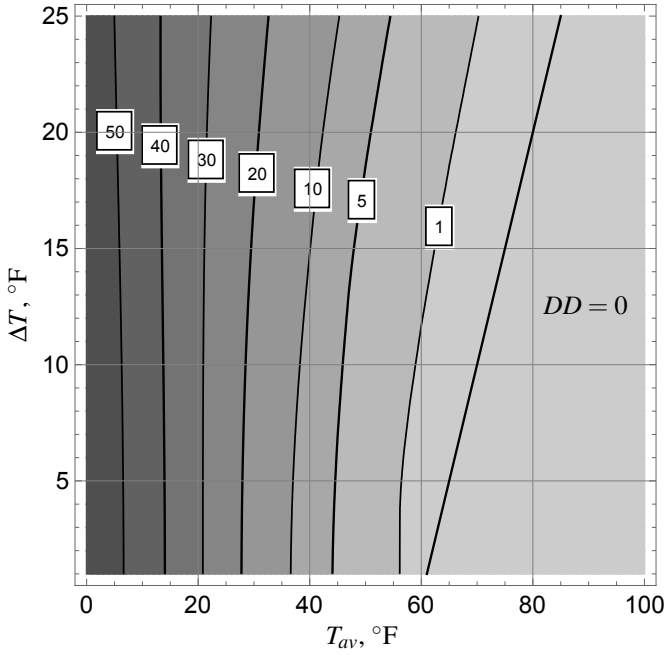

FIGURE 7. Daily heat pump equivalent degree-days.

The next step in the computational process was to compute 10-year cases of daily temperatures. To enable the Monte Carlo analyses that determine cost distributions, we chose to compute three sets of 1,000,000 cases each. Two sets were for use in computing cost distributions independently for the heat pump and gas heating systems, and the third set was for computing the differential cost distributions. For the differential cost computation, each case used the same sequence of temperatures for both heating systems.

Once we had simulated cases of daily temperatures, T_{av} and ΔT, we next used these cases together with the tabular data illustrated in Figs. 6 and 7 to obtain a degree-day table including degree-day records for both systems, both uncorrelated and correlated. This table enabled cost computations for each system for each of the 1,000,000 sets of cases, thus providing three independent sets of cost distributions. First, we obtained cost distributions for each system independently as shown in Fig. 8, which shows the heat pump system to have marginally lower expected cost than the gas system, but with considerable overlap in the probability distributions of cost. We then used these data to obtain the difference distribution shown in Fig. 9. Together, Figs. 8 and 9 show a distinct cost advantage for the heat pump system with the cost of the heat pump system lower than the cost of the gas system with probability 0.656. We then computed these same distributions using the third set of temperature data where the costs of both heating systems were determined from the same temperature set, that is, these results are correlated through the variables T_{av}, ΔT, r and i. The results are shown in Fig. 10.

Interestingly, comparing this figure to Fig. 8, we see that the addition of information relating to the correlation between these alternatives of the above four variables has obtained the result that the gas heating system is favored, contradicting the result

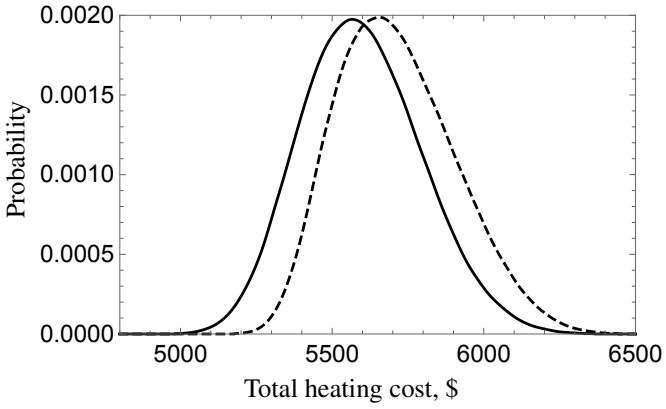

FIGURE 8. Estimated distributions of heating costs for heat pump (solid line) and gas (dashed line) systems computed independently.

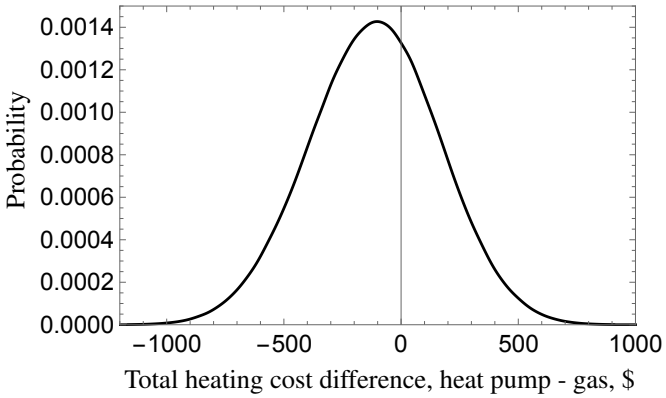

FIGURE 9. Difference distribution of heating costs for heat pump minus gas systems computed independently.

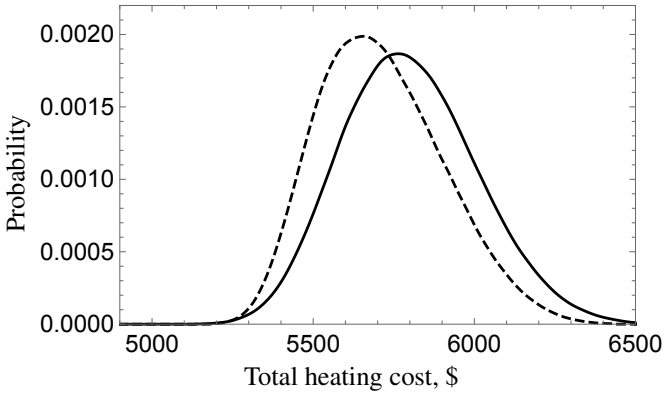

FIGURE 10. Estimated distributions of heating costs for heat pump (solid line) and gas (dashed line) systems computed accounting for correlation in weather and inflation.

shown in Fig. 8. Again, we plot the difference in cost between the two systems, Fig. 11, now using the correlation information, and we obtain the conclusion that the gas heating system has probability 0.6664 of being the cost-effective alternative.

With the above calculations completed, we can move on to an evaluation of the joint probability distributions of the costs of the heat pump and gas heating systems accounting for the condition, $J(A) - J(B) = h(\mathbf{x}, \mathbf{y}, \mathbf{z})$. This distribution, in the form of a 3-dimensional histogram, is shown in Fig. 12. Despite the fact that this histogram made use of 1,000,000 cases, each involving a 10-year cost simulation, it is evident that perhaps ten times more simulations would be necessary to obtain a reasonably smooth result in the vicinity of the most likely point. Upon completion of

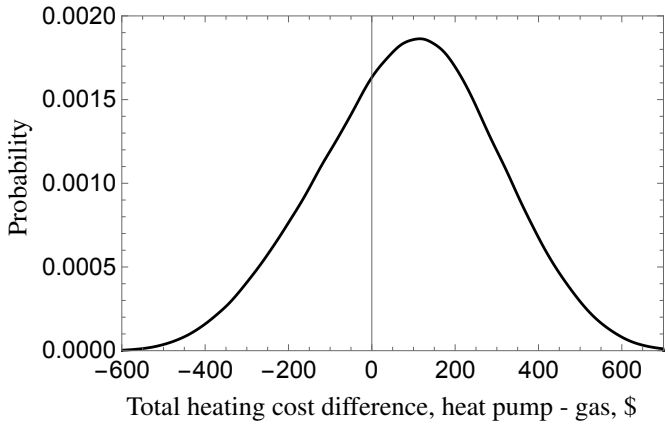

FIGURE 11. Difference distribution of heating costs for heat pump minus gas systems accounting for correlation.

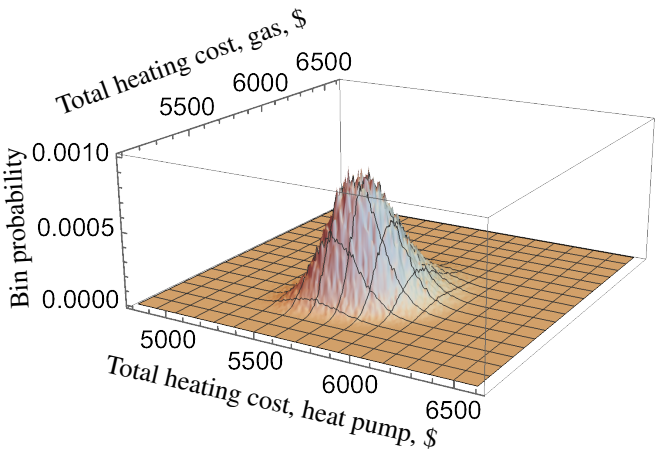

FIGURE 12. Joint distribution of heating costs for heat pump and gas systems accounting for correlation (bin size $\$10 \times \$10$).

the preliminary computations, which consumed about 20 hours of computer time, the final computation to obtain the results of Fig. 12 took an additional 4 hours of computation on a desktop computer with a processor speed of 3.3 GHz. It is easy to see that problems of this sort can be limited by computing capability and time.

At this point, it remains only to determine the expected utilities of the two heating system alternatives and, from that, the differential utility. To do this, we must first compute the marginal probabilities of each system, that is, the probabilities of total heating costs for each system as would be seen when projected onto the axes of Fig. 12. The histograms of these probabilities would appear as edge views of Fig. 12. In other words, they are computed as the sum of the bin counts of the histogram perpendicular to the axis of the system in question divided by the total of all bin counts (1,000,000 in this case). These probabilities are shown in Fig. 13. It is from these distributions that we can compute the expected utilities of the two systems. For the gas heat system we obtain $E\{u_G\} = 8.37612$, and for the heat pump system we obtain $E\{u_G\} = 8.36972$, confirming that the gas system would be preferred over the heat pump system consistent with the data of the example.

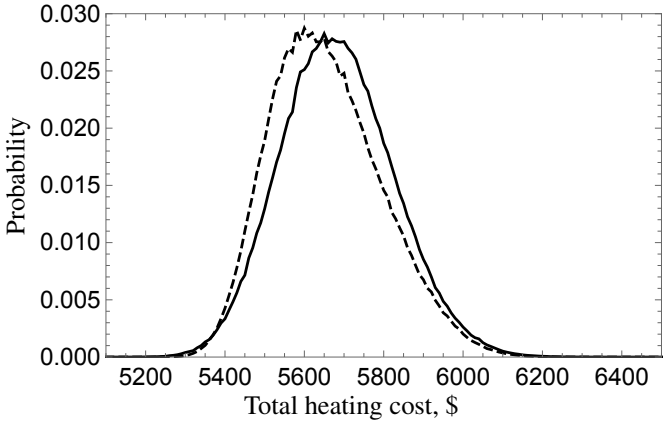

FIGURE 13. Marginal distributions of heating costs for heat pump (solid line) and gas (dashed line) systems accounting for correlation.

6 Conclusions

The objective of the research presented here is to provide a method for the introduction of additional information into the analysis of alternatives in otherwise ambiguous decision situations. In particular, we show that information may be derived from correlation among uncertain variables relating to different alternatives that is not available when analyzing the alternatives independently. In the example problem involving the evaluation of two alternative heating systems, while there is considerable uncertainty in the weather and future energy prices, significantly effecting our ability to accurately predict total system costs, the weather will be the same regardless which system is chosen, and prices for different forms of energy will be highly correlated. Introducing this information into the analysis provides additional clarity and, in the case presented, it even flipped the decision.

We show that information that emerges from a comparative evaluation of alternatives may derive from correlations among variables in a difference model (performance of alternative *A* minus performance of alternative *B*), it may arise because of the similarity of modeling assumptions made independently in the analyses of the alternatives, or it may come from a source that is unique to the performance difference between the alternatives such as a difference estimate or measurement.

We see that, for a risk neutral decision maker, a probability distribution on performance differences among alternatives is all that is needed to support the decision. However, for a decision maker who is not risk neutral, it is necessary to determine utility differences, referred to in this work as the *differential utility*, among the alternatives. This is because the utility difference depends on the decision maker's overall level of "satisfaction" (for example, net wealth) given the risky performance outcomes of the alternatives. Thus, it is necessary to compute a joint probability distribution of the performance measures of the alternatives and their respective differences. This can lead to significant computational intensity, possibly requiring days of computing. This is because we did not find a way to efficiently solve Eq. 9, which requires enforcing the condition $J(A) - J(B) = h(\mathbf{x}, \mathbf{y}, \mathbf{z})$. Because of the transcendental nature of these terms, there is no simple way to ensure that every Monte Carlo simulation meets this condition. Indeed, the condition is met only by a small percentage of the simulation cases, and these are the only useful cases.

While the use of Monte Carlo simulation greatly reduces the mathematical complexity, it does so at the cost of computational time. Indeed, the computational time required even for a problem as simple as that presented here can easily run into days or even months on a typical desktop computer. Thus, we found it necessary to make a number of simplifying assumptions and to break the overall analysis into a number of steps designed to save computing time. For example, we created tables of quantities that would require repetitive computation, thus converting a computational procedure into a table lookup using readily available indices. This is the case with the computation of degree-days for both heat pump and gas systems. By assuming that daily temperature profiles are sinusoidal, degree-days depend only on average daily temperature and maximum deviation from the average temperature. This computational simplification alone may have cut the overall computational time in half. Another simplification dealt with the creation of a set of randomly generated daily temperature profiles. By allowing daily temperature profiles to be created independent of day-to-day correlations in weather and allowing the sampling process to adequately represent each year's worth of daily profiles enabled another computational efficiency. This should be apparent when one realizes that the simulations presented here required the generation of approximately 3.65×10^9 such profiles. Unfortunately, tricks such as these are peculiar to the particular problem at hand, making it difficult to generalize the computational procedure as pictured in Fig. 3.

As a result of the above complexities, while differential utility can be quite helpful in providing clarity in the case of decisions where a clear dominance of one alternative over others is not otherwise available, the computational effort needed may diminish the value of the analysis. This said, it would appear that reliance more on analytical solutions and less on Monte Carlo simulation could prove to dramatically reduce computational times while retaining most of the mathematical convenience of the Monte Carlo approach and render differential utility an available protocol in otherwise difficult decision situations. It is suggested that this would be a beneficial focus of future work.

ACKNOWLEDGMENT

This work has been supported by the National Science Foundation under award CMMI-1923164.

REFERENCES

[1] Hazelrigg, G. A., and Busch, A. C., 1986. "Setting navigational performance standards for aircraft flying over the oceans". *Transportation Research,* **20A**(5), pp. 373–383.

[2] International Civil Aviation Organization, 1976. Report of the limited North Atlantic regional air navigation meeting (1976). Tech. Rep. Doc 9182, LIM NAT, Montreal, Canada.

[3] Reich, P. G., 1966. "Analysis of long-range air traffic systems: Separation standards–i". *Journal of Navigation,* **19**, pp. 88–98.

[4] M. Mitici, A. P. B., 2018. "Mathematical models for air traffic conflict and collision probability estimation". *IEEE Transactions on Intelligent Transportation Systems,* **20**, pp. 1052–1068.

[5] J. F. Shortle, Y. Xie, G. L. D., 2004. "Simulating collision probabilities of landing airplanes at nontowered airports". *Simulation,* **80**, pp. 21–31.

[6] Howard, R. A., and Abbas, A. E., 2016. *Foundations of Decision Analysis*. Pearson, Essex, England.

[7] Hazelrigg, G. A., 2012. *Fundamentals of Decision Making for Engineering Design and Systems Engineering*. G.A. Hazelrigg, http://engineeringdecisionmaking.com/, Vienna, VA.

[8] Hazelrigg, G. A., and Saari, D. G., 2020. "Toward a theory of systems engineering". Proceedings of the ASME 2020 IDETC/CIE. Paper IDETC2020-22004.

[9] Anon., 2014. White paper #30, hybrid (dual fuel) - gas heat and air source heat pump. Tech. rep. https://docplayer.net/16141975-Hybrid-dual-fuel-gas-heat-and-air-source-heat-pump.html.

[10] National Oceanic and Atmospheric Administration, 2020. Daily summaries station details. `https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USW00014734/detail`. Minimum and maximum daily temperatures for Newark Liberty International Airport.