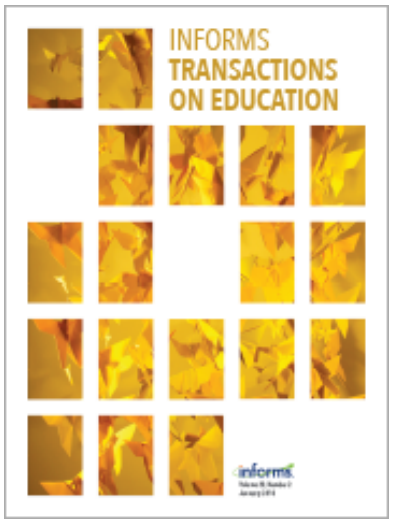
This article was downloaded by: [45.82.243.160] On: 14 July 2023, At: 23:50
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA



INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Monte Carlo Enhancement via Simulation Decomposition: A "Must-Have" Inclusion for Many Disciplines

Mariia Kozlova, Julian Scott Yeomans

To cite this article:
Mariia Kozlova, Julian Scott Yeomans (2022) Monte Carlo Enhancement via Simulation Decomposition: A "Must-Have" Inclusion for Many Disciplines. INFORMS Transactions on Education 22(3):147-159. https://doi.org/10.1287/ited.2019.0240

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2020 The Author(s)

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

http://pubsonline.informs.org/journal/ited

INFORMS TRANSACTIONS ON EDUCATION
Vol. 22, No. 3, May 2022, pp. 147–159
ISSN 1532-0545 (online)

Monte Carlo Enhancement via Simulation Decomposition: A "Must-Have" Inclusion for Many Disciplines

Mariia Kozlova,[a] Julian Scott Yeomans[b]

[a] School of Business and Management, LUT University, 53850 Lappeenranta, Finland; [b] Operations Management and Information Systems Area, Schulich School of Business, York University, Toronto, Ontario M3J 1P3, Canada

Contact: mariia.kozlova@lut.fi, https://orcid.org/0000-0002-6952-7682 (MK); syeomans@schulich.yorku.ca (JSY)

Received: December 18, 2019
Revised: June 30, 2020
Accepted: August 14, 2020
Published Online in Articles in Advance: December 7, 2020

https://doi.org/10.1287/ited.2019.0240

Copyright: © 2020 The Author(s)

Abstract. Monte Carlo (MC) simulation is widely used in many different disciplines in order to analyze problems that involve uncertainty. Simulation decomposition has recently provided a simple, but powerful, advancement to the standard Monte Carlo approach. Its value for better informing decision making has been previously shown in the investment-analysis field. In this paper, we demonstrate that simulation decomposition can enhance problem analysis in a wide array of domains by applying it to three very different disciplines: geology, business, and environmental science. Further extensions to such disciplines as engineering, natural sciences, and social sciences are discussed. We propose that by incorporating simulation decomposition into pedagogical practices, we expect students to significantly advance their problem-understanding and problem-solving skills.

Open Access Statement: This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "*INFORMS Transactions on Education*. Copyright © 2020 The Author(s). https://doi.org/10.1287/ited.2019.0240, used under a Creative Commons Attribution License: https://creativecommons.org/licenses/by/4.0/."

Funding: Financial support from the Strategic Research Council of Finland [Grant 313396/MFG40–Manufacturing 4.0]; the Foundation of Economic Education, Finland [Grant 16-8940]; and the Natural Sciences and Engineering Research Council of Canada [Grant OGP0155871] is gratefully acknowledged.

Supplemental Material: The Excel model for simulation decomposition is available at https://doi.org/10.1287/ited.2019.0240.

Keywords: Monte Carlo simulation • simulation decomposition • SimDec • uncertainty analysis

1. Introduction

Monte Carlo simulation methods have been used in a wide variety of scientific and applied disciplines to provide better insights into the uncertainties of future outcomes (Shonkwiler and Mendivil 2009, Zio 2013). For example, Monte Carlo methods have been considered in such diverse areas as the behavior of molecules (Van der Straaten et al. 2005), CO_2 emissions (New and Hulme 2000), ambulance demand (Fulton et al. 2009), CPU performance (Srinivasan et al. 2006), investment profitability (Arnold and Yildiz 2015), and seal populations (Warren et al. 1997). The basic idea underlying Monte Carlo simulation is straightforward: (i) The initial conditions/variable settings are altered multiple times within predefined intervals or distributions; (ii) with each change to the conditions, the resulting impacts to the model are calculated and recorded; and (iii) the outcomes of interest are subsequently summarized in histograms or some other forms of probability distribution. The histograms offer a visual representation of outcome uncertainties. The histograms also provide a significant quantity of information, including the minimum and maximum possible outcomes, the probability-weighted mean (i.e., the expected value), the shape of the distribution, and the likelihood that the outcome falls within a specific interval range. Frequently, the Monte Carlo simulation study concludes with the production of the graphical output. Sometimes, however, the simulation analysis is repeated for a number of scenarios, requiring the need for several different outputs to be displayed on separate histograms.

Simulation decomposition (SimDec) has recently provided a simple, but powerful, advancement to the standard Monte Carlo approach (Kozlova et al. 2016). The basic concept behind SimDec is to decompose (and color-code) the final output distribution into groups of outcomes resulting from multivariable combinations of different states of the input variables. The implementation of the method involves: (i) deciding which variables to choose for decomposition and which states to use for decomposition; (ii) recording states of chosen input variables through the simulation alongside the outcome values; and (iii) color-coding the resulting histogram in accordance with the combination of states from which each part of the histogram resulted. Because

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

the contribution of the state combinations to the overall output is easy to portray visually, SimDec can reveal previously unidentified connections between the multivariable combinations of inputs on the outputs. The decomposition has been shown to provide deeper insights into the uncertainty surrounding the problem, to assist decision making when actionable variables are chosen for decomposition, and to better understand the interplay of different sources of uncertainty on the distribution of outcomes (Kozlova et al. 2016, Kozlova and Yeomans 2019). The SimDec approach is completely generalizable to any Monte Carlo model with negligible additional computational overhead.

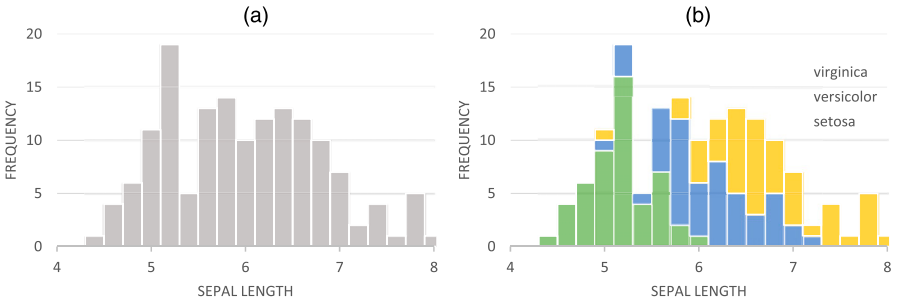
The concept behind applying decomposition to Monte Carlo simulation can be illustrated with the use of a stacked bar chart. In Figure 1, a histogram of the sepal lengths from the widely used data set of iris flowers (keywords "iris data" or "iris dataset" in ScienceDirect return 2,710 studies) is shown in graph (a), but is subsequently "decomposed" (or color-coded) by species in graph (b). The color-coding decomposition operation clearly supplies the insight that, even though their length intervals overlap, different species of iris possess different sepal lengths. Contrasting the one-series histogram (Figure 1(a)) to the decomposed multiple-series graph (Figure 1(b)) reveals a visually distinct relationship between the category (species) and the variable of interest (sepal length). Note that the histogram series on Figure 1(b) is stacked, not overlaid, which permits the displaying of all the data without anything being hidden, while simultaneously preserving the shape of the original overall histogram. Analogously, simulation decomposition can expose relationships between multivariable combinations of input variables and the outcomes of Monte Carlo simulation. Furthermore, the more complex the simulation model is, the more valuable the contribution of the decomposition becomes. Although stacked histograms and bar charts have been widely used across all domains of academia and industry, the idea of decomposing Monte Carlo simulation results has remained largely overlooked. To the best of our knowledge, multivariable simulation decomposition has only been applied to a couple of investment cases (Kozlova et al. 2016, Kozlova and Yeomans 2019); single-factor decomposition has been used in memristor performance analysis (García-Redondo et al. 2012) and has been alluded to sparingly in some very narrowly focused commercial software products.

In this paper, we will provide a series of examples from different disciplines illustrating where the use of simulation decomposition can prove extremely beneficial. These simplified examples can be readily used in pedagogical practices. To facilitate the adoption, we also provide an Excel model for SimDec and explain its usage. A list of possible applications that would benefit from decomposition is suggested at the end of the paper. This list should not be considered collectively exhaustive, but, rather, as a source of inspiration to extend decomposition to other problems that the reader might be likely to come across in their professional life.

Among the benefits of introducing SimDec to education is the opportunity for students to more deeply analyze various problems that they may encounter. This introduction will provide an incentive for them to better understand the interplay of variables in uncertain outcomes via a more systematic, meticulous approach to thinking. In other words, it is not only useful for solving specific types of problems faced by future practitioners, but also simultaneously to enhance the overall problem-solving skills of students. Furthermore, it is the high degree of applicability of simulation decomposition to a wide spectrum of "real-world" decision-making environments that makes an introduction to, and facility with, the concept imperative.

The remainder of the paper is structured as follows. First, we introduce the logic behind simulation decomposition in greater detail. Then, we demonstrate the Excel model and provide user instructions. We subsequently use the Excel model to apply SimDec to various examples taken from different disciplines. Finally, a summary, discussion, and concluding comments close the paper.

Figure 1. (a) Distribution of Sepal Length and (b) The Same Histogram Color-Coded by Species



Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

2. The Method

Decomposing the histogram of iris sepal lengths is straightforward because the data set already contains the values of the output variables (sepal lengths) and the corresponding array of categories. In contrast, Monte Carlo simulations usually provide only a vector of output-variable values. Consequently, SimDec requires some additional steps that involve decisions on how to perform the decomposition (which variables to consider and how to divide them into ranges/states) and recording the input variable states that produce the specific output values. The algorithm for SimDec is explained below and visualized in Figure 2.

Step 1. Identify the uncertain input variables $\tilde{V}_i$ and their probability distributions or ranges if uniform distributions are assumed. This step corresponds to the first step of a "normal" Monte Carlo simulation.

Step 2. Break down the variables $\tilde{V}_i$ into a relevant set of states, $\tilde{A}_{ij}$, for the problem-at-hand (for example, low–high, optimistic–expected–pessimistic, etc.). If an input variable is not relevant for the proposed decomposition, the variable can be left out in this step or, equivalently, said to possess only a single state $\tilde{A}_{i1}$ corresponding to the full range of its values.

Step 3. Define numeric boundaries $\tilde{B}_{ij}$ corresponding to each of the identified states $\tilde{A}_{ij}$ of each variable $\tilde{V}_i$. The boundaries can be defined by natural thresholds or dependency characteristics, be estimated by experts, or simply assumed to divide the variable value range into equal parts.

Step 4. Determine all possible combinations of the states $\tilde{A}_{ij}$ of the variables $\tilde{V}_i$. Each combination refers to a multivariable group $\tilde{G}_k$ that will correspond to a portion of the decomposition of the final probability distribution. The total number of these multivariable groups, $\tilde{G}_k$, will be a combinatorial function of the number of states $\tilde{A}_{ij}$ of each variable $\tilde{V}_i$.

Step 5. Record the states $\tilde{A}_{ij}$ of the variables $\tilde{V}_i$ while running Monte Carlo simulation together with the output values. Based on the resulting combinations of states, define the corresponding group $\tilde{G}_k$ that matches every outcome. The identification of the correct subdistribution for each iteration is achieved by mapping the randomly generated values of each individual input to their corresponding states $\tilde{A}_{ij}$, then mapping these states into the appropriate multivariable scenario combination $\tilde{G}_k$.

Step 6. Build a stacked histogram, where groups $\tilde{G}_k$ form the series of the chart. Derive required statistics for the full distribution and each group—for example, probability weighted mean, standard deviation, minimum, and maximum.

Figure 2. Schematic Representation of the Simulation Decomposition Approach (Kozlova and Yeomans 2019)

STEP 1 STEP 2 STEP 3

Variable $\tilde{V}_i$	Range	State $\tilde{A}_{ij}$	Boundary $\tilde{B}_{ij}$
$\tilde{V}_1$: X	[0, 100]	$\tilde{A}_{11}$: Low	$\tilde{B}_{11}$: [0, 30)
		$\tilde{A}_{12}$: Medium	$\tilde{B}_{12}$: [30, 60)
		$\tilde{A}_{13}$: High	$\tilde{B}_{13}$: [60, 100]
$\tilde{V}_2$: Y	{0,1}	$\tilde{A}_{21}$: True	$\tilde{B}_{21}$: {1}
		$\tilde{A}_{22}$: False	$\tilde{B}_{22}$: {0}

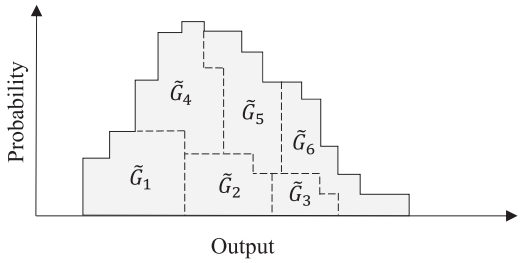
STEP 4

Group $\tilde{G}_k$	$\tilde{V}_1$ state	$\tilde{V}_2$ state	Combination of states
$\tilde{G}_1$	$\tilde{A}_{11}$: Low	$\tilde{A}_{21}$: True	$\{\tilde{A}_{11}, \tilde{A}_{21}\}$
$\tilde{G}_2$		$\tilde{A}_{22}$: False	$\{\tilde{A}_{11}, \tilde{A}_{22}\}$
$\tilde{G}_3$	$\tilde{A}_{12}$: Medium	$\tilde{A}_{21}$: True	$\{\tilde{A}_{12}, \tilde{A}_{21}\}$
$\tilde{G}_4$		$\tilde{A}_{22}$: False	$\{\tilde{A}_{12}, \tilde{A}_{22}\}$
$\tilde{G}_5$	$\tilde{A}_{13}$: High	$\tilde{A}_{21}$: True	$\{\tilde{A}_{13}, \tilde{A}_{21}\}$
$\tilde{G}_6$		$\tilde{A}_{22}$: False	$\{\tilde{A}_{13}, \tilde{A}_{22}\}$

STEP 5

Simulation iteration	Output	$\tilde{V}_1$ state	$\tilde{V}_2$ state	Resulting group $\tilde{G}_k$
1	45	$15 \in \tilde{A}_{11}$	$1 \in \tilde{A}_{21}$	$\tilde{G}_1$
2	100	$79 \in \tilde{A}_{13}$	$0 \in \tilde{A}_{22}$	$\tilde{G}_6$
3	90	$75 \in \tilde{A}_{13}$	$0 \in \tilde{A}_{22}$	$\tilde{G}_6$
4	20	$7 \in \tilde{A}_{11}$	$1 \in \tilde{A}_{21}$	$\tilde{G}_1$
.....				
10000	75	$32 \in \tilde{A}_{12}$	$1 \in \tilde{A}_{21}$	$\tilde{G}_3$

STEP 6



Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

Representing the SimDec with a stacked histogram permits a visualization of the overall output distribution from Monte Carlo simulation simultaneously overlaid with the contributions to this output distribution from different group combinations of the input variables. Namely, this explicitly allows an immediate visual tracing of how the variables chosen for the decomposition affect the outcome. If external sources of uncertainty are chosen for the decomposition variables, then the results will provide an understanding of how the outcome would behave under different uncertainty scenarios. Conversely, if partially controlled factors are selected for the decomposition, the results become a roadmap for decision making by depicting the effect of steering the input variables in one direction or another.

3. The Model

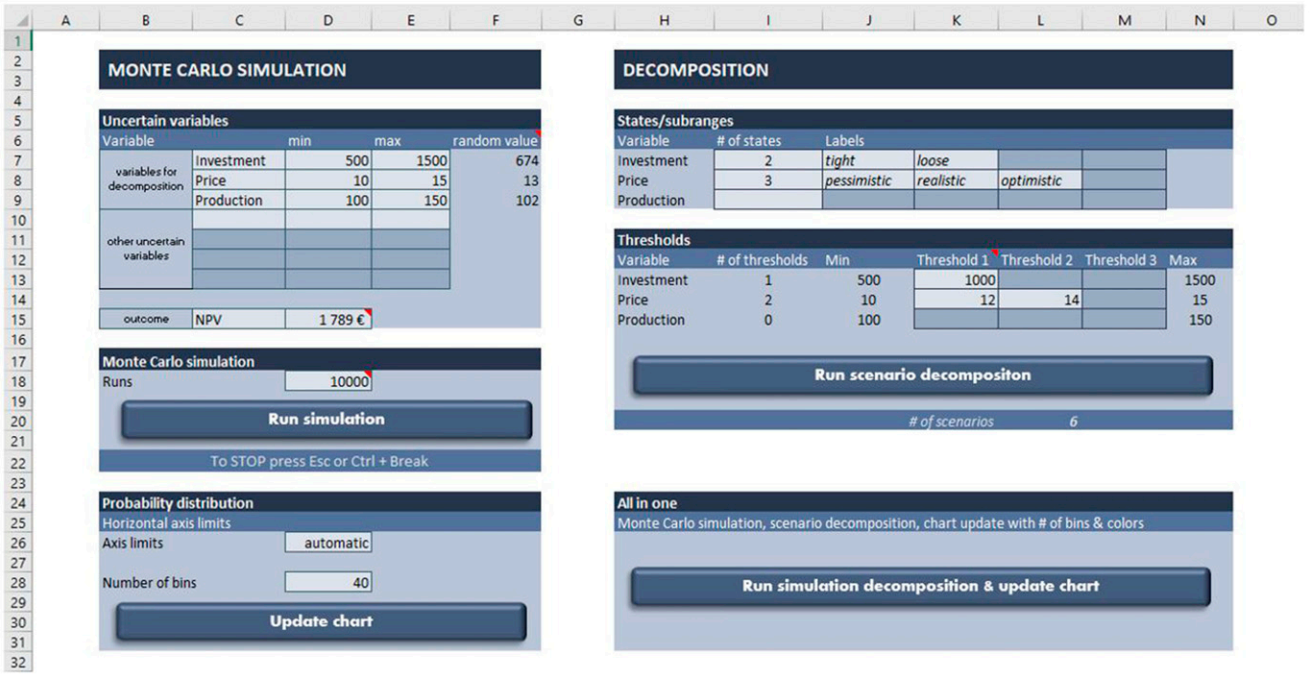
The logic of simulation decomposition has been previously implemented in Matlab (Kozlova et al. 2018a, b). In conjunction with this article, we provide an Excel model and explain its structure and usage. For its basis, an existing Monte Carlo model (Christie 2018) was taken and expanded with a histogram and simulation decomposition.

The main sheet, where one can tune the SimDec and obtain the results, is named "Monte Carlo." The model also performs operations on the auxiliary hidden sheets "MC_hist" and "MC_data." These hidden sheets should not be altered and are, therefore, inaccessible to the user. The first block (Figure 3, left) provides the means for performing classical Monte Carlo simulation. Uncertain variables and their ranges are specified in the area C7:E13 (Step 1); the model samples random numbers from a uniform distribution between the user-specified ranges. The resulting randomized values F7:F13 should be linked as inputs to the user model (problem-specific, and the calculation must be added by the user). Note, in this illustrated model structure, only the first three variables can be used for decomposition. The outcome from the user model (i.e., the output of interest), which is connected to the randomly generated inputs, should be linked to cell D15 on the "Monte Carlo" sheet. This value will be recorded for the histogram.

The number of iterations for the Monte Carlo must be specified in D18. Depending upon the inherent complexity of the user model, a Monte Carlo simulation with a high number of iterations can take a considerable amount of time. For a relatively smooth output distribution, we recommend at least 10,000 runs. The status bar (gray bar at the bottom of the Excel window) will show the progress of the simulation. In addition, one can update the desired display of the histogram by specifying a custom number of bins and axis limits. The "Run simulation" button initiates a simple Monte Carlo simulation performed without decomposition and updates to the chart. The "Update chart" button updates only the chart using the specified number of bins and axis limits without running the simulation again.

The decomposition block (Figure 3, right) is meant for setting up the SimDec. In the area I7:M9, the user must decide on the number of states for the chosen

Figure 3. A Snapshot #1 of the Model



Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

variables and assign labels to them (Step 2). When assigning labels and boundaries, the values must be set in increasing order of magnitude. Conditional formatting rules automatically display how many boundaries to specify in the area K13:M15 (Step 3) based upon the chosen number of states. The upper and lower boundaries are defined automatically based on the specified limits of the variable ranges. The "Run scenario decomposition" button uses the outputs of the most recently run Monte Carlo simulation and decomposes them into groups in accordance with specified rules (Steps 4 and 6). Note that the values of input variables are recorded automatically with simple Monte Carlo simulation (Step 5); therefore, the decomposition can be performed or renewed without running the Monte Carlo simulation again.

Running the simulation decomposition updates the graph and prepares/updates the area for the legend and color selection U8:Y_ (the values are dependent on number of groups) (Figure 4). Here, in column V, the user can assign the scenario/group titles, and in column U choose desired colors for every group. The "Update colors" button transfers the colors specified in the legend onto the chart. Different orderings of input variables and color choices can be used to enhance the visual clarity of the decomposed distribution.

All of the previous operations can be run consecutively by pressing the "Run simulation decomposition and update chart" button (Figure 3, bottom right). Selecting this option will (i) repeat the Monte Carlo simulation, (ii) use the new decomposition inputs (if any), and (iii) update the histogram. Statistics of the overall distribution are shown in the range P4:T6.

If a distribution other than a uniform is required, the user can rewrite the formulation in the random-number generator F7:F13. Common types of probability distributions can be created with Excel functions. If a custom distribution is required, the user can opt for either Visual Basics for Application (VBA) coding or for importing the values from such a distribution into Excel and randomly picking the values from that vector.

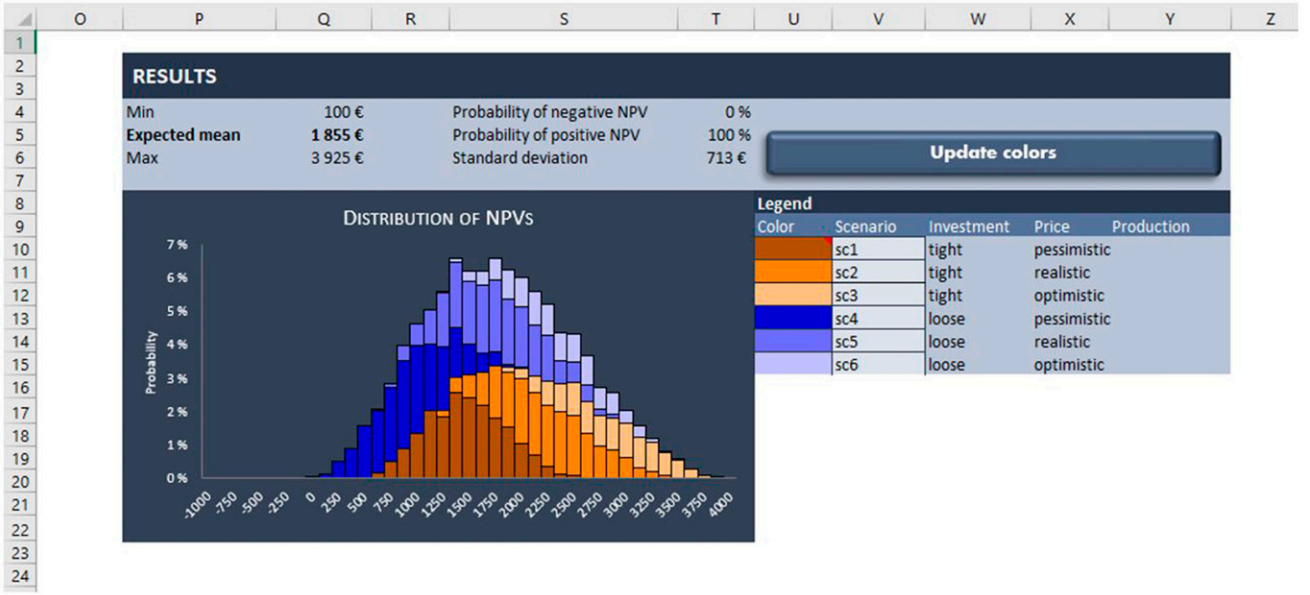
4. SimDec and Other Sensitivity-Analysis Techniques

As a method that aims to explain output variation, simulation decomposition falls under the aegis of sensitivity analysis.

Traditionally, a simple sensitivity analysis is performed factor-by-factor and represented by either spider diagrams or tornado charts. Whereas spider diagrams are limited to the same variations in relative magnitude of the input variables—for example, ±50%—tornado charts more realistically portray situations with asymmetrically different variations between the various inputs. Nevertheless, the main drawback of such analyses is that the influence of only one factor at a time is checked, while, in reality, several factors can underperform simultaneously, resulting in an outcome not neatly captured by the analysis.

Break-even analyses also aim to investigate the effects of different factors one by one. However, contrary to sensitivity analysis, a break-even analysis seeks to determine the critical input values that drive the system to an inappropriate output value. In investment analysis, the profitability threshold is usually

Figure 4. A Snapshot #2 of the Model



Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

one such break-even target—for example, checking what maximum amount of investment cost the project profitability can sustain. In software, robustness testing is designed to check how many faults the system can endure before crashing. In engineering, stress testing reveals the critical loads that lead to a breaking point. Although insightful, break-even analyses are also limited to a one-factor-at-a-time check, that might not necessarily properly reflect multifaceted systems or operation conditions.

Scenario analysis overcomes the one-factor limitation and, instead, proposes checking the outcome of the system in different possible future developments, usually comprising simultaneous variations of multiple inputs. However, the composition of scenarios depends on individual choices and could, therefore, conceivably omit realistic possible input combinations unintentionally.

Design of experiments is a broad field devoted to understanding the sensitivity of a system. Central to this field is an experiment which conditions are varied. Its underlying principles readily extend into mathematical and simulation modelling. The core questions pursued are how to determine the key variables and how to better reflect their influence. The representation of results is commonly portrayed by response surfaces, where the effect of up to two-at-a-time input/explanatory variables on the outcome/response variable is depicted. The number of surfaces required corresponds to the number of unique pair combinations—namely, $(n(n-1))/2$, where n is the number of input variables. In a small example, four input variables would require six separate response surfaces—a quantity already sufficiently large to pose potential perception challenges to a decision maker. Although SimDec is designed for simulations, its logic could easily be extended for visualizing experimental effects from multiple input variables on a frequency (rather than a probability) distribution.

The closest existing approach to SimDec is overlay charts that display results from different simulations in a single representation. Some commercial Monte Carlo packages enable using overlay charts—for example, Oracle's Crystal Ball and Palisade's @RISK, which are frequently used in business schools. The output from simulation decomposition differs from overlay charts in a number of distinct ways. First, the simulation-decomposition process automatically scales the display in the output chart according to the proportional likelihoods of the different scenarios, whereas overlay charts do not. Second, simulation decomposition is more computationally efficient, as it requires only a single simulation run to evaluate all N scenarios simultaneously, whereas overlay charts must perform N runs. Third, by default, the simulation-decomposition scenarios encapsulate the universal set of the simulation, implying that it is not possible to omit any potentially crucial scenario. Fourth, overlay charts require repetitive model modifications for each truncated input-scenario simulation, whereas simulation decomposition permits multiple scenario runs without the need for modifying the underlying/original simulation model. Finally, in comparison with overlay charts, the visualization of simulation decomposition (i) does not hide any information, (ii) allows a clearer visual comparison of scenario interdispositions, and (iii) preserves the readability when displaying numerous scenarios, simultaneously.

5. SimDec Application Examples

Here, we provide several examples of applying simulation decomposition to various problems that involve a great deal of uncertainty. The problems are gathered from geology, business, and environmental science and concern, in particular, (i) assessing oil reserves and the effect of different exploration options, (ii) evaluating investment-project profitability under different construction schedules, and (iii) estimating the carbon footprint of pallets for developing policy measures. All three examples could serve as teaching material.

5.1. Geology

One problem in geology that possesses a significant degree of uncertainty is the estimation of oil (or gas, or any other) reserves when exploring new fields. The amount of oil underground depends on the properties of the reservoir, its areal extent, thickness, and porosity. During the exploration phase, all three properties can only be known with imprecision, and a decision on whether to proceed with production will depend on an estimation of total resources. In order to cut down the imprecision around this estimate, further exploratory measures, such as exploration drilling and additional seismic studies, are undertaken.

Here, we describe an example of an arbitrary oil field possessing a complex geological structure with a reserve estimation based upon the current knowledge state (A). It is possible to resolve additional uncertainties by either drilling an additional exploration well (B) or by conducting a three-dimensional (3D) seismic investigation (C). The expert estimates of reservoir properties and of the uncertainty of these estimates is provided in Table 1. Because of unknown distributions, by using a standard Bayesian argument, the uncertain variables are assumed to be uniformly distributed. The uncertainty is specified as a percentage of the variation of the variable estimation (e.g., the current range of the areal extent estimations is $5 \pm 15\%$ km^2) (column (A) of Table 1). Experts believe that drilling the additional well will reduce the uncertainty in all factors down by 5% (column (B) of Table 1). The experts estimate that the 3D seismic study will reduce both the areal extent and

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

Table 1. Expert Estimation of Input Values for Oil-Reserve Calculation

		Uncertainty		
Item	Value	(A) Current, %	(B) After an exploration well, %	(C) After 3D seismic, %
Areal extent, km^2	5	±15	±10	±7.5
Thickness, m	20	±40	±35	±20
Porosity factor	0.2	±20	±15	±20
Oil density, t/m^3	0.875	—	—	—
Correction factor	1.15	—	—	—

thickness of reservoir uncertainties by half, but that this analysis would provide no additional understanding of the porosity (column (C) of Table 1). The question, then, is how to proceed with the exploration: drill several wells or conduct seismic works?

The oil reserve is calculated as

$$R = S * h * \mu * \rho * k,$$

where R is oil reservoir, S is areal extent, h is thickness, μ is porosity factor, ρ is oil density, and k is the inverse of the oil-formation volume factor.

For decomposition, we take all three uncertain factors and divide each one into two states of equal interval size—designating each state as either "low" or "high." The results are presented in Figure 5.

The current knowledge about the reserves (Figure 5(a)) is apparently mostly affected by uncertainty in the thickness of the reservoir. It can be seen in the figure that the group of scenarios 1–4 is shifted significantly from the group 5–8, while dispersion of other groups—for example, 1 and 2 versus 3 and 4 or 1 and 3 versus 2 and 4—is relatively close. Drilling an extra well (Figure 5(b)) reduces the overall uncertainty only slightly (the distribution becomes tighter). However, the 3D seismic study (Figure 5(c)) reduces the uncertainty surrounding the most influential factor, thickness. Groups 1–4 and 5–8 shift toward each other, and the whole distribution becomes much tighter. Because of this, the porosity is contributing to the variation in the rest of the results relatively more than to the other factors.

Here, SimDec reveals that the most contributing factor to uncertainty in the current situation is thickness. The decision to proceed with 3D seismic should be undertaken to diminish this uncertainty.

If a standard Monte Carlo simulation without simulation decomposition had been used, Figure 5(a) would not exist to show that the thickness was, in fact, the most crucial factor leading to the decision to conduct the seismic study. Furthermore, Figure 5(c) would not be available to indicate that porosity becomes the next crucial factor contributing to uncertainty. Without simulation decomposition, such findings would only be revealed by analyzing multiple different scenarios, individually, and by sequentially analyzing different factors in iterative simulation runs. Given the structural complexities inherent in oil reservoir analysis, a "real world" case would not necessarily be this straightforward, as, for example, an extra well could actually introduce significant additional complexities that confound the initial overall reserve estimate process.

5.2. Business

Investment analysis represents a classic business problem containing uncertainty that frequently employs Monte Carlo simulation. Investment profitability is a crucial factor for the decision to invest or not. Profitability is affected by multiple sources of uncertainty, such as demand and the state of the economy in the future.

Here, we present a simplified investment case. The aim is to analyze which strategy is better under price uncertainty—(i) enter the market with uncertain prices, (ii) enter bilateral agreement with a fixed price equal to its current level, or (iii) choose governmental support. Table 2 presents assumptions for the investment case.

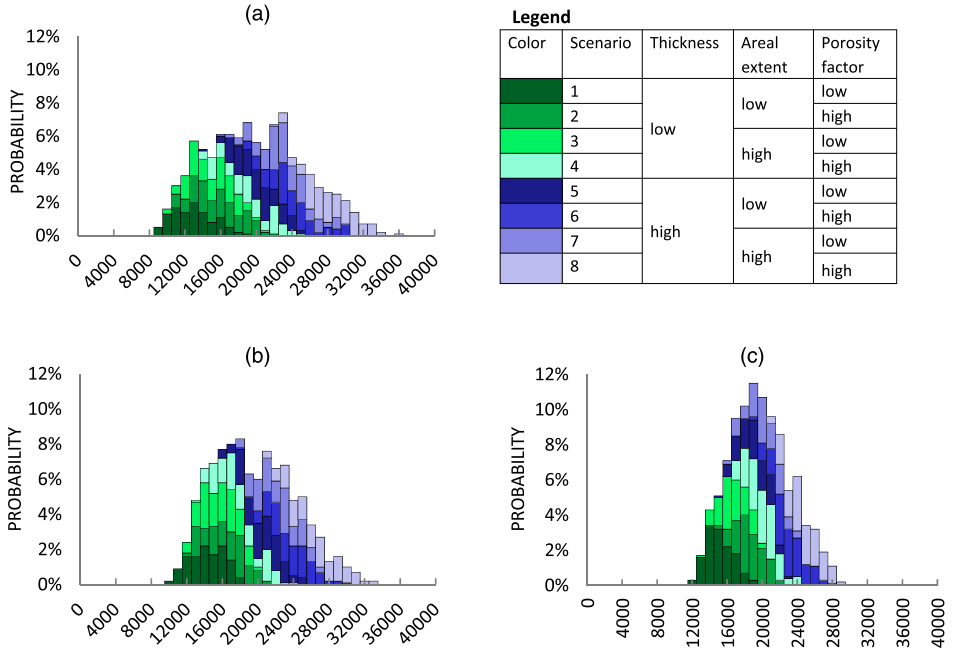
The governmental support incentivizes higher production and offers a fixed price of \$7,000 (7 K\$)/product if the annual production volume is lower than 150; 8 K\$/product if the production is between 150 and 200; and 9 K\$/product if it exceeds 200. The price uncertainty is modeled through an annual growth percentage that makes the price a dynamic variable in the calculation. The profitability of the investment is represented by its net present value (NPV), which is calculated as:

$$NPV = \sum_{t=0}^{T} \frac{CF_t}{(1+d)^t},$$

where NPV is the net present value; t is a year index from 0 to $T = 10$, CFt is annual cash flow in year t (i.e., total revenues minus total costs), and d is a discount rate.

The NPV distribution is decomposed by the two manageable variables, investment cost and production volume. The simulation is performed three times for the three strategies employed. The states of the investment cost are [2,000; 2,300], [2,300; 2,600], and [2,600; 3,000] and for the production [100; 150], [150; 200], and

Figure 5. Distribution of Possible Oil Reserves (Kt)



Note. (a) Current understanding; (b) after an exploration well; (c) after 3D seismic.

[200; 250]. The price growth and operation costs are not decomposed. The results are shown in Figure 6.

The investment in the market with a volatile price seems quite a risky endeavor, with the largest portion of profitability distributed in the negative zone (Figure 6(a)). The decomposition here shows that, in the negative zone, changes neither in investment cost nor in production volume would significantly affect the situation. However, the scenarios with high production volume possess higher potential if prices develop favorably. Removing price uncertainty via a bilateral contract with fixed price reduces the uncertainty in project profitability as well (Figure 6(b)). However, neither reaching high production volumes nor reducing investment costs guarantees profitability of the investment, and every single scenario has its minimum in the negative NPV zone. The subsidy that incentivizes production performance shifts the groups of scenarios with different production volumes further apart (Figure 6(c)). The decomposition allows one to visualize that, with high production volume and nonpessimistic investment cost, the profitability of the investment can be secured, thereby enabling a unique opportunity for investors. Obviously, this specific conclusion might not hold for a different investment case.

In this example, SimDec provides critical information for decision making. The best investment alternative is to seek state support while securing higher production (e.g., by choosing a good location or technology). In this case, the investment is profitable, even with all of the other factors uncertain.

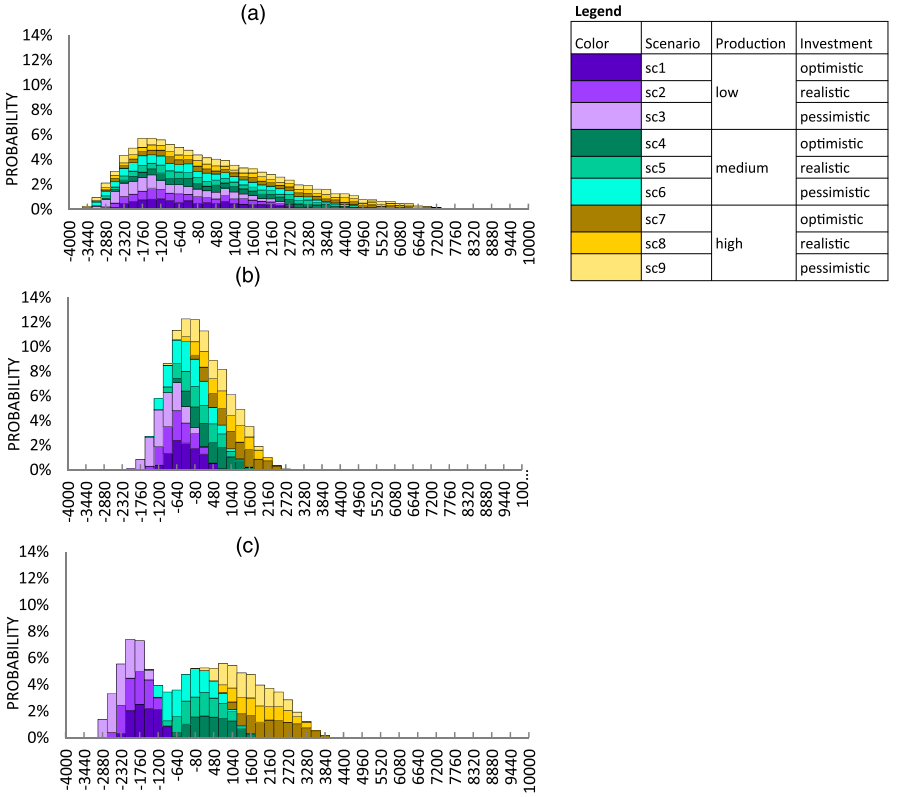
If a standard Monte Carlo simulation without decomposition had been used, the profitability distribution under the governmental support (Figure 6(c)) looks worse than under a bilateral contract (Figure 6(b)), and, therefore, a nonoptimal strategy would be selected. The unique opportunity of securing investment

Table 2. Investment Case Assumptions

	Value	
Variable	Minimum	Maximum
Investment cost, K$	2,000	3,000
Production, product/year	100	250
Current price, K$/product	8	
Price growth (annual), %	10	10
Maintenance and operation costs, K$/product	4	6
Tax, %	20	
Discount rate, %	10	
Investment lifetime, years	11	

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

Figure 6. Net Present Value (NPV) of Investment



Note. Under (a) the price uncertainty, (b) bilateral contract, and (c) governmental support.

profitability under the subsidy could not have been identified unless numerous separate simulations were run. Consequently, in this investment analysis, simulation decomposition helps to visualize the effect of strategy on risks that can be quite asymmetric. In the same fashion, simulation decomposition could be used for analyzing different hedging strategies, in which all the risks can be conveniently visualized.

5.3. Environmental Science

Life-cycle assessment studies involve an evaluation of environmental impact of a product or service from its initial resource extraction (so-called "cradle") to its manufacturing to the eventual disposal at the end of its life (so-called "grave"). Naturally, such assessments contain a high degree of variation and uncertainty, because identical products can lead very different lives (e.g., number of uses, regional variations in manufacturing and logistics, etc.).

In this section, we contrast the carbon footprints of plastic and wooden pallet systems that are used extensively in logistics worldwide. The study considers data on the pallets' raw materials, energy, and other inputs; on their use; and end-of-life. The inventory data are then used to calculate the global-warming potentials. The variations in the inventory data, which are a compilation of all inputs to the process, create uncertainty in the final estimation. For this example, all values are taken from the literature review estimates provided by Deviatkin et al. (2019). The numerical inputs are presented in Table 3, and the computational logic is described below.

Total life-cycle emissions are calculated as a sum of the emissions from different stages of the pallets' life, taking into account the functional unit size and the number of uses of a single pallet

$$EM_{total} = (EM_{production} + EM_{repair} + EM_{end_of_life}) * \frac{FU}{N} + EM_{transport} * FU.$$

Emissions for every stage are based on the global-warming potential (GWP) defined for every material or resource in accordance with the GaBi database (Thinkstep AG 2019). For wooden pallets. the footprints of every stage are the following:

$$EM_{production} = M_{timber} * GWP_{timber} + M_{nails} * GWP_{nails} + El_{wood} * GWP_{El_{wood}} + E_T * GWP_{E_T},$$

$$EM_{repair} = (M_{timber} * GWP_{timber} + M_{nails} * GWP_{nails} + El_{wood} * GWP_{El}) * 2,$$

$$EM_{transport} = (M_{timber} * 0.98 + M_{nails}) * L * 2 * GWP_{truck},$$

$$EM_{end_of_life} = GWP_{end_{wood}}.$$

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

Table 3. Input Details for the Pallets' Life-Cycle Emissions Estimation

Input	Notation	Value	Uncertainty/variation: Distribution type	Minimum (for uniform)/standard deviation (for lognormal)	Maximum (uniform)
Functional unit, # of customer trips	FU	1,000	—	—	—
Plastic pallets					
Plastic, kg	$M_{plastic}$	20	Lognormal	7	—
Electricity, kWh	$El_{plastic}$	99.8	Lognormal	117	
Number of uses	N	66	Uniform	34	250
End of life	$end_{plastic}$	Incineration	—	—	—
		Landfilling	—	—	—
Wooden pallets					
Timber, kg	M_{timber}	21.9	Lognormal	6.90	—
Nails, kg	M_{nails}	0.35	Lognormal	0.11	—
Electricity, kWh	El_{wood}	0.69	Lognormal	0.73	—
Thermal energy, MJ	E_T	4.1	Lognormal	2.5	—
Number of uses	N	20	Uniform	1	83
End of life	end_{wood}	Incineration	—	—	—
		Landfilling	—	—	—
Transportation (for both same materials)					
Transportation distance, km	L	150	Uniform	100	200
Truck type (GWP [kg/t/km])	$truck$	Big	—	—	—
		Medium	—	—	—
		Small	—	—	—

For plastic pallets, the stages' emissions are

$$EM_{production} = M_{plastic} * GWP_{plastic} + El_{plastic} * GWP_{El_{plastic}},$$

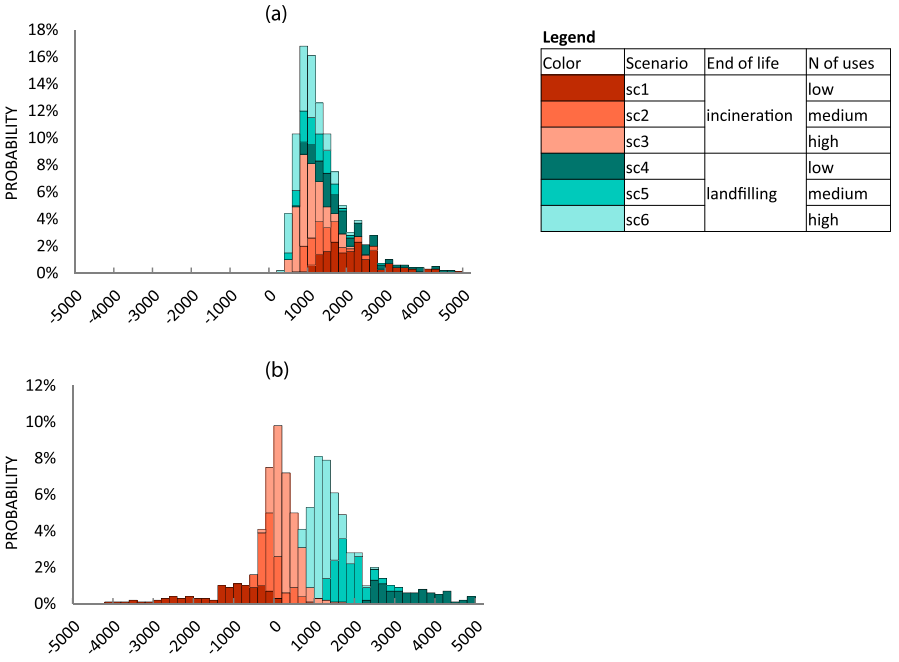
$$EM_{transport} = M_{plastic} * 0.98 * L * 2 * GWP_{truck},$$

$$EM_{end_of_life} = GWP_{end_{plastic}}.$$

The Monte Carlo simulation is run with all uncertain inputs, as specified in Table 3, and the decomposition is made by end-of-life type (burning or storing in a landfill) and the number of uses. The results are presented in Figure 7 for plastic (Figure 7(a)) and wooden (Figure 7(b)) pallets.

The footprint of the plastic pallets is not affected by the end-of-life choice—namely, whether to incinerate or to dispose of in a landfill. The environmental impact is higher the fewer times a pallet has been used.

Figure 7. Total Life-Cycle Emissions in Kg of CO2-Equivalent for 1,000 Customer Trips with (a) Plastic Pallets and (b) Wooden Pallets



Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

Therefore, the durability of plastic pallets emerges as a pivotal characteristic. For wooden pallets disposed of in a landfill, one can observe a comparable outcome. However, the effect inverts if the pallets are burned. This can be attributed to the negative emissions (or emission substitution) that results from generated electricity from burning a wooden pallet. When, for example, a plastic pallet is burned, some electricity is also produced, but the emissions from burning overcompensate this positive effect. Negative emissions at the end-of-life stage can outweigh the positive emissions from the other stages of a wooden pallet's life cycle when it is not used frequently, resulting in overall negative emissions. Therefore, for wooden pallets that are incinerated, the efficiency of use does not affect the overall environmental footprint in the same way.

Here, the simulation-decomposition analysis contributes a vital new insight into the problem by revealing how the effect of one factor reverses with respect to two groups of another factor. Without the visualization provided by simulation decomposition, this phenomenon could only have been uncovered by performing extensive computational testing of the model, while simultaneously playing with numerous different inputs.

In this case, SimDec provides critical information for policymaking decisions. For plastic pallets, longer usage should be incentivized. For wooden pallets, landfilling should be avoided, and, instead, procedures for pallet incineration should be implemented. Such a strategy will minimize the carbon footprint. A standard Monte Carlo approach would not have been able to produce enough information to provide these conclusions.

This example also explicitly illustrates that different types of probability distributions of input variables can be accounted for in the SimDec model.

6. Discussion and Conclusion

Simulation decomposition can be applied in any domain or discipline to practically any problem that involves uncertainty or imprecision. A decomposition is able to partition a Monte Carlo-simulation output distribution into components attributable to specific groupings of multivariable input combinations. Unlike other methods, the decomposition permits a concurrent visualization and projection of the combined multivariable stochastic uncertainties directly onto the overall output distribution. The SimDec approach is generally usable and can be utilized independent of the simulation context (Kozlova and Yeomans 2019).

There are multiple ways in which simulation decomposition can be utilized. García-Redondo et al. (2012) decompose memristor performance by a single variable, the level of its operation. Such decomposition allows understanding of how the output is affected by a variation in a single factor. This is especially valuable in nonlinear and complex models. Kozlova et al. (2016) decompose investment profitability by three actionable variables, investment cost, production level, and compliance with state requirements. Using only actionable factors creates a roadmap for decision making, because such decomposition visualizes where one would end up when steering the parameters in a certain direction. Kozlova and Yeomans (2019) focus on policy analysis and decompose investment profitability by the levels of two subsidy mechanisms that target one type of investment. They show how the picture would change if the two subsidies are correlated and have a cancelling-out effect on each other. Two decomposed distributions are used to demonstrate the effect of correlation.

In this paper, we introduce several more examples of using simulation decomposition. The example from geology features oil-reserve assessment. It uses three factors for decomposition that influence the estimation of reserves and depicts how the uncertainty in the estimation would change with different exploration actions in a series of distributions with the same decomposition logic. Such an approach illustrates how the effect of different uncertainty sources changes with different actions. The investment-analysis example provides an illustration of how an actionable variable (project design) and an external source of uncertainty (price growth) are used together for decomposition. Combining the two allows for a more sophisticated analysis and better-informed decision making. The example from environmental science highlights variations in ecological-footprint estimates of a product life cycle and reveals an intricate behavior in the effects arising from different life-cycle decisions.

It is possible (and even recommended) to apply simulation decomposition in different ways to a single problem. For example, for a complex investment case, it would make sense to see only external sources of uncertainty on the decomposition initially, then only actionable variables, and then, finally, the combinations of the two. Each investment case is different, and different ways of decomposing the uncertainty could enhance understanding and show a better way of managing it.

To illustrate the potentially vast applicability of SimDec, we have collected selected problems that involve uncertainty analysis and, therefore, could benefit from SimDec (see Table 4).

We start the list with a generalization attempt. Many problems can be reduced to analyzing the performance of the system. This depends on uncertain and uncontrollable factors of the supersystem or external environment and possibly, to some extent, manageable inner variations in system elements and subsystems. Although this generalization fits well, for example, investment analysis and engineering device/

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

Table 4. Selected Problems that Could Benefit from Simulation Decomposition

Discipline	Problem	Uncertain outcome	External sources of uncertainty	Inner variation, actionable variables
System theory (generalization attempt)	System analysis	System performance	Factors of the supersystem	Characteristics of system elements and/or subsystems
Meteorology	Wind speed assessment	Wind speed	Meteorological conditions	—
Engineering	Wind resource assessment	Power output	Wind speed and other meteorological conditions	Design, height, density, and spatial arrangement of turbines
	Device performance simulation	Performance output, efficiency	Factors of device environment	Elements of device design, operational modes
	Control robustness analysis	Control performance, error rates	Process variables, noise in the system	Control design
Chemistry	Pyrolysis	Mass and energy density of the resulting fuel	Natural variations in material composition	Reaction temperature and pressure, residence time, process media, particle size
Geology	Oil and gas resources/ reserve assessment	Quantity and quality of the resources/reserves	Reservoir thickness, areal extent, porosity, permeability	—
Biology	Animal population estimation	Population size	Mortality rate, pregnancy rate	Hunting
Business	Investment profitability analysis	Profitability indicator, e.g., net present value	Prices, demand, interest rate, inflation	Project costs, operation regime, project design, timing
Operations research	Scheduling	Cost	Service/work demand	Number of workers, number of shifts per worker, duration of shifts
	Logistics	Cost, distance	Product demand, breakdowns	Warehouse capacity, production volume, route, number of tracks
	Inventory	Cost	Demand, order frequency	Size, schedule, residence time
	Real options analysis	Real options value or net present value	Market uncertainty	Project costs, operation regime, project design, timing
Social sciences	Ambulance planning	Ambulance requirements	Number of patients, their condition, their location, traffic	Number and type of available vehicles
	Public policy analysis	Effectiveness, taxpayers' burden	Economic, social, political factors	Elements of the policy design

system performance, some problems in chemistry and molecular biology have the factors flipped. Namely, the environment becomes controllable, while the system itself is not. Many of the mentioned problems are nested. For example, the assessment of wind speed becomes an input to designing a wind farm to optimize power output, while the latter is one of the sources of uncertainty when analyzing wind-farm investment profitability. In fact, most of the problems end up being converted into monetary terms, one way or another, in investment valuation or cost assessment.

In educational practices, simulation decomposition can serve as an instrument to enhance students' understanding of a problem-at-hand and their problem-solving skills, in general, as well as adding a valuable takeaway instrument to their analytical toolkit for ready use in their future professional careers. The three examples in this paper have clearly demonstrated how problem understanding can be significantly enhanced through SimDec. These examples can be readily adapted to specific classroom situations by other instructors. Table 4 hints at the vast range of disciplines and problems that could easily benefit from the application of SimDec. Incorporating SimDec into pedagogical practices would also prove beneficial for equipping future specialists with this added analytical and visual tool. In academic settings, where Monte Carlo simulation modelling is frequently performed via general-purpose software (e.g., by coding in R, Matlab, or Excel VBA), specialized analytical techniques tend to be overlooked, unless they have been introduced in prior coursework. Consequently, the simulation studies cited in the first paragraph of this paper could have quite conceivably benefited and been extended by using SimDec. In industry and business, the vast majority of small- and medium-sized enterprises do not possess specialized simulation software, whereas essentially all of them would have Excel. Furthermore, even for companies already utilizing specialized simulation software, the additional understanding and deeper analysis from SimDec could prove crucial to its results and performance. All of these issues

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

could be resolved by incorporating SimDec directly into educational programs.

The authors have presented simulation decomposition as a tool within the broader topic of simulation in their introductory business classes at both the undergraduate (bachelor of business administration) and graduate (master of business administration (MBA)) levels. The students are provided with a spreadsheet model similar to that of Section 4.2 and are tasked with interpreting the resulting output charts. The SimDec concept itself has been well received, in general, but especially so by those part-time MBA students with backgrounds in the financial industries. In subsequent course offerings, we plan to expand the examples to incorporate applications from several of the other functional business disciplines. One of the authors recently introduced SimDec to their Master of Business Analytics (MBAN) VBA programming class. Given the more technical grounding of these students, there was more leeway to examine the underlying computer-modelling aspects inherent in the method. A couple of the MBAN students subsequently indicated that they plan to incorporate SimDec into their capstone project studies. Another couple of students from different universities chose to apply SimDec as an analytical tool in their theses. Perhaps the most surprising pedagogical acceptance of SimDec, thus far, has been by colleagues from the Economics Faculty of Saint Petersburg State University, who, after a guest lecture by one of the authors on decision-support systems that included a SimDec example, are now preparing to include SimDec in their textbooks. That will make SimDec a part of the analytical toolbox systematically taught to future generations of students through the established education program. Given the real-world nature of SimDec and its successful introduction so far, we plan to expand the coverage of SimDec in the subsequent instances of our courses. We strongly believe that other instructors will similarly observe the analytical benefits and "real-world" practicalities provided by SimDec and will be keen to adopt the approach into their classes also.

In this paper, simulation decomposition has been shown to be applicable to various problems from different application fields and can provide significant additional insight to the analysis of these problems by enabling an exploratory tracing of the effects of uncertainty on outcomes of interest. We believe that SimDec should become a readily incorporated addition to any simulation course, because it is easy to implement and requires only very minor additional computational resources to standard Monte Carlo simulation.

Acknowledgments

We thank Olga Timoshenko and Ivan Deviatkin for their help with building the cases for this paper. We thank the editor-in-chief, the associate editor, and two anonymous referees for contributing numerous insightful comments and helpful suggestions, which led to a substantially improved paper.

References

Arnold U, Yildiz Ö (2015) Economic risk analysis of decentralized renewable energy infrastructures—A Monte Carlo simulation approach. *Renewable Energy* 77:227–239.

Christie DS (2018) Build your own Monte Carlo spreadsheet. *Spreadsheets Ed.* 11(1):1–16.

Deviatkin I, Khan M, Ernst E, Horttanainen M (2019) Wooden and plastic pallets: A review of life cycle assessment (LCA) studies. *Sustainability* 11(20):5750.

Fulton L, McMurry LP, Kerr CB (2009) A Monte Carlo simulation of air ambulance requirements during major combat operations. *Military Medicine* 174(6):610–614.

García-Redondo F, López-Vallejo M, Ituero P, Barrio CL (2012) A CAD framework for the characterization and use of memristor models. *2012 Internat. Conf. Synthesis Model. Anal. Simulation Methods Appl. Circuit Design (SMACD)* (IEEE, Piscataway, NJ), 25–28.

Kozlova M, Yeomans JS (2019) Multi-variable simulation decomposition in environmental planning: An application to carbon capture and storage. *J. Environ. Informatics Lett.* 1(1):20–26.

Kozlova M, Collan M, Luukka P (2016) Simulation decomposition: New approach for better simulation analysis of multi-variable investment projects. *Fuzzy Econom. Rev.* 21(2):3–18.

Kozlova M, Collan M, Luukka P (2018a) Simple example for decomposition. Working paper, LUT University, Lappeenranta, Finland.

Kozlova M, Collan M, Luukka P (2018b) Matlab function for simulation decomposition. Working paper, LUT University, Lappeenranta, Finland.

New M, Hulme M (2000) Representing uncertainty in climate change scenarios: A Monte-Carlo approach. *Integrated Assessment* 1(3):203–213.

Shonkwiler RW, Mendivil F (2009) *Explorations in Monte Carlo Methods*, *Undergraduate Texts in Mathematics* (Springer Science & Business Media, Berlin).

Srinivasan R, Cook J, Lubeck O (2006) Ultra-fast CPU performance prediction: Extending the Monte Carlo approach. *2006 18th Internat. Sympos. Comput. Architecture High Performance Comput. (SBAC-PAD'06)* (IEEE, Piscataway, NJ), 107–116.

Thinkstep AG (2019) GaBi software system and database for life cycle engineering. Accessed February 1, 2020, http://www.gabi-software.com/finland/databases/gabi-database-edition-2019/

Van der Straaten T, Kathawala G, Trellakis A, Eisenberg RS, Ravaioli U (2005) BioMOCA—A Boltzmann transport Monte Carlo model for ion channel simulation. *Molecular Simulation* 31(2-3):151–171.

Warren WG, Shelton PA, Stenson GB (1997) Quantifying some of the major sources of uncertainty associated with estimates of harp seal prey consumption. Part 1: Uncertainty in the estimates of harp seal population size. *J. Northwest Atlantic Fishery Sci.* 22: 289–302.

Zio E, ed. (2013) *The Monte Carlo Simulation Method for System Reliability and Risk Analysis* (Springer, London).