ASCE

New Measure to Understand and Compare Bridge Conditions Based on Inspections Time-Series Data

Akshay Kale[1]; Brian Ricks, Ph.D.[2]; and Robin Gandhi, Ph.D.[3]

Abstract: The C+ score for US bridges on the 2017 infrastructure report card underscores the need for improved data-driven methods to understand bridge performance. There is a lot of interest and prior work in using inspection records to determine bridge health scores. However, aggregating, cleaning, and analyzing bridge inspection records from all states and all past years is a challenging task, limiting the access and reproducibility of findings. This research introduces a new score computed using inspection records from the National Bridge Inventory (NBI) data set. Differences between the time series of condition ratings for a bridge and a time series of average national condition ratings by age are used to develop a health score for that bridge. This baseline difference score complements NBI condition ratings in further understanding a bridge's performance over time. Moreover, the role of bridge attributes and environmental factors can be analyzed using the score. Such analysis shows that bridge material type has the highest association with the baseline difference score, followed by snowfall and maintenance. This research also makes a methodological contribution by outlining a data-driven approach to repeatable and scalable analysis of the NBI data set. **DOI: 10.1061/(ASCE)IS.1943-555X.0000633.**

Introduction

Highways and bridges are cornerstones of the US transportation system. This infrastructure is essential for commercial and economic activity because it supports the primary mode of transportation worldwide. According to the 2017 infrastructure report, there are over 600,000 bridges in the US (ARTBA 2017). The average age of 4 out of 10 bridges is 50 years. Approximately 39% of the bridges will soon require rehabilitation since most of them were designed for a life span of around 50 years. By 2017, about 9.1% of the nation's bridges had been designated as structurally deficient (ARTBA 2017). These statistics highlight the urgent need for innovative solutions to understand and manage bridge health. National Bridge Inventory (NBI) inspection records quantify bridge deck, substructure, and superstructure health using condition ratings (0–9), providing a rich data set that is often used to analyze bridge performance over time.

Bridge deterioration is a subject of tremendous interest to bridge engineers. Understanding factors that influence bridge health will provide insights to key stakeholders. For example, such an understanding can inform designers and engineers about building bridges that undergo deterioration and help maintainers employ smarter strategies to monitor factors that influence bridge health. Researchers have proposed several methods for evaluating the present condition of bridges, identifying influential factors for bridge condition and predicting bridges' future condition using deterioration models (Nasrollahi and Washer 2015). These methods often rely on deterministic, stochastic, and artificial intelligence (AI) models. A deterministic model is typically built using techniques like straight-line extrapolation and regression (Morcous et al. 2002). These techniques are used to evaluate the deterioration of a bridge by computing the time spent at a particular condition rating (Saeed et al. 2017) or computing hazard ratios for the pace of bridge deterioration (Wettach-Glosser et al. 2020). Stochastic models, built using Markov methods, are used to analyze bridge health by computing the probability of transition from one condition rating to another (Madanat et al. 1995; DeStefano and Grivas 1998). AI models are built using techniques such as artificial neural networks (ANNs), decision trees, and case-based reasoning (Madanat et al. 1995). AI models are effective at sheding light on the interactive effect among factors that affect bridge health (Morcous et al. 2002). In summary, methods that rely on deterministic, stochastic, or AI models add a unique perspective in understanding bridge health. However, certain aspects of these computational models can influence the study of factors that affect bridge health. One such aspect is how age affects bridge performance. The rate of bridge deterioration is observed to be different at different ages (Huang et al. 2009). Yet many recent approaches to understanding bridge health do not account for this difference when developing stochastic models based on Markov methods (Chang et al. 2019; Saeed et al. 2017; Assaad and El-adaway 2020).

Next, factors that affect bridge health usually interact with each other (Barreto and Howland 2006). These factors may also vary with region and type of bridge (Saeed et al. 2017; Huang et al. 2009). In addition, the study objectives may influence the selection of factors being studied or the amount of data available in the study of these factors. As a result, prior studies provide inadequate or fragmented evidence to decide which factors are influential for bridge health.

We propose a new measure to understand and compare bridge conditions based on inspections of time-series data. This measure accounts for the effect of age on the rate of deterioration. It does this by aggregating the differences in the condition rating of a bridge with respect to a baseline of condition ratings at each age. The baseline is an average of condition ratings computed using all available annual NBI inspection records since 1992, from all states and for bridge records available at every age value.

[1]Ph.D. Student, Univ. of Nebraska at Omaha, PKI 362, 6001 Dodge St., Omaha, NE (corresponding author). Email: akale@unomaha.edu

[2]Assistant Professor, Univ. of Nebraska at Omaha, PKI 174H, 6001 Dodge St., Omaha, NE. Email: bricks@unomaha.edu

[3]Professor, Univ. of Nebraska at Omaha, PKI 177A, 6001 Dodge St., Omaha, NE. Email: rgandhi@unomaha.edu

Note. This manuscript was submitted on March 3, 2020; approved on March 26, 2021No Epub Date. Discussion period open until 0, 0; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Infrastructure Systems***, © ASCE, ISSN 1076-0342.

There are three main contributions of this new approach. First, our algorithm, which is based on readily available condition ratings, provides time-series results, capturing how the condition of an individual bridge changes over time. Second, our algorithm is not geographically bound and allows for the analysis of bridges within a region and across the US. Third, our method assigns a single health score for each bridge in the NBI. This score, which we call the baseline difference score (BDS), follows a normal distribution. Hence, using statistical methods, we can examine what factors best explain its variance. Our analysis demonstrates the use of bridge inventory, environment, and regional factors obtained from the NBI and multiple external data sources. The findings are based on inspection records of over 600,000 bridges in the United States reported every year since 1992 in the NBI data set. Between 1992 and 2017, this amounts to over 17 million inspection records.

In this study, we cleaned and transformed the NBI data extracted from the Federal Highway Administration (FHWA) and integrated external data sources such as Centers for Disease Control and Prevention (CDC) data and weather data. Also, we tested commonly cited factors and unexplored factors as determinants of the bridge condition. Our results were counterintuitive regarding average daily traffic (ADT) and average daily truck traffic (ADTT). We observed a relatively low association of ADT and ADTT with bridge performance, a finding that is surprising in light of several prior studies (Chang et al. 2019; Assaad and El-adaway 2020; Kim and Yoon 2010; Hasan and Elwakil 2020; Nasrollahi and Washer 2015). Lastly, we made our research reproducible by adopting a cloud-based work environment using DEEDS (datacenterhub.org; Smarti data set).

Prior Work

There are several challenges in modeling bridge health. Prominent challenges include incomplete bridge survey records, lack of reporting of improved bridge condition ratings, and identifying interactive effects among influential factors.

Incomplete records of bridges are one of the biggest challenges in modeling bridge health. Survey records of bridges are available only from 1992 to 2019; bridges built before 1992 have no prior records. To address the problem of incomplete and missing bridge condition records, Nasrollahi and Washer (2015) devised a method to estimate bridge condition ratings. The method computes the amount of time a bridge has spent in a particular condition rating using 20 years of bridge survey records. This method of measuring bridge condition ratings is also known as the time in condition rating (TICR). Fleischhacker et al. (2020) extended the use of TICR to develop a scalable method, which was applied to study the health of over 150,000 bridge decks nationwide. Under this method, TICR is used with Bayesian analysis to predict future health conditions. Similar to TICR, the hazard ratio (HR) was developed by Wettach-Glosser et al. (2020), which describes a bridge's probability of staying in a particular condition. An increase in HR greater than one indicates an increase in the likelihood of bridge condition rating decrease (deterioration). In contrast, a decrease in HR to less than one indicates a reduction in the likelihood of a bridge condition rating decrease. Wettach-Glosser et al. (2020) applied this method to study the performance of 5,242 bridge decks in Oregon.

In addition to the lack of complete records, the condition rating of bridges can be subjective because bridge inspectors manually inspect them. Often, the condition rating of bridges may increase or decrease over a period of time. However, it is difficult to understand whether improvements in bridges' condition are due to a subjective rating or intervention such as repair, reconstruction, or rehabilitation. Accounting for maintenance activities in bridge health is considered crucial because the absence of maintenance activity in modeling bridge health may not provide a bridge condition's true deterioration rate. Often, missing reports on a bridge's intervention history in the survey records can further exacerbate the problem. Researchers have attempted to mitigate bias caused by missing reports on bridge intervention through analyzing a nonincreasing subset of bridge records. Bolukbasi et al. (2004) studied records from 1976–1998 on 2,601 bridges in Illinois. The author proposed two different methods to account for unrecorded repairs and replacement: assuming that consecutive inspections should not increase and taking the duration between the consecutive inspections as a means to construct deterioration curves. Other researchers, such as Saeed et al. (2017), addressed unrecorded repairs and replacement by introducing the Bridge Intervention Matrix (BIM), which maps transitions in condition ratings to interventions that might have caused it. The BIM takes into account improvements in consecutive condition ratings. We use BIM in this study to observe maintenance events.

Various factors may contribute to bridge health; hence, it is essential to consider how factors interact to model bridge health (Morcous et al. 2002). Kim and Yoon (2010) applied a deterministic approach to identify latent variables that affect bridges' deterioration by studying 5,289 bridge records from 2006 and 2007 in North Dakota. Saeed et al. (2017) studied 5,600 Indiana bridges from 1992 to 2014 to account for the interactive effect among factors to model bridge deterioration; the study included a marginal effect analysis to identify influential factors in bridge health, while other researchers have used stochastic models, such as the Markov decision process (MDP) (Scherer and Glagola 1994; Saeed et al. 2017; Cesare et al. 1992; DeStefano and Grivas 1998). These probabilistic models are effective at accounting for the latent relationship among factors. In 1995, Madanat et al. (1995) studied 5,700 bridges from Indiana to develop a MDP model for predicting the future condition of bridges. DeStefano and Grivas (1998) further developed MDP methods and updated transition probabilities to account for previous bridge conditions.

In our analysis of prior literature, we observed the use of deterministic, stochastic methods and AI methods to understand bridge health. Chang et al. (2019) implement a combination of Markov chain stochastic and deterministic modeling of bridge condition using logistic regression. They also apply a linear regression model for identifying influential factors that affect bridges' condition in Wyoming. Researchers have also adopted AI techniques to predict future bridge health. For instance, Huang (2010) implemented ANNs using 942 observations from the Wisconsin Pontis Bridge Management System (BMS) data set to predict Wisconsin bridges' future deterioration. Case-based reasoning (CBR) is an AI technique to predict future bridge health. Morcous et al. (2002) built a CBR model using 521 observations of bridges obtained from the Canadian province of Quebec to predict the future condition of bridges. Kim and Yoon (2010) conducted a GIS analysis to predict the future condition of bridges. Chang et al. (2019) also implemented classification trees to model bridge health and identify influential factors. Assaad and El-adaway (2020) developed ANN models and k-nearest neighbor models by studying 19,269 bridges in Missouri. The authors report 91.44% accuracy in predicting the future condition of bridges. Other research efforts to predict future bridge conditions include simulation techniques. Hasan and Elwakil (2020) modeled the impact of influential factors on bridge decks using Monte Carlo simulations based on 1992–2018 California bridges. The simulation of bridge decks provides a probability of future condition ratings of decks.

In the development of various methods to model bridge health, researchers identified various influential factors. We use the general categories of factors by Kim and Yoon (2010) to report influential factors (Table 1). Our literature review revealed that some of the most commonly reported influential variables were traffic (Chang et al. 2019; Assaad and El-adaway 2020; Kim and Yoon 2010; Hasan and Elwakil 2020; Nasrollahi and Washer 2015), age (Chang et al. 2019; Saeed et al. 2017; Assaad and El-adaway 2020), structure length (Chang et al. 2019; Saeed et al. 2017; Assaad and El-adaway 2020; Hasan and Elwakil 2020), and interventions such as replacement (Saeed et al. 2017; Kim and Yoon 2010). Other than a commonly reported influential factors, Morcous (2011) found highway class and region to be the most influential factors in understanding Quebec bridges' future condition. Chang et al. (2019) found that physical attributes, such as the functional classification of routes and lanes on structures in records from the period 1992–2014, were the most influential factors. Assaad and El-adaway (2020) found that deck width, number of spans, operating rating, inventory rating, and structural evaluation as influential factors in modeling bridge condition. Kim and Yoon (2010) found that materials, such as concrete, prestressed concrete, and steel, and structural systems, such as slab and temperature, were helpful in understanding bridges' condition. Hasan and Elwakil (2020) found that inspection frequency, degree of skew, structure length were influential factors in simulating bridge condition ratings. In general, we observed that consideration of environmental factors in prior literature was limited. This may be due to a lack of data integration.

Table 1. Categorization of influential factors

Category	Factor	Description
Physical	Design load	Design load
	No. of spans	No. of spans in main unit
	Width	Width of deck
	Year built/age	Year of construction
	ADT/ADTT	Average daily volume of vehicles
	Replacement	Widening, replacement, and rehabilitation of bridge
	Replaced length	Length of structure improvement
	Deck/substructure/superstructure condition	Condition ratings of deck, substructure, and superstructure
Region	State	State
	County	County
	US Census region	Region that includes states, defined by US Census
Structural-material type	Material type	Kind of material used such as concrete, steel, wood, or timber
	Structure type	Kind of structural system of bridge such as slab, girder/beam, and truss
Environmental	Precipitation	Annual mean precipitation in inches or millimeters
	Snowfall	Annual mean snowfall
	Temperature	Mean temperature
	Over water	Bridge over waterway
Service	Highway	Bridge service on highway
	Railroad	Bridge service on railroad
	Interstate	Bridge is part of Interstate
	Toll	Toll bridge
	Average farm size	Average farm size of size of census track where bridge located
	Population	Population of census track where bridge is located
	Maintenance	Maintenance responsibility of bridges

Several studies do rely on augmenting the NBI data set with additional data collected by their respective state DOT (Bolukbasi et al. 2004; Saeed et al. 2017; Nasrollahi and Washer 2015; Kim and Yoon 2010; Morcous 2011).

In reviewing the literature, we found that the data used are often limited to a single state or region. There is a lack of insight in understanding maintainers' role in bridge health. Such studies provide both limited information on other states' bridges (Bolukbasi et al. 2004; Nasrollahi and Washer 2015; Saeed et al. 2017) and are conducted on relatively small bridges compared to the full NBI data set.

Data

Our research method can be divided into six phases: data collection, data cleaning, data transformation, factor selection, BDS computation, and identifying significant factors. These phases are summarized as a flowchart in Fig. 1.

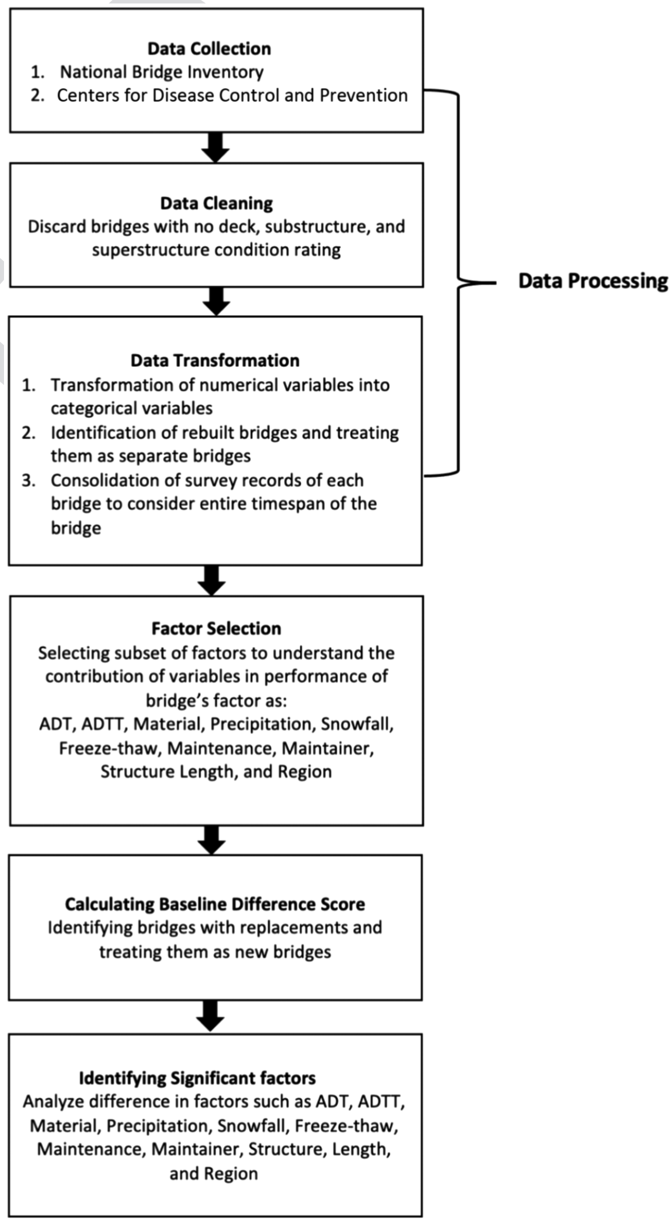

Fig. 1. Research methodology of study.

Data Collection

To accomplish the goals of this research, we identified three data sets: the NBI, snowfall, and freeze-thaw data from the National Oceanic and Atmospheric Administration (NOAA) (supplied by FHWA) and precipitation data from the CDC. A script is used to download the NBI data set from the FHWA website for all 52 states over the period 1992–2017. This includes over 17 million inspection records for more than 600,000 bridges in the US. The snowfall and freeze-thaw data were supplied by LTBP InfoBridge, a centralized national repository for bridge information. The precipitation data downloaded from the CDC span the years 1992–2011.

Data Cleaning

Data cleaning is essential to assure reliable analysis in the later stages of this research study. We created a clean, integrated data set consisting of information about bridges from the NBI and weather data from the CDC. We now describe the data cleaning and transformation strategy applicable to three data sets identified in the data collection step.

In the NBI data set, we observed errors, missing data, and inconsistencies in complying with the FHWA coding guide (FHWA 1995). As expected from any large data collection effort, the NBI inspection records aggregated from different state DOTs by the FHWA face many issues. In particular, the early years of NBI-related data collection has a large percentage of data that is incomplete or nonconformant to the coding guide. For example, we encountered missing condition ratings, incorrect longitude and latitude, unrecorded reconstruction, and rehabilitation dates. These missing data, if left unaddressed, affect the reliability of any findings reported from data analysis. For example, when a bridge is repaired, reconstructed, or rehabilitated, its condition improves. Therefore, missing information makes it difficult to account for reconstruction and rehabilitation as a factor that can improve bridge condition.

In addition to condition ratings, we also examined other fields, such as latitude, longitude, and structure type associated with bridge inspection records in the NBI. For these fields, we addressed several data-related inconsistencies as follows:

- Discard inspection records with missing values for deck, substructure, and superstructure.
- Convert longitude and latitude to degrees and minutes.
- Detect changes in year-built record for a bridge over different inspection records in the NBI. The change is assumed to be a bridge replacement. For every change, reintroduce the bridge with a new identifier (original structure number with a segment suffix) to preserve bridge life-cycle consistency.
- Maintain a log of invalid values and repeated records. A log promotes transparency of any updates to the original data for the purpose of cleaning.

Our observations regarding missing records are summarized by region in Table 2. These statistics suggest that the Northeast region has the highest number of missing records, followed by the West and South. The Midwest region has the highest number of clean records. On average, 60% of the records are missing.

While a complete listing of the number of missing records per state is available in the Appendix, we only discuss the summary statistics. Midwestern states, Nebraska (73%), and Iowa (77%) have the highest number of records available. In contrast, Northeastern states, such as Connecticut (4%), New Hampshire (11%), and Massachusetts (14%), have the smallest number of survey records for analysis. Since each bridge may have multiple records, the number of inspection records considered in the study is higher than the number of bridges. As a result, Table 2 gives the percentage of missing inspection records, not the number of missing bridges.

Table 2. Summary of missing records

Data processing summary			
Region	Total records	No. of studied records	Percentage of studied records(%)
Northeast	1,913,728	606,759	31.7
Midwest	7,124,522	3,683,689	51.7
South	6,177,021	2,648,231	42.8
West	2,859,731	1,115,204	38.9

Data Transformation

In the data cleaning phase, we observed that the year built field had some inconsistencies. For example, bridges may appear unnaturally aged or newly built. Based on our discussion with subject matter experts from the Nebraska DOT, we learned that when an inspection record shows a more recent year built than previous inspection records, it often means the bridge was rebuilt. To account for this inconsistency, we split a bridge's condition rating time series into segments that share the same year built. Bridge segments are identified by adding a segment number as a suffix to the bridge structure number. Each segment is then tracked separately in the data set and associated with its own BDS.

Factor Selection and Preparation

Factor selection was driven by observed popularity in prior literature. To avoid bias, in addition to frequently tested factors in the literature, we also included factors not commonly associated with bridge condition ratings as well as environment and geographic factors. The full list of factors considered in our research study includes ADT, ADTT, maintainer, maintenance, material, structure length, precipitation, snowfall, freeze-thaw, and region.

We used ANOVA to examine the association of each factor with the BDS. To perform this analysis, we converted all independent factors into categorical variables. The conversions were informed by methods from earlier literature and the NBI coding guide. For instance, Morcous (2011) provides a guide for converting ADT into very light, light, medium, and very heavy. For factors such as precipitation, snowfall, and freeze-thaw, we created two categories after removing outliers: the largest 10% of values and the lowest 10% of values.

To check for normality, the Shapiro-Wilk test was performed on small and large samples for each factor level. While several methods exist for determining the appropriate sample size, there is limited guidance in the literature related to appropriate sample sizes for NBI data sets. For small-sample analysis, statisticians consider 100 to be considered a good minimum sample size. For large sample analysis, factors are downsampled with a random selection such that there is equal representation from all factor levels. Thus, a large sample size is determined by the factor level with the smallest number of observations.

The results of normality tests for all factor levels are reported in Tables 3–5. For large samples, the tests reject the null hypothesis that the samples came from a normally distributed population, with 95% confidence for all factors. On the other hand, all small-sample tests fail to reject the null hypothesis for all factors with 95% confidence. With large sample sizes, it is expected that the Shapiro-Wilk test will detect even trivial departures from the null hypothesis. This phenomenon explains the divergent results between

Table 3. Factor division criteria and normality test results

Summary of data set				
Factors	Sample size small, large	Criteria	Category	Normality p-value
ADT	100, 112K	ADT < 100	Very light	0.5, 0.0
(NBI Item 27)	100, 112K	100 ≤ ADT ≤ 1,000	Light	0.3, 0.0
Unit: Volume of vehicles	100, 112K	1,000 ≤ ADT ≤ 5,000	Moderate	0.4, 0.0
	100, 112K	ADT > 5,000	Heavy	0.4, 0.0
ADTT	100, 112K	ADTT < 100	Light	0.2, 0.0
(NBI Item 27)	100, 112K	100 ≤ ADTT ≤ 500	Moderate	0.2, 0.0
Unit: volume of vehicles	100, 112K	ADTT > 500	Heavy	0.2, 0.0
Maintainer	100, 43K	NBI Item: 21 = 02	County	0.3, 0.0
(NBI Item 21)	100, 43K	NBI Item: 21 = 01	State	0.3, 0.0
	100, 43K	NBI Item: 21 = 03	Town	0.3, 0.0
	100, 43K	NBI Item: 21 = 04	Municipal	
And city	0.3, 0.0			
Material	100, 105K	NBI Item: 43A = 1	Concrete	0.2, 0.0
(NBI Item 43A)	100, 105K	NBI Item: 43A = 3	Steel	0.2, 0.0
	100, 105K	NBI Item: 43A = 5	Prestressed	0.2, 0.0
	100, 105K	NBI Item: 43A = 7	Wood	0.1, 0.0
Material (age constant)	100, 8.6K	NBI Item: 43A = 1	Concrete	0.3, 0.0
(NBI Item 43A)	100, 8.6K	NBI Item: 43A = 3	Steel	0.2, 0.0
	100, 8.6K	NBI Item: 43A = 5	Prestressed	0.2, 0.0
	100, 8.6K	NBI Item: 43A = 7	Wood	0.1, 0.0

Table 4. Continued. Factors division criteria and normality test results

Summary of data set				
Factors	Sample size small, large	Criteria	Category	Normality *p*-value
Structure length	100, 45K	6.1–8.5 (m)	Very short	0.2, 0.0
(NBI Item 49)	100, 45K	58.9–38421 (m)	Very large	0.2, 0.0
Precipitation	100, 56K	0.43–1.23 (mm)	Very low	0.1, 0.0
(CDC)	100, 56K	3.76–7.64 (mm)	Very high	0.1, 0.0
Unit: Mean daily precipitation				
Freeze-thaw	100, 56K	47–60	Very low	0.1, 0.0
(NOAA)	100, 56K	93–95	Very high	0.1, 0.0
Unit: Mean No. of freeze-thaw days				
Snowfall	100, 56K	12–14.3	Very low	0.2, 0.0
(NOAA)	100, 56K	61.8–66	Very high	0.2, 0.0
Unit: Mean No. of snowfall days				
Maintenance	100, 27K	0	Very low	0.2, 0.0
Unit: No. of intervention (repair, reconstruction, and reconstruction)	100, 27K	3–5	Very high	0.2, 0.0
Region	100, 105K	See Table 5	West	0.1, 0.0
(Derived from NBI Item 1)	100, 105K	See Table 5	South	0.2, 0.0
	100, 105K	See Table 5	Northwest	0.2, 0.0
	100, 105K	See Table 5	Midwest	0.3, 0.0

Table 5. Grouping of bridges into regions with respect to each bridge's state

Region	States
Northeast	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont, New Jersey, New York, and Pennsylvania
Midwest	Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota
South	Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, Washington, DC, West Virginia, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma, and Texas
West	Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Washington, Utah, Wyoming, Alaska, California, Hawaii, and Oregon

small- and large-sample normality (Anderson et al. 2000). In addition to sample sizes, we must also ensure that each sample is of a consistent quality with respect to factors such as maintenance. In the context of NBI records, we have observed that the time series of condition ratings are of varying lengths depending on the age of the bridge. Thus, a bias may arise if we select bridge samples with only a few or with many condition ratings in their time series. For example, a divergence in the length of the time series may occur for newly built bridges or old bridges.

With varying time-series lengths, factors such as the number of maintenance events can be misleading—bridges with fewer records will have fewer maintenance events. Thus, to analyze maintenance, we only select bridge samples with lengths of their time series between the first and third quartile of all possible lengths. Our resulting sample contains bridges with a minimum of 13 and a maximum of 26 observations in their condition rating time series. To derive the number of maintenance events, we then use the BIM (Saeed et al. 2017). It determines a maintenance event by observing an increase in the condition of a bridge. Finally, the samples are categorized into the top 10% of bridges with maintenance and bottom 10% of bridges with maintenance to observe the difference in the means of their BDSs using ANOVA. This process of sampling bridges to study the number of maintenance events does not apply to other factors, such as region, precipitation, structure length, and material, because these factors do not depend on the length of a bridge's condition rating time series.

Methodology

In this section, we describe the computation of the BDS and provide a mathematical formulation. Further, we analyze the BDS of bridges nationally to understand the method's characteristics, strengths, and limitations.

New Measure for Bridge Health

The NBI data set is updated annually with the condition ratings for bridge decks, substructures, and superstructures. Among bridge components, the superstructure is considered the backbone of a bridge (Nasrollahi and Washer 2015). Our analysis shows that the superstructure condition ratings are highly correlated to deck and substructure. Since computing a condition score using deck and substructure condition ratings would produce similar statistical results, in the rest of the paper, we only consider the superstructure condition rating for the purposes of our study.

We now outline the method for computing the BDS. The first step in the method is to compute a baseline of condition ratings, which considers the time series of bridge condition ratings from all states in the US. This baseline is the national average of bridge condition ratings, calculated using the superstructure condition rating, from ages 1 to 60. As a result, the computed baseline is a vector of length 60. We choose to end our analysis at age 60 since bridges are built for a 50-year life span. To compute the baseline, we compute the average of the condition ratings of all the available bridges at age one. Next, we compute all bridges' average condition ratings at age two; this process is continued until age 60.

Once the baseline is computed, we compute the BDS for each bridge. The condition ratings of bridges are compared against the baseline with respect to age. For instance, a bridge that has condition ratings from ages 10 to 25 is compared with the average condition rating from ages 10 to 25 in the baseline. Finally, we compute the BDS by averaging the difference between the condition ratings and baseline. This process of BDS computation is carried out for all available bridges. The magnitude of the BDS of a bridge indicates bridge performance compared to the average bridge; an increase in BDS indicates higher performance, whereas a decrease in BDS indicates lower bridge performance.

In the following subsection, we present the mathematical formulation and algorithm to compute the baseline difference score. Table 6 provides a description of the notation used in explaining the formulation.

Table 6. Description of notation

Notation	Description
A	Set of bridges ages $a \in \boldsymbol{A}$
B	Set of bridges b at age $b \in \boldsymbol{B}$
bds_b	Condition rating of bridge $b \in \boldsymbol{B}$
$\boldsymbol{C}$	Set of possible condition ratings
C_{ba}	Condition ratings of bridge b at age a
$\boldsymbol{S}$	Set of expected condition ratings for all bridges at all ages
s_a	Expected condition rating at age a
$\boldsymbol{Q}_b$	Set of all ages for which $b \in \boldsymbol{B}$
$\boldsymbol{X}_a$	List of all condition ratings of bridge at age a
X_{ai}	An ith element of sequence X_a

Formulation of Baseline Difference Score

Let a bridge be $b \in \boldsymbol{B}$, where $\boldsymbol{B}$ is the set of bridges in the NBI.

Let a bridge age be $a \in \boldsymbol{A}$, where $\boldsymbol{A}$ is the set of bridge ages in the NBI.

Let the set of possible condition ratings be $\boldsymbol{C}$, $\boldsymbol{C} \subset \mathbb{Z} = [0, 9]$.

Let the condition rating of bridge b at age a be $c_{ba} \in \boldsymbol{C}$, where $ba \in \boldsymbol{A} \times \boldsymbol{B}$.

Let the list of all condition ratings of bridges in the NBI at age a be sequence $\boldsymbol{X}_a$, where X_{a_i} is the ith element of sequence $\boldsymbol{X}_a$.

Then the expected condition rating of a bridge in the NBI at age a, s_a, is defined as follows in Eq. (1):

$$s_a = \frac{1}{|\boldsymbol{X}_a|} \sum_{i=1}^{|\boldsymbol{X}_a|} X_{a_i} \tag{1}$$

Let $\boldsymbol{S}$ be the sequence of expected condition ratings (baseline) for all bridge ages, such that $\forall a,\ a \in \boldsymbol{A} \Rightarrow s_a \in \boldsymbol{S}$.

Let $\boldsymbol{Q}_b$ be the set of all ages for which $b \in \boldsymbol{B}$ has a condition rating.

Then the baseline difference score of bridge b, bds_b, is defined as follows in Eq. (2):

$$bds_b = \frac{1}{|\boldsymbol{Q}_b|} \sum_{a \in \boldsymbol{Q}_b} c_{ba} - s_a \tag{2}$$

Analysis of Baseline Difference Scores

After computing the BDS for all the bridges in the US, we observe that it follows a normal distribution. We present summary statistics of BDSs from over 300,000 bridges in Table 7. Note that the mean and median are practically zero, and the standard deviation is close to one. It is expected that when performing a normality test on such a large sample, small deviations can cause the test to report nonnormality. Hence, we also ran the normality test with a small sample size of $100 \times 4 = 400$. The multiplier of 4 is chosen based on the maximum number of levels available among our 10 categorical factors. Fig. 2 shows the resulting box plot and histogram, which appear to follow a normal distribution. The normality test results in

Table 7. Summary statistics of distribution of BDS of bridges in US

	US
Count	334,331
Mean	−0.061
Standard deviation	0.887
Minimum	−2.50
25%	−0.617
50%	0.014
75%	0.54
Maximum	2.368

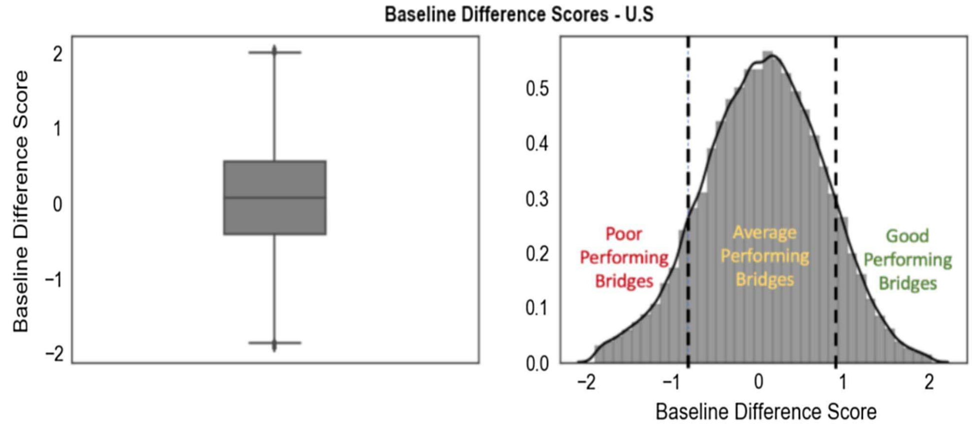

Fig. 2. Baseline difference scores—US.

Table 8. Normality test on random sample of 400 bridges

Test	Sample size	Statistic	p-value	Null hypothesis
Shapiro-Wilks	Small	4.204	0.122	Fail to reject
Shapiro-Wilks	Large	4.204	0.02	Reject

Note: The p-value suggests that the distribution of the baseline difference score of a random sample of 400 is normal.

Table 8 also confirm that the distribution of the random sample is normal.

Fig. 3 illustrates the time series of condition ratings for three bridges. Note that condition ratings for a particular bridge only take on ordinal (0–9) values. In contrast, the baseline is continuous, because it is the average condition rating over all bridges in the NBI at a particular age. Note also that the bridge condition ratings may fluctuate above and below the baseline because of improvements from interventions and deterioration of bridge components. Thus, bridges with equal BDSs could have highly varied condition ratings. To understand how condition rating time series of bridges vary with respect to BDS, using measures of standard deviation (SD), we categorized bridges as average (within plus or minus one SD), poor (less than one SD), and good (greater than one SD) based on their BDS. Thus, rather than using absolute thresholds, the interpretation of BDS for a population of bridges is established based on statistical significance.

It is essential to understand the characteristics of the time series of bridges with statistically significant poor, average, and good BDSs. For example, bridges with statistically significant poor BDSs are expected to have a time series of condition ratings that stay below the national baseline time series. Fig. 3 illustrates such time-series comparisons. In this example, the x-axis is the age and the y-axis the condition rating (superstructure). The blue line is the average condition rating of all bridges nationally, i.e., the baseline. The red, green, and yellow lines are time series that belong to actual bridges in the NBI with statistically significant poor, average, and good BDSs, respectively. While the figure only provides anecdotal evidence, we performed additional analysis. We observe that for bridges with good BDSs, their time series of condition ratings spends on average 80% of the time above the baseline. On the other hand, for bridges with poor BDSs, their time series of condition ratings spends on average 80% of the time below the baseline. For average performing bridges, their time series of condition ratings spend about 50% of the time above and below the baseline, respectively. This finding suggests that a good BDS, i.e., a statistically

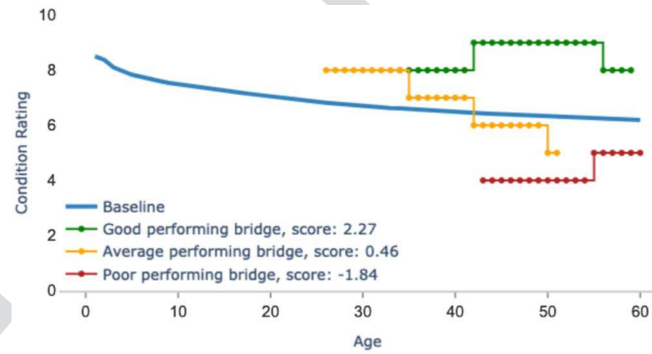

Fig. 3. Comparison of good, average, and poor performing bridges against blue national baseline. x-axis is bridge age and y-axis is bridge condition.

significant positive score, is not a result of large variances that are spread on either side of the baseline but a collection of small variances throughout the time series of the bridge that are positive.

The BDS is a sum of the differences among two condition rating time series aligned by age. While a single score for bridge performance is desirable, the summation operation in BDS computation cannot capture the consistency (or variation) of bridge conditions with respect to the national baseline over time. As a result, even a bridge with condition ratings above the national baseline at a given age can have a BDS of zero or even negative owing to its past performance. BDS is not a measure of the current bridge state but a longitudinal performance over time. It is possible for two bridges with very different condition ratings over time to have a BDS of zero: a bridge whose condition rating perfectly matches the baseline and a bridge whose score is above the baseline half the time and below the baseline the other half. The sum of differences in BDS computation also prevents detecting significant rises and declines in bridges' condition ratings. However, empirical observation in the NBI data set suggests that condition ratings often do not rise or decline significantly.

Since the BDS follows a normal distribution, ANOVA can be used to test the significance of various factors (e.g., ADT) on the BDS of bridges. For every factor, we categorized bridges into several groups (e.g., low, moderate, and high ADT) as appropriate and then performed one-way ANOVA to find the difference between the mean BDS of the groups. The degree of association

Table 9 Description for magnitudes of Cohen's d

Effect size	Cohen's d
Very small	0.01
Small	0.20
Medium	0.50
Large	0.80
Very large	1.20
Huge	2.00

Source: Data from Sawilowsky (2009).

between the factors and BDS is measured using effect size. In ANOVA, a commonly used measure of effect size is eta squared. However, eta squared is not easy to interpret. Therefore, we use Cohen's d—a standardized measure that makes it easy to interpret the effect size between two means. Eta squared values from ANOVA are converted to Cohen's d using a conversion tool (DeCoster 2009). Table 9 provides a description of the magnitude of Cohen's d and the corresponding effect sizes. In the following section, we will explore the effect of the selected factors on BDS.

Results

Our results are summarized in Tables 10 and 11. Table 10 lists the significance of ANOVA tests for each of the 11 factors, while Table 11 shows the effect size for each factor when tested individually. Notice that in Table 10, for a small sample size, only material (i.e., the bridge material type), snowfall, and maintenance explains the differences in BDS. Since the type of material used in bridge construction generally improves with time, we also tested the material factor, while keeping age constant. This factor is listed as material (age constant) in Tables 10 and 11. Correspondingly, the effect size of material (age constant) was the largest (Cohen'd = 0.54) in both the large and small sample analysis (Table 11). Snowfall has the second-highest association with BDS (Cohen'd = 0.45), followed by a close third in maintenance (Cohen'd = 0.44). These top three factors exhibit a medium association with BDS. Precipitation, freeze-thaw, maintainer, ADTT, and region have a small association with BDS. Both ADT and structure length have a very small association with BDS.

Upon further investigation of material categories, we observed that younger bridges were primarily made of concrete and prestressed concrete. We also expected younger bridges to perform better than older bridges. Therefore, to account for this bias related to age, we performed a post hoc analysis with a balanced sample of young bridges from four categories of bridges: steel, wood or timber, prestressed concrete, and concrete. This post hoc analysis further revealed a difference in BDS with respect to the material of the bridge, even when age is kept constant (e.g., all young bridges).

In Fig. 4, we show another view to understand the relation of material to BDS. The figure shows the percentage of bridges with different materials above and below the BDS of zero. Prestressed concrete bridges have the highest number of bridges above the BDS of zero between the ages of 30 and 45 (young category). In the same age range, a much lower percentage of wood bridges are above the BDS of zero. Fig. 4 reveals that prestressed concrete bridges generally perform better than other material types in this analysis using BDS.

A similar analysis of precipitation reveals a counterintuitive result. In Fig. 5, the percentage of bridges in very low and very high precipitation regions above and below the BDS of zero is shown between the ages of 1 and 60. We observe that very high precipitation regions have a higher percentage of bridges above a BDS of zero compared to very low precipitation regions. We conducted a Mann-Whitney U test to observe whether this difference was statistically significant since the distributions did not follow a normal distribution. Based on the test results, we conclude that the distribution of bridges with a BDS greater than 0 in the very low precipitation region is different from bridges with a BDS greater than 0 in the very high precipitation region. Similarly, the distribution of bridges with a BDS less than 0 in very low precipitation regions is different from bridges with a BDS less than 0 in very high precipitation regions.

Table 10. Summary of Analysis—ANOVA

	Fail to reject null hypothesis	
Factor	Small sample	Large sample
ADT	True (0.8)	False (0.0)
ADTT	True (0.12)	False (0.0)
Maintainer	True (0.13)	False (0.0)
Material	False (0.01)	False (0.0)
Material (age constant)	False (0.0)	False (0.0)
Structure Length	True (0.63)	False (0.0)
Precipitation	True (0.11)	False (0.0)
Freeze-thaw	True (0.4)	False (0.0)
Snowfall	False (0.0)	False (0.0)
Maintenance	False (0.0)	False (0.0)
Region	True (0.3)	False (0.0)

Note: Statistically significant difference in baseline difference score only with respect to material.

Table 11. Summary of analysis—effect size

	Effect size	
Factor	Small sample	Large sample
ADT	Very small (0.08)	Very small (0.04)
ADTT	Small (0.20)	Very small (0.06)
Maintainer	Small (0.15)	Very small (0.12)
Material	Small (0.30)	Small (0.35)
Material (age constant)	Medium (0.54)	Medium (0.54)
Structure length	Very small (0.06)	Very small (0.02)
Precipitation	Small (0.20)	Small (0.20)
Freeze-thaw	Small (0.17)	Very small (0.05)
Snowfall	Medium (0.45)	Small (0.21)
Maintenance	Medium (0.44)	Small (0.27)
Region	Small (0.17)	Small (0.20)

Note: In the small- and large-sample analysis, the material (age constant) has the highest effect size, followed by material and precipitation on baseline difference score.

Because we are very interested in precipitation and its effect on bridge health, we conducted further analyses to observe the interaction between precipitation and material types. We know from the previous results in Fig. 4 that prestressed concrete bridges perform better than wood or timber bridges. In addition, Fig. 6 shows that the distribution of prestressed concrete bridges is low in very low precipitation regions, and the distribution of prestressed concrete bridges is high in very high precipitation regions (Fig. 6). Also, wood or timber bridges are nearly absent in high precipitation regions.

We performed a Shapiro-Wilk normality test on four categories of materials with respect to high precipitation and low precipitation regions to check for normality. The results suggest that with 95% confidence, we reject the null hypothesis with a p-value of 0.0. In other words, none of the categories of material with respect to both

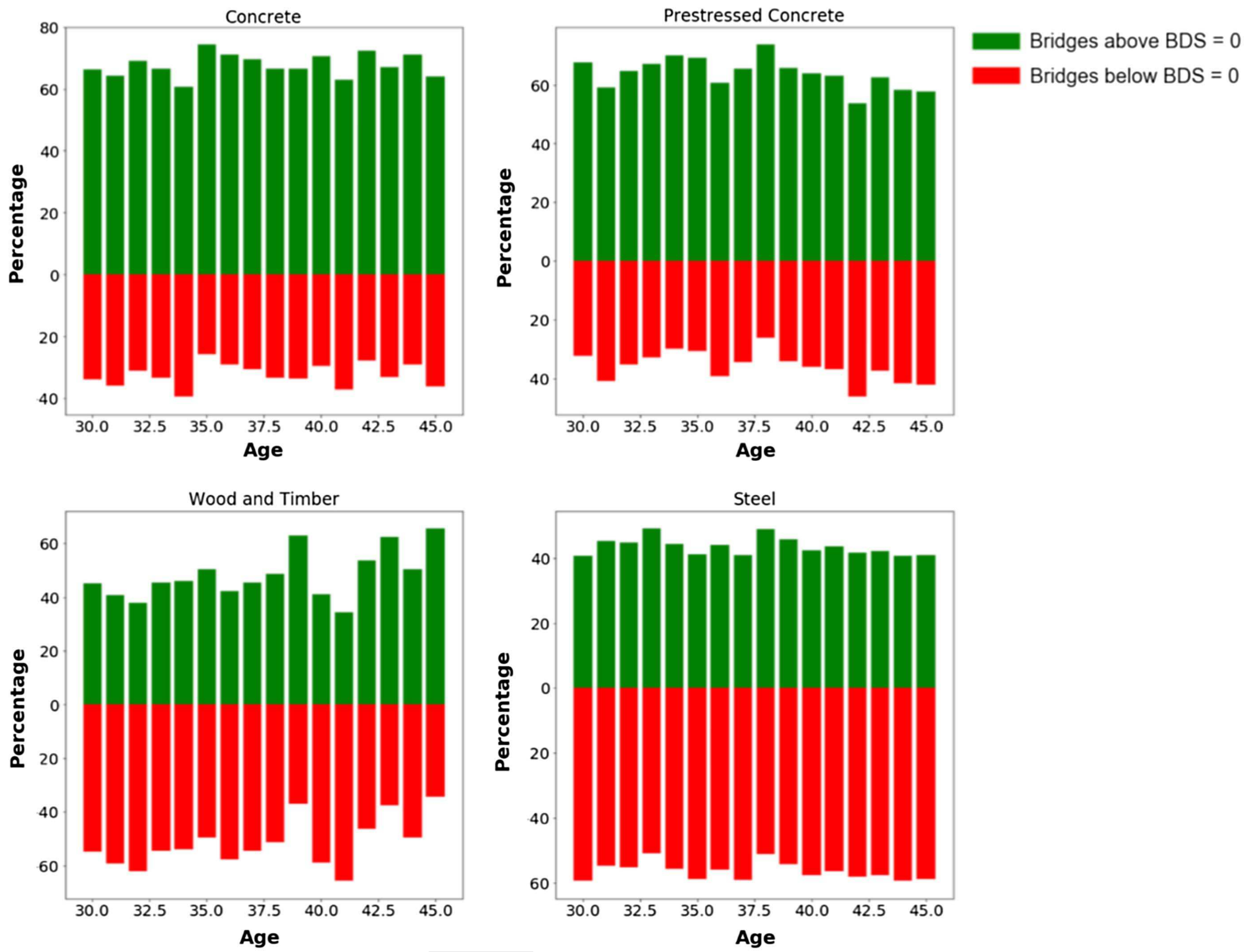

Fig. 4. Percentage of bridges above and below BDS of zero with respect to material type; note concrete and prestressed concrete bridges have higher percentages of bridges above BDS.

high precipitation and low precipitation regions follows a normal distribution. Therefore, to understand whether all material types' distributions are similar with respect to high precipitation and low precipitation, we performed a Kruskal-Wallis H test. Our result suggests that with 95% confidence and a p-value of 0.0, we can reject the null hypothesis that the BDS from the two distributions is similar. In conclusion, the distribution of the material types is different in high precipitation regions and low precipitation regions. This may explain the counterintuitive result with respect to precipitation.

Tool Support

Large data sets like the NBI are challenging to analyze and share. Therefore, in this study, we used DataCenterHub to host and explore the NBI data set. DataCenterHub is built on the HUBzero platform developed at Purdue University. In addition to providing specialized hosting services, DataCenterHub also provides user-friendly and flexible tools for exploration. DataCenterHub supports various data types, such as media (photos, drawings, and images) and data (CSV, text files, and JSON files). DataCenterHub provides a standard format for uploading large data sets and a web entry to upload and update metadata and parameter sets. DataCenterHub provides security mechanisms to protect data and supports restricted access that grants access only to select groups to see appropriate data sets with role-based constraints.

In this study, DataCenterHub consisted of Jupyter Notebooks and centrally hosted files with NBI records from all years and all states. The NBI data set is populated from CSV format files available from the FHWA website. The database is curated, cleaned, and populated using Python scripts. The Jupyter Notebook environment enabled us to execute Python scripts and display charts in a shareable and readable document. Python scripts are developed as part of this research study for data processing, such as extraction, formatting, and curating data sets. Links to the Data Cleaning and Analysis scripts are hosted on GitHub.

Discussion

This research study addresses some critical challenges with respect to data cleaning and related transformation related to NBI data. It does so in a manner that is repeatable and scalable. We overcome challenges related to incomplete or missing records, accounting for maintenance events, and explore interactive effects among factors like materials, age, and precipitation.

Incomplete and missing inspection records can be quite challenging. While computing a national baseline, missing condition rating data in individual bridges are not a concern. However, when

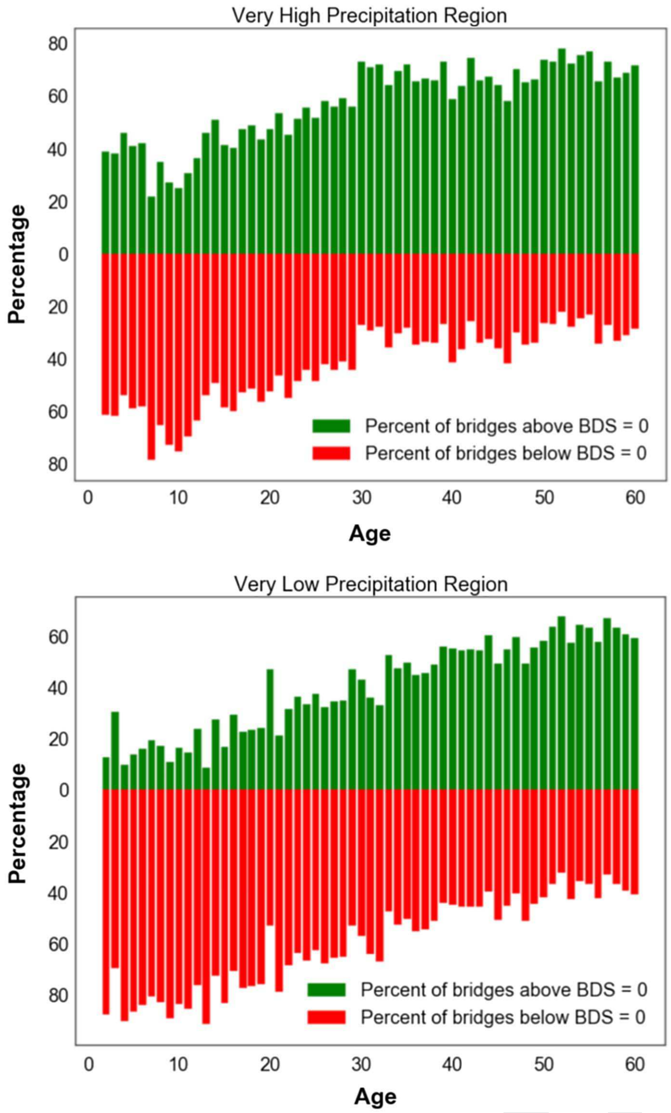

Fig. 5. Percentage of bridges above and below BDS of zero with respect to precipitation regions. There is a lower percentage of bridges above BDS in high precipitation, and there is a higher percentage of bridges above BDS in lower precipitation.

computing the BDS of a bridge, gaps in the time series of condition rating is yet to be completely addressed. In this study, if we observe a large gap in the time series of condition ratings for a bridge, we split the time series into segments and treat each segment as a separate bridge record. We observed that large gaps in time series usually corresponded to a bridge being rebuilt.

Another challenge in modeling bridge health is accounting for maintenance/intervention. In our study, we considered maintenance as a factor that influences bridge health. We did not take into account specific interventions, such as replacement, rehabilitation, and reconstruction, but this can be easily done in the future with data availability. The analysis presented in this paper is based on the observed improvement in the condition rating of a bridge that can be attributed to any of the general intervention types. Since BDS computation is independent of the factors that influence bridge health, testing additional factors presents no significant computational overhead.

Our primary objective in this study was to understand the uniform association of an individual factor with the BDS. Therefore, we did not specifically test for the interactive relationship among factors. Our analysis of interaction among factors is limited to material, age, and precipitation to further explore factors that exhibited a high association with BDS. Our results suggest that the BDS is a key performance indicator, which can complement other bridge indicators provided by previous literature. In future studies, techniques like ANNs, classification trees, and regression analysis could include BDS with other bridge attributes to study interactions among factors.

Our study included all US bridges. Our selection was not limited to any particular region. Analysis with ANOVA included both small and large samples. Using small sample sizes allowed us to watch for deflated p-values due to large sample sizes. Both small- and large-sample analysis also provided a range for the effect size of the analysis. ANOVA's necessary assumptions of a normally distributed population of scores are preserved in both small and large samples.

To compare and contrast our findings with prior efforts, Table 12 lists significant results from the current state of the art. Previous research studies generally reported that physical and service-related factors were more influential than environmental and regional factors. For example, physical factors such as traffic (or ADT) (Chang et al. 2019; Assaad and El-adaway 2020; Kim and Yoon 2010; Hasan and Elwakil 2020; Nasrollahi and Washer 2015), age (Chang et al. 2019; Saeed et al. 2017; Assaad and El-adaway 2020), and structure length (Chang et al. 2019; Saeed et al. 2017; Assaad and El-adaway 2020; Hasan and Elwakil 2020) were found to be the most influential in understanding bridge conditions in several research efforts. Chang et al. (2019) found the functional classification of a route and the number of lanes to be influential factors. Assaad and El-adaway (2020) reported that deck width, number of spans, operating rating, inventory rating, and structural evaluation were also significant factors. Finally, Hasan and Elwakil (2020) found the degree of skew and structure length to be influential factors in simulating future bridge condition ratings. However, our results shoed only a small association of the commonly reported physical factors to the BDS at a national scale. These differences present interesting opportunities for future work.

Our results do align with the findings of Kim and Yoon (2010), Hasan and Elwakil (2020), and Saeed et al. (2017). Similar to their findings, our results suggest that the material type of a bridge has a higher association with the bridge's health. We observed that prestressed concrete and concrete bridges performed better than wood or timber and steel bridges. The climate of a given region is influenced by various factors, such as latitude, elevation, topography, and prevailing winds. The cumulative effect of multiple variations in climate factors may also affect bridge condition. We included bridges' geographic regions as a factor in our test as a proxy for climate's cumulative effect and found that region had a slight association with BDS. However, we plan to perform additional tests at a more granular geographic level in future work. Contrary to popular belief, we only found a small effect of precipitation or freeze-thaw cycles on BDS, but snowfall had a medium effect. Our literature review suggested that the approaches to understanding bridge health differ in methods, objectives, and factors. The scope of various studies also differs. For example, one study may consider a particular type of bridge component, while another may focus on bridges in a specific region. We aggregate the influential factors reported from studies that used deterministic models (regression analysis), stochastic models (Markov decision processes and probit models), and AI models (ANNs and CBR) to inform factor selection in our study. However, owing to differences in our work's method, scale, and scope, comparisons are not straightforward.

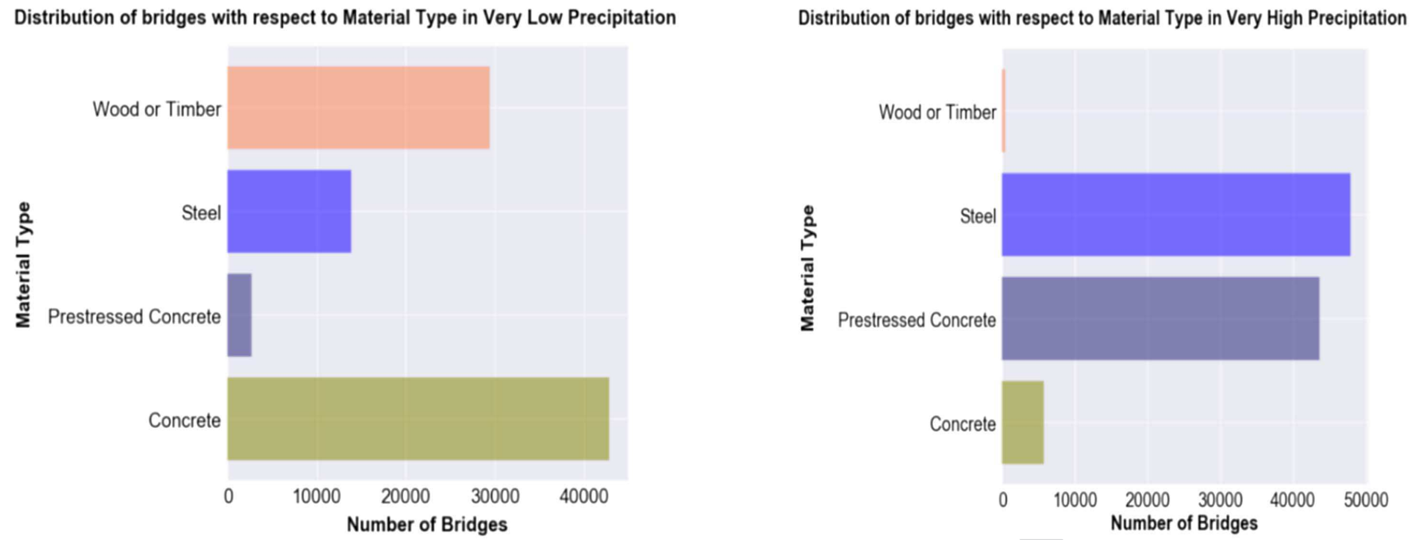

Fig. 6. Distribution of bridges with respect to material type in very low precipitation and very high precipitation regions. Note high distribution of concrete and wood or timber bridges in very low precipitation region and high distribution of steel and prestressed concrete bridges in very high precipitation region.

Table 12. Comparison of influential factors identified throughout literature

	Physical	Region	Structural and material	Environmental	Service
Case-based reasoning (Morcous et al. 2002)	✓ *	—	✓	—	—
Cox hazards model with LASSO a and stepwise regression (Wettach-Glosser et al. 2020)	✓ *	—	✓ *	✓ *	✓ *
Linear regression and Monte Carlo simulation (Hasan and Elwakil 2020)	—	✓ *	✓ *	✓ *	✓ *
Bayesian survival analysis (Fleischhacker et al. 2020)	✓ *	—	✓ *	✓ *	✓ *
Artificial neural network and *k*-nearest neighbor (Assaad and El-adaway 2020)	✓ *	—	✓	—	✓
Logistic regression and classification tree (Chang et al. 2019)	✓*	—	—	—	—
Ordered probit model (Saeed et al. 2017)	✓ *	✓	✓	✓	✓
Multiple regression and GIS (Kim and Yoon 2010)	✓ *	—	✓ *	✓ *	✓ *
Baseline difference score	✓ *	✓ *	✓ *	✓ *	✓ *

Note: A check mark indicates factor category tested; an asterisk indicates the factor category found to be influential.

Finally, we realize that our method includes many steps of varying complexity. These steps are data-intensive and require a balanced representation of data across several factors. For example, based on the discussion in section 4, we know that bridges from all states are not equally represented after data cleaning and filtration. We address the issue of imbalanced representation of data by subsampling overrepresented states. Finally, the BDS method is new. We expect that it will take time before the method is widely adopted and further improved. Overall, this study's results were validated and deemed useful by subject matter experts at several Bridging Big Data workshops (https://bridgingbigdata.github.io/pages/workshops.html).

Conclusion and Future Work

This research study introduces a new measure to understand and compare bridge conditions based on inspection time-series data called BDS. The BDS provides a single score for bridge health. The BDS follows a normal distribution, which categorizes bridges as average, poor, and good using standard deviation measures. We observed that good-performing bridges have condition ratings that are mostly above the baseline from our analysis. In contrast, poor-performing bridges have condition ratings that are below the baseline. Our use of a national baseline in BDS computation also accounts for a bridges' condition ratings with respect to age and scheduled Maintenance performed over a bridge's life.

This research study provided a reference implementation of a big data pipeline for bridge health–related data sets. The data cleaning and related transformation used in this research scale to NBI records across all US states. The strategy is also traceable to coding guidelines and cross-checks, as provided by FHWA (1995). Our data cleaning efforts only found 42% of the original data set to be suitable for analysis. A substantial number of bridges in the NBI data set have condition ratings that do not change across all reported inspections.

We examined 11 factors, including environmental factors that could influence bridge conditions. This paper found that material, snowfall, and maintenance were the three factors most associated with bridge condition. However, material, in particular concrete and prestressed concrete, was associated with better-performing bridges. Material type has the highest association with the condition of bridges, followed by snowfall and maintenance, compared to the other factors selected in this study. One could also argue that snowfall may be associated with an increased need for maintenance. We did not explore this relationship in this paper. The factor that had the smallest effect on BDS was ADT. External factors, such as precipitation, showed only a slight association with bridge condition. The association between all factors selected in this analysis and BDS varies from very small to medium, meaning individual factors cannot fully explain the condition of bridges. We also observed a strong association between material and BDS by accounting for the latent relationships with age and precipitation.

BDS can provide a measure of the performance of bridges over time. However, BDS does not take into account the variance of condition ratings over time. Therefore, a significant drop or rise in a bridge's condition ratings are not reflected in the BDS. Future research could focus on developing a complementary measure that would provide a degree of variance in condition ratings over time. Such a measure could allow for a deeper understanding of the stability of bridge performance. Also, the quality of the NBI data needs to be improved, for example, by estimating missing values that would provide more data points for analysis. In addition to missing data, bridge inspections are biennial, and the long intervals between inspections prevent the accurate characterization of bridges. Further, inspectors conduct visual inspections of bridges that are highly subjective. Data collection using Internet of Things devices could provide reliable, frequent, and objective data. Finally, as more states start collecting and publishing element-level inspection data, we expect to use more detailed condition assessments in computing bridge health scores.

Appendix. Statewise Summary of Missing Records after Data Cleaning

In Table 13, we provide a statewise analysis of the total survey records available in the NBI from 1992 to 2012; we also provide the total number and percentage of surveys considered for analysis after performing the data cleaning process.

Table 13. Table showing number of NBI inspection records available after data cleaning

State	Total survey records	Records considered in study	Percentage of records in study
Colorado	223,645	55,259	24.71
Montana	143,904	86,220	59.91
Washington	219,318	111,952	51.05
Utah	86,707	21,616	24.93
California	761,313	336,288	44.17
Hawaii	28,868	10,968	37.99
Oklahoma	623,417	330,803	53.06
Arizona	198,595	53,104	26.74
Virginia	396,744	164,877	41.56
Tennessee	545,968	138,745	25.41
South Carolina	246,494	93,626	37.98
Alabama	417,539	212,866	50.91
Louisiana	350,988	192,878	54.95
Nebraska	400,539	295,655	73.81
Illinois	714,936	362,921	50.76
Kansas	663,487	387,056	58.34
North Dakota	116,171	71,287	61.36
South Dakota	158,215	98,721	62.4
Wisconsin	386,700	225,939	58.43
Massachusetts	137,664	20,186	14.66
Maine	67,423	22,560	33.46
Rhode Island	22,591	2,521	11.16
New Jersey	216,169	76,105	35.21
Pennsylvania	665,314	267,464	40.2
Maryland	144,047	60,340	41.89
Arkansas	339,492	178,566	52.6
Wyoming	81,257	48,403	59.57
Idaho	110,415	37,871	34.3
Oregon	198,421	55,259	33.01
Nevada	45,179	13,013	28.8
Alaska	334,522	21,204	63.39
Texas	1,347,902	473,361	35.12
New Mexico	105,730	38,725	36.63
West Virginia	194,514	87,703	45.09
Kentucky	373,241	214,966	57.59
North Carolina	517,708	132,552	25.6
Georgia	405,216	186,850	46.11
Mississippi	442,724	278,356	62.87
Florida	339,786	168,565	49.61
Iowa	641,429	500,280	77.99
Indiana	484,585	260,133	53.68
Michigan	641,698	369,188	57.53
Missouri	641,698	369,188	57.53
Ohio	1,570,646	531,684	33.85
Minnesota	480,773	156,378	32.53
Connecticut	126,267	6,118	4.85
New Hampshire	82,523	12,203	14.79
Vermont	73,408	20,385	27.77
New York	522,369	179,217	34.31
Washington, DC	7,808	2,623	33.59
Delaware	28,935	11,135	38.48
Puerto Rico	61,350	25,421	41.44

Data Availability Statement

Some or all data, models, or code generated or used during the study are available in a repository or online in accordance with funder data retention policies. All code generated during the study is available at GitHub (Kale 2019). All data used during this study are available at Box (Kale 2020).

Acknowledgments

This work is partially funded by NSF Awards 1762034 and 1636805. Subject matter experts at Nebraska DOT—Bridge Division, and LTBP Infobridge —FHWA provided valuable assistance and access to data needed for this research. We also want to thank the anonymous reviewers for their thoughtful feedback in helping improve the paper. Drs. Deepak Khazanchi, Daniel Linzell, Christian Haas, Chungwook Sim, and Saeed Eftekhar Azam provided valuable guidance during this research work and provided reviews for initial drafts.

References

Anderson, D. R., K. P. Burnham, and W. L. Thompson. 2000. "Null hypothesis testing: Problems, prevalence, and an alternative." *J. Wildl. Manage.* 64 (4): 912–923. https://doi.org/10.2307/3803199.

ARTBA. 2017. *2017 infrastructure report card Bridget*. Reston, VA: ASCE.

Assaad, R., and I. H. El-adaway. 2020. "Bridge infrastructure asset management system: Comparative computational machine learning approach for evaluating and predicting deck deterioration conditions." *J. Infrastruct. Syst.* 26 (3): 04020032. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000572.

Barreto, H., and F. Howland. 2006. *Introductory econometrics: Using Monte Carlo simulation with Microsoft Excel*. Cambridge, UK: Cambridge University Press.

Bolukbasi, M., J. Mohammadi, and D. Arditi. 2004. "Estimating the future condition of highway bridge components using national bridge inventory data." *Pract. Period. Struct. Des. Constr.* 9 (1): 16–25. https://doi.org/10.1061/(ASCE)1084-0680(2004)9:1(16).

Cesare, M. A., C. Santamarina, C. Turkstra, and E. H. Vanmarcke. 1992. "Modeling bridge deterioration with Markov chains." *J. Transp. Eng.* 118 (6): 820–833. https://doi.org/10.1061/(ASCE)0733-947X(1992)118:6(820).

Chang, M., M. Maguire, and Y. Sun. 2019. "Stochastic modeling of bridge deterioration using classification tree and logistic regression." *J. Infrastruct. Syst.* 25 (1): 04018041. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000466.

DeCoster, J. 2009. "Converting effect sizes." Accessed February 2020. [25] https://www.stat-help.com.

DeStefano, P. D., and D. A. Grivas. 1998. "Method for estimating transition probability in bridge deterioration models." *J. Infrastruct. Syst.* 4 (2): 56–62. https://doi.org/10.1061/(ASCE)1076-0342(1998)4:2(56).

FHWA (Federal Highway Administration). 1995. *FHWA recording and coding guide for the nation's bridges, office of engineering*. Washington, DC: FHWA.

Fleischhacker, A., O. Ghonima, and T. Schumacher. 2020. "Bayesian survival analysis for us concrete highway bridge decks." *J. Infrastruct. Syst.* 26 (1): 04020001. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000511.

Hasan, S., and E. Elwakil. 2020. "National bridge inventory data-based stochastic modeling for deck condition rating of prestressed concrete bridges." *Pract. Period. Struct. Des. Constr.* 25 (3): 04020022. https://doi.org/10.1061/(ASCE)SC.1943-5576.0000505.

Huang, Y., R. Bird, and M. Bell. 2009. "A comparative study of the emission by road maintenance works and the disrupted traffic using life cycle assessment and micro-simulation." *Transp. Res. Part D* 14 (3): 197–204. https://doi.org/10.1016/j.trd.2008.12 [26] .003.

Huang, Y.-H. 2010. "Artificial neural network model of bridge deterioration." *J. Perform. Constr. Facil* 24 (6): 597–602. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000124.

Kale, A. 2019. "kaleoyster/thesis." Accessed January 24, 2019. https://github.com/kaleoyster/thesis.

Kale, A. 2020. "Baseline difference score." Accessed March 2, 2020. https://unomaha.app.box.com/v/baselinedifferencescore.

Kim, Y. J., and D. K. Yoon. 2010. "Identifying critical sources of bridge deterioration in cold regions through the constructed bridges in North Dakota." *J. Bridge Eng.* 15 (5): 542–552. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000087.

Madanat, S., R. Mishalani, and W. H. W. Ibrahim. 1995. "Estimation of infrastructure transition probabilities from condition rating data." *J. Infrastruct. Syst.* 1 (2): 120–125. https://doi.org/10.1061/(ASCE)1076-0342(1995)1:2(120).

Morcous, G. 2011. *Developing deterioration models for Nebraska bridges*. Lincoln, NE: University of Nebraska-Lincoln.

Morcous, G., H. Rivard, and A. M. Hanna. 2002. "Modeling bridge deterioration using case-based reasoning." *J. Infrastruct. Syst.* 8 (3): 86–95. https://doi.org/10.1061/(ASCE)1076-0342(2002)8:3(86). [27]

Nasrollahi, M., and G. Washer. 2015. "Estimating inspection intervals for bridges based on statistical analysis of national bridge inventory data." *J. Bridge Eng.* 20 (9): 04014104. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000710.

Saeed, T. U., Y. Qiao, S. Chen, K. Gkritza, and S. Labi. 2017. "Methodology for probabilistic modeling of highway bridge infrastructure condition: Accounting for improvement effectiveness and incorporating random effects." *J. Infrastruct. Syst.* 23 (4): 04017030. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000389.

Sawilowsky, S. S. 2009. "New effect size rules of thumb." *J. Mod. Appl. Stat. Methods* 8 (26): 1. [28]

Scherer, W. T., and D. M. Glagola. 1994. "Markovian models for bridge maintenance management." *J. Transp. Eng.* 120 (1): 37–51. https://doi.org/10.1061/(ASCE)0733-947X(1994)120:1(37).

Wettach-Glosser, J., A. Unnikrishnan, and T. Schumacher. 2020. "Survival analysis of concrete highway bridge decks in Oregon utilizing lasso and stepwise-variable selection." *J. Bridge Eng.* 25 (10): 04020077. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001606.