14th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP14
Dublin, Ireland, July 9-13, 2023

Effect of Bridge Data Heterogeneity on Neural Network Survival Predictive Performance

Antti Valkonen
Graduate Student, Dept. of Civil and Environmental Engineering, Princeton University, Princeton, USA

Branko Glisic
Professor, Dept. of Civil and Environmental Engineering, Princeton University, Princeton, USA

ABSTRACT: Accurate deterioration models are required for data-based bridge management. These models allow converting raw data into actionable insights to guide maintenance decisions. Survival modeling is a technique of modeling time-to-event that has been found helpful for bridge deterioration modeling purposes. Neural network-based survival models have recently shown promise for use in the field. In this work, we investigate the effect of bridge population heterogeneity on the predictive performance of such models. We study this problem in the context of the National Bridge Inventory (NBI), a dataset containing inspection data of US highway bridges. There are many structural systems, materials, deck protection systems, and loading conditions in the NBI bridge population. We hypothesize that this type of heterogeneity will influence the model performance. To test our hypothesis, we split the data into subsets and compare model performance when fitted individually to subsets. In splitting the data, we utilize two separate approaches: statistical clustering and a physics-based approach, where we split the data based on understanding the underlying deterioration mechanisms. By comparing the models fitted to different subsets of data, we can study the effect of data heterogeneity on model performance. The results of this work help further understand the potential limitations the data places on the Neural Network survival model approach. We expect this understanding to improve further development of the modeling approach.

1. INTRODUCTION

Accurate prediction of the infrastructure condition is an essential task in asset management. The predictions are important not only for the engineers who deal with asset management but also for the general public because the models are used in creating government budgets. (“Strong Infrastructure and a Healthy Economy Require Federal Investment” 2019) In the United States, the importance of deterioration modeling of highway infrastructure has been codified in the law: state DOTs are required to “identify” deterioration models for pavements and bridges in the National Highway System.(“23 CFR § 515.7” n.d.)

In this work, we study one form of deterioration modeling for concrete bridge decks: Neural Network-based survival analysis. We have, in earlier work, shown the principle of this novel modeling method. (Valkonen and Glisic 2021) In this paper, we study the effect of data heterogeneity on the prediction accuracy of the NN survival model. This is important to understand because multiple types of bridges have fundamentally different functional principles. For effective utilization of the deterioration models, it is essential to understand the effect of this fundamental data characteristic.

2. THE DATASET USED IN THIS STUDY

In the United States, State Departments of Transportation (DOTs) maintain records of the physical conditions of highway bridges. (“23 CFR § 650.315” n.d.) The physical condition of bridges and their components are measured with Condition Ratings (CR), which have numerical values between 0 and 9, corresponding to descriptions in Table 1. (*Recording and coding guide for the structure inventory and appraisal of the nation's bridges* 1995) The National Bridge Inventory (NBI) is a dataset containing condition rating data from all US Highway bridges. The data is available online starting from the year 1992. ("National Bridge Inventory - Bridge Inspection - Safety - Bridges & Structures - Federal Highway Administration" n.d.)

*Table 1: NBI Condition Ratings(*Recording and coding guide for the structure inventory and appraisal of the nation's bridges *1995)*

Rating	**Condition description**
9	Excellent Condition
8	Very Good Condition
7	Good Condition
6	Satisfactory Condition
5	Fair Condition
4	Poor Condition
3	Serious Condition
2	Critical Condition
1	"Imminent" Failure Condition
0	Failed Condition – out of service – beyond corrective action

Survival analysis requires observations of the time a bridge component (in our case deck) spends in each condition rating before an inspector assigns a lower rating. This value is referred to as the *Time-In-Condition Rating* (TICR).(Fleischhacker Adam et al. 2020) The data from the NBI database does not contain TICR values, but they must be computed from the collection of yearly CR values. In addition, the data is missing information about a phenomenon known as censoring. Censoring refers to the situation where the observation period ends before the inspector decreases the condition rating for the bridge, leading to an imperfect record of the TICR.

In conclusion, to be useful for survival analysis, the NBI data must be processed to contain both TICR and flags for censored entries. We utilize a dataset generated by Fleischhacker et al.(Fleischhacker Adam et al. 2020). This dataset is based on the NBI and processed to contain the TICR values and censoring flags. The dataset is also filtered to contain only the covariates the authors have found to be meaningful to deterioration analysis. The collection of covariates included in this dataset, their abbreviations used in this paper, and their range of values are given in Table 2.

All variables presented in Table 2 are recorded in the NBI, except SeaDist (distance from seawater), which Fleischhacker et al. (Fleischhacker Adam et al. 2020) found to be an essential parameter and converted into two categories: "sea less than 3 km away" and "sea more than 3 km away" and the climatic regions, which is also their addition. Our work is focused on studying the model we have previously developed, and because we utilized this dataset in developing the model, we will use it as it is.

2.1. Computational Approach: Survival Analysis

We want to study the effect of data heterogeneity on neural network-based survival analysis. The goal of bridge survival analysis is to derive *survival curves* that describe an individual bridge's probability of *surviving* in each condition rating over some period. The survival function is given by:

$$S(t) = 1 - F(t). \tag{1}$$

where $f(t)$ is the probability density function (PDF) of the survival TICR, and $F(t)$ is the corresponding cumulative distribution function (CDF), The graph of this function is known as the survival curve.(Aalen et al. 2008) Hazard function is another key aspect of survival analysis. The hazard function can be given using the survival function and the PDF of survival time:

$$h(t) = \frac{f(t)}{S(t)}, \qquad (2)$$

The interpretation of the hazard function is that it gives the instantaneous event rate or the probability of having the event within "a short period of time.

We study the effect of data heterogeneity on survival modeling using a neural-network model designed for this purpose in our earlier work. The model is built on top of a Python library, "nnet-survival".(Gensheimer 2020) Figure 1 shows the architecture of our neural network. To be concise, we have given only a brief overview of the necessary concepts. Still, we encourage readers to see our earlier publication for details about the model and survival analysis. (Valkonen and Glisic 2021)

Table 2: Description of Covariates and their range of values (modified from (Fleischhacker Adam et al. 2020) Reference values bolded).

Description of Covariate	**Range of Values**
Average Daily Truck Traffic	[0,56595] ,**ref=100**
Climatic Region	"Region2 - very hot", "Region3 - hot", "Region4 - average", "**Region5 - cold** ", "Region6 - very cold", "Region7 - extremely cold", "Region8 - subarctic", "Region9 - average marine", "Region10 - hot marine"
Condition Rating	CR3,CR4,CR5,CR6, **CR7**,CR8,CR9
Deck Protection Type	**"None",** "Epoxy-coated reinforcing", "Galvanized reinforcing", "Other coated reinforcing", "Cathodic protection", "Polymer impregnated", "Internally sealed", "Unknown", "Other"
Deck Type	**"Concrete cast-in-place"**, "Concrete precast panels"
Distance to Sea Water	**"Sea Less than 3 km Away",** "Sea More Than 3 km Away"
Functional Classification (NBI Item 26)	**"Rural"**, "Urban"
Maintenance Responsibility	**"State highway agency"**, "County highway agency", "Town/township highway agency", "City/municipal highway agency", "Private (other than railroad)", "State toll authority"
Structural Type	"Concrete-simple span", "Concrete-continuous", **"Steel-simple span"**, "Steel-continuous", "Prestressed concrete-simple span", "Prestressed concrete-continuous"

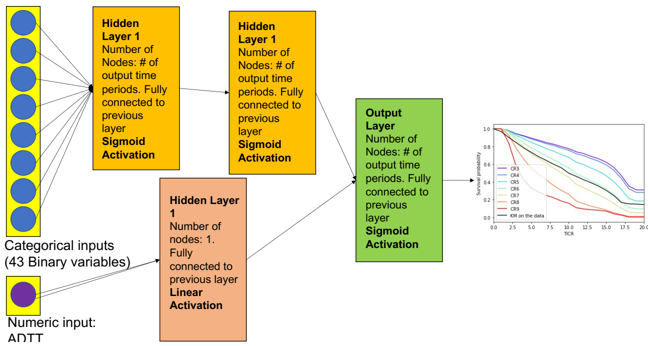

Figure 1: NN Architecture

3. EFFECT OF DATA HETEROGENEITY ON MODEL PERFORMANCE

We hypothesize that this type of heterogeneity will influence the model performance. To test our hypothesis, we split the data into subsets and compare model performance when fitted individually to subsets. In splitting the data, we utilize two separate approaches: statistical clustering and a physics-based approach, where we split the data based on understanding the underlying deterioration mechanisms. We will measure model predictive performance using Concordance Index (CI). CI is calculated through a pairwise ranking of observations: predicted *score* vs. which one is observed to fail first. The CI is the proportion of pairs where the model correctly predicted the ordering.(Harrell Jr. et al. 1996) The *score* can be any suitable measure of the risk of failure (or its inverse, propensity to survive). In evaluating model performance, we use CI calculated using the following scores:

- 1-hazard at TICR=4
- Survival function at TICR=4
- 1-hazard at TICR=9
- Survival function at TICR=9
- 1-hazard at TICR=14
- Survival function at TICR=14.

We compute the average of CIs resulting from these scores and assign this average as the performance metric for the models evaluated.

3.1. Statistical Clustering

Statistical clustering refers to a group of statistical/machine learning methods with the goal of grouping the data into *n* groups, or *clusters* as more commonly known. We utilize a method called *k-prototypes.*(Huang 1998) This method is suitable for our purposes because it allows clustering data with numerical and categorical variables. Another benefit is that the algorithm is relatively easy to understand, which we believe helps interpret results since we want to compare the results of machine learning clustering to the selection of clusters based on engineering grounds. Following (Huang 1998), we give a brief overview of the method.

The basic principle is that objects are placed into clusters based on their (dis)similarity with the cluster or more precisely the other objects in the cluster. The solution to the clustering problem minimizes aggregate dissimilarity over all clusters and objects. For the numeric variables, dissimilarity is measured by the Euclidean distance between the object and the cluster centroid. This method of using Euclidean distance is typically known as k-means clustering.(Tibshirani et al. 2009) For the categorical variables, the dissimilarity is measured by the following function:

$$d = (X, Y) = \sum_{j=1}^{m} \delta\left(x_j, y_j\right) \tag{3}$$

Where *X, Y* are vectors of *m* categorical variables and:

$$\delta\left(x_j, y_j\right) = \begin{cases} 0 \left(x_j = y_j\right) \\ 1 \left(x_j \neq y_j\right) \end{cases} \tag{4}$$

Similarly, as in the numerical variable case, the clustering is decided by minimizing aggregate dissimilarity when the objects are placed in the clusters. In the case of the numerical variables, the dissimilarity is calculated with respect to the centroid of the clusters. However, centroid does not exist for categorical variables, at least not in a conventional sense. Thus, a parallel to the centroid needs to be defined. In the context of the k-prototypes method, this is called the *mode*. The mode of a set of categorical objects $\{X_1, X_2, \dots, X_n\}$ is defined as the vector Q (of categorical variables) that minimizes:

$$\sum_{i=1}^{m} d(X_i, Q) \ , \tag{5}$$

where *d(X, Q)* defined as in (3). The clustering algorithm then consists of finding clusters that minimize the weighted sum of the Euclidean distance between objects in a cluster and the cluster centroids for the numerical variables, and the aggregate distance between the cluster objects and cluster modes for the categorical variables.

The outcome of the clustering analysis is the sets of objects for each cluster, and the cluster modes and centroids. Because we are interested in utilizing the method to study the model we developed earlier, we refrain from a detailed explanation of the algorithm. We have presented what we deemed necessary to interpret the clustering results for the purposes of this work: the general idea and the role of centroids and modes. The algorithm is implemented in the Python library "kmodes" that we utilize in this work.(de Vos 2023) Interested reader is encouraged to refer to the documentation of the library and the original publication of the clustering method. (Huang 1998; de Vos 2023)

4. PHYSICS-BASED CLUSTERING

The statistical clustering approach presented above is completely ignorant of the potential underlying engineering/physical rationale that could be used to group the data into clusters. To provide a point of comparison and further insight into the results of the statistical clustering, we split the data into clusters based on engineering judgment. As with the statistical clustering, we fit the survival model individually for each cluster to study the performance of the survival model with different clusters.

If the data heterogeneity influences the performance of the deterioration model, it is safe to assume that this is caused by differences in underlying deterioration mechanisms. Thus, to study the effect of data heterogeneity, we split the data into clusters according to factors that, in our judgment, affect the most common deterioration mechanisms in concrete bridge decks. Transportation research board publication "Nondestructive Testing to Identify Concrete Bridge Deck Deterioration" recognizes the four most important deterioration mechanisms of concrete bridge decks(Gucunski et al. 2012):

- Rebar corrosion
- Deck delamination
- Vertical cracking
- Concrete degradation.

Out of these, we want to study the effect of corrosion-related issues because our model uses covariates that are directly tied to corrosion: distance to seawater and the decks' corrosion protection measures.

5. RESULTS: STATISTICAL CLUSTERING

The clustering algorithm we have utilized requires selecting the number of clusters. This is not a straightforward task since we do not have any a priori knowledge about the number of clusters that could be expected to be found in the data. We utilize a practical approach: we use different numbers of clusters, and for each of the resulting sets of clusters, we fit the survival model. When fitting the model, we only utilize 80% of the data in the cluster at hand, leaving 20% for the computation of the CI. This way we are not testing the model's performance against data used to train it. We then evaluate the performance of each individual model by computing the CI metrics. After fitting the models and computing the CIs, we choose the number of clusters for further evaluation.

We evaluate options with 2, 3 , 4, 5, and 6 clusters by comparing the highest CI cluster of each set. The CI values are close to each other, but the number of clusters equals five yields the highest average CI on its best cluster. Hence, we choose the clustering with five clusters for further investigation. Even if this might not be the optimal way to select the number of clusters, it is good enough reasoning for us at this stage – in theory we could increase the number of clusters arbitrarily. Still, the time expenditure would be prohibitive with no guaranteed improvements. Table 3 below shows the cluster sizes and CI metrics for the selected clustering.

Table 3: Model Performance with Number of Clusters=5

CLUSTER NUMBER	CLUSTER SIZE	AVERAGE CI
0	35774	68.9%
1	2905	61.6%
2	81520	67.9%
3	11990	62.7%
4	538597	67.5%
	Maximum CI	68.9%

We study the result of clustering into five clusters by studying the compositions of the cluster in terms of the values of covariates in each cluster. We have omitted commentary on the covariates for which no significant differences could be easily observed between the clusters. Figures 2-5 show how the covariate values are distributed among the clusters. Inspecting the composition of the clusters, two observations can be made:

1) Figure 3: The Average Daily Truck traffic levels differentiate the clusters.
2) Figures 2,4, and 5: Cluster 4 Differs from the others more significantly.

The first observation is somewhat expected. The level of truck traffic is a numerical variable, and it is easy to imagine the bridges would fall in clear clusters of different average traffic levels based on their importance. The algorithm would easily recognize these clusters. However, comparing Figure 3 to Table 7 yields an interesting observation: the predictive performance of the survival model is weakest in the two clusters with the highest ADTT levels. The reason for this is unclear, considering such a correlation does not hold with the other three clusters. A more detailed study of this phenomenon could yield improvements in the accuracy of deterioration models. However, this is out of scope of this work.

The second observation shows that the largest of the clusters is also the most different. It has very low traffic levels, a more significant proportion of rural bridges than the others, a larger proportion of precast panels than the others, more observations from bridges maintained by county highway agencies, and a smaller proportion of roads maintained by toll road authorities. Because this cluster contains bridges that are in rural areas, have low traffic levels, and are maintained by smaller entities, we believe this cluster represents "non-critical" bridges. The model did well on this cluster, suggesting that this type of deterioration model is suitable for optimizing the maintenance of the large stock of "non-critical" bridges.

6. RESULTS: PHYSICS-BASED CLUSTERING

We split the data two ways, with respect to the distance to seawater, we have two clusters, sea closer than 3km and sea farther than 3km. With respect to corrosion protection, we create two clusters, one with decks with no protection and the other one with decks that have a specified protection measure in place. The data contains datapoints where the protection is unspecified; we leave those out. For these clusters, we fit the model using 80% of the data and compute the CI indices with the remaining 20%. The results are presented below in Table 4.

Figure 2: Functional Class composition of clusters.

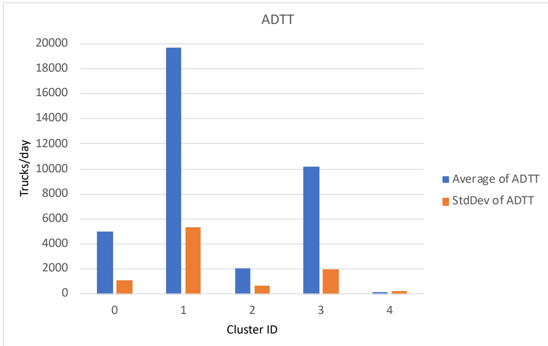

Figure 3: ADTT levels of clusters.

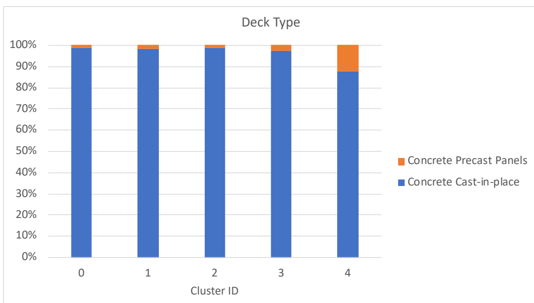

Figure 4: Deck Type composition of clusters.

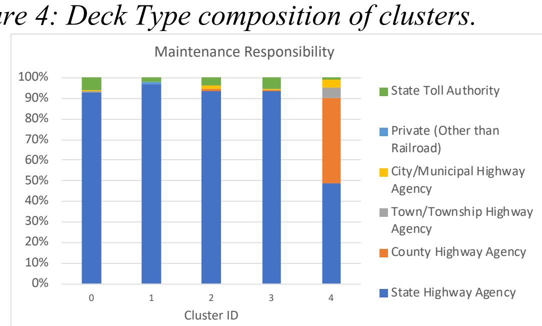

Figure 5: Maintenance responsibility composition of clusters

Table 4: Model Performance with different physics-based clusters.

Distance to seawater	Cluster Size	Average CI
Sea within 3km	6185	60.5%
Sea > 3km away	664601	67.6%
Deck Protection	**Cluster Size**	**Average CI**
None	461411	66.0%
Known protection method	107514	69.0%

From this, we can make a few interesting observations. First, compared to the results from the statistical clustering, the overall performance and the difference between clusters are similar., however. Second, and perhaps more interesting, it seems that corrosion plays some role in the predictive performance of the model, an effect that was not picked up by the statistical clustering approach. Regarding the distance to seawater, the model performs better with a cluster with a distance to seawater larger than 3km, meaning the corrosion-accelerating effect of seawater should be less. Similarly, for deck protection, the model performs better with the cluster of corrosion-protected decks. We do not have an explanation, but it seems that in more corrosion-prone situations the predictive power suffers. Considering the importance of corrosion for concrete structures, this could be an important avenue for further research. Investigating this further could improve degradation modeling significantly.

7. CONCLUSION

We studied the effect of bridge deck population heterogeneity on the prediction performance of a Neural Network based survival model. We used a statistical clustering approach and a physics-based clustering approach. Although both gave numerically similar results, there was some performance difference between the model with different clusters. However, the results from the physics-based approach imply that more aggressive corrosion creates difficulty for the model. This result has two important implications. First, the statistical approach did not pick this up, so the result shows the importance of considering the physical deterioration characteristics when designing statistical models. Second, the result suggests further research into why more aggressive corrosion interferes with deterioration prediction. Solving this could drastically increase deterioration modeling performance, considering the importance of corrosion in the deterioration of concrete structures.

8. ACKNOWLEDGMENTS

This material is, in part, based upon work supported by USDOT's Office of Research, Development & Technology Programs and Activities, #CAIT-UTC-REG 63. Any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the funding agency.

9. REFERENCES

"23 CFR § 515.7." n.d. *LII / Legal Information Institute*. Accessed December 23, 2020. https://www.law.cornell.edu/cfr/text/23/515.7.

"23 CFR § 650.315." n.d. Accessed November 12, 2020. https%3A%2F%2Fwww.govinfo.gov%2Fapp%2Fdetails%2FCFR-2020-title23-vol1%2FCFR-2020-title23-vol1-sec650-315.

Aalen, O., O. Borgan, and H. Gjessing. 2008. *Survival and Event History Analysis: A Process Point of View*. New York, NY: Springer.

Fleischhacker Adam, Ghonima Omar, and Schumacher Thomas. 2020. "Bayesian Survival Analysis for US Concrete Highway Bridge Decks." *Journal of Infrastructure Systems*, 26 (1): 04020001. American Society of Civil Engineers. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000511.

Gensheimer, M. 2020. "MGensheimer/nnet-survival." Python.

Gucunski, N., A. Imani, F. Romero, S. Nazarian, D. Yuan, H. Wiggenhauser, P. Shokouhi, A. Taffe, and D. Kutrubes. 2012. *Nondestructive Testing to Identify Concrete Bridge Deck Deterioration*. Washington, D.C.: Transportation Research Board.

Harrell Jr., F. E., K. L. Lee, and D. B. Mark. 1996. "Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors." *Statistics in Medicine*, 15 (4): 361–387. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

Huang, Z. 1998. "Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values." *Data Mining and Knowledge Discovery*, 2 (3): 283–304. New York, Netherlands: Springer Nature B.V. https://doi.org/10.1023/A:1009769707641.

"National Bridge Inventory - Bridge Inspection - Safety - Bridges & Structures - Federal Highway Administration." n.d. Accessed January 31, 2021. https://www.fhwa.dot.gov/bridge/nbi.cfm.

Recording and coding guide for the structure inventory and appraisal of the nation's bridges. 1995. Washington, D.C.: U.S. Dept. of Transportation, Federal Highway Administration.

"Strong Infrastructure and a Healthy Economy Require Federal Investment." 2019. *House Budget Committee Democrats*. Accessed March 30, 2020. https://budget.house.gov/publications/report/strong-infrastructure-and-healthy-economy-require-federal-investment.

Tibshirani, R., J. H. Friedman, and T. Hastie. 2009. *The elements of statistical learning data mining, inference, and prediction*. Springer series in statistics. New York: Springer.

Valkonen, A., and B. Glisic. 2021. "Neural network survival analysis of concrete bridge decks." *International Conference on Structural Health Monitoring of Intelligent Infrastructure: Transferring Research into Practice, SHMII*, 1635–1640.

de Vos, N. 2023. "kmodes categorical clustering library." Python.