DICEROLLER: Estimating D&D costs for the NNSA

Zachary Matheson; Dr. Charles R. Loelius; Gregory Stamp; Cash Fitzpatrick; Julie Anderson

Abstract

The National Nuclear Security Administration Office of Programming, Analysis, and Evaluation has developed a model for estimating the cost of Decontamination and Disposition (D&D) activities. This effort involved collecting and normalizing cost data from past D&D projects, and then generating a parametric cost estimating relationship to predict future D&D project costs. The resulting model, named DICEROLLER, will be used to make lifecycle cost estimates and one-for-one replacement cost estimates of capital acquisition projects.

Table of Contents

Introduction

The National Nuclear Security Administration (NNSA), a semi-autonomous organization within the U.S. Department of Energy (DOE), contributes to national and global security through nuclear deterrence, nonproliferation, counterterrorism, naval nuclear propulsion, and national leadership in science, technology, and engineering. The Office of Programming, Analysis, and Evaluation (PA&E) supports the NNSA mission by providing analytical services such as cost analyses to aid informed planning and decision-making.

One set of cost analyses provided by PA&E concerns the D&D of existing facilities owned by the NNSA. The NNSA owns facilities across the United States which date back to the Cold War. Many of these facilities are in disrepair or disuse. According to DOE policy [1], new NNSA construction must be "offset by the sale, declaration of excess, or demolition of building area of an equivalent or greater size." The cost of this one-for-one replacement is to be included in the budget justification document for the project. Additionally, PA&E is sometimes asked to provide lifecycle cost estimates for new facilities, including D&D costs.

The need to estimate D&D costs motivated PA&E to develop a CER for early-stage D&D cost estimates in the NNSA. The model, known as DICEROLLER (**D**&D **I**ntegrated **CER** for **O**ne-for-one and **L**ifecyc**L**e **E**stimate **R**anges), supports lifecycle cost estimates for capital acquisition projects, and "one-for-one" replacement cost estimates.

The DICEROLLER model

Model Objectives

Because PA&E analyses are performed early in the project, the D&D CER should be high-level. A lifecycle cost estimate created during an analysis of alternatives cannot draw on a detailed set of facility information to estimate the cost of D&D activities, since the project has not yet made it through even the conceptual design phase. In particular, the D&D cost model should depend on a small number of cost drivers, which should be easy to identify at early stages. Yet the model should cover a wide range of project scope, facility size, and costs in order to describe the variety of facilities belonging to the NNSA. Finally, the model should be firmly rooted in historic project data.

Desiderata:

- High-level for early-stage estimates
- Easy to use
- Small number of variables, which should be easy to identify at early stages
- Covers a wide range of project scope, size, costs, etc.
- Based on historic data
- AACE class 5 estimate (within -50% to +100% of the actual value) [2]

Historical Data

D&D project cost and scope data was collected from *Project Assessment & Reporting System II (PARS)* [3], NNSA *Program Management Information System, Generation 2 (G2)* [4], and the DOE Office of Environmental Management *Environmental Cost Analysis System (ECAS)* [5]. These points were cross-referenced with the DOE *Facilities Information Management System (FIMS)* [6] to obtain facility-specific information such as facility size, usage, contamination, and type of construction. This information was used to identify cost drivers. In total, 41 data points were collected, covering a wide range of project scope (see **Table 1**).

	Range
Facility Size Removed	240 ft^2 – 319,742 ft^2
Total Project Cost	$3,764 - $343,000,000
Contamination	Radiological Lead & asbestos None
Building Type	Permanent technical Permanent non-technical Simple or temporary

Table 1 Range of project cost and scope for data used in DICEROLLER

Cost data was escalated using PA&E's chosen escalation index [7] to directly compare projects from different years, and it was adjusted by a location-based factor [8] to account for different costs of doing business at the various sites across the nuclear security enterprise.

Cost Drivers

After analyzing the data, PA&E identified three cost drivers for NNSA D&D projects:

1. Facility size,
2. Contamination type, and
3. Building construction type.

Facility Size: Facility size is given in gross square feet (GSF).

Contamination Type: Each data point was binned into one of three categories, representing three different types of contamination:

1. Radiological contamination,
2. Construction material contamination (such as lead or asbestos), and
3. No contamination.

If a facility contains both radiological and asbestos contamination, it is binned under "radiological" since radiological contamination tends to have the greater impact on project cost.

Building Construction Type: Buildings are grouped by the complexity of their construction:

1. *Permanent technical facility:* Buildings in this category are generally purpose built and feature special fixtures, unusual building requirements, or custom support infrastructure.
2. *Permanent non-technical facility:* Buildings in this category resemble "standard" office buildings, with features such as a foundation, steel or wood framing, and plumbing, electrical, and HVAC systems.
3. *Temporary or simple structure:* Buildings in this category include portable facilities such as trailers, which lack a foundation, and simple structures such as warehouses, which consist mainly of empty space.

To give some examples, a college chemistry laboratory built in the 1940s would likely fall in contamination category 2 and building category 1. An office building constructed in 1980 would probably be in contamination category 3 and building category 1.

Because the costs and facility sizes represented in the dataset cover several orders of magnitude, the total project cost (TPC) and gross square footage were converted to logarithmic space (logarithm base 10), and then the regression was performed deterministically using ordinary least squares.

To convert from a log space prediction $\hat{y}_i$ to an actual dollar amount $\widehat{TPC_i}$, the value $10^{\hat{y}_i}$ is multiplied by a zero-bias factor: $\widehat{TPC_i} = ZBF \cdot 10^{\hat{y}_i}$. The purpose of a zero-bias factor, as explained in [9], is to correct for the tendency of power law-based models to overestimate. The zero-bias factor is equal to $ZBF = \sum_{i=1}^{n} \frac{1}{n} 10^{y_i - \hat{y}_i}$ where y_i is the actual value of $\log(TPC_i)$ and $\hat{y}_i$ is the prediction. In this paper, however, only the log space equations will be shown.

Treatment of Categorical Variables

The DICEROLLER model depends on two categorical variables: building type and contamination. PA&E trialed several model forms for DICEROLLER, each of which treats the categorical variables in a different way.

The first and simplest model form uses a technique called label encoding, wherein each group is assigned an integer value 1, 2, or 3. With facility size (GSF), contamination type ($Contam$), and building type ($BldgType$) known for a particular facility, the following equation estimates the project's TPC:

Equation 1: $$\boldsymbol{log\ (TPC) = \alpha + \beta \cdot log\ (GSF) + \gamma \cdot Contam + \delta \cdot BldgType}$$

where $Contam \in \{1,\ 2,\ 3\}$, $BldgType \in \{1,\ 2,\ 3\}$.

However, the label encoding method is generally not recommended as a best practice because it has well-known deficiencies. For one, it fails to account for interactions between variables. This method treats building type as independent of contamination, but in reality a building that is contaminated with radiological material, for example, is most likely to be a permanent technical facility. Another issue is that label encoding imposes an artificial spacing between levels. Because

the groups are labelled by evenly spaced integers, the model implicitly assumes – correctly or not – that the difference in D&D costs (in log space) between a building with radiological contamination and a building with asbestos contamination is the same as the difference between a building with asbestos and a building with no contamination.

An improved version of label encoding treats the labels as hyperparameters to be tuned. For DICEROLLER, bin labels 1 and 3 were preserved but the middle bin label was allowed to vary in order to minimize some predefined objective function. Such optimization removes the label encoding method's implicit assumption that bins are equally spaced. Additionally, if labels for two or more categories are allowed to vary simultaneously then this improved method will, in a very limited way, incorporate interactions between different categorical variables.

Equation 2: $\boldsymbol{log\ (TPC) = \alpha + \beta \cdot log\ (GSF) + \gamma \cdot Contam + \delta \cdot BldgType}$

where $Contam \in \{1,\ x,\ 3\}$, $BldgType \in \{1,\ y,\ 3\}$.

A third trial model form uses a technique called dummy encoding [10]. Dummy encoding involves assigning data a value of 1 if it belongs to a particular group, or 0 if not. In the case of a single categorical variable, a category with k groups introduces k-1 dummy variables; a kth label is redundant since a datapoint that does not belong to any of the first k-1 groups must automatically belong to the kth. In the case of multiple categorical variables, one creates a new dummy variable for every combination of variable groupings (e.g. buildings with contamination bin 3 and building type 2). In this way, dummy encoding makes no assumptions about interactions between variables, but rather provides a framework for such relationships to fall out automatically in the regression. The drawback to this approach is that it introduces many additional parameters to the model, which means that a large dataset is required in order to avoid overfitting.

The most general model form for the DICEROLLER CER has a total of 18 parameters (3 contamination groups times 3 building type groups, times two to account for *possible* interactions with facility size):

Equation 3: $\boldsymbol{log(TPC) = \beta_1 + \beta_2 D_{11} + \beta_3 D_{12} + \beta_4 D_{13} + \beta_5 D_{21} + \beta_6 D_{22} + \beta_7 D_{23} + \beta_8 D_{31} + \beta_9 D_{32} \sout{+ \beta_{10} D_{33}} + log\ (GSF) * (\beta_{11} + \beta_{12} D_{11} + \beta_{13} D_{12} + \beta_{14} D_{13} + \beta_{15} D_{21} + \beta_{16} D_{22} + \beta_{17} D_{23} + \beta_{18} D_{31} + \beta_{19} D_{32} \sout{+ \beta_{20} D_{33}})}$

where $D_{ij} = \begin{cases} 1\ if\ Contam = i, BldgType = j \\ 0\ otherwise \end{cases}$ and $Contam \in \{1,\ 2,\ 3\}$, $BldgType \in \{1,\ 2,\ 3\}$.

The group labels for all three methods are shown in **Table 2**.

Contamination	Building Type	Contam	Bldg	Contam	Bldg	Group							
Radiological	Technical	1	1	1	1	1	0	0	0	0	0	0	0
Lead/asbestos	Technical	2	1	2.14	1	0	1	0	0	0	0	0	0
None	Technical	3	1	3	1	0	0	1	0	0	0	0	0
Radiological	Non-technical	1	2	1	1.98	0	0	0	1	0	0	0	0
Lead/asbestos	Non-technical	2	2	2.14	1.98	0	0	0	0	1	0	0	0
None	Non-technical	3	2	3	1.98	0	0	0	0	0	1	0	0
Radiological	Temporary	1	3	1	3	0	0	0	0	0	0	1	0
Lead/asbestos	Temporary	2	3	2.14	3	0	0	0	0	0	0	0	1
None	Temporary	3	3	3	3	0	0	0	0	0	0	0	0

Table 2 Three different ways of labeling categorical variables in DICEROLLER: 1) Label Encoding, 2) Improved Label Encoding, and 3) Dummy Encoding

Results

The 41 points in the DICEROLLER dataset were split into three groups: a training set consisting of 29 points (approximately 70% of the total dataset), and validation and testing sets each consisting of 6 points (approximately 15% of the dataset). Data was assigned randomly to the three groups, except that the points with the largest and smallest costs were deliberately assigned to the training set.

After using the training set to tune the regression parameters, model verification was performed by checking that the conditions for OLS regression (homoscedasticity and normality of residuals, etc.) were satisfied. For model form 2 (**Equation 2**), an additional step was necessary to optimize the bin label hyperparameters. After minimizing log space mean-squared error (MSE) on both the training and validation datasets, the bin labels for the improved label encoding method are: $Contam \in \{1,\ 2.14,\ 3\}$, $BldgType \in \{1,\ 1.98,\ 3\}$ (see **Figure 1**). The testing data was not used for model training or selection, but is used to report the final model's score.

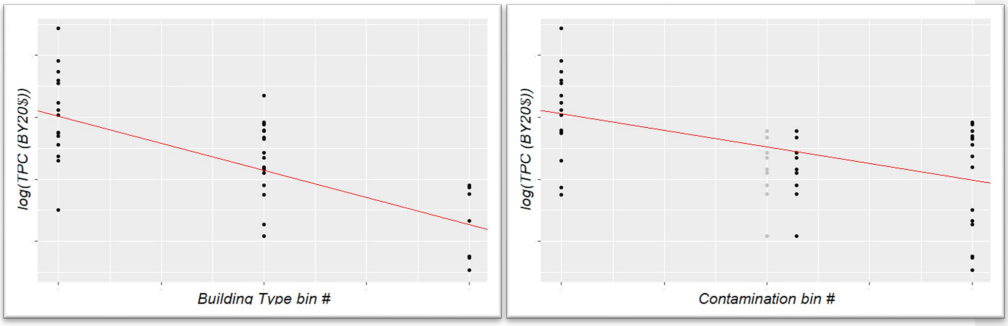

Figure 1 log(TPC) as a function of 1) building type bin # and 2) contamination bin #. Gray dots represent label encoded bin numbers {1, 2, 3}, while black dots represent improved label encoded bin numbers. Note how the black dots align more closely with the red trendline than the gray dots. Closer alignment with the trendline is associated with a better overall fit.

Commented [MZ1]: Probs redo these plots with bigger letters and maybe no TPC numbers

Commented [LC2R1]: Yes I think that's good!

For model form 3, it was necessary to remove most of the 18 terms in **Equation 3** due to statistical-insignificance in the regression. Starting from the full 18 variable model, terms were removed one at a time until all of the coefficients were statistically significant ($p < 0.05$). This procedure eliminated all but 6 parameters. At this point, the model did not satisfy the assumptions for OLS, so additional parameters were removed until the OLS assumptions were satisfied and MSE was minimized. The resulting model is shown in **Equation 4**:

Equation 4: $$\boldsymbol{log\ (TPC) = \alpha + \beta \cdot D_{NoneTemp} + log(GSF)\ (\gamma + \delta \cdot D_{RadTech})}$$

where $D_{RadTech} = \begin{cases} 1\ if\ rad\ contaminated\ technical\ facility \\ 0\ otherwise \end{cases}$

and $D_{NoneTemp} = \begin{cases} 1\ if\ uncontaminated\ temporary\ facility \\ 0\ otherwise \end{cases}$.

The resulting mean squared error for each model is shown in **Table 3**.

	Label encoding	Improved label encoding	Dummy encoding
Training	0.29	0.28	0.32
Validation	0.39	0.39	0.42
Test	0.37	0.31	0.33

Table 3 Mean squared error for each model form, arranged by subset of the data

Conclusion

Based on the results shown in ***Table 3***, model form 2 (improved label encoding) performs slightly better than model form 1 (label encoding) and model form 3 (dummy encoding) on the training and validation results. Model form 2 also has a clear meaning that is easy to communicate to stakeholders. Comparing the models' performance on the test dataset confirms that this was indeed the best choice, with a test MSE of 0.31. In all, this means that PA&E has successfully developed a model that predicts NNSA facility D&D costs to within +100%/-50% at the 70% confidence level. The model, which is based on 41 historic data points, uses three only high-level inputs: building size, contamination, and construction type.

Looking ahead, PA&E is actively soliciting additional D&D data from throughout the Department of Energy, which will be used to further develop and validate the model. In addition, PA&E is working on features that will benefit the user experience, such as a user interface that generates prediction intervals and a s-curve.

References

[1] US Department of Energy, "DOE O 430.1C Chg 2 (AdminChg), Real Property Asset Management," 17 September 2020. [Online]. Available: https://www.directives.doe.gov/directives-documents/400-series/0430.1-BOrder-c-chg2-adminchg. [Accessed 17 February 2022].

[2] AACE International, "Cost Estimate Classification System, Recommended Practice No. 18R-97," AACE International, Fairmont, WV, 2020.

[3] S. Dekker, "PARS II: Redefining Program Oversight & Assessment at the Department of Energy," in *ISPA/SCEA Joint Annual Conference and Training Workshop*, Albuquerque, NM, 2011.

[4] National Nuclear Security Administration, "NNSA's G2 Management Information System Wins Association for Enterprise Information's (AFEI) "Excellence in Enterprise Information Award"," 17 February 2016. [Online]. Available: https://www.energy.gov/nnsa/articles/nnsa-s-g2-management-information-system-wins-association-enterprise. [Accessed 17 February 2022].

[5] Department of Energy Environmental Management, "Environmental Cost Analysis System (ECAS)".

[6] DOE Office of Asset Management (MA-50), "FIMS Info," [Online]. Available: https://fims.doe.gov/fimsinfo/. [Accessed 8 December 2020].

[7] D. C. R. Loelius, C. Fitzpatrick, R. Strand and C. David E. Zimmerman, "Escalation Study for DOE NNSA's Capital Acquisition," in *2021 AACE International Conference & Expo*, Virtual, 2021.

[8] Gordian, "RSMeans," Greenville, SC, 2021.

[9] S. A. Book and N. Y. Lao, "Minimum-Percentage-Error Regression under Zero-Bias Constraints," *Proceedings of the Fourth Annual U.S. Army Conference on Applied Statistics, 21-23 October 1998,* no. November, pp. 47-56, 1999.

[10] S.-P. Hu and A. Smith, "Using Dummy Variables in CER Development," *Journal of Cost Analysis and Parametrics,* vol. 10, no. 1, pp. 76-90, October 2021.

[11] Department of Energy Office of Project Management Oversight and Assessments, "DOE Order 413.3B Chg 5 (MinChg), Program and Project Management for the Acquisition of Capital Assets," 2018.