Replication of Critical Meteorological Conditions Prediction for Expressway Bridge Pavement Icing

Madison Woo (2975355), madison.woo@uconn.edu
Olivia Kashalapov (2940517), olivia.kashalapov@uconn.edu

12/08/23

Section 1. Introduction:

The objective of this project is to replicate and analyze the findings of the original study to verify the models that they use for bridge pavement temperature prediction. By replicating this study, our goal is to understand the factors that contribute to bridge pavement temperatures during winter conditions.

This study focuses on predicting critical meteorological conditions for icy bridge decks on expressways during winter. Accurate bridge pavement temperature prediction is significant information for early warning systems to enhance traffic safety in icy conditions. The study conducted in the paper is based on meteorological data from a monitoring station on the Tuhai River Bridge in Shandong province, China, which is an important component of the Beijing-Shanghai Expressway.

The study introduces models for predicting bridge pavement temperatures during winter. These models use a smaller sample size, which is good for areas with less data. The BP neural network model was used in the study, in particular, to show better accuracy and is recommended for predicting expressway bridge deck temperatures.

A valuable characteristic of BP neural networks using nonlinear transfer functions is their ability to model complex algorithms related to nonlinear issues. An accurate BP neural network can predict the error in bridge pavement temperature during winter without understanding the internal interactions among influencing factors. This ability enhances the bridge deck icing prediction accuracy.

The goal of our project is to replicate the study using multiple linear regression to predict critical meteorological conditions of expressway bridge pavement icing, we aim to validate the predictive capabilities of these models for bridge pavement temperatures in winter conditions. This replication project will help us get a better understanding of factors influencing bridge pavement temperatures, which ensures safe expressway travel during colder months of the year.

The analysis of the data in the original study calculates the confidence intervals of the data, but it does not really check assumptions of the data. For our extension/enhancement, we will be checking the statistical assumptions and reporting it in our replication.

Section 2. Description of the Data:

In order to forecast when roads might get icy, a study was conducted that gathered and analyzed minute-by-minute weather data of traffic meteorological stations installed on expressway bridges. Data was measured from a traffic weather monitoring station on the Tuhai River Bridge, which is part of the Beijing-Shanghai Expressway. This expressway is crucial in connecting Beijing and Shanghai. The Tuhai River Bridge is a continuous rigid-frame bridge that is often used when constructing large-span bridges. The meteorological station, in Shandong Province, experiences a temperature continental climate with harsh winters. This station is the typical observation point for studying bridge icing.

By the definition of icing, if the bridge's temperature is below 0°C while it is raining or snowing, it is considered icy. The site's meteorological data gathers information such as bridge pavement temperature (°C), dew point temperature (°C), wind chill temperature (°C), wet bulb temperature (°C), relative humidity (%), air temperature (°C), and wind speed (m/s), by the time. The study focused on winter months (December, January, and February) from the past three years (2020-2022), as depicted in Table 1 below.

```
## [1] "Table 1. Meteorological data of Tuhai River Bridge."
```

```
##       Time dew_point wind_chill wet_bulb air_temp bridge_deck relative_humidity
## 1  9:32:00     -5.10      -5.94    -4.74    -4.57        0.32             96.13
## 2  9:31:00     -5.09      -6.23    -4.74    -4.57        0.30             96.13
## 3  9:30:00     -5.16      -6.32    -4.81    -4.64        0.25             96.16
## 4  9:29:00     -5.16      -6.42    -4.81    -4.63        0.20             96.08
## 5  9:28:00     -5.24      -6.56    -4.88    -4.71        0.14             96.03
## 6  9:27:00     -5.23      -6.58    -4.88    -4.71        0.08             96.09
## 7  9:26:00     -5.32      -6.62    -4.96    -4.78        0.03             96.06
## 8  9:24:00     -5.31      -6.70    -4.96    -4.78       -0.02             96.09
## 9  9:23:00     -5.41      -6.65    -5.04    -4.86       -0.06             95.98
## 10 9:22:00     -5.40      -6.46    -5.04    -4.86       -0.09             96.04
## 11 9:21:00     -5.40      -6.45    -5.04    -4.87       -0.10             96.03
## 12 9:20:00     -5.33      -6.45    -4.96    -4.79       -0.14             95.98
## 13 9:18:00     -5.33      -6.20    -4.96    -4.79       -0.18             96.00
## 14 9:17:00     -5.33      -6.19    -4.96    -4.79       -0.20             95.98
## 15 9:16:00     -5.29      -6.17    -4.92    -4.74       -0.24             95.99
##    wind_speed
## 1        0.96
## 2        1.06
## 3        1.08
## 4        1.12
## 5        1.17
## 6        1.17
## 7        1.18
## 8        1.23
## 9        1.21
## 10       1.13
## 11       1.10
## 12       1.13
```

```
## 13        1.01
## 14        1.00
## 15        0.98
```

The minute-by-minute meteorological data collected are quite consistent over short periods, as seen in Table 1. The road's adhesion coefficient drops rapidly when the bridge deck begins to freeze, even if the nearby road sections have not experienced icing yet. When a vehicle moves onto the bridge, the sudden shift in the adhesion coefficient can have critical impact on driving safety. This is why identifying the point when the bridge deck begins to ice is crucial, as stable weather conditions alone do not influence the icing state. Noticeable changes in weather factors have the potential to change the bridge deck's icing state, which is important to take into account when predicting pavement temperature and icing. In order to choose an effective sample for predicting bridge pavement temperature, it is necessary to categorize meteorological data levels to capture significant changes.

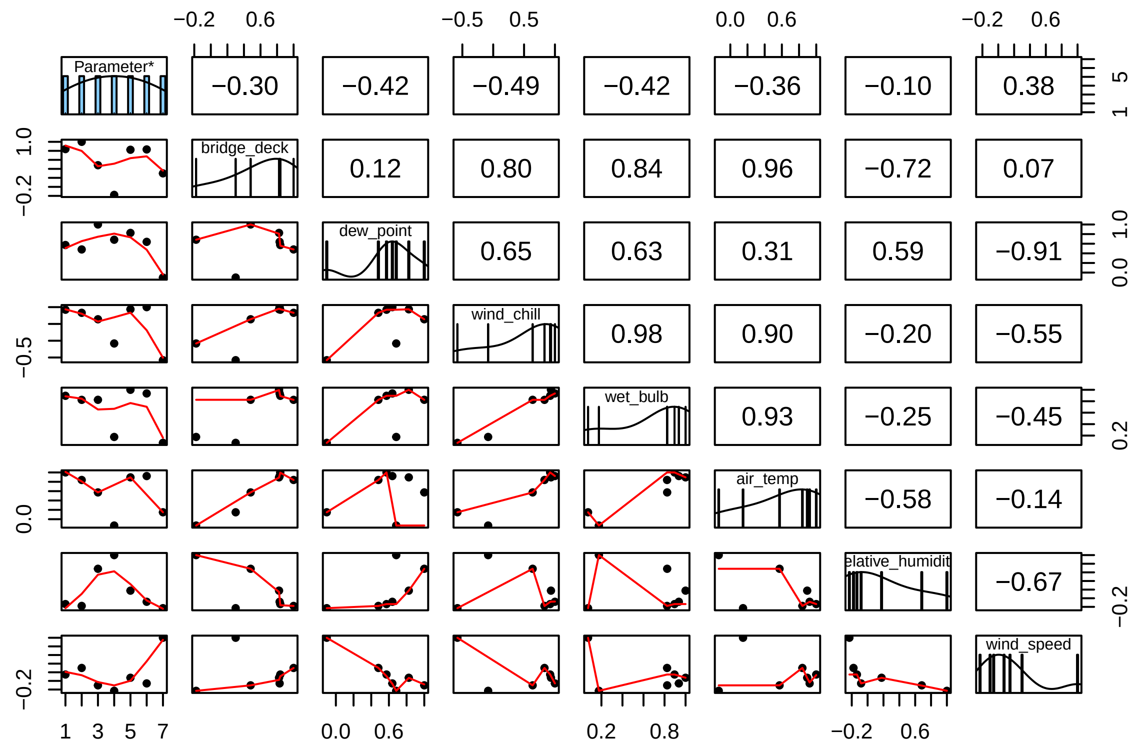
To effectively choose samples, accurately predict bridge deck temperature, and determine whether the bridge section is icing, it is important to identify the factors influencing the bridge deck temperatures. Literature analysis indicates that meteorological factors such as temperature, humidity, and wind speed affect the bridge pavement temperature. The bridge pavement temperature is the regressor or independent variable. Using multiple linear regression helps identify influencing factors by looking at the interaction between the independent variables and predicting the dependent variable.

Pearson partial correlation analysis is done to study the correlation between the variables. Meteorological factors remain consistent over short periods, and there are a few outliers in bridge pavement temperature due to external factors such as the environment and traffic. As a result, data at 10-minute intervals from three winters were chosen. After removing missing data, 11,837 sets of valid data were obtained, as shown in the raw data set. Table 2 displays the Pearson partial correlation coefficient among variables.

```
## [1] "Table 2. Pearson partial correlation analysis results."

##           Parameter bridge_deck dew_point wind_chill wet_bulb air_temp
## 1   bridge_pavement       1.000     0.481      0.831    0.824    0.839
## 2    dew_point_temp       0.481     1.000      0.639    0.825    0.573
## 3   wind_chill_temp       0.831     0.639      1.000    0.936    0.923
## 4     wet_bulb_temp       0.824     0.825      0.936    1.000    0.895
## 5 relative_humidity      -0.178     0.683     -0.082    0.176   -0.135
## 6          air_temp       0.839     0.573      0.923    0.895    1.000
## 7        wind_speed       0.300    -0.103     -0.580    0.073    0.148
##   relative_humidity wind_speed
## 1            -0.178      0.300
## 2             0.683     -0.103
## 3            -0.082     -0.058
## 4             0.176      0.073
## 5             1.000     -0.230
## 6            -0.135      0.148
## 7            -0.230      1.000

## [1] "Scatterplot Matrix 1"
```



A coefficient closer to 1 means that there is a stronger correlation between two variables. The scatterplot matrices in this replication were not included in the original study, but are there to verify and visually evaluate collinearity as opposed to the table. From Scatterplot Matrix 1 and Table 2, it can be seen that bridge pavement temperature has the highest correlation with air temperature, while air temperature is strongly correlated with wind chill and wet bulb temperatures. The dew point temperature has a high correlation with air temperature and relative humidity as well. To efficiently screen effective samples, air temperature, relative humidity, and wind speed were chosen as indicators for classifying sample grades when predicting bridge pavement temperature.

To enhance the prediction accuracy of bridge pavement temperature and icing, it is important to categorize meteorological data based on a sample grading index. After preprocessing the 11,837 data sets, it was found that the bridge deck starts freezing when the air temperature falls below 5°C, with stable freezing happening below -10°C. Using wind scale specifications and relative humidity ranges, a sample level division was established, resulting in 3,906 sample classification standards. Out of those 3,906 samples, 1,378 groups met the classification standards. A sample level division is shown in Table 3 below.

```
## [1] "Table 3. Classification criteria for valid sample grades."

##   independent_variable                                classification
## 1             air_temp -10-5 degrees C, increasing by 0.5 degrees C
## 2    relative_humidity    15% and below; 15%-100%, increasing by 5%
## 3           wind_scale           level 0-6, increasing step-by-step
##   number_of_levels
```

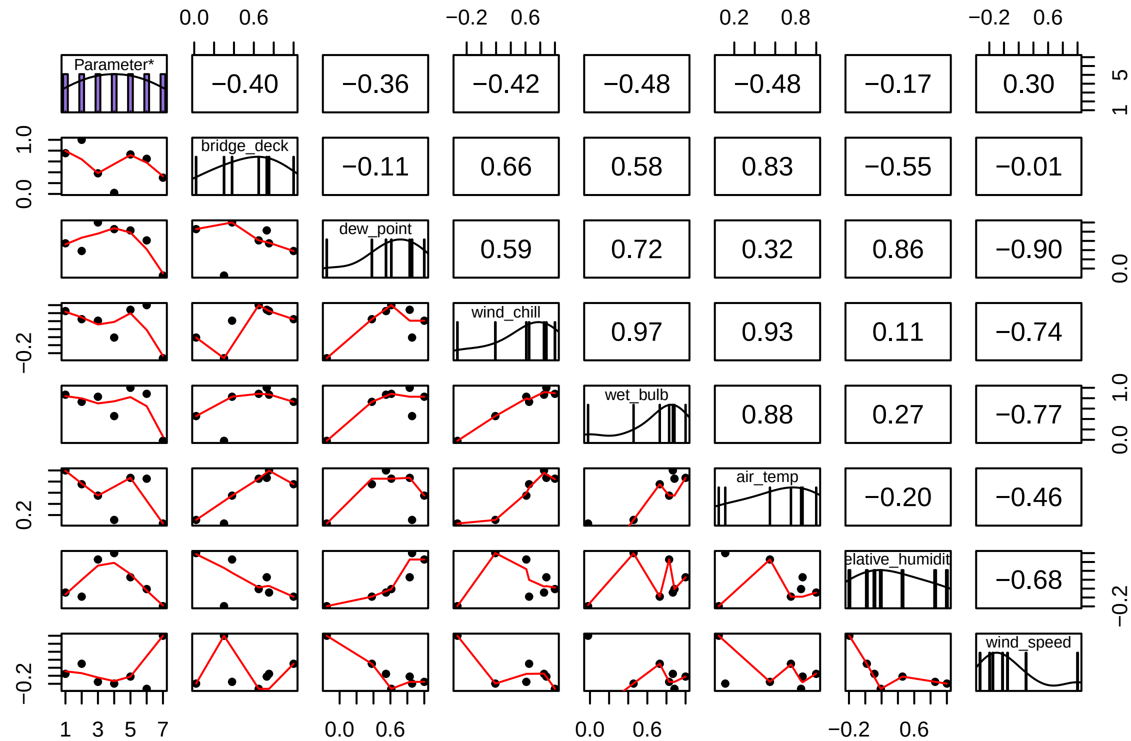
```
## 1                31
## 2                18
## 3                 7
```

It is important to check the linearity assumption among independent variables. A partial correlation coefficient close to 1 between two independent variables indicates collinearity, which can impact the accuracy of the regression model. To prevent this, collinear factors should be excluded from the analysis.

```
## [1] "Table 4. Pearson partial correlation analysis results."
```

```
##           Parameter bridge_deck dew_point wind_chill wet_bulb air_temp
## 1   bridge_pavement       1.000     0.381      0.649    0.730    0.753
## 2    dew_point_temp       0.381     1.000      0.611    0.828    0.548
## 3   wind_chill_temp       0.649     0.611      1.000    0.884    0.852
## 4     wet_bulb_temp       0.730     0.828      0.884    1.000    0.867
## 5 relative_humidity       0.019     0.856      0.190    0.456    0.112
## 6          air_temp       0.753     0.548      0.852    0.867    1.000
## 7        wind_speed       0.300    -0.152     -0.326   -0.020    0.048
##   relative_humidity wind_speed
## 1             0.019      0.300
## 2             0.856     -0.152
## 3             0.190     -0.326
## 4             0.456     -0.020
## 5             1.000     -0.195
## 6             0.112      0.048
## 7            -0.195      1.000
```

```
## [1] "Scatterplot Matrix 2"
```



Using the 1,378 valid samples from the pearson correlation analysis method, Scatterplot Matrix 2 and Table 4 shows high partial correlation coefficients among air temperature, wind chill temperature, wet bulb temperature, relative humidity, and dew point temperature. This indicates a clear collinearity trend. As a result, dew point temperature, wind chill temperature, and wet bulb temperature are excluded, while air temperature, relative humidity, and wind speed are kept in the highway bridge pavement temperature prediction model.

Section 3. Methods (or Models) and Analysis:

To enhance the accuracy of the temperature predictions, two models were created - one using multiple linear regression and another using a neutral network with a nonlinear function. The assumptions of the data is that there is linearity, homoscedasticity, errors are independent, and there is normality.

```
##
## Call:
## lm(formula = bridge_deck ~ air_temp + relative_humidity + wind_speed,
##     data = rawdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -30.362  -1.929  -0.462   1.359  33.881
##
## Coefficients:
```

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.023459   0.113296  -0.207    0.836
## air_temp           0.959380   0.005779 166.018  < 2e-16 ***
## relative_humidity -0.009337   0.001556  -5.999 2.04e-09 ***
## wind_speed         0.621215   0.017497  35.504  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.614 on 11797 degrees of freedom
## Multiple R-squared:  0.7334, Adjusted R-squared:  0.7333
## F-statistic: 1.081e+04 on 3 and 11797 DF,  p-value: < 2.2e-16
```

```
##                         2.5 %       97.5 %
## (Intercept)       -0.24553758  0.198619788
## air_temp           0.94805259  0.970707258
## relative_humidity -0.01238725 -0.006285924
## wind_speed         0.58691747  0.655512504
```

As mentioned before, model with air temperature, relative humidity, and wind speed are kept in the highway bridge pavement temperature prediction model. The numbers in this replication are different than the numbers in the original study, which will be explained more in the discussion. The model is represented as T = -0.023 + 0.959X1 - 0.009X2 + 0.621X3. T, X1, X2, and X3 represent the bridge pavement temperature (°C), air temperature (°C), relative humidity (%), wind speed (m/s) respectively.

Included in the study was the 95% confidence interval model. The original study did not show their confidence interval calculations, but we did here. Since the independence assumption and constant variance assumption is violated, the confidence interval values in this model are not correct because the standard errors are affected. This will be expanded on in the discussion component of this replication. Our 95% confidence interval model is T = -0.246 + 0.948X1 - 0.012X2 + 0.587X3. The summary of the model indicates that the coefficient of determination R^2 is 0.733, suggesting that 73.3% of the bridge pavement temperature is explained by air temperature, relative humidity, and wind speed. The F-test value is 1.081e+04 with a p-value of 2.2e-16. Since the p-value is less than the significance level of 0.05, the model parameters are suitable for estimating the entire sample, passing the significance test.

The bridge pavement temperature prediction model calculated above is based on the linear relationship between air temperature, relative humidity, wind speed and bridge pavement temperature. To improve prediction accuracy, the study uses BP neural network to predict bridge pavement temperature based on a nonlinear algorithm. Our replication focuses on multiple linear regression methods only, so we will not be including the BP neural network model.

```
##   (Intercept) air_temp relative_humidity wind_speed     adjr2       BIC
## 1           1        1                 0          0 0.7006206 -14214.77
## 2           1        1                 0          1 0.7324934 -15534.80
## 3           1        1                 1          1 0.7332844 -15561.37
```

```
##           mse           mae          VEcv
##     86.049111      8.851913 -27006.427296
```

To substitute, we have included a multiple linear regression prediction accuracy for our model. The collinear factors were previously excluded, but it can be confirmed that the best model with the remaining predictors involves the intercept, air temperature, relative humidity, and wind speed. The model with all three predictors has both the highest adjusted R^2 and the lowest BIC as seen above, which means that it is the best subset model. The predictive accuracy is shown in the code by the MSE, MAE, and VEcv values. The MSE value is 86.049111, the MAE is 8.851913, and the VEcv is -27006.427296 for m1.

Section 4. Extension/Enhancement:

For our extension/enhancement, we will be checking the assumptions of the data. The study mentions removing predictors with collinear factors, but only provides the table to check variables with high collinearity. The study does not check for any other assumptions. For our assumption of normality and Tukey's test, we used studentized residuals.

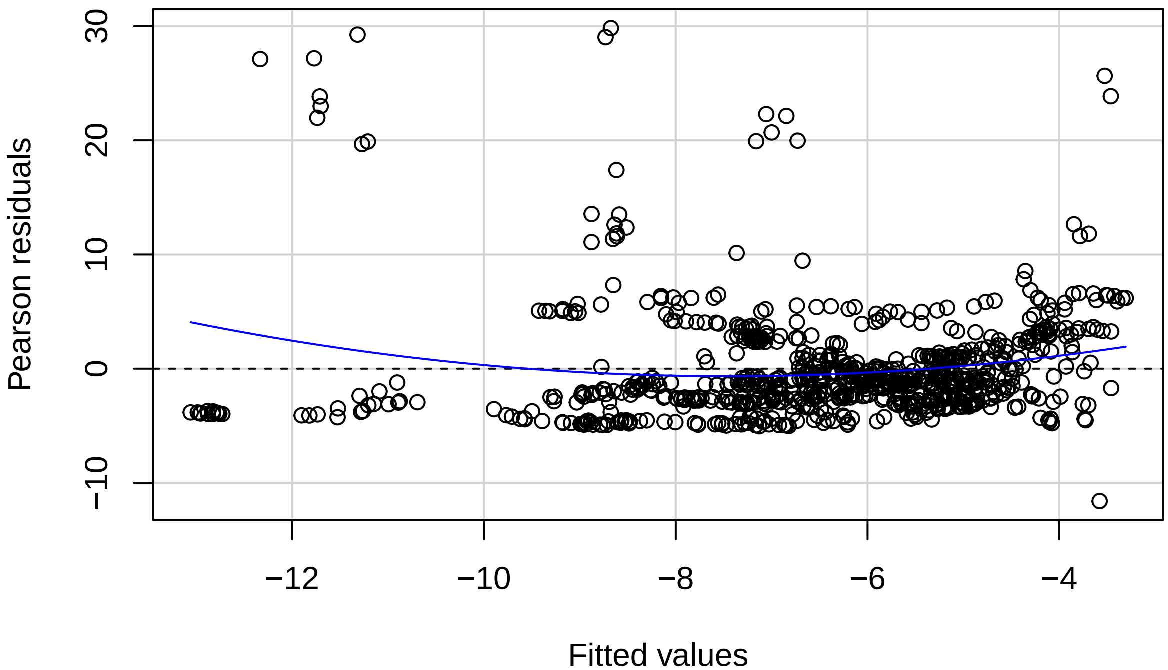
```
##          air_temp relative_humidity        wind_speed
##          1.040025          1.068348          1.071923
```

There does not appear to be an multicollinearity, so the linearity assumption is upheld. All of the VIF scores are between 1 and 1.1, which is very low. A score 5 and above would be of concern for the linearity assumption.

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 69.72656, Df = 1, p = < 2.22e-16
```

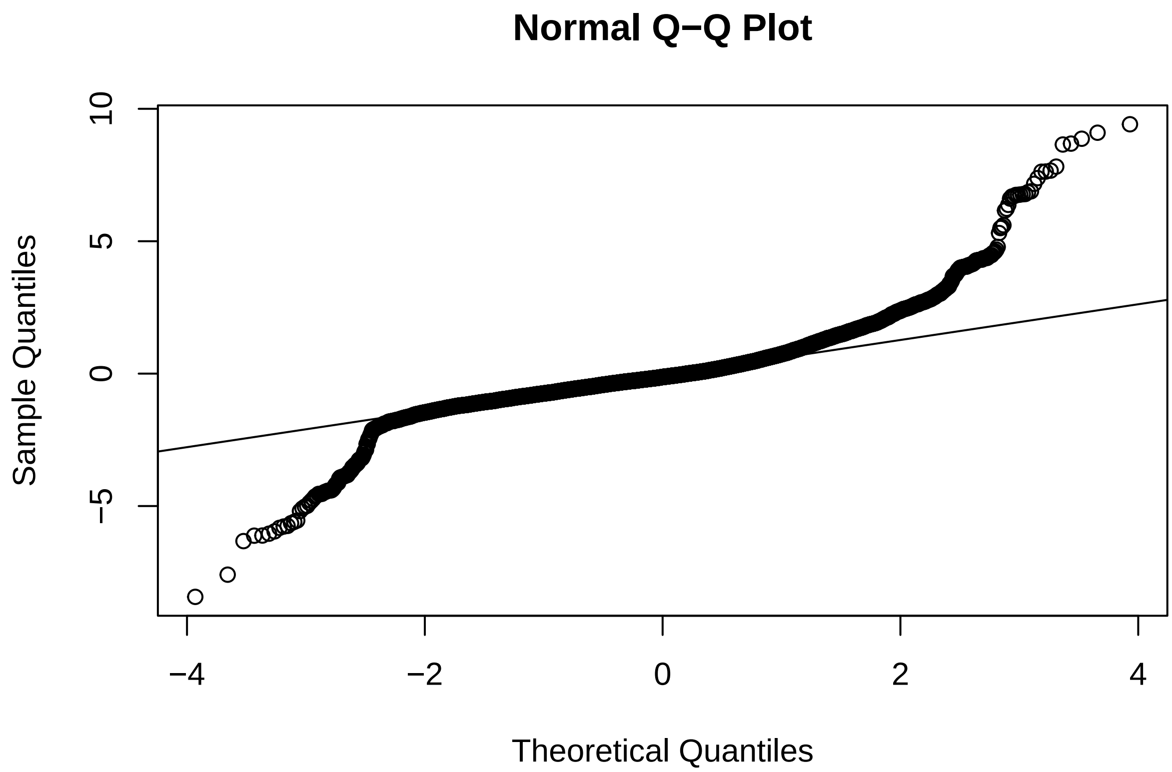
Since the p-value is 2.22e-16, which is less than the significance value of 0.05, this means that we reject that null hypothesis that there is constant variance and conclude the constant variance assumption is not reasonable with respect to the fitted values. This can also be shown in the residuals plot below. In the residuals plot, we use a model that takes a smaller sample of the data, the first 800 data points.

```
##  lag Autocorrelation D-W Statistic p-value
##    1      0.06496505      1.869917       0
##  Alternative hypothesis: rho != 0
```



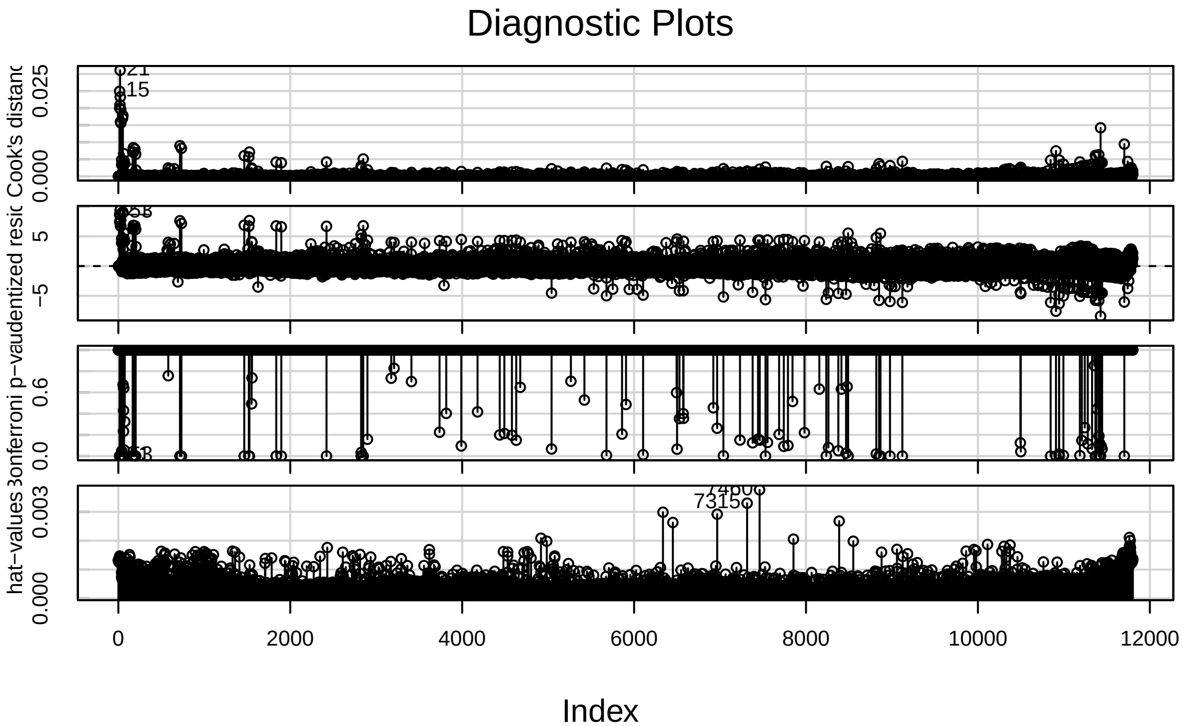

```
##                     Test stat Pr(>|Test stat|)
## air_temp               4.9183        8.848e-07 ***
## relative_humidity     17.2068        < 2.2e-16 ***
## wind_speed             1.9301          0.05362 .
## Tukey test             7.0731        1.515e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null hypothesis states that the errors are independent. Since we have a p-value of 0, which is greater than the significance value of 0.05, we fail to reject the null hypothesis that the errors are not auto-correlated with themselves. This means that the errors are not independent. The residuals plot also displays how the data is clustered towards the end and there is slight curvature with the fitted line. This shows how the variables are not independent of one another and there is not constant variance. We tested if the curvature was of concern with Tukey's test of nonadditivity. For Tukey's test, we used the model with the full data set. The Tukey test for nonadditivity has a low p-value, indicating that the curvature is a cause for concern and the mean function specified might not be as adequate.




```
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  ti
## D = 0.10764, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

For the normality test, we have a large sample size of 11,837, so we used a QQ plot and the Kolmogorov-Smirnov test. The QQ plot appears to be light-tailed, which is not normal. Only the ends deviate, so the distribution may still be somewhat normal. The Kolmogorov-Smirnov test has a low p-value of 2.2e-16, which is smaller than the significance level of 0.05. The normality assumption is not met since the p-value rejects the null hypothesis that the data is normally distributed.



There are only a few influential data points out of the large data set. The outliers appear to be at index 15, 21, 7315, 7460. These outliers affect the mean value produced by the data set.

Section 5. Discussion:

The values for the model and coefficient of determination R^2 are different in this replication than the original study. The coefficient of determination and F-statistic in this replication was higher than in the paper. Unaccounted variables or confounding factors in the original study could contribute to differences in outcomes as well. It was not clear whether or not the original study used the valid samples of the data or the full raw data set for their model, which is another reason why the results might be different. The study also did not show what those valid samples were even though they described how to classify those data points.

```
##
## Call:
## lm(formula = bridge_deck ~ air_temp + wind_speed, data = rawdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -30.101  -1.909  -0.539   1.325  34.305
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.597317   0.060799  -9.824   <2e-16 ***
```

```
## air_temp     0.963426   0.005748 167.617   <2e-16 ***
## wind_speed   0.642976   0.017143  37.508   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.62 on 11798 degrees of freedom
## Multiple R-squared:  0.7325, Adjusted R-squared:  0.7325
## F-statistic: 1.616e+04 on 2 and 11798 DF,  p-value: < 2.2e-16
```

```
##                  2.5 %     97.5 %
## (Intercept) -0.7164942 -0.4781405
## air_temp     0.9521597  0.9746929
## wind_speed   0.6093734  0.6765777
```

In our replication, we tried refitting the data (as shown above) without relative humidity. This was to see if the model values would match the ones in the study and if that is what they may have done for their model fit, but the values still did not match. This means that this is not what the original study did and the original study did not interpret their model correctly since they forgot to mention relative humidity as an explanation for the bridge pavement temperatures.

```
## [1] "Root Mean Square Error = 9.276266"
```

The root mean square error value (RMSE) being 9.276266 indicates that the model may not fit the data or make accurate predictions. The difference between the predicted value of the linear regression model and the real is greater. As seen in the BIC and adjusted R^2 calculation, the model with all three predictors has both the highest adjusted R^2 and the lowest BIC, which means that it is still the best subset model to use for bridge pavement temperature predictions. Our R^2 value of 0.733 is high which indicates high reliability and good fitting degree, indicating that our model is useful to predict the bridge pavement temperature.

All of the listed assumptions in Section 3 were tested as part of the extension/enhancement. All of the assumptions were violated except for the linearity assumption. The linearity assumption was not violated because in the study, the included a model that excluded collinear factors. Since multiple assumptions in this paper were not met, the inferences concluded in this paper are not correct. The violation of the constant variance assumption and normality assumption means that there can be inaccurate confidence intervals and p-values. The errors not being independent entails that there is a relationship between the residuals of our model and the response variables. If error terms in the regression model are positively (auto)correlated, the use of the OLS procedures has a number of important consequences: correlation will not affect the unbiasedness of OLS estimators, but it does affect their efficiency, it may seriously underestimate the true variance of the error terms, standard errors calculated according to the OLS procedure may seriously underestimate the true standard deviation of the estimated regression coefficient, confidence intervals and tests using the t and F distributions are no longer accurate.

Using data from small traffic meteorological monitoring stations, the researchers implemented a multiple linear regression method to establish a predictive model. The linear regression model did not perform too well, meaning that these predictions are not very accurate. The study recognizes the need for future research to include data from various stations to improve the models applicability, especially in regions lacking meteorological monitoring stations.

References

Han, S., Liu, Z., Xu, J., & Yan, M. (2022). Using multiple linear regression and BP neural network to predict critical meteorological conditions of expressway bridge pavement icing. PLOS ONE, 17(2). https://doi.org/10.1371/journal.pone.0263539